

## A Technique for Habit Classification of Cloud Particles

A. KOROLEV AND B. SUSSMAN

*Meteorological Service of Canada, Downsview, Ontario, Canada*

(Manuscript received 23 August 1999, in final form 9 November 1999)

### ABSTRACT

A new algorithm was developed to classify populations of binary (black and white) images of cloud particles collected with Particle Measuring Systems (PMS) Optical Array Probes (OAPA). The algorithm classifies images into four habit categories: "spheres," "irregulars," "needles," and "dendrites." The present algorithm derives the particle habits from an analysis of dimensionless ratios of simple geometrical measures such as the  $x$  and  $y$  dimensions, perimeter, and image area. For an ensemble of images containing a mixture of different habits, the distribution of a particular ratio will be a linear superposition of basis distributions of ratios of the individual habits. The fraction of each habit in the ensemble is found by solving the inverse problem. One of the advantages of the suggested scheme is that it provides recognition analysis of both "complete" and "partial" images, that is, images that are completely or partially contained within the sample area of the probe. The ability to process "partial" images improves the statistics of the recognition by approximately 50% when compared with retrievals that use "complete" images only. The details of this algorithm are discussed in this study.

### 1. Introduction

Knowledge of cloud ice particle habits is of great importance for many aspects of cloud physics. The shape of an ice particle influences its scattering properties, growth rate, and terminal fall velocity. The scattering properties of cloud ice particles may significantly affect the climate and radiation balance of the whole earth. Scattering properties of ice particles are also important for satellite retrievals and remote sensing. An ice particle's shape affects its rate of growth through the processes of vapor deposition and riming. These mechanisms define the rate of precipitation growth, which is important for weather and forecasting models. The terminal fall velocity influences the lifetime of ice clouds which may, in turn, effect the radiation budget of the earth.

In the mid 1970s Particle Measuring Systems, Inc. (PMS) developed Optical Array Probes (OAPs) for aircraft measurements of two-dimensional (2D) shadow images of cloud particles (Knollenberg 1981). Since that time, these instruments have become one of the major tools for studies of the sizes and shapes of cloud particles.

Several different algorithms have been developed for classification of cloud particle images measured by

OAP-2D. Rahman et al. (1981) used different geometrical parameters applied to artificial idealized images of ice particles and maximum likelihood Bayes family classes (Tou and Gonzales 1974). Duroure (1982) performed a radial analysis of particle image outlines on a selected dataset. A Fourier transform was then used to extract shape discrimination from the image parameters of the analyzed dataset. Hunter et al. (1984) used a specially constructed vector consisting of FFT radial frequency, size parameter, and sums of occulted pixels in image slices. Holroyd (1987) used several parameters including a product moment correlation coefficient and a dimensionless ratio  $PD_{\max}/S$  to classify ice particle images, where  $P$  is particle image perimeter,  $S$  is the square area of the image, and  $D_{\max}$  is maximum length of the image. Moss and Johnson (1994) used a pattern recognition technique similar to one described by Duroure (1982) but implemented different criteria for categorizing cloud particles. Duroure et al. (1994) used the geometrical measures  $S$  and  $P$  of individual particles integrated over a population of images to relate to particle habit and size distribution of the population. Garbrick et al. (1995) developed a method based on a neural network techniques.

The current work presents an originally developed algorithm for automated particle classification of particle images generated with a PMS OAP-2D. From a given ensemble or "team" of images this algorithm deduces the fraction of cloud particles that belong to a specific habit by analyzing the distribution of ratios of geometrical measures calculated for the entire ensemble.

---

*Corresponding author address:* Dr. Alexei V. Korolev, Cloud Physics Research Division, Meteorological Service of Canada, 4905 Dufferin Street, Downsview, ON M3H 5T4, Canada.  
E-mail: alexei.korolev@ec.gc.ca

Ratio analysis of binary image teams (RABIT) was used in this study to classify images into four categories: "irregulars," "spheres," "needles," and "dendrites." RABIT provides a number of benefits over previously implemented classification schemes: accurate and fast image recognition, processing of partial images, and easy implementation. The limitation of this method is related to that the results of habit recognition cannot be associated with features of individual images, like size, area, mass, etc., since the habit classification is accomplished for a population of images.

## 2. Instrumentation

The cloud particle images used in this study were measured by a PMS OAP-2DC. The principle of particle imaging in OAPs (Knollenberg 1981) is based on measuring a particle's shadow on a linear array of photodetectors. A He-Ne laser is projected onto the photodetectors. As the particle traverses the sample area, it moves between the laser and the photodetector and its shadow obscures a selection of the photodetectors. A photodetector element is considered obscured if the decrease in the intensity caused by the particle is greater than 50%. The detector array produces image slices at a frequency proportional to the true air speed. This generates images with the same resolution along the flight direction ( $x$  axis) and the direction of orientation of the photodiode array ( $y$  axis). The probe used in this study has a photodiode array with 32 elements, and the pixel resolution was adjusted to  $\delta = 25 \mu\text{m}$ .

## 3. Number of habit categories for particle classification

The infinite variety of shapes of natural snow particles makes their classification a difficult task. Different researchers have suggested their own classification of cloud ice particles. Nakaya (1954) used 41 categories for classification of ice particles; Zamorsky (1955) suggested 72 categories to classify cloud particles; Magono and Lee (1966) classified snow crystals into 80 categories. The low pixel resolution ( $25 \mu\text{m}$ ), the small view field (32 pixels in width), and the low gray-level resolution (black and white) makes the habit recognition of OAP-2DC imagery significantly limited, and it definitely cannot provide classification of 40–80 habit categories.

