QUARTERLY JOURNAL
OF THE
ROYAL METEOROLOGICAL SOCIETY

Vol. 128 APRIL 2002 Part B No. 582

Q. J. R. Meteorol. Soc. (2002), **128**, pp. 1043–1057

Towards quantifying droplet clustering in clouds

By R. A. SHAW*, A. B. KOSTINSKI and M. L. LARSEN
Michigan Technological University, USA

(Received 30 July 2001; revised 10 January 2002)

SUMMARY

Droplet positions in atmospheric clouds are random but possibly correlated on some scales. This 'clustering' must be quantified in order to account for it in theories of cloud evolution and radiative transfer. Tools as varied as droplet concentration power spectrum, Fishing test, and fractal correlation analysis have been used to describe the small-scale nature of clouds, and it has been difficult to compare conclusions systematically. Here we show, by using the correlation-fluctuation theorem and the Wiener–Khinchin theorem, that all of these measures can be related to the pair-correlation function. It is argued that the pair-correlation function is ideal for quantifying droplet clustering because it contains no scale memory and because of its quantitative link to the Poisson process.

KEYWORDS: Pair-correlation function Poisson process

1. INTRODUCTION AND BACKGROUND

Research activity devoted to the spatial distribution of cloud droplets at small (mm to m) scales shows no signs of subsiding (for a review see Vaillancourt and Yau 2000). Interest in this topic is driven by the importance of the droplet spatial distribution in cloud processes such as droplet growth by condensation, collision–coalescence, and radiative transfer. For example, the theory of droplet collision rates in clouds is based on the assumption that droplets are distributed in space in a random and uncorrelated fashion (e.g. Saffman and Turner 1956). Similarly the Beer–Lambert law, fundamental to the theory of radiative transfer, is based on the same assumption (e.g. Kostinski 2001).

Recent progress in this field has been facilitated by at least two factors: the development of new instruments capable of resolving small-scale features in clouds from research aircraft (e.g. Chaumat and Brenguier 2001; Gerber *et al.* 2001), and the introduction of theoretical tools suitable for describing or quantifying the spatial structure of clouds (e.g. Baker 1992; Davis *et al.* 1999; Jaczewski and Malinowski 2000; Kostinski and Jameson 2000; Pinsky and Khain 2001). It is on the second point that we wish to focus in this paper. More specifically, our goal is to show that many of the tools used for quantifying the spatial distribution of droplets are closely related, and can be linked to fundamental probability theory through the pair-correlation function. To that end, we restate the importance of examining the small-scale structure of clouds from the perspective of random processes, attempt to clarify some of the essential notions and tools used in the field, derive various relations between commonly used clustering signatures and provide illustrations using both data and simulations.

* Corresponding author: Department of Physics, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA. e-mail: rashaw@mtu.edu

(*a*) *Clustering described*

Difficulties sometimes arise over differences in terminology used when describing the nature of the small-scale structure of clouds. Therefore we will attempt to clarify the terminology of the theory of random processes used in this paper. To begin with, we adopt the view that the position of a cloud droplet is a random variable. Given that most clouds are turbulent, this follows the long-accepted tenet of turbulence studies that turbulent systems should be viewed from a stochastic perspective. Part of the reasoning for this is that it would be impossible to measure initial conditions with sufficient precision to predict the details of the evolution of a turbulent flow or, for that matter, the trajectories of individual cloud droplets—and even if it were possible, we would have little reason to know such details.

While perhaps not as widely accepted in the cloud physics community, it seems reasonable that the small-scale features of a cloud (e.g. metre-scales and below) must be viewed in the same way, essentially as random processes governed by various probability laws. A crucial aspect of this problem is that randomness of the variable under consideration, whether it is the position of molecules in an ideal gas or velocity in a turbulent fluid, *does not preclude the existence of correlations*; certainly, for example, the existence of spatial and temporal correlations in the (random) velocity field is a defining characteristic of turbulence. Any variable which cannot be predicted is considered here to be a random variable.

In the light of these statements, we will use the following definitions (several of these refer to the number of droplets in a volume, which is a countable random variable not typically encountered in turbulence studies).

1. Randomness is understood to mean, qualitatively, 'unpredictable' and does not imply a lack of correlations or, in the case of numbers of cloud particles in volume bins, randomness does not have to imply that the random variable is Poisson-distributed.
2. We will use the term 'perfectly random' for describing a random, countable variable that is described by a Poisson probability density function (pdf) with no correlations on any scale (the Poisson process is discussed below).
3. The word 'homogeneous' is taken to mean statistical homogeneity or, roughly, that the statistical properties of the data do not change with position (the temporal analogue being statistical stationarity). While it has been common to use the term 'inhomogeneous' when describing clouds containing 'clustered' (not perfectly random) droplet distributions, we feel that this usage is best avoided for reasons detailed below.

(*b*) *Clustering quantified*

Figure 1 is a schematic illustration of the three most important models of a counting random process and of the notion of the pair-correlation function. Figure 1(a) provides the standard of perfect randomness for dilute systems of particles. It is 'perfect' because the following conditions are ensured: (i) all particle positions are uniformly and identically distributed random variables; (ii) the process is statistically homogeneous so that statistical moments are independent of the choice of origin; (iii) particle positions are statistically independent of each other. These conditions define the Poisson process whose statistics are characterized by the Poisson distribution on *any scale*. That is, the (random) number of particles $N = N(V)$, in a test volume V, is distributed according to:

$$p(N) = \frac{\overline{N}^N \exp(-\overline{N})}{N!}, \tag{1}$$

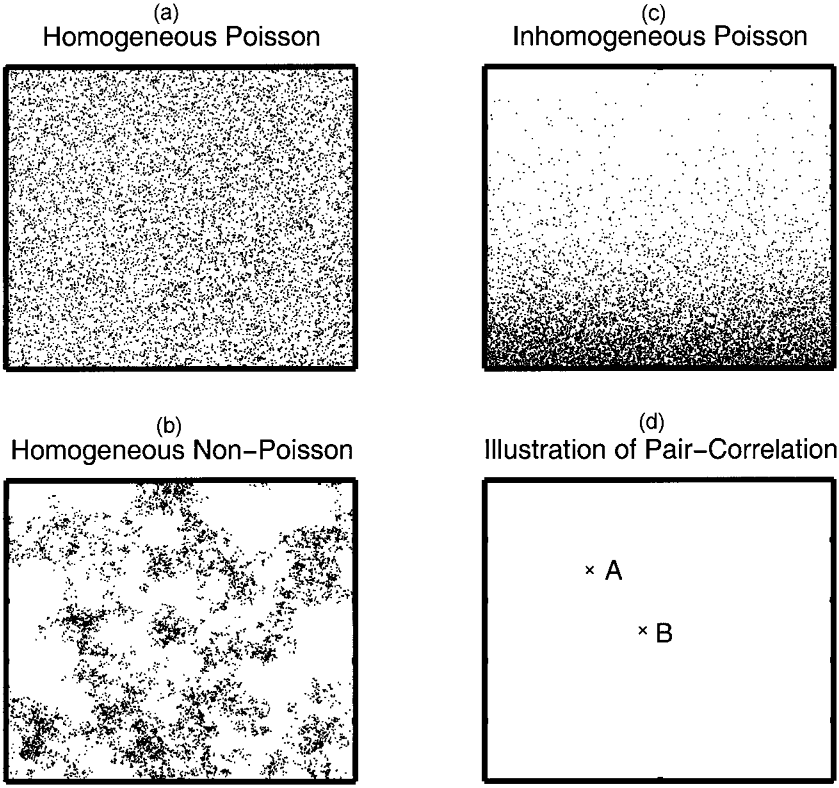

Figure 1. (a) Statistically homogeneous Poisson process. Particle positions are uniformly, identically and independently distributed random variables. This is the 'perfect randomness' standard for dilute systems of particles. (b) Statistically homogeneous but spatially correlated (not Poisson) random process. Particle positions are uniformly distributed but not independent random variables. Clumps exist but their 'centres' are uniformly distributed random variables. (c) Statistically inhomogeneous Poisson process. Our example is an ideal gas of molecules in the gravitational field. Note that there are no clumps but the distribution parameter, number density ($\overline{N}$), is a *deterministic* (exponentially decreasing) function of height. (d) Pair-correlation function. Point A is placed completely at random (uniformly distributed anywhere in the box). However, the *conditional* probability for the position of point B is 'biased' by the (positive) pair-correlation with point A. This results in clumps as in (b).

where $\overline{N} = \overline{N(V)}$ is the mean number of particles in V. (Symbols are defined in the appendix.) It is important to note that the validity of the Poisson distribution on some spatial scale V does not imply the Poisson process. For example, positive and negative deviations from the Poisson process on smaller spatial-scales can cancel each other and result in Poisson statistics on longer scales as was shown by Kostinski and Shaw (2001).

In order to broaden our model we can either relax the statistical independence or the statistical stationarity assumptions. Let us discuss the resulting random processes, in turn. The statistically homogeneous but spatially correlated random process is depicted in Fig. 1(b). This is no longer a Poisson random-process, because lack of statistical independence in particle positions allows the formation of clumps (or voids) which, in turn, causes deviations from the Poisson distribution on spatial-scales comparable with the clump size. Particle positions are not independent random variables. Note that the positions of clump 'centres' are unpredictable—in fact, they are uniformly distributed random variables. The 'patchiness' is quantified by the pair-correlation function η, specifically defined as a deviation from perfect randomness (Poisson process; see Eq. (3)).

A statistically inhomogeneous Poisson process is illustrated in Fig. 1(c). For simplicity we choose an ideal gas comprising molecules in a gravitational field. Note that there are no clumps but, unlike the homogeneous process, the distribution parameter, number density ($\overline{N}$), is a *deterministic, hence predictable* (exponentially decreasing) function of height. Therefore, the particle count can still be given by the Poisson law but with a variable process rate λ:

$$p(N;\, L_1,\, L_2) = \frac{(\int_{L_1}^{L_2} \lambda(z)\, \mathrm{d}z)^N}{N!} \exp\left(- \int_{L_1}^{L_2} \lambda(z)\, \mathrm{d}z \right), \tag{2}$$

where z is, for example, the height between L_1 and L_2, while $\overline{N} = \int_{L_1}^{L_2} \lambda(z)\, \mathrm{d}z$. Note that this distribution still has the Poissonian property that the mean $\overline{N}$ and variance $\overline{(\delta N)^2}$ are equal.

In order to model cloud droplets via an inhomogeneous Poisson process, one has to treat the local mean number density as a deterministic function. This requirement applies even when attempting to model clouds as any statistically inhomogeneous random process (not necessarily Poisson). Insofar as the mean number density of cloud particles is not a predictable function, we choose the homogeneous correlated random-process approach. This is similar to the introduction of the concept of homogeneous isotropic turbulence where 'structure' functions are defined in terms of the turbulent velocity correlations.

Figure 1(d) is our attempt to introduce the key notion of a pair-correlation idea intuitively. The random process is statistically homogeneous (stationary) but correlated (not Poisson). Point A is placed completely at random (uniformly distributed anywhere in the box). However, the *conditional* probability for the position of point B is 'biased' by the pair-correlation with point A. If the bias (pair-correlation function) is positive, particle clumps result as in Fig. 1(b). Physically, this could be due to the fact that both points belong to the same vortex. If the bias is negative ('repulsion' or 'dead space' artifacts), the resulting particle distribution is more uniform than perfect randomness.

Based on the preceding discussion, we will assume that droplets are distributed in a cloud in a random fashion—understanding, of course, that randomness does not necessarily imply the absence of correlations. As mentioned before, 'perfect randomness' will be defined as randomness without correlations at any scale. This lack of correlations can be considered from at least three perspectives. First, for randomly spaced, uncorrelated droplets, the probability of any given droplet being located at a particular position in the cloud volume is constant throughout the volume (uniform pdf). Second, the probability of finding a given number of droplets in a volume element is given by the Poisson distribution. Third, when projected into one dimension, the inter-droplet spacing is exponentially distributed. In principle, any of these three equivalent statements can be used to 'test' the perfect randomness of a cloud. The difference is in the choice of the random variable: particle position, number of droplets in a volume, or droplet spacing.

