DETERMINISTIC CHAOS VERSUS STOCHASTICITY IN ANALYSIS AND MODELING OF RAINFALL STRUCTURE

Demetris Koutsoyiannis and Demetris Pachakis
Department of Water Resources, Faculty of Civil Engineering,
National Technical University of Athens, Greece.

Extended abstract

In recent years, new methods for time series analysis, derived from chaos theory, yielded some fascinating results. Among them, the fact that deterministic processes may sometimes be statistically indistinguishable from random noise. Currently, discussion is going on about how to characterize a process and how to recognize whether it is stochastic or deterministic. A popular method for revealing the underlying dynamics, if any, from a time series, is the phase-space reconstruction via time delay embedding. According to this method, the state of a system, as seen from an observable, is approximated by a number of previous values of this observable. A critical issue is how many of these values are necessary, in order to capture efficiently the evolution of the system. In other words, how many variables we need to describe the phase-space in which the phenomenon evolves. The least necessary number of these variables is related to the dimension of the attractor of the system, if such an attractor is identified. The capacity and correlation dimension quantify the nature of the attractor. A variety of studies have shown that in deterministic processes there exists a finite dimension [dk1]while in stochastic processes there is no such a finite value. Therefore the estimation of this dimension provides a way of detecting determinism in a time series.

In recent studies, it was found that the correlation dimension of rainfall time series is finite, whereas simple stochastic models (such as ARMA) that are often used to simulate the rainfall process exhibit very different behavior. This indicates the presence of deterministic chaos in the rainfall process. Despite this, the stochastic modeling of the rainfall process still dominates in operational hydrology as it gives sufficient solutions to most problems.

Our objective in this study is similar to that of several previous studies: to detect deterministic chaos in historic rainfall series and compare the results with those obtained from synthetic series generated by stochastic models. The difference is that we use a more detailed stochastic model to simulate the rainfall process instead of the simple ARMA models. Specifically, we use the scaling model of storm hyetograph†. This is a simple, parameter parsimonious model for the temporal evolution of rainfall, which limits the need for a stationarity assumption only within the same storm event. The basic hypothesis behind it, is that the process of instantaneous rainfall intensities within a storm is a self-similar (simple scaling) process that scales with the duration of the storm. The model explains reasonably well the dependence of the total storm depth and of the internal structure of a storm on the storm duration. To form a complete rainfall generator, this model was combined with a stochastic process describing storm arrivals, durations and total depths. It is noteworthy that self-similarity that characterizes the stochastic scaling model is often recognized behind chaotic patterns. However, at the generation stage the model behaves like a purely linear stochastic model, which for each storm adapts its parameters according to the storm duration.

The historic data set used for the study consists of incremental rainfall depths, measured every quarter of an hour, for an observation period of six years (1984-89), at a station in Ortona, Florida. The synthetic data set was generated using the scaling model, after proper fit. The comparison of the two sets is done by means of time delay embedding and capacity and correlation dimension estimates of the two time series.

† Koutsoyiannis, D. and E. Foufoula-Georgiou, A scaling model of storm hyetograph, *Water Resour. Res.*, 29(7), 2345-2361, 1993

Introduction

- Studies of the structure of **particular storm events** (*Rodriguez-Iturbe et al*., 1989; *Sharifi et al*., 1990; *Rodriguez-Iturbe*, 1991) provided evidence that the temporal evolution of a storm may be characterized as a deterministic chaotic process with low-dimensional strange attractor.
- Similar results are obtained from simultaneous study of **several events** of the same meteorological convective character (*Tsonis*, 1992, p. 169; *Tsonis et al*., 1993).
- The results are not conclusive for **continuous rainfall records at a certain time resolution**:
  - *Rodriguez-Iturbe et al.* (1989), and *Rodriguez-Iturbe* (1992) do not detect low-dimensional chaotic dynamics in weekly rainfall data of Genoa covering a period of 148 years.
  - *Jayawardena and Lai* (1994) detect high-dimensional chaotic behavior (for embedding dimension between 30 and 40) in daily rainfall at three rain gages in Hong Kong covering a period of 11 years. They conclude that rainfall series could be better modeled by methods of chaos theory, such as time delay embedding, than by traditional stochastic models, such as ARMA.
- The detection of chaotic behavior in a rainfall time series has led many researchers to interpret rainfall as a deterministic process rather than a stochastic process. However:
  - The boundary between a deterministic and a stochastic process is not clear.
  - Stochastic models can incorporate deterministic components, if any. In addition, deterministic models are not generally free of random noise.
  - There are difficulties in the operational use of deterministic models, whereas stochastic models have been used operationally for several purposes (e.g. simulation).
  - Deterministic models do not necessarily improve predictions due to sensitive dependence on initial conditions. This becomes clearer in the case of high-dimensional chaotic behavior.
- A different approach was suggested by *Koutsoyiannis and Foufoula-Georgiou* (1993) who took advantage of revealed scaling properties in rainfall data to build a **stochastic scaling model of storm hyetograph**. This model describes and parametrizes a **population of storms**, not the structure of a particular storm.

