2007

Spatial distributions of aerosol particles: Investigation of the Poisson assumption

Michael L. Larsen
University of Nebraska at Kearney, LarsenML@gmail.com

Available online at www.sciencedirect.com
ScienceDirect
Aerosol Science 38 (2007) 807 – 822
Journal of Aerosol Science
www.elsevier.com/locate/jaerosci

Spatial distributions of aerosol particles: Investigation of the Poisson assumption

Michael L. Larsen[a, b, *]

[a] Department of Physics and Physical Sciences, University of Nebraska at Kearney, Kearney, NE, USA
[b] Atmospheric Sensing Branch, Battlefield Environment Division, Army Research Laboratory, Adelphi, MD, USA

Received 1 May 2007; received in revised form 25 June 2007; accepted 27 June 2007

Abstract

Atmospheric particulates are known to be distributed non-uniformly in space due to spatially localized sources and incomplete mixing. Many studies have extensively examined the variability of aerosol particle concentrations on diurnal, seasonal, and annual timescales. Receiving much less attention, however, are aerosol particle spatial distributions on shorter ($\lesssim$5 min) temporal scales. Knowledge regarding the spatial distribution on these temporal scales is vital to fully understand and predict the properties of microphysical processes like activation, coagulation, and collision/coalescence, as well as some large-scale phenomena like radiation attenuation through a thick yet dilute layer of particles. Traditional treatments of microphysical phenomena often implicitly assume that particles are perfectly randomly distributed (obey Poisson statistics) on short temporal scales, even though this assumption may be inconsistent with observations made on longer scales. This study seeks to explicitly resolve the statistics of particle positions on a variety of temporal scales (over five orders of magnitude) to determine if and when the assumption of perfect spatial randomness can be justified in nature. Some implications of the results are discussed.

Keywords: Aerosol spatial distribution; Poisson statistics; Concentration fluctuations; Variance/mean ratio; Pair-correlation function

1. Introduction and motivation

Given a collection of aerosol particles and the open-ended invitation to "measure the system", most investigators begin and end their study with a careful measurement of the particle size distribution. There is nothing inherently improper about this approach; particle size can be directly measured by a number of different instruments and remains an important quantity to identify. (Particle size is the single biggest factor in determining how radiation interacts with a particle distribution and also suggests information regarding particulate origin and composition.) Nevertheless, knowing the size distribution of particles with good accuracy is not always sufficient, depending on the quantity or phenomena of interest. To choose a simple example, calculation of radiation transmission through a particle laden turbid medium requires more information about particle number, composition, orientation, and morphology than a measure of relative abundances at each size can identify.

* Department of Physics and Physical Sciences, University of Nebraska at Kearney, Kearney, NE, USA. Tel.: +1 301 593 1656; fax: +1 301 394 4797.
E-mail address: LarsenML@gmail.com.

0021-8502/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jaerosci.2007.06.007

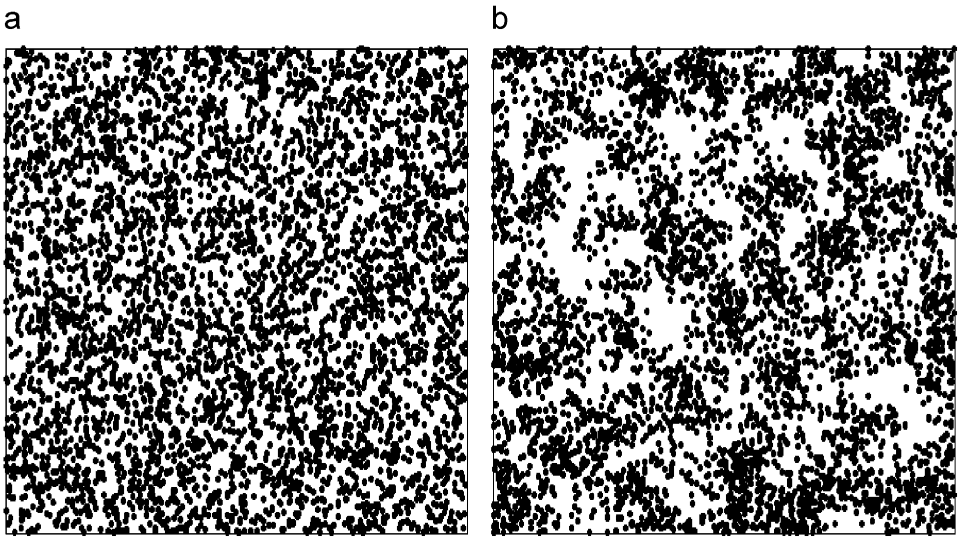

Fig. 1. A cartoon showing two hypothetical two-dimensional representations of the spatial distribution of aerosol particles. The left-hand panel (a) shows "particles" (represented by individual dots) distributed according to a Poisson distribution; there are no correlations between the particles and the underlying particle concentration is constant. The right-hand panel (b) shows the same number of particles distributed according to a statistically homogeneous but non-Poisson (statistically correlated) distribution. Assuming the box size represents a typical volume used to calculate concentration, no difference between systems (a) and (b) would be detected. In panel (b), though the concentration of particles at "box-scales" is identical to that in panel (a), the sub-sample variability exceeds that normally expected—due to the presence of spatial correlations. Despite identical concentration and mean inter-particle spacing in panels (a) and (b), the distribution in panel (b) would have a higher coagulation or collision rate than the distribution in panel (a). This result is qualitatively consistent with the predicted behavior from Kasper (1984), and can be understood geometrically—panel (b) has a larger number of particles close enough to be in physical contact with each other, thus enhancing the collision or coagulation rate.

One property seldom examined in even the most comprehensive studies of atmospheric aerosol particles is the statistical spatial distribution of the particles in the medium. Rather than examining this property in detail, most studies merely measure the total number of particles arriving at each of several sizes in consecutive temporal intervals, divide out the sampling volume, and interpret as a number concentration. Typically these measurements are taken at separations of at least several meters in space and several minutes (or substantially more) in time in order to minimize fluctuations (or "errors") due to sampling. Although such measurements may lead to an estimate of the fluctuations regarding total loading in a modestly sized volume, little information about the immediate environment of each particle is gleaned from these types of measurements. The local environment immediately surrounding each particle, however, does matter for many of the processes aerosol particles undergo (e.g. collision/coalescence and coagulation, activation) as well as processes that particles influence (e.g. radiation attenuation).

A particularly direct example arises when considering coagulation, which requires physical interaction between two particles in contact. The influence of spatial concentration inhomogeneities on coagulation rate was first studied by Kasper (1984); this study explicitly calculated the increase in coagulation rate due to particle number inhomogeneities. Later, Larsen, Cantrell, Kannosto, and Kostinski (2003) discussed how the deviations from classical coagulation or collision rates are likely to occur in even statistically homogeneous—yet correlated—distributions. (The idea may be initially counter-intuitive, but Fig. 1 illustrates how a hypothetical homogeneous and random-yet-correlated statistical spatial structure may enhance collision rate from geometrical considerations alone.)[1]

The "classical" coagulation and collision rates alluded to above—as well as classical predictions for activation, radiative transfer, and other phenomena—rely on the (usually implicit) assumption that individual particle positions

[1] The difference between statistically homogeneous-yet-correlated distributions with constant underlying concentration and statistically inhomogeneous distributions with varying concentration is extremely subtle, but requires an entirely different set of mathematical tools to describe the system. This work stays within the framework of statistically homogeneous media due to the natural analogue to the usually assumed Poisson distribution, which is statistically homogeneous by construction. Much more extensive discussion of these and other considerations can be found in Chapter 2 of Larsen (2006) and Appendix A of Larsen, Kostinski, and Tokay (2005).

are independent on sub-sample scales. Given a constant concentration of particles c in a volume, the probability of finding k particles in subvolume v is assumed to follow the relationship:

$$p_k(v) = \frac{(cv)^k \exp(-cv)}{k!}, \tag{1}$$

which is the probability distribution function for a Poisson distribution. In words, it is assumed that within the volume the particles are distributed as randomly as randomness allows.

The assumption that aerosol particles naturally follow a Poisson distribution is still under debate. There is a fair amount of literature supporting the Poisson formulation. Herdan (1960) and the sources cited therein (Badger, 1946; Fairs, 1943; Scrase, 1935) argue that Poisson statistics are appropriate for suspended particles in a fluid. Additional evidence for Poisson statistics (which can be physically interpreted as perfect spatial randomness) can be found in Green (1927) and Chapman and Ruhf (1955) (the latter of which cites the conclusions from Student, 1907). However, at least two more recent studies (Larsen et al., 2003; Preining, 1983) conclude that the Poisson distribution does not adequately describe all collections of particles in all size ranges, no matter what time-scale or volume is of interest.

The range of applicability of many earlier studies is questionable and, in some cases, known to be limited; many of the cited investigations actually determine the statistics of particles in liquid, which could change the statistical structure substantially. Two of the studies (Green, 1927; Scrase, 1935) rely on particle activation prior to measurement and result in the measurements being dominated by the small particles present. One of the studies (Badger, 1946) *assumes* all deviations from Poisson statistics must be due to experimental error without physical justification.

