Hydrology and Earth System Sciences, 4(3), *407–417* (2000) © EGS

Hydrology & Earth System Sciences

Dynamics of monthly rainfall-runoff process at the Göta basin: A search for chaos

B. Sivakumar[1], R. Berndtsson[2], J. Olsson[3], K. Jinno[3] and A. Kawamura[3]

[1]Department of Land, Air & Water Resources, University of California, Davis, CA 95616, USA
[2]Department of Water Resources Engineering, Lund Institute of Technology, Lund University, P.O. Box 118, S-221 00 Lund, Sweden
[3]Institute of Environmental Systems, Faculty of Engineering, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
e-mail for corresponding author: sbellie@ucdavis.edu

Abstract

Sivakumar *et al.* (2000a), by employing the correlation dimension method, provided preliminary evidence of the existence of chaos in the monthly rainfall-runoff process at the Göta basin in Sweden. The present study verifies and supports the earlier results and strengthens such evidence. The study analyses the monthly rainfall, runoff and runoff coefficient series using the nonlinear prediction method, and the presence of chaos is investigated through an inverse approach, i.e. identifying chaos from the results of the prediction. The presence of an optimal embedding dimension (the embedding dimension with the best prediction accuracy) for each of the three series indicates the existence of chaos in the rainfall-runoff process, providing additional support to the results obtained using the correlation dimension method. The reasonably good predictions achieved, particularly for the runoff series, suggest that the dynamics of the rainfall-runoff process could be understood from a chaotic perspective. The predictions are also consistent with the correlation dimension results obtained in the earlier study, i.e. higher prediction accuracy for series with a lower dimension and vice-versa, so that the correlation dimension method can indeed be used as a preliminary indicator of chaos. However, the optimal embedding dimensions obtained from the prediction method are considerably less than the minimum dimensions essential to embed the attractor, as obtained by the correlation dimension method. A possible explanation for this could be the presence of noise in the series, since the effects of noise at higher embedding dimensions could be significantly greater than that at lower embedding dimensions.

Keywords: Rainfall-runoff; runoff coefficient; chaos; phase-space; correlation dimension; nonlinear prediction; noise

Introduction

The recent shift in complex hydrological problems, such as real-time flood and drought forecasting, management of water resources, pollution transport and soil water infiltration, necessitates accurate modelling of the rainfall-runoff process in a region. Even though, the past few decades witnessed the proposal of a wide variety of approaches and the development of a large number of models to understand the dynamics of the rainfall-runoff process, a unified approach to the problem is still missing. This is due to, (1) the considerable temporal and spatial variability exhibited by the rainfall-runoff process; and (2) the limitation in the availability of 'appropriate' mathematical tools to exploit the dynamics underlying the rainfall-runoff process.

The considerable spatial and temporal variability exhibited by the rainfall-runoff process is due to the various physical mechanisms (acting on a huge range of temporal and spatial scales) that govern the dynamics of the process. The rainfall-runoff process depends not only on the space-time distribution of the rainfall occurrence, but also on the kind and the state of the basin, which, in turn, depend on climatic conditions and vegetation states. Therefore, what is really important is a unified description of the complex behaviour of the dynamical system arising from the coupling of all its components. Although not all components are complex in themselves, the size of the space-time domain, the number of individual processes involved, and the fact that almost all of them present some degree of nonlinearity, make the total resulting rainfall-runoff process highly complicated. For such a complex system, the only possibility for realistic modelling seems to be that only a few of the various mechanisms become prevalent in the process, so that the system dynamics are simplified with a corresponding reduction in the number of the effective degrees of freedom. Therefore, the notion of chaos theory, i.e. seemingly complex behaviour could be the result of simple determinism influenced by only a few nonlinear interdependent variables, and the related methods of nonlinear dynamics could contribute to an understanding of the rainfall-runoff dynamics.

Applications of chaos theory to understand the dynamics of hydrological processes, particularly rainfall and runoff, have been gaining momentum (e.g. Hense, 1987; Rodriguez-Iturbe *et al.*, 1989, 1991b; Sharifi *et al.*, 1990; Islam *et al.*, 1993; Tsonis *et al.*, 1993; Berndtsson *et al.*, 1994; Jayawardena and Lai, 1994; Georgakakos *et al.*, 1995; Koutsoyiannis and Pachakis, 1996; Porporato and Ridolfi, 1996, 1997; Puente and Obregon, 1996; Sangoyomi *et al.*, 1996; Liu *et al.*, 1998; Wang and Gan, 1998; Sivakumar, 2000; Sivakumar *et al.*, 1999a, b, 2000b, c; Stehlik, 1999; Krasovskaia *et al.*, 1999; Jayawardena and Gurung, 2000). Though the primary objective of those studies was to investigate the existence of chaos in the processes, attempts were also made for prediction (e.g. Jayawardena and Lai, 1994; Porporato and Ridolfi, 1996, 1997; Liu *et al.*, 1998; Sivakumar *et al.*, 1999a, b, 2000c; Jayawardena and Gurung, 2000), noise reduction (e.g. Porporato and Ridolfi, 1997; Sivakumar *et al.*, 1999b; Jayawardena and Gurung, 2000) and disaggregation (e.g. Sivakumar *et al.*, 2000b).

Past studies investigating chaos in rainfall and runoff processes have been limited either to the rainfall process alone (e.g. Hense, 1987; Rodriguez-Iturbe *et al.*, 1989; Sharifi *et al.*, 1990; Islam *et al.*, 1993; Tsonis *et al.*, 1993; Berndtsson *et al.*, 1994; Jayawardena and Lai, 1994; Georgakakos *et al.*, 1995; Koutsoyiannis and Pachakis, 1996; Puente and Obregon, 1996; Sivakumar *et al.*, 1999a, b, 2000b) or to the runoff process alone (e.g. Jayawardena and Lai, 1994; Porporato and Ridolfi, 1996, 1997; Liu *et al.*, 1998; Wang and Gan, 1998; Stehlik, 1999; Krasovskaia *et al.*, 1999; Jayawardena and Gurung, 2000; Sivakumar *et al.*, 2000c), but not to rainfall-runoff processes as a whole in a basin. As a result, none of these studies provides information regarding the possibility of the existence of chaotic behaviour in the joint rainfall-runoff process in a basin.

Sivakumar *et al.* (2000a) investigated the possibility of understanding the dynamics of the joint rainfall-runoff process from a chaotic dynamical perspective. Monthly rainfall and runoff series observed over a period of 131 years (January 1807–December 1937) at the Göta basin in Sweden were analysed separately and jointly (using the runoff coefficient). The underlying assumption behind investigating the two series separately was that the individual behaviour of the dynamics of the rainfall (input) and the runoff (output) processes could provide important information about the behaviour of the dynamics of the joint rainfall-runoff process. The runoff coefficient (given by the ratio of runoff to rainfall) was considered as a parameter connecting rainfall and runoff and, therefore, to represent better the rainfall-runoff process as a whole. The correlation dimension method (e.g. Grassberger and Procaccia, 1983a, b), one of the fundamental methods developed in the field of chaos theory, was employed to investigate the existence of chaos. The finite correlation dimensions obtained for the rainfall, runoff and runoff coefficient series indicated the presence of chaos, providing preliminary evidence regarding the possible existence of chaos in the dynamics of the rainfall-runoff process (see below for details).

