Automatic Classification of Clouds on METEOSAT Imagery: Application to High-Level Clouds

Michel Desbois, Geneviève Sèze, Gerard Szejwach

Automatic Classification of Clouds on METEOSAT Imagery: Application to High-Level Clouds

MICHEL DESBOIS, GENEVIEVE SEZE AND GERARD SZEJWACH[1]

Laboratoire de Météorologie Dynamique du CNRS, École Polytechnique-91128 Palaiseau Cedex, France

(Manuscript received 19 August 1981, in final form 29 November 1981)

ABSTRACT

A statistical classification method based on clustering on three-dimensional histograms is applied to the three channels of the METEOSAT imagery [Visible (VIS)–Infrared Window (IR)–Infrared Water Vapor (WV)]. The results of this classification are studied for different cloud cover cases over tropical regions. For high-level cloud classes, it is shown that the bidimensional histogram IR-WV allows one to deduce the cloud top temperature even for semi-transparent clouds.

1. Introduction

The problem of cloud classification from satellite imagery has been studied by scientists applying different methods according to the purpose of the research. In most operational uses, visual cloud classification (nephanalysis) is done by trained operators from VIS and IR imagery.

Automatic methods are primarily used for cartography of the cloud-top level, for example in ESOC (European Space Operation Center). For cloud classification itself, methods using typical classes defined by meteorologists have been tried, for example, in the French Center for Spatial Meteorology. The methods work relatively well in pure cloud cases, but fail with intermediate classes or mixed-cloud layers.

However, for research on mesoscale meteorology or for numerical modeling, there is a need for more precise and entirely objective classification. This classification essentially refers to the radiative properties of the clouds, which have a direct effect on the physics of the model, i.e., the approach of Shenk *et al.* (1976), Reynolds and Vonder Haar (1977) and of Chen *et al.* (1980). Two- or three-dimensional histograms are extracted from the imagery, the major problem being how to separate classes on these kinds of diagrams. Most methods are based on the determination of thresholds in the different channels (Shenk *et al.*, 1976). At the European Space Operation Center, which monitors METEOSAT, at least two methods have been tried (Bowen *et al.*, 1979): successive separations on monodimensional histograms and an interactive method where the operator himself chooses visually the classes on a representation of the histogram on a TV screen.

At the international scale, it has been proposed (Smith *et al.*, 1981) to archive all the meteorological imagery satellite data in the form of simplified bidimensional histograms (the aim is to keep the radiative information).

In all the classifications using VIS and IR window channels only, an indeterminancy rises for semi-transparent clouds. This indeterminancy can be overcome by using another IR channel, e.g., the water vapor channel (5.7–7.1 μm) in the case of METEOSAT (Cayla, 1978; Szejwach, 1982). Note that this channel is also taken into account by the European Space Agency in cloud-top temperature determination.

In this paper, we propose to use both the information of the water-vapor channel and a statistical method of three-dimensional histogram partition to obtain better classifications for mesoscale cases in tropical regions. We will first describe the histogram partition method (clustering method), then show and discuss some examples of analyses, and finally propose an application for the determination of semi-transparent cloud-top temperatures.

2. Clustering method

Clustering methods have the advantage of forming natural data groupings, without *a priori* classification. The classes which are found depend on the set of data analyzed. Clusters are not limited by straight lines or plane surfaces, as it is the case in threshold methods. This is particularly well adapted to histograms of cloud coverage, where it can be seen that the grouping of points can hardly be attributed to

[1] Present affiliation: Goddard Laboratory for Atmospheric Sciences, NASA, Goddard Space Flight Center, Greenbelt, MD.

0021-8952/82/030401-12$07.00

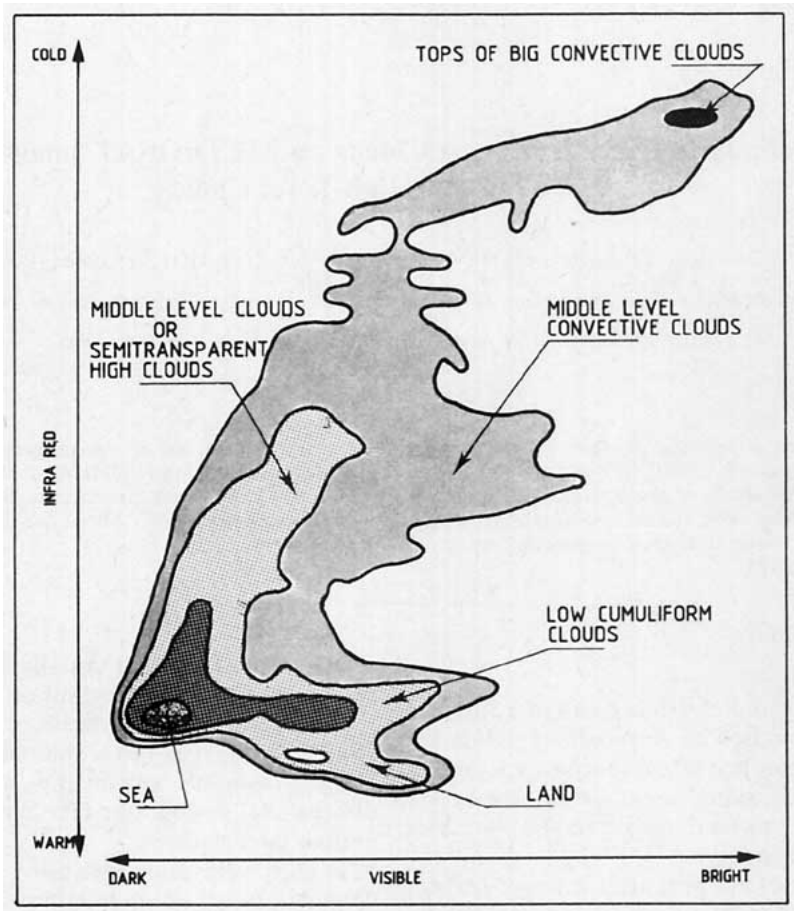


FIG. 1. Example of a VIS-IR bidimensional histogram where classes can be identified visually.

rectangular areas, but better to oblong shapes with major axes not parallel to the axes of the diagram.

The method used here, the dynamic cluster method, has been presented by Diday and Simon (1976). It has been adapted here for the particular case of METEOSAT, working on the three channel spectral representation of the image segments (VIS-IR-WV). In other words, classes must be separated on three-dimensional histograms. We do not use any training set, i.e., there is no *a priori* classification of any of the points of the histogram. It is well known (Bizzari, 1976; Desbois, 1978) that in simple cases two-dimensional histograms show (Fig. 1) that every well-defined type of cloud can be associated with a specified spectral signature on the diagram. Although general cases of histograms are not so simple, it can be assumed that every kind of cloud is represented in the spectral domain by a compact subset of the projection of the image segment in this domain; i.e., all the points of the same subset are close, according to the Euclidian distance, and all the subsets are disjoint.

