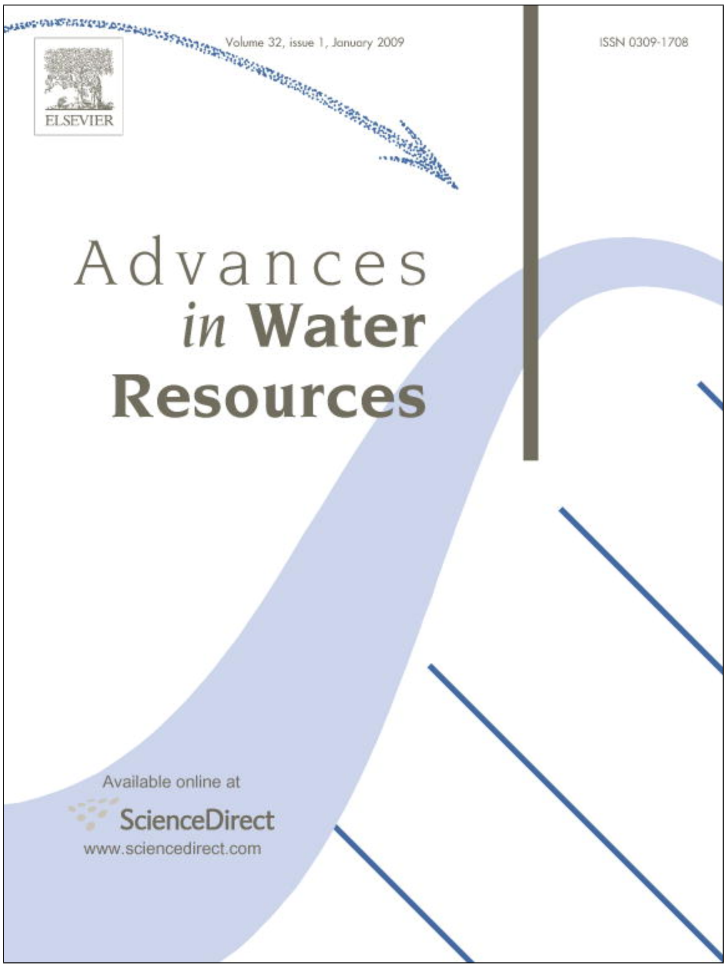
Volume 32, issue 1, January 2009
ISSN 0309-1708
ELSEVIER
Advances
in Water
Resources
Available online at
ScienceDirect
www.sciencedirect.com

Advances in Water Resources 32 (2009) 41–48

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Evidence for inherent nonlinearity in temporal rainfall

Stéphane G. Roux [a,b], V. Venugopal [c], Kurt Fienberg [d], Alain Arneodo [a,b], Efi Foufoula-Georgiou [d,*]

[a] Université de Lyon, F-69000 Lyon, France
[b] Laboratoire de Physique, Ecole Normale Superieure de Lyon, 46 Allée d' Italie, 69364 Lyon Cédex 07, France
[c] Centre for Atmospheric and Oceanic Sciences, Indian Institute of Science, Bangalore 560012, India
[d] St. Anthony Falls Laboratory, Department of Civil Engineering, University of Minnesota, 2 Third Avenue SE, Minneapolis, MN 55414, USA

ARTICLE INFO

Article history:
Received 14 February 2008
Received in revised form 22 September 2008
Accepted 24 September 2008
Available online 10 October 2008

Keywords:
Rainfall
Multiscaling
Surrogates
Nonlinearity
Wavelets
Multifractals

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. [22,33,16,37,11,38]). 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 ([23,17,11]; among others). However, in a recent study Ferraris et al. [14] 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. [36] for hypothesis testing in nonlinear time series analysis (see also [7]). 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 the need for a nonlinear model can be objectively assessed.

In the aforementioned study, Ferraris et al. [14] 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 the 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. Sapozhnikov and Foufoula-Georgiou [31] also found that a nonlinear transform of a linear underlying model (an exponentiation of a Langevin model) could adequately preserve the scaling properties of spatial rainfall.

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

* Corresponding author. Tel.: +1 612 626 0369; fax: +1 612 624 4398.
E-mail address: efi@umn.edu (E. Foufoula-Georgiou).

0309-1708/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2008.09.007

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 ([38] 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. [33,17,8,11]).

