The Classification of Ambiguous Ice Particle Shadowgraphs by Consensus

ROSEMARY M. DYER AND MORTON GLASS
Air Force Geophysics Laboratory, Hanscom Air Force Base, MA 01731

HERBERT E. HUNTER*
Adapt Corporation, Reading, MA 01867

11 October 1984 and 10 April 1985

ABSTRACT

A major impediment to the development of computer algorithms for the automatic classification of ice particle types found in the atmosphere as measured by a Particle Measuring System two-dimensional probe is the difficulty of obtaining training data. This is especially true when, as is usually the case, the particle shapes do not correspond to any of the pure crystal types found in textbooks.

This paper presents the results of testing such a training set. Sources of bias among human observers include the effect of training and previous familiarity with the data, fatigue, and particle orientation, as well as subjective differences among observers. The deviation of individual human observers from the classifications arrived at by consensus indicates an upper bound to the accuracy possible in automated classification schemes.

1. Introduction

The many forms ice crystals and snowflakes take are known to be dependent on the temperature and moisture content of the environment in which the particles grow and through which they fall. The observation and characterization of ice crystal types are therefore important to understanding cloud growth and precipitation processes.

One method of observing cloud and precipitation size particles is the airborne optical array probe (Knollenberg, 1976). This instrument is capable of collecting very large data sets, and produces two-dimensional (2-D) shadow images of individual particles. Two versions of the probe were available to us: the "cloud" probe with a 25 μm resolution, and the "precipitation" probe with a 200 μm resolution. The cloud probe was used in this study. Many problems have arisen in the analysis of data obtained in this manner. In this paper we will deal specifically with the difficulty of interpreting the images to identify the ice particle type.

We became aware of this problem when we tried to derive a training set for the construction of computer algorithms that would permit the automated classification of ice particle type (Hunter *et al.*, 1984). Deriving computer algorithms for any form of pattern recognition requires that there be a sufficient number of samples for which the identification is known. In identifying ice particle types from their shadowgraphs, the problem is that a significant number of the images to be classified are so ambiguous that there is no consensus as to their correct identification. Human classification is highly dependent (because of the tedious nature of the task) on the individual's physical and mental condition at the time the classification is made. Indeed, one of the reasons for computer algorithms is to eliminate the inconsistency that arises when the observer is fatigued or approaches the data with certain unconscious biases. However, in order to derive the algorithms, it is first necessary to produce a set of data on which to train the machine. Unless great care is taken in defining the "true" particle classifications, the algorithms derived from the training set will be either meaningless or contain a systematic bias. In any event the ambiguity of the images places an upper limit on the accuracy to be obtained from computer identification algorithms. This can be no higher than the accuracy of an unbiased, unfatigued observer.

2. The data

The training data set consisted of 403 particularly good examples of ice particle types. Care was taken to ensure that the particles were evenly balanced among the four classifications used: 1) plates and spheres; 2) needles; 3) dendrites; and 4) columns and bullets. An example of each of these particle types is given in Fig. 1. These broad crystal classifications were chosen so that the differences evident in their characteristics could be most clearly discerned in the shadow images produced by the 2-D probe. Even for this data set, there were some particle identifications for which the two authors–meteorologists (Dyer and Glass) were not in

* Current affiliation: Nichols Research, Huntsville, AL 35802.

