Scale dependence of spatiotemporal intermittence of rain

Prasun K. Kundu[1] and Ravi K. Siddani[2]

Received 30 September 2010; revised 4 May 2011; accepted 27 June 2011; published 20 August 2011.

[1] It is a common experience that rainfall is intermittent in space and time. This is reflected by the fact that the statistics of area- and/or time-averaged rain rate is described by a mixed distribution with a nonzero probability of having a sharp value zero. In this paper we have explored the dependence of the probability of zero rain on the averaging space and time scales in large multiyear data sets based on radar and rain gauge observations. A stretched exponential formula fits the observed scale dependence of the zero-rain probability. The proposed formula makes it apparent that the space-time support of the rain field is not quite a set of measure zero as is sometimes supposed. We also give an explanation of the observed behavior in terms of a simple probabilistic model based on the premise that rainfall process has an intrinsic memory.

Citation: Kundu, P. K., and R. K. Siddani (2011), Scale dependence of spatiotemporal intermittence of rain, *Water Resour. Res., 47*, W08522, doi:10.1029/2010WR010070.

1. Introduction

[2] Irregular variation of rain over a wide range of space-time scales has long presented an intriguing challenge to theoretical modeling efforts. Because of the difficulty that one encounters in predicting the precise atmospheric conditions that determine the occurrence of rain, models of rainfall variability often employ a statistical framework in which rain is described in terms of a random variable governed by an underlying probability distribution. The intermittent nature of rain in both space and time is a matter of common experience. In a statistical description this is reflected by the fact that the distribution of the rain rate is a mixed distribution with a finite nonzero probability of zero rain. [*Kedem and Chiu*, 1987; *Bell*, 1987].

[3] Experimental measurements generally yield precipitation data as averages over space-time scales determined by the resolution achieved in a particular measurement method. Radar observations provide a near-instantaneous rainfall image with horizontal spatial resolution of the order of 1 km. Rain gauge and disdrometer measurements, on the other hand, yield rain rate estimates at a point with a time resolution of the order of 1 min. Although from a microscopic point of view rain is particulate in nature, under normal rainy conditions the rain rate $R(\mathbf{x}, t)$ can be treated as a continuous stochastic field when averaged at these scales. This continuum approximation is not unlike the one employed in many areas of traditional physics, for example, in kinetic theory for describing transport phenomena in dilute gases.

[4] A fundamental property of rainfall statistics is the dependence of the probability distribution of rain on the space-time averaging scale [*Kundu and Siddani*, 2007]. Power law scaling of the moments of rain rate, or equivalently, the tail of the distribution empirically observed over a significant range of scales, has been the main motivation behind the popular multifractal models of rain in terms of a multiplicative random cascade process. Many variants of such models exist in the literature that describe spatial, temporal, or full space-time multiscaling [e.g., *Gupta and Waymire*, 1990, 1993; *Schertzer and Lovejoy*, 1987; *Lovejoy and Schertzer*, 1990; *Over and Gupta*, 1996; *Marsan et al.*, 1996; *Menabde et al.*, 1997]. These models have been extensively used to simulate rainfall process. Their usefulness comes from the prospect of their being able to faithfully capture the subgrid variability by extrapolating the rainfall process statistically from the scales of observational resolution down to the much smaller scales that are relevant for hydrological applications.

[5] A particular feature of rainfall that is just as fundamental is the frequent occurrence of zeroes in a precipitation field, a property referred to as intermittence. Intermittence contributes significantly to the observed dependence of the rainfall statistics on the averaging area or time. Moreover, the statistical properties of the zero-rain regions must themselves be scale dependent [*Kundu and Siddani*, 2007] since coarse graining of a precipitation field by averaging over adjacent space-time regions of zero and nonzero rain produces a region that is rainy when viewed on a coarser scale. Despite being an essential aspect of the statistics of precipitation, the distribution of rainfall zeroes by themselves has received relatively limited attention among theoretical investigators. The presence of zeroes causes the probability distribution of rain to be described by a mixed distribution where there is a nonzero probability of rain rate assuming the sharp value zero. An early effort to model monthly rainfall totals in an arid region in terms of a distribution suggested by *Fisher and Cornish* [1960] that includes the zeroes of rain goes back to *Öztürk* [1981] (who attributes it to *Buishand* [1977]; see also *Revfeim* [1982]). The model describes rainfall as a Poisson process of instantaneous

[1]Joint Center for Earth Systems Technology, University of Maryland Baltimore County and NASA Goddard Space Flight Center, Baltimore, Maryland, USA.
[2]Merck & Co., Inc., Kenilworth, New Jersey, USA.

Copyright 2011 by the American Geophysical Union.
0043-1397/11/2010WR010070

showers that are mutually independent, with the event size negative-exponentially distributed. The temporal distribution of dry and wet events and its scaling property have been studied by *Hubert and Carbonnel* [1989], *Olsson et al.* [1993], and *Schmitt et al.* [1998]. Power law scaling behavior of the rainy area fraction in gridded radar data over a limited range of spatial scales was recognized by *Kedem and Chiu* [1987], *Over and Gupta* [1994], and *Kundu and Bell* [2003]. Dependence of the tail quantiles and moments of the duration of wet and dry events on spatial scale in radar data has been investigated by *Pavlopoulos and Gupta* [2003]. In the context of a multifractal cascade model often used to simulate a rain field [e.g., *Over and Gupta*, 1994, 1996; *Paulson and Baxter*, 2007], the zeroes are generated by allowing the cascade generator to have zero value with a finite probability, a theoretical construct commonly referred to as the beta model, which was initially utilized in theory of fluid turbulence to represent intermittence of turbulent flow. The effect of intermittence on the estimation of the parameters of the so-called universal multifractal model [*Schertzer and Lovejoy*, 1987; *Lovejoy and Schertzer*, 1990], which does not naturally include zero rain, was investigated by *de Montera et al.* [2009]. Recently, *Martínez et al.* [2007] and *Lana et al.* [2010] have explored various aspects of the distribution of dry spell lengths over the Iberian Peninsula and Europe. They quantified the distribution of zero-rain gaps at the daily scale in terms of its lacunarity, a concept originated by *Mandelbrot* [1982], and attempted to reproduce the observed lacunarity from synthetic time series generated by suitably constructed deterministic or random Cantor sets. They also investigated the possibility of persistence or long memory in the time series of dry spell lengths by examining the Hurst scaling exponent from the scaling behavior of the so-called *R*/*S* statistic [*Beran*, 1994] and applied considerations of dynamical systems theory to characterize the underlying random process that potentially generates the time series.

[6] In the present paper we investigate the scale dependence of intermittence for both spatial and temporal averages of rainfall from a phenomenological point of view. Our study leads to a simple empirical formula expressing the dependence of the zero-rain probability (which measures the fractional area or time for which the rain rate is zero) on the averaging area or time. Our finding is based on a detailed exploration of large multiyear precipitation data sets produced from ongoing radar and rain gauge measurements in Florida that were collected as part of the ground validation (GV) program under the Tropical Rainfall Measuring Mission (TRMM) undertaken jointly by NASA and Japan Aerospace Exploration Agency (JAXA).

[7] The outline of the paper is as follows. In section 2 we briefly explain the theoretical framework employed in describing the statistics of dry and wet events. In section 3 we present a simple probabilistic model that yields an explicit formula for the zero-rain probability as a function of the time interval for a rain gauge time series. The model we propose is based on the premise that the rainfall process is intrinsically non-Markovian, i.e., a process with memory. A similar formula also holds in the two-dimensional case of spatial radar data. Section 4 gives an account of our analysis of the various radar and gauge data sets. In section 5 we then present the results of data analysis and discuss them in the context of other related investigations. The paper is concluded in section 6 with a summary of the findings and some directions for future work. Appendix A is devoted to the proofs of some mathematical formulas, and Appendix B describes the statistical tests conducted to assess the spatial homogeneity of the Florida gauge network.

2. Theoretical Framework

[8] In this section we describe the theoretical framework employed in our data analysis. Let

$$r_{AT}(\mathbf{x},t) = (1/AT)\int_{A(\mathbf{x})} \mathrm{d}^2\mathbf{x}' \int_{t-T/2}^{t+T/2} \mathrm{d}t' R(\mathbf{x}',t') \tag{1}$$

denote the area-time average of the instantaneous point rain rate field over an area $A(\mathbf{x})$ centered at the point $\mathbf{x}$ and over a time interval T centered at the instant t. Radar and rain gauge measurements capture the statistics of the random variable r_{AT} in two distinct spatiotemporal regimes, the large A, small T and the large T, small A regimes. The statistical properties of the rain rate field are assumed to be stationary in both space and time.

2.1. Temporal Statistics of Dry and Wet Periods in Rain Gauge Data

[9] First, we consider the statistics of time-averaged rain rate data obtained from a set of rain gauges located within a region in which the spatial statistics is assumed to be homogeneous. Let $r_T \equiv r_T(t)$ denote the time average of the rain rate measured by a gauge over a time interval T centered at the instant t. Each rain gauge provides a time series $\{r_{Ti};\ i = 1, 2, \ldots\}$ of time-averaged rain rates at equally spaced points. In this paper we are interested only in occurrence or nonoccurrence of rain; the actual numerical value of the rain rate is irrelevant. Consequently, it is convenient to introduce a rain indicator field $\xi_T \equiv \xi_T(t)$, which is a binary random variable equal to 1 when $r_T > 0$ and 0 when $r_T = 0$. In accordance with our assumption, the statistics is regarded as stationary, i.e., independent of the choice of t; ξ_T generates a stochastic process of "trials" with a binary outcome: rain or no rain. The probabilities of nonzero rain ("success") and zero rain ("failure") in a randomly chosen time interval T are given by $p_1(T) \equiv \langle \xi_T \rangle = \Pr[r_T > 0]$ and $p_0(T) = 1 - p_1(T) = \langle 1 - \xi_T \rangle = \Pr[r_T = 0]$, respectively. An important aspect of the process is that unlike the familiar Bernoulli trials, here the successive trials are not independent. The dependency arises through the space-time correlation inherent in the physical precipitation process and can be expressed through the joint probabilities for two randomly chosen time intervals of length T separated by a lag τ:

$$\begin{aligned}
p_{11}(T;\tau) &\equiv \langle \xi_T \xi'_T \rangle = \Pr[r_T > 0, r'_T > 0],\\
p_{10}(T;\tau) &\equiv \langle \xi_T (1-\xi'_T) \rangle = \Pr[r_T > 0, r'_T = 0],\\
p_{01}(T;\tau) &\equiv \langle (1-\xi_T)\xi'_T \rangle = \Pr[r_T = 0, r'_T > 0],\\
p_{00}(T;\tau) &\equiv \langle (1-\xi_T)(1-\xi'_T) \rangle = \Pr[r_T = 0, r'_T = 0],
\end{aligned} \tag{2}$$

where a prime indicates evaluation at instant $t + \tau$ instead of t, i.e., $\xi'_T = \xi_T(t + \tau)$ and so on. They satisfy the elementary relations

$$\begin{aligned} p_{11}(T;\tau) + p_{10}(T;\tau) &= p_1(T) \\ p_{01}(T;\tau) + p_{00}(T;\tau) &= p_0(T), \end{aligned} \tag{3}$$

with $p_0(T) + p_1(T) = 1$. For simplicity, we assume the statistics to be stationary, i.e., independent of the choice of t. Consequently, $\langle \xi_T \rangle = \langle \xi'_T \rangle$, which implies the time symmetry property

$$p_{01}(T;\tau) = p_{10}(T;\tau). \tag{4}$$

The degree of dependence can be quantified by the lagged autocorrelation function of the indicator field

$$\gamma(T;\tau) = \frac{\langle \xi_T \xi'_T \rangle - \langle \xi_T \rangle^2}{\langle \xi_T^2 \rangle - \langle \xi_T \rangle^2}. \tag{5}$$

Using the basic definitions above and the fact that $\langle {\xi_T}^2 \rangle = \langle \xi_T \rangle$, the lagged autocorrelation $\gamma(T;\ \tau)$ can be written (dropping the arguments to simplify notation) in the following equivalent forms:

$$\gamma = \frac{p_{00} - p_0^2}{p_0 - p_0^2} = \frac{p_{11} - p_1^2}{p_1 - p_1^2} = -\frac{p_{01} - p_0 p_1}{p_0 p_1}. \tag{6}$$

As rain decorrelates (as would be expected for large τ), $\gamma \to 0$, and consequently, the joint probabilities factorize, i.e., $p_{00} \to {p_0}^2$, $p_{11} \to {p_1}^2$, and $p_{01} \to p_0\, p_1$.

