Pertanika J. Sci. & Technol. 26 (4): 1777 - 1788 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Rebuilding Hydrological Data with ANN or GA Methods: Case Study-Dez Reservoir, Western Iran

Adib, A.

Civil Engineering Department, Engineering Faculty, Shahid Chamran University of Ahvaz, Ahvaz, Iran

ABSTRACT

Access to sufficient and confident hydrometric data is necessary for water resources management. Most of the Iran's hydrometric stations do not have sufficient data. The method of producing synthetic data should use probability concepts and retains main characteristics of the data, too. In this research, synthetic hydrometric data are generated by the monthly and annual Markov chain method at the Telezang station in the upstream of the Dez River. Using the discharge of the driest day and the wettest day of each month and the generated monthly hydrometric data, the probable highest and lowest daily discharge for each month was calculated. At the end, artificial neural network was trained with a number of observed and generated hydrometric data. The results of artificial neural network were compared with a number of observed hydrometric data which were not used in training of the network. The training of artificial neural network (ANN) with the generated hydrometric data can improve results of network. For more improvement of the results of network, genetic algorithm (GA) is used in its training and optimizing its parameters. The GA method can reduce the MSE (mean of square error) by 97% that of ANN.

Keywords: Artificial neural network, Dez River, genetic algorithm, Markov chain, Telezang station

INTRODUCTION

The forecasting of drought periods is an important task for water resources management. Therefore, accessibility to sufficient and accurate climatic and hydrometric data is necessary. Because of the shortage of hydrometric stations and inaccuracy of their data in developing countries, using of methods for generation of data is essential.

Occurrence of drought is a usual and destructive phenomenon in the Middle East countries. Application of a stochastic method is appropriate for prediction of drought, wet periods and their duration.

ARTICLE INFO

Article history:
Received: 01 November 2017
Accepted: 27 July 2018
Published: 24 October 2018

E-mail addresses:
arashadib@yahoo.com

ISSN: 0128-7680
e-ISSN: 2231-8526

This method must generate synthetic flow discharge data by considering observed flow discharge data (by saving of stochastic characteristics of observed data as mean, variance and governing stochastic distribution). Markov chain method is one of the suitable methods for this purpose. A new method must be utilized for verification of the results of the Markov chain method. This new method can generate synthetic flow discharge data. The generated data with the new method should be compared to observed flow discharge data and generated flow discharge data by Markov chain method. In this regard, artificial neural network is an appropriate method and has necessary stated characteristics.

Hydrologists utilized different Markov chain methods for generating daily and monthly flow discharge in rivers. Xu et al. (2001) utilized MACP (Markov Auto- Correlation Plus) method for generation of daily, monthly and annual flow discharge in the Wupper River in Germany. They studied about spatial correlation between two adjacent stations. Also Xu et al. (2003) used MCCP (Markov Cross- Correlation Plus) method for forecasting daily stream flow of the Wupper River in Germany. They determined wet and dry periods. Aksoy (2003) utilized Markov chain method for prediction of daily stream flow. He determined wet and dry days in different watersheds. Szilagyi et al. (2006) used hybrid Markov chain method for forecasting daily stream flow of the Tisza River and its top branches (Szamos, Bodrog, & Kraszna) in Hungary. They utilized observed data from 1951 to 2000 and considered two dry and wet states. They evaluated transfer probability between two states (dry to dry, wet to dry, dry to wet and wet to wet). Sarlak et al. (2009) predicted annual stream flow of the Goksu River in Turkey based on oscillation of sea water level in North Atlantic Ocean. Stošić et al. (2012) applied Monte Carlo Markov Chain (MCMC) for prediction of discharge and velocity profile in the Exu River and the Capibaribe River. These rivers are in the northeast of Brazil.