In the case of METEOSAT, the spectral domain is three-dimensional (VIS-IR-WV). The method can be summarized as follows:

1) A given number N of classes is chosen.
2) A set of F points is then chosen randomly for each class. This set is called kernel of the class.
3) For each kernel, the center of gravity and the variance are calculated.
4) Classes are constructed: each point is affected to the class of which the center of gravity of the kernel is the closest. During this process, if the number of elements of one class becomes very small, the class is suppressed.
5) The center of gravity and the variance of the classes are calculated again.
6) New kernels of F points are again defined.
7) A new iteration is begun from 3).

Iterations are carried out until the distance between the centers of gravity of a class and the corresponding kernel remains constant.

In this application to METEOSAT, the initial number of classes was $N = 15$, and the number of elements in one kernel $F = 30$.

We worked essentially on image segments of 200 × 200 pixels (40 000), taking a sampling (learning set) of only 8000 points to reduce the computation time. At the end of the process, each pixel of the original image segment was assigned to the class to which it was the closest.

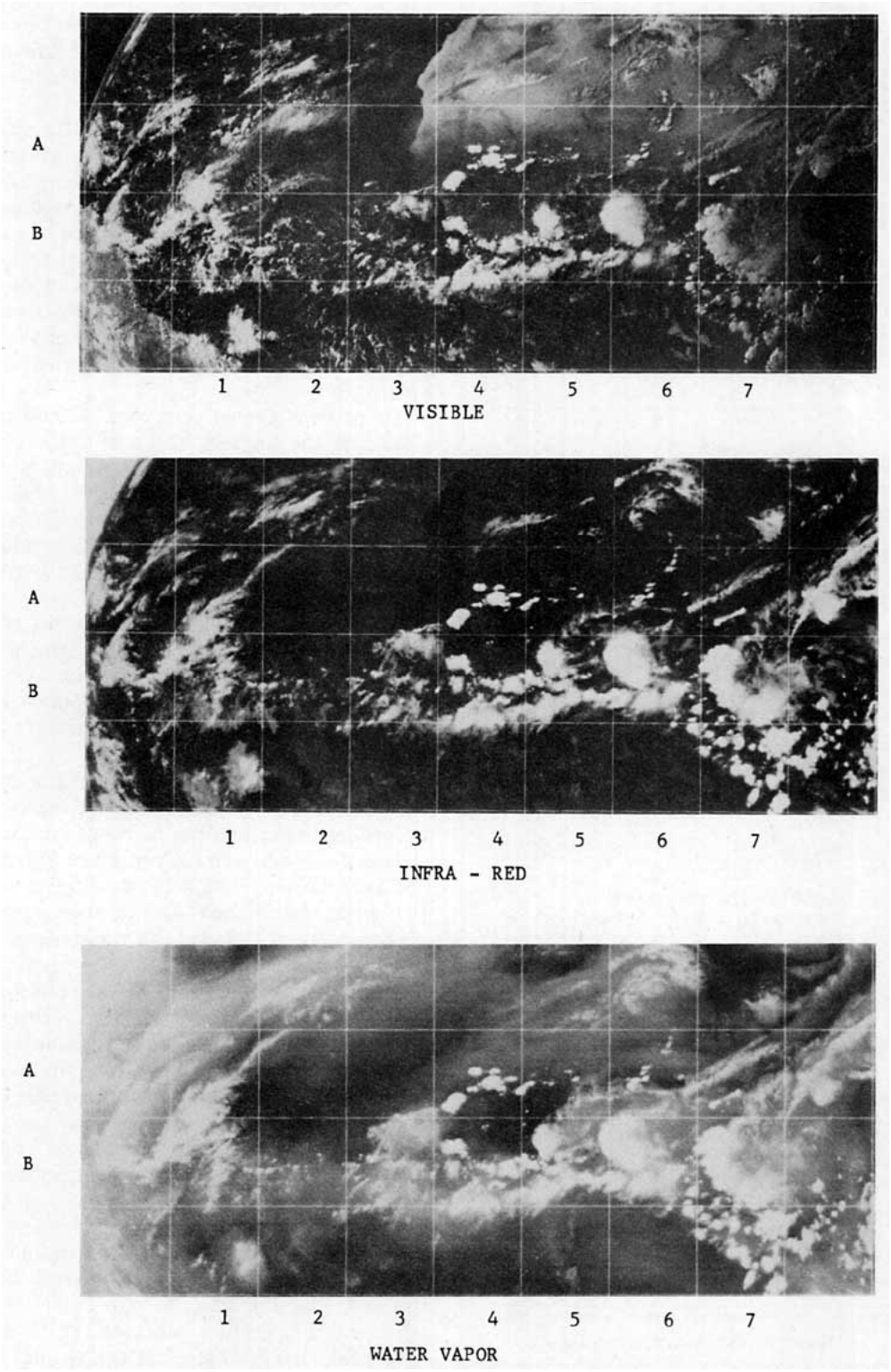


FIG. 2. METEOSAT image of 1530 GMT 23 May 1979 in the tropical regions. The three channels (VIS-IR-WV) are shown with a 200 × 200 pixel mesh.

3. Examples of applications

The method was applied to METEOSAT images in the tropical regions over Africa and the Atlantic Ocean. No correction was applied for the visible, but only for restricted areas where the solar angle can be taken as constant. The region studied is shown in Fig. 2 (3 channels) for the image at 1530 GMT 23 May 1979. The squares are 200 × 200 pixels (IR pixels of METEOSAT measure 5 km at the subsatellite point). Fifteen squares have been analyzed on this particular image.