[10] Finally, since two consecutive zero-rain intervals of length T separated by an interval $\tau \le T$ overlap to form a zero-rain interval of length $T + \tau$, it immediately follows that

$$p_{00}(T;\tau) = p_0(T + \tau)(\tau \le T). \tag{7}$$

[11] For $\tau > T$, the intervals of zero rain are disjoint. Consequently, one must have $p_{00}(T;\ \tau) > p_0(T + \tau)$ in order to allow the possibility of rain occurring during the interval that separates them.

[12] It should be noted that although, in principle, T and τ can be treated as continuous variables, in practice, precipitation data are usually available only as a discrete time series at equidistant points. The spacing between two consecutive points is a certain nonzero time interval θ whose value is dictated by the finite temporal resolution of the experiment. All allowed values of T and τ are therefore integer multiples of θ. Thus, for a randomly selected time interval $[0,\ T]$ of length $T = n\theta$ consisting of n consecutive time intervals θ, the probability of zero rain can be expressed as

$$p_0(T) \equiv p_0(n\theta) = \Pr\left[\xi^{(n)} = 0,\ \ldots,\ \xi^{(1)} = 0\right], \tag{8}$$

where $\xi^{(j)} \equiv \xi_\theta(t_j)$ $(j = 1,2,\ \ldots,\ n)$ denotes the indicator for the interval j centered at $t_j = (j - 1/2)\theta$.

[13] Another quantitative measure of intermittence is the lacunarity of a rain rate time series [see, e.g., *Martínez et al.*, 2007]. Let $P(k,\ n)$ denote the probability that there are k zero-rain gaps of length θ (i.e., $\xi_\theta = 0$) in an interval of total length $T = n\theta$. The first and second moments of k with respect to the probability distribution $P(k,\ n)$ are

$$M_1(n) = \sum\nolimits_{k=1}^{n} kP(k,n), M_2(n) = \sum\nolimits_{k=1}^{n} k^2 P(k,n),$$

and the lacunarity (at the scale n) is defined as the quotient $\Lambda_\theta(n) = M_2(n)/M_1^2(n)$. If the trials generated by the binary variable ξ_θ are independent, i.e., represent Bernoulli trials, then $P(k,\ n)$ is simply the binomial distribution:

$$P(k,n) = \binom{n}{k} [p_0(\theta)]^k [p_1(\theta)]^{n-k},$$

where the symbol in parentheses on the right-hand side denotes the binomial coefficient. In particular, $P(n,n) \equiv p_0(T) = [p_0(\theta)]^n$, as expected. For the binomial distribution, we have $\Lambda_\theta(n) = 1 + p_1(\theta)/np_0(\theta)$.

[14] A related quantity of interest in hydrology is the distribution of the duration of a dry or wet spell, the so-called residence time [see, e.g., *Schmitt et al.*, 1998]. The probability of occurrence of a continuous dry spell of duration $T = n\theta$ is given by the conditional probability

$$\Pi_n = \begin{cases} \Pr\left[\xi^{(n+1)} = 1, \xi^{(n)} = 0,\ \ldots, \xi^{(1)} = 0 \middle| \xi^{(0)} = 1\right]; & n > 0 \\ \Pr\left[\xi^{(1)} = 1 \middle| \xi^{(0)} = 1\right]; & n = 0 \end{cases}. \tag{9}$$

It can be readily expressed in terms of $p_0(T)$. It is not difficult to show that Π_n are given by the formula (see Appendix A for a proof)

$$\Pi_n = \begin{cases} \{p_0(n\theta) - 2p_0[(n+1)\theta] + p_0[(n+2)\theta]\}/p_1(\theta) & (n > 0) \\ [1 - 2p_0(\theta) + p_0(2\theta)]/p_1(\theta) & (n = 0) \end{cases}. \tag{10}$$

We can easily check that they satisfy the normalization condition

$$\sum\nolimits_{n=0}^{\infty} \Pi_n = 1.$$

Now, we consider the dry residence time, i.e., the length of a dry spell separating the occurrence of rain, regarded as a new random variable $\mathbf{T}$. Then, by definition, $\Pi_n = \Pr[\mathbf{T} = n\theta]$. The probability $P_n = \Pr[\mathbf{T} \ge n\theta]$ that a dry spell of duration greater than or equal to $T = n\theta$ $(n > 0)$ occurs is obtained by partially summing the Π_m over all $m \ge n$:

$$P_n \equiv \sum\nolimits_{m=n}^{\infty} \Pi_m = \{p_0(n\theta) - p_0[(n+1)\theta]\}/p_1(\theta)(n > 0). \tag{11}$$

Regarding θ as infinitesimal, we can reexpress equation (11) as (prime denotes derivative with respect to T)

$$P_n = -p_0'(n\theta)\theta/p_1(\theta) \quad (n > 0), \tag{12}$$

where it is assumed implicitly that $p_0(T)$ is differentiable. However, passage to the continuous limit is conceptually somewhat subtle. Formally, it involves taking the double

limit $\theta \to 0$, $n \to \infty$, with $T = n\theta$ held fixed at a finite (nonzero) value. One would naturally like to consider the probability density function (pdf) $\phi(T)$ of $\mathbf{T}$, which is now regarded as a continuous random variable. In the continuous limit the quantity Π_n is to be identified with $\phi(T)\theta = \Pr[T \leq \mathbf{T} \leq T + \theta]$, the probability that $\mathbf{T}$ assumes values lying between $n\theta$ and $(n + 1)\theta$. Recognizing that the numerator in equation (10) is proportional to the second derivative of $p_0(T)$, it follows that

$$\phi(T) = p_0''(T)[\theta/p_1(\theta)]_0 \quad (T > 0), \tag{13}$$

where $[\ ]_0$ denotes the limit of vanishingly small θ. The quantity P_n then goes over to the complimentary cumulative distribution function (CDF), also called the survival function, which represents the falloff behavior of the upper tail:

$$1 - \Phi(T) \equiv \int_T^\infty \phi(T')dT' = -p_0'(T)[\theta/p_1(\theta)]_0 \quad (T > 0). \tag{14}$$

An aspect of (14) that might appear puzzling at first is the explicit dependence on the resolution θ that survives the limiting procedure. However, a little reflection shows that this is a necessary consequence of the definition of Π_n via equation (9). Imagine downscaling the temporal resolution by setting $\theta \to \theta' = \theta/\lambda$, $n \to n' = \lambda n$, (λ = integer > 1), leaving T unchanged. This scale transformation does not leave Π_n and P_n unchanged, indicating that they must depend on both T and θ. The θ dependence drops out in the special case in which $p_1(\theta)$ is proportional to θ for small θ, yielding an unambiguous result in the continuous limit. As we shall see later, this indeed occurs if the rainfall zeroes follow a Poisson process. Unfortunately, the Poisson property does not appear to hold for the rainfall data we have examined. In this case our heuristic limiting process indicated in equations (13) and (14) is, strictly speaking, ill defined. We should point out that the continuous limit is a mathematical abstraction since in practice the distribution $\Phi(T)$ is only inferred from a histogram of the discrete finite resolution data.

[15] One can choose to quantify zero-rain intermittence in terms of Π_n (or, equivalently, its continuous version $\phi(T)$) instead of $p_0(T)$ (or its spatial analog $p_0(L)$; see section 2.2), as we have done in this paper. However, our motivation for considering $p_0(T)$ comes from the fact that it is this quantity that naturally appears if one considers the full pdf of rain rate $f(r_T)$, which is a scale-dependent mixed distribution with a Dirac δ function singularity ("atom") at zero [*Kundu and Siddani*, 2007]: $f(r_T) = p_0(T)\delta(r_T) + p_1(T)g(r_T)$, where $g(r_T)$ is the pdf conditional on nonzero rain.

[16] It is, in principle, possible to characterize the temporal zeroes solely in terms of $\phi(T)$, the distribution of the interdrop intervals T, if a high-resolution measuring device is available that can register individual raindrops [*Lavergnat and Golé*, 1998]. However, the question arises whether intermittence is a distinguishable aspect of the temporal gaps, i.e., whether it can be meaningfully separated from the inherent discreteness of the rainfall phenomenon at the scale of individual drops. Clearly, the temporal gaps at the small end of the scale are to be interpreted as the interdrop intervals within a rain episode. On the other hand, the large zero-rain intervals correspond to genuine gaps separating distinct rain episodes. A central question is whether there is a clear-cut separation between the two regimes. It is also important to recognize the interplay between space and time scales. The distribution of the temporal gaps necessarily depends on the spatial observation area.

[17] One preliminary way to explore the issue is to consider the mean interdrop interval during rainy conditions. Consider, for example, observations of $\theta = 1$ min averaged precipitation by a rain gauge of cross sectional area a. The average number of drops striking the gauge per unit time is $Ra/\bar{v}$, where $R \equiv r_\theta$ is the 1 min averaged rain rate measured by the gauge and $\bar{v}$ is the mean volume of a drop defined with respect to the corresponding 1 min drop size distribution function. The quantity $\tau_* = \bar{v}/Ra$ can be taken as a crude measure of the mean interdrop arrival time that quantifies intrinsic temporal discreteness. For light to moderate rain $R \approx 1$ mm h^{-1}, we take the mean drop diameter to be $\sim$1 mm [*Tokay and Short*, 1996]. For a rain gauge of collection area $a \approx 300$ cm^2 we obtain the estimate $\tau_* \approx 0.06$s. Typically for a disdrometer, $a \approx 50$ cm^2, so that $\tau_* \approx 0.4$s. Since $\tau_* << \theta$ in each case, one can argue that a zero of rainfall at scale θ is, in fact, a manifestation of intermittence. Representing rain mathematically as a continuous random process at this scale is justified even at the small spatial scales of a rain gauge or disdrometer.

2.2. Spatial Statistics of Dry and Wet Areas in Radar Data

[18] In a similar manner one can set up a framework for studying intermittence in spatially averaged radar rainfall data. A radar precipitation data set typically consists of a sequence of radar scans, each gridded at a certain spatial resolution l. Radar-retrieved rain rates near the surface have an implicit vertical averaging or, equivalently, averaging over a time scale determined by the mean vertical fall speed of raindrops. Each image can be regarded as an array of $l \times l$ pixels in which the average (near-instantaneous) rain rate is specified. The rain field at the pixel scale can be "coarse grained" by averaging the rain rate over $n \times n$ adjacent pixels to obtain a new rain field at scale $L = nl$. Let r_L denote the average rain rate in an $L \times L$ square grid box $A_L(\mathbf{x})$ centered at the point $\mathbf{x}$ at time t. One can then analogously introduce a rain indicator field $\xi_L \equiv \xi_L(\mathbf{x})$ equal to 1 when $r_L > 0$ and 0 when $r_L = 0$ and define quantities such as the probability of zero rain $p_0(L) = \Pr[r_L = 0]$ in a box $A_L(\mathbf{x})$, the joint probability of zero rain $p_{00}(L; s) = \Pr[r_L = 0, r'_L = 0]$ in a pair of boxes $A_L(\mathbf{x})$ and $A_L(\mathbf{x}')$ separated by a distance s, and so on. The formulation of section 2.1 applies to analogous spatial quantities as well under the assumption of homogeneity and isotropy of the spatial statistics.

2.3. Scale Dependence of the Zero-Rain Probability

[19] The coarse-graining process described in section 2.2 defines a unique deterministic way of carrying out an upscaling of the rain field as well as its associated indicator field. Downscaling, i.e., going from a coarser to a finer resolution, is obviously not unique since information about the structure of the precipitation field is destroyed by averaging. Rather, it can only be defined in a probabilistic manner so that certain key statistical properties of the rain field are preserved under its action.

[20] The zero-rain probabilities $p_0(T)$ and $p_0(L)$ (or, equivalently, the probabilities of nonzero rain $p_1(T)$ and

$p_1(L)$) must necessarily depend on the integration scale. In fact, it is easy to see that $p_1(T)$ and $p_1(L)$ are monotonically increasing functions of their arguments. Consider, for example, aggregating a rain rate time series through a scale transformation $T \to T' = \lambda T$ ($\lambda > 1$). When averaged together, a rainy segment and a nonrainy segment will result in a longer segment with nonzero rain, and consequently, $p_1(\lambda T) > p_1(T)$ when $\lambda > 1$, so that $p_1(T)$ is a monotonically increasing function of T. One expects $p_1(T)$ to approach unity (i.e., $p_0(T) \to 0$) as $T \to \infty$ (since a rain gauge "must experience rain sometime"). If we surmise that gaps in rainfall time series exists at all scales, then it would be natural to also expect that $p_1(T) \to 0$ as $T \to 0$. However, as we shall find later in section 5, the data at the available resolution suggest that at small scales, $p_1(T)$ tends to a small but nonzero value. These observations apply to $p_1(L)$ in the spatial case as well. One additional topological feature in that case is the result of the two-dimensional geometry of rain maps: one can have zeroes as "voids" embedded within a rainy area as well as open areas separating patches of rain. The upscaling of the rain indicator field can be elegantly formulated as a suitable morphological transformation of rain images [*Kumar and Foufoula-Georgiou*, 1994]. The size distribution of the zero-rain voids as a function of scale can then be expressed in terms of the outcome of such a transformation. However, the reverse operation of downscaling is inherently undefined unless augmented by an additional probabilistic element: a realistic model of the spatial distribution of voids at different scales, i.e., quantities such as $p_{00}(L;\, s)$. A downscaling algorithm needs to make specific statistical assumptions in order to generate an ensemble of downscaled fields compatible with a given rain field at coarse resolution. We refrain from pursuing this problem further in this paper and restrict ourselves simply to the search for a realistic form for the marginal distributions $p_0(T)$ and $p_0(L)$ on the basis of empirical evidence.