Also, some of the hydrologists used ANN for generation of monthly flow discharges in rivers. Anmala et al. (2000) applied a feed forward ANN and a recurrent ANN for prediction of runoff in three watersheds of Kansas, USA. The inputs layer of ANN included monthly precipitation and temperature and the output layer included monthly runoff. Cigizoglu (2003) utilized combination of artificial neural network and autoregressive–moving-average (ARMA) model for forecasting of monthly stream flow of Karahacili hydrometric station on the Goksu River in Turkey (east of Mediterranean Sea). He used ~~of~~ perceptron ANN and trained it with the observed and the data generated by ARMA model. The MSE of the ANN trained with the generated data by ARMA model was less than that of the ANN trained with the observed data. Keskin and Taylan (2010) used combined McCulloch and Pitts ANN and adaptive neuro-fuzzy inference system (ANFIS) methods for prediction of monthly stream flow of the Alara River in Turkey. They concluded that ANN could predict monthly stream flow better than ANFIS. Raman and Sunilkumar (2010) applied a perceptron ANN for generation of monthly stream flow of main river of Bharathapuzha watershed in India. They concluded that ANN could predict monthly stream flow better than ARMA model. Ochoa-Riveria et al. (2002) used hybrid artificial neural network for generation of monthly flow discharge. Results of their network were more accurate than results of ARMA model. Ahmed and Sarma (2007) applied ANN, ARMA and Thomas-Fiering models for prediction of monthly stream flow of the Pagladia River (a top branch of the Brahmaputra River) in Himalayan region of India. They observed that ANN could predict monthly stream flow more accurately than other methods. Lee and Kang (2016) used ANN for simulating daily flow discharge in the Bocheong-cheon watershed (in the centre of South Korea). Parsaie et al. (2017) applied ANFIS method in prediction of flow discharge in the compound open channel. The coefficient of determination (0.98) and root mean square error (0.029) illustrated the accuracy of ANFIS method. Young et al. (2015) combined two hybrid models and Hydrologic Engineering Center- Hydrologic Modeling System (HEC-HMS) model. They linked this combined model to genetic algorithm neural network (GANN) and ANFIS and applied it in the Laonong Creek basin in southern Taiwan. This method had a high accuracy in prediction of hourly runoff discharge. Khan et al. (2016) applied ANN for forecasting of flow discharge and water surface elevation in the Ramganga River catchment of the Ganga Basin (in India). The mean square errors (MSE) of flow discharge and water surface elevation were 0.046 and 0.012.

The biggest gap and problem of the above researches is that they use only one method which reduces the accuracy of generated data. Therefore, some of researchers combined the Markov chain and ANN methods for forecasting monthly and annual flow discharges, rather than using only one of these methods. Therefore, they covered the gap between the applications of each method separately. For example, one can refer to the following researches:

Adib and Mahmoodi (2017) utilized the Markov chain method for prediction of flow discharge in Idenak hydrometric station, located at the Marun River. This river is located in southwest of Iran. They predicted suspended sediment load using ANN and generated flow discharges by the Markov chain method. Also, they optimized the parameters of ANN with the GA and the GA could reduce the Normalized Mean Square Error (NMSE) by 20%. Rezaeianzadeh et al. (2016) used the Markov chain method and ANN for prediction of inflow discharge to the Doroodzan reservoir in south of Iran. A combination of the two methods increased the accuracy of forecasting the droughts and flow discharges. The results of this combined method are more reliable too.

In this research, a perceptron ANN was utilized for prediction of mean monthly and annual discharge in hydrometric stations. For training of this network, observed and generated data by Markov chain method were used. The GA method was used in optimization of parameters of ANN for improving ANN method.

MATERIALS AND METHODS

The Telezang Hydrometric Station of the Dez River

A hydrometric station was considered on the Dez River (Telezang station), for generating the synthetic data. This station is at the upstream of the dams constructed on the Dez River. Therefore, these dams cannot regulate discharge of fluvial flow in this station. This hydrometric station was constructed in 1955 and its elevation is 468 m above sea level. The area of its watershed is 16130 km^2. It is in 48°46’3’’ E and 32°49’19’’ N. The mean of annual flow discharge, precipitation and temperature are 249.569 CMS, 76.2 mm and 33.6°C, respectively, in the Telezang hydrometric station. The vicinity of this station is shown in Figure 1 and the stochastic parameters of different months are illustrated in Table 1. The mean and variance of monthly and annual flow discharges in this table pertain to data of several years (1977-2015).