To contrast that viewpoint, other discrete objects in atmospheric science have been recently observed to deviate from a Poisson distribution. Although recently a hot-topic for discussion, it is now generally accepted that rain arrivals (Jameson & Kostinski, 2000; Jameson, Kostinski, & Kruger, 1999; Kostinski & Jameson, 1997), cloud droplets (Kostinski & Jameson, 2000; Marshak, Knyazikhin, Larsen, & Wiscombe, 2005; Shaw, Kostinski, & Larsen, 2002), and other inertial particles in a turbulent fluid (Balkovsky, Falkovich, & Fouxon, 2001; Elperin, Kleeorin, L'vov, & Rogachevskii, 2002) exhibit non-Poisson statistics.

Recent theoretical and simulation studies suggest that aerosol particles also likely exhibit departures from perfect spatial randomness (Chun, Koch, Rani, Ahluwalia, & Collins, 2005; Elperin, Kleeorin, Liberman, L'vov, & Rogachevskii, 2007), though whether the clustering is substantial for smaller, less inertial particles may still be in question.

Finally, there is a substantial amount of literature on the spatio-temporal distribution of passive scalars in turbulent flows (see Batchelor, 1959; Corrsin, 1951 for two of the most famous papers investigating these ideas and Warhaft, 2000 for a review). The widely accepted notion of a power spectrum of concentration fluctuations that decays as a power-law is inconsistent with the perfect spatial randomness viewpoint.

This study approaches the conflicting viewpoints experimentally. We utilize instrumentation unavailable in the early 20th century to determine whether airborne aerosol particles obey Poisson statistics and explore the implications of the observed result.

2. Experimental setup

As in Larsen et al. (2003), a CLiMET CI-8060 optical particle counter was used to make the measurements. This instrument measures the integrated scattering intensity of collimated white light in order to discriminate the size of individual aerosols into five size bins ranging from 0.3 to 10 μm diameter.

The built-in maximum time-resolution of the instrument is 1 Hz (this instrument was designed for air quality monitoring and resolving time-scales shorter than 1 s normally is not required). To obtain better time-resolution to completely test the Poisson hypothesis over a wide range of temporal scales, the analog photodiode signal was used along with custom software to identify within about 4 μs precision when each particle arrived in the sensing region. Use of the analog signal also allowed instrumental calibration with a variety of standardized spheres, permitting size discrimination to greater precision than normally reported by the instrument.

The instrument was placed on the roof of the Physics building at New Mexico State University in Las Cruces, New Mexico, as part of a field study investigating the behavior of the atmospheric boundary layer. The instrument was placed inside a wooden box and the sample was drawn from a vertically oriented metal tube directly connected to the instrument inlet. Data were collected intermittently between January 22 and February 6, 2007. Some data from inside the building were also acquired.

The preparation of each data-set involved taking 5 min of raw voltages from the analog output of the instrument. Then, a processing algorithm used the recorded voltages to determine the precise arrival times and size estimate for each particle detected in the 5 min interval. Upon completing the processing, the acquisition algorithm was allowed to acquire the next 5 min sample. A total of 643 data-sets were acquired with the instrument utilized in this study, resulting in the arrival time and size-stamping of approximately 4.4×10^8 particles.

3. Data analysis and results

3.1. Basic trends

A casual perusal of the acquired data-sets coarsened to 1 Hz resolution reveals that over each 5 min period there often were bursts or trends to the underlying particle concentration. (Two examples selected at random are shown in Fig. 2.) Such bursts and trends suggest that not all of the data-sets have statistical properties consistent with a Poisson distribution at 1 Hz scales.

This observation is not new. The conjecture that aerosol particles are spatially distributed perfectly randomly is not commensurate with the many existing observations of diurnal, seasonal, and annual trends in background aerosol concentration. Expectation of near-constancy is also inconsistent with the fluctuation properties expected of a passive scalar in a turbulent flow.

This investigation sets aside the lack of uniformity over longer spatial scales and focuses on finer structure—the domain where one might be more tempted to apply the results of, e.g., Scrase (1935), Badger (1946), Herdan (1960). We are testing the hypothesis that, *when the background concentration is "constant"*, the aerosol particles are distributed perfectly randomly (i.e. following Poisson statistics).[2] Care must be taken in interpreting this statement, however, as an overly literal readings could result in an impossible condition.

For example, if the mean particle inter-arrival time is defined to be τ, it is not possible for particle concentrations to be "constant" on time-scales of duration, say, $\tau/2$. Each interval will have either 0, 1, 2, etc. particles in it. The mean will be 1/2, but since no interval can have half a particle, the "concentration" is not technically constant no matter how the particles are spatially distributed (even if they form a rigid lattice). Similarly, it is far too stringent to require *every* interval of length 100τ to have *exactly* 100 particles. Even if the perfectly random assumption were to be valid, there is only about a 4% chance of seeing *exactly* 100 particles in such an interval (from Eq. (1)).

As we push the notion of concentration to smaller and smaller scales, we naturally expect more sample-to-sample variability due to shot noise alone; a standard deviation of $1/\sqrt{N}$ becomes non-negligible for small N. Consequently, we must take care not to require too stringent conditions for "constant concentration". (Friedlander, 1977, p. 8 specifically warns against applying the notion of "concentration" independent of limiting to proper scale.)

Previous studies often rely on a medium being "well mixed" or "physically homogeneous" to ensure statistical homogeneity. Although less precise than a formal mathematical description of constant concentration, this is a condition that may actually occur in nature or the laboratory. In such a medium, the claim made then suggests that there exist *a range of scales* where, practically speaking, deviations from Poisson statistics are negligible. The existence of both a minimum and maximum scale for this condition is not surprising given that the mechanisms that create and mix aerosol particles are finite in size and duration. In the case of data-sets like those shown in Fig. 2, the maximal temporal scale for Poisson statistics may be no more than a fraction of a second.

Observations of 5 min data-sets suggest that the longest temporal scale with concentration steady enough to test the Poisson hypothesis may sometimes (often in our data) be shorter than 5 min. As an observational conclusion, this has limited utility (although a valid conclusion for the specific conditions measured, little can be definitively said about aerosols not in the lower boundary layer, not in New Mexico, not in the 0.5–10 μm size range, and not measured during the dates of the experiment). Nevertheless, to the best of the author's knowledge, no guidance in the literature currently exists as to where the breakpoint between "concentrations are nearly constant" and "concentrations exhibit trends and other deterministic changes" for ambient aerosols occur.

Given the assumptions in some of the airborne-pathogen literature (e.g., from Nicas & Hubbard, 2002, "the ambient exposure level C_A is defined as a constant airborne concentration of pathogens over a T hour interval"), it seems that

[2] Such a statistical description is not without precedent in the atmospheric sciences. Recently, Larsen et al. (2005) found evidence for some (rare) events of homogeneous rain with Poisson-like statistics.

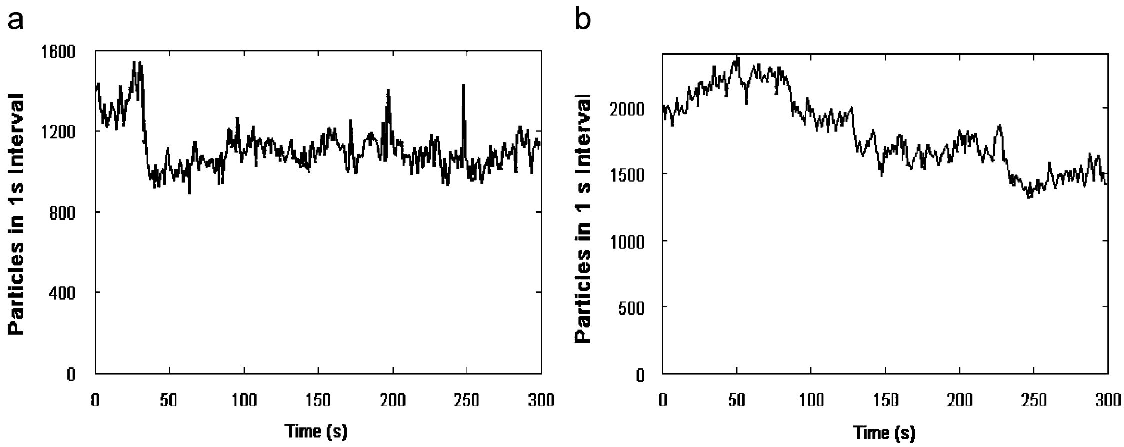

Fig. 2. Two 5 min samples of 1-Hz concentration data from the field study at New Mexico State University. These plots are for random 5-min intervals not selected for more detailed study, due to the obvious inhomogeneities that suggest an underlying changing concentration. In panel (a), the first 40 s of data have a higher concentration than the remaining fraction of the data-set. In panel (b), a smooth trend of changing concentration is evident. In both cases, the data can be concluded to be statistically non-stationary. Although this indicates non-Poisson statistics, in principle the data could be an inhomogeneous Poisson process (characterized by being piece-wise Poisson, with concentration a function of time). Based on visual inspection alone, it can be determined that these plots cannot be described by a homogeneous Poisson distribution and likely have a changing underlying concentration.

identifying the time-scales over which concentration can appreciably change as shorter than an hour could be of benefit for a variety of studies.