[21] It is possible to derive a simple analytic expression for $p_0(T)$ by making suitable probabilistic assumptions regarding rainfall intermittence. We describe such a model in section 3.

3. A Simple Probabilistic Model

[22] Consider the probability of occurrence of precipitation during a small time period θ conditioned on a previous dry period of duration of at least T. It is expressed by the conditional probability

$$\pi_{01}(T;\theta) \equiv \Pr[\xi_T(\theta) = 1 | \xi_T(0) = 0] = p_{01}(T;\theta)/p_0(T). \quad (15)$$

We hypothesize that $\pi_{01}(T;\theta)$ depends on T according to a simple power law form,

$$\pi_{01}(T;\theta) = CT^{\chi-1}\theta + o(\theta) \quad (16)$$

as $\theta \to 0$, where C is a constant. Validity of the hypothesis will be checked a posteriori in section 5 by a detailed data analysis to verify its main consequence, equation (18). We shall find that the fit to observational data yields $\chi < 1$, implying that the probability $\pi_{01}(T;\theta)$ decreases with the increase of T. Physically, this indicates a sort of "persistence of drought": the longer a dry spell persists, the less likely it is to end. A proper justification of equation (16) is likely rooted in detailed microphysics of the precipitation process, which is far beyond the scope of this paper.

[23] Since $\pi_{00}(T;\theta) = 1 - \pi_{01}(T;\theta) = p_{00}(T;\theta)/p_0(T)$, application of equation (7) yields a functional equation for the zero-rain probability $p_0(T)$:

$$p_0(T+\theta) = p_0(T)\left[1 - CT^{\chi-1}\theta - o(\theta)\right].$$

The limit $\theta \to 0$ results in the differential equation

$$\frac{\mathrm{d}\ln p_0(T)}{\mathrm{d}T} = -CT^{\chi-1}. \quad (17)$$

Integrating with the initial condition $p_0(T) \to \alpha_0$ as $T \to 0$, one immediately obtains the solution

$$p_0(T) = \alpha_0 \exp[-(T/T_0)^{\chi}], \quad (18)$$

provided the constant C is identified via the relation $C = \chi T_0^{-\chi}$. The type of function defined by equation (18) is often referred to as a (Kohlrausch-Williams-Watts) stretched exponential function. In probability theory it usually arises in the context of the Weibull distribution, a member of the family of generalized extreme value distributions.

[24] The function $p_0(T)$ determined by our model is a nonanalytic function of T: the origin $T = 0$ is, in fact, an essential singularity of $p_0(T)$. It should be recognized that the limiting behavior of $p_0(T)$ as $T \to 0$ is experimentally inaccessible. As pointed out in section 2.1, although mathematically one can let T become arbitrarily small, in practice, θ, the smallest possible value of T, is nonzero and is dictated by the finite temporal resolution of the measurement. Moreover, the intrinsic discreteness of rain at the scale of individual drops ultimately prevents θ from becoming arbitrarily small. The limit $T \to 0$ is thus an idealized limit that cannot be reached in reality. Note that the value of the multiplicative constant α_0, which, of course, cannot exceed unity, is not determined by this argument.

[25] It is worth noting that when $\chi \neq 1$, $\pi_{01}(T;\theta)$ depends on T, the duration of the prior dry period. Clearly, this is a reflection of the non-Markovian nature of the precipitation process. To see this explicitly, consider a dry period $[0, T]$ of length $T = n\theta$ followed by a wet period θ. (Recall that a Markov process is one in which, given the present state, the future state is independent of the past.) Markov property of the sequence $\{\xi_\theta^{(j)}\}$ implies

$$\begin{aligned}\pi_{01}(n\theta;\theta) &= \Pr\left[\xi_\theta^{(n+1)} = 1 \middle| \xi_\theta^{(n)} = 0, \ldots, \xi_\theta^{(2)} = 0, \xi_\theta^{(1)} = 0\right] \\ &= \Pr\left[\xi_\theta^{(n+1)} = 1 \middle| \xi_\theta^{(n)} = 0\right] \\ &= \pi_{01}(\theta;\theta)\end{aligned}$$

for all n. But in view of equation (16) this is possible if and only if $\chi = 1$. Incidentally, the special case $\alpha_0 = 1$, $\chi = 1$ would naturally arise if the statistics of "events" corresponding to occurrence of rain ($\xi_\theta = 1$) during a small time interval θ with probability $p_1(\theta) = \theta/T_0$ could be modeled in terms of a simple Poisson process [see, e.g., *Ross*, 1996] with the quantity $\lambda = 1/T_0$ identified as the rate parameter. A Poisson process is memoryless since $\gamma(\theta;\theta) = 0$; that is, rainfall indicators in two consecutive time intervals θ are

independent. In the general case, the autocorrelation $\gamma(T;\tau)$ is nonzero, signaling the presence of memory. The asymptotic falloff rate of $\gamma(T;\tau)$ for large τ dictates the nature of the memory [*Beran*, 1994]. A long memory is characterized by the condition that the sum $S_n \equiv \sum_{m=1}^{\infty} \gamma(n\theta; m\theta)$ is divergent. Typically, power law decay in τ indicates long memory since it satisfies this condition. On the other hand, for an exponentially decaying $\gamma(T;\tau)$, as is the case for the familiar Markov and autoregressive moving average processes, the quantity S_n converges to a finite value, and the process is said to have short memory. We have not quantitatively explored the exact falloff property of $\gamma(T;\tau)$ in the present work and therefore cannot draw any firm conclusion regarding how the memory in rainfall should be classified. However, if the leading falloff behavior is a stretched exponential, as seems likely to be the case in view of our results in section 5, then standard convergence tests point toward a convergent S_n and thus a short-range memory.

[26] For area-averaged rain rate r_L estimated from radar data an expression for the probability of zero rain $p_0(L)$ can be derived in a form similar to equation (18) by simply extending the previous arguments to two dimensions. We obtain

$$p_0(L) = \alpha_0' \exp\left[-(L/L_0)^{\chi'}\right], \tag{19}$$

where we have introduced a new set of parameters, α_0', χ', and L_0. The special case $\alpha_0' = 1$, $\chi' = 2$ considered by *Gupta and Waymire* [1993] arises if the zeroes in contiguous grid boxes are independent. It corresponds to the Poisson case $\alpha_0 = 1$, $\chi = 1$ noted earlier in this section.

[27] The zero-rain probabilities $p_0(T)$ and $p_0(L)$ can be estimated from precipitation data. These estimates at different scales allow us to directly test the model prediction. In section 4 we describe an empirical study that examines the scale dependence of $p_0(T)$ and $p_0(L)$ on the temporal and spatial scales from rain gauge and radar observations of rainfall.

4. Data Analysis

[28] The temporal statistics of dry and wet periods were obtained from a 10 year long (1997–2006) rain gauge data set (TRMM standard product 2A-56) collected as part of the TRMM GV program. The data consist of estimates of 1 min rainfall accumulations from a network of 300+ tipping bucket (TB) rain gauges located throughout Florida (Figure 1), some of which are within the Melbourne radar field of view (FOV). The data were divided into 3 month long seasons: spring (March–May), summer (June–August), autumn (September–November), and winter (December–February). The time series from each gauge was subdivided into time intervals of length T ranging from 1 min to 5 days. The probability of encountering a dry period of length T, namely, $p_0(T)$, was estimated as the ratio of the number of intervals with zero rain to the total number of intervals in the time series for all the available gauges during that period. In order to get a reasonably large sample we decided to restrict the upper limit of T to about 5 days. Our 3 month long data series were not single continuous data streams

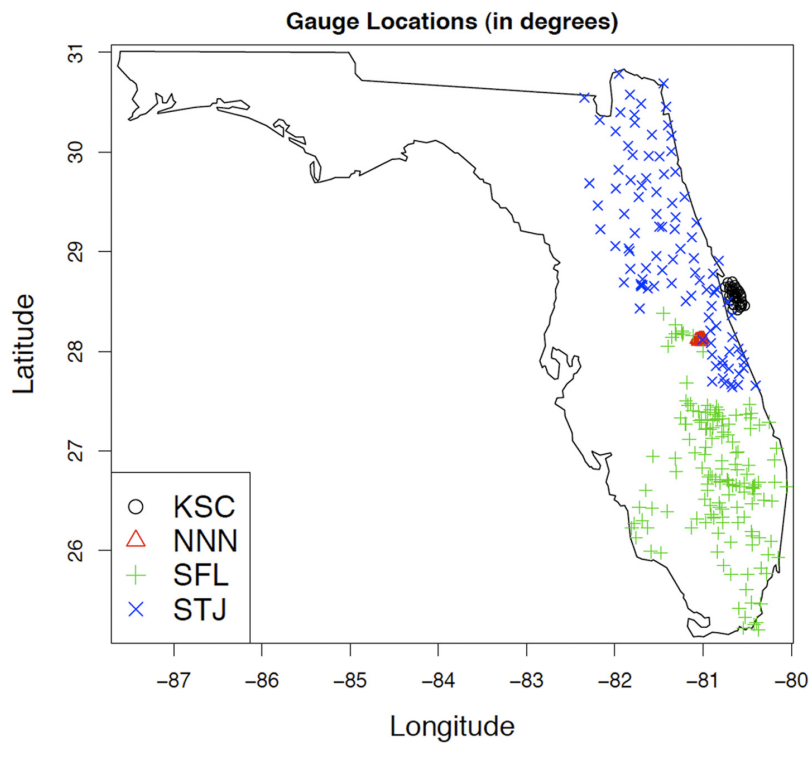

Figure 1. Map showing the locations of the tipping bucket rain gauges in Florida operated as part of the Tropical Rainfall Measuring Mission (TRMM) ground validation program.

but were assembled from monthly data sets with occasional missing data, making the identification of genuine long dry intervals ambiguous. For each season, only those gauges that had data for all 3 months were considered in the calculations. In a straightforward partitioning of the time series into disjoint intervals the number of samples decreases in inverse proportion to the value of T, leading to a progressively larger sampling error in the estimate of $p_0(T)$ with increasing T. In order to address this issue to some extent, for increasing values of T we oversampled the data by considering staggered overlapping intervals. One should note, however, that rainfall is correlated in time, and consequently, the sample count greatly overestimates the "effective" number of samples in the data that can be regarded as independent.

[29] There is a complicating aspect of TB rain gauge measurements that makes the statistics at small T (less than about 7–10 min) somewhat unreliable [*Habib et al.*, 2001]. In reality, the TB gauge data consist of a sequence of tip times when a certain fixed amount of rain accumulation occurs, causing the bucket to tip. The time series of rain rates is obtained by interpolation of the raw tip data to equal time intervals of 1 min using a cubic spline algorithm [*Wang et al.*, 2008]. This interpolation method leads to approximations when isolated single and double tips occur at low rain rates. Hence, the estimates of $p_0(T)$ at small scales can conceivably be affected by the artifacts of this algorithm.

[30] In order to explore the behavior of $p_0(T)$ at small T with greater confidence, we also examined precipitation data from laser disdrometers [*Loffler-Mang and Joss*, 2000]. These devices estimate rain rate from direct optical detection of raindrops by a sensor and are capable of providing accurate rain rates at time scales of order 1 min and higher. We utilize the rainfall time series from a particle size velocity (PARSIVEL) optical disdrometer for three seasons (spring, summer, and autumn), collected during an experimental campaign in 2004 at the NASA Wallops Flight Facility located in Wallops Island, Virginia. The experimental resolution attained in the disdrometer data is 1 min. Even though the laser disdrometer provides estimates of precipitation during snow, the winter season data were not used in the present analysis.

