Stoch Environ Res Risk Assess
DOI 10.1007/s00477-016-1343-2

ORIGINAL PAPER

Deterministic simulation of highly intermittent hydrologic time series

Mahesh L. Maskey[1] · Carlos E. Puente[1] · Bellie Sivakumar[1,2]

Abstract Application of a deterministic geometric approach for the simulation of highly intermittent hydrologic data is presented. Specifically, adaptations of the fractal-multifractal (FM) method and a Cantorian extension are advanced in order to simulate rainfall records measured at the daily scale and encompassing a water year. It is shown, using as case studies 2 years of rainfall sets gathered in Laikakota, Bolivia and Tinkham, Washington, USA, that the FM approach, relying on only at most 8 parameters, is capable of closely preserving either the whole record's histogram (therefore including moments), the whole data's Rényi entropy function and/or the maximum number of consecutive zero values present in the sets, resulting in suitable rainfall simulations, whose overall features and textures are similar to those of the observed sets. The study hence establishes the possibility of simulating highly intermittent sets in time in a deterministic and holistic way as a novel parsimonious methodology to supplement available stochastic frameworks.

Keywords Rainfall · Simulation · Intermittency · Fractals · Fractal-multifractal approach · Inverse problem · Particle swarm optimization

Abbreviations

NSH	Nash–Sutcliffe statistic for histograms
PS90	Percent histogram mass in simulated sets corresponding to 90% in observed data
NZR	Number of zeroes in observed and (simulated) rainfall sets
EZR	Percent error of zeros preserved by FM simulation
NSE	Nash–Sutcliffe statistic for entropies
MCZ	Maximum consecutive zeros in records and (simulated) sets
NSA	Nash–Sutcliffe statistic for autocorrelations

1 Introduction

As rainfall exhibits considerable spatial and temporal variability, its proper modeling is a tremendously challenging task. Apart from the physically-based methodologies that approximate the physics of the process using mass and energy conservation laws coupled with cloud microphysics, the procedures most employed to simulate rainfall may be classified into two categories: (i) those based on classical stochastic notions such as moments of the probability distributions of rain, and (ii) those that use, in addition, chaotic, fractal and multifractal features in the records.

In the first category, there are, for instance, various kinds of models based on: (a) Poisson or other point processes (e.g. Neyman–Scott, Bartlett-Lewis) coupled with instantaneous or rectangular rain pulses (e.g. Eagleson 1978; Kavvas and Delluer 1981; Rodríguez-Iturbe et al. 1987; Vandenberghe et al. 2011; Kaczmarska et al. 2014), and (b) Markov chain representations that account for some of dynamics of the process (e.g. Haan et al. 1976; Geng et al. 1986; Jothityangkoon et al. 2000; Garbrecht and Zhang 2012). These approaches, given their explicit assumptions and approximations, turn out to be capable of

✉ Carlos E. Puente
cepuente@ucdavis.edu

[1] Department of Land, Air & Water Resources, University of California, Davis, CA 95616, USA

[2] School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

Published online: 15 November 2016

preserving distinct, but not all, (statistical) features of the records.

In the second category, there are, for example, models based on: (a) chaotic representations of rainfall (e.g. Rodriguez-Iturbe et al. 1989; Tsonis et al. 1993; Sivakumar 2000; Sivakumar et al. 2001), (b) multiplicative cascades that progressively divide energies and masses (e.g. Lovejoy and Schertzer 1985, 1990; Schertzer and Lovejoy 1987; Gupta and Waymire 1993; Tessier et al. 1993; Veneziano and Langousis 2005, 2010; Veneziano et al. 2006, 2009; Langousis and Veneziano 2007; Langousis et al. 2009, 2013), and (c) fractal-multifractal constructs that represent observations as a fractal transformation of a turbulence-related multifractal input (e.g. Puente 1996; Puente and Obregón 1996; Obregón et al. 2002a, b; Cortis et al. 2009, 2013; Huang et al. 2012; Maskey et al. 2015). Although all of these notions do provide reasonably-looking rainfall outcomes that preserve some statistical/morphological characteristics of the records, the ones based on fractal-multifractal constructs, unlike those using chaotic and stochastic notions, may be more easily conditioned to mimic specific realizations of rainfall patterns. This feature of such ideas provides the motivation of the present study, which is to further explore the capabilities of the so-called fractal-multifractal FM approach, not as a technique for encoding information, but as a mechanism to define sensible simulations (based on appropriate statistical information) of the highly intermittent rainfall sets available when recording the process, say at the daily scale over a year.

Specifically, here we demonstrate, for the first time, that the original FM approach (Puente 1996) and its Cantorian extension (Huang et al. 2013; Maskey et al. 2015), coupled with suitable adaptations that allow studying highly-intermittent daily rainfall sets, may be used to easily obtain rainfall simulations that preserve, not only whole general statistics of the records (complete histogram and entropy functions, distribution of zeroes, etc.) but also the overall implicit complexity (and textures) of rainfall sets, hence establishing that these notions do supplement other simulation algorithms.

The organization of this paper is as follows. Section 2 reviews the original fractal-multifractal approach, the Cantorian extension, and adaptations used in order to generate highly intermittent sets. Section 3 explains the strategy used in order to find suitable rainfall simulations via appropriate inverse problems that aim to fit relevant statistical information. Section 4 presents various simulations obtained fitting histograms, entropies and consecutive zeros in the observed sets using as case studies a year of records each as gathered in Laikakota, Bolivia and Tinkham, Washington State, USA, including a complete statistical analysis of the results. Finally, Sect. 5 summarizes the results and presents the conclusions of the research.

2 Fractal multifractal approach

2.1 The original approach

Based on the notion of fractal interpolation functions and their related invariance multifractal measures (Barnsley 1988; Puente 1996), the so-called fractal-multifractal (FM) approach is defined by computing projections of such measures over the range of the functions, which, end up possessing textures and shapes that resemble those encountered in a variety of applications (Puente 2004). The construction relies on the iteration of simple affine mappings (via the so-called 'chaos game') defined such that the unique attractor of the iterations—that is, the invariant set—passes by a pre-specified set of points in the plane.

To illustrate the ideas, consider the two affine maps

$$w_1\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.44 & 0 \\ 1.11 & -0.58 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \quad (1)$$

$$w_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.56 & 0 \\ 1.08 & -0.64 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0.46 \\ 0.56 \end{pmatrix}, \quad (2)$$

associated with the interpolating points {(0,0), (0.44, 0.54), (1,1)}, and iterated, say, 2^{14} times such that w_1 is used 51% of the time and w_2 the reminder 49%, following the independent tosses of a biased coin. As seen in Fig. 1, these iterations induce a fractal function f, having a fractal dimension $D \approx 1.26$ (top left) and two unique invariant measures over the two coordinates: dx (below) over the x-axis and dy (to the right) over the y-axis. While dx comes from a multiplicative cascade with length scales of 0.44 and 0.56 and redistributions 0.51 and 0.49 (Mandelbrot 1989) and exhibits a rather repetitive appearance, the derived measure dy, found considering the input textures on crossing over y, gives rise to a highly intermittent set reflecting the geometry of the fractal function, which, resembles, for instance (after a 90 degree flip), daily rainfall time series as encountered in nature.

Mathematically, a fractal interpolating function f: $x \to y$, passing through $N+1$ points in x, $\{(x_n, y_n)|x_0 < x_1 \ldots < x_N\}$ is defined such that its graph $G = \{(x, f(x)|x \in [0,1]\}$, having a fractal dimension $1 \le D < 2$, is the unique fixed point of N simple maps of the form:

$$w_n\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_n & 0 \\ c_n & d_n \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_n \\ f_n \end{pmatrix}, \; n = 1, \ldots, N, \quad (3)$$

that is, $G = w_1(G) \cup w_2(G) \cup \cdots \cup w_N(G)$.

In these equations, the coefficients d_n are the vertical scaling parameters, $|d_n| < 1$, and the other entries in the matrices: a_n, c_n, e_n and f_n are defined based on the scalings and the interpolating points based on initial conditions:

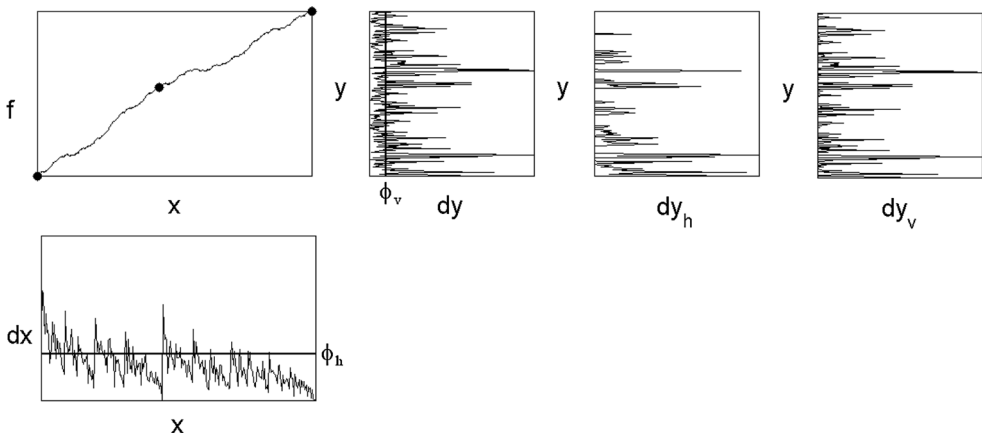

Fig. 1 The original FM approach: from an multinomial multifractal dx, to a derived projection dy, via a fractal interpolating function from x to y, followed by two adaptations: a projection dy_h, found pruning dx below a horizontal threshold ϕ_h, and a projection dy_v found pruning dy below a vertical threshold ϕ_v

$$w_n\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix},\ w_n\begin{pmatrix} x_N \\ y_N \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \tag{4}$$

leading to two simple sets of simultaneous linear equations per map (Barnsley 1988).

