Hydrol. Earth Syst. Sci. Discuss., 10, 14331–14354, 2013
www.hydrol-earth-syst-sci-discuss.net/10/14331/2013/
doi:10.5194/hessd-10-14331-2013

Hydrology and Earth System Sciences
Discussions

Downstream prediction using a nonlinear prediction method

N. H. Adenan[1] **and M. S. M. Noorani**[2]

[1]Universiti Pendidikan Sultan Idris, Perak, Malaysia
[2]Universiti Kebangsaan Malaysia, Selangor, Malaysia

Received: 1 October 2013 – Accepted: 6 November 2013 – Published: 22 November 2013

Correspondence to: N. H. Adenan (nurhamiza.adenan@gmail.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The estimation of river flow is significantly related to the impact of urban hydrology, as this could provide information to solve important problems, such as flooding downstream. The nonlinear prediction method has been employed for analysis of four years of daily river flow data for the Langat River at Kajang, Malaysia, which is located in a downstream area. The nonlinear prediction method involves two steps; namely, the reconstruction of phase space and prediction. The reconstruction of phase space involves reconstruction from a single variable to the m-dimensional phase space in which the dimension m is based on optimal values from two methods: the correlation dimension method (Model I) and false nearest neighbour(s) (Model II). The selection of an appropriate method for selecting a combination of preliminary parameters, such as m, is important to provide an accurate prediction. From our investigation, we gather that via manipulation of the appropriate parameters for the reconstruction of the phase space, Model II provides better prediction results. In particular, we have used Model II together with the local linear prediction method to achieve the prediction results for the downstream area with a high correlation coefficient. In summary, the results show that Langat River in Kajang is chaotic, and, therefore, predictable using the nonlinear prediction method. Thus, the analysis and prediction of river flow in this area can provide river flow information to the proper authorities for the construction of flood control, particularly for the downstream area.

1 Introduction

Urbanization and urban growth are essential factors for planners and policy makers because the urbanization pattern has major implications for the hydrological processes. Urbanization can have various effects on certain hydrological problems, such as flood prevention in urban areas, allocation of adequate resources in terms of water quality and quantity, and waterborne waste disposal (Hall, 1984). Thus, the problems of

urban hydrology involve flood and pollution prevention. However, our study only considers flood prevention based on the river flow prediction. If an undeveloped area is then transformed into a developed area, the conditions of the soil structure will be disturbed (Hall, 1984). These factors can change the magnitude of the river flow. The volume of runoff will increase significantly with the increase in the magnitude of the river flow due to the impervious areas and the lack of drainage. Hence, downstream flooding problems exist in urban areas. There are several methods that can be used to estimate the river flow in a watershed that is located in an urban area, such as the empirical and physical process methods. Referring to the empirical method for urban hydrology research, the behaviour of river flow in the downstream area is important to provide accurate information for the whole river flow (Viesmann and Lewis, 1996). This information can help in planning, development and flood prevention of the downstream area.

The Langat River, which is one of the longest rivers in the state of Selangor, Malaysia, is used as a case study. This research focuses on the downstream area at Kajang, which is well-known for experiencing flood hazards. Figure 1 shows the four gauging stations along the Langat River. The Langat River flows from east to southeast, which is from Lui River to Kajang. The total length of the upstream and downstream is about 34.4 km and the downstream area has been identified as a flood risk area (Mohammed et al., 2011). Checkpoint 1 (station number 3118445) is located at the Lui River gauging station (upstream) and Checkpoint 2 (station number 2917401) is located at the Kajang gauging station (downstream). The Langat River at the Kajang gauging station has been used for the river flow analysis and prediction using the nonlinear prediction method. This area had a population of 229 655 people in 2000, which increased to 342 657 people in 2010 (Department of Irrigation and Drainage Malaysia, 2005). The increase in population in this area reflects the development in the Kajang area. Furthermore, the study area is adjacent to an industrial area and pig farms. Flooding in this area can cause damage to the industrial area and pollution in the Langat River basin. Thus, studies of the downstream area (Kajang) are important to provide information

about the flow downstream. This study was conducted at this point so that the release of water from Checkpoint 2 could be estimated for a certain length of time. The results of this study could help to identify the preventive measures that could be undertaken in this downstream area.

The analysis and prediction of river flow could provide the information about the dynamics of the river flow system. However, the flow of the river is not dependent on rainfall alone. The characteristics of an area, such as shape, slope, land, soil structure and climate change, can also affect the flow of the river in an area (Viesmann and Lewis, 1996). Thus, the application of stochastic methods is often used to analyse complex natural conditions, such as the river flow. Developments in the study of nonlinear time series analysis is growing with some revolutionary methods. One particular method that provides important findings is known as chaos theory, which explains that a complex system can be analysed by deterministic methods that use a minimum number of the system's variables (Islam and Sivakumar, 2002). Several decades ago, a number of studies were performed to obtain information on characterizing, modelling and predicting hydrological phenomena as a deterministic system (e.g. Jayawardena and Lai, 1993; Sivakumar, 2000; Ghorbani et al., 2010). The results showed that the river flow prediction and other hydrological processes are in good agreement with the actual data values (Sivakumar, 2003; Regonda et al., 2005; She and Yang, 2010; Khatibi et al., 2012). In addition, prediction using chaos theory can reveal the number of variables that affect the dynamics of the river flow.