[31] Our spatial statistics analysis is based on two gridded radar precipitation data sets. One of them is a relatively small one, only a little over 3 months long (November 1992 to February 1993), consisting of radar images from two shipborne Doppler radars during the Tropical Ocean Global Atmosphere–Coupled Ocean Atmosphere Response Experiment (TOGA-COARE) [*Short et al.*, 1997] that was studied in an earlier work [*Kundu and Bell*, 2003]. The other one is a large 6 year long data set extending from 1999 to 2004 constructed from the images from a ground-based radar located at the TRMM ground validation site near Melbourne, Florida (80.65°W, 28.11°N), made publicly available by NASA (TRMM standard product 2A-53). The TOGA-COARE data allow us to examine precipitation in a purely oceanic environment. Each image from this data set is a 278 × 278 array representing a square grid assembled from the two partially overlapping circular radar FOVs (labeled TOGA and MIT). A typical Melbourne radar (MELB) image consists of a 151 × 151 array containing rain rate data from a single circular radar FOV. The spatial grid resolution is 2 km in each case. In order to reduce the uncertainties of radar attenuation with distance, the statistics were collected from the 128 × 128 km^2 areas (i.e., 64 × 64 pixel arrays) centered at the radar location subdivided into $L \times L$ squares, with $L = 2, 4, 8, \ldots, 128$ km. Only those boxes in a rain map were used for which at least 95% of the box had valid data. This was done in order to exclude boxes, especially those at the smaller scales of 4, 8, and 16 km located near the center, which occasionally suffered from data dropout because of ground clutter.

[32] The conversion of a rain field to a binary indicator field implicitly assumes a threshold applied to define rainfall zeroes for both temporal and spatial data at the finest resolution scale. The estimates of zero-rain probabilities $p_0(T)$ and $p_0(L)$ are therefore somewhat sensitive to the choice of this threshold. For the purpose of this work we simply accept the thresholds that were applied in defining the TRMM standard products and the other data sets. It is possible to relax them and explore the threshold dependence of the zero-rain probabilities, something we do not pursue here.

5. Results and Discussion

5.1. Temporal Intermittence

[33] Our central finding is that the zero-rain probabilities $p_0(T)$ estimated from rain gauge data are well described by equation (18), namely,

$$p_0(T) = \alpha_0 \exp[-(T/T_0)^{\chi}],$$

over a remarkably large range of T between 1 min and several days, where T_0 is a characteristic time scale and α_0 and χ are dimensionless parameters. Note that for large T, equation (18) predicts that $p_0(T) \to 0$, $p_1(T) \to 1$, as expected. On the other hand, for sufficiently small T, we have the approximation

$$p_1(T) \approx (1 - \alpha_0) + \alpha_0 (T/T_0)^{\chi}. \tag{20}$$

Clearly, the parameter α_0 can be interpreted as the limit $p_0(T \to 0+)$ and must lie in the range $0 < \alpha_0 \leq 1$. If $\alpha_0 = 1$, $-\ln p_0(T)$ has a simple power law dependence on T and so does $p_1(T)$ at small T, $p_1(T) \propto T^{\chi}$, consistent with a simple fractal picture of the support of the rain rate field. When $\alpha_0 < 1$, for timescales $T << (1 - \alpha_0)^{1/\chi} T_0$, $p_1(T)$ approaches a small but nonzero value $1 - \alpha_0$, implying that when resolved at sufficiently small scales, the temporal support of precipitation is not a set of measure zero as it would be for a true fractal set. There is still an approximate power law behavior for the wet fraction $p_1(T)$ in our model but only for timescales T within the intermediate range

$$(1 - \alpha_0)^{1/\chi} T_0 << T << T_0. \tag{21}$$

[34] In Figure 2 we show representative sample plots of $p_1(T)$ versus T for several seasons, illustrating the quality of fit to equation (18). In fitting the data the parameter α_0 was first selected by trial and error for overall best fit, including the small T regime, and the remaining two parameters T_0 and χ were then fitted by the usual equal weighted least

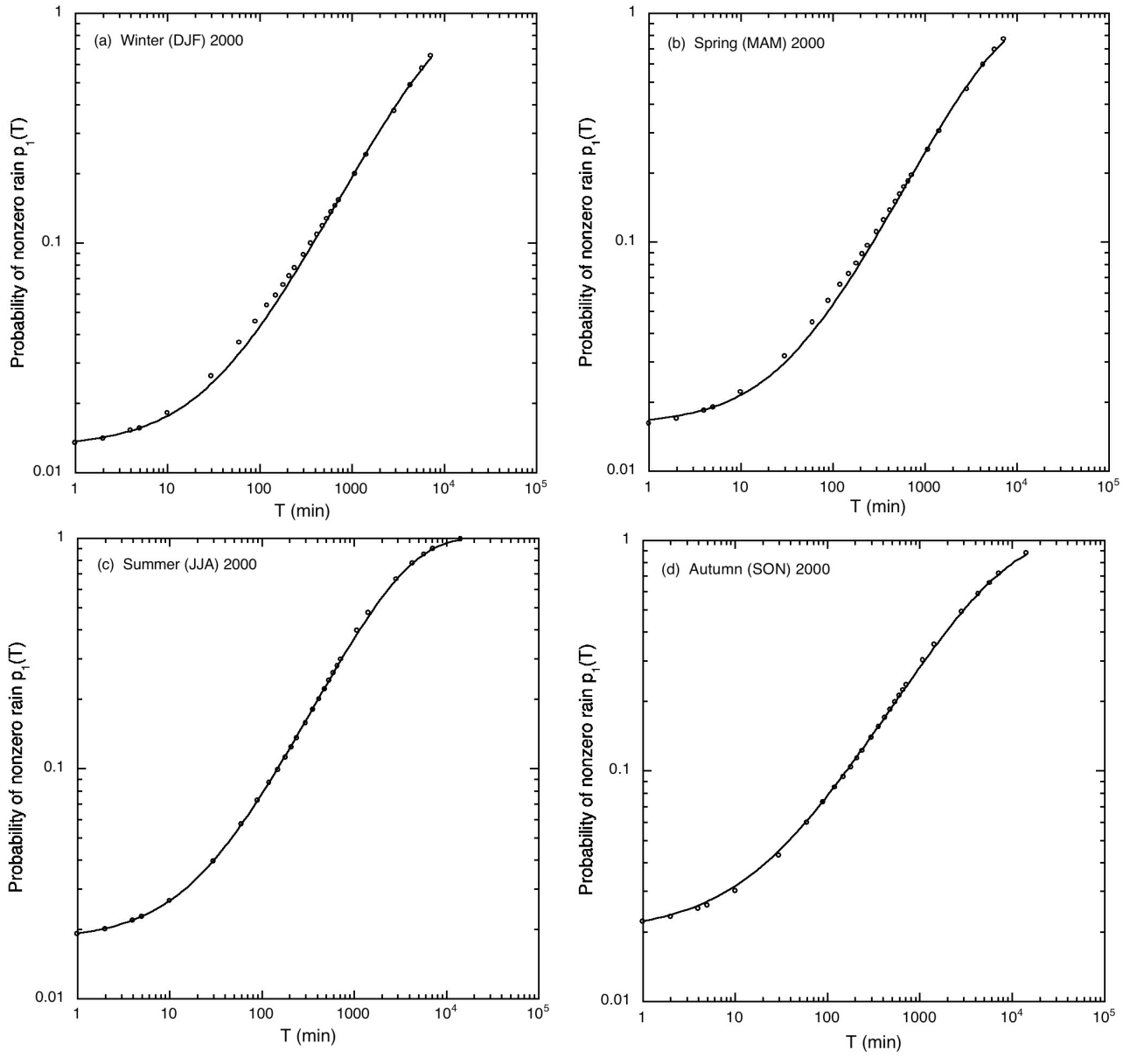

Figure 2. Plots of the probability of nonzero rain $p_1(T)$ at different temporal scales T estimated from the Florida tipping bucket rain gauge data (TRMM standard product 2A-56) superimposed on the curves predicted from the model formula equation (18) for four seasons spanning the period December 1999 to November 2000 with the parameters listed in Table 1: (a) winter (December 1999 to February 2000), (b) spring (March to May 2000), (c) summer (June to August 2000) and (d) autumn (September to November 2000). The fits to the model for all the other years analyzed (not shown) are of comparable quality.

squares method. (Alternatively, one could have estimated the parameters from a least squares fit to the $-\ln p_0(T)$ versus T plot. This would have affected the estimates by making the fit disproportionately sensitive to data in the large T regime. However, these data points are determined from progressively smaller samples and therefore suffer from potentially larger finite sample effect. Fitting the $p_1(T)$ versus T plot effectively assigns less weight to the large T regime.) From Figure 2 it is apparent that there is a departure from power law dependence at small T less than about 30 min. This signals the presence of an "inner scale" for the breakdown of the temporal fractal structure of rain. For the Florida TB gauge data the estimated best fit values of α_0 range between 0.965 and 0.991, with mean and standard error of 0.980 $\pm$ 0.001. The time scale T_0 denotes the value of T for which the probability $p_0(T)$ drops to α_0/e; its value is found to exhibit large seasonal variations, being in the range of 1700–3000 min during summer, 3400–12,200 min during spring, 2600–5500 min during autumn, and 4000–7000 min during winter. The index χ was found to lie in

Table 1. Model Parameters for the Florida TB Gauge Data

Season[a]	α_0	χ	T_0 (min)
MAM 1997	0.975	0.824	3420
JJA 1997	0.978	0.882	1995
SON 1997	0.985	0.768	4639
DJF 1998	0.967	0.834	3992
MAM 1998	0.985	0.742	12,174
JJA 1998	0.982	0.815	3026
SON 1998	0.972	0.728	3711
DJF 1999	0.989	0.863	6229
MAM 1999	0.990	0.797	7130
JJA 1999	0.972	0.817	2169
SON 1999	0.972	0.662	3359
DJF 2000	0.987	0.812	7187
MAM 2000	0.984	0.835	4899
JJA 2000	0.982	0.844	2666
SON 2000	0.980	0.705	5275
DJF 2001	0.991	0.850	7455
MAM 2001	0.985	0.777	9091
JJA 2001	0.972	0.805	2047
SON 2001	0.965	0.682	2578
DJF 2002	0.981	0.810	5027
MAM 2002	0.988	0.806	7073
JJA 2002	0.970	0.809	1774
SON 2002	0.983	0.810	3637
DJF 2003	0.980	0.863	4828
MAM 2003	0.981	0.765	4478
JJA 2003	0.973	0.818	1705
SON 2003	0.984	0.715	4497
DJF 2004	0.975	0.843	5181
MAM 2004	0.990	0.711	11,942
JJA 2004	0.977	0.830	1823
SON 2004	0.971	0.795	3264
DJF 2005	0.983	0.746	7032
MAM 2005	0.975	0.838	5126
JJA 2005	0.974	0.761	2196
SON 2005	0.980	0.656	3862
DJF 2006	0.983	0.821	6230
MAM 2006	0.988	0.844	6276
JJA 2006	0.976	0.795	2321
SON 2006	0.989	0.757	5493

[a]MAM, March–May; JJA, June–August; SON, September–November; DJF, December–February.

Table 2. Model Parameters for the NASA Wallops Flight Facility Data

Season	α_0	χ	T_0 (min)
MAM 2004	0.947	0.712	3274
JJA 2004	0.941	0.714	3232
SON 2004	0.966	0.594	9675

the narrow range of 0.66–0.88, with mean and standard error of 0.79 ± 0.01 for the entire Florida TB gauge data set. The central values for each season are listed in Table 1. The disdrometer data from the NASA Wallops Flight Facility in Virginia also yield estimates that are broadly consistent with these results (Figure 3). For the three seasons during 2004, we obtain estimates of α_0 in the range of 0.941–0.966, χ in the range of 0.59–0.71, and the time scale T_0 in the range of 3200–9700 min. The details are given in Table 2. Intuitively, the time scale T_0 represents the characteristic time between two consecutive storm events. As we have seen in section 3, the fact that $\chi \neq 1$ indicates the presence of memory in the rainfall process. We should also point out that the parameter estimation is dependent on the rainfall detection threshold and is therefore device sensitive.