It is important to note, however, that the observation of the finite correlation dimension can be taken only as a preliminary indicator of the presence of chaos and not as strong evidence, because finite correlation dimensions could be observed even for linear stochastic processes (e.g. Osborne and Provenzale, 1989). On the other hand, insufficient data size and presence of noise can also influence significantly the correlation dimension estimation. For example, a small data set may result in a significant underestimation of the correlation dimension, whereas the presence of noise may overestimate the dimension. Therefore, the results obtained using the correlation dimension method must be verified and the existence (or non-existence) of chaos must be substantiated with additional evidence. The present study is aimed at verifying the earlier results and, hence, strengthening the evidence regarding the existence of chaos in the rainfall-runoff process, by employing a very promising chaos identification method, the nonlinear prediction method, to the above three time series.

Regarding the limitations of the correlation dimension method, a detailed discussion of the important issues in the application of the method to hydrological time series, e.g. data size, noise, delay time, etc., has been made in a study by Sivakumar (2000). Also, Sivakumar *et al.* (2000c) investigated the reliability of the correlation dimension estimation in short hydrological (runoff) time series, by evaluating its accuracy using the prediction results obtained from two methods: (1) the phase-space reconstruction method; and (2) the artificial neural networks technique. The results reveal that the accuracy of the correlation dimension depends primarily on whether the length of the time series would be sufficient to represent the changes that the system undergoes over a period of time, rather than the data size in terms of the number of values in the time series, suggesting that the correlation dimension could be a reliable indicator of low-dimensional chaos in short hydrological time series. The authors believe that the rainfall and runoff series analysed in the present study, with a 131-year monthly record, is sufficient to represent the changes in the system with time.

The organization of this paper is as follows. Firstly, a brief account of the driving forces behind the application of chaos theory to understand the dynamics of the rainfall-runoff process is presented, by providing a qualitative discussion on why and where deterministic components could be expected in the rainfall-runoff system. Secondly, some preliminary evidence, including the results presented by Sivakumar *et al.* (2000a), regarding the existence of chaos in the rainfall-runoff process is provided. Thirdly, the nonlinear prediction analyses of the rainfall-runoff process are given and the results discussed. Finally, the conclusions drawn from the present study and the scope for further research are presented.

Rainfall-runoff dynamics: determinism versus stochasticity

The system governing rainfall-runoff process may be seen as made up of a cascade of coupled components, each determining in some way the state of the others. When a great number of variables is involved in a system, a high-dimensional dynamics (practically indistinguishable from a purely stochastic system) is usually expected. In this case the system dimension, i.e. its number of active degrees of freedom, is so high as to preclude practically any kind of deterministic description. However, the intensification of the relative importance of some mechanisms may allow an accurate low-dimensional deterministic description. The successes obtained by hydrologists with physically-based models of the rainfall-runoff process support this possibility. Furthermore, recent developments of dynamical system theory have shown that also for systems governed by infinite dimensional (i.e. partial differential) equations the dynamics may develop in a lower, finite dimensional attractors (Cross and Hohenberg, 1993), and phenomena of generalised synchronisation may take place in systems made up of interacting nonlinear sub-systems (e.g. Carroll and Pecora, 1993). All these observations suggest that a low dimensional deterministic description of the rainfall-runoff dynamics may be possible. Hence, a brief qualitative discussion is provided below as to why and where deterministic components could be expected in the rainfall-runoff system.

Climate produces the input of the rainfall-runoff transformation and also determines in many ways the state of the basin with a sort of "parametric forcing" on vegetation cover, soil saturation, etc. Many studies have suggested the possible presence of low-dimensional deterministic components in the dynamics of the global climate and its external forcing (e.g. Matsumoto *et al.*, 1995). Other works have conjectured that low-dimensional chaotic dynamics can originate from the nonlinear interaction of climate and large water bodies, e.g. oceans, land with its relative soil moisture content and large ice-masses (e.g. Salzman, 1983; Nicolis, 1989; Tsonis and Elsner, 1990, 1997). The interaction between climate and river basins is not uni-directional. Due to the land-atmosphere coupling, the basin exerts a feedback on the climate through its orography, soil moisture content, vegetation cover, etc. Recent studies have reported clues of low-dimensional dynamics in the response of large basins where connections with the local climate are likely to be present (e.g. Sangoyomi *et al.*, 1996). Interaction between soil-moisture and atmospheric dynamics are able to produce chaotic behaviour on different time scales by feedback mechanisms related to the recycling of soil moisture through evaporation (Rodriguez-Iturbe *et al.*, 1991a, b). For basins having hydrological regimes influenced also by snowmelt or by perennial glaciers additional delays and feedbacks with the climate dynamics are present. Even if the actual dynamics of the atmosphere at a meteorological time-scale is more likely of the kind of spatio-temporal chaos of a quite high dimension, the coupling with the basin with its various feedbacks could also give rise to recurrent low-dimensional components.

Once these mechanisms have produced the input (i.e. rainfall) to the system, the final step of the process is the complex synthesis performed by the basin during the rainfall-runoff transformation. The strong low-pass filtering action of the basin, while smoothing out some of the space-time complexity of rainfall, could also indicate the low-dimensional components originating from both climate and rainfall-runoff transformation. Low-dimensional components are introduced by the very action of the basin; in addition to the influence of climate and meteorological dynamics, runoff time series bears the fingerprint of the basin characteristics (topography, geology, channel-network geometry, vegetation, human actions, etc.). For example, groundwater systems behave as a nonlinear dissipative system with a few degrees of freedom (e.g. Brandes *et al.*, 1998), and recession curves usually show quite simple behaviour (Tallaksen, 1995); Chiu and Huang (1970) successfully proposed a low-dimensional (2nd order) nonlinear ODE for the falling limbs of the hydrograph, which reproduced measurements closely.

Preliminary evidence of chaos in rainfall-runoff dynamics

DATA USED

Monthly rainfall and runoff series observed at the Göta basin in Sweden are analysed separately and jointly (using the runoff coefficient) to investigate the possible existence of chaotic behaviour in the rainfall-runoff process. The Göta basin, located in the south of Sweden between 55° and 60°N and 12.9° and 16°E, is about 50132 sq. km in extent, with a lake percentage of 18.6%. The climate in this region varies between boreal and more temperate, without frequently recurring permanent snow cover during winter.

Rainfall and runoff time series observed over a period of 131 years (January 1807–December 1937) are investigated. The runoff represents runoff from natural and unregulated conditions. The runoff coefficient is calculated as the ratio of runoff to rainfall with a concentration time of 6 months. Possible reasons for choosing a concentration time of as long as 6 months are (1) the large size and the flat nature of the basin; (2) the presence of a large percentage of lakes; and (3) the occurrence of snow and ice over a 3–4 month period. (The autocorrelation functions obtained for the rainfall and