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. [15]. The sampling interval of the data is 5 s, and the duration of the rain is approximately 9 h. The rainfall data is shown in Fig. 1. It can be seen with the naked eye that the intensity is quite variable, with a mean of 2.7 mm/h but peaks above 10 mm/h. The scaling properties of this rainfall time series were previously analyzed by Venugopal et al. [38], in which the presence of multiscaling was documented within a range of scales between approximately 4 min and 1 h. Note that the upper limit of this scaling range is much less than the storm duration (9 h) but is of the order of the duration of the storm pulses (1–2 h), which can be identified in Fig. 1a. Within this range of scaling, the rainfall data was found to show significant intermittency and deviation from monoscaling; furthermore, it showed long-range dependence in the fluctuations, that was found 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 [36,35]. The surrogates series $\{s_n\}$ is assumed to be generated by a process of the form

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

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. M and N are the orders of an autoregressive (first term) and moving average (second term) model, respectively.

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 method proposed by Schreiber and Schmitz [35], known as the iterative amplitude adjusted Fourier transform (IAAFT) method. This is a modification of the earlier amplitude adjusted Fourier transform (AAFT) method [36], 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 [35].

An ensemble of 100 surrogates were generated for the rainfall data using the IAAFT method (sufficient for a 95% significance level

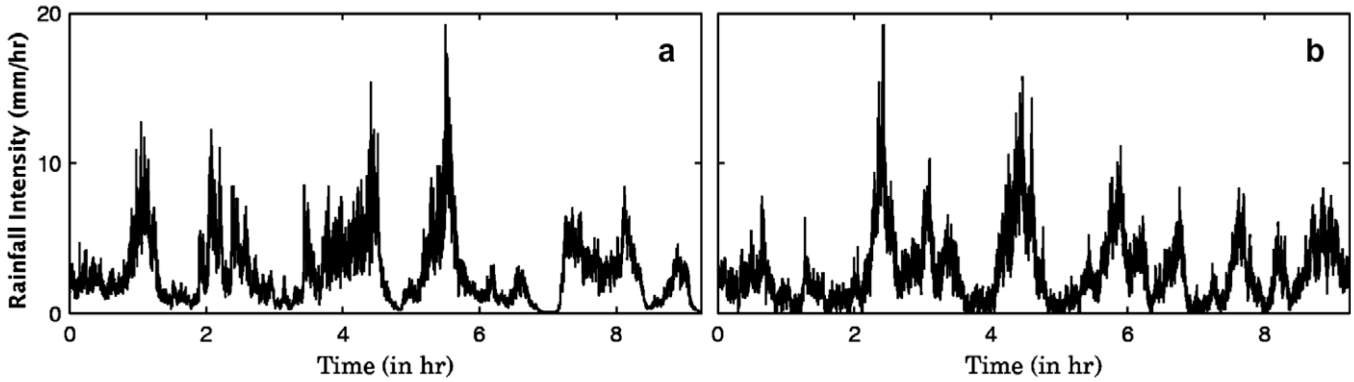

Fig. 1. (a) The observed rainfall series of May 3, 1990 in Iowa city; (b) one surrogate time series which preserves the pdf and power spectrum of the original series.

for a one-sided test). An example of a surrogate series for the rainfall data examined in this study can be seen in Fig. 1b. Visually, there is an obvious 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 analyzing 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 [29,19,28], is summarized, 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 [13] and two-point magnitude correlation analysis [2,3], 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. [38]. For completeness, a brief summary of each method is presented here, and for a more in-depth discussion, the reader is referred to Venugopal et al. [38] and references therein.

4.1. Method of moments

The original multifractal formalism was developed for measures in the context of dynamical systems [e.g. [18,9]], with a generalization for functions provided by Muzy et al. [26,28]. While multifractal distributions (including measures and functions) are often described by the spectrum of Hölder exponents h also called $D(h)$ singularity spectrum [5], 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 [29,1]). 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

$$Ƶ(q,a) \sim a^{\tau(q)}, \tag{2}$$

where $Ƶ(q,a)$ is a partition function corresponding to the statistical moment of order q, estimated from the observations as

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

$T(x,a)$ are the so-called multiresolution 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 [29]. However, working with first order increments has several limitations: they cannot detect singularities of Hölder exponent h greater than 1, and they do not remove higher order non-stationary trends (first order increments remove only constant-level trends) [1]. 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. [5] and Muzy et al. [26,28].

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 [25], Daubechies [10] or Mallat [24]. 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 ($\int_{-\infty}^{+\infty} x^q g^{(N)}(x)dx = 0, 0 \leqslant q < N$), thus satisfying our need to remove higher order nonstationarities (polynomial trends), if present, from the data. These derivatives of a Gaussian have been used extensively to study the behavior of multifractal functions (e.g. [28,1]).

