Deterministic chaos versus stochasticity in analysis and modeling of point rainfall series

Demetris Koutsoyiannis and Demetris Pachakis

Department of Water Resources, Faculty of Civil Engineering, National Technical University, Athens, Greece.

Abstract. The differences between historic rainfall data and synthetic data obtained by a stochastic rainfall model are investigated using nonlinear analysis tools devised for description and characterization of chaotic behavior. To achieve this goal, a 6-year point rainfall record with a time resolution of one quarter of hour is studied. A stochastic model capable of preserving important properties of the rainfall process, such as intermittency, seasonality and scaling behavior, is fitted to this data set and a synthetic time series of equal length is generated. For both data sets the correlation dimension is calculated for various embedding dimensions by the time delay embedding method. However, the applicability of this method in estimating dimensions proves limited due to the domination of voids (dry periods) in a rainfall record at a fine time resolution. Thus, in addition to time delay embedding, a Cantorian dust analogue method is developed and used to estimate dimensions. Results of both methods show that there is no substantial difference in behavior between the synthetic and the historic records. Moreover, no evidence of low-dimensional determinism is detected in the sets examined.

1. Introduction

Recent studies revealed that nonlinear dynamical systems may produce time series indistinguishable from random noise by typical statistical methods, although they are fully deterministic with few degrees of freedom. Such systems, called chaotic, are very sensitive to small perturbations in the initial conditions and exhibit instability due to their nonlinear nature. Because of the inefficiency of the standard statistical methods to describe and model those systems, new methods were devised to measure and characterize them using concepts from nonlinear dynamical systems analysis and information theory. Furthermore, new methods have been established capable of revealing the underlying deterministic dynamics of systems with seemingly random behavior. Such methods are summarized by *Gershenfeld and Weigend* [1993] and *Tsonis* [1992] among others. In hydrology, researchers have tried to implement those new methods in rainfall and streamflow records of short timescales seeking underlying nonlinear deterministic laws. The results so far vary, depending on the type of the examined data.

Studies in the structure of individual storm events [*Rodriguez-Iturbe et al.*, 1989; *Sharifi et al.*, 1990; *Rodriguez-Iturbe*, 1991] provided evidence that the temporal evolution of a particular storm event may be characterized as a deterministic chaotic process with a low-dimensional strange attractor. Similar results were obtained from simultaneous study of several events of the same meteorological convective character [*Tsonis*, 1992, p. 169; *Tsonis at al.*, 1993].

For continuous rainfall records at a certain time resolution there are no conclusive results yet. *Rodriguez-Iturbe et al.*, [1989] and *Rodriguez-Iturbe* [1991] do not detect low-dimensional chaotic dynamics in weekly rainfall data of Genoa, Italy, covering a period of 148 years. On the contrary, *Jayawardena and Lai* [1994] detect high-dimensional chaotic behavior (with embedding dimension between 30 and 40) in daily rainfall at three rain gauges in Hong Kong covering a period of 11 years. They conclude that rainfall time series could be better modeled by methods of nonlinear dynamics, such as phase-space reconstruction via time delay embedding,

than by traditional stochastic models, such as autoregressive moving average (ARMA). We note, however, that these conclusions must be taken with criticism, as, given the results of *Tsonis et al.* [1993], the embedding dimension is too high for so few data points to allow a claim of robustness.

In the present, discussion is going on whether the temporal evolution of rainfall is better modeled as a deterministic process rather than a stochastic process. However, the distinction between a deterministic and a stochastic process is not so clear in practice. For example, random number generators, used for simulation of stochastic processes, are obviously deterministic. Furthermore, stochastic models can incorporate deterministic components, if any, and deterministic models very often need to add a random component to fit the evolution of a physical system more appropriately. Another fact that should be taken into account is that, due to sensitive dependence to initial conditions and given the inherent uncertainty of the standard measurements, the horizon of feasibly reliable forecasting through a deterministic model can be limited dramatically. The latter becomes very obvious in the case of chaotic systems.

In practice, stochastic models of rainfall are well established and very useful tools for several engineering applications, from short-term prediction to long-term simulation of rainfall and rainfall-driven processes such as runoff. On the other hand, applications of chaotic deterministic methods, such as the above referenced, are currently limited to detection of nonlinear dynamics and short-term prediction. However, the new methods based on the chaos literature may potentially reduce the value of stochastic models, if they can provide evidence that their results differ significantly from historic data. This is attempted for example in the work by *Jayawardena and Lai* [1994] mentioned above.

Indeed, the real rainfall process in a short timescale (e.g., hourly or finer) differs significantly from synthetic signals generated by simple stochastic models, like white noise or ARMA processes. It is quite widely understood that rainfall is not effectively represented by these linear models and it is not our purpose to provide more evidence on that. The objective of this paper is to investigate if there are essential differences that distinguish a historic rainfall record from a synthetic record generated by a well-structured stochastic model, capable of preserving important properties of rainfall such as intermittency, seasonality, and scaling behavior. For this comparison, we use typical methods of nonlinear analysis that have successfully shown the difference between stochastic and chaotic deterministic time series in other fields. Other relevant issues discussed in this paper are: how typical descriptors of chaotic behavior can be used to characterize the rainfall process, how proper these methods of nonlinear analysis are to analyze rainfall data and whether there are any characteristic timescales in a continuous rainfall record or not.

The methodology adopted consists of the following steps:

Selection of a historic data set. This was a 6-year (1984-1989) record of incremental rainfall depths, measured every quarter of an hour at station Ortona Lock 2, Florida, USA. The record measurements were given with a depth resolution of 1 mm.

Adoption and calibration of a stochastic model. The Scaling Model of Storm Hyetograph [*Koutsoyiannis and Foufoula-Georgiou*, 1993] coupled with an alternating renewal model [see e.g., *Ross*, 1993, p. 326] for storm occurrence were adopted to simulate the rainfall process.

Generation of a synthetic record. A synthetic record of equal length to the historic (6 years) and equal time resolution (one-quarter hour) was generated using the stochastic model.

Comparison of historic and synthetic records. Various descriptors of chaotic dynamics were calculated and compared for both the synthetic and the historic data.

The paper is outlined as follows. In section 2 we summarize the adopted stochastic model and its calibration to the study data set. In section 3 we summarize several topics of the chaos literature that are used in our application. In sections 4 and 5 we describe the computations and their results and compare the synthetic and historic series using tools of the chaos literature. In section 6 we draw some conclusions. There is also an appendix, where we give details about the generation of the synthetic record using the adopted stochastic model.

2. Description and Calibration of the Stochastic Model

2.1. Summary of the Scaling Model of Storm Hyetograph

The scaling model of storm hyetograph [*Koutsoyiannis and Foufoula-Georgiou*, 1993] is a stochastic model that parameterizes a population of storms, taking advantage of scaling properties revealed in rainfall data. It describes the temporal evolution of rainfall intensity within a storm in continuous time. It can also be applied in discrete time to generate incremental storm depths at a given time interval for a storm with given total depth and duration.

The main hypothesis of the model, is that the process of instantaneous rainfall intensities within storms, in terms of their distribution function, is a self-similar (simple scaling) process with a scaling exponent H. In mathematical terms, if D denotes the duration of the storm and $\xi(t, D)$ denotes the instantaneous rainfall intensity at time t ($0 \le t \le D$), then

$$\{\xi(t, D)\} \overset{d}{=} \{\lambda^{-H} \xi(\lambda t, \lambda D)\} \tag{1}$$

where the above equality is in terms of the finite-dimensional probability distribution, and λ is any positive number.

A secondary hypothesis of the model is that the process $\xi(t, D)$ is stationary within the same storm event or within storm events of the same duration, that is,

$$\{\xi(t, D)\} \overset{d}{=} \{\xi(t + \tau, D)\}, \quad 0 \le t, t + \tau \le D \tag{2}$$

On the basis of these hypotheses, the statistics of the rainfall intensity at any time t within a storm of any duration D can be deduced. The mean value is

$$E[\xi(t,D)] = c_1 D^H \tag{3}$$

and the product moment

$$E[\xi(t, D)\xi(t + \tau, D)] = \phi(\tau/D)D^{2H} \tag{4}$$

where τ is a time lag, $c_1 \equiv E[\xi(t/D, 1)]$ and $\phi(\tau/D) \equiv E[\xi(t/D, 1)\, \xi((t + \tau)/D, 1)] = k\,(\tau/D)^{\beta}$, whereas k and β are parameters.

