2013, Scienceline Publication

Journal of Civil Engineering and Urbanism

Volume 3, Issue 4: 197-200 (2013) (Received: May 07, 2013; Accepted: July 17, 2013; Published: July 30, 2013) ISSN-2252-0430

Forecasting Daily Stream Flows of Vaniar River Using Genetic Programming and Neural Networks Approaches

Kiyoumars Roushangar[1], Fatemeh Vojoudi Mehrabani[2*], and Mohammad Taghi Alami[3]

[1]*Assist. Professor, Dept. of Civil Hydraulics Engineering, University of Tabriz, Tabriz, Iran*
[2]*Dept. of Civil Hydraulics Engineering, University of Tabriz, Tabriz, Iran*
[3]*Assoc. Professor, Dept. of Civil Hydraulics Engineering, University of Tabriz, Tabriz, Iran*

*Corresponding author's Email address: fa.vojoudi@gmail.com

ABSTRACT: This study compares three different artificial intelligence approaches, namely, gene expression programming (GEP), and artificial neural networks (ANN), in daily as well as monthly stream flow forecasting. Daily stream flow data from Vaniar River in the Northwestern Iran were used. Coefficient of determination, root mean square error and scatter index were used to compare simulation results. The study demonstrates that the optimal results were obtained from the triple-input models including stream flows of current and two previous days and that the GEP model performed better than the ANN model in daily stream flow forecasting.

Keywords: Gene Expression Programming, Neural Networks, Forecasting

ORIGINAL ARTICLE

INTRODUCTION

Accurate predictions of stream flow and, consequently, accurate flood forecasts with sufficient lead time are of great importance for protecting vulnerable areas and reducing flood damages. Such predictions are also important for water quality estimates and management as well for fluvial sediment transport studies.

Artificial Intelligence methods have been widely applied for modeling such complicated problems. Numerous applications of artificial neural networks (ANNs) have been addressed in literature (e.g. Smith and Eli, 1995; Minnes and Hall, 1996; Tayfur, 2002; Kisi, 2004a; Kisi, 2004b; Kişi, 2005; Cigizoglu and Kisi, 2005; Kişi, 2006a; Kişi, 2006b; Kişi, 2007)

GP was first proposed by Koza (1992), and is particularly suitable where interrelationships among relevant variables are poorly understood; a theoretical analysis is constrained by assumptions and therefore their solutions are of limited use; and there is a large amount of data in computer readable forms requiring tedious processing. GEP (Gene Expression Programming) is comparable to GP yet evolves computer programs of different sizes and shapes encoded in linear chromosomes of fixed lengths. The chromosomes are composed of multiple genes, each gene encoding a smaller subprogram. As a result GEP surpasses the old GP system in 100-10,000 times (Koza, 1992; Ferreira, 2001b). GP has been applied for rainfall-runoff modeling (Drecourt, 1999; Savic, Walters, and Davidson, 1999; Aytek and Alp 2008), suspended sediment modeling (Aytek and Kisi, 2008), for predicting short term groundwater level fluctuations (Shiri and Kisi, 2011a), estimating daily pan evaporation (Shiri and Kisi, 2011b), wind speed prediction (Kisi, Shiri, and Makarynskyy, 2011a), and predicting daily lake level variations (Kisi, Shiri, and Nikoofar, 2012).

The present study aims at modeling daily stream flow values using GP and ANNs approaches as well as the inter-comparison of the obtained results through using these approaches.

USED METHODOLOGIES

Artificial Neural Networks

ANNs are basically parallel information-processing systems. The internal architecture of ANNs is similar to the structure of a biological brain with a number of layers of fully interconnected nodes or neurons. Each neuron is connected to other neurons by means of direct communication links, each with an associated weight. The neural network usually has two or more layers of neurons in order to process nonlinear signals. The input layer admits the incoming information, which is processed by the hidden layer(s), and the output layer presents the network result. During the learning process, the weights of the interconnections and the neural biases are adjusted in trial and error procedures, to minimize the errors. Two-layer feed-forward networks were employed in this study, with a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer. The hidden-layer-node numbers of each model were determined after an iterative process, because there is not yet a definite theoretical background for determining the interconnections of neurons.

