Statistical distributions of water content and sizes for clouds above Europe

R. P. H. Berton

O.N.E.R.A., B.P. 72, F-92322 Châtillon Cedex, France, E-mail: berton@onera.fr

Abstract. The statistics of quantities involved in the synthesis of cloud scenes have been investigated from an original data base. Frequency distributions of ice and water content (IWC), horizontal and vertical sizes (L and H), and top temperatures (T) of clouds above Europe have been derived for nine types of clouds (As, Cb, Ci, Cg, LwCg, OrCg, Cs, Ns, Sc). It appears that the cumulated frequency plots can be well fitted with log-normal or Weibull profiles, and that for IWC and T cloud types can be split into two or three classes according to slopes in logarithmic coordinates. Cross-correlation coefficients between IWC, L, H and T have been also derived. Implications for the physics of the cloud build-up processes are briefly outlined. Critical analysis and comparison of other published results are proposed.

Key words: Meteorology and atmospheric dynamics (climatology)

1 Introduction

In order to test the performances of algorithms for the detection of a target in very different environments (clouds, forests, crops, sea, mountains...) subject to a wide variety of situations (source of illumination, position of observer...) it is often necessary to simulate these backgrounds by means of simple models. The synthesis of such natural scenes requires a good knowledge of the relevant parameters in the models (size, reflectivity...) and their frequencies of occurrence in various situations (type and coverage). Among such parameters, the geometrical dimensions and quantities reflecting the microphysical state (temperature, density of particles) are particularly important.

There are various ways for synthesizing the texture of a cloudy sky seen from space, most of them relying on Fourier transforms (Anjyo, 1991; Tessendorf et al., 1992), fractal cascades (Lévesque, 1991) or Mandelbrot-Weierstrass function (Jennane et al., 1997). The algorithms involved in these methods are generally parametrized by non-dimensional exponents (spectral slope, Hurst parameter, lacunarity parameter), microphysical quantities (temperature and water contents) controlling radiation transfer in the cloud, and geometrical sizes used in the computation of optical thicknesses. In the visible, the reflectance ρ of a cloud can be related to the liquid water content (LWC) via an observational relationship (Coakley and Snider, 1989). The cloud albedo α has been found to be a function of LWC and the observing angle θ (Minnis et al., 1990), and optical parameters have been given a parametrization in term of the ice and water content (IWC) in ice or mixed-phase clouds (Sun and Shine, 1994). Moreover, IWC has been shown to be locally an increasing function of the temperature in cirrus clouds (Heymsfield and Platt, 1984).

The statistical description of cloud systems has been actually addressed for about three decades (Plank, 1969), and information has been gained mainly about *cumulus* and *cirrus* clouds (Warner and Grumm, 1984; Joseph, 1985; First ISCCP Regional Experiment, FIRE II, 1989; Ansmann et al., 1993). Various parameters have been analyzed such as: height (Lopez, 1977; Chatterjee et al., 1991), horizontal size (Lopez, 1977), average distance (Joseph, 1985), horizontal area, area coverage and ascent rate (Warner and Grumm, 1984), droplet size (Pruppacher and Klett, 1980), total water content (Mazin, 1995). The *log-normality* of occurrence distributions has been rather well verified for these structures (Lopez, 1977; Mazin, 1995).

The present work tries to extend these conclusions by dealing with the information about *nine types of clouds*, namely As, Cb, Ci, Cg, LwCg, OrCg, Cs, Ns, Sc (see Table 2 for the meaning of these abbreviations), accumulated during five campaigns (Gayet et al., 1992) spread over ten years (Table 1). Depending on the kind of data analysis to be performed, clouds can be classified

Table 1. Experiment campaigns

Campaign	Location	Season	Type of cloud	Number of observations
PEP 79	North Spain	Spring	Cg, Sc, Ns	291
PEP 81	North Spain	Winter–Spring	Cg, Sc, Ns	616
FRONTS 84	S-W France	Summer	Cg, Cb	574
PREICE 89	Paris, France	Winter	Ci, Sc	129
ICE 89	North Sea	Autumn	Ci	345

either according to morphological features such as shape, altitude and height (Roulleau and Trochon, 1958), or according to multispectral criteria, such as brightness temperatures at 11 μm and 12 μm wavelength (Inoue, 1987; Rossow and Schiffer, 1991; Desbois et al., 1982), complemented with information from channels at visible and infrared wavelengths near 0.6 μm, 1 μm and 3.8 μm (Bankert, 1994). The morphological classification adopted here is due to a trained observer participating in the measurement flight.

After a description of the available data and a detailed account of error sources (Sect. 2), the analysis procedure will be explained (Sect. 3). The log-normal (Sect. 4) and Weibull (Sect. 5) distributions fitted here to our data will be compared to those of existing published works. A study of the correlation between all these parameters is then carried out (Sect. 6). These results will then be briefly discussed in terms of physical processes (Sect. 7).

2 Data

The present analysis uses aircraft data described in Gayet et al. (1992). It is composed of in situ measurements of thermodynamical and microphysical quantities along the trajectory of the aircraft, and additional information. The former include pressure, air and dewpoint temperatures, and the latter gather liquid water content, droplet concentrations measured with a Johnson-Williams probe. The additional data comprises the geographic location, the meteorological configuration (synoptic situation, air mass) and cloud characteristics (cloud type, temperature and levels of top and base, particle types, precipitation). Among these data a subset has been extracted in which the parameters of concern here are: the ice and water content IWC, the vertical thickness H, the top temperature T, and horizontal sizes L. The original sampling frequency of probe measurements was 1 Hz, but the values used in the present analysis are averages along homogeneous segments of about one minute or five kilometres (assuming an aircraft velocity $\mathbf{V} \approx 300$ km/h or 83 m/s).

These experiments were actually devised in order to study the effects of water erosion and they focused on particles larger than 50 μm. The ice and water content IWC, derived from particle counts may be therefore underestimated, and result in a systematic error. We shall consider here a precision for IWC of 10% due to instrumental noise, though the systematic error may be substantially larger (Larsen et al., 1998) and probably

shift the profiles of occurrence distributions to higher values. Nevertheless, since the discussion is based mainly on the slopes of cumulative distributions, this effect will not be taken into account in the present work.

The vertical thickness H was determined by the difference

$$H = h_u - h_l \tag{1}$$

between the cloud top and base heights h_u and h_l estimated visually by the pilot with reference to the plane's altimeter. The relative error for h_u and h_l, due to the observer's skill, is taken as 20%, so that we derive from Eq. (1) that the relative error for H is also equal to 20%.

The top temperature T was derived from an estimate of the top height h_u previously estimated. The error for T is therefore the sum of an error on h_u and an error on the reference temperature T_u at the height h_u, thus

$$\delta T = \frac{\mathrm{d}T}{\mathrm{d}z} \delta h_u + \delta T_u \tag{2}$$

where the vertical temperature gradient $\mathrm{d}T/\mathrm{d}z$ has the approximately constant value 9.8 K/km in the dry troposphere (Liou, 1992). With a relative accuracy of 20% on the top height, this yields a first term $\delta T \approx 2h_u$ (h_u in km, δT in K), and a value of a few degrees. The second term can be estimated as being not larger than a few 0.1 K and therefore considered negligible compared to the first term.

It should be emphasized that a horizontal length L corresponds to the distance flown by the plane through a *homogeneous* region of a cloud, and not necessarily to its full diameter; it may represent the core extension or the boundary thickness, or a part of them. In agreement with previous experiments (Jeck, 1988), homogeneous regions have been defined here as domains where either liquid water or ice or a mixture of ice and liquid water is found along the aircraft trajectory in the cloud. The accuracy of L-measurements can be estimated from the definition as a function of the flight time τ and the speed $\mathbf{V}$ of the aircraft

$$L = V\tau \tag{3}$$

as the error estimate

$$\mathrm{d}L = V\delta\tau + \tau\delta V \ , \tag{4}$$

or else the relative error estimate (L in meters)

$$\frac{\delta L}{L} = 0.1 + \frac{100}{L} \tag{5}$$

if we retain the following values: $\mathbf{V} = 100$ m/s, $\delta\mathbf{V}/\mathbf{V} = 0.1$, $\delta\tau = 1$ s.

These data have been used here to calculate at first discretized frequency distributions by means of the relation

$$f_i = \frac{1}{N}\{\text{number of samples such that:}$$

$$i\Delta u < u \le (i+1)\Delta u\} \tag{6}$$

where Δu and N denote respectively the sampling interval and the total number of samples. Discrete cumulated frequency distributions F have been then derived such that

$$F_j = \sum_{i \le j} f_i \ . \tag{7}$$

From the continuous expression of $F(u)$

$$F(u) = \int_{-\infty}^{u} f(v)\mathrm{d}v \ , \tag{8}$$

by elementary variational calculus it is straightforward to derive a formula for the accuracy δF of F-values, given the accuracy δu of u-measurements

$$\delta F = F'(u)\,\delta u \ . \tag{9}$$

Application of Eq. (9) to analyzed data will provide error bars on displays with the corresponding scaling of axes.

