Predictability and Chaotic Nature of Daily Streamflow

Article *in* Australian Journal of Water Resources · January 2013

DOI: 10.7158/W12-024.2013.17.1

Dhanya C T
Indian Institute of Technology Delhi

D Nagesh Kumar
Indian Institute of Science

Predictability and chaotic nature of daily streamflow*

CT Dhanya
Department of Civil Engineering, Indian Institute of Technology Delhi, New Delhi, India

D Nagesh Kumar†
Department of Civil Engineering, Indian Institute of Science, Bangalore, India

ABSTRACT: *The predictability of a chaotic series is limited to a few future time steps due to its sensitivity to initial conditions and the exponential divergence of the trajectories. Over the years, streamflow has been considered as a stochastic system. In this study, the chaotic nature of daily streamflow is investigated using autocorrelation function, Fourier spectrum, correlation dimension method (Grassberger-Procaccia algorithm) and false nearest neighbour method. Embedding dimensions of 6-7 obtained, indicate the possible presence of low-dimensional chaotic behaviour. The predictability of the system is estimated by calculating the system's Lyapunov exponent. A positive maximum Lyapunov exponent of 0.167 indicates that the system is chaotic and unstable with a maximum predictability of only 6 days. These results give a positive indication towards considering streamflow as a low dimensional chaotic system than as a stochastic system. Prediction is done using local polynomial method for a range of embedding dimensions and delay times. The uncertainty in the chaotic streamflow series is reasonably captured through the ensemble approach using local polynomial method.*

KEYWORDS: Streamflow; chaos; correlation dimension; Lyapunov exponent; nearest neighbour; non-linear prediction; local polynomial prediction.

REFERENCE: Dhanya, C. T. & Nagesh Kumar, D. 2013, "Predictability and chaotic nature of daily streamflow", *Australian Journal of Water Resources*, Vol. 17, No. 1, pp. 1-12, http://dx.doi.org/10.7158/W12-024.2013.17.1.

1 INTRODUCTION

The development of various climate models that numerically integrate an adequate set of mathematical equations of physical laws governing the climatic processes marked a major breakthrough in the routine weather prediction. The mathematical equations in these climate models form a non-linear dynamical system in which an infinitesimally small uncertainty in the initial conditions will grow exponentially even under a perfect model, leading to a chaotic behaviour (Smith et al, 1998). Such sensitivity of any deterministic system to a slight change in the initial conditions leads to a vast change in the final solution and is often known as "butterfly effect" in the field of weather forecasting (Lorenz, 1972). Hence, Earth's weather can be treated as a chaotic system with a finite limit in the predictability, arising mainly due to the incompleteness of initial conditions. The exponential growth with time of an infinitesimal initial uncertainty ∂_0 is given by the highest Lyapunov exponent λ (Wolf et al, 1985; Rosenstein et al, 1993). Hence, the separation or uncertainty after Δt time steps ahead is given as $\partial_{\Delta t} \cong e^{\lambda \Delta t}\partial_0$. The predictability of a chaotic system is therefore limited (i) due to the indefiniteness in the initial conditions (given a perfect model) and also (ii) due to the imperfection of the model.

Modelling of many weather phenomena have been done so far employing the concept of stochastic systems. However, a large number of studies

* Reviewed and revised version of paper originally presented at the 34th International Association for Hydro-Environment Engineering and Research (IAHR) World Congress (incorporating the 33rd Hydrology and Water Resources Symposium and 10th Conference on Hydraulics in Water Engineering), 26 June to 1 July 2011, Brisbane, Queensland.

† Corresponding author Prof Nagesh Kumar can be contacted at nagesh@civil.iisc.ernet.in.

employing the science of chaos to model and predict various hydrological phenomena have emerged only in the past decade (Elshorbagy et al, 2002; Islam & Sivakumar, 2002; Jayawardena & Lai, 1994; Porporato & Ridolfi, 1996; 1997; Puente & Obregon, 1996; Rodriguez-Iturbe et al, 1989; Liu et al, 1998; Sangoyomi et al, 1996; Sivakumar et al, 1999; 2001; Sivakumar, 2001; Shang et al, 2009; Wang & Gan, 1998; Dhanya & Kumar, 2010; 2011a). Most of these studies dealt with scalar time series data of various hydrological phenomena like rainfall, runoff, sediment transport, lake volume, etc. In these cases, since neither the mathematical relations nor the influencing variables are known, the state space in which the variable is lying is reconstructed from the time series itself using phase space reconstruction method by Takens (1981).

The phase space reconstruction provides a simplified, multi-dimensional representation of a single-dimensional non-linear time series. According to this approach, given the embedding dimension m and the time delay τ, for a scalar time series X_i where $i = 1, 2, ..., N$, the dynamics can be fully embedded in m-dimensional phase space represented by the vector:

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

Now, the dynamics can be interpreted in the form of an m-dimensional map f_T such that $Y_{j+T} = f_T(Y_j)$ where Y_j and Y_{j+T} are vectors of dimension m; Y_j being the state at current time j and Y_{j+T} being the state at future time $j+T$. The approximation of f_T can be done using either a global or a local non-linear model.

The dimension m can be considered as the minimum number of state variables required to describe the system, which is commonly estimated through Grassberger-Procaccia algorithm (GPA) (Grassberger & Procaccia, 1983a), and the false nearest neighbour (FNN) method (Kennel et al, 1992). The optimum τ is usually determined using either autocorrelation function or the mutual information method (Frazer & Swinney, 1986). For an exponential autocorrelation function, delay time can be chosen as the lag time at which the autocorrelation crosses the zero line (Holzfuss & Mayer-Kress, 1986) or falls below the threshold value e^{-1} (Tsonis & Elsner, 1988). Another method is to take τ as the lag time corresponding to the first minimum of the mutual information function.

The outcomes of these studies affirm the existence of low-dimensional chaos, thus indicating the possibility of only short-term predictions. Better predictions can be obtained using the chaotic approach since it takes into account the dynamics of the irregular hydrological phenomena from a chaotic deterministic view, thereby reducing the model uncertainty. Also, the dynamic approach employing chaotic theory outperforms the traditional stochastic approach in prediction (Jayawardena & Gurung, 2000). Most of these studies rely only on the low correlation dimension as a measure of the chaotic nature of the time series and as an estimate of embedding dimension. Osborne & Provenzale (1989) claimed that a low correlation dimension can also be observed for a linear stochastic process. Hence, it is advised to assess the chaotic nature and to determine the embedding dimension and delay time by employing a variety of methods (Islam & Sivakumar, 2002; Dhanya & Kumar, 2010; 2011b). Since different methods will give slightly different embedding dimensions and delay times for a single series, one should opt for an ensemble of predictions with a set of these parameters in order to capture the uncertainty in parameter estimation (Dhanya & Kumar, 2010).

