11-1-2005

Observations and analysis of uncorrelated rain

Michael L. Larsen
Michigan Technological University

Alexander Kostinski
Michigan Technological University

Ali Tokay
NASA Goddard Space Flight Center

Follow this and additional works at: https://digitalcommons.mtu.edu/physics-fp
Part of the Physics Commons

Observations and Analysis of Uncorrelated Rain

MICHAEL L. LARSEN AND ALEXANDER B. KOSTINSKI

Department of Physics, Michigan Technological University, Houghton, Michigan

ALI TOKAY

Joint Center for Earth Systems Technology, University of Maryland, Baltimore County, Baltimore, and NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 9 July 2004, in final form 3 May 2005)

ABSTRACT

Most microphysical models in precipitation physics and radar meteorology assume (at least implicitly) that raindrops are completely uncorrelated in space and time. Yet, several recent studies have indicated that raindrop arrivals are often temporally and spatially correlated. Resolution of this conflict must begin with observations of perfectly uncorrelated rainfall, should such "perfectly steady rain" exist at all. Indeed, it does. Using data with high temporal precision from a two-dimensional video disdrometer and the pair-correlation function, a scale-localized statistical tool, several ~10–20-min rain episodes have been uncovered where *no* clustering among droplet arrival times is found. This implies that (i) rain events exist where current microphysical models can be tested in an optimal manner and (ii) not all rain can be properly described using fractals.

1. Introduction

In most microphysical theories regarding raindrop evolution it is implicitly assumed that the distribution of raindrops is perfectly random (i.e., the rain is perfectly steady or uncorrelated on all scales). This lack of correlation is a basis for incoherent scattering (addition of intensities) often employed in radar meteorology. In addition, under "equilibrium" conditions, Z–R relations should take a linear form (Atlas and Chmela 1957; List 1988; Jameson and Kostinski 2001; Steiner et al. 2004). Furthermore, uncorrelated rain is implicitly assumed in most descriptions of evolving drop size distributions (Kostinski et al. 2005). Finally, an important practical problem in rain microphysics is accounting for sampling bias in rainfall estimation. A study recently submitted by Uijlenhoet et al. (2005) gives an analytic solution to this problem but, again, only for uncorrelated rain. Clearly, these examples demonstrate that the associated theories need to be tested against measurements in uncorrelated rain; measurements in correlated rain do not suffice. But what exactly is perfectly uncorrelated rain and does it exist in nature?

Following a suggestion in Jameson and Kostinski (2002) we call the uncorrelated rain "perfectly steady" when it is statistically stationary and devoid of correlations at any scale. We render the latter notion precise below by employing the pair-correlation function. Despite several studies on the microphysical texture of rain, no empirical study to date has identified rain events that are perfectly steady. Here we supply evidence supporting the existence of such rain.

We begin with a brief literature review. While the proper description of rainfall's microstructure is still a matter of some debate, most studies agree that raindrops are usually clustered or correlated in some way. This is sometimes used to justify a fractal description of rain. Lovejoy et al. (2003) examined well-resolved three-dimensional images of rain fields, concluding that rain shows fractal scaling down to about 40 cm. (At terminal velocity of the drops, this distance would correspond to a time scale of a few tenths of a second.) Lavergnat and Golé (1998) used modified disdrometer data to show power-law (fractal) correlated behavior on temporal scales above 0.5 s. In an earlier study,

Corresponding author address: Michael L. Larsen, Department of Physics, Michigan Technological University, 1400 Townsend Dr., Houghton, MI 49931.
E-mail: mllarsen@mtu.edu

Zawadzki (1995) used a Joss–Waldvogel disdrometer to give evidence of fractal behavior down to scales around 1 s. However, these scaling approaches encountered a break from "simple" scaling (Lovejoy and Schertzer 1995) on scales of order 1 s, and resolution was not sufficient to determine whether the change at 1 s was a mere scale break to a new fractal dimension for shorter times, an artifact of poor instrumental resolution, a change to perfect randomness, or something else entirely. A nonfractal approach based on theory of stationary random processes was used in Kostinski and Jameson (1997) who demonstrated clustering on scales from 1 min to 0.01 s (Jameson and Kostinski 2000).

The only evidence to date for nonclustered rain behavior (to the best of our knowledge) can be found in Kostinski and Jameson (1997) and Uijlenhoet et al. (1999). Both of these studies, however, found random (Poisson distributed) behavior at only one sampling resolution and only for specific raindrop sizes. Such observations are insufficient to classify rain as perfectly steady. The observation of raindrop clustering on progressively smaller time scales prompted Jameson and Kostinski (2002) to conjecture that small-scale clustering may be nearly ubiquitous and that so-called perfectly steady rain (rain devoid of clustering on any scale for *all* drop sizes) might not occur in nature. In summary, thus far neither classical nor fractal methods have given evidence for a purely random distribution of drops regardless of drop size.[1]

In this paper, we present observations that strongly suggest that uncorrelated rain *does*—at least occasionally—occur in nature. Perfectly random behavior has been verified for all drop sizes over time scales spanning five orders of magnitude (0.6 ms to 1 min). These data (i) conclusively demonstrate that not all rain obeys fractal statistics and (ii) serve as a guide to find data suitable for validation of theoretical models discussed above.

2. Theoretical background

a. *Perfectly steady or uncorrelated rain*

The primary result of this paper is the positive identification and observation of perfectly steady rain segments; the primary tool for data analysis is the pair-correlation function. Therefore, we begin with a brief review of the essentials. Following the suggestion in Jameson and Kostinski (2002), rain is called perfectly steady if it is (i) statistically stationary (constant or "steady" rainfall rate) and (ii) there are no correlations between raindrops. Since here we are primarily concerned with the microstructure of rain where the fundamental discreteness is important, we test for stationarity of the actual droplet counts (regardless of size) rather than for a continuous variable such as the rainfall rate.

The two requirements above imply that droplet arrivals in steady rain constitute a *Poisson process*—a statistical series that, regardless of chosen scale, results in a Poisson distribution. A statistical tool examining data for scale-dependent departure from a Poisson process (pure randomness) would be ideal for checking whether a given rain event is steady; we have such a tool in the pair-correlation function.

b. *The pair-correlation function*

The pair-correlation function (pcf) is a tool devised for determining the magnitude of the deviation from a Poisson distribution at a given spatial or temporal scale. In previous work the pcf has often been introduced in the language of statistical homogeneity and spatial statistics but, given the format of disdrometer data, here we present it in the language of wide-sense stationarity and arrival times. For the purposes of this paper, however, we use the notions of homogeneous (for spatial data) and stationary (for temporal data) wholly interchangeably. It is one of the features of the pair-correlation function that one can alternate between a temporal and spatial description changing little but the notation.

To define the pair-correlation function, we begin with a sample containing N raindrops detected in total time T. We then select two small disjoint time intervals, $d\tau_1$ and $d\tau_2$, separated by time t. If the sample were a realization of a Poisson process, the probability of having each time interval occupied is $N(d\tau_i/T)$ (assuming $d\tau_i$ is small enough to ensure there is at most one raindrop in the interval). Additionally, for a Poisson process the occupations of the two time intervals are independent so we can conclude that the joint probability density function is given by the product of the individual densities, for example,

$$p_{\mathrm{poiss}}(1, 2) = \left(\frac{N}{T}\,d\tau_1\right)\left(\frac{N}{T}\,d\tau_2\right) = c^2 d\tau_1 d\tau_2, \qquad (1)$$

with c being the arrival frequency N/T (the temporal equivalent of number concentration), and still assuming that each subinterval is sufficiently small to ensure it contains at most one raindrop.