Objectives, methodology and data used

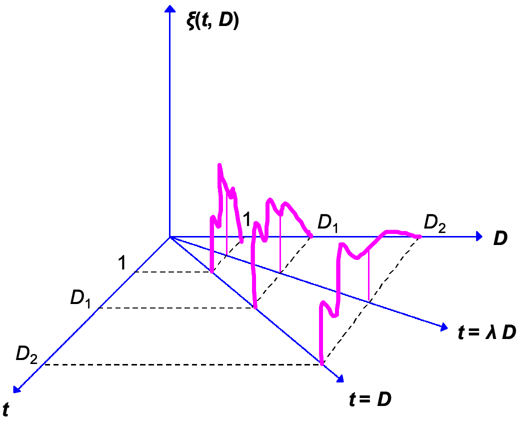
- It is not a critical issue to distinguish the real rainfall process from simple stochastic models such as white noise or ARMA processes. Obviously, there are differences between the real process and this kind of models.
- The **important question** is if there are essential differences that distinguish the real rainfall process from a well-structured stochastic model, capable of preserving important properties of the rainfall process such as intermittency, seasonality, scaling behavior etc.
- **Other questions** relevant to this issue are:
  - How can typical descriptors of chaotic behavior, such as capacity, information and correlation dimensions, and typical methods, such as time delay embedding, can be used to characterize the rainfall process?
  - Are there any characteristic scales in a continuous rainfall record, or not?
- The **methodology** adopted consists of the following:
  - Selection of a historic data set (six years (1984-89) of incremental rainfall depths, measured every quarter of an hour at station Ortona Lock 2, Florida, USA).
  - Adoption and calibration of a stochastic model (the scaling model of storm hyetograph coupled with an alternating renewal model for modeling rain durations and dry times).
  - Generation of a synthetic record with an equal length (six years) using the stochastic model.
  - Computations and comparisons of various descriptors of chaotic dynamics for both the historic and the synthetic data sets.

Summary of the scaling model of storm hyetograph

- **Main hypothesis**: $\{\xi(t,D)\} =_d \{\lambda^{-H}\xi(\lambda t, \lambda D)\}$
  where $\xi()$ = instantaneous rainfall intensity, D = duration of the event, t = time ($0 \le t \le D$), and H = scaling exponent.
- **Secondary hypothesis**: Weak stationarity (= stationarity within the event), resulting in

$$E[\xi(t,D)] = c_1 D^H$$

$$E[\xi(t,D)\xi(t+\tau,D)] = \phi(\tau/D)D^{2H} \text{ where } \varphi(\tau/D) = k(\tau/D)^{-\beta}$$



- **Statistics of total depth**, Z

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

$$Var[Z] = c_2 D^{2(H+1)}, \text{ where } c_2 = 2k/[(1-\beta)(2-\beta)] - c_1^2$$

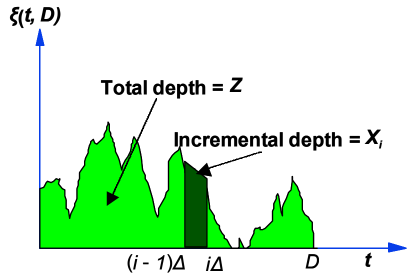
- **Statistics of incremental depth**, X

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

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

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

where $\delta = \Delta / D$, $f(m,\beta) = \frac{1}{2}\left[(m-1)^{2-\beta} + (m+1)^{2-\beta}\right] - m^{2-\beta} \quad (m > 0)$