The aim of this paper is to analyse the chaotic behaviour and predictability of a streamflow series employing various techniques. Autocorrelation method is used for preliminary investigation to identify chaos and also to determine the delay time for the phase space reconstruction. Optimum embedding dimension is determined using correlation dimension and FNN algorithms. Phase space prediction is done using local polynomial method and the model performance and convergence is analysed.

2 PREDICTABILITY AND CHAOTIC NATURE

A variety of techniques have emerged for the identification of chaos which include correlation dimension method (Grassberger & Procaccia, 1983a), FNN algorithm (Kennel et al, 1992), non-linear prediction method (Farmer & Sidorowich, 1987), Lyapunov exponent method (Kantz, 1994), Kolmogorov entropy (Grassberger & Procaccia, 1983b), and surrogate data method (Theiler et al, 1992). In this study, correlation dimension, FNN method and Lyapunov exponent are employed to analyse the chaotic nature of the time series.

2.1 Lyapunov Exponent

One of the basic characteristics of a chaotic system is the unpredictability due to the sensitive dependence on initial conditions. The divergence between the trajectories emerging from very close initial conditions will grow exponentially, hence making the system difficult to predict even after a few time steps. Lyapunov exponent gives the averaged information of divergence of infinitesimally close trajectories and thus the unpredictability of the system. Let s_{t1} and s_{t2} be two points in two trajectories in state space such that the distance between them is $\|s_{t1} - s_{t2}\| = \partial_0 << 1$. After Δt time steps ahead, the distance $\partial_{\Delta t} \cong \|s_{t1+\Delta t} - s_{t2+\Delta t}\|$, $\partial_{\Delta t} << 1$, $\Delta t >> 1$ follows an exponential relation with initial separation ∂_0, ie. $\partial_{\Delta t} \cong e^{\lambda \Delta t} \partial_0$, where λ is the Lyapunov exponent (Kantz, 1994). Since the rate of separation is different for various orientations of initial separation vector, the total number of Lyapunov exponents is equal to the number of dimensions of the phase space defined, ie. a spectrum of exponents will be available. Among them, the highest (global) Lyapunov exponent need only be considered, as it determines the total predictability of the system.

Many algorithms have been developed to calculate the maximal Lyapunov exponent (Wolf et al, 1985; Rosenstein et al, 1993; Kantz, 1994). The exponential divergence is examined here using algorithm introduced by Rosenstein et al (1993). For calculating the maximum Lyapunov exponent, one has to compute:

$$S(\Delta t) = \frac{1}{N}\sum_{t_0=1}^{N}\ln\left(\frac{1}{\left|U\left(s_{t_0}\right)\right|}\sum_{s_t \in U\left(s_{t_0}\right)}\left|s_{t_0+\Delta t} - s_{t+\Delta t}\right|\right) \quad (2)$$

where s_{t_0} are reference points or embedding vectors, and $U\left(s_{t_0}\right)$ is the neighbourhood of s_{t_0} with diameter ξ. For a reasonable range of ξ and for all embedding dimensions m which is larger than some minimum dimension m_0, if $S(\Delta t)$ exhibits a linear increase, then its slope can be taken as an estimate of the maximal Lyapunov exponent λ.

The exponential divergence of the nearby trajectories and hence an unstable orbit (chaos) is indicated by a positive λ. Negative Lyapunov exponents are characteristic of dissipative or non-conservative systems. Their orbits attract to a stable fixed point or periodic orbit. Zero Lyapunov exponents are exhibited by conservative systems for which the orbit is a neutral fixed point. For more details, refer Kantz & Schreiber (2004).

2.2 Correlation dimension method

In correlation dimension method, the correlation integral $C(r)$ is estimated using the GPA (Grassberger & Procaccia, 1983a) which uses the reconstructed phase space of the time series as given in equation (1). According to this algorithm, for an m-dimensional phase space, the correlation integral $C(r)$ is given by:

$$C(r) = \lim_{N\to\infty}\frac{2}{N(N-1)}\sum_{\substack{i,j\\(1\le i<j\le N)}} H\left(r - \left|Y_i - Y_j\right|\right) \quad (3)$$

where H is the Heaviside function, with $H(u) = 1$ for $u > 0$ and $H(u) = 0$ for $u \le 0$, where $u = (r - | Y_i - Y_j |)$, r is the radius of the sphere centred on Y_i or Y_j, and N is the number of data. For small values of r, the correlation integral holds a power law relation on r, $C(r) \sim r^d$, where d is the correlation dimension of the attractor. The correlation exponent or the dimension, d can be calculated from the slope of the plot of $\log C(r)$ versus $\log r$.

For a chaotic series, the correlation exponent saturates to a constant value on increasing the embedding dimension m and the nearest integer above that saturation value indicates the number of variables necessary to describe the evolution in time. On the other hand, if the correlation exponent increases without reaching a constant value on increase in the embedding dimension, the system under investigation is generally considered as stochastic. This is because, contrary to the low dimensional chaotic systems, stochastic systems acquire large dimensional subsets of the system phase space, leading to an infinite dimension value.

However, the sole presence of finite, non-integer dimension correlation dimension is not sufficient to indicate the presence of a strange attractor. Osborne & Provenzale (1989) opposed the traditional view that stochastic processes lead to a non-convergence of the correction dimension by demonstrating that "coloured random noises" characterised by a power law power spectrum exhibit a finite and predictable value of the correlation dimension. While the saturation of correlation dimension in low dimensional dynamic systems is due to the phase correlations, for the above mentioned stochastic systems it is mainly due to the shape of the power spectrum (power law). Hence, it would be worthwhile to repeat the correlation dimension on first numerical derivative and phase randomised signal of the original data, to distinguish low dimensional dynamics and randomness (Provenzale et al, 1992). In the case of stochastic systems, due to the change in the spectral slope on differentiation, the correlation dimension of the differentiated signal will be much larger than that of the original signal. For low dimensional dynamic systems, correlation dimension will be almost invariant.

Phase randomised signal of the original data can be obtained by generating stochastic surrogate data of the same Fourier spectra as that of the original data. The Fourier phases are then randomised and are uniformly distributed. In the case of phase randomised data, the correlation dimension will be the same as that of the original data, provided the convergence of the dimension is forced only by the shape of the power spectrum and not due to any low-dimensional dynamics.

Another approach is to compute the correlation integral of the first (numerical) derivative of the signal. If the system is stochastic, then the correlation dimension of the differentiated signal will be much larger than that of the original signal. This behaviour is due to the change in the spectral slope on differentiation. But for low dimensional dynamic systems, correlation dimension will be almost invariant, since the saturation is in fact due to the chaotic nature and not due to the power law shape of power spectrum. For more details, refer Kantz & Schreiber (2004).

