A statistical analysis for pattern recognition of small cloud particles sampled with a PMS-2DC probe

A. Fouilloux, J. Iaquinta, C. Duroure, F. Albers

A statistical analysis for pattern recognition of small cloud particles sampled with a PMS-2DC probe

A. Fouilloux[1], J. Iaquinta[2], C. Duroure[1], F. Albers[3]

[1] LAMP, 24 avenue des Landais, F-63177 Aubière Cedex, France
[2] Currently on leave from LAMP
[3] GKSS Research Center, Max-Planck Strasse, D-21502 Geesthacht, Germany

Abstract. Although small particles (size between 25 μm and 200 μm) are frequently observed within ice and water clouds, they are not generally used properly for the calculation of structural, optical and microphysical quantities. Actually neither the exact shape nor the phase (ice or water) of these particles is well defined since the existing pattern recognition algorithms are only efficient for larger particle sizes. The present study describes a statistical analysis concerning small hexagonal columns and spherical particles sampled with a PMS-2DC probe, and the corresponding images are classified according to the occurrence probability of various pixels arrangements. This approach was first applied to synthetic data generated with a numerical model, including the effects of diffraction at a short distance, and then validated against actual data sets obtained from in-cloud flights during the pre-ICE'89 campaign. Our method allows us to differentiate small hexagonal columns from spherical particles, thus making possible the characterization of the three dimensional shape (and consequently evaluation of the volume) of the particles, and finally to compute e.g., the liquid or the ice water content.

1 Introduction

Size spectra obtained from in situ measurements within clouds (Liou, 1992; Kinne and Liou, 1989) show that the number of small particles (size ranging from 25 μm to 200 μm) is very large, and that it is important for the determination of structural, optical and microphysical quantities such as liquid or ice water contents (LWC or IWC). In addition, since small ice crystals may drasti-

cally affect the radiative properties of cirrus clouds (Takano *et al.*, 1992), their shape can not be simply approximated by spheres or cylinders. Actually the scattering properties of ice crystals (with hexagonal structure or even a more complex shape) differ significantly from that of perfect spheres (as given by the Mie theory), and consequently the spherical assumption induces quite large errors for the calculation of bulk quantities such as the liquid water path or optical thickness of the cloud.

The main problem, causing these inaccurate estimations, is that the three dimensional shape of the particles can not be deduced from images obtained with the PMS-2DC (Particle Measuring Systems, Boulder, CO, USA) probe by applying the existing pattern recognition algorithms (Darlison *et al.*, 1988; Duroure, 1982; Duroure *et al.*, 1994). Indeed, the same configuration can correspond either to a spherical particle or a hexagonal column. Our first objective is therefore to classify the images of these hydrometeors according to their occurrence probabilities. Nevertheless, we have to keep in mind that for PMS-2DC measurements the errors are largest with particle sizes smaller than 200 μm and may reach 85% (Korolev *et al.*, 1990). Thus, the estimation of the corresponding shape will intrinsically contain non removable uncertainties.

In the first part of this study, we present a numerical model, including the effects of Fresnel's diffraction, specifically intended to simulate the sampling of hexagonal columns and spheres. In a second step, this scheme was used to create several synthetic data sets considered afterwards in a statistical analysis. The objective was to define a new automatic pattern recognition algorithm in order to differentiate the images of small hexagonal columns from those of spherical particles. Our method was validated against actual data obtained during an in-cloud flight of the pre-ICE'89 (International Cirrus Experiment) campaign (Gayet *et al.*, 1990; Gayet *et al.*, 1993). We propose a simple way to approximately determine the volume of the particles and then the liquid or ice water content.