The total storm depth over duration D,

$$Z = \int_0^D \xi(s,D)\, ds \tag{5}$$

has mean

$$E[Z]= c_1 D^{H+1} \tag{6}$$

and variance

$$\mathrm{Var}[Z]= c_2 D^{2(H+1)} \tag{7}$$

where $c_2 = 2k/[(1-\beta)(2-\beta)]-c_1^2$. To convert the model from continuous time to discrete time we use the incremental storm depth precipitated in the time interval $((i-1)\Delta, i\Delta)$ of length Δ:

$$X_i = \int_{(i-1)\Delta}^{i\Delta} \xi(t,D)\, dt, \quad i = 1, 2, \ldots, k \tag{8}$$

where k is the smallest integer that is greater than or equal to D/Δ. The mean of X_i is

$$E[X_i]= c_1 \delta D^{H+1} \tag{9}$$

and its variance is

$$\mathrm{Var}[X_i]=[(c_2 + c_1^2)\delta^{-\beta} - c_1^2]\delta^2 D^{2(H+1)} \tag{10}$$

where $\delta = \Delta/D$. The covariance between two incremental depths within the same storm is given by

$$\mathrm{Cov}[X_i, X_j]=[(c_2 + c_1^2)\delta^{-\beta} f(|j-i|,\beta) - c_1^2]\,\delta^2 D^{2(H+1)} \tag{11}$$

where $f(m,\beta) = (1/2)[(m-1)^{2-\beta} + (m+1)^{2-\beta}] - m^{2-\beta}$ for integer time lag between the two incremental storm depths $m > 0$. The derivations of the above equations are given by *Koutsoyiannis and Foufoula-Georgiou* [1993].

In brief, the scaling model of storm hyetograph has only four parameters: the scaling exponent H, the mean value parameter c_1, the variance parameter c_2, and the correlation decay parameter β. These parameters can be easily estimated from the data after partitioning the storm events into classes of certain duration as described in detail by *Koutsoyiannis and Foufoula-Georgiou* [1993]. Thus H and c_1 are directly estimated from (6) by least squares, and c_2 from (7). The parameter β is estimated from the lag-one autocorrelation function of the incremental storm depths at any chosen time increment Δ. In this paper we have introduced an improvement in the estimation of β, given by the equation

$$\beta = 1 - \frac{\ln(E[X_i X_{i+1}]/E[X_i^2]+1)}{\ln 2} \tag{12}$$

which is a consequence of (9) to (11). The latter equation allows the direct estimation of β.

The scaling model can simulate the evolution of the rainfall within any rainstorm, given its total duration D. To make a complete rainfall generator, it should be combined with a model capable of simulating the durations D of rainstorms and the dry times B between rainstorms. Here we have used the well-known alternating renewal process [see e.g., *Ross*, 1993, p. 326] for this purpose. The details and the use of this model are described in the appendix.

2.2. Fitting of the Scaling Model to the Data

The scaling model was fitted to the Ortona data set that was given with a time resolution Δ = 0.25 hours and depth

resolution 1 mm. All storms of the 6-year period with durations greater than Δ were used except for those including corrupted recordings or missing data. There were 426 events with durations ranging from 0.5 to 35 hours and 37 events with durations less or equal than Δ that were modeled separately. As usual in similar situations [*Huff*, 1967; *Restepo-Posada and Eagleson*, 1982; *Koutsoyiannis and Foufoula-Georgiou*, 1993], events were allowed to include periods of zero rainfall lasting less than a critical time c. Initially, the approach suggested by *Koutsoyiannis and Foufoula-Georgiou* [1993], which is similar to that of *Restepo-Posada and Eagleson*, [1982] was tried to determine c. According to this approach, several trial values of c are investigated and finally the value that leads to storm arrivals forming a Poisson process is chosen. However, in our case this method gave a value of c greater than 24 hours, which is too high given that the average storm duration is about 4 hours. Once the Poisson process is not a structural postulation of the entire method, we finally selected a smaller value, that is, c = 7 hours, which is about 1.5 times the average storm event duration. This value is equal to that adopted by *Koutsoyiannis and Foufoula-Georgiou* [1993] for another data set and very close to the arbitrary value suggested by *Huff* [1967], that is, 6 hours.

The fitting procedure was based on the separation of the storms into classes according to their duration. Initially, the storms were grouped in two seasons, the dry season (October through May) and the wet season (June through September). The parameter estimation procedure was performed separately for each season. The parameter sets estimated were similar for the two seasons and thus, for simplicity, a unique parameter set was estimated using all events of both seasons. The final parameters are H = −0.449, c_1 = 8.74, c_2 = 85.68, and β = 0.246 (corresponding to units of millimeters and hours). Notable is the departure of H from zero, which indicates the departure of the process from stationarity, as it implies a strong dependence of the intensity process on storm duration. Figure 1 and Figure 2 depict the model fitting to the historic data regarding the total and incremental storm depths, respectively. In these figures the comparison of the synthetic and historic data is given in terms of the mean values and standard deviations of total and incremental depths. The modeled statistics refer to the final unique parameter set. The statistics of the historic data refer to the unified data set of all storms as well as the seasonal subsets for the dry and wet period. Overall, these figures indicate that the scaling hypothesis is consistent with the data examined and justify the use of a unique parameter set. Below we will perform other comparisons using tools of the chaos literature.

Figure 1

Figure 2

The time resolution of 0.25 hours and the depth resolution of 1 mm may have some implications in the model parameters. However, our past experience of the model application shows that the parameters and the general fitting of the model are not essentially affected when increasing the time resolution from 0.25 to 1 or 2 hours, nor when increasing the depth resolutions from 0.25 mm to 1 mm. In the latter case, some bias is introduced to the mean values because of the round-off of the very small depths to 1 mm, which affects the value of c_1.

The parameters of the alternating renewal model were allowed to vary through the different months of the year as is more specifically described in the appendix.

Copy editor: Please keep it as a separate paragraph.

3. Basic Concepts of the Chaos Literature

3.1. Descriptors of Chaotic Behavior

In this subsection we describe briefly several important topics of the chaos literature that are used in the next sections to characterize, quantify, and compare the rainfall time series. Here the description is generalized for any set A that is a subset of an n-dimensional metric space with a normalized measure μ defined on its Borel field. In our case, for $n = 1$, this set may represent all possible values of the incremental rainfall depth $X(t)$ for an arbitrary time interval t ($t = 1, 2, \ldots$) of length Δ, which is the set of positive real numbers R^+. It may also represent all values in a certain observed time series of incremental rainfall depths, in which case A is a finite subset of R^+. Accordingly, for $n > 1$, the set may represent the so-called time-delayed vectors, formed from the scalar time series of incremental rainfall depths,

$$\mathbf{X}=\left\{X(t),X(t-\tau),X(t-2\tau),\cdots,X(t-(n-1)\tau)\right\} \quad (13)$$

where τ is the time delay.

Let us consider a partition of the set A into $M(\varepsilon)$ boxes (hypercubes) A_1, A_2, ..., $A_{M(\varepsilon)}$ with length scale (i.e., edge length of each hypercube) ε. The entropy $I_1(\varepsilon)$ for this partition is defined by

$$I_1(\varepsilon)=-\sum_{i=1}^{M(\varepsilon)} p_i \ln p_i \quad (14)$$

where p_i is the measure of the part of the set A contained in the ith hypercube, that is, $p_i = \mu(A_i)$, such that

$$\sum_{i=1}^{M(\varepsilon)} p_i = 1 \quad (15)$$

If the set A consists of N observed values (points in the n-dimensional space) and N_i of them are contained in the ith hypercube A_i, then $p_i = N_i / N$. Accordingly, if this set is the sample space of a vector of random variables $\mathbf{X}$ then each hypercube A_i represents an event and $p_i = \Pr(\mathbf{X} \in A_i)$ where $\Pr(\,)$ denotes probability.