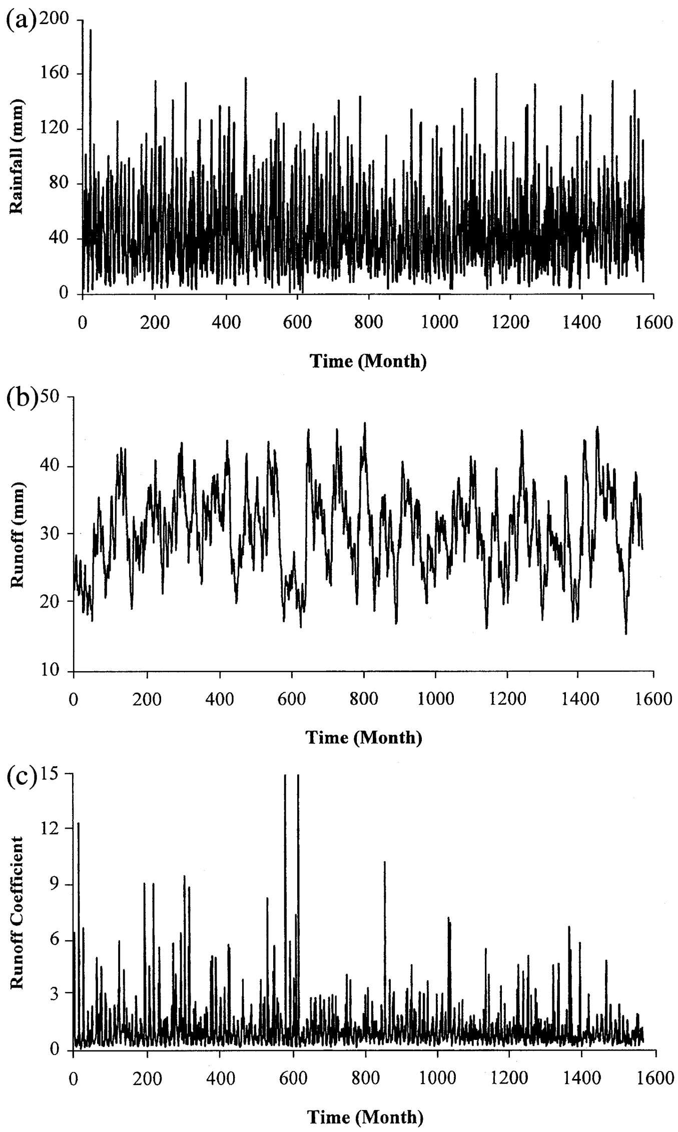

Fig. 1. *Time Series Plot: (a) monthly rainfall series from Göta River; (b) monthly runoff series from Göta River; and (c) monthly runoff coefficient series from Göta River.*

runoff coefficient series also support the selection of a concentration time of 6 months).

Figures 1(a) to 1(c) show the variations in the monthly rainfall, runoff, and runoff coefficient series observed at the Göta basin, and Table 1 presents some of the important statistics for the three series. Though a visual inspection of the three series indicates significant peaks every few years, the seemingly irregular behaviour of the three series does not indicate anything regarding the presence (or absence) of chaotic behaviour and, therefore, additional tools are required. In regard to the runoff coefficient, several values of the runoff coefficient have values greater than 1.0 (Figure 1(c)) and the (long-term) mean is greater than 1.0 (Table 1). This certainly contradicts the acceptable definition of runoff coefficient, which should always be less than 1.0. Though one reason for the above problem could be an inappropriate selection of the concentration time, it is important to note that similar observations are also made when several other concentration times are used. A cross-correlation analysis between rainfall and runoff also indicates a concentration time of about 6 months. On the other hand, the median, which describes the data better than the mean, is less than 1.0 (about 0.74). Therefore, the problem seems to lie inherently with the resolution of the original data considered. With a low-resolution monthly series, with a basin that is large and flat with a large percentage of lakes and also the occurrence of snow and ice over a 3–4 month period, such a problem may occur, irrespective of the concentration time used. Therefore, it may be necessary to find another parameter, connecting rainfall and runoff, to represent better the joint rainfall-runoff process.

AUTOCORRELATION FUNCTION AND PHASE-SPACE DIAGRAM

Before employing any specific chaos identification technique, such as the correlation dimension method or the nonlinear prediction method, useful information regarding the possible presence of chaos in the rainfall, runoff, and runoff coefficient series can be obtained by looking at, for example, the autocorrelation function plot and the phase-space diagram. Figure 2 shows the variation of the autocorrelation function against the lag time for the monthly rainfall, runoff, and runoff coefficient series observed at the Göta basin. The figure indicates that the runoff series shows

Table 1. Statistics of rainfall, runoff and runoff coefficient data observed at the Göta River basin.

Statistic	Rainfall	Runoff	Runoff Coefficient
Number of data	1572	1572	1567
Mean	45.61 mm	30.53 mm	1.12
Standard deviation	28.81 mm	6.15 mm	1.37
Coefficient of variation	0.63	0.20	1.22
Maximum value	192.50 mm	46.30 mm	23.62
Minimum value	1.00 mm	15.40 mm	0.11
Number of zeros	0	0	0

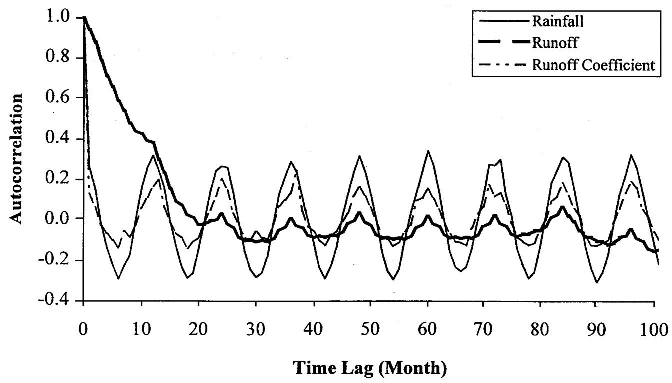

Fig. 2. *Autocorrelation Function for monthly rainfall, runoff, and runoff coefficient series from Göta River.*

some kind of exponential decay in the autocorrelation function up to a lag time of about 20 months, whereas those for the rainfall and runoff coefficient series fall off somewhat abruptly with lag times between 3 and 6. The exponential decay of the autocorrelation function observed for the runoff series may be an indication of chaotic behaviour of the runoff process. With respect to the rainfall and the runoff coefficient series, the delay, though small, in the autocorrelation function seems to indicate that the two series are not stochastic (Sivakumar *et al.*, 2000a).

With regard to the study of the attractor (a geometric object which characterises the long-term behaviour of a system in the phase-space), a useful tool of analysis is the projection of the attractor of the (scalar) time series X_i, where $i = 1, 2, \ldots, N$, reconstructed in a (suitable) embedding dimension (m), using the method of delays, according to

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

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; and τ is a delay time taken to be some suitable multiple of the sampling time Δt (Packard *et al.*, 1980; Takens, 1981). Figures 3(a) to 3(c) show, respectively, for the rainfall, runoff, and runoff coefficient series, the phase-space plots constructed in two-dimensions ($m = 2$) with $\tau = 1$, i.e. the projection of the attractor on the plane $\{X_i, X_{i+1}\}$. The projection yields a well-defined structure for the runoff series, whereas those for the rainfall and runoff coefficient series are, in order, less and less clear, suggesting that the runoff series might yield the lowest dimension, followed by rainfall and runoff coefficient series, respectively. This is verified below using the correlation dimension results.

