Journal of Civil Engineering and Urbanism

Volume 2, Issue 2: 51-55 (2012) (Received: November 23, 2011; Accepted: February 16, 2012; Published: March 30, 2012) ISSN-2252-0430

Local Prediction in River Discharge Time Series

Mohammad Ali Ghorbani[1*], Rasoul Daneshfaraz[2], Hadi Arvanagi[1], Ali Pourzangbar[3], Seyed Mahdi Saghebian[4], Shahab Kaveh Kar[1]

[1] *Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran*
[2] *Department of Civil Engineering, Faculty of Engineering, University of Maragheh, Maragheh, Iran*
[3] *Department of Civil Engineering, Iran University of Science & Technology, Narmak, Tehran, Iran*
[4] *Department of Civil Engineering, Islamic Azad University, Ahar Branch, Ahar, Iran*

*Corresponding author's Email: ghorbani@tabrizu.ac.ir

ABSTRACT: In this paper, chaotic behavior of the daily river discharge time series from the Karoon River, during January 1999-December 2004 is investigated. The phase space, which describes the evolution of the behavior of a nonlinear system, is reconstructed using the delay embedding theorem suggested by TAKENS. The delay time used for the reconstruction is chosen after examining the first minimum of the average mutual information (AMI) of the data. It is found that a delay time of 40 days and the sufficient embedding dimension is estimated using the false nearest neighbor algorithm which has a value of 8 for the river flow time series. Based on these embedding parameters we calculate the average divergence rate of nearby orbits given by the largest Lyapunov exponent. The largest Lyapunov exponent 0.0255 for is estimated. In this study the local prediction model has been applied to predict daily discharge time series. In this prediction model, the dynamics of the system are described step by step locally in the phase space, the results are quite satisfactory.

Keywords: Chaos theory; Karoon River; Local Prediction; Lyapunov exponent; Time series

ORIGINAL ARTICLE

INTRODUCTION

Study of river flow is important for designing, exploitation and study of water supply systems. River flow processes is dynamic, nonlinear, extremely complex, and are affected by several interconnected physical variables, so that ddifferent methods including hydrologic modeling, time series analysis, artificial neural networks, fuzzy logic, neuro-fuzzy, genetic programming and recently chaos theory are used for river flow modeling.

Inspite of the previous study on river flow have essentially employed the concept of a stochastic process, recent studies have indicated that even simple deterministic systems, influenced by a few nonlinear interdependent variables, might give rise to very complicated structures (i.e. deterministic chaos). Therefore, it is now believed that the dynamic structures of the seemingly complex processes, such as river flow variations, might be better understood using nonlinear deterministic chaotic models than the stochastic ones. The investigation of the existence of chaos in hydrological processes has been of much interest lately. The outcomes of the investigations are very encouraging as they provided evidence regarding the existence of low-dimensional chaos implying the possibility of accurate of modeling and short-term predictions.

A chaotic system is defined as a deterministic system in which small changes in the initial conditions may lead to completely different behavior in the future. Signal from the chaotic system is often, at first sight, indistinguishable from a random process, despite being sensitive to initial conditions) behaviour of many systems was observed by many researchers for a number of decades, but was first described as such by Lorenz (Wilks (1991)). During the past two decades, the theory of chaos showed its applicability in solving a wide class of problems in many areas of natural sciences. The discovery that very simple deterministic systems can produce seemingly irregular time series pushed researchers to try identifying such systems and apply chaos theory in order to predict their behaviour. However, chaotic signal analysis is still a novel approach in many areas related to civil engineering and to water-related problems in particular. In literature, many researchers have investigated the stream flow modelling with chaos theory. The papers by Jayawardena & Lai (1994); Porporato & Ridolfi (1997); Krasovskaia *et al.* (1999); Stehlik (1999); Sivakumar *et al.*(2002) have shown the presence of low-dimensional deterministic behaviour in the stream flow process . Islam & Sivakumar (2002), Lisi & Villi (2001), Liu *et al.* (1998) have suggested the possibility of accurate stream flow predictions using nonlinear deterministic approaches. Elshorbagy *et al.*(2002) has performed noise reduction and missing data estimation Qingfang & Yuhua (2007) has developed a new local linear prediction model for chaotic stream flow series.

