# A Bayesian Method for Upsizing Single Disdrometer Drop Size Counts for Rain Physics Studies and Areal Applications

A. R. Jameson

*Abstract*—**A time or spatial series of drop counts is but one realization of a multiple stochastic process. In this paper, a method is presented that extracts more of the information contained in the time series of 1-min Joss–Waldvogel disdrometer counts in rain than a simple analysis of the magnitudes of the counts would provide. This is done by greatly increasing the size of a data set using a Bayesian analysis of drop count measurements in 17 size bins. Using the empirical copula statistical technique of probability density function transformations, a 1391-min time series of drop counts was expanded to the equivalent of 40 000 min. This dramatic increase in sample size permits a deeper characterization of the rain. Using this single disdrometer, it also allows one to translate these counts into a 200 × 200 grid filled at each point with drop size distributions of mean drop concentrations consistent with the observed statistical properties of the rain. Such a field can be used for remote sensing studies of the effect of partial beam filling and for algorithm development. Moreover, since there is nothing unique to this set of drop counts, this approach can be applied to any other set of count data, including snow and clouds.**

*Index Terms*—**Algorithms, atmospheric measurements, estimation, meteorology, multiplication, radar applications, rain, random media, remote sensing, statistics.**

## I. INTRODUCTION

THE purpose of this paper is to more fully exploit the information contained in a time series of observations from a single disdrometer by expanding the size of the observed set of data *without any assumptions of a parametric form of the size distributions or of their spatial/temporal structure*. There are two reasons for wanting to do this. One is to better understand the physics of the precipitation (illustrated toward the end of this paper), and the other is to enable the application of point measurements using a disdrometer to the interpretation of observations by a remote sensing device over an area. Both aspects are important to geoscience remote sensing.

First, however, it must be realized that, for every drop size, each measurement in the time series represents one correlated draw from a distribution of drop counts associated with a mean value drawn from a distribution of mean values. To get a handle on the distribution of these mean values, a Bayesian analysis of the counts at each drop-bin size is used to define the probability density function (pdf) of the mean values for each size bin. In addition we, of course, observe the autocorrelation function for each size bin, well described by the exponential function for these observations. It is shown next that with these two sets of information, we can expand the size of the data set in a manner consistent with the observations of the counts and the observed autocorrelation functions.

Such an expansion of the size of a data set serves two purposes. First, it reduces or eliminates the spurious effects of a small sample size in the characterization of the rain [11]. This will be illustrated next. Second, the data set can be used to simulate a giant grid of instantaneous measurements by instruments at each grid point.

There are other approaches for filling a grid with point estimates of drop size distributions (DSDs) and their associated integral properties. However, these are not equivalent to increasing the size of a data set of counts, because they either assume specific parametric forms for the DSDs [18] or they go directly to the simulations of the bulk parameters [4], [14]. They are also not instantaneous, so that these approaches have the unfortunate consequence of filtering the data either in time and/or space so that such approaches obscure the direct interpretation in terms of rain physics. We wish to avoid that here.

However, that is not meant to imply that such approaches do not serve a purpose, such as assigning plausible DSDs to a field of radar observations, for example. While such assignments may not be correct because of underlying assumptions, it is something that cannot be done using the approach defined in this paper. However, the objective of this paper is to use measurements from a single instrument without any additional information from other observations such as in [18]. That is, one cannot compare the results from the method presented here to a field of observations, because to do so, one would need a technique to invert an observed data field to recover the base field of random numbers appropriate to an observed data field. There is no known method for doing that. Hence, one must be satisfied that this technique preserves the important statistical information (pdfs ands correlations) for other purposes and not for the purpose of comparing to a field of observations. If one wishes to compare directly to a field of observations, one needs to use one of the approaches aforementioned instead.

In this paper, a method is presented for upscaling a time series of 1-min disdrometer observations, which were collected during a convective/stratiform rain event. This differs from previous work [6], in which a spatial series of ice cloud observations along a flight path were directly upscaled to large-scale domains. In this paper, a time series must be converted to have a spatial interpretation. Such a time series, however, is really just one realization of a multiple statistical process. That is, the observed counts may be considered to be random draws from random (but correlated) mean values. There is, then, a great deal of information contained within the time series aside from just the magnitude of the counts themselves. For example, correlations in counts give scale information out to 1/2 the observation interval, whereas Bayesian analyses provide information about the mean values of counts contributing to the observations. As follows, a method is developed to access some of this additional information. This approach not only provides detailed size distributions at each grid point, useful for probing remote sensing physics, but these distributions can be also integrated to yield "point" estimates of upscaled quantities, such as the rainfall rate $(R)$, radar reflectivity factor $(Z)$, the total number of particles $(C_{\text{tot}})$, and the mean drop size $(\overline{D})$ (equivalent to the slope of an exponential fit to the distribution).

The approach used here is presented in the next section, followed by a brief description of the set of disdrometer observations used as the example. Some results are presented in the final section.

## II. Theory and Method Applied to a Set of Data

Here, we first focus strictly on the application of Bayesian statistics to counting. As argued in detail in [5], for each observed count of particles of a specified size, the application of the Bayesian approach under the assumption of Poisson counting statistics leads to the distribution of mean values of counts per unit observation interval, $C$, for particles of a specified size given by

$$P(C|n, D) \propto \frac{C^n}{n!} e^{-C} \tag{1}$$

where $n$ is the observed number of counts, $C$ is the mean value of the distribution of counts per unit observational interval and the vertical bar denotes conditioning. (As discussed in [5, p. 2015] and [19, pp. 15–20], the Poissonian assumption is not critical here and, for example, as calculations using a Gaussian distribution showed.) Every time there is a new count, there is a new distribution defined by (1). As shown in [5], for a particular particle size, the secret of the success of the Bayesian approach, then comes from overlaying each of these $P(C|n)$ for each observation of $n$. That is

$$P(C|D) = \frac{\sum_{i=1}^{T} \frac{C^{n_i}}{n_i!} P(n_i|D) e^{-C}}{\sum_{C=0}^{C_{\max}} \sum_{i=1}^{T} \frac{C^{n_i}}{n_i!} P(n_i|D) e^{-C}} \tag{2}$$

where $T$ is the total number of observations, $P(n_i|D)$ is the probability of observing $n$ particles of size $D$ at the $i$th observation, and the denominator is a normalizer of the distribution. The most frequent $C$ then naturally emerge as the most likely values surrounded by the entire pdf of $C$. This significantly differs from histograms, which only show the frequency distribution of *observed* counts $N$ often at low resolution (coarse bins) in order to achieve enough samples with an outcome that is heavily dependent upon the one realization sampled and the total number of samples. $P(C)$, on the other hand, is the observed pdf of the *mean* values, which can emerge quite rapidly from the observations. This is particularly useful in the discussion of DSDs, which describe the mean concentrations as a function of drop diameter.

