On the Recovery of 3D Spatial Statistics of Particles from 1D Measurements: Implications for Airborne Instruments

MICHAEL L. LARSEN

Department of Physics and Astronomy, College of Charleston, Charleston, South Carolina

CLARISSA A. BRINER

Department of Physics, University of Colorado Boulder, Boulder, Colorado

PHILIP BOEHNER

Department of Scientific Computing, Florida State University, Tallahassee, Florida

(Manuscript received 30 December 2013, in final form 21 May 2014)

ABSTRACT

The spatial positions of individual aerosol particles, cloud droplets, or raindrops can be modeled as a point processes in three dimensions. Characterization of three-dimensional point processes often involves the calculation or estimation of the radial distribution function (RDF) and/or the pair-correlation function (PCF) for the system. Sampling these three-dimensional systems is often impractical, however, and, consequently, these three-dimensional systems are directly measured by probing the system along a one-dimensional transect through the volume (e.g., an aircraft-mounted cloud probe measuring a thin horizontal ''skewer'' through a cloud). The measured RDF and PCF of these one-dimensional transects are related to (but not, in general, equal to) the RDF/PCF of the intrinsic three-dimensional systems from which the sample was taken. Previous work examined the formal mathematical relationship between the statistics of the intrinsic three-dimensional system and the one-dimensional transect; this study extends the previous work within the context of realistic sampling variability. Natural sampling variability is found to constrain substantially the usefulness of applying previous theoretical relationships. Implications for future sampling strategies are discussed.

1. Introduction

Many three-dimensional physical systems are often described using the mathematics of discrete point processes. To name just a few examples, the mathematics and statistical methods of spatial point processes are used extensively in astrophysics (e.g., Martínez and Saar 2002), atmospheric physics (e.g., Larsen 2006), biology (e.g., Young et al. 2001), and a large and growing community use this methodology to describe particle-laden turbulent fluid flow (e.g., Shaw 2003; Bateson and Aliseda 2012; Devenish et al. 2012; Saw et al. 2012).

Practical considerations often require sampling of these three-dimensional systems by measuring particle positions within a thin ''pencil beam'' transect through the volume. For example, an imaging sampler can identify particles passing a fixed point in a wind tunnel under laminar flow conditions (e.g., Saw et al. 2012), or a cloud particle sensor can identify the positions of ambient cloud droplets as an airplane traverses a cloud with a sensor on its wing (e.g., Brenguier et al. 1998). These ''skewers'' through the three-dimensional volume often comprise the only dataset available to try to characterize the three-dimensional systems of interest.

One of the most commonly used statistical tools to describe spatial point processes is the pair-correlation function $\eta(r)$ [or its related statistic, the radial distribution function $g(r) = \eta(r) + 1$] (see, e.g., Shaw et al. 2002). The pair-correlation function (PCF) yields a direct scale-localized measure of the deviation from a homogeneous Poisson point process (e.g., see Larsen 2012). However, it was noted by Holtzer and Collins (2002) that the radial distribution function (and, consequently,

Corresponding author address: Michael L. Larsen, Dept. of Physics and Astronomy, College of Charleston, 66 George St., Charleston, SC 29424.
E-mail: larsenml@cofc.edu

DOI: 10.1175/JTECH-D-14-00004.1

the PCF) of the sampled one-dimensional transect does not precisely match the intrinsic radial distribution function/PCF of the three-dimensional system.

Holtzer and Collins found a mathematical way of relating the measured radial distribution function (RDF) of the transect [$g_{1D}(z)$] to the intrinsic RDF of the system [$g_{3D}(r)$]. This relationship (which takes the form of an integral equation) is not generally invertible; $g_{1D}(z)$ can be found from $g_{3D}(r)$, but $g_{3D}(r)$ cannot be generally found from $g_{1D}(z)$. Holtzer and Collins did find, however, that $g_{3D}(r)$ could be found from $g_{1D}(z)$ if a general functional form for $g_{3D}(r)$ could be assumed. They then confirmed their results through direct numerical simulation (DNS) of a particle-laden fluid with Stokes numbers 0.4 and 0.7.

The results of Holtzer and Collins (2002) are beginning to be used fairly frequently in some communities (e.g., Siebert et al. 2010; Bateson and Aliseda 2012; Devenish et al. 2012; Saw et al. 2012). Our goal in this manuscript is to explore the utility of the Holtzer and Collins result to realistic one-dimensional transects through a random but correlated distribution of particles/events, keeping in mind a particular application: in situ measurement of cloud droplets via a wing-mounted imaging sampler. We find that sampling variability may prevent us from learning as much about $g_{3D}(r)$ from $g_{1D}(z)$ as we might have hoped or expected, even using sample sizes that suggest plenty of data have been acquired.

2. Theoretical tools

In this section, we briefly introduce the basic notation used in the rest of this manuscript. As noted in the introduction, the tools and methods described in this paper appear in a wide variety of different disciplines, with each community developing its own usage patterns, nomenclature, and symbolic convention. The interdisciplinary nature of this approach allows for wide applicability for this analysis, but—if care is not taken—it can result in simple misunderstandings that take quite some time to sort out [e.g., see Kostinski (2001), Borovoi (2002), Kostinski (2002), Shaw et al. (2002), Baker and Lawson (2010), and Larsen (2012) for reasonably simple misunderstandings that could have been avoided with a careful exposition]. In an effort to be as widely understood as possible, we proceed somewhat tutorially.

a. *Marked point processes*

In the introduction, it was pointed out that the mathematics of discrete point processes is applicable to a wide variety of different physical systems spanning many different disciplines. The development in this manuscript is applicable to any system where a discrete number of particles (or events) can be uniquely associated with spatial locations inside some known volume. As a simple example, if N cloud droplets are distributed inside a unit cube, then each cloud droplet can be said to be located at a position $\langle x_i, y_i, z_i \rangle$ with i running from 1 to N and each coordinate taking on some value between 0 and 1. This basic ability to associate each particle or event with a unique point in space is what we refer to here as a point process.

