Chaos in rainfall: variability, temporal scale and zeros

B. Sivakumar

ABSTRACT

A recent study on rainfall observed at the Leaf River basin reports that the presence of a large number of zeros in the data significantly underestimates the correlation dimension. The present study attempts to verify such a claim, by making predictions and comparing the results with the correlation dimensions. A nonlinear prediction method, which uses the concept of reconstruction of a single-variable series in a multi-dimensional phase space to represent the underlying dynamics, is employed. Correlation dimension analysis of only the non-zero rainfall series is also carried out for further verification. Rainfall data of four different temporal resolutions (or scales), i.e. daily, 2-day, 4-day and 8-day, are analyzed. The predictions for the finer-resolution (i.e. higher-resolution) rainfall are found to be better than those obtained for the coarser-resolution (i.e. lower-resolution) rainfall and seem to be consistent with the variability vs. predictability logic in a deterministic sense, i.e. higher prediction accuracy for data with lower correlation dimension and vice versa. An important implication of this result is that the presence of (a large number of) zeros in the rainfall data may not always result in an underestimation of the correlation dimension. The correlation dimensions estimated for the non-zero rainfall series are not significantly different when compared to those obtained for series including zero values, supporting the above. These results suggest that the low correlation dimensions for rainfall (in particular finer-resolution ones that commonly have a large number of zeros), as reported by past studies, could well be, or at least closer to, the actual dimensions of the rainfall processes studied.

Key words | chaos, correlation dimension, data size, noise, rainfall, scale, zeros

B. Sivakumar
Department of Land, Air and Water Resources,
University of California,
Davis, CA 95616,
USA
E-mail: *sbellie@ucdavis.edu*

INTRODUCTION

Investigation of the possible presence of low-dimensional chaotic behavior in rainfall dynamics has been of much interest lately (e.g. Rodriguez-Iturbe *et al.* 1989; Jayawardena & Lai 1994; Berndtsson *et al.* 1994; Sivakumar *et al.* 1999a, 2001; Sivakumar 2001). Though a variety of methods are available to identify the existence of chaos in a time series, the correlation dimension method has been the most widely used one in rainfall (and other hydrologic) time series, either as a method of proof or as a method of preliminary evidence. However, due to the potential limitations of this method, criticisms on its application to rainfall (and other hydrologic) time series and the reported results are very common [see, for instance, Ghilardi & Rosso (1990), Sivakumar (2000), Schertzer *et al.* (2002), and Sivakumar *et al.* (2002) for discussions]. In addition to the problems of insufficient data size, sampling frequency and presence of noise, the analysis of rainfall (and other hydrologic) time series, finer-resolutions in particular, and the outcomes may also be (significantly) influenced because of the presence of zeros. For instance, in the presence of a large number of zeros, the reconstructed hypersurface in phase space will tend to a point and may result in a significant underestimation of the correlation dimension (e.g. Tsonis *et al.* 1994).

A recent study by Sivakumar (2001), investigating the behavior of rainfall dynamics at four different resolutions (daily, 2-day, 4-day and 8-day) observed at the Leaf River basin, Mississippi, USA, reports that the presence of a large number of zeros indeed significantly influences the correlation dimension estimation (underestimation). Having obtained a lower correlation dimension for the finer-resolution rainfall (e.g. daily) compared to the coarser-resolution rainfall (e.g. 8-day), the study claims that the correlation dimension of the former is significantly underestimated due primarily to the presence of a large number of zeros. Such a claim (rather than that the correlation dimension of the coarser-resolution rainfall is overestimated) is made based on the following observations:

(1) the coefficient of variation (or variability) of the finer-resolution rainfall is higher than that of the coarser-resolution rainfall, suggesting that the correlation dimension of the former must be higher than that of the latter;
(2) the possibility of underestimation of correlation dimension due to insufficient data size (in the finer-resolution data) can be eliminated, as the finer-resolution rainfall consist of more (at least twice the) number of data than the coarser-resolution ones; and
(3) the possible overestimation of the correlation dimension of coarser-resolution data due to the presence of higher level of noise may be considered only small, as noise has only very little influence on the correlation dimension estimation (Sivakumar *et al.* 1999b).

The significance of the above reasons and, hence, the claim regarding the influence of the presence of zeros lie entirely on the statistical parameter used to measure the variability of the rainfall series and to compare with the correlation dimension. The justification of the use of the coefficient of variation to measure the variability lies in the fact that it is a very common and reliable statistical tool. Further justification of its use with respect to rainfall of different resolutions may be that, in general, the finer-resolution rainfall yields a higher coefficient of variation (higher variability) than the coarser-resolution rainfall, which is consistent with the assumption that, in general, the former is less predictable than the latter. Whether or not such an assumption is valid for every situation remains a question, though its validity in certain (or even in a large number of) cases cannot be excluded. On the other hand, one possible way to verify the accuracy of the correlation dimensions is by making predictions and comparing the results, as time series with lower correlation dimensions should, in general, yield better predictions than time series of higher correlation dimensions, from a deterministic dynamic point of view.