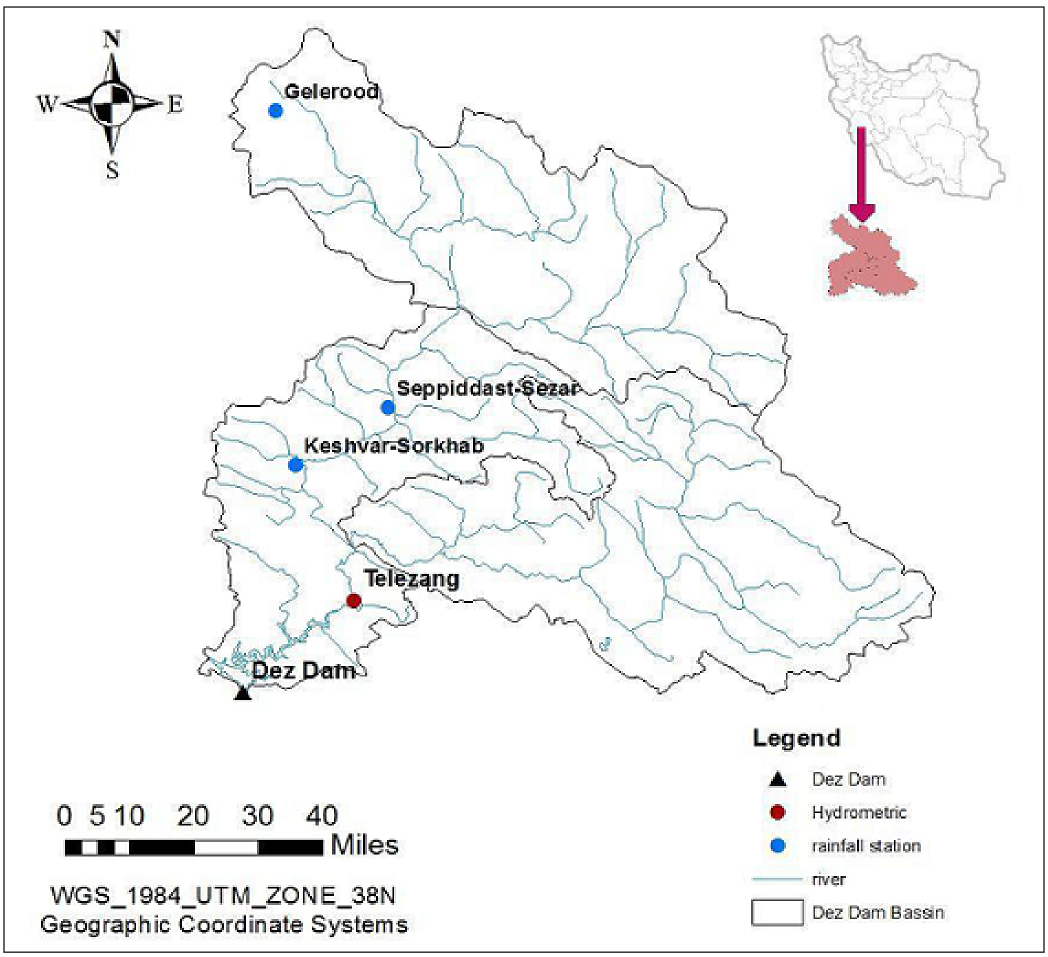

Figure 1. The region of Telezang Station- Banihabib et al. (2017)

Table 1
Stochastic parameters of the Telezang station

Month	Mean (CMS)	Variance (CMS^2)	Governing stochastic distribution
Jan	206.512	14390.96	Log PersonIII
Feb	293.9	33002.18	Log PersonIII
Mar	418.575	68039.38	Log PersonIII
Apr	580.092	77971.41	Log PersonIII
May	510.13	64629.52	Log PersonIII
Jun	276.282	14813.83	Log PersonIII
Jul	165.122	4328.768	Log PersonIII
Aug	110.712	1881.624	Log PersonIII
Sep	81.535	762.552	Log PersonIII
Oct	68.643	393.278	Log normal 2 Para
Nov	101.608	4767.981	Log PersonIII
Dec	181.712	21875.2	Log PersonIII
Year	249.569	8915.936	Log PersonIII

METHODS

In this research, monthly and annual Markov chain methods were applied.

This method uses (1) for annual Markov chain method:

$$Q_{i+1} = \overline{Q} + r_Q(1)(Q_i - \overline{Q}) + t_{i+1} S_Q (1 - r_Q(1)^2)^{1/2} \quad [1]$$

Where: Q is flow discharge (CMS), S is the standard deviation of flow discharges (CMS), r is correlation coefficient between successive flow discharges and t is the value of t-distribution.