The point process may be statistically homogeneous or inhomogeneous. Statistical homogeneity in this context is a bit complicated; formally (see, e.g., Feller 1966), a system is considered homogeneous if and only if the moments of the distribution do not depend on spatial position. However, *any* finite discrete system ultimately *must* fail this formal test. This observation has been made in several other studies; the basic argument is most fully summarized in Larsen (2012). In brief, any finitely measured system has assumed 0 mean outside the volume of interest but nonzero mean within the volume of interest (this can, for some analysis methods, be overcome with the use of periodic boundary conditions). Similarly, for sufficiently small scales, discreteness will prevent measurements of the mean in disjoint volumes from being the same. Disentangling this inevitable sampling variability from the formal criterion for statistical homogeneity has thus far proven to be intractable without accidentally removing finitely sized sampled subsystems drawn from homogeneous parent populations from the class of "homogeneous systems" (see, e.g., Wunsch 1999). In short, by adhering to the formal definition, we categorize *all* finite-measured systems as inhomogeneous. This seems contrary to the intent of the term—it seems that, a constant-mean Poisson distribution constrained to be inside a subvolume should be statistically homogeneous; by the formal definition, it is not. Attempts to broaden the notion of statistical homogeneity have been attempted (see, e.g., Larsen et al. 2005; Anderson and Kostinski 2010), but no completely satisfactory method seems to have yet been developed.

This discussion of statistical homogeneity is relevant because the tool used here to describe departures from perfect spatial randomness (the PCF and/or the RDF) formally requires a statistically stationary dataset to allow for meaningful physical interpretation (see, e.g., Shaw et al. 2002; Larsen et al. 2005; Larsen 2006). This study is able to avoid the question since all systems explored are computationally simulated and therefore statistically homogeneous by construction; the use of the methods utilized here on experimental data, however, is a bit more complicated.

b. The PCF and RDF

It is stated above that the tool used to characterize deviations from perfect spatial randomness is the PCF $[\eta(r)]$ or, equivalently, the RDF $g(r) \equiv \eta(r) + 1$. Following the convention used in Holtzer and Collins (2002), we will identify the RDF when measured for the three-dimensional system as $g_{3D}(r)$ and the RDF of a one-dimensional transect through the three-dimensional distribution as $g_{1D}(z)$. Similarly, we will introduce $\eta_{3D}(r)$ and $\eta_{1D}(z)$ for the PCFs of the 3D system and 1D transect, respectively. For simplicity, our analysis from this point forward will be written in terms of the PCF; converting results to RDF involves just a careful addition of unity.

Physically, the PCF is a measure of the scale-localized deviation from perfect spatial randomness (see, e.g., Larsen 2012). Within the context of discrete random variables (which is the domain of interest here, since we are talking about discrete points or events within a volume), "perfect spatial randomness" is taken as a homogeneous Poisson distribution with constant mean λ_{3D}. Therefore, in a three-dimensional volume of size V, there are a total of $\lambda_{3D}V = N$ events.

A Poisson distribution is characterized by three basic properties (see, e.g., Cramér and Leadbetter 2004): (i) the probability of k events found in a volume V depends on k and the magnitude of V, but not the location of V (statistical homogeneity); (ii) the events occurring in disjoint volumes are mutually independent random variables (independence); and (iii) the probability of more than one event occurring in a small volume dV is $\mathcal{O}(dV)$ as $dV \to 0$ (noncoincidence).

Based on this definition, the probability of finding one event in each of two infinitesimally small, nonoverlapping volumes—each of volume dV and separated by a distance r—can be calculated as

$$p_{r,P}(1,2) = (\lambda_{3D}dV)(\lambda_{3D}dV). \tag{1}$$

In general for a homogeneous but not necessarily Poisson distribution, we broaden this expression to

$$p_r(1,2) = (\lambda_{3D})^2(dV)^2[1 + \eta_{3D}(r)], \tag{2}$$

where $\eta_{3D}(r)$ is the PCF evaluated at separation distance r. This expression, then, identifies the enhanced or lowered joint probability of finding particles/events in infinitesimal volumes located r from each other.

Calculation of $\eta_{3D}(r)$ from a given dataset is straightforward. Note that

$$\frac{p_r(1,2)}{p_{r,P}(1,2)} = 1 + \eta_{3D}(r). \tag{3}$$

Thus, to find $\eta(r)$, one needs to count the number of particles/events separated by scale $[r, r + dr)$ and divide by the number of particles/events that would be separated by scale $[r, r + dr)$ in a Poisson distribution with the same total number of particles/events and associated with the same volume.

c. The impact of sampling on the PCF and the RDF

This paper explores how $\eta_{3D}(r)$ and $g_{3D}(r)$ are altered when a one-dimensional transect is used to explore an inherently three-dimensional system.

After "skewering" a three-dimensional system, a one-dimensional spatial dataset is obtained. The general notion of the PCF extends as expected down to one dimension. For simplicity, let the original three-dimensional system be within a unit cube and the one-dimensional transect have dimensions $\delta \times \delta \times 1$ with $\delta \ll 1$. Although this is still technically a three-dimensional volume, we will assume that coordinates along the two dimensions of extent δ are not known, so that particle/event positions are characterized only through the linear position along the direction that extends the farthest.

If we allow λ_{1D} to be the number of particles/events per unit length on this transect, then we can define the probability of finding two particles separated by a linear distance z through

$$p_z(1,2) = (\lambda_{1D})^2(dz)^2[1 + \eta_{1D}(z)] \tag{4}$$

and thus allowing us to calculate

$$\frac{p_z(1,2)}{p_{z,P}(1,2)} = 1 + \eta_{1D}(z). \tag{5}$$

In general, if a three-dimensional volume is homogeneous, isotropic, and statistically stationary, it may be characterized by some PCF $\eta_{3D}(r)$. Holtzer and Collins (2002) found a theoretical link between $g_{3D}(r)$ and $g_{1D}(z)$. Modifying their result to the geometry introduced above and writing their result in terms of the PCF, we can represent their result as

$$\eta_{1D}(z) = \left\{\frac{1}{\delta^4}\int_0^\delta\int_0^\delta\int_0^\delta\int_0^\delta \left\{\eta_{3D}\left[\sqrt{z^2 + (x - x_\circ)^2 + (y - y_\circ)^2}\right] + 1\right\} dx\, dx_\circ\, dy\, dy_\circ\right\} - 1. \tag{6}$$

Holtzer and Collins go on to carefully rewrite this in polar coordinates, but our numerical comparison is actually more easily completed with the quadruple integral as shown.