The cloud types and the number of observations for each of them are displayed in Table 2. It should be kept in mind that the frequencies shown there *do not represent* a measure of the spatial or temporal coverage by these various kinds of clouds, but merely reflect the interest of the scientists involved in the campaigns or the relative success in measuring such or such type of cloud. Note that no cirrocumulus (Cc) or stratus (St) were present in this sample, and that the data about altocumulus (Ac) and cumulus (Cu) include too few cases to allow the determination of corresponding frequency distributions. Nevertheless, Table 2 shows that all of the cumulonimbus (Cb) were observed during FRONTS 84,

Table 2. Number of observations for each type of cloud

Abbreviation	Type of cloud	Number of observations	Frequency
Ac	Altocumulus	1	0.0005
As	Altostratus	213	0.109
Cb	Cumulonimbus	234	0.120
Cg	Cumulus congestus	644	0.329
LwCg	Long-wave cumulus congestus	30	0.015
OrCg	Orographic cumulus congestus	29	0.015
Ci	Cirrus	186	0.095
Cs	Cirrostratus	136	0.070
Cu	Cumulus	2	0.001
Cw	Cumulus wave	22	0.011
Ns	Nimbostratus	89	0.046
Sc	Stratocumulus	369	0.189
Total		1955	1.000

and all of cirrus (Ci) during PREICE 89 and ICE 89, and therefore they constitute two homogeneous subsets of data.

The first two moments of these measured distributions, average μ and centered standard deviation σ, can be estimated directly by means of the classical expressions

$$\mu = \sum_i f_i u_i \Delta u_i \quad \sigma^2 = \sum_i f_i (u_i - \mu)^2 \Delta u_i \tag{10}$$

3 Fitting distributions

The variables IWC, H, L, $T+273$ can be fitted with usual distribution functions. Because these quantities are essentially positive, only frequency distribution functions which, in the limit of accuracy, are equal to zero for negative frequencies need to be considered. Such distributions are summarised in Table 3, with their frequency density function $f(u)$, their cumulative frequency function $F(u)$, their average μ and centered standard deviation σ (Johnson and Kotz, 1970; Nakagami, 1960; Bischoff and Chytil, 1969) where $g(u,v)$, $erf(u)$ and $K_a(u)$ denote respectively the incomplete gamma function, the error function and the modified Bessel function of second kind and order a (Abramowitz and Stegun, 1972).

The log-normal distribution has been quite widely used with success in the analysis of clouds corresponding to different situations (Houze and Cheng, 1977; Lopez, 1976, 1977; Cahalan and Snider, 1990; Chatterjee et al., 1991; Mazin, 1995). Moreover, its cumulative frequency distribution can be easily found in tables or computed with good accuracy from analytical expressions (Bagby, 1995). Nevertheless, as we shall see, the fitting is worse for extreme values, and in that respect, Weibull distributions may be more adequate. We shall therefore proceed to analyze our data according to these two distributions.

In this work, *instead of using the conventional log-probability paper* for plotting the graph of a random variable U under study, we chose to plot the data in the following coordinates systems.

For the log-normal analysis we plot in abscissa $x = \log u$, where log denotes the decimal logarithm, and in ordinate the quantity $z = F_N^{-1}(y)$, where y represents the cumulated frequencies of U and F_N^{-1} denotes the inverse of the cumulated frequency distribution function $F_N(u)$ of the standard normal distribution. This way of display enables one to calculate directly the parameters of the distributions since the points corresponding to the data, with coordinates $Ln\, u$ and $F_N^{-1}(y)$ will be aligned along a straight line, known as a Henry line (Johnson and Kotz, 1970) the equation of which is

$$z = \frac{Ln\, u - m}{b} \tag{11}$$

where m and b denote the parameters of the normal distribution (Table 3). In practice, the decimal logarithm is used, so that Eq. (11) becomes

Table 3. Some usual statistical distributions

Distribution	$f(u)$	$F(u)$	μ	σ
Log-normal	$\frac{1}{\sqrt{2\pi}\,b\,u}\exp\left\{-\frac{(Ln\,u-m)^2}{2b^2}\right\}$	$\frac{1}{2}\left[\mathrm{erf}\left(\frac{Ln\,u-m}{b\sqrt{2}}\right)+1\right]$	$\exp\left(m+\frac{b^2}{2}\right)$	$e^{m+\frac{b^2}{2}}\sqrt{e^{b^2}-1}$
Gamma	$\frac{1}{\Gamma(a)b}\left(\frac{u}{b}\right)^{a-1}\exp\left(-\frac{u}{b}\right)$	$\frac{1}{\Gamma(a)}\,\gamma\left(a,\frac{u}{b}\right)$	ab	$b\sqrt{a}$
Nakagami	$\frac{2\,a^a}{\Gamma(a)\,b}\left(\frac{u}{b}\right)^{2a-1}\exp\left\{-a\left(\frac{u}{b}\right)^2\right\}$	$\frac{1}{\Gamma(a)}\,\gamma\left[a,a\left(\frac{u}{b}\right)^2\right]$	$\frac{b}{\sqrt{a}}\frac{\Gamma\left(a+\frac{1}{2}\right)}{\Gamma(a)}$	$b\sqrt{1-\frac{\Gamma\left(a+\frac{1}{2}\right)^2}{a\,\Gamma(a)^2}}$
Rayleigh	$2\frac{u-m}{b^2}\exp\left\{-\left(\frac{u-m}{b}\right)^2\right\}$	$1-\exp\left\{-\left(\frac{u-m}{b}\right)^2\right\}$	$m+b\frac{\sqrt{\pi}}{2}$	$b\sqrt{1-\frac{\pi}{4}}$
Weibull	$\frac{a}{b}\left(\frac{u-m}{b}\right)^{a-1}\exp\left\{-\left(\frac{u-m}{b}\right)^a\right\}$	$1-\exp\left\{-\left(\frac{u-m}{b}\right)^a\right\}$	$m+b\,\Gamma\left(1+\frac{1}{a}\right)$	$b\sqrt{\Gamma\left(1+\frac{2}{a}\right)-\Gamma\left(1+\frac{1}{a}\right)^2}$
K	$\frac{2}{\Gamma(a)b}\left(\frac{u}{2b}\right)^a K_{a-1}\left(\frac{u}{b}\right)$		$b\sqrt{\pi}\frac{\Gamma\left(a+\frac{1}{2}\right)}{\Gamma(a)}$	$2b\sqrt{a-\frac{\pi}{4}\frac{\Gamma\left(a+\frac{1}{2}\right)^2}{\Gamma(a)^2}}$

$$z = \frac{2.303}{b}x - \frac{m}{b} \quad . \tag{12}$$

This Henry line has therefore a slope p and an intercept q ($Ln\,u = 0$) determined through a least squares minimization and from which m and b can be eventually evaluated as

$$m = -2.303\,\frac{q}{p} \qquad b = \frac{2.303}{p} \quad . \tag{13}$$

For the Weibull analysis, we plot in abscissa $x = \log u$ and in ordinate the quantity $z = \log[-\log(1-y)]$ where y represents the cumulated frequencies of U. If the distribution is of Weibull type, the graph is therefore a straight line whose equation writes

$$z = \log[-\log(1-y)] = ax - a\log b - 0.3622 \tag{14}$$

where a and b are respectively the shape and the scale parameters (Table 3). From the slope p and intercept q determined through a least squares minimization, a and b can be evaluated as

$$a = p \qquad b = 10^{-\frac{q+0.3622}{p}} \quad . \tag{15}$$

It is important to estimate the expression of errors on probability plots as functions of errors on measurements, taking into account the scaling of the axes. The expressions derived from Eq. (9) after some algebra in the two cases studied can be cast into the unique form

$$\delta z = 10^{-x}\frac{\mathrm{d}z}{\mathrm{d}x}\,\delta u \quad . \tag{16}$$

The resulting errors on the regression coefficients p and q have been calculated in Appendix A3.