Studies on river flow analysis and prediction in Malaysia have been done and improved for a variety of purposes, such as providing information for flood prevention. Several methods, such as support vector machine method (Shabri and Suhartono, 2012), neural network model (Ahmad and Juahir, 2006) and hydrodynamic modelling (Ghani et al., 2010), have been used for river flow prediction. However, several methods have yet to be explored for the purpose of river flow prediction in Malaysia, such as chaos theory, Bayesian methods and wavelet methods. River flow prediction using chaos theory involves a single variable (river flow data) albeit there are other dominant

variables affecting river flow prediction. Meanwhile, the Bayesian and Wavelet methods are dependent on a number of dominant variables, such as rainfall, temperature and soil type. To the best of our knowledge, this is the first attempt to use the for the analysis and prediction of river flow in Malaysia.

2 Nonlinear prediction method

The nonlinear prediction method (NLP) of chaos theory is used to analyse river flow and predict the future value of the flow. There are two steps in NLP – phase space reconstruction and prediction. Reconstruction of the phase space uses observed data (one-dimensional) to build the m-dimensional phase space that reflects the dynamics of the river flow (Abarbanel, 1996; Adenan and Noorani, 2013). A scalar time series $x(t)$ forms a one-dimensional time series:

$$\{x_i\} = \{x_1, x_2, x_3, \ldots, x_N\} \tag{1}$$

where N is the total number of points in the time series that can be transformed into m-dimensional vectors:

$$\boldsymbol{Y}_t = \left\{x_t, x_{t+\tau}, x_{t+2\tau}, \ldots, x_{t+(m-1)\tau}\right\} \tag{2}$$

where τ is an appropriate time delay and m is a chosen embedding dimension (Abarbanel, 1996; Tongal and Berndtsson, 2013).

Referring to Eq. (2), the value of τ and m are needed to reconstruct the phase space. In this study, τ has been predetermined. Selection of the appropriate τ is important during the reconstruction of the phase space. The most optimal value of τ can provide a separation of neighbouring projections with respect to the dimension of the phase space. If the value of τ is too small, the coordinates of the phase space cannot properly describe the dynamics of the system. Meanwhile, information on trajectories in the phase space will diverge if the value of τ is too big (Sangoyomi et al., 1996; Islam and

Sivakumar, 2002). The optimal value of m in phase space reconstruction can describe the topology of the attractor. The number of dimensions in the reconstructed phase space is equal to the number of columns in the matrix resulting from the embedding parameters in the time series. If the number of columns is insufficient, it cannot reflect the phase space dynamics of the system. Therefore, the selection of the preliminary parameter pair (τ, m) is important to reflect the dynamics of the phase space.

2.1 Determination of preliminary parameter pair (τ, m)

Previous studies on the river flow prediction showed that when a condition of time delay $\tau = 1$ is used in phase space reconstruction, the results gave good predictions (Sivakumar, 2002, 2003). Thus, in this study, the time delay $\tau = 1$ is used. The embedding dimension m is calculated using the correlation dimension and false nearest neighbour method (FNN). There are two models to be considered, Model I and Model II, which involve different combinations of preliminary parameter pairs for the reconstruction of the phase space. Model I involves $\tau = 1$, and m is the result of the calculation from the correlation dimension; and Model II involves the combination $\tau = 1$, and m is the result of the calculation of FNN. A comparison of the prediction results is conducted to distinguish the strength of the two models.

The correlation dimension method is the most fundamental method in the study of chaotic time series for proving the presence of chaotic behaviour in hydrological studies (Jayawardena and Lai, 1994; Martins et al., 2011; Khatibi et al., 2012). For a given distance r, the main idea of the correlation function $C(r)$ is related to the shortest distance of the vectors $\boldsymbol{Y}_t$. Here the Euclidean distance is used to calculate the distance between points on the vector space:

$$d_s(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = \|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|. \tag{3}$$

