River Flow Prediction Using Nonlinear Prediction Method

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

***Abstract*—**River flow prediction is an essential to ensure proper management of water resources can be optimally distribute water to consumers. This study presents an analysis and prediction by using nonlinear prediction method involving monthly river flow data in Tanjung Tualang from 1976 to 2006. Nonlinear prediction method involves the reconstruction of phase space and local linear approximation approach. The phase space reconstruction involves the reconstruction of one-dimensional (the observed 287 months of data) in a multidimensional phase space to reveal the dynamics of the system. 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) have been employed to compare prediction performance for nonlinear prediction method, ARIMA and SVM. Prediction performance comparisons show the prediction results using nonlinear prediction method is better than ARIMA and SVM. Therefore, the result of this study could be used to develop an efficient water management system to optimize the allocation 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 in providing information about the flow of the river for water allocation. Therefore, river flow prediction method that could produce accurate prediction is important to provide information for optimal water management. River flow is a continuous phenomenon. Irregular pattern in the river flow data show that river is a complex 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]. Then, a few decades ago, stochastic hydrology approach is widely used in the hydrology analysis [4].

Development of research on river flow prediction is growing with reference to some research on river flow prediction. Gene-expression programming [5], fuzzy logic [6], hydrodynamic modeling [7], autoregressive integrated moving average 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 have been used in Malaysia. All the methods described have been using a number of variables that affect river flow prediction. Therefore, a transitional approach will be done by just analyzing a time series (river flow data) for river flow prediction. The approach involves the chaotic theory.

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

There are a number of studies have been carried out by applying the principles of chaos theory in hydrology time series data to prove the chaotic behavior of the hydrological system [10]-[12]. The study 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 similar prediction with the actual data values [13]-[15]. Apart from being able to provide accurate prediction results, chaotic theory approach can reveal the number of variables that influence the flow of a river in the 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 involving 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 refereeing at the trajectories in the phase space. Attractor of a system can be shown on the trajectories of a system. The trajectories focus on a particular sub-space called attractor. Observations on the plot attractor in phase space can give information about chaotic behavior. A scalar time series $x(t)$ form 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: nurhamiza.adenan@gmail.com, hamieza@ fsmt.upsi.edu.my).

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

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

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

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

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 predetermined value while the value of m is varied. 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 phase space will diverge if the value is too large, and dynamic τ on chaotic system [16], [17]. Previous studies on the river flow prediction showed that when a condition of time delay τ = 1 is used in phase space reconstruction, the result ensemble good prediction [13], [14]. Thus in this study, the time delay τ = 1 is used. Meanwhile, the optimal value of m- embedding dimensional phase space can describe the attractor topology. In this study, m-dimensional varied (m = 2, 3, 4, ..., 10) to find the best set of dimensions that can provide a 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. 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 is $\mathbf{Y}_M$. Value $\mathbf{Y}_M$ and $\mathbf{Y}_{M+1}$ are used to satisfy a linear equation $\mathbf{Y}_{M+1} = A\mathrm{Y}_M + B$. The constant value of A and B is calculated using the least square 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

Kinta River catchment area comprises the entire 2540km^2 covering the Kinta River in the eastern state of Perak and is located at latitude 04' 19' 20' and longitude 101' 04' 30'. Kinta River is an important water source in the state because this river is the main source of drinking water and irrigation. Kinta dam is capable to supply 639 000m^3 of water each day and will be able to meet the demand of water consumption by the year 2020 [9]. Thus, water resource management is important to ensure the water supply of Kinta Dam can be allocated productively to the user. Topology of the catchment area consists of forest cover in the hills 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 which are Pari River with 245km^2, Raia River 250km^2 and Kampar River 430km^2.

