

# Detection and predictive modeling of chaos in finite hydrological time series

S. Khan<sup>1</sup>, A. R. Ganguly<sup>2</sup>, and S. Saigal<sup>1</sup>

<sup>1</sup>Civil and Environmental Engineering, University of South Florida, Tampa, FL, USA

<sup>2</sup>Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Received: 31 August 2004 – Revised: 22 December 2004 – Accepted: 23 December 2004 – Published:

Part of Special Issue “Nonlinear deterministic dynamics in hydrologic systems: present activities and future challenges”

**Abstract.** The ability to detect the chaotic signal from a finite time series observation of hydrologic systems is addressed in this paper. The presence of random and seasonal components in hydrological time series, like rainfall or runoff, makes the detection process challenging. Tests with simulated data demonstrate the presence of thresholds, in terms of noise to chaotic-signal and seasonality to chaotic-signal ratios, beyond which the set of currently available tools is not able to detect the chaotic component. The investigations also indicate that the decomposition of a simulated time series into the corresponding random, seasonal and chaotic components is possible from finite data. Real streamflow data from the Arkansas and Colorado rivers are used to validate these results. Neither of the raw time series exhibits chaos. While a chaotic component can be extracted from the Arkansas data, such a component is either not present or can not be extracted from the Colorado data. This indicates that real hydrologic data may or may not have a detectable chaotic component. The strengths and limitations of the existing set of tools for the detection and modeling of chaos are also studied.

## 1 Introduction

The presence of nonlinear dynamics and chaos has strong implications for predictive modeling and the analysis of dominant processes in any discipline. The existence of chaotic behavior has been demonstrated in diverse areas ranging from turbulence (e.g. Abarbanel, 1994), weather or climate (e.g. Lorenz, 1969; Fraedrich, 1986, 1987; Essex et al., 1987) and geophysics (e.g. Hense, 1987; Wilcox et al., 1991; Lorenz, 1996; Porporato and Ridolfi, 1996; Sivakumar, 2004), to biology or medicine (e.g. Almog et al., 1990), finance (e.g. Hsieh, 1991; Trippi, 1995; Cornelis, 2000), and electrical circuits (e.g. Yim et al., 2004). The presence of chaos in

hydrology has been suggested by previous researchers (e.g. Rodriguez-Iturbe et al., 1989; Islam et al., 1993; Jayawardena and Lai, 1994; Stehlik, 1999; Sivakumar, 2000, 2003, 2004; Sivakumar et al., 2000, 2001a,b, 2002a,c; Elshorbagy, 2001; Elshorbagy et al., 2001, 2002a,b; Islam and Sivakumar, 2002). The ability to detect and model chaotic behavior from finite hydrologic time series has recently been debated (e.g. Schertzer et al., 2002; Sivakumar et al., 2002b).

Characterization of chaos from real-world observations is known to be a difficult problem in nonlinear dynamics (e.g. Fraedrich, 1986; Theiler et al., 1992; Basu and Foufoula-Georgiou, 2002). The complexity was highlighted in the context of climate models by Wang (2004), who demonstrated that sensitivity to initial conditions may become less apparent when the randomness in internal atmospheric variables begins to dominate. Fundamental questions still remain unanswered in these areas, for example the ability to detect chaos from a finite time series with random and seasonal components, the ability to decompose a time series into these components, and the corresponding implications for predictive modeling. However, addressing these questions is critical for hydrology. This can be gauged from the wealth of hydrologic literature in areas like complexity analysis (e.g. Bras and RodriguesIturbe, 1976; Tessier et al., 1996; Douglas and Barros, 2003), predictability (e.g. Islam and Sivakumar, 2002) and nonlinear predictive modeling (e.g. Kuligowski and Barros, 1998a,b; Kim and Barros, 2001; Ganguly and Bras, 2003).

This paper investigates the ability of nonlinear dynamical tools to detect, characterize, and predict chaos from finite hydrologic observations, using both simulated and real time series. Realistic simulated data is generated by contaminating chaotic signals with random and seasonal components, while real streamflow data are used from the Arkansas and Colorado rivers. The correlation dimension method is used for detecting the possible presence of chaos. Nonlinear predictive models, namely the phase-space reconstruction (PSR) and artificial neural networks (ANN) are employed for time series decomposition and prediction. This study develops

several new insights and interesting results. The presence of thresholds for the detect-ability of chaos is demonstrated, specifically when a chaotic signal is mixed with random or seasonal signals, or a combination thereof. These thresholds can be expressed in terms of the relative dominance of the chaotic component compared to the random or seasonal components. The ability to decompose a time series into the contributions from the individual components (random, seasonal and chaotic) is shown. The corresponding implications for predictive modeling and characterizing the nonlinear dynamics are highlighted. Real streamflow data analysis provides additional insights. First, not all hydrologic time series contain chaotic components which can be detected and modeled. Second, for certain finite hydrologic time series, the presence of chaos can indeed be detected, isolated from random and seasonal components, and utilized for predictive modeling.

The rest of the paper is organized as follows. Section 2 presents the tools and methods employed in this study. The simulated and real hydrological data are discussed in Sect. 3. Sections 4 and 5 present and discuss the results obtained with simulated and real hydrological time series, respectively. The summary and conclusions of this paper are presented in Sect. 6.

## 2 Tools and methods

### 2.1 State of the art and Literature Review: Tools and Concepts

The theoretical concepts underlying the methodologies for the detection and modeling of nonlinear dynamical and chaotic components, as well as their implementation, are available in the literature (e.g. Takens, 1980; Sivakumar et al., 1998; Hilborn, 2000; Islam and Sivakumar, 2002). The present study exploits the correlation dimension method, as well as nonlinear predictive models like PSR and ANN. These tools have been used widely for the identification and modeling of nonlinear dynamics and low-dimensional chaos in short hydrological time series (e.g. Jayawardena and Lai, 1994; Sivakumar et al., 2002c; Islam and Sivakumar, 2002). Contributions include refinement of these tools in terms of either the underlying methodologies or the applications. A brief discussion of the tools and the present modifications is described in this section.

#### 2.1.1 Correlation dimension

An algorithm suggested by Grassberger and Procaccia (1983) is the most commonly used method in hydrology to characterize chaotic attractors. In a time series of a single dynamic variable,  $z_i$ , where  $i=1, 2, \dots, N$ , a vector in an  $m$ -dimensional phase space may be given as  $\mathbf{Z}_j=(z_j, z_{j-\tau}, \dots, z_{j-(m-1)\tau})$ , where  $j=\frac{m\tau}{\Delta t}, \dots, N$ ;  $\Delta t$  is the time interval;  $\tau$  is the delay time; and  $m$  is the embedding

dimension of the state space. The correlation sum  $C(r)$  is expressed as

$$C(r) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^{i-1} \Theta(r - \|\mathbf{Z}_i - \mathbf{Z}_j\|), \quad (1)$$

where  $\Theta$  is the Heaviside step function;  $N$  is the number of points in the time series; and  $r$  is the radius of a sphere with its center at either of the current points,  $z_i$  or  $z_j$ . The relation between correlation sum  $C(r)$  given by Eq. (1) and correlation exponent  $\nu$  is expressed as

$$C(r) = \lim_{r \rightarrow \infty} cr^\nu, \quad (2)$$

where  $c$  is a constant and  $\nu = \lim_{r \rightarrow \infty} \frac{\ln C(r)}{\ln r}$ . Since a real data set consists of a finite number of points, there always exists a minimum distance,  $d_{\min}$ , between the trajectory points. When  $r < d_{\min}$ , the correlation sum  $C(r)$  is zero and no longer scales with  $r$ . Therefore, Eq. (2) with limit  $r \rightarrow \infty$  cannot be used directly to determine the correlation exponent. In an alternative technique, a plot of  $\ln C(r)$  vs.  $\ln r$  graph is obtained. Several methods are available to determine the correlation exponent from this plot. The slope of the scaling region, where the curve can be approximated as a straight line, gives the correlation exponent  $\nu$ . In the present study, the least squares method is used to fit a line using a moving window of size  $d$  to plot the slope vs.  $\ln r$  curve. The adjacent slope values lying close to a horizontal portion of the curve are averaged to determine the correlation exponent. For the system to be chaotic, the correlation exponent should increase up to a certain point and then saturate with an increase in the embedding dimension. To calculate the saturation value of the correlation exponent, the acceptable error is defined as  $\varepsilon = p \cdot (\max_{C_i} - \min_{C_i})$ , where  $p$  is an acceptable percentage, and  $\max_{C_i}$  and  $\min_{C_i}$  are the respective maximum and minimum correlation exponents in a set of  $M$  values. In this study, the values of  $p$  and  $M$  are taken as 1% and 20, respectively. The difference between the adjacent values of the correlation exponent is given as  $\delta = C_j - C_i$ , where  $j=i+1$ . If  $\delta \leq \varepsilon$ , then  $\Delta = \frac{|C_M - C_i|}{M-i}$ . If  $\Delta \leq \varepsilon$ , then  $C_i$  is the saturation value of the correlation exponent. The nearest integer above the saturation value of the correlation exponent gives the correlation dimension of the attractor. It has been proposed (e.g. Sugihara and May, 1990; Sivakumar et al., 2002c) that the correlation dimension may be related to the minimum number of variables required to extract a multidimensional description of the dynamical system. If  $\delta > \varepsilon$  and saturation never occurs, the presence of chaos cannot be confirmed.