The simple observation that aerosol concentrations seem to be variable in a deterministic way over intervals as short as 5 min require us to modify the original goal. No longer is the question as simple as "Are aerosol particles distributed in a perfectly random manner in space and time?" Rather, we need to ask "Over what range of scales (if any) can aerosol particles be described by a Poisson distribution?"

Alternatively, placing this inquiry into a practical context, if air quality or concentration is measured on a nominal (say 5 min) interval, it would be useful to find out how often (if ever) theoretically based estimates of microphysical processes would hold using an ensemble-averaged (measured) concentration. This amounts to verifying the Poisson assumption for any sub-interval duration appropriate to the process of interest.

With a single observational data-set, the frequency of Poisson behavior cannot be determined definitively. The distance from source, strength of mixing, particle size, geographical location, and a host of other variables likely influence the relevant statistics. As a first step, however, we will analyze the data at hand while keeping in mind that one ought not to overgeneralize from one set of observational data.

3.2. Fixed scale investigations

For the remainder of this paper, we actively seek out the best possible candidates for an interval of at least 5 min that can be described by a homogeneous Poisson distribution. Although most of the data-sets have 1 Hz concentration profiles similar to those seen in Fig. 2, there were a non-negligible fraction that had concentrations that were visually trendless, approximately constant, and devoid of sudden peaks. Qualitatively dividing the data-sets into broad categories of "spiky", "trended", "spiky and trended", and "stationary-looking", about 15–20% appear more or less stationary (about half of the indoor data and about 10% of the outdoor data). The two examples we will use in our analysis for the rest of this paper are shown in Fig. 3. These two intervals were visually most similar to a perfect horizontal straight-line when plotting the 1 s concentration data; if a 5-min interval obeying Poisson statistics occurred at NMSU during the field study, these two data-sets would likely be the best candidates.[3]

Visual similarity of 1 Hz measurements to a trendless line is necessary but insufficient to determine whether particles obey Poisson statistics. Two other properties easily tested are (i) the probability of observing k particles in a time

[3] Similar analysis was completed on several more of the stationary-looking data-sets with comparable results.

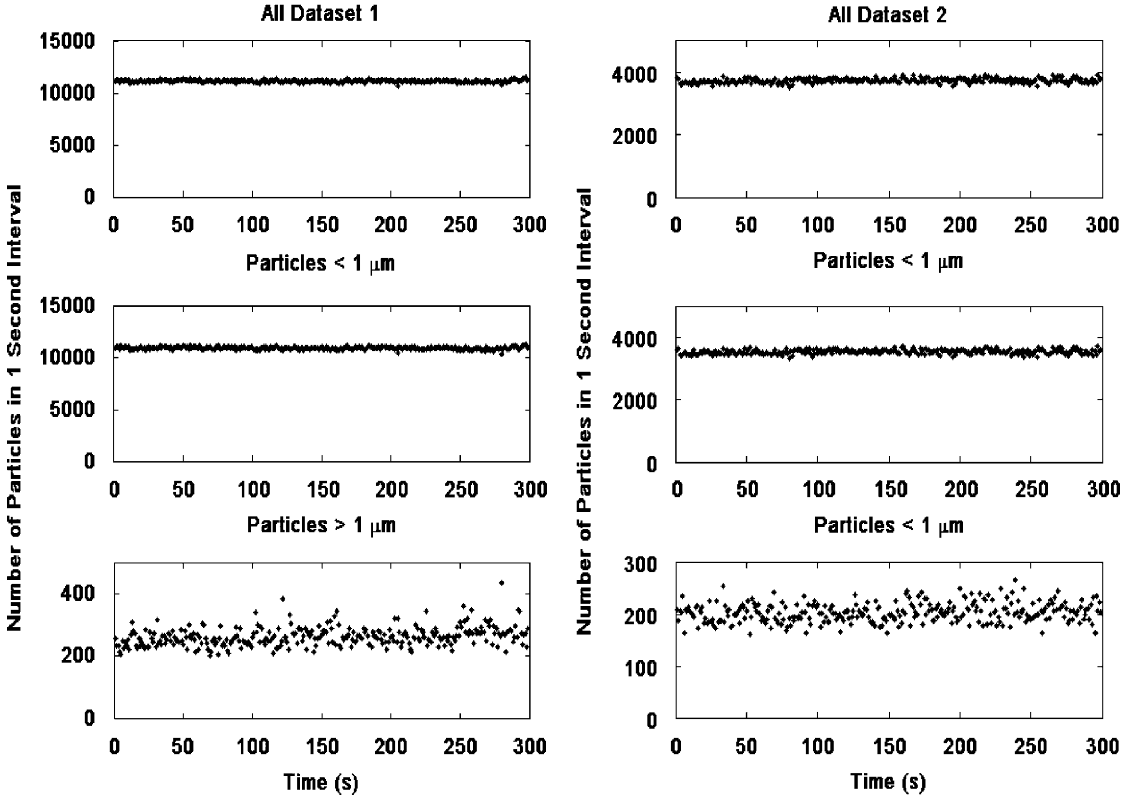

Fig. 3. Counts of particles observed in two 5 min periods by the CLiMET CI-8060 particle counter on the roof of the Physics building at New Mexico State University on January 30th, 2007. Throughout this paper, Dataset 1 (shown on the left-hand side) corresponds to the 5-min period starting at 10:41 AM (local time) and dataset 2 (shown on the right) corresponds to the 5-min interval beginning at 12:33 AM. In each plot, the total number of particles observed in each second is recorded along the y-axis. To try and determine some size-dependent information, subpopulations of smaller (less than 1 μm) and larger (greater than 1 μm) particles are examined independently in the second and third rows of each panel, respectively. These particular 5-min periods were chosen for analysis because there was no visually discernable trend or anomalous inhomogeneities in the entire 5-min interval.

interval t must follow the relation given in Eq. (1), substituting observation time interval t for v and observed particles per unit time λ for c and (ii) the variance of counts in time interval t should be equal to the mean for the same interval (λt). These properties are explicitly tested in Fig. 4.

To generate Fig. 4, each 1 s count from Fig. 3 was recorded and binned into a histogram. After plotting the results in bar-form, a computer-generated Poisson simulation was run with the same total number of particles. The histogrammed result of the simulation is superposed on the figure with dots to show the shape of the histogram as a truly Poisson distribution may appear. Finally, each panel has the observational variance/mean ratio in its title.

What conclusions can be drawn from Fig. 4? Both data-sets exhibit some deviations from perfect Poisson behavior. The histograms for particles less than 1 μm in diameter appear to approximate the results from the simulation fairly well in both data-sets. Similarly (since particles less than a micron make up the vast majority of all the data), the histograms for the entire data-sets shown in the first row look qualitatively similar as well. However, explicit calculation of the variance/mean ratios show some variability in excess of that expected for a Poisson distribution. All of the deviations from simulation are present in much larger degree for particles larger than 1 μm in size.

Some of the uncertainty in drawing conclusions from the figure may be due to the fact that only 300 independent measurements of 1 s duration are available in a 5 min measurement. This keeps the values for both the measured and simulated system fairly modest along the y-axis. Also, viewing a histogram this way really only tells us about the accuracy of the Poisson hypothesis for averaging time t equal to 1 s (a temporal scale related to a sample-volume larger than is relevant for many microphysical processes).