With this in mind, the present study attempts to predict the same (four) rainfall series (daily, 2-day, 4-day and 8-day) observed at the Leaf River basin, studied by Sivakumar (2001) for the correlation dimension estimation. A nonlinear prediction method, which uses the concept of reconstruction of a single-variable series in a multi-dimensional phase space to represent the underlying dynamics (same concept as the one used in the correlation dimension estimation), is employed. The correlation dimensions of only the non-zero rainfall series are also estimated for further verification of the results.

The organization of this paper is as follows. First, the correlation dimension and nonlinear prediction methods are briefly explained. Next, analyses of the four rainfall series and the results are presented. A comparison between the correlation dimension results and the prediction results is also made. Finally, important conclusions drawn from the present study are provided.

METHODS

Correlation dimension method

The goal of determining the dimension of a time series is that the dimensionality furnishes information on the number of dominant variables present in the evolution of the corresponding dynamic system. Dimension analysis will also reveal the extent to which the variations in the time series are concentrated on a subset of the space of all possible variations. Correlation dimension is a measure of the extent to which the presence of a data point affects the position of the other points lying on the attractor in phase space (or coordinate system).

The correlation dimension method uses the correlation integral (or function) to distinguish between low-dimensional and high-dimensional systems. Herein, the Grassberger–Procaccia correlation dimension algorithm (Grassberger & Procaccia 1983) is employed. The algorithm uses the concept of phase-space reconstruction for representing the dynamics of the system from an available time series. Using a (single-variable) time series X_i, where $i = 1, 2, \ldots, N$, the (multi-dimensional) phase space can be reconstructed using the method of delays, according to (Takens 1981):

$$Y_j = (X_j, X_{j+\tau}, X_{j+2\tau}, \ldots, X_{j+(m-1)\tau}) \tag{1}$$

where $j = 1, 2, \ldots, N - (m-1)\tau/\Delta t$, m is the dimension of the vector Y_j, also called the embedding dimension, τ is a delay time and Δt is the sampling time. For an m-dimensional phase space, the correlation function $C(r)$ is given by

$$C(r) = \lim_{N\to\infty} \frac{2}{N(N-1)} \sum_{\substack{i,j \\ (1\le i<j\le N)}} H(r - |Y_i - Y_j|) \tag{2}$$

where H is the Heaviside step function, with $H(u) = 1$ for $u > 0$, and $H(u) = 0$ for $u \le 0$, where $u = r - |Y_i - Y_j|$, and r is the radius of sphere centered on Y_i or Y_j. If the time series is characterized by an attractor, then the correlation function $C(r)$ and the radius r are related according to

$$\underset{\substack{r\to 0 \\ N\to\infty}}{C(r)} \approx \alpha r^{\nu} \tag{3}$$

where α is a constant and ν is the correlation exponent or the slope of the log $C(r)$ versus log r plot. The slope is generally estimated by a least square fit of a straight line over a certain range of r, called the scaling region. The presence/absence of chaos in the time series can be identified using the correlation exponent (ν) versus embedding dimension (m) plot. If ν saturates after a certain m and the saturation value is low, then the system is generally considered to exhibit low-dimensional deterministic behavior. The saturation value of ν is defined as the correlation dimension (d) of the attractor. On the other hand, if ν increases without bound with increase in m, the system under investigation is generally considered as stochastic.

Nonlinear prediction method

The nonlinear prediction method also employs the concept of phase-space reconstruction. A phase-space reconstruction in a dimension m (Equation (1)) allows one to interpret the underlying dynamics in the form of an m-dimensional map f_T, that is:

$$Y_{j+T} = f_T(Y_j) \tag{4}$$

where Y_j and Y_{j+T} are vectors of dimension m, describing the state of the system at times j (current state) and $j + T$ (future state), respectively. The task now is to find an appropriate expression for f_T (e.g. F_T) to predict the future. There are several approaches for determining F_T, which may broadly be grouped under two categories: (1) global approximation; and (2) local approximation. In this study, a local approximation approach (e.g. Farmer & Sidorowich 1987) is employed.