We then define the pair-correlation function through the relation

[1] The very use of fractal methods implicitly includes the assumption that such rain does not exist (see the discussion of homogeneity in appendix A).

$$p_t(1, 2) = c^2 d\tau_1 d\tau_2 [1 + \eta(t)], \tag{2}$$

with $\eta(t)$ being the value of the pcf at separation time t, and $p_t(1, 2)$ the probability that both subintervals are occupied in the dataset being examined. Note that $\eta(t)$ can take any value larger than -1 (which corresponds to mutual exclusion). If the pair-correlation function evaluated at t is greater than zero, it implies that if a raindrop is encountered at a certain time t_o, there is an enhanced probability of finding another raindrop at time $t_o \pm t$.

Perhaps more accessible than the formal definition above is the "heuristic" definition of the pair-correlation function. If we examine the ratios of the probabilities used above, we find that

$$\frac{p_t(1, 2)}{p_{\text{poiss}}(1, 2)} = \frac{d(t)}{r(t)} = [1 + \eta(t)], \tag{3}$$

where $d(t)$ is the number of pairs separated by time t in the dataset to be analyzed and $r(t)$ is the number of pairs expected to be separated by time t in a Poisson process with the same number of total particles and total duration as the data.[2]

The latter, heuristic definition shows particularly clearly the advantage of using the pcf in searching for uncorrelated rain; for a true Poisson process (perfectly steady rain) $d(t) = r(t)$ and the pcf is *identically 0 for all lags*. If $\eta(t)$ is found to be zero for all length scales of interest, we have sufficient evidence to establish the dataset to be a Poisson process; the rain is steady. Conversely, if we see significant departures from $\eta(t) = 0$, we find a degree of structure/clustering larger than one would expect to arise from natural, purely random variability.[3] Though the ultimate cause of the clustering may not be known, we can conclude that the rain is not steady.

One of the basic assumptions in employing the notion of the pair-correlation function (or even using the language of "correlations" instead of concentration fluctuations) is that of statistical stationarity for the dataset being examined. This point (couched in the language of spatial statistics and "homogeneity" instead of "stationarity") has been discussed in some detail elsewhere (e.g., Shaw et al. 2002; Kostinski et al. 2005), but is worth reexamining given the fundamental disagreement between the (spatially) homogeneous classical approach and the inhomogeneous fractal approach. Within this paper, we have chosen to work in the homogeneous classical framework of the pair-correlation function in part because we are searching for perfectly steady rain, which, by definition, *must be* statistically homogeneous and wide-sense stationary. Using a tool that relies on the statistical inhomogeneity of the dataset to search for homogeneous, correlation-free subsets of data would be, at the very least, counterintuitive. Note that blind calculation of the pcf for nonstationary data, however, will give nonzero values. Consequently, the detection of $\eta(t) = 0$ for all scales is sufficient to classify rain as uncorrelated and perfectly steady. A more complete discussion of homogeneity/stationarity and related concerns are presented in appendix A.

Aside from the claim of statistical stationarity, the pair-correlation function is an assumption-free tool for the scale-by-scale examination of distributions of discrete random variables. However, when trying to estimate the true pair-correlation function from given data, other constraints and assumptions are introduced due to the finite duration of all real datasets. These limitations are discussed in appendix B. We note, however, that the pcf's shown in the following section satisfy the basic criteria established in the appendix.

3. Analysis of observations

To search for perfectly steady rain, we examined 17 events recorded by a two-dimensional video disdrometer (2DVD) in 2002 at Wallops Island, Virginia. A thorough discussion of this instrument can be found in Kruger and Krajewski (2002). For the purposes of this paper, it is sufficient to note that this is an instrument that (among many other things) records arrival times of individual raindrops with 1-ms precision. The sensing area is small enough ($\sim$100 cm^2) to assume that the detected rain is distributed uniformly over the surface of the detector.[4]

To obtain sufficient drop numbers to use the pair-

[2] We call this the heuristic definition because, formally, the number of pairs expected to be exactly separated by any time interval is 0; this conundrum is resolved by the introduction of some averaging time; e.g., we count the number of particles separated by some lag $t \pm (\delta t/2)$. Note that this technicality does not plague the first definition of the function since probability density functions are used in that case.

[3] A natural question that arises from this statement is "what constitutes a significant departure from $\eta(t) = 0$?" We address this question in appendix B.

[4] For the 2DVD, the two-dimensional coordinates of each drop are recorded. We find that for all subsamples examined here, the entire surface of the detector is well represented. Reanalysis of the data using only part of the detector yields similar results to that which follow, so concerns about horizontal clustering obscuring the conclusions drawn below are unfounded. It is worth noting, however, that horizontal clustering on scales larger than the sensing area of the 2DVD is possible.

TABLE 1. Basic information about the data subsets used in this paper. For each dataset, $\overline{D}$ is the mean diameter of raindrops, N is the total number of raindrops, $\overline{\tau}$ is the mean interarrival time between raindrops, and $\overline{R}$ is the mean rain rate.

Label	Date (2002)	Duration (min)	$\overline{D}$ (mm)	N	$\overline{\tau}$ (s)	$\overline{R}$ (mm h^{-1})
Dataset 1	8 Feb	21	1.066	4492	0.2805	1.47
Dataset 2	12 Mar	31	0.831	8915	0.2086	0.69
Dataset 3	13 Mar	15	0.963	4810	0.1871	1.35
Dataset 4	20 Jan	27.5	1.069	5001	0.3300	1.11
Poisson simulation	—	20	—	6000	0.2000	—

correlation function effectively, we required that the candidates for an uncorrelated rain event be relatively persistent. Furthermore, by definition perfectly steady rain does not have large drop-number fluctuations on any scale; if significant fluctuations would occur on moderate (e.g., 1 min) time scales, the data could not be a realization of a stationary Poisson process—it is not uncorrelated. Therefore, we tagged potentially uncorrelated rain events by binning the data for the 17 datasets into 1-min drop counts, and then searched for at least 10 consecutive minutes that have at least a 10 drops per minute increment and that all have approximately the same number of drops, allowing for some natural random variability. While admittedly ad hoc, this method was found to give satisfactory results. It can be formally written as follows:

$$|q_{i+n} - q_i| \le 2\sqrt{q_i} \quad \forall n \in [1, 10], q_i \ge 10, \tag{4}$$

where q_i is the number of drops in the ith 1-min interval.

Among the 17 days of data, three subsets satisfying the above criteria were found. Additionally, there were several other subsets that nearly satisfied this requirement but showed slightly more variability in the minute-to-minute drop count statistics. The three subsets that qualified (datasets 1–3), along with one of the subsets with more variability (dataset 4) and a simulated Poisson distribution with similar parameters to the data selected, are analyzed in the rest of this paper. Details regarding the subsets are given in Table 1.

Since the 2DVD records arrival time data, the appropriate statistic to calculate is the temporal pair-correlation function. It is important to use appropriate small time increments since an increment that is too long can violate the assumption that there is at most one particle in each increment, and an increment that is too short can result in oversampling the recorded data (see appendix B).