Correspondence to: A. Fouilloux

2 Statistical analysis of synthetic data sets

2.1 Simulating the sampling of spherical particles and hexagonal columns

2.1.1 Theoretical study of particle shadow and diffraction effects. The formation of a particle shadow is a laborious problem since the diffracted, refracted and absorbed components of the light beam have to be considered. As in Korolev *et al.* (1990), we neglected the refracted component of light and assumed that the particles are completely opaque, therefore the shadow image is simply defined by the diffracted component. In addition, we represented the effect of diffraction by a three-dimensional object, e.g., hexagonal column or spherical particle, by that of its projection on a plane perpendicular to the direction of propagation of the incident wave. Moreover, we considered the Huygens-Fresnel approximation since the characteristic size of the particles (> 20 µm) is much larger than the wavelength of laser beam (i.e., 632 nm for the PMS-2DC probe). We implemented the theory initiated by Maggi-Rubinowicz (Maggi, 1888 ; Rubinowicz, 1938), and further developed by Miyamoto and Wolf (1962), in which the diffraction can be considered as the combined effect of an incident wave with a boundary wave. The amplitude $U(P)$ of the diffracted wave at a point P on the screen, as shown on Fig. 1, is given by:
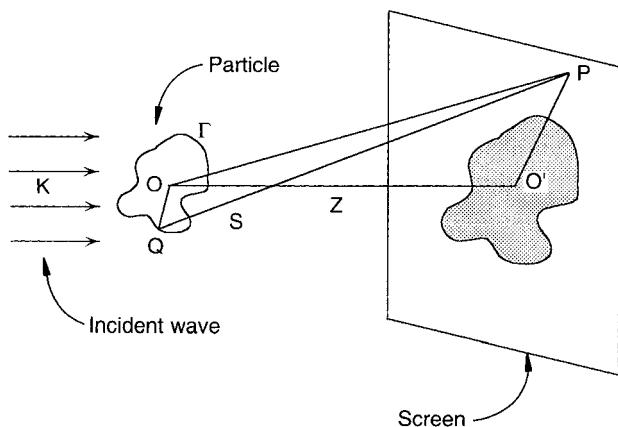
$$U(P) = U^g(P) + U^d(P) \tag{1}$$

where

$$U^g(P) = \begin{cases} e^{iKZ} & \text{when } P \text{ is in the direct beam} \\ 0 & \text{when } P \text{ is in the geometrical shadow} \end{cases} \tag{2}$$

and

$$U^d(P) = \frac{1}{4\pi} \oint_{\Gamma} e^{i\vec{k}\cdot\vec{q}} \frac{e^{iKS}\left(\vec{s} \times \vec{k}\right)}{S\left(1 - \vec{s}\cdot\vec{k}\right)} d\vec{l} \tag{3}$$

Fig. 1. Scheme representing the derivation of the boundary diffraction wave, adapted from Born and Wolf (1991)

In these equations, Z represents the distance between the particle and the screen, $K = 2\pi/\lambda$ is the wave number, λ is the wavelength of the incident collimated light beam, S is the distance between point Q (on the contour Γ) and point P (on the screen). The corresponding unit vectors are: $\vec{k}$ in the direction of the plane wave propagation, $\vec{q}$ (the radius vector of point Q) and $\vec{s}$. The integration is carried out along the contour Γ (for which the differential element of length is $d\vec{l}$), i.e., the boundary of the geometrical shadow. The quantity U^g represents the disturbance as predicted by the geometrical optics, and U^d the effect of diffraction. This approach is true only if Z is larger than $200\,\lambda/\pi$ (Papoulis, 1968), and the intensity $I(P)$ of the diffraction is then deduced from Eqs. (1–3) by:

$$I(P) = |U(P)|^2 \tag{4}$$

In the case of the PMS-2DC probe a 50% threshold level is applied for particle image sizing (Knollenberg, 1970).

The spherical particles are defined by a radius R, that can randomly change over a range of $\pm 10\%$ to consider the noise of actual data, and varies from 25 µm to 150 µm. This representation of the shape deformations is very important, not only for water droplets, but also for aggregates which must be differentiated from hexagonal particles.

The aspect ratio of the hexagonal columns (size < 200 µm), i.e., the ratio of the width w and the length L of the column, was found to be about 0.5 by Heymsfield (1972). We selected intentionally a similar value in order to be in the least advantageous situation (since the particles are very compact their image looks like almost that of a sphere), even if this is not really in agreement with the result of our actual data processing (the retrieved particle aspect ratio is 0.3 on average). Because the size of a photodiode is 25 µm, and to obtain images of less than 10 pixels, the length of the simulated columns takes values in the interval [25 µm; 300 µm]. The location of a particle of maximal dimension D within the sampling volume is chosen randomly so that Z is in the range [D; 60 mm], according to the probe characteristics.