At this point, one can already see flaws in the assumption that droplets are distributed in space in a perfectly random manner. For example, because of the finite size of droplets it is clear that the inter-droplet separation distance can never be less than one droplet diameter, implying that, strictly speaking, projected separation distances cannot be exponentially distributed (Kostinski and Shaw 2001). This observation serves as motivation for quantifying the randomness of a cloud as a function of spatial-scale. In other words we know, because of finite size effects, that clouds are not perfectly random at spatial-scales less than one droplet diameter; but we would still like to know whether they are perfectly random at mm or cm scales.

We consider a volume element $\mathrm{d}V$ sufficiently small so that it can contain only one droplet, and so that the probability of finding a droplet in that volume is $\overline{n}\,\mathrm{d}V$, where $\overline{n}$ is the mean droplet number density. If droplet positions are uncorrelated then we would expect that the probability of finding droplets in each of two volume elements $\mathrm{d}V_1$ and $\mathrm{d}V_2$ separated by distance r is $P_r(1, 2) = (\overline{n}\,\mathrm{d}V_1)(\overline{n}\,\mathrm{d}V_2)$. If, however, the probabilities are not independent we define the pair-correlation function $\eta(r)$ as (e.g. Green 1969, chapter 2):

$$P_r(1, 2) = (\overline{n}\,\mathrm{d}V_1)(\overline{n}\,\mathrm{d}V_2)\{1 + \eta(r)\}. \tag{3}$$

The pair-correlation function is zero for perfect randomness and has a lower limit of -1, e.g. for scales less than the diameter of impenetrable particles. If the pair-correlation function is greater than zero it implies that if a droplet is encountered at a given position in a cloud, there is an enhanced probability of finding another droplet distance r away. It should also be mentioned that in the fluid mechanics literature, the pair-correlation function often is referred to as the 'radial distribution function', $g(r)$. More precisely, $\eta(r) = g(r) - 1$, so that $g(r) = 1$ when no correlations are present (Sundaram and Collins 1997).

2. The pair-correlation function and other measures of clustering

The commonly used tools for describing the small-scale structure of clouds are: the Fishing test or clustering index (Baker 1992; Uhlig *et al.* 1998; Chaumat and Brenguier 2001; etc.); the correlation dimension fractal analysis (Grits *et al.* 2000; Jaczewski and Malinowski 2000); and the power spectrum of droplet concentration or liquid water content (Davis *et al.* 1999; Jeffery 2000; Gerber *et al.* 2001; Pinsky and Khain 2001). Two fundamental theorems of mathematical physics allow us to link quantitatively each of these measures to the pair-correlation function. These quantities are related as in the following schematic:

$$\left.\begin{matrix} CI \\ \\ \overline{\eta} \end{matrix}\right\} \Leftarrow \eta \Longleftrightarrow \rho \Rightarrow \left\{\begin{matrix} S \\ \\ P \end{matrix}\right., \tag{4}$$

where CI is the clustering index, $\overline{\eta}$ is the volume-averaged pair-correlation function, ρ is the autocorrelation function, S is the structure function, and P is the power spectrum of droplet concentration. As will be discussed, the clustering index is calculated from η using the Ornstein–Zernike equation, also referred to as the correlation-fluctuation theorem (Kostinski and Jameson 2000; Kostinski and Shaw 2001). Also, η is directly linked to the power spectrum of droplet concentration via the Wiener–Khinchin theorem as we show below.

(*a*) *Clustering index*

It is a property of the Poisson distribution that the variance is equal to the mean. Therefore it is natural to define CI as the variance-to-mean ratio as has been done in several fields of science and was introduced to cloud physics by Baker (1992). Hence,

$$CI(V) = \frac{\overline{(\delta N)^2}}{\overline{N}} - 1, \tag{5}$$

where N is the random number of droplets in a volume, $\overline{N}$ is the mean number of droplets, and $\overline{(\delta N)^2} \equiv \overline{(N - \overline{N})^2}$ is the variance. CI is zero for the Poisson distribution (for any given test volume V) and becomes positive when positive spatial correlations are dominant in V. It has been used by several groups to quantify the degree of droplet clustering in clouds (Baker 1992; Borrmann *et al.* 1993; Uhlig *et al.* 1998; Vaillancourt *et al.* 2000; Chaumat and Brenguier 2001). For completeness we note that the Fishing test, F, is a scaled estimator of *CI* (see Baker 1992 for details); only *CI* will be considered in this paper because the two are closely related.

An important point to be made here is that one must interpret the clustering index with caution, because of its inherent volume dependence it is not uniquely related to a single spatial-scale, but instead represents contributions from a range of scales. Indeed, the explicit link between *CI* and the scale-localized pair-correlation function is given by the correlation-fluctuation theorem (Landau and Lifshitz 1980, section 116)

$$\frac{\overline{(\delta N)^2}}{\overline{N}} - 1 = \overline{n} \int_0^V \eta(V')\,\mathrm{d}V'. \tag{6}$$

Hence, the clustering index is said to contain 'memory' of all scales within volume V (Kostinski and Shaw 2001). However, Eq. (6) gives a clear approach for obtaining a scale-localizable measure of droplet clustering from *CI*.

(*b*) *Volume-averaged pair-correlation function*

The volume-averaged pair-correlation function is defined as

$$\overline{\eta} = \frac{1}{V}\int_0^V \eta(V')\,\mathrm{d}V' = \frac{\overline{(\delta N)^2}}{\overline{N}^2} - \frac{1}{\overline{N}} = \frac{CI}{\overline{N}}, \tag{7}$$

where $\overline{N}$, $\overline{(\delta N)^2}$, CI, and $\overline{\eta}$ are all volume dependent. As with CI, $\overline{\eta}$ is straightforward to calculate, for example, simply by calculating $\overline{N}$ and $\overline{(\delta N)^2}$. However, it has the added advantage of possessing greatly reduced scale memory (or greater localization) relative to *CI*, due to the weighting of *CI* by $\overline{N}(V)$ in Eq. (7).