The correlation dimension is based on the correlation integral introduced by Grassberger (1986):

$$C(r) = \lim_{N\to\infty} \frac{2}{N(N-1)} \sum_{i,j=1}^{N} H(r - \|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|) \quad (4)$$

where H is the Heavyside function, which has the value 0 or 1 and can be defined as:

$$H(r - \|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|) = \begin{cases} 1,0 & \leq (r - \|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|) \\ 0,0 & > (r - \|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|) \end{cases} \quad (5)$$

and acts as a barrier to the Euclidean distance between two points on the attractor $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$. The correlation function $C(r)$ is calculated for the pair of points $(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ with a distance less than the radius r. In the limit to infinite amount of data ($N \to \infty$) and sufficiently small $r(r \to 0)$, the relation $C(r) \cong \alpha r^{D_2}$ is expected (Men et al., 2004). The correlation dimension D_2 and correlation exponent v can be defined as:

$$D_2 = \lim_{r\to 0} v \quad (6)$$

$$v = \frac{\delta[\log C(r)]}{\delta[\log r]} \quad (7)$$

Several steps are required to identify the value of the correlation dimension. The first step is to draw a graph $\ln C(r)$ vs. $\ln(r)$ with a given m. Then the gradient (correlation exponent v) of the m-dimensional curve values has to be determined. The gradient of the graph can be measured by the least squares method for determining the scaling. For finite data and where the value of r exceeds the diameter, there exists a saturated area of the graph. The saturated area is the scaling region. A better way to estimate the gradient is to use $\delta[\log C(r)]/\delta[\log r]$. To examine if there is a chaotic nature, the correlation exponent (slope v) vs. m-dimensional has to be plotted. If the value of the

correlation exponent is finite, low and non-integer, the system is considered to be of low dimensional chaotic nature (Men et al., 2004). If the correlation value increases without limit as m increases, the system should be studied as a stochastic system.

The false nearest neighbour method (FNN) is an effective method for finding the embedding dimension m for the reconstruction phase space. This method has been used to analyse river flow time series (Wu and Chau, 2010; Ghorbani et al., 2012). This paragraph describes how FNN is implemented. Suppose the dimension increases then the distance between the point and the nearest neighbour should not change if it is indeed the nearest neighbouring point. Computation of the distance between the point and the nearest neighbour is by the Euclidean distance.

FNN can be calculated using the following algorithm. Assume that $\boldsymbol{Y}_i = \left\{x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}\right\}$ has nearest neighbour $\boldsymbol{Y}_i^{\mathrm{NN}} = \left\{x_i^{\mathrm{NN}}, x_{i+\tau}^{\mathrm{NN}}, x_{i+2\tau}^{\mathrm{NN}}, \ldots, x_{i+(m-1)\tau}^{\mathrm{NN}}\right\}$. Then, calculate the Euclidean distance $\left\|\boldsymbol{Y}_i - \boldsymbol{Y}_i^{\mathrm{NN}}\right\|$. For all points i in vector space, equation $\frac{\left|x_{i+m\tau} - x_{i+m\tau}^{\mathrm{NN}}\right|}{\left\|\boldsymbol{Y}_i - \boldsymbol{Y}_i^{\mathrm{NN}}\right\|} > R_T$ is used and the value of false nearest neighbour can be calculated. R_T is a value between 10 and 30. In this study, the value of 15 is used (Wu and Chau, 2010). Repeat the algorithm with different embedding dimensions and the value of the false nearest neighbour that is close to zero is used as the embedding dimension.

2.2 Prediction

In this study, the prediction of river flow has been performed by using the local linear approximation method. This method was proposed by Lorenz (1969). Application of the local linear approximation method is to (1) examine whether the river flow at the downstream areas can be predicted, (2) to compare the prediction results for Models I and II. The local linear approximation method is used to predict river flow in downstream areas as follows. The first step is to reconstruct the phase space. The combination of the preliminary parameter pair (τ, m) is important for reconstruction of the phase space

because this phase space result will be used in making a prediction. The difference between Models I and II is in the reconstruction phase space. Models I and II involve $\tau = 1$ but involve different methods in determining the value of m. Model I uses the correlation dimension while FNN is employed for Model II. Assume that the reconstruction of phase space is like $\boldsymbol{Y}_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}\}$. The nearest neighbour for $\boldsymbol{Y}_t$ is required to predict $\boldsymbol{Y}_{t+1}$. Assume that the vector of the minimum distance to the nearest neighbour is $\boldsymbol{Y}_M$. Next, for the local linear approximation method, the values of $\boldsymbol{Y}_M$ and $\boldsymbol{Y}_{M+1}$ are used to satisfy the linear equations $\boldsymbol{Y}_{M+1} = A\boldsymbol{Y}_M + B$. The constants A and B are calculated using the least squares method. Thus, the predictive value $\boldsymbol{Y}_{t+1}$ can be calculated using $\boldsymbol{Y}_{t+1} = A\boldsymbol{Y}_t + B$.

2.3 Performance evaluation

The assessment of the prediction accuracy of the models for predicting the daily river flow is evaluated by using the mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (CC). The MAE, RMSE and CC are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|y_t^{\mathrm{o}} - y_t^{\mathrm{f}}\right| \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t^{\mathrm{o}} - y_t^{\mathrm{f}}\right)^2} \tag{9}$$

$$\mathrm{CC} = \frac{\frac{1}{n}\sum_{t=1}^{n}\left(y_t^{\mathrm{o}} - \overline{y}_t^{\mathrm{o}}\right)\left(y_t^{\mathrm{f}} - \overline{y}_t^{\mathrm{f}}\right)}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t^{\mathrm{o}} - \overline{y}_t^{\mathrm{o}}\right)^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t^{\mathrm{f}} - \overline{y}_t^{\mathrm{f}}\right)^2}} \tag{10}$$

where y_t^{o} is the observed and y_t^{f} is the forecast value at time t, and n is the number of data points. MAE and RMSE can provide information on the predictive ability of the

models involved. Meanwhile, the correlation coefficient CC can measure the correlation between the prediction and the observed data.

