

## A Two-Dimensional Hydrometeor Machine Classifier Derived from Observed Data

HERBERT E. HUNTER<sup>1</sup>

*ADAPT Service Corporation, Reading MA 01867*

ROSEMARY M. DYER AND MORTON GLASS

*Air Force Geophysics Laboratory, Hanscom AFB, MA 01731*

(Manuscript received 7 October 1982, in final form 26 October 1983)

### ABSTRACT

Classification algorithms have been developed to distinguish six categories of cloud ice particles. These algorithms have been incorporated in schema which, when applied to shadowgraph images produced by the Precision Measurement System laser scanning device, have demonstrated the capability of classifying with more consistency than human classifiers, and with almost no sensitivity to particle orientation.

The data used to derive the algorithms consisted of observations obtained on four separate aircraft flights. Two human classifiers, interacting with a preliminary machine classification, defined the correct answers for this training data set. The algorithms were then tested against arbitrarily selected segments from two additional flights. The ADAPT Service Corporations eigenvector, or empirical orthogonal function (EOF) technique, defined the features objectively, and the ADAPT independent eigenscreening algorithm development program related these features to the particle type.

Analysis of the performance suggests that considerable variation is to be expected, based on the set-to-set variation of the distribution of particle types between real data sets. The classification schema have been developed to allow the user to change key parameters in order to compensate for this variation.

It was concluded that the machine classification was superior to manual classification for the identification of large numbers of particles in terms of speed and consistency.

### 1. Introduction

The problem of identifying the mixture of ice particle types found in clouds has recently become important in cloud physics studies. The effects of clouds on microwave sensing in weapon systems, satellite imagery, and the propagation of millimeter waves are often functions of the shapes of cloud particles as well as their concentration and size distribution. Dyer and Barnes (1979) have presented a survey of the characteristics of ice crystals found in different regions of cloud systems, and discuss the measurement techniques used to count, size and classify the particles. Rahman *et al.* (1981a,b) discuss the identification of cloud particle shapes from two-dimensional shadowgraph data.

The problem of examining large numbers of two-dimensional images, identifying the particle type, and calculating the statistics of particle types occurring during the data collection period is particularly well suited for automated machine classification. This is due both to the large quantity of data collected, and to the fact that manual classification of particles rapidly

becomes fatiguing. A large staff of technicians would be required to classify even a modest percent of the data obtained on a single aircraft flight within a reasonable time. There is also significant disagreement among classifications made by different people using the same data, and even among classifications made by the same person at different times.

Very few automated machine techniques for the classification of cloud particle types have been reported in the open literature. Papers by Rahman *et al.* (1981a,b) and by Duroure (1982) discuss the development of classifiers for the same type of data treated in the present paper. There are several significant differences between the approaches taken by Rahman *et al.* and by Duroure and that discussed here. Rahman *et al.* used artificial idealized patterns and a maximum likelihood Bayes family of classifiers. Duroure performed a radial analysis of particle image outlines from a selected data set. A Fourier transform was then used to extract shape discrimination from the image parameters of the analyzed data set. In the present study, classification algorithms were developed based on an objective analysis of a large set of real data. Fourier coefficients represent 31 of the 99 projections used to generate our algorithms. Those Fourier coefficients were included which the classification development

<sup>1</sup> Present affiliation: Nichols Research Corp., 4040 South Memorial Parkway, Huntsville, AL 35802.

program found to be significant; the others were rejected. The present classifiers are based on the variation of the Fisher classifier, rather than the maximum likelihood Bayes family of classifiers used by Rahman *et al.* (a discussion of these classifiers may be found in several standard statistical textbooks). Finally, the present algorithms are incorporated into schema which permit the adjustment of thresholds according to the needs of the user or the characteristics of the particular data set.

## 2. The data

The data used in this study were obtained from the two-dimensional laser scanning device produced by Precision Measurements Systems, Inc. (PMS). These devices and their calibration are described extensively in the literature by Knollenberg (1970, 1975, 1976), by Heymsfield and Knollenberg (1972), by Heymsfield (1976) and by Cunningham (1978). The system consists of a laser beam illuminating a line of photo diodes. As a particle falls through the viewing volume, it shadows some of the diodes, the number and location of which are determined by the particle's size and shape. A rapid scanning system records the diodes shadowed per unit forward motion of the aircraft. This forward motion equals the minimum grid size, and hence the smallest measureable particle size ( $25 \mu\text{m}$  for the cloud probe). Fig. 1 is an unusually clear example of an unusually pure set of dendrites recorded by the two-dimensional PMS probe.

## 3. Method of deriving the algorithms

The technique used to derive the algorithms consists of four steps: preprocessing, construction of the data vector, definition of classes, and design of machine classifiers.