In practice, one often actually wishes to take a measurement of $\eta_{1D}(z)$ and infer the intrinsic expression for $\eta_{3D}(r)$; the one-dimensional transect often is the actual measurement made in the laboratory or the field, and one hopes to infer something about the real physical system from the measurements obtained. Much of the work of Holtzer and Collins (2002) is devoted to carefully exploring this question; they conclude that this expression is not, in general, invertible, but if some functional form of η_{3D} is assumed, then an inversion can successfully be completed. They verify this result by exploring a system where the distributed particles are known to have a power-law RDF.

In the remainder of this paper, we investigate Eq. (6) as applied to several other simulated systems that do not have a power-law RDF.

3. Point process models

As mentioned previously, Holtzer and Collins (2002) tested their theoretical result through a direct numerical simulation of a particle-laden fluid. The advantages of their approach is that (i) it allowed them to easily investigate both 3D to 2D and 3D to 1D downscaling, (ii) they were able to examine the response as a function of Stokes number, and (iii) they had theoretical (e.g., Balkovsky et al. 2001) and numerical (e.g., Reade and Collins 2000) reasons to suspect that $g_{3D}(r)$ obeyed a power-law form for their physical system of interest. This power-law form allowed for an analytic computation to upscale 2D and 1D results back to 3D. Here, we examine two different systems, both of which have easily formulated PCFs. However, neither of these systems has a power-law form.

a. Homogeneous Poisson process

The homogeneous Poisson process is the simplest spatial point process. As such, the simulation of a homogeneous Poisson process forms a good validation case for our computer algorithms and a basic test of our analysis methodology. Since the PCF is a direct measure of the scale-localized deviation from a homogeneous Poisson process, it is clear that, for a Poisson process, $\eta_{3D}(r) \equiv 0 \; \forall \; r$ [similarly, $g_{3D}(r) \equiv 1 \; \forall \; r$].

Appealing to Eq. (6), we find that when $\eta_{3D}(r) \equiv 0$, we calculate $\eta_{1D}(z) = 0$. Thus, the case of a homogeneous Poisson process is a nice test for a few reasons: (i) the functional forms of η_{3D} and η_{1D} are very simple; (ii) $\eta_{3D}(r)$ and $\eta_{1D}(z)$ should have the same form [similar to the Holtzer and Collins (2002) power-law test case]; and (iii) we are able to conduct a direct test of the effects of sampling, since a single skewer through the distribution should only have $N_{1D} = \delta^2 N_{3D}$ particles.

b. Matérn cluster process

The other system that we will investigate here is known as a Matérn cluster process (Matérn 1960; Martínez and Saar 2002). We choose this particular system because (i) it has an analytically tractable, closed-form expression for $\eta_{3D}(r)$; (ii) unlike the homogeneous Poisson process, the form of $\eta_{3D}(r)$ is nontrivial; (iii) this spatial point process model may actually be a decent mathematical description of some physical processes (see, e.g., Larsen 2006); and (iv) the system can be natively simulated in 1D, 2D, or 3D with some interesting properties.

This last point merits further explanation. A Matérn cluster process can be simulated in 1D, 2D, or 3D. The variable $\eta_{3D}(r)$ for a 3D Matérn system has a different functional form than $\eta_{2D}(s)$ for a 2D Matérn system or $\eta_{1D}(d)$ for a 1D Matérn system; the functional form of the PCF depends on how many dimensions in which the system is created. Further, $\eta_{1D}(d)$ for a 1D Matérn system does not appear the same as $\eta_{1D}(z)$ found from a 1D transect through a 3D Matérn distribution. Thus, a Matérn cluster process gives an interesting test to the theoretical result developed by Holtzer and Collins (2002); the 1D transect should, in principle, look different than both the 3D simulation *and* a 1D Matérn distribution.

CONSTRUCTION OF A MATÉRN CLUSTER PROCESS

The Matérn cluster process is a particular member of a class of point process models known as Neyman–Scott processes. To construct a realization of any Neyman–Scott process, the following steps are followed in order:

1) Distribute N_p particles throughout the domain V in a perfectly random (homogeneous Poisson) manner (let $N_p/V = \lambda_p$, the density of "parents.").
2) Use some discrete probability distribution with mean given by λ_D—the mean number of "daughter particles" per "parent particle"—to determine how many daughter particles each parent has. Let this discrete random number be given by n_i with i running from 1 to N_p.
3) Use some continuous (and presumably isotropic) probability density function (pdf) to independently place the n_i particles about the ith parent particle.
4) Remove the N_p parent particles from the distribution. The resulting collection of daughter particles forms the final distribution.

The Matérn process is one in which the pdf governing the number of daughters per parent (step 2) is drawn

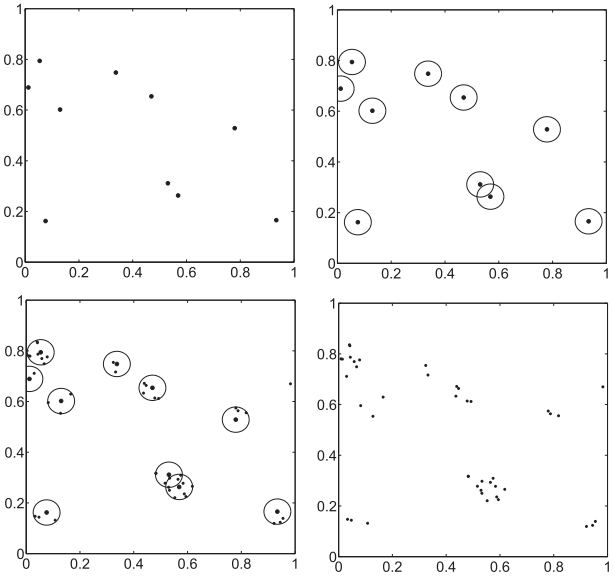

FIG. 1. A cartoon demonstrating the construction of a 2D Matérn cluster process. (top left) First, parent particles are distributed randomly throughout the volume. (top right) Then, n-dimensional spheres of R are drawn around each parent particle. Associated with each parent particle, a number of daughter particles are distributed at random inside the n-dimensional sphere. (bottom left) The number of daughter particles varies from parent to parent; they are drawn from a Poisson distribution with a user-defined mean. (bottom right) Finally, the parent particles and the spheres are removed, leaving the final distribution of particles. (Note that periodic boundary conditions are used throughout.)

