World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:7, No:11, 2013

Monthly River Flow Prediction Using a Nonlinear Prediction Method

N. H. Adenan, M. S. M. Noorani

***Abstract*—**River flow prediction is an essential tool to ensure proper management of water resources and the optimal distribution of water to consumers. This study presents an analysis and prediction by using nonlinear prediction method with monthly river flow data for Tanjung Tualang from 1976 to 2006. Nonlinear prediction method involves the reconstruction of phase space and local linear approximation approach. The reconstruction of phase space involves the reconstruction of one-dimension (the observed 287 months of data) in a multidimensional phase space to reveal the dynamics of the system. The revenue of phase space reconstruction is used to predict the next 72 months. A comparison of prediction performance based on correlation coefficient (*CC*) and root mean square error (RMSE) was employed to compare prediction performance for the nonlinear prediction method, ARIMA and SVM. Prediction performance comparisons show that the prediction results using the nonlinear prediction method are better than ARIMA and SVM. Therefore, the results of this study could be used to develop an efficient water management system to optimize the allocation of water resources.

***Keywords*—**River flow, nonlinear prediction method, phase space, local linear approximation.

I. Introduction

EFFICIENT allocation of water resources can meet the needs of water demand. Water management is closely related to river flow prediction. Accurate prediction can help to provide information about the flow of the river for water allocation. Therefore, a river flow prediction method that could produce an accurate prediction is important to provide information for optimal water management. River flow is a continuous phenomenon. The irregular patterns in the river flow data show the complexity of the river system. The system is influenced by catchment characteristics (size, slope, shape), the characteristics of the storm (and the increase in rainfall), geographic characteristics (topology, terrain, soil type) and climatic characteristics (temperature, humidity, wind) [1-3]. Since a few decades ago, the stochastic hydrology approach has been widely used in hydrology analysis [4].

The development of research on river flow prediction is growing in respect of certain research that has been undertaken. Gene-expression programming [5], fuzzy logic [6], hydrodynamic modelling [7], autoregressive integrated moving average (ARIMA) [3], artificial neural network (ANN) [3], support vector machine (SVM) [3], [8] and support vector machine smallest power (LSSVM) [3] are some approaches that have been used in Malaysia. All the methods described use a number of variables that affect river flow prediction. Therefore, a transitional approach will be undertaken by just analyzing a time series (river flow data) for river flow prediction. The approach involves the chaotic theory.

The increase in water demand is significant with population growth and the rapid economic development in certain areas. This situation is apparent in the Kinta District of Perak in Malaysia. The rapid development in this area has impacted on the water management in the Kinta District. Referring to Table I, the demand for water supply is expected to increase to 471,000 m^3 per day in 2050 compared to 277,200 m^3 per day in 2010 [9]. Therefore, river flow prediction is crucial to ensure the optimal distribution of water. A study of the monthly river flow prediction in this area has been undertaken by using autoregressive integrated moving average (ARIMA) and support vector machine (SVM) [3]. However, in this study, using chaotic theory, the monthly river flow using the same data was determined to provide comparative results on the accuracy of the predicted monthly river flow.

Various studies have been carried out by applying the principles of chaos theory to the hydrology time series data to prove the chaotic behavior of the hydrological system [10]-[12]. The studies focused on the predictive value of the time series data in the future. The results showed that river flow prediction and other hydrological processes give a similar prediction to the actual data values [13]-[15]. Apart from being able to provide accurate prediction results, the chaotic theory approach can reveal the number of variables that influence the flow of a river in an area. Thus, the dynamic behavior of the river flow prediction can help to provide information for the efficient management of hydraulic structures.

II. Material and Methods

Nonlinear prediction method involves the reconstruction of phase space using deterministic data. Then, the prediction is done on the phase space using local linear approximation method. The first step is the reconstruction of the phase space to reveal the dynamics of time series by referring to the trajectories in the phase space. The attractor of a system can be shown on the trajectories of a system. The trajectories focus on a particular sub-space called the attractor. Observations on the plot attractor in the phase space can provide information about chaotic behavior. A scalar time series $x(t)$ forms a one-dimensional time series:

N. H. Adenan is with the Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia (e-mail: hamieza@fsmt.upsi.edu.my).

M. S. M. Noorani is with the School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia (e-mail: msn@ukm.my).

$$\{x_i\} = \{x_1, x_2, x_3, ..., x_N\} \quad (1)$$

where N is the total number of time series. From this signal, we can construct the m-dimensional signal:

$$\mathbf{Y}_t = \{x_t, x_{t+\tau}, x_{t+2\tau}, ..., x_{t+(m-1)\tau}\} \quad (2)$$

where τ is an appropriate time delay and m is a chosen embedding dimension.