Figure 1 shows the pair-correlation function $\eta(t)$ as a function of separation time t for $d\tau = 0.6$ ms. Using this $d\tau$ might appear questionable since the data were only recorded with a precision of 1 ms but it is apparent that the selected subsamples [panels (a)–(c)] show no evidence of net deviations from $\eta = 0$ on temporal scales

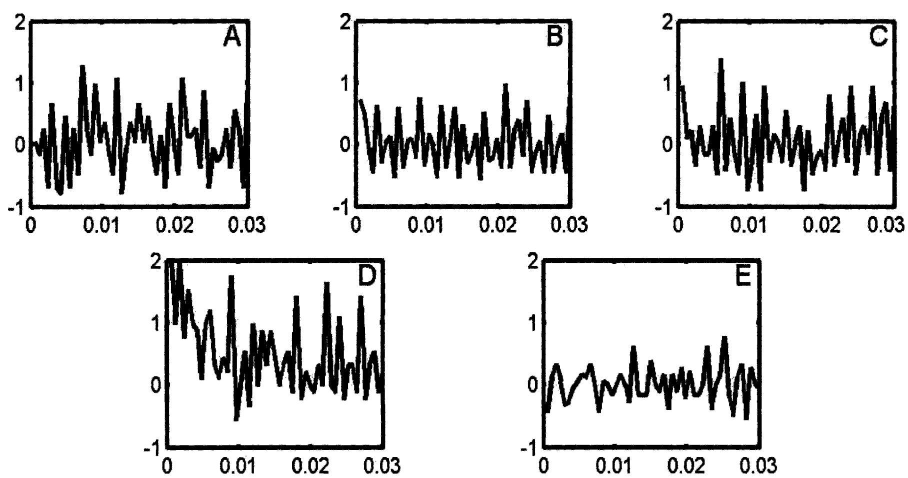

FIG. 1. The pair-correlation function evaluated as a function of temporal lag (in s) for the datasets described in Table 1. In all figures in this paper, datasets 1–4 correspond to (a)–(d) and the Poisson simulation is shown in (e). In this figure, the averaging temporal scale $d\tau = 0.6$ ms. The large fluctuations result from using a $d\tau$ smaller than the precision of how the data were recorded. Nevertheless, the behavior in (a)–(c) seems qualitatively different than in (d).

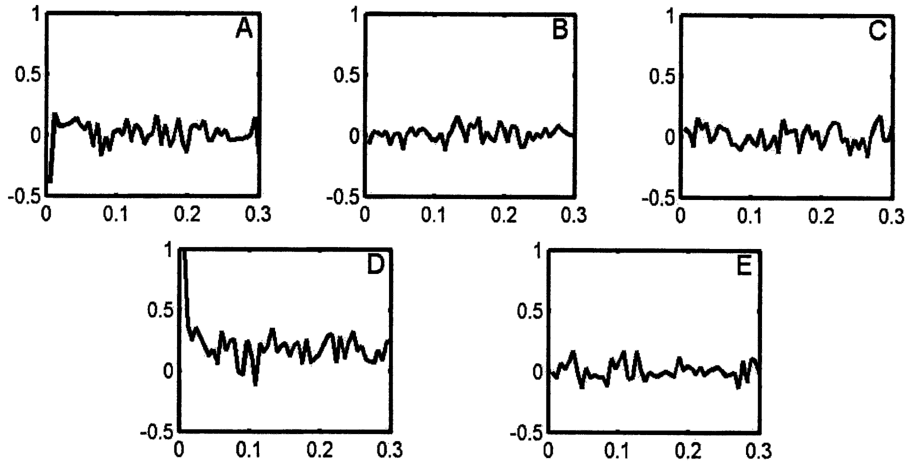

FIG. 2. As in Fig. 1 except with $d\tau = 6$ ms. Note that, from Table 2, this $d\tau$ best satisfies the inequalities established for most appropriate use of the pair-correlation function established in appendix B. Once again, we find no significant deviations from pure random behavior in (a)–(c), but a nonnegligible signal of positive clustering in (d).

of less than 0.03 s. Dataset 4 [panel (d)] seems to show a nonnegligible departure from $\eta = 0$ for very short time scales. All four curves are more noisy than that for the Poisson simulation [panel (e)] because we are sampling the data with a shorter time scale than the precision it was recorded with.

Figure 2 is similar to Fig. 1 except that we now use a $d\tau = 6$ ms ten times larger than that for Fig. 1. Note that now the fluctuations in all of the datasets are comparable to the Poisson simulation; small fluctuations of about $\eta = 0$ exist even in the Poisson simulation, due to the finite size of the sample (this is examined in some detail in appendix B). Once again, we see a departure from pure randomness for the shortest time increments only, in dataset 4. The same results are found when τ is again increased by a factor of 10–60 ms as Fig. 3 shows.

To verify that datasets 1–3 show no correlations at scales between 1 s and 1 min, we resort to histograms as illustrated in Fig. 4. We use this method instead of the pair-correlation function because, for $d\tau \gtrsim 100$ ms the probability of getting more than one particle in an interval may no longer be negligible and renders the pair-correlation function meaningless (see appendix B).

The probability distribution function for a Poisson process of observing k raindrops in time interval $d\tau$—where the mean count (μ) is $c\ d\tau$—is given by

$$p(k) = \frac{(c\,d\tau)^k}{k!}\exp(-c\,d\tau). \tag{5}$$

If the dataset in question obeys the Poisson statistics, the number of intervals with k particles should be given by the total number of intervals multiplied by $p(k)$. This is shown for each of the four datasets and the Poisson simulation for $d\tau = 1$ s and $d\tau = 5$ s as solid lines in Fig. 4. The actual recorded numbers of intervals with k drops are also shown with vertical bars. For a Poisson process, the vertical bars should be a reasonable approximation to the solid line [see panel (e) for an example]. If there is significant (positive) clustering at the scale of interest, the vertical lines should have longer, more pronounced tails.

This method of detecting scale-dependent deviations from a Poisson process has been used before; Green (1927) and Scrase (1935) used this approach to search for deviations from pure randomness in the spatial positions of aerosol particles. Although perhaps more intuitive than the pair-correlation function, this method is also more qualitative and perhaps not as reliable. [Using other methods, Preining (1983) and Larsen et al. (2003) independently demonstrated results that contradicted the findings of both of the earlier studies above.] These caveats aside, Fig. 4 shows excellent agreement between datasets 1–3 and the Poisson frequency expected. For time scales greater than 5 s, the subsets are too small to have enough subdivisions to estimate the distribution function well even for the Poisson simulation.