from a Poisson distribution and the placement of the daughter particles (step 3) is equiprobable in a sphere of radius R about the parent particle. Specifically, $p(r)dr = 3r^2dr/R^3$ with $0 \le r \le R$ for a three-dimensional simulation [in a 1D Matérn simulation, this pdf is $p(r)dr = dr/2R$ with $-R \le r \le R$, leading to a system that does not have the same structure as a one-dimensional transect through a three-dimensional Matérn system]. A cartoon of the creation of a two-dimensional Matérn distribution constructed in a unit square with periodic boundary conditions is presented in Fig. 1.

It is worth noting that, by construction, the Matérn cluster process is statistically homogeneous. (Although the system is clearly clustered, the spatial position of each cluster is not knowable a priori, and each point in the volume is equally likely to be a cluster center. Further, the clusters are mutually independent of each other. This passes the test for statistical homogeneity on physical grounds.)

The PCF for a Matérn cluster process depends, as noted earlier, on whether the distribution is generated in one, two, or three dimensions. The relationship for a one-dimensional simulation derived from expressions given in Stoyan et al. (1987) is

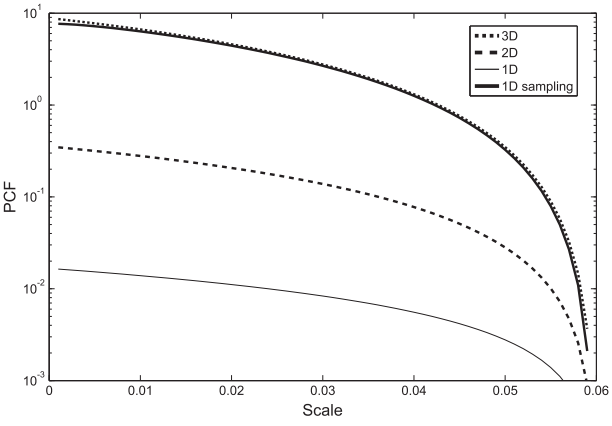

FIG. 2. A plot of the theoretical PCF for Matérn systems analyzed in different dimensions. For this theoretical system, it was assumed that the relationships given in Eqs. (6)–(9) are valid. Plotted are systems with $\lambda_d = 1000$, $R = 0.03$, and $\delta = 0.01$. We clearly see that η_{3D}, η_{2D}, and η_{1D} differ for this system. On this semilogarithmic axis, the sampled pseudo-1D skewer (represented by the dark solid line) appears very close to η_{3D}. This gives some hope that a 1D transect may be a good measure to approximate the true statistical structure of a 3D system. Only the domain $0 < r < 2R$ is shown, as $\eta(r) = 0$ for $r > 2R$.

$$\eta_{1D}(d) = \frac{1}{2R\lambda_p}\left(1 - \frac{d}{2R}\right). \tag{7}$$

In two dimensions, this becomes

$$\eta_{2D}(s) = \frac{2}{\pi^2 R^2 \lambda_p}\left[\arccos\left(\frac{s}{2R}\right) - \frac{s}{2R}\sqrt{\left(1 - \frac{s^2}{4R^2}\right)}\right]. \tag{8}$$

Finally, the expression in three dimensions is

$$\eta_{3D}(r) = \frac{3}{8\pi R^6 \lambda_p}\left(R - \frac{r}{2}\right)^2\left(2R + \frac{r}{2}\right). \tag{9}$$

Note that, for all three of these theoretical PCFs, the formulas above are valid for scales less than $2R$ only; for scales greater than $2R$, $\eta_{1D}(d) = \eta_{2D}(s) = \eta_{3D}(r) = 0$. Figure 2 shows these three theoretical PCFs as a function of scale. Also shown is the PCF expected for a 1D transect of the 3D system [as computed using Eq. (6)]. Note that there is a clear difference between the PCF of a 1D Matérn distribution and the PCF of a 1D transect of a 3D Matérn distribution.

4. Simulation

To investigate the relationship proposed by Holtzer and Collins (2002) summarized in Eq. (6) above, we developed a numerical simulation. This, unlike the DNS of Holtzer

and Collins, is a static simulation; the particles are merely placed according to some algorithm and then the PCF is computed.

Studies in atmospheric science are sometimes completed with only a few thousand particles comprising the dataset from a one-dimensional transect (see, e.g., Jameson et al. 1998; Jameson and Kostinski 2000; Kostinski and Jameson 2000; Larsen et al. 2005; Baker and Lawson 2010; Larsen 2012). Similarly, studies in two dimensions have been conducted in other fields with only a few tens of thousands of particles (see, e.g., Young et al. 2001). We wanted to explore the influence that sampling variability may have on one-dimensional transects with 10^3–10^4 particles contained in the transect to match these practically used cases.

The simulation parameters were constrained by multiple considerations. We wanted to generate volumes that (i) were small enough to be simulated effectively in 3D, so that $\eta_{3D}(r)$ could be numerically calculated in a reasonable amount of time; (ii) could be "skewered" to form one-dimensional transects of size $\delta \times \delta \times 1$ where there would still be 10^3–10^4 particles enclosed in each transect on average; and (iii) allowed $\delta \ll 1$, so that transects were pseudo-1D.

The first consideration (size of the 3D system) required us to keep the total number of particles $N \sim 1 \times 10^7$, given our available hardware and code to calculate $\eta_{3D}(r)$. To ensure approximately 10^3–10^4 particles per transect, this then set the range of δ to be nominally $0.01 < \delta < 0.1$, with a preference (for the sake of simulating a "skewer" aspect ratio) to stay closer to $\delta \sim 0.01$.

The transects through the three-dimensional volume were constructed to travel parallel to one of the three faces of the cube with the midpoint of the cross-sectional area chosen at random between δ and $1 - \delta$ along each of the two directions perpendicular to the direction of propagation for the transect, thus giving a random skewer of the distribution.