Figure 2 illustrates the influence of the distance between the particle and the screen on the diffraction pattern for a sphere of radius $R = 70$ µm and a hexagonal column of length $L = 190$ µm. When the screen is far from the particle ($Z = 2.0$ m), the diffracted image corresponds roughly to the projected shadow. However, when the distance decreases ($Z = 0.02$ m), a hole appears in the center of the sphere and the column is split in to two parts (C_1 and C_2 on Fig. 2). The point here is that the radius of the outer circle is larger than the radius R of the sphere, and that the length of C_1 and C_2 can be much smaller than L while the distance between these two parallel lines is generally larger than the width of the actual column.

For the two types of particles we have used Marshall-Palmer size distributions that correspond to experimental spectra derived from in situ measurements. This

Sphere of radius *R* = 70 μm

Z = 0.02 m   Z = 2 m

Column of length *L* = 190 μm

Z = 0.02 m   Z = 2 m
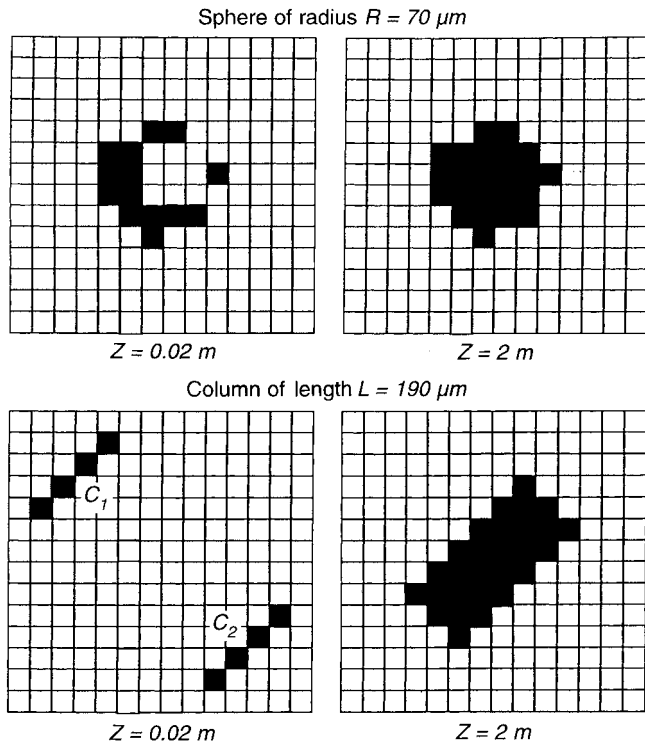
Fig. 2. Examples of diffracted images for spheres (of radius $R = 70$ μm) and hexagonal columns (of length $L = 190$ μm) while increasing the distance Z between the particle and the screen from 0.02 to 2.0 m

option has a small influence on the results considering that the analysis is conducted separately for the configurations with different numbers of pixels.

2.1.2 Modeling of the TAS clock frequency. The representativeness of the sampled images depends on both the true air speed (TAS) clock frequency and the lapse rate of the photodiodes. For instance, if the TAS clock frequency is not set precisely to the exact aircraft airspeed, then the corresponding images will be elongated or contracted. The mean pixel aspect ratio $\Delta x/\Delta y$ calculated for two data sets obtained during experiments within the pre-ICE and EUCREX (EUropean Cloud and Radiation EXperiment) program (Gayet *et al.*, 1990; Raschke *et al.*, 1990) are 0.9 and 2.4, respectively. Here, Δx characterizes the nominal pixel size (25 μm) defined by the array of the photodiodes, Δy is the pixel size along the flight direction (determined by the TAS clock). The mean value of $\Delta x/\Delta y$ strongly depends on the characteristics of the probe and on the aircraft airspeed. When the images are elongated, a pre-processing can be applied in order to compress the information to a ratio value of 1.0, but when the loss of material is too large (for a pixel ratio much larger than 1.0, which is typically the case for the EUCREX data), our approach is not applicable (it is actually the case with all the existing methods even for large particles). Finally the first column of each simulated image is eliminated in order to reproduce the triggering of the PMS-2DC probe.