METHODOLOGY

Reconstruction of phase space

To cite this paper: Ghorbani, M.A., Daneshfaraz, R., Arvanagi, H., Pourzangbar, A., Saghebian, SM., Kaveh Kar, Sh. 2012. Local Prediction in River Discharge Time Series. *Journal. Civil Eng. Urban.* 2(2): 51-55.
Journal homepage: http://www.ojceu.ir/main/

The first step in the process of chaos theory is reconstructing the dynamics in phase space. The concept of phase-space is a powerful tool for characterizing dynamic system, because with a model and a set of appropriate variables, dynamics can represent a real-world system as the geometry of a single moving point .A method for reconstructing phase-space from a sight time series has been presented by Takens (1981).The time series is assumed to be generated by a nonlinear dynamic system with m degrees of freedom. It is therefore necessary to construct an appropriate series of state vectors Y_t with delay coordinates in the m-dimensional phase space:

$$Y_t = \{X_t, X_{t-\tau}, X_{t-2\tau}, ..., X_{t-(m-1)\tau}\} \qquad (1)$$

where τ is referred to as the delay time and for a digitized time series is a multiple of the sampling interval used, while m is termed the embedding dimension. If the dynamics of the system can be reduced to a set of deterministic laws, the trajectories of the system converge towards the subset of the phase space, called the attractor.

The time delay τ can be defined by means of an autocorrelation function or, as used in this study, the average mutual information method (Fraser & Swinney, 1986). This method defines how the measurements $X(t)$ at time t are connected in an information theoretic fashion to measurements $X(t+\tau)$ at time $t+\tau$ □ (Abarbanel, 1996). The average mutual information is defined as:

$$I(\tau) = \sum_{X(i),X(i+\tau)} P(X(i), X(i+\tau)) \log_2 \left[\frac{P(X(i), X(i+\tau))}{P(X(i))\, P(X(i+\tau))} \right] \qquad (2)$$

Where i is total number of samples. $P(X(i))$ and $P(X(i+\tau))$ are individual probabilities for the measurements of $X(i)$ and $X(i+\tau)$. $P(X(i), X(i+\tau))$ is the joint probability density for measurements $P(X(i))$ and $P(X(i+\tau))$. The appropriate time delay τ is defined as the first minimum of the average mutual information $I(\tau)$. Then the values of $X(i)$ and $X(i+\tau)$ are independent enough of each other to be useful as coordinates in a time delay vector but no so independent as to have no connection with each other at all.

A technique to estimate the optimal embedding dimension m is by looking for false neighbours in phase space. The False Nearest Neighbour (FNN) method proposed by Kennel *et al.* (1992) was used to determine the minimal sufficient embedding dimension m.

Lyapunov exponents

Another technique to determine the presence of chaotic behaviour is the largest Lyapunov exponent, which measures the divergence of nearby trajectories in the phase space. Thus, a positive Lyapunov exponent is a strong indicator of chaos. The largest Lyapunov exponent λ_l is defined as:

$$\lambda_1 = \frac{1}{t_M - t_0} \sum_{k=1}^{M} \log_2 \frac{L'(t_k)}{L(t_{k-1})} \qquad (3)$$

Where M is the number of replacement steps, $L(t_{k-1})$ is the Euclidean distance between the point $\{X(t_{k-1}), X(t_{k-1}-\tau), ..., X(t_{j-1-(m-1)}\tau)\}$ and its nearest neighbour, and $L'(t_k)$ is the evolved length of $L(t_{k-1})$ at time t_k.

Before computing the largest Lyapunov exponent, the dimension m of the phase space has to be determined. The inverse of largest Lyapunov exponent ($1/\lambda_l$) determines the average horizon of predictability for the system (Rosenstein *et al.*, 1993).

Local prediction

A correct phase-space reconstruction in a dimension m facilities an interpretation of the underlying dynamics in the form of an m-dimensional map f_T, according to $Y_{J+T} = f_T(Y_j)$ (4)