### a. Preprocessing

The purpose of the initial preprocessing is to reject artifacts and trivial cases. The most common reasons for rejection are as follows:

1) Particle is less than three diodes in size. Particles smaller than three diodes do not provide sufficient information for classification, and may be treated as spheres for many applications.

2) More than one particle is present. Multiple particles are rejected primarily to make the classification problem more tractable. The argument can also be made that multiple particles are often the result of the breakup of a larger single particle.

3) In the first analysis, particles entering broadside were rejected. Such particles can be included in the analysis by rotating them  $90^\circ$ . In the applications of these algorithms thus far, the authors have found that particles rejected because of their orientation represent a small percentage of the total.

Streakers (usually water shedding across the upstream edge of the probe and across the aperture) have not been rejected in the preprocessing stage, but have been retained as a particle class.

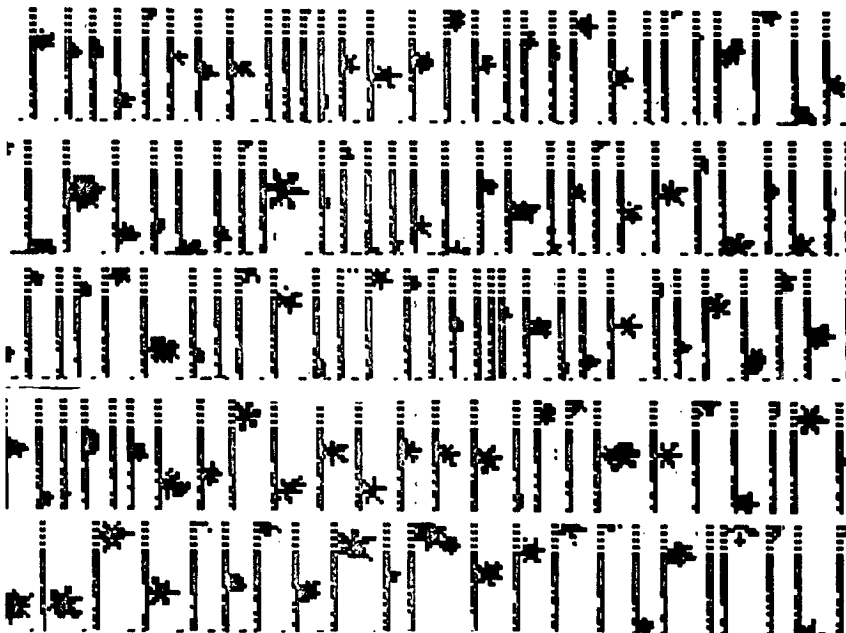


FIG. 1. An example of the shadowgraphs obtained from an airborne PMS two-dimensional probe.

A preprocessing option is to rotate all particles until their ratio of maximum width to length is minimized. Originally this was done to investigate the sensitivity of the algorithms to the particle orientation. As will be discussed later, particle orientation was found to have no effect on the performance of the algorithms for the data analyzed thus far.

*b. Construction of the data vectors*

The second step in the analysis is to construct a linear data vector from the two-dimensional binary arrays of shadowed and exposed diodes.

Figure 2 illustrates how the two-dimensional spectrometer data were converted into a data vector. This figure demonstrates the procedure for two different particles. The upper portion of the figure is the data vector corresponding to the shadow image shown in

the lower portion of the figure. The particle shadows some of the diodes in a 32 by  $N$  array as it passes through the field of view. Each of the X's in the lower figure indicate the location of a shadowed diode. The shadowgraphs of the particles were relocated so that the origin is in the lower left corner of the image shown in Fig. 2, and at least one shadowed diode lies on the first vertical column (abscissa value = 1).

To derive a single numerical value for each column, each shadowed diode was treated as a binary bit of value 1, and each unshadowed diode was treated as a zero. This bit structure was then interpreted as a binary number at each horizontal location. For the purposes of this study, the length  $N$ , of the particle was arbitrarily limited to a maximum of 64. The first 64 numbers in the upper figure are the natural logarithms of the integer resulting from the bit structure of the corresponding column of diodes, normalized to unity square mag-