An example relevant to cloud physics illustrates the difference in scale memory between *CI* and $\overline{\eta}$ (see Kostinski and Shaw 2001 for additional details). We consider a perfectly random distribution of cloud droplets, except that the droplets possess some finite size, say diameter D. For this scenario the pair-correlation function is $\eta(r) = -1$ for $r \le r_\circ$ and $\eta(r) = 0$ for $r > r_\circ$ (assuming the system is dilute). Using Eq. (7) we see that for $r > r_\circ$, $\overline{\eta(r)} = -(r_\circ/r)^3$ and $CI(r) = -\overline{n}(4/3)\pi r_\circ^3$. Clearly, the memory of droplet finite size in $CI(r)$ persists for all scales, whereas this memory dies out as r^{-3} for $\overline{\eta(r)}$.

Another important consideration is that $\overline{\eta}$ can be related to other approaches for quantifying droplet clustering in clouds, as will be explained in the following section.

(*c*) *Correlation dimension and fractal analysis*

Jaczewski and Malinowski (2000) have analysed two-dimensional (2-D) droplet spacing data by calculating an average concentration, defined as the average number of droplets in a circle of radius r, with the condition that the circle is centred on a droplet: $C(r) = \overline{N(r)}/\pi r^2$. The data are assumed to scale as $C(r) \sim r^{D-2}$, where D is called the 'correlation dimension' and is used to quantify the degree of droplet clustering. Grits *et al.* (2000) have used a similar approach, centring circles on a particle, counting the number of neighbouring droplets falling in the circle, etc. These measures of droplet clustering have a clear physical interpretation in the context of the pair-correlation function and its fundamental definition. For example, if a volume V is chosen at random we would expect to find $\overline{N} = \overline{n}V$ particles in the volume. If, however, the volume is *conditioned* to be centred on a particle we would expect the number of other particles in the volume to be:

$$\overline{N}_{\mathrm{p}} = \overline{n}V + \overline{n}\int_0^V \eta(V')\,\mathrm{d}V', \tag{8}$$

where the additional term on the right side of Eq. (8) is a result of correlations in droplet positions. We note that the correlation term can be positive or negative, so that $\overline{N}_{\mathrm{p}}$ can be greater or less than $\overline{N}$.

Adopting the 2-D geometry of Jaczewski and Malinowski (2000) and Grits *et al.* (2000) and assuming the collection of droplets is distributed in a random and uncorrelated fashion, we would expect that $\overline{N(r)} = \overline{n}\pi r^2$, where $\overline{n}$ is the 2-D mean number density of droplets. If, however, correlations exist at scales less than or equal to r, then $\overline{N(r)}$ will be enhanced. This enhancement can be quantified since $C(r)$ is related to the *conditional* probability of finding a droplet at distance r from an existing droplet, which, from Eq. (3) is $P_r(1|2) = \overline{n}\,\mathrm{d}V\{1 + \eta(r)\}$. This must be integrated because *all* droplets between 0 and r are counted, so it follows that $C(r)$ can be written as:

$$C(r) = \overline{n}\left(1 + \frac{1}{\pi r^2}\int_0^r \eta(r')2\pi r'\,\mathrm{d}r'\right) = \overline{n}(1 + \overline{\eta(r)}). \tag{9}$$

Clearly, the conditional concentration $C(r)$ is closely related to the volume-averaged pair-correlation function, which has been shown to be a useful approximation to η, in the sense that it is has less scale memory than the clustering index or Fishing test (Kostinski and Shaw 2001). In fact, it is possible to generalize this statement based on the power law assumption of Jaczewski and Malinowski (2000). To do this, let us consider a population of droplets that obeys the power law $\eta(r) = cr^{-\alpha}$ at scales from r down to the finite size of the droplets, which will be taken as $r_\circ$ (this scale being equal to the diameter of the droplets). For $r < r_\circ$ we know that $\eta(r) = -1$. Let us also take the 1-D case for simplicity and calculate $\overline{\eta(r)}$, which will allow us to calculate $C(r)$ using Eq. (9). The integration yields:

$$\overline{\eta(r)} = -\frac{c}{\alpha - 1}r^{-\alpha} - \frac{r_\circ}{r}\left(1 - \frac{c}{\alpha - 1}r_\circ^{-\alpha}\right). \tag{10}$$

Using Eq. (10) it can be shown that $\overline{\eta(r)}$ has the same slope (in log–log space) as $\eta(r)$ *only* for $\alpha < 1$. For $\alpha > 1$ the memory of droplet finite size ($r_\circ$) dominates the power law exponent of $\overline{\eta(r)}$ at *all scales*. The implication, therefore, is that the power law exponent of $\overline{\eta(r)}$, and therefore $C(r)$, must be interpreted carefully because it can have completely different physical causes. In one regime it closely resembles the actual slope of $\eta(r)$ and in another regime it is dominated at all scales by the finite size of the droplets.

(d) *Autocorrelation function*

From the fundamental definition of $\eta(r)$ given in Eq. (3) it is possible to write:

$$\eta(r) = \frac{P_r(1, 2)}{(\overline{n}\,\mathrm{d}V)^2} - 1. \tag{11}$$

The joint probability $P_r(1, 2)$ can be calculated in practice by dividing a data series into volumes sufficiently small so as to contain only 1 or 0 droplets, the volumes being separated by a length r. For a 1-D data series the joint probability is calculated as $\overline{\{n(r_\circ)\,\mathrm{d}V\}\{n(r_\circ + r)\,\mathrm{d}V\}}$. Because $(\overline{n}\,\mathrm{d}V)^2 = \overline{N}^2$ we can write:

$$\eta(r) = (1/\overline{N}^2)\overline{N(r_\circ)N(r_\circ + r)} - 1, \tag{12}$$

which we may consider to be an 'operational definition'. This is a powerful result because it provides a direct link to the traditional autocorrelation function, defined as (e.g. Frisch 1995, chapter 4):

$$\rho(r) = \overline{n'(r_\circ)n'(r_\circ + r)}/\overline{n'^2}, \tag{13}$$

where $n' \equiv n - \overline{n}$ is the 'fluctuating' component. We note that, strictly speaking, $\rho(r)$ is defined for continuous variables. It is, therefore, quite a subtle step to translate the continuous language of $\rho(r)$ to the 'counting' language of $\eta(r)$. If lag $r = 0$ is ignored, however, this translation can be made via:

$$\eta(r) = \rho(r)\overline{n'^2}/\overline{n}^2, \tag{14}$$

keeping in mind that N in Eq. (12) is not the fluctuating component N' but, rather, is the full variable. Indeed, the -1 in Eq. (12) serves to remove the mean so that Eq. (14) holds.

