Evidence for Inherent Nonlinearity in Temporal Rainfall

Stéphane G. Roux[1,2], Venugopal, V.[3], Kurt Fienberg[4],
Alain Arneodo[1,2], and Efi Foufoula-Georgiou[4]

[1] Université de Lyon F- 69000 Lyon, France

[2] Laboratoire de Physique
Ecole Normale Superieure de Lyon
46 Allée d́ Italie
69364 Lyon Cédex 07, France

[3] Centre for Atmospheric and Oceanic Sciences
Indian Institute of Science
Bangalore 560012, India

[4] St. Anthony Falls Laboratory
University of Minnesota
2 Third Avenue SE
Minneapolis, MN 55414, USA.

January, 2008

(submitted to *Advances in Water Resources*)

ABSTRACT

We examine the underlying structure of high resolution temporal rainfall by comparing the observed series with surrogate series generated by a invertible nonlinear transformation of a linear process. We document that the scaling properties and long range magnitude correlations of high resolution temporal rainfall series are inconsistent with an inherently linear model, but are consistent with the nonlinear structure of a multiplicative cascade model. This is in contrast to current studies that have reported for spatial rainfall a lack of evidence for a nonlinear underlying structure. The proposed analysis methodologies, which consider two-point correlation statistics and also do not rely on higher order statistical moments, are shown to provide increased discriminatory power as compared to standard moment-based analysis.

1 Introduction

Empirical evidence has repeatedly demonstrated that both temporal and spatial rainfall fields exhibit multiscaling (or multifractal) behavior, by which it is meant that the statistical moments of the fluctuations of rainfall have a power-law dependence on scale, with the power-law exponents varying as a non-linear function of moment order [e.g. *Lovejoy and Mandelbrot*, 1985; *Schertzer and Lovejoy*, 1987; *Gupta and Waymire*, 1990; *Veneziano et al.* 1996; *Deidda et al.*, 1999; *Venugopal et al.*, 2006]. One question that is raised by these findings is what type of stochastic model could reproduce the observed statistics. The presence of multiscaling has often been associated with an underlying multiplicative cascade as the model to generate rainfall or another preserved quantity related to rainfall [*Lovejoy and Schertzer*, 1991; *Gupta and Waymire*, 1993; *Deidda et al.*, 1999; among others]. However, in a recent study *Ferraris et al.* [2003] concluded that the observed scaling statistics of spatial rainfall could be reproduced by a linear model that was subjected to an invertible nonlinear transformation, as opposed to an inherently nonlinear model. The question of underlying nonlinearity of an observed series can be approached through the use of surrogate series, first introduced by *Theiler et al.* [1992] for hypothesis testing in nonlinear time series analysis [see also *Basu and Foufoula-Georgiou*, 2002]. To test for inherent nonlinearity in this framework, one stochastically generates a number of synthetic sequences that retain as many of the properties of the original data as possible but are derived from a linear model. Here the retained properties are the probability density function (pdf) and the linear correlation structure (or power spectrum) of the original series. Although the surrogates from a linear Gaussian model would have a Gaussian pdf, the original pdf is reconstructed by applying an invertible nonlinear transform as the final step in the surrogate generation process. This is what is meant by *inherent* or *underlying* linearity of a series: that even if there is nonlinearity present it is the result of an invertible nonlinear transform of a linear process. Finally, by comparing (with any pertinent test) the original series with the ensemble of the surrogate series, the presence of inherent nonlinearity and as the need for a nonlinear model can be concluded.

In the aforementioned study, *Ferraris et al.* [2003] found that the (multi)scaling statistical properties of spatial rainfall (using two-dimensional spatial rainfall fields from the GATE campaign) could not be distinguished from those of their surrogates. The metrics used to examine the scaling properties and test for nonlinearity were based on the log-log linear slopes of the statistical moments up to eighth order. As a result they concluded that a meta-Gaussian model (i.e. an invertible transform of a linear Gaussian model) would be adequate to reproduce the spatial structure of rainfall. A nonlinear transform of a linear underlying model (an exponentiation of a Langevin model) has been found also to adequately preserve the scaling properties of spatial rainfall by *Sapozhnikov and Foufoula-Georgiou* [2007].

In this paper we provide evidence that a model with linear underlying dynamics subjected to an invertible nonlinear transformation is not consistent with high-resolution temporal rainfall observations. This is shown through the examination of one- and two-point magnitude coefficient analysis, extensively used in turbulence and recently introduced for the analysis of scaling properties in rainfall [*Venugopal et al.* 2006 and references therein]. We conclude that a nonlinear model with long-range (power-law decaying) dependencies would be more consistent with the observations, such as a multiplicative cascade model, which has frequently been used as a stochastic model of rainfall [e.g. *Schertzer and Lovejoy*, 1987; *Rebora et al.*, 2006; *Gupta and Waymire*, 1993; *Carsteanu and Foufoula-Georgiou*, 1996; *Deidda et al.*, 1999].

The structure of this paper is as follows: in the following section we briefly outline the rainfall data used in this study; in section 3 the surrogate data series are introduced and the surrogate generation process is described; in section 4 the data analysis methods are presented with emphasis on the new methodologies based on magnitude cumulant analysis. The results of the analysis are presented in section 5, and conclusions drawn in section 6.

2 Rainfall observations

The high resolution temporal rainfall series analyzed here was collected during a storm on the 3rd of May, 1990, at the Iowa Institute of Hydraulic Research, University of Iowa. The instrumentation and the meteorological conditions are reported in *Georgakakos et al.* [1994]. The sampling interval of the data is 5s, and the duration approximately 9hr. The rainfall data is shown in Figure 1a. It can be seen with the naked eye that the intensity is quite variable, with a mean of 2.7 mm/hr but a peak above 10 mm/hr. The scaling properties of this rainfall time series were previously analyzed by *Venugopal et al.* [2006], in which the presence of multiscaling was documented within a range of scales between approximately 4 minutes and 1 hour. Note that the upper limit of this scaling range is much less than the storm duration (9 hrs) but is of the order of the duration of the storm pulses (1-2 hr), which can be identified in Figure 1a. Within this range of scaling, the series was found to show significant intermittency and deviation from monoscaling, as well as having long-range dependencies in the fluctuations, and to be consistent with a lognormal cascade model.

3 Surrogates

To investigate which type of model is consistent with the observed statistical structure of the rainfall data, and specifically to test the null hypothesis of inherent linearity with an invertible nonlinear transformation, we employ the so-called surrogate series [*Schreiber and Schmitz* 1996].The surrogates series $\{s_n\}$ is assumed to be generated by a process of the form

$$s_n = S(x_n), \quad x_n = \sum_{i=1}^{M} a_i x_{n-i} + \sum_{i=0}^{N} b_i \eta_{n-i}, \tag{1}$$

where S could be any invertible nonlinear function, $\{x_n\}$ is the underlying linear process, $\{a_n\}$ and $\{b_n\}$ are constant coefficients, and $\{\eta_n\}$ is white Gaussian noise.