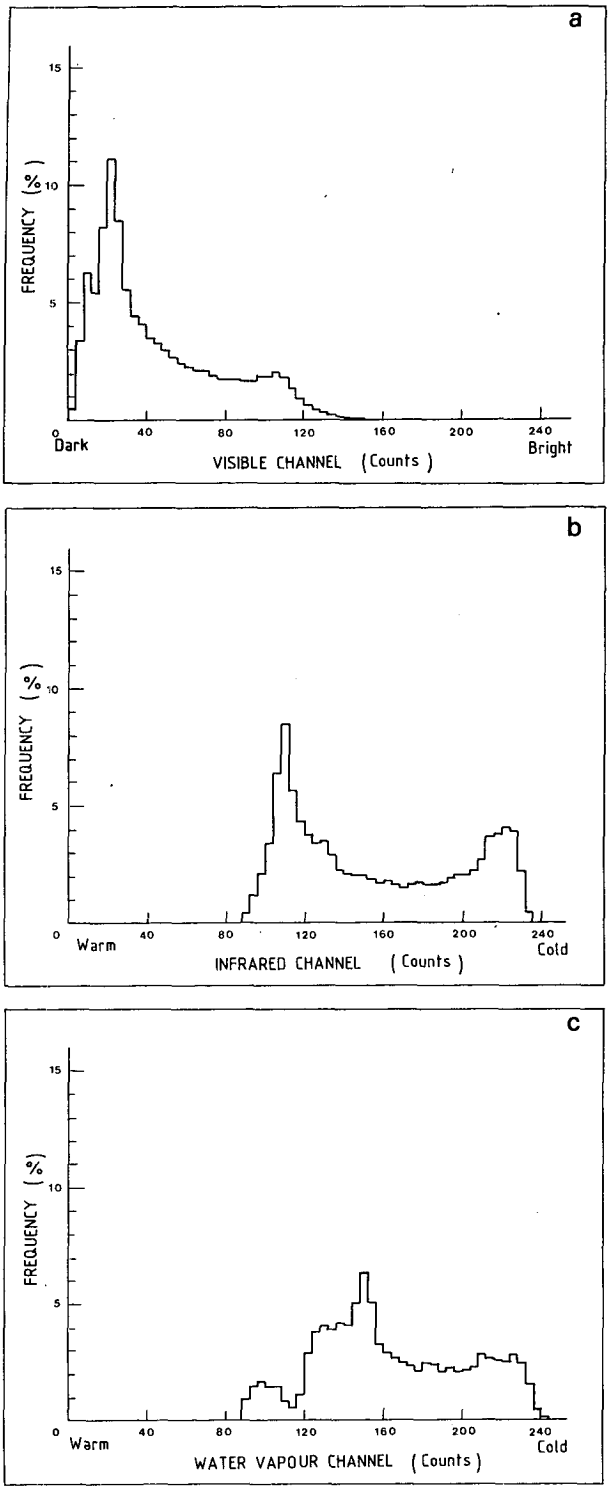


FIG. 3. Mono-dimensional histograms (VIS-IR-WV) for square B5.

Due to the initial character of this study, we did not study the transferability of the technique to higher latitudes, with various sun angles; this is the object of a further study now in progress. However, it has already been noticed that the sun angle problem is at least partially solved by the method itself; thus, for contiguous areas with the same kind of cloud cover but different sun angles (near the terminator), we found the same classes with a self-adjustment of the visible (general shift of the visible values).

To prevent the effects of a bad initial random choice of the kernels, three of these choices were carried out for each of the segments studied before the application of the algorithm. This could be avoided easily, taking an initial choice with *a priori* values corresponding to a general approximated classification; however, we tried here to test the method for entirely automatic classification.

The algorithm is then applied to the three choices on 8000 pixels chosen in the segment of 40 000 points. This number of 8000 points was determined after testing which showed that 6000 points was a limit to the satisfactory reproductibility of the classification.

Then, if there is a significant difference between the three classifications obtained, the best is taken, according to the number of classes and the value of the distances between the centers of gravity of classes and kernels (see Section 2) (the smaller the distance, the better the choice). At this stage, generally, 5–7 classes remain. Finally, the classification is applied on all 40 000 pixels.

As an example we take the case of segment B5 on Fig. 2. This is a complex case of a disturbed area in the intertropical convergence zone, in the region of the Ivory Coast. Figs. 3–6 give the statistical description of this image segment and the results of the analysis (mono- and bi-dimensional histograms, representation of the classes obtained on bi-dimensional histograms, and the resulting classification of the pixels on the image).

Mono-dimensional histograms (Fig. 3) differ significantly from one another, which shows that the information in the three channels will be useful for the final classification. In particular, WV shows a tri-modal distribution, whereas the IR is bi-modal. Two peaks are also seen on the visible, but the one corresponding to the brightest part of the picture is much smaller than the other.

Bi-dimensional histograms (Fig. 4) show stretched and continuous shapes, on which only two or three classes can be separated visually. It appears also that WV and IR are the most correlated channels, and that VIS and WV are the least correlated.

In this case, five classes have been separated by the algorithm (Table 1): three have very small variances and correspond to very definite kinds of clouds: class 1, big opaque convective clouds; class 4, low clouds; class 5, ground or sea. The two other classes show slightly greater variances and correspond (class 2) to the edges of big convective masses and (class 3) to semi-transparent high clouds (cirrus) escaping from the anvils. If we look at the bi-dimensional histograms VIS-IR for the individual classes (Fig. 5), it can be seen that classes 1, 4 and 5 correspond effectively to the clusters which could be identified visually, whereas classes 2 and 3 correspond to intermediate regions difficult to separate visually. On these diagrams, classes seem to recover themselves slightly: in fact, they are separated on the 3D histogram, but not on its projections on 2D. On the 2D histograms of class 5, it can also be seen that two classes could still be separated from the VIS scale. This separation has not been performed by the method because the kernels of the two classes were too close to one another.

The examination of the restituted picture (Fig. 6) of the classes confirms this analysis; for example, low clouds are found ahead of the squall line northwest of the square, and on the sea at the southeastern part. For the high clouds, it can be seen that they end abruptly to the west of the squall line, whereas they become more and more diffuse (semi-transparent) at the eastern end.

For segment B2, six classes are obtained (Table 2 and Fig. 7), three of them being very homogeneous: classes 4, 5 and 6. Classes 5 and 6, with 44 and 32% coverage respectively, are two classes of sea, separated from the variations observed on the water vapor channel. Class 4 represents small low clouds (cumulus), or edges of bigger clouds (the pixels of this class are probably only partially covered with low clouds). There is only a very little amount (2%) of high cloud tops (class 1), which explains the large variance. Class 2 represents semi-transparent high clouds (cirrus) (low visible value) whereas class 3 shows lower clouds with or without semi-transparent cloud overcast (two distinct maxima can be seen inside this class on IR and WV, whereas VIS remains at relatively high values). The separation has not been performed for the same reason as for class 5 of segment B5.

For segment A6 (Table 3 and Fig. 8), six classes are obtained: class 1 again corresponds to the tops of high convective clouds or thick cirrus; class 2 to thin cirrus; class 3 to middle-level clouds (note that

TABLE 1. Classification obtained for zone B5 of Fig. 2. The table gives the percentage of coverage of the analysis window by each class, the center of gravity and the variance of each class. Units are numerical counts of the METEOSAT radiometers.

Class	Percent coverage	VIS average (S.D.)	IR average (S.D.)	WV average (S.D.)
1	18	100 15	219 10	218 14
2	18	56 13	195 13	191 15
3	17	28 9	151 14	162 13
4	7	54 16	134 13	129 16
5	38	18 8	111 9	132 19

TABLE 2. As in Table 1 but for zone B2.