4. Discussion and conclusions

The analysis given in the previous section presents convincing evidence that uncorrelated rain does exist; datasets 1–3 show no significant deviations from Poisson statistics on time scales from 0.6 ms to 1 min. Although we only have three identified events at this

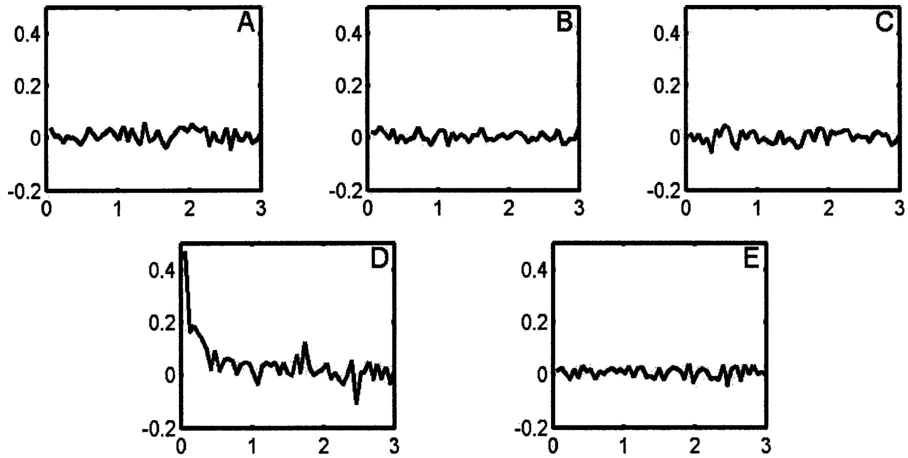

FIG. 3. As in Figs. 1 and 2 but with $d\tau = 60$ ms. Yet again, we see what appears to be perfectly random behavior for (a)–(c), with correlations appearing in (d). Note that here, as in Fig. 2, (a)–(c) show fluctuations about 0 around the same magnitude as the perfectly random simulation.

point, additional observations can be made. It seems that when uncorrelated rain does occur, it can be fairly persistent—lasting to about half of an hour in one location. However, it may also be rather rare (at least at Wallops Island, Virginia; the uncorrelated rain found corresponds to only about 1% of the total fraction of raindrops observed in the 17 events). Yet, these events are ideal for study of radar reflectivity–rainfall rate relations, drop-size distribution (DSD) evolution, and theories of collision coalescence. For example, during homogeneous and steady rain, Atlas and Chmela (1957), List (1988), and Jameson and Kostinski (2001) noted that relationships between radar reflectivity factor and rainfall rate (Z–R relations) should be physical (rather than merely statistical) and linear. Although the data we have analyzed do not persist long enough for verification of this fact, if a disdrometer "farm" were to measure a rain event that was classified as steady, a linear relationship should be found when scatterplotting Z versus R for sufficiently long time intervals. (Typical Z–R studies examine a best-fit line through at least several hundred data points; to gather a similar number of data points with each sample measuring over a long enough interval to accurately approximate the DSD in each measurement, many disdrometers would have to work in tandem.)

It is also notable that raindrop size distributions converge to the true distribution (should one exist) most rapidly when being measured in uncorrelated rain. Studies of cloud droplet clustering (e.g., Knyazikhin et al. 2005) and aerosol clustering (e.g., Preining 1983) seem to indicate that atmospheric constituents exhibit correlations that depend (among other things) on the size of the particle of interest; as such, estimates of the droplet size distributions are destined to be biased if the sample DSD is taken from correlated rain. It is only in uncorrelated rain that one can get the optimal "$\sqrt{N}$" convergence that we often rely on for statistical accuracy.

Recently, Kostinski et al. (2005) suggested that the classical coagulation equation may be insufficient for a full description of drop growth by collision and coalescence. However, the assumptions made in establishing droplet evolution described by this equation are least in error for a perfectly steady, correlation-free collection of drops. In short, if classical coalescence theory can ever be used, it should be able to most accurately model the times immediately after observed uncorrelated rain. Careful examination of DSDs surrounding periods of uncorrelated rain could give better insight into whether raindrop clustering significantly influences raindrop growth.

In the introduction of this paper, it was noted that many of the fractal approaches to rain microstructure identify a deviation from simple scaling at scales of order 1 s. The studies here have not shown any type of signal occurring at those scales. However, the implicit assumption of inhomogeneity in the fractal approach to the statistical microstructure of rain may be in question. Appendix C offers one possible reason for not seeing signatures on time scales of 1 s, but the key point is that there is no compelling reason to expect rain's microstructure to follow a fractal scaling. The observation of perfectly steady rain—by construction of a Poisson process—gives evidence that not all rain is fractal.

Despite the observations of perfectly steady rain reported here, most rain is *not* uncorrelated. It has frequently been argued that this is quite possibly due to convection and turbulence. Additionally, Kostinski et

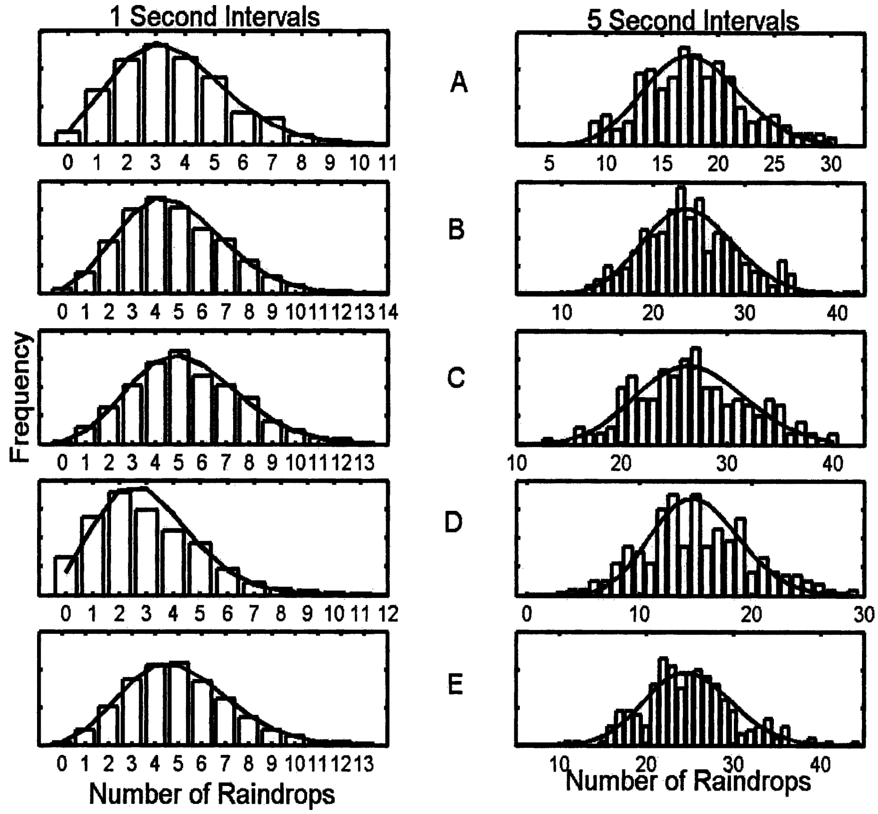

FIG. 4. The qualitative test outlined in the text. Given are histograms of droplet counts in increments of (left) 1 and (right) 5 s. The solid black line shows the values expected if the dataset were to be perfectly random. Though all of these curves look like quite good approximations, the long tails in the right-hand column of (d) might be some evidence of non-Poissonian behavior. The inconclusive nature of this method points to the utility of using the pair-correlation function to classify deviations from pure randomness, when possible. As noted in the text, the analyzed subsets of data do not persist long enough to do this sort of analysis for time intervals of more than about 5 s.

al. (2005) suggested that another origin of the correlated behavior found in most rain may arise from a nonequilibrium between fragmentation and coalescence, even in so-called steady conditions. While at times in error, conventional microphysical treatments of rainfall nevertheless rely—at least implicitly—on the perfect spatial and temporal randomness of raindrops. If we want to test current theories of rain microphysics, we must begin with examining rain events that are nearly perfectly steady. Hopefully, the data and methods used above will serve as a first step toward that goal.

