Hydrofractals '03

An international conference on fractals in hydrosciences

Monte Verità, Ascona, Switzerland, 24-29 August 2003

On embedding dimensions and their use to detect deterministic chaos in hydrological processes

Demetris Koutsoyiannis

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

Introduction: The context

- **1989-present:** Several investigators discover low dimensional deterministic chaos in hydrologic processes
- This is a manifestation of a wider trend, for which other investigators expressed their scepticism, e.g.,
  - "… the desire for finding a chaotic attractor has led to a naïve application of the analysis methods; as a result, the number of claims on the presence of strange attractors in vastly different physical, chemical, biological and astronomical systems has grown (exponentially?)" (Provenzale et al., 1992)
  - "… most (if not all) of these claims have to be taken with much caution" (Grassbrerger et al., 1991)
- **August 2001:** Koutsoyiannis, D., Are hydrologic processes chaotic? (Unpublished)
- **February 2002**: Schertzer, D., I. Tchguirinskaia, S. Lovejoy, P. Hubert, H. Bendjoudi & M. Larchvêque, Which chaos in the rainfall–runoff process? Hydrol. Sci. J., 47(1), Discussion to Sivakumar, B., R. Berndtsson, J. Olsson & K. Jinno, Evidence of chaos in the rainfall–runoff process, Hydrol. Sci. J., 46(1), 131–146, 2001.

Studies that have investigated chaos in hydrologic processes

Item #	Reference	Data type	Location	Time scale	Data size	Time delay used	Embedding dimension	Attractor dimension
1	Jayawardena and Lai (1994)	streamflow	2 stations in Hong Kong	daily	5840-7300	2-3	10-20	0.45-0.65
2	Jayawardena and Lai (1994)	rainfall	3 stations in Hong Kong	daily	3650-4015	2	30-40	0.95-2.54
3	Sivakumar et al. (1998, 1999)	rainfall	6 stations in Singapore	daily	10958	7-20	12-13	1.01-1.03
4	Tsonis et al. (1993); Tsonis (1992, p. 168)	raingauge tip times	not reported	0.01 mm	2200	not needed	5	2.2
5	Sharifi et al. (1990)	raingauge tip times	Cambridge, Massachusetts	0.01 mm	3316-4000	4-134	8-10	3.35-3.75
6	Sangoyomi et al. (1996)	lake volume	Great Salt Lake	biweekly	3750	9	8	3.4
7	Rodriguez-Iturbe et al. (1989); Rodriguez-Iturbe (1991)	storm	Boston	15 s	1990	8-12	5	3.78
8	Porporato and Ridolfi (1997)	streamflow	Dora Baltea (tributary of Po)	daily	14246	1	<10	<4
9	Sivakumar et al. (2000, 2001)	runoff	Göta, Sweden	monthly	1572	20	10	5.5
10	Sivakumar et al. (2000, 2001)	rainfall	Göta, Sweden	monthly	1572	3	10	6.4
11	Wang and Gan (1998)	streamflow	6 rivers in Canadian Prairies	daily	3044-30316	40-180	10	3 (interpreted to be 7-9)
12	Sivakumar et al. (2000, 2001)	runoff coefficient	Göta, Sweden	monthly	1572	3	13	7.8
13	Rodriguez-Iturbe et al. (1989)	rainfall	Genoa	weekly	7722	not mentioned	up to 8	no convergence
14	Wilcox et al. (1991)	runoff	Reynolds Mountain, Idaho	daily	8800	2-16	up to 20	no convergence
15	Koutsoyiannis and Pachakis (1996)	rainfall	Ortona, Florida	0.25 h to 24 h	70 000 - 2214	96-12	up to 32	no convergence

The notion of an attractor

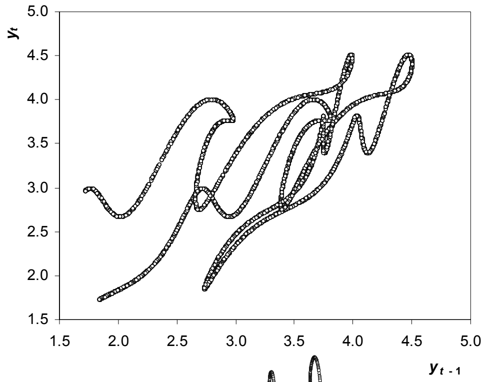
- A dynamical system in discrete time:
  $\mathbf{x}_{n+1} = \mathbf{S}_1(\mathbf{x}_n)$, $n \in \mathrm{I}$, $\mathbf{x}_n \in \mathcal{R}^m$
- Expression of the system trajectory $\mathbf{x}_n$ through time delayed vectors of a single observable y_n:
  $\mathbf{x}_n := [y_n, y_{n-\tau}, \ldots, y_{n-(m-1)\tau}]^T$, $n, \tau \in \mathrm{I}$
- An attractor: a set $A \subseteq \mathcal{R}^m$ that is invariant under the dynamical evolution ($\mathbf{S}_1(A) = A$)
- Basic property of an attractor (if it is not a fixed point or a limit cycle): It is nonintersecting ($\mathbf{x}_{n_1} \neq \mathbf{x}_{n_2}$ for all $n_1 \neq n_2$)
- Whitney's (1936) embedding theorem (generalized for fractal objects by Sauer et al., 1991): A D-dimensional object can be embedded in an m-dimensional Euclidean space if $m \geq 2D + 1$

Demonstration of Whitney's embedding theorem

2-dimensional delay representation of a series of 10 000 points generated from a linearly routed logistic equation

Attractor dimension:
$D = 1$ (known from theory and clearly shown in Figure – A line)

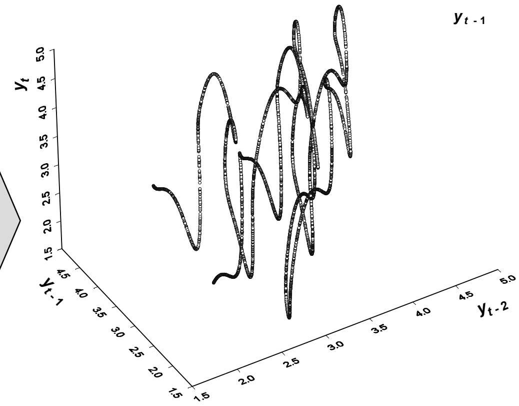
Embedding dimension:
$m = 2 < 2D + 1 = 3$ => Intersecting