Class	Percent coverage	VIS average (S.D.)	IR average (S.D.)	WV average (S.D.)
1	2	124 24	192 17	179 20
2	7	37 11	150 12	148 13
3	5	82 14	145 18	142 16
4	4	50 9	124 7	129 12
5	44	21 5	119 6	135 7
6	32	26 4	112 3	107 9

TABLE 3. As in Table 1 but for zone A6.

Class	Percent coverage	VIS average (S.D.)	IR average (S.D.)	WV average (S.D.)
1	5	68 23	203 15	196 20
2	7	34 8	162 14	174 11
3	5	47 15	147 15	145 12
4	12	26 8	119 10	153 13
5	22	25 8	89 8	139 10
6	45	34 5	65 9	126 9

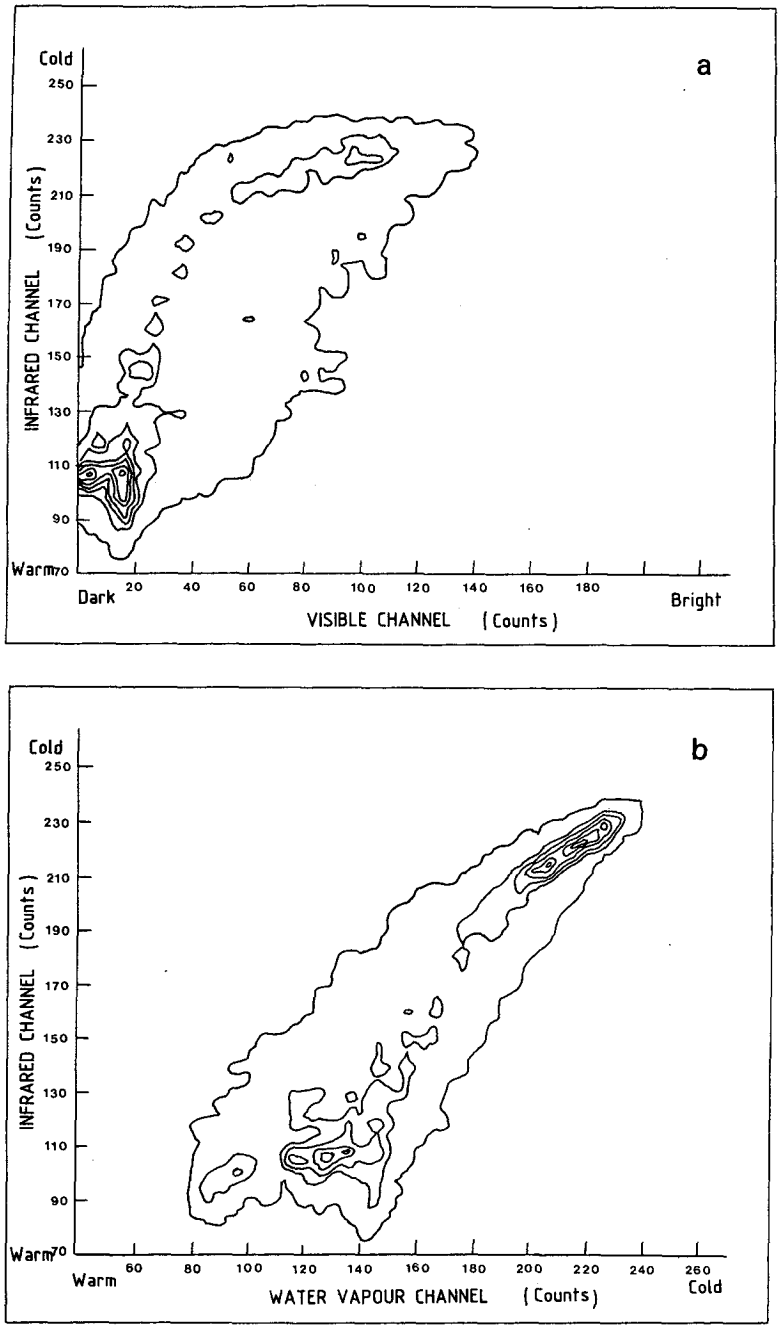


FIG. 4: Bi-dimensional histograms for square B5: IR-VIS; IR-WV; VIS-WV.

the WV value is very close to the value without clouds); class 4 to very thin cirrus (colder WV, but warmer IR than class 3); classes 5 and 6 are two types of ground, separated by their IR and VIS values (due to vegetal cover differences). Note that in this case, VIS values are smaller than for the other segments, due to the vicinity of the terminator (sunset).

These three examples, as are the other cases studied in this image or on other images, show that the classes separated correspond effectively to cloud types on the pictures. The three channels are effectively useful independently for the separation of the classes. From one segment to another, different classes are found, according to the cloud types which are present, but also to the extent of the cloud cover:

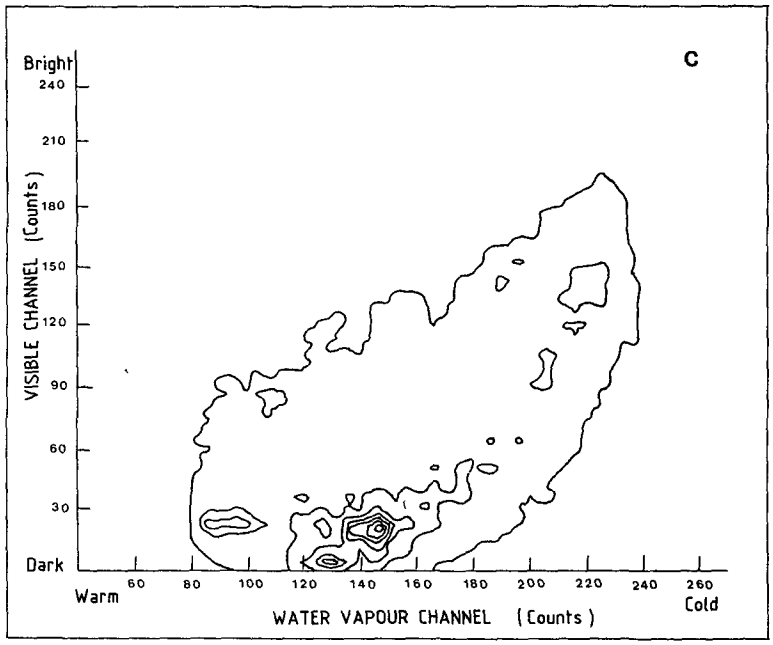


FIG. 4. (*Continued*)

the discrimination between different low-cloud classes is better when there are no high clouds, for example. Morever, it is impossible to identify a class if there is less than ~2% of cloud coverage with this class. However, if contiguous segments are taken (Fig. 9), it can be seen that there is no big discontinuity at the edges of the squares: sometimes, two classes merge from one square to the other; sometimes, a new class appears. This shows that the method separates the same kinds of clouds from one square to another, depending only of the variety of clouds which are present. Moreover, we have made the classification on larger squares (400 × 400), taking one pixel out of two; the same principal classes are found—classes with very small percentages of coverage disappear and classes with small separations merge. At least in tropical regions, as the variety of clouds in 400 × 400 pixels segments are greater, it seems that the final number of 6–7 classes obtained by our method is a better representation of the distribution.