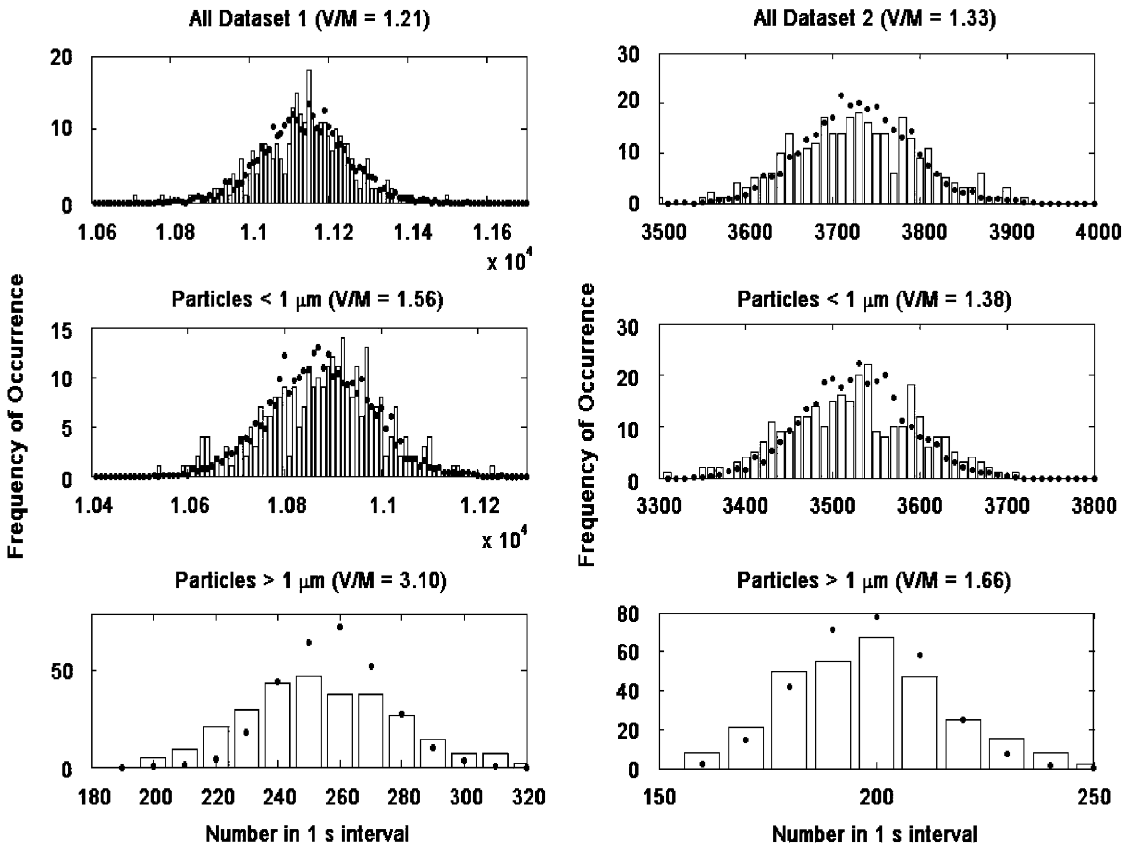

Fig. 4. Histograms of the data displayed in Fig. 3. The bar graph shows the actual number of 1 s measurements in the 5 min period with the specified number of particles. If particle arrivals are perfectly independent, the resulting histogram is expected to follow a Poisson distribution. The dots on the figure demonstrate this with simulated data. In the title for each panel, the observed variance/mean ratio is recorded. (V/M for a Poisson distribution is always equal to 1, independent of scale). When the actual histogram is broader (e.g. has longer tails) than the simulated data, we find V/M ratios substantially greater than unity.

Similar plots using averaging times t equal to 100 and 10 ms are presented in Figs. 5 and 6, respectively. These allow histogramming of the same data into many more disjoint measurements, and approaching scales of more relevance to physical interaction.

We limit ourselves to examining figures of temporal interval 10 ms or longer; a timescale of 1 ms, for example, would allow an average of less than one large ($> 1\,\mu$m) particle per temporal interval. Though in principle this is perfectly acceptable, the ability to visually distinguish between Poisson and non-Poisson distributions with very low means is difficult and can easily be misleading (see, e.g., Larsen et al., 2003).

Merely using the three orders of magnitude plotted in Figs. 4–6 we do begin to see some trends. Most obviously, the subset of smaller ($< 1\,\mu$m) particles always match the Poisson simulation substantially better than the larger ($> 1\,\mu$m) particles, no matter the temporal scale. It is also apparent that shorter temporal intervals demonstrate smaller variance/mean ratios and better histogram agreement with a Poisson simulation. Though the data histograms for the 1 s data-sets show slightly longer tails than expected for a Poisson system, the agreement for the shorter time-scales appears excellent.

One is likely tempted to conclude that, in fact, there is a transition from Poisson behavior to non-Poisson with increasing scale. The exact transition point may be in question—depending on some sort of tolerance parameter—but nominally would happen on some scale around 1 s. For these particular data-sets, then, the analysis presented so far suggests that perhaps the Poisson assumption is adequate for temporal scales less than 1 s duration, but non-Poisson statistics must be used to describe longer scales.

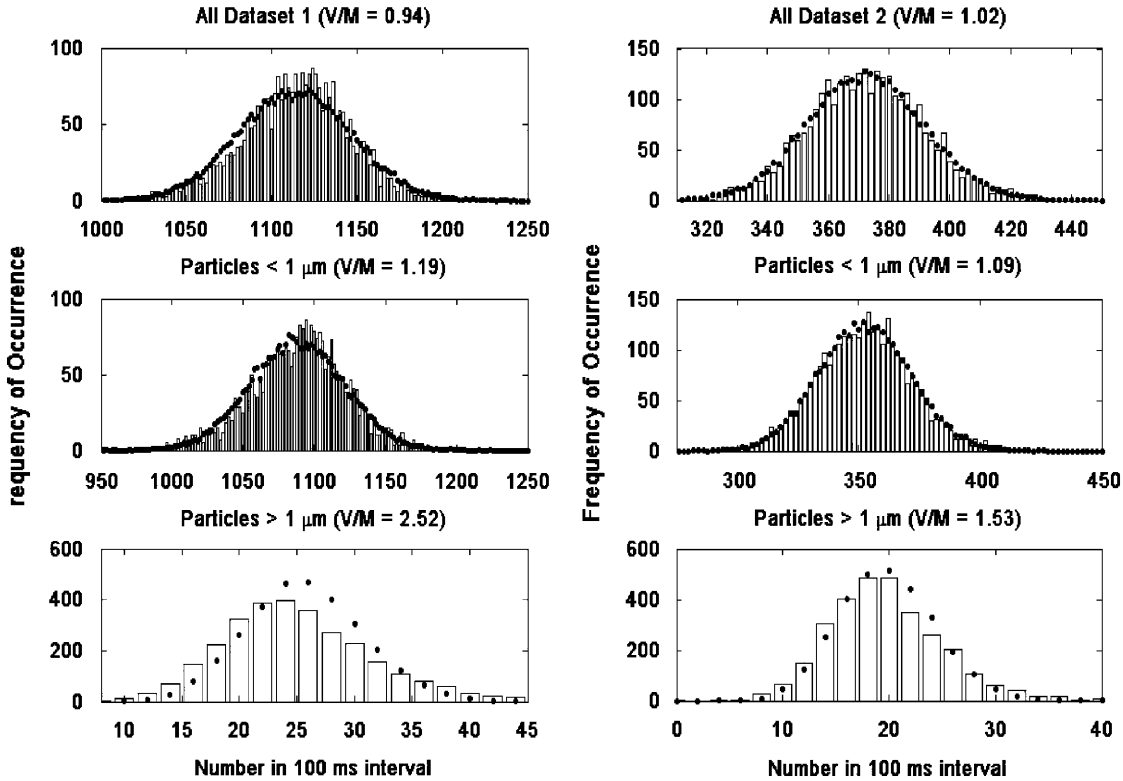

Fig. 5. The same type of plot as Fig. 4, except that instead of 300 measurements of 1 s, the historgram is made up of 3000 measurements of 100 ms each. Note that the V/M ratio is closer to unity in each of the six panels than when the averaging scale of 1 s was used (as in Fig. 4).

We have seen that the appearance of perfect spatial randomness (Poisson statistics) on one scale does not necessarily indicate similar behavior on other scales. Given this observation, it may be premature to conclude that for all scales less than our nominal transition scale of 1 s the behavior must be perfectly random. For microphysical phenomena like collision/coalescence and activation, the *immediate* environment of each particle influences the process. For these data-sets, can we argue that there is no deviation from perfect randomness at time-scales comparable to the inter-arrival time at the detector? *No!*

3.3. Scale-dependent investigations

In Figs. 4–6 we used the variance/mean ratio to help quantify the departure from Poisson statistics, noting that this ratio should be near unity for Poisson statistics. It is observed that this ratio is found to change as a function of time-scale; this suggests that using the variance/mean ratio as a final measure of "does the system exhibit perfect spatial randomness" may be misleading; one only can argue for agreement with properties of a Poisson system at the averaging scale used. Fig. 7 explicitly plots this ratio for the data under investigation as a function of scale.

The figure reveals non-trivial scale-dependent behavior in all of the data-sets. As noted earlier, finding a unity variance/mean ratio at one scale does not necessarily imply that the underlying distribution obeys Poisson statistics at all scales. For example, for particles larger than 1 μm in diameter from dataset 1 using an averaging time of nominally 1 s, we find a variance/mean ratio of around 4. *The exact same collection of particles measured at the exact same times* display a variance/mean ratio of about 1.5 when using 1 ms as the averaging scale.

No data-set can be justifiably classified as perfectly random by measuring the variance/mean ratio at a single spatial or temporal scale.