The number of categories in an automated processing of 2D images of cloud particles should be a compromise between the potential of the probe to identify the images (i.e., resolution, number of gray levels, occurrence of out-of-focus images, etc.) and the problem under solution. For example, for the calculation of the radiative properties of clouds, the role of pyramids and cups [C1a and C1b in the Magono-Lee (1966) classification] is negligible, since these type of ice particles rarely occur in the atmosphere. On the other hand, the effect of col-

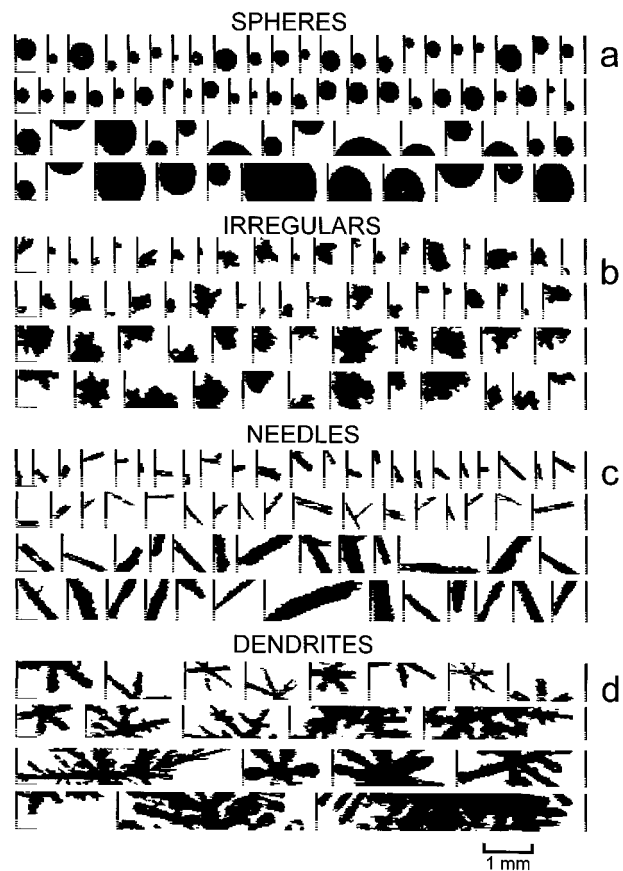


FIG. 1. Classification of PMS OAP-2DC cloud particle images: (a) spheres, (b) irregulars, (c) needles and columns, and (d) dendrites.

umns and bullet rosettes on the radiative balance of the earth may be significant because of their high frequency of occurrence in cirrus clouds.

Different authors used different numbers and types of habit categories in their automated classifiers of OAP images. Rahman et al. (1981) recognized seven classes of particle habits: drop, graupel, stellar crystal, spatial dendrite, column, capped column, and plate. Duroure (1982) classified particles in eight habit categories: drop, graupel, stellar crystal, spatial dendrite and aggregate, plate, and needle and column. Hunter et al. (1984) used six classes to categorize OAP-2D images: dendrite, needle, column or bullet, plate or sphere, streaker, and miscellaneous. Holroyd (1987) and Garbrick et al. (1995) categorized particle habits in eight classes: sphere, hexagonal, graupel, linear, oriented, dendrite, aggregate, and irregular. Moss and Johnson (1994) distinguished nine categories for habit recognition: droplet, large droplet, small droplet, small ice, graupel, stellar, symmetric, column, and aggregate.

The present algorithm classifies 2D images into four habit categories: spheres, irregulars, needles, and dendrites (Fig. 1). The number of habit categories were deliberately reduced compared to other classifiers in order to achieve two important objectives. First, all habit

categories included in the analysis should be equally well recognized throughout the entire considered size range. Second, misclassification of images due to large number of habit categories would be avoided. Since binary imagery do not describe the third particle dimension, the classification of particle habit from 2D images contains certain ambiguity. The suggested four habit categories derived from OAP-2D imagery should be interpreted as follows.

- 1) The sphere category (Fig. 1a) includes particles that generate circular images. Such particles could be liquid drops or quasi-spherical ice particles, like graupel or frozen drops. In some cases flat solid particles, like hexagonal plates, may also give quasi-circular images. However, as will be discussed in section 5a, the effect of plates on the frequency of occurrence of spherical particles is small.
- 2) The needle category (Fig. 1c) includes images having elongated quasi-rectangular projections with an aspect ratio  $c/a > 3$ . Here,  $c$  is the length along the  $c$  axis, and  $a$  is the length along the  $a$  axis. In the majority of cases such projections are produced by needles or columns. Rosettes with three to five bullets also fall into the category needles. Planar crystals, like plates or dendrites, may also give a needlelike projection when viewed edge on, although the effect of such particles on the statistics of habit occurrence in this study is negligible (see section 5b).
- 3) The dendrite category (Fig. 1d) includes dendrites, stellar crystals, aggregates of dendrites, and spatial dendrites.
- 4) The irregular category (Fig. 1b) includes particles having an irregular or random shape that does not fall into any of the above three categories.

#### 4. Algorithm and data processing

##### a. Basics of the algorithm

The present algorithm classifies the particle habits based on an analysis of dimensionless ratios of simple geometrical measures such as the  $x$  and  $y$  dimensions, perimeter, and image area. The key to the suggested method lies in considering these ratios for an ensemble of images. It is assumed that each habit category has a unique distribution (or, basis distribution) for a given ratio. For an ensemble of images constructed from a mixture of different habits, the distribution of a particular ratio will be a linear superposition of basis distributions of individual habits. If the basis distributions are known, then the fraction of each habit for an ensemble of 2D images can be found by solving the inverse problem.

In the present analysis, the following measurable parameters of 2D images were used (Fig. 2):  $N_x$ , the number of pixels in the  $x$  direction (perpendicular to the photodiode array, i.e., along the flight direction);  $N_y$ , the

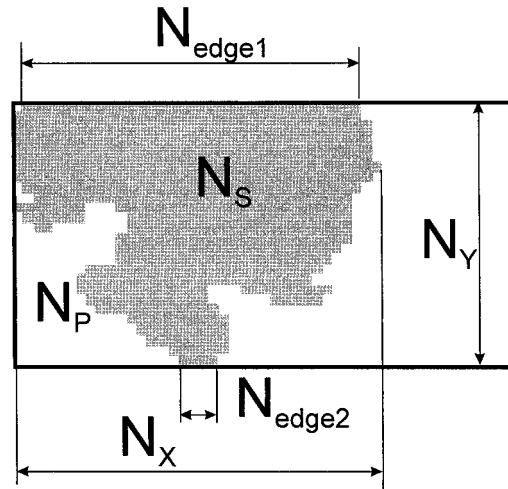


FIG. 2. Simple measurable parameters used for the image habit recognition.

number of pixels in the  $y$  direction (parallel to the photodiode array);  $N_{\text{total}}$ , the total number of shadowed pixels;  $N_{\text{edge}}$ , the total number of triggered edge pixels;  $N_{\text{edge1}}$ ,  $N_{\text{edge2}}$ , the number of triggered edge pixels on each side of the particle image; and  $N_p$ , the number of perimeter pixels as defined in section 4c(3). The measurable parameters can be converted into their corresponding geometrical parameters as follows:  $D_x = \delta N_x$  the  $x$  length,  $D_y = \delta N_y$  the  $y$  length,  $S = \delta^2 N_{\text{total}}$  the particle shadow image area, and  $P = \delta N_p$  the particle image perimeter. Here  $\delta$  is the pixel resolution of the OAP-2DC. It is also assumed that the resolution in  $x$  and  $y$  directions are the same. If this assumption is not valid, an appropriate correction should be introduced in the above equations.