[35] An approximate power law scaling $p_1(T) \sim T^c$ in rain gauge data has been known for some time [*Hubert and Carbonnel*, 1989; *Olsson et al.*, 1993; *Schmitt et al.*, 1998], with $c = 1 - D$ being identified with the codimension of the fractal temporal support of the rainfall events. In particular, *Schmitt et al.* [1998] report a value $c \approx 0.45$ for a 29 year long time series recorded at Uccle (Belgium) with a siphon pluviograph between 1 January 1967 and 31 December 1995 for time scales T between 10 min (the sampling time of the recording instrument) and about 3.5 days. In comparing the value of c with our χ it should be remembered, however, that they do not quite represent the same quantity. In our model, power law scaling holds only in the

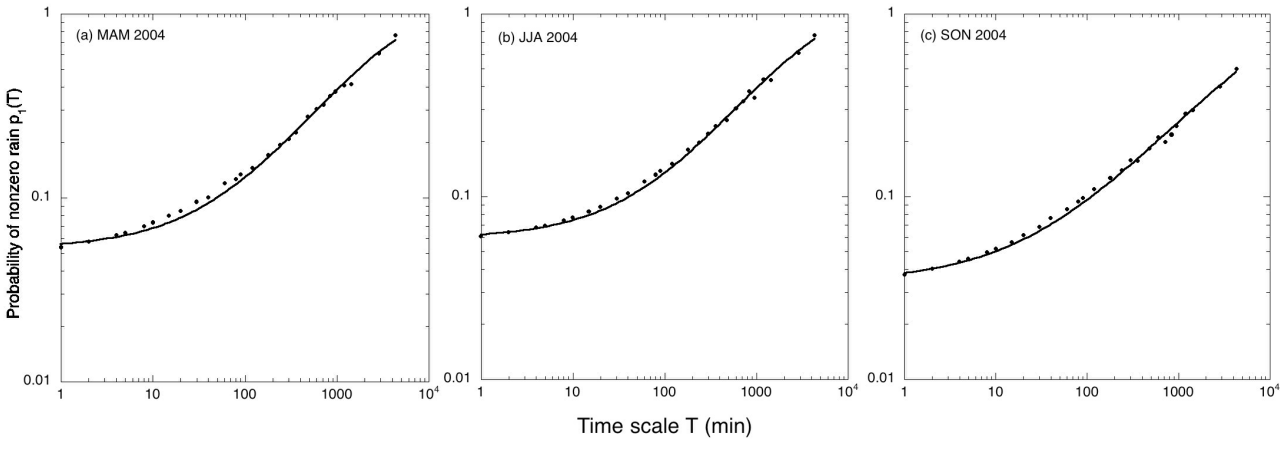

Figure 3. Plots of $p_1(T)$ as a function of time scale T from the precipitation data from an optical disdrometer (PARSIVEL 4) at NASA Wallops Flight Facility (courtesy of Ali Tokay) for three seasons, (a) spring (March–May), (b) summer (June–August), and (c) autumn (September–November) during 2004.

limited range of time scales determined by the inequality (21) and does not extend to arbitrarily small T even in principle. The lower bound of the range (the "inner scale" referred to above) also exhibits some seasonality that arises mainly from the seasonal dependence of T_0. From the parameter values for the Florida TB gauge data given in Table 1 we find that this inner scale lies in the range 9.9–38.0 min, with mean and standard error of 21.3 $\pm$ 1.3 min, for the summer and fall seasons and 18.4–66.8 min, with mean and standard error of 38.9 $\pm$ 3.2 min, for the winter and spring seasons. For the single disdrometer data set from the NASA Wallops Flight Facility whose parameters are listed in Table 2, the lower bound lies in the range 32.6–61.4 min, which is reasonably consistent with the range displayed by the Florida data.

[36] Sometimes a distinction is made between an "internal" and an "external" intermittency [see, e.g., *Bernardara et al.*, 2007]. The former is meant to describe the frequency of dry intervals within a rain event, while the latter is meant to represent the frequency of dry periods between rain events. One introduces a time scale T_{m} separating the two regimes that is reasonably expected to depend upon the climatic conditions in the geographical area considered. *Bernardara et al.* [2007] have discussed a rainfall model based on statistics from two long rainfall time series from Florence (1962–1986, 5 min resolution) and Milan, Italy (1988–2000, 1 h resolution), divided into 1 year long segments. Assuming T_{m} to be equal to 6 h, they describe the CDF of the residence times of dry spells $\Phi(T)$ by a generalized Pareto distribution [*Pickands*, 1975] for $T > T_{\mathrm{m}}$ with a power law falloff of the right tail. (This corresponds to a continuous version of our P_n.) They give an asymptotic falloff of the survival function $(1-\Phi) \sim (T-T_{\mathrm{m}})^{-\beta}(\beta \approx 4)$ for values of T up to about 70 and 35 days for the Florence and Milan data sets, respectively. From equation (14) it follows that in our model the survival function has the form

$$1-\Phi(T) \equiv \int_T^{\infty} \phi(T')\mathrm{d}T' \sim \mathrm{const}T^{\chi-1}\exp[-(T/T_0)^{\chi}] \quad (22)$$

away from the origin. This is quite different from the power law falloff found by *Bernardara et al.* [2007]. It should be noted that the apparent failure of the normalization of the pdf $\phi(T)$ when $\chi < 1$ because of the singularity of $p_0'(T)$ at $T = 0$ does not signal any serious conceptual problem. It merely reflects an artifact caused by the limitation of our continuum approximation in the neighborhood of the origin, as evidenced by the special form of Π_n in equation (10), which is regular at $n = 0$. We have tested the form of the distribution $\Phi(T)$ for the four seasons of 2000 depicted in Figure 2 by fitting the functional dependence on T to the form (22) with the parameters χ and T_0 listed in Table 1. The multiplicative factor was estimated by least squares fit. The observed and model-predicted forms of the survival function $1-\Phi(T)$ are shown in Figure 4. The plots show that for time scales T greater than about 50–100 min, the T dependence expressed by equation (22), rather than the power law form suggested by *Bernardara et al.* [2007], fits the data quite well. Note that in Figure 4 we have extended the range of the residence time beyond the time scale considered in Figure 2. The additional structure ("shoulder") of $p_0'(T)$ noticeable at T around 100 min is less conspicuous in the $p_1(T)$ versus T plots of Figure 2 (since integration tends to smooth out variations) and cannot be described by our model. It remains to be seen whether this is a general feature of intermittence. The disagreement at small T is at least partly attributed to the fact that the singularity in equation (22) at small T is an artifact of the passage to the continuous limit. On the other hand, as the listing of T_0 in Table 1 clearly shows, the rainfall climate in Florida has considerable seasonal dependence. Because of this seasonal variation it is clearly untenable to regard the statistics of a time series longer than about 90 days as stationary, and the validity of the distribution for T larger than about 15–30 days is therefore questionable. The issue of temporal stationarity merits a further in-depth study.

[37] In an interesting paper, *Lavergnat and Golé* [1998] studied temporal intermittence directly in terms of the distribution of the interdrop arrival time. Backed by data from a long time series of high-resolution observation of individual drops by a single optical disdrometer, they proposed to consider the drop time series as a renewal process. They showed that the distribution of the interdrop time interval T can be fit over a range of time scales spanning 9 decades by what they call a complete bi-Pareto model, a six-parameter probability distribution that arises as a mixture of an exponential and a gamma distribution. The survival function in the model has the form

$$1-\Phi(T) = e^{-\lambda_0\tau}\left[(1-q)(1+T/\tau_1)^{-\alpha} + q(1+T/\tau_2)^{-\beta}\right].$$

From a least squares fit to the histogram of drop counts, Lavergnat and Golé estimate the parameters to be $\tau_1 = 0.133$ s, $\tau_2 = 1.22$ s, $\alpha = 1.35$, $\beta = 0.68$, $q = 0.011$, and $1/\lambda_0 = 10^5$ s. They interpret the first component as coming from the temporal gaps within a rainy episode and the second component from the temporal spacing between distinct episodes that accounts for the longer time intervals. This interpretation is obscured by the fact that the parameters τ_1 and τ_2 differ by less than an order of magnitude, so that there is no clear separation between the two terms. Introduction of the short time scales represented by τ_1 and τ_2 serves to regularize the function at the origin. The pdf peaks at about 0.2 s, which is not too different from our estimates of the mean interdrop interval τ_* obtained in section 2. The asymptotic form of the tail of the CDF for large T is to be compared to the form given by equation (22) in our model. While the ability of the model of Lavergnat and Golé to fit the data over a 9 decade time scale is indeed impressive, the large number of adjustable parameters seems a rather steep price to pay. They also indicated that their parameters corresponding to the shorter time scale component may depend on the collection area a. Also, there was a considerable amount of noise in the data (their Figure 3a) at time scales larger than about 10 min. There is no analog to the short time scales τ_1 and τ_2 in our model, which is validated by data over a much smaller range of scales between 1 and 10^4 min. The single constant α_0 subsumes the unknown behavior of the temporal intermittence at the sub-1 min scale. It may also plausibly depend on the collection area a.

[38] As another caveat, let us add that we have also implicitly assumed spatial homogeneity of rainfall statistics

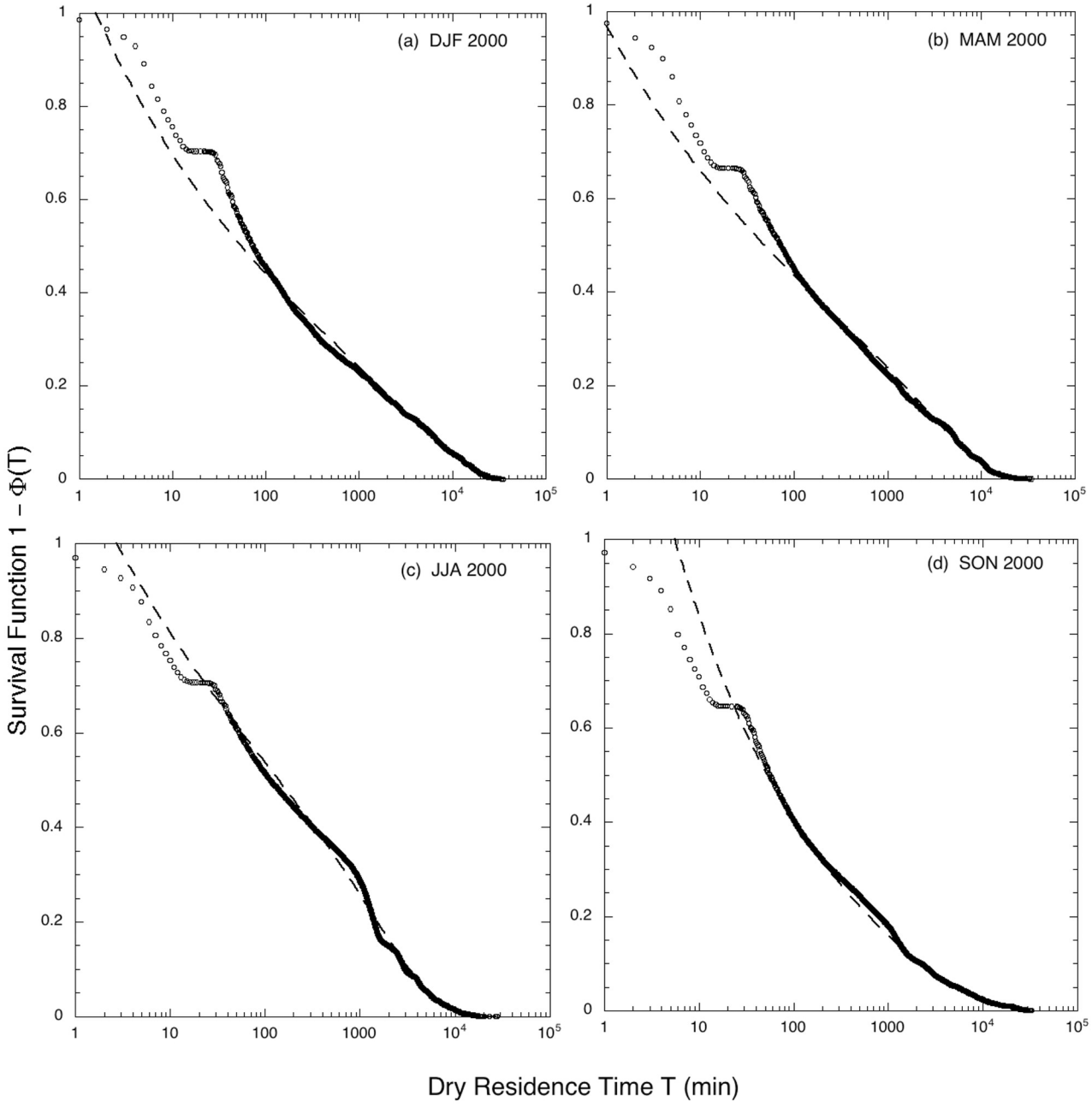

Figure 4. Plots of the survival function corresponding to the distribution of the dry residence time T estimated from the Florida tipping bucket rain gauge data (TRMM standard product 2A-56) superimposed on the curves predicted from the model formula equation (22) with the parameters χ and T_0 taken from Table 1: (a) winter (December 1999 to February 2000), (b) spring (March–May 2000), (c) summer (June–August 2000), and (d) autumn (September–November 2000).

over all of Florida so that all the gauges could be pooled together. So far, we are unaware of any evidence of major regional dependence of the rain statistics in Florida, but this should be investigated further. For our present purpose, it is sufficient to test the spatial homogeneity of the rainfall indicator field. We here present evidence of spatial homogeneity for rain at the smallest accessible time scale $\theta = 1$ min. Let p_{1j} denote the probability of nonzero rain for a particular gauge j ($j = 1, 2, \ldots, m$) at the temporal scale θ. The quantity $p_1(\theta)$ is the "true" value of the probability as estimated from all the gauges. We undersample the time series $\xi_\theta(t)$ by choosing the samples to be sufficiently far apart in time so as to minimize the influence of the time correlation. This allows us to effectively treat them as independent Bernoulli trials. It appears that for temporal resolution $\theta = 1$ min, choosing samples with time separation $\tau = 60$ min is adequate. The test of spatial homogeneity can now be formulated as a multiple comparison of a set of binomial

proportions since independent drawings of the rain indicator obey the binomial distribution. We have carried out statistical tests of the null hypotheses $H_{0j} : p_{1j} = p_1(\theta)$ ($j = 1, 2, \ldots, m$) using standard methods of multiple comparison (see Appendix B for details) for the four seasons in 2000 that are depicted in Figure 2. If a sufficiently large number of hypotheses H_{0j} is rejected out of a total of m, it is deemed as evidence against spatial homogeneity. We used two distinct methods of multiple testing, the Holm-Bonferroni (HB) test [*Holm*, 1979] based on the family-wise error rate (FWER) $\bar{\alpha}$ and the Benjamini-Hochberg (BH) test [*Benjamini and Hochberg*, 1995] based on the false discovery rate (FDR) δ. The first aims to control the FWER, i.e., the overall Type I error rate (false positive) for the gauge system as a whole. A type 1 error occurs when H_{0j} is rejected even when it is true. The BH test, on the other hand, fixes the acceptable level of FDR, which is the average proportion of rejected hypotheses H_{0j} that are actually true. Our main reason for choosing these particular tests among the many that are available is that they apply to the case when the tests are mutually dependent. Because of (generally positive) spatial correlation among the gauges, the null hypotheses H_{0j} are also expected to be dependent.

