Aerosol Science & Technology

ISSN: 0278-6826 (Print) 1521-7388 (Online) Journal homepage: https://www.tandfonline.com/loi/uast20

Detection of Spatial Correlations among Aerosol Particles

Michael L. Larsen , Will Cantrell , Jonna Kannosto & Alexander B. Kostinski

To cite this article: Michael L. Larsen , Will Cantrell , Jonna Kannosto & Alexander B. Kostinski (2003) Detection of Spatial Correlations among Aerosol Particles, Aerosol Science & Technology, 37:6, 476-485, DOI: 10.1080/02786820300958

To link to this article: https://doi.org/10.1080/02786820300958

Published online: 30 Nov 2010.

Submit your article to this journal

Article views: 148

View related articles

Citing articles: 9 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=uast20

Aerosol Science and Technology 37: 476–485 (2003)
© 2003 American Association for Aerosol Research
Published by Taylor and Francis
0278-6826/03/$12.00 + .00
DOI: 10.1080/02786820390126402

Detection of Spatial Correlations among Aerosol Particles

Michael L. Larsen, Will Cantrell, Jonna Kannosto, and Alexander B. Kostinski
Department of Physics, Michigan Technological University, Houghton, Michigan, USA

Most approaches for evaluating rates of fundamental processes in aerosol science depend upon the implicit assumption that aerosol particles are independently and identically distributed in space. The validity of this assumption has not been examined in several decades, despite the fact that the presence of correlations can be shown to alter theoretical expressions significantly for such phenomena as attenuation of electromagnetic radiation and coagulation rate. Here, we provide evidence that the classification of the positions of aerosol particles as a Poisson process—even under stationary conditions—is often in error. In particular, small-scale aerosol clustering is experimentally demonstrated. Some consequences of these findings are discussed.

INTRODUCTION AND MOTIVATION

The assumption of statistically uniform and independently distributed particle positions (perfect randomness) is ubiquitous in aerosol physics in general, and in aerosol sampling, in particular (e.g., Hinds 1999, p. 226). However, this fundamental assumption has been recently questioned in several areas complementary to aerosol physics, e.g., sedimentation in a laminar flow (Lei et al. 2001) and cloud physics (Kostinski and Jameson 2000). Therefore, in this article, our aim is to extract information on the spatial distribution of aerosol particles in a controlled laboratory experiment—to test the fundamental assumption of perfect randomness.

To that end, we view particle detection events as elements of a time series, then use tools of statistical physics to obtain spatial information. This procedure is motivated by recent studies which have questioned the validity of assuming that the spatial positions of cloud droplets are statistically independent. Processes which depend upon interactions between individual cloud droplets can be sensitive to correlations in spatial position. This is perhaps even more relevant in aerosol science, where chemical interaction as well as bulk physical behavior can be important. This potentially broader regime of influence motivates our study here.

An example of one of the processes with dependence on spatial distribution is the Beer-Lambert law of exponential attenuation of radiation. As shown in Kostinski (2001), the Beer-Lambert law using the mean concentration as the relevant parameter is strictly followed only when the material causing scattering and/or absorption is distributed in a completely random manner at all scales. Shifrin (1988) also discusses how the Beer-Lambert law is violated in dispersive systems, concluding that "absorption depends on the quantity of the absorbing substance traversed and also on how this substance is dispersed." The nature of the deviation (faster or slower than exponential) depends on the qualitative nature of the correlations, as was shown in Shaw et al. (2002a).

Another pertinent motivation for the study of aerosol spatial correlations was theoretically addressed by Kasper (1984), who noted that correlations (which he calls "inhomogeneities") in the spatial distribution of particles can cause an enhanced coagulation rate that is proportional to the degree of spatial correlation.

In addition, global climate models, radiative transfer theories, LIDAR studies, and even our conception of how of a size distribution evolves could be affected if an investigation reveals that other aerosol particles, besides cloud droplets, also tend to "cluster" in space.

There are undoubtedly other consequences of the existence of spatial correlations in aerosol particle positions. Nevertheless, the most commonly used forms of the distribution of aerosol particles implicitly assume that the particles are distributed in an uncorrelated (Poisson) manner at all scales. Indeed, the only experimental investigations that we have found (Green 1927; Scrase 1935) have concluded that particles are Poisson distributed. In the remainder of this paper, we demonstrate that we have found results that contradict this classification, which

Received 31 July 2002; accepted 9 December 2002.
This work was supported in part by the NSF grant ATM-0106271. We would like to thank Glenn Shaw at the University of Alaska, Fairbanks, for the loan of the optical particle counter used in the experimental component of this investigation.
Jonna Kannosto is now at Technological University of Tampere, Physics Department, Finland.
Address correspondence to Michael L. Larsen, Michigan Technological University, Department of Physics, Houghton, MI 49931. E-mail: mllarsen@mtu.edu

is consistent with the expectation of at least one other study (Balkovsky et al. 2001).

TIME SERIES ANALYSIS OF DISCRETE DATA

Before delving into detailed experimental analysis, we feel that a brief exposition on the nature of counting data may be warranted. Since our experiment involved temporal detection of *individual* aerosol particle arrivals in a sampling volume, the statistical methods utilized here may be unfamiliar in a field dominated by (continuous) concentration measurements.

Our fundamental tool—the pair-correlation function—measures scale-localized deviations from a Poisson process. The Poisson probability density function (pdf), that is, the probability of finding k particles in a time increment τ, is given by

$$p(k) = \frac{\overline{N}^k \exp(-\overline{N})}{k!}, \qquad [1]$$

where the mean count is $\overline{N(\tau)}$.

This formulation is, in some ways, a concentration-oriented approach as it is valid for any volume/timescale τ; one only has to adjust $\overline{N}$ accordingly. There are at least two other ways to approach the Poisson process which are more explicitly motivated by its inherent lack of memory. The first of these, the fact that interevent spacings are exponentially distributed, will not be used here. However, the Poisson process also fundamentally rests on the assumption that *particle counts for any two nonoverlapping volume elements are independent of each other*. This last interpretation naturally lends itself to two-point statistics in which the volume elements are shrunk to differential size (where "concentration" is no longer a meaningful quantity). For a more details see Shaw et al. (2002b).