Osborne and Provenzale (1989) observed that the finite correlation exponent achieved using the correlation dimension method is not a good indicator of the presence of chaos since linear stochastic processes may also yield a finite correlation exponent. The determination of correlation exponent is greatly influenced by several factors including limited data size, the presence of noise, delay time, and the presence of a large number of zeros in the data set. Hydrological time series is finite, contaminated with noise, and may contain a

large number of zeros. A finite and small data set produces a smaller scaling region. This may not be sufficient to calculate the slope of the  $\ln C(r)$  vs.  $\ln r$  curve and may result in an underestimation of the correlation exponent. A large scaling region may be better delineated when a large data set is used which, in turn, results in a better estimation of the slope. Nerenberg and Essex (1990) suggested that the minimum number of data points required for the correct estimation of the correlation exponent is  $N_{\min}=10^{2+0.4m}$ , where  $m$  is the embedding dimension. The presence of noise may affect the scaling behavior and may tend to make the slope of the  $\ln C(r)$  vs.  $\ln r$  plot larger for small values of  $r$  resulting in an overestimation of the correlation exponent. If the delay time,  $\tau$ , is too small, the phase space may contain very little information and may result in an underestimation of the correlation exponent. If  $\tau$  is too large, the phase space may miss out nearby diverging trajectories resulting in an overestimation of the correlation exponent. The presence of a large number of zeros in the time series produces a phase space with limited information about the underlying dynamics and results in an underestimation of the correlation exponent.

### 2.1.2 Artificial Neural Networks (ANNs)

ANNs have been used widely for modeling nonlinear hydrologic processes such as rainfall-runoff, streamflow, ground-water management, water quality simulation, and precipitation (e.g. ASCE, 2000). A prediction is expressed by a function,  $\hat{\mathbf{f}}_{\text{NN}}$ , that maps the input  $(z_t, z_{t-\tau}, \dots, z_{t-(m-1)\tau})$  to the expected output  $\hat{z}_{t+T}$ , i.e.  $\hat{z}_{t+T} = \hat{\mathbf{f}}_{\text{NN}}(z_t, z_{t-\tau}, \dots, z_{t-(m-1)\tau})$ , where  $\tau$  and  $m$  are the delay time and number of variables required to model the system, respectively. The present study employs the multi-layer perceptron (MLP) as a nonlinear function approximator in the context of time series observations (e.g. Weigend and Gershenfeld, 1994; Zhang et al., 1998). For a  $m$ -dimensional phase space, an appropriate ANN architecture could comprise  $m$  inputs and one output. The number of hidden layers and corresponding hidden nodes would then depend on the complexity of the functional form to be modeled. In practical applications, more than two hidden layers are seldom used, while a single hidden layer is usually deemed adequate. The choice of the optimal number of hidden nodes is still an open research topic, although methods based on information criteria and other approaches have been proposed (e.g. Bishop, 1995). For the purposes of this study, a single hidden layer with  $\text{integer}(m/2)$  hidden nodes is employed. The ANN is trained using the commonly used backpropagation algorithm. This algorithm repeatedly computes an error between the output of the network and the desired output and feeds this error back to the network. The error input is used to adjust the weights until the error is minimized. In this study, two-third of the data set is used for training and the remaining one-third is used for predictions. The accuracy of the prediction is calculated in terms of the goodness of fit statistics, correlation coefficient ( $CC$ ), and error statistics, mean squared error ( $MSE$ ), between the original data and the pre-

dicted data.

As the number of input nodes in the ANN is increased, the prediction skills increase up to a certain point and then become constant. A large number of input nodes reflects the use of additional lagged variables to model the time series. When the number of inputs is too low, the information content in the lagged values is not adequately captured. This results in a higher bias and hence lower prediction skills. When the number is too large, the functional form to be modeled grows complex, leading to a larger error variance and corresponding decay in skills. This bias-variance tradeoff (and related issues) usually results in an optimal number of nodes where the skills attain a maxima. In certain cases, the number of input values representing the number of lagged variables of a time series at which the skills attain a maxima have been equated with the number of variables influencing the dynamical system (e.g. Sivakumar et al., 2002c).

### 2.1.3 Phase-Space Reconstruction (PSR) prediction

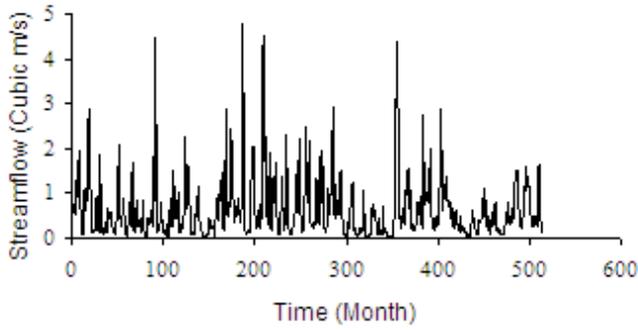
The PSR (e.g. Farmer and Sidorowich, 1987) has been used widely for predicting chaotic time series in hydrology. The time series with a single dynamical variable is embedded in a  $m$ -dimensional state space. The dynamical system can be interpreted in the form of a nonlinear function  $\mathbf{F}_T$ , i.e.  $\mathbf{Z}_{i+T} = \mathbf{F}_T \mathbf{Z}_i$ , where  $\mathbf{Z}_i$  and  $\mathbf{Z}_{i+T}$  are the current state space and the future state space, respectively. In order to predict  $\hat{z}(i+T)$  using the current state  $\mathbf{Z}_i$ , a function  $\hat{\mathbf{F}}_T$  that maps a  $m$ -dimensional state space to a single scalar variable of the system is employed as

$$\hat{z}_{i+T} = \hat{\mathbf{F}}_T \mathbf{Z}_i. \quad (3)$$

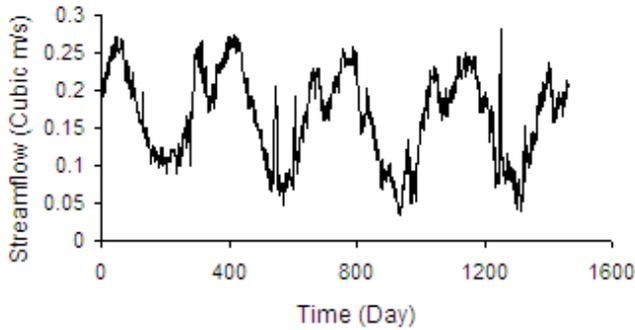
In the present study, the local approximation approach suggested by Farmer and Sidorowich (1987) is used to determine  $\hat{\mathbf{F}}_T$ . This approach reduces the complexity of this function by dividing its domain into many neighborhoods and estimating an approximation map for each neighborhood. As suggested by Porporato and Ridolfi (1996),  $\hat{\mathbf{F}}_T$  can be modeled as a linear function for best results. Two-third of the data set is used for calibrations and the remaining one-third for predictions. The function of Eq. (3) is modeled by using a linear regression model as

$$\begin{bmatrix} z_{11+T} \\ z_{22+T} \\ \dots \\ z_{kk+T} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{11-\tau} & \dots & z_{11-(m-1)\tau} \\ z_{22} & z_{22-\tau} & \dots & z_{22-(m-1)\tau} \\ \dots & \dots & \dots & \dots \\ z_{kk} & z_{kk-\tau} & \dots & z_{kk-(m-1)\tau} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix}, \quad (4)$$

where,  $(z_{11}, \dots, z_{11-(m-1)\tau}), \dots, (z_{kk}, \dots, z_{kk-(m-1)\tau})$  are the  $k$  states nearest to the current state,  $\mathbf{Z}_i$ , and are computed based on the Euclidean distance measure  $\|\mathbf{Z}_i - \mathbf{Z}_j\|$ , where  $\mathbf{Z}_j = (z_{jj}, \dots, z_{jj-(m-1)\tau})$ ;  $b_1, \dots, b_m$  are the function coefficients; and  $z_{11+T}, \dots, z_{kk+T}$  are the  $k$  predicted data points. Using function coefficients,  $b_i$ , from Eq. (4), the predicted value for the current state is given as  $\hat{z}_{i+T} = \sum_{r=1}^m z_{i-(r-1)\tau} b_r$ , where  $\tau$  is the delay time. In the present study, the values employed for  $k$  and  $\tau$  are 25 and 1, respectively. The prediction accuracy is represented in terms



**Fig. 1.** Monthly streamflow time series observed at the Arkansas river.



**Fig. 2.** Daily streamflow series observed at the Colorado river.

of  $CC$  and  $MSE$  between the original data and the predicted data. It has been suggested (e.g. Sivakumar et al., 2002c) that the embedding dimension at which the  $CC$  vs.  $m$  curve saturates may indicate the minimum number of variables required to capture the dynamics of the system.

### 3 Data description

The effectiveness of the methods described in Sect. 2.1 in terms of their ability to detect the presence of low-dimensional chaos, as well as for short-term predictions, is investigated. This is accomplished through an analysis of both simulated and real hydrologic data.