Gene Expression Programming (GEP)

The procedure for GEP modeling of stream flow is as follows. The first step is selecting the appropriate fitness function, which may take various shapes. Here, the root relative square error (RRSE) was selected as fitness function. The second step consists of choosing the set of terminals T and the set of functions F, to create the chromosomes. In the present case, the terminal set includes stream flow valuesQ_{i-2}, $Q_{i-1}$and Q_i, where Q_i

To cite this paper: F. Vojoudi, K. Roushangar, M.T. Alami.2013. Forecasting Daily Stream Flows of Vaniar River Using Genetic Programming and Neural Networks Approaches. *J. Civil Eng. Urban.*,3 (4): 197-200.
Journal homepage http://www.ojceu.ir/main/

denotes the stream flow at *ith* time step. Different mathematical functions were utilized (for building the pars tree), including basic arithmetic operators (+, -, *, /) and basic mathematical functions ($\sqrt{\ }$, $\sqrt[3]{\ }$, ln(x), e^x , x^2 , x^3) as well as trigonometric functions (*sin, cos and arctg*). The third step is choosing the chromosomal architecture; where the length of head h=8 and three genes per chromosomes were employed. The fourth step is selecting the linking function. Here, the sub-trees were linked by addition. The fifth and final step is to choose the genetic operators. The parameters used in each simulation were as follows; number of chromosomes: 30; head size: 8; number of genes: 3; linking function: addition; fitness function error type: root relative squared error; mutation rate: 0.044; inversion rate: 0.1; one point recombination rate: 0.3; two point recombination rate: 0.3; gene recombination rate: 0.1, gene transposition rate: 0.1, insertion sequence transposition rate: 0.1, root insertion sequence transposition: 0.1.

Used Data

Daily stream flow data used in the study were from the Vaniar River in the Northwest of Iran From the 13-years (September1997-Spetember2010) worth of stream flow records were the first 10 years of data (75% of the whole data set) were used for training the models, and the remaining 3 years of records (25% of the whole data set) were reserved for testing process. Figure 1 displays the time series of stream flow values for the study period.

Table 1 depicts some selected statistics of the stream flow data. In the table, the terms X_{mean}, X_{min}, X_{max}, S_d, C_v and C_{sx} denote the mean, minimum, maximum, standard deviation, coefficient of variation and skewness coefficient, respectively. Figure1 displays the time series of the observed stream flows. C_{sx} values in Table 1 indicate that the stream flow data shows scattered distribution.