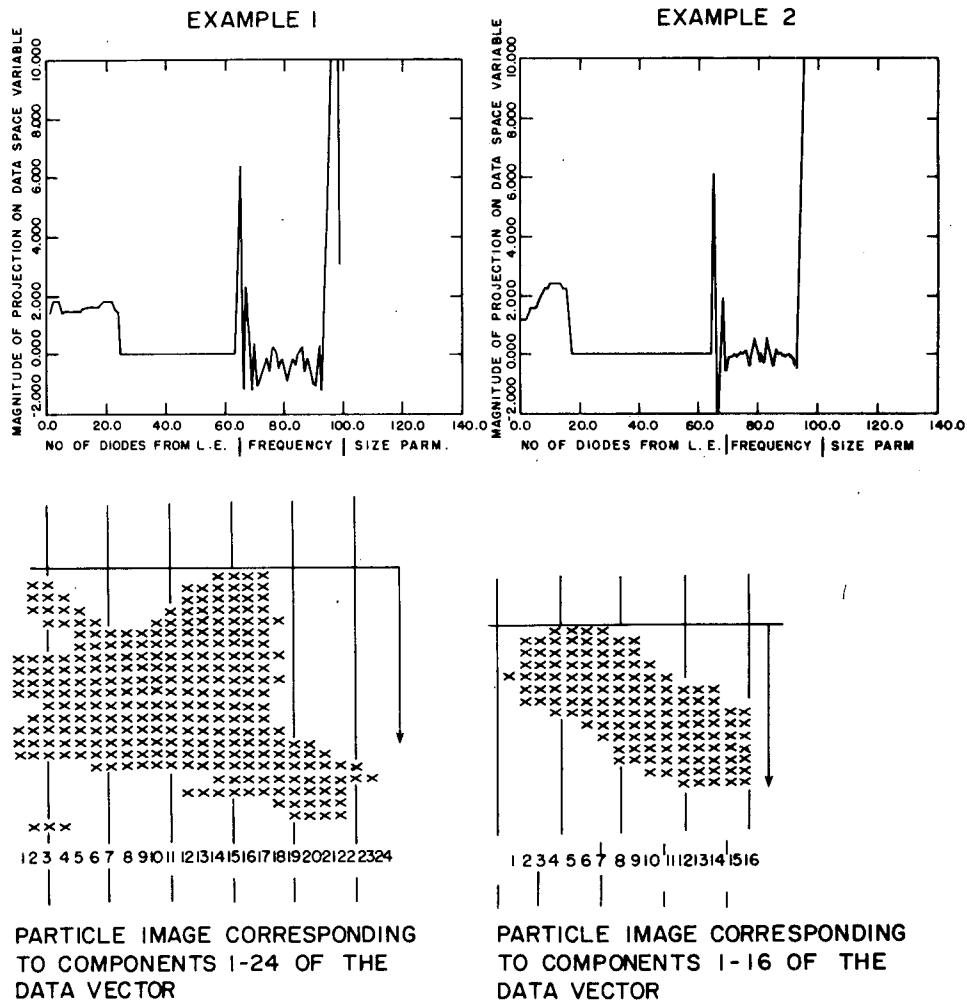


FIG. 2. Illustration of the relation between the image obtained from the PMS two-dimensional probe (lower) and the data vector space (upper). The particle on the left was classified as a dendrite by the human observers, while that on the right was classified as a column. LE is *leading edge* (leftmost side of particle image).

nitude. The next 30 numbers are the frequency spectra obtained when the 64-point bit structure was processed through a fast Fourier transform. The first two bins of the frequency spectra have been deleted to eliminate the DC term. The 95th and 96th values in the data vector are the square magnitude of the time history and the square magnitude of the data history. Points 97 and 98 are the projections of the maximum length and the maximum width. The 99th point is the index value at which the particle had its maximum width.

### c. Definition of classes

The training data used consisted of patterns obtained from the selected parts of four different flights. The samples from these flights were selected because the great majority of the patterns included in each of these samples were similar in character and each appeared to typify a particular crystal type. Manual classification was done for three data sets. These sets may be called, first, the training set (403 examples), second, the manual reference set (201 examples), and finally, the proof test set (1500 examples). Except for the proof test set, manual classifications were made at least twice for each of these data sets. The "correct" identifications for this study were based on a team effort designed to minimize the problem of inconsistencies in manual classification of ambiguous patterns.

After completing their individual analyses of the 201 cases in the manual reference set, the two authors located at AFGL made a joint examination of a total of 2104 images of which the 403 training images and the 201 manual images were a part. These images were arranged in an order such that the authors were unaware of which examples were part of the training set, which were part of the manual reference set, and which were being examined for the first time. A period of time not exceeding one hour was set aside each day to work on this task. A longer period of time would result in fatigue, which in turn could cause inconsistencies in identification. When these images were classified, a comparison was made between the manual classification and that obtained from the computer, using the preliminary algorithms based on the original 403 example training set. Inconsistencies between manual and machine classification were then resolved, and the result was taken as the truth data, that is, the correct classification. It should be remarked that a different group of people would probably produce a somewhat different set of "correct" classifications. This makes it difficult to compare our results with Duroure, whose the "truth" set was classified by only one observer, or with Rahman *et al.* who used idealized data.

### d. Design of machine classifiers

The classifiers developed using the data vectors derived from the shadowgraphs are Fisher classifiers. The

data vectors were processed through an eigenvector transformation, and then the Fisher classifier was obtained through an independent eigenscreening scheme. A description of the ADAPT independent eigenscreening approach to deriving classifiers such as the Fisher classifier has been described by Hunter *et al.* (1981).