3 Description of data

Langat River is one of the longest rivers in Selangor and its river basin is transboundary, inasmuch as it crosses three states – Selangor, Negeri Sembilan, and the Federal Territory of Kuala Lumpur and Putrajaya (Department of Irrigation and Drainage Malaysia, 2011).The Langat River flows from Mount Nuang in Hulu Langat district to the Straits of Malacca in Kuala Langat. The Langat River catchment area covers a total of 1815 km^2 and is located between latitude 2°40′152″ N and 3°16′15″N, and longitude 101°19′20″ E and 102°1′10″ E (Juahir et al., 2011). There are two water reservoirs located in this area – Langat Dam and Semenyih Dam. Langat Dam was built with an area of 54 km^2 and Semenyih Dam has an area of 41 km^2. Both of these dams were built to deliver water for domestic and industrial use. In addition, the Langat Dam is also used to generate electricity for the use of residents in the vicinity of the Langat Valley. There are several towns and villages built along the Langat River – Cheras, Semenyih, Dengkil and Kajang. Since 1976, Langat River has also been acknowledged to be an area that regularly suffers flooding.

The variations of daily river flow data for Checkpoint 2 are shown in Fig. 2. The irregular patterns in data for Kajang River show that the river in this area is a complex system. The overall data were taken from the Department of Irrigation and Drainage Malaysia. Missing data constitute about 0.018 % and were filled using the results of linear interpolation calculation. The statistical parameters of the data cover a period of four years (January 2002 to December 2005) and are shown in Table 1.

4 Results and Discussion

River flow prediction using NLP involves the reconstruction of the phase space and prediction. Thus, the discussion of the findings is divided into two parts. The first part is to determine the parameters for the reconstruction of the phase space for Models I and II. Meanwhile, the description of the prediction results are discussed in the second part.

The phase diagram can provide information about the dynamics of a system through the trajectories in the phase space. The trajectories that are of interest focus on a subspace called the attractor. In addition, the observation of attractor trajectories in the phase space can provide information about the chaotic behaviour of the system. Hence, the phase diagram and the observation data involved are plotted. Figure 3 shows the phase diagram in two and three dimensions with $\tau = 1$. The trajectories in the phase space can indicate the presence of chaotic behaviour of the data (Sivakumar, 2002). Referring to Fig. 3, the trajectories of the attractor are clearly shown in the two phase diagrams. Thus, the data involved in this analysis are chaotic. Therefore, the dynamics of the system can be studied using chaos theory without involving stochastic methods.

This study involved data from January 2002 to December 2005 (1433 days). Three years of data are used in the reconstruction of the phase space to predict the behaviour one year ahead. Reconstruction of the phase space is based on the embedding dimension. In Model I, the embedding dimension is based on the calculation of the correlation dimension. Graph $\ln C(r)$ vs. $\ln(r)$ in Fig. 4a shows the behaviour of the correlation function v vs. radius r for the increasing m-dimensional. In general, the increasing value of the m-dimensional gradient occurs at the beginning of the curve from left $m = 1$ to right $m = 10$. Meanwhile, the graph of the correlation dimension estimation, the relationship between the correlation exponent v for different values of m is shown in Fig. 4b. The relationship between the value of correlation dimension d_2 and m-dimension can be seen in Fig. 4c, which is a graph of d_2 vs. m. The value of the correlation dimension

increased as the value of the m-dimensional increased. The increase in m-dimension can be seen up to a scaling region where the correlation dimension is saturated. The situation in which the value for the correlation dimension is saturated might indicate the existence of deterministic dynamics in the system. The saturated conditions for the d_2 value is in the interval $(2.5, 3)$. The saturation value for d_2 is known as the correlation dimension attractor (Sivakumar, 2000). In general, the sufficient condition for the value of the smallest integer m is m greater than $2D_2$ (Wu and Chan, 2010). Thus, the value of $m = 6$ is related to the Langat River flow time series in Kajang. The correlation dimension d_2 is finite and shows low levels of correlation dimension. Hence, Sungai Langat is a chaotic and deterministic system. Model I involves a combination of preliminary parameters $(1, 6)$ in the phase space reconstruction. Model II involves the calculation of m using FNN to find a combination of the preliminary parameters for RPS. Figure 5 shows the percentage of false nearest neighbours vs. m. Thus, the optimal value for the embedding dimension identified is $m = 14$. Model II involves a combination of parameters $(1, 14)$ for the reconstruction phase space.