Where Y_j and Y_{j+T} are vectors of dimension m, describing the state of the system at times j (e.g. current state) and $j+T$ (e.g. future state), respectively. The problem then is to find an appropriate expression for f_T (i.e. F_T). Local approximation entails the subdivision of the f_T domain into many subsets (neighbourhoods), each of which identifies some approximations F_T, valid only in that same subset. In other words, the dynamics of the system is described step by step locally in the phase-space. By considering a time series of a single variable, it is possible to reconstruct the phase space. Before applying reconstruction procedure it is necessary to have some information, embedding dimension, delay time, etc., concerning the attractor .One of the independent coordinates mentioned above is taken as the time series itself. The remaining coordinates are formed by its ($m-1$) lagged time series shifted by ($m-1$) multiples of the correlation time τ, at which correlation between coordinates become zero. It is assumed that the time series data are generated from a chaotic dynamical system in the ν-dimensional space (ν is dimension of attractor). In this m-dimensional space, prediction is performed by estimating the change of X_i with time. Considering the relation between the points X_t and X_{t+p} at time p later on the attractor is approximated by function F as

$$X_{t+p} \cong F(X_t) \qquad (5)$$

In this prediction method, the change of X_t with time on the attractor is assumed to be the same as those of nearby points, $(X_{T_h}, h = 1,2,...,n)$. Herein, X_{t+p} is determined by the dth order polynomial $F(X_t)$ as fallows

$$x_{t+p} \cong f_0 + \sum_{k_1=0}^{m-1} f_{1k_1} X_{t-k_1\tau} + \sum_{\substack{k_2=k_1 \\ k_1=0}}^{m-1} f_{2k_1k_2} X_{t-k\tau_1} X_{t-k_2\tau} + ... + \sum_{\substack{k_d=k_{d-1} \\ \vdots \\ k_2=k_1 \\ k_1=0}}^{m-1} f_{dk_1k_2...k_d} X_{t-k_1\tau} X_{t-k_2\tau} ... X_{t-k_d\tau} \qquad (6)$$

Using n of X_{T_h} and $\boldsymbol{X}_{T_{h+p}}$ for which the values are already known, the coefficients f are determined by solution of the following equation:

$$x \cong Af \qquad (7)$$

where

$$\boldsymbol{x} = (\boldsymbol{x}_{T_{1+p}}, \boldsymbol{x}_{T_{2+p}}, ..., \boldsymbol{x}_{T_{n+p}}) \qquad (8)$$

$$f = (f_0, f_{10}, f_{11}, ..., f_{1(m-1)}, f_{200}, ..., f_{d(m-1)(m-1)...(m-1)}) \qquad (9)$$

and A is the $n \times (m+d)!/m!d!$ Jacobian matrix which in its explicit form is

$$A=\begin{bmatrix} x_{T_1} & x_{T_{1-\tau}} & \cdots & x_{T_{1-(m-1)\tau}} & x_{T_1}^2 & \cdots & x_{T_{1-(m-1)\tau}}^d \\ x_{T_2} & x_{T_{2-\tau}} & \cdots & x_{T_{2-(m-1)\tau}} & x_{T_2}^2 & \cdots & x_{T_{2-(m-1)\tau}}^d \\ \vdots & & \vdots & & \vdots & & \vdots \\ x_{T_n} & x_{T_{n-\tau}} & \cdots & x_{T_{n-(m-1)\tau}} & x_{T_n}^2 & \cdots & x_{T_{n-(m-1)\tau}}^d \end{bmatrix} \quad (10)$$

In order to obtain a stable solution, the number of rows in the Jacobian matrix A must satisfy the relation

$$n \geq \frac{(m+d)!}{m!d!} \quad (11)$$

As stated by *Porporato & Ridolfi (1997),* even though in the case F are first degree polynomials, the prediction is nonlinear, because during the prediction procedure every point $x(t)$ belongs to a different neighbourhood and is therefore defined by different expressions for f (Kocak,1997).

STUDY AREA AND DATA USED

Karoon River, which has a watershed area of 58,180 km^2 and is located in southwest of the I.R. of Iran in Khuzestan province is chosen for this study. The river lies between the city of Ahwaz (31^o 20' N, 48^o 41' E) and the Bahmanshir River (30^o 25' N, 48^o 12' E), which is about 190 km in length. The Karoon river is a meandering river which supplies water for the irrigation of sugarcane cultivation projects, as well as other agricultural lands. Near the Persian Gulf, it splits into two rivers, the Bahmanshir and the Arvand. These two rivers flow into the Persian Gulf (Fig.1). For the present investigation river flow data observed over a period of 6 year (January 1999-December 2004) are considered. Fig. 2 shows the variations of daily river flow time series and Table 1 presents some of the important statistics of the time series.