The procedure used may be summarized briefly as follows. The training data vectors were transformed to their eigenvector space [that is, to the optimum empirical orthogonal function (EOF) representing the data]. This transformation is then transformed to the eigenvector space. A screening process performed on the projections of each of the EOFs is then used to develop the Fisher classifier. The screening decision to keep or reject a given eigendirection (that is, a feature of the original pattern) is based on an unbiased performance estimate using a modification of Lackenbrach and Mickey's (1968) "one-out" method. Thus, this approach differs from the conventional screening in two ways: the use of the components in eigenvector space as independent variables, and the use of an unbiased performance estimate as opposed to the normal procedure of using dependent or biased test results to make the screening decisions.

Five separate algorithms were developed to classify the particle as a dendrite, a needle, a column or bullet, a plate or sphere, or as a streaker. These algorithms consisted of vectors whose dot product with the data vector produced the detection statistic. The machine classification scheme was built around these five algorithms, each of which was of the same mathematical form. Each algorithm consisted of a vector onto which all of the data were projected. This vector was selected to maximize the ratio of the interclass dispersion to the intraclass dispersion, using the Fisher criteria for the first four algorithms. The fifth (streaker detection) algorithm was formulated by examination of the projection of the data on the first two eigendirections. These five algorithms were used to generate likelihood ratios. The likelihood ratios were based on the detection statistics obtained by applying each of the five algorithms to the training set, using a modification of the one-out method of Lackenbrach and Mickey (1968). The algorithm displaying the highest likelihood ratio was used to define the class to which the shadowgraph belonged.

## 4. Performance of individual algorithms

Table 1 presents a summary of performance of the algorithms. The first column names the algorithm and associates it with one of three groups of algorithms. The next two columns define the members of the classes used in the development of the algorithm. The fourth and fifth columns contain the Fisher parameter and the equivalent probability error, respectively. The performance is given in terms of the Fisher parameter

TABLE 1. Performance of Fisher classification algorithms used for design of classification schema.

Algorithm	Definition of classes*		Performance	
	Class I	Class II	Parameter	Probable error
<i>a. Individual class detectors</i>				
Dendrite detector	Den	Streakers & Misc	0.783	0.095
Needle detector	Ned	Streakers & Misc	0.647	0.06
Column detector	Col	Streakers & Misc	1.59	0.27
Plate detector	Plate	Streakers & Misc	0.80	0.10
<i>b. Paired class</i>				
Column and needle detector	Col & Ned	Streakers & Misc	0.79	0.10
Column and plate detector	Col & Plate	Streakers & Misc	1.30	0.16
Column and dendrite detector	Col & Den	Streakers & Misc	2.00	0.30
<i>c. Classification algorithms</i>				
Dendrite vs needle	Den	All but Den & Ned	0.424	0.02
Dendrite vs column	Den	All but Den & Col	0.685	0.07
Dendrite vs plate	Den	All but Den & Plate	0.966	0.15
Needle vs column	Ned	All but Ned & Col	0.891	0.13
Needle vs plate	Ned	All but Ned & Plate	0.535	0.03
Plate vs column	Plate	All but Plate & Col	0.784	0.095

\* Den, dendrite; ned, needle; col, column.

and an equivalent probability of error. The Fisher parameter is the parameter which is minimized by the Fisher classifier, and is simply the ratio of the sum of the standard deviations of the two classes used to develop the algorithm (the within-class variation), to the distance between the means of the two classes (the between-class variation) when the classes are projected onto the optimal separation direction. The equivalent probability of error is included in Table 1 because the nonlinearity of the Fisher parameter makes it difficult to grasp its physical meaning. We define the equivalent probability error as the probability of making any error in the classification which would be associated with a particular Fisher parameter for a classification where the standard deviation of Class I equals the standard deviation of Class II. This definition allows a unique relationship between a probability of error and the Fisher parameter. In general, the probability of error will depend on the Fisher parameter and on the relative sizes of the standard deviations of the two classes, as well as on the threshold selected. All of the performances given in Table 1 are based on the Lackenbrach and Mickey (1968) one-out method, with the modification that we applied it to groups of observations rather than to single observations.

In examining Table 1, it is useful to realize that the human classifiers had error probabilities ranging between 0.25 and 0.65 when compared with the correct classification, and between 0.50 and 0.84 when each person was compared with his own performance at a different time.

## 5. Definition of classification schema

The information presented in Table 1 was used to select two schema for combining the algorithms into a classification decision procedure suitable for automation. Fig. 3 is a diagram of the two schema developed for this study. Since the thresholds used in each of these schema are easily modified, they provide a mechanism for optimizing the schema for a specific data set.