Two parameters have to be determined, the time delay τ and embedding dimension m. In this study, τ is a predetermined value while the value of m varies. The most optimal value of τ can provide a separation of neighboring projections in any dimension embedded in the phase space. If the value of τ is too small, the coordinates of the phase space cannot describe the dynamics of the system. Meanwhile, information on projections in the phase space will diverge if the value is too large [16], [17]. Previous studies on river flow prediction showed that when the condition for time delay $\tau = 1$ is used in phase space reconstruction, the result provides good prediction [13], [14]. Thus in this study, the time delay $\tau = 1$ is used. Meanwhile, the optimal value for the m-embedding dimensional phase space can describe the attractor topology. In this study, the m-dimensional was varied (m = 2, 3, 4... 10) to find the best set of dimensions that can provide good prediction results.

Reconstruction of the phase space is $\mathbf{Y}_t = \{x_t, x_{t+\tau}, x_{t+2\tau}, ..., x_{t+(m-1)\tau}\}$. To predict $\mathbf{Y}_{t+1}$, the nearest neighbor(s) to $\mathbf{Y}_t$ are searched. The Euclidean distance between $\mathbf{Y}_t$ and the vectors before $\mathbf{Y}_i$ (i = 1, 2...t-1) is calculated. Let, the minimum distance to the nearest neighbor be $\mathbf{Y}_M$. The values $\mathbf{Y}_M$ and $\mathbf{Y}_{M+1}$ are used to satisfy the linear equation $\mathbf{Y}_{M+1} = A\mathrm{Y}_M + B$. The constant values of A and B are calculated using the least squares method. Thus, the predictive value $\mathbf{Y}_{t+1}$ can be calculated using $\mathbf{Y}_{t+1} = A\mathrm{Y}_t + B$. Evaluation of prediction performance has been done by using correlation coefficient (CC) and root mean square error (RMSE).

III. Description of Data

The Kinta River catchment area comprises the entire 2540 km^2 covering the Kinta River in the eastern state of Perak and is located at latitude 4.1° and longitude 101.0166667°. The Kinta River is important because it is the main source of water for drinking and irrigation in the state. The Kinta dam is able to supply 639,000 m^3 of water each day and has the ability to meet the demands of water consumption by the year 2020 [9]. Thus, water resource management is important to ensure that the supply of water from the Kinta Dam is properly allocated to the user. The topology of the catchment area consists of forest, which covers the hills in the north and south. Land use in the Kinta Valley consists of urban development, former unproductive mines and agriculture, rubber, oil palm and fruit trees [18]. There are three streams that contribute to the Kinta River flow system – Pari River with 245km^2, Raia River 250km^2 and Kampar River 430km^2.

Tables I and II show that the water demand and population are expected to increase in the Kinta and Kampar region. Thus, the river flow of the section of Kinta River studied in this paper is suitable for providing information about the river flow. Along the Kinta River, the river flow is measured at several measurement stations. Thus, the river flow station in Tanjung Tualang (station number: 4310401) was analyzed. The location of the station is shown in Fig. 1. River flow data that dates back to 1973 from the Department of Drainage and Irrigation are available. However, this study only used monthly data from October 1976 to July 2006, which has 0.06% missing data. The missing data were replaced with the results from the computation of the linear interpolation method. The basic statistics for river flow data are presented in Table III.

TABLE I
WATER DEMAND IN KINTA AND KAMPAR REGION [9]

Water demand (m^3 per day)	2010	2015	2020	2030	2040	2050
Domestic	188,800	205,700	223,500	234,500	258,900	269,500
Industrial	51,900	56,600	61,300	81,600	112,000	147,600
Commercial	27,400	30,000	32,800	36,100	38,800	40,400
Institutional	9,100	9,900	10,900	12,000	12,900	13,500
Total	277,200	302,200	328,500	373,200	422,600	471,000