Figure 1. Location of the Karoon River

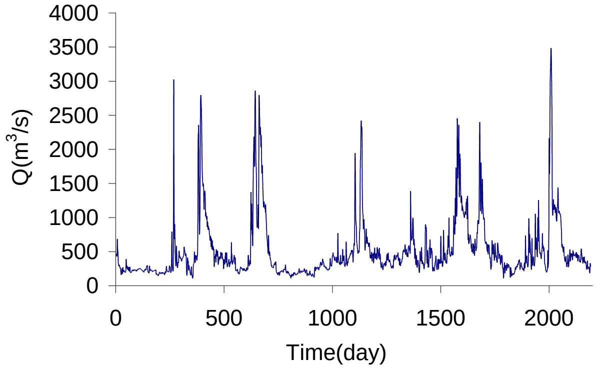

Figure 2. Daily river time series at the Karoon River (1999-2004)

ANALYSIS, RESULTS AND DISCUSSION

Determination of reconstruction parameters

In order to reconstruct the original phase space, we first estimate reconstruction parameters, the delay times τ and embedding dimension m. We calculate AMI using time lags of 1-100 days. The time series exhibit AMI shows well-defined first minima at time lag 40 days (Fig. 3). The method used for the determination of the sufficient embedding dimension is based on the calculation of the percentage of false nearest-neighbors for the time series. For the rest of the data considered the application of the method shows that the estimation value of embedding dimension is 8 (Fig. 4).

Table 1. Statistics of daily river flow data from Karoon River

Statistic	Daily river flow (m^3/s)
Number of Data	2192
Mean	519.517
Standard deviation	464.865
Maximum value	3485.83
Minimum value	107
Skewness	2.8134
Kurtosis	9.3953

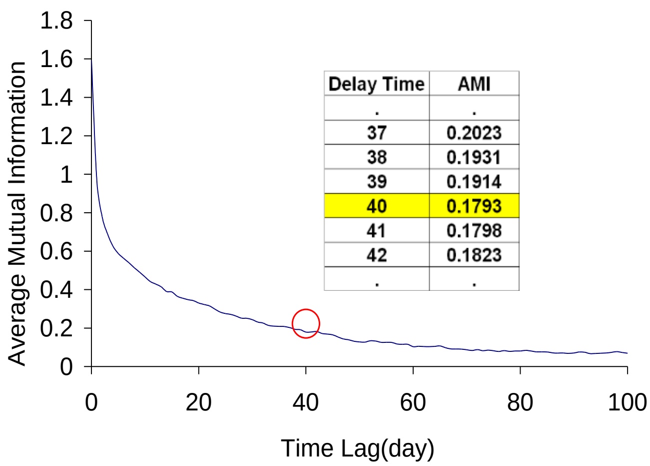

Figure 3. Mutual information function of daily discharge time series

To cite this paper: Ghorbani, M.A., Daneshfaraz, R., Arvanagi, H., Pourzangbar, A., Saghebian, SM., Kaveh Kar, Sh. 2012. Local Prediction in River Discharge Time Series. *Journal. Civil Eng. Urban.* 2(2): 51-55.
Journal homepage: http://www.ojceu.ir/main/

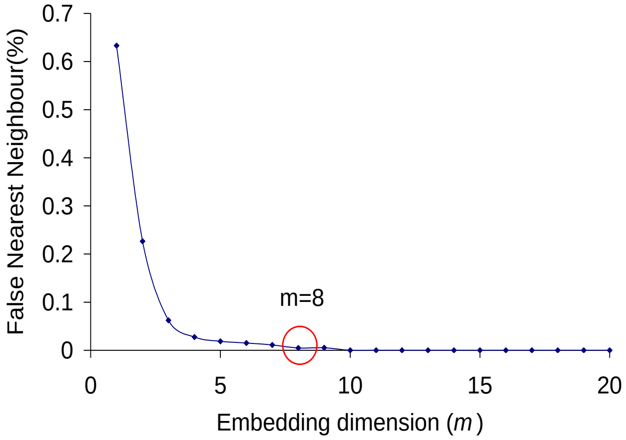

Figure 4. Percentage of false nearest neighbour in embedding dimension of daily time series

Estimation of the largest Lyapunov exponent

The curves for the stretching factor λ_l versus the number of points *N* show the expected linear increase/flat regions (Fig.5) with some fluctuations Superimposed on the linear part of the curve. The slope value corresponding to the largest Lyapunov exponent is obtained after the least-squares line fit for the discharge series and is found to be 0.0255.