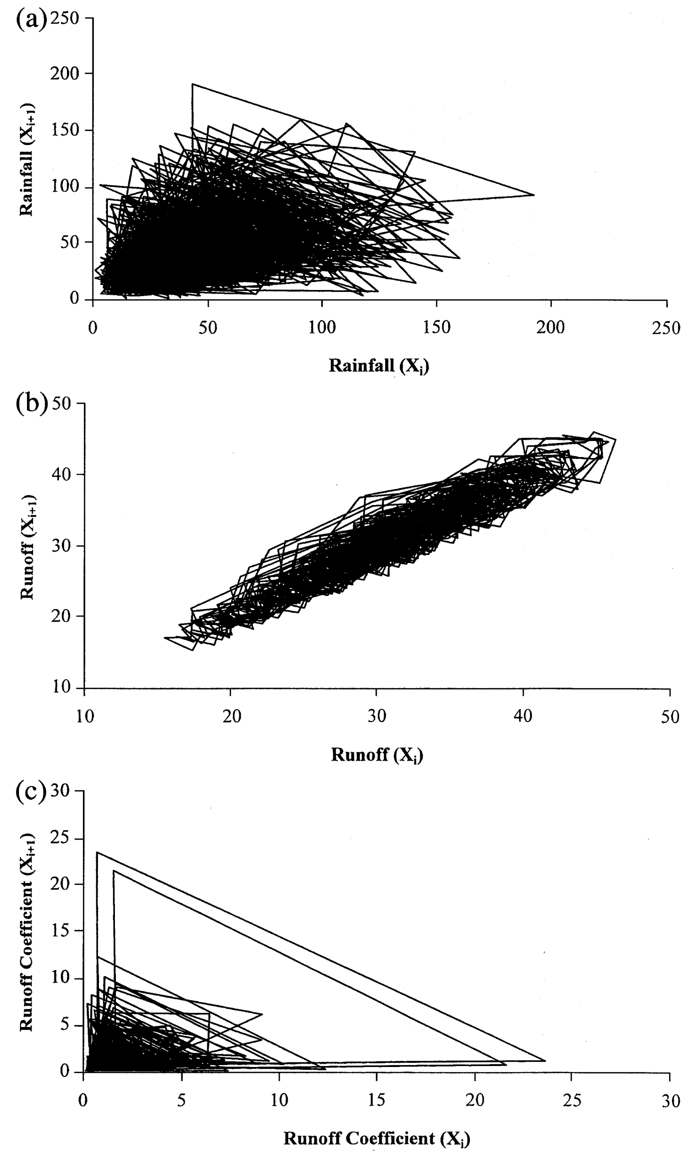

Fig. 3. *Phase-space Plot: (a) monthly rainfall series from Göta River; (b) monthly runoff series from Göta River; and (c) monthly runoff coefficient series from Göta River.*

CORRELATION DIMENSION METHOD

The correlation dimension is estimated by employing the Grassberger-Procaccia algorithm (Grassberger and Procaccia, 1983a, b), which uses the reconstruction of the phase-space according to Eqn. (1). For an m-dimensional phase-space, the correlation function $C(r)$ is given by

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

where H is the Heaviside step function, with $H(u) = 1$ for $u > 0$, and $H(u) = 0$ for $u \le 0$, where $u = r - |\mathbf{Y}_i - \mathbf{Y}_j|$, r is the radius of sphere centred on $\mathbf{Y}_i$ or Y_j, and N is the number of data points. If the time series is characterised by an attractor, then for positive values of r, the correlation function $C(r)$ and radius r are related according to

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

where α is constant; and ν is the correlation exponent or the slope of the log $C(r)$ versus log r plot. If the correlation

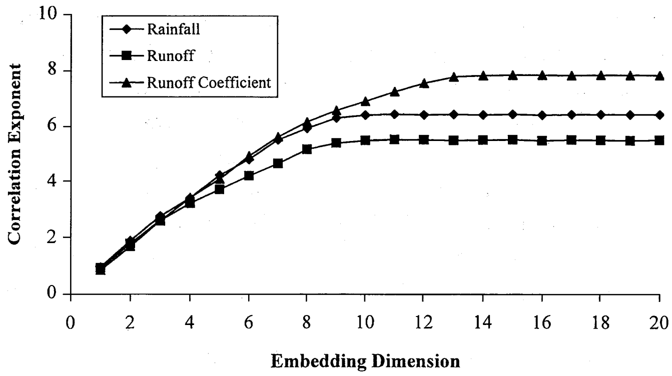

Fig. 4. *Relationship between Correlation Exponent and Embedding Dimension: monthly rainfall, runoff and runoff coefficient series.*

exponent saturates with an increase in the embedding dimension, then the system is generally considered to exhibit chaos (e.g. Fraedrich, 1986). The saturation value of the correlation exponent is defined as the correlation dimension of the attractor. The nearest integer above the saturation value provides the minimum number of phase-space or variables necessary to model the dynamics of the attractor. If the correlation exponent increases without bound with increase in the embedding dimension, then the system under investigation is considered as stochastic.

The correlation functions and the exponents are computed for the monthly rainfall, runoff, and runoff coefficient series. The delay time for the phase-space reconstruction is computed using the autocorrelation function method, and is taken as the lag time at which the autocorrelation function first crosses the zero line (e.g. Holzfuss and Mayer-Kress, 1986). The first zero value of the autocorrelation function is at lag times 3, 20, and 3 respectively for the three series (Fig. 2) and, therefore, these values are used in the phase-space reconstruction (see Sivakumar (2000) for details on the selection of the delay time).

Figure 4 shows the correlation dimension results, i.e. the relationship between the correlation exponent values and the embedding dimension values, obtained for the three series (Sivakumar *et al.*, 2000a). For all the three series, the correlation exponent value increases with the embedding dimension up to a certain value, and then saturates beyond that value. The saturation of the correlation exponent beyond a certain embedding dimension value is an indication of the existence of deterministic dynamics. The saturation values of the correlation exponent for the rainfall, runoff, and runoff coefficient series are 6.4, 5.5, and 7.8 respectively, (Table 2.) The finite correlation dimensions obtained for the three series indicate that they may exhibit chaotic behaviour. The existence of chaos in the rainfall (input) and runoff (output) series suggests that the rainfall-runoff process may also exhibit chaotic behaviour. The existence of chaos in the runoff coefficient (a parameter connecting rainfall and runoff) supports the above.

The correlation dimension estimates obtained above indicate that the minimum number of variables essential to understand and model the dynamics of the rainfall, runoff, and runoff coefficient processes is 7, 6, and 8 respectively. On the basis of the dimension results and, hence, the number of variables essential to model the dynamics of the three processes, the prediction of the runoff process may be easier and better compared to the rainfall and runoff coefficient processes. With the encouraging results obtained using the correlation dimension analysis regarding the existence of chaos in the three series analysed above, an attempt is made to test the possibility of (accurate) predictions. To this effect, the nonlinear prediction method is employed and is discussed below.

Nonlinear prediction method

INTRODUCTION

The role of nonlinear prediction in the study of dynamical systems is two-fold. Firstly, it is linked to the practical possibility of making reliable (short-term) forecasts in complex systems. Within nonlinear deterministic dynamics, nonlinear prediction enables better forecasts than those obtained with statistical methods, e.g. autoregressive methods, autoregressive moving average methods (e.g. Jayawardena and Lai, 1994; Jayawardena and Gurung, 2000), because of its capacity to pinpoint the nonlinear aspect of the phenomenon. Secondly, it is an important investigative tool for the dynamics of a natural phenomenon, in particular, for detecting the presence of deterministic chaos because the possibility to make forecasts on a time

Table 2. Correlation dimension and Nonlinear prediction results: Monthly rainfall, runoff and runoff coefficient data observed at Göta River basin.