In the local approximation approach, the f_T domain is subdivided into many subsets (neighborhoods), each of which identifies some approximations F_T, valid only in that subset. In this way, the underlying system dynamics are represented step by step locally in the phase space. The identification of the sets in which to subdivide the domain is done by fixing a metric $\| \|$ and, given the starting point Y_j from which the forecast is initiated, identifying neighbors Y_j^p, $p = 1, 2, \ldots, k$, with $j^p < j$, nearest to Y_j, which constitute the set corresponding to Y_j. The local functions can then be built, which take each point in the neighborhood to the next neighborhood: Y_j^p, to Y_{j+1}^p. The local map F_T, which does this, is determined by a least squares fit minimizing

$$\sum_{p=1}^{k} \|Y_{j+1}^p - F_T Y_j^p\|^2. \tag{5}$$

The local maps may be learned in the form of local averaging (e.g. Farmer & Sidorowich 1987) or local polynomials (e.g. Abarbanel, 1996). The present study uses local polynomials, and predictions are made forward from a new point Z_0 using these local maps. For the new point Z_0, the nearest neighbor in the training set is found, denoted as Y_q. Then the evolution of Z_0 is found, which is denoted as Z_1 and is given by

$$Z_1 = F_q(Z_0) \tag{6}$$

Then the nearest neighbor to Z_1 is found and the procedure is repeated to predict the subsequent values. The accuracy of prediction may be evaluated using any of the statistical measures, such as correlation coefficient and root mean square error.

ANALYSIS, RESULTS AND DISCUSSION

Study area and data

In the present study, rainfall time series observed at the Leaf River basin in the state of Mississippi are analyzed. The climate in this region is humid subtropical, characterized by short, mild temperate winters, long, hot summers, and rainfall that is fairly evenly distributed throughout the year. The mean annual rainfall at the basin is about 1350 mm. March is the wettest month with a mean rainfall of about 160 mm, while October is the driest month with a mean rainfall of about 80 mm. Further details about the climatic conditions at the basin can be found in Sivakumar *et al.* (2001). For the present investigation, rainfall data of four different resolutions, daily, 2-day, 4-day and 8-day, observed over a period of 25 years (January 1963–December 1987) are used. Figures 1(a–d) show the variations of these four series, respectively, and Table 1 presents some of their important statistics.

Phase-space reconstruction

As a first step in the analysis, the phase spaces for the four rainfall series are reconstructed according to Equation (1). Examples of phase-space reconstructions for the daily, 2-day, 4-day and 8-day series are presented in Figs. 2(a–d), respectively. These figures show reconstructions in two dimensions ($m = 2$) and with delay time $\tau = 1$, i.e. the projection of the attractor on the plane $\{X_i, X_{i+1}\}$. For each of the four series, the projection does not seem to yield a clearly defined attractor, but only occupies a large space in the phase-space diagram. This, unfortunately, does not allow any useful interpretation regarding the variability of rainfall relative to the four resolutions. Also, a comparison of the four projections does not seem to indicate any definitive (increasing or decreasing) trend in the rainfall variability, from finer to coarser scale of observation.

It is important to note that an appropriate τ for phase-space reconstruction is necessary because only an optimum τ gives the best separation of neighboring trajectories within the minimum embedding space, whereas an inappropriate τ may lead to underestimation or overestimation of correlation dimension (Havstad & Ehlers 1989). Many researchers have addressed the issue of τ selection, and also proposed different methods and guidelines. Popular among them are the autocorrelation function (Holzfuss & Mayer-Kress 1986), the mutual information (Fraser & Swinney 1986) and the correlation integral (Leibert

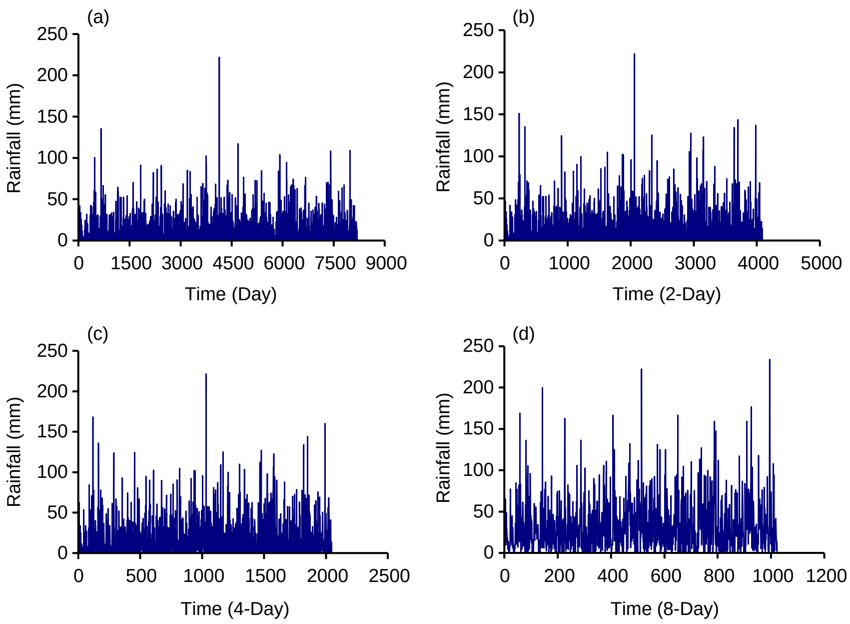

Figure 1 | Time series plots for rainfall data of different resolutions from the Leaf River basin: (a) daily; (b) 2-day; (c) 4-day and (d) 8-day.