3-dimensional delay representation of the same series

Attractor dimension:
$D = 1$

Embedding dimension:
$m = 3 = 2D + 1$ => Non-intersecting

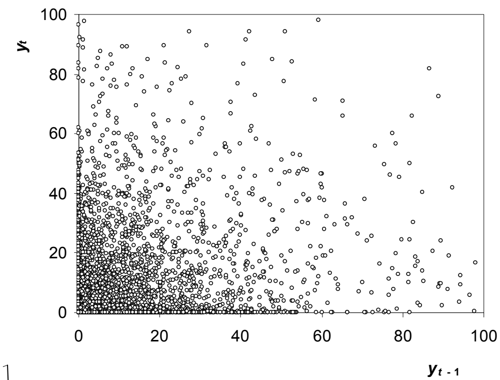


Whitney's embedding theorem applied to hydrological series

2-dimensional delay representation of a series of 10 000 daily rainfall depths (Vakari raingage, W. Greece)

Embedding dimension: $m = 2$

Attractor dimension: $D = ?$



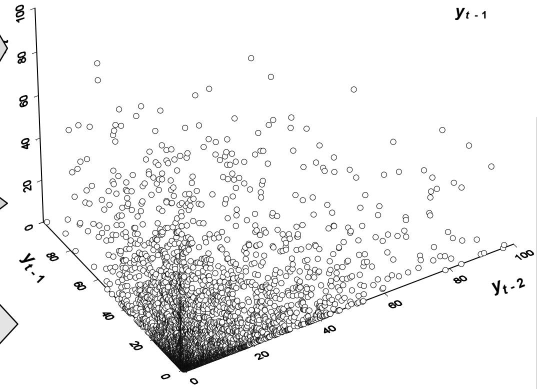
3-dimensional delay representation of the same rainfall series

Embedding dimension: $m = 3$

Attractor dimension: $D = ?$

Question: Does this look like a line with $D = 1$ or smaller?

Answer: If $D \leq 1$, then intersections would not occur in $m = 3$ dimensions. But intersections occur, ergo $D > 1$



Whitney's embedding theorem applied to hydrological series (2)

Question: In some papers analysing hydrological series (daily rainfall, daily streamflow) the attractor dimension was estimated as $D = 1$ or smaller, down to 0.45. For these estimations, it was necessary to use embedding dimensions m as high as 10 up to 40. What does this mean?

Answer: Clearly, if $D \leq 1$, then a dimension $m = 3$ would suffice to embed the attractor. Thus, something was wrong in the estimation procedure followed

Possible sources of errors:

- What was estimated must not be the topological dimension of the trajectory (to be discussed later)
- As the accuracy of estimation decreases with increased embedding dimension m, one may need to use high m to make calculations inaccurate enough so as to get "good" wrong results (to be discussed later)

Conclusion: The result $D \leq 1$ cannot be acceptable

Seeking for a minimum acceptable attractor dimension in a rainfall series

Question: Can we obtain a rough estimate of the minimum acceptable attractor dimension D, when analysing a daily rainfall series, without doing any calculation?

Answer: Daily rainfall series contains dry periods

Let k be the length of the longest dry period

Set $n = 1$ the day when this dry period starts, so that the rainfall depths y_n for $n = 1$ to k are all zero

Assume that the rainfall at the examined location is the outcome of a deterministic system whose attractor can be embedded in $\mathcal{R}^m$ for some integer m. This attractor is reconstructing using delay embedding with delay τ

Furthermore, assume that $m < (k - 1) / \tau + 1$. Then, there exist at least two delay vectors with all their components equal to zero. Namely, the vectors:

$\mathbf{x}_k = [y_k, y_{k-\tau}, y_{k-2\tau}, \ldots,]^T$, $\mathbf{x}_{k-1} = [y_{k-1}, y_{k-1-\tau}, y_{k-1-2\tau}, \ldots, y_{k-1-(m-1)\tau}]^T$

both will be zero ($\mathbf{x}_k = \mathbf{x}_{k-1} = \mathbf{0}$)

Seeking for a minimum acceptable attractor dimension in a rainfall series (2)

Answer (continued):

In that case, $\mathbf{x}_k = \mathbf{S}_1(\mathbf{x}_{k-1}) = \mathbf{S}_1(\mathbf{0}) = \mathbf{0}$, and since the system is deterministic, it will result in $\mathbf{x}_n = \mathbf{0}$ for any $n > 0$ (since $\mathbf{x}_{k+1} = \mathbf{S}_1(\mathbf{x}_k) = \mathbf{S}_1(\mathbf{0}) = \mathbf{0}$, etc.)

That is, given that rainfall is zero for a period k, it will be zero forever, which means that the attractor is a single point

This of course is absurd and thus the embedding dimension should be $m \geq (k-1) / \tau + 1$

Now, Whitney's embedding theorem tells that the attractor should have dimension

$D \geq (m-1)/2$ and, hence, $D_{\min} = (k-1) / 2\tau$

Example: In Athens, Greece, in a 132-year rainfall record we have a dry period with length $k = 120$ days (4 months)

If we assume a 'safe' delay $\tau = 10$, then $m \geq 12$ and $D_{\min} = 6$ (like the largest of estimates published in the literature) – Could be $D = \infty$

Question: How many data points do we need to study a $D = 6$ attractor?

Answer: So many that such a study is impossible (to be discussed later)

Typical procedure to estimate an attractor dimension

- It is an iterative procedure
  - For successive m we attempt to reconstruct the attractor using time delay vectors $\mathbf{x}$ of size m
  - For each m we estimate the attractor dimension $D(m)$
  - If beyond some m^* the attractor dimension remains constant, i.e., $D(m) = D(m^*) = D$, then the attractor dimension is D and the required embedding dimension is m^*
- The dimension of an object is determined in terms of the generalised entropy. If the object is spanned with hypercubes of edge length ε, there is a sequence of entropies $I_q(\varepsilon)$ and thus a sequence of dimensions
  $D_q = \lim_{\varepsilon \to 0} I_q(\varepsilon) / \ln(\varepsilon)$, $q = 0, 1, 2, \ldots$.
- The dimension implied in the embedding theorems is the topological (box counting or capacity) dimension D_0
- The dimension used in typical calculations is the correlation dimension D_2

Estimation of the correlation dimension

- For $q \geq 2$, the estimation of dimension may be based on the so-called correlation sums (rather than generalised entropies)