##### b. Image groups

The classification begins with a break down of 2D images into four groups: “tiny,” “rejected partials,” “complete,” and “accepted partials.”

- 1) The tiny images are defined as images that trigger less than  $N_{\text{total}}^*$  pixels. In the frame of this work the threshold value was set  $N_{\text{total}}^* = 20$  pixels. The images with number of pixels less than  $N_{\text{total}}^*$  provide a limited habit resolution and they were removed from analysis. Examples of tiny images are shown in Fig. 3a. Recalling that OAP-2DC does not record the first image slice, the total particle image would contain about 23–25 pixels. Such images correspond somewhat to five pixels or  $125 \mu\text{m}$  in diameter for particles having a quasi-spherical shape. If a particle image has a noncircular shape, the linear size of the image for an equivalent area particle will be larger than that for a circular one. Thus, the particle size  $D = 125 \mu\text{m}$  can be considered as a threshold size for the image recognition analysis in the present

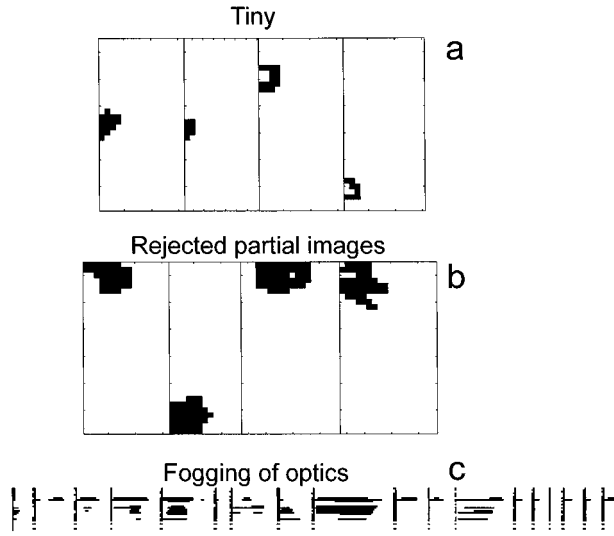


FIG. 3. Examples of rejected OAP-2DC particle images: (a) tiny, (b) rejected partial, and (c) degraded images due to optics fogging.

- study. In general case the threshold  $N_{\text{total}}^* = 20$  pixels can be replaced by any larger value. However,  $N_{\text{total}}^* = 20$  pixels should be considered as the lower limit used for habit recognition (section 5a).
- 2) The group of rejected partials consists of particles where the largest part of image is assumed to be out of the image frame. Such images were defined as  $N_{\text{total}} \leq 180$ , and  $N_{\text{edge}} \geq 4$  and  $N_y < 32$ . Preliminary analysis showed that the recognition of such images gives poor results. Figure 3b shows an example of rejected partials. Images containing empty slices inside the shadowed zone along the  $y$  axes (flight direction) were also rejected (Fig. 3c). Such images usually occur in the case of fogging and in some special cases probe malfunctions. There are a number of other known OAP-2D artifacts caused by electromagnetic noise, shedding, etc. Eliminating these artifacts is a matter of introducing special image filters, which are not discussed in this study.
  - 3) The third group, complete images, includes the images with the number of triggered pixels  $N_{\text{total}} > N_{\text{total}}^*$  with the total number of edge pixels  $N_{\text{edge}} < 4$ . Examples of complete images are shown in each first two strips in Figs. 1a–c.
  - 4) The accepted partial images were defined as 2D images with  $N_{\text{total}} > 180$  pixels and  $N_{\text{edge}} \geq 4$  or  $N_y = 32$ , which is maximum number of pixels in the  $y$  direction. The criterion  $N_{\text{total}} > 180$  was found empirically, and can be changed either side. However, the partial images with lower  $N_{\text{total}}$  yield less habit recognition. Examples of accepted partial images are shown in Fig. 1d and in each bottom two strips in Figs. 1a–c.

The main ensemble has now been sorted into four groups. Of these four, a retrieval will be performed on

the complete and accepted partial groups. Further for brevity accepted partial images will be referred to as partial images.

### c. Basis distributions

The next step of the retrieval is the calculation of the distribution of geometrical ratios for the complete and partial images. Fifteen candidates to ratios to be used in the recognition were considered for both complete and partial images:  $D_x D_y / S$ ;  $(D_x^2 + D_y^2) / S$ ;  $[\max(D_x^2, D_y^2) / S]$ ;  $D_x D_y / P^2$ ;  $(D_x^2 + D_y^2) / P^2$ ;  $[\max(D_x, D_y)] / P$ ;  $D_x / D_y$ ;  $[\min(D_x, D_y)] / [\max(D_x, D_y)]$ ;  $P_{\text{circ}} / P$ ;  $S_{\text{circ}} / S$ ;  $S_{\text{circ}} / P^2$ ;  $P_{\text{circ}}^2 / S$ ;  $P^2 / S$ ;  $[P(D_x + D_y)] / S$ ; and  $(P_{\text{circ}} \sqrt{S_{\text{circ}}}) / P \sqrt{S}$ . Here,  $S_{\text{circ}}$  and  $P_{\text{circ}}$  are expected image area and perimeter, respectively, in assumption that the images are circles [section 4c(2)]. Basis distributions and retrieval matrices (as described in section 4d) were constructed for each of these ratios and tested on special samples of images of particle measured in natural clouds. The ratios that provide the best habit recognition were selected for the analysis. The rest of the ratios were not used.

#### 1) COMPLETE IMAGES

For complete images, the ratio

$$R_1 = \frac{\pi(N_x^2 + N_y^2)}{8N_{\text{total}}} = \frac{\pi(D_x^2 + D_y^2)}{8S} \quad (1)$$

was found to have the most unique features to characterize each category of habits. Perfect spheres give circular images,  $R_1 = 1$ . Due to imperfections of images caused by discretization (Korolev et al. 1998) or the probe electronics, the value of  $R_1$  for real spherical particles may be different from 1. Therefore, the images from an ensemble of spherical particles return a narrow frequency distribution of  $R_1$  centered around  $R_1 = 1$ . Needles and columns are implicitly defined as rectangles having a length-to-width ratio ( $c/a$ ) of between 3 to 8. Therefore, for needles the ratio  $R_1 \approx \pi c/8a$  ranges from 1 to 3. Due to their unique shape, dendrites and irregular particles will also return a unique distribution for the ratio  $R_1$ .