[39] The results presented in Table 3 show the total number of gauges tested along with the numbers rejected in each case, with the FWER $\bar{\alpha}$ and the FDR δ each set at 0.05. We also carried out the tests by resampling the series at smaller time separations τ ranging between 20 and 60 min and found that the number of gauges rejected by either method goes down steadily. We attribute this to the decreasing effect of time correlation until the samples presumably become nearly independent. We did not extend the range of τ beyond 60 min as this reduces the number of samples to the point that the normal approximation to the binomial distribution becomes inapplicable for more and more gauges. We notice that, in general, there are fewer rejections arising from the FWER method than the FDR method. We attribute this to the greater statistical power of the FDR-based test compared to the FWER-based test. The statistical power of a test is defined as $(1 - \beta)$ where β is the type 2 error rate. A type 2 error (false negative) occurs when a null hypothesis H_{0j} is accepted even when it is actually false, a possibility that we consider to be more insidious in the context of our problem. So it is possible that some of the H_{0j} that are accepted by the FWER criterion are actually false and are correctly rejected by the FDR criterion. Fixing the value of δ at 0.05 means that, on average, only 5% of the null hypotheses rejected by the FDR-based test (about 9–10 gauges for our data sets) are actually true. Thus, we prefer the FDR method to the FWER method. An inspection of Table 3 shows the assumption of spatial homogeneity to be a reasonable one. Moreover, our approach provides a diagnostic test of spatial homogeneity separately for each individual gauge. Future work can proceed with isolating the suspected gauges and examining the data from them for the probable cause of inhomogeneity.

Table 3. Results of Spatial Homogeneity Testing

Season	Total Gauges m	Null Hypotheses Rejected: Holm-Bonferroni Test	Null Hypotheses Rejected: Benjamini-Hochberg Test
DJF 2000	189	17	35
MAM 2000	180	14	43
JJA 2000	195	10	19
SON 2000	184	16	27

Table 4. Model Parameters for the Melbourne Radar Data

Season	α'_0	χ'	L_0 (km)
MAM 1999	0.989	1.026	418.2
JJA 1999	0.981	1.111	144.0
SON 1999	0.981	0.948	149.9
DJF 2000	0.982	1.073	294.4
MAM 2000	0.991	1.020	358.3
JJA 2000	0.988	1.138	167.5
SON 2000	0.986	1.040	219.8
DJF 2001	0.992	1.126	359.5
MAM 2001	0.984	0.906	397.4
JJA 2001	0.980	1.027	157.0
SON 2001	0.975	1.067	130.3
DJF 2002	0.976	0.965	306.7
MAM 2002	0.991	1.236	243.3
JJA 2002	0.976	0.995	135.1
SON 2002	0.988	1.118	176.6
DJF 2003	0.976	0.845	523.5
MAM 2003	0.984	1.049	254.1
JJA 2003	0.978	1.027	153.1
SON 2003	0.985	1.064	163.2
DJF 2004	0.972	0.928	406.9
MAM 2004	0.994	1.039	358.5
JJA 2004	0.984	1.040	173.2
SON 2004	0.969	1.039	157.0

5.2. Spatial Intermittence

[40] A similar analysis of the gridded radar data shows that formula (19) represents the scale dependence of $p_0(L)$ fairly well over the range of spatial scales studied, namely, $L = 2$–128 km. As in the temporal case, for the Melbourne radar data the parameters α'_0, χ', and L_0 are estimated for 3 month long seasons during the 6 year period 1999–2004. The results are listed in Table 4. The estimated best fit values of α'_0 range between 0.969 and 0.994, with mean and standard error 0.983 ± 0.001. The exponent χ' lies in the range 0.85–1.24, with mean and standard error 1.04 ± 0.02. We notice that χ' is substantially smaller than the critical value 2, indicating "persistence of drought" in the spatial dependence as well. Sample plots of $p_1(L)$ versus L illustrating the quality of the fit are exhibited in Figure 5. Unlike in the temporal case, the asymptotic behavior at small spatial scales implied by equation (19) is barely discernable in Figure 5. However, the recently available higher-resolution radar data may allow a more decisive test of this feature in the near future. Again, there seems to be some seasonal dependence in the length parameter L_0 for the Melbourne radar data: the values obtained for the winter and spring seasons are generally larger than for the summer and autumn seasons. Plots for the TOGA-COARE data are shown in Figure 6. Despite the smaller size of the data set the quality of the fit to equation (19) is excellent. The parameters α'_0, χ', and L_0 for the six data subsets are listed in Table 5. The values of the length-scale parameter L_0 are much smaller than for the Melbourne radar data, indicating widespread rain in the TOGA-COARE area.

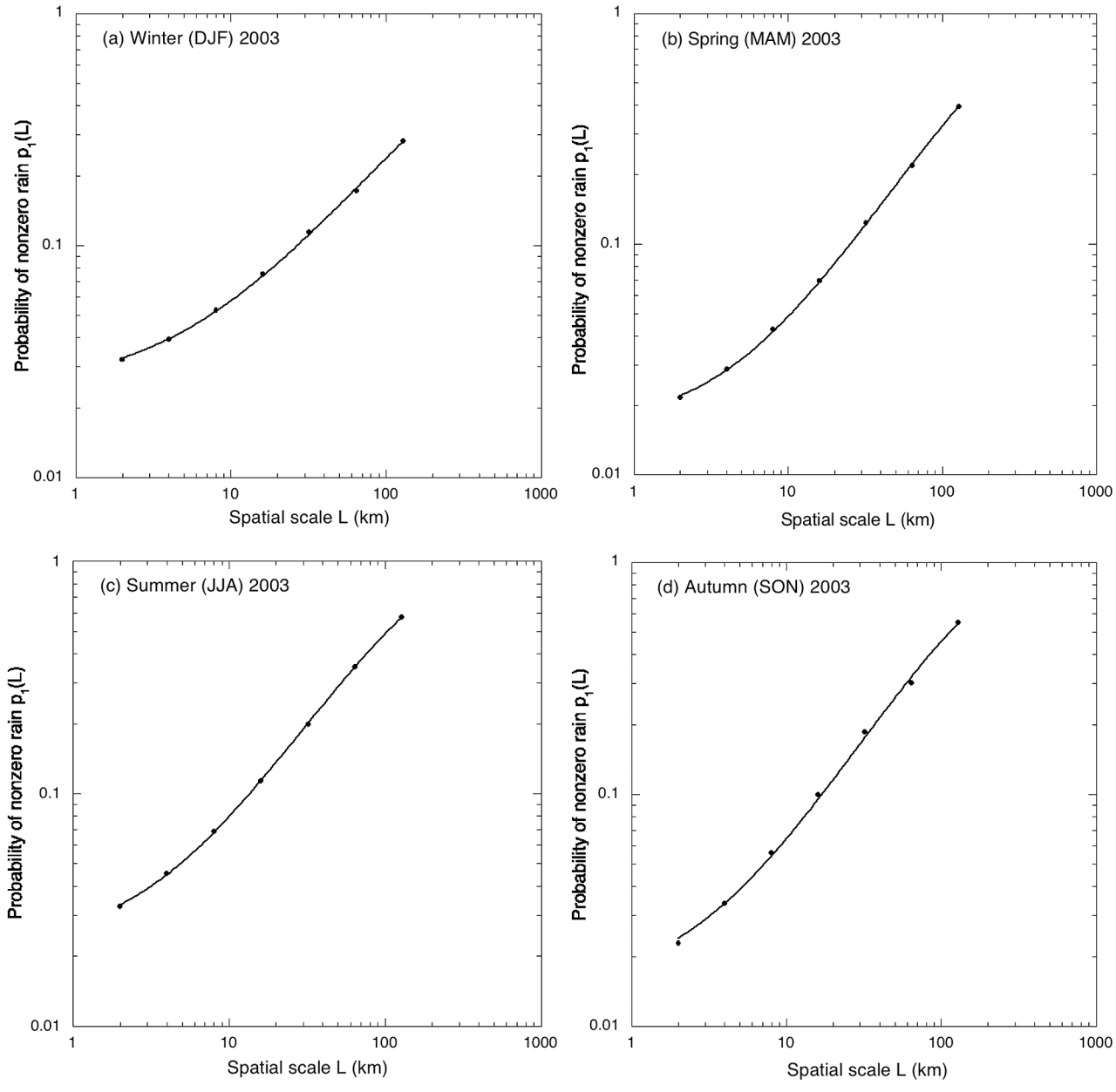

Figure 5. Plots of $p_1(L)$ for the Melbourne radar data (TRMM standard product 2A-53) as a function of spatial averaging scale L for four seasons spanning the period December 2002 to November 2003 with the parameters listed in Table 4: (a) winter (December 2002 to February 2003), (b) spring (March–May 2003), (c) summer (June–August 2003) and (d) autumn (September–November 2003). The fits to the model for all the other years analyzed (not shown) are of comparable quality.

6. Concluding Remarks

[41] In this paper we have explored the phenomenon of intermittence in space-time rainfall. From empirical analyses of rain gauge and radar data we have obtained simple formulas for the probabilities of occurrence of zero rain, $p_0(T)$ and $p_0(L)$, in time- and area-averaged precipitation data. We have also presented a probabilistic model of rainfall intermittence that leads to a simple explanation of the empirical formula suggested by data. The non-Markovian nature of the model suggests the presence of memory in the stochastic process underlying precipitation. The phenomenon of persistence of drought implied in the observed stretched exponential behavior of the zero-rain probability begs for a physical explanation. It will be very interesting to see if it can be understood in terms of the underlying physical mechanism of the formation and evolution of the rainfall process.