Table 1 | Statistics of rainfall data of different resolutions in the Leaf River basin

Parameter	Daily	2-day	4-day	8-day
Number of data	8192	4096	2048	1024
Mean	4.03	8.06	16.12	32.24
Standard deviation	10.47	15.61	22.08	31.90
Variance	109.46	243.56	487.62	1017.62
Coefficient of variation	2.60	1.94	1.37	0.99
Maximum value	221.52	221.52	221.52	234.03
Minimum value	0.00	0.00	0.00	0.00
Number of zeros	4467 (54.53%)	1633 (39.87%)	412 (20.12%)	62 (6.05%)

& Schuster 1989). However, the autocorrelation function method has widely been used in hydrologic studies and τ has been chosen as the lag time at which the autocorrelation function first crosses the zero line or attains 0.1 or 0.5.

Even though these selections have been shown to provide reliable results (see, for instance, Sivakumar (2000) for details), there is neither a consensus nor a clear-cut guideline. The τ values resulting from the autocorrelation function (or any other) method may sometimes not have a physical (or perceptual or conceptual) relation to the system dynamics under investigation. For example, the first zero value of the autocorrelation function attained for the above daily, 2-day, 4-day and 8-day rainfall series (figures not shown) is at lag times of 2 (2 days), 5 (10 days),

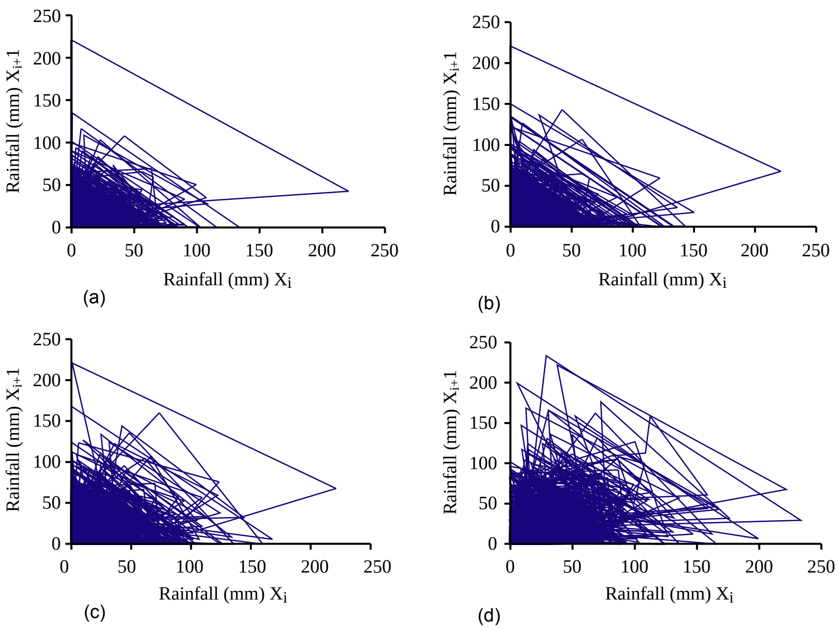

Figure 2 | Phase-space plots for rainfall data of different resolutions from the Leaf River basin: (a) daily; (b) 2-day; (c) 4-day and (d) 8-day.

3 (12 days) and 5 (40 days), respectively (Table 2). Whether, and how, these lag time values, reflecting a time length of 2–40 days, translate into the actual physical mechanisms that take place in the rainfall dynamics is not known. While this issue warrants further investigation, for the present purpose, the above values are used for phase-space reconstruction.

Correlation dimension

The correlation dimensions of the above four rainfall series are computed using the Grassberger–Procaccia algorithm, explained earlier (see also Sivakumar (2001) for details of intermediate results). Figure 3 shows the correlation dimension results (relationship between the correlation exponent values, ν, and the embedding dimension values, m) for the four rainfall series. As can be seen, for all four series, the correlation exponent value increases with the embedding dimension up to a certain point and then saturates beyond that point. The saturation of the correlation exponent beyond a certain embedding dimension value may be an indication of the existence of deterministic dynamics. The saturation values of the correlation exponent (or correlation dimension) for the four rainfall series are, respectively, about 4.82, 5.26, 6.42 and 8.87 (Table 2). The finite correlation dimensions obtained for the four rainfall series may be an indication that all four series exhibit chaotic behavior (even though the correlation dimensions seem to be on the higher side).