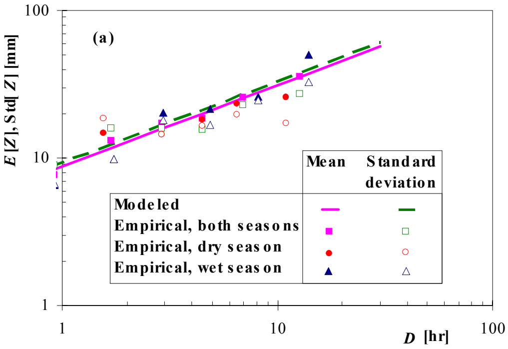


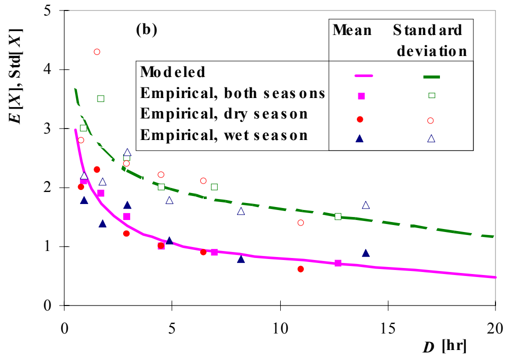
- **Model parameters**

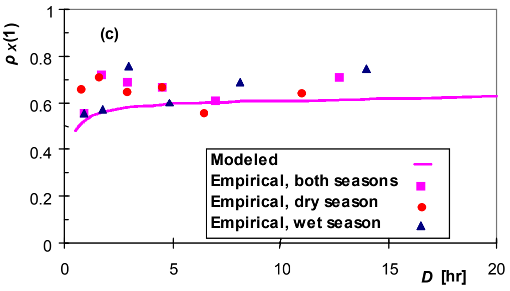
H	scaling exponent	estimated from $E[Z] = c_1 D^{H+1}$ (by least squares)
c_1	mean value parameter	
c_2	variance parameter	estimated from $c_2 = Var[Z]/D^{2(H+1)}$
β	correlation decay parameter	estimated from $\beta = 1 - \dfrac{\ln\left(E[X_i X_{i+1}]/E[X_i^2] + 1\right)}{\ln 2}$

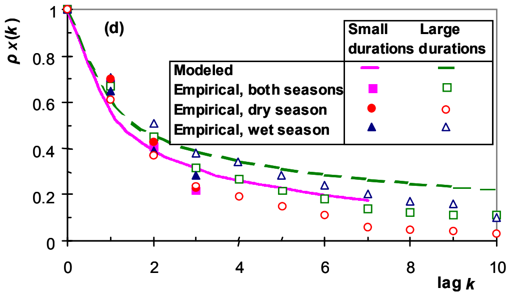
Fitting of the scaling model of storm hyetograph

- The scaling model was fitted to the Ortona data set using a time resolution Δ =1/4 hr.
- All storm events of the six-year period with durations greater than Δ were used (426 events with durations ranging from 1/2 to 35 hr). Other 37 events with durations Δ (or less) were modeled separately. Events were allowed to include periods of zero rainfall less than 7 hr (≈ 1.5 times the average duration).
- The fitting procedure was based on the separation of storms into six classes according to their durations as described by *Koutsoyiannis and Foufoula-Georgiou* (1993).
- Analysis was first performed by separating storms into two seasons (wet: Jun.-Sep.; dry: Oct.-May).
- For simplicity a unique set of parameters was finally adopted for both seasons (H = –0.449, c_1 = 8.74, c_2 = 85.68, β = 0.246).
- Notable is the departure of H from zero, which indicates the departure of the process from stationarity.
- The comparison of the model with data of wet, dry and both seasons is given in Figures **(a)** (statistics of total depth), **(b)** (statistics of incremental depth), **(c)** (lag-1 correlation coefficient of incremental depth) and **(d)** (lag-*k* correlation coefficient of incremental depth).