4 Log-normal analysis

4.1 Horizontal size L

Figure 1 shows the cumulative distributions of L obtained for all of the nine types of clouds, and fitted with log-normal curves. Except for the Ns data which are too scarce, the other data points seem, *in the limit of error bars displayed on each plot*, to be log-normally distributed and to fall in the same category.
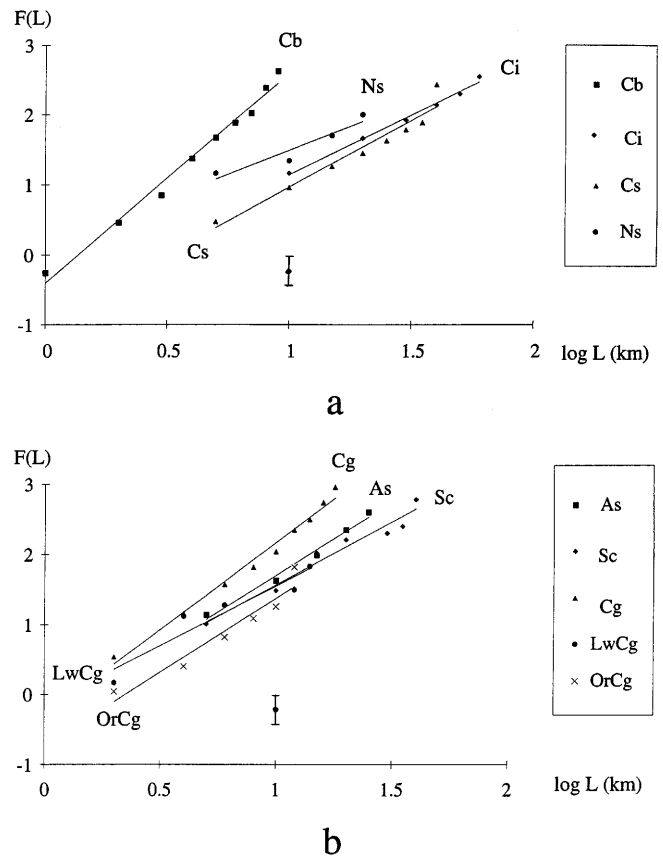

Fig. 1a, b. Cumulative frequency distributions of size (L, in km) and log-normal fitting: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg, LwCg, OrCg. The error bar displayed is calculated using Eqs. (16) and (3). Its magnitude is roughly constant in the range of sizes studied

As has been said in Sect. 2, the sizes L are not representative of the total diameter of clouds, but of homogeneous horizontal segments. In that respect, only the largest value of L measured for a fixed cloud type gives an idea of its horizontal size.

The statistical distribution of *full* chord lengths along a direction at angle θ with the vertical can be determined in the simple case of an axisymmetrical ellipsoid (Su and Pomraning, 1994)

$$\begin{cases} f(L) = \frac{1}{2R_{\text{eff}}^2} & \text{if} \quad 0 \le L \le 2R_{\text{eff}} \\ f(L) = 0 & \text{if} \quad 2R_{\text{eff}} < L \end{cases} \quad (17)$$

where the effective radius R_{eff} is such that

$$\frac{1}{R_{\text{eff}}^2} = \frac{\cos^2 \theta}{A^2} + \frac{\sin^2 \theta}{B^2} \quad (18)$$

with semi-axes A (vertical) and B (horizontal). For a horizontal chord, we have $\theta = 0$ and thus $R_{\text{eff}} = B$.

4.2 Thickness H

Figure 2 shows the cumulative distributions of H obtained also for all of the nine types of clouds, and fitted with log-normal curves. The estimation of the cumulated value through the log-normal scaling from numerical tables is too inaccurate in the distribution tails, so that we generally *omitted the extreme value*. Despite this reduction of data, the fitting seems globally satisfying, though the estimated errors are large. Because of this, the fact that a first group (Ci, Cb, Ns) has larger slopes ($p > 6$), and consequently smaller scale parameters ($b \le 0.4$), is highly uncertain. Moreover,

these types of clouds will be found to form a separate class according to other quantities.

Our results for cumulus and cirrus can be compared with published values analyzed in Appendices A1 and A2 respectively. For *cumulus* clouds (Chatterjee *et al.*, 1991), despite our large errors scale parameters are about the same but our means are one half or one third of theirs. This may also reflect the fact that the cumuli analyzed by these authors were tropical clouds, with much larger vertical extension than our middle-latitude clouds. For *cirrus* clouds (Uttal *et al.*, 1993), means m are about the same, but our scale parameter b is smaller.

4.3 Ice and water content IWC

Figure 3 shows the cumulative distributions of *IWC* obtained again for all of the nine types of clouds, and fitted with log-normal curves. This time, errors are small enough to ensure a good fitting in general, although small deviations appear for large values of *IWC*, which can be explained by the fact that large values are missing (Lopez, 1977). It is interesting to notice that again two families of clouds can be distinguished according to the slope of the fitted log-normal lines, and appearing in
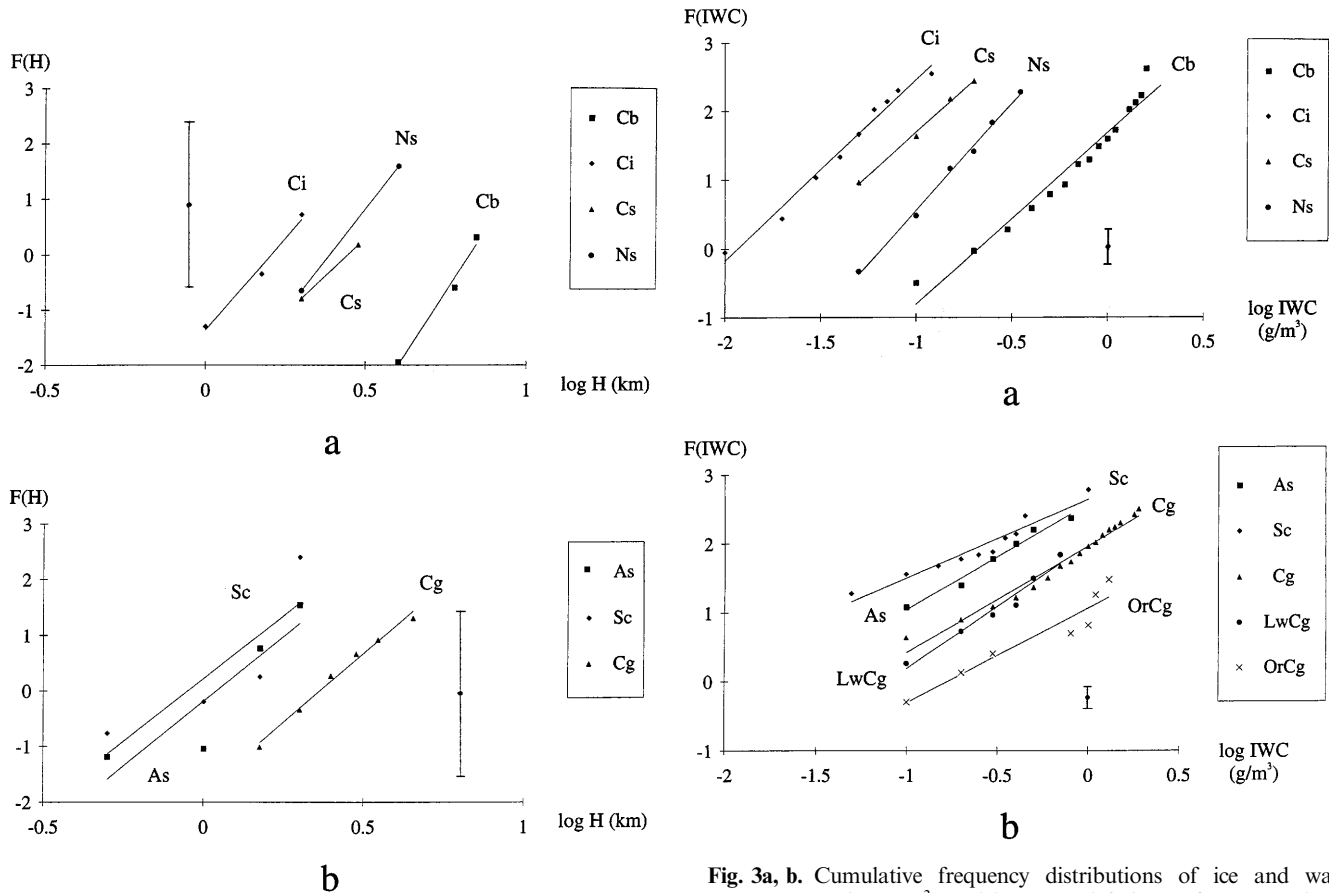

Fig. 2a, b. Cumulative frequency distributions of thickness (H, in km) and log-normal fitting: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg. Application of formula (16) with an accuracy of 20% on heights yields rather large errors of about ±1.5 on distributions

Fig. 3a, b. Cumulative frequency distributions of ice and water content (*IWC*, in g m^{-3}) and log-normal fitting: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg, LwCg, OrCg. The error bar displayed is calculated using Eq. (16) and a relative accuracy of 10% on IWC. Its magnitude is roughly constant in the range of water contents studied

Fig. 3a, b. In the first group (Cb, Ci, Cs, Ns), slopes are steeper ($p \geq 2.5$) denoting a smaller scale parameter ($b < 1$). In the second group (As, Sc, Cg, LwCg, OrCg), smaller slopes ($p \leq 1.5$) correspond to larger scale parameters ($b > 1$).

4.4 Top temperature T

Figure 4 shows the cumulative distributions of T obtained for all of the nine types of clouds, and fitted with log-normal curves. Among the two families previously found in the case of H and IWC, only Cs and Ns exhibit a much larger slope ($p > 100$), and therefore a smaller scale parameter ($b \leq 0.02$). A more detailed account of the confidence levels will be carried out in the section devoted to Weibull analysis.