Table 2 | Correlation dimension and nonlinear prediction results for rainfall data of different resolutions in the Leaf River basin

Parameter	Daily	2-day	4-day	8-day
Rainfall with Zeros				
(i) Correlation dimension				
Delay time	2	5	3	5
Correlation dimension	4.82	5.26	6.42	8.87
(ii) Nonlinear prediction				
Correlation coefficient	0.783	0.276	0.332	0.339
Root mean square error	7.211	13.752	21.363	32.623
Rainfall without Zeros				
Coefficient of variation	1.59	1.36	1.14	0.93
Correlation dimension	5.92	6.62	8.16	9.46

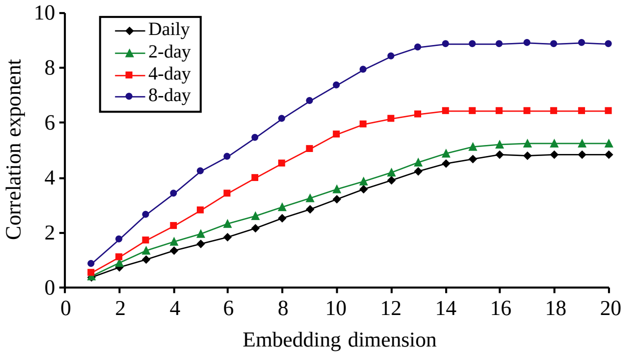

Figure 3 | Correlation dimension results for rainfall data of different resolutions in the Leaf River basin: relationship between correlation exponent (ν) and embedding dimension (m).

Nonlinear prediction

The nonlinear prediction method with the local polynomial approach, explained earlier, is now employed for the four rainfall series. The predictions are made for only one-step ahead (i.e. lead time = 1). Embedding dimensions from 1–10 are used for reconstructing the phase space. It is found, in general, that the number of neighbors needed to obtain the optimal results increases with increasing embedding dimension (see, for instance, Sivakumar *et al.* (1999a, 2001) for details). The measures of prediction accuracy are the correlation coefficient (CC) and the root mean square error ($RMSE$). In addition to these two measures, the time series plots and the scatter diagrams are also used to select the best results among those obtained for the above cases.

Figures 4(a) and (b), for instance, show the prediction results in terms of the correlation coefficient and RMSE, respectively, against the embedding dimension for the four rainfall series, and Table 2 presents the best prediction results for these series. As can be seen, reasonably good predictions are obtained for the daily rainfall series ($CC = 0.783$), whereas the predictions for the remaining three series ($CC = 0.276$, 0.332 and 0.339, respectively) are significantly worse compared to that of the daily series. These results seem to indicate that data aggregation in

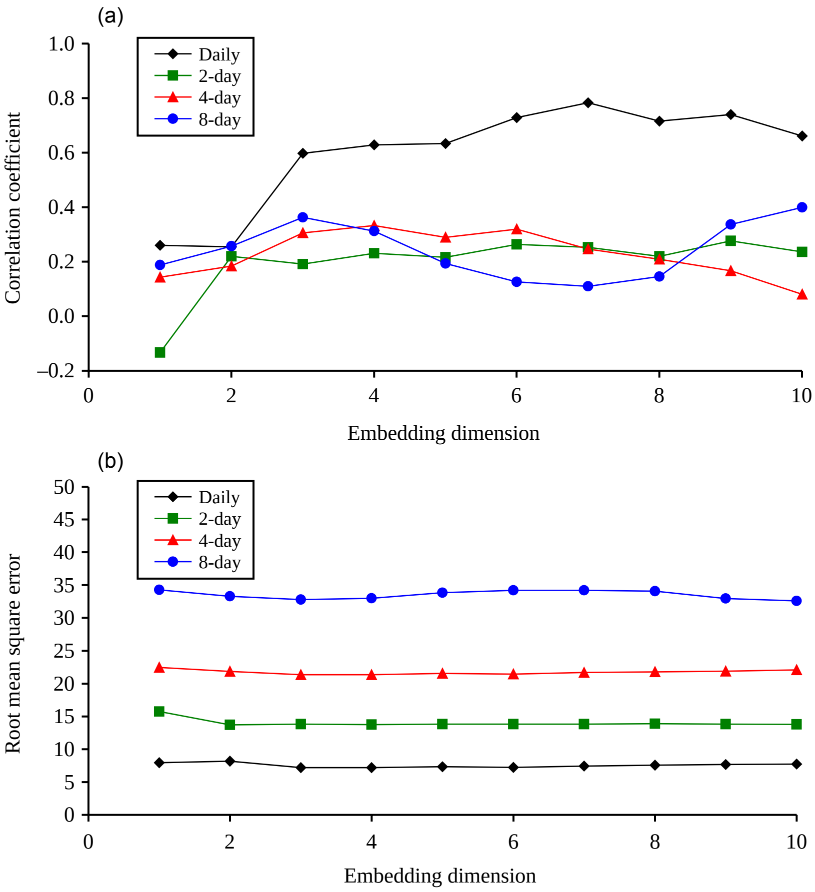

Figure 4 | Nonlinear prediction results for rainfall data of different resolutions in the Leaf River basin: (a) correlation coefficient (*CC*) versus embedding dimension (*m*); and (b) root mean square error (*RMSE*) versus embedding dimension (*m*).

temporal scale yields less accurate predictions, even though such an interpretation is not entirely accurate.