The pair correlation function, $\eta(t)$, can be introduced with the following demonstration. Take two nonoverlapping volume elements dV_1 and dV_2in V that correspond to nonoverlapping sampling time increments dt_1 and dt_2 in T, each small enough that the probability of finding more than one particle in each volume element is negligibly small. Then, for any isolated volume element, the probability of finding a particle in time dt is $N\,dt/T \equiv \overline{n}\,dt$, where $N$is the total number of particles in T. For the Poisson probability distribution, the probability of finding a particle in volumes dt_1 and dt_2, which are separated by some time t, is given by

$$P(1,2) = \overline{n}^2(dt_1)(dt_2), \qquad [2]$$

since the probabilities are independent of each other. The pair-correlation function $\eta(t)$ is then defined via the expression (Green 1969)

$$P_t(1,2) = \overline{n}^2(dt_1)(dt_2)(1+\eta(t)), \qquad [3]$$

where P_t indicates the probability of codetection in time intervals dt_1 and dt_2.

We use $\eta(t)$ because of its physical meaning as the scale-localized (function of t only) deviation from a Poisson process as demonstrated in Figure 1. (For more details on pair-correlation function, see Landau and Lifshitz 1980; Kostinski and Shaw 2001).

Although physically intuitive, this formulation of the pair-correlation function does not have an obvious interpretation in the more "traditional" viewpoint with N (number of counts in a macroscopic volume) as the fundamental random variable. It seems such an extension *should* exist, for though the introduction of differential volumes $dV_1$and dV_2 and the limitation that $\overline{n} \ll 1$ is convenient for developing a well-defined pdf for $P_t(1,2)$, nothing in the above formulation relies on physical phenomena at differential scales. An extension to arbitrary volume sizes can, indeed, be made by appealing to the correlation-fluctuation theorem (Landau and Lifshitz 1980; Ornstein and Zernike 1914):

$$\frac{\overline{(\delta N)^2}}{\overline{N}} - 1 = \overline{n}\int_0^t \eta(t')\,dt', \qquad [4]$$

where $\overline{(\delta N)^2}$ is the variance of the counts in the volume. By rewriting this equation, we can define the volume-averaged pair correlation function:

$$\overline{\eta(t)} \equiv \frac{1}{t}\int_0^t \eta(t')\,dt' = \frac{\overline{(\delta N)^2}}{(\overline{N})^2} - \frac{1}{\overline{N}}. \qquad [5]$$

Rearranging the above equation shows that the mean of counts in a (potentially macroscopic) volume is related to the variance by

$$\overline{(\delta N)^2} = \overline{N} + \overline{\eta(t)}\,\overline{N}^2. \qquad [6]$$

This is a particularly fortuitous relationship, as $\overline{(\delta N)^2}$ and $\overline{N}$ are easy statistical quantities to obtain and work with.

Recall that one of the often utilized properties of the Poisson distribution is that the variance equals the mean. Consequently, from the above equation, something is Poisson distributed on scale t if $\overline{\eta(t)}$ vanishes for that t. A Poisson process, however, is characterized by $\eta(t) \equiv 0$ for all t. These two classifications are subtly, but critically, different. Regimes of positive correlation and negative correlation of equal magnitude and duration, both existing at scales less than t, can result in a Poisson distribution at scale t. However the existence of any correlations on any scale at all necessarily prevents classification of a distribution as a realization of a Poisson process. This differentiation can identify differences in how aerosol size distributions have evolved.

Calculating $\eta(t)$ directly from the definition proves to be a daunting task. In practice, we compute $\eta(t)$ from the operational formulation given in Kostinski and Jameson (2000):

$$\eta(t) \equiv \frac{(\overline{N(t+\tau)N(\tau)} - \overline{N}^2)}{\overline{N}^2} = \frac{\overline{N(t+\tau)N(\tau)}}{\overline{N}^2} - 1. \qquad [7]$$

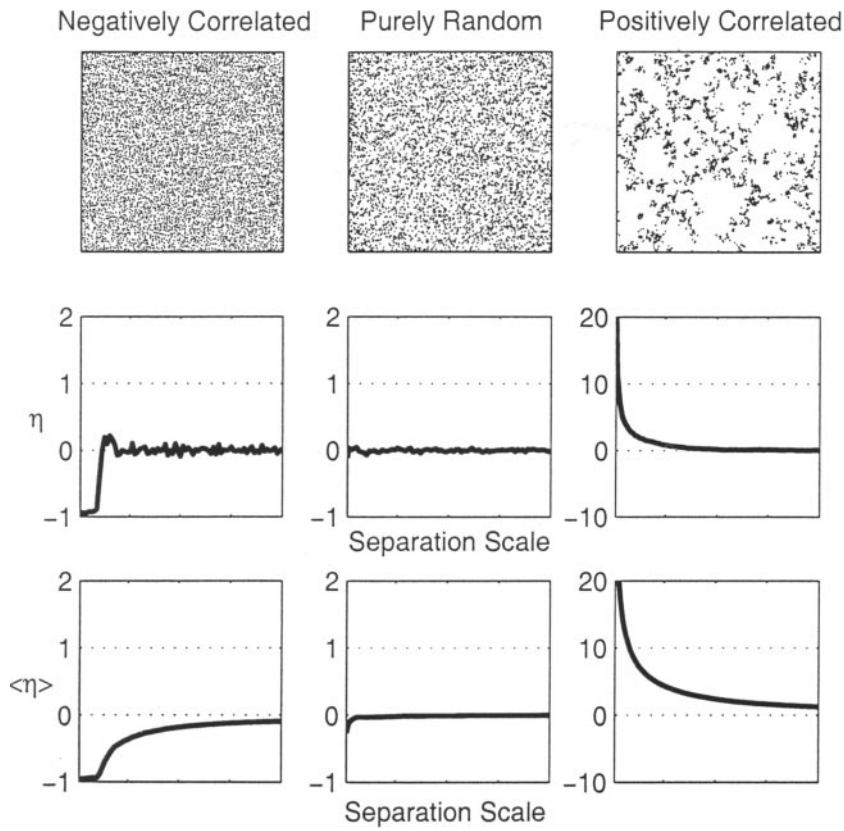

Figure 1. Three simulations. The left column examines a negatively correlated medium (i.e., a medium where the variance of counts for a sample volume is smaller than the mean count in the sample volume). The middle column depicts a purely random (Poisson) process, and the right column a positively correlated medium. The first row of the figure shows the (in these simulations) two-dimensional sample. Note that in each distribution the amount of particles is the same. The second row plots the pair-correlation function for each distribution. The x-coordinate corresponds to the distance (or time) lag between the volumes of interest (i.e., t in Equation (3)). In the last row, we have plotted the volume-averaged pair-correlation function for each distribution. This function, as noted in the text, has extensive "memory" of smaller spatial scales, thus leading to a long $1/t$-like decay from any localized fluctuation in η.