The wavelet-based multifractal analysis thus consists in estimating the partition function $Ƶ(q,a)$ using the wavelet coefficients $T_\psi(x,a)$ as the multiresolution coefficients, $T(x,a)$, in Eq. (3). It is important to note that there are two wavelet-based approaches to calculate the partition function: (i) use all the wavelet coefficients at all scales; or (ii) use only those wavelet coefficients which constitute the local modulus maxima at each scale and can be chained together to form lines of maxima from the largest to the smallest scale. The CWT-based method cannot estimate the partition function for $q < 0$ since the estimate can potentially diverge (e.g., see [38]). To alleviate this concern, we choose to use the second method based on the Wavelet Transform Modulus Maxima (WTMM), introduced by Muzy et al. [27,28]. In other words, the estimation of the partition function is reformulated as

$$Ƶ(q,a) = \sum_{l\in\mathscr{L}(a)} [|T_\psi(x,a)|]^q, \tag{5}$$

where $q \in \mathscr{R}$, $\mathscr{L}(a)$ is the set of all maxima lines that satisfy: $l \in \mathscr{L}(a)$, if $\forall a' \leqslant a, \exists (x,a') \in l$. In order to avoid numerical instabilities in the computation of the partition function, especially those that are likely to arise for $q < 0$, Muzy et al. [27,28] suggested that the value of the wavelet transform modulus at each maximum be replaced by the supremum value along the corresponding maxima line at scales smaller than a.

We do not delve into the rigorous mathematical formalism of the reformulation here, but suggest that the reader refers to the original work of Muzy et al. [27,28] or to the more recent (and related) work of Venugopal et al. [38], in which it was illustrated that WTMM-based estimation of the partition function is a more appropriate and robust method, compared to CWT-based estimation, to analyze high-resolution temporal rainfall. For the rest of the manuscript, the WTMM-based approach is assumed, unless otherwise specified. Once the partition function is calculated using Eq. (5), the scaling exponents $\tau(q)$ are estimated from Eq. (2). It is reminded that a linear $\tau(q)$ indicates monoscaling and a nonlinear $\tau(q)$ indicates multiscaling.

In the context of testing the consistency of a data set with a linear model (possibly with a subsequent invertible nonlinear transform) vs. 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. [14] (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 positive ($q > q_+^* > 0$) and negative ($q < q_-^* < 0$) moments due to the so-called multifractal phase transition. This could be due to the fact that the observed multifractal field is actually the result of an integral over an underlying cascade process [34], or simply due to the inherent property of a multiplicative cascade to produce only a limited range of singularity strengths [20,21]. Specifically, Venugopal et al. [38] found that for the temporal rainfall observations examined here, $| q_\pm^* |$ 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 $|q|$ 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 only a few parameters. This approach is known as the magnitude cumulant analysis [13], and we propose here that it can form the basis for a more powerful test for determining the compatibility of a series with a linear model.

4.2. Magnitude cumulant method

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

$$\begin{aligned} C_1 &= M_1, \\ C_2 &= M_2 - M_1^2, \\ C_3 &= M_3 - 3M_2M_1 + 2M_1^3, \\ C_4 &= M_4 - 4M_3M_1 - 3M_2^2 + 12M_2M_1^2 - 6M_1^4. \\ &\ldots \end{aligned} \tag{6}$$

Consider the qth-order moment of the modulus of the wavelet coefficients, $|T_\psi(x,a)|$, as defined in Eq. (3). Retaining only the dependence on scale a for the sake of brevity and rewriting the coefficients as $|T_a|$, it can be shown [13] that:

$$-D_f \ln(a) + \sum_{n=1}^{\infty} C_n(a)\frac{q^n}{n!} \sim \tau(q)\ln(a), \tag{7}$$

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.,

$$\begin{aligned} C_1(a) &\equiv \langle \ln|T_a|\rangle \sim c_1 \ln(a), \\ C_2(a) &\equiv \langle \ln^2|T_a|\rangle - \langle \ln|T_a|\rangle^2 \sim -c_2\ln(a), \\ C_3(a) &\equiv \langle \ln^3|T_a|\rangle - 3\langle \ln^2|T_a|\rangle\langle \ln|T_a|\rangle + \langle \ln|T_a|\rangle^3 \sim c_3\ln(a), \\ &\ldots \end{aligned} \tag{8}$$

It is then easy to see from Eqs. (7) and (8) that:

$$\begin{aligned} \tau(q) &= -D_f\,\frac{q^0}{0!} + \sum_{n=1}^{\infty}\left[\frac{C_n(a)}{\ln(a)}\right]\frac{q^n}{n!}, \\ &= -c_0 + c_1 q - c_2 q^2/2! + c_3 q^3/3! \cdots, \end{aligned} \tag{9}$$

where the coefficients $c_n > 0$ are estimated as the slope of $C_n(a)$ vs. $\ln(a)$ ($n = 1,2,3,\ldots$), and $c_0 = D_f$ (see also [38] for the proof).