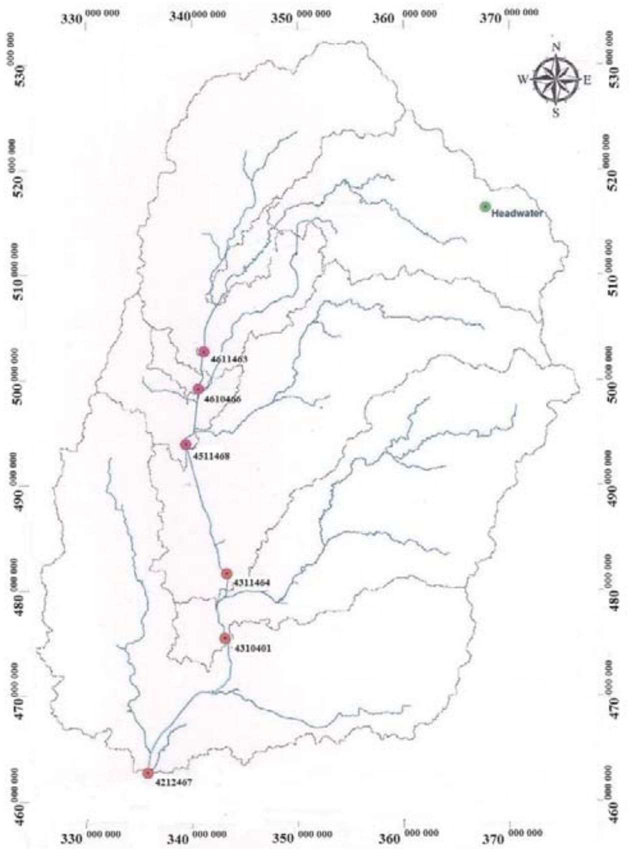

Fig. 1 Station location

TABLE II
POPULATION GROWTH IN KINTA AND KAMPAR REGION [9]

Year	Population
2010	807 000
2015	843 000
2020	880 000
2030	928 000
2040	966 000
2050	987 000

TABLE III
STATISTICS OF RIVER FLOW SERIES AT TANJUNG TUALANG STATION

Number of data	Average	Max	Min	Standard deviation	Skew	Kurtosis
360	78.68	651.08	9.41	44.21	2.079	8.246

IV. RESULT AND DISCUSSION

Fig. 2 illustrates the two-dimensional phase diagram showing the attractor plot using time delay $\tau = 1$. The presence of the attractor in the phase space can show the chaotic behavior of the data [13]. Referring to the figure, the attractor trajectories are a reasonably well defined region and the plot shows that the river flow is chaotic. Therefore, this data can be predicted with confidence using the nonlinear prediction method and the chaotic theory approach. The study included data from October 1976 to July 2006 (360 months). The monthly river flow data for 24 years (287 months, 80% of the data) are used for the reconstruction of the phase space to predict the next 72 months. The phase space is built by using different embedding dimensions from 2 to 10. Revenue prediction for the 72 months, including the correlation coefficient (*CC*) and root mean square error (RMSE), is shown in Table IV. Referring to Table IV, the prediction accuracy was seen when $m = 6$ (m_{opt}); hence, $m = 6$ (m_{opt}). The presence of an optimum value embedding dimension shows the low dimensional chaotic behavior of the river flow dynamics [19]. Thus, there are two indications that the observed data is chaotic. The existence of an attractor of the trajectories in the phase space is shown in Fig. 2 and the prediction result for m_{opt}. The value of m_{opt} is low, hence the observed data are categorized as having low-dimensional chaotic behaviour. The scatter plot for the best prediction result, $m = 6$, is shown in Fig. 3.

The analysis of the comparison for river flow prediction performance is done by comparing the prediction results for nonlinear prediction method, ARIMA and SVM. The performance evaluation for ARIMA and SVM was taken from previous studies involving analysis of the same data [3]. Table V shows the performance evaluation of ARIMA and SVM. By using nonlinear prediction method, the correlation coefficient (*CC*) value is 0.586 with $m = 6$ (m_{opt}), as shown in Table III. The lowest RMSE value for $m = 6$ (m_{opt}) is 19.467. Fig. 3 shows the best prediction accuracy, which is using $m = 6$ and $\tau = 1$. While the prediction results for the correlation coefficient (*CC*) show that the ARIMA and SVM methods are 0.525 and 0.565, respectively, and the RMSE values of ARIMA and SVM are 21.783 and 16.715, respectively. The comparative analysis shows that the nonlinear prediction method gives better accuracy than ARIMA and SVM models.