One is likely to conclude from Fig. 7 that most clustering present occurs at temporal scales 10 ms and longer. This, too, is misleading. If the variance/mean ratio is near unity at temporal scale T_0 and not near unity at temporal scale

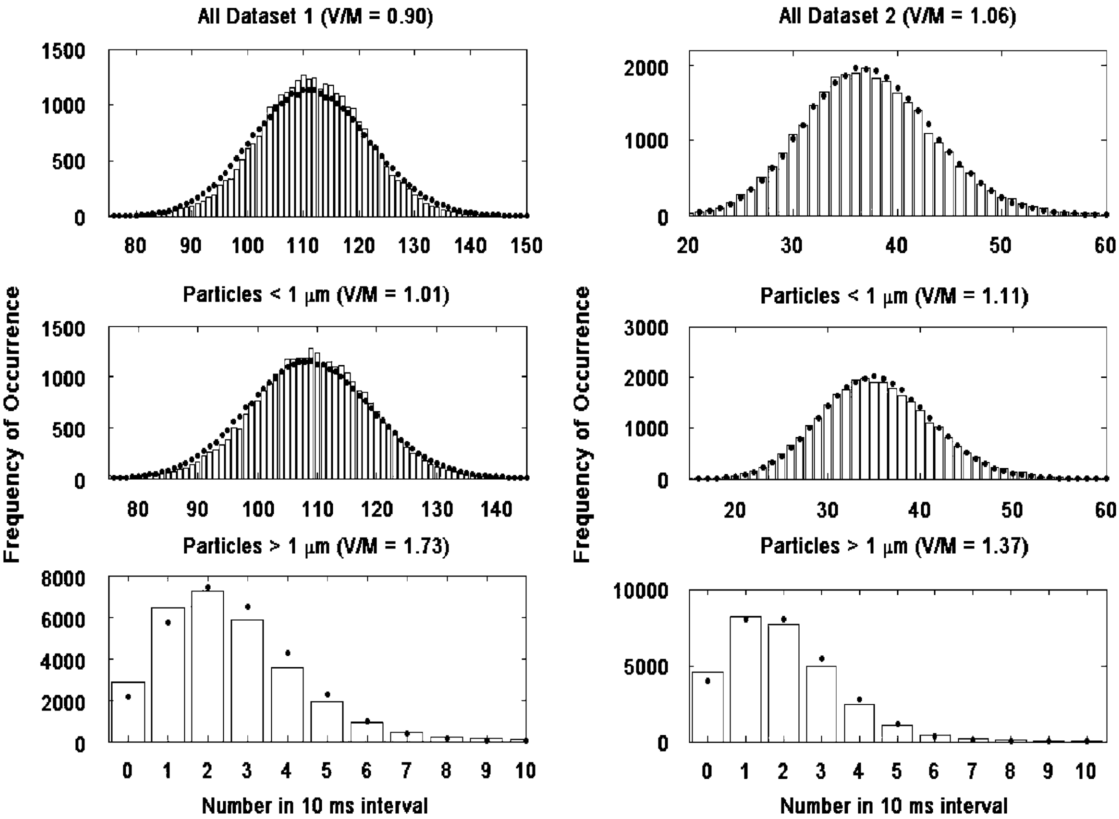

Fig. 6. The same plot as Figs. 4 and 5, but with time-scale 10 ms. Once again, comparison with Figs. 4 and 5 demonstrates the V/M ratio seems to decrease as a function of decreasing time-scale. Note the excellent qualitative agreement between the simulated Poisson process and the observed data at this timescale, especially for data-set 2.

T_1, it *does not imply* that the non-Poisson statistical structures are between T_0 and T_1. The variance/mean ratio, though a convenient easy to calculate statistical quantity, is not a scale localized measure of departures from perfect spatial randomness (see, e.g., Larsen, 2006; Shaw et al., 2002).

Another difficulty in using the variance/mean ratio stems from the realization that the variance/mean ratio for a physical system is rarely *exactly* unity. Any data-set is finite and the variance and mean estimates from the observed data are subject to sampling variability. It then becomes a meaningful question to ask "what departure from a variance/mean ratio of unity is significant?" Careful consideration may reveal that a variance/mean ratio of 1.2, for example, is in some cases a significant departure from a Poisson expectation and in other cases close enough to be within sampling error due to the finite nature of the data. This point is explored in more detail in Appendix A.[4]

The best tool to examine scale-localized departures from perfect spatial randomness in a statistically stationary setting is the pair-correlation function ($\eta(t)$). Heuristically, $\eta(t)$ quantifies the enhanced probability of observing a particle at time $t_0 + t$ given the presence of a particle detected at t_0. A positive pair-correlation function at some scale t directly demonstrates a net clustering effect on scale t and says nothing about scales larger or smaller than t. Formal definitions and analytic relationships connecting more commonly used statistics can be found in Cox and Isham (1980), Landau and Lifshitz (1980), Kostinski and Jameson (2000), Shaw et al. (2002), and Larsen (2006).[5]

[4] Corrections can be made to the simple computation of variance/mean ratios to account for this scale-dependence (see, e.g., the Fishing statistic approach used by Baker, 1992). However, the physical scale of clustering remains independent of the scale where the variance/mean ratio deviates most from the Poisson expectation. Although the Fishing statistic, for example, is a valid metric to determine *whether* a system obeys Poisson statistics at all scales, it can be deceptive to use this statistic to identify *at which scales* any clustering effects occur.

[5] The pair-correlation function is also sometimes referred to as the "radial distribution function" when used when quantifying departures from perfect spatial randomness in fluids, especially DNS studies (see, e.g., Sundaram & Collins, 1997).

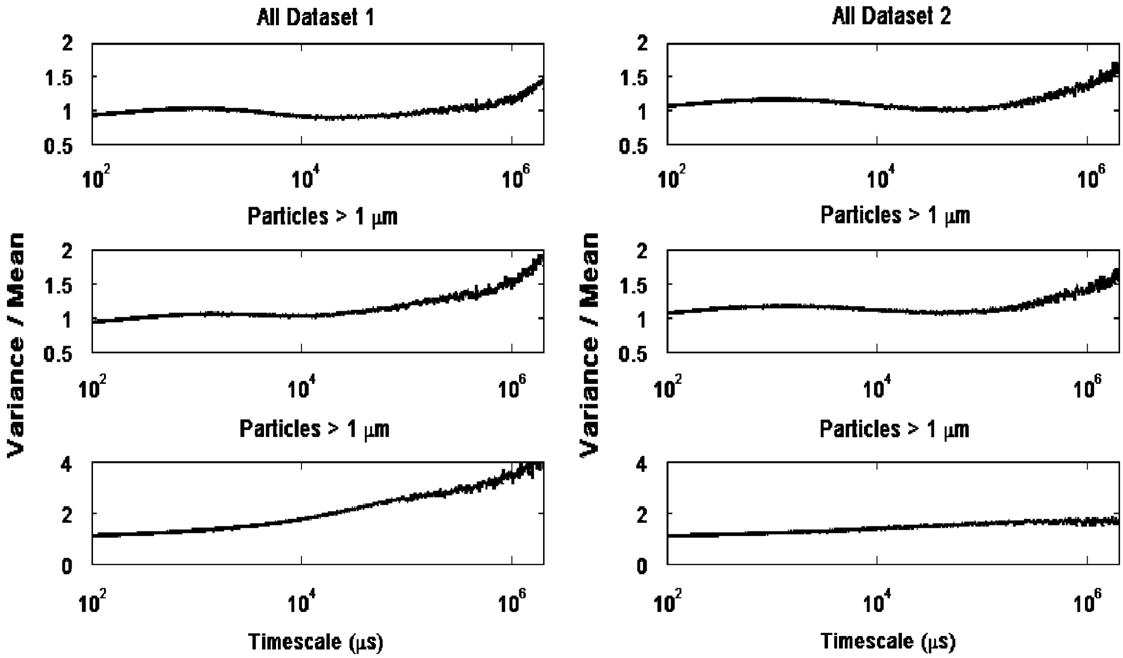

Fig. 7. Examination of the variance/mean ratio as a function of time-scale for each of the data-sets and subpopulations. Note that for a Poisson distribution the variance/mean ratio should be nearly unity for the entire domain. (To avoid sampling issues, a maximal time-scale of 2 s was used on the plots, even though the observations were 300 s long.) Extension or analysis of the sub ms domain with the histogram technique as used in Figs. 4–6 becomes complicated as the mean inter-arrival time is of the same order as the measurement duration. Better techniques are available in this time domain and are discussed in the text.

The pair-correlation function can be written directly in terms of interarrival time data like that acquired by the instrument. The technique, previously identified by Picinbono and Bendjaballah (2005), was used to compute the estimates of the pair-correlation functions plotted in Fig. 8; the technique is outlined in Appendix B.

Fortunately, via the so-called correlation-fluctuation theorem (explicitly derived in Cox & Isham, 1980; Larsen, 2006), we can also relate the pair-correlation function to the more familiar variance/mean ratio:

$$\left(\frac{\text{Variance}}{\text{Mean}}\right)_T - 1 = \frac{2\lambda}{T}\int_0^T (t - t')\eta(t')\,\mathrm{d}t'. \tag{2}$$

The presence of the weighting term in the integrand $(t - t')$ demonstrates the limited utility in using the variance/mean ratio to infer scale-localized clustering characteristics. Recall that $\eta(t)$ is the scale-localized measure of clustering (or, more precisely, correlations). We then see that the clustering at small scales ($\eta(t)$ for small t) becomes weighted more as the upper limit of the integrand increases; the deviations from unity we see in Fig. 7 may be largely due to correlations existing on very small (previously unresolvable) temporal scales.