2.2 Results of the statistical analysis

Note that the results of our model are only presented for a mean pixel aspect ratio $\Delta x/\Delta y$ equal to 1.0 in the rest of this study. Indeed, the computation of diffraction effects, even using the Maggi-Rubinowicz approximation, is very time consuming, and we had to restrain our study. Nevertheless, this work can be easily extended to other values of the mean pixel aspect ratio as mentioned in section 2.1.2.

In order to discriminate between the shapes of hexagonal columns from those of spherical particles, we sampled about 5000 particles for each data set. Many pixel configurations appear, especially the non-related arrangements that are due to the diffraction effects (see Fig. 2) which produce 'hollow' images (i.e., with a hole in the center of the spheres or where columns are split in two parallel parts). A configuration will be considered as non-related when there is at least one pixel of the image for which the 8-neighbors are not activated. For pattern recognition, we define the following categories:

- related and oriented vertically (referred hereafter as *RV*)
- related and oriented horizontally (*RH*)
- related and non-related with no particular orientation (*RD*1 and *RD*2 for 2-pixels, *U* otherwise)
- non-related and oriented vertically (*NRV*)
- non-related and oriented horizontally (*NRH*)

On Fig. 3a are represented the occurrence probabilities for six 2-pixel configurations (*RV*, *RH*, *RD*1 for the first diagonal and *RD*2 for the second one, *NRV* and *NRH*). For spherical particles the percentage of occurrence of *RV* is very different from that of *RH*, while in the case of hexagonal columns the percentages of occurrence of *RV* and *RH* are close to each other (12.1% and 9.9%). The same differences are seen for the non-related arrangements with 22.4% of *NRV* and 0.0% of *NRH* for spherical particles, and 32.1% of *NRV* and 24.0% of *NRH* for hexagonal columns.

The analysis of 3-pixel arrangements shown on Fig. 3b gives similar results. In these conditions, the two percentages of occurrence corresponding to *RH* and *RV*, which are only slightly different for both spherical particles and hexagonal columns, are nevertheless rather different from one another. For non-related arrangements, the contrast between *NRV* and *NRH* values is very large in case of spheres and small for hexagonal columns. The discrimination would obviously be easier if we only consider non-related-images. We therefore defined a new index, α, to discriminate the two types of particles as follows:

$$\alpha = \frac{PNRV}{PNRH} \tag{5}$$

where *PNRH* is the probability of non-related horizontal configurations and *PNRV* that of the vertical ones (we do not use the related configurations).

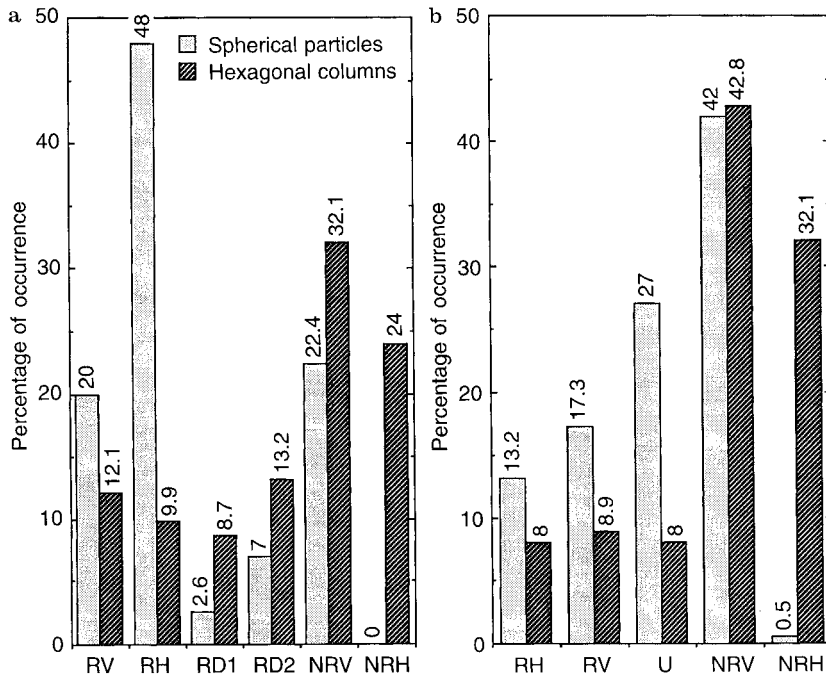The values of α obtained for 2-, 3- and more pixel arrangements are summarized in Table 1. This indicates