Requiring little effort in the implementation, which simply provides a collection of (x, y) points via arbitrary iterations guided by a pseudo-random "coin"—as included in the Matlab code in the Online Appendix http://puente.lawr.ucdavis.edu/omake/serraSimul2016/serraSimul2016.html, the computed histograms over x and y become stable for large enough numbers of iterations (2^{14}) and yield an input multifractal histogram dx that gets automatically transformed via a fractal interpolating function f (found plotting the x vs. y values) into an output (derived) histogram dy.

As shown in the figure, adjacent to dy, there are two other graphs, named dy_h and dy_v, that exhibit similarities with dy, but that contain distinct and additional zero values when seen from the y (time) axis. Such graphs also resemble the intermittencies encountered in rainfall, including high peaks and number of zeros, and are defined as follows. On the one hand, dy_h is found by considering a horizontal threshold, ϕ_h, as shown superimposed on the input dx, and transforming only the mass above it via the same function f. Such a derived measure (shown normalized, so that its total mass is one unit) may then be interpreted as the output produced by sufficiently large eddies (in x), as it happens when rainfall is formed aided by turbulence. On the other hand, the set dy_v (shown on the far right) is obtained by using a vertical threshold, ϕ_v, (seen superimposed to dy) below which the output becomes zero. As shown normalized, this set reflects what is customarily done in practice when "traces" of rainfall less than a value, and in consonance with the inherit accuracy of the measurements, are neglected.

At the end, both kinds of sets dy_h and dy_v represent the adaptations of the FM approach that shall be used later on in order to simulate complex rainfall records gathered at the daily scale. As a summary, the implied derived measures depend on the following parameters: (a) the interpolating points by which a fractal interpolating function passes, (b) the vertical scalings d_n, (c) the frequencies used to do the iterations, and (d) a horizontal or vertical threshold that defines zero-rain values. If the first and last interpolating points are fixed, as done herein using the points (0,0) and (1,1) as in the example shown in Fig. 1, the total number of parameters for an FM representation based on two affine maps is 6. It ought to be noted that, as dx and dy are found as histograms of iterations, dy_v is more easily found than dy_h, for the latter requires pruning the chaos game and, hence, is computationally a more expensive procedure. In order to aid the reader fully appreciate the FM approach, the aforementioned Online Appendix includes an interactive demo based on Fig. 1, which allows seeing the geometric effects of parameter variations when employing two maps and the notion of a vertical threshold.

2.2 Extension to a Cantorian representation

Instead of using affine maps that induce a fractal interpolating function, it is possible to modify the approach so that the iterations generate more general "attractors" that explicitly contain periods of inactivity (Huang et al. 2013). Such may be done using N maps as in Eq. (3), but using new initial conditions:

$$w_n\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_{2n} \\ y_{2n} \end{pmatrix},\quad w_n\begin{pmatrix} x_{2N-1} \\ y_{2N-1} \end{pmatrix} = \begin{pmatrix} x_{2n+1} \\ y_{2n+1} \end{pmatrix},\quad n = 1, \ldots, N, \tag{5}$$

such that the end-points $\{(x_{2n}, y_{2n}), (x_{2n+1}, y_{2n+1})\}$ that define the range of map w_n, contain gaps over x, yielding hence a Cantor set (Mandelbrot 1982). The representations having gaps over x on their mapping end-points give rise to Cantorian attractors (Huang et al. 2013).

As an example, Fig. 2 shows what is found while using two affine maps, with domains from 0 to 1, whose end-points are $\{(0,0),(0.09,-2.64)\}$ and $\{(0.56,-2.13),(1,1)\}$ and corresponding to the maps:

$$w_1\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0.09 & 0\\ -2.69 & 0.05\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}0\\0\end{pmatrix} \quad (6)$$

$$w_2\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0.44 & 0\\ 2.41 & 0.72\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}0.56\\ -2.13\end{pmatrix}, \quad (7)$$

which are iterated, once again 2^{14} times, following a biased "coin" leading to percent usage of w_1 and w_2 of 27 and 73%, respectively.

As seen, this representation, containing in its attractor a noticeable gap from 0.09 to 0.56 and yielding an attractor with fractal dimension $D \approx 0.72$, produces yet an interesting projection dy that exhibits high intermittency and contains many zero values. As described before, adjacent to dy, there are two plausible adaptations of the idea via sets dy_h and dy_v that correspond to the horizontal and vertical thresholds explained earlier. As before, dy_v is more easily computed than dy_h, for the latter requires pruning the iterations.

At the end, this Cantorian extension also produces normalized derived measures over y that depend on the following parameters: (a) the end-points associated with the affine maps, (b) the vertical scalings d_n, (c) the frequencies used on the iterations, and (d) a horizontal or vertical threshold. If the first point of the first end-point and second point on the last end-point are fixed, as done herein using the points (0,0) and (1,1) as in the example shown in Fig. 2, the total number of parameters for an FM representation based on two overlapping affine maps is 8. For more information on this Cantorian extension, the interested reader may access the Online Appendix http://puente.lawr.ucdavis.edu/omake/serraSimul2016/serraSimul2016.html, which also includes a Matlab code for such a generalization and an interactive demo for the notions based on the pattern in Fig. 2.

3 Simulation strategy

As elaborated in Sect. 2, both FM approaches, the original methodology and a generalization yielding Cantorian attractors, may produce highly intermittent sets that mimic rainfall patterns gathered daily and over a water year. In the reminder of this article, it is shown how the FM approaches may be used to faithfully simulate those kinds of rainfall records by solving a suitable inverse problem that defines FM parameters based on close fits of appropriate statistical information.

Although an FM pattern may be computed in hardly any time once the parameters are known, the inverse problem for finding a sought simulation is rather complex. This is, in part, due to the dimensionality of the parameter space (six dimensions for the original approach with a threshold and eight parameters for Cantorian representation and a threshold, for models having two affine maps) and because there may be several possible solutions. Based on our experience in dealing with this issue (Huang, et al. 2013; Maskey et al. 2015), the generalized particle swarm optimization algorithm of Fernández Martínez et al. (2010) is employed here.

As highly intermittent sets typically exhibit very little correlation, the strategy adopted for finding the simulations herein, of sets that indeed have rather small autocorrelations, is based on trying to capture two other relevant statistics, namely, histograms and entropy functions of the records. As such, the objective functions to be minimized are the root mean square errors of such attributes, given by:

$$\in_{HIS}=\sqrt{\frac{1}{10}\sum_{i=1}^{10}\left(h_i-\hat{h}_i\right)^2} \quad (8)$$

for histograms, where h_i and $\hat{h}_i$ are, respectively, the ith observed and FM-fitted histograms of the highly intermittent time series when using 10 bins, extending from the minimum to the maximum of the observed records, while

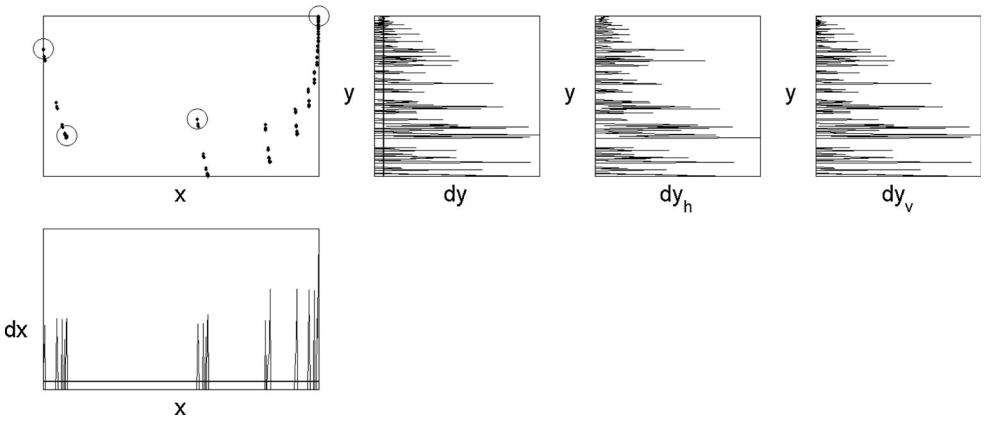

Fig. 2 A generalized FM approach: from a Cantorian texture dx, to a projection dy, via a disperse attractor from x to y, followed by two adaptations: a projection dy_h, found pruning dx below a horizontal threshold ϕ_h, and a projection dy_v found pruning dy below a vertical threshold ϕ_v

leaving FM extreme values beyond the records' maximum (if any) in the largest bin; and

$$\in_{ENTRO} = \sqrt{\frac{1}{n_e}\sum_{i=1}^{n_e}(e_i - \hat{e}_i)^2} \quad (9)$$

for entropy functions, where e_i and $\hat{e}_i$ are, respectively, the ith observed and FM-fitted Rényi entropy values corresponding to a generic exponent weight q_i, as defined by $e_i = E(q_i) = 1/(1-q_i)\log\left(\sum_{j=1}^{M} p_j^{q_i}\right)$, for q_i between 0.1 and 5.1, and in increments of 0.1 for a total of $n_e = 51$ values, where p_j is the jth rainfall (observed or FM-simulated) value and M is the total number of days in the records.