Thus having access to the coefficients c_n (from linear-log regression of the cumulants of $\ln|T_a|$ vs. scale), one can estimate the functional form for $\tau(q)$. For instance, if $c_n = 0$, $n \geqslant 2$, 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 [38] 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 second order cumulant. Note that the magnitude cumulant analysis implicitly assumes that the $\tau(q)$ spectrum does not display any nonanalyticity.

Remark. Before we proceed with the discussion of two-point statistics, we elaborate on some computational issues surrounding the estimation of $\tau(q)$ using the "WTMM with supremum" method. This method (which enables us to estimate $\tau(q)$ for $q < 0$) has a minor disadvantage of not being able to estimate singularities of negative Hölder exponent. The issue of singularities with $h < 0$ is important because Venugopal et al. [38] illustrated that the high resolution temporal rainfall considered in this work has values of h ranging between -0.7 and 1.5 (e.g., see Fig. 13 of [38]). Recalling that integration adds 1 to the singularity Hölder exponents of a signal, we address this deficiency of the "WTMM with sup" method, by working with the cumulative (integral) of the rain. Consequently, $c_1^{\text{cumulative}}$, the first-order cumulant slope (Eq. (8)) becomes $c_1^{\text{orig}} + 1$. (All other order cumulants remain the same.) It is easy to see that this affects the spectrum of scaling exponents $\tau(q)$ (owing to Eq. (9)), i.e., $\tau^{\text{cumulative}}(q) = \tau^{\text{orig}}(q) + q$. Thus, in order to retrieve $\tau(q)$ of the original signal, we estimate $\tau^{\text{cumulative}}(q)$ and subtract q.

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 (e.g., [33,17,11]). 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 [2,3] as

$$\begin{aligned} \mathscr{C}(a,\Delta x) = \langle (\ln|T_a(x)| - \langle \ln|T_a(x)|\rangle)(\ln|T_a(x+\Delta x)| \\ - \langle \ln|T_a(x)|\rangle)\rangle \end{aligned} \tag{10}$$

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

$$\mathscr{C}(a,\Delta x) \sim \ln \Delta x, \quad \Delta x > a, \tag{11}$$

then long-range dependence is inferred. For a random multiplicative cascade on a dyadic tree Arneodo et al. [2,3] showed that:

$$\mathscr{C}(a,\Delta x) \sim -c_2 \ln \Delta x, \tag{12}$$

where the proportionality coefficient c_2 is the same as the one defined in Eq. (8), i.e.,

$$\mathscr{C}(a,\Delta x = 0) \equiv C_2(a) \sim -c_2 \ln(a). \tag{13}$$

We reiterate that the presence of multifractality does not necessarily imply either long-range dependence or a multiplicative cascade structure [4]. 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 also 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. It provides (i) a single numerical statistic that can be compared between rainfall and its surrogates, and (ii) additional information about the underlying structure of a model that is consistent with the observations. Armed with these statistical methods of unravelling 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 100 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 [38] 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 Fig. 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 min and 1 h. Plotting the scaling exponents $\tau(q)$ for rain and its surrogates (Fig. 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 100 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. [14] obtained for spatial rainfall data using the method of moments. Note that for both rainfall and surrogates, $\tau(q = 0) = -D_f = -1$ which implies that both signals are singular almost everywhere.

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 an average, the $\tau(q)$ for the surrogate series in Fig. 3 seems to be closer to a straight line (specially for $-1 \leqslant q \leqslant 1.5$), 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.