  $C_q(\varepsilon) = N^{-q}$ {nr. of q-tuples $(\mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_q})$ with all $\| x_{j_s} - x_{j_r} \| < \varepsilon$}
- This is due to the relationship $C_q(\varepsilon) \approx \exp[(1 - q)\, I_q(\varepsilon)]$
- This enables calculation of $D_2(m)$ (for some m) with the following algorithm
  1. Calculate the correlation sum $C_2(\varepsilon, m)$ for several values of ε
  2. Make a log-log plot of $C_2(\varepsilon, m)$ vs. ε and a semi-log plot of the local slope $d_2(\varepsilon, m) := \Delta[\ln C_2(\varepsilon, m)]/\Delta[\ln \varepsilon]$ vs. ε, and locate a region with constant slope, known as a scaling region
  3. Calculate the slope of the scaling region, which is the estimate of the correlation dimension $D_2(m)$ of the set for the embedding dimension m

Example

Correlation sums $C_2(\varepsilon, m)$ and their local slopes $d_2(\varepsilon, m)$ vs. length scale ε for embedding dimensions m = 1 to 8 calculated from a series of 10 000 independent random values with Pareto distribution with κ = 1/8.

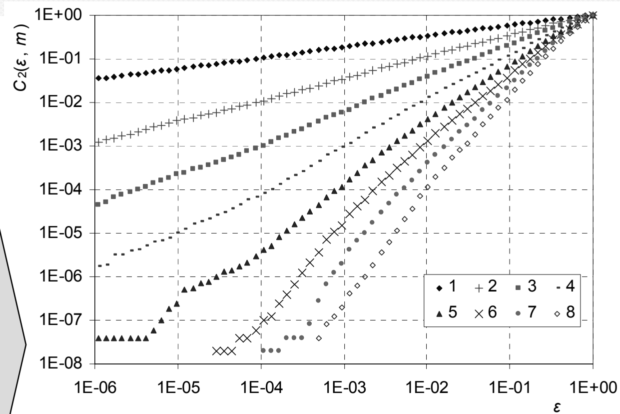
Pareto distribution and density:

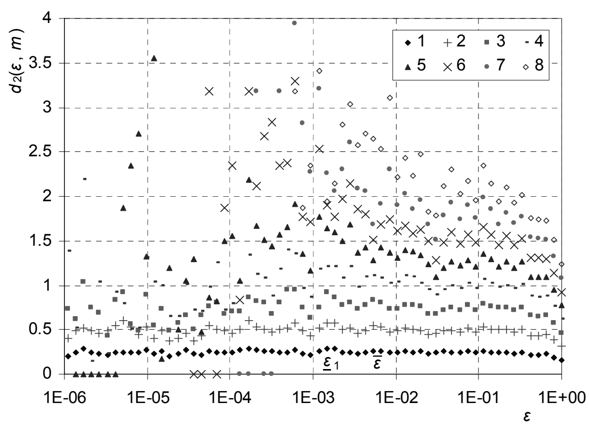
$F(y) = (y / a)^{\kappa}$, $f(y) = (\kappa / a)\,(y / a)^{\kappa - 1}$

$(0 \leq y \leq a)$

Question: We observe that $D_2(1)$ = 0.25, $D_2(2)$ = 0.5, etc. What do these results mean?

Answer: In fact $D_0(m) = m$ (space filling set) but it happens here $D_2(m) < D_0(m) = m$ (significant underestimation of dimensions)





Correlation dimension vs. capacity dimension and the effect of an asymmetric density function

- Rule: $D_2(m) = D_0(m)$
- The rule is valid only for square-integrable density functions $f(y)$, i.e., those whose square integral over their domain A is finite ($\int_A f^2(y)\, dy < \infty$)
- In asymmetric, J-shaped densities, this integral can be infinite
- In this case $D_2(1) = 2 + 2 \lim_{\varepsilon \to 0} [\varepsilon f'(\varepsilon)] / f(\varepsilon)] < 1 = D_0(1)$
- Hydrological time series on fine time scales have asymmetric, J-shaped densities

Example: In densities commonly used in hydrological processes, like:

Gamma

$f(y) = [1 / a\, \Gamma(\kappa)]\ (y / a)^{\kappa - 1}\ e^{-y / a}$

Weibull

$f(y) = (\kappa / a)\ (y / a)^{\kappa - 1} \exp[-(y / a)^{\kappa}]$

Pareto

$f(y) = (\kappa / a)\ (y / a)^{\kappa - 1}$

it is shown that

$D_2(1) = \min(1, 2\kappa)$

That is, for $\kappa < ½$, $D_2(1) < 1$

This explains why in the previous example ($\kappa = 1/8$)

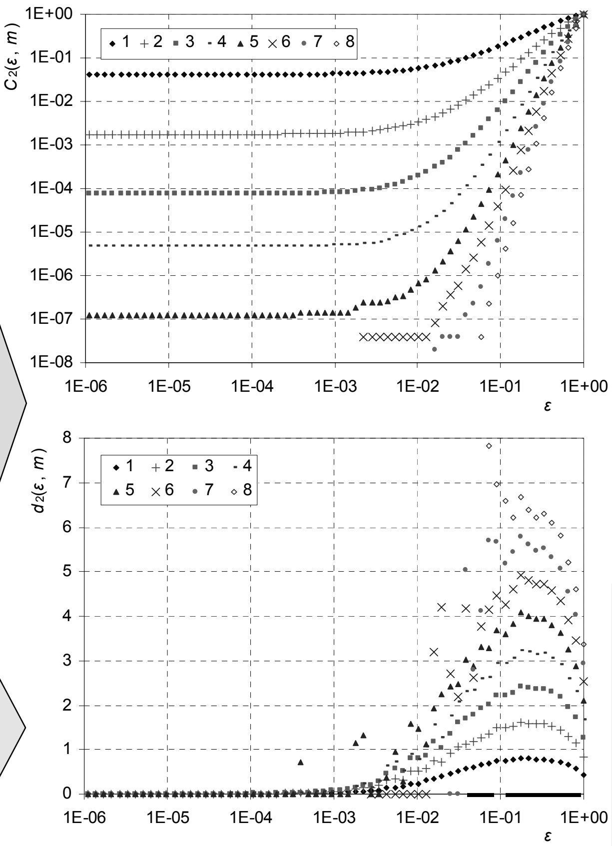
$D_2(m) = 0.25\, m < m$

The effect of intermittency

- It can be shown that if a process y_n is strictly intermittent, i.e., if $\mathrm{Prob}(y_n = 0) > 0$, then $D_2(1) = 0$ **exactly**
- This can be extended to many dimensions, i.e., if $\mathrm{Prob}(\mathbf{x}_n = \mathbf{0}) > 0$,where $\mathbf{x}_n$ is the m-dimensional delay vector, then $D_2(m) = 0$ **exactly**
- Hydrological processes like rainfall (at a time scale monthly or finer) and runoff (in ephemeral streams) may be strictly intermittent
- In such cases, calculation of correlation dimension has no meaning at all, as it says nothing about the capacity dimension of the set under study
- Ignorance of the effect of intermittency is a potential source of erroneous results