Comparison of correlation dimension and prediction results

The correlation dimension, d, of a time series represents, in a way, the degree of variability or irregularity of the values in the series. A time series with a higher variability in the values provide a higher correlation dimension, whereas a lower correlation dimension would be the result of lower variability. With regards to this, the correlation dimension results obtained for the four rainfall series (Table 2) indicate that the daily rainfall (with $d = 4.82$) exhibits the lowest variability, whereas the 8-day rainfall (with $d = 8.87$) exhibits the highest variability. The 2-day and 4-day rainfall (with $d = 5.26$ and $d = 6.42$, respectively) exhibit, in order, higher and higher variabilities with respect to the daily series. A comparison of the degree of variability of the four series based on the correlation dimension (Table 2) and the coefficient of variation (Table 1) indicates an inverse relationship between the two. That is, the time series with the highest coefficient of variation (i.e. daily) yields the lowest correlation dimension and vice versa. Such an observation is contrary to the concept of the correlation dimension (or the coefficient of variation), at least from a deterministic dynamic point of view.

It is not known, at this stage, what causes such a problem, but one possible explanation could be that the correlation dimension of the finer-resolution rainfall series is (significantly) underestimated due to the presence of a large number of zeros, as claimed by Sivakumar (2001). This is because, as explained previously, in the presence of a large number of zeros (or any other single value) in a time series, the reconstructed hypersurface in phase space will tend to a point (e.g. Tsonis *et al.* 1994). As can be seen in Table 1, the finest-resolution (daily) rainfall consists of 54.53% of zeros compared to only 6.05% of zeros in the coarsest resolution (8-day) rainfall. The above explanation may also be supported by the facts that neither the underestimation of the correlation dimension of finer-resolution rainfall due to insufficient data size nor the (significant) overestimation of the coarser-resolution rainfall due to the higher level of noise is possible (see Sivakumar (2001) for details). However, one has to be cautious in providing such an interpretation, as the influence of the presence of zeros is not entirely known yet.

It is appropriate to note, at this point, that since the correlation dimension of a time series represents the degree of variability of the series, a time series with a lower correlation dimension should, in all probability, provide better predictions than the one with a higher correlation dimension, at least from a deterministic dynamic point of view. This suggests that, for example, the daily rainfall series (with $d = 4.82$) should yield significantly better predictions than that of the 8-day series (with $d = 8.87$). The prediction results obtained (Fig. 4 and Table 2) for the four rainfall series seem to reflect this point, as the prediction accuracy obtained for the daily rainfall series ($CC = 0.783$) is significantly better than the one obtained for 8-day series ($CC = 0.339$); see also the scatter diagrams, depicting the predicted versus observed rainfall values, for the daily and 8-day series (Fig. 5), for a better representation of this, including over-predictions (for low rainfalls) and under-predictions (for high rainfalls) in these cases (the "averaging" effect of the neighbor searching procedure is also seen). The results obtained for the four series seem to show, though not as consistently as one would like to see, an inverse relationship between the prediction accuracy and the correlation dimension. Such an inverse relationship between the correlation dimension and prediction accuracy seems to be consistent, from the above point of view of variability vs. predictability. This, in turn, seems to suggest that the correlation dimensions obtained for the four rainfall series could at least be closer to the actual correlation dimensions of the series, if not accurate. This may be an indication that the presence of a large number of zeros may not always result in an underestimation of the correlation dimension. An important implication of such an observation is that the correlation dimensions obtained in the present study, and those reported by past studies for other rainfall series, could well be closer to the actual correlation dimensions of the rainfall dynamics.