4.1 River flow prediction for Model I and Model II

The combination of preliminary parameters for Model I is $(1, 6)$ while for Model II it is $(1, 14)$. Thus, for both models, the combination of the preliminary parameters (τ, m) has been applied to construct the phase space. Figure 6 and Table 2 provide a summary of the river flow prediction results in terms of MAE, RMSE and CC. Overall, the results show good performance prediction for chaos theory in predicting the future value of the river flow for the downstream area. Referring to Table 2, a comparison of prediction performance shows that the prediction results for Model II are better than Model I. The correlation coefficient for Model II (0.6360) is slightly higher compared to Model I (0.6103). Thus, analysis and prediction of the Langat River can provide information in which the selection of a combination of preliminary parameters in the reconstruction phase space is essential for better prediction results. In this study, Model II uses FNN

to calculate the embedding dimension m and is more appropriate than the correlation dimension method.

5 Conclusions

Analysis and prediction for testing the presence of chaotic behaviour in daily river flow data recorded at Langat River involving the station at Kajang, Selangor, Malaysia, has been performed. The station is located in the downstream area, which is a flood prone area. The analysis was carried out on the river flow data for a period of 4 yr (2002–2005). The focus of this study was to identify the chaotic behaviour of the river flow data in the downstream area and determine whether the river flow can be predicted when chaotic behaviour of the river exists downstream. Chaos theory, together with NLP, were used in the analysis. The reconstruction phase space clearly shows the existence of a chaotic attractor. Hence, the data involved in this analysis are chaotic. Next, was the attempt to make a prediction for one year ahead with the observed data using the results of the reconstruction of the phase space for three years. Two combinations of preliminary parameters were used. Model I used $\tau = 1$ for which m is the result of the calculation of the correlation dimension, while Model II used $\tau = 1$ for which m is the result of FNN calculation. Using these methods, the optimal combination for Model I was (1,6) and for Model II it was (1,14). The overall prediction results showed that both models could give a good prediction for the river flow downstream. However, the combination of the preliminary parameters for Model II using the FNN algorithm provided a better prediction result than Model I, which used the correlation dimension. The results showed that Langat River in Kajang, which is in the downstream area, is chaotic and predictable using NLP. Therefore, the results of the analysis and prediction of river flow in the downstream area could provide information on river flow for the authorities to take appropriate control of the downstream flooding.

Acknowledgements. The authors are grateful to Universiti Kebangsaan Malaysia for providing financial support via the grant UKM-DIP-2012-31.

References

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

Adenan, N. H. and Noorani, M. S. M.: Behaviour of daily river flow: Chaotic?, Proceedings of the 20th National Symposium on Mathematical Sciences: Research in Mathematical Sciences: A Catalyst for Creativity and Innovation, Malaysia, 221–228, 2013.

Ahmad, Z. and Juahir, H.: Neural Network Model for Prediction of Discharged from the Catchments of Langat River, Malaysia, IIUM Engineering Journal, 7, 25–34, 2006.

Department of Irrigation and Drainage Malaysia: Langat River Integrated River Basin Management Study, Final Report, Technical Studies Part 1 of 4, 2005.

Department of Irrigation and Drainage Malaysia: Review of the National Water Resources (2000–2050) and Formulation of National Water Resources Policy, Selangor, Federal Territory of Kuala Lumpur and Putrajaya, 2011.

Ghani, A. A., Ali, R., Zakaria, N. A., Hasan, Z. A., Chang, C. K., and Ahmad, M. S. S.: A Temporal Change Study of the Muda River System Over 22 Years, International Journal of River Basin Management, 8, 25–37, 2010.

Ghorbani, M. A., Kisi, O., and Aalinezhad, M.: A probe into the chaotic nature of daily streamflow time series by correlation dimension and largest Lyapunov methods, Appl. Math. Modell., 34, 4050–4057, 2010.

Ghorbani, M. A., Daneshfaraz, R., Arvanagi, H., and Pourzangbar, A.: Local Prediction in River Discharge Time Series, Online Journal of Civil Engineering and Urbanism, 2, 51–55, 2012.

Grassberger, P.: Do climatic attractors exist?, Nature, 323, 609–612, 1986.

Hall, M. J.: Urban Hydrology, Elsevier Applied Science Publishers, New York, 1984.

Islam, M. N. and Sivakumar, B.: Characterization and prediction of runoff dynamics: a nonlinear dynamical view, Adv. Water Res., 25, 179–190, 2002.

Jayawardena, A. W. and Lai, F.: Chaos in hydrological time series, Proceeding of the Yokohama Symposium – Extreme Hydrological Events: Precipitation, Floods and Droughts, Yokohama, 59–66, 1993.

Jayawardena, A. W. and Lai, F.: Analysis and prediction of chaos in rainfall and streamflow time series, J. Hydrol., 153, 23–52, doi:10.1016/0022-1694(94)90185-6, 1994.

Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I. T., Armi, A. S. M., Toriman, M. E., and Mokhtar, M.: Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques, Environ. Monit. Assess., 173, 625–641, 2011.