Fig. 3a, b. Configuration of **a** 2-pixels and **b** 3-pixels obtained with the model

Table 1. Values of α for different pixel configurations obtained using the numerical model

	Number of pixels			
	2-pixels	3-pixels	4-pixels	5-pixels
$\alpha_{columns}$	1.3	1.3	1.7	2.0
$\alpha_{spheres}$	$+\infty$	58.6	39.5	14.5

clearly that a small value of α is representative of hexagonal columns whereas spherical particles present larger values (that is to say $\alpha > 5.0$). In addition, we can see that the difference between spheres and hexagonal columns increases in terms of α with the number of pixels, but on the other hand, the proportion of non-related images decreases. The pattern recognition procedure is very tedious for more than 10 pixels, where for instance, the classical algorithm developed by Duroure (1982) becomes efficient.

Finally, we also checked the automatic pattern recognition procedure for $\Delta x/\Delta y$ much larger than 1.0 and found that it can not be applied for a mean pixel aspect ratio larger than 1.5. Therefore, the EUCREX data set ($\Delta x/\Delta y = 2.4$) is totally unusable, which is not surprising since the aircraft airspeed was very high (about 200 m/s).

2.3 Estimation of the volume of the particle

With our model, the geometrical characteristics of the particles (either spheres or hexagonal columns) are perfectly known. It is therefore possible to derive some kind of statistical 'transfer function' associating each pixel configuration with a mean particle size (for both

cases) and find the corresponding average volume value. Such theoretical relationships are given in detail in Table 3a (for spherical particles) and Table 3b (for hexagonal ice crystals), where the characteristic dimension of the particles is tabulated with the standard deviation for each of the two categories defined already.

Surprisingly, for a given number of pixels (for instance with the 2-pixel configurations *RH* and *NRV*) the dimension of the columns can change drastically (in this case by a factor of two), since both small and large particles can produce diffraction images classified in the same category. Furthermore, the number of pixels is not directly proportional to the size of the particle (i.e., a simple computation of the volume would be completely wrong), and the standard deviation is never larger than 2 pixels. As a consequence, one can simply estimate the order of magnitude of the ice or liquid water contents by multiplying the average volume (given in our tables) by the corresponding concentration and mass density of the hydrometeors.

3 Statistical analysis of actual data

3.1 Images with 2 pixels

The approach was validated against two actual data sets obtained during aircraft measurements on 17 February 1989 (pre-ICE'89) for which altostratus clouds were observed. This case was particularly interesting because spherical particles were located at the bottom of the cloud and hexagonal crystals in the upper part. Figure 4 shows an overview of the two types of particle shapes sampled with the PMS-2DC probe within the cloud, with a first group of spherical particles and a second one of hexagonal crystals. We assumed that the small

Spherical particles

Hexagonal columns

0 1 2 mm
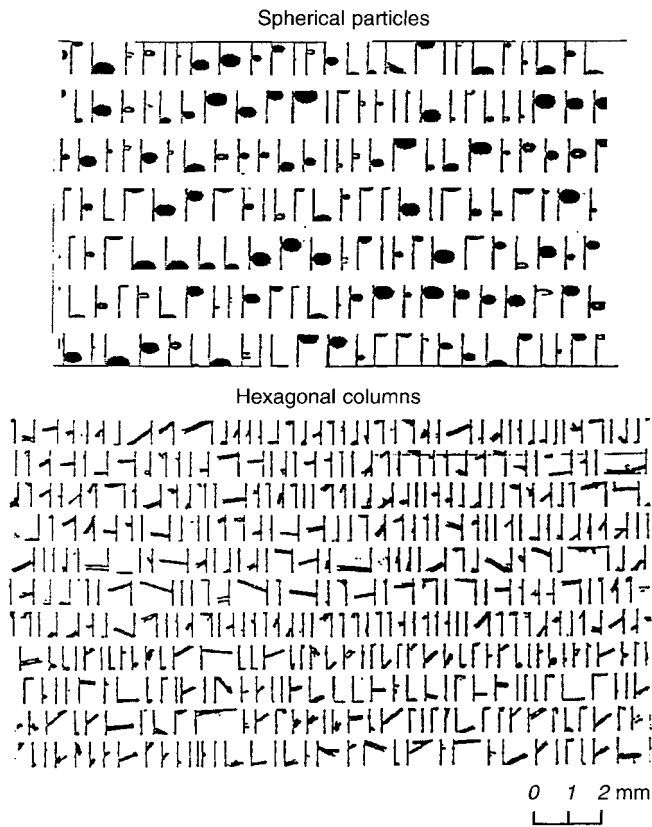
Fig. 4. Spherical particles and hexagonal crystals sampled with the PMS-2DC probe (17 February 1989)