Figure 4a shows three basis distributions of the ratio  $R_1$  for complete spheres, irregulars, and needles for a training set of images like in Fig. 1 measured by OAP-2DC and classified by eyeball. Notice that although all distributions overlap, each particle category leaves a unique distribution signature. Due to rare occurrence of complete dendrites the basis distribution for dendrites is absent (see section 5c).

#### 2) PARTIAL IMAGES

The partial images were classified in a different way. The classification begins by assuming that the particle image is a perfect circle. Knowing the number of trig-

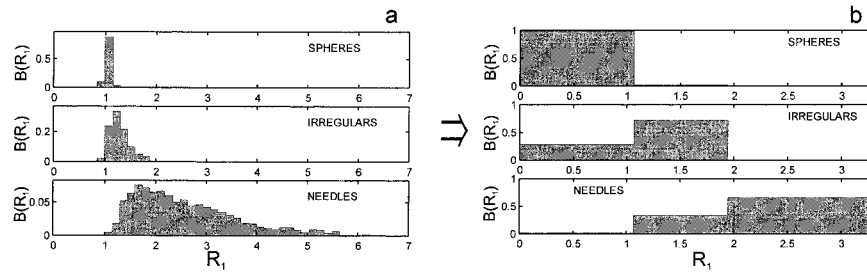


FIG. 4. (a) Basic distributions of the dimensionless ratio  $R_1 = [\pi(D_x^2 + D_y^2)]/8S$  for training set of complete spheres, irregulars, and needles measured by OAD-2DC in natural clouds. (b) The same distributions as in (a) after discretizing. The distribution in (b) is not normalized on the width of the intervals  $[R_j, R_{j+1})$ .

gered edge pixels on the image  $N_{\text{edge1}}$  and  $N_{\text{edge2}}$ , the perimeter of the circle can be found as

$$P_{\text{circ}} = 2(\theta_2 - \theta_1)r. \tag{2}$$

Here,

$$r = \sqrt{\frac{x_1^2}{4} + \left(\frac{y}{2} + \frac{x_2^2 - x_1^2}{8y}\right)^2}, \quad \theta_1 = \pi - 2 \arctan\left(\frac{x_1}{2z}\right),$$

$$\theta_2 = \pi - 2 \arctan\left[\frac{x_2}{2(z+y)}\right], \quad z = \frac{x_1^2 - x_2^2}{8y} - \frac{y}{2} + r,$$

$x_1 = \delta N_{\text{edge1}}$ ,  $x_2 = \delta N_{\text{edge2}}$ , and  $y = \delta N_y$ . Figure 5 provides explanation for calculation of the perimeter of a partially imaged circle. Note that Eq. (2) is true for all possible positions of a circle in the viewfield of the probe, that is, when  $x_1 \neq 0$  and  $x_2 \neq 0$  (case in Fig. 5);  $x_1 = 0$  and  $x_2 \neq 0$ ; or  $x_1 \neq 0$  and  $x_2 = 0$ ; or  $x_1 = 0$  and  $x_2 = 0$  (complete circle).

The assumed circle image perimeter  $P_{\text{circ}}$  is then compared with the measured one  $P$ , that is,

$$R_2 = \frac{P}{P_{\text{circ}}}. \tag{3}$$

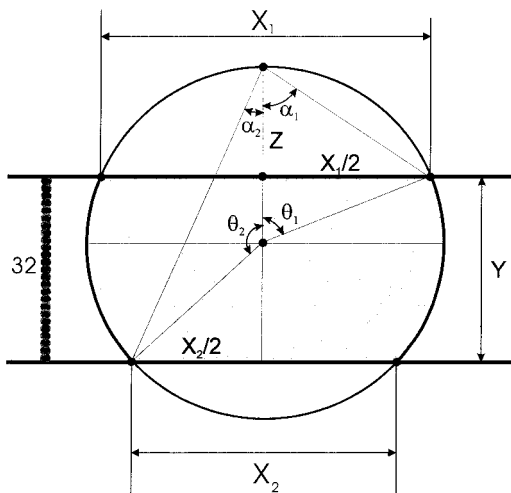


FIG. 5. Calculation of the perimeter of a partially imaged circle.

Figure 6a shows four basis distributions of the ratio  $R_2$  for partial spheres, irregulars, needles, and dendrites for a training set of images measured by OAP-2DC in natural clouds.

### 3) PERIMETER CALCULATION

The calculation of perimeter of discrete image is not that straightforward as that of  $N_{\text{total}}$ ,  $N_x$ , and  $N_y$ . The perimeter of a discrete image may be defined in several different ways (Duda and Hart 1973), therefore resulting in an ambiguity in its calculation. The way how the perimeter is calculated effects the ratio  $R_2$  [Eq. (3)] and hence basis distribution and retrieval matrices.

In this study for perimeter calculation a specially developed algorithm was used. The calculation of the perimeter may be considered as a five-step process schematically shown in Fig. 7. During the first step the whole image is shifted one pixel up, and then the resulting image is summed with the original image A. The result of such summing will be an image consisting of ones and twos. The algorithm automatically converts twos into zeroes, that gives a binary image B (Fig. 7). The second step is similar to the first one, but the image A is shifted one pixel left and then summed with the original image A. This results in the image C. The third step is summing the images B and C giving the image D. The fourth step generates the image E by shifting the image B one pixel right and summing with C. The final step consists in getting the image F by summing the images E and D and converting ones into zeroes. This procedure results in the image F where the pixels in horizontal and vertical legs are represented by twos, and the diagonal pixels consist of threes (Fig. 7). Therefore, the image perimeter can be estimated as

$$P = \delta N_p = \delta(N_2 + \sqrt{2}N_3). \tag{4}$$

Here  $N_2$  and  $N_3$  are number of twos and threes, respectively, in the image F. Note that in this definition of the perimeter [Eq. (4)] number of perimeter pixels  $N_p$  is not necessarily an integer number. For partial images the edge pixels corresponding to  $N_{\text{edge1}}$ ,  $N_{\text{edge2}}$  are not included in the perimeter.