(*e*) *Structure function*

Although to our knowledge it has not often been used in cloud physics, in Eq. (4) we included the structure function, S, because of its prominence in the field of turbulence and its connections to the power spectrum. Qualitatively, S characterizes self-coherence as a function of spatial-scale r. For example, S is used as a measure of the correlation in velocity or passive scalar fields in turbulent flows (e.g. Frisch 1995, chapter 5). The second-order structure function is defined as:

$$S(r) = \overline{\{n'(r_\circ) - n'(r_\circ + r)\}^2}, \tag{15}$$

and is directly related to the autocorrelation and pair-correlation functions via:

$$S(r) = 2\overline{n'^2}(1 - \rho(r)) = 2(\overline{n'^2} - \overline{n}^2\eta(r)). \tag{16}$$

Both $\rho(r)$ and $S(r)$ are related to the power spectrum, as discussed in the next section.

(*f*) *Power spectrum*

The Wiener–Khinchin theorem provides a quantitative link between the autocorrelation function of a data series and the power spectral density of the data series (Reif 1965, section 15.15):

$$P(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \rho(r)\,\mathrm{e}^{-\mathrm{i}kr}\,\mathrm{d}r. \tag{17}$$

It is possible, therefore, to relate the well-developed concepts of Fourier analysis to the pair-correlation function, via its links to the autocorrelation function and the structure function. For example, the 'enhanced variance' at small scales observed by Davis *et al.* (1999) and Gerber *et al.* (2001) may be related to the peak in the pair-correlation function observed at similar scales by Kostinski and Shaw (2001). A direct quantitative comparison of the data is not possible because Davis *et al.* (1999) and Gerber *et al.* (2001) used continuous liquid water content data while Kostinski and Shaw (2001) used discrete droplet counts. Another approach that can be interpreted in the context of Eq. (17) is the Fourier decomposition work of Pinsky and Khain (2001) using droplet concentration data. The authors break their time series into two regimes, separated by an arbitrary scale, with the large scales being removed by fitting and subtracting Fourier components. The remaining small-scale concentration fluctuations are analysed by calculating a power spectrum and then taking an inverse Fourier transform of this spectrum. Considering Eq. (17), it would appear that the end result is a calculation of the autocorrelation function of high-pass-filtered droplet concentration. Hence, the work of Pinsky and Khain (2001) is based on a fundamentally different point of view, that of an inhomogeneous Poisson process (Fig. 1(c)). On the other hand, our point of view is that of a homogeneous non-Poisson (correlated) process (Fig. 1(b)).

A specific example will serve to make the link between $\rho(r)$ and $P(k)$ explicit. Under certain conditions (e.g. Frisch 1995) it is possible to relate the functional form of the power spectrum for a certain range of wave numbers to the functional form of the autocorrelation function over a similar range of lags. If the power spectrum has the form k^{-n}, and $1 < n < 3$, then the autocorrelation function will scale as $1 - r^{n-1}$. For example, the autocorrelation function:

$$\rho(r) = c\, \mathrm{e}^{-\alpha|r|}, \tag{18}$$

(e.g. this kind of model might be justified in the context of correlating droplet velocity to local fluid velocity) has the Fourier transform pair:

$$P(k) = \frac{c}{\pi\alpha}\left(\frac{1}{1+(k/\alpha)^2}\right). \tag{19}$$

In the limit $(k/\alpha)^2 \gg 1$, Eq. (19) has a power law form, $P(k) \sim k^{-2} = k^{-6/3}$. In this limit it follows that $\alpha|r| \ll 1$ so that Eq. (18) has the form $\rho(r) \sim 1 - \alpha|r|$, as expected. It is possible, therefore, to extend the various power law exponents encountered in the study of scalars in turbulent flows, for example $-5/3$ for the inertial subrange, to the shape of the pair-correlation function (via the operational definition of η). Furthermore, we may speculate that an exponent of -1 for the power spectrum of droplet concentration, which might be expected for the viscous convective subrange due to the relatively small diffusivity of cloud droplets (e.g. Batchelor 1959; Jeffery 2000), will result in a *flat* $\eta(r)$ at that range of scales. (Note that one must consider an exponent that *approaches* -1, since -1 is outside the range of the approximation being used.) Indeed, data from a statistically homogeneous region in a cumulus cloud show that $\eta(r)$ is nearly flat at scales between 10^{-1} and 1 m. (See Figs. 3 and 4, described in section 3.) Below 10^{-1} m, however, there is significantly more clustering than would be expected; possible causes of this have been discussed elsewhere (Kostinski and Shaw 2001).

3. Illustration

To illustrate the relationships described in the last section, we will examine simulated time series as well as actual cloud probe data. We begin with the notion of the pair-correlation function and then proceed to discuss other clustering signatures. In Fig. 2 we show an example of binned arrival positions of droplets detected by the Météo-France Fast Forward Scattering Spectrometer Probe (Fast FSSP) during the Small Cumulus Microphysics Study (SCMS). This probe records the arrival time of each cloud droplet that enters the detection region of the probe. This is converted to a spatial position by multiplying the elapsed time by the aircraft velocity to obtain a distance from the first

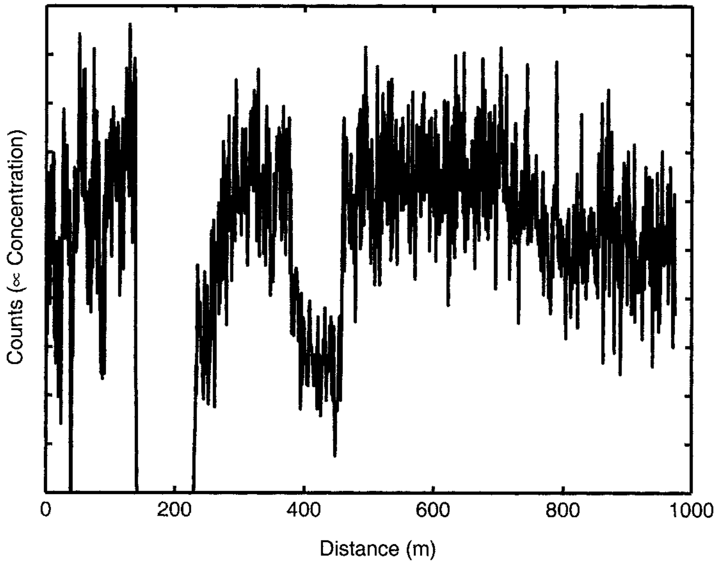

Figure 2. Droplet counts for a traverse of a cumulus cloud. These measurements were taken by the Météo-France Fast Forward Scattering Spectrometer Probe during the Small Cumulus Microphysics Study.

droplet (with the relative separation known to within approximately 10 μm, based on the instrument clock speed).