particles located in the upper (lower) part were also hexagonal columns (spheres), and then conducted the discrimination analysis. The first step consists of compressing the data in order to reach a mean pixel aspect

ratio equal to 1.0 (note that $\Delta x/\Delta y = 0.9$ for the initial data set), but we found that this transformation does not disturb the process since $\Delta x/\Delta y$ was already close to the ideal 1.0 value.

Figure 5a displays (with the same representation as for Fig. 3a), the occurrence probabilities of 2-pixel configurations including the two groups of non-related arrangements (horizontal *NRH* and vertical *NRV*, respectively). The analysis indicates that non-related images represent about 25.0% of the total data. For spherical particles, most of the non-related configurations are vertical because of the combined effects of diffraction and triggering. For hexagonal columns, the occurrence probabilities of non-related horizontal and vertical arrangements are rather similar with 12.7% and 9.5%, respectively. This lack of distinction can be explained by the fact that, in natural clouds, small hexagonal columns may be randomly oriented in a three-dimensional space because of Brownian motion (which is the underlying assumption of our model). As for the simulated images, the recognition scheme is based on the emergence of differences between the occurrence probabilities corresponding to non-related configurations (actually there are almost only non-related vertical configurations for spherical particles). Indeed, the index α is equal to 1.3 for the hexagonal data set and 13.0 for the spherical particles (see Table 2).

Table 2. Values of α for different pixel configurations determined from aircraft measurements

	Number of pixels			
	2-pixels	3-pixels	4-pixels	5-pixels
$\alpha_{columns}$	1.3	2.2	6.0	1.7
$\alpha_{spheres}$	13.0	$+ \infty$	9.0	29.2
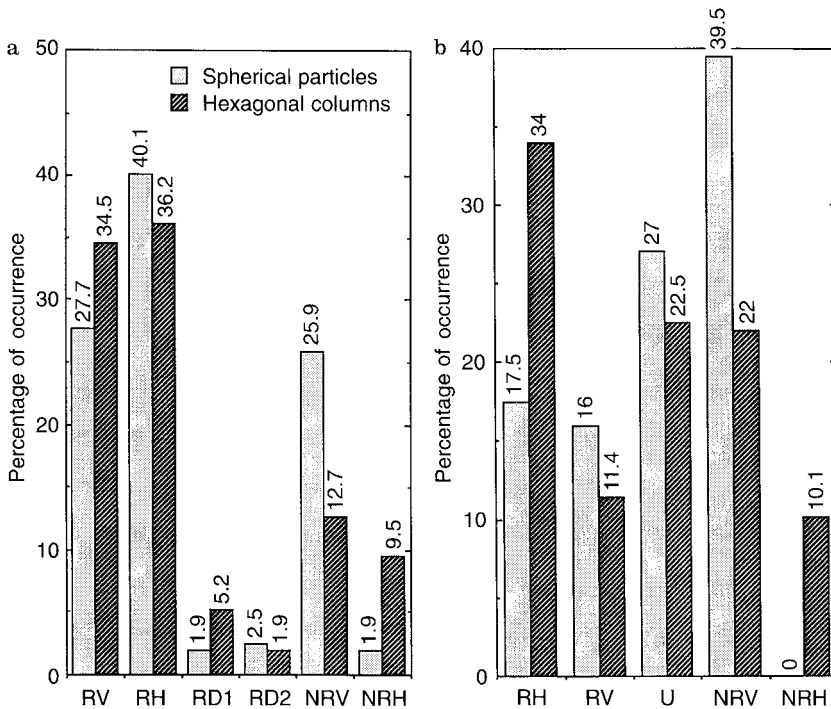
Fig. 5a, b. Configuration of **a** 2-pixels and **b** 3-pixels obtained in the case of actual data (pre-ICE'89)

Table 3a. Theoretical average radius R and standard deviation for spherical particles (μm).