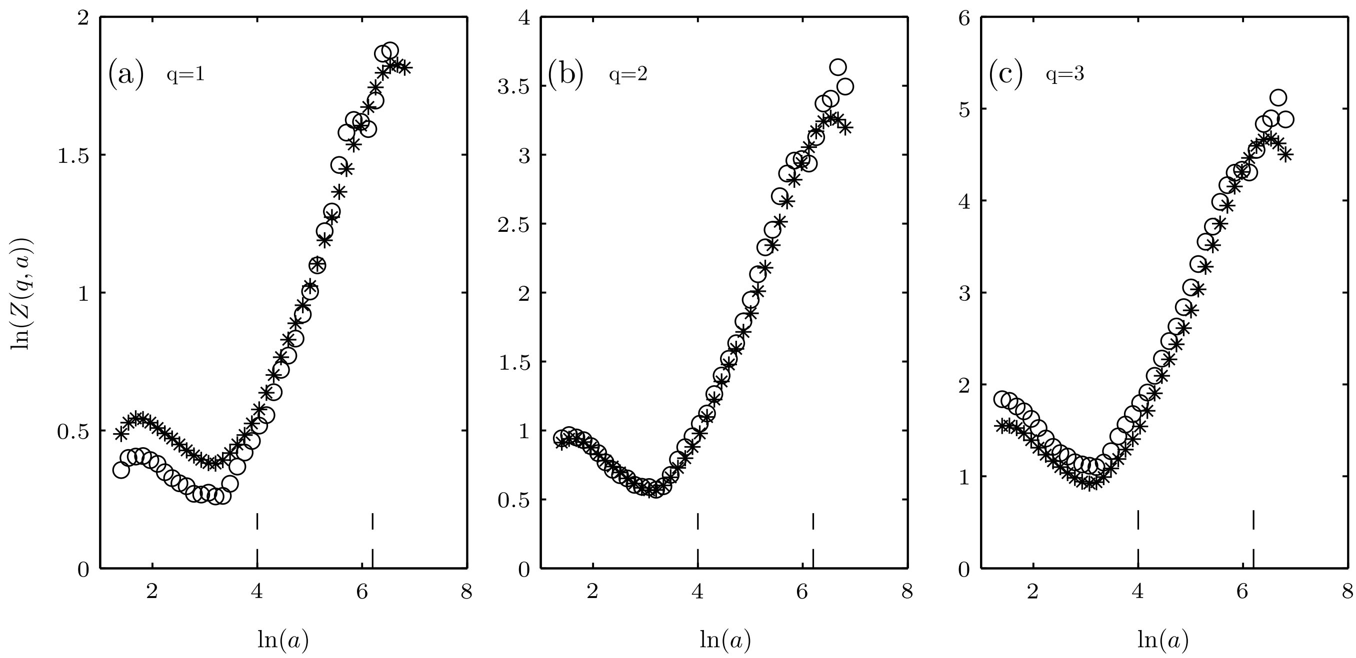

Fig. 2. Moments of the intensity of rainfall (∘) and of surrogates (∗); as explained in the text, the WTMM method was applied on the rainfall cumulative using the $g^{(3)}$ wavelet. Shown are the first, second and third moments. The estimates for the surrogates are based on 100 realizations. The vertical dashed lines delimit the range of scales (expressed in 5 s unit) used for the linear regression estimate of $\tau(q)$.

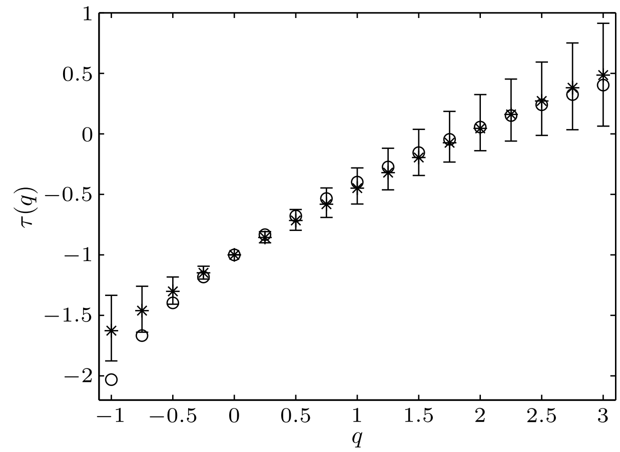

Fig. 3. $\tau(q)$ curve for rain (∘) and surrogates (*). The estimates for the surrogates are the average value for 100 realizations. Error bars represent the 5% and 95% levels for the surrogate series.

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 Fig. 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 $-c_2 = -0.26 \pm 0.04$ for rainfall and $-c_2 = -0.06 \pm 0.03$ 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). Fig. 5 shows the frequency histogram for the values of c_2 estimated from a realization of 100 surrogates. The c_2 value of 0.26 estimated from the rainfall observations is greater than the largest c_2 of 0.18 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. (Note that the calculation of the confidence intervals is done via the procedure of inverting the empirical cumulative distribution function which, in turn, is estimated using the traditional plotting position approach.)