Basic concepts of chaos theory

- **Generalized entropy** $I_q(\varepsilon)$ (*Rényi*, 1970) of a set (subset of an *n*-dimensional metric space) partitioned into $M(\varepsilon)$ boxes (=hypercubes) with length scale ε:

$$I_q(\varepsilon) = \frac{1}{1-q} \ln \sum_{i=1}^{M(\varepsilon)} p_i^q, S_q(\varepsilon) = \exp[-I_q(\varepsilon)] = \left(\sum_{i=1}^{M(\varepsilon)} p_i^q \right)^{\frac{1}{q-1}}, \quad q \neq 1$$

$$I_1(\varepsilon) = -\sum_{i=1}^{M(\varepsilon)} p_i \ln p_i, S_1(\varepsilon) = \exp[-I_1(\varepsilon)] = \exp\left(\sum_{i=1}^{M(\varepsilon)} p_i \ln p_i \right), q = 1$$

where p_i is a measure of the part of the set contained in the *i*th box, such that $\sum_{i=1}^{M(\varepsilon)} p_i = 1$

For a set consisting of *N* (*n*-dimensional) points: $p_i = N_i / N$, where N_i = the number of points contained in the *i*th box

- **Generalized dimensions** of the set (*Grassberger*, 1983)

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

If $S_q(\varepsilon)$ is a power law then

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

- **Special cases:**

$q = 0:$	$I_0(\varepsilon) = \ln M'$	$S_0(\varepsilon) = 1/M'$	$D_0 = \lim_{\varepsilon \to 0} \frac{-\ln M'}{\ln \varepsilon}$	capacity (or fractal) dimension
$q = 1:$	$I_1(\varepsilon) = -\sum_{i=1}^{M(\varepsilon)} p_i \ln p_i$	$S_1(\varepsilon) = \exp\left(\sum_{i=1}^{M(\varepsilon)} p_i \ln p_i \right)$	$D_1 = \lim_{\varepsilon \to 0} \frac{-\sum_{i=1}^{M(\varepsilon)} p_i \ln p_i}{\ln \varepsilon}$	information dimension
$q = 2:$	$I_2(\varepsilon) = -\ln \sum_{i=1}^{M(\varepsilon)} p_i^2$	$S_2(\varepsilon) = \sum_{i=1}^{M(\varepsilon)} p_i^2$	$D_2 = \lim_{\varepsilon \to 0} \frac{\ln \sum_{i=1}^{M(\varepsilon)} p_i^2}{\ln \varepsilon}$	correlation dimension

where M' = the number of boxes that intersect the set.

- **Correlation integral of order-*q*** (integer $q \geq 2$) for a set consisting of *N* points (*Grassberger*, 1983)

$$C_q(\varepsilon) = N^{-q} \left\{ \text{number of } q\text{-tuples } \left(x_{j_1}, \ldots, x_{j_q}\right) \text{with all } \left| x_{j_s} - x_{j_r} \right| < \varepsilon \right\}$$

Basic property: $C_q(\varepsilon) \approx S_q(\varepsilon)$

- **Special case: Correlation integral of order-2** or simply **correlation integral**

$$C_2(\varepsilon) = N^{-2} \left\{ \text{number of pairs } \left(x_j, x_l\right) \text{with } \left| x_j - x_l \right| < \varepsilon \right\}$$

This is calculated more easily and accurately than $S_2(\varepsilon)$ (*Grassberger & Procaccia*, 1983; *Grassberger*, 1983)

- **Takens** (1981) **embedding theorem:** For a scalar time series $X(t)$ obtained from a *D*-dimensional deterministic system, the vector with time-delayed coordinates $\{X(t), X(t + \tau), \ldots, X(t + (n - 1)\tau)\}$, $n \geq 2(D + 1)$, will trace out a trajectory that is a smooth coordinate transformation of the attractor of the original dynamical system. If the dynamical system has an attractor of a particular dimension, the embedded trajectory will have the same dimension.
- **Application of the theorem**: A dynamical system's reconstruction by time-delay embedding provides a method for detecting determinism in a time series and revealing the underlying dynamics, if any, of the system. The method is applied for various values of the embedding dimension *n*, and for each *n* the dimension D_q is calculated for some *q*. If D_q becomes invariant for increasing *n*, there is evidence that:
  - The system is deterministic rather than stochastic (for simple stochastic processes there is no finite (saturation) dimension as *n* increases).
  - The system's attractor has been identified and quantified by its dimension D_q. This is related to the number of time-delayed values that are necessary, in order to capture efficiently the evolution of the system (i.e., the variables we need to describe the phase-space in which the phenomenon evolves).