4. Application to the determination of semi-transparent cloud-top temperature

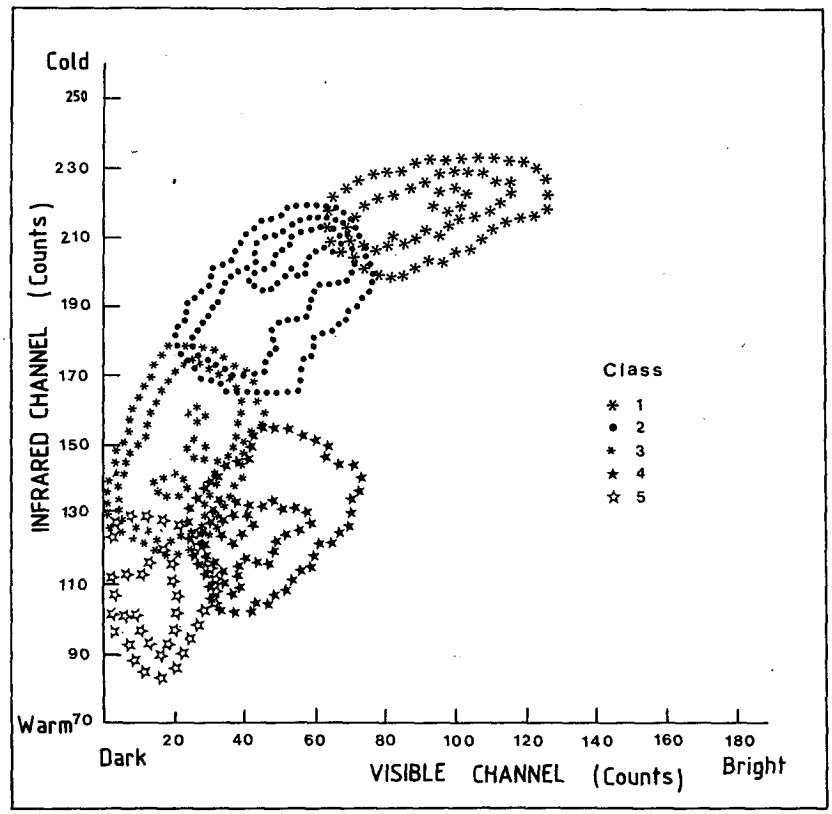
We have seen in the various examples of Section 3 that there is always a significant percentage of cloud coverage constituted by semi-transparent clouds. It is very useful to determine accurately the top temperature of this kind of cloud, since such clouds play an important role in the radiation budget and are often used as wind tracers for high tropospheric levels. The top temperature of cirrus clouds must be determined from the radiance in the IR window channel, assuming an emissivity. This is often done, rather crudely, from the visible image (Liou, 1973; Platt *et al.*, 1980; Shenk and Curran, 1973). More recently, Szejwach (1980) proposed a method using both IR and WV channels of METEOSAT. A similar kind of method has been proposed by Cayla (1978) and is used in the European Space Operation Center with fairly good success.

We have tried here to use such a method on *classes* identified as semi-transparent clouds and not on a chosen zone containing only semi-transparent clouds, as was done in previous studies.

The example shown here (Fig. 10) is a segment of 120 × 120 pixels containing high clouds, low clouds and sea. The results of the dynamic clusters method gave six classes:

1. Thick cirrus	3% of the area
2. Thinner Cirrus	6% of the area
3. Very thin cirrus	3 % of the area
4. Sea with very light haze	22% of the area
5. Low clouds	17% of the area
6. Sea	45% of the area

The cirrus has been decomposed into three classes—those with very thick contour, and with two classes of edges, more and more semi-transparent. The cor-



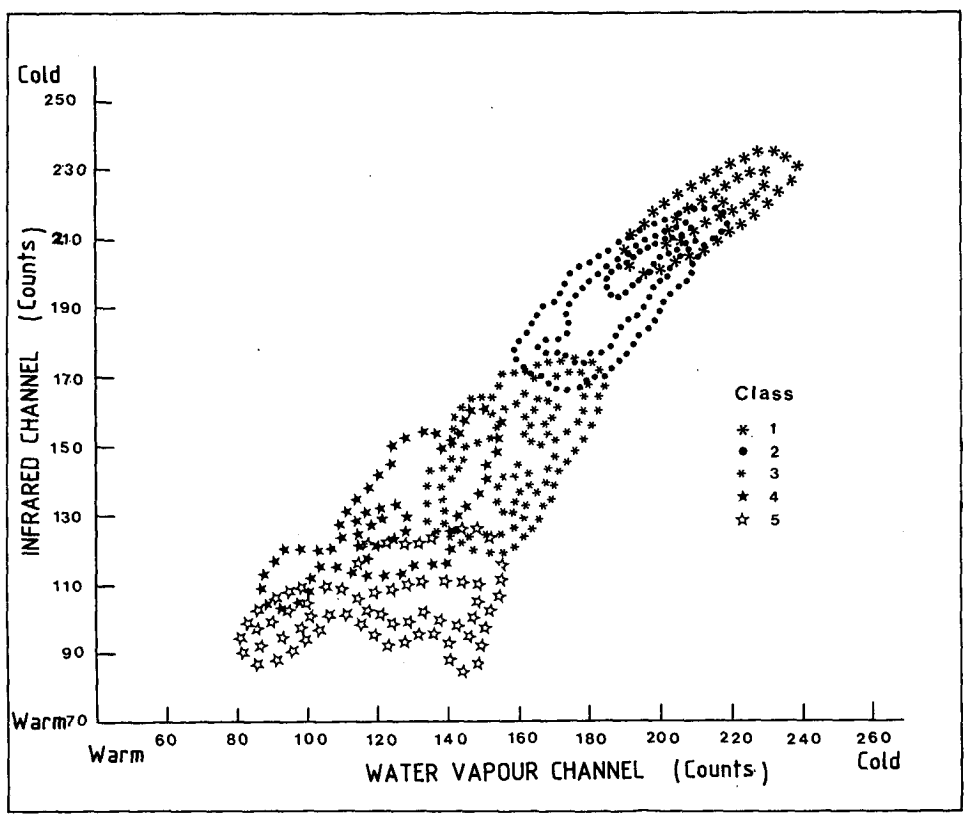


FIG. 5. Classes obtained by the dynamic cluster method for square B5, represented on bi-dimensional histograms IR-VIS and IR-WV.