Acknowledgments. This work was supported in part by NSF Grant ATM01-0106271 (AK and ML), the Michigan Space Grant Consortium (ML), and the National Defense Science and Engineering Grant Fellowship Program (ML). One of the authors (AT) wishes to acknowledge the assistance of Paul Bashor of Computer Science Corporation and Katherine R. Wolff of Science Systems Applications, Inc., for 2DVD data collection and compilation.

APPENDIX A

On Statistical Stationarity

The assumption of statistical stationarity (or statistical homogeneity for spatial data) of the underlying dataset is explicit when using the pair-correlation function. How legitimate is this assumption? Such a question does not cast doubt on the conclusions reported here since a deviation from $\eta = 0$ would be found in data that are not statistically stationary. Nevertheless, such objections bring up the broader question of how to determine whether data are more appropriately described by statistically stationary or nonstationary methods.

Formally, a process is wide-sense stationary if its expected value is constant and its autocorrelation depends only on the time difference between events (Papoulis 1965, p. 302). The first property—the mean count over a time interval remains constant—is usually the key assumption in atmospheric science and is notoriously difficult to verify since there will always be some natural variability of this mean count. In particular, if the sample time interval is too short, fluctuations that are possibly a result of correlations alone can cause an estimate of the mean count to vary throughout the sample—even when the dataset is rigorously stationary. Examples of the pitfalls of using too critical and severe limitations for stationarity are given in Wunsch (1999).

Rather than addressing the stationarity of a dataset in some objective fashion, it is usually simply assumed that it either is or is not stationary from the outset. This is potentially misleading,[A1] even though as Bras and Rodriguez-Iturbe (1993, p. 4) point out, homogeneity is "the first and most common assumption [in hydrology and other geophysical sciences]." As alluded to earlier, however, a fractal description of a system cannot be considered statistically homogeneous (Pietronero 1987; Coleman and Pietronero 1992; Martinez and Saar 2001) even though fractal sets do have some properties of homogeneity/stationarity. Quoting from p. 264 of Pietronero (1987):

> For fractal structures therefore the number density function is highly singular and even its average is not a well defined quantity. The system is instead characterized by the mass-length relation and the conditional density. These properties are homogeneous in a conditional sense. In fact, [the mass–length relation and the conditional density] hold from every point of the system considered as origin. This means that every point of the system has the same type of environment but this is not true any more for a point that does not belong to the system. . . . This implies an asymmetry between points belonging to the system and points not belonging to the system but homogeneity for all points of the system.

Less formally, the essential point above is that wide-sense stationarity requires the number density to be constant (Papoulis 1965) whereas in a fractal system the number density is not even well defined (Pietronero 1987). In particular, a Poisson process is nonfractal. Using a fractal description is inappropriate for a physically homogeneous (stationary) system, just as using a homogeneous description of an inhomogeneous process is in error (Pietronero 1987; Coleman and Pietronero 1992).

The formal definitions for stationary and homogeneity can be quite precise, but when the governing distributions are not known a priori we have little guidance as to when one can classify the data as stationary. Here, we opt to make a short list of properties that a statistically homogeneous set of observations should exhibit. Doing so gives necessary—not sufficient—criteria for a dataset to pass and appropriately be labeled as stationary. For a given set of data to be considered statistically stationary, one should at least verify that the necessary conditions below are satisfied. Conversely, there is at least some justification for using an inhomogeneous approach such as fractals if it can be shown one of the following conditions are violated.

First, we have that

$$\left|\lim_{T\to\infty}\int_0^T \eta(t)\,\mathrm{d}t\right| < \infty. \tag{A1}$$

This tries to assure that the limit of the integral of the pair-correlation function both exists and is finite. The final two conditions are related to the characteristic correlation scale, q. We define this scale by

$$q \equiv \{\min t\colon \eta(t) = k\}, \tag{A2}$$

where k is usually taken to be either 0 or 1. The remaining conditions, then, are

$$q \neq q(T) \quad \text{and} \tag{A3}$$

$$q \ll T, \tag{A4}$$

where T is the duration of the dataset analyzed.

Equation (A3) ensures that the characteristic correlation scale is not a function of sample duration. If one does find that $q = q(T)$, it generally indicates that there is clustering on all scales present in the sample, and an inhomogeneous approach is likely more appropriate. The final condition, Eq. (A4), is the oft-cited requirement that the correlation time must be much shorter than the duration of the dataset analyzed. If it is found that these times are close to each other, an inhomogeneous approach again may be appropriate. In the case of steadylike rain, of course, all of the above conditions are trivially satisfied since we have that $\lim_{T\to\infty}\int_0^T \eta(t)\,\mathrm{d}t = 0$, and $q = 0$.

Dataset 4 from the main text of this paper furnishes a less trivial example for the application of these tests.

[A1] A short anecdote might help motivate this claim. Recently, a paper submitted on the spatial structure of rain received two reviews. The first questioned the need for a study of this sort since it was so widely know that the microstructure of rain is perfectly random (a homogeneous Poisson process). The other reviewer claimed the manuscript was useless because it was obvious that rain is statistically inhomogeneous, clustered on all scales, and the use of a homogeneous measure of texture had no practical utility.

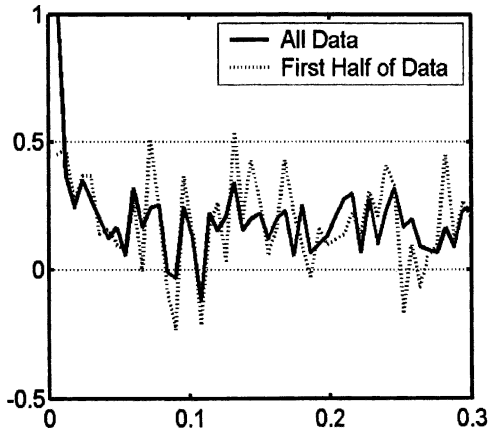

FIG. A1. A figure to help aid in understanding the necessary conditions for stationarity outlined in appendix A. Plotted is the pair-correlation function of dataset 4 against lag time (in s) for $d\tau = 6$ ms. The solid line shows the pcf as evaluated from using the entire 27.5-min subset. The dotted line shows the pcf using only the first 13 min of this subset. One could conclude that the dataset was likely not stationary if one of these curves decayed markedly faster than the others. Here, however, the decay seems comparable in the two curves (though the noisiness in the shorter subset makes it hard to tell for certain).

Figure 3 seems to verify that Eq. (A1) is satisfied. The final condition is obviously met since we see a decay to $\eta \approx 0$ well within 3 s from Fig. 3, whereas the data subset is about half an hour long (Table 1). To verify the second condition, see Fig. A1 where the pair-correlation function is plotted for $T = 27.5$ min and $T = 13$ min. (Further plots are out of the question because for shorter T the sample pcf is not well behaved; see appendix B.) We see that, though noisier, the two curves are reasonably close; the curve for shorter T does not seem to decay markedly faster so that q is independent of T. Consequently, dataset 4 passes all of the necessary conditions for homogeneity outlined above and dismissing it as "obviously" inhomogeneous is perhaps premature; while dataset 4 was found to be "not uncorrelated," the raindrop arrival statistics may be statistically stationary.