The Rényi entropy function herein is limited to positive exponents due to the presence of zero rainfall values. Such a function, which generalizes the classic concept of Shannon's entropy (when q_i tends to one), qualifies the overall roughness of the records as probed by distinct exponents that progressively amplify (as q_i. increases) large rainfall values, as it is also similarly done while computing multifractal features (Feder 1988).

In order to ensure that solutions would possess similarly-looking intermittencies, these previous objective functions $\in_{HIS}$ and $\in_{ENTRO}$ are added penalties regarding the number of zero values on the records, such that FM simulations would not have periods of inactivity that differ from observed values by more than 10%. As shall be seen, the results that follow end up satisfying such restrictions.

In addition to optimizing histograms and entropies, and as the number of consecutive zeros plays a critical role in visualizing water scarcity, results are also advanced on an attempt to capture such a feature. Specifically, this is achieved by trying to preserve the maximum number of consecutive days of no rain (setting a penalty beyond a 5-day error), while keeping the entropy and the aforementioned total number of zeros of the records and FM simulations both at the 10% range.

Once optimal simulations are identified, this study reports the Nash-Sutcliff efficiencies for histograms and entropies, values that are not considered explicitly in the optimization exercise and that serve to further qualify the goodness of the results.

4 Results and discussion

This section shows several plausible deterministic FM simulations of highly intermittent rainfall records based on those gathered in Laikakota, Bolivia (water year 1965–1966, starting September 1st for 273 days, excluding the dry season of June to August) and in Tinkham Creek, Washington, USA (National Resource Conservation Service, Station ID 899, water year 2000–2001, starting October 1st for 365 days), which were selected due to their distinct climates, clear intermittencies, and their varied distributions of zero values. The simulated sets are obtained following the strategy explained in Sect. 3, namely fitting either the records' (a) histogram—Eq. (8), (b) entropy function—Eq. (9), and (c) the maximum number of consecutive zeros, with due penalties. The patterns shown are chosen from several other close "solutions" that are not difficult to find once the optimization exercise runs are completed, based on distinct initial parameter values.

Prior to testing the notions, the observed rainfall data sets are normalized so that they integrate to one (just dividing by the sum of all values in the season). In what follows, simulations based on the classical FM approach using two maps (six parameters as in Sect. 2.1) and the extended FM method based on two maps with gaps (8 parameters as in Sect. 2.2) are shown. As a consequence, the results correspond to simulations with compression ratios (the total number of data points over the number of FM parameters) that are at least 34:1 for Laikokota and 46:1 for Tinkham. Appendix 1 includes the FM parameters corresponding to the simulations chosen for both sites.

4.1 Laikakota rainfall

Figures 3, 4 and 5 present plausible rainfall simulations at the Laikakota site, based on the three aforementioned objective functions, namely fitting histograms (Fig. 3), preserving entropy (Fig. 4) and capturing maximum consecutive zeros (Fig. 5).

As is seen, all figures share four columns, with the first one being data (real or simulations), and the other three including histograms, entropies and autocorrelations. To illustrate that both the original FM approach and the Cantorian extension do yield sensible results, the figures include two simulations based on each one of the FM procedures while using a horizontal and one a vertical threshold. Those simulations labeled A and B are based on the original FM approach (for thresholds ϕ_h and ϕ_v, respectively) and those named C and D come from the Cantorian extension (for thresholds ϕ_h and ϕ_v, respectively). While statistical information for the real data set is reported in black, the one corresponding to the simulations is shown in gray. For illustration purposes, notice that simulation labeled B on Fig. 3 corresponds to the construction shown in Fig. 1, i.e. based on the original FM method and the use of a vertical threshold.

As seen, all the simulations in all figures (shown such that the largest peak on the page is one unit) represent well the intrinsic complexity of the daily rainfall records at

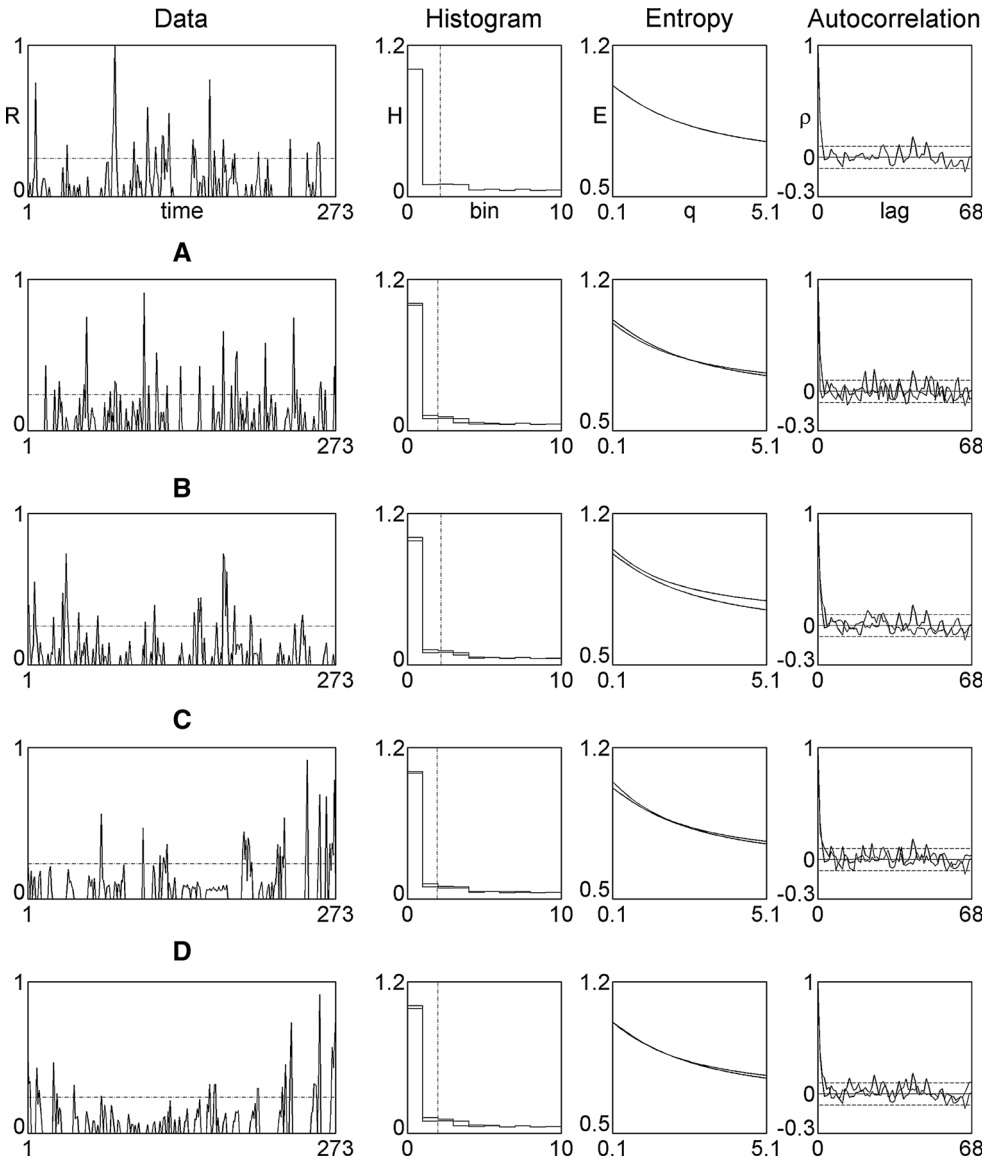

Fig. 3 A measured set of daily rainfall at Laikakota, Bolivia gathered from September 1965 to May 1966 (*top*), followed by its histogram, Rényi entropy and autocorrelation functions; and four suitable rainfall simulations (*bottom*), based on fits of the records' histogram. While simulations **A** and **B** emanate from fractal wires, **C** and **D** come from Cantorian attractors. While sets **A** and **C** use a horizontal threshold, **B** and **D** employ a vertical one

Laikakota (assumed to be gathered without error): (a) all sets include similar intermittencies, as reflected by two or three major peaks and at least six notorious intervals of no rain, (b) all simulations have similar histograms concentrated on zero values (with the first bin giving a value of one for the observed records), (c) all sets have entropy functions that decay monotonically and similarly to that of the records, and (d) all autocorrelation functions decay rather quickly towards zero and remain rather small, as reflected by 10% statistical bands and as it happens to the records.