Khatibi, R., Sivakumar, B., Ghorbani, M. A., Kisi, O., Kocak, K., and Zadeh, D. F.: Investigating chaos in river stage and discharge time series, J. Hydrol., 414–415, 108–117, 2012.

Lorenz, E. N.: Atmospheric predictability as revealed by naturally occurring analogues, J. Atmos. Sci, 26, 636–646, 1969.

Martins, O. Y., Sadeeq, M. A., and Ahaneku, I. E.: Nonlinear Deterministic Chaos in Beneu River Flow Daily TIme Sequence, Journal of Water Resource and Protection, 3, 747–757, 2011.

Men, B., Zhao, X., and Liang, C.: Chaotic analysis on monthly precipitation on Hills Region in Middle Sichuan of China, Nature and Science, 2, 45–51, 2004.

Mohammed, T. A., Al-Hassoun, S., and Ghazali, A. H.: Prediction of flood levels a streacth of the Langat River with insufficient hydrological data, Pertanika Journal of Science & Technology, 2, 237–248, 2011.

Regonda, S., Rajagopalan, B., Lall, U., Clark, M., and Moon, Y.-I.: Local polynomial method for ensemble forecast of time series, Nonlin. Processes Geophys., 12, 397–406, doi:10.5194/npg-12-397-2005, 2005.

Sangoyomi, A., Lall, L., and Abarbanel, H. D. I.: Nonlinear dynamics of the Great Salt Lake: dimension estimation, Water Resour. Res., 32, 149–159, 1996.

Shabri, A. and Suhartono: Streamflow forecasting using least-squares support vector machines, Hydrolog. Sci. J., 57, 1275–1293, doi:10.1080/02626667.2012.714468, 2012.

She, D. and Yang, X.: A New Adaptive Local Linear Prediction Method and Its Application in Hydrological Time Series, Math. Probl. Eng., 2010, 205438, doi:10.1155/2010/205438, 2010.

Sivakumar, B.: Chaos theory in hydrology: important issues and interpretation, J. Hydrol., 227, 1–20, 2000.

Sivakumar, B.: A phase-space reconstruction approach to prediction of suspended sediment concentration in rivers, J. Hydrol., 258, 149–162, 2002.

Sivakumar, B.: Forecasting monthly streamflow dynamics in the western United States: a nonlinear dynamical approach, Environ. Modell. Softw., 18, 721–728, 2003.

Tongal, H. and Berndtsson, R.: Phase-space reconstruction and self-exciting threshold modeling approach to forecast lake water levels, Stoch. Env. Res. Risk. A., online first, doi:10.1007/s00477-013-0795-x, 2013.

Viessman, W. and Lewis, G. L.: Introduction to Hydrology, HarperCollins College Publishers, New York, 760 pp., 1996.

Wu, C. L. and Chau, K. W.: Data driven for monthly streamflow time series prediction, Eng. Appl. Artif. Intel., 23, 1350–1367, 2010.

Table 1. Statistics for river flow series at Kajang station (Checkpoint 2).

Number of data	Average	Max	Min	Standard Deviation	Skew	Kurtosis
1433	8.6492	212	0.3	13.779	6.733	68.438

Table 2. Prediction performance at Kajang station (Checkpoint 2).

	MAE	RMSE	CC
MODEL I	2.6857	4.4025	0.6103
MODEL II	3.5160	5.6929	0.6360

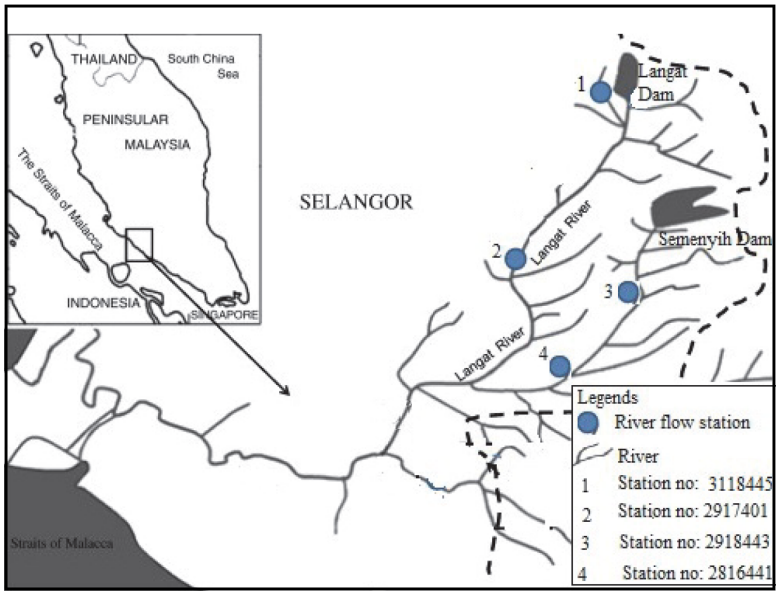

Fig. 1. Location of stations.