The application of the method requires numerous points of the time series and a proper selection of the lag τ. (Tsonis, 1992, p. 151,162; Tsonis et al., 1993).

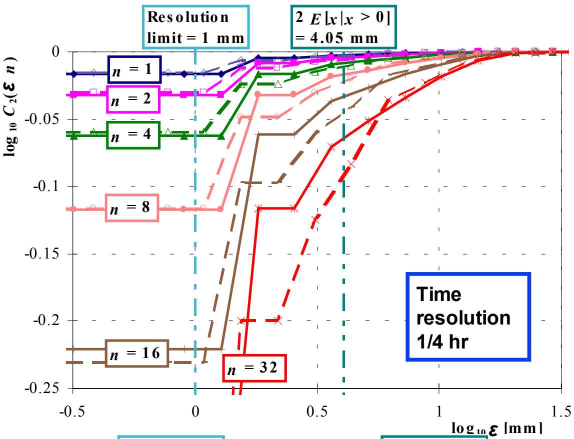
Application of the time-delay embedding method for rainfall data

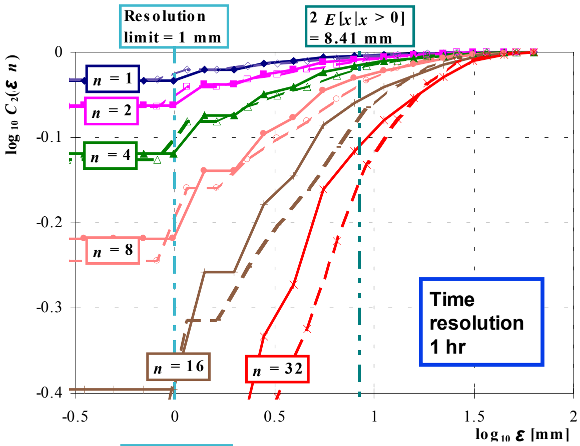
- **Data used**

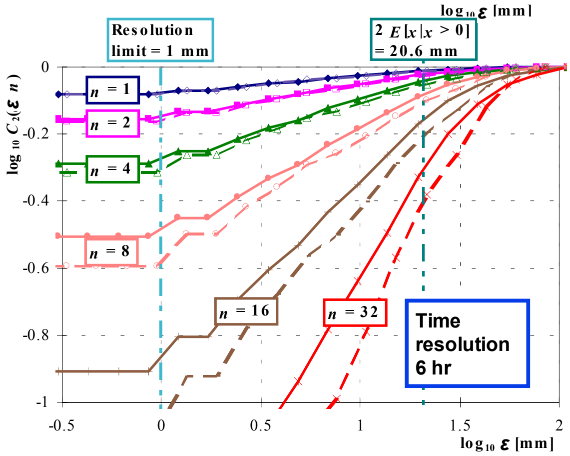
  The method was applied for both the historic and synthetic (generated by the stochastic model, see Appendix) data for four time resolutions, as shown in the table to the right, and for embedding dimensions up to 32. The results are shown in the following figures.

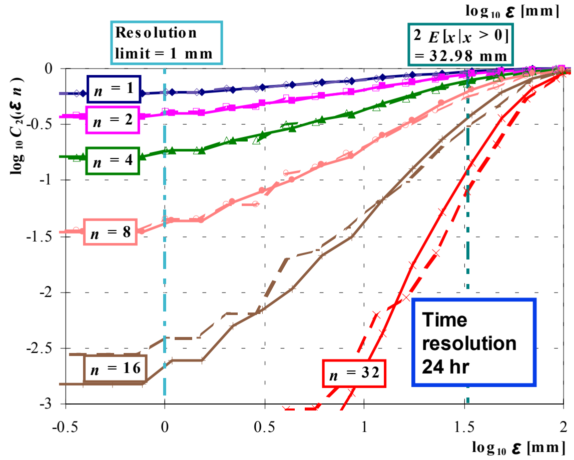
Time resolution (hr)	Historic record length (# points)	Synthetic record length (# points)	Adopted time lag (hr)
1/4	209 580*	212 607*	24
1	52 395	53 152	48
6	8 732	8 858	72
24	2 183	2 214	144