As observed looking at Figs. 3 and 4, the objective functions based on histograms and entropy functions are faithfully preserved in the shown simulations. For instance, Fig. 3 reveals that all four histogram approximations (in gray) based on the objective function given by Eq. (8)—with fulfilled restrictions as mentioned in Sect. 3—are rather close to that of the Laikakota set (in black) for the ten bins considered. Similarly, Fig. 4 shows that all four entropy functions, obtained minimizing Eq. (9) with due penalties, fully preserve the record's entropy.

Specifically and with the aid of Table 1, the simulations and statistics included in these figures may be summarized as follows. Regarding Fig. 3, the histogram fits are indeed excellent irrespective of Cantorian or non-Cantorian representations and of the usage of horizontal or vertical thresholds. This is reflected, in the top section of the table, by the very high values of the Nash–Sutcliffe statistic for the histogram (NSH), the close agreement on the histograms of the simulated sets when compared to the records at the 90th percentile (PS90), the closeness in the number of zero values in the simulations as compared to the actual set (NZR), as implied by small percent errors (EZR) that do not exceed 10% in magnitude.

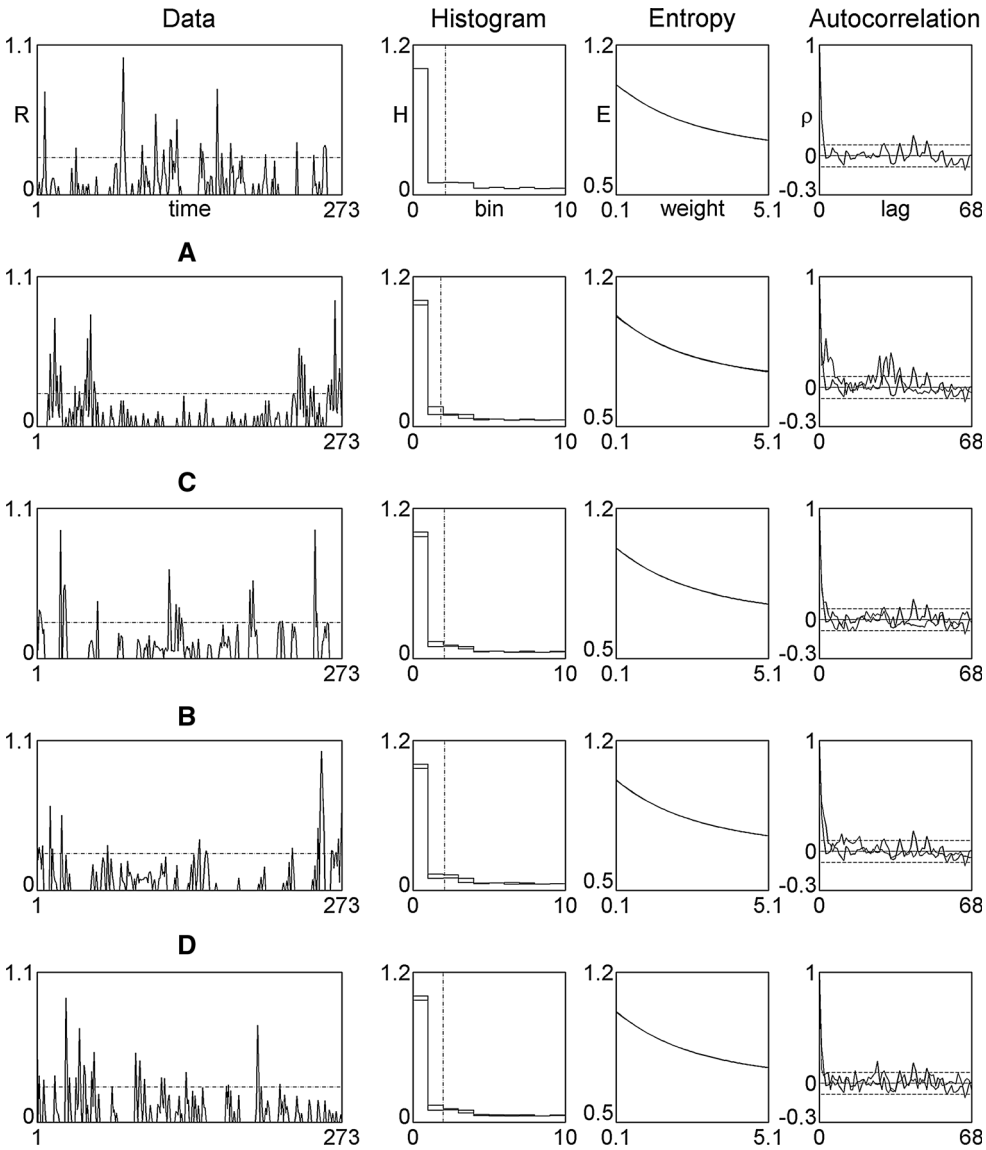

Fig. 4 A measured set of daily rainfall at Laikakota, Bolivia gathered from September 1965 to May 1966 (*top*), followed by its histogram, Rényi entropy and autocorrelation functions; and four suitable rainfall simulations (*bottom*), based on fits of the records' entropy function. While simulations **A** and **B** emanate from fractal wires, **C** and **D** come from Cantorian attractors. While sets **A** and **C** use a horizontal threshold, **B** and **D** employ a vertical one

Although the four FM simulations included in Fig. 3 (based on optimizing the record's histogram) exhibit a varying number of major peaks and diverse placements of such peaks, it is interesting to note that those labeled A, C, and D result in rather similar entropy functions (i.e. high NSE in Table 1), even though such an attribute is not optimized. Also observe, from Table 1, that the maximum number of consecutive days of zero rain (MCZ) (also not optimized but included as a penalty) are closely preserved by these simulations, as such a quantity varies from the observed 14 days by only two days, at the most. Finally, note that although the autocorrelation function of the simulations and of the Laikakota records do decay rather fast, they lead to poor (even negative) Nash–Sutcliffe values (NSA), which are due to the obvious differences between the functions even if they are related to statistically insignificant low values.

In relation to the rainfall simulations shown in Fig. 4 (i.e. based on the preservation of entropy), all the simulations, for various FM representations and thresholds, almost perfectly preserve the optimized entropy function—as reflected by the very high NSE values on the second block of Table 1. As the deviation in the number of zeros between the record and the FM simulations is used as a penalty, and as such a feature dominates the histograms, all the simulations in this category turn out to preserve very well the whole histogram of the records. This is reflected in the second block of Table 1 by the very high values of NSH, close agreement in PS90 (i.e. over 90%), and errors less than 2.9% in the number of zeros in the simulations relative to those of the Laikakota record. Overall, this set of FM simulations also exhibits various highly-intermittent geometries with similar numbers of peaks and zeros that cannot be discriminated by the naked eye. This is