Alternatively, one can use the volume-averaged pair-correlation function as an estimator for η. Special care must be taken in using $\overline{\eta}$ to give information about η, however. As can be seen from Figure 1 or by carefully examining Equation (5), $\overline{\eta}$ has what we call "memory" of shorter temporal scales. Unlike η, which depends only on t, $\overline{\eta}$ depends on all temporal scales from zero to t. Whereas η could fluctuate arbitrarily rapidly, $\overline{\eta}$ "adjusts" from a point-fluctuation in η as $1/\tau$, where τ is the separation time between the current scale and that of the point fluctuation in η.

With both $\overline{\eta}$ and the operational formulation of η, care must be taken in analysis for short time scales. Because a particle counter cannot both count and record continuously, the data must be discretized on some scale. This discretization introduces a degree of ambiguity in the precise arrival time t of the particle. As such, the differential volumes alluded to in deriving Equation (3) cannot be achieved in practice. In computing the pair-correlation function via the operational formulation given above we inherently introduce some localized averaging into the process. For most of the domain of interest this is a negligible factor since all averaging is only over the instrumental resolution lag. Below, however, we will find that most of the clumping is at lags beneath the instrument's resolution. The high value of η tabulated at the first time scale *has memory due to using the operational formulation* (see, e.g., Figure 6). For the first lag time, η and $\overline{\eta}$ as computed are identical as can be seen by looking at all graphs with these quantities in this paper. At this first lag, averaging over the instrument resolution is equivalent to the average from zero to the scale of interest.

As such, unfortunately, nothing can be said about how correlations are distributed on the scales that we cannot resolve: η could be relatively constant over that regime, or most of the correlations could come from a range much shorter than the current resolution; there is no way to know until temporal resolution can be increased. Nevertheless, a positive value for η as computed via the operational formulation above does give

acceptable evidence for net clustering effects on subresolution scales since the averaged pair-correlation function over this range is positive.

Now that these tools have been introduced, how can they be applied to aerosol science? In our concluding remarks we give instruction as to how this study can be directly applied to the determination of the coagulation rate. If a rate-governed process is only dependent on making a Poisson assumption at one given scale (like coagulation; e.g., Kasper 1984), it simply needs to be shown that η vanishes for that scale. If the process depends on having no net correlations *up to* a specific scale (like extinction; e.g., Kostinski 2001; Shaw et al. 2002a), the correct quantity to compute would be $\overline{\eta}$.

DATA ANALYSIS

We made measurements in room air of aerosol particles with a CLIMET CI-8060 optical particle counter which covers the size range 0.3 to 10 μm diameter in 5 size bins. In examining this data, we view each size bin as a time series of aerosol particle arrival times that one could, in principle, invert to obtain spatial information. (This inversion can be complex and nonunique; as such we will primarily limit ourselves to conclusions of the nature that "it does have correlations.")

To ensure a well-mixed environment we blew air across the sample inlet tube with a small fan. The data selected in this study was taken from the largest (5–10 μm) size bin. Recall that one of the requirements for η to be meaningful is that the probability of finding any particles in a unit volume is equal to the probability of finding one particle in a unit volume. The bin used in this investigation generally had a concentration on the order of 0.05 particles/measurement, assuring that this dilute assumption is justified.

The functions η and $\overline{\eta}$ were calculated for each data set. We then discarded data sets that did not pass our simple test for stationarity, as the pair-correlation function depends on the assumption of wide-sense stationarity. When appropriate, uniformlooking subsets of the data were also examined to see if any regime without correlations could be identified. All data was also plotted along with a Monte Carlo simulation of a Poisson process with the same meancount for qualitative comparison with the data. This simulation generates integer-valued random variables from the Poisson pdf with the mean fixed equal to that of the raw (real) data. (η and $\overline{\eta}$ were computed for this simulation as well for comparison purposes.)

All of the data we analyzed could loosely be broken into two categories: (1) those which displayed significant correlations over a broad range of scales, and (2) those that showed only extremely short-scale correlations. Typical data and physical interpretation for three scenarios are given below.

Figures 2, 3, and 4 display the raw time series data analyzed in the following sections. On each graph, the horizontal axis represents time elapsed since the first measurement of the sample. Within each data set, the sample exposure time was held constant. It should be noted that this data isn't rigorously continuous, as after each measurement a (uniform duration) pause in counting was necessary to write the data to disk.

Figure 5 is a histogram of the frequency of occurrence for each particle count in the most dense data analyzed here (that shown in Figure 2). (This sort of histogram analysis is comparable to the analysis done in Green 1927; Scrase 1935). Note that the vertical axis is logarithmically scaled in Figure 5.