3.4. The pair-correlation function and complications due to dead-time

A system exhibiting Poisson statistics at all temporal scales should have $\eta(t) = 0$ for all t. Likewise, for any scale with $\eta(t) = 0$, there are no correlations present *on that scale*, though there may be non-unity variance/mean ratios or deviations from the Poisson relation given in Eq. (1) resulting from correlations at shorter scales.

Examination of the observed pair-correlation functions in Fig. 8 reveal deviations from $\eta(t) = 0$. Note that, for very short temporal separation ($t \lesssim 10\,\mu\text{s}$), $\eta(t) = -1$. In fact, the pair-correlation function is less than 0 for separation times up to a few tens of microseconds.

Physically, $\eta(t) < 0$ suggests that the probability of detecting a particle at $t_0 + t$ given a particle at t_0 is less than would be expected in a Poisson distribution, and $\eta(t) = -1$ states that mutual detection at $t_0 + t$ and t_0 is impossible.

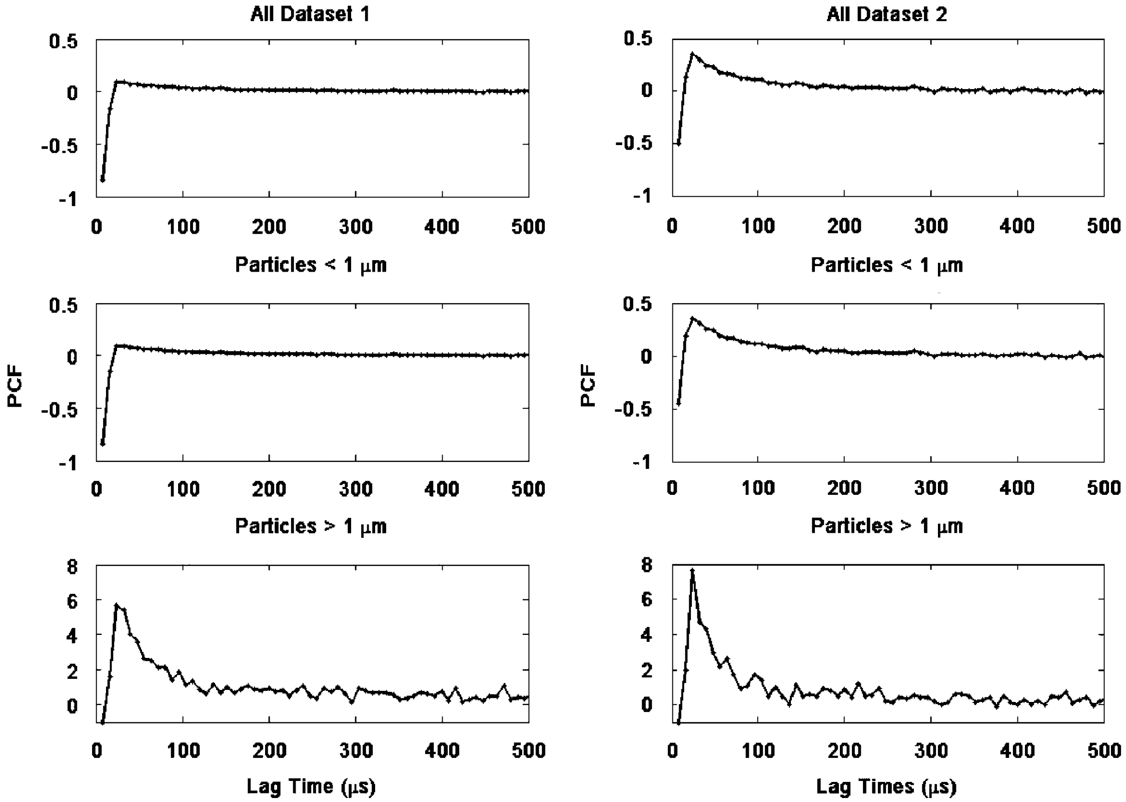

Fig. 8. The pair-correlation (PCF) as a function of lag-time for each of the data-sets and subpopulations. This *is* a scale-localized statistic, so a deviation from perfect randomness ($\eta(t) = 0$) of a particular lag-scale suggests some non-Poisson behavior (clustering) on *scale t*. The negative value for times less than about 10 μs are associated with the dead-time of the instrument after particle detection (see text). Note that even though the V/M ratio does not substantially differ from 0 until a temporal scale of nearly a millisecond (see Fig. 7), all of the actual deviations from perfect randomness occur on substantially shorter, microphysically relevant, time-scales. Note that, though $\eta(t)$ seems to be nearly equal to 0 for small particles at most lags, the departures exceed that expected due to sampling effects for *all* curves (see Appendix A).

In this case, the mutual exclusion arises from dead-time (either due to instrumental insensitivity or coincidence in the sensing volume. Although these are different physical origins, the mathematical treatment remains the same).

Recall that a particle detection in this instrument occurs if the photodiode voltage ascends above some threshold value. While the particle remains in the sampling volume of the instrument, the photodiode voltage remains above the threshold value and traces out a curve that is approximately Gaussian in time. The peak of this Gaussian shape is used to estimate the size of the particle, and the arrival time is associated with when the minimal threshold value was first exceeded. This technique *prevents the detection of any other particle until the photodiode voltage returns below the threshold value*. For this instrument in the setup used, the threshold value was nominally 20 mV. Empirical analysis of the resulting voltage traces suggest that there is a time of instrumental insensitivity to new particles (τ) lasting nominally 15 μs after the detection of a particle.

This insensitive interval (referred to as "dead-time") varies some from particle to particle, though is independent of the particle size. When analyzing a data-set subject to instrumental dead-time, one finds that $\eta(t < \tau) = -1$ during the dead-time interval. Additionally, estimation of the true number of pairs detected for a Poisson process requires accurate measurement of the total counting rate λ. When dead-time is present, the measured estimate for λ is slightly under the true counting rate. This, then, results in a spurious positive pair-correlation function for temporal scales near, but slightly longer than, $t = \tau$, as observed in the pair-correlation traces.

In a forthcoming paper, the author and a collaborator explicitly calculate the pair-correlation functions for Poisson distributions subjected to various types of instrumental dead-time. The details are beyond the scope of this paper, but—as can be seen from Fig. A.2 in Appendix A—dead-time alone cannot account for the positive deviations from

$\eta = 0$ observed in Fig. 8; real positive correlations in excess of those due to instrumental dead-time are present. In fact, drawing from Larsen (2006), it can be argued that if the system could be measured without dead-time present, the deviations from perfect randomness would be more pronounced, no matter what metric is used to quantify the deviations from Poisson behavior.

4. Discussion and conclusions

The basic question, in its modified form, seems simple enough; "In the acquired data, are there any 5 min intervals where perfect temporal randomness is observed?" The ultimate answer, too, is rather simple; no.

What complicates the issue is that there are simple statistical tools familiar to many scientists that would erroneously suggest perfectly random subintervals exist in the data. For example, the qualitative similarity between the measured and Poisson simulated histograms for submicron particles in dataset 2 shown in Fig. 6 is striking; it is hard to blame an investigator who would use such agreement to incorrectly infer Poisson statistics. (This is essentially what was done in some of the studies cited in the introduction). Similarly, that same data reveals a variance/mean ratio of 1.11—very nearly unity. Yet, the analysis in Appendix A reveals that for timescales of 10^4 μs, a value of 1.11 exceeds unity by more than we can expect due to sampling issues alone. The misuse of familiar, simple statistical tools is likely what led most of the investigations cited in the introductory section to erroneously conclude that airborne particles inevitably follow Poisson statistics at small spatial scales.

It is valid to ask about the significance of this non-Poisson behavior at short interarrival times. The significance of the measured result *critically depends on the application of interest*. If one desires to calculate how many particles are in a well-mixed room, this entire investigation is overkill; little error will be induced by extrapolating a well-considered concentration estimate over a (sufficiently large) sub-volume to the entire room. For collision rate, however, what really matters is the value of the pair-correlation function on spatial-scales comparable to the particle-size. If instrumental perturbation of the ambient environment is not significant (itself a questionable assumption), the magnitude of the correction could range from values as small as a few percent (for the small particles in the most homogeneous subsets observed) up to factors of 1000% or more (for large particles in statistically inhomogeneous data-sets with more intermittency than even those shown in Fig. 3).[6] In many cases (e.g. radiation attenuation—see Kostinski, 2001), the magnitude of the correction required due to clustering would not be known even if the pair-correlation function was precisely known for the media. Until recently, nobody knew the question was worth asking so the theoretical correction is not yet worked out.

We have seen that care must be taken in choosing appropriate statistical tools to confirm or reject the Poisson assumption. Near the end of his paper, Preining (1983) writes:

> Conditional distributions such as … the distributions of time intervals that occurred just after a particle of a certain size had arrived or a time interval of a certain duration had elapsed would give entirely new and deep insights into the structure of particulate clouds and thus provide a much needed quantitative tool in aerosol microphysics.