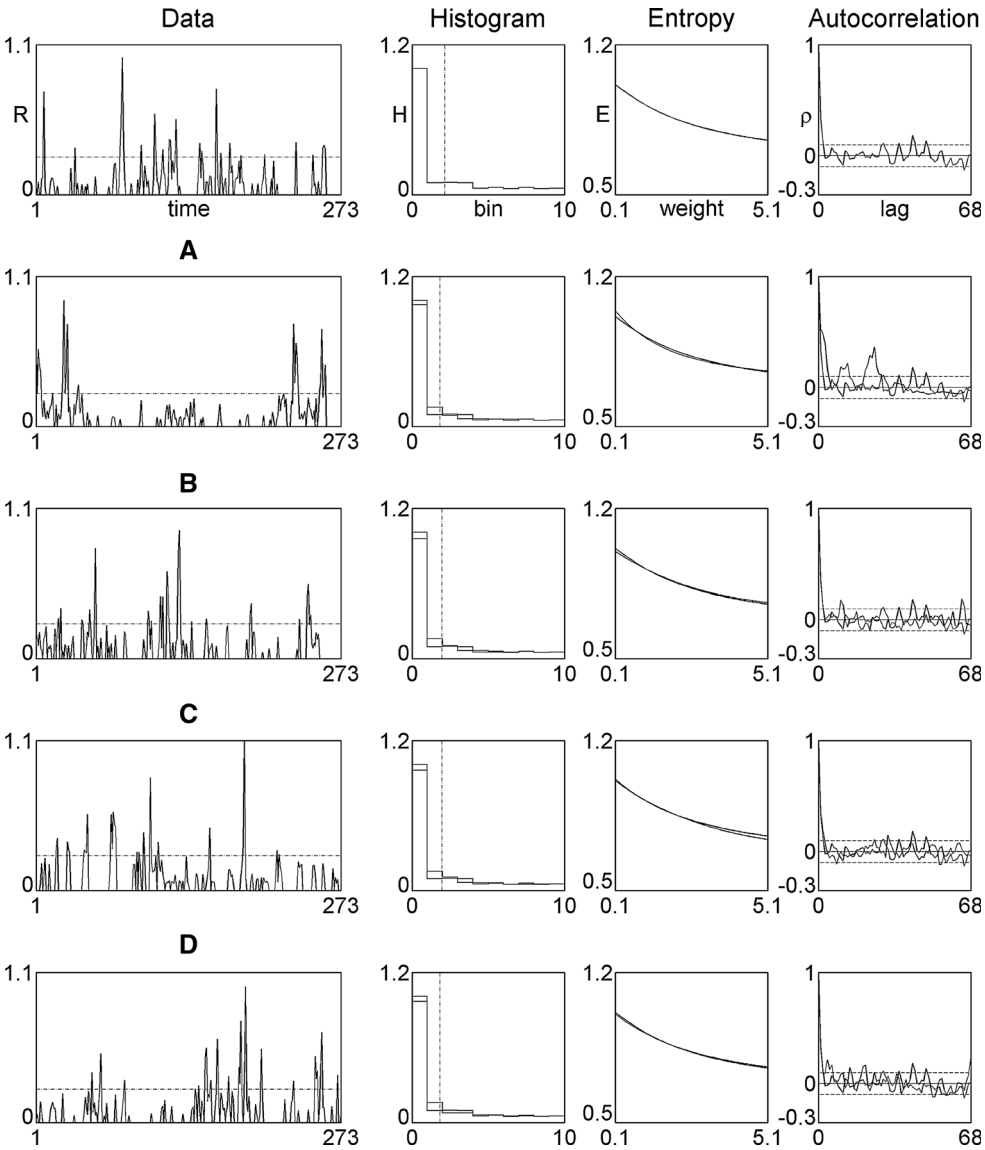

Fig. 5 A measured set of daily rainfall at Laikakota, Bolivia gathered from September 1965 to May 1966 (*top*), followed by its histogram, Rényi entropy and autocorrelation functions; and four suitable rainfall simulations (*bottom*), based on fits of the records' distribution of zeros. While simulations **A** and **B** emanate from fractal wires, **C** and **D** come from Cantorian attractors. While sets **A** and **C** use a horizontal threshold, **B** and **D** employ a vertical one

Table 1 Error statistics associated with Figs. 3, 4 and 5 for rainfall records at Laikakota, Bolivia, based on preservation of histogram (top block), entropy (middle block) and consecutive zeros (bottom block)

Simulation	NSH (%)	PS90 (%)	NZR	EZR (%)	NSE (%)	MCZ	NSA (%)
3A	99.7	90.9	174 (181)	−4.0	98.0	14 (14)	−118.2
3B	99.8	89.8	174 (162)	6.9	81.2	14 (12)	−10.4
3C	99.9	90.9	174 (157)	9.8	98.2	14 (16)	−58.9
3D	99.6	90.9	174 (174)	0.0	99.1	14 (15)	−26.9
4A	99.2	92.1	174 (169)	2.9	99.9	14 (10)	−42.4
4B	99.5	90.3	174 (170)	2.3	100.0	14 (17)	−0.9
4C	99.6	90.5	174 (172)	1.1	100.0	14 (18)	−64.4
4D	99.6	90.8	174 (171)	1.7	100.0	14 (16)	−135.3
5A	99.3	92.1	174 (156)	10.3	99.0	14 (15)	−6.7
5B	99.0	91.2	174 (181)	−4.0	99.3	14 (18)	−59.1
5C	99.2	91.2	174 (168)	3.4	99.0	14 (17)	−34.7
5D	99.1	91.9	174 (177)	−1.7	99.6	14 (17)	−10.3

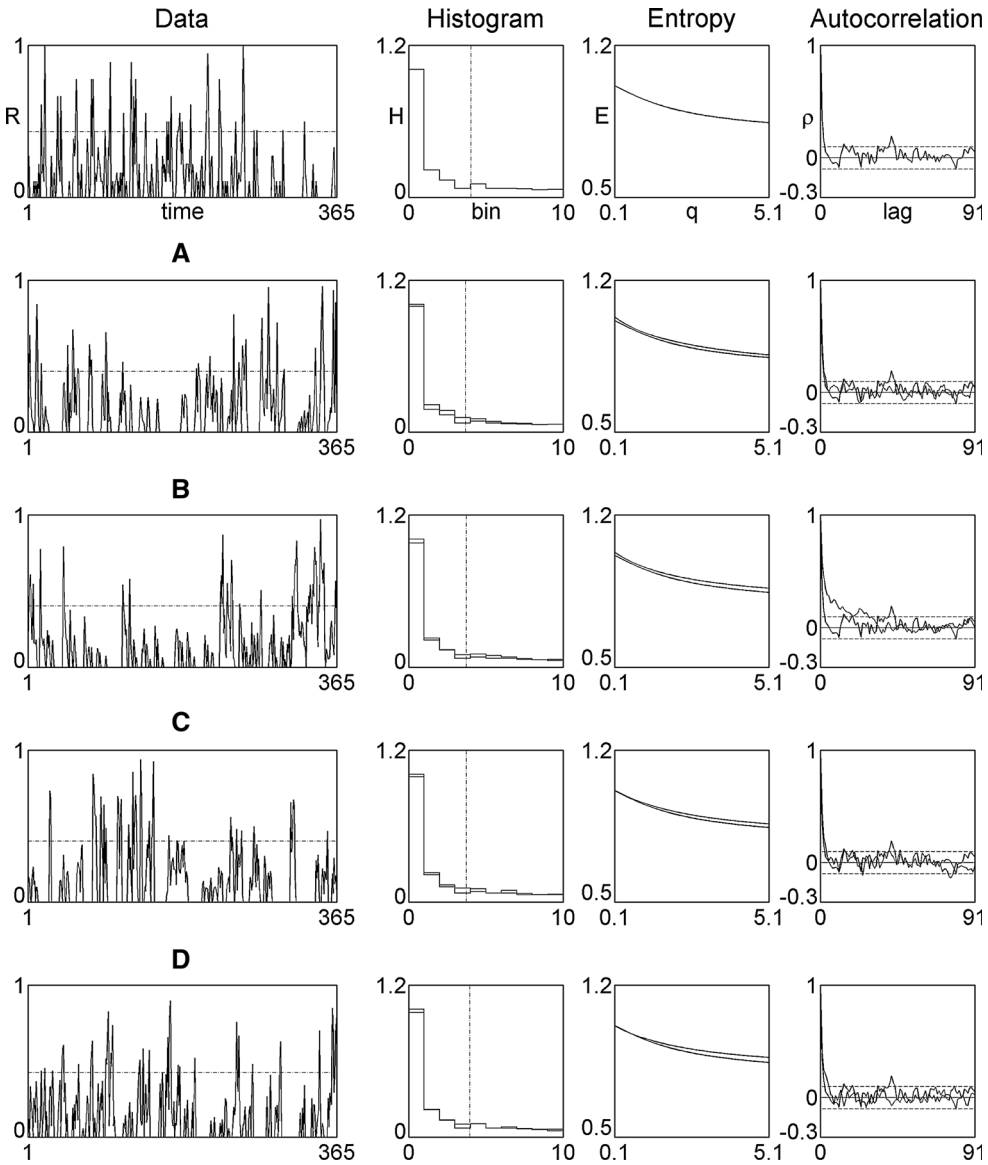

Fig. 6 A measured set of daily rainfall at Tinkham Creek, Washington gathered from October 2000 to September 2001 (*top*), followed by its histogram, Rényi entropy and autocorrelation functions; and four suitable rainfall simulations (*bottom*), based on fits of the records' histogram. While simulations **A** and **B** emanate from fractal wires, **C** and **D** come from Cantorian attractors. While sets **A** and **C** use a horizontal threshold, **B** and **D** employ a vertical one

particularly the case for simulations B, C, and D that exhibit quick decays in autocorrelation and values close to zero, unlike the somewhat "denser" simulation A, whose autocorrelation goes to zero much slower and includes a noticeable block of values above 0.25 for lags 27–32. In any case, it is indeed easy to find solutions of the inverse problem optimizing entropy and using the aforementioned penalties on number of zeros, and the maximum run of zeros present in the data may be closely maintained, as illustrated for MCZ on the second block of Table 1.

Figure 5 includes yet four other simulations, obtained by optimizing the maximum duration of consecutive days of no rain, i.e. the maximum run of zeros, with 10% penalties on entropy and total number of zeroes, as explained earlier. As seen, these new simulations are generally comparable to the ones shown in Figs. 3 and 4. Although these simulations are not necessarily better than those already reported on the optimized attribute MCZ, they turn out to provide suitable rainfall representations that may not be distinguishable from the original data set. This fact is reflected by: (i) the very close histograms obtained (which were not explicitly optimized, but via a penalty), (ii) the excellent agreement in entropy functions, and (iii) also the rather closely-related statistics, as seen in the last block of Table 1.