#### 3.1 Simulated data

Hydrological time series are always finite and may be contaminated with noise and seasonality. To obtain realistic insights, simulated data is generated by contaminating chaotic time series with noise, i.e. white and autoregressive, and seasonality. The chaotic time series is represented here by the Lorenz system of equations as

$$\frac{dx}{dt} = \beta(y - x), \quad \frac{dy}{dt} = -xz + rx - y, \quad \frac{dz}{dt} = xy - bz, \quad (5)$$

where  $\beta=10$ ;  $r=28$ ; and  $b=8/3$ . The Lorenz system is highly sensitive to the initial conditions and is, therefore, chaotic. Seasonality implies periodicity which, in turn, could

**Table 1.** Streamflow data statistics (values in  $m^3/s$ ).

Parameter	Arkansas river	Colorado river
Data points	516	1461
Number of zeros	0	0
Mean	0.6529	0.1672
Standard deviation	0.684	0.0594
Variance	0.4678	0.00353
Maximum value	4.7571	0.2819
Minimum value	0.0187	0.0334

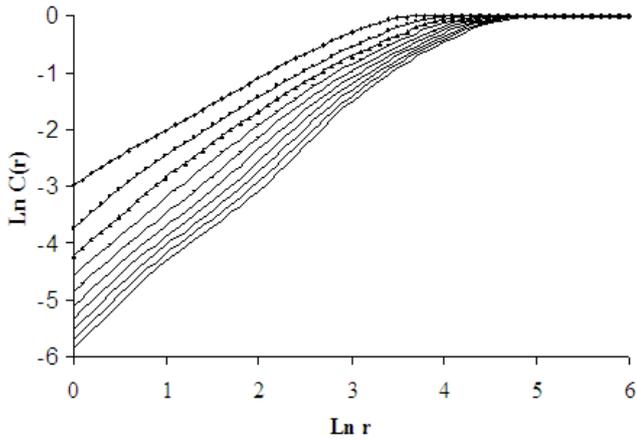
broadly include intra-annual and inter-annual cycles. For this study, seasonality is represented by a periodic function,  $x(t)=A \cos(2\pi ft)$ , where  $A$  and  $f$  are the amplitude and frequency, respectively. Noise consists of random variations in the data. White noise is generated by using normal distribution with  $\mu(\text{mean})=0$  and a user specified value of  $\sigma$ (standard deviation).

#### 3.2 Hydrologic time series

Real hydrologic series including monthly streamflow time series of the Arkansas river at Little Rock and daily streamflow time series of the Colorado river below Parker dam are investigated. The raw data is obtained from the U.S. Geological Survey site.

The presence (or absence) of low-dimensional chaos in a short hydrological (monthly streamflow) time series at the Arkansas river at Little Rock in Arkansas is first studied. The Arkansas River is the fourth longest river in the United States and is a tributary of the Mississippi which flows east and southeast through Colorado, Kansas, Oklahoma and the state of Arkansas. It is located at Latitude  $34^{\circ}45'00''$  and Longitude  $92^{\circ}16'25''$ . The associated drainage area is 158 090 square miles while the contributing drainage area is 135 849 square miles. The temperature at Little Rock ranges from a mean low of  $40^{\circ}\text{F}$  in January to a mean high of  $81^{\circ}\text{F}$  in July. The mean annual precipitation at Little Rock is 50.26 inches. The streamflow data of 43 years (October 1927–September 1970) is analyzed. The monthly streamflow time series at the Arkansas river is shown in Fig. 1.

The daily streamflow data observed at the Colorado river below Parker dam, Arizona-California is also investigated for the existence of chaotic behavior. The Colorado River flows through Colorado, New Mexico, Utah, California, Arizona and Nevada. It drains a part of the arid regions on the western slopes of the Rocky Mountains. The height, crest length and base thickness of Parker dam are 320 feet, 856 feet, and 100 feet, respectively. It is located at Latitude  $34^{\circ}17'44''$ , Longitude  $114^{\circ}08'22''$ . The associated drainage area is 182 700 square miles while the contributing drainage area is 178 700 square miles. The average temperature varies from  $43.1^{\circ}\text{F}$  in winter to  $107.5^{\circ}\text{F}$  in summer. The annual average precipitation is 6.2 inches. The daily streamflow data observed for 4



**Fig. 3.**  $\ln C(r)$  vs.  $\ln r$  plot for Lorenz ( $X$ -component) time series. The curves are shown from top to bottom in ascending order of embedding dimension,  $m=2, 4, \dots, 20$ .

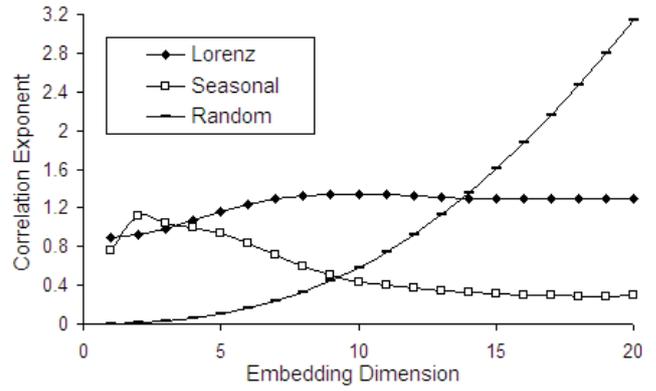
years (1 June 1959–31 May 1963) is analyzed. The variation of the daily streamflow series at the Colorado river is shown in Fig. 2.

Statistics of the streamflow data observed at the Arkansas and Colorado rivers are given in Table 1.

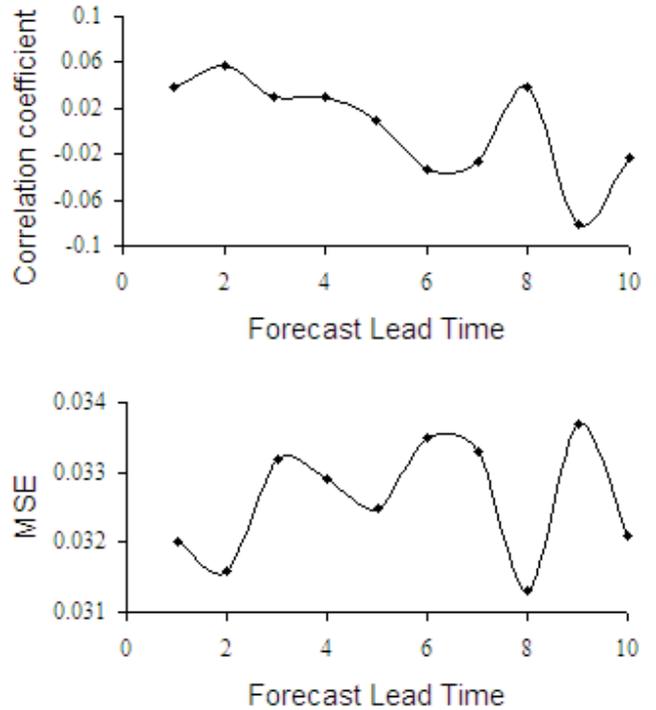
#### 4 Results with simulated data

##### 4.1 Pure chaotic, random and seasonal time series

In this study, a chaotic time series is represented by the  $X$  component of Eq. (5); a random series is a normally-distributed white noise with  $\mu=0$  and  $\sigma=0.16$ ; and seasonal series is represented by a periodic function, i.e. cosine function with a frequency of 10 Hz and an amplitude 10. Each series has 1000 data points, which is greater than the number of points required ( $10^{2+0.4*m}=650$ , where  $m=2$  for the Lorenz time series) for the correct estimation of the correlation exponent of the Lorenz time series (e.g. Nerenberg and Essex, 1990). The plot of  $\ln C(r)$  vs.  $\ln r$  described in Sect. 2.1.1 for Lorenz time series is shown in Fig. 3, while the corresponding correlation exponent vs.  $m$  plot is shown in Fig. 4. For chaotic series, the correlation exponent saturates at an embedding dimension of 10. The saturation value of the correlation exponent is 1.34, while the correlation dimension of the Lorenz attractor is 2. For random series, no saturation is observed as seen in Fig. 4. The correlation exponent shown in Fig. 4 decreases rather markedly for the seasonal series considered here, while the  $\ln C(r)$  vs.  $\ln r$  plot shows characteristic steps (e.g. Hilborn, 2000). For purposes of generality, uniform and objective methods are utilized to identify the scaling regions and to calculate the slope through the moving window technique described above. A window size of 9 is chosen and the average of ten slope values from  $\ln r=0.8$  to  $\ln r=1.7$  in increments of 0.1 are taken. These parameters are selected through trial and error. Once selected, they are kept constant for all the case studies with simulated data.

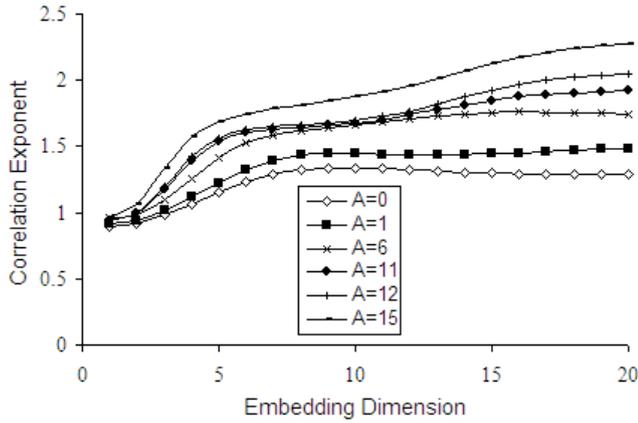


**Fig. 4.** Relation between correlation exponent and embedding dimension for Lorenz ( $X$  component), seasonal, and white noise series.