However, the temporal/spatial description of rain involves more than just $P(C)$. In particular, the occurrences of raindrops are correlated among drops of the same size, as well among drops of different sizes. The correlation among drops of the same size is described by the appropriate correlation function as observed and as described further next. In this paper, the cross correlation among drops of different sizes is not directly considered but instead arises naturally by applying the autocorrelation function for each different drop size to the same field of random numbers as described and demonstrated next. This correlation among counts can be interpreted as a description of the meteorological "structure," which, in turn, is a reflection of whatever physical processes produced it. A knowledge of the meteorology is not required to use the correlation information and for this technique to work. On the contrary, as shown below, it may, at times, shed light on some of the meteorology.

Furthermore, $P(C)$ only provides a measure of means when the counts exist. However, in general, there may be extended periods during an interval when drops of a particular size are not observed. This is particularly true for the larger drops. Hence, one must also know the fractional time, i.e., $F$, for which the counts are zero, and this varies for each size bin. This is calculated as the fraction of the observational interval, in which no counts are observed. For example, if there are ten 1-min observations containing no drops of a given size out of 100 min of observations, then $F$ is 10%. Thus, how can one combine all of these considerations in a rational manner?

Perhaps, it is easiest first to present a set of measurements and then to apply the approach. Here, we consider a 23-h rain event observed using a Joss–Waldvogel impact disdrometer having 1-min temporal resolution. These data were supplied by Prof. Carl Ulbrich (retired) at Clemson University at Spartanburg, SC during a summer rain event most likely in 2005. To get a meteorological feel of this event, the rainfall rates, i.e., $R$, were calculated for each of the 1-min samples as illustrated in Fig. 1. Obviously, $R$, varied considerably during the event with an overall average value of just under 2 mm h$^{-1}$ and a peak of just over 30 mm h$^{-1}$. However, since the purpose of this paper is to demonstrate a technique, we assume that this event is approximately stationary.

For each of the 17 drop size bins (centered at 0.35, 0.45, 0.55, 0.65, 0.77, 0.92, 1.11, 1.33, 1.50, 1.66, 1.91, 2.26, 2.58, 2.87, 3.20, 3.54, and 3.92 mm diameters), the $1/e$ correlations lengths (where $e$ is the Euler number) and the fractional empty time were calculated. These are illustrated in Fig. 2(a). As might be expected, the sparseness of drops generally increases with increasing drop size. The correlation lengths, however, vary considerably with the longest associated with the approximately 1 mm sized drops.
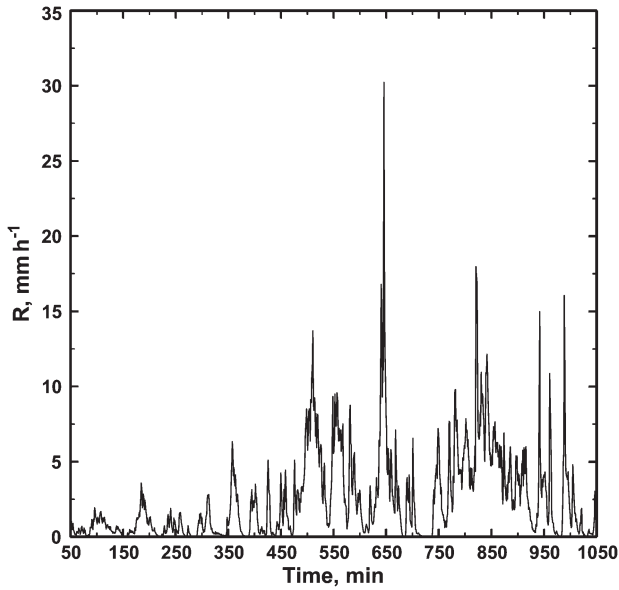
Fig. 1. Temporal history of 1-min rainfall rate calculated using the observed Joss–Waldvogel disdrometer observations used in this paper.

The distributions $P(C)$ of the mean values $C$ are then computed for each drop size using (2) and using the empirical copula transformation technique (see [1], [2], [16] and as briefly described in [12, App. B] with regard to radar signals.). Basically, in the copula technique, one forms the accumulated pdf of the random variable that one wishes to correlate. One then takes a string of unit variance, zero-mean correlated variables having the desired correlation function (often Gaussian or exponential). The accumulated pdf of the desired variable is then used to invert these correlated variables to become a correlated string of the desired variable. However, before doing this, it is first necessary to generate a correlated 2-D field of uniformly distributed random real numbers over [0,1] having the characteristic correlation appropriate to each drop size. In this paper, we arbitrarily assume that each minute corresponds to 250-m spatial translation because of the wind. (This assumption is not vital. Other assumptions are, of course, possible and if wind measurements were available, such translations would become more meaningful. However, this work is an exercise demonstrating a method, not a purposeful analysis of meteorological data.) We then envision a 50 km × 50-km square grid having 40 000 points. An exponential correlation function having the correct correlation length as given in Fig. 2(a) is then imposed on a grid of zero mean, unit variance uniformly distributed random numbers (the base field of random numbers) Because the correlation function (transformed to be a spatial correlation function under the assumption of a steady wind speed aforementioned) contains all length scales from 250 m to more than 348 km, the natural variability over all those scales is automatically included in the simulation, as long as the data remain statistically homogeneous. The correlation is then reproduced at each drop size separately using the so-called root matrix method (e.g., [13] as discussed in [8, pp. 3924 and 3925] for a 1-D field). For a 2-D field, one uses a given correlation function to form the covariance matrix, $K_\rho$, which is a symmetric matrix of the same dimension as the input data field, such that
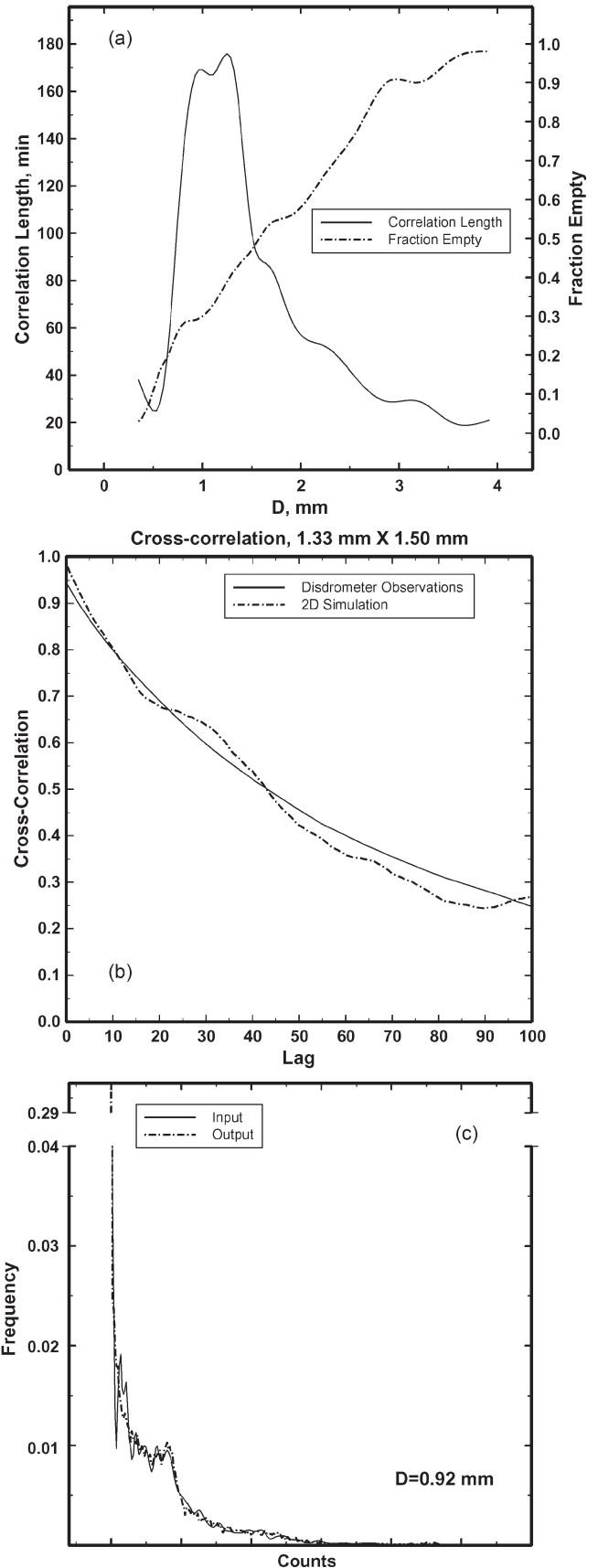