For monthly Markov chain method (1) is converted to (2) and (3).

$$Q_{i,j+1} = \overline{Q}_{j+1} + b_j (Q_{i,j} - \overline{Q}_j) + t_{i,j+1} S_{Q_{j+1}} (1 - r_j^2)^{1/2} \quad [2]$$

$$b_j = r_j \frac{S_{j+1}}{S_j} \quad [3]$$

Where: i is index of year and j is index of month.

In this research, a series of random numbers were produced using (4):

$$P_{i+1} = \mathrm{Mod}(aP_i + b)/c \quad [4]$$

Where: P is random number and *a, b* and c are constants chosen optionally by the users. The number of digits of the random number is equal to the number of digits of c.

Neural networks can be divided into two types based on their structures: feed forward networks and recurrent networks. In this research, a feed forward network is used. Application of feed forward networks is more common than of recurrent networks in water engineering.

In a feed forward network, the nodes are grouped into layers. Signals flow from the input layer through the network towards the output layer, via unidirectional connections. The nodes are connected from one layer to the next one, but not within the same layer. A multi-layer perceptron (MLP) is a feed forward network with one or more hidden layer. Given a training set of input-output data, the most common learning rule for multi-layer perceptrons is the back propagation algorithm. A neural network with such a type of learning algorithms is usually referred to as back propagation network (BPN).

The genetic algorithm method is utilized for optimizing the training step of artificial neural network. This procedure modifies parameters of artificial neural network such as momentum and learning rate. However, this method increases the time of training step, as the genetic algorithm, searches for the global optimum of fitness between results of artificial neural network and observed data. Also, the fitness between the results of this method and those of the Markov chain method is better than the fitness between the results of ordinarily trained artificial neural network and Markov chain methods.

Also mean square error (MSE) is a performance criterion in this research.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Q_{gen} - Q_{obs})^2 \qquad [5]$$

Where: n is number of data, Q_{gen} is generated streamflow discharge and Q_{obs} is observed streamflow discharge.

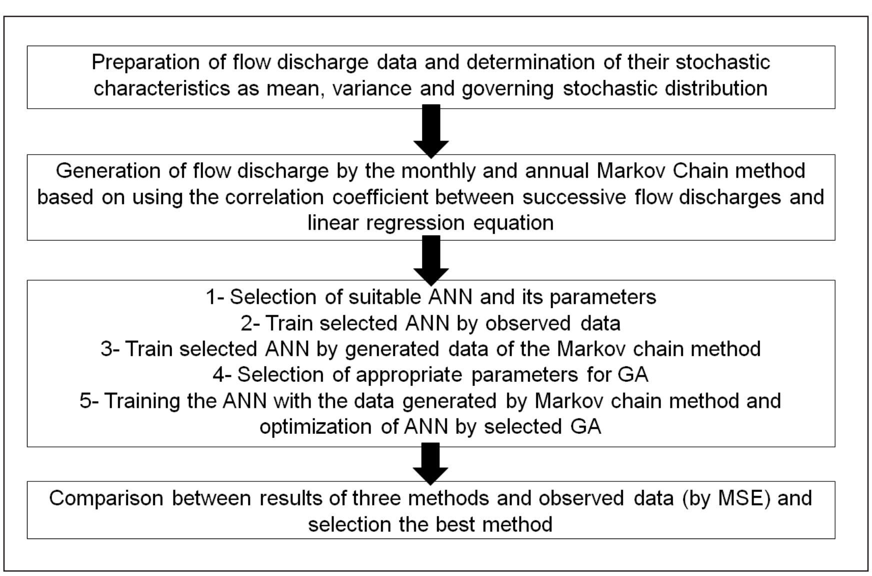

Figure 2. The Flowchart of this research methodology

RESULTS AND DISCUSSION

Results of Monthly and Annual Markov Chain Method

Eighty time series of hydrometric data were generated with the monthly and annual Markov chain method. The number of data in each time serial is equal to 50. If a data of a time series is lower than the mean of the that time series, it shows a drought period, also, if a data of a time series is greater than the mean of that time series, then it shows a wet period. In this research, 100 time series were generated with the monthly and annual Markov chain method. Each annual time serial and monthly (for 12 months) time serial has 50 data. The mean of annual and monthly generated data (50*100 data is number of generated annual data and data of each month) must be compared to annual and monthly observed data. The driest time serial, among the produced 100 time series of generated annual and monthly data, is the one with the minimum mean of its monthly flow discharge. Also, the wettest time serial is the one that with maximum mean of its monthly flow discharge. The driest and wettest annual and monthly time series should also be compared to observed annual and monthly data too.