The pair-correlation function directly characterizes such conditional distributions. Despite a directed effort to find a dataset where this conditional distribution follows Poisson statistics (exponential interarrivals), *all systems observed* exhibited some degree of non-Poisson behavior. Many of these systems were obviously non-Poisson due to statistical inhomogeneity. However, even in the statistically homogeneous intervals, deviations from Poisson statistics existed.

To study this system, a sampler with a fixed flow-rate was utilized. The instrument is not perfect and does have an inherent instrumental dead-time associated with it. It may be argued that the measurement of the existing correlations may have been magnified (due to turbulence or stretched flow through the instrument) or diminished (by dead-time effects). The data merely reveal that particle arrivals at the sensing volume of the detector cannot be adequately described a perfectly random Poisson process nor a perfectly random Poisson process subject to instrumental dead-time. In short, the two-dimensional representation on the left hand side of Fig. 1 is not appropriate for particle arrivals in the sampling volume.

[6] If collision rate is assumed to depend on the number of particles with some particular spatial separation (r), then the actual number of particles with this separation is equal to the expected number $[n(r)]$ multiplied by a correction factor $1 + \eta(r)$ where $\eta(r)$ is the spatial analogue of $\eta(t)$ (Larsen, 2006).

It may be argued that it was turbulence in the detector, for example, that caused the deviations from perfect spatial randomness. However, one must think that if turbulence in the detector could cause the deviations, why not environmental turbulence as well? The ultimate source for the deviations from perfect temporal randomness may be inertial response to turbulence, incomplete mixing of initially localized sources, electrostatic effects, or some other phenomena. The substantially larger deviation from Poisson behavior for particles larger than 1 μm may yet yield a clue to the mechanism present. No matter what the ultimate cause, however, the degree of departure should be quantified (hopefully with the aid of other, more passive, instruments) to help estimate the influence of non-Poisson statistical structure on microphysical phenomena.

There are possible (and in most cases, unestimated) impacts of very small-scale clustering effects on coagulation, collision/coalescence, radiation attenuation, activation, and even airborne pathogen risk assessment. This paper establishes that the absence of perfect spatial randomness among aerosols, despite previous investigations, may be truly ubiquitous. The extent of the effect this may have on estimating microphysical processes should at least be bounded, and a useful first step is quantifying the magnitude of the deviations from perfect randomness with the pair-correlation function.

Acknowledgements

The author would like to thank Glenn Shaw for the loan of the CLiMET particle counter used in this experiment. The author also thanks Will Cantrell, Steven C. Hill, Elena Fernandez, Dennis Garvey, Alexander B. Kostinski, Ron Pinnick, and Jim Rosen for helpful discussions and experimental support. Additional comments that substantially improved the manuscript were supplied during the peer-review process. Some of this research was performed while the author held a National Research Council Research Associateship Award at the Army Research Laboratory. The data were acquired as part of the DTRA-funded ARL Stable Boundary Layer Initiative to study Turbulence in the Stable Boundary Layer. Additional support was supplied by the College of Natural and Social Sciences at the University of Nebraska at Kearney.

Appendix A. What constitutes a “significant” departure from a variance/mean ratio of unity?

One of the properties of an idealized Poisson distribution is that the variance equals the mean. However, no finite system can perfectly approximate a Poisson distribution and, as such, deviations from a variance/mean ratio of unity are inevitable.

In examining temporally finite intervals, we actually ask whether deviations from Poisson statistics exceed those expected from a truncated idealized Poisson distribution. This requires determination of when deviations from a variance/mean ratio of unity are “significant”.

Fig. A.1 displays the observed variance/mean ratios for particles less than 1 μm in diameter for both datasets examined in the main text of the paper (the same curves are shown in the second row of Fig. 7).

To compare to a truncated Poisson system, the same variance/mean calculations were carried out on 1000 generated distributions. The distributions were generated using the native pseudo-random number generator embedded in MATLAB. The maximum value and minimum value of the variance/mean ratio for each timescale was then recorded. The grayed out sections of plot A.1 show the domain in which all of the 1000 curves lie. Roughly speaking, then, these bounds establish “confidence intervals” of at least, say, 95%.

Since this system is subject to instrumental dead-time, the same computation was done with the generated distributions subjected to a dead-time of nominally 15 μs (an empirical approximation based on observation of the photodiode outputs). This dead time, as expected (see, e.g., Larsen, 2006) lowers the variance/mean ratio and demonstrates that the departure from perfect randomness observed is more substantial than at first blush. Note that the particles less than 1 μm were used because the variance/mean ratios in Fig. 7 and the pair-correlation functions shown in Fig. 8 suggest that these two subsets are the best candidates for following a Poisson distribution.

Fig. A.2 explicitly calculates the pair-correlation function for the two datasets (of particles less than 1 μm) and compares to a typical pair-correlation function for a simulated Poisson distribution subject to non-extensible dead-time. Although Fig. 8 might lead some to believe that the departures from $\eta = 0$ above the dead-time intervals are insignificant, $\eta(t)$ exceeds 0 for scales substantially longer than one expects for a perfectly random system. Though the effect seems to decay to 0 at scales of about 300 μs, dead-time alone can only account for deviations for scales of less than about 40 μs.

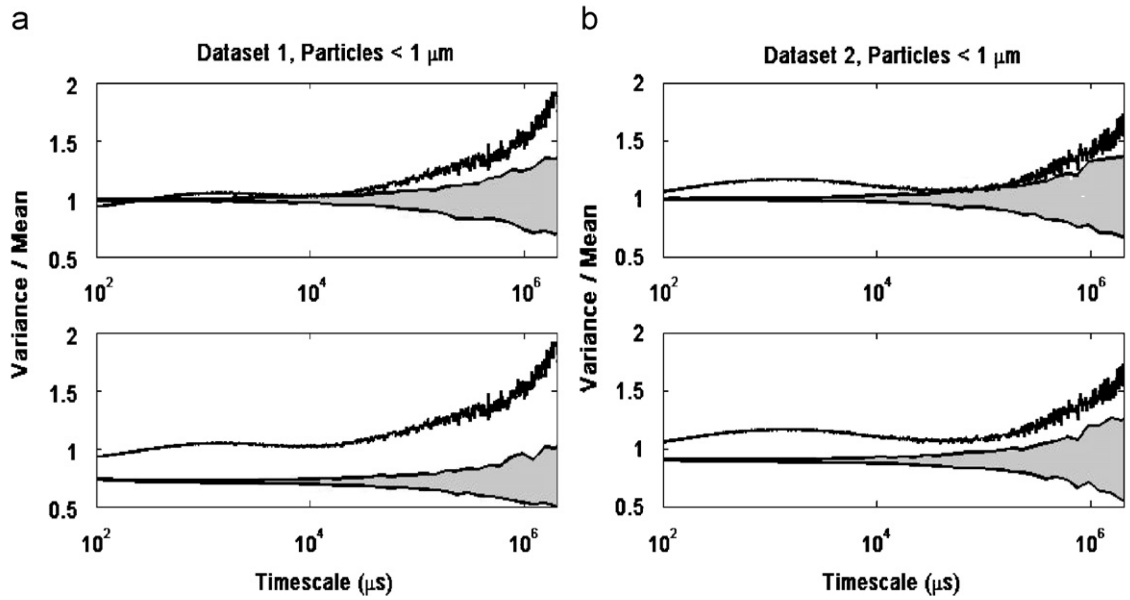

Fig. A.1. Variance/mean ratios for submicron particles examined in the main text. The solid line in the left-hand panels (a) both show the same variance/mean ratio already plotted in the middle panel of the left-hand side in Fig. 7. Likewise, the solid line on the right-hand panels (b) both show the same trace as shown in the middle panel of the right-hand side in Fig. 7. The shaded bounds demonstrate the possible range of variance/mean ratios for a Poisson distribution with the same number of total particles distributed over the same temporal intervals. These bounds were established by running numerical simulations as described in the text. The top row of both (a) and (b) shows the bounds established ignoring instrumental dead-time. The bottom row of both (a) and (b) shows the bounds expected for a Poisson distribution sampled with an instrument subject to extensible dead-time of duration $\tau = 15\,\mu s$. With or without dead-time, variance/mean ratios do fall outside of the Poisson expectation at virtually all scales. Note that in the case where dead-time is taken into account, the observed deviations are even further from the Poisson expectation. The deviations from Poisson behavior may not be most prevalent when the variance/mean ratio is furthest from unity (see main text).