Fig. 2. (a) Observed $1/e$ correlation lengths and fraction of null counts as a function of drop size, (b) example of the cross-correlation function of drop counts as observed and that from the 2-D simulation as discussed in the text, and (c) a comparison of the input and output frequencies of counts at one drop size.

$K_\rho(i,j) = \rho(0)$ and $K_\rho(i,j) = K_\rho(j,i)$ for the off-diagonal elements. One then computes $\mathbf{S}$, the square root matrix of this covariance matrix (e.g., [3]). For a zero mean, unit variance input data matrix, $\mathbf{D}$, the matrix $\mathbf{SDS^T}$ yields the correlated field where $\mathbf{S^T}$ is the transpose matrix of $\mathbf{S}$ (also see [8, pp. 3924–3925]). The resulting field of numbers, however, must then be renormalized because of distortions to the pdf introduced by this process. This is accomplished by taking the empirical pdf of these values and then applying the copula method, so that we finally end up with a list or "vector," $\mathbf{Z}$, of [0,1] uniformly distributed of properly correlated random numbers for each drop size. This vector is then used in the copula transformation described next.

Remember that the Bayesian $P(C)$ previously derived only applies to nonzero counts. Hence, this $P(C)$ is first adjusted so that the number of zero counts agrees with the empty fraction in Fig. 2(a) and the $P(C)$ remains normalized. That is, the frequency of null counts is added to $P(C)$ and the frequencies of all other counts are adjusted so that $P(C)$ integrates to unity. Hence, for example, if the null counts happened to be 50% of the observations, the remaining $P(C)$ would be halved. Of course, null counts are correlated just as any other count, which is why there are empty regions in the figures that follow. It is then used in the copula method to transform $\mathbf{Z}$ into a vector of integer $C$ values satisfying $P(C)$. Finally, this vector is then unstacked to form the $200 \times 200$ grid of values. This is done for each size bin to yield 50 km $\times$ 50 km fields. This provides "true" (in the sense of using mean values) DSDs having the correct correlation (in the sense that they reproduce the exponential fit to the observed correlation function as previously discussed), as well as the correct pdf properties, including the empty fractions at each drop size at each of the 40 000 grid points. These can then be also combined to compute integral variables such as rainfall rate and radar reflectivity factor. These results are discussed in the next section.

It is worth mentioning, however, that while seemingly adding unnecessary complexity by going from a correlated 2-D field to a vector and then reversing the process after the copula calculations, such a procedure avoids the introduction of periodic artificial correlations that occur if one just takes a 40 000 string of correlated data and tries to convert them directly into a 2-D field. It is also worth mentioning that this approach preserves long-term cross correlations among drop counts at different sizes as reported in [7] for Joss–Walvogel disdrometers and [9] for video disdrometers. An example is given in Fig. 2(b). For completeness, Fig. 3(c) illustrates an example of input and output count frequencies at $D = 0.92$ mm.

Finally, it should be recognized that every base field is a random field of numbers, each will generate its own unique output pattern. Hence, since there are an infinite number of such base fields, there are an infinite number of possible 2-D patterns. However, each one will satisfy the statistical properties found in the original set of observations.

### III. RESULTS

The results and plots here are all derived from the one simulation generated using the technique and 1391 min of
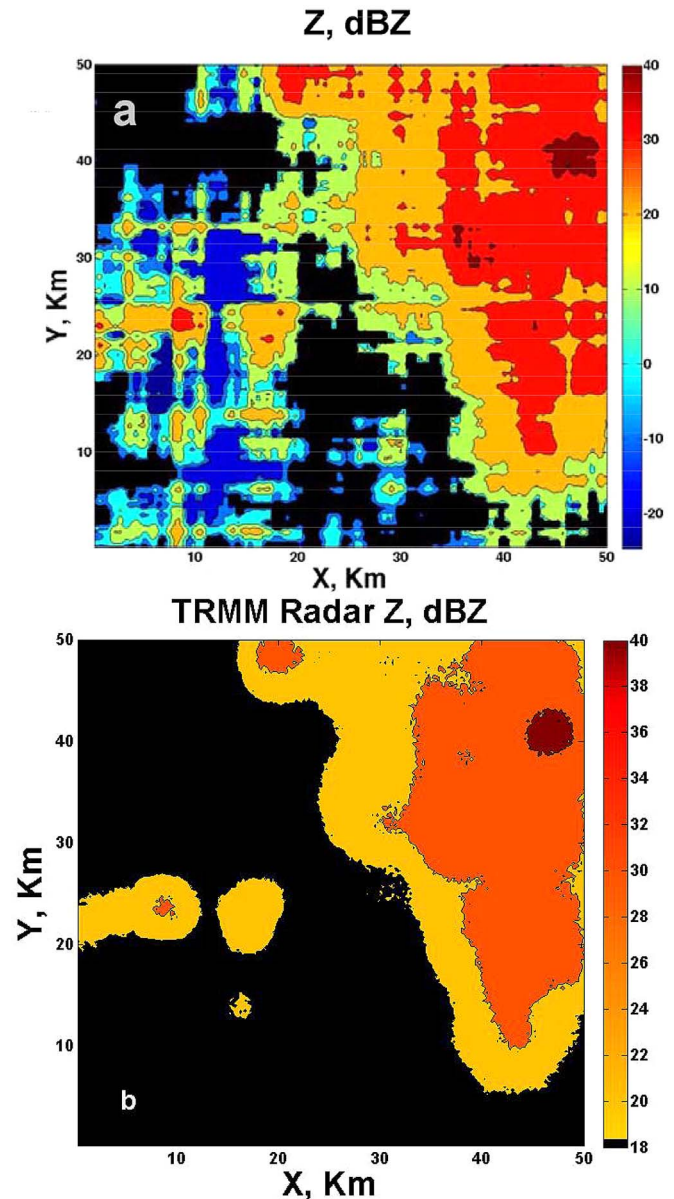