APPENDIX B

On the Use of the Pair-Correlation Function

a. When is the pair-correlation function a meaningful statistic?

As noted in section 2, analysis based on the pair-correlation function requires that the expected number of raindrops detected in time interval $d\tau$ be small enough so the probability of finding two or more drops in $d\tau$ is negligible. To appreciate the stringency of this condition, we examine a Poisson process. Then if k is the number of drops, we have from Eq. (5) that $p(k > 1) = 1 - p(k = 0) - p(k = 1)$ or, for mean count $\mu = c\, d\tau$, $p(k > 1) = 1 - \exp(-\mu)(1 + \mu)$ should be much less than unity. This yields the following inequality:

$$1 - (1 + \mu)\exp(-\mu) \ll 1, \qquad \text{(B1)}$$

which, after expanding in a power series, we find can only be true for $\mu^2 \ll 1$.

Another constraint for the pair-correlation function is easiest to understand from the heuristic description given earlier. For each time lag of interest, the number of raindrop pairs that are separated by lag $t \pm (d\tau/2)$ can be counted. Call this number $d(t)$. The "expected" number of drop pairs at this time lag can be found by taking the number of raindrops that are separated by t for a totally random (Poisson) distribution of the same duration (T) and number of drops (N). Let this be $r(t)$. Then, we have that

$$\eta(t) \sim \frac{d(t)}{r(t)} - 1. \qquad \text{(B2)}$$

It can be known a priori what $r(t)$ will be, once the total time and number of drops are specified. For a Poisson process and sufficiently small $d\tau$, the probability of finding a particle in $(t, t + d\tau] = (N\, d\tau/T)$. This is true for all t, so the probability of finding particles in $(t_1, t_1 + d\tau]$ and $(t_2, t_2 + d\tau]$ is given by the product $(N\, d\tau/T)(N\, d\tau/T)$. The number of disjoint time increments of duration $d\tau$ separated by t in the sample is given by $[(T - t)/d\tau]$, so the number of drop pairs separated by t in the sample is given by

$$r(t) = \frac{T - t}{d\tau}\left(\frac{N\, d\tau}{T}\right)\left(\frac{N\, d\tau}{T}\right) \sim \frac{N^2 d\tau}{T}, \qquad \text{(B3)}$$

with the approximation valid when $t \ll T$.

For a meaningful estimation of what $\eta(t)$ really is, $r(t)$ needs to be as large as possible. If $r(t)$ were found to be, for example, around 5, it would indicate that if five raindrop pairs were detected in the data, $\eta \sim 0$; whereas if 8 were detected, $\eta \sim 0.6$. It does not seem prudent to rely on a few events—three in this example—to classify a distribution to be significantly correlated or uncorrelated at a certain scale. As $r(t)$ increases, the fact that we are using finitely sized samples becomes less relevant. Noting that the mean count is given by $N\, d\tau/T$, we state that another requirement for computation of a meaningful pair-correlation function should be

TABLE B1. Conditioning of the pair-correlation function. As derived in appendix B, the pair-correlation function is most useful in the limit as $\mu^2 \to 0$ and $N\mu \to \infty$. Here are the values of these parameters for the plots given in Figs. 1–3.

Label	$d\tau$	μ^2	$N\mu$
Dataset 1	0.0006	4.6×10^{-6}	9.61
	0.006	4.6×10^{-4}	96.1
	0.06	4.6×10^{-2}	961
Dataset 2	0.0006	8.3×10^{-6}	25.6
	0.006	8.3×10^{-4}	256
	0.06	8.3×10^{-2}	2560
Dataset 3	0.0006	1.0×10^{-5}	15.4
	0.006	1.0×10^{-3}	154
	0.06	1.0×10^{-1}	1540
Dataset 4	0.0006	3.3×10^{-6}	9.09
	0.006	3.3×10^{-4}	90.9
	0.06	3.3×10^{-2}	909
Poisson simulation	0.0006	9.0×10^{-6}	18.0
	0.006	9.0×10^{-4}	180
	0.06	9.0×10^{-2}	1800

$$\frac{N^2 d\tau}{T} = N\mu \gg 1. \tag{B4}$$

This, along with the other requirement from Eq. (B1),

$$\mu^2 \ll 1, \tag{B5}$$

help to specify bounds for what values of $d\tau$ ought to be chosen for the data at hand.[B1]

Values of these parameters for the data used in this paper are given in Table B1. For the values of $d\tau$ chosen, it seems both conditions are met reasonably well for all datasets. All values of $d\tau$ used here follow the above constraints better than some other papers in the literature.

b. When are deviations from $\eta = 0$ significant?

When introducing the pair-correlation function, it was stated that for perfectly steady rainfall, $\eta(t) \equiv 0$ for all lag times t. However, for real rainfall that lasts a finite time, there will be small departures from $\eta(t) = 0$ detected independent of the underlying interparticle distribution [see, e.g., panel (e) in Figs. 1–3 that should have $\eta(t) \equiv 0$ by construction but do not when measured merely because the sample is finite].

An important question naturally follows: At what magnitude can one be confident that the departures from $\eta(t) = 0$ are large enough to identify the distribution in question as non-Poisson?

The treatment given here follows that in Larsen et al. (2003) and is similar to tests of significance for various correlation measurements shown in Martinez and Saar (2001, 109–112). We begin by looking at the range of possible values the pair-correlation function can reasonably take on due to the finite nature of the data only. If the measured pair correlation of a dataset falls outside of this range, we can conclude that we are seeing real (physical) variability and the rain in question is not perfectly steady. Figure B1 demonstrates this test. For each of the four datasets examined in this paper, we redisplay the pair-correlation function as a function of temporal scale for averaging time $d\tau = 6$ ms (top row, a copy of the information shown in Fig. 2) and $d\tau = 60$ ms (bottom row, from Fig. 3) with black circles. Then, for each dataset, we generate 1000 realizations of a Poisson simulation with the same intensity and duration as the original dataset. (For example, from Table 1, we find dataset 1 has 4492 particles detected over 21 min, so each realization has 4492 particles distributed perfectly randomly over an interval 21 min long.) For each of these 1000 realizations, we then calculate the pair-correlation function and record the results. The largest and smallest value for each lag t then form the light-colored envelope in each panel; the 100th largest and 100th smallest (similar in spirit to an 80% confidence interval) form the darker envelope.

How do we interpret these plots? If the dataset is perfectly random, no black dots would be outside the light-colored envelope and most of them should be inside the dark-colored envelope. Even though dataset 4 nearly passed our initial test for steadiness, it can clearly be seen that nearly all of the black dots in the upper-right-hand panel are outside of the dark envelope, and in both the upper and lower panels for dataset 4 there are values that clearly violate the possible expected range for perfectly steady rain. Similarly, the other three datasets stay in the dark envelope virtually all of the time and it seems reasonable to say they are statistically indistinguishable from one of the random simulations.[B2]

[B1] It may be worthwhile to note that both bounds are approximate as they both implicitly rely on pseudo-Poisson behavior. However, a more precise bound could be attained by verifying that both inequalities are satisfied, calculating η for the distribution in question, and then modifying the bounds given the value found for η at the first averaging scale ($d\tau$). Even for fairly clustered systems, it can be shown that the bounds given here are usually satisfactory.