2.3 False nearest neighbour method

The concept of FNN is based on the concept that if the dynamics in phase space can be represented by a smooth vector field, then the neighbouring states would be subject to almost the same time evolution (Kantz & Schreiber, 2004). Hence, after a short time into the future, any two close neighbouring trajectories emerging from them should still be close neighbours. In the present study, the modified algorithm by Hegger & Kantz (1999) in which the fraction of FNNs are computed in a probabilistic way has been used.

The basic idea is to search for all the data points which are neighbours in a particular embedding dimension m and which do not remain so, upon increasing the embedding dimension to m+1. Considering a particular data point, determine its nearest neighbour in the m^{th} dimension. Compute the ratio of the distances between these two points in the $m+1^{th}$ and m^{th} dimensions. If this ratio is larger than a particular threshold f, then the neighbour is false. When the percentage of FNNs falls to zero (or a minimum value), the corresponding embedding dimension is considered high enough to represent the dynamics of the series. For more details, refer Kantz & Schreiber (2004).

2.4 Local polynomial method

As discussed before, the state space is reconstructed using Takens theorem in equation (1). The prediction at T time steps ahead is then done by mapping the dynamics into a m-dimensional map using the function $Y_{j+T} = f_T(Y_j)$. The selection of a non-linear model for f_T can be made either globally or locally. The global approach approximates the map by working on the entire phase space of the attractor and seeking a form, valid for all points. Neural networks and radial basis functions adopt the global approach. In the second approach which works on local fitting, the dynamics are modelled locally piecewise in the embedding space. The domain is broken up into many local neighbourhoods and modelling is done for each neighbourhood separately, ie. there will be a separate f_T valid for each neighbourhood. The complexity in modelling f_T is thus considerably reduced without affecting the accuracy of prediction. One such approach is local approximation by Farmer & Sidorowich (1987), in which the prediction of Y_{j+T} is done based on values of Y_j and k nearest neighbours of Y_j. Y_{j+T} is taken as a weighted average of the k nearest neighbours.

In the present study, the prediction is done using local polynomial method. The applications of local polynomial approach have been demonstrated by various studies to analyse flood frequency (Apipattanavis et al, 2010), to predict hydrologic extremes (Lee & Ouarda, 2010), and forecasting time series (Regonda et al, 2005). The procedure for prediction using local polynomial method is as follows:

1. The state space is constructed for a specific embedding dimension and delay time.
2. A limited radius of influence or number of nearest neighbours (say $\alpha \times$ total number of points) is fixed around the current state of the system, Y_j.
3. The nearest neighbours, Y_{nn} falling near to Y_j are obtained along with their future states, Y_{nn+1}.
4. A local function is fitted between Y_{nn} and Y_{nn+1}, ie. $Y_{nn+1} = f(Y_{nn})$. In local polynomial method, the function, f is a polynomial with order greater or equal to 1.
5. Now, the current state, Y_j can be mapped into the T time steps into future using the expression $Y_{j+T} = f(Y_j) + \varepsilon$.

The mathematical details of local polynomial method can be found in Loader (1999). In this study, local polynomial is fitted using LOCFIT, which can be found from Lucent Technologies (2001).

3 DATA USED

The daily streamflow data at Basantpur station of Mahanadi basin, India, for the period June 1972 to May 2004 is considered for the present study. The location map of the Basantpur station on Mahanadi basin is shown in figure 1(a). The frequency histogram of the daily streamflow series for the study period is shown in figure 1(b). The streamflow is widely varying from 0 to 3.5×10^4 Mm^3 (million cubic metres), with maximum frequency falling in the range of 0-1000 Mm^3. Major portion of the annual streamflow is received in the monsoon months of July, August and September. The non-monsoonal flows are almost invariant, while the monsoon flows show large deviations from the mean.

4 RESULTS AND DISCUSSIONS

4.1 Preliminary investigation of chaos

As a preliminary investigation, the autocorrelation function and Fourier spectrum are plotted and are shown in figures 2(a) and 2(b) respectively. The initial exponential decay of autocorrelation function indicates that the streamflow may be of chaotic nature. The periodic behaviour of the autocorrelation function for higher lags is due to the seasonal periodicity. The power spectrum is also exhibiting a broad band form clearly visible for a large frequency range and a power law shape, ie. $P(f) \propto f^{-\alpha}$ with $\alpha \approx 1.35$. The choice of the delay time τ is made using the autocorrelation method and the mutual information method (Frazer & Swinney, 1986). In autocorrelation method, the lag time at which the autocorrelation function attains a zero value (figure 2(a)), ie. 74^{th} day is considered as the delay time. The mutual information obtained for various lag times are shown in figure 2(b). The delay time for the phase space reconstruction is the first minimum value, which is at 79^{th} day.

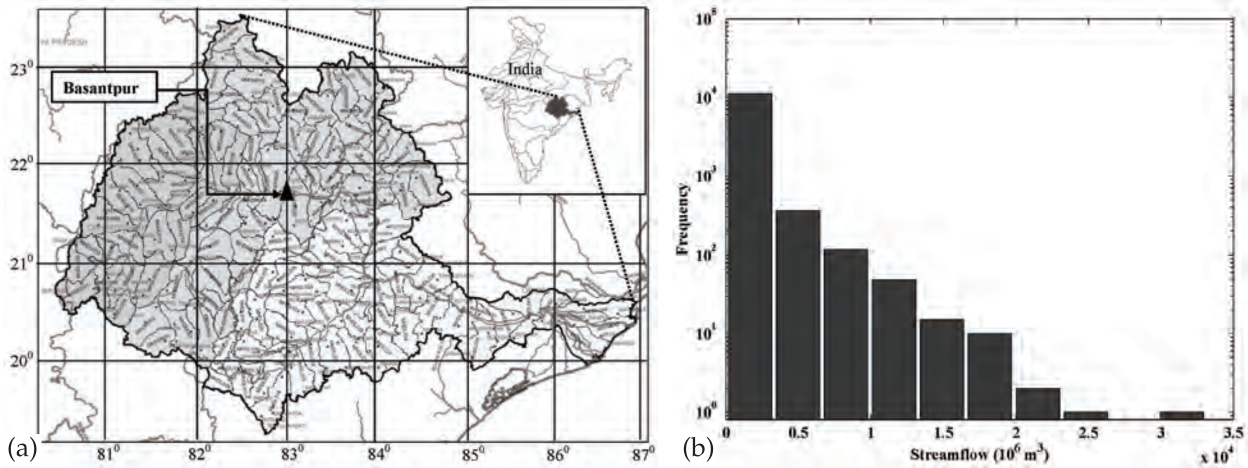

Figure 1: (a) Location map of the Basantpur station on Mahanadi basin. (b) Frequency histogram of Basantpur daily streamflow for the period June 1972 to May 2004 (frequency ordinate is in log scale).