Rényi [1970] introduced a generalization of the entropy concept by defining the entropy of order q as

$$I_q(\varepsilon)=\frac{1}{1-q}\ln\sum_{i=1}^{M(\varepsilon)} p_i^{\,q} \quad (16)$$

Definition (16) applies for $q \neq 1$. Taking the limit for $q \to 1$ and using de l'Hospital's rule we get (14).

The generalized dimension of order q of the set under examination is defined by the following equation [*Grassberger*, 1983]:

$$D_q = \lim_{\varepsilon \to 0} \frac{-I_q(\varepsilon)}{\ln \varepsilon} \quad (17)$$

For simple geometrical objects such as lines or surfaces, all D_q are equal to the integer topological dimension (1 for a line, 2 for a surface, etc.). For fractal objects, they are not necessarily integers, nor all D_q are necessarily equal to each other. These dimensions have important content when applied to certain kinds of sets. For example, if the set A is the attractor of a dynamical system, its generalized dimensions represent

the number of local directions available to the system and so they provide an estimate of the number of degrees of freedom needed to describe the state of the system [*Gershenfeld and Weigend*, 1993, p. 48].

By applying de l'Hospital's rule in (17) we get

$$D_q = \lim_{\varepsilon \to 0} \frac{d\left(-I_q(\varepsilon)\right)}{d\left(\ln \varepsilon\right)} \tag{18}$$

where $d(\ )$ is the differentiation operator. This expression is more advantageous than (17) for numerical applications as the convergence of the derivative is faster.

For low values of q we have the most frequently used dimensions. Thus, for $q = 0$ we have the so-called capacity dimension D_0, given by

$$I_0(\varepsilon) = \ln M'(\varepsilon),$$

$$D_0 = \lim_{\varepsilon \to 0} \frac{-\ln M'(\varepsilon)}{\ln \varepsilon} = \lim_{\varepsilon \to 0} \frac{d\left(-\ln M'(\varepsilon)\right)}{d\left(\ln \varepsilon\right)} \tag{19}$$

where $M'(\varepsilon)$ is the number of boxes that intersect the set A. For $q = 1$ we have the information dimension D_1, that is,

$$I_1(\varepsilon) = -\sum_{i=1}^{M(\varepsilon)} p_i \ln p_i ,$$

$$D_1 = \lim_{\varepsilon \to 0} \frac{-I_1(\varepsilon)}{\ln \varepsilon} = \lim_{\varepsilon \to 0} \frac{d\left(-I_1(\varepsilon)\right)}{d\left(\ln \varepsilon\right)} \tag{20}$$

and for $q = 2$ we have the correlation dimension D_2, that is,

$$I_2(\varepsilon) = -\ln \sum_{i=1}^{M(\varepsilon)} {p_i}^2 ,$$

$$D_2 = \lim_{\varepsilon \to 0} \frac{-I_2(\varepsilon)}{\ln \varepsilon} = \lim_{\varepsilon \to 0} \frac{d\left(-I_2(\varepsilon)\right)}{d(\ln \varepsilon)} \tag{21}$$

In practice the quantities $I_q(\varepsilon)$ are not accurately estimated from finite samples of observations using the above definitions. *Grassberger* [1983] introduced another class of parameters called correlation integrals. For integer $q \geq 2$ the correlation integral of order q is defined by

$$C_q(\varepsilon) = N^{-q} \{\text{number of } q\text{-tuples } (\mathbf{X}_{j1}, \ldots, \mathbf{X}_{jq}) \text{ with all } \|\mathbf{X}_{js} - \mathbf{X}_{jr}\| < \varepsilon\} \tag{22}$$

and has the important property

$$C_q(\varepsilon) \approx \exp[(1 - q) I_q(\varepsilon)] \tag{23}$$

Thus, for integer $q \geq 2$, we can replace $-I_q(\varepsilon)$ with $\ln C_q(\varepsilon) / (q - 1)$ in the calculation of dimensions using the above equations, since the estimation of $C_q(\varepsilon)$ is more accurate than that of $I_q(\varepsilon)$ [*Grassberger and Procaccia*, 1983; *Grassberger*, 1983]. In practice however, only the correlation integral $C_2(\varepsilon)$ for $q = 2$ is used, because the calculation of higher-order integrals is extremely time consuming. The correlation integral of order 2, or simply correlation integral, is given by the following equation that is a consequence of (22):

$$C_2(\varepsilon)=\frac{1}{N}\sum_{j=1}^{N}\frac{1}{N}\sum_{i=1}^{N}H\left(\varepsilon-\left\|\mathbf{X}_i-\mathbf{X}_j\right\|\right) \tag{24}$$

where H is the Heaviside step function, with $H(u) = 1$ for $u > 0$ and $H(u) = 0$ for $u \le 0$. For the calculation of the distance $\|\mathbf{X}_i - \mathbf{X}_j\|$, the maximum norm is usually used as it reduces the computational time [*Hübner et al.*, 1993]. The correlation integral $C_2(\varepsilon)$ expresses the average proportion of pairs of points having distance smaller than ε between them, for a certain distance ε. As a direct consequence of (23), $C_2(\varepsilon)$ is related to the entropy $I_2(\varepsilon)$ with

$$C_2(\varepsilon) \approx \exp[-I_2(\varepsilon)] \tag{25}$$

The typical steps to estimate the correlation dimension D_2 of a set of N points in an n-dimensional space using the correlation integral, are the following:

1. Calculate the correlation integral $C_2(\varepsilon)$ for several values of the distance ε using an appropriate algorithm [e.g., *Grassberger*, 1990].

2. Make a log-log plot of $C_2(\varepsilon)$ versus ε and observe whether there exists any region with constant slope, known as a scaling region [e.g., *Hübner et al.*, 1993].

3. Calculate (by least squares) the slope of the scaling region, which is the estimate of the correlation dimension of the set.

Another useful concept is the mutual information of two sets. We consider the subsets of the real numbers A and B and their product $A \times B$. Using (14), we can define the entropies $I_1^A(\varepsilon)$, $I_1^B(\varepsilon)$ and $I_1^{A\times B}(\varepsilon)$ of the sets A, B, and $A \times B$, respectively, for the partition determined by the length scale ε. Then the mutual information of the two sets is defined as the difference between the marginal and the joint entropies, that is,

$$M^{AB}(\varepsilon)=I_1^A(\varepsilon)+I_1^B(\varepsilon)-I_1^{A\times B}(\varepsilon) \tag{26}$$

If the sets A and B represent two measurable quantities x and y, their mutual information represents the amount that a measurement of x reduces the uncertainty of y. If the quantities are independent, then the mutual information is zero: no knowledge can be gained for the second quantity by knowing the first. Otherwise the mutual information is a positive number.

3.2. Phase-Space Reconstruction From a Time Series

Let us consider an m-dimensional deterministic dynamical system. The state (or phase) of the system at a certain time can be represented by a point (or position vector) with m coordinates. Successive state points trace out a trajectory that represents the evolution of the dynamical system from some known initial state through time. In dissipative systems, the trajectories converge to some subspace regardless of the initial conditions. This subspace is called attractor of the system and has topological dimension D (integer or not) less than or equal to the Euclidean dimension m of the phase-space it lies in.

According to Takens' embedding theorem [*Takens*, 1981], if $x(t)$ is a scalar time series in discrete time, obtained from a continuous time multidimensional deterministic system with an attractor contained in a manifold of dimension D, there exists an embedding dimension $n \le 2D + 1$ such that the vectors with time-delayed coordinates $\mathbf{X} = \{x(t), x(t - \tau), \ldots, x(t - (n - 1)\tau)\}$, where τ is a properly chosen time delay, will trace out a trajectory that represents a smooth coordinate

transformation (diffeomorphism) of the original trajectory of the system. Hence, the trajectory of the delay vectors will have the same topological dimension as the underlying attractor of the dynamical system.

The Takens theorem allows for the reconstruction of the dynamics of the system using a single scalar observable $x(t)$. In particular, the application of the theorem provides a method for detecting determinism in a time series and revealing the underlying dynamics, if any, of the system that produces this time series. The method is applied for several increasing values of the embedding dimension n, and for each n the dimension D_q is calculated for some q (most frequently for $q = 2$). If the system is deterministic, then D_q becomes invariant for increasing n and saturates to a constant value, which is the dimension of the underlying attractor. If the system is random no such saturation value of the dimension is observed. Thus, the detection of a saturation value of the dimension D_q provides evidence that the system is deterministic rather than stochastic.