For the Matérn systems, the parameters N_p, λ_d, and R were constrained to have $N_p\lambda_d = 1 \times 10^7$. However, beyond that constraint parameters could be freely chosen. Since general results were desired, parameters were not optimized to match any particular physical system. Therefore, combinations of these variables were chosen that gave appreciable deviations from $\eta(r) = 0$ for scales much smaller than the cube side length. [A rule of thumb suggested for the PCF in Larsen (2006) is to note that calculated values have limited utility for $r > 0.1L$, where L is the characteristic size of the measurement]. Thus, the additional requirement was made that $R < 0.05$. It was also desired to have $R > \delta$ (otherwise, much of the information associated with the position in the cross section of the one-dimensional transect, which is often not recorded, could be important), thus suggesting the use of values between $R = 0.03$ and $R = 0.05$. Finally, if N_p is chosen too low, then there is an appreciable chance that any given transect will not intersect any of the "clusters" in the system, and transects through the volume could have a nonnegligible probability of being totally devoid of particles. For N_p too large, however, the PCF approaches $\eta(r) \equiv 0$. Values of N_p between 1000 and 20 000 were ultimately used.

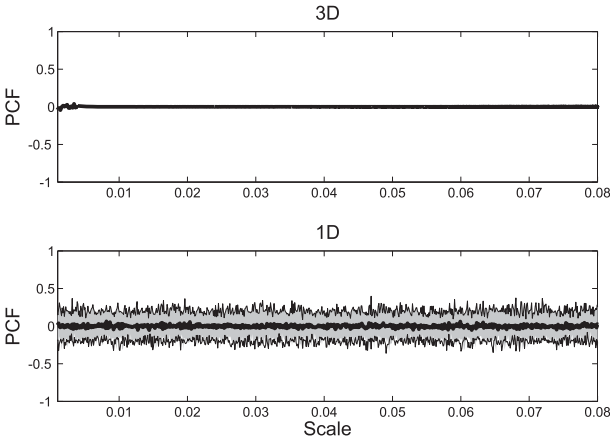

FIG. 3. An illustration of the calculated 3D and sampled 1D PCFs for a distribution of 1×10^7 particles distributed perfectly randomly in a unit cube. Both curves should theoretically be identically zero. (Top) The measured PCF as a function of scale (r) in the initial volume. (bottom) The PCF associated with pseudo-1D transects through this volume. The solid line staying near PCF = 0 is the average observed PCF as a function of scale for 30 random 1D transects. The shaded area identifies the region between the smallest and largest values of the PCF observed as a function of scale for the 30 transects.

5. Simulation results

a. Poisson distribution

Ten million particles were randomly placed inside a unit cube. The PCF $\eta_{3D}(r)$ for this system was tabulated and is shown in the top panel of Fig. 3. Note there is excellent agreement with the theoretical relationship $\eta(r) \equiv 0$; the slight deviations from zero seen for small values of r are a result of sampling noise. (Even in a system with 1×10^7 particles, the expected number of two particles being a very small distance apart is vanishingly small. This sample noise diminishes with increasing distance, since the actual volume associated in looking at a spherical shell with radius in $(r, r + dr]$ grows as r^2.) The bottom panel reveals the PCF for 30 different one-dimensional transects through this volume. The shaded area indicates the range of different PCFs observed, whereas the solid line shows the PCF

after averaging the results from the 30 transects. For these skewers, $\delta = 0.03$, which resulted in approximately 9000 particles in each transect.

We see immediately that the Holtzer and Collins result does seem to hold for this system; the predicted value of $\eta_{1D}(z) = 0 \; \forall \; z$ holds. Noteworthy, however, is that any one particular transect may see a nonnegligible deviation from this null value; it is only when averaging together all 30 transects that the expected result is obtained. This is important; in experimental circumstances only one particular transect is usually available. Given this curve, one might erroneously conclude from a single transect that a system shows deviations from perfect spatial randomness when the deviation is caused purely by sampling variability. For context, a PCF of a few tenths is comparable to values given in the literature that suggest solid evidence of non-Poisson behavior (see, e.g., Shaw et al. 2002; Larsen 2006) [a careful consideration of this concern is presented in appendix B of Larsen et al. (2005), but only within the context of an inherently one-dimensional system]. The deviations of the PCF from identically zero is brought about here by simple shot noise; the finite number of particles in each skewer through the distribution may be above or below the standard expectation, and this causes a deviation around the expected curve of $\eta_{1D}(z) = 0$.

Note that the shot noise actually manifests itself in two separate ways: (i) the number of particles that exist in the skewer [which changes $p_{z,P}(1, 2)$ in Eq. (5)] and (ii) the number of particles seen to be separated by scale z in the skewer [which changes $p_z(1, 2)$ in Eq. (5)]. This dual nature of the effect of shot noise results in a relationship between sampling fluctuations and measured PCFs that is more complicated than expected.

b. *Matérn distribution*

For the Matérn distribution, a nontrivial interplay is expected between R and δ. As clearly seen in Eq. (6), as $\delta \to 0$, $\eta_{3D}(r) \to \eta_{1D}(z)$. In a practical setting, taking the limit as $\delta \to 0$ also limits the number of particles observed and introduces sampling fluctuations.

Figure 2 shows $\eta_{3D}(r)$ (dotted line) and $\eta_{1D}(z)$ (bold solid line) of a one-dimensional transect through the three-dimensional system. These two curves look extremely similar, but the close agreement is partially due to the logarithmic scaling of the vertical axis. Figure 4 more carefully explores how $\eta_{3D}(r)$ and $\eta_{1D}(z)$ differ for a transect. Note that for $\delta \to 0$, $\eta_{3D}(r) \sim \eta_{1D}(z)$. However, also observe that appreciable differences between η_{3D} and η_{1D} can exist for scales larger than δ. (Even though the PCF is a scale-localized measure of deviations from perfect spatial randomness, there is some necessary "scale mixing" in the pseudo-one-dimensional sampling process).