5 Weibull analysis

5.1 Horizontal size L

Figure 5 shows the cumulative distributions of L obtained for seven types of clouds, and fitted with Weibull curves. On the one hand, the cloud types Cb and OrCg have larger slopes ($a \approx 1$) than the other
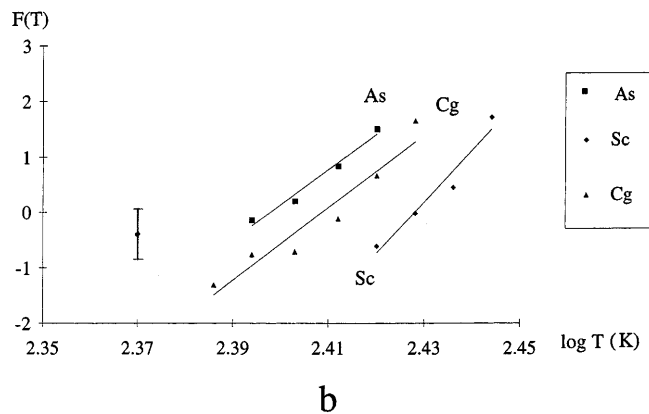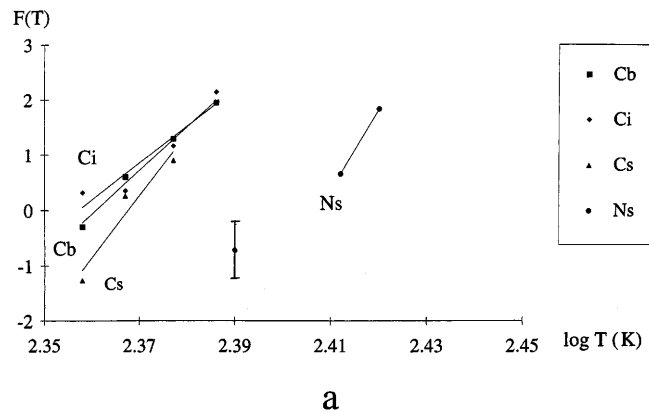

a

b

Fig. 4a, b. Cumulative frequency distributions of top temperature (T, in K) and log-normal fitting: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg. The error bar displayed is calculated using Eq. (16) and an accuracy of 2 K on temperatures
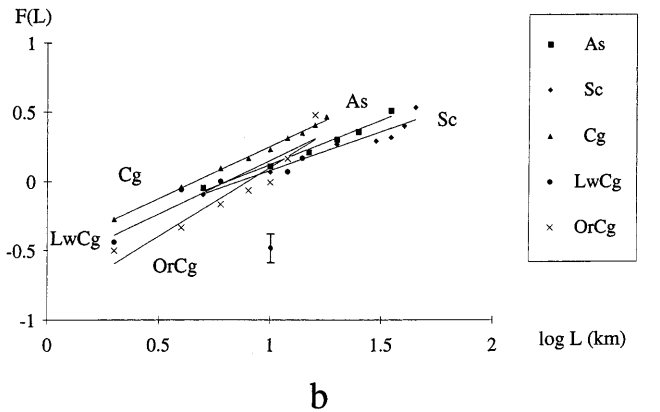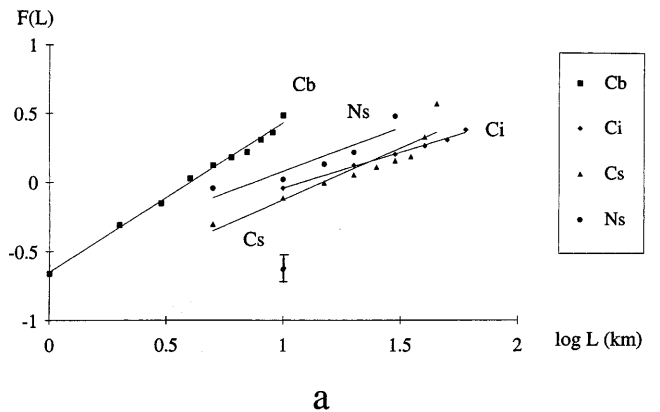

a

b

Fig. 5a, b. Cumulative frequency distributions of size (L, in km) and Weibull fitting: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg, LwCg, OrCg. The error bar displayed is calculated using Eqs. (16) and (3). Its magnitude is roughly constant in the range of sizes studied

types, and on the other hand the cirrus type ($a \approx 0.5$) smaller slopes. The errors on the slope estimated from expressions A3.5 ($\delta p \approx 0.2$) allow this distinction to some extent. There is no basic difference with the log-normal fit, either qualitative or quantitative.

An exponential distribution has been found (Plank, 1969) for the cloud base diameter L in the case of cumuli, mainly of Cg type

$$f(L) = Ae^{-\frac{L}{B}} \tag{19}$$

with $A = 1/B$ (for normalization) and B of the order of 0.2 km. This actually is a Weibull distribution with exponent $a = 1$, and it agrees particularly well with our OrCg case.

5.2 Thickness H

Figure 6 shows the cumulative distributions of H obtained also for seven types of clouds, and fitted with Weibull curves. Since data can be plotted with a few more points than in the log-normal analysis, the fitting should more reliable but the estimated errors are pretty large. In particular, the type Ns could be included into the group As, Sc, Cg because its slope is smaller. We could thus obtain the two classes:
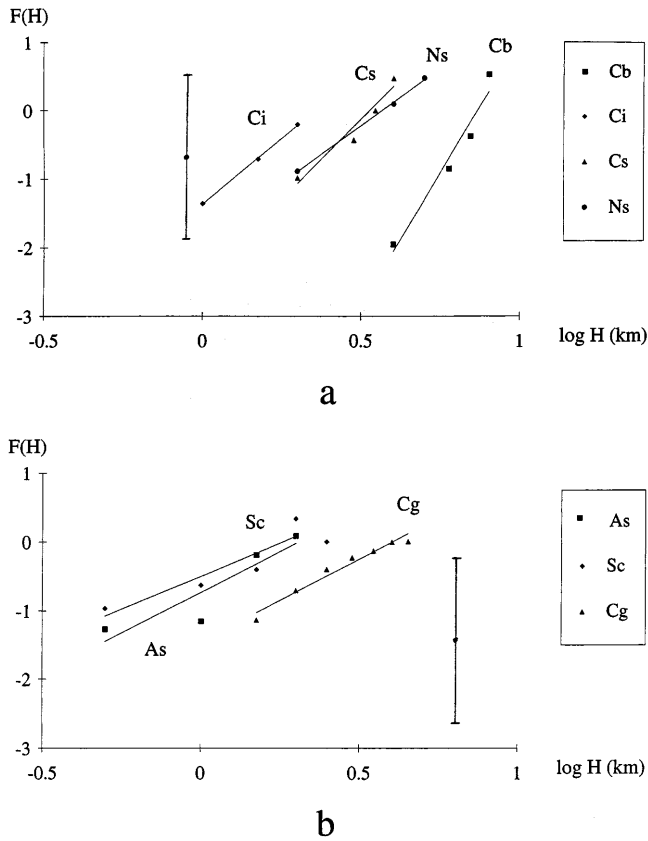
Fig. 6a, b. Cumulative frequency distributions of thickness (H, in km) and Weibull fitting: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg. Application of formula (16) with an accuracy of 20% on heights yields rather large errors of about ± 1.5 on distributions
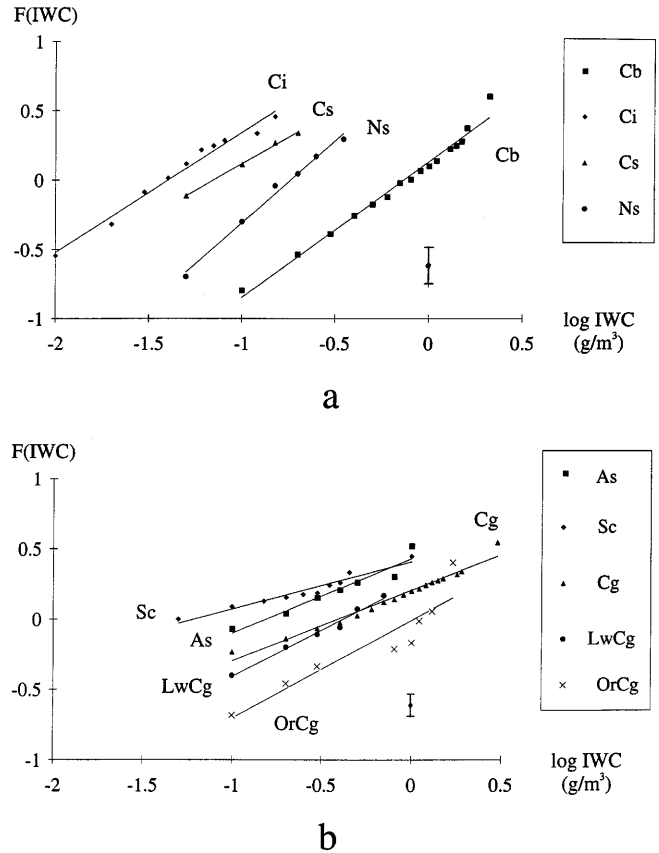

Fig. 7a, b. Cumulative frequency distributions of ice and water content (*IWC*, in g m^{-3}) and Weibull fitting: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg, LwCg, OrCg. The error bar displayed is calculated using Eq. (16) and a relative accuracy of 10% on *IWC*. Its magnitude is roughly constant in the range of water contents studied

Class A :	Cb, Ci, Cs	$a \geq 4$
Class B :	As, Sc, Cg, Ns	$a < 4$

Unfortunately, the accuracy of p estimated by means of A3.5 gives $\delta p \approx 4$. Were the errors on H smaller, this would yield nearly the same division as that found by log-normal analysis of the water content *IWC*, except for the Ns type.