The application of the time-delay embedding method requires numerous points of the time series and a proper selection of the time delay τ. These issues are discussed by *Tsonis* [1992, p. 151, 162] and *Tsonis et al.* [1993] among others. In particular, the time delay τ must be chosen so as to result in points that are not correlated to previously generated points [*Tsonis et al.*, 1993]. Commonly the time delay is chosen by means of the autocorrelation function, that is, in a way that the autocorrelation function attains a certain small value like $1/e$, 0.5 [*Schuster*, 1988] or 0.1 [*Tsonis and Elsner*, 1988]. Another approach employs the concept of mutual information [*Fraser and Swinney*, 1986]. Given that the mutual information of $x(t)$ and $x(t - \tau)$ measures the general dependence between these values, it can provide a good criterion for choosing the proper time delay τ, so that the delay coordinates used for the reconstruction be independent. If the time delay τ is chosen to coincide with the first minimum of the mutual information, then the recovered state vector $\mathbf{X}$ will consist of components that possess minimal mutual information among them. Moreover, *Fraser and Swinney* [1986] argue that mutual information may be more appropriate than the autocorrelation function in the case of nonlinear dynamics, because the autocorrelation function measures only linear dependence, whereas mutual information measures general dependence between successive points. However, neither method should be applied blindly. For example, mutual information can have early minima, possibly from periodicities (A.-A. Carsteanu, personal communication, 1996). Thus a very reassuring practice is experimenting with various values of τ and addressing possible effects of the choice of τ [*Tsonis et al.*, 1993].

4. Application of the Time-Delay Embedding Method to the Rainfall Data

As mentioned before, the time-delay embedding method for embedding dimension up to 32 was applied to both the historic and the synthetic time series generated by the stochastic model (see appendix). The synthetic record was generated with length, time resolution, and depth resolution equal to those of the historic record, that is 6 years, 0.25 hours, and 1 mm, respectively. Then, both time series (historic and synthetic) were aggregated to the coarser resolutions of 1, 6, and 24 hours. Analysis was performed for all four resolutions.

The general characteristics of the historic and synthetic data used in the analysis are shown in Table 1. The available number of points of the time series varies from about 2,200 for the 24-hour resolution to about 210,000 for the 0.25-hour resolution. However, in the calculations we did not enter more than 70,000 points in order to obtain the results in reasonable computer time. We emphasize that the correlation integral calculations are time consuming; for example, the calculation of the correlation dimension for a single embedding dimension with 210,000 points requires more than 1 day in a Hewlett-Packard 9000/730 workstation.

Table 1

For the choice of the time delay, both methods based on the autocorrelation function and on the mutual information were used, which resulted in very different values of the time delay. The former indicated a time delay varying from 1 to 6 days depending on the time resolution of the rainfall series. The latter led to time delays of about 12 days (for time resolution of 0.25 and 24 hours) or more (for intermediate time resolutions). In the subsequent investigation we used two values of the time delay for each time resolution as resulted from both methods, which are shown in Table 1 (also fixing τ of the second method for the intermediate time resolutions to 12 days). Despite the large departure of time delays estimated by each method, the results of both cases were almost indistinguishable, which means that the choice of τ within a large interval does not affect the results in our case. Thus, it is not worth to include the results of both cases in the paper and we present those for the first case only, that is, for values of τ obtained by the autocorrelation function method.

For the calculation of the correlation integral both the straightforward algorithm and a faster box-assisted algorithm proposed by *Grassberger* [1990] were implemented after conversion to ANSI C and adaptations.

The computed correlation integrals $C_2(\varepsilon, n)$ for the four time resolutions are plotted in Figure 3 through Figure 6 versus the scale length ε. The examined embedding dimensions n were 1, 2, 4, 8, 16, and 32. As shown in these figures, for time resolutions of 1, 6, and 24 hours there is roughly a scaling region extending between the depth resolution limit (1 mm) and about double the average of the nonzero depths, whereas for the finer time resolution (0.25 hours) no scaling region emerges. One may observe that, in the latter case, the depth resolution of 1 mm is too high, as compared to the average nonzero rain depth for an interval of 0.25 hours, which is about 2 mm. Apparently, this explains why no scaling region emerges for the time resolution of 0.25 hours. As we progress to coarser time resolution (Figure 4 through Figure 6) the effect of the quantization of the rainfall depth with 1-mm resolution is diminished.

Figure 3

Figure 4

Figure 5

Figure 6

Inspecting the correlation integral curves in these figures, one can see that their lower tail (as $\varepsilon \to 0 \Rightarrow \log_{10} \varepsilon \to -\infty$) is horizontal indicating zero correlation dimension. This is due to the nonzero probability of zero rainfall, which results in numerous time-delayed vectors with all coordinates being zero. This fact is a drawback of the method application to this type of data since zero rainfall measurements can reach more than 90% of the record for a fine time resolution.

In Figure 7 we have plotted the slope of the scaling region ($\Delta [\log C_2(\varepsilon, n)] / \Delta [\log \varepsilon]$), estimated by least squares, versus the embedding dimension n (logarithmic plot). We observe that this slope increases monotonically with the increase of the embedding dimension and no saturation appears. Therefore, no low dimensional determinism is detected. However, it is

Figure 7

apparent from this figure that the rainfall time series depart from white noise as the slopes are quite lower than the characteristic slope of white noise that equals the embedding dimension n.

The most important observation in Figure 3 through Figure 7 is that the results obtained from the synthetic series (dashed lines in the plots) are quite similar to those obtained from the historic series. This indicates the good performance of the stochastic model in preserving the properties of rainfall that are concerned with the entropy and correlation integrals, properties that were not explicitly fitted to the historic data.

The time delay embedding method was also applied to the series of time intervals corresponding to an increase of rainfall depth by 1 mm. The employment of this series aimed at avoiding the zero time-delayed vectors discussed above. Similar series have been also used by *Sharifi et al.* [1990] and *Tsonis et al.* [1993]. The series of time intervals is obtained by inverting the series of cumulative rainfall depth $h(t)$, that is, by determining the function $t(h)$, and then finding the time instants corresponding to an increase of h by 1 mm. Therefore, we call it "inverse rainfall series". Linear interpolation was used for inverting the series $h(t)$, which obviously introduces error for time intervals smaller than 0.25 hours; thus, our results for those small timescales are not exact. In Figure 8, we depict the results of the correlation integral analysis for this series. No clear scaling region appears in this case. The results for the synthetic series are again quite similar to those for the historic.

Figure 8

5. The Rainfall Series as Cantorian Dust

Clearly, the results described in the previous section indicate that the time delay embedding method is not ideal for continuous rainfall records with short time resolution. As discussed previously, the problem is caused by the nonzero probability of zero rainfall. Therefore, we have attempted to process the time series in a different way, again using tools borrowed from the chaos literature.

The domination of voids (dry periods) in a rainfall time series evokes the parallelism with the Cantorian dust. More specifically, we can parallel the cumulative hyetograph of a certain period with the "devil's staircase" [*Hénon*, 1988; *Schroeder*, 1991, p.167], that is, the function that maps the interval [0,1] into itself having plateaus along all void intervals of the Cantorian dust (i.e., almost everywhere). Such an analogy can provide useful characterization and quantification of a rainfall time series and can reveal the presence or absence of characteristic timescales.

The method we have formed and applied uses the series of cumulative rain depth $h(t)$ in the finest available time resolution (i.e., 0.25-hour) as a one-dimensional series. This is a nondecreasing series starting with the value 0 at $t = 0$ and ending with the value $h(s)$, that is, the total depth of the entire period s (6 years in our case). The function $h(t)$ has a plateau along any dry time period. Dimensions of any order q may be calculated by applying the definitions cited in section 3.1 using a box counting algorithm with an appropriate selection of the measure p in each box. The boxes are time intervals of equal size $\varepsilon = k\,\Delta$, where Δ is the time resolution and k an integer. Apparently, in our situation the measure p_i for the ith box cannot be determined with the common way, that is, in terms of the number of points contained in the box, normalized by the total number of points; in fact we do not have

"points" in this formulation. Instead we can set $p_i = \Delta h_i / h(s)$, where Δh_i is the incremental rainfall depth in the ith time interval. Obviously, the measures so defined add up to unity, that is, satisfy (15). For a specific ε, we can easily estimate the entropy $I_q(\varepsilon)$ using directly (14) or (16). Repeating the procedure for several values of ε we can numerically reconstruct the function $I_q(\varepsilon)$. If its plot versus $\ln \varepsilon$ is a straight line then we can estimate the dimensions D_q using (18).