[42] In section 2.1 we raised the question whether intermittence is an aspect of the temporal gaps distinguishable from inherent temporal discreteness. Our results show that

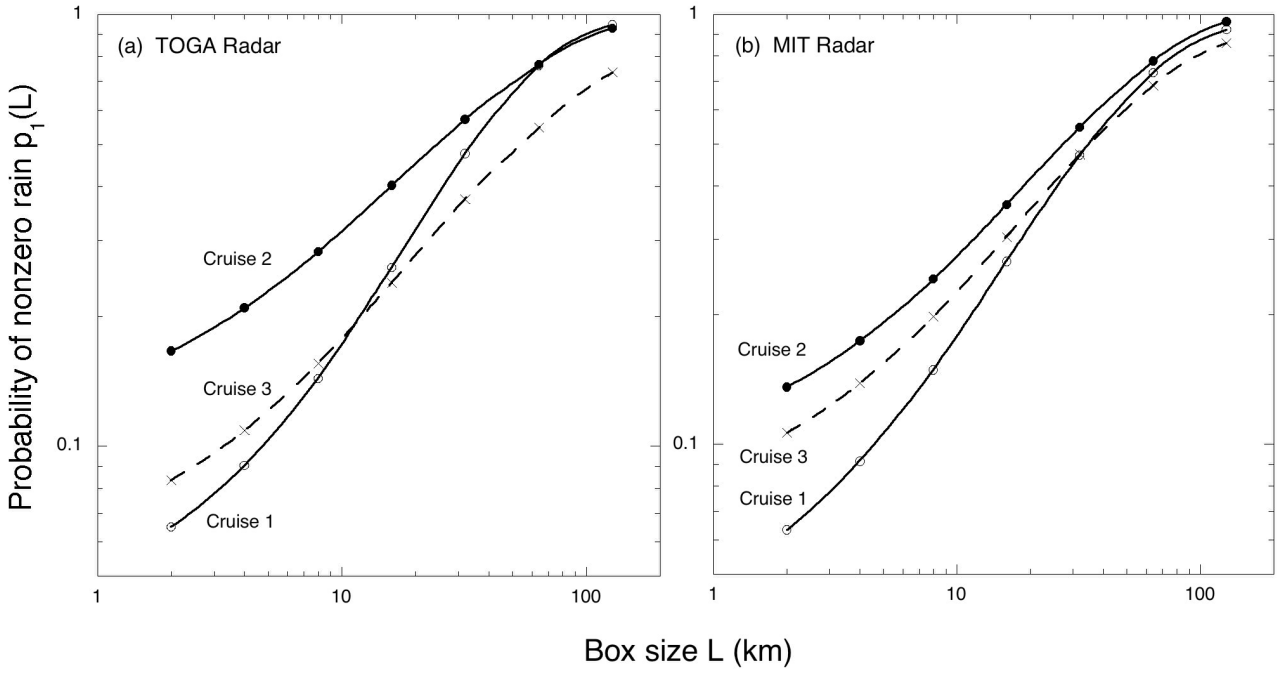

Figure 6. Plots of $p_1(L)$ versus L for the Tropical Ocean Global Atmosphere–Coupled Ocean Atmosphere Response Experiment (TOGA-COARE) data for the parameters given in Table 5: (a) TOGA radar and (b) MIT radar.

over a large range of time scales the answer is in the negative, except possibly for the additional structure seen in Figure 4. This feature, if it is a general one, may signify a demarcation between two distinct regimes. The only time scale present in our formula is the quantity T_0, which can be physically interpreted as the time scale of the dry period between two separate storm events. Of course, our work is unable to reveal the presence of very short time scales, if any, unlike the investigations of Lavergnat and Golé.

[43] The zero-rain probability is only one particular aspect of the mixed distribution governing the space-time distribution of rain that exhibits dependence on the averaging scales. Moreover, in order to fully quantify the spatio-temporal structure of rain it is not adequate to just specify the marginal distribution. Rather, one needs the joint probability distribution of rain at two different space-time points, which is also expected to exhibit nontrivial scale dependence. The intermittence is then more fully characterized by the joint probability of zero rain, i.e., quantities such as $p_{00}(T;\ \tau)$ introduced in section 2. It will be interesting to see whether the empirical behavior of such quantities can be explained in terms of simple model stochastic processes.

Table 5. Model Parameters for the TOGA-COARE Data Set

Data Set	α_0'	χ'	L_0 (km)
TOGA cruise 1	0.959	1.181	49.2
MIT cruise 1	0.970	1.070	51.6
TOGA cruise 2	0.887	0.889	45.9
MIT cruise 2	0.893	1.050	46.0
TOGA cruise 3	0.964	0.809	92.4
MIT cruise 3	0.944	0.882	59.7

Appendix A: Formula for the Residence Time Π_n

[44] The probability of occurrence of a dry spell of duration $T = n\theta$ is given by the conditional probability

$$\Pi_n = \begin{cases} \Pr\left[\xi^{(n+1)} = 1, \xi^{(n)} = 0, \ldots, \xi^{(1)} = 0 \middle| \xi^{(0)} = 1\right]; & n > 0 \\ \Pr\left[\xi^{(1)} = 1 \middle| \xi^{(0)} = 1\right]; & n = 0 \end{cases}, \quad \text{(A1)}$$

where $\xi^{(i)} = \xi_\theta{}^{(i)}$ is the indicator at the finest available temporal scale θ in the data. It can be easily expressed in terms of $p_0(T)$. Define the unconditional probability

$$\Omega_n = \begin{cases} \Pr\left[\xi^{(n+1)} = 1, \xi^{(n)} = 0, \ldots, \xi^{(1)} = 0, \xi^{(0)} = 1\right]; & n > 0 \\ \Pr\left[\xi^{(1)} = 1, \xi^{(0)} = 1\right]; & n = 0 \end{cases} \quad \text{(A2)}$$

satisfying $\Omega_n = p_1(\theta)\Pi_n$. We now establish the following result:

Theorem. Ω_n is given by the formula

$$\Omega_n = p_0(n\theta) - 2p_0[(n+1)\theta] + p_0[(n+2)\theta] \quad \text{(A3)}$$

for integers $n > 0$ and

$$\Omega_0 = 1 - 2p_0(\theta) + p_0(2\theta). \quad \text{(A4)}$$

The result is a simple consequence of the additivity of probabilities of mutually exclusive outcomes of a trial. For $n > 0$, we have identically

$$\begin{aligned}&\Pr\left[\xi^{(n)}=0,\ldots,\xi^{(1)}=0,\xi^{(0)}=1\right]\\&=\Pr\left[\xi^{(n+1)}=0,\ldots,\xi^{(1)}=0,\xi^{(0)}=1\right]\\&\quad+\Pr\left[\xi^{(n+1)}=1,\ldots,\xi^{(1)}=0,\xi^{(0)}=1\right].\end{aligned}$$

The second term is the desired quantity Ω_n. The quantity on the left-hand side can be evaluated similarly:

$$\begin{aligned}&\Pr\left[\xi^{(n)}=0,\ldots,\xi^{(1)}=0,\xi^{(0)}=1\right]\\&=\Pr\left[\xi^{(n)}=0,\ldots,\xi^{(1)}=0\right]\\&\quad-\Pr\left[\xi^{(n)}=0,\ldots,\xi^{(1)}=0,\xi^{(0)}=0\right]\\&=p_0(n\theta)-p_0[(n+1)\theta].\end{aligned}$$

The first term on the right-hand side is obtained by replacing n with $n + 1$, i.e.,

$$\begin{aligned}&\Pr\left[\xi^{(n+1)}=0,\ldots,\xi^{(1)}=0,\xi^{(0)}=1\right]\\&=\Pr\left[\xi^{(n+1)}=0,\ldots,\xi^{(1)}=0\right]\\&\quad-\Pr\left[\xi^{(n+1)}=0,\ldots,\xi^{(1)}=0,\xi^{(0)}=0\right]\\&=p_0[(n+1)\theta]-p_0[(n+2)\theta].\end{aligned}$$

Combining the three relations above, we get equation (A3).

[45] Next, we note that

$$\Omega_0 = \Pr\left[\xi^{(1)} = 1, \xi^{(0)} = 1\right] = p_{11}(\theta;\theta).$$

Equation (A4) then follows by a straightforward evaluation taking into account equations (3) and (4):

$$\begin{aligned}p_{11} &= p_1 - p_{10}\\&= 1 - p_0 - p_{01}\\&= (1-p_0) - (p_0 - p_{00}).\end{aligned}$$

It is easily verified that the quantities Π_n satisfy the normalization

$$\sum_{n=0}^{\infty} \Pi_n = 1,$$

as they should.

Appendix B: Testing Spatial Homogeneity of $p_1(T)$

[46] In this appendix we outline several methods of testing spatial homogeneity of the rain indicator field. We consider a sample of wet intervals of a specified length T from the precipitation time series observed by a rain gauge. Members of the sample are sufficiently separated in time from one another so as to be considered independent. Let p denote the probability of nonzero rain for a particular gauge and $p^* \equiv p_1(T)$ denote the "true" probability estimated from the entire gauge network. One way to frame the test of spatial homogeneity is to examine the difference $p - p^*$ for each gauge throughout the network and see whether some deviate significantly from zero. Formulated in this manner, statistical testing of spatial homogeneity for a gauge data set as a whole leads to multiple hypothesis testing, one for each gauge.

[47] First, let us consider each gauge separately. Since the samples from a gauge are treated as independent, the probability of recording M rainy intervals out of a sample of N is determined by the binomial distribution Bin(N; p). The proportion $\hat{p} = M/N$ provides a sample estimate of the probability p for each particular gauge. One can now seek a suitable confidence interval (CI) for p at a specified nominal level $(1 - \alpha)$. Each realization of the data for a particular gauge determines a confidence interval. The significance of the nominal level is that if we had a large number of such realizations and we computed the confidence intervals for each of them, the probability p for that gauge would fall in approximately $100(1 - \alpha)\%$ of these intervals (i.e., $\Pr[p \in CI] = 1 - \alpha$). If the true probability p^* falls within the calculated confidence interval, we will accept the null hypothesis $H_0 : p = p^*$ as opposed to the alternative hypothesis $H_a : p \neq p^*$ based upon the sample at a level of significance α. If H_0 is true, the gauge is "average" (analogous to a coin being "fair" with $p = 1/2$). The quantity

$$\alpha = \Pr[\text{reject } H_0 | H_0 \text{ is true}] \tag{B1}$$

denotes the false positive, or type 1 error rate for a single gauge.

[48] The simplest approach for obtaining a CI for a binomial proportion relies on the approximation of the binomial distribution Bin(N;p) by a normal distribution $\mathbf{N}(p, \sqrt{p(1-p)/N})$. Assuming asymptotic normality of the sample proportion yields the $100(1 - \alpha)\%$ CI for p, the standard (Wald) interval $\mathrm{CI_S}$:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}},$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard normal CDF. The standard CI represents the region of acceptance for p^* with a p value (observed significance level) exceeding α in the test of H_0 against H_a employing the approximately normal test statistic $Z = (\hat{p} - p)/\sqrt{\hat{p}(1-\hat{p})/N}$.

[49] It is well known that the normal approximation is a poor one, especially when p is close to 0 or 1, which is the case of interest for us. Many improvements of standard CI estimates have been proposed. See, for example, the review article by *Brown et al.* [2001], which elaborates the hazards of the normal approximation and describes a number of superior alternatives. One of their recommended alternatives is the Wilson score CI, first introduced by *Wilson* [1927]. It is determined as the range of the values of p^* satisfying the inequality

$$\frac{|\hat{p} - p^*|}{\sqrt{p^*(1-p^*)/N}} \leq z_{\alpha/2} \tag{B2}$$

and is given by the expression

$$\frac{\hat{p}+z_{\alpha/2}^2/2N}{1+z_{\alpha/2}^2/N} \pm \frac{z_{\alpha/2}}{1+z_{\alpha/2}^2/N}\left[\frac{\hat{p}(1-\hat{p})}{N}+\frac{z_{\alpha/2}^2}{4N^2}\right]^{1/2}. \quad \text{(B3)}$$

This Wilson CI, denoted by $\mathrm{CI_W}$, results as the region of acceptance from inverting the normal approximation to the family of equal tail tests of H_0 against H_a using the test statistic

$$Z(p) = (\hat{p}-p)/\sqrt{p(1-p)/N}. \quad \text{(B4)}$$

One rejects H_0 at a level of significance α if and only if p^* lies outside $\mathrm{CI_W}$.

[50] Next, we consider the problem of multiple-hypothesis testing. Suppose that we have a data set consisting of m gauges located over a certain area. Spatial homogeneity would imply that all the gauges are "average" in the sense defined here. Testing spatial homogeneity in the present context therefore entails simultaneously testing H_0 for all the gauges (analogous to testing multiple fair coins). However, an attempt to test this is expected to occasionally lead to false positives, in which one would wrongly reject H_0 even when it is true. These will tend to obscure any signal of genuine spatial inhomogeneity. In order to detect a possible signal, one needs a control on the overall type 1 error rate. There are several approaches to multiple comparisons in the literature that achieve this control. In the following we consider two of them that apply to our present problem.

[51] A traditional method consists of controlling the so-called family-wise error rate. If $\mathbf{V}$ denotes the number of type 1 errors out of m of tests of H_0, then the data set–wide significance level, the FWER for the entire group of gauges, is defined as the probability that at least one type 1 error occurs during the multiple comparisons, i.e.,

$$\bar{\alpha} \equiv \Pr[\mathbf{V} > 0] = 1 - \Pr[\mathbf{V} = 0]. \quad \text{(B5)}$$

It is the value of $\bar{\alpha}$ rather than α that is central to deciding how strong the assumption of spatial homogeneity is on the basis of the data. If all the tests of H_0 were independent, then $\bar{\alpha} = 1-(1-\alpha)^m$. If the value of $\bar{\alpha}$ is set to a preassigned value, the significance level for an individual test α is obtained from the solution $\alpha = 1-(1-\bar{\alpha})^{1/m}$(the Dunn-Šidák correction). In practice, this is not the case because of the spatial correlation among the gauges. However, for dependent tests one still has the inequality $\bar{\alpha} \le m\alpha$. Thus, for a prescribed $\bar{\alpha}$, say 0.05, if each individual test of H_0 is carried out at significance level $\alpha = \bar{\alpha}/m$, known as the Bonferroni correction, it is guaranteed that the total type 1 error rate will not exceed $\bar{\alpha}$. Unfortunately, the simple Bonferroni correction is overly conservative since as the individual gauges are tested, the number of remaining gauges decreases, and the effect of this should be taken into account. A number of improved sequential Bonferroni-type correction methods exist. We adopt the Holm-Bonferroni [*Holm*, 1979] method, which applies even when the tests are dependent. Instead of testing the hypotheses using the same confidence interval, say the Wilson CI, one now computes the individual p values for each hypothesis and compares them with an adjusted threshold. Denote by Z_j^* the test statistic $Z(p)$ evaluated at $p = p^*$ for the jth gauge ($j = 1,2, \ldots, m$) from equation (B4). Let H_{0j} denote the null hypothesis for the jth gauge and $P_j = \Pr[|Z(p)| \ge Z_j^*]$ denote the corresponding p value. First, one orders the p values from the smallest to the largest, i.e., $P_{(1)} \le P_{(2)} \le \ldots \le P_{(m)}$. The Holm-Bonferroni method consists of the following sequential steps.