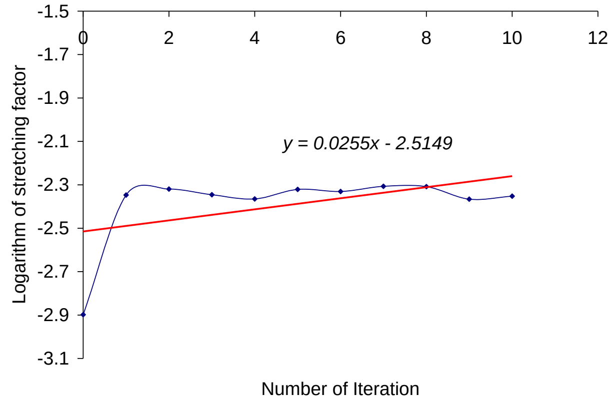

Figure 5. Estimation of the Largest Lyapunov exponent using the method of Rosenstein *et al*. (1993) of daily time series

Local Prediction

In this study, the entire data set of 6 years is divided into two parts; the first 5 years of data are used in the phase space reconstruction and predictions are made for the subsequent 1 year (2003-2004) of data.

Table 2 shows values of R^2 and *RMSE* for different Embedding dimensions in prediction. Overall reasonably good predictions, with R^2> 0.91 is achieved for all 10 embedding dimensions. However, a closer look at the statistics reveals that the best predictions are achieved when the embedding dimension is m_{opt}=3 for daily discharge time series. Fig.6 presents a comparison of the actual discharge values and the predicted ones. What is more encouraging is that even minor fluctuations present in the actual series are very well captured by the nonlinear prediction technique. Fig.7 shows the Scatter plots of observed and calculated values. Such results certainly indicate the appropriateness of the phase-space-based nonlinear prediction technique, employed herein, to understand, model and predict the discharge at the Karoon River.

Table 2. Values of R^2 and RMSE for different embedding dimensions in prediction processes

Embedding Dimension (*m*)	R^2	RMSE(m^3/s)
1	0.9331	137.61
2	0.932	137.56
3	0.9343	135.19
4	0.9197	149.64
5	0.9122	156.46
6	0.9111	157.34
7	0.9147	154.07
8	0.9173	153.93
9	0.9326	137.04
10	0.9233	147.44