Fig. 3. Contours of the radar reflectivity factors expressed as dBZ (a) over the entire grid calculated using the technique described in the text and based upon the disdrometer observations (the intrinsic field) and (b) as that field would appear to the TRMM radar (radar filtered field) after sampling at each grid point, with an 18-dBZ threshold, including signal statistics fluctuations and partial beam filling.

disdrometer observations just previously described. One motivation (but not the only one) for doing all of this is to expand the size of the disdrometer data set in a manner consistent with the disdrometer measurements. We apply the technique just previously described to a base field of random numbers for all the drop sizes. Thus, by combining the fields for all the different drop sizes, at each grid point, we end up with a DSD that can be used to calculate, for example, the radar reflectivity factor, $Z$. For example, assuming Rayleigh scattering by the drops, Fig. 3(a) is a contour plot of $Z = \sum D^6$, where the summation is over all the drops, calculated at each grid point and $Z$ is expressed in the usual units of $\mathrm{dBZ} = 10\log_{10}(Z)$.

As pointed out by a few reviewers, we note that the grid resolution and contour program sometimes produce what appear to be stretched features, sometimes vertically and sometimes horizontally. This happens because of the necessity for using a finite matrix having finite resolution. First, recall that the purpose here is to take time series observations at a point with no other data and then to try to turn them into a field of spatial values for use in validation efforts, for example. Time is not equal to space, so that one must make an assumption about how to transform a times series into a spatial field. The first step is to perform a simple transformation that a characteristic distance is simply given by a characteristic velocity times the characteristic time, in this case, 1 min. In this paper, we arbitrarily assume a characteristic velocity of around 4.17 ms$^{-1}$ so that the characteristic spatial dimension is 250 m for the 1-min sample time. However, this then simply translates time into a linear distance, so that we still only have information along one dimension. In order to get two dimensions, we need yet one more assumption. That assumption is that the correlation function is identical in each of the two orthogonal directions. The two orthogonal directions define a matrix. In Nature, this matrix can essentially be infinite with infinitesimal resolution. We do not have that luxury and, instead, must rely on finite matrices with coarser resolution. much as numerical models solve the Navier–Stokes equations use 1 km or coarser grid spacing, for example. Consequently, whereas Nature has access to full 2-D symmetric correlation through essentially infinite resolution, we must be satisfied with 2-D correlation along the orthogonal axis with coarser resolution. This means that, at times, this finite resolution 2-D correlation scheme can produce elongated features in either of the two directions. However, just as the solutions for the Navier–Stokes equations contain the physical reality of those equations even on a coarser grid, thus too our 2-D correlations, and simulations retain the physics useful for 2-D applications on a coarser grid. After all, as we shall see, the 5-km radar beam of TRMM sees some 314 of these 250-m grid points, so that axis symmetry in a 250-m radius becomes irrelevant. All that matters, really, is that we have correlated mean values assigned to each grid point in such a way that is consistent with the observations by the single instrument. Indeed, this is the only known method for taking one point observations to produce 2-D fields, which reproduce the drop correlations, drop cross correlation and pdf structures seen in the original observations over all the drop sizes. The method here produces 2-D fields of detailed drop size by drop size microphysical values from single-point measurements without assumptions about the form of the DSD or any additional information for use in studies by the precipitation physics community as illustrated later (in Fig. 12), for example.

Returning to Fig. 3(a), clearly, a large portion of the larger values are found in the upper right portion of the field, although there are additional spots scattered at other locations, as well with 24% of the values greater than 30 dBZ. In addition, Fig. 3(b) illustrates how this intrinsic disdrometer derived field of $Z$ in Fig. 3(a) would be transformed at each grid point by a radar such as the NASA Tropical Rainfall Measurement Mission (TRMM) radar after beam filtering and subject to a nominal minimum detection threshold of 18 dBZ, signal statis-
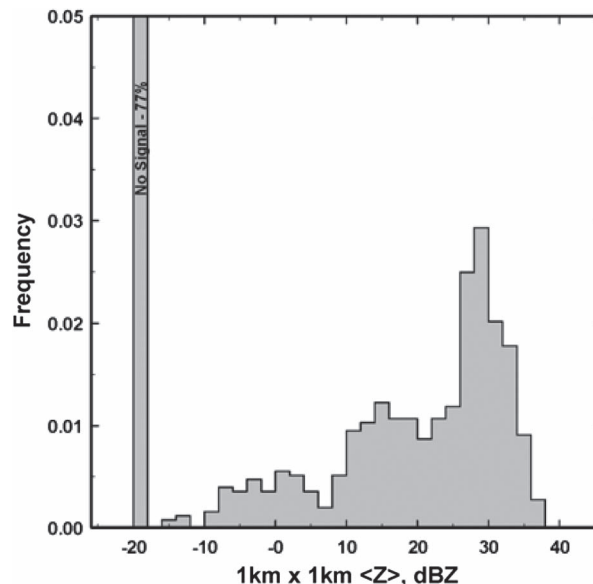


Fig. 4. Histogram of the radar mean reflectivity factors, which a radar having a 1° beam width would have measured at locations centered 1 km apart.

tics fluctuations (64 independent samples) and partial beam filling. (The signal fluctuations were simulated by drawing a sample from the Erlang distribution of mean values peaked at the original matrix mean value and corresponding to the nominal TRMM 64 independent samples as is standard in radar meteorology.) As one would expect, there are differences between what was there and what the TRMM radar might have seen. This is discussed further next.

With these results, one can also ask what a ground-based radar with a 1° beam at 60 km (i.e., nominal 1 km × 1 km) having a minimum threshold of useable $Z$ of $-20$ dBZ would measure at locations each separated from the other by 1 km (i.e., independent samples having no overlap). This is illustrated in Fig. 4 where in this case, there is a distribution of possible values, the most frequent being no signal! There is, therefore a lot of smoothing of the high-resolution field in Fig. 3(a) and a reduction of the largest values.