Remarkably, the FM approach (based on a Cantorian input or not and with distinct sets of parameters as reflected in Appendix 1) may provide deterministic sets that look as complex (random) as those of the highly-intermittent rainfall data set from Laikakota. As the shown sets are just examples of other "solutions" having similar properties, the FM methodology represents a useful repository of patterns with which to approximate rainfall at this site at the daily scale.

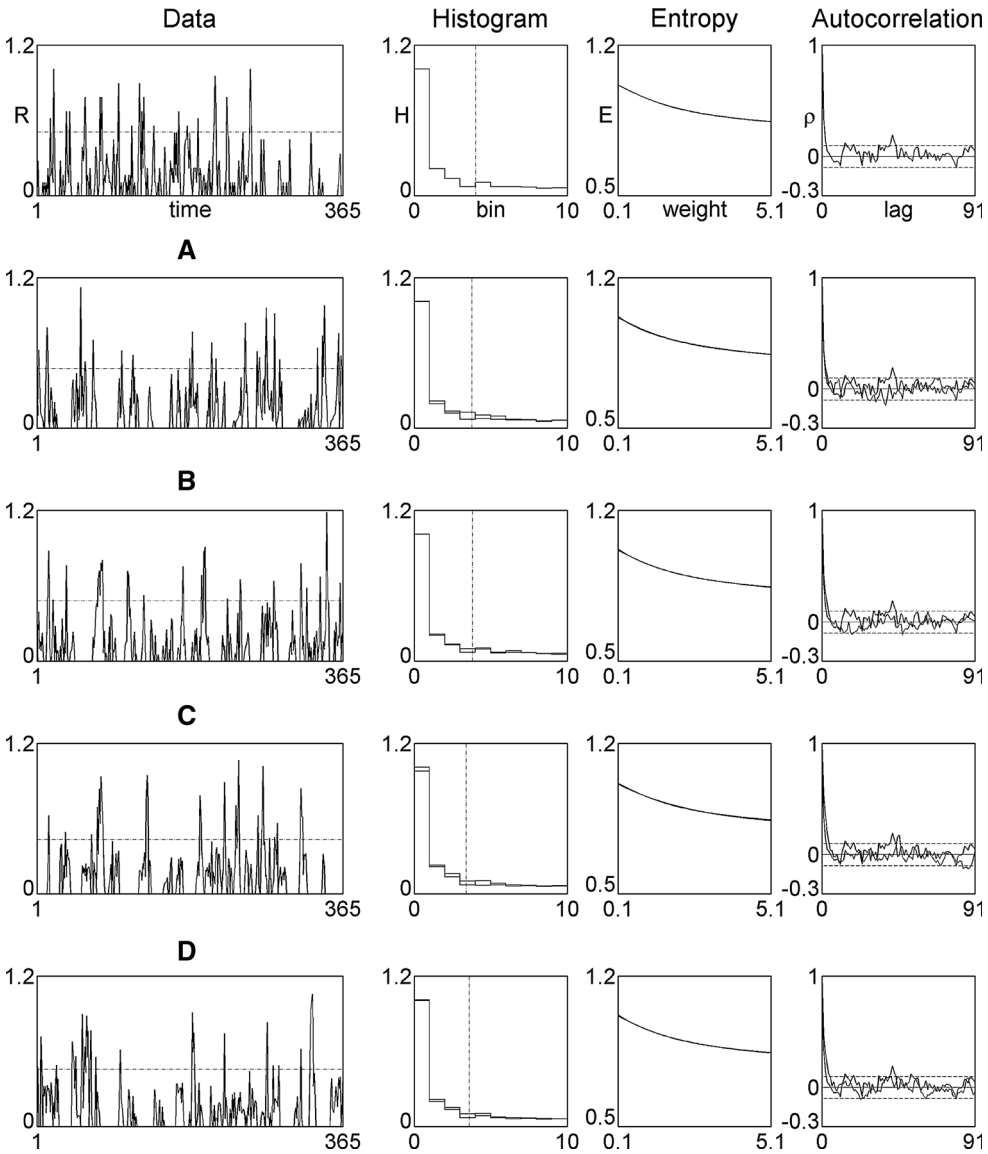

Fig. 7 A measured set of daily rainfall at Tinkham Creek, Washington gathered from October 2000 to September 2001 (*top*), followed by its histogram, Rényi entropy and autocorrelation functions; and four suitable rainfall simulations (*bottom*), based on fits of the records' entropy function. While simulations **A** and **B** emanate from fractal wires, **C** and **D** come from Cantorian attractors. While sets **A** and **C** use a horizontal threshold, **B** and **D** employ a vertical one

4.2 Tinkham rainfall

Figures 6, 7 and 8 shows some examples of the three aforementioned versions of plausible simulations for rainfall, but now for a rainfall set gathered in Tinkham, i.e. based on: (a) fittings of histograms, (b) preservation of entropies, and (c) capturing consecutive zero rainfall values. Similar to the results just presented for the Laikakota rainfall, these figures show the "real" rainfall records at Tinkham for water year 2000–2001, followed by four distinct simulations A to D, defined, as before, regarding the FM method (original and Cantorian) and aided by thresholds (horizontal and vertical). Observe, once again, and for illustration purposes, that simulation labeled C on Fig. 8 corresponds to the construction shown in Fig. 2 based on a horizontal threshold.

As for sets at Laikakota, the very top of the figures for Tinkham include the actual record and its main statistics (histogram, entropy, and autocorrelation function). Based on these, it may be surmised that the Tinkham daily rainfall time series appears to be more complicated than the Laikakota rainfall set. This set in the State of Washington, measured over the rainy season of 365 days, contains several sporadic peaks and, stunningly, only three major gaps showing no rainfall activities. Even though the histogram concentrates at low rainfall values and the autocorrelation function decays rather fast towards values close to zero (similar to the one observed for the Laikakota rainfall), the entropy function decays more slowly as a reflection of an increased level of complexity.

Figure 6, the counterpart of Fig. 3, reveals some suitable simulations based on histogram fittings for the