The two curves in Fig. 3 correspond to the pair-correlation function computed for the 'cloud core' (lower curve) and for the entire traverse (upper curve). The former is the region between 500 and 700 m in Fig. 2 which is characterized by statistically homogeneous conditions (for details see Kostinski and Shaw (2001), and appendix A of Kostinski and Jameson (2000)).

The most notable feature of Fig. 3 is the insensitivity of the entire shape of the pair-correlation function to whether the cloud core or the entire traverse is used. The two curves are separated by a vertical shift but look similar otherwise, at least qualitatively (e.g. both curves display a sharp rise at small scales, a peak near 500 μm, and a gentle decay with increasing scale). The fact that small-scale clustering is not obscured by large-scale fluctuations (sometimes described as 'inhomogeneities' in the literature) is quite encouraging, and demonstrates robustness of the pair-correlation-function approach. We also note that eventually the pair-correlation function must approach zero for both curves.

Perhaps the vertical shift can be understood as follows. The pair-correlation function is a measure of departure from the Poisson process. Consequently, the clustering at fine scales is an enhancement in the fraction of particle pairs separated, say, by 300 μm, relative to the fraction expected for the Poisson process. For the homogeneous core data, the enhancement is negative (hardly any pairs are separated by 300 μm because of probe filtering—Kostinski and Shaw (2001)), hence the pair-correlation function is negative. However, placing this fact in the context of the entire traverse causes the positive shift because cloud 'holes' such as between 175 and 225 m are included, which greatly lowers the Poissonian expectation against which the pair-correlation function is measured. Thus, the notion is ensemble-dependent, and renders subtle the physical interpretation of actual numerical values. This reasoning holds for any scale in the figure and, therefore, accounts for the upward shift illustrated in Fig. 3.

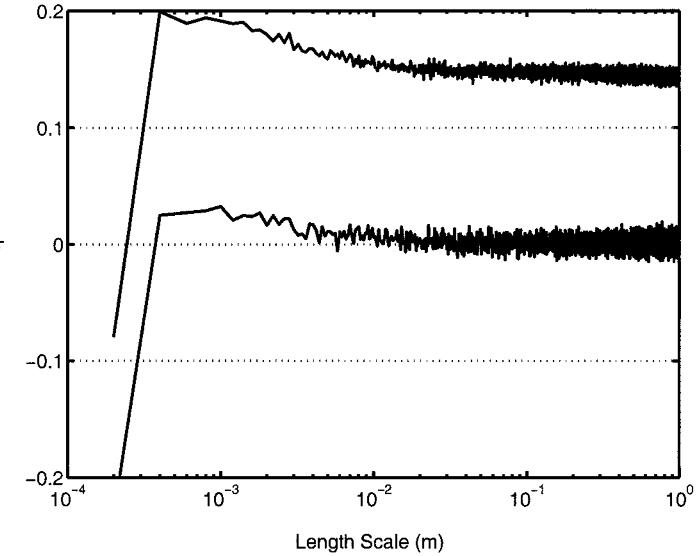

Figure 3. The pair-correlation function $\eta(r)$, calculated using the operational definition (see Eq. (12)). The top curve corresponds to all of the data shown in Fig. 2 and the bottom curve to data between approximately 500 and 700 m. The general shape of $\eta(r)$ is largely the same for the cloud core and for the entire traverse. The curve corresponding to all of the data is shifted up because the Poissonian expectation is lower when 'holes' such as that between 175 and 225 m in Fig. 2 are included in the series.

To illustrate η and its relationship to other measures of droplet clustering we have calculated several of the signatures discussed in section 2. In Fig. 4 we show plots of η, $\overline{\eta}$, CI, and P calculated for the entire cloud traverse shown in Fig. 2, and also for an 'equivalent Poisson' simulation having the same mean and number of droplets as the data. The variable ρ is not shown because it has the same shape as η, simply scaled by a constant (see Eq. (14)). The top graph in each of the four panels in Fig. 4 shows the results for the Fast FSSP data, while the bottom graph displays the same quantity calculated for a simulated Poisson process.

In Fig. 4(a), $\eta(r)$ corresponding to the Poisson process is zero at all scales, as expected. For the cloud traverse $\eta(r)$ displays a clear enhancement of droplet clustering at scales below several cm. The enhancement becomes stronger with decreasing scale until it is cut off due to the finite instrument resolution resulting from the FSSP optical design. Note also that at lags greater than 1 m, a slow linear decay is observed which, although not shown, eventually reaches zero at about 60 metres.

The volume-averaged pair-correlation function $\overline{\eta(r)}$ is shown in Fig. 4(b). The clustering can no longer be completely 'localized' or attributed to a given scale. Indeed, because of the volume averaging, it has memory of sub-scales including those affected by the probe filtering. This is reflected in the peak delay and slower (than η) descent. Note also that the negative tail can be fitted approximately with the simple model described in section 2.