**Fig. 5.** The variation of  $CC$  and  $MSE$  with forecast lead time for white noise with  $\sigma=0.16$ .

While this ensures uniformity, there is a possibility that judgmental approaches on each series might have yielded differing values for the scaling region and smoothing windows. However, one purpose of the paper is to motivate the development of tools that can be utilized in an automated fashion and that do not rely on user judgment or subjective considerations. The results with predictive modeling approaches like PSR and ANN indicate significant prediction skills, i.e.  $CC=1$  and  $MSE=0$ , for forecast lead time from one to ten at  $m=10$  for the chaotic and the seasonal series. For random series, the predictive skills fluctuate around a lower value and show only a slightly decreasing trend as shown in Fig. 5.



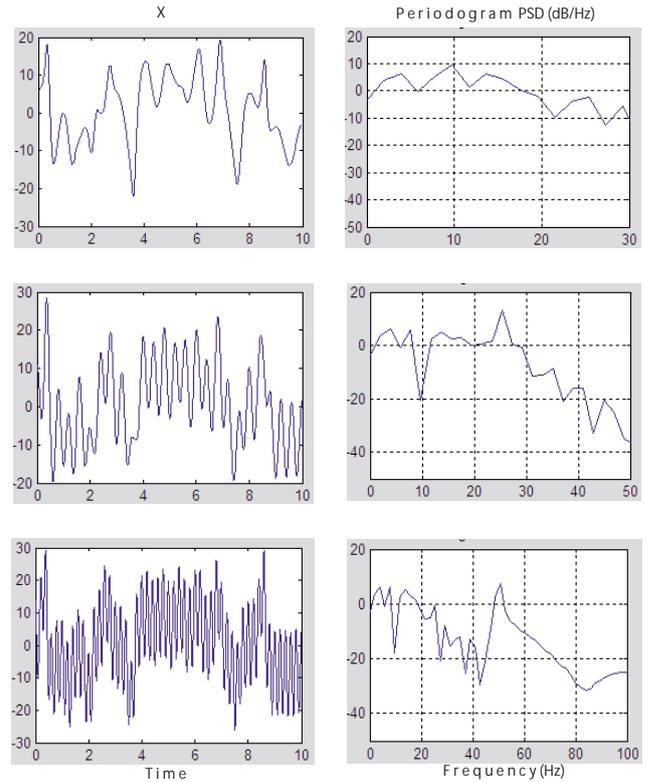
**Fig. 6.** Relation between correlation exponent and embedding dimension for mixed time series. Series includes Lorenz X-component and seasonality with  $f=10$  Hz and different amplitudes.

## 4.2 Mixed time series

As described earlier and indicated by previous researchers (e.g. Sugihara and May, 1990; Koutsoyiannis and Pachakis, 1996; Elshorbagy, 2001), hydrologic and other real systems tend to generate observables that have random and seasonal components, in addition to any nonlinear deterministic (or chaotic) signal that may be present. Simulated data are generated using a mixture of these components to understand the ability to identify, characterize and quantify chaos from amidst seasonality and noise through the commonly used tools for nonlinear dynamics and chaos.

### 4.2.1 Mixture of chaotic and seasonal series

The ability to distinguish chaotic signals, when mixed with seasonal signals of varying amplitudes, is depicted by correlation exponent vs.  $m$  plots shown in Fig. 6. These plots are obtained using the correlation dimension method. First, it is seen that as the seasonal component is increased, i.e. the amplitudes are made higher, the saturation value of the correlation exponent also increases. Second, there appears to be a threshold value, expressed as the seasonality to chaotic-signal ratio, for the degree of seasonality beyond which the chaotic component cannot be distinguished any longer. The degree of seasonality is expressed in terms of the amplitude of the seasonal component for fixed value of the chaotic Lorenz series. From eye estimation, for amplitudes of 1 and 6 in Fig. 6, the correlation exponent appears to saturate with an increase in the embedding dimension. Based on objective techniques established in Sect. 2.1.1, it can be demonstrated that the saturation indeed occurs and the correlation exponent either decreases, or remains constant, or increases within tolerable limits, till an amplitude of eleven. However, for amplitudes larger than 11, the increase in correlation exponent no longer remains within the tolerable limits, thus indicating the presence of a possible threshold. While the presence of a threshold is indicated, the exact value can de-



**Fig. 7.** Mixed times series (Lorenz X-component and seasonality) and its periodograms showing the variation of power spectral density (PSD) with frequency. Top: Lorenz X-component and seasonality with  $f=10$  Hz and  $A=5$ . Middle: Lorenz X-component and seasonality with  $f=25$  Hz and  $A=10$ . Bottom: Lorenz X-component and seasonality with  $f=50$  Hz and  $A=13$ .

pend to some extent on how the criteria for saturation and the corresponding tolerances are defined. It may also depend on the number of data points used for the analysis. It is further observed that the amplitude of 11 corresponds to the highest integer value of the seasonal amplitude for which  $\sigma$  of the seasonal component is lower than that of the chaotic series. The  $\sigma$  values for the pure chaotic series, pure seasonal series with amplitude 11, and pure seasonal series with amplitude 12, are 8.31, 7.78, and 8.48, respectively. For an amplitude of 11, the seasonality to chaotic-signal ratio, i.e.  $\frac{\sigma_s}{\sigma_c}$ , is 0.94. Thus, the results: (a) appear to indicate the presence of a threshold value, i.e. 0.94, for the degree of seasonality beyond which chaotic components may no longer be distinguishable at least through the use of the correlation exponent based approaches, and (b) suggest that the threshold may occur at the point where the seasonality to chaotic-signal ratio roughly equals unity.

The separation of the nonlinear deterministic signal from the seasonality is next considered. Frequency domain methods are applied to identify the dominant frequencies of the seasonal component. This implicitly assumes that the seasonal component, if present, in a simulated or real hydrologic series, is dominant enough to be identifiable through the periodogram estimate. Figure 7 shows chaotic-signal

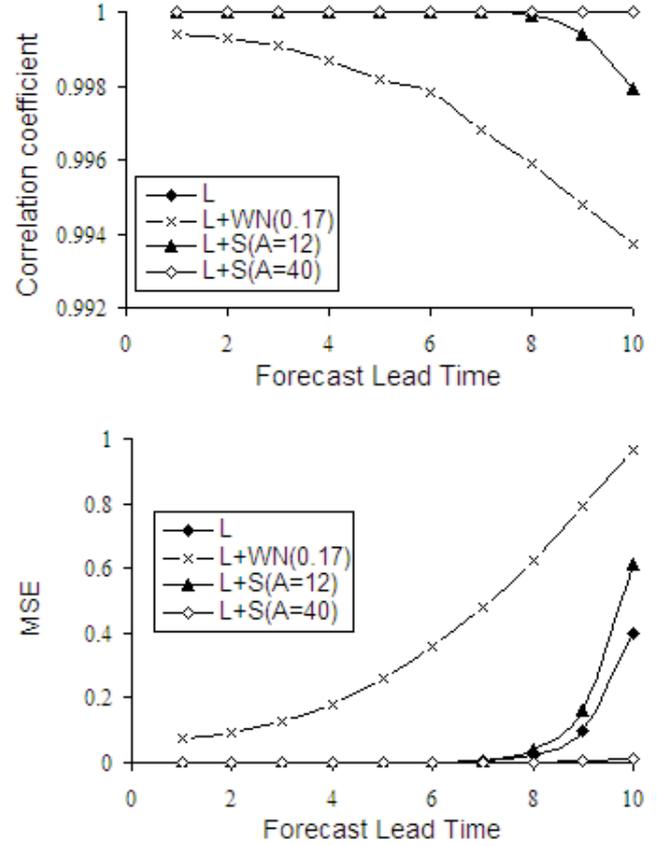
**Table 2.** Separation of white noise from a mixture of chaotic, seasonal and white noise series using the PSR with  $m = 10$ .