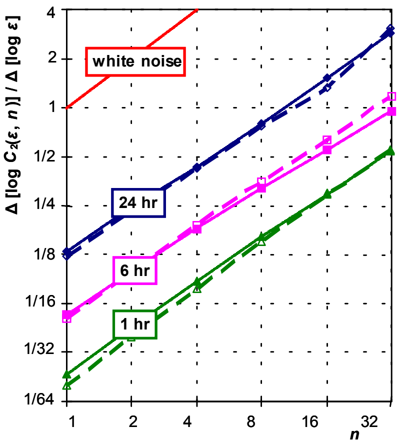
*70 000 points were used for correlation dimension calculations





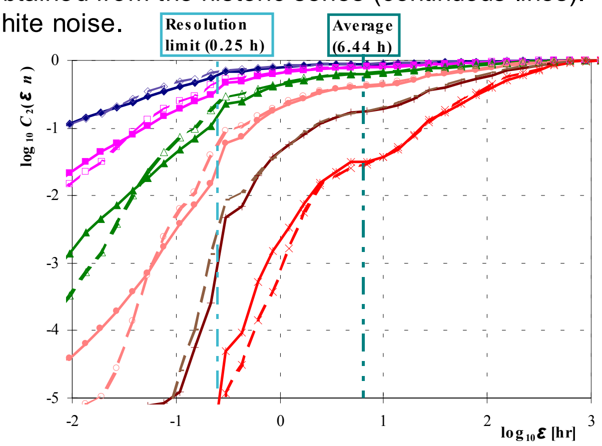






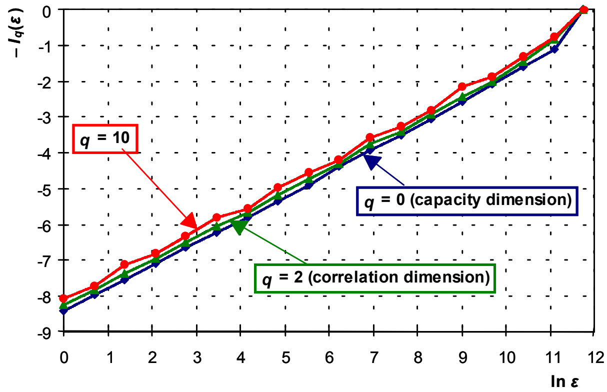
- **An interpretation of the results**
  - The horizontal line in the lower tail of all curves (Figures above) indicates zero correlation dimension. This is due to the nonzero probability of zero rainfall, which results in numerous time-delayed vectors with all coordinates zero.
  - There is (roughly) a scaling region (except for the case of 1/4 hr resolution) extended between the depth resolution limit (1 mm) and about two times the average of nonzero depths.
  - The slope of this scaling region, estimated by least squares, is an increasing function of embedding dimension (logarithmic plot to the left). No saturation value appears.
  - The results of the analysis of synthetic series (dashed lines) are quite similar with those obtained from the historic series (continuous lines). Both depart from white noise.

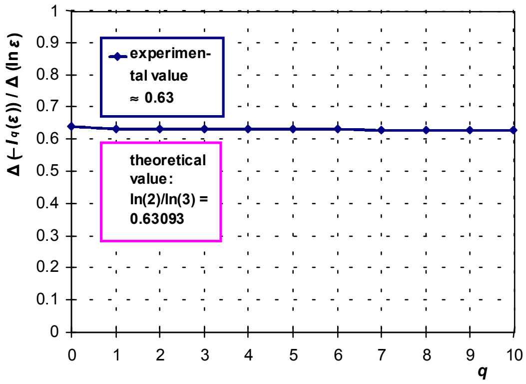
- **Application to the inverse series** (Figure to the right).
  - The inverse series represents time intervals corresponding to an increase of rainfall depth by 1 mm. Linear interpolation was used to inverse the series, which obviously introduces error for intervals less than 1/4 hr.
  - The results of the synthetic series (dashed lines) are again quite similar with those obtained from the historic series (continuous lines).
  - No clear scaling region appears here.