[B2] It is also noteworthy that we tried this test for $d\tau = 0.6$ ms as well. We found that, due to oversampling, the envelopes for this averaging time display a periodic shape and the deviations observed from the envelope were sporadic for all of the data observed. Because of this, we once again wish to caution the reader not to take the results of Fig. 1 as quantitatively meaningful.

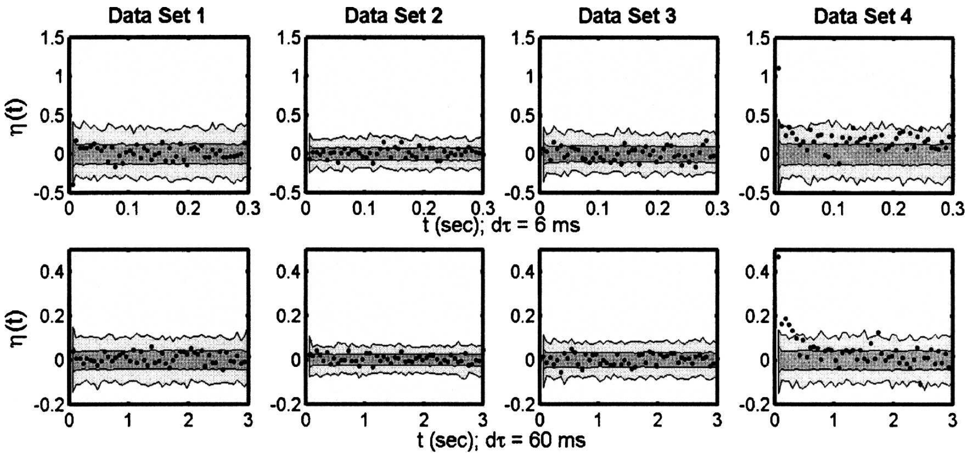

FIG. B1. A plot investigating the significance of departures from $\eta \equiv 0$. In each panel, the pair-correlation function is plotted as a function of time (in s). From left to right, the columns show results for datasets 1–4; the top row has $d\tau = 6$ ms (as in Fig. 2) and the bottom row has $d\tau = 60$ ms (as in Fig. 3). The measured pcf is shown with black circles. In each panel the shaded envelopes show the boundaries of Poisson simulations with the same parameters as the dataset in question. If the data in question were Poissonian, no departure from the lightly shaded region would be expected and most of the calculated values of η would reside in the darker envelope. See text for more details.

APPENDIX C

On Fractal Analysis of Rain Events

When reviewing the literature in the introduction, it was noted that several independent studies had determined that rain arrival statistics all deviate from simple scaling behavior at temporal scales of order 1 s. These methods seemed sensible after results from other studies that examined longer time intervals (see, e.g., Lavergnat and Golé 1998). Here we argue that such fractal methods for studying rain microstructure are questionable. Specifically we demonstrate a similar "scale break" behavior even though the rain is *homogeneous and nonfractal*.

There are nearly as many different ways to assign fractal dimensions to a system as there are investigators assigning them. For simplicity, we will use the box-counting dimension as employed by Wiscombe et al. (2003) and similar to the method used by Zawadzki (1995).

For a one-dimensional system with fractal dimension D, the number $[N(t)]$ of nonempty disjoint intervals of duration t can be shown to follow the relation

$$[N(t)] \propto t^{-D}. \qquad \text{(C1)}$$

As a numerical example, then, if we break the data into segments of duration τ, we get some number of nonempty intervals $[N(\tau)]$. If we then break the same data into segments of duration $\tau/2$, we find that $[N(\tau/2)] = 2^D[N(\tau)]$. As the distribution approaches pure randomness, we would expect that halving the interval length would nearly double the number of nonempty intervals and, hence, $D = 1$. Finding the dimension of the system requires plotting $\ln[N]$ as a function of $\ln(t)$. A least squares linear fit to the result will give a slope of $-D$.

No real system exhibits fractal behavior over *all* scales. Rather, the claim is made that raindrop arrivals exhibit fractal behavior over a range of scales $t_0 < t < t_H$. If the system is divided into intervals of time smaller than t_0, all particles already occupy their own interval; no new nonempty intervals are revealed. Consequently, $[N(t < t_0)] = N_{\text{total}}$; the slope of $\log[N(t)]$ in this range will always be 0. Similarly, in many systems there is a maximal scale t_H above which the data appear homogeneous ($D = 1$). For finite systems, t_H is the time scale large enough to guarantee the presence of at least one particle in each interval. Doubling the interval size halves the number of total intervals present and also halves the number of nonempty intervals since every interval is occupied. For $t \geq t_H$, the number of nonempty intervals is equal to the number of intervals and the slope of $\log[N(t)]$ in this range will always be -1. A value for D between 0 and 1 is reported if there is a linear portion of the regression of $\log[N(t)]$ in the range $t_0 < t < t_H$. However, there may be some curvature for $t \sim t_0$ and $t \sim t_H$, where the slope of the curve is changing.

Curves of $\log[N(t)]$ as a function of $\log t$ are given for the five datasets used throughout this paper in Fig. C1. The solid line in each marks $D = 1$. A strong argument

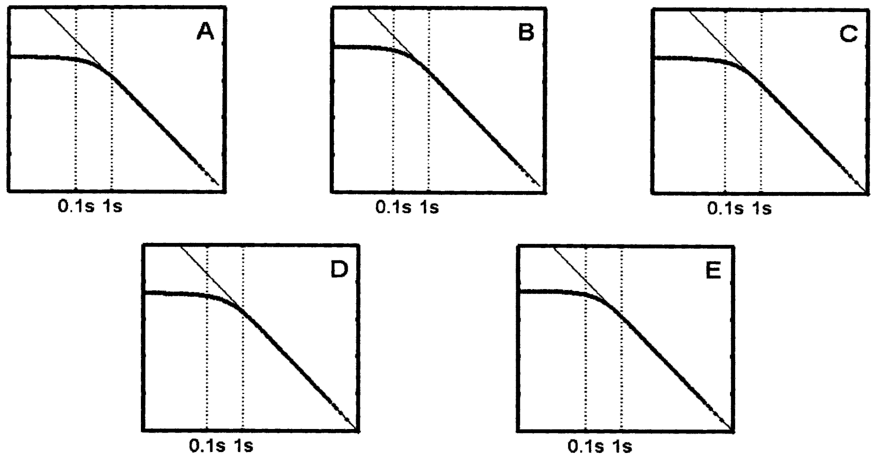

FIG. C1. Plots of $\log[N(t)]$ (y axis) as a function of $\log t$ (x axis) for the five datasets used in this paper. The solid line in each panel shows $[N(t)] = t^{-1}$. Note that we see a departure from homogeneous behavior for all curves at scales between 0.1 and 1 s—the same range reported in the literature, and roughly corresponding to the mean interarrival time of the drops. Although (d) (non-Poisson) does seem to show a departure at a slightly more lengthy temporal scale, no evidence of a persisting linear segment between the horizontal line and line of slope -1 is seen (which would be required if it were to be argued that rain followed simple scaling).

can be made for the observation of two linear subsections of this plot; for $t \gtrsim 1$ s, the slope of the best-fit line is -1. For $t \lesssim 0.1$ s, the slope of the best-fit line is 0. However, to call these systems fractal with $t_0 = 0.1$ s and $t_H = 1$ s would be erroneous. In none of these graphs is there any evidence for a linear segment for $t_0 < t < t_H$; all exhibit continuous curvature.[C1] The excellent agreement between the slope of -1 for $t > 1$ s gives *no* evidence for fractal behavior; we would see excellent agreement in this range for *any* reasonable distribution of points [including the homogeneous, nonfractal Poisson simulation shown in panel (e)].