The results of monthly and annual Markov chain method are illustrated in Table 2.

Table 2
Comparison between generated time series and observed data in the Telezang Station

Month	Difference between the mean of discharges of generated the driest time serial and the mean of observed data (%)	Difference between the mean of discharges of generated the wettest time serial and the mean of observed data (%)	Difference between the mean of discharges of generated time series and the mean of observed data (%)
Jan	-15.26	14.22339	-0.71667
Feb	-17.782	16.15243	-1.01837
Mar	-17.498	15.33035	-1.07149
Apr	-14.829	9.971349	-1.17826
May	-13.805	9.750652	-1.05659
Jun	-6.3949	4.624261	-0.74236
Jul	-4.0243	2.957207	-0.35731
Aug	-4.8369	2.801864	-0.48956
Sep	-4.1209	2.60379	-0.26246
Oct	-8.9011	7.380214	-0.7881
Nov	-17.18	29.75947	-0.872
Dec	-16.315	35.02355	0.310381
Year	-14.005	12.16097	-1.162

Results of Daily Markov Chain Method

In this research, a new method was applied for determination of the driest and wettest days in each month. The method is based on the following procedure:

The discharge of the driest day generated in a particular month is equal to the product of observed discharge of the driest day in that particular month and the ratio of the average of monthly discharges of the generated driest time series and the average of the observed data of corresponding time series.

$$Q_{Gen\ the\ driest\ day} = Q_{Obs\ the\ driest\ day} * (Q_{Avg\ of\ gen\ the\ driest\ time\ series} / Q_{Avg\ of\ obs\ data}) \quad [6]$$

Also, the discharge of the generated wettest day in each specific month is equal to product of discharge of the observed wettest day in that particular month and the ratio of the mean of discharges of the generated wettest time series and the mean of the observed data in that time series.

$$Q_{Gen\ the\ wettest\ day} = Q_{Obs\ the\ wettest\ day} * (Q_{Avg\ of\ gen\ the\ wettest\ time\ series} / Q_{Avg\ of\ obs\ data}) \quad [7]$$

The results of daily Markov chain method are illustrated in Table 3.

Table 3
Generated discharges of the driest and wettest day in the Telezang Station

Month	Generated discharge of the driest day (CMS)	Generated discharge of the wettest day (CMS)
Jan	46.61	3213.1
Feb	60.02	5216.4
Mar	84.98	3658.28
Apr	114.1291	3179.27
May	113.78	3492.27
Jun	78.44	1011.72
Jul	57.2	519.93
Aug	44.73	306.35
Sep	36.34	226.75
Oct	33.71	164.29
Nov	34.78	4130.24
Dec	49.54	6115.22

Estimation of the Longest Drought Period

The longest observed drought period in the Dez watershed was from 1957 to 1968 (11 years). The Markov chain method could generate it. The longest generated drought period is, also, equal to 11 years. The longest drought period is the longest number of consecutive years that their annual flow discharges are less than the mean annual flow discharge of the same time series data. It means that the annual flow discharge from 1957 to 1968 was less than 249.569 CMS (mean observed annual flow discharge in the Telezang station).

Results of ANN and ANN Trained with the GA

The GA is a popular method in civil and water engineering for optimization of nonlinear problems. Training with the GA method can improve the results of ANN and reduce its MSE. This method can easily be linked to numerical models and ANN, using MATLAB toolboxes. While regular gradient-based technique can not optimize the non linear problems.

The objective function of the GA method is:

$$\text{Objective function: } \quad Min\sum_{i=1}^{m}(\text{Output of ANNs - Desired output})^2 \qquad [8]$$

In order to stop the training process of ANN, a convergence criterion must be considered. This criterion is shown in below:

$$\text{Abs (Output of ANN- Desired output)} < \text{error tolerance} \qquad [9]$$

The training of ANN will be terminated if this criterion is satisfied for all of the outputs of ANN.