Demonstration of the effect of intermittency

Correlation sums $C_2(\varepsilon, m)$ and their local slopes $d_2(\varepsilon, m)$ vs. length scale ε for embedding dimensions m = 1 to 8 calculated from a series of 10 000 independent random values, 80% of which are generated from the uniform distribution and the remaining are zeros (located at random)



Observation: Clearly, the example verifies the theoretical result $D_2(m) = 0$ (unless one goes to large scales, $\varepsilon > 10^{-2}$)
Note: Still $D_0(m) = m$ (a space filling set)

The effect of wide-sense intermittency

- Wide-sense intermittency is met when a process shifts among different regimes not necessarily going to zero state
- Streamflow series display this kind of intermittency (low flows, regular flows, floods)
- For such kinds of intermittency, Graf von Hardenberg et al. (1997) have shown that the standard algorithms fail to estimate correctly the dimensions of processes, while giving no warning of their failure
- In addition, they demonstrated that the standard algorithms, applied on a time series from a composite chaotic system with randomly driven intermittency, estimates a very small dimension (e.g. $D_2 = 1$ or smaller) although the actual dimension of the system is infinite due to the random character of intermittency

Recovery from the effect of wide-sense intermittency

- Graf von Hardenberg et al. (1997) proposed ways to refine the algorithm so as to obtain correct results
- The simplest of them is to filter the data by excluding all the delay vectors **x** having at least one component $y_i < c$, where c an appropriate cutoff value that leaves out all "off" data points of the intermittent time series
- This simple algorithm was proven very effective
- In addition, it was found appropriate to recover from the effects of strict-sense intermittency and asymmetric distribution function, as well
- However, it reduces dramatically the number of data points, especially for large embedding dimensions (to be discussed further later)

The effect of data size

Question: What is the sufficient data size (N_{min}) to accurately estimate correlation dimension $D(m)$ for embedding dimension m?

Typical answer:

1. There is the formula due to Smith (1988)

   $N_{min} = 42^m$

   but this results in too many data points (e.g. 10^8 and 10^{16} points for m = 5 and 10)
2. Then, there is the formula due to Nerenberg and Essex (1990)

   $N_{min} = 10^{2 + 0.4\,m}$

   but this still results in many points (e.g. 10^4 and 10^6 points for m = 5 and 10)
3. Since I do not have so many points I can do with fewer (just as many as I have)
4. Several examples have demonstrated good performance with fewer points

Comment: Demonstration is not a proof

Statistical estimation of the required data size

- Most studies have attempted to show that a time series originates from a low-dimensional deterministic system rather than a stochastic system
- In this case, it is natural to make the null hypothesis that it originates from a stochastic system and then to reject this hypothesis
- Under this null hypothesis, the correlation sum for any length scale ε and any embedding dimension m is
  $C_2(\varepsilon, m) = [C_2(\varepsilon, 1)]^m$
- Since in a stochastic system, $C_2(\varepsilon, m)$ expresses a probability (the probability that the distance of two points is less than ε), the required data size $N_{\min}$ can be easily estimated by classical statistical techniques

Statistical estimation of the required data size (2)

The statistical result is

$N_{\min} = \sqrt{2}\,(z_{(1+\gamma)/2} / c)\,[C_2(\bar{\varepsilon}, 1)]^{-m/2}$

where

z_a the a-quantile of the standard normal distribution,

γ a confidence coefficient

c the acceptable relative error in the estimation of $C_2(\varepsilon, m)$

$\bar{\varepsilon}$ the upper limit of the scaling area for embedding dimension 1 (or the highest possible length scale that suffices to adequately estimate $D_2(1)$, meaning that for $\varepsilon > \bar{\varepsilon}$, $d_2(\varepsilon, 1)$ is not constant)

Example: For $\gamma = 0.95$ or $z_{(1+\gamma)/2} = 1.96$, $c = 3\%$ and $C_2(\bar{\varepsilon}, 1) = 0.15$ we obtain

$N_{\min} \approx 10^{2 + 0.4\,m}$

i.e., the Nerenberg and Essex formula

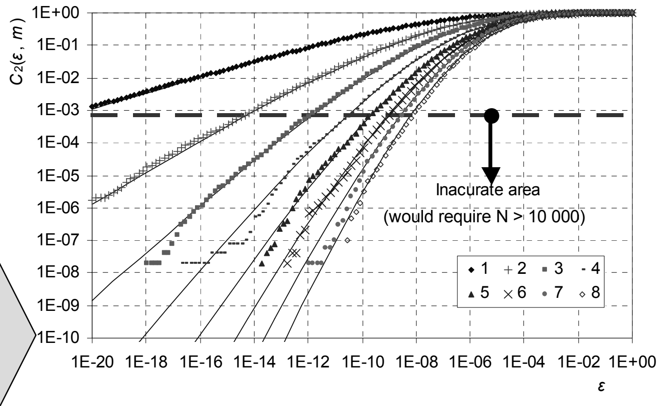
Demonstration of the statistical estimation of the required data size

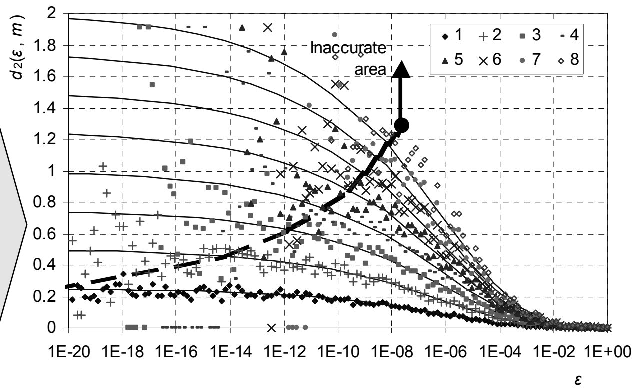
Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ from a series of 10 000 independent random points from the Weibull distribution with $\kappa = 1/8$

Comments: Here dimensions are known from theory ($D_2(m) = 0.25\ m$); also $C_2(\varepsilon, m)$ and $d_2(\varepsilon, m)$ can be computed from theory (blue solid curves)

Clearly, $\overline{\varepsilon} \le 10^{-20}$ (error 2%), so $N_{\min} = 30^{1.65 + m}$ (close to Smith)

For $m = 1, 2, 5, 10$: $N_{\min} = 8\ 350$, $252\ 000$, 6.9×10^9, 1.7×10^{17}





A complete procedure for the typical problem

The typical problem: For given data set of size N, if nothing is known for the system dynamics, up to which embedding dimension m can $D_2(m)$ be estimated accurately?