The comparison with other data (Uttal *et al.*, 1993) shows discrepancies (Appendix A2) that can be accounted for by the lack of accuracy in our data. Actually, our shape parameter is four times larger than theirs ($a \approx 1$), whereas our scale parameter is slightly smaller than theirs ($b \approx 3$).

5.3 Ice and water content IWC

Figure 7 shows the cumulative distributions of *IWC* obtained again for all of the nine types of clouds, and fitted with Weibull curves. Without any difference, the two classes found by log-normal analysis are found again here:

Class A :	Cb, Ci, Cs, Ns	$a > 0.7$
Class B :	As, Sc, Cg, LwCg, OrCg	$a \leq 0.7$

Using the expressions A3.5 we find an accuracy of about 0.2 on the slope p, and add some degree of confidence to this result.

5.4 Top temperature T

Figures 8 and 9 show the cumulative distributions of T obtained for all of the nine types of clouds, and fitted with Weibull curves, with $m = 0$ and $m = 227$ K respectively. The three-parameter distribution ($m = 227$ K) yields smaller shape parameters a by about one order of magnitude than those found with a two-parameter distribution ($m = 0$). Due to the small dispersion of temperature data in the case $m = 0$ (Fig. 8), the absolute error δp calculated with formula A3.5 amounts several thousands and does not allow any conclusion. The situation is much more favourable for $m = 227$ (Fig. 9), since the relative dispersion is larger, and therefore the error $\delta p \approx 0.2$ is reasonably smaller than the slope itself ($p \approx 0.5$).

Whereas the log-normal analysis did not show a clear discrimination between two classes, here the As, Cg, Ns and Sc types have steeper slopes than Ci, Cb, Cs; this corresponds fairly well to the two *IWC* or H classes A and B.
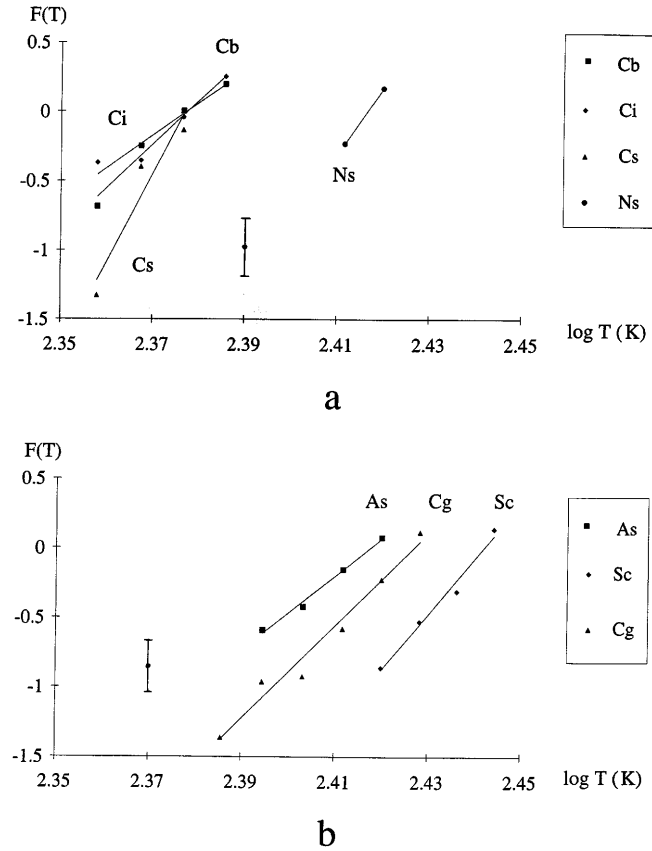
Fig. 8a, b. Cumulative frequency distributions of top temperature (T, in K) and Weibull fitting with $m = 0$: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg. The error bar displayed is calculated using Eq. (16) and an accuracy of 2 K on temperatures
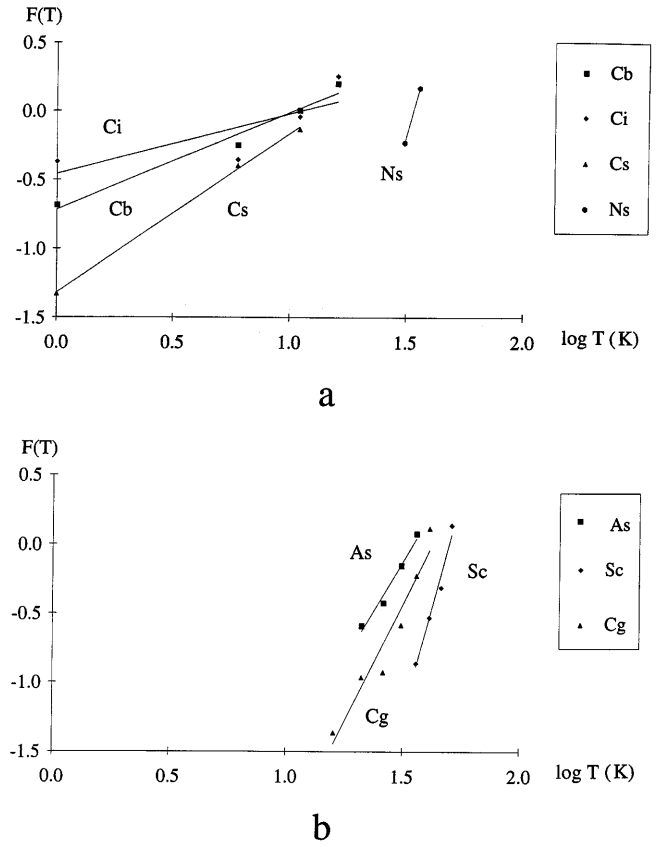
Fig. 9a, b. Cumulative frequency distributions of top temperature (T, in K) and Weibull fitting with $m = 227$: **a** for types Cb, Ci, Cs, Ns; **b** for types As, Sc, Cg

6 Correlations

If we want to use the distributions previously derived in order to build up realistically a cloudy sky scene by random drawing of relevant quantities, say H, L, IWC and T, it is also important to know whether these are correlated or not. In other words, certain combinations of these parameters may not be valid because of cross-correlations. In such a case, the knowledge of cross-distributions $f(H, L)$, $f(H, IWC)$, $f(H, T)$, $f(L, T)$ and so on, would be greatly appreciated.

Unfortunately, our data are not equally sampled for H, L, IWC and T, since for instance more data points are available for IWC, and thus it is not possible at this stage to derive cross-distributions. Nevertheless, we can have an idea of this situation by estimating the cross-correlation coefficient, defined for two random variables X and Y by

$$\rho_{XY} = \frac{\overline{(X - \mu_X)(Y - \mu_Y)}}{\sigma_X \, \sigma_Y} \qquad (20)$$

or equivalently, if we set $Z = XY$,

$$\rho_{XY} = \frac{\mu_Z - \mu_X \mu_Y}{\sigma_X \, \sigma_Y} \; . \qquad (21)$$

Results are displayed in Table 4 for the pairs (L, H), (H, T), (L, T), (L, IWC), (T, IWC). Note that a negative correlation coefficient simply means that the slope of the least-square line fitting the data set is also negative. Moreover, the dispersion of the data around this line is described by the deviation of ρ from unity. This can be simply shown using a model (Appendix A4).

A correlation between the mean ice and water content $\overline{IWC}$ (horizontal average) and L or H has been already pointed out in *tropical cumulus clouds* (Ackerman, 1959), and the dependencies found by the author can be written as

$$\overline{IWC} = 0.485\,L + 0.705 \qquad (22)$$

and

$$\overline{IWC} = 0.866\,H + 0.368 \qquad (23)$$

with $\overline{IWC}$ in g/m³ and H, L in km. The ranges of these parameters are 0.9–2.3 g/m³ for IWC, 0.3–3.3 km for L and 0.5–2.3 km for H. Actually, our values of the correlation coefficient for cumulus congestus (Cg) are also positive for both pairs (IWC, L) and (IWC, H), which result is quite consistent with the positive slopes in Eqs. (22) and (23), 0.485 and 0.866 respectively, although smaller. The correlation between L and IWC is more natural for convective clouds, in which an

Table 4. Correlation coefficients

Cloud type	(L,H)	(H,T)	(T,L)	(L,IWC)	(T,IWC)	(H,IWC)
As	0.24	0.0077	0.046	0.088	0.21	0.68
Cb	0.021	−0.69	−0.094	0.10	0.075	−0.016
Ci	−0.24	0.38	−0.14	0.16	0.15	0.10
Cg	−0.051	−0.82	0.0031	0.069	−0.25	0.25
Cs	−0.029	−0.27	−0.16	0.37	−0.032	−0.11
Ns	−0.090	−0.81	−0.024	0.10	0.17	−0.094
Sc	0.063	−0.20	0.12	0.12	−0.16	0.030

observed H-L correlation (Ackerman, 1959) could be explained by the isotropy of turbulence.