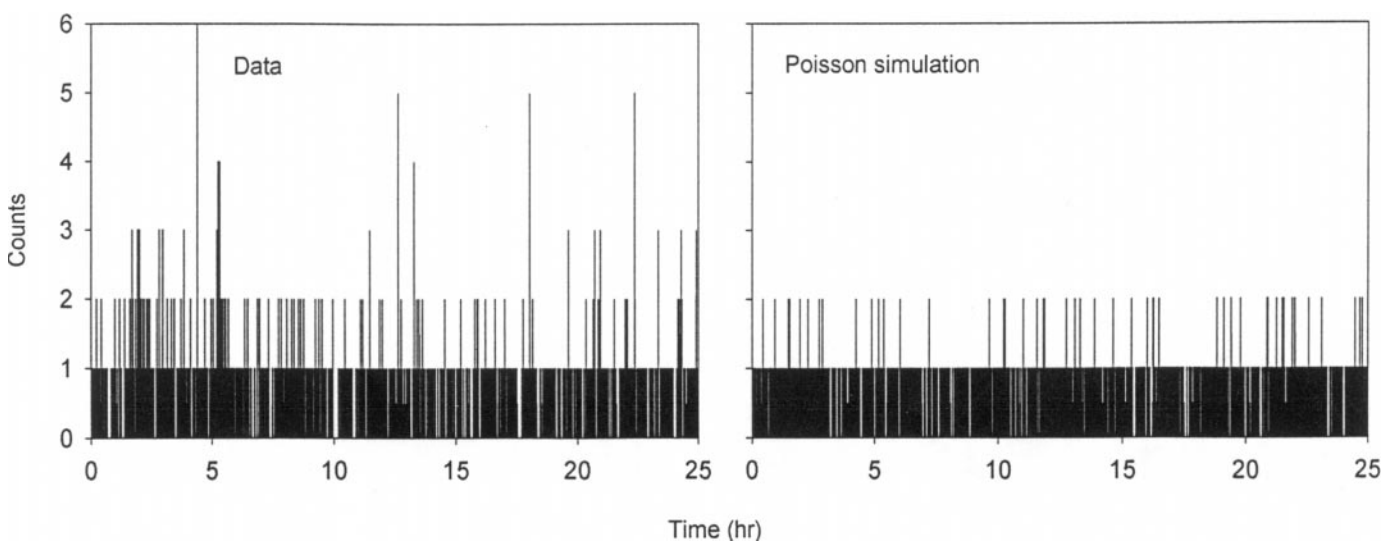

Figure 2. The raw data analyzed in the section "First Evidence of Correlations." For this data, each data point is a particle count for 5 s. Though the flow continued uninterrupted, the measurement process was paused for 1 s to write information to disk, followed by another measurement of 5 s duration, etc. The mean is only 0.09 counts/measurement. The left panel is raw data, the right panel a Poisson simulation generated using Equation (1).

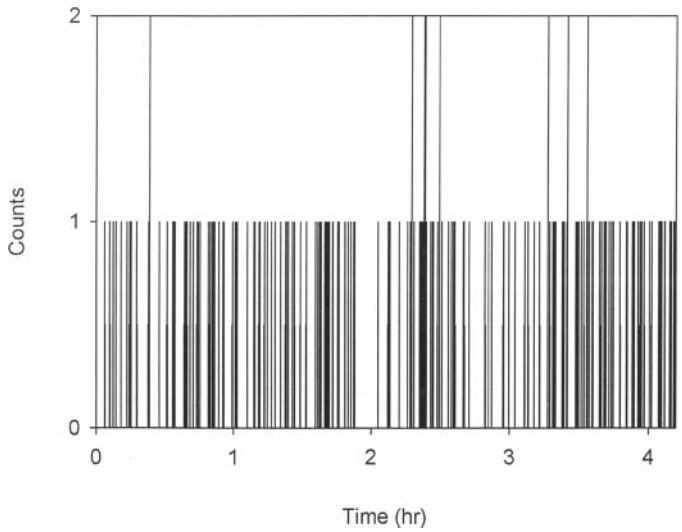

Figure 3. Similar to Figure 2, this shows the segment of the larger data set analyzed in the section "Evidence for Poisson Distribution." Note that here measurements were taken every second. Therefore, the data is sparser; the mean count here being only 0.025 counts per measurement.

Note also the long tails of the observed data when compared to the Poisson simulation. Even given this, however, less than one percent of the measurements contain more than one detection, thus assuring that η as defined by Equation (3) is meaningful.

Without making any rigorous statements, a casual inspection of Figures 2, 3, and 4 lend credence to the claim of stationarity in the wide sense. Our criterion for stationarity was based on plotting the "local" mean and variance of the distribution and verifying that these values fell within two standard deviations of the global values for these variables. This assures that not too much of the fluctuation in the data set was isolated in one area of the distribution. When combined with visual evidence and after verification that the autocorrelation function decays to zero, we argue that we have sufficient evidence for the steady-state stationary condition in these distributions. The reason this is such a sensitive point is because there are many stationarity criterion that would erroneously identify a stationary distribution with

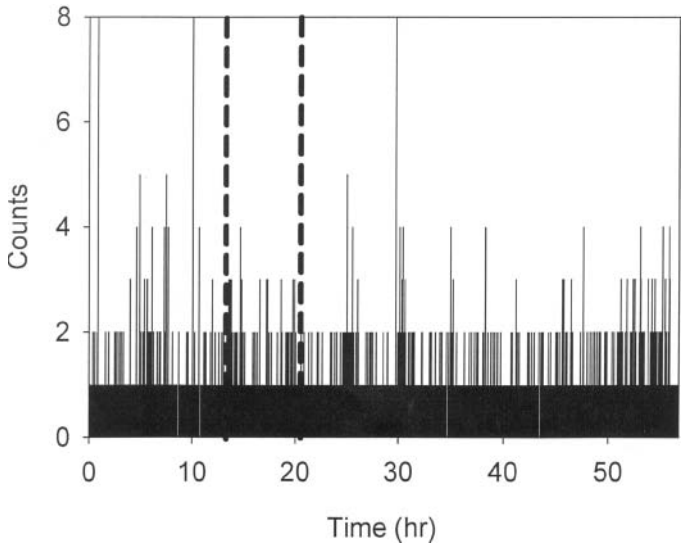

Figure 4. As in Figure 3, each measurement in this data had a duration of only 1 s. Results are analyzed in the section "Large-scale Variability Propagation." The dotted lines specify the boundary for the subset analyzed therein. The mean count (per measurement) for the entire duration was 0.044 and 0.038 for the subset.