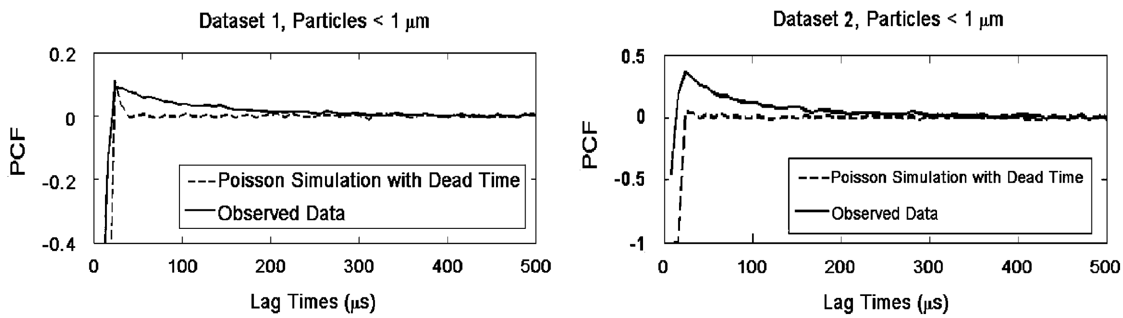

Fig. A.2. Pair-correlation functions (PCF) for the sub-micron data analyzed in the main text. The solid lines display the empirically computed pair-correlation functions (a copy of the trace found in row two of Fig. 8). The dashed lines demonstrate the computed pair-correlation function of a simulated Poisson distribution with the same number of particles over the same interval, subjected to a non-extensible dead-time of 15 μs. (A non-extensible dead-time better matches the empirically observed pair-correlation functions than an extensible dead-time. See Larsen (2006) for a discussion why this would be so.) In the left-hand panel, we observe that though the "peak-value" of the pair-correlation function is of the order we would expect for a Poisson distribution subjected to dead-time, the empirical decay to zero takes substantially longer and suggests true non-Poisson behavior that cannot be attributed to the dead-time effects alone. Deviations from a dead-time altered Poisson process are more pronounced in the right hand panel for all scales, suggesting that the system is more correlated in an absolute sense. (This is somewhat surprising given that the histograms have data-set 2 more visually similar to a Poisson distribution.)

Appendix B. Using inter-arrival distributions to calculate the pair-correlation function

As noted in the text, this link between the pair-correlation function and waiting-time distribution functions is due to Picinbono and Bendjaballah (2005), with some additions as found in Larsen (2006).

For a statistically homogeneous system with rate λ, one can use the pair-correlation function to write the probability of observing a particle at time $t + t_0$ given an observation at t_0. The associated probability density function is given by

$$p_t(t_0 + t|t_0) = \lambda[1 + \eta(t)]. \tag{B.1}$$

However, if a particle is observed at $t_0 + t$ it must be the nth particle posterior to the particle at t_0 where n is some positive integer.

Let $f_k(t)$ be the probability density that the kth particle posterior to a particle at t_0 (i.e. the kth nearest neighbor) is located at $t_0 + t$. Then, assuming co-located particles are impossible and using the observation in the above paragraph, we can write

$$p_t(t_0 + t|t_0) = \sum_{k=1}^{\infty} f_k(t). \tag{B.2}$$

Combining the two equations:

$$\eta(t) = -1 + \frac{1}{\lambda}\sum_{k=1}^{\infty} f_k(t). \tag{B.3}$$

Each of the $f_k(t)$ can be estimated from the observed inter-arrival distributions, thus allowing a computationally simple way to compute the pair-correlation function from particle arrival time data. (This function empirically is found to converge very rapidly, especially if pair-correlation function is desired for temporal scales up to just a few mean inter-arrival times. In practice, the upper limit of the sum can safely be set to a nominal value. For completeness, $k = 50$ was used for the upper limit of the sum in this study, even though no appreciable difference between $k = 10$ and 50 can be seen.)

References

Badger, E. H. M. (1946). Particle counts in the ultramicroscope. *Nature*, *157*, 480.

Baker, B. A. (1992). Turbulent entertainment and mixing in clouds: A new observational approach. *Journal of the Atmospheric Sciences*, *49*, 387–404.

Balkovsky, E., Falkovich, G., & Fouxon, A. (2001). Intermittent distribution of inertial particles in turbulent flows. *Physics Review Letters*, *86*, 2790–2793.

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

Chapman, H. M., & Ruhf, R. C. (1955). Dust counting reliability. *American Industrial Hygiene Association Quarterly*, *16*, 201–209.

Chun, J., Koch, D. L., Rani, S. L., Ahluwalia, A., & Collins, L. R. (2005). Clustering of aerosol particles in isotropic turbulence. *Journal of Fluid Mechanics*, *536*, 219–251.

Corrsin, S. (1951). On the spectrum of isotropic temperature fluctuations in an isotropic turbulence. *Journal of Applied Physics*, *22*, 469–473.

Cox, D. R., & Isham, V. (1980). *Point processes*. London: Chapman & Hall.

Elperin, T., Kleeorin, N., Liberman, M. A., L'vov, V. S., & Rogachevskii, I. (2007). Clustering of aerosols in atmospheric turbulent flow. *Environmental Fluid Mechanics*, *7*, 173–193.

Elperin, T., Kleeorin, N., L'vov, V.S., & Rogachevskii, I. (2007). Clustering instability of the spatial distribution of inertial particles in turbulent flows. *Physical Review E, 66*, 036302 1–16.

Fairs, G. L. (1943). The use of the microscope in particle size analysis. *Chemistry and Industry*, *62*, 374–378.

Friedlander, S. K. (1977). *Smoke, dust, and haze*. New York: Wiley.

Green, H. L. (1927). On the application of the Aitken effect to the study of aerosols. *Philosophical Magazine*, *4*, 1046–1069.

Herdan, G. (1960). *Small particle statistics* (2nd Rev. Ed., 418pp). New York: Academic Press.

Jameson, A. R., & Kostinski, A. B. (2000). Fluctuation properties of precipitation. Part VI: Observations of hyperfine clustering and drop size distribution structures in three-dimensional rain. *Journal of the Atmospheric Sciences*, *57*, 373–388.

Jameson, A. R., Kostinski, A. B., & Kruger, A. (1999). Fluctuation properties of precipitation. Part IV: Fine scale clustering of drops in variable rain. *Journal of the Atmospheric Sciences*, *56*, 82–91.

Kasper, G. (1984). On the coagulation rate of aerosols with spatially inhomogeneous particle concentrations. *Journal of Colloid and Interface Science*, *102*(2), 560–562.

Kostinski, A. B. (2001). On the extinction of radiation by a homogeneous but spatially correlated random medium. *Journal of the Optical Society of America, A*, *18*, 1929–1933.

Kostinski, A. B., & Jameson, A. R. (1997). Fluctuation properties of precipitation. Part I: Deviations of single size drop counts from a Poisson distribution (1997). *Journal of the Atmospheric Sciences*, *54*, 2174–2186.

Kostinski, A. B., & Jameson, A. R. (2000). On the spatial distribution of cloud particles. *Journal of the Atmospheric Sciences*, *57*, 901–915.

Landau, L. D., & Lifshitz, E. M. (1980). *Statistical physics* (3rd ed., 544pp). Oxford: Pergamon Press.

Larsen, M. L. (2006). *Studies of discrete fluctuations in atmospheric phenomena* (220pp). Ph.D. dissertation, Michigan Technological University.

Larsen, M. L., Cantrell, W., Kannosto, J., & Kostinski, A. B. (2003). Detection of spatial correlations among aerosol particles. *Aerosol Science and Technology*, *37*, 476–485.

Larsen, M. L., Kostinski, A. B., & Tokay, A. (2005). Observation and analysis of uncorrelated rain. *Journal of the Atmospheric Sciences*, *62*, 4071–4083.

Marshak, A., Knyazikhin, Y., Larsen, M. L., & Wiscombe, W. J. (2005). Small-scale drop-size variability: Empirical models for drop-size-dependent clustering in clouds. *Journal of the Atmospherics Sciences*, *62*, 551–558.

Nicas, M., & Hubbard, A. (2002). A risk analysis for airborne pathogens with low infectious doses: Application to respirator selection against *Coccidioides immitis* spores. *Risk Analysis*, *22*, 1153–1163.

Picinbono, B., & Bendjaballah, C. (2005). Characterization of nonclassical optical fields by photodetection statistics. *Physical Review A, 71,* 0.13812 1–12.

Preining, O. (1983). Optical single-particle counters to obtain the spatial inhomogeneity of particulate clouds. *Aerosol Science and Technology*, *2*, 79–90.

Scrase, F. J. (1935). The sampling errors of the Aitken nucleus counter. *Quarterly Journal of the Royal Meteorological Society*, *61*, 367–379.

Shaw, R. A., Kostinski, A. B., & Larsen, M. L. (2002). Towards quantifying droplet clustering in clouds. *Quarterly Journal of the Royal Meteorological Society*, *128*, 1043–1057.

"Student" (1907). On the error of counting with a haemacytometer. *Biometrica*, *5*, 351–360.

Sundaram, S., & Collins, L. R. (1997). Collision statistics in an isotropic particle-laden turbulent suspension. Part I. Direct numerical simulations. *Journal of Fluid Mechanics*, *335*, 75–109.

Warhaft, Z. (2000). Passive scalars in turbulent flows. *Annual Review of Fluid Mechanics*, *32*, 203–240.