A correlation between H and L', where L' denotes the cloud base diameter, has been shown to follow a power law in the case of cumuli (Plank, 1969)

$$H = a \, L'^{\beta} \tag{24}$$

with $a \approx 0.5$ and $\beta \approx 0.6$. Clouds are modelled as axisymmetrical paraboloids. By taking the logarithm of relation (24), we see from Appendix A4 that the correlation between Log H and Log L' should have the same sigh as β. The correlation coefficient $\rho_{\mathrm{Log}H\mathrm{Log}L}$ estimated from our Cg data (-0.048) is slightly different from the coefficient ρ_{HL} (-0.051) appearing in Table 4, and negative. Since the exponent β in Eq. (24) is positive, Plank's result is not confirmed by our data.

7 Discussion

The determination of statistical distributions of quantities L, H, IWC and T according to the type of cloud has been carried out quite satisfyingly for IWC, and more coarsely for L, H and T, in the limit of error bars displayed on each plot. The reliability of the derived log-normal or Weibull parameters actually depends not on the sample size M (see Table 2) but on the number of statistical classes N which could be distinguished, and which corresponds to the number of points on the cumulated curves. Since IWC is a local parameter, it has been sampled along the aircraft trajectory so that this number N is in general large, even for rare clouds. On the opposite, L, H and T are global parameters, common to several positions of the aircraft, so that N may be small though M is large. Moreover, even when N is large enough, measurement errors reduce the number of effective data (e.g., the thickness H for As) and the degree of confidence on output parameters p and q.

In order to make ideas more precise in the most favourable case of IWC, we have plotted both distributions (log-normal an Weibull) together with the original data for each type of cloud (Fig. 10). Note that the data are sampled in finite-width classes, say Δu, and therefore the corresponding theoretical distribution to be plotted, $\tilde{f}(u)$, is equal to $f(u)$ integrated between u and $u + \Delta u$

$$\tilde{f}(u) = \int_{u}^{u+\Delta u} f(t)\mathrm{d}t = F(u + \Delta u) - F(u) \; , \tag{25}$$

which is nothing but the increment of the cumulative distribution function $F(u)$ in the same interval. For data with few points, the fit is about the same with log-normal and Weibull distributions (As, Cg, LwCg, OrCg, Cs, Sc). For data with high sampling rates, the fit is better with a Weibull distribution than with a log-normal one (Cb, Ci, Ns).

The mathematical nature of the distribution describing the statistical behaviour of IWC, L, H, T or other variables is related to the physical processes happening in clouds (Chatterjee *et al.*, 1991). General theory says that a log-normal distribution is obtained as soon as the random variable X follows the law of proportionate effects (Aitchison and Brown, 1973). Of course, the present analysis used only one distribution of each kind at a time, but it has been shown that a mixture of several log-normal distributions with different parameters may also account for observed data (Kayano and Shimizu, 1994).

On the opposite Weibull distribution, which seems to describe better our data, has no simple physical explanation. Nevertheless, in the context of electromagnetic scattering by a random medium (Blacknell, 1994; Jakeman, 1987) it has been shown to be an approximation of a K distribution (see Table 3). Actually, for a given Weibull-distributed random variable with parameters a and b it is straightforward to find the closest K distribution with parameters a' and b' by just equating their first and second moments μ and σ

$$b\Gamma\left(1 + \frac{1}{a}\right) = b'\sqrt{\pi}\,\frac{\Gamma\left(a' + \frac{1}{2}\right)}{\Gamma(a')} \qquad b\Gamma\left(1 + \frac{2}{a}\right) = 4a'b'^2 \tag{26}$$

which yields an implicit equation

$$\Gamma\left(1 + \frac{1}{a}\right)\Gamma\left(1 + \frac{2}{a}\right)^{-1} = \frac{\sqrt{\pi}}{2\sqrt{a'}}\Gamma\left(a' + \frac{1}{2}\right)\Gamma(a')^{-1} \; . \tag{27}$$

A plot of a' versus a (Fig. 12) clearly shows the close relationship of both distributions. In particular, when $a = 1$, then $a' = 1/2$ exactly.

A K distribution can be thought of as the distribution of a compound stochastic process. Let us consider a process X with distribution $p(x, y)$ whose average y itself is a random variable Y which obeys a distribution $q(y, z)$ with average z. Then the resulting process will admit the distribution $r(x, z)$ defined by

$$r(x, z) = p(x, y) \wedge q(y, z) = \int_{0}^{\infty} p(x, y)q(y, z)dy \; , \tag{28}$$

and the average easily found to be

$$\langle r(x, z) \rangle = z \; . \tag{29}$$

It is interesting to notice that if $p(x, y)$ is an exponential and $q(y, z)$ a gamma function, then $r(x, z)$ is a K distribution. Gamma distributions arise in natural structures such as vegetation canopies, irregular terrain, ocean surfaces, cloud boundaries, atmospheric layers, which are essentially fractal (Mandelbrot, 1982).
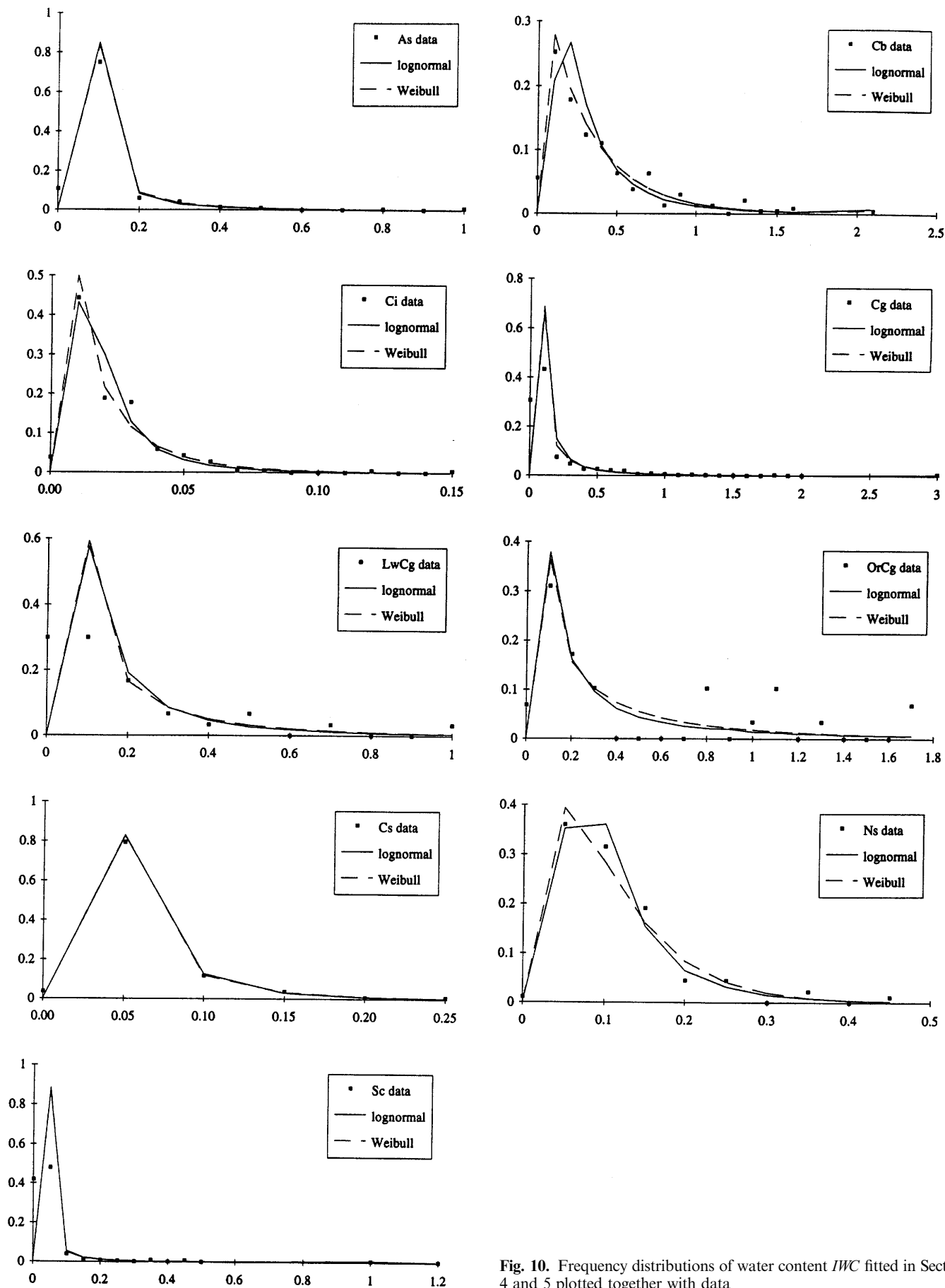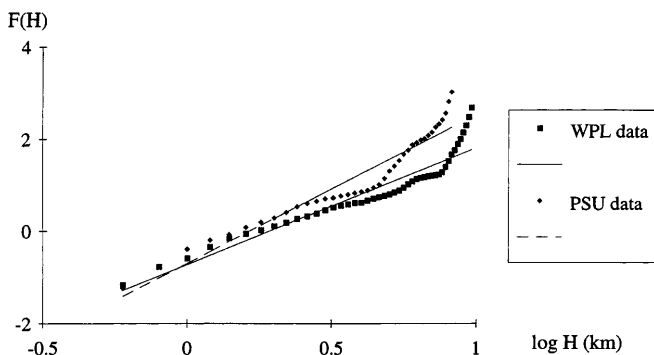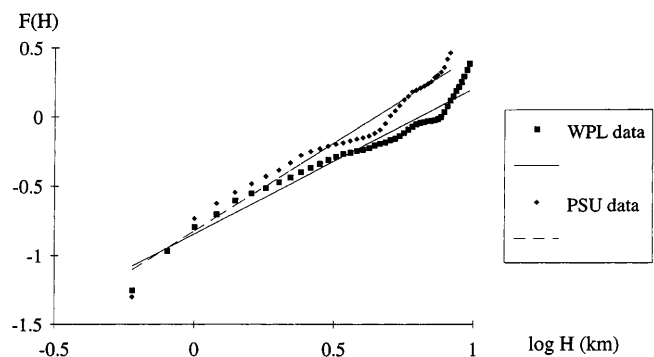
Fig. 10. Frequency distributions of water content *IWC* fitted in Sects. 4 and 5 plotted together with data