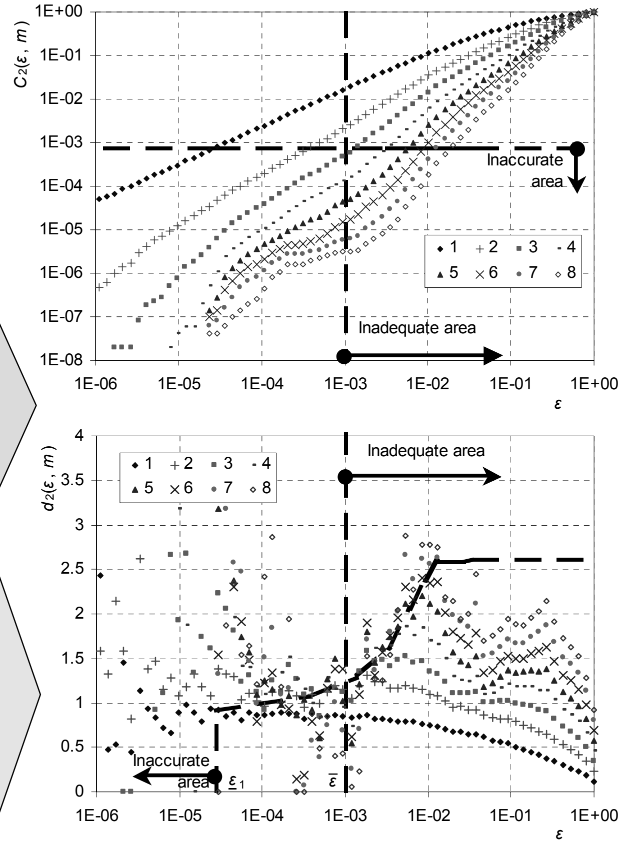
The procedure

- Make plots of $C_2(\varepsilon, m)$ and $d_2(\varepsilon, m)$ for several m
- In the plot of $d_2(\varepsilon, 1)$ (i.e., for $m = 1$) locate a region where $d_2(\varepsilon, 1)$ becomes constant and relatively smooth. Set $\overline{\varepsilon}$ the upper limit of this area (above which $d_2(\varepsilon, 1)$ is not constant) and $\underline{\varepsilon}_1$ the lower limit (below which $d_2(\varepsilon, 1)$ becomes too rough)
- From the plot of $C_2(\varepsilon, 1)$ determine $C_2(\underline{\varepsilon}_1, 1)$
- Set $C_2(\underline{\varepsilon}_m, m) = C_2(\underline{\varepsilon}_1, 1)\ (N_1 / N_m)^2$ and determine $\underline{\varepsilon}_m$ for each m (N_1 and N_m are the actual data size for embedding dimensions 1 and m which can be different)
- For those m in which $\underline{\varepsilon}_m \le \overline{\varepsilon}$, determine $D_2(m)$ as the average $d_2(\varepsilon, m)$ on the interval $(\underline{\varepsilon}_m, \varepsilon)$
- For those m in which $\underline{\varepsilon}_m > \overline{\varepsilon}$, $D_2(m)$ cannot be determined

Demonstration of the complete algorithm: An example involving asymmetry and autocorrelation

Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from a series of 10 000 autocorrelated random values having approximately Pareto distribution with $\kappa = 0.44$

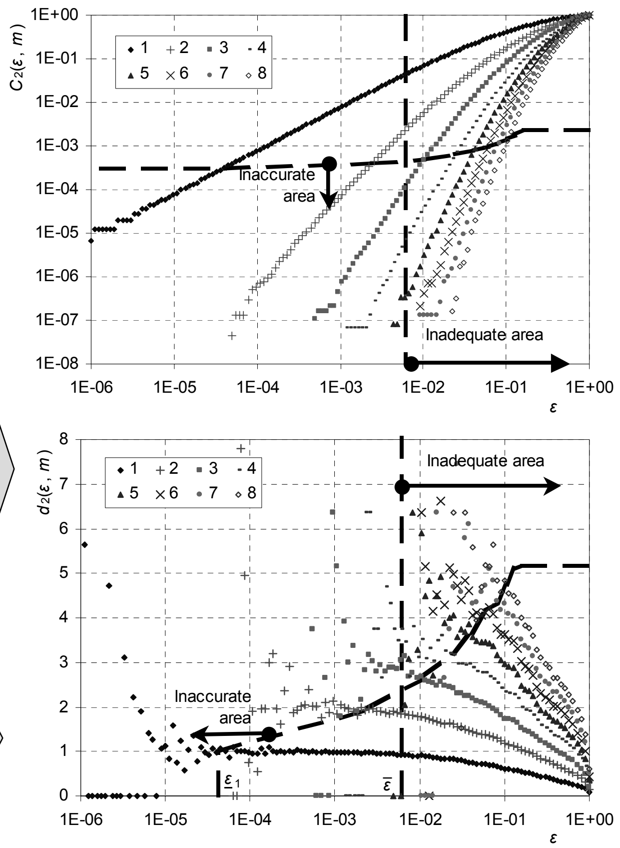
Comments:
1. Maximum $m = 2$ for adequate estimation of $D_2(m)$
2. The synergistic effects of asymmetry and autocorrelation lead to the conclusion that $D = 1$
3. However, it is known that $D_2(m) = 0.88\ m$, so that $D = \infty$



An example involving asymmetry and autocorrelation Recovery from the effect of asymmetry

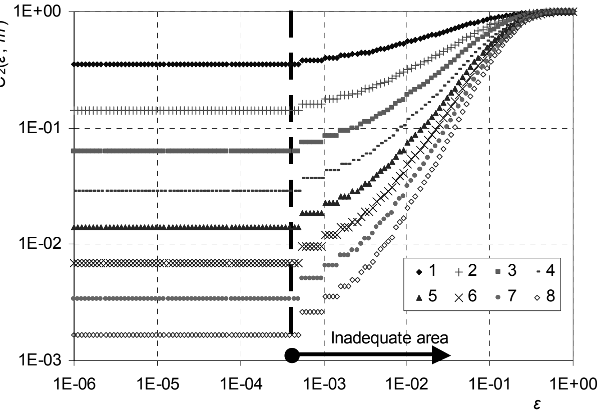
Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from the same series as in the previous figure but excluding points having at least one coordinate smaller than 0.01 (following the procedure by Graf von Hardenberg et al.)