Hypothesis testing is typically performed by evaluating some test statistic, or measure of nonlinearity, for both the original series and an ensemble of surrogate series. The results for the ensemble of surrogates provide the distribution of the test statistic that would be produced by an inherently linear process. This allows the establishment of confidence intervals for the rejection of the null hypothesis based on the value of the test statistic computed from the original series. To generate surrogates that maintain the pdf and correlation structure (and hence power spectrum) of the original data, we use the surrogate generation method proposed by *Schreiber and Schmitz* [1996], known as the iterative amplitude adjusted Fourier transform (IAAFT) method. This is a modification of the earlier amplitude adjusted Fourier transform (AAFT) method [*Theiler et al*. 1992], that iteratively adjusts both pdf and linear correlation structure to minimize their deviation from the original series. The generation process proceeds in the following way:

1. Randomly shuffle the data points of the original series $\{r_n\}$ to destroy any correlation or nonlinear relationships, while keeping the pdf unchanged. The reshuffled series is the starting point for the iteration $\{s_n^{(0)}\}$.

2. Take the Fourier transform of the current series $\{s_n^{(i)}\}$, and adjust the amplitudes to recreate the power spectrum of the original data. Keep the phases unchanged. Perform inverse Fourier transform.

3. The pdf will no longer be correct. Transform the data to the correct pdf by rank ordering and replacing each value with the value in the original series ($\{r_n\}$) with the same rank. This gives the updated series $\{s_n^{(i+1)}\}$.

4. Repeat steps 2 and 3 until the discrepancy in the power spectrum is below a threshold, or the sequence stops changing (reaches a fixed point).

In this manner a surrogate data series can be created with an identical pdf and optimally similar power spectrum to the original series. Any underlying nonlinear structure, which in Fourier space would be embodied by correlations in the phase, is destroyed, since only the absolute value (or power) of the Fourier coefficients is retained, whereas the phases are randomized by the shuffling of the series. For a heuristic argument for the convergence of the algorithm, see *Schreiber and Schmitz* [1996].

An ensemble of 20 surrogates were generated for the rainfall data using the IAAFT method (sufficient for a 95% significance level for a one-sided test). An example of a surrogate series for the rainfall data examined in this study can be seen in Figure 1b. Visually there is a striking resemblance with the original series, but in the following sections we will document striking differences, as inferred by some statistical tests that compare the scale dependence and long-range dependence of the original and surrogate series.

4 Methods of Analysis

In this section we review three different wavelet-based methods of analysing the scaling properties of a data series, with the goal of comparing the scale dependence of the rainfall data with that of its surrogates. Firstly the method of moments, or partition function approach, which looks directly at the scaling of the moments of the fluctuations [*Parisi and Frisch*, 1985; *Holschneider*, 1989], is summarised, and its limitations in the context of nonlinearity detection or comparison with surrogate data are discussed. Then two alternate test statistics to examine the scale dependent structure of the series are described: magnitude cumulant analysis [*Delour et al.*, 2001] and two-point magnitude correlation analysis [*Arneodo et al.*, 1998a, 1998b], which examine the one- and two-point statistics, respectively, of the magnitude coefficients of the rainfall fluctuations. Both of these were first developed for the analysis of fluid turbulence and have recently been applied to temporal rainfall series by *Venugopal et al.* [2006]. For completeness, brief summaries of the methods are presented here, but for a more in-depth discussion, readers are referred to *Venugopal et al.* [2006] and references therein.

4.1 Method of Moments

The original multifractal formalism was developed for measures in the context of dynamical systems (e.g., *Collet et al.*, 1997; *Halsey et al.*, 1986], with a generalization for functions provided by *Muzy et al.* [1991, 1994]. While a multifractal sequence is often described by the spectrum of singularities [*Bacry et al.*, 1993], in this work we focus on the alternate description, consisting of the scaling exponents of the statistical moments (these two descriptions are equivalent and related by a Legendre transform [*Parisi and Frisch*, 1985; *Arneodo et al.*, 1995]). More explicitly, the scale-dependence of fluctuations in a time series is described by the scaling exponent function $\tau(q)$, since for a scaling process

$$\mathcal{Z}(q, a) \sim a^{\tau(q)}, \tag{2}$$

where $\mathcal{Z}(q,a)$ is the statistical moment of order q, estimated from the observations as

$$\mathcal{Z}(q,a) = \frac{1}{N(a)} \sum_{x}^{N(a)} |T(x,a)|^q. \tag{3}$$

$T(x,a)$ are the so-called multi-resolution coefficients that capture the fluctuations in the time series at the scale a, and $N(a)$ is the total number of observations at that scale. The simplest choice for multiresolution coefficients $T(x,a)$ is to take first order increments, giving rise to what are known as structure functions [*Parisi and Frisch*, 1985]. However, working with first order increments has several limitations: they cannot detect singularities of order greater than 1, and they do not remove higher order non-stationary trends (first order increments remove only constant-level trends) [*Arneodo et al.*, 1995]. An alternative approach that avoids these limitations is to define $T(x,a)$ as the wavelet coefficients generated by the continuous wavelet transform (CWT), using wavelets of increasing-order vanishing moments, as shown by *Bacry et al.* [1993] and *Muzy et al.* [1991, 1994].

The continuous wavelet transform of a function $f(x)$ can be defined as:

$$T_\psi(x,a) = \frac{1}{a}\int f(x')\psi\left(\frac{x'-x}{a}\right)dx', \quad a>0, x \in R, \tag{4}$$

where a is the scale parameter, x is the location, and ψ defines a family of wavelets. For a general background on wavelets, see *Meyer*, [1992]; *Daubechies*, [1992]; *Mallat*, [1998]. For our analysis, we use as wavelets (apart from a normalizing constant) the successive derivatives of a Gaussian function $g^{(N)}(x) = \frac{d^N}{dx^N}e^{-x^2/2}$, which have N vanishing moments ($\int_{-\infty}^{+\infty} x^q g^{(N)}(x)dx = 0,\ 0 \leq q < N$), thus satisfying our need to remove higher order nonstationarities (polynomials trends) from the data, if such are present. These derivatives of a Gaussian have been used extensively to study the behavior of fractal functions [e.g. *Muzy et al.,* 1994; *Arneodo et al.,* 1995].

The wavelet-based multifractal analysis thus consists of estimating the statistical moments $\mathcal{Z}(q,a)$ using the wavelet coefficients $T_\psi(x,a)$ as the multiresolution coefficients, $T(x,a)$, in equation (3). Finally the scaling exponents $\tau(q)$ are estimated from equation (2). It is reminded that a linear $\tau(q)$ indicates monoscaling and a nonlinear $\tau(q)$ indicates multiscaling. Note that in this work, we focus on the estimate of $\tau(q)$ for $q>0$ only. This explains that we do not use the more sophisticated wavelet transform modulus maxima method (WTMM) ([*Muzy et al.*, 1991, 1994; *Arneodo et al*, 1985]) which was shown to be well suited to compute $\tau(q)$ for both positive and negative values of q. We have checked ([*Venugopal et al*, 2006]) that the results reported here for $q>0$ are consistent with the ones obtained with the WTMM method.