Statistic	Rainfall	Runoff	Runoff Coefficient
Correlation dimension	6.4	5.5	7.8
Correlation coefficient	0.785	0.995	0.567
Root mean square error	20.204 mm	0.668 mm	0.682 mm
Coefficient of efficiency	0.447	0.985	0.220
Optimal dimension	3	2	6

series is connected with the type of dynamics which generated the time series. The identification of chaos using the prediction results themselves can, therefore, be termed as an inverse approach.

The application of the nonlinear prediction method in the present study, therefore, is two-fold: (1) to investigate whether the dynamics of the rainfall-runoff process could be predicted reliably using a method based on the concept of chaos theory; and (2) to detect the possible presence of chaos in the rainfall-runoff process using the prediction results themselves. The former assumes that the dynamics of the rainfall-runoff process exhibits chaotic behaviour, and in the latter the validity of such an assumption is verified. One advantage of the nonlinear prediction method to the rainfall, runoff, and runoff coefficient (or any other hydrological) series is that it does not require a large data size and can provide reasonably good results even when the data size is small. Another advantage lies in the fact that the inverse approach allows one not only to detect the presence of chaos but also to verify the results obtained using other methods. For example, the optimal embedding dimension obtained from this method can be compared to the minimum number of dimensions essential to model the dynamics (the nearest integer higher than the correlation dimension), as obtained using the correlation dimension method.

METHODOLOGY

The nonlinear prediction method uses the concept of phase-space for reconstructing the attractor of the time series (using its past history). For a scalar time series X_i, where $i = 1, 2, \ldots, N$, the phase-space can be reconstructed using the method of delays according to Eqn. (1). Once the attractor has been correctly reconstructed in phase-space of dimension m, it is possible to interpret the 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 (In real situations, however, the optimal embedding dimension for reconstruction is not known *a priori* and, therefore, different embedding dimensions have to be used and the optimal dimension should be chosen based on the prediction results).

The problem then is to find an appropriate expression for f_T (e.g. F_T). There are several possible approaches for determining F_T, broadly divided into two categories: (1) global; and (2) local. By means of the first, an attempt is made to approximate the map (Eqn. 4), working globally on all the attractor and seeking a map F_T valid at every point of it. Local approximation (e.g. Farmer and Sidorowich, 1987), on the other hand, entails the subdivision of the f_T domain into many subsets, each of which identifies some approximations F_T, valid only in that same subset. In this way, the dynamics of the system are described step by step locally in the phase-space. This choice leads to a considerable reduction in the complexity of the representation F_T, without lowering the quality of the forecast, to the point that, for the very short term, it generally provides better results than those obtainable by global methods.

In the present study, the local approximation approach is employed for the prediction of the three time series considered. The identification of the sets in which to subdivide the domain can be done in several ways; the usual one entails fixing a metric $\| \ \|$, then, given the starting point Y_j from which the forecast is initiated, identifying neighbours Y_j^p, $p = 1, 2, \ldots, k$, with $j^p < j$, nearest to Y_j, which constitute the set corresponding to the point Y_j. With this, the local functions can then be built, which take each point in the neighbourhood to the next neighbourhood: Y_j^p to Y_{j+1}^p. The local map F_T, which does this, is determined by a least squares fit minimising

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

The local maps are learned in the form of local polynomials (e.g. Abarbanel, 1996), and the predictions are made forward from a new point Z_0 using these local maps. For the new point Z_0, the nearest-neighbour in the learning or training set is found, which is denoted as Y_q. Then the evolution of Z_0 is found, which is denoted as Z_1 and is given by

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

and then the nearest-neighbour to Z_1 is found, and the procedure is repeated to predict other values. In this study, the above methodology is used for local polynomial perdictions, as implemented in the software cspW (Randle Inc., 1996).

The accuracy of prediction can be evaluated using any of the standard statistical measures, such as the correlation coefficient. In this study, the correlation coefficient (CC) between the predicted values and the observed values is used as a main tool to determine the accuracy of prediction. A correlation coefficient of 1 is considered as a perfect prediction whereas a value of 0 refers to no relationship between the predicted and the observed values. The root mean square error ($RMSE$) and the coefficient of efficiency (E^2) are also used to measure the prediction accuracy. The time series plots and the scatter diagrams are also used to choose the best prediction results, among a large combination of results achieved with the different embedding dimensions.

The prediction results themselves can be used to detect the presence of chaos in the time series. This can be done by checking the accuracy of prediction against the embedding dimension. The concept behind the use of embedding dimension is that if the dynamics is chaotic, then the prediction accuracy will increase (to reach its best) with the increase in the embedding dimension up to a certain point, called the optimal embedding dimension (m_{opt}), and remain close to its best for embedding dimensions higher than m_{opt}. On the other hand, for stochastic series, there would be no increase in the prediction accuracy with an increase in the embedding dimension and the accuracy would remain the same for any value of the embedding dimension (e.g. Casdagli, 1989).

RESULTS AND DISCUSSION

In the present study, the first 1440 points from each of the monthly rainfall, runoff and runoff coefficient series are used for the phase-space reconstruction (i.e. training or learning set) to predict the subsequent 80 points. In this study, only one-step ahead (i.e. lead-time = 1) predictions are made. Table 2 presents a summary of the prediction results achieved for the three series. The measures of prediction accuracy are the correlation coefficient (CC), the root mean square error ($RMSE$), and the coefficient of efficiency (E^2). The results presented in Table 2 are those obtained at the optimal embedding dimension for each of the three series, also presented in Table 2. In addition to the above three criteria, the time series plots and the scatter diagrams are also used to select the optimal embedding dimension.

Table 2 shows that, the prediction results are best for the runoff series, followed by the rainfall and the runoff coefficient series respectively. Figures 5(a) to 5(c) compare, using time series plots, the observed and the predicted values for the rainfall, runoff and runoff coefficient series respectively. For the runoff series, the predicted values are in close agreement with the observed values. In the case of the rainfall and runoff coefficient series, though the predicted values are not in good agreement with the observed values, the trends (rises and falls) in the values seem to be fairly well captured. The reasonably good prediction results achieved for the three series using the nonlinear prediction method seem to indicate the suitability of the method (or any other nonlinear dynamical approach) to model and predict the dynamics of the three series and, hence, the rainfall-runoff process.

The prediction results and the correlation dimension results obtained for the rainfall, runoff and runoff coefficient series (Table 2) indicate an inverse relationship between the two. The best prediction results are achieved for the (runoff) series with the lowest correlation dimension (5.5) and vice-versa. Such an inverse relationship between the dimensions and the prediction results for the three series is consistent with the concept of the dimension analysis presented earlier, i.e. a higher dimension is an indication of a more complex process, which is more difficult to model and predict than a less complex process, which is recognised by a lower dimension. These observations suggest the usefulness of the correlation dimension is only as a preliminary indicator to identify the behaviour of the rainfall-runoff (or any other dynamical) system.