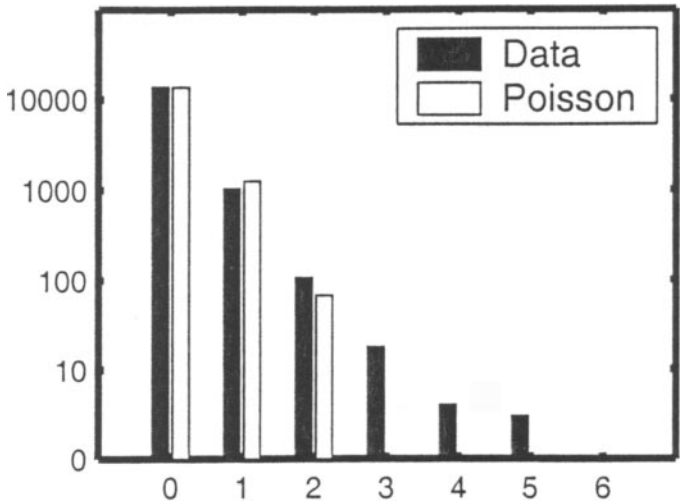

Figure 5. A histogram of the data shown in Figure 2. The black bars correspond to the frequency of occurrence for each value in the data set. The white bars show the frequency of occurrence for a Poisson process simulation with the same mean as the data. Note that only about 0.9% of the data measurements contain >1 detection, indicating that the small dt assumption required for the pair-correlation function to be meaningful under the definition given in Equation (3) is legitimate. Nevertheless, one consistently sees longer tails in the data than in Poisson process simulations. This gives some basis for the hypothesis that correlations on time-scales shorter than the measurement duration exist. See the text for more details.

correlations (what we have) as a distribution with a deterministic trend (see, for example, Wunsch 1999).

First Evidence of Correlations

The pair-correlation function and the volume-averaged pair correlation function for the data shown in Figure 2 are plotted in Figure 6. This graph shows a constant positive value for η that persists from $\sim$10 to 100 s and a region of correlations at short time lags. Recall that η is a measure of obtained "successes" over expected "successes," thus a value of 2 means that the probability of finding two particles that distance apart is thrice the probability of the same event if the particles were Poisson distributed on that scale. In this light, these magnitudes are somewhat surprising.

The constant, positive value of η through time scales is discussed in more detail below but here we note that the high values of η and $\overline{\eta}$ at short lags could be critical to many physical processes. It might be tempting to discount this as a numerical anomaly, but this feature shows up in every data set we analyze. Furthermore, it does not appear when running the Poisson simulation, which generates data in a form exactly analogous to the experiment. A close investigation reveals that almost all of this short-scale correlation is coming from the first (and sometimes second) lag scale. The decay in η is quite abrupt, and Figure 7 indicates that essentially all of the memory in $\overline{\eta}$ can be traced back to that first calculated value.

The observed phenomenon is exactly what one would expect if the clumping principally exists on scales lower than the resolution size. This conclusion is supported by the rest of our findings as well.

Evidence for Poisson Distribution

The data in Figure 8 comes from a subset of a larger data set given in Figure 3 where the sample time is 1 s, not 5 s as in Figure 2. This subset was chosen because it appeared to contain no "clumps." In other words, the data looked as if it were Poisson distributed. Indeed, as far as η is concerned, there is no significant difference between the data and the Poisson simulation for most of the domain plotted.

Although it still persists for quite some time, $\overline{\eta}$ decreases to a steady value much sooner here than in Figure 6. Furthermore, one could argue an asymptotic approach to zero is reached already around 50 s, well over an order of magnitude faster than the data for the section above, "First Evidence of Correlations." Consequently, an argument can be made that we are seeing a *Poisson distribution* at scales corresponding to $t \geq \sim 50$ s. It is not, however, a Poisson process, as correlations for very short times exist.

Large-Scale Variability Propagation

In the above two sections we have shown evidence for correlations on several scales. The correlations on short scales (one or two lags) can be anticipated on physical grounds due to a number of phenomena; two typical mechanisms that might be at work are inertial response to high vorticity from turbulence and correlations induced by spatially localized sources. However the second type of correlation, that which persists over timescales up to half an hour, may seem anomalous. An explanation of the meaning of a constant value of η persisting over long scales needs to be established.