Org. $\sigma_{wn}$	Cal. $\sigma_{wn}$	% error	MSE	CC
0.075	0.0910	21.3	0.0083	0.9999
0.1	0.1254	25.4	0.0157	0.9999
0.1	0.2318	15.9	0.0537	0.9997
0.33	0.3608	9.33	0.1301	0.9992
1.0	1.1985	19.85	1.4349	0.9913
2.0	2.6403	32.01	6.9706	0.9587
3.0	4.1341	37.8	17.0917	0.9022
5.0	6.5895	31.8	43.4134	0.7806
10.0	12.624	26.24	159.3665	0.4948

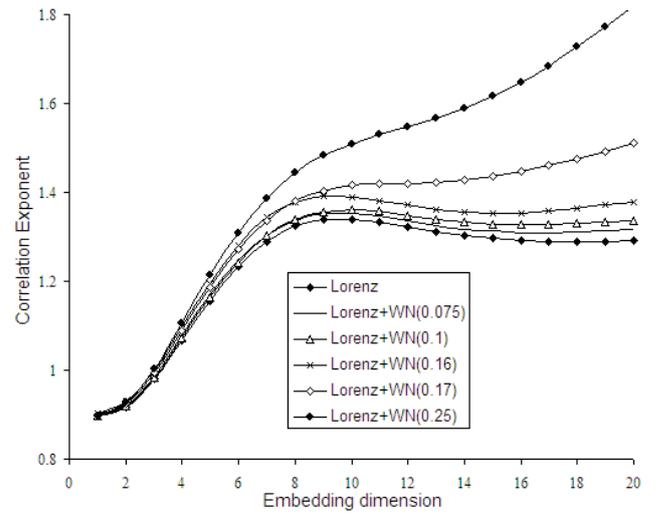
mixed with seasonal signals of different frequencies and amplitudes, and the corresponding periodogram estimates. As shown in Fig. 7, the dominant frequency is observed for each simulation. The amplitudes selected here represent the minimum values of the amplitudes, corresponding to the given frequencies, above which a peak can be distinguished in the periodogram. Thus, seasonal components of higher frequencies need higher amplitudes to make them identifiable. Once the frequencies are obtained, curve fitting techniques can be utilized to obtain the amplitudes of the seasonal component. The remaining component is the non-seasonal portion, which in this case would be the chaotic component. It is noted again that when extending this observation to real data, the non-seasonal component needs to be further decomposed into random and deterministic components. The deterministic component is next analyzed for chaotic signals. The separation of random and deterministic (possibly chaotic) signals is discussed in Sect. 4.2.2. Table 2 provides an example where the white noise component has been isolated from a series which is a mixture of chaotic signal, seasonality, and white noise. The original and recovered standard deviations of the white noise are shown along with the corresponding errors. Once the deterministic component, i.e. chaotic + seasonality, is isolated from the mixture, the seasonal component is separated by fitting a periodic curve with the dominant frequency found using frequency domain analysis. The remaining chaotic component is compared with the original Lorenz series used for the simulation. It is seen that as the white noise component in the mixture increases, the prediction skills represented by statistics  $CC$  and  $MSE$  decrease. The PSR, with the embedding dimension of 10, is used for multi-step ahead forecasting of a mixture of chaotic series and seasonality ( $f=10, A=12$ ; and  $f=10, A=40$ ). Figure 8 shows the  $CC$  vs. Forecast lead time and  $MSE$  vs. Forecast lead time plots. Good prediction skills with  $CC=1$  and  $MSE=0$  for a mixture of chaotic and seasonal series are observed.

#### 4.2.2 Mixture of chaotic and random series

The ability to separate the deterministic component from a mixture of chaotic and random series is studied. Correla-



**Fig. 8.** Variation of  $CC$  and  $MSE$  with forecast lead time for chaotic and mixed series. L, WN and S stand for Lorenz, white noise and seasonality, respectively.



**Fig. 9.** Correlation exponent vs. embedding dimension plot for mixed time series consisting of Lorenz ( $X$ -component) and white noise.

tion exponent vs.  $m$  plot obtained using the correlation dimension method is shown in Fig. 9. Each curve represents a different value of  $\sigma$  for the white noise component that

**Table 3.** Separation of white noise from a mixture of Lorenz (X component) and white noise using the PSR with  $m=10$ .

$\sigma_{wn}$	$\sigma_{wn}$ (cal)	% error	$\mu$ (norm fit)	$\mu_{ci}$ (norm fit)	$\sigma_{ci}$ (norm fit)	$MSE$	$CC$
0.075	0.0861	14.8	-0.0012	-0.0066-0.0041	0.0825-0.0901	0.0074	0.9999
0.1	0.1066	6.6	-0.0022	-0.0088-0.0045	0.1021-0.1115	0.0114	0.9999
0.2	0.2008	0.4	0.0	-0.0135-0.0115	0.1924-0.2101	0.0403	0.9997
0.33	0.3502	6.12	-0.0011	-0.0229-0.0208	0.3355-0.3664	0.1226	0.9991

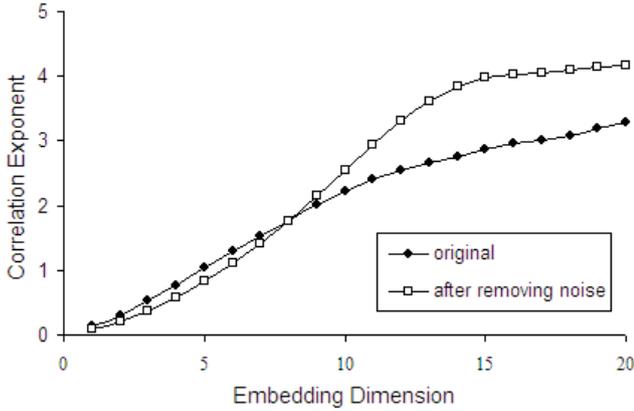
**Table 4.** Separation of white noise from a mixture of Lorenz (X component) and white noise using the ANN with  $m=10$ .

$\sigma_{wn}$	$\sigma_{wn}$ (cal)	% error	$\mu$ (norm fit)	$\mu_{ci}$ (norm fit)	$\sigma_{ci}$ (norm fit)	$MSE$	$CC$
0.075	0.0871	16.1	0.0	-0.0054-0.0054	0.0834-0.0911	0.0076	0.9999
0.1	0.1129	12.9	0.0	-0.007-0.007	0.1081-0.1181	0.0138	0.9999
0.2	0.2238	11.9	0.0	-0.014-0.014	0.2144-0.2341	0.05	0.9996
0.33	0.3541	7.3	0.0	-0.0221-0.0221	0.3392-0.3704	0.1391	0.9991

is mixed with the same underlying chaotic signal. Several properties are observed. First, the saturation value of the correlation exponent increases as the white noise component ( $\sigma$ ) increases. Second, a threshold value appears to exist for the noise to chaotic-signal ratio, beyond which the latter cannot be distinguished through the use of this method. The correlation exponent does not saturate when the  $\sigma$  of the white noise component becomes more than 0.17, which roughly corresponds to a threshold value of 1/49 for the noise to chaotic-signal ratio. Similar to the case of mixture of chaos and seasonality described in the previous section, the results presented here appear to suggest the existence of a threshold. The actual value of the threshold may vary depending on the predefined saturation criteria, tolerances and possibly on the number of data points. However, the results do indicate that a threshold value may exist, beyond which the chaotic component can not be determined from a mixture of chaotic-signal and white noise. The threshold value appears surprisingly low (just two percent in terms of the ratio of the standard deviations) for a chaotic-signal contaminated with pure random noise. Besides the white noise series, a mixture of chaotic series and colored noise generated from autoregressive process of order one is studied (results not shown). The same threshold value of 1/49 beyond which the chaotic component can not be determined from a mixture of chaotic-signal and colored noise is observed.

The separation of the nonlinear deterministic signal from the random component is now considered. The separation is obtained through the use of predictive modeling strategies that have demonstrated value for chaotic systems. Several assumptions are made for this purpose. First, it is assumed that the functional form encapsulated by the trained nonlinear predictive models can be utilized to model the deterministic component of the time series. Thus, a decomposition of the series into a deterministic and a random component is realized. Second, it is assumed that the

chaotic-signal is contained within the deterministic component and adequately modeled by the predictive models. Thus, if a functional form  $\hat{\mathbf{f}}_{\text{NN}}$  modeled by the ANN exists, where  $\hat{z}_{t+T} = \hat{\mathbf{f}}_{\text{NN}}(z_t, z_{t-\tau}, \dots, z_{t-(m-1)\tau}) + \epsilon$ , the ANN can be trained using the entire data set to find  $\hat{\mathbf{f}}_{\text{NN}}$ . Once trained, the estimate of  $\hat{z}_t$  provided by the ANN is assumed to represent the deterministic component, and the residual is assumed to represent noise. This assumption holds as long as the functional form modeled by the predictive model is valid. The second assumption implies that the chaotic-signal, if any, is contained within the deterministic signal isolated using the predictive modeling-based decomposition strategy. It is noted, however, that this approach does not distinguish between chaotic and non-chaotic determinism. Once the deterministic component has been isolated, the presence of chaos can be determined using the correlation dimension method. For the simulated data, a mixture of chaotic-signal and white noise is considered. Thus, the deterministic signal that can be isolated is assumed to be chaotic, provided the signal has been adequately modeled. The separated components can also be compared with the original signals from which the simulations are generated. Table 3 shows the comparison when the PSR is used as the predictive model, while Table 4 shows the same for the ANN. The recovered noise is further tested for normality. A perfect recovery would be indicated by a passing of this test with a mean that is statistically indistinguishable from zero and a standard deviation indistinguishable from that of the original white noise. Once the deterministic component has been isolated, it is compared with the original chaotic series used for the simulation. The prediction skills are related to  $MSE$  and  $CC$  values obtained from the predictive models. The results presented in Tables 3 and 4 indicate that for the simulated data, a fairly good degree of separation can be achieved. An embedding dimension of 10 is used at which the saturation value of the correlation exponent of the Lorenz time series is obtained. When these