Fig. 11a, b. Cumulative frequency of cirrus thickness H (after Uttal et al., 1993) fitted with: **a** log-normal distributions; **b** Weibull distributions
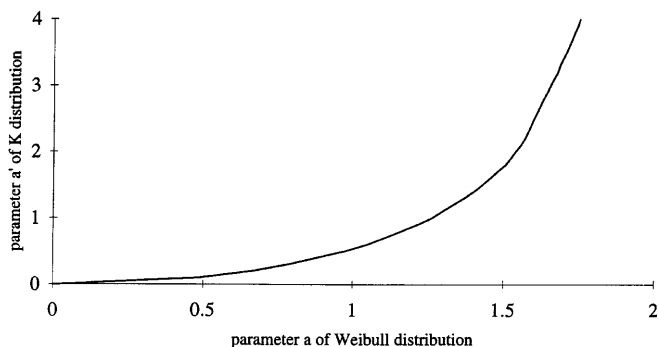

Fig. 12. Plot of parameter a' of K distribution versus parameter a of Weibull distribution

Therefore it would make sense to use K-distributions to describe further data more accurate than those dealt with here, although such a method would be heavier, because K functions of non-integer order have no analytic expression.

Appendix A1: analysis of height of cumuli

The results of the authors (Chatterjee et al., 1991) are summarised in Table A1 (hot season) and Table A2 (monsoon season).

Appendix A2: analysis of thickness of cirrus

The histograms published by the authors (Uttal et al., 1993) have been here cumulated and fitted with log-normal distributions (Fig. 11a) or Weibull distributions (Fig. 11b), for two kinds of data: WPL (Wave Propagation Laboratory at NOAA) and PSU (Pennsylvania State University). The fitting is satisfying in the interval $\log H \leq 0.6$, and better for WPL's data than for PSU's. Parameter values are listed in Table A3 for log-normal fitting and in Table A4 for Weibull fitting.

Appendix A3: errors on regression coefficients p and q

The parameters p and q are determined from the fitting of a straight line to the data

Table A1. Parameters of fitted log-normal distributions for the height H (km) of cumuli estimated from p and q (hot season)

Region	p	q	m	b
Bombay	10.92	−11.096	2.340	0.211
Calcutta	7.407	−6.600	2.052	0.311
Hyderabad	6.019	−5.037	1.927	0.383
Gauhati	7.668	−5.798	1.741	0.300
Delhi	5.394	−4.126	1.762	0.427

Table A2. Parameters of fitted log-normal distributions for the height H (km) of cumuli estimated from p and q (monsoon season)

Region	p	q	m	b
Bombay	9.6	−8.856	2.125	0.240
Calcutta	7.835	−6.481	1.905	0.294
Hyderabad	6.261	−4.769	1.754	0.368
Gauhati	8.052	−5.532	1.582	0.286
Delhi	5.586	−4.273	1.762	0.412

Table A3. Parameters of fitted log-normal distributions for the thickness H (km) of cirrus estimated from p and q

Data	p	q	m	b
WPL	2.530	−0.721	0.656	0.910
PSU	3.229	−0.695	0.496	0.713

Table A4. Parameters of fitted Weibull distributions for the thickness H (km) of cirrus estimated from p and q

Data	p	q	a	b
WPL	1.049	−0.844	1.049	2.879
PSU	1.267	−0.821	1.267	3.621

$$z = px + q \tag{A1}$$

by minimizing the cost function

$$S = \frac{1}{N}\sum_{i=1}^{N}(z_i - px_i - q)^2 \tag{A2}$$

that is to say by solving the system

$$\begin{cases} \frac{\partial S}{\partial p} = 2(\bar{X^2}p + \bar{X}q - \overline{XZ}) = 0 \\ \frac{\partial S}{\partial q} = 2(\bar{X}_{p+q} - \bar{Z}) = 0 \end{cases} \tag{A3}$$

to yield the following solutions provided $\sigma_X \neq 0$

$$\begin{cases} p = \frac{\overline{XZ} - \bar{X}\bar{Z}}{\sigma_x^2} \\ q = \frac{\bar{X^2}\bar{Z} - \bar{X}\,\overline{XZ}}{\sigma_x^2} \end{cases} . \tag{A4}$$

A variational calculus on these expressions, with the assumption that $\delta x_i \equiv \delta x$ and $\delta z_i \equiv \delta z$, leads to error estimates

$$\begin{cases} \delta p = \frac{2}{\sigma_x^2}[|\bar{Z} - 2p\bar{X}|\delta X + |\bar{X}|\delta Z] \\ \delta q = \left|p + 4\frac{q\bar{X}}{\sigma_x^2}\right|\delta X + \delta Z \end{cases} . \tag{A5}$$

Appendix A4: cross-correlation coefficients

The cross-correlation coefficient of two random variables X and Y may be defined by Eq. (21), or

$$\rho_{XY} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_X \sigma_Y} \tag{A6}$$

with

$$\begin{cases} \sigma_X = \sqrt{\overline{(X - \bar{X})^2}} = \sqrt{\bar{X^2} - \bar{X}^2} \\ \sigma_Y = \sqrt{\overline{(Y - \bar{Y})^2}} = \sqrt{\bar{Y^2} - \bar{Y}^2} \end{cases} . \tag{A7}$$

Now, if we display the data as a (X, Y)-plot, and suppose a linear relationship between X and Y

$$Y = aX + B \tag{A8}$$

where a is a constant and B represents a random variable which may be correlated with X or not. Then we have

$$\begin{cases} \bar{Y} = a\bar{X} + \bar{B} \\ \overline{XY} = a\bar{X^2} + \overline{BX} = a(\sigma_X^2 + \bar{X}^2) + \bar{X}\bar{B} + \rho_{XB}\sigma_X\sigma_B \\ \sigma_Y = \sqrt{\overline{\{a(X - \bar{X}) + (B - \bar{B})\}^2}} \\ \quad = \sqrt{a^2\sigma_X^2 + \sigma_B^2 + 2a\rho_{XB}\sigma_X\sigma_B} \end{cases} \tag{A9}$$

where the correlation coefficient ρ_{XB} between X and B is defined as in Eq. (A6).