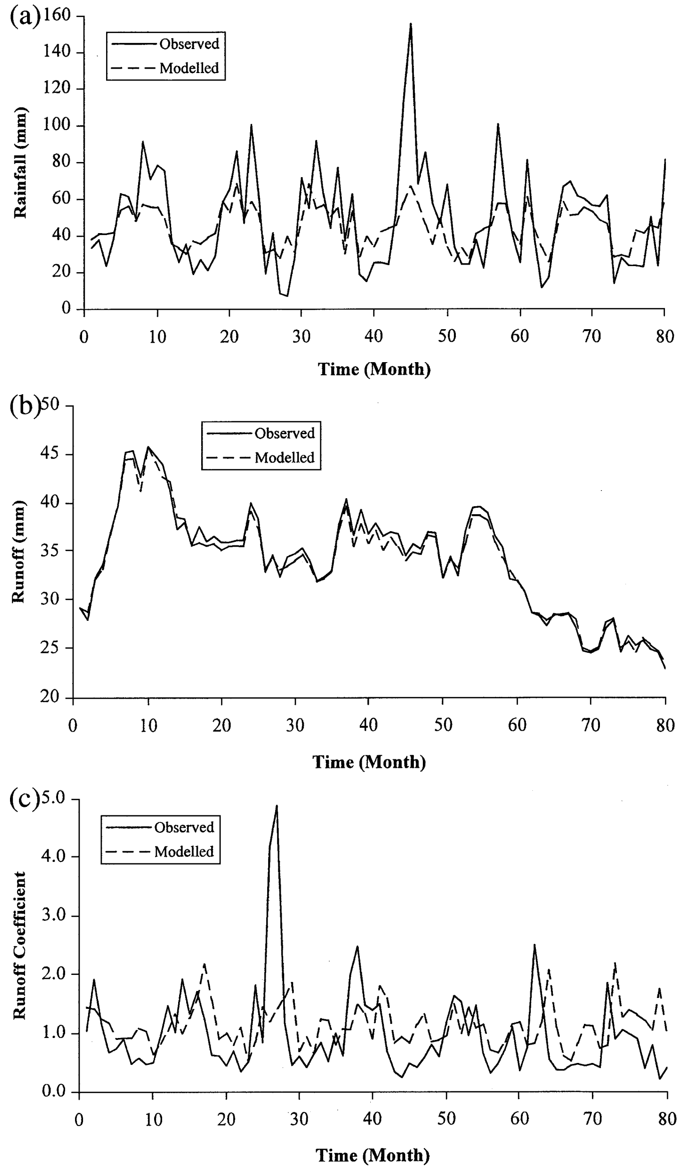

Fig. 5. *Comparison between Time Series Plot of Predicted and Observed Values: (a) monthly rainfall series from Göta River; (b) monthly runoff series from Göta River; and (c) monthly runoff coefficient series from Göta River.*

The results above regarding the existence of chaos can be supported further by employing the inverse chaos identification approach, explained earlier. Figures 6(a) to 6(c) show the variation of the correlation coefficient against the embedding dimension for the rainfall, runoff and runoff coefficient series respectively. In the case of the rainfall and runoff series, the correlation coefficient increases with the embedding dimension up to $m = 2$ and 3 respectively and then decreases when the dimension is increased further, whereas a saturation of the correlation coefficient beyond a certain embedding dimension, $m = 6$, is observed for the runoff coefficient series. The presence of optimal embedding dimension values, $m_{opt} = 2$, 3, and 6 respectively, for

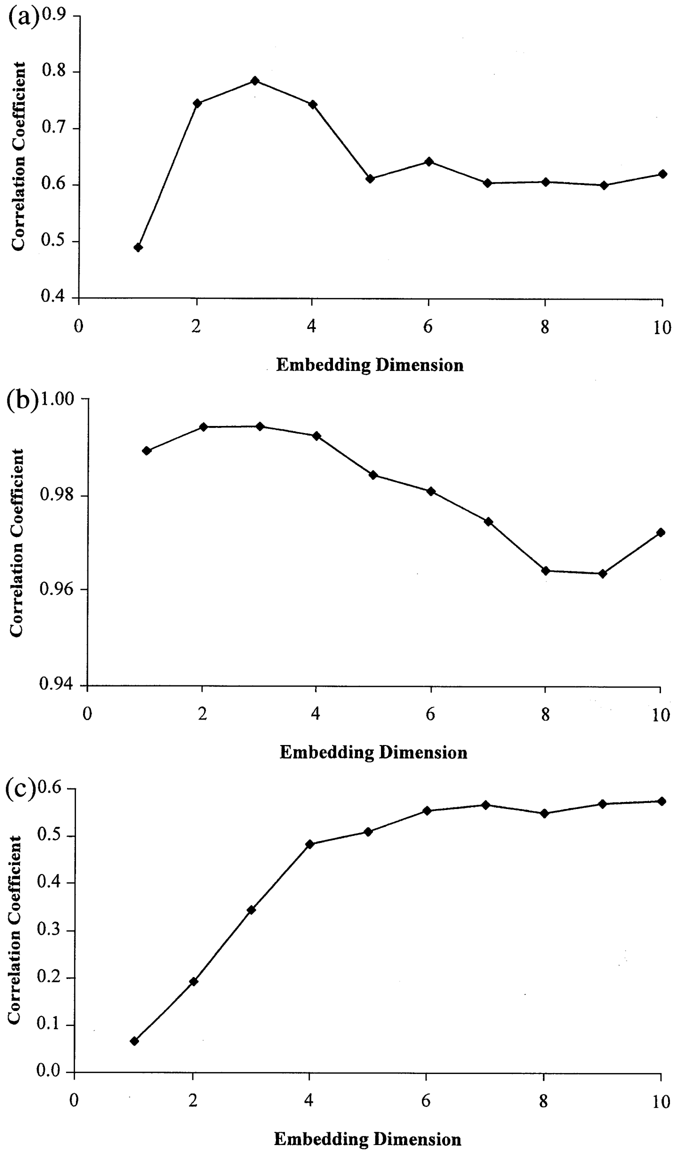

Fig. 6. *Relationship between Correlation Coefficient and Embedding Dimension: (a) monthly rainfall series from Göta River; (b) monthly runoff series from Göta river; and (c) monthly runoff coefficient series from Göta River.*

the rainfall, runoff and runoff coefficient series indicates the possible presence of chaos in the three series. Optimal embedding dimensions are observed also with respect to the RMSE values and the E^2 values (figures not shown), providing further support to the above results regarding the presence of chaos. Furthermore, the optimal embedding dimensions obtained for the three time series with respect to *RMSE* and E^2 (3, 2, and 6 respectively) are consistent with those obtained with respect to the correlation coefficient values.

A brief discussion about the presence of chaos and the prediction results is now in order. According to the concept of chaos theory, for a chaotic time series with an attractor dimension, d, (1) accurate short-term predictions can be achieved when it is embedded in a sufficient phase-space, m_{opt} or higher; and (2) the prediction accuracy will remain constant for any embedding dimension higher than m_{opt}. The prediction results achieved in the present study raise important questions, since: (1) the prediction results are far from accurate, particularly for the rainfall and runoff coefficient series; and (2) the prediction accuracy does not remain constant beyond the optimal embedding dimension, rather decreases when the embedding dimension is increased further.

A possible explanation for such observations is the presence of noise in the time series. Noise is one of the most prominent limiting factors for the predictability of deterministic chaotic systems. Noise limits the accuracy of predictions in three possible ways: (1) the prediction error cannot be smaller than the noise level, since the noise part of the future measurement cannot be predicted; (2) the values on which the predictions are based are themselves noisy, inducing an error proportional to and of the order of the noise level; and (3) in the generic case, where the dynamical evolution has to be estimated from the data, this estimate will be affected by noise (Schreiber and Kantz, 1996). In the presence of these three effects, the prediction error will increase faster than linearly with the noise level.