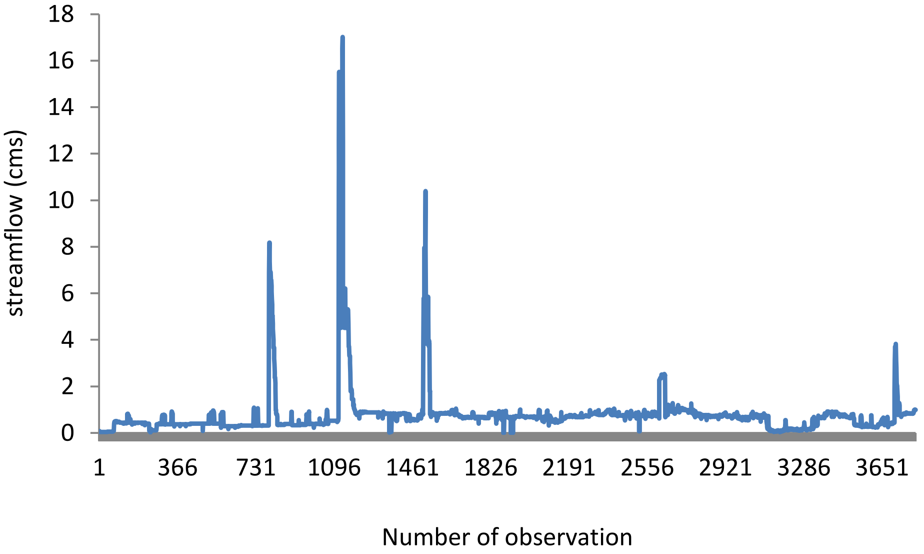

Figure1. Time series of observational stream flow values

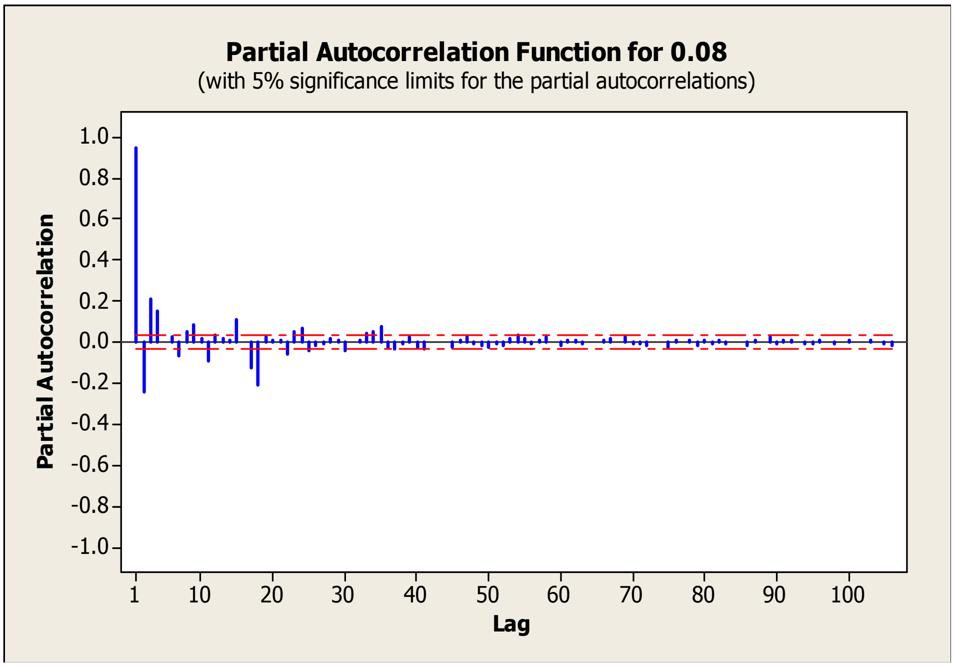

Figure2. PACF of the observed stream flow records

Table1. Statistical parameters of used stream flow records

	Training Period	Testing Period	Validation period
X_{mean}	0.87	0.81	0.56
X_{min}	0.00	0.00	0.04
X_{max}	17.01	2.53	3.83
S_d	1.43	0.32	0.4
C_v	1.65	0.39	0.71
C_{sx}	5.19	3.67	3.17

RESULTS AND DISCUSSIONS

Three statistical evaluation criteria were used to assess the model performance, namely, the Coefficient of Determination (R^2), Root Mean Square Error (*RMSE*) and Scatter Index (*SI*), expression for which are as follows:

$$R^2 = \left(\frac{\sum_{i=1}^{n} \left(Q_{io} - \overline{Q}_o\right)\left(Q_{iM} - \overline{Q}_M\right)}{\sqrt{\sum_{i=1}^{n} \left(Q_{io} - \overline{Q}_o\right)^2 \sum_{i=1}^{n} \left(Q_{iM} - \overline{Q}_M\right)^2}} \right)^2 \qquad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(Q_{io} - Q_{iM}\right)^2} \qquad (9)$$

$$SI = \frac{RMSE}{\overline{Q}_o} \qquad (10)$$

whereQ_{io} and Q_{ie} denote the observed and estimated stream flows, and $\overline{Q}$ represents the mean (observed) stream flow.

The numbers of lags were selected according to the partial auto-correlation function (PACF) of the stream flow data (Figure2). Figure2 clearly indicates that the first four lags have a significant effect on Q_{t+1}. Therefore, the following input combinations were employed:

(i) Q_t
(ii) Q_{t-1}, Q_t
(iii) Q_{t-2}, Q_{t-1}, Q_t

The input variables present the previously recorded stream flows, while the output variable corresponds to the stream flow at time *i+1* as well as *i+30*.

ANN models

Table 2 exhibits the final structure of the used ANNs, e.g. 1-5-1 denotes an ANN comprising 1 input, 5 hidden and 1 output nodes respectively. The number of hidden layer nodes of each ANN model has been determined by trial and error. Table 2 also gives validation statistics for each developed ANN; the double-input ANN model produces better results than the other input combinations with higher correlation and lower error values.

GEP models

The GEP model was applied for predicting stream flows and the validation statistics are demonstrated in Table3 for the optimal input combination (double-input model). It is clear from the table that the GEP model has the lowest RMSE and SI and the highest R^2 for the both prediction intervals.