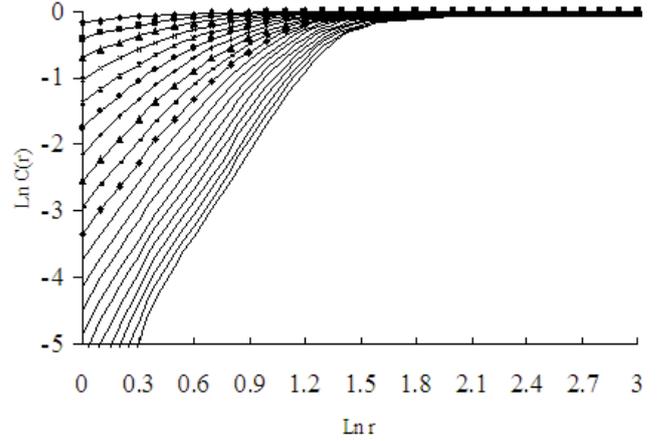
**Fig. 10.** Correlation exponent vs. embedding dimension for monthly streamflow series at the Arkansas river.

concepts are generalized to the real data, the nature of the deterministic component is not known a priori and needs to be verified using tools like correlation exponent methodologies. A mixture of chaotic series and white noise ( $\sigma=0.17$ ) is analyzed for multi-step ahead forecasting using the PSR with the embedding dimension of 10. Figure 8 shows the  $CC$  vs. Forecast lead time and  $MSE$  vs. Forecast lead time plots. It is observed that  $CC$  decreases and  $MSE$  increases as the forecast lead time increases. Due to randomness, the accuracy of the prediction for a chaotic series mixed with white noise decreases with an increase in the forecast lead time.

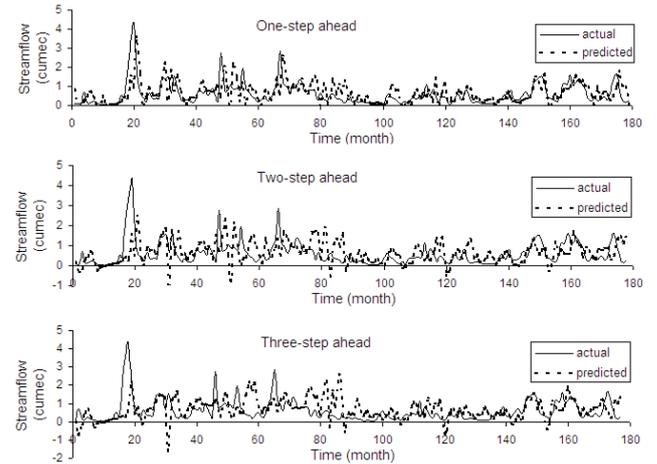
## 5 Analysis with hydrologic time series

### 5.1 Arkansas river

The monthly streamflow time series at the Arkansas river is shown in Fig. 1 and is analyzed using the correlation dimension method to detect the presence (or absence) of low-dimensional chaos in the system. The methodology used for the analysis is described in Sect. 2.1.1. A moving window of size 3 units is used to compute the slope. The number of  $\text{Ln } r$  values for which the slope is available and constant is limited for this data. The slope at a middle range (around  $\text{Ln } r=0.8$ ) is used as the correlation exponent. Figure 10 shows the correlation exponent vs.  $m$  plot for the original monthly streamflow series which does not saturate based either on an eye estimation or the objective techniques described in Sect. 2.1.1. This indicates that either the system is stochastic or the chaotic component, if present, is dominated by noise or seasonality. The time series is analyzed using the PSR with  $m=10$  to isolate the deterministic component from the original series. The methodology is similar to that used for simulated data, and is described in detail in Sect. 4.2. The mean and standard deviation of the non-deterministic component, i.e. noise, separated using the PSR are  $-0.0097$  and  $0.7159$ , respectively. The mean and standard deviation of the original series and the series after the removal of noise, i.e. deterministic series, are  $\mu=0.6529$ ;  $\sigma=0.6833$

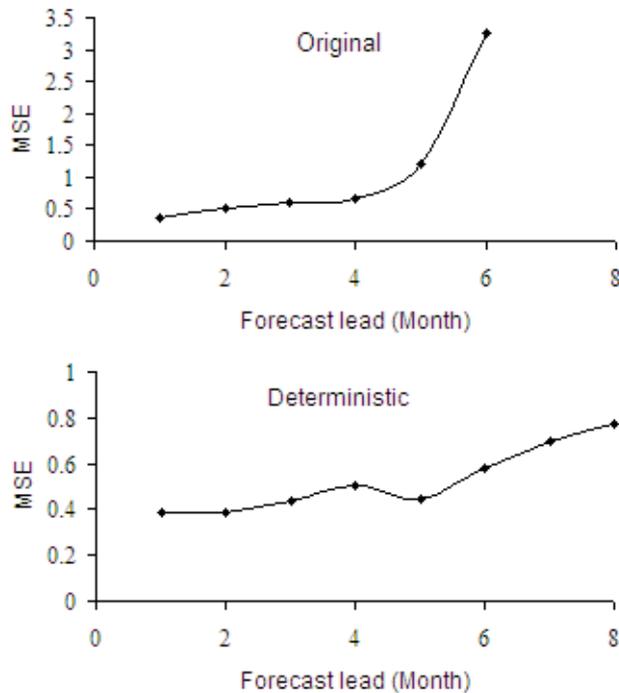


**Fig. 11.**  $\text{Ln } C(r)$  vs.  $\text{Ln } r$  plot for the series, after removing noise, observed at the Arkansas river. The curves are shown in ascending order of embedding dimension,  $m=1, 2, \dots, 20$  from top to bottom.



**Fig. 12.** Multistep ahead predictions for the Arkansas river streamflow. Top: one-step ahead predictions. Middle: two-step ahead predictions. Bottom: three-step ahead predictions.

and  $\mu=0.5908$ ;  $\sigma=0.6564$ , respectively. In the present example, the ratio of  $\sigma_n$  ( $0.7159$ ) and  $\sigma_c$  ( $0.6529$ ) is  $1.09$  which is greater than the threshold value (i.e.  $0.02$ ) obtained for the simulated data in Sect. 4.2.2. An attempt to fit a periodic function to the deterministic series to isolate seasonal behavior is not made since no dominant frequency in the series is observed from frequency domain analysis. The deterministic series is analyzed using the correlation dimension method with the same embedding dimension, i.e. 10, and window size, i.e. 3, to examine a low-dimensional chaos in the system. Figure 11 shows the  $\text{Ln } C(r)$  vs.  $\text{Ln } r$  plot for the series. The correlation exponent vs.  $m$  plot in Fig. 10 for the deterministic series shows that the correlation exponent saturates after an embedding dimension of 15 based on the objective techniques. The saturation value of the correlation exponent is  $3.97$ . The number of variables required to model the streamflow dynamical system would be 4. The low value of

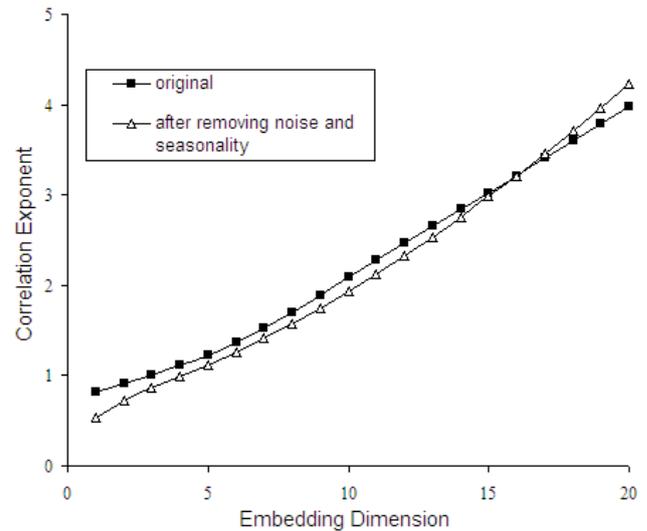


**Fig. 13.** Monthly streamflow data at the Arkansas river: Variation of *MSE* with forecast lead time for the original and deterministic data. The deterministic data is obtained after removing noise from original data.

correlation dimension suggests the possible presence of low-dimensional chaos in the streamflow dynamics. Figure 12 shows one-step, two-step, and three-step ahead predictions for the original monthly streamflow series at the Arkansas river. Figure 13 shows *MSE* vs. Forecast lead time plots for the original and deterministic series. The *MSE* values at Forecast lead time of 7 and 8 are 27.22 and 315.27, respectively. For the original series, *MSE* values are high and increase as the forecast lead time increases indicating the presence of non-deterministic component confirming the results obtained earlier using the PSR. This, in turn, decreases the prediction accuracy with an increase in the lead time. For the deterministic series, the prediction accuracy is significantly improved, i.e. the *MSE* values are very low as compared with the original series values, indicating that the deterministic series is indeed chaotic. This confirms the results from the correlation dimension method.