In the context of testing the consistency of a data set with a linear model (possibly with a subsequent invertible nonlinear transform) versus a nonlinear one, comparison is often performed between the $\tau(q)$ curve estimated from the observed data series with that estimated from the surrogates. Note that this was essentially the approach of *Ferraris et al* [2003] (although their multiresolution coefficients were not wavelet-based). There are two potential drawbacks of this approach. The first is the need to estimate higher-order moments from the data in order to accurately define the shape of the $\tau(q)$ curve. The use of higher-order moments is problematic not only for statistical reasons (a large number of data points is needed for accurate estimation), but also by the inherent degeneracy of higher moments (greater than a critical order value $q^{\star}$) due to the so-called multifractal phase transistion. This could be due to the fact that the observed multifractal field is actually the result of an integral over an underlying cascade process [*Schertzer and Lovejoy*,1992], or simply due to the inherent property of a multiplicative cascade to produce only a limited range of singularity strengths [*Lashermes et al*., 2004; *Lashermes*, 2005]. Specifically, *Venugopal et al*. [2006] found that for the temporal rainfall observations examined here, $q^{\star}$ was approximately 3, and that while the $\tau(q)$ curve was well-estimated for $q < 3$, it degenerated to a linear curve for moment-order greater than 3, due to the inherent multifractality of the field (specifically the limits on singularity strength), rather than any limitation of sample size.

The second drawback of using the method of moments for nonlinearity detection is simply that there is no clear test statistic for comparing the two $\tau(q)$ curves (for observations and surrogates), given that these curves have confidence intervals that vary with moment order. This problem is exacerbated by the first drawback mentioned above, i.e., the inability to accurately estimate $\tau(q)$ for higher order moments.

There is an alternate approach, however, for examining the scaling properties of a data set, which avoids the reliance on higher order moments and parameterizes the $\tau(q)$ curve with a only few parameters. This approach is known as the magnitude cumulant analysis [*Delour et al*., 2001], and we propose here that it can form the basis for a more powerful test for detetermining the compatibility of a series with the linear model.

4.2 Magnitude Cumulant Method

Let X be a random variable, $\mathcal{P}(x)$ its probability density function, $\Phi_{\mathcal{P}}(k)$, its moment generating function (i.e., the Fourier transform of $\mathcal{P}(x)$), and M_n its n^{th}-order moment. M_n can be estimated as the n^{th}-derivative of $\Phi_{\mathcal{P}}(k)$ at $k = 0$. The cumulant generating function of a random variable is defined as $\Psi_{\mathcal{P}}(k) = \ln \Phi_{\mathcal{P}}(k)$, and the cumulants C_n of X (similar to the moments) can be estimated as the n^{th}-derivative of $\Psi_{\mathcal{P}}(k)$ at $k = 0$. The moments and cumulants of X can in turn be related as:

$$C_1 = M_1,$$

$$
\begin{aligned}
C_2 &= M_2 - M_1^2, \\
C_3 &= M_3 - 3M_2M_1 + 2M_1^3, \\
C_4 &= M_4 - 4M_3M_1 - 3M_2^2 + 12M_2M_1^2 - 6M_1^4. \\
&\cdots
\end{aligned} \tag{5}
$$

Consider the q^{th}-order moment of the modulus of the wavelet coefficients, $|T_\psi(x,a)|$, as defined in Eqn. (3). Retaining only the dependence on scale a for the sake of brevity and rewriting the coefficients as $|T_a|$, it can be shown [*Delour et al*. 2001] that:

$$
-D_f \ln(a) + \sum_{n=1}^{\infty} C_n(a) \frac{q^n}{n!} \sim \tau(q) \ln(a), \tag{6}
$$

where D_f is the fractal dimension of the support of singularities, and $C_n(a)$ are the cumulants of the so-called magnitude coefficients, $\ln|T_a|$, i.e.,

$$
\begin{aligned}
C_1(a) &\equiv \langle \ln|T_a| \rangle \sim c_1 \ln(a), \\
C_2(a) &\equiv \langle \ln^2|T_a| \rangle - \langle \ln|T_a| \rangle^2 \sim -c_2 \ln(a), \\
C_3(a) &\equiv \langle \ln^3|T_a| \rangle - 3\langle \ln^2|T_a| \rangle \langle \ln|T_a| \rangle + \langle \ln|T_a| \rangle^3 \sim c_3 \ln(a), \\
&\cdots
\end{aligned} \tag{7}
$$

It is then easy to see from equations (6) and (7) that:

$$
\begin{aligned}
\tau(q) &= -D_f \frac{q^0}{0!} + \sum_{n=1}^{\infty} \left[\frac{C_n(a)}{\ln(a)} \right] \frac{q^n}{n!} \\
&= -c_0 + c_1 q - c_2 q^2/2! + c_3 q^3/3! \cdots
\end{aligned} \tag{8}
$$

where the coefficients $c_n > 0$ are estimated as the slope of $C_n(a)$ vs. $\ln(a)$ ($n = 1, 2, 3 \cdots$), and $c_0 = D_f$ (see also *Venugopal et al*. [2006] for the proof).

Thus having access to the coefficients c_n (from linear-log regression of the cumulants of $\ln|T_a|$ versus scale), one can estimate the functional form for $\tau(q)$. For instance, if $c_n = 0, n \geq 2$, then the given function is a monofractal, since $\tau(q)$ is linear. A quadratic estimate for $\tau(q)$, on the other hand, which signifies a multifractal, will require two regression fits to estimate c_1, and c_2. For the temporal rainfall data it was previously found [*Venugopal et al*., 2006] that $c_n = 0$ for $n > 2$, and as such a quadratic $\tau(q)$ was sufficient, and only two parameters were required. Thus, in relation to the standard structure function or wavelet-based multifractal formalism based on the method of moments, the cumulant-based estimation of the multifractal attributes could be considered more efficient, as it requires fewer regression fits to determine the shape of the spectrum of scaling exponents. More pertinently, it does not require the use of higher order moments and produces (for rainfall) just two relevant parameters, c_1 and c_2, whose distribution can be found for the surrogate data series and compared to the values for the original series. This method of course requires the convergence of at least the 2nd order cumulant. Note that the magnitude cumulant analysis implicitely assumes that the $\tau(q)$ spectrum does not display any nonanalyticity.

4.3 Two-point Magnitude Correlation Analysis

The analysis methods that we have presented thus far (method of moments and magnitude cumulant analysis) can be categorized as one-point statistics . However, it is known that for a multifractal the one-point statistics do not provide all possible information about the underlying mechanism that might have given rise to the multifractal: the two-point statistics will carry further information.

For example, one of the common models used to generate a multifractal field is a multiplicative cascade [*Schertzer and Lovejoy*, 1987; *Waymire and Gupta*, 1993; *Deidda et al*., 1999]. It is evident from a multiplicative cascade construction that the "parent" (large scale) would somehow be related to the children (any small scale). Thus, examining two-point statistics will allow us to evaluate if there exists any dependence between scales, and, if so, whether it follows any particular behavior. The hypothesis testing with surrogates remains necessary, as it allows us to assess the levels to which such long-term dependencies could occur through an inherently linear process, and hence set confidence levels on any conclusions we might make about the underlying cascade structure.