In order to validate the method we have applied it first to the devil's staircase. In that case the function $h(t)$ is a staircase that rises only at those values t that belong to the Cantor set. We know that all dimensions of this set are equal to $\ln(2) / \ln(3) = 0.63093\ldots$. This theoretical result can be also obtained empirically by the proposed method. To construct the function $h(t)$ we have used the algorithm described by *Schroeder* [1991, p.169]: for any given value t we write t as a ternary number and convert it into a binary fraction, replacing any digits 2 up to the first 1 (reading from left to right) by the digit 1; we keep the first 1 (if any) and write 0 for all following digits to the right; the obtained number is $h(t)$ written in binary form. With this algorithm we constructed 209,000 values of $h(t)$ at equidistant values of t (a number equal to the available intervals of the rainfall data set). Then we used the method described in the previous paragraph to estimate dimensions. The results, shown in Figure 9 and Figure 10 are in perfect agreement with the theoretical expectations. The numerically calculated entropies I_q plot as straight lines against $\ln \varepsilon$. All dimensions D_q, up to $q = 10$, estimated by the method are equal to about 0.63.

Figure 9

Figure 10

The application of this method to both the historic and the synthetic rainfall data sets gave the results shown in Figure 11 and Figure 12. As one can observe in Figure 11, the curve for each generalized entropy I_q, has a slope approaching 1 for large ε. This slope becomes low (close to 0.2-0.3) for intermediate values of ε, and then, for low ε it rises again, yet remaining significantly less than 1. The change of slope with ε indicates the presence of characteristic timescales in a rainfall time series, as opposed to the Cantorian dust, where the single straight line with the unique slope suggests the absence of characteristic scales. To quantify the characteristic timescales we have plotted in Figure 11 the vertical lines corresponding to the mean rainfall duration $E[D]$ and the mean rainfall interarrival time $E[T]$. These lines define three distinct regions of timescales: the coarse ($\varepsilon > E[T]$), the intermediate ($E[D] < \varepsilon < E[T]$) and the fine ($\varepsilon < E[D]$) timescales. The slope within each region is approximately constant, whereas different regions have different slopes. Thus, we can interpret these regions as distinct scaling areas. The slopes of $-I_q$ versus $\ln \varepsilon$, are about 0.5 for the fine timescale area and about 0.3 for the intermediate timescale area. These values indicate that the subset of wet time intervals looks "thinner" in the intermediate timescale than in the fine (and, obviously, the coarse) timescale.

Figure 11

Figure 12

The different slopes for changing q, estimated by the method of least squares for each scaling region and shown in Figure 12, indicate a multifractal structure for both the historic and the synthetic data sets. We must comment that the slopes of the distinct regions are not dimensions in a strict sense. The definition of dimensions demands that $\varepsilon \to 0$ and, thus, only the slope of the fine scale area may be interpreted as a dimension. Yet the estimation of this slope is affected by the time resolution limit ($\varepsilon_{min} = \Delta = 0.25$ hours), which implies uncertainty. This uncertainty is responsible for the increasing

slope in the $\varepsilon < E[D]$ plot (Figure 12) for q increasing from 0 to 1. It is known that D_q is a nonincreasing function of q, as the generalized entropy $I_q(\varepsilon)$ is also a nonincreasing function of q for all $q \geq 0$ [*Grassberger*, 1983]. As shown in Figure 11, the latter rule is never broken. However, when the estimator of D_q based on the slope of $I_q(\varepsilon)$ is applied, the empirical results may not strictly obey the rule that D_q is nonincreasing with q. This is the case in the $\varepsilon < E[D]$ plot of Figure 12 for $q = 0$ and 1. However, in all other cases this rule is not broken.

A final remark, which is very important for the objective of this study, is that the results for the synthetic data agree well with those of the historic data, as is apparent in both Figure 11 and Figure 12. It is impressive that even the break of the rule that D_q is nonincreasing with q, discussed in the previous paragraph, occurs simultaneously for both the historic and synthetic data at the same point.

6. Conclusions

The main conclusion of this investigation is that a synthetic continuous rainfall series generated by a well-structured stochastic rainfall model may be practically indistinguishable from a historic rainfall series, even if we use tools of chaotic dynamics theory to characterize and compare the two rainfall series. This conclusion defends the use of stochastic models in Engineering Hydrology for simulation purposes. We must highlight, though, that our conclusion is constrained by the limited data used (only a 6-year record from Ortona, Florida) and by our macroscopic view of the time series as a continuous record that includes numerous dry periods. This result may not be applicable for a single rainfall event, in which case many researchers have detected low-dimensional chaotic dynamics. In our case, no determinism has been detected in the examined continuous rainfall series from time resolutions of 0.25 to 24 hours and dimensions up to 32. Obviously however, both the historic and synthetic rainfall records are distinguished from white noise.

Phase-space reconstruction via time delay embedding, applied to continuous rainfall records with low time resolution, does not provide effective characterization of the time series. The obstruction is caused by the nonzero probability of having zero rainfall in a short time interval, resulting in many zero time-delayed vectors. Moreover, the application of the time delay embedding method is extremely time consuming. We note characteristically that the computational time needed to generate the synthetic rainfall series in our application was about 3 orders of magnitude smaller than that required to apply the time delay embedding method. It is anticipated that time delay embedding may be applied without problems for timescales considerably larger than the mean dry time (e.g., monthly timescale), in which case the problem of nonzero probability of zero rainfall is eliminated. However, this would require records of hundreds of years in order to obtain reliable results.

The proposed Cantorian dust analogue eliminates the problem of nonzero probability and gives useful characterization of rainfall time series. This method reveals the presence of two characteristic timescales in a rainfall series, which may be quantified by the average rainfall duration and the average rainfall interarrival time. In addition, the analogue quantifies the examined rainfall series with a fractal dimension of about 0.5 for short timescales. However, a multidimensional exten-

sion of this method using time-delayed vectors is not straightforward.

Appendix: Generation of Synthetic Data by the Scaling Model

The model for the generation of the synthetic time series consists of two separate modules: the alternating renewal model and the scaling model of storm hyetograph. The former is used to generate storm durations and dry times and the latter to generate incremental depths within a storm. The scaling model and its parameter estimation has been discussed in section 2. The alternating renewal model consists of a set of distribution functions for storm durations and dry times with their associated parameters. The dry times and rain durations are assumed independent variables. Different parameters were estimated for each month using the historic data. Specifically, the exponential distribution was assumed for the rain durations of every month. The single parameter of the exponential distribution was estimated from the mean storm duration of each month. For the dry times of the dry season (October through May) a Weibull distribution was assumed and its parameters were estimated by the method of moments. For the dry times of the wet season (June through September) it was found that no typical simple distribution function in the literature was suitable for the data and a two-segment Weibull distribution was adopted, which agreed perfectly with the data. This is described by the equation

$$F_B(b) = \begin{cases} 1 - \exp[-\kappa_1 (b - c)^{\lambda_1}], & b \le \xi \\ 1 - \exp[-\kappa_2 (b - c)^{\lambda_2}], & b \ge \xi \end{cases} \tag{27}$$

where b is the dry time, κ_1, κ_2, λ_1, λ_2 and ξ are parameters estimated from the data and c is the minimum allowed dry time (in our case 7 hours). As noted before, dry times smaller than c are included in storm events.

The generation of the time series takes place in three phases. In phase A we apply the alternating renewal model for the temporal location of an event, that is, we generate a rain duration and the consecutive dry time using the adopted distributions for each variable and the appropriate parameters for the month where the event is located.