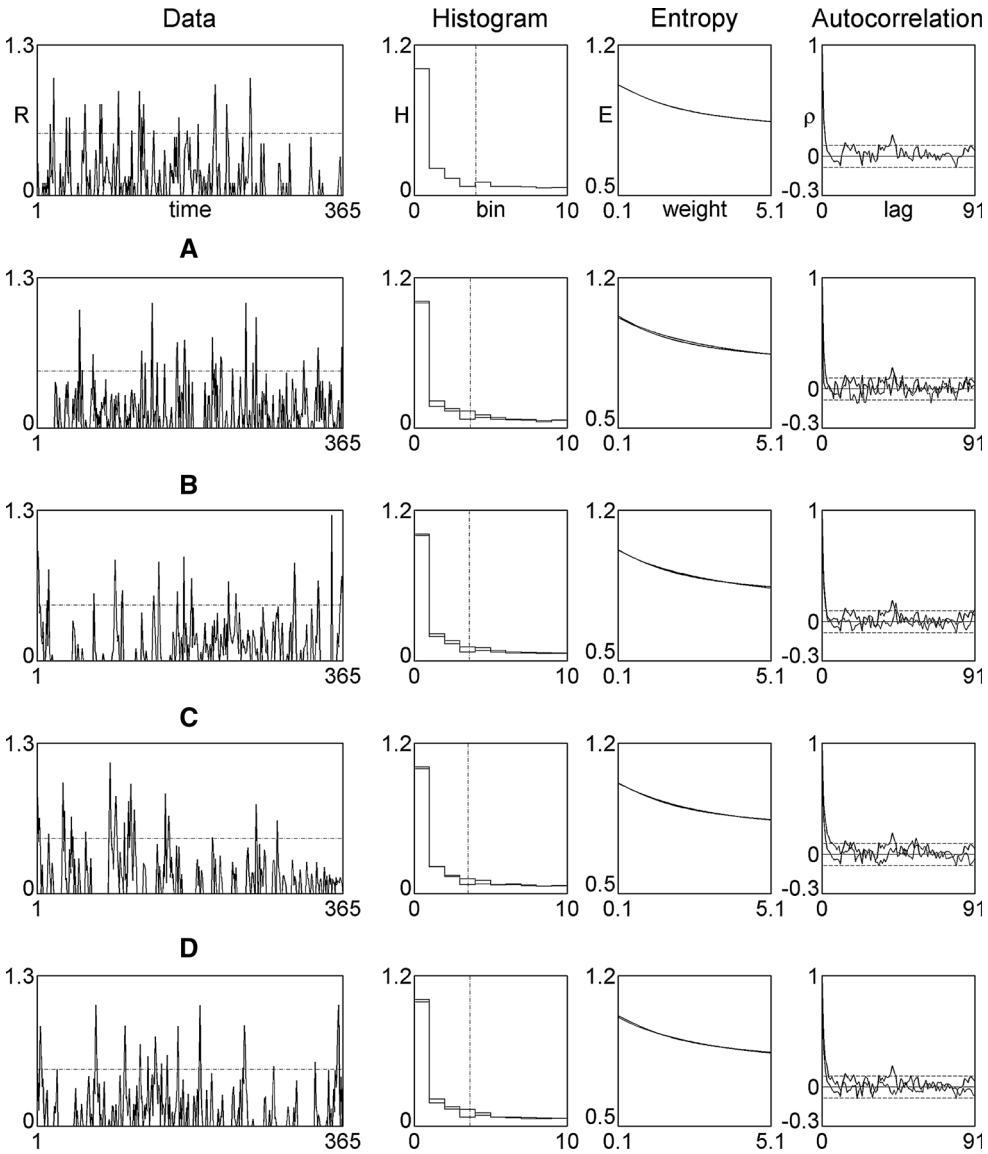

Fig. 8 A measured set of daily rainfall at Tinkham Creek, Washington gathered from October 2000 to September 2001 (*top*), followed by its histogram, Rényi entropy and autocorrelation functions; and four suitable rainfall simulations (*bottom*), based on fits of the records' distribution of zeros. While simulations **A** and **B** emanate from fractal wires, **C** and **D** come from Cantorian attractors. While sets **A** and **C** use a horizontal threshold, **B** and **D** employ a vertical one

Tinkham rainfall (i.e. via Eq. (8) plus penalties): A and B come from fractal wires, C and D from "Cantorian" representations, A and C use a horizontal threshold, and B and D a vertical threshold. As seen, since the given time series is rather intricate, the simulations also turn out to be rather intermittent. Overall, simulation D (based on a Cantorian construction and vertical threshold) appears to be closer to the "spirit" of the data more so than the others, but all simulations do preserve very well the histogram employed in the objective function.

The first block in Table 2 includes relevant statistical information for the FM simulations in Fig. 6. As seen, the Nash–Sutcliffe for the objective function used (NSH) is indeed very close to 100%, leading to close agreements between the record and simulations at the 90% level (PS90), numbers of zero values in records and simulations (NZR), and, consequently, small errors in zero values (EZR), which, as seen, have the aforementioned simulation D as best. Observe that the Nash-Sutcliffe values for entropy (NSE), although not optimized, are all rather high (all above 83%), with the curious observation that the best simulation to the naked eye D is the one that is worse. Notice that the maximum consecutive number of zero rains (MCZ, not optimized) are rather close (even if they happen at distinct locations within the record and simulations) and that the Nash-Sutcliffe values for autocorrelations (NSA) are rather low (once again, and as found for the records in Bolivia).

Figure 7, the counterpart of Fig. 4, contains some selected simulations based on optimizing square errors in entropy function between the record and the FM projections, i.e. Equation (9) plus penalties on zero values, as explained in Sect. 3. As seen, all these simulations, while exhibiting noticeable intermittencies that happen

Table 2 Error statistics associated with Figs. 6, 7 and 8 for rainfall records at Tinkham, Washington, based on preservation histogram (top block), entropy (middle block) and consecutive zeros (bottom block)

Simulation	NSH (%)	PS90 (%)	NZR	EZR (%)	NSE (%)	MCZ	NSA (%)
6A	99.3	91.6	200 (180)	10.0	93.3	21 (24)	−7.4
6B	99.6	91.0	200 (184)	8.0	88.3	21 (17)	−16.6
6C	99.6	91.3	200 (197)	1.5	91.5	21 (22)	−16.6
6D	99.8	90.3	200 (201)	−0.5	83.2	21 (20)	−38.6
7A	99.4	91.3	200 (191)	4.5	99.9	21 (24)	−55.1
7B	99.8	90.8	200 (193)	3.5	100	21 (20)	−21.3
7C	99.5	92.4	200 (203)	−1.5	99.8	21 (22)	−1.9
7D	99.7	91.6	200 (193)	3.5	100	21 (21)	20.7
8A	99.1	92.1	200 (191)	4.5	98.4	21 (20)	−109.0
8B	99.6	91.8	200 (195)	2.5	99.6	21 (22)	−46.1
8C	99.6	92.4	200 (202)	−0.5	99.7	21 (19)	−16.4
8D	99.2	92.1	200 (204)	−2.0	99.6	21 (19)	−31.8

throughout the year, closely fit not only the entropy function but also the histogram of the record as well. This is reflected in the second block of Table 2, which exhibits excellent agreements in all attributes, even if the placements of consecutives zeros do not match, as expected on simulations.

As done for the Laikakota rain record, Fig. 8 presents a further attempt to find additional suitable FM simulations that preserve the maximum duration of no rain activities. As seen and as corroborated by the third block in Table 2, this notion, aided by relevant penalties (see Sect. 3), also produces plausible scenarios that statistically preserve relevant features present in the highly intermittent Tinkham set. Notice, for instance, the similarly-looking FM sets containing sharp concentrations of rain, sets that reflect the complexity in the original set, as also found in Figs. 6 and 7.

Overall, and in accordance with the Laikakota set, both FM approaches (the original one and a Cantorian extension, while based in distinct parameters as shown in Appendix 1) provide sensible deterministic simulations that possess similar complexities as the observed record in Tinkham. As the shown sets are just examples of other "solutions" having similar properties, the FM methodology represents a useful deposit of patterns with which to approximate rainfall at this site at the daily scale. As illustrated in Figs. 6, 7 and 8, the FM methodology provides a collection of patterns with which to model rainfall at the daily scale at the Tinkham site.

5 Conclusions

The present study has explored the capabilities of the fractal-multifractal (FM) procedure to serve as a suitable tool for simulating highly-intermittent rainfall records. The original FM approach (Puente and Obregón 1996) as well as its Cantorian extension (Huang et al. 2013; Maskey et al. 2015), coupled with the notion of thresholds (either in the input *dx* or in the output *dy*), have been employed to daily rainfall sets observed at Laikakota in Bolivia and at Tinkham in Washington State, USA. The results indicate that these, ultimately simple notions, may be used to produce reasonable "realizations" of the highly-intermittent rainfall records at both sites. The results also show that it is not difficult to employ the FM methodology, coupled with a particle swarm optimization algorithm, to find deterministic sets that preserve the entire rainfall record's histogram, the entire entropy function, and the maximum number of consecutive zeros in the records.

Realizing that neither histograms nor entropies fully characterize a given rainfall set, it has been shown that up to 8 FM parameters are sufficient to produce possible simulations having compression ratios that are at least 34:1 and 46:1 for the Laikakota and Tinkham rainfall data, respectively. Although it may be thought that a representation having 6 or 8 parameters is not parsimonious enough when compared with, say, random multiplicative cascades and random Neyman–Scott models of rain that require at most 5 parameters, it ought to be emphasized that such stochastic techniques are typically not capable of preserving complete statistical features as in the simulations reported herein. Given the recent reports of precipitation measurements being plagued with nontrivial errors (e.g. Lanza and Vuerich 2009), it may be said that all representations shown, with rather high Nash-Sutcliffe statistics for the key attributes near 100%, are indeed acceptable (parsimonious) renderings of what nature produces.

The results herein certainly give further credence to the suitability of the FM approach and its variants for modeling rainfall (and other geophysical) records, which clearly substantiate the notion that complex geophysical patterns

may be holistically described in a deterministic way (Puente and Sivakumar 2007). Certainly, the FM approach and extensions are computationally rather simple, and as such it is a useful tool for synthesis and simulation of complex and seemingly-random sets. As the FM methodology may indeed be given a physical interpretation as a realization of a non-trivial multiplicative cascade having variable multipliers (Cortis et al. 2013), it is envisioned that the encoding of sets at other specific sites (not reported here) may result in a "physical" interpretation of the FM parameters that may guide how to further refine the suitable simulations. Application of the ideas to various sites corresponding to distinct climatic conditions shall be necessary in order to fully evaluate how FM parameters vary and how unique (or not) such representations may be.

Acknowledgements The research leading to this article was supported by a JASTRO Award provided to the first author by the University of California, Davis. We are thankful to Ministerio de Medio Ambiente y Agua, Bolivia for providing rainfall records gathered at Laikakota and also to the team of National Resource Conservation Service for the availability of rainfall records in its web portal. Bellie Sivakumar acknowledges the financial support from the Australian Research Council (ARC) through the Future Fellowship Grant awarded to him (FT110100328). Comments and suggestions by anonymous reviewers helped improve the manuscript and are gratefully acknowledged.

Appendix 1: Parameter values for FM representation of Figs. 3–8

See Tables 3 and 4.