Specifically, the two-point correlation function of the magnitude coefficients (log of the wavelet coefficients), defined [*Arneodo et al*. 1998a, 1998b] as:

$$\mathcal{C}(a, \Delta x) = \langle\ (\ln|T_a(x)| - \langle \ln|T_a(x)|\rangle)\ (\ln|T_a(x+\Delta x)| - \langle \ln|T_a(x)|\rangle)\ \rangle, \tag{9}$$

can provide information about the space-scale (or time-scale) structure that underlies the multifractal properties of the considered signal. For example, if $\mathcal{C}(a, \Delta x)$ is logarithmic in Δx and independent of scale a provided that $\Delta x > a$, i.e.,

$$\mathcal{C}(a, \Delta x)\ \sim \ln \Delta x \qquad \Delta x > a\ , \tag{10}$$

then long-range dependence is inferred. For a random multiplicative cascade on a dyadic tree *Arneodo et al*. [1998a, 1998b] showed that:

$$\mathcal{C}(a, \Delta x) \sim\ -c_2 \ln \Delta x \tag{11}$$

where the proportionality coefficient c_2 is the same as the one defined in Eqn. (7)), i.e.,

$$\mathcal{C}(a, \Delta x = 0) \equiv C_2(a) \sim -c_2 \ln(a). \tag{12}$$

We reiterate that the presence of multifractality does not necessarily imply either long-range dependence or a multiplicative cascade structure [*Arneodo et al.,* 1999]. For instance, one can have scaling in $C_2(a)$, and if the (log-linear) slope is non-zero, it suggests the presence of multifractality. In addition to that, if $C(a, \Delta x)$ decreases to zero rapidly, it would suggest that there is no long-range dependence; or, if $C(a, \Delta x)$ changes linearly with $\ln \Delta x$, then it suggests long-range dependence. All these cases can be judged in relation to the surrogates of the data, that are known not to have a cascade construction, and can be used to show the range of correlation that could be produced with an underlying linear mechanism. When the slope of $C(a, \Delta x)$, vs $\ln a$, $\Delta x > a$, is equal to c_2 (which is the slope of $C_2(a)$ vs $\ln a$), and also significantly different than the slope of the inherently linear surrogates, we can infer that the underlying mechanism which gave rise to the multifractal, is a multiplicative cascade.

Thus the two-point magnitude correlation provides a second test that is not dependent on higher order moments, which provides a single numerical statistic that can be compared between rainfall and its surrogates, and can provide additional information about the underlying structure of a model that is consistent with the observations. Armed with these statistical methods of examining in detail the multiscale statistical structure of a series, we can now begin to examine the scaling properties of the rainfall and its surrogates.

5 Results

5.1 Application of the Method of Moments for comparison with Surrogates

The scale-dependence of the rainfall data and the twenty surrogate series, of the same length as the data, generated by the IAAFT method described in section 3, was analyzed by the method of moments. It was previously found [*Venugopal et al*., 2006] that an analyzing wavelet with $N = 3$ vanishing moments was required to remove the nonstationarities from the rainfall time series and correctly estimate its scaling properties. Thus the $g^{(3)}$ wavelet was used for the analysis of all data series for consistency, although using a lower order was not seen to affect the results for the surrogate data. It can be seen in Figure 2 that the moments of the wavelet coefficients are log-log linear, implying thus scaling-invariance for both the rainfall and surrogate data series between the scales of approximately 4 minutes and 1 hour. Plotting the scaling exponents $\tau(q)$ for rain and its surrogates (Figure 3) shows that although there is some difference between the scaling exponents for rain and the ensemble averaged values for the surrogates, the spread amongst the 20 surrogate series is certainly large enough to encompass the observed scaling exponents of rainfall. Hence, from the method of moments it would be difficult to reject the null hypothesis of an underlying linear generating process. This is the same conclusion that *Ferraris et al.* [2003] obtained for spatial rainfall data using the method of moments.

However, as outlined in section 4.1, the method of moments has limitations, and the cumulant analysis methods will be shown in the next section to exhibit higher discriminatory power. Finally, note that on average the $\tau(q)$ for the surrogate series in Figure 3 seems to be closer to a straight line, i.e. closer to a monofractal, than the rainfall data, although still showing a slight curvature. We will see that this tendency becomes more apparent as we consider the results from the other tests.

5.2 Magnitude Cumulant Analysis

Similarly, the wavelet-based magnitude cumulant analysis was performed on both the rainfall and surrogate time series. The first, second and third-order cumulants are displayed as a function of the scale a in Figure 4. While the rainfall and surrogates have overlapping third-order cumulants with slopes $c_3 \approx 0$, it can be seen that there is a clear difference in the second order cumulant, with the mean slope being $-c_2 = -0.26$ for rainfall and $-c_2 = -0.03$ for the surrogates. Thus the cumulant analysis shows an even more marked tendency for the surrogates to be closer to monoscaling, that is, the surrogates have a more linear $\tau(q)$ and lower intermittency as measured by the parameter c_2 (but still not having a $c_2 = 0$ that would indicate a perfectly monoscaling field). Figure 5 shows the frequency histogram for the values of c_2 estimated from a realization of 20 surrogates. The c_2 value of 0.26 estimated from the rainfall observations is greater than the largest c_2 of 0.08 for the surrogates. Since the c_2 of rainfall is greater than the largest value for the surrogates, this test would allow us to clearly reject (with a 95% confidence level) the null hypothesis of an inherently linear generating process.

Now consider the first order cumulant, which from Figure 4 is also linearly dependent on $\ln(a)$, with a slope of c_1. We have seen that the surrogates have significantly reduced c_2 compared to the rainfall series, which has a relatively high c_2. But the surrogate generation is designed to preserve the value of the power spectrum, and hence the scaling properties of the second moment of the fluctuations, which is to say that it preserves $\tau(2)$. So if one were to start with a parabolic $\tau(q)$ (as is the case with a lognormal cascade, and as has been observed for temporal rainfall [*Venugopal et al.* 2006]), i.e., $\tau(q) = c_1 q - c_2 q^2/2$, then the surrogate operation, by design, preserves $\tau(2)$. In other words,

$$2c_1^R - 2c_2^R = \tau^R(2) = \tau^S(2) = 2c_1^S - 2c_2^S \tag{13}$$

where the superscript R denotes the original rainfall and S the surrogates. Cancelling the factor of 2, we see $c_1^R - c_2^R = c_1^S - c_2^S$.

In our case the c_1 of rainfall is 0.64 from the slope of the first order cumulant. If we assume that c_2^S tends to 0, we would predict that the slope of the first order cumulant of the surrogates should be $c_1^S = c_1^R - c_2^R$, or 0.64-0.26=0.38 in our case. If we take the mean c_2^S of the surrogates as 0.03 (from our analysis), we get instead $c_1^S = 0.38 + 0.03 = 0.41$. Direct fitting of the slopes of the first order cumulant gives $c_1^S = 0.40 \pm 0.4$, closely matching either of these estimates.