In phase B we calculate the statistics of total and incremental depths for each event of given duration D, and formulate the generating scheme. This phase includes the following steps:

B1. We calculate the statistics of the total depth $E[Z]$ and $\mathrm{Var}[Z]$ (equations (6) and (7)) and $E[\mathbf{X}]$ and $\mathrm{Cov}[\mathbf{X}, \mathbf{X}]$ (equations (9), (10) and (11)), where $\mathbf{X}$ denotes the vector $[X_1, X_2, \ldots, X_k]^T$, and assume a gamma conditional distribution function for Z and X_i, given the duration D.

B2. We formulate a sequential generating scheme as $\mathbf{X} = \mathbf{\Omega}\,\mathbf{V}$, where $\mathbf{\Omega}$ is a matrix of coefficients and $\mathbf{V}$ is a vector of independent variates, which are assumed to have approximately a three-parameter gamma distribution.

B3. We estimate the parameters of the generating scheme, that is, the coefficient matrix $\mathbf{\Omega}$, and the statistics of V_i. For the former we use the equation $\mathbf{\Omega}\,\mathbf{\Omega}^T = \mathrm{Cov}[\mathbf{X}, \mathbf{X}]$ and then we perform lower triangular decomposition to obtain $\mathbf{\Omega}$ from $\mathbf{\Omega}\,\mathbf{\Omega}^T$. For the latter we use the equations

$$E[V_i] = \frac{1}{\omega_{ii}}\left\{ E[X_i] - \sum_{j=1}^{i-1} \omega_{ij}\, E[V_j] \right\} \tag{28}$$

$$\mathrm{Var}[V_i] = 1 \tag{29}$$

$$\mu_3[V_i] = \frac{1}{\omega_{ii}^3}\left(\mu_3[X_i] - \sum_{j=1}^{i-1} \omega_{ij}^3 \mu_3[V_j] \right) \tag{30}$$

where $\mu_3[X_i]$ and $\mu_3[V_i]$ are the third central moments of X_i and V_i, respectively. Here $\mu_3[X_i]$ is obtained from the assumed gamma conditional (on D) distribution of X_i. The procedure of phase B is documented in detail elsewhere [*Koutsoyiannis and Tsakalias*, 1992; *Koutsoyiannis*, 1994; *Mamassis et al.*, 1994].

Finally, in phase C we generate the sequence of incremental depths for each event. This phase has also three steps:

C1. We generate the total depth Z, from the assumed gamma distribution.

C2. We apply the sequential generating scheme to obtain an initial sequence of incremental depths $\mathbf{X}'$.

C3. We determine the final adjusted sequence using the adjusting procedure

$$X_i = \left(X_i' / \sum_{j=1}^{k} X_j' \right) Z \tag{31}$$

The documentation and proof of the appropriateness of the adjusting procedure is given elsewhere [*Koutsoyiannis*, 1994].

Acknowledgments. We would like to thank N. Mamassis for his assistance in data processing and A. Manetas for his help in programming. We also thank U. Hübner (Phys.-Techn. Bundesanstalt, D-3300 Braunschweig, Germany) for providing us with a FORTRAN code used for comparisons and checking of our programs, and A. Fraser for providing us with a code used for the mutual information calculations. We are grateful to the Guest Editor E. Foufoula-Georgiou, A.-A. Carsteanu, and two anonymous reviewers for their constructive comments. Computational resources were provided by the National Technical University of Athens (Central Computer Department and HYDROSCOPE Computer Network).

References

Fraser, A., and H. Swinney, Independent coordinates for strange attractors from mutual information, *Phys. Rev. A*, *33*(2) 1134-1140, 1986.

Gershenfeld, N. A., and A. S. Weigend, The future of time series: Learning and understanding, in *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, pp. 1-70, SFI Stud. in the Sci. of Complex., Proc. Vol. XV, Addison-Wesley, Reading, Mass. 1993.

Grassberger, P., Generalized dimensions of strange attractors, *Phys. Lett.*, *97*A(6), 227-230, 1983.

Grassberger, P., An optimized box-assisted algorithm for fractal dimensions, *Phys. Lett. A*, *148*(1,2), 63-66, 1990.

Grassberger, P., and I. Procaccia, Characterization of strange attractors, *Phys. Rev. Lett.*, *50*(5), 346-349, 1983.

Hénon, M., Chaotic scattering modelled by an inclined billiard, *Physica*, *D* 23, 132-156, 1988.

Hübner, U., C. O. Weiss, N. Abraham, and D. Tang, Lorenz-like chaos in NH_3-FIR lasers, in *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, SFI Studies in the Sciences of Complexity, Proc. Vol. XV, pp. 73-105, Addison-Wesley, 1993.

Huff, F. A., Time distribution of rainfall in heavy storms, *Water Resour. Res.*, *3*(4) 1007-1019, 1967.

Jayawardena, A. W., and F. Lai, Analysis and prediction of chaos in rainfall and stream flow time series, *J. Hydrol., 153*, 23-52, 1994.
Koutsoyiannis, D., A stochastic disaggregation method for design storm and flood synthesis, *J. Hydrol., 156*, 193-225, 1994.
Koutsoyiannis, D., and E. Foufoula-Georgiou, A scaling model of storm hyetograph, *Water Resour. Res., 29*(7), 2345-2361, 1993.
Koutsoyiannis, D., and G. Tsakalias, A disaggregation model for storm hyetographs, paper presented at the 3rd meeting of the AFORISM project, EC, DG XII, Athens, Greece, 1992.
Mamassis, N., D. Koutsoyiannis, and E. Foufoula-Georgiou, Stochastic rainfall forecasting by conditional simulation using a scaling model, XIX EGS General Assembly, Grenoble, abstract, *Ann. Geophys., 12*, Suppl. II, part II, C324,1994.
Rényi, A., *Probability Theory*, North-Holland, Amsterdam,1970.
Restepo-Posada, P. J., and P. S. Eagleson, Identification of independent rainstorms, *J. Hydrol., 55*, 303-319, 1982.
Rodriguez-Iturbe, I., Exploring complexity in the structure of rainfall, *Adv. Water Resour., 14*(4), 162-167, 1991.
Rodriguez-Iturbe, I., B. F. de Power, M. B. Sharifi, and K. P. Georgakakos, Chaos in Rainfall, *Water Resour. Res., 25*(7), 1667-1675, 1989.
Ross, S. M., *Introduction to Probability Models*, 556 pp., Academic, San Diego, Calif., 1993.
Schroeder, M., *Fractals, Chaos and Power Laws*, 429 pp., W. H. Freeman, New York, 1991.
Schuster, H. G., *Deterministic Chaos*, VCH, New York, 1988.
Sharifi, M. B., K. P. Georgakakos, and I. Rodriguez-Iturbe, Evidence of deterministic chaos in the pulse of storm rainfall, *J. Atmos. Sci., 45*(7), 888-893, 1990.
Takens, F., Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young, lecture notes in Mathematics, 898, pp. 336-381, Spinger-Verlag, New York, 1981.
Tsonis, A. A., *Chaos: From Theory to Applications*, 274 pp., Plenum, New York, 1992.
Tsonis, A. A., and J. B. Elsner, The weather attractor over very short time scales, *Nature*, 33, 545-547, 1988.
Tsonis, A. A., J. B. Elsner, and K. Georgakakos, Estimating the dimension of weather and climate attractors: Important issues on the procedure and interpretation, *J. Atmos. Sci., 50*(15) 2249-2555, 1993.

D. Koutsoyiannis and D. Pachakis, Department of Water Recourses, Faculty of Civil Engineering, National Technical University, Heroon Polytechniou 5, GR-157 80 Zografou, Greece. (e-mail: dk@hydro.civil.ntua.gr; dpach@hydro.civil.ntua.gr)

(Received October 16, 1995; revised April 16, 1996;
accepted April 25, 1996.)

Paper number 96JD01389.
0148-0227/96/96JD-01389$09.00

Tables

Table 1. Length of the Time Series and Time Delays Adopted for the Four Resolutions

Time Resolution, hr	Historic Record Length, Number of points	Synthetic Record Length, Number of points	Time Delay Estimated From Autocorrelation Function, days	Time Delay Estimated From Mutual Information, days
0.25	209,580[1]	212,607[1]	1	12
1	52,395	53,152	2	12
6	8,732	8,858	3	12
24	2,183	2,214	6	12

[1] 70,000 points were used for calculations.