Comments:
1. Maximum $m = 2$ for adequate estimation of $D_2(m)$
2. $D_2(1) = 1$, $D_2(2) = 2$, so no saturation



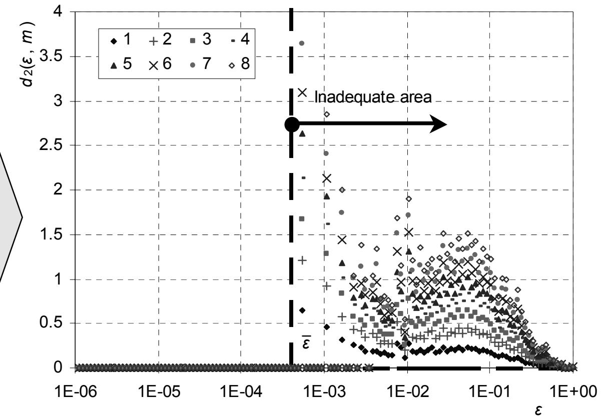
Real world examples 1. A daily rainfall series

Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from the daily rainfall series at the Vakari raingage (N = 11 476; intermittent; skewness = 4.59; chosen τ = 12)



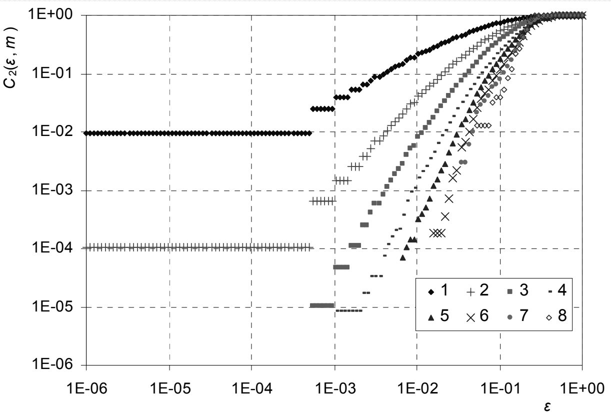
Comments:

1. Due to the presence of zeros, $D_2(m) = 0$ for all m
2. Figure says nothing about the capacity dimension of the 'attractor'
3. If we incorrectly ignored the small ε and instead, chose ε in the region 0.01-0.1, we would estimate a small D (≤ 1.5)



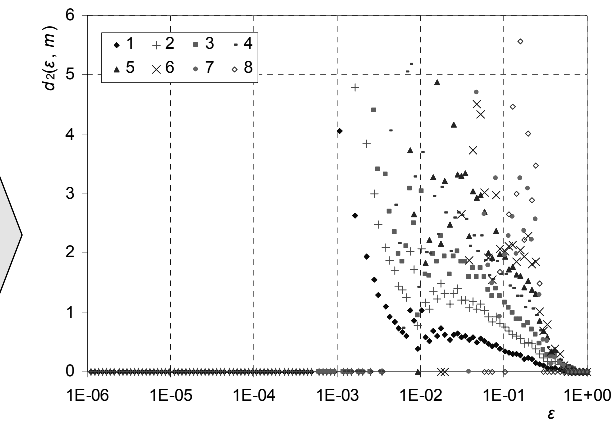
Real world examples 1. A daily rainfall series (continued)

Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from the same daily rainfall series as in the previous figure but excluding points with zero values



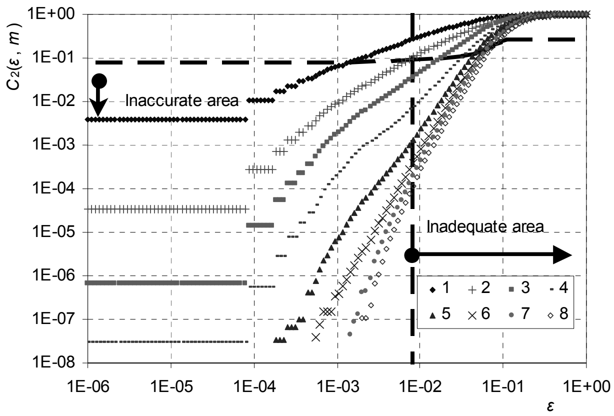
Comments:

1. Again $D_2(m) = 0$ due to "measurement induced intermittency" (e.g. 5% of the values are recorded as 0.1 mm, 4% as 0.2 mm, etc.)
2. It becomes much more difficult to obtain an inaccurate estimate of a small D



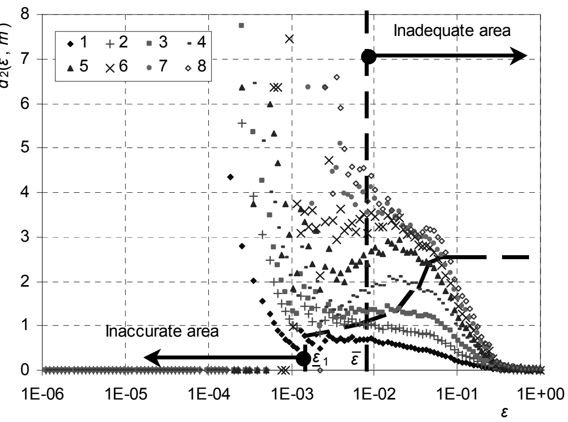
Real world examples 2. A storm data series

Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from a storm data series at Iowa ($N = 9679$ data points measured every 10 s; skewness = 4.83 corresponding to $\kappa = 0.40$ for Gamma distribution; high autocorrelation; chosen $\tau = 500$)