### *a. One-step classification schema*

The one-step schema shown at the top of Fig. 3 represents the simplest approach to using the algorithms. It consists of processing the unknown particle's data vector through the five detection algorithms and using the projection of the particle on each of the five Fisher directions to calculate a likelihood ratio that the particle belongs to each of the five classes. The likelihood ratios are then compared with a threshold, and if none of them exceed this threshold, the particle is considered unclassifiable, and is put in the miscellaneous category. If one or more of the likelihood ratios exceeds the threshold, the particle is associated with the detection algorithm for which the likelihood ratio is the greatest.

This method results in a significant reduction in the computation required to apply the algorithm. It also provides for considerable additional flexibility, since both the statistics of the training data and the char-

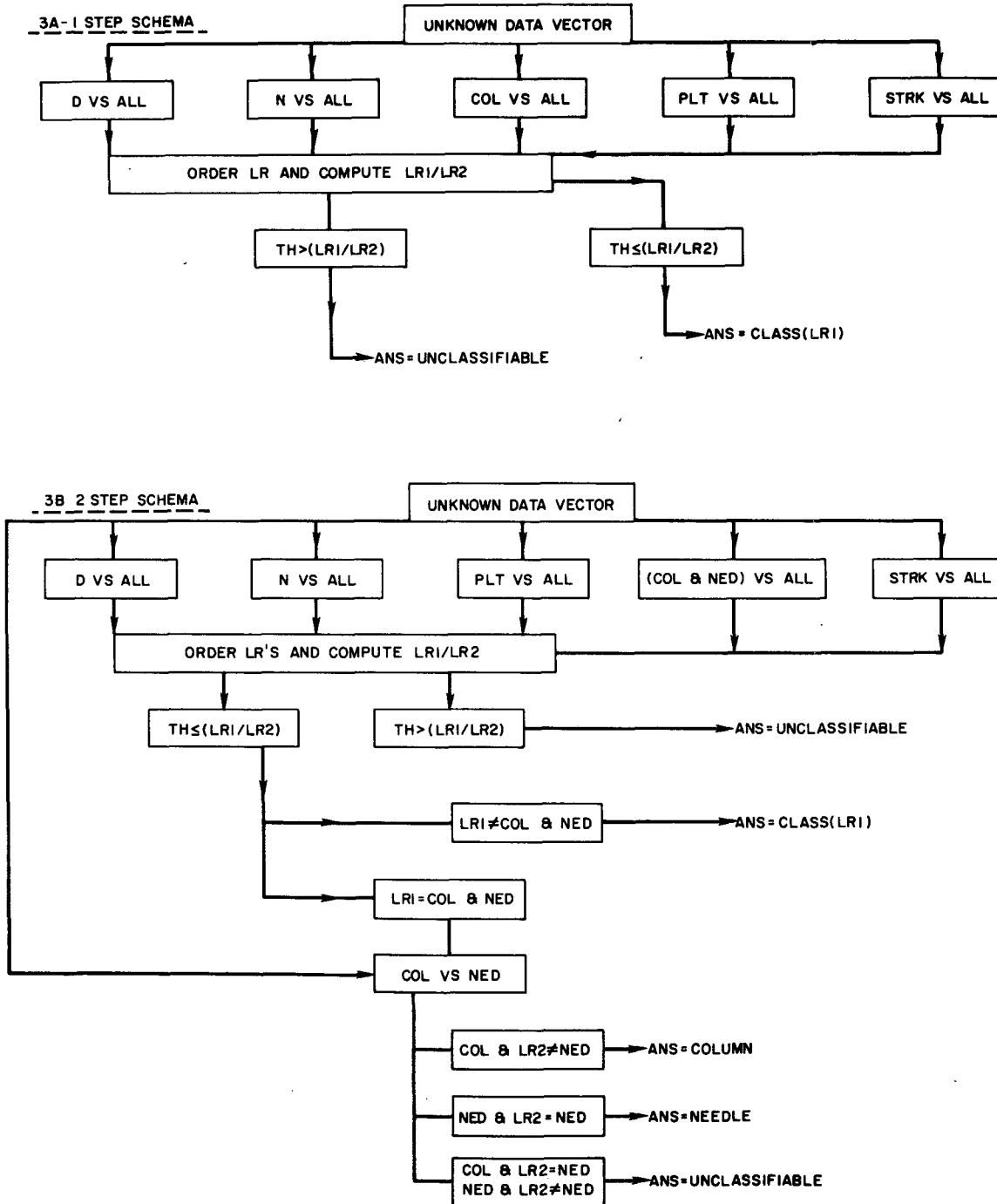


FIG. 3. Flow-diagram illustrating two possible procedures used in applying the individual algorithms to the problem of classifying a given particle. In each procedure (one-step and two-step), the algorithms are applied in turn, and the likelihood ratio (LR) computed. The maximum likelihood ratio (LR1) is compared with the next higher one (LR2). If this ratio exceeds a predetermined threshold, then the particle is classified as that producing LR1.

acteristics of the individual algorithm can be modified by an appropriate input. Thus, the algorithms can be adjusted to include specific *a priori* information, without rederiving the entire algorithm.