The rainfall rate $R$ can be also computed using the DSD at each grid point, i.e., $R = (\pi \rho_w/6) \sum_{D_{\min}}^{D_{\max}} V_t(D)D^3$ where $D_{\min}$ and $D_{\max}$ are the minimum and maximum drop diameter at each grid point, respectively, $\rho_w$ is the density of water and $V_t$ is the terminal fallspeed of the drop of diameter $D$. These are plotted in Fig. 5. As with $Z$, the larger values of $R$ occur in the upper right portion of the figure. For comparison, the histograms of the observed and 2-D simulated $R$ are plotted in Fig. 6. The two histograms are quite similar although the simulations have about 4% more frequent lighter rains than the observations most likely because of randomness in the simulation process.

With these two variables and this expanded data set, one can also compute the so-called $Z–R$, relation as given in Fig. 7. In addition, the $Z–R$ relation corresponding to the individual disdrometer observations is plotted. The fits to the 2-D simulated mean values and the disdrometer observations are quite consistent with historical values and are quite similar. However, they do differ. Moreover, the $Z–R$ relation for the 2-D values is
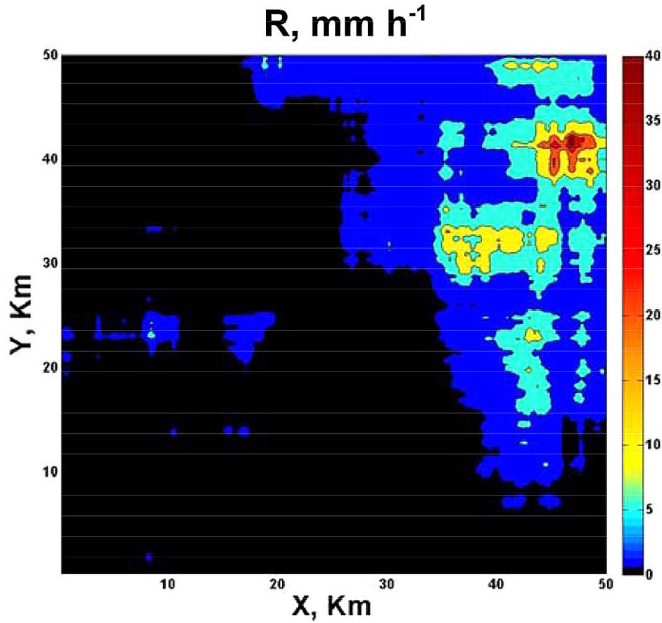
Fig. 5.   Contour plot of the rainfall rate in mmh$^{-1}$. Note some of the similarities to Fig. 8.
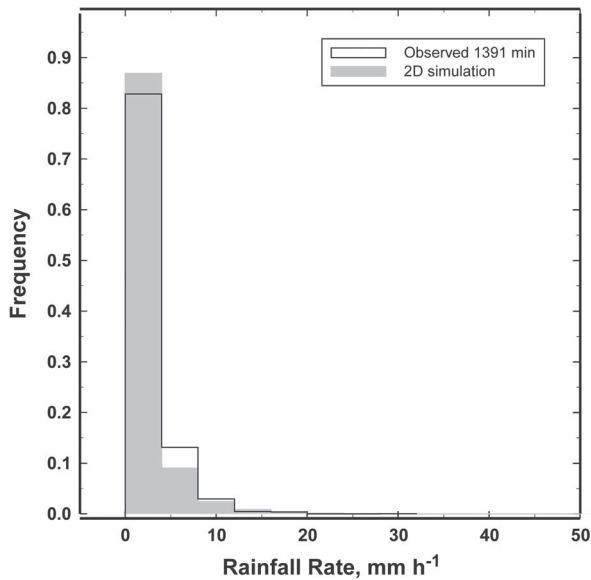


Fig. 7.   (Black) $Z-R$ relations for the 2-D simulation, (red) the 1391 observations, and (green) what the TRMM radar would have observed simulation. $\rho$ is the magnitude of the correlation, whereas $\sigma_\varepsilon$ is the standard deviation of the error. Note that there are differences as those discussed in the text.



Fig. 6.   Histograms of the observed versus the 2-D simulation rainfall rates.

statistically "tighter" in that the standard deviation of the error $(\sigma_\varepsilon)$ is only about 1/3 that for the relation computed using only the individual observations. Hence, the fuller expression of the data set reduces statistical uncertainty in part, because we are using values computed from the DSDs of mean concentrations rather than the usual random draws from a set of means as occurs in short-term observations often used when formulating such relations and, in part, from the much larger sample size. Furthermore, unlike for the individual samples, in the 2-D values several smaller $Z-R$ relations are apparent as shorter line segments cross at an angle with respect to the mean fit. The straight line segments are reminiscent of straight line relations associated with physically (not simply statistically) meaningful DSD at those locations [10].
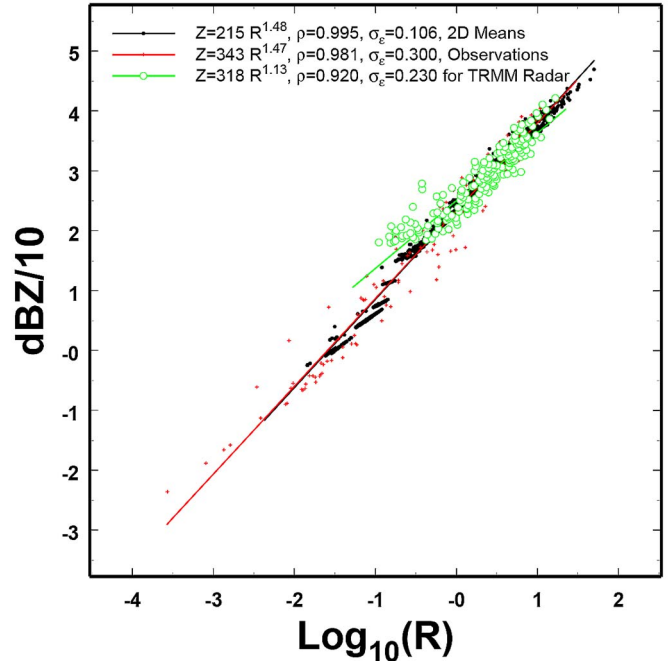
In addition, the expanded set of disdrometer values can be used to explore what the TRMM radar might have seen, for example, if it had observed this simulation field. That is, we convolve the beam with the 2-D grid values of both $R$ and $Z$ (with signal fluctuations). To improve visibility while retaining the structure, only every 500th point (80 representative values) are plotted in Fig. 7. This illustrates how the TRMM $Z-R$ relation (derived using all 18 000 values greater than 18 dBZ as required by the TRMM radar) differs from the other two because of the threshold, beam effects, signal statistics fluctuations, and partial beam filling.