Three methods of training, used in this research, include:

1- Training of ANN with the observed data

2- Training of ANN with the data generated by the Markov chain method

3- Training of ANN with the data generated by the Markov chain method and optimization of the parameters of ANN with the GA method

Sixty percent of data were used for training of ANN. Also 15 percent of the data were used in validation of ANN and the remaining 25 percent were utilized in testing process of ANN. The best architecture and parameters of perceptrons ANN are selected using the trial error method. This network has one input node (the mean discharge of previous month or year), one output node (the mean of discharge of present month or year) and one hidden layer with two nodes. The momentum and learning rate of network are 0.6 and 0.1 respectively. The transfer function is assumed tangent hyperbolic. This network utilizes from back propagation for training and error tolerance of ANN is 0.01.

The GA method optimizes momentum and learning rate of network. The GA method results in a momentum and learning rate of 0.55 and 0.08, respectively. The characteristics of the applied GA in this research are:

Rate of crossover=0.8, Type of mutation= Uniform, Type of crossover= Heuristic, Selection method= Stochastic universal sampling, Number of generations= 3000, Population of each generation= 120. Also, mutation rates for different generations are:

Mutation rate=0.3   if (no of generation<700)

Mutation rate= (-0.295/1300)*(no of generation-700) +0.3 if (700<no of generation<2000)

Mutation rate=0.005   if (no of generation>2000)

On average, the training of ANN with synthetic data reduces the mean of square error (MSE) by 48%, whilst the training of ANN with synthetic data followed by optimization of ANN with the GA method further reduces the MSE by 97% overall. For example while the MSE of annual flow discharge in the training stage of a regular ANN is 198182.88 CMS^2, it is 4999.59 CMS^2 for NN+GA+MARKOV, showing a 97.48% reduction, 100-(4999.59*100/198182.88)=97.48%. The MSE of different training methods of ANN are presented in Tables 4 and 5 for training and testing stages respectively. Also in this research, a recurrent network with similar parameters and architecture to applied MLP network was used. But, the MSE of the recurrent network was almost twice the MSE of applied MLP network. Therefore MLP network was selected as superior network.

Table 4
Comparison of MSE of different training methods of ANN at the Telezang Station (training stage)

Month	MSE (CMS)^2		
	ANN+GA+MARKOV	ANN+ MARKOV	ANN
Jan	5364.44	98912.1	177597.03
Feb	22058.71	262275.84	447604.5
Mar	16271.18	397336.8	802419.6
Apr	39124.15	908614.8	1758405.9
May	16949.89	435803.1	886870.2
Jun	4170.12	100122	203548.2
Jul	1707.82	42817.17	84591.91
Aug	877.9	21666.73	41993.81
Sep	586.9	14789.3	19200.11
Oct	419.92	7333.62	14061.47
Nov	714.98	18190.14	33286.65
Dec	4347.4	73732.38	134681.82
Year	4999.59	102597.9	198182.88

Table 5
Comparison of MSE of Different Training Methods of ANN in the Telezang Station (testing stage)

Month	MSE (CMS)^2		
	ANN+GA+MARKOV	ANN+ MARKOV	ANN
Jan	6200.732	112017.6	195701.1
Feb	28060.53	293814.6	501693.5
Mar	16040.68	469974.5	968437.9

Table 5 *(continue)*

Month	MSE (CMS)^2		
	ANN+GA+MARKOV	ANN+ MARKOV	ANN
Apr	45594.35	958903.2	1836703
May	25781.85	571080.6	1135659
Jun	7039.602	121709.7	236454.5
Jul	2589.605	44962.59	85645.8
Aug	1211.123	22064.8	41277.78
Sep	640.8639	13046.96	24454.46
Oct	582.5766	9404.678	18058.33
Nov	720.6559	16845.31	30839.54
Dec	5274.558	70838.36	124535.5
Year	6763.516	126010.8	239793.2

CONCLUSION

In this research, it was proven that the Markov chain method could generate hydrometric data. This method can predict the duration of drought period precisely. Also, it produces both dry and wet time series. By generation of dry time series, managers and designers can prepare plans for water resource management under critical conditions and meet the water demand. In the other hand, to control the flood damages, designers can use wet time series generated by Markov chain method. The mean of generated time series data and observed data are very close. This proves the accuracy of Markov chain method of data production.