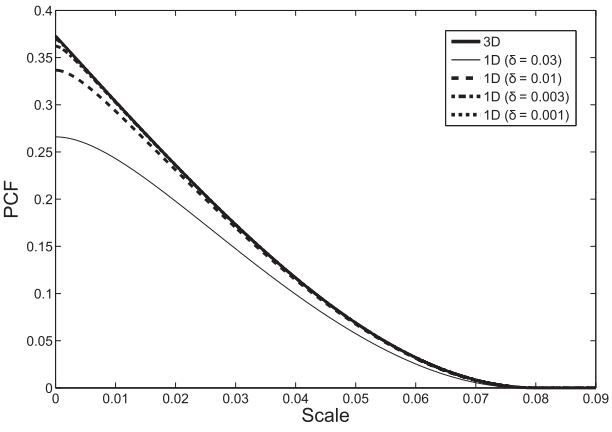

FIG. 4. A plot of the theoretical PCF for a 3D Matérn system. The parameters for the Matérn system were 20 000 parents, 500 daughters per parent, and a radius around each parent for daughter placement of $R = 0.04$. This plot shows the theoretical 3D PCF for this system, along with a curve computed via Eq. (6) (solved with a computer algebra system) showing the expected relationships for a 1D transect of dimensions $\delta \times \delta \times 1$. Four different values for δ are shown, clearly indicating that deviations of $\eta_{1D}(r)$ from $\eta_{3D}(r)$ become more pronounced as δ increases. What this curve does not indicate, however, is that sampling statistics become sparser as δ decreases. These effects inevitably compete and, in fact, $\delta \to 0$ does not become an ideal estimator for η_{3D} in practice like this plot might suggest.

The curves shown in Fig. 2 are theoretical, based on the expression for $\eta_{3D}(r)$ given in Eq. (9) and the transformation from Holtzer and Collins (2002) given in Eq. (6). To computationally explore the competing effects of trying to simultaneously make δ small [so that $\eta_{3D}(r) \approx \eta_{1D}(z)$, allowing for more detailed insight into η_{3D} from the inversion from η_{1D}] and to make δ large (in order to increase the sample size), an analysis similar to that given for the skewered Poisson process was completed.

The results are shown in Fig. 5, constructed very similarly to Fig. 3. Here, parameters are shown that seemed particularly revealing with $R = 0.03$, $N_p = 1000$, $\lambda_d = 1 \times 10^4$, and $\delta = 0.01$. The top panel clearly shows that $\eta_{3D}(r)$ takes on the expected form from Eq. (9). However, we find in the bottom panel that $\eta_{1D}(z)$ differs noticeably from the theoretical expression predicted by Eq. (6).

Not only do we see an appreciable range of fluctuations about the expected curve (identified by the shaded region, similar to that shown in Fig. 3), but there is also a difference when comparing the average behavior and the theoretical expression expected from Eq. (6).

Figure 6 investigates this further; here, $\eta_{1D}(z)$ is plotted for each of the 30 transects (thin lines) along

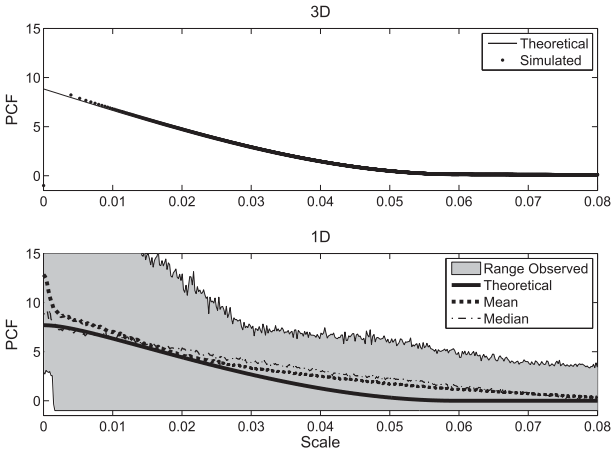

FIG. 5. A plot of the PCF for a 3D Matérn simulation. For this system, there were 1000 parent particles and an average of 10 000 daughters per parent. The entire distribution was constructed inside a unit cube (with periodic boundary conditions applied where necessary). Each daughter particle was placed equiprobably inside a sphere of $R = 0.03$ centered around the associated parent particle. (top) The theoretical PCF as a function of spatial scale for this system. The dots indicate the measured PCF for the realization made in this study. (bottom) The PCFs for 1D transects through this volume. The theoretical solid curve is based on the expression derived in Holtzer and Collins (2002) and repeated in Eq. (6). The shaded region indicates the envelope of observed values for 30 different pseudo-1D transects through the simulation; these transects were of dimensions $0.01 \times 0.01 \times \delta$ and, therefore, contained on average 1000 particles. The dotted line and the dotted–dashed line show the mean and median, respectively, of these 30 transects.

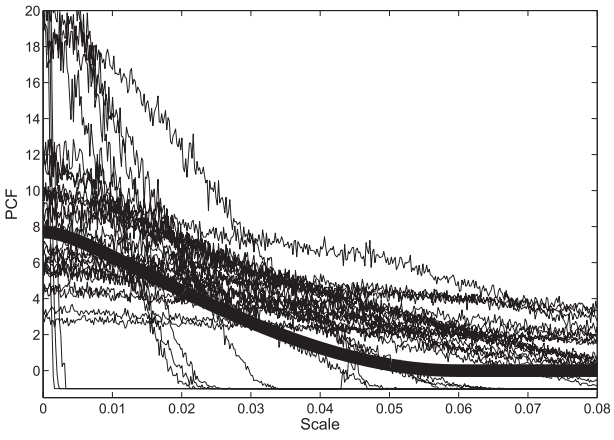

FIG. 6. A plot of the PCF for 30 different pseudo-1D transects through a 3D Matérn simulation. For this system, there were 1000 parent particles and 10 000 daughters per parent. The entire distribution was put inside a unit cube (periodic boundary conditions were applied where necessary). The radius around each parent in which all daughters were placed was of size $R = 0.03$. The thin black lines indicate the actual measured PCFs for the transect, while the dark solid line indicates the theoretical curve based on the expression derived in Holtzer and Collins (2002).

with the expected theoretical mean predicted by Eq. (6) using $\eta_{3D}(r)$ described by Eq. (9) and verified in the top panel of Fig. 5. Clearly *none* of the individual transects really matches the expected 1D transect behavior well.