Now consider the first order cumulant, which from Fig. 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 [38]), i.e., $\tau(q) = -1 + c_1 q - c_2 q^2/2$, then the surrogate operation, by design, preserves $\tau(2)$. In other words,

$$-1 + 2c_1^R - 2c_2^R = \tau^R(2) = \tau^S(2) = -1 + 2c_1^S - 2c_2^S, \qquad (14)$$

where the superscript R denotes the original rainfall and S the surrogates. Removing the -1 ($-D_f$) and 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 0.69 from the slope of the first order cumulant (Fig. 4a). 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 $0.69 - 0.26 = 0.43$ in our case. If we take the mean c_2^S of the surrogates as 0.06 (see Fig. 5), we get instead $c_1^S = 0.43 + 0.06 = 0.49$ (following Eq. (14)). Direct fitting of the slopes of the first order cumulant gives $c_1^S = 0.51 \pm 0.03$, 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). To reiterate, the c_2 value of rainfall being significantly higher than the maximum c_2 of the 100-surrogate ensemble indicates that the rainfall time series is not generated by an underlying linear

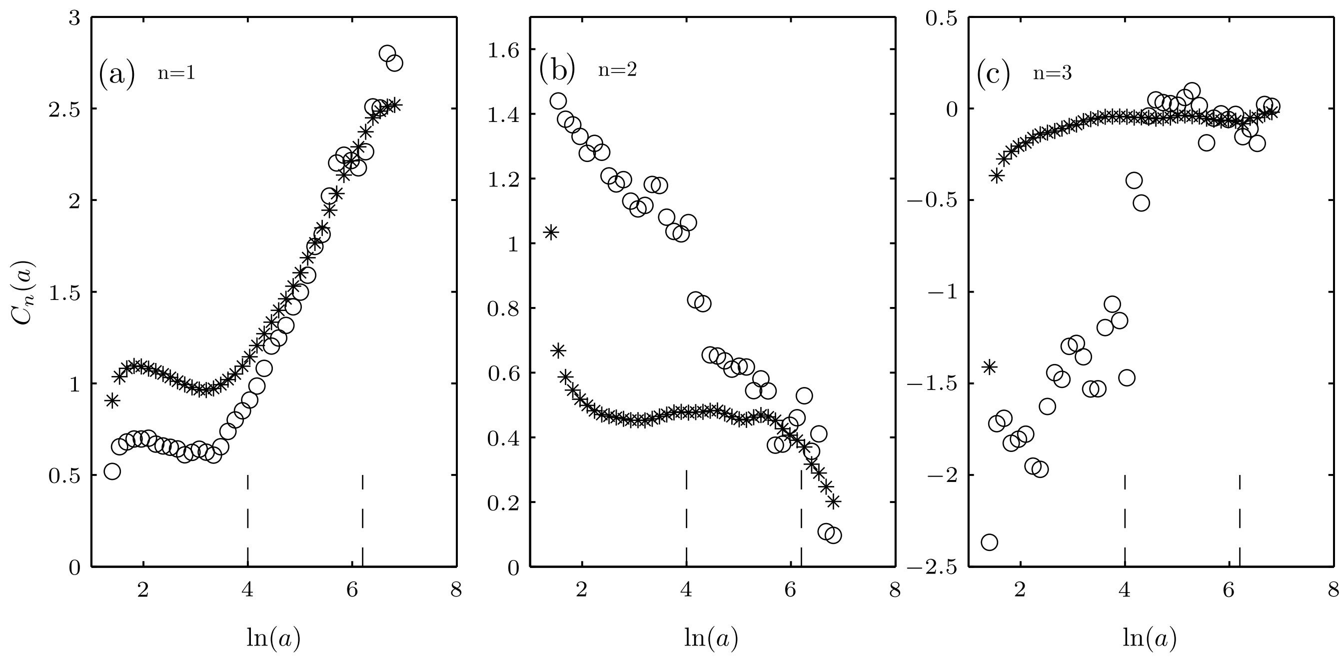

Fig. 4. Cumulants of the rainfall intensity and of surrogates. Plots (a), (b) and (c), show the first, second and third cumulants respectively, for the rainfall (∘) and surrogates (★). The estimates for the surrogates are the average value for 100 realizations. The vertical lines indicate the scaling range over which parameter estimation was performed.

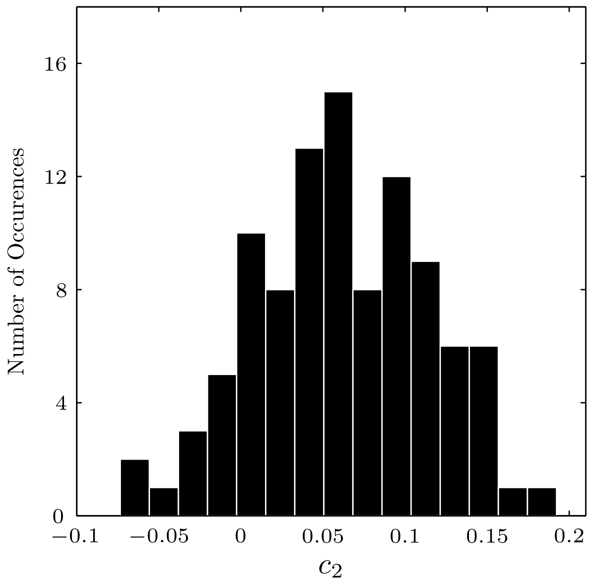

Fig. 5. Histogram of c_2 values estimated from the magnitude cumulant analysis of 100 surrogate data series.