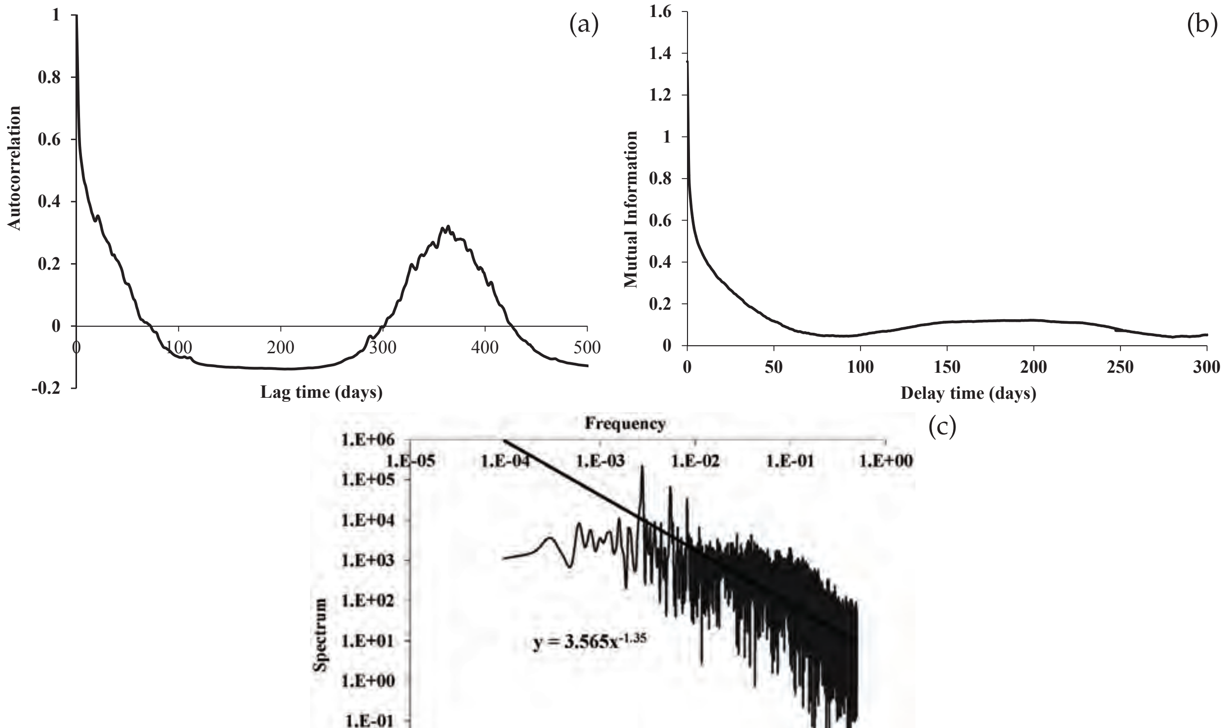

Figure 2: (a) Autocorrelation function, (b) variation of mutual information with lag time, and (c) Fourier spectrum of Basantpur streamflow data.

4.2 Determination of predictability: Lyapunov exponent

Lyapunov exponent provides a measure of the exponential growth due to infinitesimal perturbations. The maximal Lyapunov exponent is calculated employing the algorithm by Rosenstein et al (1993) which is based on the nearest neighbour approach. The variation of $S(\Delta t)$ with time t for Basantpur station at dimensions m = 4 to 6 is shown in figure 3. The slope of the linear part of the curve gives the maximum Lyapunov exponent. A positive slope of around 0.167 confirms the exponential divergence of trajectories and hence the chaotic nature of the daily streamflow. The inverse of the Lyapunov exponent defines the predictability of the system, which is around 7 days.

4.3 Determination of embedding dimension

4.3.1 Correlation dimension method

The correlation integral $C(r)$ is calculated according to GPA for embedding dimensions 1 to 40. A plot of correlation integral $C(r)$ versus radius r on a log-log scale for embedding dimensions m = 1 to 40 is shown in figure 4(a). For each of the embedding dimensions, slope of $C(r)$ versus r over the clear scaling region gives the corresponding correlation exponent. The variation of the correlation exponent with the embedding dimension is shown in figure 4(b). The correlation exponent is increasing with embedding dimension and reaching a constant saturation value at embedding dimension $m \geq 18$, which is an indication of the existence of chaos in the streamflow series. The saturation value is slightly different for different regions. The saturation value of 5.21 at an embedding dimension m = 18 indicates that the number of variables dominantly influencing the streamflow dynamics is approximately 6. The low correlation dimension also suggests the possible presence of low-dimensional chaotic behaviour.

The power spectrum of the Basantpur streamflow series is showing a power law behaviour with $\alpha \approx 1.35$ as shown in figure 2(b). Since the convergence of the correlation dimension can also be exhibited by some stochastic series due to its power law behaviour of power spectrum, it is recommended to perform the correlation dimension method on the first derivative and the phase randomised data of the original signal. A comparison of the variations of correlation exponent with embedding dimension for the first derivative of data, phase randomised data and original data are shown in figure 4(c). While the variation of correlation exponent of first derivative is almost identical to that of the original data with almost the same saturation value, the correlation dimension of the phase randomised data set is not converging at all. This eliminates the possibility of linear correlations forcing the saturation of correlation exponent and thereby confirms the presence of a low dimensional strange attractor in the streamflow series.

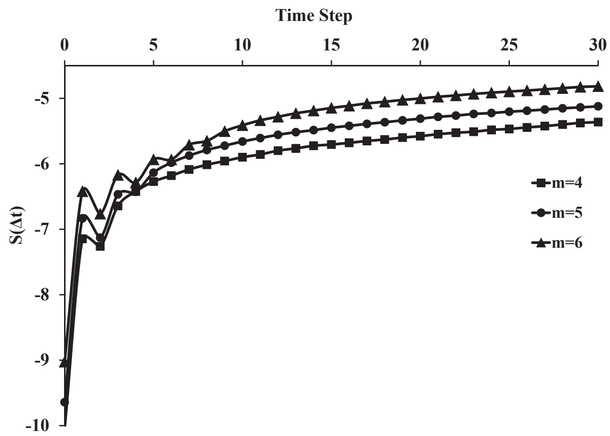

Figure 3: Variation of $S(\Delta t)$ with time for various embedding dimensions.