6. Identifying and interpreting results

It is important to note that this example, while computationally generated, is not completely unrealistic (though it is an extreme example, so that the illustration is clear). The generated point process is, by construction, statistically homogeneous and used an appropriate statistical tool to explore the system.

There have been some attempts in the past to quantify whether a measurement of the PCF can be physically interpreted, for example, the efforts made in Larsen et al. (2005). This work argued that, for the PCF to be physically meaningful, $N\mu \gg 1$ and $\mu^2 \ll 1$. For these transects, $\mu = 0.2$ and $N = 1000$, thus having $N\mu = 200$ and $\mu^2 = 0.04$ which, when compared to the real experimental data explored in the same paper, seems right in line with realistic data parameters [it is important to note that the authors of Larsen et al. (2005) were quite clear that these guidelines constituted necessary—not sufficient—conditions for sampling of a system].

The conundrum for this system is clear. Figure 6 shows that there is a huge amount of sampling variability from transect to transect. Few (if any) of the transects through the system match the value for $\eta_{1D}(z)$ that they "should," based on the theoretical treatment given in Holtzer and Collins (2002), though the previously published criteria for finding a meaningful value of η based on a dataset as published in Larsen et al. (2005) have been satisfied. One can say that the criteria for using the PCF should be strengthened, but how? Particularly troubling is that these issues are not merely tied to the PCF; since the PCF is analytically linked to many of the other means of characterizing deviations from perfect randomness, including (but not limited to) fractal dimension, power spectral density/structure function, autocorrelation, clustering index, and Fishing statistic (see, e.g., Shaw et al. 2002; Larsen 2012), the problems we observe here are likely to exist independently of the specific measure used to characterize the structure.

In their earlier study, Holtzer and Collins (2002) saw much better agreement between the theoretically predicted and simulated values of $\eta_{1D}(z)$. We believe there are several explanations why the disagreement with the theoretical expression observed in this study was not seen in the earlier study. First, the Holtzer and Collins (2002) study had an RDF that took on a power-law form. Given the context (inertial particles in a turbulent fluid), this was a reasonable assumption, but it may have also

introduced symmetries due to the scale invariance of the functional form that aided in rapid convergence with small sample sizes. Second, it appears that Holtzer and Collins (2002) obtained their curves for $\eta_{1D}(z)$ by simultaneously observing 65 536 parallel and nonoverlapping transects through the distribution. Therefore, sampling considerations did not exist in their study. This was not stated explicitly, but it appears the same number of total particles were used to calculate $\eta_{3D}(r)$ and $\eta_{1D}(z)$. Although valid in a DNS setting, this is seldom the case in an experimental context.

This still leaves the practical question of how to ensure that a sampled transect can give meaningful results for a 3D system. One might hope that these issues would be mitigated through the use of a 3D measurement from an instrument like, for example, the Holographic Detector for Clouds (HOLODEC) developed by cloud particle measurements (see, e.g., Fugal et al. 2004; Fugal and Shaw 2009). However, one finds that simulations of a similarly sized 3D volume, including nominally the same number of total particles, do not necessarily do much better in accurately recreating the intrinsic PCF (and, in some cases, the 3D-estimated PCF is far worse). It is clear that, for this system, several thousand particles are insufficient to sample a 3D system—no matter the specific geometry applied.

Previous work (e.g., Larsen 2006, 2012) has revealed that several thousand particles *are* often sufficient to recreate the general shape of the PCF of a time series, so we are left with a cautionary tale for investigators—the number of particles needed to develop a physically meaningful PCF depends not only on N and μ (as suggested in Larsen et al. 2005) but also on other parameters, including (but not necessarily limited to) the number of dimensions the system was initially distributed within prior to sampling. A 1D time series has different sampling criteria for meaningful use than a 3D system sampled with a 1D transect.

Even knowing that the system initially was distributed in three dimensions is not likely to be enough. In the Matérn system explored with these simulations, there are a number of different characteristic length scales that can come into play. Some of them are based on the instrumental sampling properties (e.g., δ), some of them are derivable from the sampled data itself (e.g., the mean interparticle detection distance μ^{-1}), but some are not knowable a priori. [In this system, the physical size of the spherical volume—the daughter particles are distributed inside—and the mean distance between ''parents'' could have some influence on the size of sample needed for convergence to the expected value of $\eta_{1D}(z)$ as predicted by Holtzer and Collins (2002)]. This general theme highlighting an interplay between multiple scales has recently been observed in a completely separate context (e.g., Larsen and Clark 2014). Unfortunately, deconvolving the influence of these different length scales into concrete sampling recommendations (even when these scales can be identified) has proven to be a substantial task and may be highly dependent on specific physical and instrument sampling properties.

7. Conclusions

The PCF (and, by extension, the RDF) has been repeatedly shown to be useful in characterizing spatial and temporal distributions of particles or events; the variable has direct physical meaning, gives a scale-localized measure of clustering, and is easy to calculate from a dataset. Further, the assumptions that go into using and calculating the PCF are straightforward. Given the simulations run here, however, we urge extreme caution in tying the measurement of the PCFs of one-dimensional transects to properties of three-dimensional systems; far more data may be required than traditionally used for analysis in time series data. A firm grasp on minimal sampling criteria is not straightforward and may rely on both physical and instrumental parameters that should be well characterized before inversion is attempted.

Acknowledgments. This research was supported by a Cottrell College Science Award from the Research Corporation for Scientific Advancement and through NSF Grant AGS-1230240. Additionally, the authors thank the anonymous reviewers, who had insightful comments and suggestions for the first draft of this manuscript, especially in regard to the physical interpretation of the results.

REFERENCES

Anderson, A., and A. B. Kostinski, 2010: Reversible record breaking and variability: Temperature distributions across the globe. *J. Appl. Meteor. Climatol.,* **49,** 1681–1691, doi:10.1175/2010JAMC2407.1.

Baker, B., and R. Lawson, 2010: Analysis of tools used to quantify droplet clustering in clouds. *J. Atmos. Sci.,* **67,** 3355–3367, doi:10.1175/2010JAS3409.1.