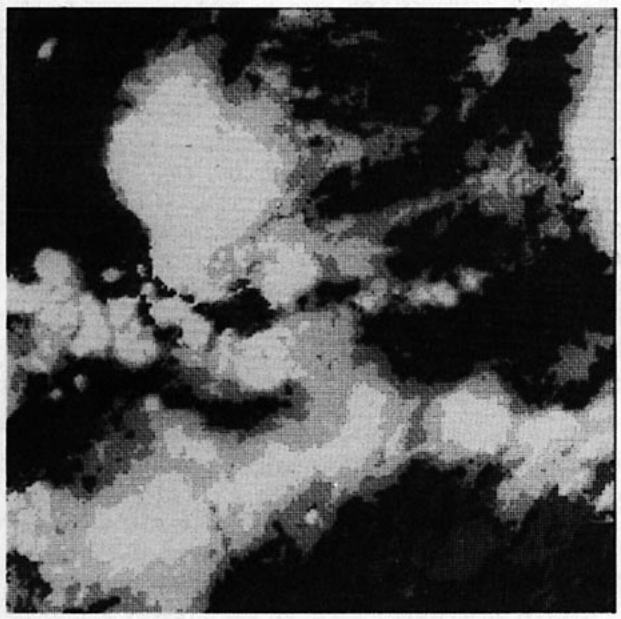


FIG. 6. Spatial distribution of the classes for square B5: white = class 1, dark = class 5.

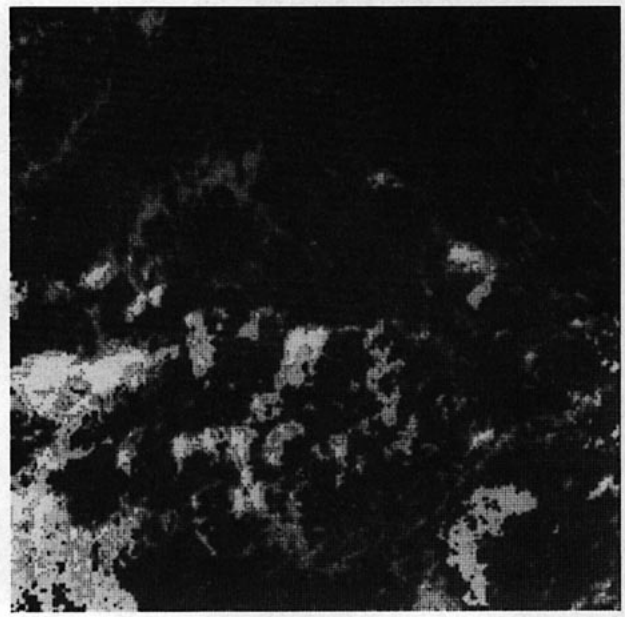


FIG. 7. Spatial distribution of the classes for square B2: white = class 1, dark = class 6.

responding bi-dimensional histograms are given in Fig. 11.

On these histograms, we have superposed the curve $F(T) = (x, y)$ such as $x = F_{IR}(T)$, $y = F_{WV}(T)$, where F_{IR} and F_{WV} are functions which associate to the temperature T of a blackbody a radiometric value x in the IR channel and y in the WV channel (inverse of the calibration function).

It has been shown by Szejwach (1982) that the intersection between this curve and the principal axis of the histogram gives a good approximation of the temperature of the upper surface of the cloud. The regression line corresponding to each class of cirrus was calculated and drawn on Fig. 11 for each corresponding histogram. It can be seen that the intersection of the straight line with the curve $F(T) = (x, y)$ is nearly the same in each of the three cases. The corresponding temperature is 226 K for class 1, and 228 K for classes 2 and 3. These values may appear a little warm for tropical cirrus, but this may be due to the fact that this cirrus was taken northwest of segment A2 on the picture (i.e., in the middle of the Atlantic, above 20°N). This may be due also to a relatively bad calibration of the METEOSAT radiometer. The most important result is that we found very similar cloud-top temperatures from classes giving very different radiances in the infrared, water vapor and visible, the three channels being used to perform the class separation. In this case, this is physically significant: the three classes are part of the same cloud, in regions with different optical thickness.

The same method was also applied to the other squares studied in this paper. In regions with convective clouds, the first class (top of clouds) surrounds the curve $F(T) = (x, y)$, which was expected for clouds radiating as blackbodies. The dispersion around this curve can be attributed to—noise on the image, specially on the water-vapor channel (the signal is weak for low temperatures and the signal-to-noise ratio becomes significant) and a non-perfect

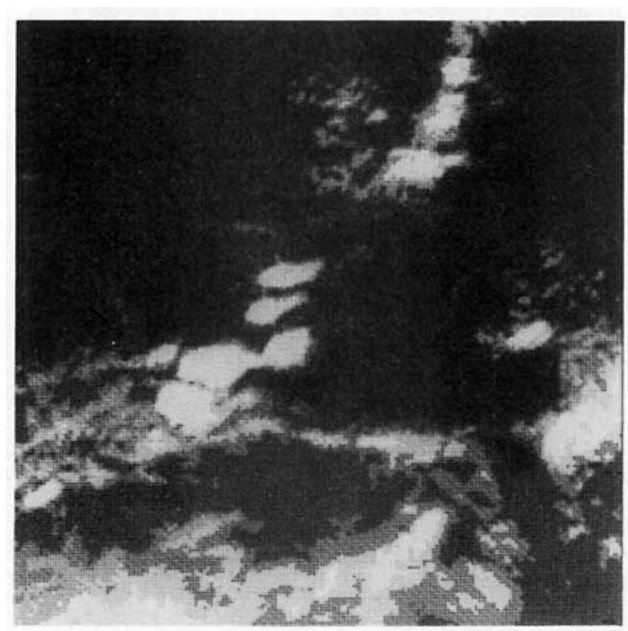


FIG. 8. Spatial distribution of the classes for square A6: white = class 1, dark = class 6.