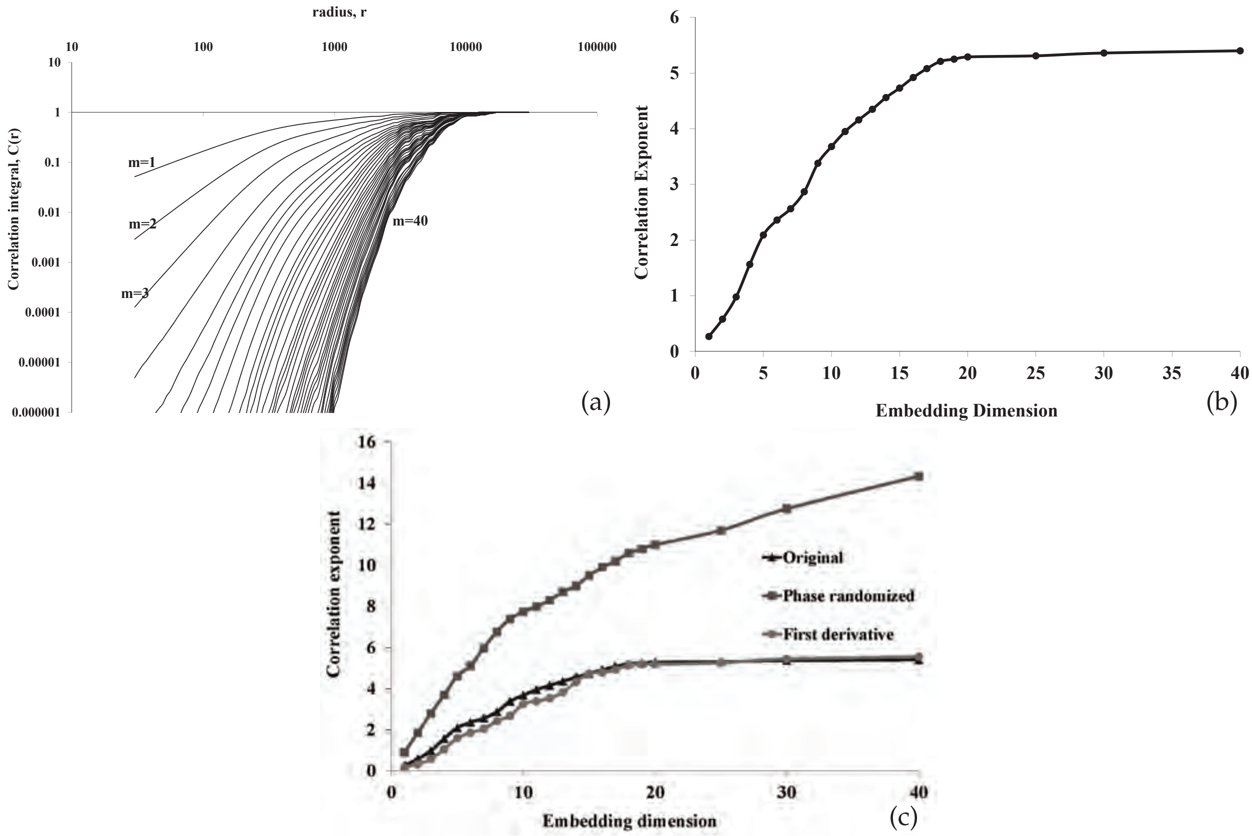

Figure 4: (a) Variation of correlation integral with radius on a log-log scale for embedding dimensions from 1 to 40; (b) variation of correlation exponent with embedding dimension; and (c) variation of correlation exponent with embedding dimension for original data, phase randomised data and first derivative of data.

4.3.2 *False nearest neighbour method*

The modified FNN algorithm by Hegger & Kantz (1999) is applied on the streamflow series. The threshold value *f* is fixed at 5. The variation of the fraction of FNNs for different embedding dimensions is shown in figure 5. The fraction of nearest neighbours is falling to a minimum value at an embedding dimension of 7, indicating that minimum 7 variables are necessary to explain the entire system. This is in close agreement with the value obtained by the correlation dimension method.

4.3 Local polynomial prediction

An ensemble of predictions is done using local polynomial approach with a range of embedding dimension from 4 to 6 and delay time from 71 to 80. State space is constructed for the specified range of embedding dimension and delay time. The order of the polynomial is fixed as 2. Training of the model is done for each combination (of embedding dimension and delay time) taking daily streamflow data for the period June 1972 to May 2003. Finally, ensemble of prediction is obtained for the year (June 2003 to May 2004). The mean daily ensemble predictions for lead time = 1 is shown in figure 6. A comparison of cumulative probability distributions (CDF) of ensembles and the observed series for lead time = 1 is shown in figure 7. The ensemble is able to catch the observed streamflow probabilities well with its range. Hence it can be concluded that the local polynomial model is able to reproduce the observed variations in the streamflow time series with much less uncertainty.

The prediction accuracy in terms of correlation between the observed and the ensemble for various lead times is demonstrated in figure 8. As the lead time increases, the accuracy of prediction is decreasing. This is also evident from figure 9 which demonstrates the variation in absolute deviation (box plots) for various lead times. In addition, from figure 8, it can be also observed that the uncertainty in predictions measured in terms of the width of the box plots (maximum/minimum), shown as bar plots in figure 9, is steadily increasing with respect to lead time. This shows the even though, local polynomial

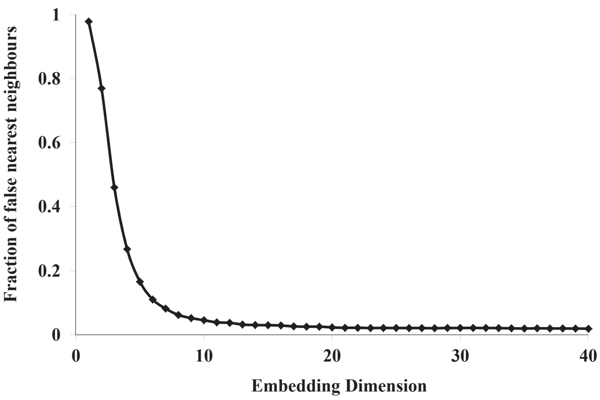

Figure 5: Variation of fraction of false nearest neighbours with embedding dimension.

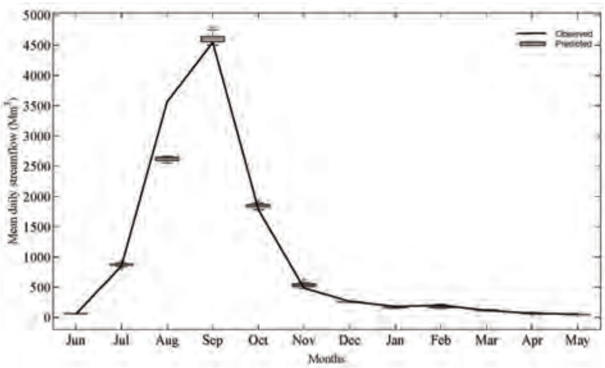

Figure 6: Ensemble of mean daily streamflow for embedding dimension = 6 and delay time = 74.