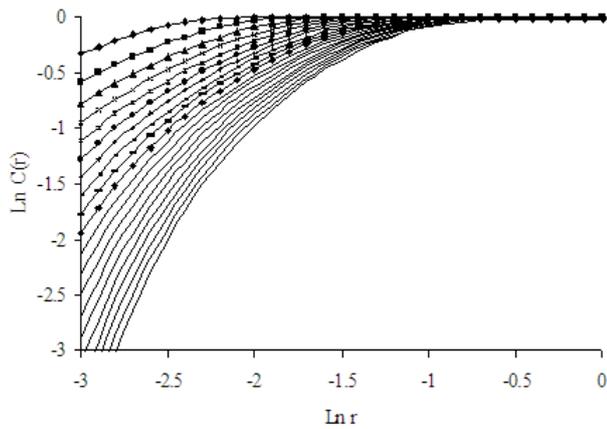
## 5.2 Colorado river

The daily streamflow series observed at the Colorado river below Parker dam is shown in Fig. 2. This series is analyzed for low-dimensional chaotic behavior using the correlation dimension method. The slope is calculated by considering a moving window of size 3 units. The correlation exponent is the slope value at  $\ln r = -2.2$ . The correlation exponent vs.  $m$  plot for the original series is shown in Fig. 14. Based on an eye estimation and the objective techniques described in



**Fig. 14.** Correlation exponent vs. embedding dimension plot for daily streamflow series at the Colorado river.

Sect. 2.1.1, it is seen that the correlation exponent does not reach a saturation value. This indicates that either the system is stochastic or it contains some dominant noise/seasonality component. To separate the non-deterministic component, the original series with  $\mu=0.1672$  and  $\sigma=0.0594$  is analyzed using the PSR with  $m=10$ . The mean and the standard deviation of the non-deterministic component, i.e. noise, separated using the PSR are 0.0 and 0.0099, respectively. The statistics for the deterministic component that might contain chaos and seasonality are given as  $\mu=0.1669$  and  $\sigma=0.0593$ . The ratio of  $\sigma_n$  (0.0099) and  $\sigma_c$  (0.0593) is 0.167 which is greater than the threshold value of 0.02, obtained for the simulated data in Sect. 4.2.2. The series, after the removal of noise, is analyzed using the frequency domain method to determine the dominant frequency. No peak is observed in the periodogram. However, based on an eye estimation a possible seasonal behavior in the data is observed. To isolate and remove the possible presence of seasonality, a periodic function is chosen to fit the data such that the *MSE* is minimized. This function is obtained using a trial and error procedure. The mean and the standard deviation of the series after the removal of noise and seasonality, i.e. deterministic series, according to the methods described earlier are 0.0469 and 0.0328, respectively. The deterministic series is analyzed using the correlation dimension method with the same embedding dimension, i.e. 10, and window size, i.e. 3. Figure 15 shows the  $\ln C(r)$  vs.  $\ln r$  plot for the deterministic series. The correlation exponent vs.  $m$  curve shown in Fig. 14 for the deterministic series does not saturate based on an eye estimation and the objective techniques indicating that the system is stochastic. Figure 16 shows one-step, two-step, and three-step ahead predictions for the original daily series at the Colorado river. The *MSE* vs. Forecast lead time plots for the original and deterministic series are shown in Fig. 17. The prediction accuracy is high, i.e. *MSE* values

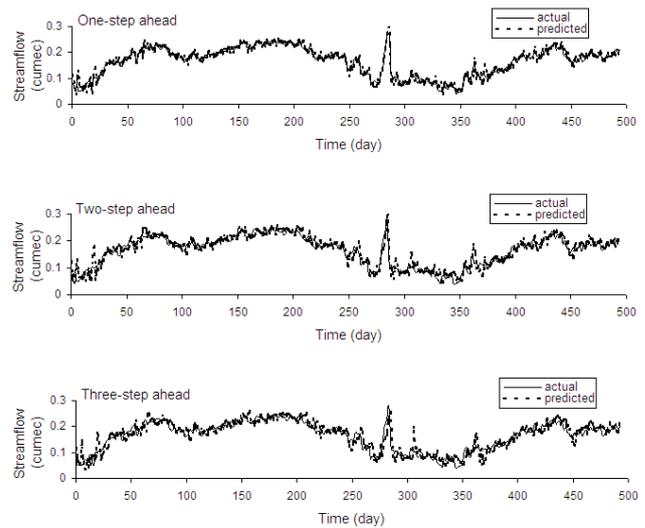


**Fig. 15.**  $LnC(r)$  vs.  $Ln r$  for daily streamflow series, after removing noise and seasonality, at the Colorado river. The curves are shown in ascending order of embedding dimension,  $m=1, 2, \dots, 20$  from top to bottom.

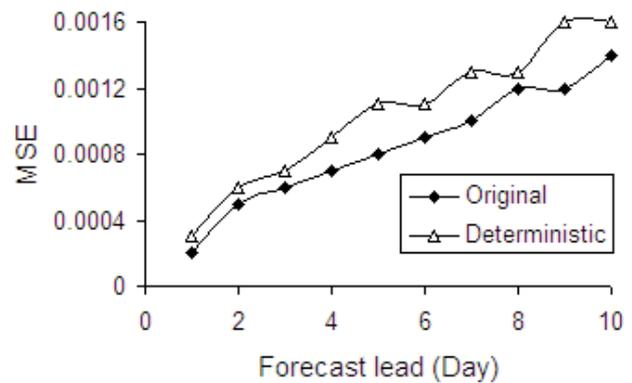
are low, due to the presence of seasonality in the data. For the original series, the  $MSE$  increases with an increase in the forecast lead time. The prediction accuracy does not change significantly for the deterministic series, i.e.  $MSE$  values are very close to the original series values, indicating the absence of chaos in the deterministic series. This indicates that the system is stochastic confirming the results from the correlation dimension method. The streamflow measurement gauge is located at a distance of 100 feet downstream of the dam (USGS, 2004). In addition, the flow at the dam is regulated based on requirements (BoR, 2004). These, in turn, may depend on various factors on a day-to-day basis leading to significant randomness in the observations. The measured flow at the dam, thus, reflects a variety of human controls implying that natural variability no longer is the dominant factor. This is the most likely cause for the loss of nonlinear dynamical information in the time series.

## 6 Summary and conclusions

The strengths of the use of nonlinear dynamical tools in conjunction with statistical time and frequency domain methodologies were demonstrated in this work. This study investigated the ability to detect and model chaos from finite hydrologic observations, especially when randomness and seasonality are present. The ability to detect and model nonlinear dynamical and chaotic components, from finite real-world observations, is likely to have significant implications for scientific understanding and predictive modeling in multiple disciplines. This research represents a step forward in these directions. The results of the present investigation demonstrated the presence of thresholds, expressed in terms of noise to chaotic-signal and seasonality to chaotic-signal ratios. The ability to detect chaos from observations depends on whether the chaotic component in the hydrologic time series is dominant enough to satisfy the thresholds. It



**Fig. 16.** Multistep ahead predictions for the Colorado river streamflow data. Top: one-step ahead predictions. Middle: two-step ahead predictions. Bottom: three-step ahead predictions.



**Fig. 17.** Daily streamflow data at the Colorado river: Variation of  $MSE$  with forecast lead time for the original and deterministic data. The deterministic data is obtained after separating white noise and seasonality from the original data.

was shown that the overall time series can be decomposed into the contributing random, seasonal and chaotic components. Time series was decomposed using nonlinear predictive modeling for separating the chaotic component, statistical methods for characterizing random data, and frequency domain approaches for isolating seasonality. This has direct implications for a scientific understanding of hydrologic phenomena and the dominant processes that may be present as well as in the development of predictive models. For example, it was shown that the chaotic component, once detected and isolated, can be better predicted in the short-term through nonlinear models like ANN and PSR. The insights obtained from simulated data were used to interpret the results of real streamflow data from the Arkansas and Colorado rivers. It was observed that a chaotic component can be detected, iso-

lated and utilized for improved predictive modeling from finite hydrologic time series like the Arkansas data. However, it was seen that for certain hydrologic data like the Colorado river data, this may not be possible. This may be due to the absence of chaotic/nonlinear dynamical component in the data. If the data observed at the Colorado river does contain chaotic/nonlinear dynamic component, then it may be completely dominated by randomness introduced as a result of the stochastic mode of dam operation on a daily basis.

*Acknowledgements.* Most of this work was done when the second author was on a visiting faculty appointment at the University of South Florida. The authors would like to thank two anonymous referees for their helpful comments.