Therefore, by inserting Eq. (A9) into Eq. (A6) we get

$$\rho_{XY} = \frac{a\sigma_X + \rho_{XB}\sigma_B}{\sqrt{a^2\sigma_X^2 + \sigma_B^2 + 2a\rho_{XB}\sigma_X\sigma_B}} . \tag{A10}$$

Then if X and B are perfectly uncorrelated ($\rho_{XB} = 0$), Eq. (A10) simplifies into

$$\rho_{XY} = \frac{a\sigma_X}{\sqrt{a^2\sigma_X^2 + \sigma_B^2}} . \tag{A11}$$

This latter expression clearly shows that $|\rho_{XY}|$ is smaller than 1, and that the smaller σ_B is, the larger ρ_{XY} will be. Moreover, the sign of ρ_{XY} is the same as that one of a, and it reflects the trend of the slope of the regression line which can be derived from Eq. (A8) by least-squares fitting. Let us set the equation of such a line to be

$$y = \alpha x + \beta . \tag{A12}$$

and define the cost function as

$$S = \frac{1}{N}\sum_{i=1}^{N}(y_i - \alpha x_i - \beta)^2 \tag{A13}$$

that is to say, because of Eq. (A12)

$$S = \frac{1}{N}\sum_{i=1}^{N}[(a - \alpha)x_i + b_i - \beta]^2 . \tag{A14}$$

The coefficients α and β are computed from the minimum of S at the point (a, b)

$$\begin{cases} \frac{\partial S}{\partial \alpha} = 2\left[\bar{X^2}(\alpha - a) + \bar{X}(\beta - \bar{B})\right] = 0 \\ \frac{\partial S}{\partial \beta} = 2[\bar{X}(\alpha - a) + \beta - \bar{B}] = 0 \end{cases} . \tag{A15}$$

This system has a non-zero determinant simply equal to σ_X^2, and therefore it admits the solution

$$\alpha = a \qquad \beta = \bar{B} . \tag{A16}$$

The equation of the regression line and the corresponding value of the cost function are finally

$$y = ax + \bar{B} \tag{A17}$$

and

$$S = \sigma_B^2 . \tag{A18}$$

From Eq. (A11) it can be noticed that whenever the dispersion is negligible ($\sigma_B \ll a\sigma_X$) then ρ_{XY} is close to 1, and on the opposite whenever the dispersion is important ($\sigma_B \gg a\sigma_X$) then the correlation coefficient writes $\rho_{XY} = a\sigma_X/\sigma_B$.

This simple model shows a direct connection between the dispersion of data about the regression line and the correlation coefficient.

References

Abramowitz, M., and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs and mathematical tables*, National Bureau of Standards, Applied Math. Series, **55**, 1972.

Ackerman, B., The variability of the water contents of tropical cumuli, *J. Meteorol.*, **16**, 191–198, 1959.

Aitchison, J., and J. A. C. Brown, *The lognormal distribution, with special reference to its uses in economics*, Cambridge University Press, Cambridge, 1973.

Anjyo, K.-I., Semi-globalization of stochastic spectral synthesis, *Vis. Comp.*, **7**, 1–12, 1991.

Ansmann, A., J. Bösenberg, G. Brogniez, S. Elouragini, P. H. Flamant, K. Klapheck, H. Linn, L. Menenger, W. Michaelis, M. Riesebell, C. Senff, P.-Y. Thro, U. Wandinger, and C. Weitkamp, Lidar network observations of cirrus morphological and scattering properties during the International Cirrus Experiment 1989: the 18 October 1989 Case study and statistical analysis, *J. Appl. Meteorol.*, **32**, 1608–1622, 1993.

Bagby, R. J., Calculating normal probabilities, *Am. Math. Mon.*, **102**(1), 46–49, 1995.

Bankert, R. L., Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network, *J. Appl. Meteorol*, **33**, 909–918, 1994.

Bischoff, K., and B. Chytil, A note on scintillation indices, *Planet. Space Sci.*, **17**, 1059–1066, 1969.

Blacknell, D., Comparison of parameter estimators for K-distribution, *IEE Proc. Radar Sonar Navig.*, **141**(1), 45–52, 1994.

Cahalan, R. F., and J. B. Snider, Inhomogeneities of stratocumulus liquid water. Fire Science Results 1989, Ed. D. S. McDougal, *NASA CP-3079*, 21–25, 1990.

Chatterjee, R. N., P. Prakash, and K. Ali, Frequency distributions of the heights of convective clouds in different regions in India, *Atmos. Res.*, **26**, 445–454, 1991.

Coakley, J. A., Jr, and J. B. Snider, Observed reflectivities and liquid water content for marine stratocumulus, *Proc. AMS*

Symp. on the role of clouds in atmospheric chemistry and global climate 30 Jan.–3 Feb., Anaheim, CA, 175–177, 1989.

Desbois, M., G. Sèze, and G. Szejwach, Automatic classification of clouds on METEOSAT imagery: application to high-level clouds, *J. Appl. Meteorol.*, **21,** 401–412, 1982.

Gayet, J.-F., V. Giraud, A. Delannoy, and H. Larsen, Description of a French cloud microphysical data bank: application to global change, *Proc. WMO Workshop on Cloud Microphysics and Applications to Global Change*, 10–14 Aug. 1992, Toronto, Canada, 188–193, 1992.

Heymsfield, A. J., and C. M. R. Platt, A parameterization of the particle size spectrum of ice clouds in terms of the ambient temperature and ice water content, *J. Atmos. Sci.*, **41,** 846–855, 1984.

Houze, R. A., and C. P. Cheng, Radar characteristics of tropical convection observed during GATE: mean properties and trends over the summer season, *Mon. Weather Rev.*, **105,** 964–980, 1977.

Inoue, T., A cloud type classification with NOAA 7 split-window measurements, *J. Geophys. Res.*, **92**(D4), 3991–4000, 1987.

Jakeman, E., Heirarchical scattering models, scattering and propagation in random media, *AGARD CPP-419*, **13**/1–11, 1987.

Jeck, R., A new data base of ice particle size spectra for altitudes up to 10 km, *Proc. 10th Int. Cloud Physics Conf.*, Aug. 15–20, 1988, Bad Homburg, F.R.G., 1988.

Jennane, R., R. Harba, and G. Jacquet, Synthèse précise d'images fractales, *16th symposium GRETSI*, 15–19 Sept. 1997, Grenoble, 487–490, 1997.

Johnson, N. L., and S. Kotz, *Distributions in statistics. Continuous univariate distributions* – vol 1, John Wiley – vol 2. Houghton Mifflin, 1970.

Joseph, J. H., The morphology of fair weather cumulus cloud fields as remotely-sensed from satellites and some applications, *Adv. Space Res.*, **5**(6), 213–216, 1985.

Kayano, K., and K. Shimizu, Optimal thresholds for a mixture of lognormal distributions as the continuous part of the mixed distribution, *J. Appl. Meteorol.*, **33,** 1543–1550, 1994.

Larsen, H., J.-F. Gayet, G. Febvre, H. Chepfer, and G. Brogniez, Measurement errors in cirrus cloud microphysical properties, *Ann. Geophysicae.*, **16,** 266–276, 1998.

Lévesque, M. P., Generation of IR sky background images, *SPIE*, **1486,** 200–209, 1991.

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

Lopez, R. E., Radar characteristics of the cloud populations of tropical disturbances in the northwest Atlantic, *Mon. Weather Rev.*, **104,** 268–283, 1976.

Lopez, R. E., The lognormal distribution of cumulus cloud populations, *Mon. Weather Rev.*, **105,** 865–872, 1977.

Mandelbrot, B. B., *The fractal geometry of nature*, W. H. Freeman, San Francisco, 1982.

Mazin, I. P., Cloud water content in continental clouds of middle latitudes, *Atmos. Res.*, **35,** 283–297, 1995.

Minnis, P., C. W. Fairall, and D. F. Young, Intercomparisons of GOES-derived cloud parameters and surface observations over San Nicolas Island, Fire Science Results 1989, Ed. D. S. McDougal, *NASA CP-3079*, 27–31, 1990.

Nakagami, M., *The m-distribution – A general formula of intensity distribution of rapid fading*, Statistical Methods in Radio Wave Propagation, Ed. W. G. Hoffman, Pergamon, Oxford, 1960.

Plank, V. G., The size distribution of cumulus clouds in representative Florida populations, *J. Appl. Meteorol.*, **8,** 46–67, 1969.

Pruppacher, H. R., and J. D. Klett, *Microphysics of clouds and precipitation*, D. Reidel, Dordrecht, 1980.

Rossow, W. B., and R. A. Schiffer, I.S.C.C.P. data products, *Bull. Am. Meteorol. Soc.*, **72**(1), 2–20, 1991.

Roulleau, J., and R. Trochon, *Météorologie générale, vol. 2* (*Stabilité verticale de l'atmosphére. Vent et turbulence atmosphérique. Nuages et précipitations*), Gauthier-Villars, Paris, 1958.

Su, B., and G. C. Pomraning, A stochastic description of a broken cloud field, *J. Atmos. Sci.*, **51**(13), 1969–1977, 1994.

Sun, Z., and K. P. Shine, Studies of the radiative properties of ice and mixed-phase clouds, *Q. J. R. Meteorol. Soc.*, **120,** 111–137, 1994.

Tessendorf, J., D. Weston, and L. Taylor, Structure of spatial spectra of simulated cloud scenes at infrared wavelength, *SPIE*, **1687,** 499–507, 1992.

Uttal, T., S. M. Shaver, E. E. Clothiaux, and T. P. Ackerman, Cloud boundaries during FIRE II, *26th Int. Conf. on Radar Meteorol.*, 24–28 May 1993, Norman, OK, American Meteorological Society, 573–575, 1993.

Warner, C., and R. H. Grumm, Cloud distributions in a Bay-of-Bengal monsoon depression, *Mon. Weather Rev.*, **112,** 153–172, 1984.