Table 3 "Wire" representations

Figure	X_1	Y_1	d_1	d_2	p_1	Threshold (%)
3A	0.409	1.952	−0.999	−0.486	0.993	2.3
3B	0.444	0.538	−0.575	−0.624	0.489	11.3
4A	0.923	−1.210	−0.999	−0.148	0.988	5.7
4B	0.883	3.010	0.104	−0.608	0.414	2.6
5A	0.623	2.626	−0.110	−0.999	0.091	12.7
5B	0.645	0.580	−0.606	−0.826	0.670	19.7
6A	0.938	−1.304	−0.673	0.188	0.617	4.0
6B	0.821	4.143	0.598	0.293	0.515	17.2
7A	0.938	−1.160	−0.576	0.224	0.613	51.3
7B	0.045	4.170	−0.720	−0.252	0.679	17.1
8A	0.981	−0.260	−0.472	−0.380	0.405	0.3
8B	0.459	−2.050	−0.840	−0.056	0.787	10.1

Without loss of generality, the following values are fixed: $X_0 = 0$, $X_2 = 1$; $Y_0 = 0$, $Y_2 = 1$

Table 4 Cantorian representations

Figure	X_1	X_2	Y_1	Y_2	d_1	d_2	p_1	Threshold (%)
3C	0.074	0.453	−1.519	−0.900	−0.399	−0.242	0.257	15.3
3D	0.159	0.399	−0.640	3.024	−0.608	0.051	0.513	4.1
4C	0.077	0.472	0.270	−0.810	−0.600	0.062	0.406	2.6
4D	0.137	0.611	−4.150	−2.000	0.004	0.726	0.258	6.2
5C	0.655	0.914	1.000	2.360	−0.114	0.473	0.636	2.7
5D	0.557	0.556	0.650	0.555	−0.540	−0.538	0.550	18.2
6C	0.265	0.568	2.570	0.638	−0.096	−0.619	0.506	16.5
6D	0.375	0.778	0.305	4.720	0.757	0.092	0.640	18.0
7C	0.196	0.564	−4.430	−1.427	0.567	−0.210	0.594	8.8
7D	0.582	0.800	−5.000	0.341	0.238	−0.590	0.457	4.82
8C	0.086	0.560	−2.640	−2.130	0.054	0.716	0.270	1.5
8D	0.477	0.568	−0.230	3.603	−0.712	0.222	0.779	13.3

Without loss of generality, the following values are fixed: $X_0 = 0$, $X_3 = 1$; $Y_0 = 0$, $Y_3 = 1$

References

Barnsley MF (1988) Fractals everywhere. Academic Press, San Diego

Cortis A, Puente CE, Sivakumar B (2009) Nonlinear extensions of a fractal-multifractal approach for environmental modeling. Stoch Environ Res Risk Assess 23(7):897–906

Cortis A, Puente CE, Huang HH, Maskey ML, Sivakumar B, Obregón N (2013) A physical interpretation of the deterministic fractal-multifractal method as a realization of a generalized multiplicative cascade. Stoch Environ Res Risk Assess 28(6):1421–1429

Eagleson PS (1978) Climate, soil, and vegetation, 2, The distribution of annual precipitation derived from observed storm sequences. Water Resour Res 14(5):713–721

Feder J (1988) Fractals. Plenun Press, New York

Fernández Martínez JL, García Gonzalo E, Fernández Álvarez JP, Kuzma HA, Menéndez Pérez CO (2010) PSO: A powerful algorithm to solve geophysical inverse problems: Application to a 1D-DC resistivity case. J Appl Geophys 71(1):13–25

Garbrecht JD, Zhang JX (2012) Generating synthetic daily precipitation realizations for seasonal precipitation forecasts. J Hydrol Eng 19(1):252–264

Geng S, de Vries FWP, Supit I (1986) A simple method for generating daily rainfall data. Agr Forest Meteorol 36(4):363–376

Gupta VK, Waymire EC (1993) A statistical analysis of mesoscale rainfall as a random cascade. J Appl Meteorol 32(2):251–267

Haan CT, Allen DM, Street JO (1976) A Markov chain model of daily rainfall. Water Resour Res 12(3):443–449

Huang HH, Puente CE, Cortis A (2012) Geometric harnessing of precipitation records: reexamining four storms from Iowa City. Stoch Environ Res Risk Assess 27(4):955–968

Huang HH, Puente CE, Cortis A, Fernández Martínez JL (2013) An effective inversion strategy for fractal–multifractal encoding of a storm in Boston. J Hydrol 496:205–216

Jothityangkoon C, Sivapalan M, Viney NR (2000) Tests of a space-time model of daily rainfall in southwestern Australia based on nonhomogeneous random cascades. Water Resour Res 36(1):267–284

Kaczmarska J, Isham V, Onof C (2014) Point process models for fine-resolution rainfall. Hydrolog Sci J 59(11):1972–1991

Kavvas ML, Delleur JW (1981) A stochastic cluster model of daily rainfall sequences. Water Res Res 17(4):1151–1160

Langousis A, Veneziano D (2007) Intensity-duration-frequency curves from scaling representations of rainfall. Water Res Res. doi:10.1029/2006WR005245

Langousis A, Veneziano D, Furcolo P, Lepore C (2009) Multifractal rainfall extremes: theoretical analysis and practical estimation. Chaos, Solitons Fractals 39(3):1182–1194

Langousis A, Carsteanu AA, Deidda R (2013) A simple approximation to multifractal rainfall maxima using a generalized extreme value distribution model. Stoch Environ Res Risk Assess 27(6):1525–1531

Lanza LG, Vuerich E (2009) The WMO field intercomparison of rain intensity gauges. Atmos Res 94(4):534–543

Lovejoy S, Schertzer D (1985) Generalized scale invariance in the atmosphere and fractal models of rain. Water Resour Res 21(8):1233–1250

Lovejoy S, Schertzer D (1990) Multifractals, universality classes and satellite and radar measurements of cloud and rain fields. J Geophys Res 95(3):2021–2034

Mandelbrot BB (1982) The fractal geometry of nature. Freeman, San Francisco

Mandelbrot BB (1989) Multifractal measures especially for the geophysicist. In: Scholz CH, Mandelbro MM (eds) Fractals in geophysics. Birkhanser, Basel, pp 1–42

Maskey ML, Puente CE, Sivakumar B, Cortis A (2015) Encoding daily rainfall records via adaptations of the fractal multifractal method. Stoch Environ Res Risk Assess. doi:10.1007/s00477-015-1201-7

Obregón N, Puente CE, Sivakumar B (2002a) Modeling high resolution rain rates via a deterministic fractal–multifractal approach. Fractals 10(3):387–394

Obregón N, Sivakumar B, Puente CE (2002b) A deterministic geometric representation of temporal rainfall. Sensitivity analysis for a storm in Boston. J Hydrol 269(3–4):224–235

Puente CE (1996) A new approach to hydrologic modelling: derived distribution revisited. J Hydrol 187:65–80

Puente CE (2004) A universe of projections: may Plato be right? Chaos, Solitons Fractals 19(2):241–253

Puente CE, Obregón N (1996) A deterministic representation of temporal rainfall: result for a storm in Boston. Water Resour Res 32(9):2825–2839

Puente CE, Sivakumar B (2007) Modeling hydrologic complexity: a case for geometric determinism. Hydrol Earth Syst Sc 11:721–724

Rodríguez-Iturbe I, Cox DR, Isham V (1987) Some models for rainfall based on stochastic point processes. Proc R Soc Lond A 410(1839):269–288 (The Royal Society)

Rodríguez-Iturbe I, Febres de Power B, Sharifi MB, Georgakakos KP (1989) Chaos in rainfall. Water Resour Res 25(7):1667–1675

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

Sivakumar B (2000) Chaos theory in hydrology: important issues and interpretations. J Hydrol 227(1–4):1–20

Sivakumar B, Berndtsson R, Olsson J, Jinno K (2001) Evidence of chaos in the rainfall-runoff process. Hydrolog Sci J 46(1):131–145

Tessier Y, Lovejoy S, Schertzer D (1993) Universal multifractals: Theory and observations for rain and clouds. J Appl Meteorol 32:223–250

Tsonis AA, Elsner JB, Georgakakos KP (1993) Estimating the dimension of weather and climate attractors: important issues about the procedure and interpretation. J Atmos Sci 50(15):2549–2555

Vandenberghe S, Verhoest NEC, Onof C, De Baets B (2011) A comparative copula-based bivariate frequency analysis of observed and simulated storm events: a case study on Bartlett-Lewis modeled rainfall. Water Resour Res 47(7):1–16

Veneziano D, Langousis A (2005) The areal reduction factor: a multifractal analysis. Water Resour Res. doi:10.1029/2004WR003765

Veneziano D, Langousis A (2010) Scaling and fractals in hydrology. In: Sivakumar B, Berndtsson R (eds) Advances in data-based approaches for hydrologic modeling and forecasting. World Scientific, Singapore

Veneziano D, Langousis A, Furcolo P (2006) Multifractality and rainfall extremes: A review. Water Resour Res. doi:10.1029/2005WR004716

Veneziano D, Langousis A, Lepore C (2009) New asymptotic and preasymptotic results on rainfall maxima from multifractal theory. Water Resour Res. doi:10.1029/2009WR008257