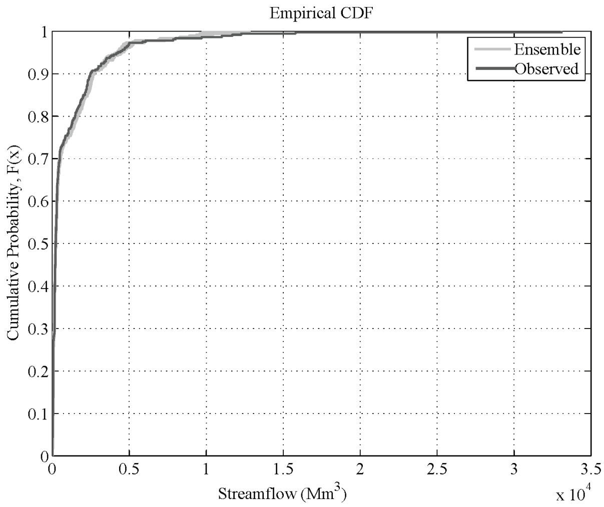

Figure 7: Cumulative density functions of daily streamflow. The ensemble CDFs and also the observed streamflow CDF are shown.

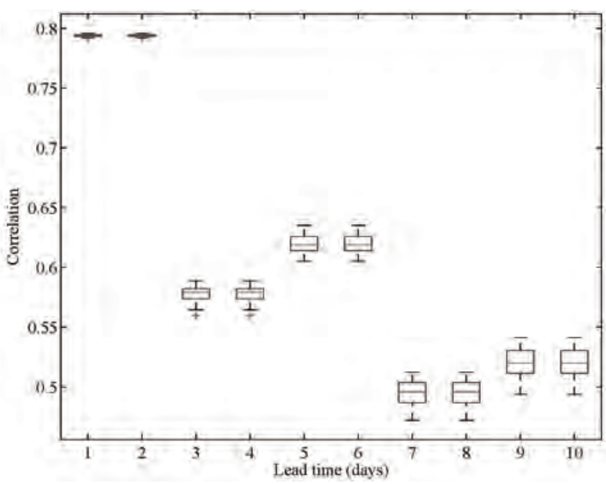

Figure 8: Box plots of correlation between observed and ensemble for various lead times.

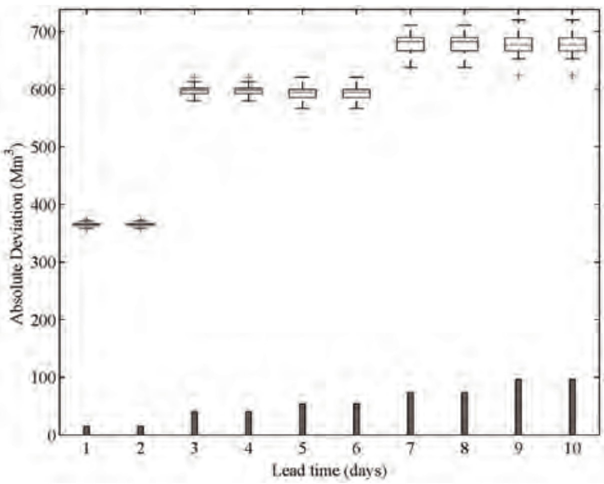

Figure 9: Box plots of absolute deviation between observed and ensemble for various lead times. The widths of box plots are shown as bar plots.

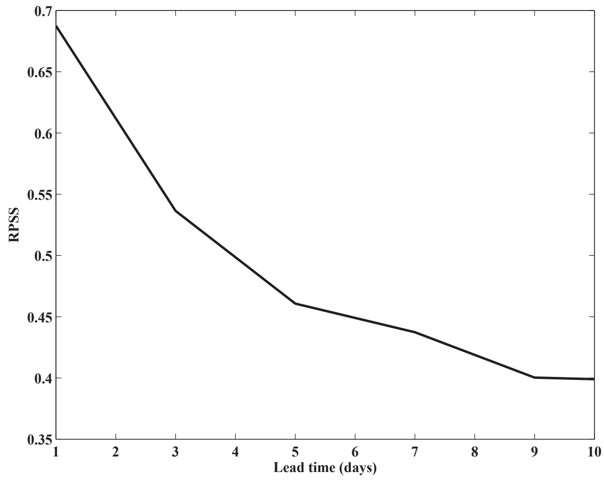

Figure 10: Rank probability skill scores of the ensemble predictions for various lead times.

method is able to capture the inherent dynamics in streamflow series, the chaotic nature limits the predictability of streamflow series, with increase in lead time.

The skill of the prediction is assessed with reference to the climatological (observed) values as the control or reference forecasts. The quality of the ensembles generated using local polynomial prediction is ascertained using two measures: (i) rank probability skill score (RPSS) and (ii) rank histograms (Wilks, 2005).

To compute RPSS, the dataset is divided into *n* categories. The RPSS is calculated as the sum of the squares of the difference of the cumulative probabilities of each of the predicted-observed data pair. RPS is given by:

$$RPS = \sum_{i=1}^{n} \left(P_i - O_i\right)^2 \tag{4}$$

where P_i is the cumulative probability of the forecast for category i and O_i is the cumulative probability of the observation for category i. Similarly, the RPS value for the climatological dataset is also computed. Finally, RPS score is calculated as:

$$RPSS = 1 - \frac{\overline{RPS}}{\overline{RPS_{clim}}} \tag{5}$$

where $\overline{RPS}$ is the mean RPS of all observation-forecast pairs and $\overline{RPS_{clim}}$ is the mean RPS of climatological forecast. An RPSS value of 1.0 indicates a perfect forecast and a negative value indicates an output worse than climatology. An RPSS of 0.0 implies no improvement in skill over the reference climatological forecast, RPS_{clim}. For more details, refer Wilks (2005) and Dhanya & Kumar (2010).

The RPSS values for the ensembles predictions for different lead times are computed by dividing the dataset into 10 categories based on the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th and 90th percentile values derived from the observed dataset. The overall RPSS values for the daily rainfall for different lead times are shown in figure 10. Positive RPSS values for all the lead times indicate a better forecast than the climatological forecast. However, the skill of the predictions decreases as the lead time increases, which again is due to the chaotic nature of the streamflow series as discussed before.

Rank histogram is used to evaluate the reliability and probable predictability of the targeted parameter by the ensembles. Let there be *n* observation forecast pairs and n_{ens} ensemble forecasts corresponding to each observation. Assuming the ensembles and also the observations are having the same probability distribution, for each ensemble forecast-observation pair, the rank of the observation is likely to take any of the values $i = 1, 2, 3, ..., n_{ens} + 1$. These ranks of the observation for each of the *n* data points are plotted to generate the rank histogram. A flat histogram indicates an ideal rank histogram whereas an U-shaped one indicates ensemble members from a less variable distribution. On the other hand, a dome shape histogram indicates too large ensemble spread. For more details, refer Wilks (2005) and Dhanya and Kumar (2010).