Tables I and II show that water demand and population is expected to increase in the Kinta and Kampar region. Thus, the river flow of the Kinta River studied in this area is suitable to assist in providing information about the river flow. Along the Kinta River, river flow measured at several measurement stations. Thus, a river flow station in Tanjung Tualang (station number: 4310401) were analyzed. Station location is shown in Fig. 1. River flow data from Department of Drainage and Irrigation is available from 1973. However, this study involved only monthly data from October 1976 to July 2006. Involved data involved has 0.06% missing data. The missing data has been replaced with the results from the computation of linear interpolation method. 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..8	205.7	223.5	234.3	258.9	269.5
Industrial	51.9	56.6	61.3	81.6	112.0	147.6
Commercial	27.4	30.0	32.8	36.1	38.8	40.4
Institutional	9.1	9.9	10.9	12.0	12.9	13.5
Total	277.2	302.3	328.5	373.1	422.7	470.9

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

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

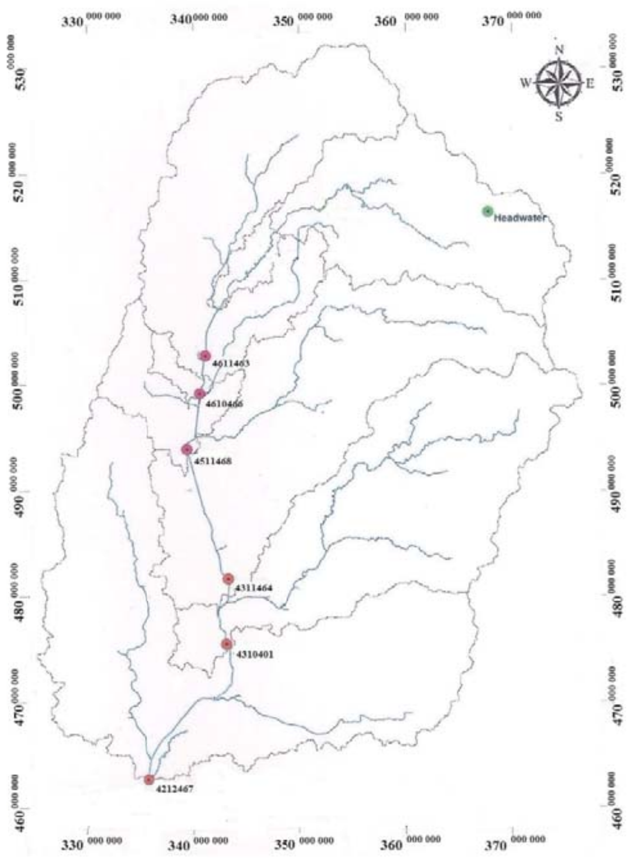

Fig. 1 Station location

IV. Result and Discussion

Fig. 2 shows the two-dimensional phase diagram showing the attractor plot using time delay $\tau = 1$. The presence of attractor in the phase space can show chaotic behavior of the data [13]. Referring to the figure, the attractor trajectories are reasonably well defined region and the plot shows the river flow is chaotic. Therefore, it is assured that this data can be predicted by using nonlinear prediction method using chaotic theory approach. The study involved data from July 1976 to October 2006 (360 months). 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 involving the correlation coefficient (*CC*) and root mean square error (RMSE) are 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 low dimensional chaotic behavior on river flow dynamics [19]. Thus, there are two evidence that the observed data is chaotic. There are exist attractor of trajectories in phase space showed in Fig. 2 and the prediction result have m_{opt}. The value of m_{opt} is low, hence the observed data are categorized as low-dimensional chaotic behavior.

Analysis of river flow prediction performance comparison is done by comparing the prediction results for nonlinear prediction method, ARIMA and SVM. The performance evaluation for ARIMA and SVM are 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 prediction results for the correlation coefficient (*CC*) shows that ARIMA and SVM method are 0.525 and 0.565. RMSE value of ARIMA and SVM is 21.783 and 16.715. The comparative analysis shows the nonlinear prediction method gave better accuracy than ARIMA and SVM model.