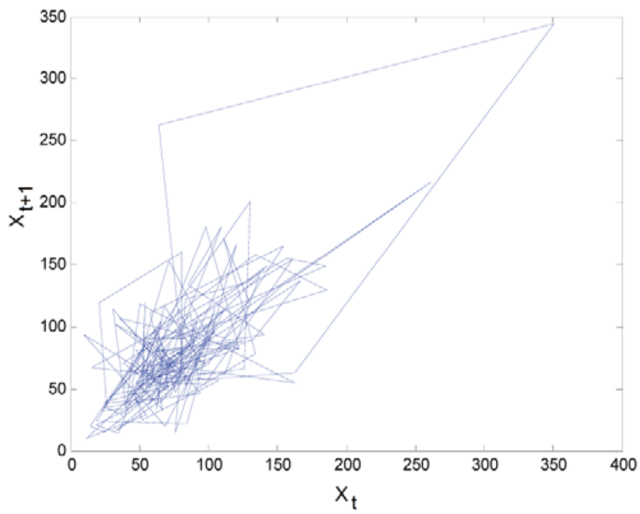

Fig. 2 Phase space

TABLE IV
MONTHLY RIVER FLOW PREDICTION USING LOCAL LINEAR APPROXIMATION METHOD AT TANJUNG TUALANG

m	*CC*	RMSE
2	0.563	22.932
3	0.454	22.114
4	0.440	22.799
5	0.480	21.863
6	0.586	19.467
7	0.450	22.242
8	0.458	23.897
9	0.473	22.174
10	0.432	23.479

TABLE V
THE PERFORMANCE RESULTS OF ARIMA AND SVM METHODS AT TANJUNG TUALANG STATION [3]

m	ARIMA	SVM
CC	0.525	0.565
RMSE	21.783	16.715

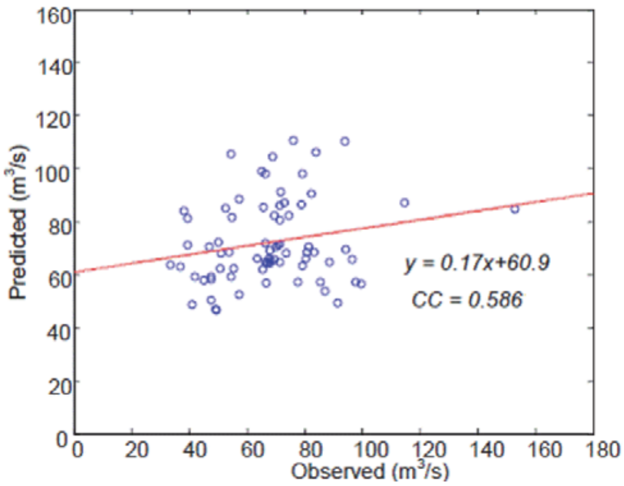

Fig. 3 Predicted and observed river flow using linear approximation method with $m = 6$

V. CONCLUSION

The observed river flow data at Tanjung Tualang from October 1976 to July 2006, a period of 30 years, is used for analysis and prediction. The behavior of the chaotic time series is analyzed by using nonlinear prediction method, which involves the reconstruction of the phase space of one dimension (river flow data for 287 months) in a multi-dimensional phase space. Phase space reconstruction is able to show the river flow dynamics. Then, prediction for the next 72 months is undertaken using local linear approximation method using the results from the reconstruction of the phase space. The prediction results using the nonlinear prediction method are compared with ARIMA and SVM. The results show that the nonlinear prediction method provides better prediction results than the ARIMA and SVM. Thus, nonlinear prediction method is recommended for long-term prediction.

ACKNOWLEDGMENTS

This work would not have been possible without the generous support of Universiti Pendidikan Sultan Idris and the Ministry of Higher Education Malaysia.