Balkovsky, E., G. Falkovich, and A. Fouxon, 2001: Intermittent distribution of inertial particles in turbulent flows. *Phys. Rev. Lett.,* **86,** 2790–2793, doi:10.1103/PhysRevLett.86.2790.

Bateson, C. P., and A. Aliseda, 2012: Wind tunnel measurements of the preferential concentration of inertial droplets in homogeneous isotropic turbulence. *Exp. Fluids,* **52,** 1373–1387, doi:10.1007/s00348-011-1252-6.

Borovoi, A. G., 2002: On the extinction of radiation by a homogeneous but spatially correlated random medium. *J. Opt. Soc. Amer.,* **19A,** 2517–2520, doi:10.1364/JOSAA.19.002517.

Brenguier, J.-L., T. Bourrianne, A. A. Coelho, J. Isbert, R. Peytavi, D. Travarin, and P. Weschler, 1998: Improvements of droplet

size distribution measurements with the fast-FSSP (forward scattering spectrometer probe). *J. Atmos. Oceanic Technol.,* **15,** 1077–1090, doi:10.1175/1520-0426(1998)015<1077:IODSDM>2.0.CO;2.

Cramér, H., and M. R. Leadbetter, 2004: *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications.* Dover Books on Mathematics, Dover Publications, 368 pp.

Devenish, B. J., and Coauthors, 2012: Droplet growth in warm turbulent clouds. *Quart. J. Roy. Meteor. Soc.,* **138,** 1401–1429, doi:10.1002/qj.1897.

Feller, W., 1966: *An Introduction to Probability Theory and Its Applications.* Vol. 2, Wiley, 669 pp.

Fugal, J. P., and R. A. Shaw, 2009: Cloud particle size distributions measured with an airborne digital in-line holographic instrument. *Atmos. Meas. Tech.,* **2,** 259–271, doi:10.5194/amt-2-259-2009.

——, ——, E.-W. Saw, and A. V. Sergeyev, 2004: Airborne digital holographic system for cloud particle measurements. *Appl. Opt.,* **43,** 5987–5995, doi:10.1364/AO.43.005987.

Holtzer, G. L., and L. R. Collins, 2002: Relationship between the intrinsic radial distribution function for an isotropic field of particles and lower-dimensional measurements. *J. Fluid Mech.,* **459,** 93–102, doi:10.1017/S0022112002008169.

Jameson, A. R., and A. B. Kostinski, 2000: Fluctuation properties of precipitation. Part V: On the distribution of rain rates—Theory and observations in clustered rain. *J. Atmos. Sci.,* **57,** 373–388, doi:10.1175/1520-0469(2000)057<0373:FPOPPV>2.0.CO;2.

——, ——, and R. A. Black, 1998: The texture of clouds. *J. Geophys. Res.,* **103,** 6211–6219, doi:10.1029/98JD00081.

Kostinski, A. B., 2001: On the extinction of radiation by a homogeneous but spatially correlated random medium. *J. Opt. Soc. Amer.,* **18A,** 1929–1933, doi:10.1364/JOSAA.18.001929.

——, 2002: On the extinction of radiation by a homogeneous but-spatially correlated random medium: Reply to comment. *J. Opt. Soc. Amer.,* **19A,** 2521–2525, doi:10.1364/JOSAA.19.002521.

——, and A. R. Jameson, 2000: On the spatial distribution of cloud particles. *J. Atmos. Sci.,* **57,** 901–915, doi:10.1175/1520-0469(2000)057<0901:OTSDOC>2.0.CO;2.

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

——, 2012: Scale localization of cloud particle clustering statistics. *J. Atmos. Sci.,* **69,** 3277–3289, doi:10.1175/JAS-D-12-02.1.

——, and A. S. Clark, 2014: On the link between particle size and deviations from the Beer–Lambert–Bouguer law for direct transmission. *J. Quant. Spectrosc. Radiat. Transfer,* **133,** 646–651, doi:10.1016/j.jqsrt.2013.10.001.

——, A. B. Kostinski, and A. Tokay, 2005: Observations and analysis of uncorrelated rain. *J. Atmos. Sci.,* **62,** 4071–4083, doi:10.1175/JAS3583.1.

Martínez, V. J., and E. Saar, 2002: *Statistics of the Galaxy Distribution.* CRC Press, 432 pp.

Matérn, B., 1960: Spatial variation. *Medd. Statens Skogsforskningsinst.,* **49,** 1–144.

Reade, W. C., and L. R. Collins, 2000: Effect of preferential concentration on turbulent collision rates. *Phys. Fluids,* **12,** 2530–2540, doi:10.1063/1.1288515.

Saw, E.-W., R. A. Shaw, J. P. L. C. Salazar, and L. R. Collins, 2012: Spatial clustering of polydisperse inertial particles in turbulence: II. Comparing simulation with experiment. *New J. Phys.,* **14,** 105031, doi:10.1088/1367-2630/14/10/105031.

Shaw, R. A., 2003: Particle-turbulence interactions in atmospheric clouds. *Annu. Rev. Fluid Mech.,* **35,** 183–227, doi:10.1146/annurev.fluid.35.101101.161125.

——, A. B. Kostinski, and M. L. Larsen, 2002: Towards quantifying droplet clustering in clouds. *Quart. J. Roy. Meteor. Soc.,* **128,** 1043–1057, doi:10.1256/003590002320373193.

Siebert, B., S. Gerashchenko, A. Gylfason, K. Lehmann, L. R. Collins, R. A. Shaw, and Z. Warhaft, 2010: Towards understanding the role of turbulence on droplets in clouds: In situ and laboratory measurements. *Atmos. Res.,* **97,** 426–437, doi:10.1016/j.atmosres.2010.05.007.

Stoyan, D., W. S. Kendall, and J. Mecke, 1987: *Stochastic Geometry and Its Applications.* Wiley Series in Probability and Statistics, Wiley, 345 pp.

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, doi:10.1175/1520-0477(1999)080<0245:TIOSCR>2.0.CO;2.

Young, W. R., A. J. Roberts, and G. Stuhne, 2001: Reproductive pair correlations and the clustering of organisms. *Nature,* **412,** 328–331, doi:10.1038/35085561.