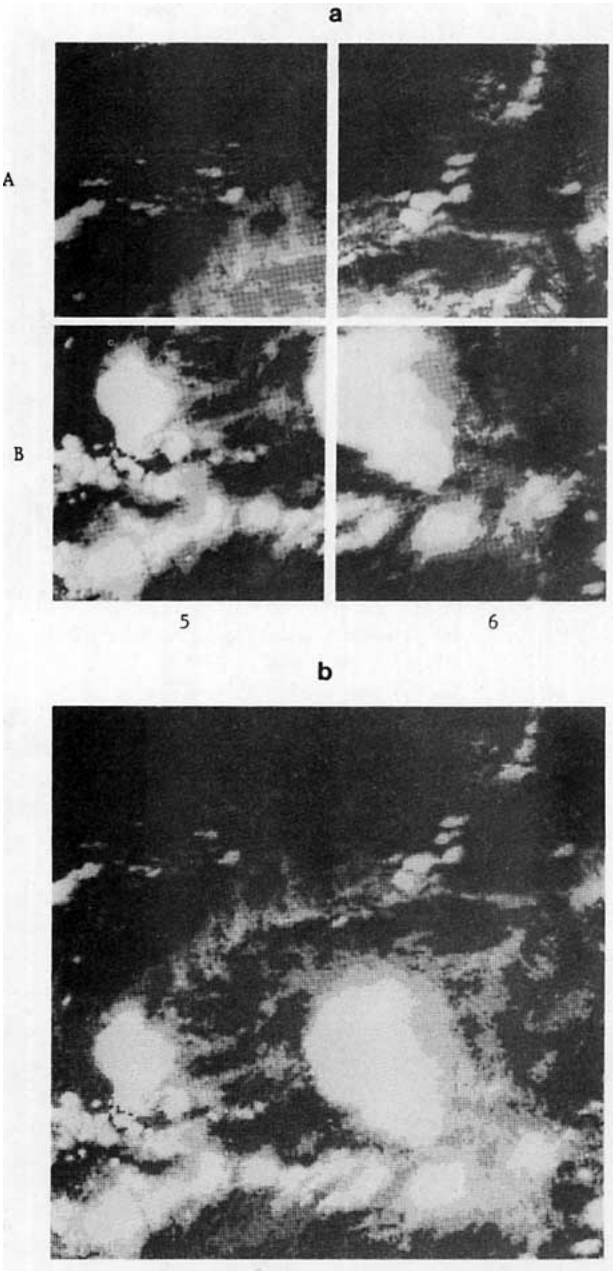


FIG. 9. Results of the classification for (a) four adjacent squares of 200 × 200 pixels and (b) a large square of 400 × 400 pixels covering the same area.

alignment of the two images IR and WV (the pixels are not covering themselves exactly, from a fraction of a pixel). Moreover, the calibration of the METEOSAT radiometer, specially for the water vapor channel, is not very accurate, and the curve $F(T) = (x, y)$ is also approximate (Roulleau *et al.*, 1980; Bériot *et al.*, 1982). Nevertheless, the present method of cloud-top temperature determination remains valid, as it allows correction of the apparent temperatures of quantities much greater than the above determinations. In the same regions of convective clouds, the second and sometimes third class are semi-transparent clouds which permit application of the method—it is generally found that the cirrus temperature is a little above that of class 1, but there are some cases of obviously bad determination due to partial cloud cover underlaying the high-level clouds; this method requires an homogeneous IR background to work properly.

5. Conclusions

The dynamic clusters method proves to be useful to extract objective information for cloud classification from the three channels of METEOSAT. However, there remains a need of *in-situ* data for validation of the classifications obtained. On another hand, the method still consumes too much computer time for operational uses: at the present stage, it takes 55 s on a Cyber 750 computer to perform a classification on a square of 200 × 200 pixels, with a sampling of 8000 pixels. If the initial choice of kernels is done from a pre-established classification (for example, from the classification on a preceding image), the number of iterations decreases, and it is also possible to reduce the sampling. Thus for 6000 points, the time becomes 13 s. Much time could still be saved by other ways such as taking a less severe criterion for stopping the iterations, starting from

FIG. 10. Spatial distribution of the classes for the case study of a cirrus (120 × 120 pixels).

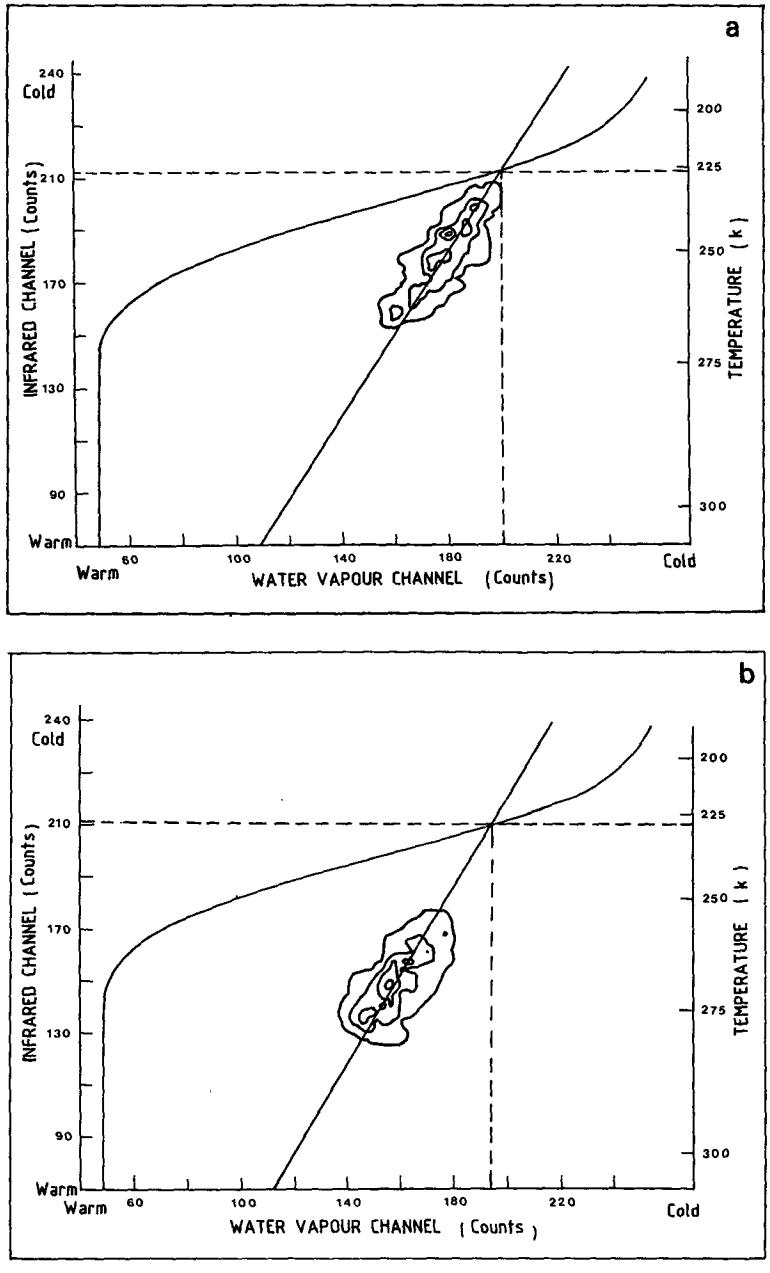


FIG. 11. Determination of the cirrus top temperature for the three classes of cirrus (see text for explanation).

much larger areas (for example, 400 × 400 pixels) with the same initial sampling of 8000 points, or using an array processing computer. For operational uses, certain natural cloud classes could be found using clustering for rather large regions and by seasons. Thereafter, cloud types could be assigned automatically without clustering.