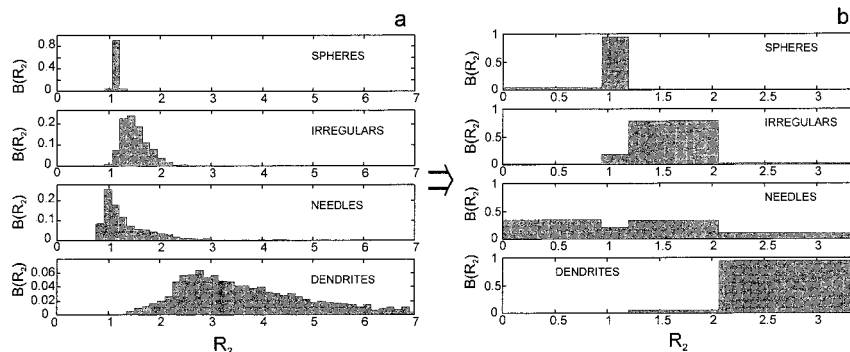


FIG. 6. (a) Basic distributions of the dimensionless ratio  $R_2 = P/P_{\text{circ}}$  for training set of partial spheres, irregulars, needles, and dendrites measured by OAD-2DC in natural clouds. (b) The same distributions as in (a) after discretizing. The distributions in (b) are not normalized on the width of the intervals  $[R_j, R_{j+1})$ .

The essence of the above algorithm and Eq. (4) is that it approximates the binary image by horizontal ( $0^\circ$ ), vertical ( $90^\circ$ ) and diagonal ( $45^\circ$ ) legs. As a result it gives reasonably accurate estimation of a perimeter of a binary image. For the image shown on Fig. 7 the number of perimeter pixels  $N_p$  estimated from the above algorithm is 37.4. The perimeter of the actual image drawn through the middle part of perimeter pixels would be 38.6 pixels. The image perimeter estimated from direct counting of perimeter pixels gives the most inaccurate result of 28 pixels.

Another advantage of this algorithm is that it uses fast computer operations, like shifting and summing, which take much less CPU time compared to checking “edge” conditions for each pixel of the image.

*d. Construction of a retrieval matrix*

Let  $M(R)$  be the measured distribution of a specific ratio  $R$  for an ensemble of measured images such that  $M(R)dR$  is the fraction of particles whose ratio lies between  $R$  and  $R + dR$ . The distribution  $M(R)$  can be represented as a linear superposition of basis distributions  $B_i(R)$ , as

$$M(R) = \sum_{i=1} f_i B_i(R), \tag{5}$$

where  $f_i$  is the fraction of particles in the  $i$ th habit subject to the normalization constraint  $\sum_{i=1} f_i = 1$ . Note that  $M(R)$  and  $B_i(R)$  are normalized to unity. The measured distribution  $M(R)$  may differ from the actual distribution due to the Poisson statistics of images in a sample. The shapes of particles in some cloud zones may also not perfectly match those used in the training sample. Therefore, the local basis distributions may be different from that used in the algorithm  $B_i(R)$ . Due these statistical uncertainties the solution of Eq. (5) may become ill posed for the cases when the distributions  $M(R)$  and  $B_i(R)$  are continuous or when they are broken into a number of intervals larger than the number habit categories. One way to obtain a unique solution to Eq. (5) for  $f_i$  is to discretize the domain of the distributions into a number of intervals  $[R_j, R_{j+1})$  equal to the number of image categories, thus generating the matrix  $\mathbf{B}$  with elements

$$B_{ij} = \int_{R_j}^{R_{j+1}} B_i(R) dR, \tag{6}$$

and the vector  $\mathbf{M}$  with components

$$M_j = \int_{R_j}^{R_{j+1}} M(R) dR. \tag{7}$$

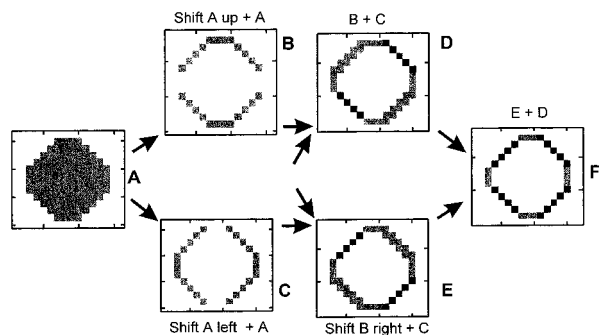


FIG. 7. Calculation of a perimeter of a discrete binary image. Gray pixels indicate ones, black pixels show twos in the images A, B, C, D, E. Gray pixels show twos, black pixels show threes in the image F.

The boundaries of the intervals  $[R_j, R_{j+1})$  were chosen by performing an optimization of  $\mathbf{B}$  that minimized its condition number. The condition number is the ratio of the largest singular value of  $\mathbf{B}$  to the smallest (Dongarra et al. 1979). The conditioning number measures the sensitivity of the solution of a system of linear equations to errors in the data. Large condition numbers indicate a nearly singular matrix. Figures 4b and 6b shows the result of discretizing of the distributions  $B_i(R)$  for complete and partial images. The numerical values of the elements of the matrices  $\mathbf{B}$  and intervals  $[R_j, R_{j+1})$

found as a result of the above optimization are shown in appendix A.

After the matrix  $\mathbf{B}$  is constructed the solution of the discrete form of Eq. (5) can be found as

$$\mathbf{f} = \mathbf{B}^{-1}\mathbf{M}. \quad (8)$$

Here,  $\mathbf{B}^{-1}$  is the inverse of  $\mathbf{B}$  and  $\mathbf{f}$  is a  $4 \times 1$  vector of fractions of images belonging to different habit categories. For example, if  $\mathbf{f} = (0.5, 0.2, 0.05, 0.25)$ , then the percentage composition of the sample would be 50% for spheres, 20% for irregulars, 5% for needles, and 25% for dendrites.

The final step would be to sum the vectors  $\mathbf{f}_c$  and  $\mathbf{f}_p$  for complete and partial images, respectively. This is done by weighting each vector with respect to the number of complete ( $N_c$ ) and partial ( $N_p$ ) particles it represents:

$$\mathbf{f} = \frac{N_c \mathbf{f}_c + N_p \mathbf{f}_p}{N_c + N_p}. \quad (9)$$

In order for this procedure to be effective, the size of the 2D image sample should be statistically significant. If only a small number of images (less than 100) are used to calculate  $\mathbf{M}$ , then the errors, related to the Poisson statistics in  $\mathbf{M}$ , may be 20%, if the fraction of habits are evenly distributed between four categories. The subsequent retrieval procedure may only increase these errors in the final result  $\mathbf{f}$ . A sample size of five buffers ( $32 \times 1024$  pixels) usually represents 200 to 400 particles acceptable for image analysis. Therefore, five buffers is large enough to accurately describe the fractional composition of particle shapes, while at the same time being sufficiently small so as to suitably display the spatial variations of particle habits in a cloud.