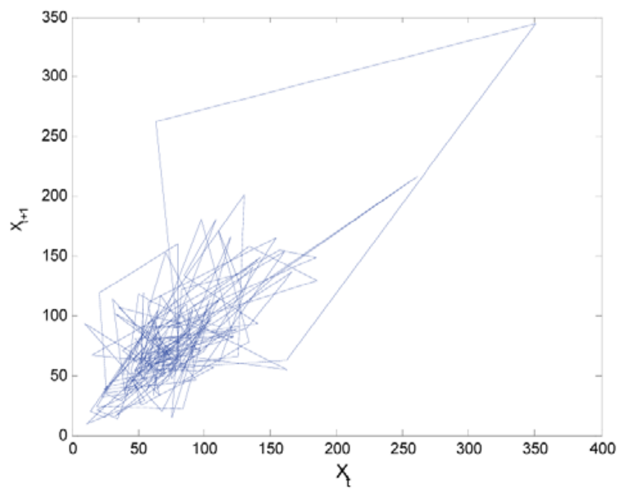

Fig. 2 Phase space

TABLE IV
MONTHLY RIVER FLOW PREDICTION USING LOCAL LINEAR APPROXIMATION METHOD AT TANJUNG TUALANG STATION FROM JULY 1976 UNTIL DECEMBER 2006

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 FROM JULY 1976 UNTIL DECEMBER 2006 [3]

m	ARIMA	SVM
CC	0.563	22.932
RMSE	21.783	16.715

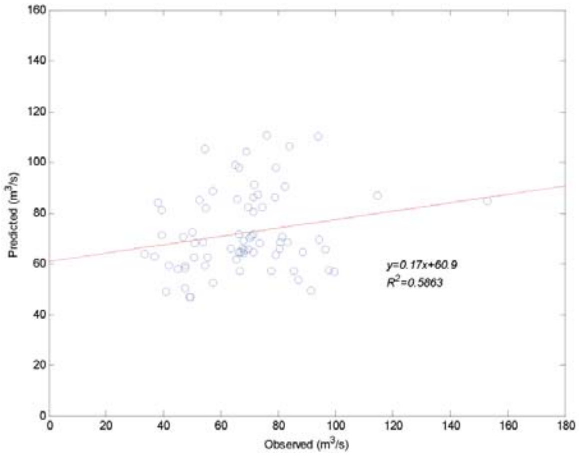

Fig. 3 Predicted and observed river flow using linear approximation method

V. CONCLUSION

Observed river flow data at Tanjung Tualang, from July 1976 to October 2006 that involved 24 years is used for analysis and prediction. The behavior of chaotic time series is analyzed by using nonlinear prediction method. Nonlinear prediction method involves reconstruction of the phase space of one dimension (river flow data for 287 months) in multi-dimension of phase space. Phase space reconstruction can show the river flow dynamics. Then, prediction is done using local linear approximation method using the results of reconstruction phase space to predict the next 72 months. The prediction results using nonlinear prediction method are compared with ARIMA and SVM. The results showed nonlinear prediction method provides a better prediction results than the ARIMA and SVM. Thus, nonlinear prediction method is recommended for long-term prediction.

ACKNOWLEDGMENT

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

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. TabiosIII 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 Programming for the Development of a Stage-Discharge 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 streamflow forecasting using 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 Temporal Change Study of the Muda River System Over 22 Years," *International Journal of River Basin Management*, vol. 8, pp. 25-37, 2010.

[8] Z.A. Zakaria and A. Shabri, "Streamflow Forecasting at Ungaged Sites 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, pp. 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 the Bachelor Degree of Education Mathematics from the Universiti Pendidikan Sultan Idris, Malaysia, in 2008, and the Master of Science in Mathematics from the Universiti Teknologi Malaysia, Malaysia, in 2010. Now, she is pursuing Ph.D. degrees 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 nonlinear prediction method, chaos theory and hydrology.

Mohd Salmi Md Noorani (M. S. M Noorani) was born in Kelantan, Malaysia, in 1961. He received the 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 where he was an Assistant Professor, became 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 a President of the Persatuan Sains Matematik Malaysia (2001 until now) and an active fellow for Southeast Asian Mathematical Society (2005 until now), American Mathematical Society for Industrial and Applied Mathematics (USA) and Mathematical Association of America.