Thus the cumulant analysis shows that the surrogates are much closer to monofractal (lower c_2) than the original rainfall series, and that the c_1 is reduced to compensate for this and maintain the original $\tau(2)$ value (and slope of the preserved power spectrum). The c_2 value of rainfall being significantly higher than the maximum c_2 of the surrogate ensemble indicates that the rainfall time series is not generated by an underlying linear Gaussian process, but has an inherently nonlinear structure. Finally, we will see that the two-point magnitude correlation confirms this result.

5.3 Two Point Magnitude correlation

The two-point magnitude correlation analysis described in section 4.3 was performed on the rainfall and surrogate time series, and the resulting magnitude correlation functions $\mathcal{C}(a, \Delta x)$ are shown in Figure 6 as a function of the displacement Δx. Once again there is a clear difference between the rainfall series that shows a very gradual decay of the correlation, i.e. a long range dependence in the magnitude coefficients, with a slope of -0.27 in the scaling range.. Recall that *Arneodo et al.*[1998a, 1998b] show that a multifractal cascade process on a wavelet dyadic tree leads to a slope of $\mathcal{C}(a, \Delta x)$ that is given by $-c_2$ (equation (11)), which is the same c_2 estimated by the cumulant analysis. So therefore this slope of -0.27, is consistent with the c_2 of 0.26 that was estimated in the previous section from scaling of the second order cumulant. Note that before we use this correspondence as an argument that the underlying generation process is consistent with a multiplicative cascade, we should check the magnitude correlation of the surrogates, to ensure that this long-range dependencies could not have, by chance, been generated by an underlying linear process. Observing Figure 6 however shows that the correlations in the magnitude of the surrogates, are in general much lower than those of the rainfall, and on average have a slope in the scaling range of just -0.02. Once again the comparison between the properties of the rainfall and its surrogates indicate that the rainfall series have an underlying structure which significantly differs from the linear structure of the surrogates.

6 Discussion and Conclusion

The question as to what models give rise to statistics which are consistent with the rich multiscaling structure of temporal and spatial rainfall observations, is a long debated one. Although it is understood that there might not exist a unique model which reproduces the statistics of the observed data, it is still of interest to at least be able to infer some of the basic characteristics required for a model to match the observations, such as deterministic versus stochastic and linear vs. nonlinear dynamics. This paper focuses on whether a distinction can be made between the rainfall series generated by a linear stochastic process subjected to an invertible nonlinear transformation, and a rainfall series generated by an inherently nonlinear process, such as a multiplicative cascade.

Our study was motivated by the previous work of *Ferraris et al.* [2003], which reported that multifractal cascades or other nonlinear stochastic processes might not be necessary to reproduce the observed spatial rainfall statistics, and that a nonlinear filtering of a linear autoregressive process suffices. In this paper we report distinct differences between the statistical properties of temporal rainfall and a nonlinearly filtered linear process (exogenous nonlinearity), while we report similarities between rainfall and a multiplicative cascade process (inherent nonlinearity). We attribute our ability to depict differences between linear and nonlinear structures, in the more powerful testing methodology we employ, based on magnitude cumulant analysis instead of method of moments. The magnitude cumulant analysis offers two main advantages: firstly it avoids the use of higher order moments, known to suffer from statistical convergence problems when estimated from small samples and also to exhibit a theoretically degenerate behaviour - called a linearization effect - in known models such as both purely multiplicative cascades [e.g. *Lashermes* 2008] and fractionally integrated cascades [e.g. *Schertzer et al.*, 1993]. Secondly, the magnitude cumulant analysis parameterizes the multiscaling in a few parameters (c_1 and c_2 here) and makes statistical inference easier [see also *Basu et al.*, 2007]. In addition, we employ a two-point correlation analysis which adds significantly to the ability to clearly depict differences in the linear and nonlinear dynamics. By way of comparison, we demonstrated that using only the scaling exponents directly estimated from the statistical moments of the fluctuations, it was impossible to reject the inherent linearity hypothesis for temporal rainfall data, just as was concluded for spatial data by *Ferraris et al.* [2003].

As a concluding remark, we can note that determining the consistency of the data series with a linear (or nonlinear) underlying structure could serve as a model diagnostic for stochastic simulation or downscaling models of rainfall. For example, the results in this paper would indicate that for high resolution temporal rainfall the statistical structure (on the order of minutes to hours) would not be well reproduced by a linear model, subject to any invertible transformation. Rather, an inherently nonlinear model structure would be necessary.

7 Acknowledgements

This work was supported by a NASA-GPM award NNX07AD33G and NSF award EAR-0120914 to the National Center for Earth-surface Dynamics, an NSF Science and Technology Center.

References

[ ] Arneodo, A., E. Bacry and J. F.Muzy (1995), The thermodynamics of fractals revisited with wavelets, *Physica A, 213*, 232-275.

[ ] Arneodo, A., E. Bacry, S. Manneville and J. F. Muzy (1998a), Analysis of random cascades using space-scale correlation functions, *Phys. Rev. Lett., 80*, 708-711.

[ ] Arneodo, A., E. Bacry and J. F. Muzy (1998b), Random cascades on wavelet dyadic trees, *J. Math. Phys., 39*, 4142-4164.

[ ] Arneodo, A., S. Manneville, J. F. Muzy and S. G. Roux (1999), Revealing a lognormal cascading process in turbulent velocity statistics with wavelet analysis, *Phil. Trans. R. Soc. of Lond. A, 357*, 2415-2438.

[ ] Bacry, E., J. F. Muzy and A. Arneodo (1993), Singularity spectrum of fractal signals from wavelet analysis : Exact results, *J. Stat. Phys. 70*, 635-674.

[ ] Basu, S. et al. (2007), Estimating intermittency exponent in neutrally stratified atmospheric surface layer flows: A robust framework based on magnitude cumulant and surrogate analyses, *Physics of Fluids*, 19, 115102.

[ ] Basu, S., and E. Foufoula-Georgiou (2002), Detection of nonlinearity and chaosticity in time series using the transportation distance function, *Phys. Lett. A, 3001*, 413-424.

[ ] Cârsteanu, A. and E. Foufoula-Georgiou (1996), Assessing dependence among weights in a multiplicative cascade model of temporal rainfall, J. Geophysical Research, 101(D21), 26363-26370.

[ ] Collet, P., J.L. Lebowitz, and A. Porzio (1987), The dimension spectrum of some dynamical systems, *J. Stat. Phys.*, 47, 609-644.

[ ] Daubechies, I. (1992), *Ten Lectures on Wavelets*, CBMS-NSF Reg. Conf. Ser. Appl. Math., vol. 61, SIAM, Philadelphia, Pa.

[ ] Deidda, R., R. Benzi, and F. Siccardi, (1999), Multifractal modeling of anomalous scaling laws in rainfall, *Water Resour. Res., 35*, 1853-1867.

[ ] Delour, J., J. F. Muzy and A. Arneodo (2001), Intermittency of 1D velocity spatial profiles in turbulence: A magnitude cumulant analysis, *The Euro. Phys. Jour. B, 23*, 243-248.

[ ] Ferraris, L., V. Gabellani, V. Parodi, N. Rebora, J. von Hardenberg and A. Provenzale (2003), Revisiting multifractality in rainfall fields, *J. Hydrometeorol., 4*, 544-551.