REFERENCES

[1] J.R.M. Hosking and J.R. Wallis, *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press. UK, 1997.
[2] A. Jain and A.M. Kumar, “Hybrid neural network models for hydrologic time series forecasting,” *Applied Soft Computing*, vol. 7, pp. 585-592, 2007.
[3] A. Shabri and Suhartono, “Streamflow Forecasting using Least-squares Support Vector Machines,” *Hydrological Sciences Journal*, vol. 57, pp. 1275-1293, 2012.
[4] J.D. Salas, G.Q. Tabios III and P. Bartolini, “Approaches to Multivariate Modeling of Water *Resources Time Series,” Water Resource Bulletin*, vol. 4, pp. 683-708, 1985.
[5] H.M. Azamatulla, A.A. Ghani, C.S. Leow, C.K. Chang and N.A. Zakaria, “Gene-Expression P*rogramming for the Development of a Stage-Disch*arge Curve of the Pahang River,” *Water Resource Management*, vol. 25, pp. 2901-2916, 2011.
[6] Y. Mohd Shafiek, Dr., J. Hishamuddin and H. Sobri, *“Daily* stream flow forecasting usi*ng simplified rule-based fuzzy logic system,” The Journal of the Institution of Engineers, Malaysia, vol.* 66, pp. 23-28, 2005.
[7] A.A. Ghani, R. Ali, N.A. Zakaria, Z.A. Hasan, C.K. Chang and M.S.S. Ahamad, “A T*emporal Change Study of the Muda River System Over 22 Years,” International Journal of River Basin Management, v*ol. 8, pp. 25-37, 2010.
[8] Z.A. Zakaria and A. Shabri, “*Streamflow Forecasting at Ungaged Si*tes Using Support Vector Machines,” *Applied Mathematical Sciences*, vol. 6, pp. 3003-3014, 2012.
[9] Department of Irrigation and Drainage Malaysia, Review of National Water Resources (2000-2050) and Formulation Water Resources Policy. 2011. Kuala Lumpur.
[10] A.W. Jayawardena and F. Lai, “Analysis and prediction of chaos in rainfall and streamflow time series,” *Journal of Hydrology*, vol. 153, pp. 23-52, 1994.
[11] W. Wang, P.H. Van Gelder and J. Vrijling, “Is the streamflow process chaotic?,” in *International Symposium Stochastic Hydraulic*, Madrid, Spain 2005 pp. 162-164.
[12] W.W. Ng, US.Panu and W.C. Lennox, “Chaos based analytical techniques for daily extreme hydrological observations,” *Journal of Hydrology*, vol. 842, pp. 17-41, 2007.
[13] B. Sivakumar, “A phase-space reconstruction approach to prediction of suspended sediment concentration in rivers,” *Journal of Hydrology*, vol. 258, pp. 149-162, 2002.
[14] B. Sivakumar, “Forecasting monthly streamflow dynamics in the western United States: a nonlinear dynamical approach,” *Environmental Modelling & Software*, vol. 18, pp. 721-728, 2003.
[15] R. Khatibi, B. Sivakumar, M.A. Ghorbani, O. Kisi, K. Kocak and D.F. Zadeh, “Investigating chaos in river stage and discharge time series,” *Journal of Hydrology, vol. 414-415, p*p. 108-117, 2012.
[16] A. Sangoyomi, L. Lall and H.D.I. Abarbanel, *“Nonlinear Dynamics of the Great Salt Lake: Dimension Estimation,” Water Resource Research*, vol. 32, pp. 149-159, 1996.
[17] M.N. Islam and B. Sivakumar, “Characterization and prediction of runoff dynamics: a nonlinear dynamical view,” *Advances in Water Resources*, vol. 25, pp. 179–190, 2002.
[18] N.M. Gazzaz, M.K. Yusoff, M.F. Ramli, A.Z. Aris and H. Juahir, “Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques,” *Marine Pollution Bulletin*, vol. 64, pp. 688-698, 2012.
[19] M. Casdagli, “Nonlinear prediction of chaotic time series,” *Physica D: Nonlinear Phenomena*, vol. 35, pp. 335–356, 1989.

Nur Hamiza Adenan (N. H. Adenan) was born in Kedah, Malaysia, in 1984. She received her Bachelor Degree of Education Mathematics from the Universiti Pendidikan Sultan Idris, Malaysia, in 2008, and the Master of Science in Mathematics from Universiti Teknologi Malaysia in 2010. She is currently pursuing her Ph.D. degree in Applied Mathematics at Universiti Kebangsaan Malaysia.

In 2008, she joined the Department of Mathematics, Universiti Pendidikan

Sultan Idris as a Tutor, and, in 2010, became a lecturer. Her current research interests include the nonlinear prediction method, chaos theory and hydrology.

Mohd Salmi Md Noorani (M. S. M. Noorani) was born in Kelantan, Malaysia, in 1961. He received his Bachelor degree, Master of Science and Ph.D. degrees in Mathematics from the University of Warwick, United Kingdom.

In 1983, he joined the School of Mathematical Sciences, Universiti Kebangsaan Malaysia, as a tutor, and, in 1986, became a lecturer. Since 1986, he has been with the School of Mathematical Sciences, initially as an Assistant Professor, before becoming an Associate Professor in 1996, and a Professor in 2006. His current research interests include functional analysis, topology and dynamic systems. Prof. Mohd Salmi is the President of the Persatuan Sains Matematik Malaysia (2001 until now) and an active fellow for the Southeast Asian Mathematical Society (2005 until now), American Mathematical Society for Industrial and Applied Mathematics (USA) and Mathematical Association of America.