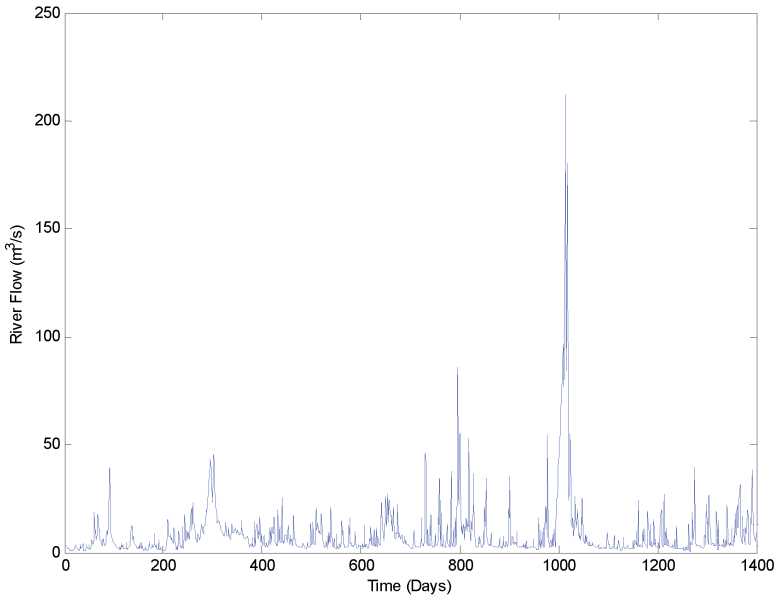

Fig. 2. Variations in data for Kajang station (Checkpoint 2).

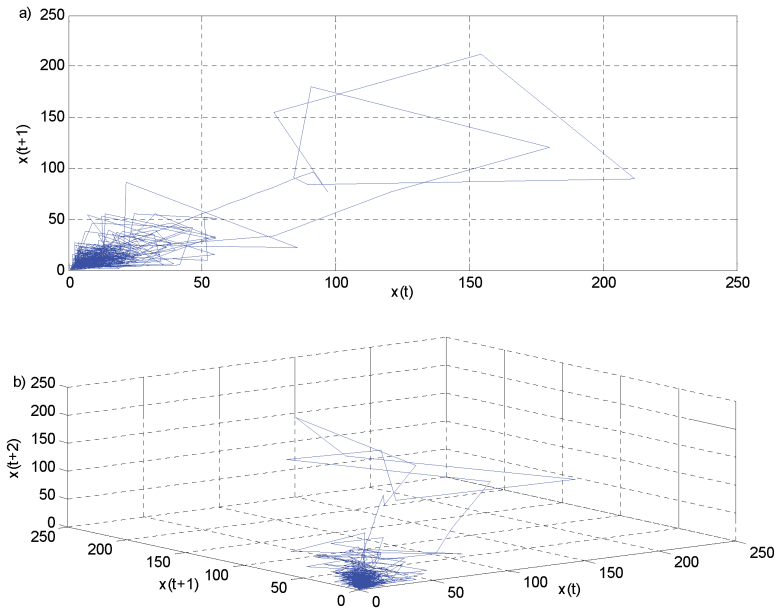

Fig. 3. Phase space diagram.