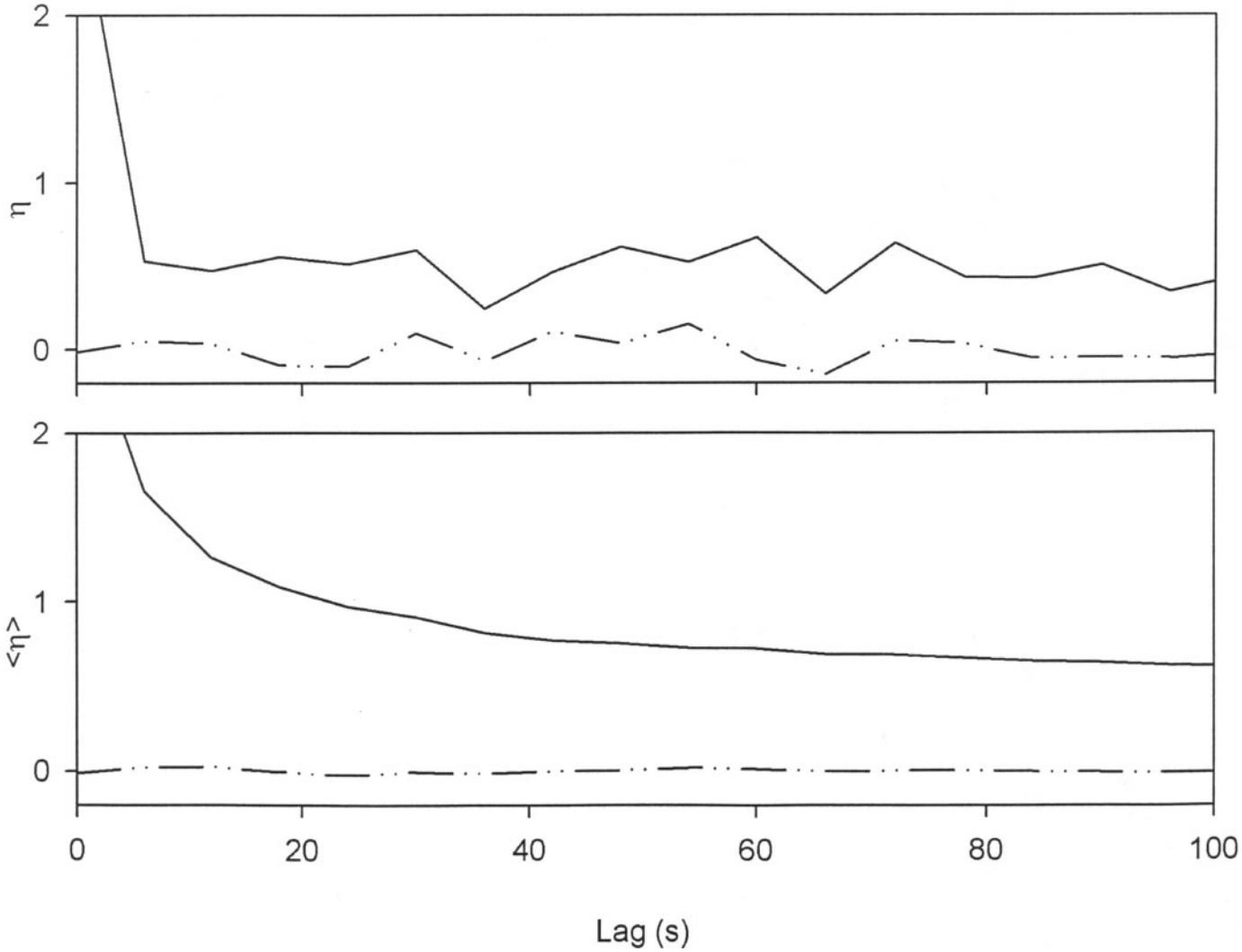

Figure 6. The pair-correlation function and its average for data that displays evidence of persistent and subresolution correlations. Here the upper (solid) line in both plots corresponds to data while the lower (dashed) line is that of a simulated Poisson process. Despite the appearance here, there is a very slow decay in η for the real-data curve. It decays approximately linearly to zero at time $t \approx 1/2$ hr, at which point it is virtually indistinguishable from the Poisson simulation. The experimental data used to compute these curves is given in Figure 2.

To that end, we have graphs for two distributions, shown in Figure 9. The first (on the lefthand side) displays statistics for the entire traverse of the data taken from Figure 4. On the righthand side we have the statistics from a uniform-looking subset of that distribution (specified by the region between the dotted lines in Figure 4). To better establish that the graph for η on the lefthand side of Figure 9 does show nonnegligible super-Poisson correlations through all length scales and that the figure on the righthand side does not, we have plotted bounds for the Poisson simulation instead of a specific value. These curves were generated by calculating the statistics from 100 independently generated Poisson processes, all with the same mean as the data in the appropriate range. Then, these curves were chosen so that, for each plotted point, exactly half of the Poisson simulations were inbetween these bounds (with equal amounts above and below the given bounds). A "rule-of-thumb" approximation, then, is that a Poisson process should stay between these curves a non-negligible fraction of the time. Barring Poisson behavior at all scales, for spatial scales that data would be Poisson *distributed*, the curve should also lie between these two curves much of the time. The volume-averaged pair-correlation function is displayed in the same manner as previously (compared to one realization of a Poisson process), since the long memory of $\overline{\eta}$ prevents the curves from overlapping.

One can draw the conclusion, based on the above explanation, that since the curve representing all data (upper left) almost never enters the regime specified by the "Poisson bounds," this data is not Poisson distributed for any statistically

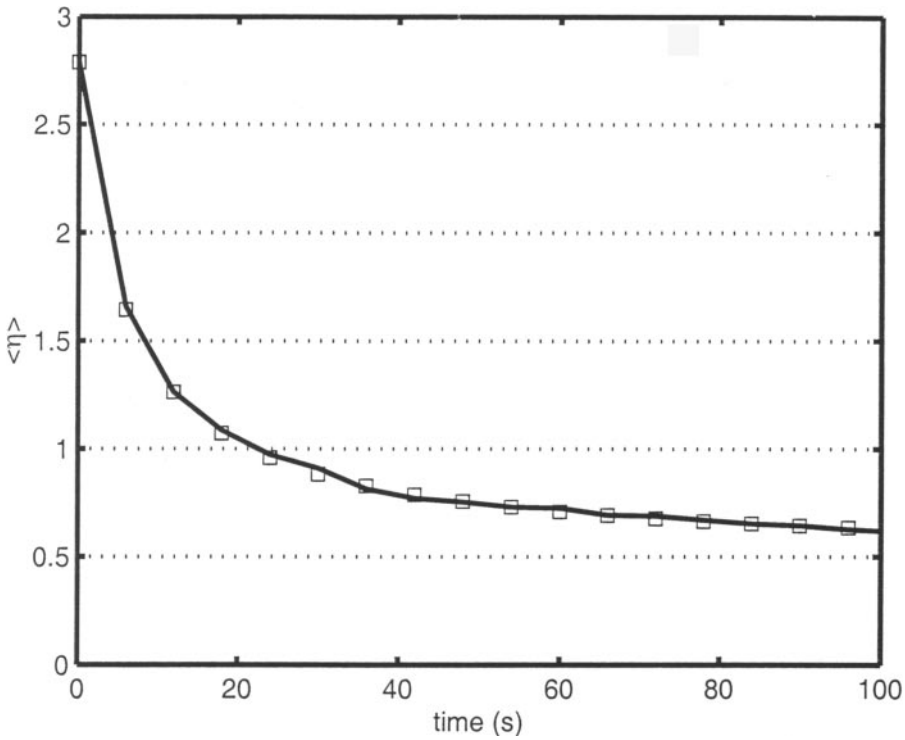

Figure 7. The solid line above represents the same $\overline{\eta}$ shown in Figure 6. The boxes indicate an approximation to this curve calculated using only the first point for η and then setting $\eta \equiv 0.5$ (its local constant value) after that. (Note the $1/t$ nature of the decay.)

significant length of time shown, thus specifying a phenomena similar to what is observed in Figure 6. It seems, however, that the curve on the righthand side is Poisson distributed at all scales except the very smallest, the sort of distribution observed in the preceding section. What is occurring here? Both of the phenomena that we observed in the above sections happen concurrently. One of these is an effect reflecting some sort of short-range correlation, whereas the other is due to large-scale variations in the data that do not surface when examining the subset.