The inability to obtain accurate prediction results for the rainfall, runoff and runoff coefficient series could be due to the presence of noise in these series, particularly in the first and the last. The rainfall measurement is influenced generally by a large number of factors, such as wind, wetting, evaporation, gauge exposure, instrumentation and human error in reading the rainfall data. The data used in this study are also influenced by a certain imprecision due to the averaging of those observed at two different stations in the basin and the rounding errors resulting from the conversion of the daily to monthly data. Added to this, and most importantly, is the intrinsic much higher erratic nature of the rainfall process, compared to the runoff process. On the other hand, the level of noise in the runoff coefficient series is believed to be much higher due to the presence of noise both in the rainfall and runoff series, as the coefficient is taken as the ratio of runoff to rainfall. The use of (incorrect) concentration times in the computation of the runoff coefficient also increases the noise level, perhaps significantly. The presence of significantly higher levels of noise in the runoff coefficient and the rainfall series, not to forget their higher variabilities, could certainly result in the much less accurate prediction results, when compared to those obtained for the runoff series. However, the prediction results can be improved considerably if the noise is removed or reduced (Porporato and Ridolfi, 1997; Sivakumar *et al.*, 1999b; Jayawardena and Gurung, 2000). Regarding the optimal embedding dimension, it may seem surprising that the accuracy of the prediction decreases at embedding dimensions higher than the optimal embedding dimension, where potentially more information–more data–is summarised in each m-dimensional point. A possible reason for this may be the contamination of nearby points in the high-dimensional embedding with points whose earlier coordinates (at low embedding dimensions) are close but whose recent coordinates (at high embedding dimensions) are distant (e.g. Sugihara and May, 1990).

One more important aspect referring to the presence of noise needs to be addressed. For each of the three time series studied, a discrepancy is observed between the dimension, d, obtained using the correlation dimension method (5.5, 6.4, and 7.8 respectively) and the optimal embedding dimension, m_{opt}, obtained using the nonlinear prediction method (2, 3, and 6 respectively). On one hand, since noise in a time series may result in an overestimation of the correlation dimension, the actual correlation dimensions of the rainfall, runoff and runoff coefficient series are believed to be somewhat lower than the ones obtained above. On the other hand, as explained above, the optimal embedding dimensions obtained in the nonlinear prediction method could also be somewhat lower than the actual ones. Therefore, the removal or reduction of noise could partially offset these problems, yielding much closer dimensions to the actual ones, than those obtained above.

Conclusions and scope for further research

The present study followed the research undertaken earlier by Sivakumar *et al.* (2000a), in which some preliminary evidence of the existence of chaos in the monthly rainfall-runoff process at the Göta basin in Sweden was collected. Techniques, ranging from fundamental statistical ones providing important information on the general behaviour of the process to more sophisticated ones capturing the (nonlinear) intricate details of the process, were employed to rainfall, runoff and runoff coefficient time series. Particular emphasis was given to the (nonlinear) prediction aspects of the rainfall-runoff process. The results achieved in this study augment those obtained previously by Sivakumar *et al.* (2000a). The existence of a chaotic component, which constitutes the framework of the dynamics under investigation, appears increasingly evident. Several kinds of analysis support its existence consistently, both qualitatively and quantitatively. The variations between the three time series, observed from time series, phase-space and autocorrelation plots, were well represented by the corresponding variations in the correlation dimensions and also the prediction results. For instance, the lower variability in the time series plot and a well-defined projection of the time series in the phase-space, were represented by a lower correlation dimension and a higher prediction accuracy. The reasonably good prediction results achieved for the three series (runoff, in particular) using a prediction method based on the concept of chaos theory indicate the suitability of a chaotic approach for understanding, modelling and predicting the underlying dynamics of the processes.

Although the present study yielded convincing evidence regarding the existence of chaotic components in the rainfall-runoff dynamics, their assessment is not simple. Due to the inherent limitations of the identification methods (e.g. assumptions of infinite and noise-free time series) and the problems in hydrological time series (e.g. finite and noisy time series), the results presented here allow only partial conclusions. New and more refined methods applied to time series of better quality are required to confirm and generalise the results. The possible extensions of the present analysis could be: (1) identification of a parameter, connecting rainfall and runoff, which could represent the rainfall-runoff process better than runoff coefficient; (2) use of other chaos identification methods to confirm the results obtained previously regarding the existence of chaos in the rainfall-runoff process; (3) application of nonlinear noise reduction techniques to obtain better quality data (e.g. Sivakumar *et al.*, 1999b); (4) independent analysis of the other variables that influence the rainfall-runoff process, such as temperature (e.g. Jinno *et al.*, 1995); (5) recovery of one variable in the rainfall-runoff system, which may be difficult to measure but is of direct physical interest, from a relatively easily measureable variable in the same dynamical system (e.g. Abarbanel *et al.*, 1994); and (6) multivariate time series analysis of the variables involved in the rainfall-runoff system (e.g. Cao *et al.*, 1998). Investigations along these lines are underway, details of which will be reported elsewhere.

Acknowledgements

This study was funded by the Swedish Research Council for Engineering Sciences and the Swedish Natural Science Research Council. The first author wishes to thank the Nils Hörjel foundation for granting a scholarship for his research stay at the Department of Water Resources Engineering, Lund University, Sweden. The authors thank Demetris Koutsoyiannis, Jiri Stehlik and an anonymous reviewer for their constructive criticisms and useful suggestions, which resulted in a more accurate and complete presentation of the work.

References

Abarbanel, H.D.I., 1996. *Analysis of observed chaotic data.* Springer-Verlag, New York, USA.

Abarbanel, H.D.I., Carroll, T.A., Pecora, L.M., Sidorowich, J.J. and Tsimring, L.S., 1994. Predicting physical variables in time-delay embedding, *Phys. Rev. E*, **49**, 1840–1853.

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

Brandes, D., Duffy, C.J. and Cusumano, J.P., 1998. Stability and damping in a dynamical model of hillslope hydrology, *Water Resour. Res.*, **34**, 3303–3313.

Cao, L., Mees, A. and Judd, K., 1998. Dynamics from multivariate time series, *Physica D*, **121**, 75–88.

Carrol, T.L. and Pecora, L.M., 1993. Cascading synchronized chaotic systems, *Physica D*, **67**, 126–140.

Casdagli, M., 1991. Chaos and deterministic versus stochastic nonlinear modeling, *J. Roy. Stat. Soc. B*, **54**, 303–324.