Moreover, this TRMM $Z-R$ relation differs from either of those used in the official TRMM algorithms for convective rain ($Z = 146R^{1.54}$) and for stratiform rain ($Z = 292R^{1.53}$), regardless of how one classifies this simulated rain. Taking all 121 statistically independent samples (that is, all samples separated by one beam width) leads to differences as illustrated in Fig. 8. This figure shows how that this technique may be useful for evaluating the estimates of the various TRMM algorithms by providing a method for a more accurate assessment of the variability associated with TRMM radar estimates. The TRMM simulated "observations" calculated from the expanded disdrometer observations fall mostly (but not entirely) between the results for the two TRMM algorithms. However, the point to be made here is simply to illustrate how the approach for extracting much more information from the disdrometer measurements might be useful for algorithm development of future space-borne instruments and not to directly analyze $Z-R$ relations. That is not the purpose of this paper.

Another important parameter related to the DSD is the total concentration of particles ($C_{\text{tot}}$) as plotted in Fig. 9. Once
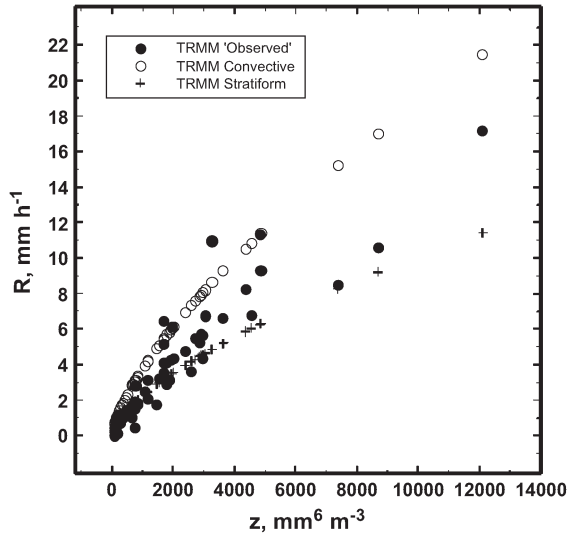
Fig. 8. "Observed" TRMM $Z$ and $R$ (i.e., the statistically independent beam-weighted averages) are compared with results using the TRMM convective and stratiform $Z$–$R$ relations as discussed further in the text.
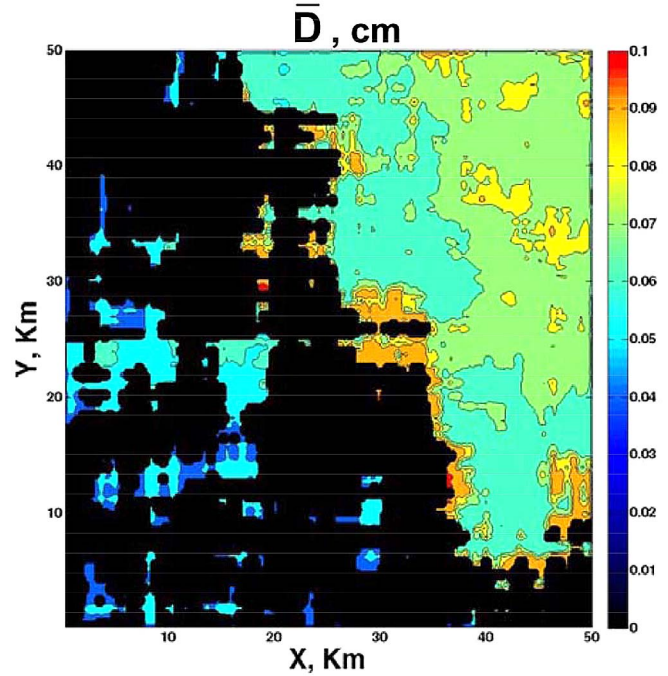


Fig. 10. Contour plot of the mean drop size in cm. Its variability indicates the variability in the simulated DSD.
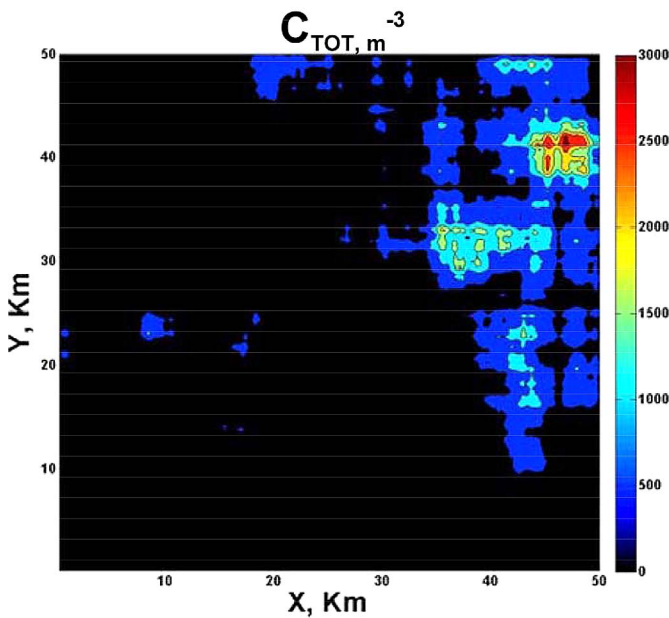


Fig. 9. Contour plot of the total crop concentration in m$^{-3}$. Note the strong similarity to Fig. 5.
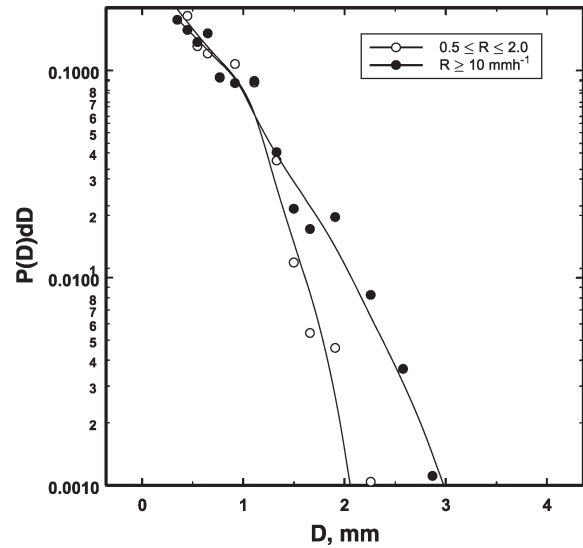


Fig. 11. Mean pdfs each over different ranges of rainfall rate intensities. On average, larger drops are found more frequently in heavier rain, consistent with many past observations.

again, the highest concentrations of particles are in the upper right part of the grid, and the overall structure is quite similar to that of the rainfall rate. We discuss this more momentarily.