Table2. Validation statistics of ANN models

	Model Structure	R^2	$RMSE(m^3/s)$	*SI*
$\boldsymbol{Q_{i+1}}$				
Q_t	1-5-1	0.930	0.084	0.25
Q_{t-1}, Q_t	2-5-1	0.930	0.075	0.23
Q_{t-2}, Q_{t-1}, Q_t	3-5-1	0.920	0.090	0.27
Q_{t-3}, Q_{t-2}, Q_{t-1}, Q_t	4-5-1	0.920	0.100	0.30

Table3. Validation statistics of ANN and GEP models for daily and monthly predictions

		R^2	$RMSE(m^3/s)$	*SI*
GEP				
Q_{i+1}		0.940	0.072	0.21
	Q_{i+30}	0.870	0.30	1.30
	ANNs			
	Q_{i+1}	0.930	0.075	0.22
	Q_{i+30}	0.865	0.31	1.3

CONCLUSIONS

The accuracies of three different artificial intelligence techniques were inter-compared in forecasting daily stream flows. The GEP and ANN models were applied to daily stream flow data of Vaniar River in the Northwestern Iran. The input combinations were determined according to the partial auto-correlation function. Double-input GEP and ANN models including stream flows of current and one-immediate previous day showed the best accuracy for each model for the both daily and monthly prediction intervals, with the mean R2 and RMSE of 0.935 and 0.073 m3/s for daily prediction interval and 0.867 and 0.30 m3/s for monthly prediction interval.

REFERENCES

Aytek A. and Alp M. (2008) "An application of artificial intelligence for rainfall runoff modeling", Journal of

Earth Systems and Sciences, Vol. 117, No.2, pp145-155.

Aytek A. and Kisi O. (2008) "A genetic programming approach to suspended sediment modeling", Journal of Hydrology, Vol. 351, pp 288-298.

CigizogluH.K. and Kisi O. (2005) "Flow prediction bt three backpropagation techniques using k-fold partitioning of neural network training data", Nordic Hydrology, Vol. 36, No. 1, pp 49-64.

Drecourt J.P. (1999) "Application of neural networks and genetic programming to rainfall runoff modeling", D2K technical report 0699-1-1, Danish hydraulic institute, Denmark.

Ferreira C. (2001a) "Gene expression programming in problem solving", In: 6th Online World Conference on Soft computing in Industrial Applications (invited tutorial).

Ferreira C. (2001b) "Gene expression programming: a new adaptive algorithm for solving problems", Complex Systems, Vol. 13, No. 2, pp 87-129.

Kisi O, Shiri J, Makarynskyy O. 2011a. Wind speed prediction by using different wavelet conjunction models. International Journal of Ocean and Climate systems, Vol. 2, NO. 3, pp 189-208.

Kisi O, Shiri J, Nikoofar B. (2012) "Forecasting daily lake levels using artificial intelligence approaches", Computers & Geosciences 41, 169-180.

Kisi O. (2004a) "River flow modeling using artificial neural networks", ASCE J. Hydrol. Eng., Vol. 9, NO. 1, pp 60-63.

Kisi O. (2004b) "Multi layer perceptions with Levenberg-marquardt training algorithm for suspended sediment concentration prediction and estimation", Hydrological Sciences Journal, pp 1025-1040.

Kişi Ö. (2005) "Suspended sediment estimation using neuro-fuzzy and neural network approaches", Hydrological Sciences Journal, Vol. 50, No. 4, pp 683-696.

Kişi Ö. (2006a) "Generalized regression neural networks for evapotranspiration modeling", Hydrological Sciences Journal, Vol. 51, No. 6, pp 1092-1105.

Kişi Ö. (2006b) "Evapotranspiration estimation using feed forward neural networks" Nordic Hydrology, Vol. 37, No. 3, pp 247-260.

Kişi Ö. (2007) "Evapotranspiration modeling from climate data using a neural computing technique", Hydrological Processes, Vol. 21, No. 6, pp 1925-1934.

KozaJ.R. (1992) "Genetic Programming: On The Programming of Computers by Means of Natural Selection", The MIT Press, Cambridge, MA, 840 pp.

Minnes, A.W., Hall, M.J., 1996. Artificial neural networks as rainfall-runoff models. Hydrological Sci. J. 41(3), 399-418.

Savic, A.D., Walters, A.G., Davidson, J.W. (1999) A genetic programming approach to rainfall-runoff modeling. Water Resour. Manage.13, 219-231.

Shiri J, Kisi, O. 2011b. Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data in the Khozestan Province (Southwestern Iran). ASCE Journal of Irrigation and Drainage Engineering 137(7): 412-425.

Shiri J. and Kisi O. (2011a) "Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations", Computers & Geosciences, 37(10), 1692-1701.

Smith, J., Eli, R.N. 1995. Neural networks models of rainfall-runoff process. J. Water Resour. Plann. Manage. 121(6), 499-508.

Tayfur, G., 2002. Artificial neural networks for sheet sediment transport. Hydrol. Sci. J. 4(6), 879-892.