#### e. Some results of image processing

The algorithm was tested on a large dataset collected during four aircraft projects in different weather conditions and climatic zones: the Beaufort Arctic Storm Experiment (Canadian Arctic, September–October 1994); the Canadian Freezing Drizzle Experiment I (Newfoundland, March 1995); the Canadian Freezing Drizzle Experiment 3 (Great Lakes, December 1997–February 1998); and FIRE.ACE (Canadian and U.S. Arctic, April 1998). The results of recognition analysis and habit statistics were summarized in a study by Korolev et al. (2000, hereafter KIH).

Figure 8 shows an example of habit recognition from flight on 23 January 1998 during the Canadian Freezing Drizzle Experiment 3. Figures 8b–e show the fraction of irregulars, dendrites, needles, and spheres, respectively, derived from OAP-2DC data with help of the developed algorithm. It is seen that dendrites and needles occur in localized zones or cells, where the fraction of needles and dendrites may reach up to 90%–100%. These cells are usually surrounded by regions with ir-

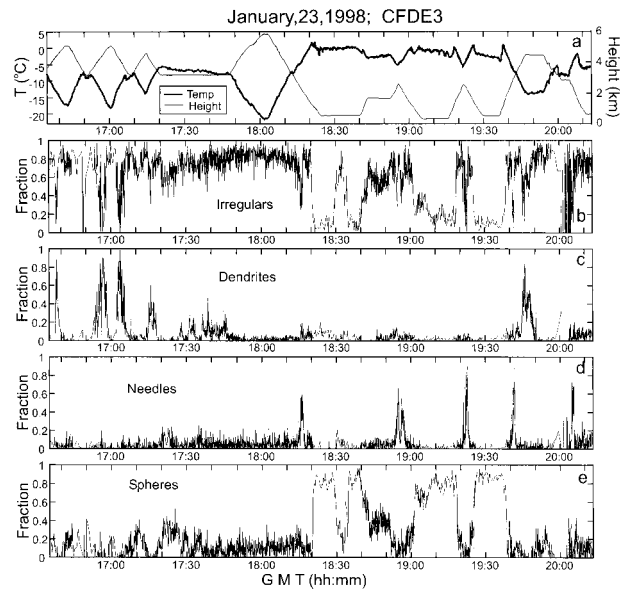


FIG. 8. Changes of cloud particle habits during flight on 23 Jan 1998 (CFDE III): (a) changes of temperature and altitude. Spatial changes of the fraction of (b) irregulars, (c) dendrites, (d) needles, and (e) spheres.

regular shaped particles. Figure 9 shows the OAP-2DC images measured in zones with high percentage of (a) spheres, (b) irregulars, (c) dendrites, and (d) needles. A comparison of results of habit recognition by the present algorithm with those obtained by a human classifier shows that on average the error does not exceed 10%–15%.

Quantifying accuracy in any classification scheme for binary images of hydrometeors is rather subjective. Ice particles often come in hybrid forms whose shapes fall in between possible habits. Low resolution and large variety of shapes of cloud particles contribute heavily to uncertainty in classification. Even human classifiers may not agree on the classification of a specific particle. Ice particles are not some absolute object that necessarily fall into a preconceived category. Therefore, there will always be some degree of uncertainty in any classification scheme.

## 5. Discussion

### a. Threshold number of pixels for habit analysis

Different habit categories require some minimum number of pixels to be properly classified. Some authors claim to be able to recognize hexagonal plates starting from images having three to five pixels in size (Duroure 1982; Holroyd 1987; Garbrick et al. 1995). However, one may only be able to resolve a hexagonal structure with images having at least 12 to 14 pixels in “diameter.” Plates with smaller sizes would fall into another category other than “plate.” Figure 10 shows two discrete binary images generated from two hexagonal

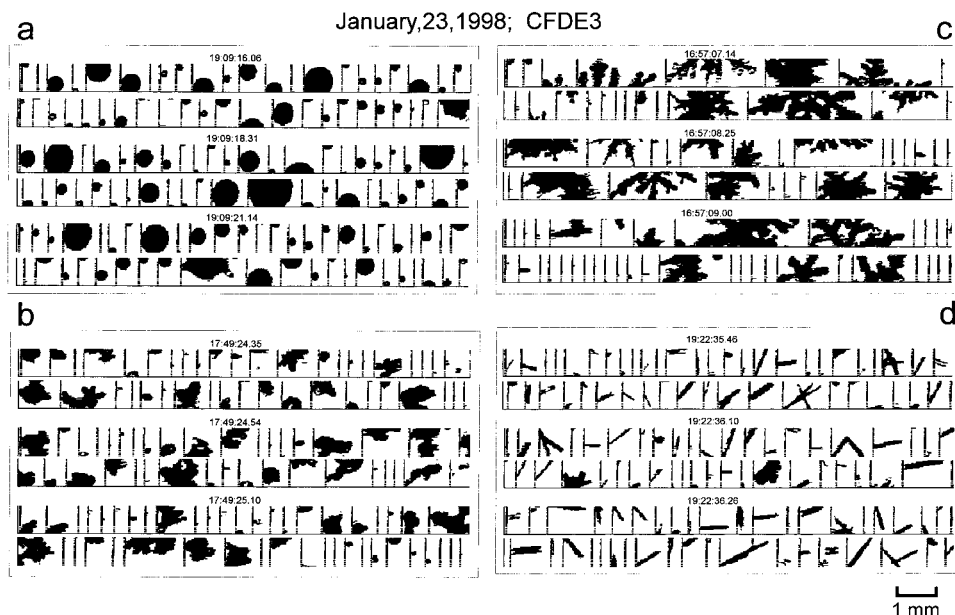


FIG. 9. Samples of OAP-2DC images measured in zones of (a) spheres, (b) irregulars, (c) dendrites, and (d) needles during the CFDE III flight on 23 Jan 1998 (Fig. 7). The measurement times are indicated at the top of each 2D buffer ( $32 \times 1024$  image slices).

plates having the same orientation but different sizes: 8 pixels (Fig. 10a) and 14 pixels (Fig. 10b) in diameter. It is clearly seen that the discrete image for the smaller plate (Fig. 10a) does not have visual features of hexagonal plate and it will be classified as irregular or sphere. Therefore, the fraction of plates calculated for an ensemble of particles would be a priori biased. For this reason the category of plates was not included into our analysis. Observations of ice particles at the ground level conducted by Zamorsky (1955) showed that the frequency of occurrence of plates is 2.8%. KIH estimated the fraction of plates in polar and midlatitude clouds as 0.4%. This is consistent with a simple visual analysis of OAP-2DC data: on average plates occur rarely in OAP-2DC imagery. Therefore the exclusion of hexagonal plate category from the recognition analysis presumably would not affect the particle habit statistics.