Edited by: B. Sivakumar

Reviewed by: two referees

## References

- Abarbanel, H. D.: *Nonlinear Systems*, VCH Publishers, New York, 1994.
- Almog, Y., Oz, O., and Akselrod, S.: Correlation dimension estimation: can this nonlinear description contribute to the characterization of blood pressure control in rats?, *IEEE Transactions on Biomedical Engineering*, 46, 5, 535–537, 1990.
- ASCE: Task Committee on Applications of Artificial Neural Networks in Hydrology, II, Hydrologic applications, *J. Hydrol. Eng.*, 5, 2, 124–137, 2000.
- Basu, S. and Foufoula-Georgiou, E.: Detection of nonlinearity and chaos in time series using the transportation distance function, *Phys. Lett. A*, 301, 413–423, 2002.
- Bishop, C. M.: *Neural Networks for pattern recognitions*, Oxford University Press, 1995.
- BoR: Bureau of Reclamation, private communication, 2004.
- Bras, R. L. and Rodriguez-Iturbe, I.: Rainfall generation: A non-stationary time varying multidimensional model, *Water Resour. Res.*, 12, 450–456, 1976.
- Cornelis, A. L.: Visualization of Chaos for Finance Majors, in 2000 Finance Educators Conference: Finance Education in the New Millennium, Proceedings of the 2000 Annual Conference, Deakin University, Burwood, Victoria, Australia, 187–226, 2000.
- Douglas, E. and Barros, A. P.: Probable Maximum Precipitation Estimation Using Multifractals: Application in the Eastern United States, *J. Hydrometeorology*, 4, 6, 1012–1024, 2003.
- Elshorbagy, A.: Noise reduction approach in chaotic hydrologic time series revisited, *Canadian Water Resources Journal*, 26, 4, 537–550, 2001.
- Elshorbagy, A., Panu, U., and Simonovic, S.: Analysis of cross-correlated chaotic streamflows, *Hydrological Sciences Journal*, 46, 5, 781–794, 2001.
- Elshorbagy, A., Simonovic, S. P., and Panu, U. S.: Estimation of missing streamflow data using principles of chaos theory, *Journal of Hydrology*, 255, 1–4, 123–133, 2002a.
- Elshorbagy, A., Simonovic, S. P., and Panu, U. S.: Noise reduction in chaotic hydrologic time series: Facts and doubts, *Journal of Hydrology*, 256, 3–4, 147–165, 2002b.
- Essex, C., Lookman, T., and Nerenberg, M. A. H.: The climate attractor on short time scales, *Nature*, 326, 64–66, 1987.
- S. Khan et al.: Detection and predictive modeling of chaos
- Farmer, J. D. and Sidorowich, J. J.: Predicting Chaotic Time Series, *Phys. Rev. Lett.*, 59, 845–848, 1987.
- Fraedrich, K.: Estimating the dimensions of weather and climate attractors, *J. Atmos. Sci.*, 43, 419–432, 1986.
- Fraedrich, K.: Estimating weather and climate predictability on attractors, *J. Atmos. Sci.*, 44, 722–728, 1987.
- Ganguly, A. R. and Bras, R. L.: Distributed quantitative precipitation forecasting combining information from radar and numerical weather prediction model outputs, *J. of Hydrometeorology*, American Meteorological Society, 4, 6, 1168–1180, 2003.
- Grassberger, P. and Procaccia, I.: Measuring the strangeness of strange attractors, *Physica D*, 9, 189–208, 1983.
- Hense, A.: On the possible existence of a strange attractor for the southern oscillation, *Beitr. Phy. Atmos.*, 60, 1, 34–47, 1987.
- Hilborn, R. C.: *Chaos and Nonlinear Dynamics*, Oxford University Press, 2000.
- Hsieh, D.: Chaos and nonlinear dynamics: Applications to financial markets, *Journal of Finance*, 46, 1839–1878, 1991.
- Islam, M. N. and Sivakumar, B.: Characterization and prediction of runoff dynamics: a nonlinear dynamical view, *Adv. Water Resour.*, 25, 179–190, 2002.
- Islam, S., Bras, R. L., and Rodriguez-Iturbe, I.: A possible explanation for low correlation dimension estimates for the atmosphere, *J. Appl. Meteorol.*, 32, 203–208, 1993.
- Jayawardena, A. W. and Lai, F.: Analysis and prediction of chaos in rainfall and streamflow time series, *J. Hydrol.*, 153, 23–52, 1994.
- Kim, G. and Barros, A. P.: Quantitative Flood Forecasting Using Multisensor Data and Neural Networks, *J. Hydrology*, 246, 45–62, 2001.
- Koutsoyiannis, D. and Pachakis, D.: Deterministic chaos versus stochasticity in analysis and modeling of point rainfall series, *J. Geophys. Res.*, 101, 26 441–26 451, 1996.
- Kuligowski, R. J. and Barros, A. P.: Localized Precipitation Forecasts from a Numerical Weather Prediction Model Using Artificial Neural Networks, *Weather and Forecasting*, 13, 4, 1194–1204, 1998a.
- Kuligowski, R. J. and Barros, A. P.: Using Artificial Neural Networks to Estimate Missing Rainfall Data, *J. of American Water Resources Association*, 34, 6, 1–11, 1998b.
- Lorenz, E. N.: Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmos. Sci.*, 26, 636–646, 1969.
- Lorenz, E. N.: *The Essence of Chaos (The Jessie and John Danz Lecture Series)*, University of Washington Press, 1996.
- Nerenberg, M. A. H. and Essex, C.: Correlation dimension and systematic geometric effects, *Phys. Rev. Lett. A*, 42, 12, 7065–7074, 1990.
- Osborne, A. R. and Provenzale, A.: Finite correlation dimension for stochastic systems with power-law spectra, *Physica D*, 35, 357–381, 1989.
- Porporato, A. and Ridolfi, L.: Clues to the existence of deterministic chaos in river flow, *Int. J. Mod. Phys. B*, 10, 1821–1862, 1996.
- Rodriguez-Iturbe, I., De Power, F. B., Sharifi, M. B., and Georgakakos, K. P.: Chaos in rainfall, *Water Resour. Res.*, 25, 7, 1667–1675, 1989.
- Schertzer, D., Tchiginskaya, Y., Lovejoy, S., Hubert, P., Bendjoudi, H., and Larchevque, M.: Which chaos in the rainfall-runoff process?, *Hydrological Sciences Journal*, 47, 1, 139–149, 2002.
- Sivakumar, B.: Chaos theory in hydrology: important issues and interpretations, *J. Hydrol.*, 227(1–4), 1–20, 2000.
- Sivakumar, B.: Forecasting monthly streamflow dynamics in the western United States: A nonlinear dynamical approach, *Environmental Modeling and Software*, 18, 8/9, 721–728, 2003.

- Sivakumar, B.: Chaos theory in geophysics: Past, present and future, *Chaos, Solitons and Fractals*, 19, 2, 441–462, 2004.
- Sivakumar, B., Liong, S. Y., Liaw, C. Y., and Phoon, K. K.: Evidence of chaotic behavior in singapore rainfall, *J. Am. Water Resour. Assoc.*, 34, 301–310, 1998.
- Sivakumar, B., Berndtsson, R., Olsson, J., Jinno, K., and Kawamura, A.: Dynamics of monthly rainfall-runoff process at the Gota basin: A search for chaos, *Hydrology and Earth System Sciences*, 4, 3, 407–417, 2000.
- Sivakumar, B., Berndtsson, R., Olsson, J., and Jinno, K.: Evidence of chaos in rainfall-runoff process, *Hydrological Sciences Journal*, 46, 1, 131–146, 2001a.
- Sivakumar, B., Berndtsson, R., and Persson, M.: Monthly runoff prediction using phase-space reconstruction, *Hydrological Sciences Journal*, 46, 3, 377–387, 2001b.
- Sivakumar, B., Berndtsson, R., Olsson, J., and Jinno, K.: Discussion on Analysis of cross-correlated chaotic streamflows by Elshorbagy et al., *Hydrological Sciences Journal*, 47, 3, 523–527, 2002a.
- Sivakumar, B., Berndtsson, R., Olsson, J., and Jinno, K.: Reply to Which chaos in the rainfall-runoff process? by Schertzer et al, *Hydrological Sciences Journal*, 47, 1, 149–158, 2002b.
- Sivakumar, B., Persson, M., Berndtsson, R., and Uvo, C. B.: Is correlation dimension a reliable indicator of low-dimensional chaos in short hydrological time series?, *Water Resources Research*, 38, 2, doi:10.1029/2001WR000333, 31–38, 2002c.
- Stehlik, J.: Deterministic chaos in runoff series, *J. Hydrol. Hydromech*, 47, 271–287, 1999.
- Sugihara, G. and May, R. M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, 344, 734–741, 1990.
- Takens, F.: Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, Lecture notes in Mathematics, 898, Springer Verlag, New York, 366–381, 1980.
- Tessier, Y., Lovejoy, S., Hubert, P., Schertzer, D., and Pecknold, S.: Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions, *J. geophysical Res.*, 101, D21, 26 427–26 440, 1996.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D.: Testing for Nonlinearity in Time Series: The Method of Surrogate Data, *Physica D*, 58, 77–94, 1992.
- Trippi, R. R.: *Chaos & Nonlinear Dynamics in the Financial Markets: Theory, Evidence and Applications*, Irwin Professional Publishing, 1995.
- USGS: private communication, 2004.
- Wang, G.: A conceptual modeling study on biosphere-atmosphere interactions and its implications for physically based climate models, *Journal of Climate*, 17, 13, 2572–2583, 2004.
- Weigend, A. S. and Gershenfeld, N. A.: *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley, Reading, MA, 1994.
- Wilcox, B. P., Seyfried, M. S., and Matison, T. H.: Searching for chaotic dynamics in snowmelt runoff, *Water Resour. Res.*, 27, 6, 1005–1010, 1991.
- Yim, G. S., Ryu, J. W., Park, Y. J., Rim, S., Lee, S. Y., Kye, W. H., and Kim, C. M.: Chaotic behaviors of operational amplifiers, *Physical Review E*, 69, 2004.
- Zhang, G., Patuwo, B. E., and Hu, M. Y.: Forecasting with artificial neural networks: The state of the art, *Int. J. Forecasting*, 14, 35–62, 1998.