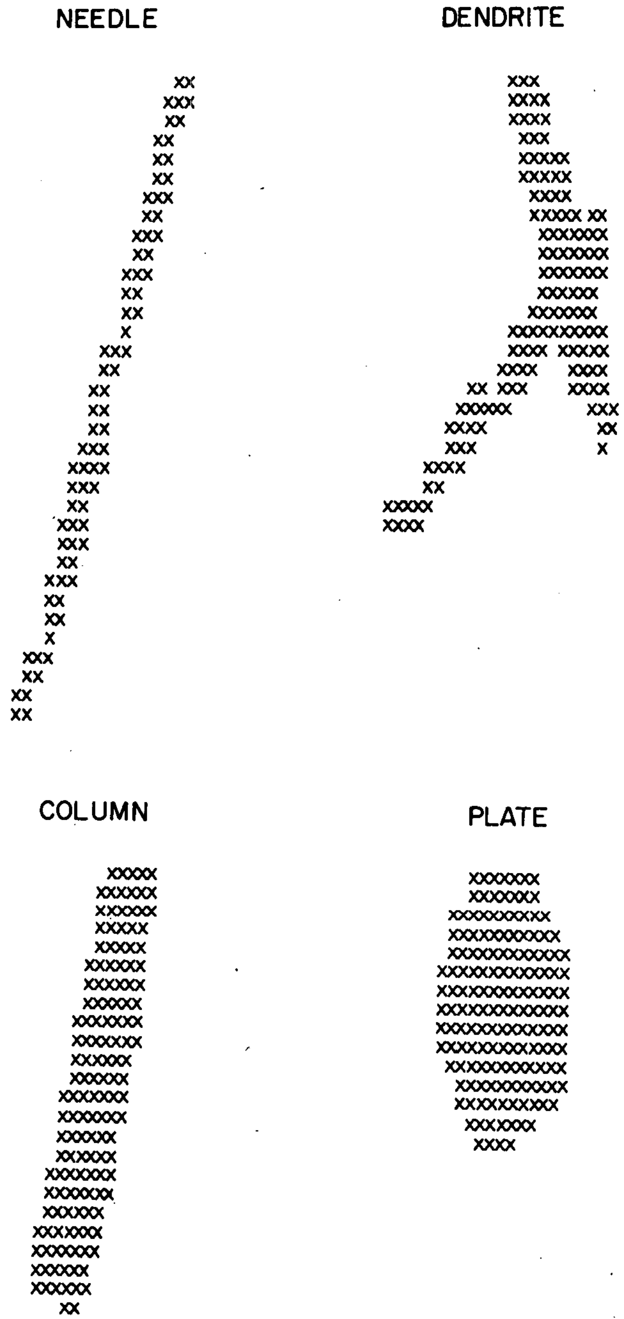

FIG. 1. Typical shadowgraphs of ice crystals as recorded by an airborne optical array probe. Each X represents an occluded square diode with an effective dimension of 25 micrometers. The patterns shown here represent each crystal type considered, and caused no disagreement among the classifiers.

complete accord. The computer scientist then derived various algorithms. The one-step schema resulted in machine classifications that agreed with the true identification for an average of 75% of the plates and spheres, 86% of the needles, 73% of the dendrites, and 70% of the columns; that is, an average agreement of 76%.

Realizing that the training data set was not representative of the particle images often observed in the atmosphere, we then examined a set of 201 randomly chosen particles. Dyer and Glass each classified the particles independently. They then collaborated to determine the "correct" classifications. This procedure is described in detail in Hunter *et al.* (1984). In addition, Hunter, four meteorologists and four technicians were asked to classify the particles manually. The distinction between the two groups is a matter of differences in formal education in meteorology (particularly in cloud physics).

Two of the meteorologists had experience of a similar nature in this work. They routinely were aboard the aircraft while the data collection flights were being made. Their assignment during the flights included classifying particle types for notation on the data logs. The other two meteorologists, and all four of the technicians, received some instruction on the particular characteristics of each of the crystal types from one of the three authors before beginning their classifications.

3. Results

The agreement matrix relating the classifications of the 11 individuals with the correct identification and with each of their fellow classifiers is given in Table 1. The first column provides a classifier number, and the second column gives a descriptive title for the classifier. "Correct 1" is the identification arrived at by collaboration. "Computer 13" is the identification obtained by applying the algorithms. Meteorologists and technicians are labelled "Met" or "Tech," respectively. The letter D, G or H indicates that the classifier received some instruction from one of the authors.