For classification of needles, depending on the  $a/c$

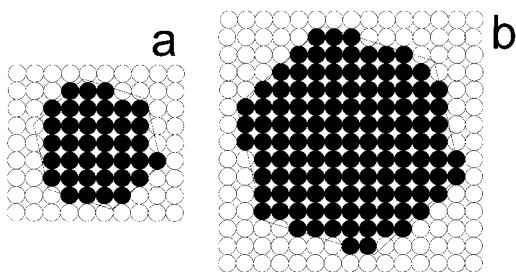


FIG. 10. Discrete binary images resulted from two hexagonal plates having the same orientation but different size: (a) 8 pixels and (b) 14 pixels in diameter. A pixel in the binary images was considered occulted if 50% or more of its area was shadowed.

ratio, an image having three to four pixels all in a line would be enough for proper classification. Therefore, the images of needles with 20 pixels in total would more than adequately describe the needle/column habit. An image having 8–10 pixels in linear size and about 20–25 pixels in total may adequately represent the stellar habit. The operational definition of a spheres is not trivial. On one hand, it depends on uncertainty related to discretization of the image (Korolev et al. 1998). On the other hand, it depends on how accurate we want to describe a deviation of the image from circular in a recognition analysis. The irregularities on the surface of the sphere should be about pixel size to generate a deviation of the image from circular, and thus generate a reason to put it in the “irregular” category. In this study we deliberately decreased the threshold  $N_{\text{total}}^*$  down to 20 pixels in order to keep the number of accepted images large. KIH analyzed about  $3 \times 10^7$  particle images measured by OAP-2DC. Approximately 73% of the images were rejected as tiny with  $N_{\text{total}} < 20$  pixels. A decrease of  $N_{\text{total}}^*$  may dramatically decrease the number of accepted particles that may make the recognition analysis meaningless. Using the threshold  $N_{\text{total}}^* = 20$  allows RABIT to distinguish needles and dendrites, though it leaves some uncertainties in the separation of irregulars and spheres for small images ( $20 < N_{\text{total}} < 60$ ). From a retrieval matrix **B** for complete images (appendix A) the value of this uncertainty may be estimated as 20%–25%. For larger images this uncertainty decreases to few percent or less. So, the threshold  $N_{\text{total}}^* = 20$  gives reasonably accurate results for recognition analysis of the above four habit categories.

*b. The effect of particle orientation on habit statistics*

Image orientation can be a severe impediment to image classification. The illuminating laser beam of the OAP-2DC installed on the NRC Convair-580 was purposely oriented vertically to combat orientational effects. Since an ice particle with the Reynolds number  $Re > 10$  (List and Schemenauer 1971) falls with its largest cross section plane oriented horizontally, it is expected that the OAP-2DC measures the largest projection of the particle, that is, when the particle image exhibits its best orientation for habit classification. From laboratory studies (List and Schemenauer 1971; Hallett et al. 2000), columns fall with their  $c$  axis being horizontal, while for plates and dendrites the  $c$  axis is vertical. As mentioned above, the planar ice particles, like plates and dendrites, may give a rectangular projection, which may be recognized as a needle or a column. In the case of aircraft measurements this may occur if the ice particle changes its natural orientation due to wind shear caused by the aircraft and the local flow around the instrument. If this were the case, we would observe on a regular basis in zones of dendrites a constant fraction of images classified as needles due to reorientation of dendrites. Similarly in the zones of needles, a certain fraction of circular would be observed all the time. However, such a coexistence of shapes is not observed (e.g., see Figs. 8c,d), and the effect of reorientation of planar and columnar ice particles on the habit statistics is estimated as no larger than a few percent.

*c. Observation of “complete” dendrites*

The growth of plates along the  $a$  axis is changed to dendritic growth when the terminal fall velocity exceeds a certain threshold ( $Re > 0.5$ ) (Alena et al. 1990). Since the terminal fall velocity is related to a plate size, there is a threshold plate size for dendritic growth. For typical atmosphere conditions it is about  $200 \mu\text{m}$ . A particle starts to exhibit stellar or dendritic features only when it reaches some  $400\text{--}600 \mu\text{m}$  in size. This explains why in OAP imagery the frequency of occurrence of complete dendrites is significantly less than that of partial dendrites. For this reason we were unable to obtain statistically significant training set of complete images of dendrites, and, thus they were not included in the retrieval matrix **B**.

*d. Dependence of basis distribution on particle size*

The dependence of particle habits on size is discussed in detail in the recent study by KIH. In this work the particle habit were classified for three size groups:  $D > 125$ ,  $D > 250$ , and  $D > 500 \mu\text{m}$ . This classification was done using three different retrieval matrixes derived for each size range. The comparison of basis distributions showed that the differences between habit bins do not exceed 5%. This allows one to make a conclusion about the generality of the retrieval matrix for the particle size range  $D > 125 \mu\text{m}$  (or  $N_{\text{total}} > 20$  pixels).

## 6. Conclusions

The following results were obtained in the frame of this work.

- A new cloud particle habit classifier (RABIT) was developed. The algorithm derives the fraction of particles having a certain habit from a distribution of dimensionless ratios extracted from a population of images.
- The RABIT provides classification of images into four habit categories: spheres, irregulars, needles, and dendrites.
- The threshold number of pixels  $N_{\text{total}}^*$  in the image for the classification of particle habits into four categories was estimated as about 20. An increase of  $N_{\text{total}}^*$  provides better habit resolution.
- The algorithm provides habit classification of both complete and partial images. This has certain advantages over previously implemented classification schemes. The processing of partial images helps to improve the statistics of the recognition analysis: the accepted partial images are approximately 35% of all accepted images (or 50% of complete images; KIH).