Chiu, C. and Huang, J.T., 1970. Nonlinear time varying model of rainfall-runoff relation, *Water Resour. Res.*, **6**, 1277–1286.
Cross, M.C. and Hohenberg, P.C., 1993. Pattern formation outside of equilibrium, *Rev. Mod. Phys.*, **65**, 851–1112.
Farmer, J.D. and Sidorowich, J.J., 1987. Predicting chaotic time series, *Phys. Rev. Lett.*, **59**, 845–848.
Fraedrich, K., 1986. Estimating the dimensions of weather and climate attractors, *J. Atmos. Sci.*, **43**, 419–432.
Georgakakos, K.P., Sharifi, M.B. and Sturdevant, P.L., 1995. Analysis of high-resolution rainfall data. In: *New Uncertainty Concepts in Hydrology and Water Resources*, Z.W. Kundzewicz, (Ed.). Cambridge Univ. Press, New York, USA, 114–120.
Grassberger, P. and Procaccia, I., 1983a. Measuring the strangeness of strange attractors, *Physica D*, **9**, 189–208.
Grassberger, P. and Procaccia, I., 1983b. Characterization of strange attractors, *Phys. Rev. Lett.*, **50**, 346–349.
Hense, A., 1987. On the possible existence of a strange attractor for the southern oscillation, *Beitr. Phys. Atmosph.*, **60**, 34–47.
Holzfuss, J. and Mayer-Kress, G., 1986. An approach to error-estimation in the application of dimension algorithms. In: *Dimensions and Entropies in Chaotic Systems*, G. Mayer-Kress, (Ed.). Springer-Verlag, New York, USA, 114–122.
Islam, S., Bras, R.L. and Rodriguez-Iturbe, I., 1993. A possible explanation for low correlation dimension estimates for the atmosphere, *J. Appl. Meteor.*, **32**, 203–208.
Jayawardena, A.W. and Gurung, A.B., 2000. Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach, *J. Hydrol.*, **228**, 242–264.
Jayawardena, A.W. and Lai, F., 1994. Analysis and prediction of chaos in rainfall and stream flow time series, *J. Hydrol.*, **153**, 23–52.
Jinno, K., Xu, S., Berndtsson, R., Kawamura, A. and Matsumoto, M., 1995. Prediction of sunspots using reconstructed chaotic system equations, *J. Geophys. Res.*, **100**, 14773–14781.
Koutsoyiannis, D. and Pachakis, D., 1996. Deterministic chaos versus stochasticity in analysis and modeling of point rainfall series, *J. Geophys. Res.*, **101**, 26441–26451.
Krasovskaia, I., Gottschalk, L. and Kundzewicz, Z.W., 1999. Dimensionality of Scandinavian river flow regimes, *Hydrol. Sci. J.*, **44**, 705–723.
Liu, Q., Islam, S., Rodriguez-Iturbe, I. and Le, Y., 1998. Phase-space analysis of daily streamflow: characterization and prediction, *Adv. Wat. Resour.*, **21**, 463–475.
Matsumoto, M., Berndtsson, R., Jinno, K., Kawamura, A. and Xu, S., 1995. Relationship between surface temperature and sunspot cycle length, *Annales Geophys. Suppl.*, **13**, part II, C467.
Nicolis, C., 1989. Long-term climatic variability and chaotic dynamics, *Tellus A*, **39**, 1–9.
Osborne, A.R. and Provenzale, A., 1989. Finite correlation dimension for stochastic systems with power-law spectra, *Physica D*, **35**, 357–381.
Packard, N.H., Crutchfield, J.P., Farmer, J.D. and Shaw, R.S., 1980. Geometry from a time series, *Phys. Rev. Lett.*, **45**, 712–716.
Porporato, A. and Ridolfi, L., 1996. Clues to the existence of deterministic chaos in river flow, *Int. J. Mod. Phys. B*, **10**, 1821–1862.
Porporato, A. and Ridolfi, L., 1997. Nonlinear analysis of river flow time sequences, *Water Resour. Res.*, **33**, 1353–1367.
Puente, C.E. and Obregon, N., 1996. A deterministic geometric representation of temporal rainfall: Results for a storm in Boston, *Water Resour. Res.*, **32**, 2825–2839.
Randle Inc., 1996. *cspW, Tools for dynamics, Applied nonlinear sciences, LLC*, Randle Inc., Great Falls, VA, USA.
Rodriguez-Iturbe, I., De Power, F.B., Sharifi, M.B. and Georgakakos, K.P., 1989. Chaos in rainfall, *Water Resour. Res.*, **25**, 1667–1675.
Rodriguez-Iturbe, I., Entekhabi, D. and Bras, R.L., 1991a. Nonlinear dynamics of soil moisture at climatic scales, 1, Stochastic analysis, *Water Resour. Res.*, **27**, 1899–1906.
Rodriguez-Iturbe, I., Entekhabi, D., Lee, J.S. and Bras, R.L., 1991b. Nonlinear dynamics of soil moisture at climatic scales, 2, Chaotic analysis, *Water Resour. Res.*, **27**, 1907–1915.
Salzman, B., 1983. Climatic system analysis. In: *Advances in Geophysics*, **25**, Academic Press, New York.
Sangoyomi, T.B., Lall, L. and Abarbanel, H.D.I., 1996. Nonlinear dynamics of the Great Salt Lake: Dimension estimation, *Water Resour. Res.*, **32**, 149–159.
Schreiber, T. and Kantz, H., 1996. Observing and predicting chaotic signals: Is 2% noise too much? In: *Predictability of Complex Dynamical Systems*, Yu.A. Kravtsov, J.B. Kadtke, (Eds.) Springer Series in Synergetics, Springer-Verlag, Berlin, Germany, 43–65.
Sharifi, M.B., Georgakakos, K.P. and Rodriguez-Iturbe, I., 1990. Evidence of deterministic chaos in the pulse of storm rainfall, *J. Atmos. Sci.*, **47**, 888–893.
Sivakumar, B., 2000. Chaos theory in hydrology: important issues and interpretations, *J. Hydrol.*, **227**, 1–20.
Sivakumar, B., Liong, S.Y., Liaw, C.Y. and Phoon, K.K., 1999a. Singapore rainfall behaviour: Chaotic? *J. Hydrol. Engng. ASCE*, **4**, 38–48.
Sivakumar, B., Phoon, K.K., Liong, S.Y. and Liaw, C.Y., 1999b. A systematic approach to noise reduction in hydrological chaotic time series, *J. Hydrol.*, **219**, 103–135.
Sivakumar, B., Berndtsson, R., Olsson, J. and Jinno, K., 2000. Evidence of chaos in rainfall-runoff process, *Hydrol. Sci. J.* **6** (in press).
Sivakumar, B., Sorooshian, S., Gupta, H.V. and Gao, X., 2000b. A chaotic approach to rainfall disaggregation, *Water Resour. Res.* (in press).
Sivakumar, B., Persson, M., Berndtsson, R. and Uvo, C.B., 2000c. Is correlation dimension a reliable indicator of low-dimensional chaos in short hydrological time series? *Water Resour. Res.* (submitted).
Stehlik, J., 1999. Deterministic chaos in runoff series, *J. Hydrol. Hydromech.*, **47**, 271–287.
Sugihara, G. and May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, **344**, 734–741.
Tallaksen, L.M., 1995. A review of baseflow recession analysis, *J. Hydrol.*, **165**, 349–370.
Takens, F., 1980. Detecting strange attractors in turbulence. In: D.A. Rand, L.S. Young *Dynamical Systems and Turbulence*, (Eds.), Lecture Notes in Mathematics, 898, Springer-Verlag, Berlin, Germany, 366–381.
Tsonis, A.A. and Elsner, J.B., 1990. Multiple attractors, fractal basins and long term climatic dynamics, *Beitr. Phys. Atmosph.*, **63**, 171–176.
Tsonis, A.A. and Elsner, J.B., 1997. Global temperature as a regulator of climate predictability, *Physica D*, **108**, 191–196.
Tsonis, A.A., Elsner, J.B. and Georgakakos, K.P., 1993. Estimating the dimension of weather and climate attractors: important issues about the procedure and interpretation, *J. Atmos Sci.*, **50**, 2549–2555.
Wang, Q. and Gan, T.Y., 1998. Biases of correlation dimension estsimates of streamflow data in the Canadian prairies, *Water Resour. Res.*, **34**, 2329–2339.