Work must be done for the adaptation of such methods of clustering to operational uses, but they can already be used in research studies: for example, as the classification is based on the radiative properties of the clouds, it seems appropriate to use it for atmospheric circulation models. Therefore, the method is now being extended to larger areas, to introduce the resulting analysis in the experimental general circulation model of the Laboratoire de Météorologie Dynamique.

It is also intended to use this method for mesoscale

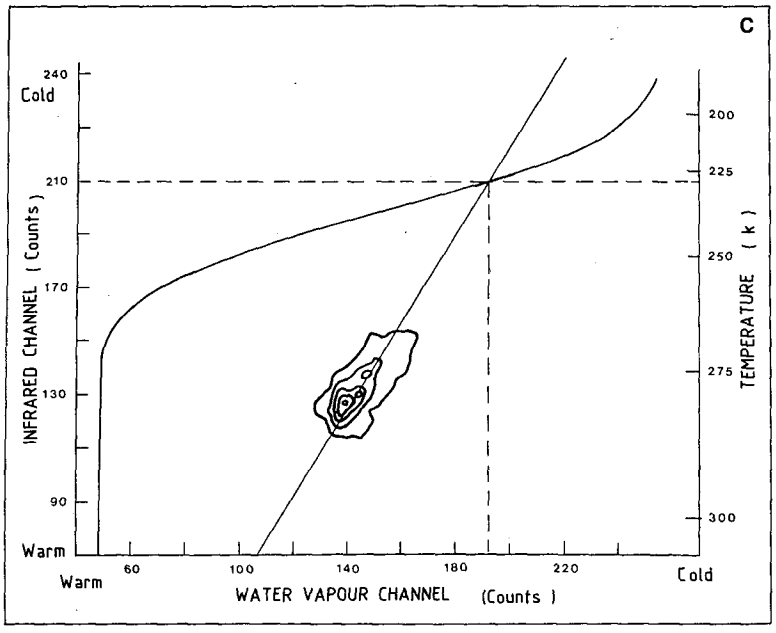


FIG. 11. (*Continued*)

studies, especially when a network of *in-situ* measurements is available.

As far as cirrus-top temperature determination is concerned, the good separation of semi-transparent clouds by the algorithm makes it possible to apply successfully the IR-WV histogram method.

REFERENCES

Bériot, N., N. A. Scott, A. Chedin and P. Sitbon, 1982: Calibration of geostationary satellites infrared radiometers using the TIROS-N vertical sounder. *J. Appl. Meteor.*, **21,** 84–89.

Bizzari, B., and C. Tomassini, 1976: Retrieval of information from high-resolution images. *Proc. Symp. Meteorological Observations from Space: Their Contribution to FGGE, COSPAR XIX,* Philadelphia, 140–144.

Bowen, R. A., L. Fusco, J. Morgan and K. O. Roska, 1979: Operational production of cloud motion vectors (satellite winds) from METEOSAT image data: Use of data from meteorological satellites. ESA, No. 143.

Cayla, F. R., 1978: Détermination de la température des cirrus semi-transparents. *La Météorologie,* **15,** Décembre, 63–67.

Chen, T. S., L. L. Stowe, V. R. Taylor and P. F. Clapp, 1980: Classification of clouds using THIR data from Nimbus-7 satellite. *International Radiation Symposium, Extended Abstracts,* Fort Collins, 315–317.

Desbois, M., 1978: METEOSAT image processing: horizontal and vertical velocities from water vapour images. *Proc. XXIXth Congress of the International Astronomical Federation,* Dubrovnik.

Diday, E., and J. C. Simon, 1976: Clustering analysis. *Digital Pattern Recognition,* K. S. Fu, Ed. Springer-Verlag 47–94.

Liou, K. N., 1973: Transfer of solar irradiance through cirrus cloud layers. *J. Geophys. Res.,* **78,** 1409–1418.

Poc, M. M., M. Roulleau, N. A. Scott and A. Chedin, 1980: Quantitative studies of METEOSAT water-vapor channel data. *J. Appl. Meteor.,* **19,** 868–876.

Platt, C. M. R., D. R. Reynolds and N. L. Abshire, 1980: Satellite and lidar observation of the albedo, emittance and optical depth of cirrus compared to model calculations. *Mon. Wea. Rev.,* **108,** 195–204.

Reynolds, D. W., and T. H. Vonder Haar, 1977: A bi-spectral method for cloud parameter determination. *Mon. Wea. Rev.,* **105,** 446–457.

Roulleau, M., M. M. Poc, and A. Chedin, 1980: Response of METEOSTAT water-vapor channel. *J. Phys. Lett.,* **41,** L99–L101.

Shenk, W. E., and R. J. Curran, 1973: A multi-spectral method for estimating cloud top heights. *J. Appl. Meteor.,* **12,** 1213–1216.

——, R. T. Holub and R. A. Neff, 1976: A multispectral cloud type identification method developed for tropical ocean areas with Nimbus-5 MRIR measurements. *Mon. Wea. Rev.,* **104,** 284–291.

Smith, E. A., T. H. Vonder Haar and J. Graffy, 1981: The impact of GOES satellite data compaction on the estimates of cloud parameters. *Report of a Workshop on Clouds in Climate,* NASA Goddard Institute for Space Studies, 192–196.

Szejwach, G., 1980: Contribution ā l'étude du rayonnement infrarouge des nuages. Application à la reconstitution tridimensionnelle de la couverture nuageuse. Thèse de Doctorat d'Etat, Université P. et M. Curie (Paris), 210 pp.

——, 1982: Determination of semi-transparent cirrus cloud temperature from infrared radiances. Application to METEOSAT. *J. Appl. Meteor.,* **21,** 384–393.

Tarpley, J. D., R. N. Green, D. B. Miller and P. M. Waters III, 1976: Cloud parameters from geostationary meteorological satellites. *Proc. Symp. Meteorological Observations from Space: Their Contribution to FGGE, COSPAR XIX,* Philadelphia, 289–293.