Figure Captions

Figure 1. Mean ($E[Z]$) and standard deviation (Std[Z]) of the total storm depth (Z) versus storm duration (D).

Figure 2. Mean ($E[X]$) and standard deviation (Std[X]) of the incremental, one-quarter hour, storm depth (X) as a function of the storm duration (D).

Figure 3. Log-log plot of correlation integral C_2 versus the radius ε for embedding dimensions n up to 32 and for time resolution of the series equal to one-quarter hours. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 4. Log-log plot of correlation integral C_2 versus the radius ε for embedding dimensions n up to 32 and for time resolution of the series equal to 1 hour. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 5. Log-log plot of correlation integral C_2 versus the radius ε for embedding dimensions n up to 32 and for time resolution of the series equal to 6 hours. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 6. Log-log plot of correlation integral C_2 versus the radius ε for embedding dimensions n up to 32 and for time resolution of the series equal to 24 hours. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 7. Slope of the scaling region of Figures 3-6 versus embedding dimension, also compared with the correlation dimension of white noise (logarithmic plot). The solid lines represent the historic data and the dashed lines the synthetic.

Figure 8. Log-log plot of correlation integral C_2 of the inverse rainfall series versus the radius ε for embedding dimensions n up to 32. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 9. Generalized entropy I_q of Cantorian dust as a function of the interval length ε, as was numerically calculated to validate the proposed Cantorian dust analogue method.

Figure 10. Generalized dimensions D_q of Cantorian dust as estimated numerically from Figure 9 for orders q up to 10, in comparison with the theoretical dimensions.

Figure 11. Generalized entropy I_q of the rainfall series, calculated with the Cantorian dust analogue method, as a function of the interval length ε. The solid lines represent the historic data and the dashed lines represent the synthetic.

Figure 12. Slopes of the three distinct scaling regions of Figure 11 for orders q up to 10. The solid lines represent the historic data and the dashed lines represent the synthetic.

Figure 1. Mean ($E[Z]$) and standard deviation (Std[Z]) of the total storm depth (Z) versus storm duration (D).

Figure 2. Mean ($E[X]$) and standard deviation (Std[X]) of the incremental, one-quarter hour, storm depth (X) as a function of the storm duration (D).

Figure 3. Log-log plot of correlation integral C_2 versus the radius ε for embedding dimensions n up to 32 and for time resolution of the series equal to one-quarter hours. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 4. Log-log plot of correlation integral C_2 versus the radius ε for embedding dimensions n up to 32 and for time resolution of the series equal to 1 hour. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 5. Log-log plot of correlation integral C_2 versus the radius ε for embedding dimensions n up to 32 and for time resolution of the series equal to 6 hours. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 6. Log-log plot of correlation integral C_2 versus the radius ε for embedding dimensions n up to 32 and for time resolution of the series equal to 24 hours. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 7. Slope of the scaling region of Figures 3-6 versus embedding dimension, also compared with the correlation dimension of white noise (logarithmic plot). The solid lines represent the historic data and the dashed lines the synthetic.

Figure 8. Log-log plot of correlation integral C_2 of the inverse rainfall series versus the radius ε for embedding dimensions n up to 32. The solid lines represent the historic data and the dashed lines the synthetic.

Figure 9. Generalized entropy I_q of Cantorian dust as a function of the interval length ε, as was numerically calculated to validate the proposed Cantorian dust analogue method.

Figure 10. Generalized dimensions D_q of Cantorian dust as estimated numerically from Figure 9 for orders q up to 10, in comparison with the theoretical dimensions.

Figure 11. Generalized entropy I_q of the rainfall series, calculated with the Cantorian dust analogue method, as a function of the interval length ε. The solid lines represent the historic data and the dashed lines represent the synthetic.

Figure 12. Slopes of the three distinct scaling regions of Figure 11 for orders q up to 10. The solid lines represent the historic data and the dashed lines represent the synthetic.