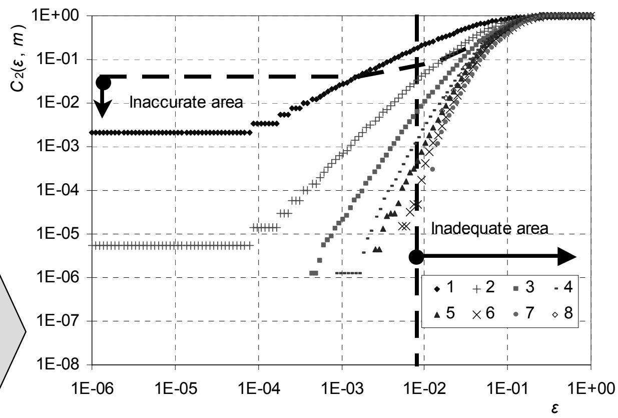
Comments:

1. Again "measurement induced intermittency" (e.g., 217 values are 0.09 mm/h, 169 are 0.08 mm/h, etc.)
2. Ignoring intermittency area, maximum $m = 2$, $D_2(1) = 0.69$ and $D_2(2) = 1.00$
3. Results do not support nor prohibit the existence of low-dimensional determinism



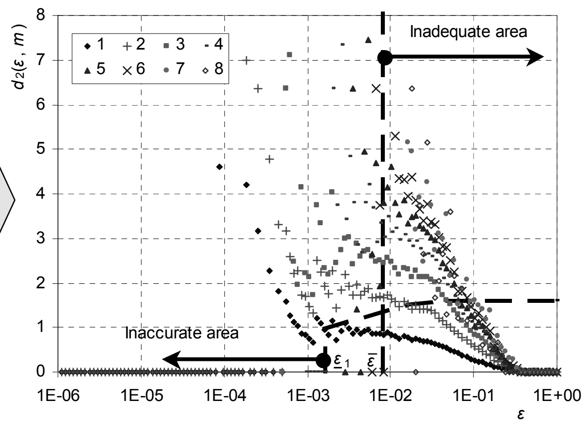
Real world examples 2. A storm data series (continued)

Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from the same storm series as in the previous figure but excluding points having at least one coordinate smaller than 0.01



Comments:

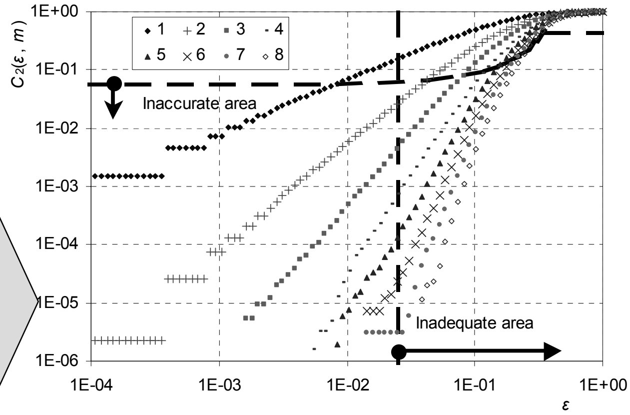
1. Ignoring intermittency area, maximum $m = 1$ (for adequate estimation of $D_2(m)$), whereas $D_2(1) = 1$
2. Results do not indicate low-dimensional determinism

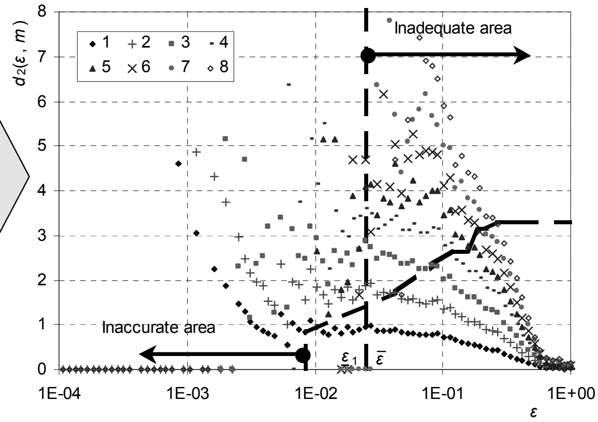


Real world examples 3. A monthly rainfall series

Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from the monthly rainfall series at Athens excluding zero points (N = 1586; intermittent; skewness = 1.75; chosen τ = 1)

Comments:
1. Maximum m = 1 (for adequate estimation of $D_2(m)$), whereas $D_2(1) = 1$
2. Results do not indicate low-dimensional determinism