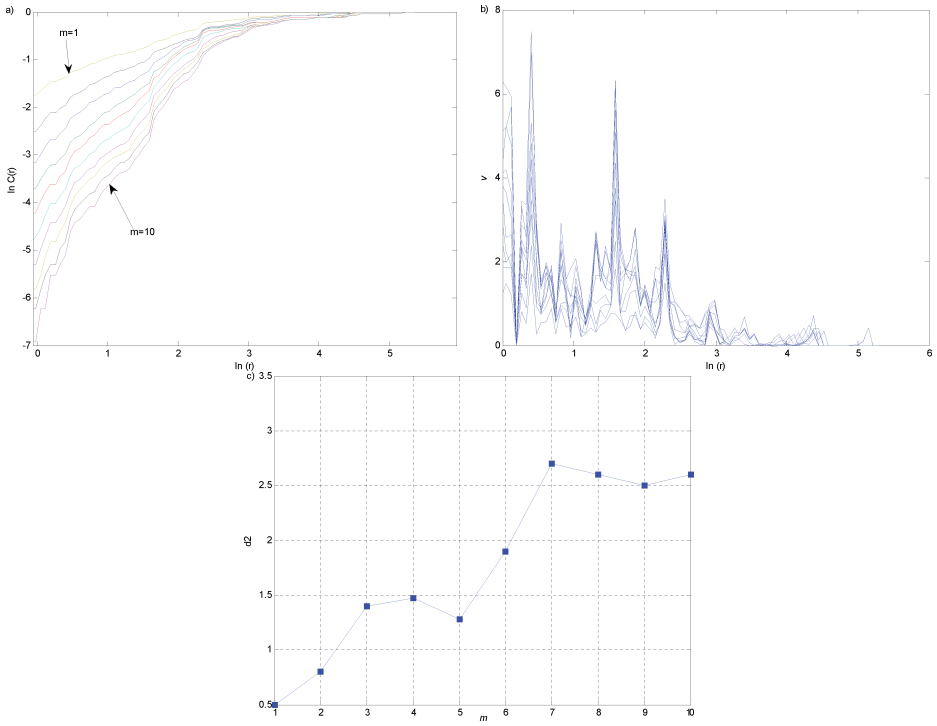

Fig. 4. $\ln C(r)$ vs. $\ln r$ for **(a)**; the estimation of correlation dimension (d_2) for $\tau = 1$ and m is at the interval of $[1, 10]$ (increasing from bottom to top in each pane) **(b)**; relationship between d_2 and m for daily river flow of Langat River **(c)**.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

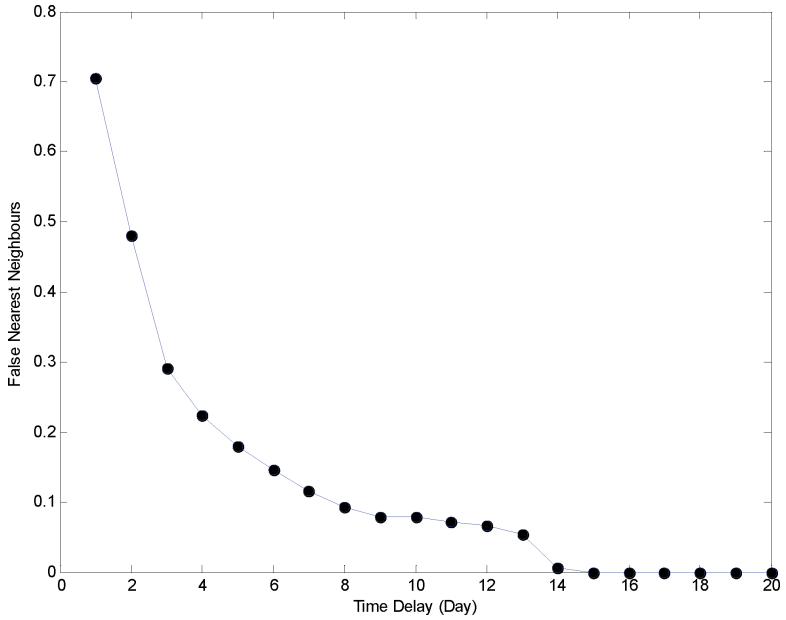

Fig. 5. False nearest neighbour at Kajang station when τ = 1 and R_T = 15.

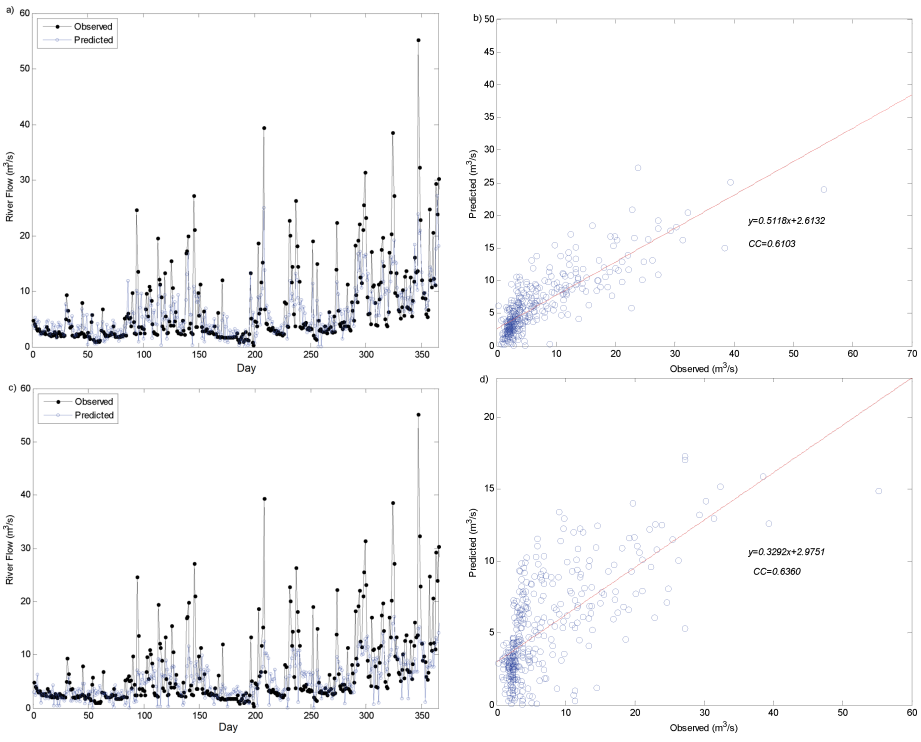

Fig. 6. Comparison of time series and scatter diagrams of prediction results and observed data for **(a–b)** Model I (1, 6) **(c–d)** and Model II (1, 14).