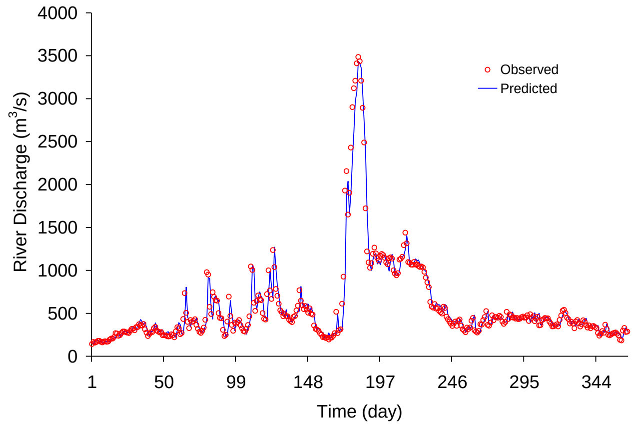

Figure 6. Comparison between time series plots of predicted and observed values

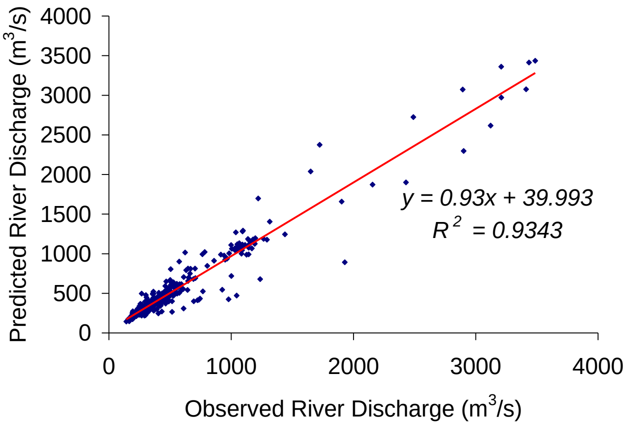

Figure 7. Scatter plot of observed and predicted values of daily time series

CONCLUSION

This paper describes a series of analytic techniques for discerning and investigating chaotic behaviours in the discharge dynamics. We have analysed daily discharge over 6 years (1999-2004) in the Karoon River, using the techniques based on phase space reconstruction. In the study, TISEAN package (Hegger *et al.*, 1999) has been used to calculate the mutual information function and the false nearest neighbour dimension. The phase space of the discharge series is reconstructed using embedding parameters. These are the delay time and the embedding dimension which were calculated as 40,8 for discharge data, respectively. The results have shown that chaotic characteristics obviously exist in the discharge due to the positive largest Lyapunov exponent 0.0255.

In This study the local prediction model has been applied to discharge. In this prediction model, the dynamics of the system are described step by step locally in the phase space. The predicted values are in good agreement with the observations.

REFERENCES

1. Wilks, D.S. (1991). Representing serial correlation of meteorological events and forecasts in dynamic decision-analytic models. Monthly Weather Review .119:1640-1662.
2. Jayawardena, A.W. & Lai, F. (1994). Analysis and prediction of chaos in rainfall and stream flow time series, Journal of Hydrology, 153, 23-52.
3. Porporato, A. & Ridolfi L.(1997).Nonlinear analysis of river flow time sequences, Water Resources Research, 33(6), 1353-1367.
4. Krasovskaia I., Gottsehalk L. & Kundzewicz, Z.W.(1999). Dimensionality of Scandinavian river flow regimes.Hydrol. Sci. J. 44 (5), 705-723.
5. Stehlík, J.(1999). Deterministic chaos in runoff series. Journal of Hydrology and Hydrodynamics, 47(4), 271-287.
6. Stehlík, J. (1999). Deterministic chaos in runoff series. Journal of Hydrology and Hydrodynamics, 47(4), 271-287.
7. Sivakumar, B., Jayawardena, A.W. & Fernando, T.M.G.H. (2002). River flow forecasting: Use of phase-space reconstruction and artificial neural networks approaches, Journal of Hydrology.265 (1-4), 225-245.
8. Islam, M.N. & Sivakumar, B. (2002). Characterization and prediction of runoff dynamics: a nonlinear dynamical view .Advances in Water Resources.25, 179-190.
9. Lisi, F. & Villi, V. (2001).Chaotic forecasting of discharge time series: A case study. Journal of the American Water Resources Association. 37(2), 271-279.
10. Liu, Q., Islam, S., Rodriguez-lturbe, I. & Lee, Y. (1998). Phase-space analysis of daily stream flow: characterization and prediction. Adv. Wat. Resour. 21, 463-475.
11. Elshorbagy, A., Simonovic, S.P. & Panu, U.S.(2002a).Noise reduction in chaotic hydrologic time series: facts and doubts, Journal of Hydrology, 256, 147-165.
12. Elshorbagy, A., Simonovic, S.P. & Panu, U.S. (2002b). Estimation of missing stream flow data using principles of chaos theory, Journal of Hydrology, 255, 123-133.
13. Qingfang, M. & Yuhua, P.(2007).A new local linear prediction model for chaotic time series", Physics Letters A. 370: 465-470.
14. Takens, F.(1981).Detecting strange attractors in turbulence, in Lectures Notes in Mathematics, edited by D. A. Rand and L. S. Young, 898, 366-381, Springer-Verlag, New York.
15. Fraser, A. M. & Swinney, H.L. (1986). Independent coordinates for strange attractors from mutual information. Physical Review A, 33(2), 1134-1140.
16. Abarbanel, H.D.I. (1996). Analysis of observed chaotic data. Springer – Verlag, New York, 272 pp.
17. Kennel, M., Brown, R. & Abarbanel, H.D.I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A .45:3403–3411.
18. Rosenstein, M.T., Collins, J.J. & deluca C.J. (1993). A practical method for the calculating largest Lyapunov exponents from small datasets. Physica D; 65:117-34.
19. Koçak, K. (1997). Application of Local Prediction Model to Water Level Data". A satellite Conference to the 51st ISI Session in Istanbul, Turkey: Water and Statistics, Ankara-Turkey, pp: 185-193.
20. Hegger, R., Kantz, H. & Schreiber, T. (1999). Practical implementation of nonlinear time series methods: The TISEAN package.Chaos. 9:413-435.