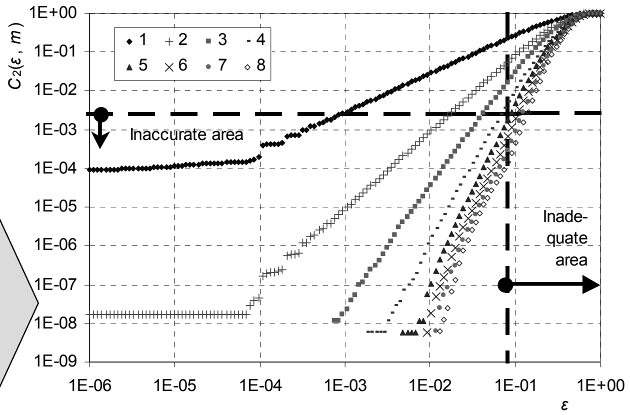


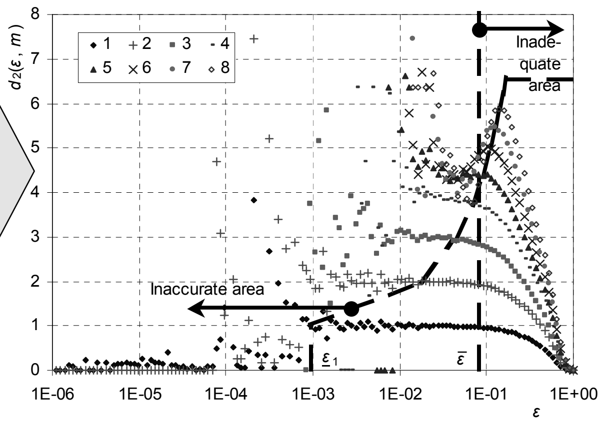


Real world examples 4. A hourly relative humidity series

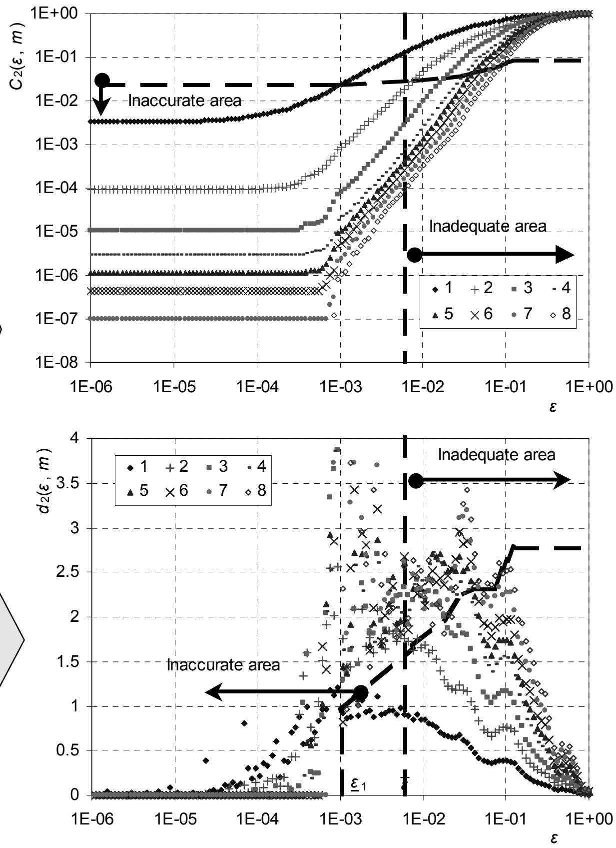
Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from the hourly relative humidity series at Athens (N = 18 888; no intermittency; no skewness; high autocorrelation; chosen τ = 108)

Comments:
1. Maximum m = 4 (for adequate estimation of $D_2(m)$),
2. $D_2(m) = m$, for m = 1 to 4
3. Clearly, there is no low-dimensional determinism





Real world examples 5. A daily streamflow series



Correlation sums $C_2(\varepsilon, m)$ and slopes $d_2(\varepsilon, m)$ calculated from the discharge series at Ali Efenti gage at Pinios River (N = 8 246 - 1435 missing values; wide-sense intermittency; skewness; high autocorrelation; chosen $\tau = 94$)

Comments:
1. Maximum $m = 1$ (for adequate estimation of $D_2(m)$), whereas $D_2(1) = 1$
2. Results do not indicate low-dimensional determinism

Conclusions

- Studies reporting the discovery of low-dimensional chaotic deterministic dynamics in hydrological systems (using time delay embedding and correlation dimension) may be misleading and flawed
- Specific peculiarities of hydrological processes on fine time scales, such as asymmetric J-shaped densities, intermittency, and high autocorrelation, are synergistic factors that can lead to misleading conclusions regarding presence of (low-dimensional) deterministic chaos
- The required size to accurately estimate chaotic descriptors of hydrological processes, as quantified by statistical reasoning, is so tremendous that cannot be met in hydrological records
- In light of the theoretical analyses and arguments, procedures are proposed to recover from misleading results
- Typical real-world hydrometeorological time series, such as relative humidity, rainfall, and runoff, are explored and none of them is found to indicate the presence of low-dimensional chaos

Made in Cefalonia

This presentation is available on line at http://www.itia.ntua.gr/e/docinfo/584/

References

Graf von Hardenberg, J., F. Paparella, N. Platt, A. Provenzale, E. A. Spiegel, and C. Tesser, Missing motor of on-off intermittency, Physical Review E, 55(1), 58-64, 1997b.

Grassberger, P., T. Schreiber, and C. Schaffrath, Nonlinear time sequence analysis, Int. J. Bifurcation and Chaos, 1, 521, 1991.

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., and D. Pachakis, Deterministic chaos versus stochasticity in analysis and modeling of point rainfall series, *Journal of Geophysical Research-Atmospheres*, 101(D21), 26444-26451, 1996.

Nerenberg, M. A. H., and C. Essex, Correlation dimension and systematic geometric effects, *Phys. Rev. A*, 42, 7065-7074, 1990.

Porporato, A., and L. Ridolfi, Nonlinear analysis of river flow time sequences, *Water Resour. Res.*, 33(6), 1353-1367, 1997.

Provenzale, A., L. A. Smith, R. Vio and G. Murante, Distinguishing between low-dimensional dynamics and randomness in measured time series, *Physica D*, 58, 31-49, 1992.

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.

Sangoyomi, T. B., U. Lall, and H. D. I. Abarbanel, Nonlinear dynamics of the Great Salt Lake: dimension estimation, *Water Resour. Res.*, 32(1), 149-159, 1996.

Sauer, T., J. Yorke, and M. Casdagli, Embedology, *J. Stat. Phys.*, 65(3/4), 579-616, 1991.

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.

Sivakumar, B., S.-Y. Liong, and C.-Y. Liaw, Evidence of chaotic behavior in Singapore rainfall, *J. Am. Water Resour. Assoc.*, 34(2) 301-310, 1998.

Sivakumar, B., S.-Y. Liong, C.-Y. Liaw, and K.-K. Phoon, Singapore rainfall behavior: chaotic? *J. Hydrol. Eng.*, ASCE, 4(1), 38-48, 1999.

Sivakumar, B., R. Berndtsson, J. Olsson, K Jinno, and A. Kawamura, Dynamics of monthly rainfall-runoff process at the Göta basin: A search for chaos, *Hydrology and Earth System Sciences*, 4(3), 407-417, 2000.

Sivakumar, B., R. Berndtsson, J. Olsson, and K Jinno, Evidence of chaos in the rainfall-runoff process, *Hydrological Sciences Journal*, 46(1), 131-145, 2001.

Smith, L. A., Intrinsic limits on dimension calculations, *Phys. Lett. A*, 133, 283-288, 1988.

Tsonis, A. A., *Chaos: From Theory to Applications*, 274 pp., Plenum, New York, 1992.

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.

Wang, Q., and T. Y. Gan, Biases of correlation dimension estimates of streamflow data in the Canadian prairies, *Water Resour. Res.*, 34(9), 2329–2339, 1998.

Whitney, H., Differentiable manifolds, *Ann. Math.*, 37, 645, 1936.

Wilcox, B. P., M. S. Seyfried, and T. H. Matison, Searching for Chaotic Dynamics in Snowmelt Runoff, *Water Resour. Res.*, 27(6), 1005-1010, 1991.