[ ] Georgakakos, K. P., A. A. Cârsteanu, P. L Sturdevant and J. A. Cramer (1994), Observation and analysis of midwestern rain rates, *J. Appl. Meteorol., 33*, 1433-1444.

[ ] Gupta, V. K., and E. Waymire (1993), A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteor.*, 32(2), 251-267.

[ ] Gupta, V. K., and E. Waymire (1990), Multiscaling properties of spatial rainfall and river flow distributions, *J. Geophys. Res.*, 95 (D3), 1999-2009.

[ ] Halsey, T.C., M.H. Jenson, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman (1986), Fractal measures and their singularities: The characterization of strange sets, *Phys. Rev. A*, 33, 1141-1151.

[ ] Holschneider, M. (1989), L'analyse d'objets fractals et leur transformation en ondelettes, Ph. D. Thesis, Univ. of Aix-Marseille II, France.

[ ] Lashermes, B., P. Abry, and P. Chanais (2004) New insights into the estimation of scaling exponents, *Int. J. Wavelets Multiresolut. Info Processing*, 2, 297-523.

[ ] Lashermes, B. (2005), Analyse multifractale pratique: coefficients dominants et ordres critiques. Applications a la turbulence pleinement developpee. Effets de nombre de Reynolds fini, Ph.D. Thesis, Ecole Normale Superieure de Lyon, Lyon, France.

[ ] Lovejoy, S. and B.B. Mandelbrot (1985), Fractal Properties of Rain, and a Fractal Model, *Tellus*, 37A(3), 209-232.

[ ] Lovejoy, S. and D. Schertzer (1991), Multifractal analysis techniques and the rain and cloud fields from 10^{-3} to 10^{6}m, in *Non-Linear Variability in Geophysics : Scaling and Fractals*, edited by D. Schertzer and S. Lovejoy, Kluwer Acad., Norwell, Mass.

[ ] Mallat, S. (1998), A Wavelet Tour in Signal Processing, Academic Press, New York.

[ ] Meyer, Y. (1992), *Wavelet and Applications*, Springer, Berlin.

[ ] Muzy, J. F., E. Bacry and A. Arneodo (1991), Wavelet and Multifractal Formalism for Singular Signal: Application to Turbulence Data, Phys. Rev. Lett. 67, 3515-3518.

[ ] Muzy, J. F., E. Bacry and A. Arneodo (1994), The multifractal formalism revisited with wavelets, *International J. Bifurcation and Chaos, 4*, 245-302.

[ ] Parisi, G. and U. Frisch (1985), On the singularity structure of fully developed turbulence, in *Turbulence and Predictability in Geophysical Fluid Dynamics*, edited by M. Ghil, R. Benzi and G. Parisi, North -Holland, Amsterdam,84-88.

[ ] Rebora, N., L. Ferraris, J.v-Hardenberg., and A. Provenzale (2006), RainFARM: Rainfall downscaling by a filtered autoregressive model, Journal of Hydrometeorology, 7(4) 724-738.

[ ] Sapozhnikov, V. and E. Foufoula-Georgiou (2007), An exponential Langevin-type model for rainfall exhibiting spatial and temporal scaling, in Nonlinear dynamics in geosciences, A. Tsonis and J. Elsner eds., Springer-Verlag, New York.

[ ] Schertzer, D., S. Lovejoy, D. Lavalle (1993), Generic multifractal phase transitions and self-organized criticality, *Cellular Automata: Prospects in astrophysical applications*, J.M. Perdang and A. Lejeune eds., World Scientific, p 216-227.

[ ] Schertzer, D. and S. Lovejoy (1987), Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes, J. Geophys. Res., 92(D8), 9693-9714.

[ ] Schertzer, D. and S. Lovejoy (1992), Hard and Soft Multifractal processes: Physica A, 185, 187-194.

[ ] Schreiber, T. and A. Schmitz (1996), Improved surrogate data for nonlinearity tests, *Phys. Rev. Lett.*, 77, 635-638.

[ ] Theiler, J., S. Eubank, A. Longtin, B. Galdikian, and J.D. Farmer (1992), Testing for nonlinearity in time series: the method of surrogate data, *Physica D, 58*, 77-94.

[ ] Veneziano, D., R.L. Bras, and J.D. Niemann (1996), Nonlinearity and self-similarity of rainfall in time and a stochastic model, *J. Geophys. Res.*, 101(D21), 26371-26392.

[ ] Venugopal, V., S. G. Roux, E. Foufoula-Georgiou and A. Arneodo (2006), Revisiting multifractality of high-resolution temporal rainfall data using a wavelet-based formalism, *Water. Resour. Res. 42*, doi: W06D1410.1029/2005WR004489.

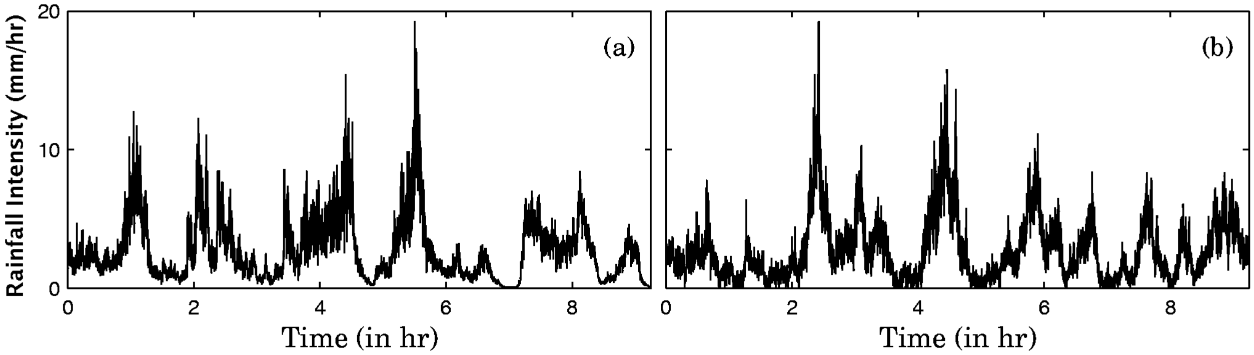

Figure 1: (a) The observed rainfall series of May 3, 1990 in Iowa city; (b) one surrogate time series which preserves the pdf and power spectrum of the original series.