1. If $P_{(1)} \ge \bar{\alpha}/m$, accept all m hypotheses H_{0j}; otherwise, reject H_{01}.
2. If $P_{(2)} \ge \bar{\alpha}/(m-1)$, accept H_{0j} ($j \ge 2$); otherwise, reject H_{02}.
3. Continue the process until one reaches the first j such that $P_{(j)} \ge \bar{\alpha}/(m-j+1)$ and accept all the remaining hypotheses.

[52] The multiple testing method based on fixing the FWER has a number of conceptual disadvantages when applied to the present problem. These stem from the facts that some of the H_{0j} may actually be false and, moreover, how many are false is unknown. Instead of a family-wise false positive rate, we can also consider a multiple comparison based on the so-called false discovery rate first introduced by *Benjamini and Hochberg* [1995]. One defines the FDR as the probability that a rejected null hypothesis is actually true, i.e.,

$$\delta = \Pr[H_0 \text{ is true} | \text{reject } H_0]. \quad \text{(B6)}$$

In the context of our problem, controlling δ is more important since $\mathbf{R}$, the total number of rejections of H_0, is a random variable whose value is known from the result of the test carried out on the data set. The quantity δ can be identified with the expectation value $E[\mathbf{V}/\mathbf{R}]$. Other closely related quantities include the positive FDR $E[\mathbf{V}/\mathbf{R}|\mathbf{R} > 0]$ [*Storey*, 2002], the proportion of false positives $E[\mathbf{V}]/E[\mathbf{R}]$ [*Fernando et al.*, 2004], and the posterior error rate $\Pr[\mathbf{V} = 1 \mid \mathbf{R} = m = 1]$ [*Morton*, 1955]. For a preselected value of the FDR δ, the Benjamini-Hochberg procedure is to reject all hypotheses H_{0j} whose p values are bounded by

$$P_{(j)} \le j\delta/m. \quad \text{(B7)}$$

[53] *Benjamini and Yekutieli* [2001] have proved that this bound also holds for tests with positive dependency, which we expect to be largely the case in view of the generally positive spatial correlation among the gauges. (They have also shown that a somewhat weaker bound holds for negative dependency.) If μ denotes the number of rejected null hypotheses out of a total of m, the expected number of null hypotheses that are correctly rejected is, on average, $(1-\delta)\mu$. The latter should be small compared to m if spatial homogeneity is to be deemed acceptable. In case a large number of gauges are rejected by this test, the evidence for spatial homogeneity is weakened.

[54] In general, a statistical test also has type 2 error, or false negatives, where H_0 is accepted even when it is false. Clearly, this is an undesirable and yet unavoidable aspect of hypothesis testing. The FDR-based test has the advantage that, in general, it has a greater statistical power $(1-\beta)$ compared to the FWER-based test, where β is the type 2 error rate.

[55] **Acknowledgments.** We are grateful to Ali Tokay of JCET, University of Maryland Baltimore County (UMBC), and GSFC for providing the disdrometer data and also for several stimulating conversations regarding microscopic behavior of rainfall. We also thank David Wolff of Science Systems and Applications, Inc., and GSFC for helping us with the TRMM 2A-56 radar data from the Melbourne radar and Anindya Roy of Department of Mathematics and Statistics, UMBC, for reading the first draft of the manuscript and for useful comments. Insightful comments from three anonymous reviewers helped correct a number of errors and improve the manuscript. This research was supported by the NASA Precipitation Measurement Missions (PMM) program.

References

Bell, T. L. (1987), A space-time stochastic model of rainfall for satellite remote sensing studies, *J. Geophys. Res.*, *92*, 9631–9643, doi:10.1029/JD092iD08p09631.

Benjamini, Y., and T. Hochberg (1995), Controlling the false discovery rate: A practical and powerful approach to multiple testing, *J. R. Stat. Soc., Ser. B*, *57*(1), 289–300.

Benjamini, Y., and D. Yekutieli (2001), The control of the false discovery rate in multiple testing under dependency, *Ann. Stat.*, *29*, 1165–1188.

Beran, J. (1994), *Statistics of Long Memory Processes*, Chapman and Hall, New York.

Bernardara, P., C. De Michele, and R. Rosso (2007), A simple model of rain in time: An alternating renewal process of wet and dry states with a fractional (non-Gaussian) rain intensity, *Atmos. Res.*, *84*, 291–301.

Brown, L. D., T. T. Cai, and A. DasGupta (2001), Interval estimation for binomial proportion, *Stat. Sci.*, *16*(2), 101–133, doi:10.1214/ss/1009213286.

Buishand, T. A. (1977), Stochastic modeling of daily rainfall sequences, *Meded. Landbouwhogesch. Wageningen*, 77–3.

de Montera, L., L. Barthès, C. Mallet, and P. Golé (2009), The effect of rain–no rain intermittency on the estimation of the universal multifractals model parameters, *J. Hydrometeorol.*, *10*(2), 493–506.

Fernando, R. L., D. Nettleton, B. R. Southey, J. C. M. Dekkers, M. F. Rothschild, and M. Soller (2004), Controlling the proportion of false positives in multiple dependent tests, *Genetics*, *166*, 611–619.

Fisher, R. A., and E. A. Cornish (1960), The percentile points of distributions having known cumulants, *Technometrics*, *2*, 209–225.

Gupta, V. K., and E. C. Waymire (1990), Multiscaling properties of spatial rainfall and river flow distributions, *J. Geophys. Res.*, *95*, 1999–2009, doi:10.1029/JD095iD03p01999.

Gupta, V. K., and E. C. Waymire (1993), A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteorol.*, *32*, 251–267.

Habib, E., W. F. Krajewski, and A. Kruger (2001), Sampling errors of tipping-bucket rain gauge measurements, *J. Hydrol. Eng.*, *6*(2), 159–166.

Holm, S. (1979), A simple sequential rejective multiple test procedure, *Scand. J. Stat.*, *6*, 65–70.

Hubert, P., and J.-P. Carbonnel (1989), Dimensions fractals de l'occurrence de plui en climat soudano-sahélien, *Hydrol. Continent.*, *4*, 3–10.

Kedem, B., and L. S. Chiu (1987), Are rain rate processes self-similar?, *Water Resour. Res.*, *23*, 1816–1818, doi:10.1029/WR023i010p01816.

Kumar, P., and E. Foufoula-Georgiou (1994), Characterizing multiscale variability of zero intermittency in spatial rainfall, *J. Appl. Meteorol.*, *33*, 1516–1525.

Kundu, P. K., and T. L. Bell (2003), A stochastic model of space-time variability of mesoscale rainfall: Statistics of spatial averages, *Water Resour. Res.*, *39*(12), 1328, doi:10.1029/2002WR001802.

Kundu, P. K., and R. K. Siddani (2007), A new class of probability distributions for describing the spatial statistics of area-averaged rainfall, *J. Geophys. Res.*, *112*, D18113, doi:10.1029/2006JD008042.

Lana, X., M. D. Martínez, C. Serra, and A. Burgueño (2010), Complex behaviour and predictability of the European dry spell regimes, *Nonlinear Processes Geophys.*, *17*, 499–512.

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

Loffler-Mang, M., and J. Joss (2000), An optical disdrometer for measuring size and velocity of hydrometeors, *J. Atmos. Oceanic Technol.*, *17*, 130–139.

Lovejoy, S., and D. Schertzer (1990), Multifractals, universality classes and satellite and radar measurements of cloud and rain fields, *J. Geophys. Res.*, *95*, 2021–2034, doi:10.1029/JD095iD03p02021.

Mandelbrot, B. B. (1982), *The Fractal Geometry of Nature*, W. H. Freeman, San Francisco, Calif.

Marsan, D., D. Schertzer, and S. Lovejoy (1996), Causal space-time multifractal processes: Predictability and forecasting of rain fields, *J. Geophys. Res.*, *101*, 26,333–26,346, doi:10.1029/96JD01840.

Martínez, M. D., X. Lana, A. Burgueño, and C. Serra (2007), Lacunarity, predictability and predictive instability of the daily pluviometric regime in the Iberian Peninsula, *Nonlinear Processes Geophys.*, *14*, 109–121.

Menabde, M., D. Harris, A. Seed, G. Austin, and D. Stow (1997), Multiscaling properties of rainfall and bounded random cascades, *Water Resources Res.*, *33*, 2823–2830.

Morton, N. E. (1955), Sequential tests for the detection of linkage, *Am. J. Hum. Genet.*, *7*, 277–318.

Olsson, J., J. Niemczynowicz, and R. Berndtsson (1993), Fractal analysis of high-resolution rainfall time series, *J. Geophys. Res.*, *98*, 23,265–23,274, doi:10.1029/93JD02658.

Over, T. M., and V. K. Gupta (1994), Statistical analysis of mesoscale rainfall: Dependence of a random cascade generator on large-scale forcing, *J. Appl. Meteorol.*, *33*, 1526–1542.

Over, T. M., and V. K. Gupta (1996), A space-time theory of mesoscale rainfall using random cascades, *J. Geophys. Res.*, *101*, 26,319–26,331, doi:10.1029/96JD02033.

Öztürk, A. (1981), On the study of a probability distribution for precipitation totals, *J. Appl. Meteorol.*, *20*, 1499–1505.

Paulson, K. S., and P. D. Baxter (2007), Downscaling of rain gauge timeseries by multiplicative beta cascade, *J. Geophys. Res.*, *112*, D09105, doi:10.1029/2006JD007333.

Pavlopoulos, H., and V. K. Gupta (2003), Scale invariance of regional wet and dry durations of rain fields: A diagnostic study, *J. Geophys. Res.*, *108*(D8), 8387, doi:10.1029/2002JD002763.

Pickands, J. (1975), Statistical inference using extreme order statistics, *Ann. Stat.*, *3*, 119–131.

Revfeim, K. F. A. (1982), Comments on "On the study of a probability distribution for precipitation totals," *J. Appl. Meteorol.*, *21*, 1942–1945.

Ross, S. M. (1996), *Stochastic Processes*, 2nd ed., John Wiley, New York.

Schertzer, D., and S. Lovejoy (1987), Physical modeling and analysis of rain and clouds by anisotropic scaling of multiplicative processes, *J. Geophys. Res.*, *92*, 9693–9714, doi:10.1029/JD092iD08p09693.

Schmitt, F., S. Vannitsem, and A. Barbosa (1998), Modeling of rainfall time-series using two-state renewal processes and multifractals, *J. Geophys. Res.*, *103*, 23,181–23,193, doi:10.1029/98JD02071.

Short, D. A., P. A. Kucera, B. S. Ferrier, J. A. Gerlach, S. A. Rutledge, and O. W. Thiele (1997), Shipborne radar rainfall patterns in TOGA-COARE, *Bull. Am. Meteorol. Soc.*, *78*, 2817–2836.

Storey, J. D. (2002), A direct approach to false discovery rates, *J. R. Stat. Soc., Ser. B*, *64*, 479–498.

Tokay, A., and D. A. Short (1996), Evidence from tropical raindrop spectra of the origin of rain from stratiform versus convective clouds, *J. Appl. Meteorol.*, *35*, 355–371.

Wang, J., B. L. Fisher, and D. B. Wolff (2008), Estimating rain rates from tipping-bucket rain gauge measurements, *J. Atmos. Oceanic Technol.*, *25*, 43–56, doi:10.1175/2007JTECHA895.1.

Wilson, E. B. (1927), Probable inference, the law of succession, and statistical inference, *J. Am. Stat. Assoc.*, *22*, 209–212.

P. K. Kundu, Joint Center for Earth Systems Technology, University of Maryland Baltimore County, 5523 Research Park Drive, No. 320, Baltimore, MD 21228, USA. (prasun.k.kundu@nasa.gov)

R. K. Siddani, Merck & Co., Inc., 2000 Galloping Hill Rd., Kenilworth, NJ 07033, USA.