*b. Two-step classification schema*

Figure 3b is a modification of the simple schema shown in Fig. 3a. An additional algorithm has been

added, and one of the algorithms has been replaced with two algorithms whose combined performance is better than the single algorithm. In comparing the two-step schema with the one-step schema, note first that the column detector has been replaced by an "either column or needle" detector. If the likelihood ratio for this new algorithm is less than the threshold value, or if the maximum likelihood ratio is associated with any of the other four algorithms, the procedure and results are identical to the one-step schema. However, if the maximum likelihood ratio is associated with the columns and needles class, the data vector is then processed through the columns versus needles classification algorithm. The results of this classification are then compared with the second largest likelihood ratio, and a decision is reached as to whether the particle should be called a column, a needle or miscellaneous.

## 6. Performance of classification schema

The performance of the raw algorithms was summarized in Table 1 in terms of the Fisher parameter and its associated equivalent probability of error. However, the performance of interest is that of the combination of these algorithms for the task of identifying the particle type associated with the given data vector. Although the Fisher parameter is an excellent measure of the performance of the individual algorithm, it is not suitable for evaluating the schema, which involve the combined performance of several algorithms. A more efficient measure of the performance of a combination of algorithms is the confusion matrix.<sup>2</sup> In the confusion matrix, the vertical axis is the actual class of the particle. The horizontal axis is the class to which the particle has been assigned by the classifier. Thus, the diagonal of this matrix is simply the number of correct classifications for each algorithm.

Table 2a presents the confusion matrix for the application of the one-step algorithm to the 403 training cases. Examining the first row of the matrix, we note that 46 particles were correctly identified as dendrites, none of the particles which should have been classified as dendrites were incorrectly identified as needles or as streakers, but four were incorrectly identified as columns, eleven were incorrectly identified as plates, and two were classified as miscellaneous. Similarly, one can get the exact performance of the algorithms for each of the particle classes by examining the other rows of the confusion matrix. It is generally useful to normalize the confusion matrix by dividing each row by the total number of particles of that class. The normalized confusion matrix corresponding to Table 2a is shown in Table 2b. Most were correctly identified,

<sup>2</sup> The confusion matrix is a measure of how well particles (actual) in each class (dendrites, needles, etc.) are identified (called) by the computer algorithm.

TABLE 2. Confusion matrix for group out testing of the training set of cloud particles.

Actual class	Called class					
	1	2	3	4	5	6
<i>a. Confusion matrix based on counts</i>						
1	46.00	0.0	4.00	11.00	0.0	2.00
2	1.00	90.00	11.00	0.0	0.0	3.00
3	6.00	8.00	62.00	3.00	0.0	10.00
4	8.00	5.00	9.00	90.00	0.0	8.00
5	1.00	0.0	0.0	0.0	2.00	0.0
6	1.00	10.0	6.0	4.00	0.0	2.00
<i>b. Normalized confusion matrix</i>						
1	0.73	0.0	0.06	0.17	0.0	0.03
2	0.01	0.86	0.10	0.0	0.0	0.03
3	0.07	0.09	0.70	0.03	0.0	0.11
4	0.07	0.04	0.07	0.75	0.0	0.07
5	0.33	0.0	0.0	0.0	0.67	0.0
6	0.04	0.43	0.26	0.17	0.0	0.09

while 6% of them were incorrectly classified as columns, 17% as plates, and 3% as miscellaneous. Table 3 contains the confusion matrices for the 201-case manual reference set of particles.

The confusion matrix provides a complete analysis of the performance of each algorithm. However, it is a rather cumbersome way to compare schema. One simple solution is to consider only the diagonal of the confusion matrix. This will tell us the percent of correctly identified particles in each class.

A comparison of the performance of nine separate schema when applied to the training set data is given in Table 4. The first column identifies the schema used. The next three columns list the thresholds used

TABLE 3. Confusion matrix for the manual reference set of cloud particles.

Actual class	Called class					
	1	2	3	4	5	6
<i>a. Confusion matrix based on counts</i>						
1	37.00	0.0	1.00	4.00	0.0	0.0
2	3.00	16.00	5.00	2.00	0.0	1.00
3	46.00	2.00	32.00	10.00	0.0	3.00
4	7.00	0.0	3.00	13.00	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0
6	1.00	4.00	2.00	7.00	0.0	2.00
<i>b. Normalized confusion matrix</i>						
1	0.88	0.0	0.02	0.10	—	0.0
2	0.11	0.59	0.19	0.07	—	0.04
3	0.49	0.02	0.34	0.11	—	0.03
4	0.30	0.0	0.13	0.57	—	0.0
5	0.0	0.0	0.0	0.0	—	0.0
6	0.06	0.25	0.13	0.44	—	0.13

TABLE 4. Comparison of schema performance using confusion matrix diagonal on 403 case training set.