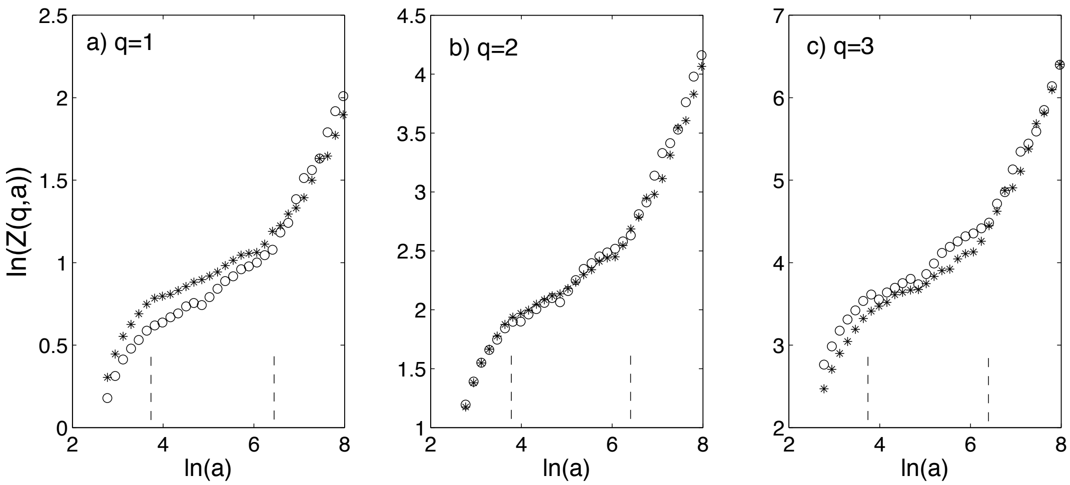

Figure 2: Moments of the intensity of rainfall (o) and of surrogates (*) analyzed with the $g^{(3)}$ wavelet. Shown are the first, second and third moments.

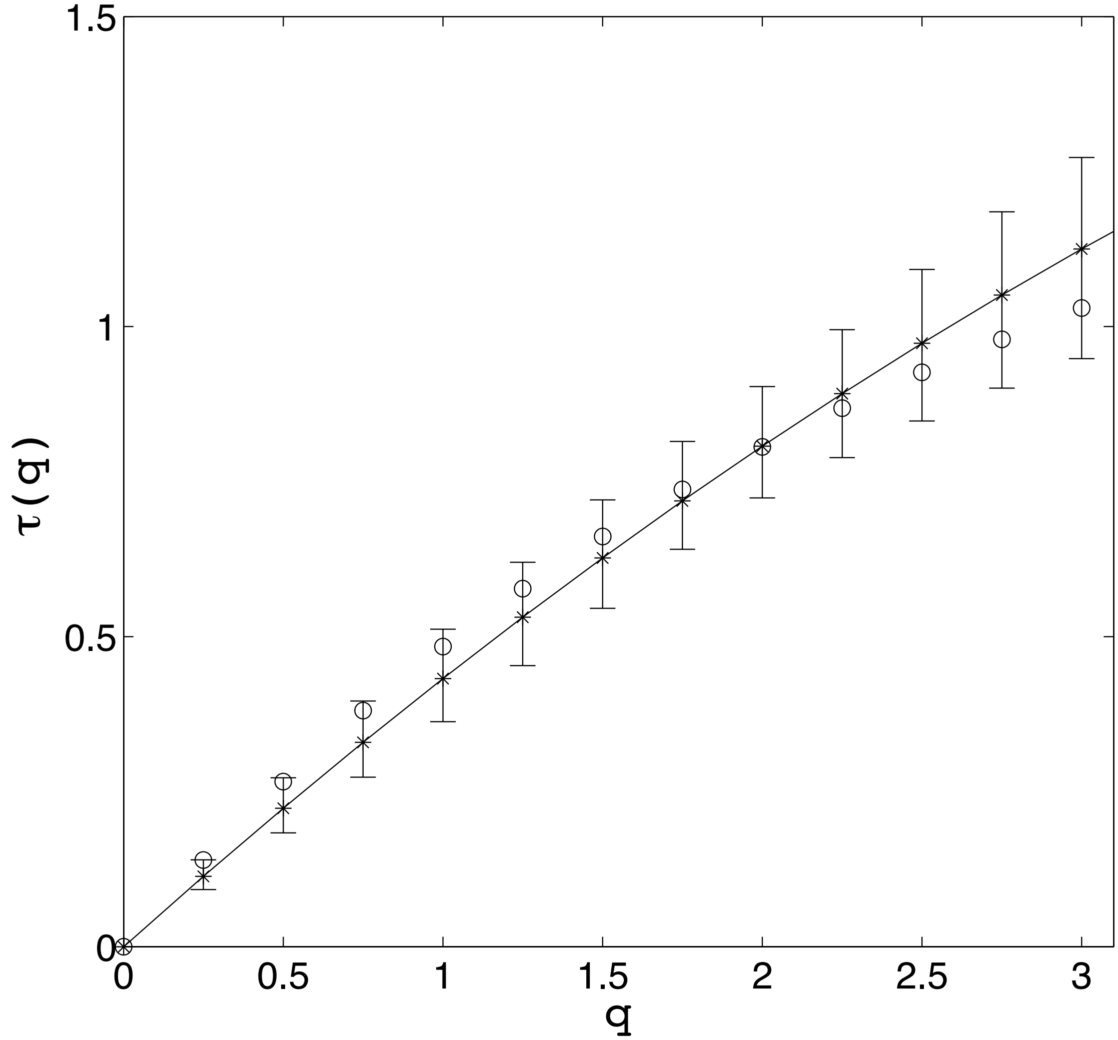

Figure 3: $\tau(q)$ curve for rain (o) and surrogates (*). Error bars represent the 5% and 95% levels for x the surrogate series.

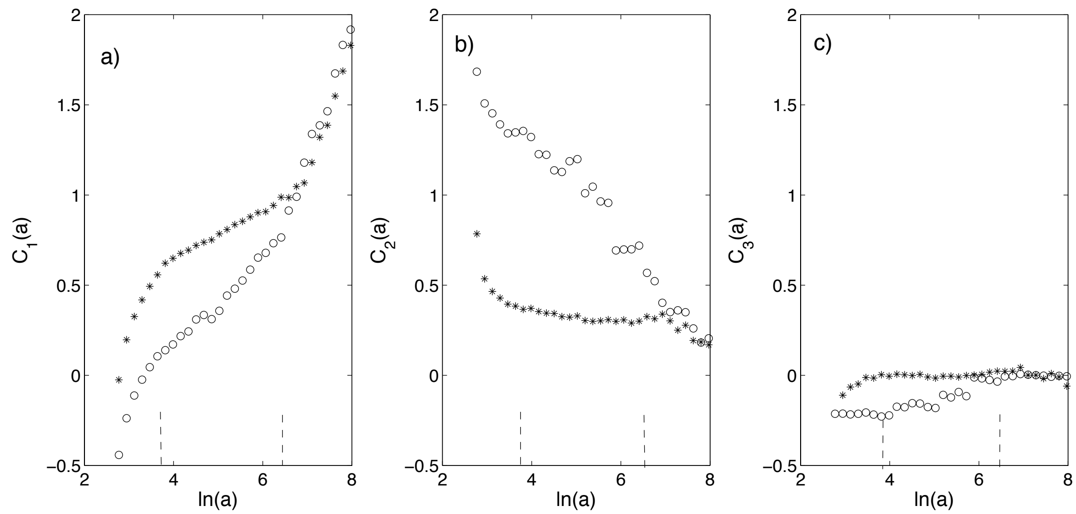

Figure 4: Cumulants of the rainfall intensity and of surrogates analyzed with the wavelet $g^{(3)}$. Plots (a), (b) and (c), show the first, second and third cumulants respectively, for the rainfall (o) and surrogates ($\star$). The vertical lines indicate the scaling range over which parameter estimation was performed.