Several interesting observations can be made after examining Table 1. For example, Met-A and Met-B are the two meteorologists with the most experience classifying particle types during data gathering aircraft flights. Yet, their proportion of correct identifications were lower than those of all other classifiers except the meteorologist trained by Glass. Interviews with all three meteorologists elicited the information that in the field, where they generally fly for a period of time at constant altitude, they expect to see particles of only one or two types during a given time interval. Therefore, when they were classifying the test data they tended to identify "strings" of particles, placing four or five or more patterns in the same category, until they came to an image that deviated blatantly from the previous ones. They could not adjust their classification system to deal with each image in isolation from its neighbors. The exception to this was the meteorologist trained by Dyer. He had received the one additional piece of information that the images he was being asked to identify came from several aircraft flights, and had been chosen at random and shuffled. His proportion of correct identifications was 0.66, compared with 0.44, 0.46 and 0.34 for the other three meteorologists.

TABLE 1. Agreement matrix comparing performance of manual and machine classification of 201 randomly chosen particles. To find the correlation between any two observers, first find their positions in the left hand column (for example, Glass (3) and Tech:D (9). The correlation is found at the junction of the column corresponding to one and the row corresponding to the other, (In this example column 3, row 9 has the value 0.48). The computer generated algorithm are seen in this table to have a 0.53 correlation with the "correct" identification.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Correct	1.00												
2. Dyer	0.67	1.00											
3. Glass	0.73	0.60	1.00										
4. Hunter	0.66	0.67	0.57	1.00									
5. Met-A	0.44	0.47	0.52	0.41	1.00								
6. Met-B	0.46	0.48	0.48	0.43	0.53	1.00							
7. Met:G	0.34	0.42	0.37	0.34	0.52	0.48	1.00						
8. Met:D	0.66	0.61	0.61	0.68	0.41	0.46	0.29	1.00					
9. Tech:D	0.53	0.49	0.48	0.58	0.38	0.44	0.33	0.58	1.00				
10. Tech:G1	0.66	0.68	0.56	0.72	0.39	0.40	0.27	0.65	0.53	1.00			
11. Tech:G2	0.65	0.65	0.67	0.68	0.49	0.49	0.43	0.66	0.52	0.66	1.00		
12. Tech:H	0.65	0.66	0.60	0.81	0.42	0.42	0.35	0.64	0.52	0.67	0.65	1.00	
13. Computer	0.53	0.39	0.36	0.38	0.20	0.25	0.14	0.44	0.32	0.40	0.33	0.37	1.00

This experience led to a closer look at the effect of the type and degree of instruction on the results. Table 2 contains the pertinent information, abstracted from Table 1. It should be noted here that Hunter (at that time) had been instructed by Dyer, and that the technician instructed by Hunter was involved only in the derivation of the algorithms and programming them on the computer, not with the meteorological data of which the ice particle images were only a part.

In her instructions to both the meteorologist and the technician, Dyer showed examples of each classification, and stated only that each image was to be treated separately, the ensemble of images not representing any particular data-gathering flight. Glass provided a minimum of instruction to the meteorologist, but went into some detail with his instructions to the technicians, providing several examples of ambiguous shadowgraphs, and indicating how he would classify them. The last column of Table 2, showing the percent frequency of agreement between teacher and pupil, is more an indication of the teacher's influence on the pupil's selections than it is of any clarity in the images.

Three of the classifiers went over the particle images a second time, after a two to three week interval. The results of their second look are presented in Table 3. The third column of this table shows the percentage of particles for which the observer did not change his identification after a second look. Note that the observer designated "Met B" changed his identification for exactly half the particles. This dramatic change is most likely due to a conscious effort to consider each of the shadowgraphs in isolation, rather than part of an ensemble of particles obtained during a single, constant altitude, aircraft pass. However, it is worth noting that Hunter and the technician instructed by Hunter, both of whom had been spending full time on the problem of automating the identification of the ice particles, were still not in perfect agreement in assignments of particle classification. The column labeled "% agreement" compares the identification of each particle on each occasion. The high correlation here is an indication of the extent to which subjective factors affect particle identification on separate occasions.