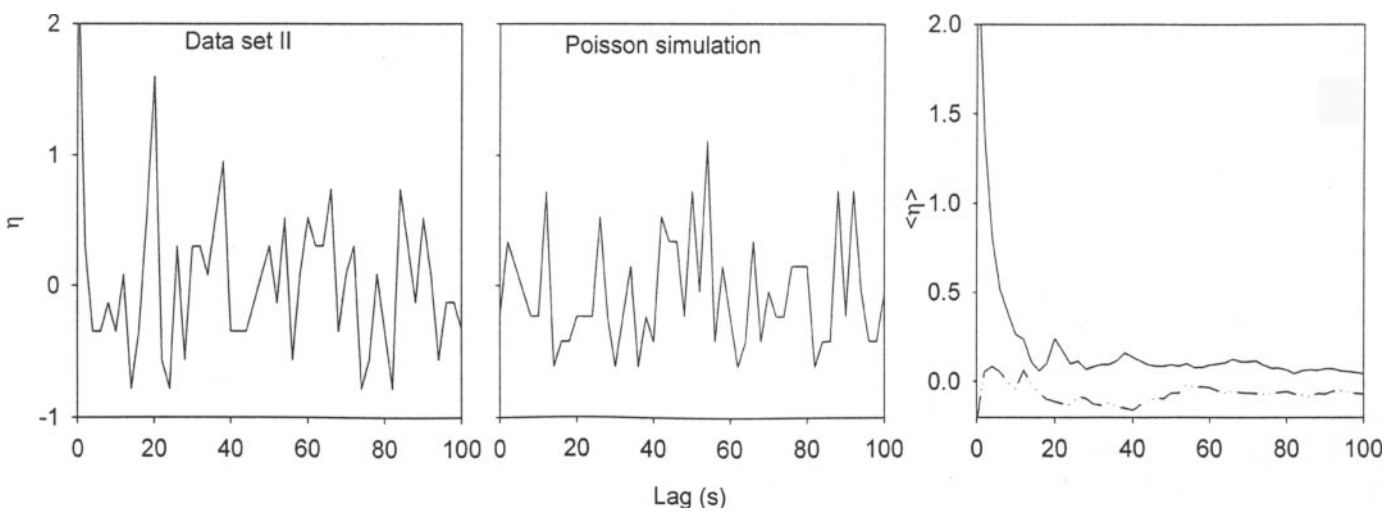

Figure 8. Comparison of data from a near-Poisson process to a Poisson simulation. The left panel is η for the data, the middle panel is η for the simulation, and the right panel is $\overline{\eta}$ for both (simulation is the dashed lower line). Note that the peak for $\eta(0)$ still exists for the data, despite increased resolution; this is indicative of the fact that most clustering is at scales still below the counter resolution. On all three graphs, the independent variable is time in seconds. The raw data for these statistics is given in Figure 3.

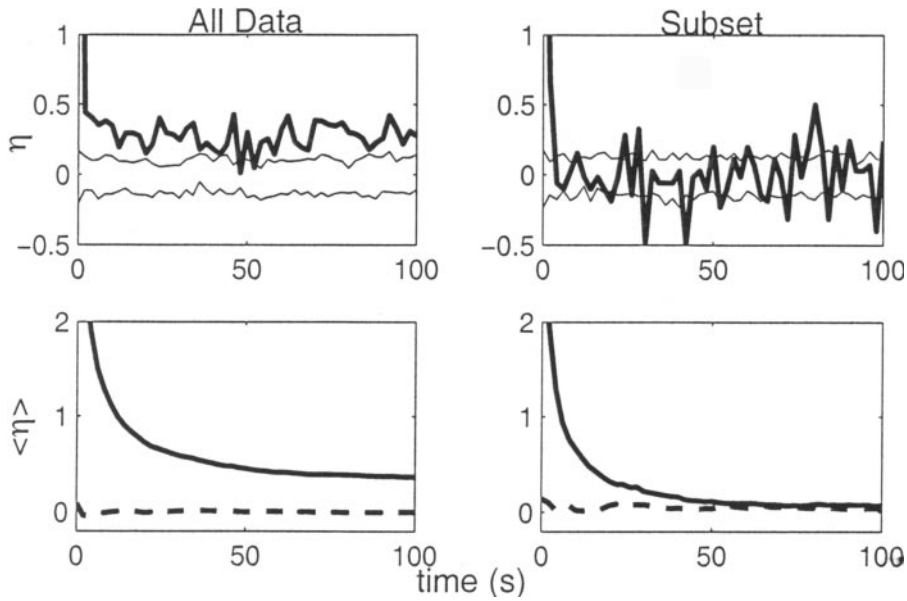

Figure 9. The statistics for the data-set shown in Figure 4. The lefthand panels are the statistics for the whole traverse, and the righthand panels are the statistics for the subset identified by the dotted lines. Instead of comparing the upper curves (displaying η) to one simulated realization of a Poisson process, we have displayed the original data set and "Poisson bounds" (specified by the lighter lines), as described in the text. In the curves for $\overline{\eta(t)}$, the dashed curves are the Poisson process simulations.

Conceptually, correlations can be looked at two ways. $\eta(\tilde{T}) > 0$ can indicate that the existence of a particle at time $t = t_0$ biases the probability of obtaining a particle at time $t = \tilde{T} + t_0$ by a factor of $1 + \eta(\tilde{T})$. Though this is how we have looked at the problem so far, it is equally legitimate to think of the absence of particles. If you have an empty volume element dV_1, then $\eta(\tilde{T}) > 0$ indicates that you have an enhanced probability of finding another empty volume element dV_2 at temporal separation $\tilde{T}$ from dV_1. (If more double detections than expected are made, and the same concentration is held throughout, it implies that there are less occurrences than expected of detecting only one particle in the two volumes.) It is this effect that causes the enhancement for η at all visible scales in Figure 6 and the lefthand side of Figure 9.

To illustrate this further, imagine a distribution that is initially a realization of a Poisson process. Now take this distribution and mix in gaps—extended areas of no particles. These gaps must be distributed in an ergodic manner so that one is still left with a stationary, random distribution. (To do this, keep the gaps of sufficient size to be non-negligible in length compared to the distribution as a whole.) Recall $\eta(t)$ can be defined as the number of realizations of a joint detection in volumes $dV_1$and dV_2 detected a time t apart, divided by the Poisson expectation for this quantity with unity subtracted off. By mixing in large gaps, few of these joint realizations are lost (especially for small t) since we are in a sparse medium and we mixed the gaps in randomly. Though some detections have been eliminated, the expected value for a Poisson distribution is lowered because the global mean value is decreased due to the gaps. As such, the same amount (or actually slightly fewer) mutual detections can still yield a value for η that is positive because of the lower expectation. Therefore, *a persistent, constant value of η can indicate large-scale variability*. Indeed, a cursory analysis of the data-set that generated the lefthand side of Figure 9 reveals a period of over 200 consecutive measurements without detection of a particle, corresponding to a gap of around 0.2% the length of the distribution.