The rank histograms of ensemble predictions for different lead times are obtained and compared with the climatological rank histogram. The ensemble prediction rank histograms are relatively flatter than the climatology rank histogram. The rank histograms of climatology and ensemble prediction for lead time 10 days are shown in figure 11. As can be seen from the histograms, climatology is biased towards the lower ranks, ie. climatology is unable to capture the low rainfall values in the prediction period. The ensemble histograms are relatively flat over the middle region, but biased towards the lowest

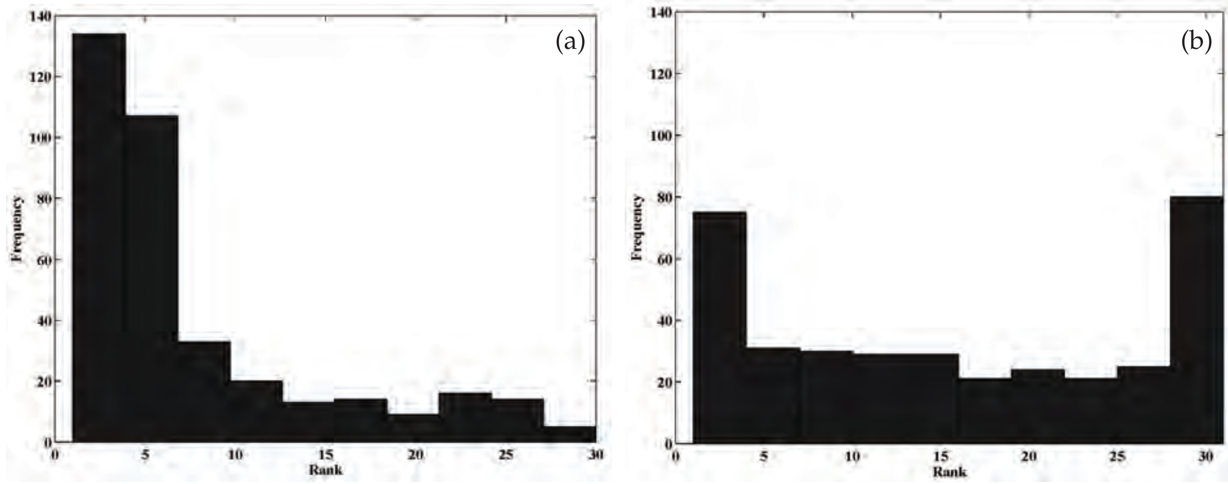

Figure 11: Rank histograms of (a) climatology and (b) ensemble predictions for lead time 10 days.

and highest ranks which indicates the difficulty in capturing the extremes. Even so, the overall performance of ensemble prediction is better when compared to the climatology as the reference forecast.

5 CONCLUSIONS

The recent interest in non-linear dynamics and also chaos theory has drawn attention towards considering streamflow as a chaotic system which is much sensitive to initial conditions and short term predictability. The present study was aimed at analysing the chaotic nature of streamflow series using different techniques. The daily streamflow data at Basantpur station of Mahanadi basin, India, for the period June 1972 to May 2004 is used for the study.

The positive Lyapunov exponents of the three regions confirm the unpredictability of the systems. The predictability of daily streamflow series is limited to only 7 days. The behaviour of streamflow dynamics was investigated using correlation dimension method with GPA. The clear scaling region in the $C(r)$ versus r plot on a log-log scale and also attaining a correlation exponent saturation value of 5.21 indicate a low dimensional chaotic behaviour of the streamflow series.

Since coloured random noises also exhibit a finite correlation dimension value, the above method is repeated on phase randomised data and on first derivative of the streamflow series. The correlation dimensions of phase randomised data are not converging, while those of first derivative are almost same as of the original data. This elucidates that the saturation of correlation dimension is not due to the inherent linear correlation in the data; but because of the low dimensional chaotic dynamics present in the data. Since one should not confirm the chaotic nature based on the correlation dimension method alone, FNN method is also employed to determine the optimum embedding dimension. The fraction of false nearest neighbours is falling to a minimum value at an embedding dimension of 7, which indicates that the optimum embedding dimension of the streamflow series is 7. These results suggest that the seemingly irregular behaviour of streamflow process can be better explained though a chaotic framework than through a stochastic representation. The chaotic nature of streamflow series is modelled through a local polynomial approach and prediction is done for various lead times. The skill of the ensemble predictions are compared with the climatology using rank probability skill score and rank histograms. The local polynomial predictions have shown an overall skilled performance than climatology reference. It has been observed that the ensemble approach using local polynomial model is able to capture the dynamics of chaotic streamflow series reasonably well for various lead times.

REFERENCES

Apipattanavis, S., Rajagopalan, B. & Lall, U. 2010, "Local Polynomial-Based Flood Frequency Estimator for Mixed Population", *J. Hydrol. Eng.*, Vol. 15, No. 9, pp. 680-691, doi: 10.1061/(ASCE)HE.1943-5584.0000242.

Dhanya, C. T. & Kumar, D. N. 2010, "Nonlinear ensemble prediction of chaotic daily rainfall", *Advances in water resources*, Vol. 33, pp. 327-347.

Dhanya, C. T. & Kumar, D. N. 2011a, "Multivariate Nonlinear Ensemble Prediction of Daily Chaotic Rainfall with Climate Inputs", *Journal of Hydrology*, Vol. 403, No. 3-4, pp. 292-306.

Dhanya, C. T. & Kumar, D. N. 2011b, "Predictive Uncertainty of Chaotic Daily Streamflow using Ensemble Wavelet Networks Approach", *Water Resources Research*, Vol. 47, No. 6, pp. 327-347, doi: 10.1029/2010WR01017333.

Elshorbagy, A., Simonovic, S. P. & Panu, U. S. 2002, "Noise reduction in chaotic hydrologic time series:

facts and doubts", *J. Hydrol.*, Vol. 256, No. 3/4, pp. 845-848.

Farmer, J. D. & Sidorowich, J. J. 1987, "Predicting chaotic time series", *Phys. Rev. Lett.*, Vol. 59, pp. 845-848.

Frazer, A. M. & Swinney, H. L. 1986, "Independent coordinates for strange attractors from mutual information", *Phys. Rev. A*, Vol. 33, No. 2, pp. 1134-1140.

Grassberger, P. & Procaccia, I. 1983a, "Measuring the strangeness of strange attractors", *Physica D*, Vol. 9, pp. 189-208.

Grassberger, P. & Procaccia, I. 1983b, "Estimation of the Kolmogorov entropy from a chaotic signal", *Phys. Rev. A*, Vol. 28, pp. 2591-2593.

Hegger, R. & Kantz, H. 1999, "Improved false nearest neighbor method to detect determinism in time series data", *Phys. Rev. E*, Vol. 60, pp. 4970-4973.