In Fig. 4(c) the cumulative nature of CI is clearly seen. For example, it increases monotonically in the region where η is a positive but constant value. Comparison with Fig. 4(a) confirms that the relatively small magnitude of CI at small scales does not necessarily imply a lack of clustering at those scales. Similarly, the relatively large magnitude of CI at larger scales is a manifestation of clustering at smaller scales 'adding up'. As suggested earlier, $\overline{\eta}$ gives a much more accurate view of scale-dependent

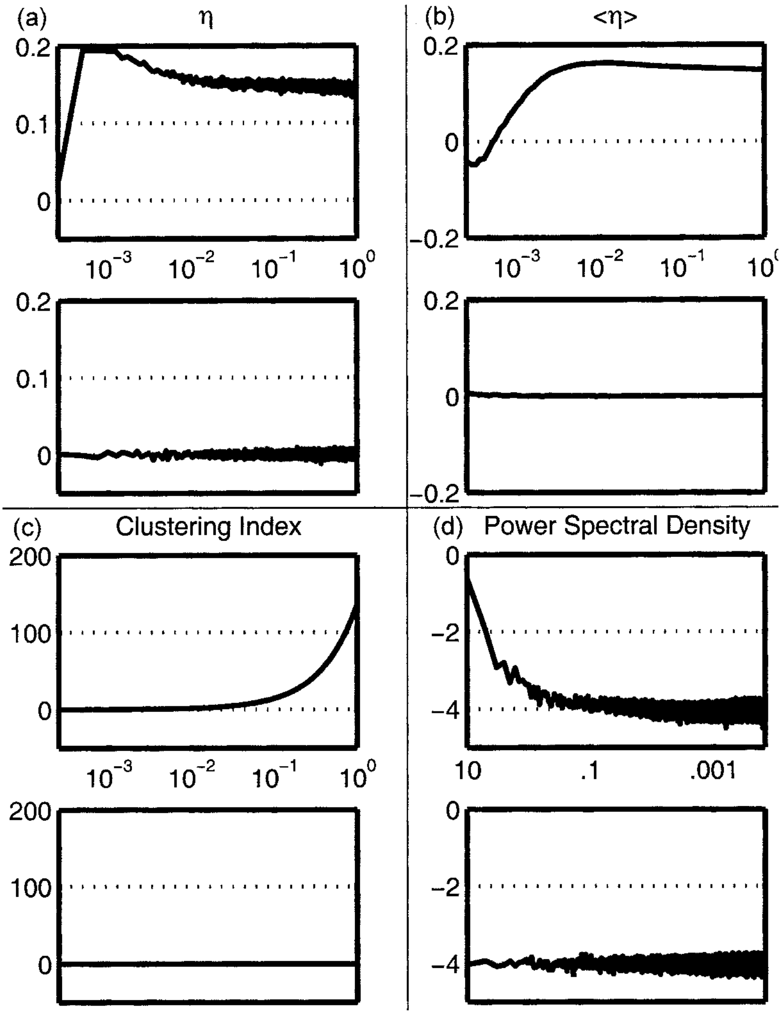

Figure 4. Illustration of the clustering signature calculations: pair-correlation function, η; its average, $\overline{\eta}$; the clustering index, *CI*; and power spectral density, *P*. Horizontal scales are given as distances (m). In the vertical scale of sub-plot (d) -4, -2, and 0 correspond to 10^{-4}, 10^{-2} and 1, respectively. In each of the four panels, the upper graph displays actual data while the lower graph displays the equivalent Poisson process (with the same mean and number of points as those of data): (a) shows the pair correlation function η calculated via the operational definition (Eq. (12)); (b) is $\overline{\eta}$; (c) is the clustering index; and (d) shows the power spectral density.

clustering than does *CI*. Fortunately, $\overline{\eta}$ can be obtained from *CI* simply by weighting it by $\overline{N}(V)$, leading to the greatly reduced scale memory (or greater localization) of $\overline{\eta}$.

In Fig. 4(d) it is clear that, as expected, the uncorrelated series have a 'white noise' (flat) spectrum. The cloud data are 'red', but there is no clear $-5/3$ slope. Spectral slopes, however, are quite sensitive to details of the data series, in particular the statistical homogeneity, so it is difficult to make solid conclusions based on limited data.

The plots in Fig. 4 illustrate the relative merits of various signatures of droplet clustering. Each variable, η, $\overline{\eta}$, *CI*, and *P*, contains scale-dependent information and its interpretation must be based on the theory underlying its definition. We would argue that for a countable variable such as droplets in clouds, the pair-correlation function provides the most natural, scale-localized, and physically meaningful measure of correlations.

4. Concluding remarks

We have attempted to show that seemingly independent tools used by various groups to study droplet clustering in clouds are related. The links between these tools are based on the correlation-fluctuation theorem and the Wiener–Khinchin theorem. In the context of these theorems we argue that the pair-correlation function has a particularly clear physical interpretation for a discrete system such as a cloud. This interpretation is based on the theory of counting random processes, and makes no ad hoc assumptions about the physical mechanisms possibly responsible for droplet clustering.

In one of the early papers which sparked an interest in this topic in the field of cloud physics, Baker (1992) wrote: “A useful method of analysis is one that can detect and characterize inhomogeneity of the droplet concentration by comparing the measured distributions with expected distributions”. The ‘inhomogeneity’ referred to is presumably not statistical inhomogeneity but rather the existence of correlations—the ‘clumpiness’ of the cloud. In fact, statistical homogeneity is *required* if the powerful mathematical tools discussed here are to be used in a meaningful way. Baker went on to introduce the clustering index to cloud physics, and quantified its statistical significance with the Fishing test.

As we have shown, the clustering index is actually a concentration-weighted integral of the pair-correlation function η. This is a two-point notion rather than an integrated quantity, so that difficulties related to scale memory are avoided and scale-dependent correlations can be localized. The pair-correlation function provides exactly what Baker alluded to: it quantifies clustering by comparing measured spatial distributions with a standard of perfect randomness and it does so in a scale-localizable manner. The conditional-probability approaches used by several groups can also be understood in terms of the volume-averaged pair-correlation function, and their scale memory can be evaluated thereby. Finally, the pair-correlation function is shown to have a fundamental link to the power spectrum of droplet concentration, allowing connections to be made to previous work in this area.

We have argued that the pair-correlation function η has certain properties that make it ideal for quantifying droplet clustering. These can be summarized as: (i) η satisfies the correlation-fluctuation theorem; (ii) η can be used to derive other commonly used measures of droplet clustering; (iii) η is scale localized; (iv) η can be easily interpreted based on the theory of random processes; and (v) η can be extended to the theory of radiative transfer, collision–coalescence, and rain characterization (e.g. Jameson and Kostinski 2000; Kostinski 2001; Shaw *et al.* 2002; Sundaram and Collins 1997).