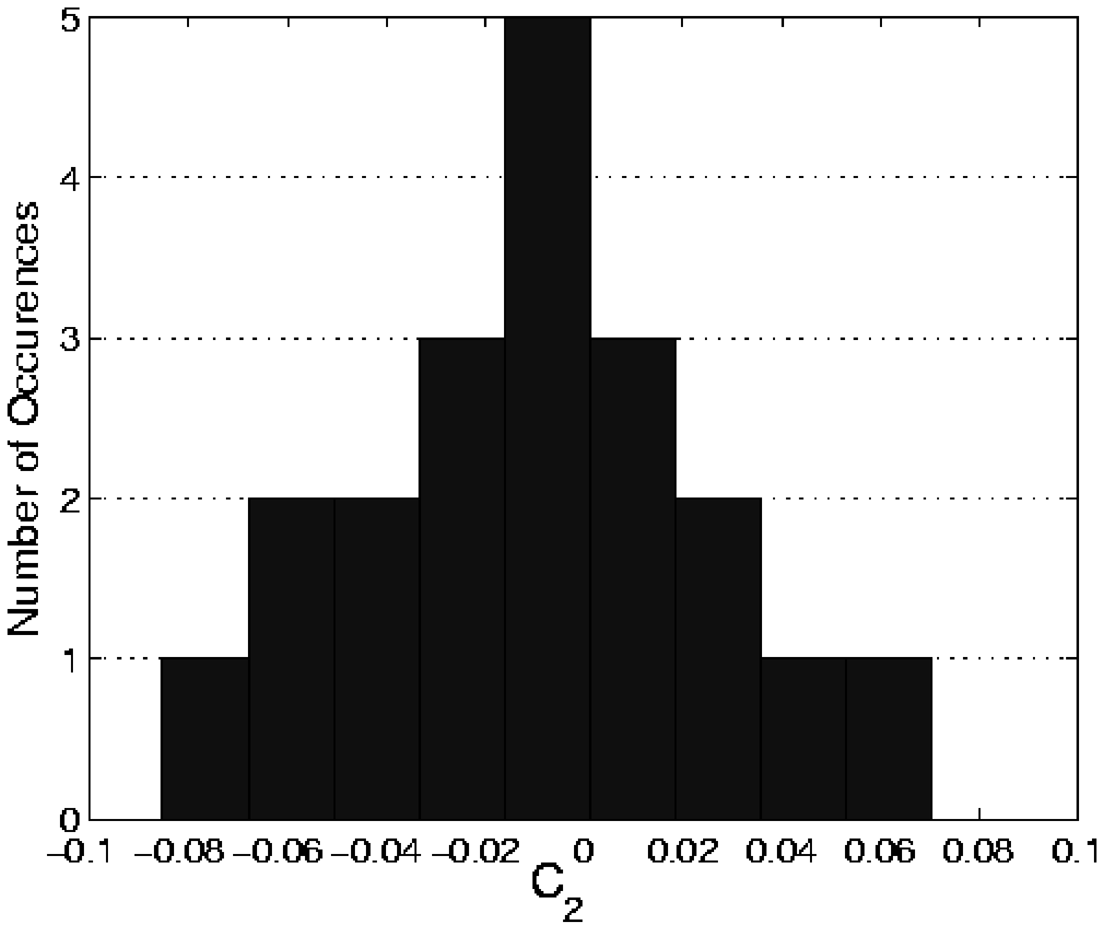

Figure 5: Histogram of c_2 values estimated from the magnitude cumulant analysis of 20 surrogate data series

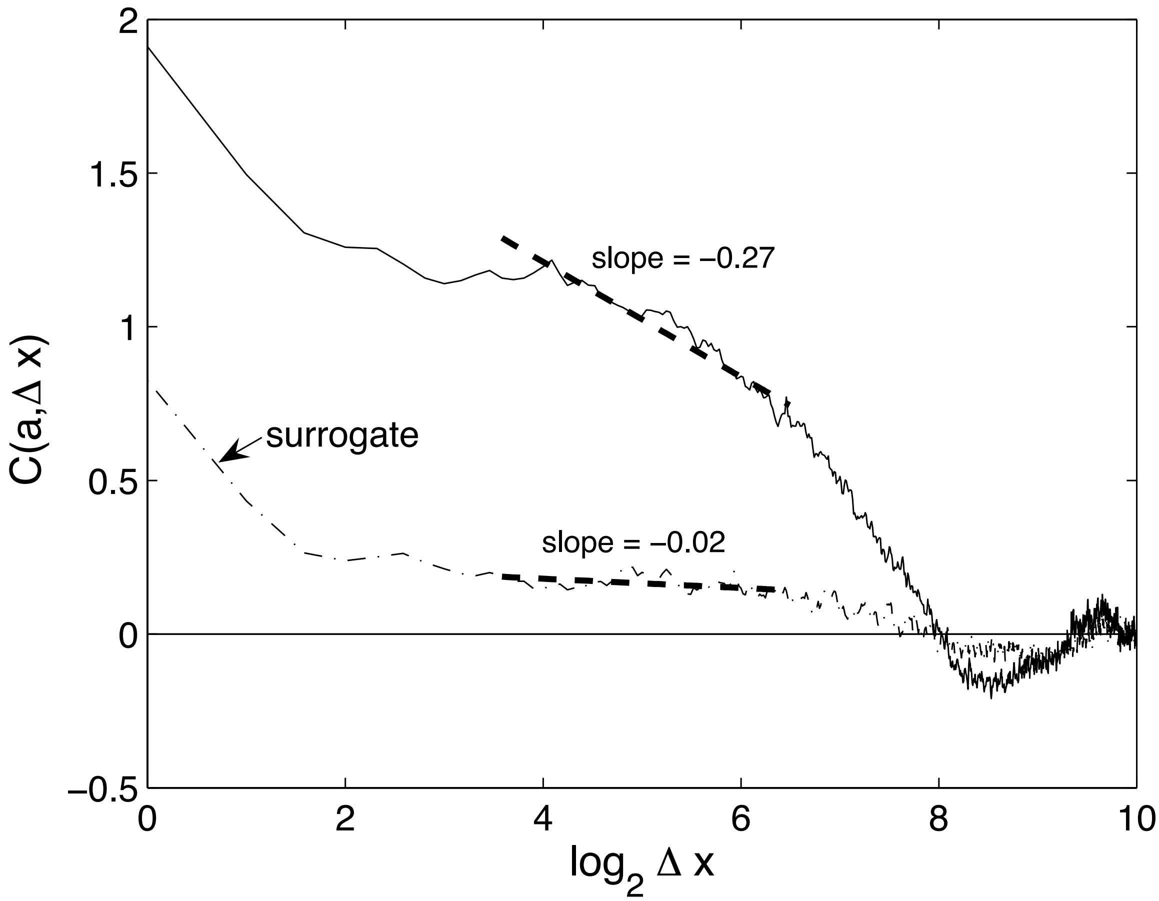

Figure 6: Two point magnitude correlation of the rainfall series (solid line) and its surrogates (dashed curve) computed at scale $a =?$.