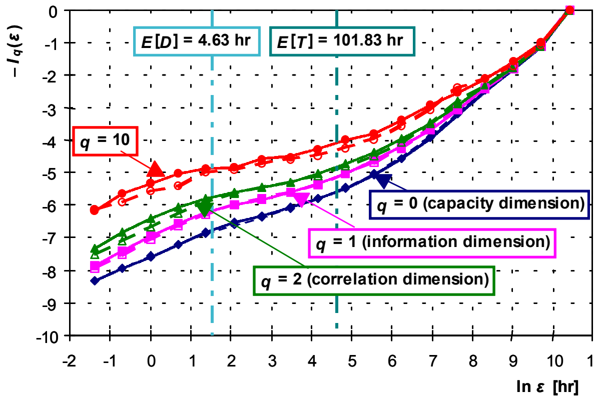
The rainfall series as Cantorian dust

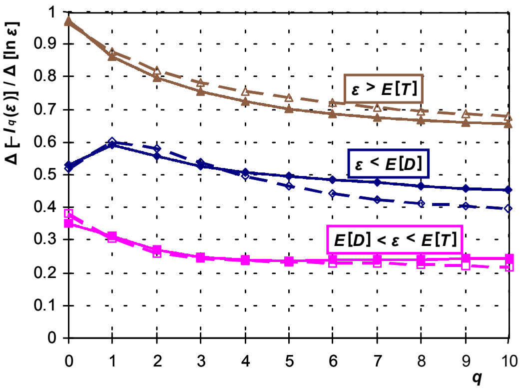
- The domination of voids (dry periods) in a rainfall time series evokes the parallelism with the Cantorian dust. More specifically, we can parallel the cumulative hyetograph of a certain period with the “devil’s staircase”, (*Schroeder*, 1991, p. 167), i.e., the function that maps the interval [0, 1] into itself having plateaus along all void intervals between the Cantorian dust (i.e., almost everywhere). Such an analogy can provide useful characterization and quantification of a rainfall time series and can reveal characteristic time scales.
- Dimensions may be easily calculated for this analogue by a box counting algorithm, where the boxes are time intervals of equal size $\varepsilon = \Delta t$. The measure p_i for the ith box must be set $p_i = \Delta h_i / h$, where Δh_i is the incremental rainfall depth in the ith box and h is the total depth of the entire period.
- To verify the method we have applied it for the devil’s staircase using up to 209 000 boxes (a number equal to the available intervals of the rainfall data set). The results shown below are in perfect agreement with the theoretical expectations.





- The application of the method with both the historic and synthetic rainfall data sets gave the results shown below (continuous lines correspond to historic data and dashed lines to synthetic). We observe that:
  - For each generalized entropy I_q, there exist three distinct regions with different slopes. The borders of these regions are approximately represented by the mean rain duration and the mean interarrival time.
  - The dimensions of the short time scale (left) area are about 0.5 and those of the intermediate time scale (middle) area are about 0.3.
  - The results for synthetic data agree well with those of the historic data.





Conclusions and discussion

- No determinism has been detected in the historic continuous rainfall record examined for time resolutions from 1/4 to 24 hr and for embedding dimensions up to 32.
- No essential differences have been detected between chaotic descriptors of the historic rainfall series and the synthetic data obtained by a well-structured stochastic model based on the scaling model of storm hyetograph.
- Both the historic and synthetic series are (obviously) distinguished from white noise.
- There are difficulties in applying the time-delay embedding method to continuous rainfall records of short time scale, owing to the nonzero probability of zero rainfall in a short time interval. It is anticipated that the method may be applied without problems for time scales considerably larger than the mean dry time (e.g., monthly time scale), but this would require records of hundreds of years to obtain reliable estimates.
- The Cantorian dust analogue of rainfall indicates the presence of two characteristic time scales in a rainfall series, which are the average rain duration and the average interarrival time.
- In addition, this analogue quantifies the examined rainfall series (both historic and synthetic) with a fractal dimension of about 0.5 for short time scales.
- The Cantorian dust analogue eliminates the problem of nonzero probability of zero rainfall; however the method's application with time-delayed vectors is not straightforward.