ANN trained with the data generated by Markov chain method and optimized with the GA is the best method for evaluation of correctness, (verification), of generated data. The data produced by this network have the best fitness to observed data. The MSE of this method was very low. Also, the MSE of ANN trained with the data generated by Markov chain method was less than the MSE of ANN that trained with the observed data. This proved that the variation of data generated by Markov chain method was considerably more than that of the observed data, and training ANN with the data generated improves the performance of networks.

REFERENCES

Adib, A., & Mahmoodi, A. (2017). Prediction of suspended sediment load using ANN GA conjunction model with Markov chain approach at flood conditions. *KSCE Journal of Civil Engineering, 21*(1), 447-457.

Ahmed, J. A., & Sarma, A. K. (2007). Artificial neural network model for synthetic streamflow generation. *Water Resources Management, 21*(6), 1015-1029.

Aksoy, H. (2003). Markov chain-based modeling techniques for stochastic generation of daily intermittent streamflows. *Advances in Water Resources, 26*(6), 663-671.

Anmala, J., Zhang, B., & Govindaraju, R. S. (2000). Comparison of ANNs and empirical approaches for predicting watershed runoff. *Journal of Water Resources Planning and Management- ASCE, 126*(3), 156-166.

Banihabib, M. E., Ahmadian, A., & Jamali, F. S. (2017). Hybrid DARIMA-NARX model for forecasting long-term daily inflow to Dez reservoir using the North Atlantic Oscillation (NAO) and rainfall data. *GeoResJ, 13*, 9-16.

Cigizoglu, H. K. (2003). Incorporation of ARMA models into flow forecasting by artificial neural networks. *Environmetrics, 14*(4), 417-427.

Keskin, M. E., & Taylan, E. D. (2010). Artificial intelligent models for flow prediction: a case study on Alara Stream. *Journal of Engineering Science and Design, 1*(1), 8-13.

Khan, M. Y. A., Hasan, F., Panwar, S., & Chakrapani, G. J. (2016). Neural network model for discharge and water-level prediction for Ramganga River catchment of Ganga Basin, India. *Hydrological Sciences Journal, 61*(11), 2084-2095.

Lee, D. H., & Kang, D. S. (2016). The application of the artificial neural network ensemble model for simulating streamflow. *Procidia Engineering, 154*, 1217-1224.

Ochoa-Riveria, J. C., Garcia-bartual, R., & Andreu, J. (2002). Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks. *Hydrology and Earth System Sciences, 6*(4), 641–654.

Parsaie, A., Yonesi, H., & Najafian, S. (2017). Prediction of flow discharge in compound open channels using adaptive neuro fuzzy inference system method. *Flow Measurement and Instrumentation, 54*, 288-297.

Rezaeianzadeh, M., Stein, A., & Cox, J. P. (2016). Drought forecasting using Markov chain model and artificial neural networks. *Water Resources Management, 30*(7), 2245-2259.

Raman, H., & Sunilkumar, N. (1995). Multivariate modeling of water resources time series using artificial neural networks. *Hydrological Sciences Journal, 40*(2), 145–163.

Sarlak, N., Kahya, E., & Anwar Bég, O. (2009). Critical drought analysis: case study of Göksu River (Turkey) and North Atlantic oscillation influences. *Journal of Hydrologic Engineering- ASCE, 14*(8), 795-802.

Stošić, B. D., Silva, J. R. S., Filho, M. C., & Cantalice, J. R. B. (2012). Optimizing river discharge measurements using Monte Carlo Markov Chain. *Journal of Hydrology, 450*, 199-205.

Szilagyi, J., Balint, G., & Csik, A. (2006). Hybrid, Markov Chain-based model for daily streamflow generation at multiple catchment sites. *Journal of Hydrologic Engineering- ASCE, 11*(3), 245-256.

Xu, Z. X., Schumann, A., & Brass, C. (2001). Markov autocorrelation pulse model for two sites daily streamflow. *Journal of Hydrologic Engineering- ASCE, 6*(3), 189-195.

Xu, Z. X., Schumann, A., & Li, J. (2003). Markov cross-correlation pulse model for daily streamflow generation at multiple sites. *Advances in Water Resources, 26*(3), 325-335.

Young, C. C., Liu, W. C., & Chung, C. E. (2015). Genetic algorithm and fuzzy neural networks combined with the hydrological modeling system for forecasting watershed runoff discharge. *Neural Computing and Applications, 26*(7), 1631-1643.