Number of pixels	Configuration				
	RH	*RV*	*U*	*NRH*	*NRV*
2	55.9 ± 2.1	30.5 ± 3.4	36.7 ± 3.0	–	37.04 ± 1.8
3	40.0 ± 1.8	36.2 ± 1.6	38.6 ± 3.8	47.3 ± 7.1	45.3 ± 5.8
4	45.0 ± 7.0	39.0 ± 3.6	36.2 ± 1.6	42.4 ± 3.6	53.2 ± 4.9
5	62.6 ± 5.8	38.6 ± 1.9	42.3 ± 0.8	49.4 ± 4.4	50.8 ± 0.8

Table 3b. Theoretical average length L and the corresponding standard deviation (in μm) for hexagonal columns (note that the volume V of the particle is given by $V = \frac{3\sqrt{3}}{32}L^3$)

Number of pixels	Configurations				
	RH	*RV*	*U*	*NRH*	*NRV*
2	83.4 ± 29.3	87.1 ± 32.1	124.0 ± 24.0	109.3 ± 24.6	156.3 ± 38.4
3	129.5 ± 31.3	144.3 ± 24.5	113.6 ± 30.6	118.7 ± 25.8	104.4 ± 34.2
4	150.8 ± 33.7	144.5 ± 15.1	118.2 ± 19.5	106.7 ± 33.9	105.6 ± 34.8
5	128.8 ± 34.6	126.9 ± 23.6	144.1 ± 20.1	130.7 ± 30.1	129.5 ± 42.8

3.2 Images with at least 3 pixels

As for the previous simulations, the identification procedure becomes more and more efficient when the number of pixels in the image increases. For 3-pixels, the differentiation can also be made with the related configurations, as shown on Fig. 5b, since the horizontal configuration *RH* appears with an occurrence probability of about 34.0% for hexagonal columns and only 17.5% for spherical particles, moreover, the vertical configuration *RV* is more probable in the case of spheres. Nevertheless, this discrimination is not identical to that of the simulated images, and the retrieval of the volume of the particles will be erroneous. The recognition procedure can be unequivocally implemented by using the non-related arrangements which are only vertical with spherical particles (α is infinite) contrary to columns (where $\alpha = 2.2$).

When considering 4-pixel configurations, the identification becomes difficult for actual data (whereas it was still possible with simulated images!) since the values of α are abnormally of the same order of magnitude for both spheres and columns (values of 6.0 and 9.0 respectively as indicated in Table 2). No other easy-to-use differentiation criteria could be established, especially with the related configurations (which are extremely different), for many reasons, including the high complexity of the three-dimensional crystals shapes, the possible presence of some hydrometeors presenting a solid core surrounded by liquid water, the perturbation due to supercooled droplets or even much larger intruders among the sampled particles.

The analysis was also conducted for 5- to 10-pixel configurations, and the differentiation was always feasible. However, this kind of work becomes tiresome for more than 6-pixel arrangements, practically the effect of diffraction is much less perceptible for large particles, and therefore, because the total number of non-related configurations decreases, our statistical approach is not relevant.

The defined index α seems to be a simple criteria to discriminate spherical particles from hexagonal columns, and we can summarize the step-by-step procedure as follows:

1. Compute the mean pixel aspect ratio $\Delta x/\Delta y$ of the data set
2. Compress the images if $\Delta x/\Delta y$ is much lower than 1.0
3. Classify all the configurations in 3 categories: non-related vertical (*NRV*), non-related horizontal configurations (*NRH*) and the other ones
4. Calculate α, and if the value is larger than 5.0 then the particles are mostly spherical otherwise the particles have a high probability to be mostly hexagonal columns.
5. Retrieve the characteristic dimension of the particle using tables generated with the numerical model.