Schema	Thresholds				Diagonal values					
	LR*	Dendrite	Second step	Data rotated	Dendrite	Needle	Column	Plate	Streak	Misc.
1. 1-step (nominal)	1.0A	0	—	no	0.73	0.86	0.70	0.75	0.67	0.09
2. 1-step (best)	1.05	0	—	no	0.73	0.89	0.72	0.80	0.33	0
3. 2-step (best)	1.05	0	-7.5	no	0.73	0.93	0.71	0.80	0.33	0
4. 2-step	1.05	0	-6.5	no	0.73	0.98	0.61	0.80	0.33	0
5. 2-step	1.05	0	-8.5	no	0.73	0.91	0.74	0.80	0.33	0
6. 1-step (best/rotated)	1.05	0	—	yes	0.70	0.93	0.65	0.78	0.33	0
7. 2-step (best/rotated)	1.05	0	-7.5	yes	0.70	0.98	0.64	0.78	0.33	0
8. 1-step (weak dendrite)	1.05	-2.2	—	no	0.51	0.90	0.75	0.85	0.33	0
9. 2-step (weak dendrite)	1.05	-2.2	-7.5	no	0.51	0.94	0.74	0.85	0.33	0

\* Likelihood ratio.

to modify the schema. These are: the threshold on the likelihood ratio (LR); the threshold on the dendrite detection algorithm; and the threshold on the second step of the 2-step algorithm for separating needles from columns. The fifth column of Table 4 indicates whether or not the particles were rotated. The final six columns are the diagonal of the confusion matrix.

The first case shown in Table 4 is the one-step algorithm for which the confusion matrix was given in Table 2. The likelihood ratio was set at an absolute value of 1.0 (indicated as 1.0A in the table). The dendrite threshold was set at zero, corresponding to the minimum total errors value, using the Anderson and Bahadur (1962) approach. Since this is a one-step algorithm, the second step threshold is not applicable. Comparing the last six columns in this case with Table 2b, we note that they correspond to the diagonal of the normalized confusion matrix.

Table 5 has the same format as Table 4, and compares the performance of the schema when applied to the manual reference set. These are 201 cases which were not part of the training set, but which were obtained from typical flights, rather than unusually good flights. In fact, further analysis revealed that these flights

represent particles which were less well defined than the average case. Dividing the sum of the diagonal by 201 shows that the average machine performance was 50% correct classification. The manual classification of these particles by ten different people varied between 35 and 75% correct identification.

Table 6 presents the confusion matrix for the application of the best one-step algorithm to the 1500-case proof test set. The performance on this larger set lies between those of the training and the manual reference sets.

### 7. Conclusions

Two schema for the machine classification of cloud particles as observed by PMS two-dimensional probes have been developed and analyzed. These schema are based on detection and classification algorithms derived from real data, and have incorporated into them flexibility in the form of variable critical thresholds. This permits a tailoring of the algorithms to suit the needs of the individual user or the characteristics of the data. The performance of these schema using the nominal settings for the thresholds has been found superior to that of manual classification.

TABLE 5. Comparison of schema performance using confusion matrix diagonal on 201 case manual reference set.

Schema	Thresholds				Diagonal values					
	LR*	Dendrite	Second step	Data rotated	Dendrite	Needle	Column	Plate	Streak	Misc.
1. 1-step (nominal)	1.0A	0	—	no	0.88	0.59	0.34	0.57	—	0.13
2. 1-step (best)	1.05	0	—	no	0.86	0.67	0.42	0.61	—	0
3. 2-step (best)	1.05	0	-7.5	no	0.86	0.74	0.40	0.61	—	0
4. 2-step	1.05	0	-6.5	no	0.86	0.74	0.39	0.61	—	0
5. 2-step	1.05	0	-8.5	no	0.86	0.67	0.42	0.61	—	0
6. 1-step (best/rotated)	1.05	0	—	yes	0.86	0.70	0.39	0.61	—	0
7. 2-step (best/rotated)	1.05	0	-7.5	yes	0.86	0.74	0.40	0.61	—	0
8. 1-step (weak dendrite)	1.05	-2.2	—	no	0.60	0.67	0.58	0.74	—	0
9. 2-step (weak dendrite)	1.05	-2.2	-7.5	no	0.60	0.70	0.59	0.74	—	0

\* Likelihood ratio.



TABLE 6. Confusion matrix for 1500 object test set.