Figure 1 contains shadowgraphs for which all the observers agreed on the identification. Examples of images for which there was disagreement are shown in Fig. 2. In each of these cases, opinion was evenly divided between two possible categories, with many classifiers opting for the miscellaneous (i.e., unclassifiable) category. A closer examination of these particles shows why there is such difficulty in getting agreement on their classifications. Since we are dealing only with shape, size should not influence the choice of particle

TABLE 2. Comparison between teacher and pupil in identification exercise.

	Scores		
Teacher–pupil	Teacher	Pupil	% Agreement
Dyer–Met:D	0.67	0.66	61
Dyer–Tech:D	0.67	0.53	49
Glass–Met:G	0.73	0.34	37
Glass–Tech:G-1	0.73	0.66	56
Glass–Tech:G-2	0.73	0.65	67
Hunter–Tech:H	0.66	0.65	81

TABLE 3. Effect of time on the identification exercise.

	Percent correct		
Classifier	First time	Second time	Percent agreement between first and second times
Hunter	66	74	84
Met B	46	61	50
Tech:H	65	66	71

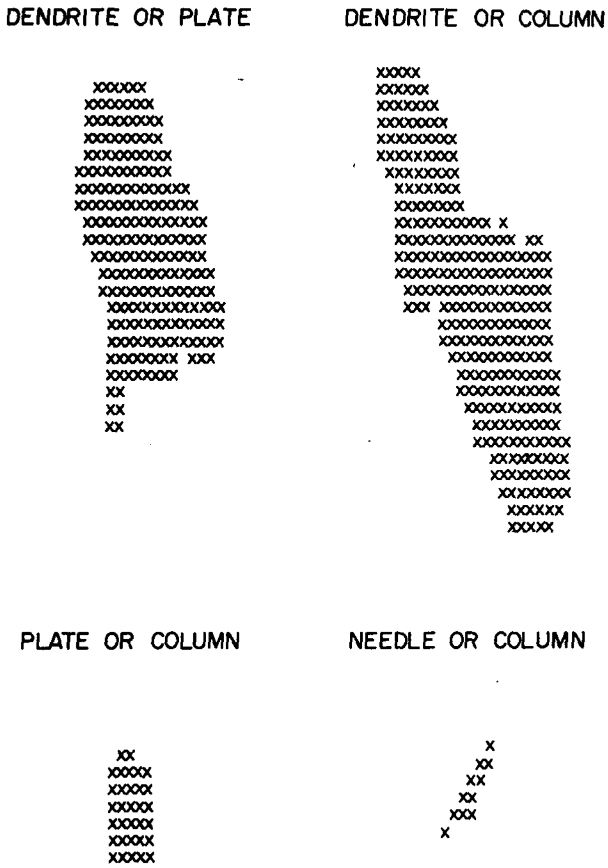

FIG. 2. Examples of ambiguous shadowgraphs. Particle classification by human observers was divided between the two possibilities indicated for each image. Compare with the patterns shown in Fig. 1.

type. However, a discontinuity on a large particle may indicate the broken arm of a dendrite, while on a smaller particle the same shape discontinuity might be due to the finite size of the pixels. The size of a particle in relation to the size of the discontinuity is a subjective parameter; one person would classify the particle as a dendrite while another mentally filled in the discontinuity, classifying the particle as a plate or column. Similarly, the length-to-width ratio for which one identifies an object as a needle or a column varies from individual to individual. It may also vary according to the orientation of the particle for different human classifiers. Indeed, our results show that this ratio may vary for the same individual at different times.