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

Islam, M. N. & Sivakumar, B. 2002, "Characterization and prediction of runoff dynamics: a nonlinear dynamical view", *Adv. Wat. Resour.*, Vol. 25, pp. 179-190.

Jayawardena, A. W. & Gurung, A. B. 2000, "Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach", *J. Hydrol.*, Vol. 228, pp. 242-264.

Jayawardena, A. W. & Lai, F. 1994, "Analysis and prediction of chaos in rainfall and streamflow time series", *J. Hydrol.*, Vol. 153, pp. 23-52.

Kantz, H. 1994, "A robust method to estimate the maximal Lyapunov exponent of a time series", *Phys. Lett. A*, Vol. 185, pp. 77-87.

Kantz, H. & Schreiber, T. 2004, *Nonlinear Time Series Analysis*, 2nd edition, Cambridge University Press, Cambridge, UK.

Kennel, M. B., Brown, R. & Abarbanel, H. D. I. 1992, "Determining embedding dimension for phase space reconstruction using a geometric method", *Phys. Rev. A*, Vol. 45, pp. 3403-3411.

Lee, T. & Ouarda, T. B. M. J. 2010, "Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes", *Journal of Geophysical Research*, Vol. 115, No. D13107, doi:10.1029/2009JD012801.

Liu, Q., Islam, S., Rodriguez-Iturbe, I. & Le, Y. 1998, "Phase-space analysis of daily streamflow: characterization and prediction", *Adv. Wat. Resour.*, Vol. 21, pp. 463-475.

Loader, C. 1999, *Local regression and likelihood*, Springer, New York.

Lorenz, E.N. 1972, "Predictability: Does the flap of a Butterfly's wings in Brazil set off a Tornado in Texas?", AAAS section on environmental Sciences New Approached to Global weather: GARP (The Global Atmospheric research Program), 139th meeting, American Association for the Advancement of Science.

Lucent Technologies, 2001, "LOCFIT: Local Regression and Likelihood", http://cm.bell-labs.com/cm/ms/departments/sia/project/locfit.

Osborne, A. R. & Provenzale, A. 1989, "Finite correlation dimension for stochastic systems with power law spectra", *Physica D*, Vol. 35, pp. 357-381.

Porporato, A. & Ridolfi, L. 1996, "Clues to the existence of deterministic chaos in river flow", *Int. J. Mod. Phys. B*, Vol. 10, No. 15, pp. 1821-1862.

Porporato, A. & Ridolfi, L. 1997, "Nonlinear analysis of river flow time sequences", *Water Resour. Res.*, Vol. 33, No. 6, pp. 1353-1367.

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

Puente, C. E. & Obregon, N. 1996, "A deterministic geometric representation of temporal rainfall: results for a storm in Boston", *Water Resour. Res.*, Vol. 32, No. 9, pp. 2825-2839.

Regonda, S. K., Rajagopalan, B., Lall, U., Clark, M. & Moon, Y.-I. 2005, "Local polynomial method for ensemble forecast of time series", *Nonlinear Processes in Geophysics*, Vol. 12, pp. 397-406.

Rodriguez-Iturbe, I., de Power, F. B., Sharifi, M. B. & Georgakakos, K. P. 1989, "Chaos in rainfall", *Water Resour. Res.*, Vol. 25, No. 7, pp. 1667-1675.

Rosenstein, M. T., Collins, J. J. & De Luca, C. J. 1993, "A practical method for calculating largest Lyapunov exponents from small data sets", *Physica D*, Vol. 65, pp. 117-134.

Sangoyomi, T., Lall, U. & Abarbanel, H. D. J. 1996, "Nonlinear dynamics of the Great Salt Lake: dimension estimation", *Water Resour. Res.*, Vol. 32, No. 1, pp. 149-159.

Shang, P., Xu, N. & Kamae, S. 2009, "Chaotic analysis of time series in the sediment transport phenomenon", *Chaos, Solitons & Fractals*, Vol. 41, No. 1, pp. 368-379.

Sivakumar, B. 2001, "Rainfall dynamics at different temporal scales: A chaotic perspective", *Hydrol. Earth System Sci.*, Vol. 5, No. 4, pp. 645-651.

Sivakumar, B., Liong, S. Y., Liaw, C. Y. & Phoon, K. K. 1999, "Singapore rainfall behavior: chaotic?", *J. Hydrol. Eng.*, Vol. 4, No. 1, pp. 38-48.

Sivakumar, B., Berndtsson, R., Olsson, J. & Jinn, K. 2001, "Evidence of chaos in the rainfall-runoff process", *Hydrol. Sci. J.*, Vol. 46, No. 1, pp. 131-145.

Smith, L. A., Ziehmann, C. & Fraedrich, K. 1998, "Uncertainty dynamics and predictability in chaotic systems", *Q.J.R. Meteorol. Soc.*, Vol. 125, pp. 2855-2886.

Takens, F. 1981, "Detecting strange attractors in turbulence", *Lectures Notes in Mathematics*, Rand, D. A. & Young, L. S. (editors), Springer-Verlag,Berlin, Germany.

Theiler, J., Eubank, S., Longtin, A., Galdikian, B. & Farmer, J. D. 1992, "Testing for nonlinearity in time series: The method of surrogate data", *Physica D*, Vol. 58, pp. 77-94.

Tsonis, A. A. & Elsner, J. B. 1988, "The weather attractor over very short timescales", *Nature*, Vol. 333, pp. 545-547.

Wang, Q. & Gan, T. Y. 1998, "Biases of correlation dimension estimates of streamflow data in the Canadian prairies", *Water Resour. Res.*, Vol. 34, No. 9, pp. 2329-2339.

Wilks, D. S. 2005, *Statistical Methods in the Atmospheric Sciences – An Introduction*, Academic Press, Inc.

Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, A. 1985, "Determining Lyapunov exponents from a time series", *Physica D*, Vol. 16, pp. 285-317.

C DHANYA

Dr C. T. Dhanya is an Assistant Professor in the Department of Civil Engineering, Indian Institute of Technology Delhi, New Delhi, India. Her research interests include climate hydrology, climate change, non-linear dynamics and chaos theory, and applications of data mining algorithms in water resources engineering.

D NAGESH KUMAR

Prof D. Nagesh Kumar is a Professor in the Department of Civil Engineering, Indian Institute of Science, Bangalore. Earlier he worked at IIT, Kharagpur, and NRSC, Hyderabad, India. His research interests include climate hydrology, climate change, water resources systems, artificial neural networks, evolutionary algorithms, fuzzy logic, and remote sensing and geographic information systems applications in water resources engineering. He has co-authored two text books titled *Multicriterion Analysis in Engineering and Management*, published by PHI, New Delhi, and *Floods in a Changing Climate: Hydrologic Modeling*, published by Cambridge University Press, UK.