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 $\mathscr{C}(a, \Delta t)$ are shown in Fig. 6 as a function of the logarithm of the displacement Δt. 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 −0.26 in the scaling range. Recall that Arneodo et al. [2,3] show that a multifractal cascade process on a wavelet dyadic tree leads to a slope of $\mathscr{C}(a, \Delta t)$ that is given by $-c_2$ (Eq. (12)), which is the same c_2 estimated by the cumulant analysis. So therefore this slope of −0.26, is consistent with the c_2 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 Fig. 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 −0.06. 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.

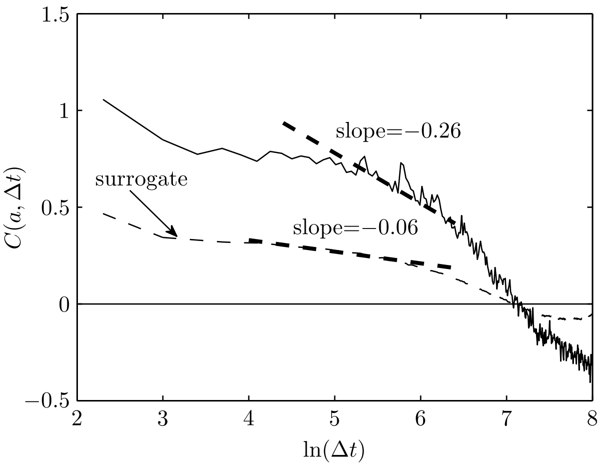

Fig. 6. Two point magnitude correlation $\mathscr{C}(a, \Delta t)$ vs. $\ln(\Delta t)$ for the rainfall series (solid line) and its surrogates (dashed curve) computed at scale $a \simeq 40$ s.

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 vs. 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 partially motivated by the previous work of Ferraris et al. [14], 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, to 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 behavior – called a linearization effect – in known models such as both purely multiplicative cascades (e.g. [21]) and fractionally integrated cascades (e.g. [32]). 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 [6]). 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. [14].

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 (e.g. [12,30]). 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.