In a Poisson process simulation for the same mean and duration as the data given in Figure 4, 51 sequences were found with more than 100 consecutive zeros. In the actual data there were 109 such occurrences. The elimination of these excess "holes" alone would bring the mean for the entire distribution up by ~10%.

CONCLUDING REMARKS

Given the above analysis, we conclude that correlations do exist between aerosol particles at subresolution scales. Though some of the enhanced values of the pair-correlation function at lags beyond a few seconds are due to large-scale fluctuations, the peaks at the shortest time scales (first lag value corresponding to 2 s) are a real phenomenon. Note that the existence of spatial correlations does not necessarily imply a physical interaction between particles, nor does it preclude "randomness," e.g., see Kostinski and Jameson (2000). Aerosol particles are inevitably stochastically distributed, just not via a Poisson process. The question as to whether or not a stochastic (e.g., turbulent) system is able to completely decorrelate an initially correlated

suspension of particles is still open to debate, and the time-scales of measurement here are far too short for diffusion to smoothen the gradients. As such, the origins for these correlations could even be in the aerosol production process.

So why is it that previous studies (Green 1927; Scrase 1935) found no evidence of clustering? The sample size of Scrase (1935) was only 0.25 cm^3, roughly 5×10^{-4} our sample size. Furthermore, we made calculations for a range of data; both other studies only had relevant information for one spatial scale. Because of this, their analysis was limited to an analogue of $\overline{\eta}$. As can be inferred from Equation (5), the existence of a Poisson distribution on one scale can be a result of canceling positive and negative correlations on shorter scales. Since this spatial scale is below our current resolution, it is also possible that there happens to be no correlations on this scale, and all of the correlations we observe exist on scales between their resolution and our first calculated value. Finally, in both Green (1927) and Scrase (1935), aerosol particles were identified by condensing water vapor onto an already present aerosol particle, thus growing it to a size where it can be reliably counted. (To do this, Green (1927) used a Wilson cloud chamber and Scrase (1935) used an Aitken nucleus counter). However, in the mildest case, Scrase (1935) was nucleating on all aerosols $\geq\sim$390 nm. This is a much broader particle size range than we examined, and examination of such a wide range could cause correlations to be masked.

Earlier, we made the claim that the pair-correlation function contains information about deviations from the coagulation rate. Indeed, Kasper (1984) defined a quantity $\beta_i = \overline{n^2}/(\overline{n})^2$, which identifies the change in coagulation rate due to concentration "inhomogeneities" (correlations, in our terminology), where n is a particle concentration based off a scale large enough where the concept of Brownian coagulation is meaningful (i.e., $n > 1$). We can express the coagulation enhancement via the pair-correlation function as $\beta_i \equiv (\eta/\rho) - 1$, where ρ is the (density) autocorrelation function. (See Shaw et al. (2002b) for more details about the relationship between η and ρ.) Hence, the pair-correlation function affects coagulation rate.

The sign of $\overline{\eta}$ (for absorbers) is associated with different radiation attenuation regimes (faster or slower than exponential), as is demonstrated in Kostinski (2001) and Shaw et al. (2002a). In this study, all macroscopic scales satisfy $\overline{\eta} > 0$ for aerosol particles, which would indicate slower than exponential extinction. It is still an open question as to whether or not extinction through such a medium must necessarily converge to exponential decay with a "modified" optical depth. At the very least, as observed in Kostinski (2002), a nonzero $\overline{\eta}$ implies at least a transition regime where Beer's law cannot be safely assumed for propagation through such a medium.

In summary, data analysis reveals large-scale correlations as well as intriguing subresolution clustering. This evidence is sufficiently strong for dismissing the assumption that the spatial distribution of aerosol particles is perfectly random. Therefore, any treatment of the aerosol counts in a unit volume as a realization of a Poisson process cannot be completely accurate. This could alter theoretical predictions for such processes as coagulation and extinction.

REFERENCES

Balkovsky, E., Falkovich, G., and Fouxon, A. (2001). Intermittent Distribution of Inertial Particles in Turbulent Flows, *Physics Review Letters* 86:2790–2793.

Green, H. L. (1927). On the Application of the Aitken Effect to the Study of Aerosols, *Philosophical Magazine* 4:1046–1069.

Green, H. S. (1969). *The Molecular Theory of Fluids*, Dover, New York.

Hinds, W. C. (1999). *Aerosol Technology*, 2nd ed., Wiley, New York.

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(8):1929–1933.

Kostinski, A. B. (2002). On the Extinction of Radiation by a Homogeneous but Spatially Correlated Random Medium: Reply to Comment, *Journal of the Optical Society of America A* 19(12):2521–2525.

Kostinski, A. B., and Jameson, A. R. (2000). On the Spatial Distribution of Cloud Particles, *Journal of the Atmospheric Sciences* 57(7):901–915.

Kostinski, A. B., and Shaw, R. A. (2001). Scale-Dependent Droplet Clustering in Turbulent Clouds, *Journal of Fluid Mechanics* 434:389–398.

Landau, L. D., and Lifshitz, E. M. (1980). *Statistical Physics*, 3rd ed., Pergamon Press, Oxford.

Lei, X., Ackerson, B. J., and Tong, P. (2001). Settling Statistics of Hard Sphere Particles, *Physical Review Letters* 86(15):3300–3303.

Ornstein, L. S., and Zernike, F. (1914). Accidental Deviations of Density and Opalescence at the Critical Point of a Single Substance, *Proc. Akad. Sci.* 17:793–806.

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., and Lanterman, D. D. (2002a). Super-Exponential Extinction in a Negatively Correlated Random Medium, *Journal of Quantitative Spectroscopy and Radiative Transfer* 75:13–20.

Shaw, R. A., Kostinski, A. B., and Larsen, M. L. (2002b). Towards Quantifying Droplet Clustering in Clouds, *Quarterly Journal of the Royal Meteorological Society* 128:1043–1057.

Shifrin, K. S. (1988). *Physical Optics of Ocean Water*, American Institute of Physics, New York.

Wunsch, C. (1999). The Interpretation of Short Climate Records, with Comments on the North Atlantic and Southern Oscillations, *Bulletin of the American Meteorological Society* 80(2):245–255.