Appendix: Generation of synthetic data by the scaling model

- Phase A: Application of the alternating renewal model for temporal location of events.
  - A1: Generation of dry time from a Weibull distribution for the dry season (Oct.-May) and a two-segment Weibull distribution for the wet season (Jun.-Sep.), using different parameter sets for each month.
  - A2: Generation of rain duration from an exponential distribution, using different parameters for each month.
- Phase B: Calculation of statistics of total and incremental depths for each event (*Koutsoyiannis and Tsakalias*, 1992; *Koutsoyiannis*, 1994; *Mamassis et al.*, 1994).
  - B1: Calculation of $E[Z]$, Var[Z], E[**X**], Cov[**X**, **X**], μ_3[**X**] from the equations of the scaling model.
  - B2: Formulation of a sequential generating scheme as $\mathbf{X} = \mathbf{\Omega}\,\mathbf{V}$, where $\mathbf{\Omega}$ is a matrix of coefficients and **V** is a vector of independent variates, assumed (approximately) three-parameter gamma distributed.
  - B3: Estimation of parameters of the generating scheme, i.e.,
    - a. Coefficient matrix: $\mathbf{\Omega\Omega}^T = \mathrm{Cov}[\mathbf{X}, \mathbf{X}] \Rightarrow \mathbf{\Omega}$ by lower triangular decomposition.
    - b. Statistics of V_i:

$$\omega_{ii} E[V_i] = E[X_i] - \sum_{l=1}^{i-1} \omega_{il} E[V_l]$$
$$Var[V_i] = 1$$
$$\omega_{ii}^3 \mu_3[V_i] = \mu_3[X_i] - \sum_{l=1}^{i-1} \omega_{il}^3 \mu_3[V_l]$$

- Phase C: Generation of the sequence of incremental depths for each event (*Koutsoyiannis and Tsakalias*, 1992; *Koutsoyiannis*, 1994; *Mamassis et al.*, 1994).
  - C1: Generation of total depth Z, assumed two-parameter gamma distributed.
  - C2. Application of the sequential procedure to obtain an initial sequence of incremental depths X'.
  - C3. Determination of the final (adjusted) sequence: $X_i = \left(X'_i / \sum_{j=1}^{k} X'_j\right) Z$

Acknowledgments
We thank N. Mamassis for his assistance in data processing and A. Manetas for his help in programming. We also thank U. Huebner for providing us with FORTRAN programs, used for preliminary comparison. Computational resources were provided by the National Technical University of Athens (Central Computer Department and HYDROSCOPE Computer Network).

References
Grassberger, P., Generalized dimensions of strange attractors, *Physics Letters*, 97A(6), 227-230, 1983.
Grassberger, P., An optimized box-assisted algorithm for fractal dimensions, *Physics Letters A*, 148(1, 2), 63-68, 1990.
Grassberger, P. and I. Procaccia, Characterization of strange attractors, *Physical review letters*, 50(5), 346-349, 1983.
Jayawardena, A. W. and F. Lai, Analysis and prediction of chaos in rainfall and stream flow time series, *Journal of Hydrology*, 153, 23-52, 1994.
Koutsoyiannis, D., A stochastic disaggregation method for design storm and flood synthesis, *Journal of Hydrology*, 156, 193-225, 1994.
Koutsoyiannis, D. and E. Foufoula-Georgiou, A scaling model of storm hyetograph, *Water Resources Research*, 29(7), 2345-2361, 1993.
Koutsoyiannis, D. and G. Tsakalias, A disaggregation model for storm hyetographs, Presentation at the 3rd meeting of the AFORISM project, Athens, 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 in *Annales Geophysicae*, Vol. 12, Supplement II, Part II, 324, 1994.
Rényi, A., *Probability theory*, North-Holland, Amsterdam, 1970.
Rodriguez-Iturbe, I., B. F. de Power, M. B. Sharifi and K. P. Georgakakos, Chaos in rainfall, *Water Resources Research*, 25(7), 1667-1675, 1989.
Rodriguez-Iturbe, I., Exploring complexity in the structure of rainfall, *Advances in Water Resources*, 14(4), 162-167, 1991.
Schroeder, M., *Fractals, chaos and power laws: Minutes from an infinite Paradise*, Freeman & Co., New York, 1991.
Sharifi, M. B., K. P. Georgakakos and I. Rodriguez-Iturbe, Evidence of deterministic chaos in the pulse of storm rainfall, *Journal of Atmospheric Sciences*, 45(7) 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, 336-381, Springer-Verlag, Berlin, 1981.
Tsonis, A. A., *Chaos: From theory to applications*, Plenum Press, New York, 1992.
Tsonis, A. A., J. B. Elsner and K. P. Georgakakos, Estimating the dimension of weather and climate attractors: Important issues about the procedure and interpretation, *Journal of Atmospheric Sciences*, 50(15) 2449-2555, 1993.