If one mistakenly assumes that these distributions really are fractal, it is not surprising we see t_H within an order of magnitude of 1 s. A numerical example might be helpful. In a perfectly random (Poisson) distribution of drops with mean interarrival time 0.2 s, the probability of a 1-s time interval being devoid of drops is given by $\exp(-5) \sim 0.7\%$; hence, you will seldom (if ever) see an interval of this length unoccupied. However, the probability of a 0.1-s time interval being devoid of drops is given by $\exp(-1/2) \sim 60.6\%$. Nearly all of the "action" happens in only one order of magnitude. *This is not enough of a range to accurately classify and identify fractal behavior.* The proper assignment of a fractal dimension for this rain would require identifying a linear segment somewhere in the range $t_0 < t < t_H$ and fitting a least squares line through it. No quantitative difference in any of the panels shown in the figures can be made with any confidence.

If the spatiotemporal structure of rain really is fractal in nature, then some other methods must be employed to try and determine the dimension of the system. From looking at the plots of the pair-correlation function, it is obvious that there is a difference between panels (a)–(c) and (d), yet the fractal approach used here does not capture that behavior. It seems that the break in "simple scaling" previously observed in the rain literature may be due to effects similar to the one described above—they are artifacts of data analysis methodology.

An adequate description of the microstructure of rainfall may require different methods and techniques than studies that concentrate on relatively long temporal scales. Because the scales of interest are close to the boundaries of the scaling regime, it seems that reliance on logarithmic plots may be ill-advised for the purpose of careful, precise textural descriptions in rainfall. Furthermore, since we demonstrated in this work that not all rain is statistically nonstationary, we argue definite justification of the inhomogeneous behavior of a dataset *must* be given before adopting an inhomogeneous data analysis strategy. (See appendix A for an example of how this could be justified.)

[C1] Continuous curvature *is* understandable within the statistically stationary framework of correlation theory.

REFERENCES

Atlas, D., and A. C. Chmela, 1957: Physical-synoptic variations of drop-size parameters. *Proc. Sixth Radar Conf.,* Boston, MA, Amer. Meteor. Soc., 21–29.

Bras, R. L., and I. Rodriguez-Iturbe, 1993: *Random Functions and Hydrology.* Dover, 559 pp.

Coleman, P. H., and L. Pietronero, 1992: The fractal nature of the universe. *Physica A,* **185,** 45–55.

Green, H. L., 1927: On the application of the Aitken effect to the study of aerosols. *Philos. Mag.,* **4,** 1046–1069.

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

——, and ——, 2001: Reconsideration of the physical and empirical origins of *Z–R* relations in radar meteorology. *Quart. J. Roy. Meteor. Soc.,* **127,** 517–538.

——, and ——, 2002: When is rain steady? *J. Appl. Meteor.,* **41,** 83–90.

Knyazikhin, Y., A. Marshak, M. L. Larsen, W. J. Wiscombe, J. V. Martonchik, and R. B. Myneni, 2005: Small-scale drop size variability: Impact on estimation of cloud optical properties. *J. Atmos. Sci.,* **62,** 2569–2581.

Kostinski, A. B., and A. R. Jameson, 1997: Fluctuation properties of precipitation. Part I: On deviations of single-size drop counts from the Poisson distribution. *J. Atmos. Sci.,* **54,** 2174–2186.

——, M. L. Larsen, and A. R. Jameson, 2005: The texture of rain: Exploring stochastic micro-structure at small scales. *J. Hydrol.,* in press.

Kruger, A., and W. F. Krajewski, 2002: Two-dimensional video disdrometer: A description. *J. Atmos. Oceanic Technol.,* **19,** 602–617.

Larsen, M. L., W. Cantrell, J. Kannosto, and A. B. Kostinski, 2003: Detection of spatial correlations among aerosol particles. *Aerosol Sci. Technol.,* **37,** 476–485.

Lavergnat, J., and P. Golé, 1998: A stochastic raindrop time distribution model. *J. Appl. Meteor.,* **37,** 805–818.

List, R., 1988: A linear radar reflectivity–rainrate relationship for steady tropical rain. *J. Atmos. Sci.,* **45,** 3564–3572.

Lovejoy, S., and D. Schertzer, 1995: Multifractals and rain. *New Uncertainty Concepts in Hydrology and Water Resources,* E. Kundzewicz, Ed., Cambridge University Press, 62–103.

——, M. Lilly, N. Desaulniers-Soucy, and D. Schertzer, 2003: Large particle number limit in rain. *Phys. Rev. E,* **68,** 025301.

Martinez, V. J., and E. Saar, 2001: *Statistics of the Galaxy Distribution.* Chapman and Hall, 432 pp.

Papoulis, A., 1965: *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, 583 pp.

Pietronero, L., 1987: The fractal structure of the universe: Correlations of galaxies and clusters and the average mass density. *Physica A,* **144,** 257–284.

Preining, O., 1983: Optical single-particle counters to obtain the spatial inhomogeneity of particulate clouds. *Aerosol Sci. Technol.,* **2,** 79–90.

Scrase, F. J., 1935: The sampling errors of the Aitken nucleus counter. *Quart. J. Roy. Meteor. Soc.,* **61,** 367–379.

Shaw, R. A., A. B. Kostinski, and M. L. Larsen, 2002: Towards quantifying droplet clustering in clouds. *Quart. J. Roy. Meteor. Soc.,* **128,** 1043–1057.

Steiner, M., J. A. Smith, and R. Uijlenhoet, 2004: A microphysical interpretation of radar reflectivity–rain-rate relationships. *J. Atmos. Sci.,* **61,** 1114–1131.

Uijlenhoet, R., J. N. M. Stricker, P. J. J. F. Torfs, and J.-D. Creutin, 1999: Towards a stochastic model of rainfall for radar hydrology: Testing the Poisson homogeneity hypothesis. *Phys. Chem. Earth,* **24B,** 747–755.

——, J. M. Porrà, D. Sempere Torres, and J.-D. Creutin, 2005: Analytical solutions to sampling effects in drop size distribution measurements during stationary rainfall: Estimation of bulk rainfall variables. *J. Hydrol.,* in press.

Wiscombe, W., Y. Knyazikhin, and A. Marshak, 2003: A new look into the treatment of small-scale drop variability in radiative transfer. *Proc. 13th Atmospheric Radiation Measurement (ARM) Science Team Meeting,* Broomfield, CO, ARM. [Available online at http://www.arm.gov/publications/proceedings/conf13/extended_abs/wiscombe-wj.pdf.]

Wunsch, C., 1999: The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull. Amer. Meteor. Soc.,* **80,** 245–255.

Zawadzki, I., 1995: Is rain fractal? *New Uncertainty Concepts in Hydrology and Water Resources,* E. Kundzewicz, Ed., Cambridge University Press, 104–108.