Another characteristic of the DSD is the mean drop size, which is the diameter weighted sum over all the drops divided by the total drop concentration at each grid point. This is plotted in Fig. 10. Clearly, the DSD vary. While there are pockets of larger drops embedded within the upper right region associated with the greater values of $Z$ and $R$, interestingly the fringes of the region contain some of the largest mean drop sizes arising from narrow size distributions containing a few large drops. This is reminiscent of what can happen on the boundaries of some storms because of size sorting. It is also interesting,

however, that the overall features in the figure show greater similarity with $Z$ than with $R$.

Nevertheless, in a mean sense, the DSD appear to be different in lighter as opposed to heavier rain as illustrated in Fig. 11. The two DSDs are normalized with respect to the total concentration so that this figure is representative of the pdf of $D$. Obviously, the heavier rains are associated with greater frequency of larger drops consistent with relations derived in many past studies, such as those of [15] and [17], for example.

However, while the contour plots of $Z$ and mean drop size show similarities, they are different from those of $R$ and $C_{\text{tot}}$. Another way of saying this is that apparently the driver in
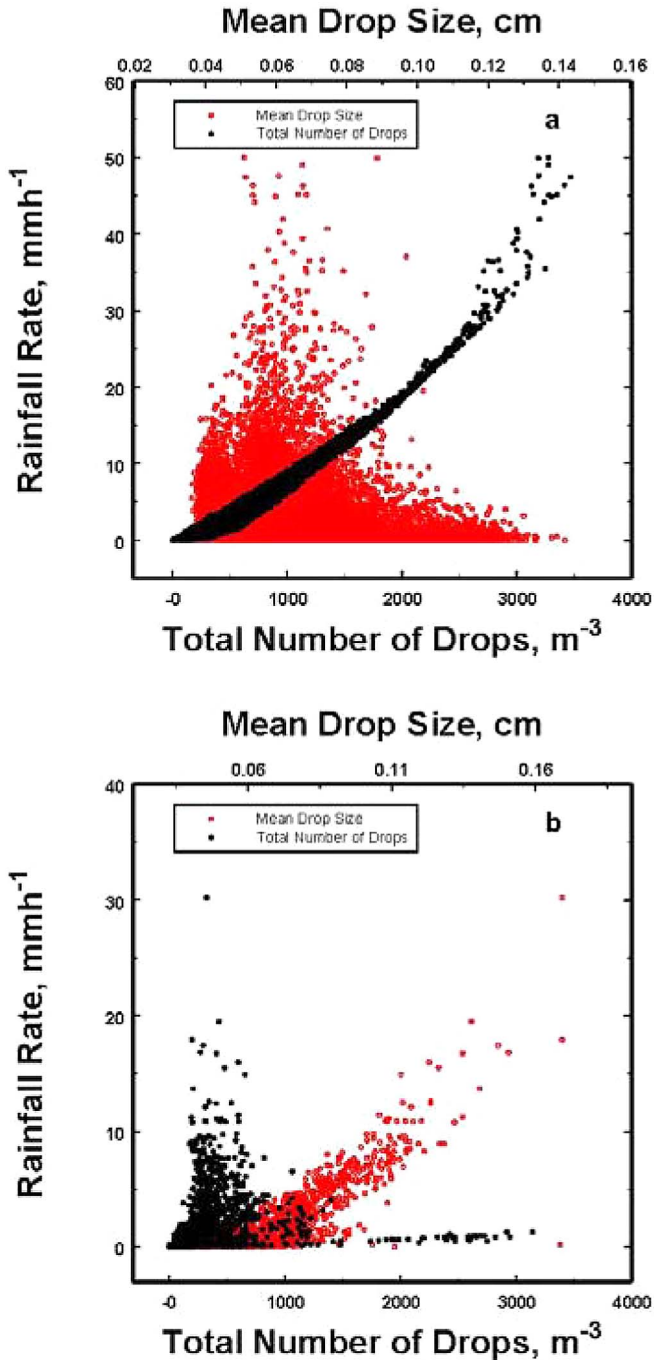
Fig. 12.   Scatter diagrams between the total drop concentration ($C_{tot}$) and the mean drop size to the rainfall rate, $R$, for (a) the 2-D means and (b) for the 1391 individual observations. This illustrates the difference between using mean values as opposed to limited single observations for probing relationships among variables as discussed in the text.

the rainfall rates is not so much the drop size (DSD) as the total number of drops. This is apparent in Fig. 12(a), in which the high correlation between $C_{tot}$ and $R$ is clear, whereas the correlation between mean drop size $\overline{D}$ and $R$ is much, much weaker. This is consistent with the findings in Jameson and [10], in which the number of drops is the most dominant variable in $R$.

These results appear to differ from those one would find just using the individual observations as illustrated in Fig. 12(b)

where it is $\overline{D}$ rather than $C_{tot}$ that appears to be most strongly correlated to $R$. This serves to illustrate a fundamental advantage of this 2-D approach, namely, that of using **mean** values rather than using individual observations. The reason is that an individual measurement is the result of one more random draw from the random (but correlated) mean values so that they contain one more layer of randomness than in Fig. 12(a). This additional randomness can greatly reduce or even destroy any evidence of correlations among parameters that might exist. However, there is another perverse problem with individual observations, as well. When the observations contain fewer than 10 000–100 000 drops, artificial correlations appear in some of the relations among parameters as discussed in [11] and as further explored in [20]. In particular, Fig. 3 in [11] shows just the sort of artificial "relationship" between $\overline{D}$ and $R$ illustrated in Fig. 12(b). It may well be conjectured that the historic emphasis on the importance of drop size rather than drop concentration to $R$ is, in part, a consequence of the limited sample sizes by most historic instruments used to study rain. On the other hand, Fig. 12(a) suggests that one could imagine taking an estimate of the mean drop size (either *a priori* or, for example, perhaps by using polarization radar), converting $Z$ to $C_{tot}$ and then using $C_{tot}$ to estimate $R$.

## IV. SUMMARY

Atmospheric phenomena, including precipitation, occur over an extremely broad range of spatial and temporal scales all of which contribute to varying degrees to the meteorology and climate of the Earth. It would not be surprising, then, to find that the properties of rain over large scales are intimately related at some level to their structure on smaller scales.

It is also true that most of what is known about rain comes from instrumentation best suited to measurements on smaller scales. Using the technique described here, the size of a data set can be expanded to better characterize the rain and to extend the applicability of a set of point measurements to cover extended areas in a manner consistent with the observed statistical properties of the rain.

A time or spatial series of counts is but one realization of a multiple stochastic process. The method presented here extracts more of the information contained in the time series of 1-min Joss–Waldvogel disdrometer counts in rain than a simple analysis of the magnitudes of the counts would provide. This is done by greatly increasing the size of the data set using a Bayesian analysis of drop count measurements in 17 size bins and the copula statistical technique of pdf transformations, a 1391-min time series of drop counts was expanded to the equivalent of 40 000 min. This is particularly important for studying the physics of rain since it provides a much fuller expression of all of the information contained within the correlations and $P(C)$ than is given by just the one realization actually observed. Specifically, consistent with past research [10] it was shown using the 40 000 values that the rainfall rate is much more strongly correlated to the total number of drops than to the mean drop size in diametric contrast to what one would conclude just using the set of 1391 observations.