4 Summary and conclusions

In this study, we have built a model, including the effects of Fresnel's diffraction, allowing us to simulate the sampling of small hexagonal columns and spherical particles with a PMS-2DC probe. This complete model, which was found to be rather realistic, was successfully implemented to create several synthetic databases and to define a simple method in order to discriminate these two types of particles. Our algorithm determines an index named α defined as the ratio of the occurrence probabilities of non-related vertical and horizontal arrangements.

We have shown that it is possible to recognize automatically the shape of hexagonal columns and spherical particles (with a characteristic size < 200 μm) by counting the occurrence of the various pixel configurations. It was also emphasized that images sampled with a very high aircraft airspeed were not useable because of the loss of information due to the distortion of the pixels. The usage of the criteria α as a tool for the classification of cloud particles according to their shape was validated with actual data sets obtained during the Pre-ICE'89 field experiment.

We tabulated several theoretical relationships between the characteristics of the digitized images (i.e., the

number of pixels and the various arrangements) and the volume of spheres or hexagonal columns producing the same pattern after being diffracted on a screen, decreasing the unavoidable volume error for small particles. These results are encouraging, and it should be interesting to extend the applicability field of this statistical algorithm to a large number of different data sets obtained within various types of clouds (a very hard task) in order to refine the determination of structural, optical and microphysical quantities.

References

Born, M., and E. Wolf, *Principles of optics, electromagnetic theory of propagation interference and diffraction of light*, 6th edn., Pergamon Press, 808 pp. 1991.

Darlison, A. G., and P. R. A. Brown, The use of automatic particles recognition to improve the determination of bulk quantities from PMS-2D probe data in cirrus, *Proc. X*th Int. Cloud Physics Conf., Bad Homburg, FRG, August 15–20, 1988.

Duroure, C., Une nouvelle méthode de traitement d'hydrométéores donnée par les sondes bidimensionnelles. *J. Rech. Atmos.*, **16,** 71–74, 1982.

Duroure, C., H. R. Larsen, H. Isaka, and P. Personne, 2D image population analysis, *J. Rech. Atmos.*, **34,** 195–205, 1994.

Gayet, J. F., P. R. A. Brown and F. Albers, Results of the pre-ICE experiment. 42 pp. (Report available from LAMP, Université Blaise Pascal, 63177 Aubière Cedex, FRANCE), 1990.

Gayet, J. F., P. R. A. Brown, and F. Albers, A comparison of in-cloud measurements obtained with six PMS-2DC probes, *J. Atmos. Ocean Tech.*, **10,** 180–194, 1993.

Heymsfield, A., Ice crystal terminal velocities, *J. Atmos. Sci.*, **29,** 1348–1357, 1972.

Kinne, S., and K. N. Liou, The effects of the non-sphericity and size distribution of ice crystals on the radiative properties of cirrus clouds, *Atmos. Res.*, **24,** 273–284, 1989.

Knollenberg, R. G., The optical array: an alternative to scattering and extinction for airborne particle size determination, *J. Appl. Meteorol.*, **9,** 86–103, 1970.

Korolev, A. V., S. V. Kuznetsov, Y. E. Makarov, and V. S. Novikov, Evaluation of measurements of particle size and sample area from optical array probes, *J. Atmos. Ocean. Tech.*, **8,** 514–522, 1990.

Liou, K. N., *Radiation and cloud processes in the atmosphere, theory, observation, and modeling*, Oxford University Press, 1992.

Maggi, G. A., *Annali di Matem*, **16,** 21, 1888.

Miyamoto, K. and E. Wolf, *J. Opt. Soc. Am.*, **52,** 615–626, 1962.

Papoulis, A., *System and transforms with application in optics*, MacGraw-Hill, 1968.

Raschke, E., J. Schmetz, J. Heintzenberg, R. Kandel, and R. Saunders, The International Cirrus Experiment (ICE). A joint European effort, *ESA J.*, **14,** 193–199, 1990.

Rubinowicz, A., *Phys. Rev.*, **54,** 931, 1938.

Takano, Y., K. N Liou, and P. Minnis, The effect of small ice crystals on cirrus infrared radiative properties, *J. Atmos. Sci.*, **49,** 1487–1493, 1992.