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

[1] Arneodo A, Bacry E, Muzy JF. The thermodynamics of fractals revisited with wavelets. Physica A 1995;213:232–75.
[2] Arneodo A, Bacry E, Manneville S, Muzy JF. Analysis of random cascades using space-scale correlation functions. Phys Rev Lett 1998;80:708–11.
[3] Arneodo A, Bacry E, Muzy JF. Random cascades on wavelet dyadic trees. J Math Phys 1998;39:4142–64.
[4] Arneodo A, Manneville S, Muzy JF, Roux SG. Revealing a lognormal cascading process in turbulent velocity statistics with wavelet analysis. Phil Trans R Soc Lond A 1999;357:2415–38.
[5] Bacry E, Muzy JF, Arneodo A. Singularity spectrum of fractal signals from wavelet analysis: exact results. J Stat Phys 1993;70:635–74.
[6] Basu S, Foufoula-Georgiou E, Lashermes B, Arneodo A. Estimating intermittency exponent in neutrally stratified atmospheric surface layer flows: a robust framework based on magnitude cumulant and surrogate analyses. Phys Fluids 2007;19:115102. doi:10.1063/1.2786001.
[7] Basu S, Foufoula-Georgiou E. Detection of nonlinearity and chaoticity in time series using the transportation distance function. Phys Lett A 2002;301:413–24.
[8] Cârsteanu A, Foufoula-Georgiou E. Assessing dependence among weights in a multiplicative cascade model of temporal rainfall. J Geophys Res 1996;101(D21):26363–70.
[9] Collet P, Lebowitz JL, Porzio A. The dimension spectrum of some dynamical systems. J Stat Phys 1987;47:609–44.
[10] Daubechies I. Ten lectures on wavelets. Philadelphia: SIAM; 1993.
[11] Deidda R, Benzi R, Siccardi F. Multifractal modeling of anomalous scaling laws in rainfall. Water Resour Res 1999;35:1853–67.
[12] Deidda R. Rainfall downscaling in a space–time multifractal framework. Water Resour Res 2000;36(7):1779–84.
[13] Delour J, Muzy JF, Arneodo A. Intermittency of 1D velocity spatial profiles in turbulence: a magnitude cumulant analysis. The Euro Phys J B 2001;23:243–8.
[14] Ferraris L, Gabellani V, Parodi V, Rebora N, von Hardenberg J, Provenzale A. Revisiting multifractality in rainfall fields. J Hydrometeorol 2003;4:544–51.
[15] Georgakakos KP, Cârsteanu AA, Sturdevant PL, Cramer JA. Observation and analysis of midwestern rain rates. J Appl Meteorol 1994;33:1433–44.
[16] Gupta VK, Waymire E. Multiscaling properties of spatial rainfall and river flow distributions. J Geophys Res 1990;95(D3):1999–2009.
[17] Gupta VK, Waymire E. A statistical analysis of mesoscale rainfall as a random cascade. J Appl Meteorol 1993;32(2):251–67.
[18] Halsey TC, Jenson MH, Kadanoff LP, Procaccia I, Shraiman BI. Fractal measures and their singularities: the characterization of strange sets. Phys Rev A 1986;33:1141–51.
[19] Holschneider M. L'analyse d'objets fractals et leur transformation en ondelettes, Ph.D. thesis, Univ. of Aix-Marseille II, France; 1989.
[20] Lashermes B, Abry P, Chanais P. New insights into the estimation of scaling exponents. Int J Wavelets Multiresolut Info Proc 2004;2:297–523.
[21] Lashermes B. 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; 2005.
[22] Lovejoy S, Mandelbrot BB. Fractal properties of rain, and a fractal model. Tellus 1985;37A(3):209–32.
[23] Lovejoy S, Schertzer D. Multifractal analysis techniques and the rain and cloud fields from 10^{-3} to 10^{6} m. In: Schertzer D, Lovejoy S, editors. Non-linear variability in geophysics: scaling and fractals. Norwell, MA: Kluwer Acad; 1991. p. 111–44.
[24] Mallat S. A wavelet tour in signal processing. New York: Academic Press; 1998.
[25] Meyer Y. Wavelet and applications. Berlin: Springer; 1992.
[26] Muzy JF, Bacry E, Arneodo A. Wavelet and multifractal formalism for singular signal: application to turbulence data. Phys Rev Lett 1991;67:3515–8.
[27] Muzy JF, Bacry E, Arneodo A. Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method. Phys Rev E 1993;47:875–84.
[28] Muzy JF, Bacry E, Arneodo A. The multifractal formalism revisited with wavelets. Int J Bifurc Chaos 1994;4:245–302.
[29] Parisi G, Frisch U. On the singularity structure of fully developed turbulence. In: Ghil M, Benzi R, Parisi G, editors. Turbulence and predictability in geophysical fluid dynamics. Amsterdam: North-Holland; 1985. p. 84–8.
[30] Rebora N, Ferraris L, von Hardenberg J, Provenzale A. RainFARM: rainfall downscaling by a filtered autoregressive model. J Hydrometeorol 2006;7(4):724–38.
[31] Sapozhnikov V, Foufoula-Georgiou E. An exponential Langevin-type model for rainfall exhibiting spatial and temporal scaling. In: Tsonis A, Elsner J, editors. Nonlinear dynamics in geosciences. New York: Springer-Verlag; 2007.
[32] Schertzer D, Lovejoy S, Lavallée D. Generic multifractal phase transitions and self-organized criticality. In: Perdang JM, Lejeune A, editors. Cellular automata: prospects in astrophysical applications. World Scientific; 1993. p. 216–27.
[33] Schertzer D, Lovejoy S. Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. J Geophys Res 1987;92(D8):9693–714.
[34] Schertzer D, Lovejoy S. Hard and soft multifractal processes. Phys A 1992;185:187–94.
[35] Schreiber T, Schmitz A. Improved surrogate data for nonlinearity tests. Phys Rev Lett 1996;77:635–8.
[36] Theiler J, Eubank S, Longtin A, Galdikian B, Farmer JD. Testing for nonlinearity in time series: the method of surrogate data. Phys D 1992;58:77–94.
[37] Veneziano D, Bras RL, Niemann JD. Nonlinearity and self-similarity of rainfall in time and a stochastic model. J Geophys Res 1996;101(D21):26371–92.
[38] Venugopal V, Roux SG, Foufoula-Georgiou E, Arneodo A. Revisiting multifractality of high-resolution temporal rainfall data using a wavelet-based formalism. Water Resour Res 2006;42. doi:W06D1410.1029/2005WR004489.