Moreover, this increase in sample size also permits one to translate these counts into a $200 \times 200$ grid filled at each point with DSDs of *mean* drop concentrations consistent with the observed statistical properties of the rain, thus permitting a translation of a point measurement to an area. Under the assumption of constant wind velocity, each 1-min value was interpreted as being representative of a measurement over 250 m. (Other assumptions are possible, of course, and had wind observations been available, a more relevant spatial interpretation of these temporal observations would have been possible. However, this assumption is adequate for the purpose of this paper, which is to demonstrate a methodology.)

A square 2-D grid of $200 \times 200$ 250 m "points" was filled with uniformly distributed random numbers having zero mean and unit variance. For each drop size, this field was correlated using the different observed correlation distance for each drop size and using the root matrix correlation method. As explained in the text, these fields were then adjusted and the empirical copula method was used to generate 2-D fields of mean drop concentrations at each drop size over the 40 000 grid points. Combining these fields overall drop sizes produced the DSD of mean concentrations at each grid point having the correct correlations and distributions of mean values of the drops as actually observed. These DSDs can then be integrated appropriately to yield the radar reflectivity factors, rainfall rates, total particle concentrations, and mean drop sizes over the entire grid. The results appear realistic, and the DSDs are consistent with the observed statistical properties. In that sense, they automatically include the "meteorology" vis-á-vis the correlation function and $P(C)$.

## References

[1] E. Frees and E. Valdez, "Understanding relationships using copulas," *North Amer. Actuarial J.*, vol. 1, no. 2, pp. 1–25, Jan. 1998.

[2] C. Genest and J. Mackay, "The joy of copulas: Bivariate distributions with uniform marginals," *Amer. Stat.*, vol. 40, no. 4, pp. 280–283, Nov. 1986.

[3] G. Golub and C. Van Loan, *Matrix Computations*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, p. 143.

[4] J. Jaffrain and A. Berne, "Quantification of the small-scale spatial structure of the raindrop size distribution from a network of disdrometers," *J. Appl. Meteorol. Climatol.*, vol. 51, no. 5, pp. 941–953, May 2012.

[5] A. R. Jameson, "A new characterization of rain and clouds: Results from a statistical inversion of count data," *J. Atmos. Sci.*, vol. 64, no. 6, pp. 2012–2028, Jun. 2007.

[6] A. R. Jameson and A. J. Heymsfield, "Upscaling ice size distributions over two-dimensional domains for large scale applications," *Meteorol. Atmos. Phys.*, vol. 123, no. 1/2, pp. 93–103, Jan. 2014.

[7] A. R. Jameson and A. B. Kostinski, "Fluctuation properties of precipitation. Part II: Reconsideration of the meaning and measurement of raindrop size distributions," *J. Atmos. Sci.*, vol. 55, no. 1, pp. 283–294, Jan. 1998.

[8] A. R. Jameson and A. B. Kostinski, "Fluctuation properties of precipitation. Part V: Distribution of rain rates—Theory and observations in clustered rain," *J. Atmos. Sci.*, vol. 56, no. 11, pp. 3920–3932, Nov. 1999.

[9] A. R. Jameson and A. B. Kostinski, "Fluctuation properties of precipitation. Part VI: Observations of hyper-fine clustering and drop size distribution structures in three-dimensional rain," *J. Atmos. Sci.*, vol. 57, no. 3, pp. 373–388, Feb. 2000.

[10] A. R. Jameson and A. B. Kostinski, "Reconsideration of the physical and empirical origins of Z-R relations in radar meteorology," *Q. J. R. Meteorol. Soc.*, vol. 127, no. 572, pp. 517–538, Jan. 2001.

[11] A. R. Jameson and A. B. Kostinski, "Spurious power-law relations among rainfall and radar parameters," *Q. J. R. Meteorol. Soc.*, vol. 128, no. 583, pp. 2045–2058, Jul. 2002.

[12] A. R. Jameson and A. B. Kostinski, "The effect of clustering on the uncertainty of differential reflectivity measurements," *J. Appl. Meteorol. Climatol.*, vol. 47, no. 11, pp. 2816–2827, Nov. 2008.

[13] G. E. Johnson, "Constructions of particular random processes," *Proc. IEEE*, vol. 82, no. 2, pp. 270–285, Feb. 1994.

[14] G. Lee, A. W. Seed, and I. Zawadzki, "Modeling the variability of drop size distributions in space and time," *J. Appl. Meteorol. Climatol.*, vol. 46, no. 7, pp. 742–756, Jun. 2007.

[15] J. S. Marshall and W. M. K. Palmer, "The distributions of raindrops with size," *J. Meteorol.*, vol. 5, no. 4, pp. 166–169, Aug. 1948.

[16] R. B. Nelsen, *An Introduction to Copulas*. New York, NY, USA: Springer-Verlag, 1999, ch. 1/2, p. 7105.

[17] R. S. Sekhon and R. C. Srivastava, "Doppler radar observations of drop-size distributions in a thunderstorm," *J. Atmos. Sci.*, vol. 28, no. 6, pp. 983–994, Sep. 1971.

[18] M. Schleiss, J. Jaffrain, and A. Berne, "Stochastic simulation of intermittent DSD fields in time," *J. Hydrometeorol.*, vol. 13, no. 2, pp. 621–637, Apr. 2012.

[19] D. S. Sivia, *Data Analysis: A Bayesian Tutorial*. New York, NY, USA: Oxford Univ. Press, 1996, ch. 1–5, pp. 1–128.

[20] M. Steiner, J. A. Smith, and R. Uijlenhoet, "A microphysical interpretation of radar reflectivity-rain rate relationships," *J. Atmos. Sci.*, vol. 61, no. 10, pp. 1114–1131, May 2004.

**A. R. Jameson** received the B.S. degree in geophysical sciences from the University of Chicago, Chicago, IL, USA, in 1968. Following the Vietnam War, he received the M.S. and Ph.D. degrees from the University of Chicago, in 1972 and 1977, respectively.

His professional career began as a Research Scientist with the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in the State Water Survey in 1978. In 1985, he joined the Applied Research Corporation, Landover, MD, USA as a Senior Scientist. In 1995, he joined RJH Scientific, Inc. based in San Diego, CA, USA as a Senior Scientist, where he is currently employed. His research includes the evolution of hail in storms, radar polarization techniques, and applications to the measurement of rainfall, non-Rayleigh signal statistics, and the stochastic structures of precipitation and clouds and their impact on radar coherent backscatter from precipitation.

Dr. Jameson is a Fellow of the American Meteorological Society, a member of the American Geophysical Union, and a member of the American Association for the Advancement of Science.