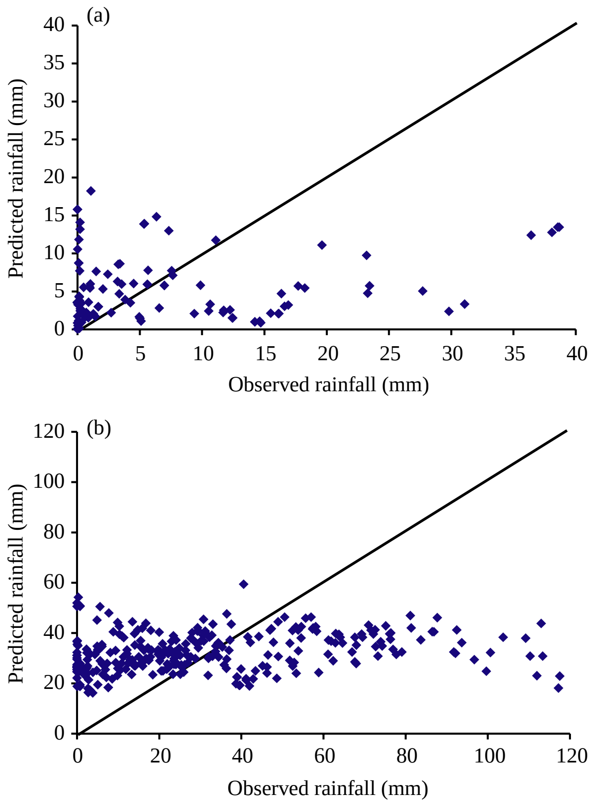

Figure 5 | Scatter plot comparisons between predicted and observed rainfall values: (a) daily and (b) 8-day.

Having said the above, a relevant question is whether the removal of zeros from the rainfall series would alter the results and, hence, the above conclusion. To this effect, the correlation dimensions for the four series are estimated after removing the zero values. The correlation dimension results thus obtained for the non-zero values are presented in Fig. 6 and Table 2. As the results indicate, there is still an inverse relationship between the correlation dimension and the coefficient of variation, though there are some differences (increases) in the correlation dimensions. Such a trend is no different than the one observed when the zero values are also included in the correlation dimension estimation. On the other hand, whether or not the increase in the correlation dimensions is due to the removal of zeros is

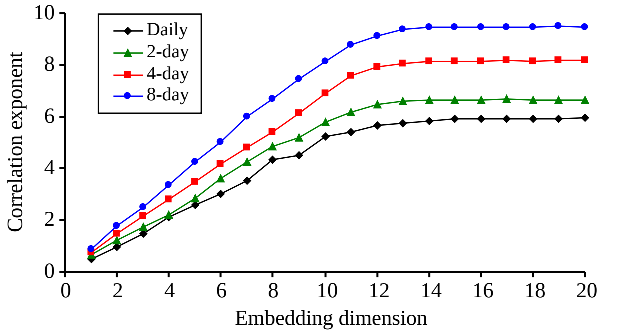

Figure 6 | Correlation dimension results for rainfall data of different resolutions without zero values: relationship between correlation exponent (ν) and embedding dimension (m).

difficult to answer. It is important to note, however, that the removal of zeros from a time series is questionable, as a zero value is indicative of, and equally important to understand (in particular, identify and predict), how the dynamics of the system are evolved (see Sivakumar *et al.* (2001) for details).

Finally, a brief discussion on the inability of the nonlinear prediction method to yield accurate results is in order now. One possible reason for this could be the presence of noise in the rainfall series, particularly in the coarser-resolution ones. The rainfall measurement is generally influenced by a large number of factors, such as wind, wetting, evaporation, gage exposure, instrumentation and human error in reading the rainfall data. Added to this, and more importantly, is the intrinsic erratic nature of the rainfall process. The coarser-resolution data used in this study is also influenced by a certain imprecision due to the round-off errors resulting from the conversion of the daily data to the particular resolution. The possible presence of significantly higher levels of noise (or noise-to-signal ratio) in the coarser-resolution rainfall series could be an important reason for the much less accurate prediction results, when compared to those obtained for the daily rainfall series. There are also certain other factors that may cause the above situation by influencing the predictions of one or more of the above series. These may include insufficient data size at coarser resolutions (though small data size may not be as serious as a problem as it is believed to be; see Sivakumar *et al.* (2002) for details), improper reconstruction of phase space due to inappropriate selection of τ and m and loss of temporal structure due to data aggregation. Investigation of the influence of all these factors is crucial to verify and/or strengthen the present results and interpretations.

CONCLUSIONS

An important limitation in the application of the correlation dimension method in rainfall time series could be the presence of a large number of zeros, since in its presence the reconstructed hypersurface in phase space will tend to a point and may result in an underestimation of the correlation dimension. A recent study by Sivakumar (2001) reported such a problem, based on the lower correlation dimension obtained for the finer-resolution rainfall (with higher coefficient of variation) than that of the coarser-resolution rainfall (with lower coefficient of variation). An attempt was made in the present study to verify such a claim by making predictions of rainfall data for four different successively doubled resolutions: daily, 2-day, 4-day and 8-day, respectively. The correlation dimensions of the non-zero rainfall series were also estimated for further verification.