While the statistical approach of RABIT provides certain benefits over previous algorithms, it also has certain limitations. Because groups of images and not individual images are classified, details specific to one image cannot be extracted. For example, size distributions for particular habit categories cannot be obtained. Nonetheless, RABIT provides a fast and accurate classification algorithm that is easy to implement due to its mathematical simplicity.

In spite of the roughness of classification scheme, the suggested four habit categories cover the major habits of particles in troposphere natural clouds: spherical, columnar, planar, and irregular. The use of RABIT for classification of cloud particles may be valuable for satellite and remote sensing retrievals, parameterization of GCM models, and the understanding ice formation in the atmosphere.

*Acknowledgments.* The authors appreciate discussion and useful comments of George Isaac of Atmospheric Environment Service (Canada) and John Hallett of Desert Research Institute (USA). Special thanks to Sergei Kuznetsov of Central Alerological Observatory (Russia). The authors are indebted to two anonymous reviewers for their valuable suggestions and their useful criticism of the original manuscript.

## APPENDIX

### Retrieval Matrices for Complete and Partial Images

The ratio  $R_i$  was split into the three following intervals:  $[0, 1.07)$ ,  $[1.07, 1.94)$ , and  $[1.94, \infty)$ . These intervals were found using the optimization procedure de-

scribed in section 4d and they correspond to peaks in  $B_i(R)$  distributions for spheres, irregulars, and needles, respectively. The vectors corresponding to these intervals are

(0.9803 0.0197 0) for spheres,  
 (0.2745 0.7235 0.0020) for irregulars, and  
 (0.0074 0.3333 0.6593) for needles.

Since dendrites and aggregates almost never appear as complete images at 25- $\mu\text{m}$  resolution and 32-element photodiode array in OAP-2DC, this category of images was excluded from the analysis of complete images (section 5c). The few dendrites and aggregates that may appear as complete images will not affect the result of recognition much.

For partial images the intervals for the ratio  $R_2$  were chosen as [0, 0.94), [0.94, 1.16), [1.16, 2.06); and [2.06,  $\infty$ ).

The vectors corresponding to these intervals are

(0.0467 0.9467 0.0067 0) for spheres,  
 (0.0117 0.1780 0.7869 0.0234) for irregulars,  
 (0.3493 0.2059 0.3346 0.1103) for needles,  
 (0 0 0.0591 0.9409)  
 for dendrites/aggregates.

The above vectors compose two square matrices:  $3 \times 3$  and  $4 \times 4$ . As it is seen these matrices are well defined and they can form inverse matrices.

Note that these results are not necessarily transferable to other PMS OAP-2D probes. The natural orientation of dendrites and needles may be changed by the wind shear around the plane fuselage in the vicinity of OAP-2D, thus changing the appearance of measured images. The OAP-2D imagery of dendrites and needles may be different if the orientation of the laser beam is not vertical. If the laser beam is directed horizontally, the above matrices are not applicable at all. In general, new retrieval matrices should be created for different probe configurations.

## REFERENCES

- Alena, T., J. Hallett, and C. P. R. Saunders, 1990: On the facet-skeletal of snow crystals: Experiments in high and low gravity. *J. Cryst. Growth*, **104**, 539–555.
- Dongarra, J. J., J. R. Bunch, C. B. Moler, and G. W. Stewart, 1979: *LINPACK: Users' Guide*. Society for Industrial and Applied Mathematics, 320 pp.
- Duda, R. O., and P. E. Hart, 1973: *Pattern Classification and Scene Analysis*. Wiley, 487 pp.
- Durore, C., 1982: Une nouvelle methode de traitement des images d'hydrometeores donnees par les sondes bidimensionnelles. *J. Rech. Atmos.*, **6**, 71–84.
- , H. R. Larsen, H. Isaka, and P. Personne, 1994: 2D image population analysis. *Atmos. Res.*, **34**, 195–205.
- Garbrick, D., V. Chandrasekar, and R. Xiao, 1995: Neural network based classification procedure for 2D-PMS ice crystal images. *Conf. on Cloud Physics*, Dallas, TX, Amer. Meteor. Soc., 59–64.
- Hallett, J., W. P. Arnott, M. P. Bailey, and J. T. Hallett, 2000: Ice crystals in cirrus. *Cirrus*, D. K. Lynch, Ed., Oxford University Press, in press.
- Holroyd, E. W., 1987: Some techniques and uses of 2D-C habit classification for snow particles. *J. Atmos. Oceanic Technol.*, **4**, 498–511.
- Hunter, H. E., R. M. Dyer, and M. Glass, 1984: A two-dimensional hydrometeor machine classifier derived from observed data. *J. Atmos. Oceanic Technol.*, **1**, 28–36.
- Knollenberg, R. G., 1981: Techniques for probing cloud microstructure. *Clouds, Their Formation, Optical Properties, and Effects*, P. V. Hobbs and A. Deepak, Eds., Academic Press, 495 pp.
- Korolev, A. V., J. W. Strapp, and G. A. Isaac, 1998: Evaluation of accuracy of PMS Optical Array Probes. *J. Atmos. Oceanic Technol.*, **15**, 708–720.
- , G. A. Isaac, and J. Hallett, 2000: Ice particle habits in stratiform clouds. *Quart. J. Roy. Meteor. Soc.*, in press.
- List, R., and R. S. Schemenauer, 1971: Free-fall behavior of planar snow crystals, conical graupel and small hail. *J. Atmos. Sci.*, **28**, 110–115.
- Magono, C., and C. Lee, 1966: Meteorological classification of natural snow crystals. *J. Fac. Sci. Hokkaido Univ. Ser. 7*, **2**, 321–335.
- Moss, S. J., and D. W. Johnson, 1994: Aircraft measurements to validate and improve numerical model parameterization of ice to water ratios in clouds. *Atmos. Res.*, **34**, 1–25.
- Nakaya, U., 1954: *Snow Crystals, Natural and Artificial*. Harvard University Press, 510 pp.
- Rahman, M. M., E. A. Quincy, R. G. Jacquot, and M. J. Magee, 1981: Feature extraction and selection for pattern recognition of two-dimensional hydrometeor images. *J. Appl. Meteor.*, **20**, 521–535.
- Tou, J. T., and R. C. Gonzales, 1974: *Pattern Recognition Principles*. Addison-Wesley, 377 pp.
- Zamorsky, A. D., 1955: *Ice in the Atmosphere*. Akademia Nauk SSSR, 277 pp.