Actual class	Called class					
	1	2	3	4	5	6
<i>a. Confusion matrix based on counts</i>						
1	214.00	8.00	37.00	92.00	2.00	39.00
2	7.00	196.00	19.00	0.0	1.00	3.00
3	120.00	13.00	142.00	13.00	1.00	54.00
4	66.00	12.00	56.00	293.00	0.0	41.00
5	0.0	0.0	0.0	0.0	1.00	0.0
6	4.00	38.00	5.00	15.00	0.0	8.00
<i>b. Normalized confusion matrix</i>						
1	0.55	0.02	0.09	0.23	0.01	0.10
2	0.03	0.87	0.08	0.0	0.0	0.01
3	0.35	0.04	0.41	0.04	0.0	0.16
4	0.14	0.03	0.12	0.63	0.0	0.09
5	0.33	0.0	0.0	0.0	1.00	0.0
6	0.06	0.54	0.07	0.21	0.0	0.11

Results presented in Tables 4, 5 and 6 show that even the variation between real sets of data can be expected to have significant effects on the proper design of the classification schema. Thus, it is extremely unlikely that algorithms derived from artificial training data will yield algorithms of practical use in the classification of PMS two-dimensional data obtained under normal circumstances.

The classification schema developed here make use of the Fisher classification scheme rather than the more complicated maximum likelihood and Bayes family of classifiers. Admittedly, the Fisher classifier is a simplification, and may produce only approximate results compared with the more exact classifiers. However, the problems associated with the definition of good training sets and the lack of consistency of data between data sets undoubtedly result in greater imprecision. The significant savings in computation time, both in the development of the classifiers and in their application, justifies the use of these computationally more efficient algorithms.

The basis for human classification appears to differ significantly from that for the machine classifications developed in this study. One indication of this difference is the relatively small number of "correct" identifications of the miscellaneous class, illustrated by the last column of Tables 4 and 5.

Comparison of the performance of the schema on

rotated and unrotated particles shows that the procedures developed in this study are not sensitive to particle orientation, and therefore it is not necessary to rotate any of the particles prior to applying the algorithms. This also indicates that the method used to construct the data vectors provides shape information sufficiently uncoupled from orientation information that they may be separated by a linear algorithm. These algorithms have been applied successfully to data taken within the melting layer (Dyer and Glass, 1982).

## REFERENCES

- Anderson, T. W., and R. R. Bahadur, 1962: Classification into two multivariate normal distribution with different covariance matrices. *Ann. Math. Statist.*, **33**, 420-431.
- Cunningham, R. M., 1978: Analysis of particle spectra data for optical array (PMS) 1D and 2D sensors. *Preprints 4th Symp. Meteorological Observations.*, Denver, Amer. Meteor. Soc., 345-350.
- Durooure, C., 1982: Une nouvelle methode de traitement des images d'hydrometeores donnees par les sondes bidimensionnelles. *J. Rech. Atmos.*, **6**, 71-84.
- Dyer, R. M., and A. A. Barnes, 1979: The microphysics of ice clouds—a survey. AFGL-TR-79-0103, 38 pp. [AD-A077020].
- , and M. Glass, 1982: Observed changes in ice crystal type in thick stratiform clouds. *Preprints Conf. Cloud Physics*, Chicago, Amer. Meteor. Soc., 201-202.
- Heymsfield, A. J., 1976: Utilization of aircraft size spectra measurements and simultaneous Doppler radar measurements to determine the physical structure of clouds. *J. Atmos. Sci.*, **29**, 1358-1366.
- , and R. G. Knollenberg, 1972: Properties of cirrus generating cells. *J. Atmos. Sci.*, **29**, 1358-1366.
- Hunter, H. E., E. R. Rodgers and W. E. Shenk, 1981: An objective method for forecasting tropical cyclone intensity using Nimbus-5 electrically scanning microwave radiometer measurements. *J. Appl. Meteor.*, **20**, 137-145.
- Knollenberg, R., 1970: The optical array: An alternative to scattering or extinction for airborne particle size determination. *J. Appl. Meteor.*, **9**, 86-103.
- , 1975: The response of optical array spectrometers to ice and snow: a study of probe size to crystal mass relationships. AFGL-TR-75-0494, 70 pp. [AD-A020276].
- , 1976: The response of optical array spectrometers to ice and snow: a study of 2-D probe area-to-mass relationships. AFGL-TR-76-0273, 35 pp. [AD-A034741].
- Lackenbrach, P. A., and M. R. Mickey, 1968: Estimation of error rates in discriminant analysis. *Technometrics*, **10**, 11-17.
- Rahman, M. M., R. G. Jacquot, E. A. Quincy and R. E. Stewart, 1981a: Two-dimensional hydrometeor image classification by a statistical pattern recognition algorithm. *J. Appl. Meteor.*, **20**, 536-546.
- , E. A. Quincy, R. G. Jacquot and M. J. Magee, 1981b: Feature extraction and selection for pattern recognition of two-dimensional images. *J. Appl. Meteor.*, **20**, 521-535.