Acknowledgements

We thank J.-L. Brenguier, Météo-France for providing the Fast FSSP data and D. Lanterman for helpful discussions. This material is based on work supported by the National Science Foundation under Grant Nos. ATM-0106271 (ABK) and ATM-9984294 (RAS).

Appendix

Key variables used

C	conditional concentration
CI	clustering index
F	Fishing test

g	radial distribution function
k	wave number
n	instantaneous droplet number density (random variable)
$\overline{n}$	mean droplet number density
n'	fluctuating component of n
N	number of droplets in volume V (random variable)
$\overline{N}$	mean number of droplets in V
$\overline{(\delta N)^2}$	variance in number of droplets in V
$p(N)$	probability density of droplet counts in V
P	power spectrum
$P_r(1, 2)$	joint probability of droplet separation by r
r	independent variable (length-scale) for all measures of clustering (η, ρ, etc.)
S	structure function
V	averaging volume associated with scale r
η	pair-correlation function
$\overline{\eta}$	volume-averaged pair-correlation function
ρ	autocorrelation function

Occasionally, measures of clustering (η, ρ, etc.) are written so as to make the functional dependence on length-scale, r, explicit (e.g. $\eta(r)$).

REFERENCES

Baker, B. A. 1992 Turbulent entrainment and mixing in clouds: A new observational approach. *J. Atmos. Sci.*, **49,** 387–404

Batchelor, G. K. 1959 Small-scale variation of convected quantities like temperature in a turbulent fluid. Part I: General discussion and the case of small conductivity. *J. Fluid Mech.*, **5,** 113–133

Borrmann, S., Jaenicke, R. and Neumann, P. 1993 On spatial distributions and inter-droplet distances measured in stratus clouds with in-line holography. *Atmos. Res.*, **29,** 229–245

Chaumat, L. and Brenguier, J. L. 2001 Droplet spectra broadening in cumulus clouds. Part II: Microscale droplet concentration heterogeneities. *J. Atmos. Sci.*, **58,** 642–654

Davis, A. B., Marshak, A., Gerber, H. and Wiscombe, W. J. 1999 Horizontal structure of marine boundary layer clouds from centimeter to kilometer scales. *J. Geophys. Res.*, **104,** 6123–6144

Frisch, U. 1995 *Turbulence: The legacy of A. N. Kolmogorov*. Cambridge University Press, Cambridge, UK

Gerber, H., Jensen, J. B., Davis, A. B., Marshak, A. and Wiscombe, W. J. 2001 Spectral density of cloud liquid water content at high frequencies. *J. Atmos. Sci.*, **58,** 497–503

Green, H. S. 1969 *The Molecular Theory of Fluids*. Dover Publications, New York, USA

Grits, B., Pinsky, M. and Khain, A. 2000 'Formation of small-scale droplet concentration inhomogeneity in a turbulent flow as seen from experiments with an isotropic turbulence model'. Pp. 138–141 in volume 1 of Proceedings of the 13th international conference on clouds and precipitation, Reno, NV, USA, 14–18 August 2000. Sponsored by the International Commission on Clouds and Precipitation of the International Association of Meteorology and Atmospheric Sciences

Jaczewski, A. and Malinowski, S. P. 2000 'Investigations of droplet clustering on small scales in a laboratory cloud—clear air mixing'. Pp. 134–137 in volume 1 of Proceedings of the 13th international conference on clouds and precipitation, Reno, NV, USA, 14–18 August 2000. Sponsored by the International Commission on Clouds and Precipitation of the International Association of Meteorology and Atmospheric Sciences

Jameson, A. R. and Kostinski, A. B. 2000 Fluctuation properties of precipitation. Part VI: Observations of hyperfine clustering and drop size distribution structures in three-dimensional rain. *J. Atmos. Sci.*, **57,** 373–388

Jeffery, C. A. 2000 Effect of particle inertia on the viscous-convective subrange. *Phys. Rev. E*, **61,** 6578–6585

Kostinski, A. B. 2001 On the extinction of radiation by a homogeneous but spatially correlated random medium. *J. Opt. Soc. Am.*, **A 18,** 1929–1933

Kostinski, A. B. and Jameson, A.R. 2000 On the spatial distribution of cloud particles. *J. Atmos. Sci.*, **57,** 901–915

Kostinski, A. B. and Shaw, R. A. 2001 Scale-dependent droplet clustering in turbulent clouds. *J. Fluid Mech.*, **434,** 389–398

Landau, L. D. and Lifshitz, E. M. 1980 *Statistical physics*. Butterworth Heinemann, Oxford, UK

Pinsky, M. and Khain, A. P. 2001 Fine structure of cloud droplet concentration as seen from the Fast-FSSP measurements. Part 1: Method of analysis and preliminary results. *J. Appl. Meteorol.*, **40,** 1515–1537

Reif, F. 1965 *Fundamentals of statistical and thermal physics*. McGraw-Hill, New York, USA

Saffman, P. G. and Turner, J. S. 1956 On the collision of drops in turbulent clouds. *J. Fluid Mech.*, **1,** 16–30

Shaw, R. A., Kostinski, A. B. and Lanterman, D. D. 2002 Super-exponential extinction of radiation in a negatively correlated random medium. *J. Quant. Spectrosc. Radiat. Transfer*, in press

Sundaram, S. and Collins, L. R. 1997 Collision statistics in an isotropic particle-laden turbulent suspension. Part 1. Direct numerical simulations. *J. Fluid Mech.*, **335,** 75–109

Uhlig, E., Borrmann, S. and Jaenicke, R. 1998 Holographic *in-situ* measurements of the spatial droplet distribution in stratiform clouds. *Tellus*, **50B,** 377–387

Vaillancourt, P. A. and Yau, M. K. 2000 Review of particle–turbulence interactions and consequences for cloud physics. *Bull. Am. Meteorol. Soc.*, **81,** 285–298

Vaillancourt, P. A., Yau, M. K. and Grabowski, W. W. 2000 'Preferential concentration and growth of cloud droplets'. Pp. 155–188 in volume 1 of Proceedings of the 13th international conference on clouds and precipitation, Reno, NV, USA, 14–18 August 2000. Sponsored by the International Commission on Clouds and Precipitation of the International Association of Meteorology and Atmospheric Sciences