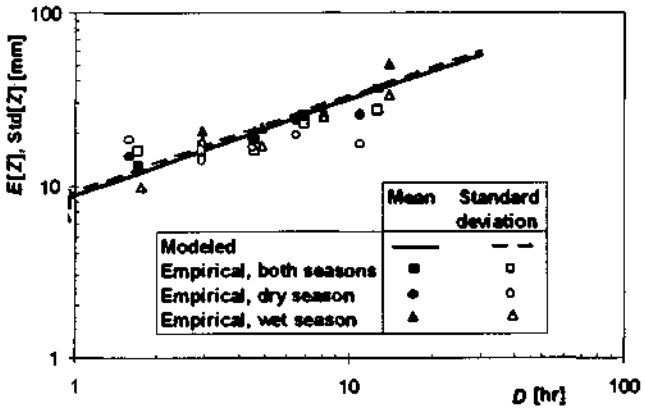

Figure 1

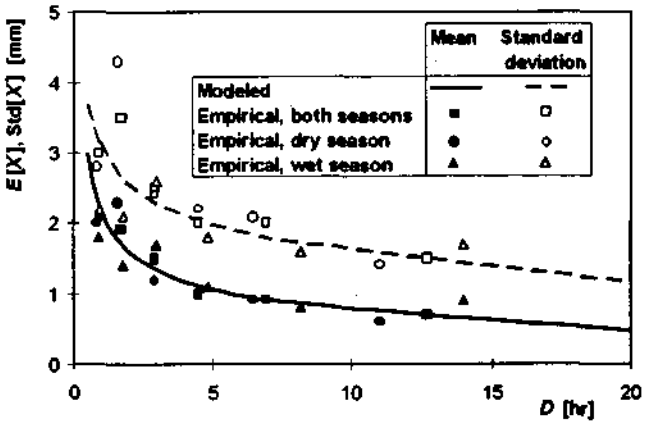

Figure 2

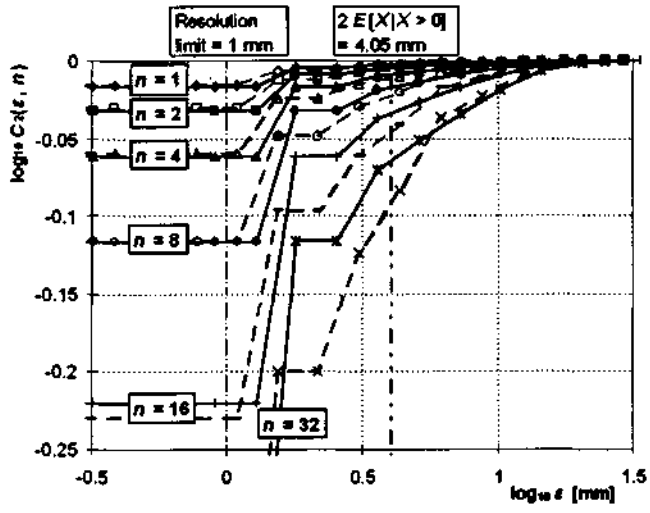

Figure 3

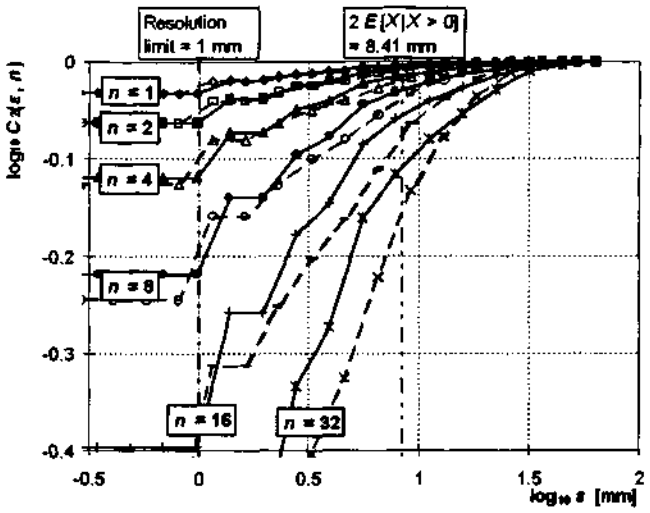

Figure 4

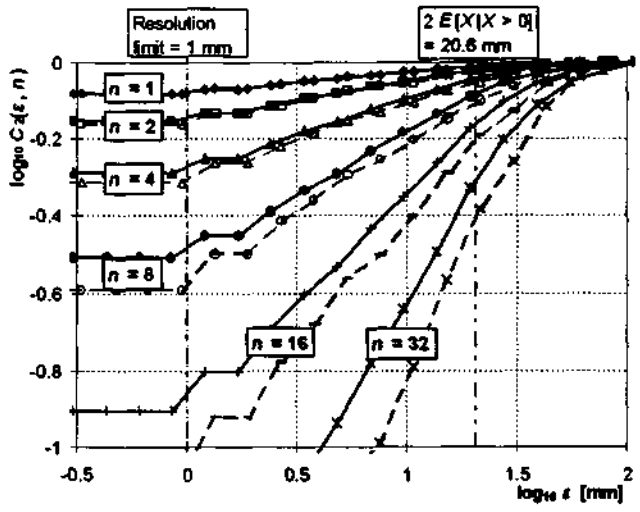

Figure 5

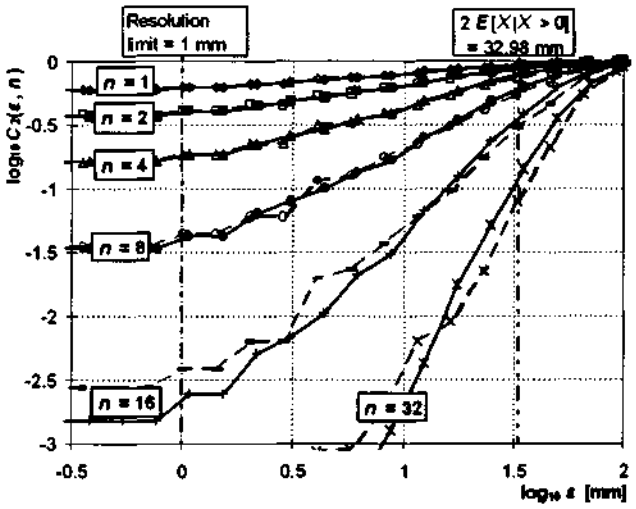

Figure 6

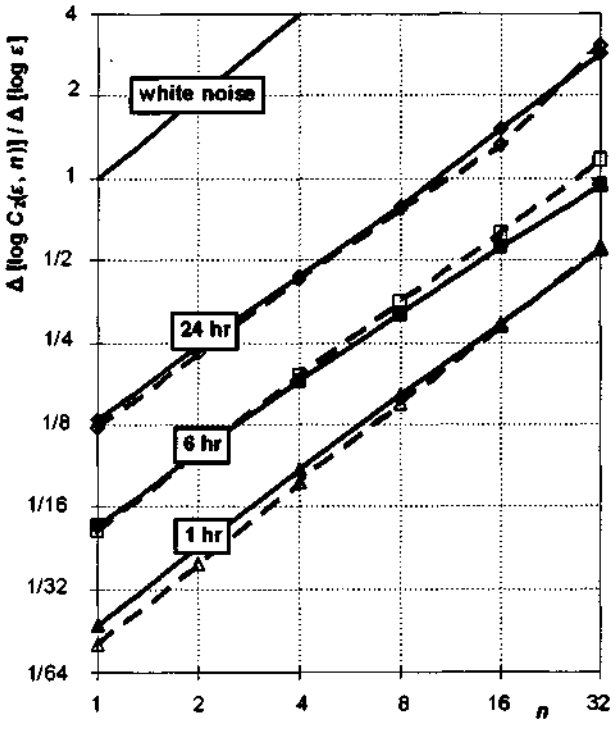

Figure 7

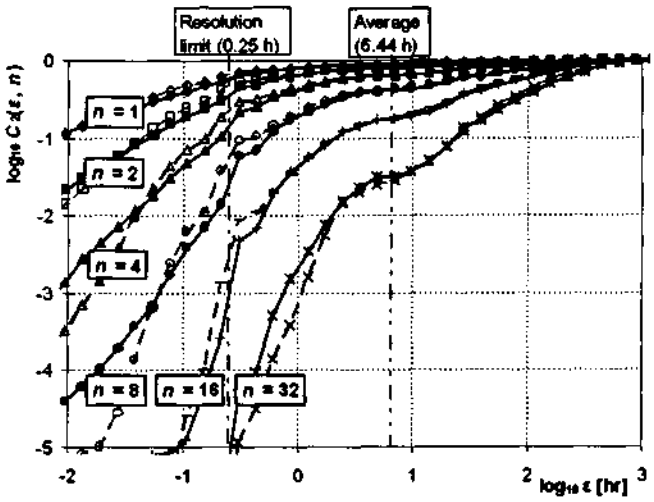

Figure 8

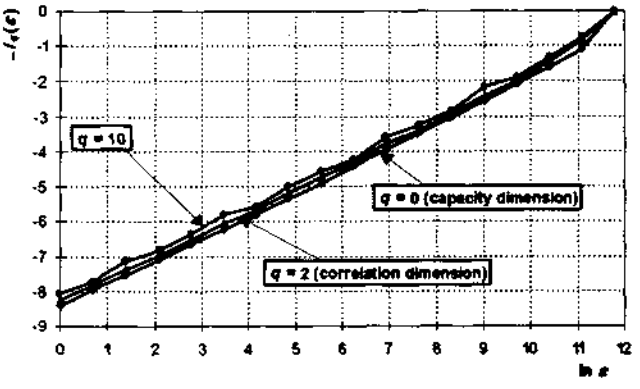

Figure 9

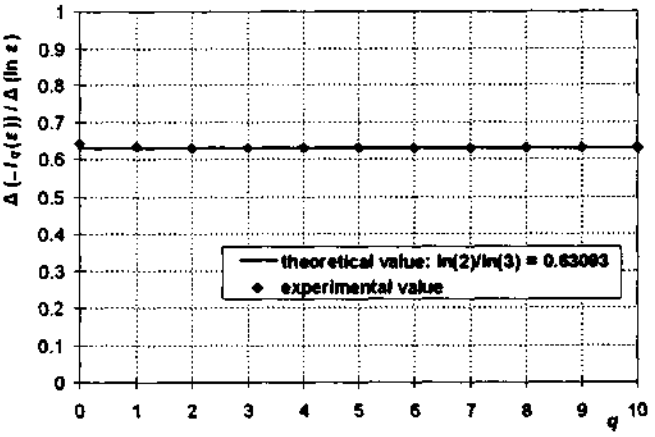

Figure 10

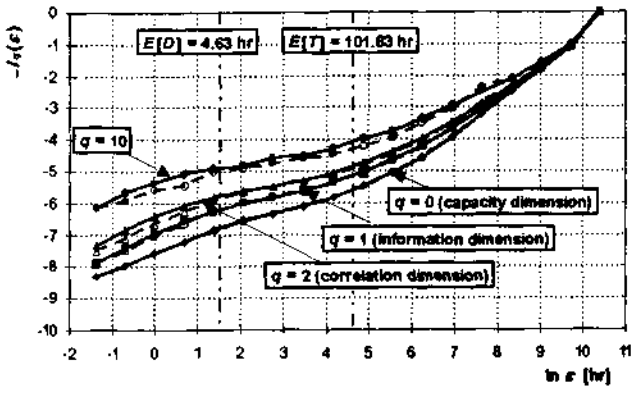

Figure 11

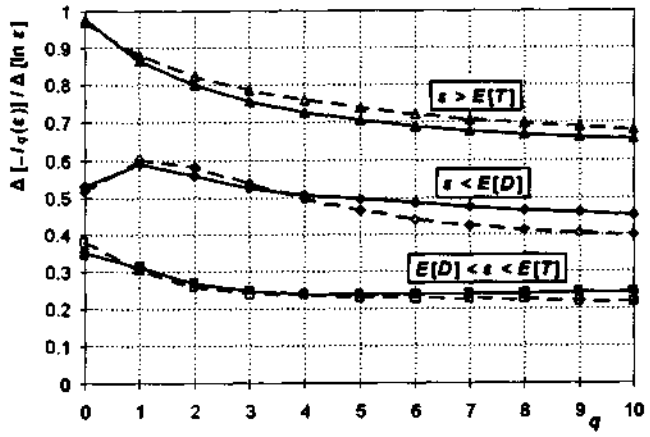

Figure 12