The prediction results obtained for the four rainfall series indicated an inverse relationship between the prediction accuracy and the correlation dimension. Such an inverse relationship between the dimension and prediction seems to be consistent with the variability vs. predictability logic, from a deterministic dynamic perspective. This, in turn, suggests that the correlation dimensions obtained for the four rainfall series from the Leaf River basin may be closer to the actual correlation dimensions and seems to eliminate the possibility of an underestimation of the correlation dimension in the finer-resolution rainfall due to the presence of a large number of zeros. Also, the correlation dimensions obtained for the non-zero rainfall series were not significantly different from the ones obtained for series including zero values, thus supporting the above. All these results imply that the correlation dimensions reported by past studies for several rainfall series (having zeros), finer-resolution ones in particular, could indeed be closer to the actual dimensions of the rainfall dynamics. However, further investigations on the effects of data quantity and quality and the parameters involved in the correlation dimension and nonlinear prediction methods are necessary to verify and/or strengthen the present results. Use of other nonlinear and chaos identification methods, such as surrogate data method, false nearest-neighbor algorithm and Lyapunov exponent method, would also help, in one way or another, this cause. Investigations along these directions are underway, details of which will be reported in future papers.

REFERENCES

Abarbanel, H. D. I. 1996 *Analysis of Observed Chaotic Data*. Springer-Verlag, New York.

Berndtsson, R., Jinno, K., Kawamura, A., Olsson, J. & Xu, S. 1994 Dynamical systems theory applied to long-term temperature and precipitation time series. *Trends Hydrol.* **1**, 291–297.

Farmer, D. J. & Sidorowich, J. J. 1987 Predicting chaotic time series. *Phys. Rev. Lett.* **59**, 845–848.

Frazer, A. M. & Swinney, H. L. 1986 Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33** (2), 1134–1140.

Grassberger, P. & Procaccia, I. 1983 Measuring the strangeness of strange attractors. *Physica D* **9**, 189–208.

Ghilardi, P. & Rosso, R. 1990 Comment on "chaos in rainfall" by Rodriguez-Iturbe *et al. Wat. Res. Res*. **26** (8), 1837–1839.

Havstad, J. W. & Ehlers, C. L. 1989 Attractor dimension of nonstationary dynamical systems from small data sets. *Phys. Rev. A* **39** (2), 845–853.

Holzfuss, J. & Mayer-Kress, G. 1986 An approach to error-estimation in the application of dimension algorithms. In *Dimensions and Entropies in Chaotic Systems* (ed. G. Mayer-Kress), pp. 114–122. Springer-Verlag, New York.

Jayawardena, A. W. & Lai, F. 1994 Analysis and prediction of chaos in rainfall and stream flow time series. *J. Hydrol.* **153**, 23–52.

Leibert, W. & Schuster, H. G. 1989 Proper choice of the time delay for the analysis of chaotic time series. *Phys. Lett. A* **141**, 386–390.

Rodriguez-Iturbe, I., De Power, F. B., Sharifi, M. B. & Georgakakos, K. P. 1989 Chaos in rainfall. *Wat. Res. Res.* **25** (7), 1667–1675.

Schertzer, D., Tchiguirinskaia, I., Lovejoy, S., Hubert, P. & Bendjoudi, H. 2002 Which chaos in the rainfall-runoff process? A discussion on "Evidence of chaos in the rainfall-runoff process" by Sivakumar *et al. Hydrol. Sci. J.* **47** (1), 139–147.

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

Sivakumar, B. 2001 Rainfall dynamics at different temporal scales: a chaotic perspective. *Hydrol. Earth Syst. Sci.* **5** (4), 645–651.

Sivakumar, B., Berndtsson, R., Olsson, J. & Jinno, K. 2002 Reply to "Which chaos in the rainfall-runoff process" by Schertzer *et al*. *. Hydrol. Sci. J.* **47** (1), 149–158.

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

Sivakumar, B., Phoon, K. K., Liong, S. Y. & Liaw, C. Y. 1999b A systematic approach to noise reduction in chaotic hydrological time series. *J. Hydrol.* **219** (3/4), 103–135.

Sivakumar, B., Sorooshian, S., Gupta, H. V. & Gao, X. 2001 A chaotic approach to rainfall disaggregation. *Wat. Res. Res.* **37** (1), 61–72.

Takens, F. 1981 Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence. Lecture Notes in Mathematics 898* (ed. D. A. Rand & L. S. Young), pp. 366–381. Springer-Verlag, Berlin.

Tsonis, A. A., Triantafyllou, G. N., Elsner, J. B., Holdzkom, J. J. II. & Kirwan, A. D. Jr. 1994 An investigation of the ability of nonlinear methods to infer dynamics from observables. *Bull. Am. Meteorol. Soc.* **75** (9), 1623–1633.