A study of the use of the classification "miscellaneous," or "unclassifiable," also demonstrates the effect of human personality and training on the identification problem. All of the participants in this study were told that if they really could not decide between two or more possible categories for a particular particle, they were to classify the particle as miscellaneous. The percent of miscellaneous particles varied between 4.7 (for tech:G-2) and 47.8 (for Met:G). The three meteorologists with the lowest percentage of correct classifications (Met-A, Met-B, and Met-G) had the highest percentages of particles classified miscellaneous. Again, this reflects their experience classifying particle types during data gathering flights. Because of the stress due to the requirement for real-time identification during a flight, there was a tendency to group all the particles into one type (a tendency discussed above), or to refer to all the particles as "junk," "miscellaneous aggregates," or simply "miscellaneous."

Even though these meteorologists knew that the purpose of the exercise was to provide training data so the computer could classify the particles on a particle by particle basis, they injected their subjective bias into the analysis. The technicians, on the other hand, tended to force the images into one or another of the four designated classifications, whether they fit or not. The classifiers with notably low percentages of miscellaneous particles were all technicians. For comparison, the percentage "miscellaneous" by the two meteorologist–authors of this paper were 12 and 13. The category with the least degree of agreement among the various classifiers was the miscellaneous group.

4. Conclusions

Some of the sources of inconsistency in particle identification are readily apparent, and should be avoided in the future. These include:

1) The (variable) influence of the instructions. This is evidenced by the anomalous performance of Met-D, the only meteorologist who was informed of the random nature of the data set, and by the difference in the performance of Met-B, after he was apprised of this fact;

2) The effect of the (again variable) desire of some participants to get a high score. This is especially noticeable among the technicians, who were reluctant to classify a particle as miscellaneous (essentially admitting defeat), and who, in some instances asked for further instruction;

3) The change in an individual's identification after more familiarity with the correct answers. This is demonstrated by Hunter's higher percentage of agreement with the test set after a period of time. The danger of too much instruction leading to conformity, but not necessarily to the best identification, will have to be considered;

4) The effect of orientation and size on the human classifiers. This varies from individual to individual, and with time for a given individual.

Beyond illustrating ideas well known to psychologists—that personality and individual biases influence the identification process—this paper demonstrates the importance of recognizing that snow particle shapes are not necessarily similar to well-defined crystal types. Snow particles observed in storm systems are likely to have encountered varying temperature and humidity conditions during their evolution. Their growth histories and ultimate shapes reflect this. The identification problem is further complicated by the mechanical processes of shattering and agglomeration, which produce significant numbers of particles having an apparently random variety of shapes.

Characterizing a particle by its 2-D shadowgraph according to standard crystal classifications has the goal of understanding the microphysical environment principally associated with its growth history. Therefore, identifying the predominant particle shapes in a region of the cloud would give information on the microphysical processes. These characteristics can be identified only by the use of an objective process from which as much individual biases as possible have been removed, and which efficiently examines a statistically representative sample. Use of a replicator with its ability to display particle shape in fine detail could add additional information on the microphysical environment in which the particles evolve. However, this would be impractical as an aid in identifying shadowgraphs, since it would require capture of the same particles viewed by the probe. Identification of replicated particles also suffer from individual biases.

Classification algorithms based on idealized crystals are not recommended because they result in the rejection of most particles occurring in nature. Except for the rare well-defined particle (which invariably attracts the analyst's attention), the 2-D shadowgraphs represent a set of coarse analogs of shape for which there is no "ground truth"; that is, there is no way to obtain the *true* identification of the particle. For this reason, a technique based on consensus rather than individual interpretation is suggested.

REFERENCES

Hunter, H. E., R. M. Dyer and M. Glass, 1984: A two-dimensional hydrometer machine classifier derived from observed data. *J. Atmos. Oceanic Technol.* **1**, 28–36.

Knollenberg, R., 1976: Three new instruments for cloud physics measurements: The 2-D spectrometer, the foreward scattering spectrometer probe, and the active scattering aerosol spectrometer. *Proc., Int. Conf. on Cloud Physics, Boulder,* Amer. Meteor. Soc., 554–561.