A comparison between artificial neural network method and nonlinear regression method to estimate the missing hydrometric data

J. Bahrami, M. R. Kavianpour, M. S. Abdi, A. Telvari, K. Abbaspour and B. Rouzkhash

ABSTRACT

Missing values are a common problem faced in the analysis of hydrometric data. The need for complete hydrological data, especially hydrometric data for planning, development and designing hydraulic structures, has become increasingly important. Reasonably estimating these missing values is significant for the complete analysis and modeling of the hydrological cycle. The major objective of this paper is to estimate the missing annual maximum hydrometric data by using artificial neural networks (ANN). Sixteen stations, with 28 years of measurements, in the catchment area of the Sefidroud watershed in the north of Iran were selected for this investigation. Comparison between the results of ANN and the nonlinear regression method (NLR) illustrated the efficiency of artificial neural networks and their ability to rebuild the missing data. According to the coefficient of determination (R^2) and the root mean squared value of error (RMSE), it was concluded that ANN provides a better estimation of the missing data.

Key words | artificial neural network, missing hydrometric data, nonlinear regression

J. Bahrami (corresponding author)
M. R. Kavianpour
Civil and Structural Engineering Department,
K. N. Toosi University Technology,
Tehran, Iran
E-mail: *Jbahrami@uok.ac.ir*

J. Bahrami
M. S. Abdi
Civil Engineering Department,
University of Kurdistan,
PO Box 416, Sanandaj, Iran

A. Telvari
Department of Civil Engineering,
Islamic Azad University,
Ahvaz, Iran

K. Abbaspour
Eawag, Swiss Federal Institute for Aquatic Science and Technology,
Ueberlandstrasse 133,
P. O. Box 611, 8600 Duebendorf, Switzerland

B. Rouzkhash
Iran Water Resources Management Company,
No. 81, Felestin Ave., Tehran, Iran

NOTATION

R^2	coefficient of determination
Q_{max}	maximum annual flow discharge
ANN	artificial neural network
NLR	nonlinear regression
RMSE	root mean squared value of error
NRM	normal ratio method
CM	correlation method
MLR	multi-linear regression
SRC	sediment rating curve
SSC	suspended sediment concentration
CCANN	cascade correlation artificial neural network
ANFIS	adaptive neuro-fuzzy system
I/O	input/output
B-P	back propagation
MLP	Multi-Layer Perceptron
NMSE	normalized mean square error

doi: 10.2166/hydro.2010.069

INTRODUCTION

To design hydraulic structures, such as dams and bridges, adequate evaluation of floods based on hydrometric data is essential. To do this, it is necessary to access a complete and large enough set of data. In most cases, the records of hydrological processes face some missing observations or may not be enough to cover the required period of data. These gaps in the data might be related to a number of

factors, such as interruption of measurements due to equipment failure, extreme natural effects due to hurricanes or landslides, human activities in the forms of war or civil unrest, and mishandling of observed records. Most hydrological models do not tolerate missing observations and thus data in-filling techniques have evolved to deal with incomplete datasets (Elshorbagy *et al.* 2000).

Attracted by the importance of estimating missing data, hydrological researchers have adopted and developed various models and techniques to deal with this problem. The efforts are devoted to extend short measurements by adding lengthy segments of estimated data, and to recover the gaps of observations (Elshorbagy *et al.* 2000). There are some methods used for rainfall data gap filling and/or flow data gap filling such as the interpolation technique (Linsley *et al.* 1988). Two well-known methods which are commonly used for filling flow data gaps are the normal ratio method (NRM) and especially the correlation method (CM) between the gauging stations. One of the well-known methods which is commonly used for filling the data gaps is called the nonlinear regression method (NLR) (Dastorani *et al.* 2010).

Also, in the last few decades, many types of data-driven techniques and models have been developed, which reflect the inherently stochastic nature of hydrological processes. This has led to an increasing interest in the ANN and fuzzy logic techniques (Dastorani *et al.* 2010). In many instances, the relationships between the predicting and predicted variables are not truly linear, and the nonlinear forms of their relationships are very difficult to be known. In those cases, neural network modeling, which is a computational approach inspired by studies of the brain and nervous systems in biological organisms, can be a well-adapted alternative for non-parametric solution of this problem (Bin & Takase 2006). This technique has been successfully applied for many hydrological and hydraulic simulations such as flood forecasting, rainfall–runoff modeling, streamflow forecasting and water level prediction (Bustami *et al.* 2007). Hsu *et al.* (1995) have shown that the ANN approach provides a better representation of the rainfall–runoff relationship of a medium-sized basin than does the ARMAX approach or the Sacramento soil moisture model. Rajaee *et al.* (2009) studied ANN, multi-linear regression (MLR) and conventional sediment rating curve (SRC) models for daily simulation of suspended sediment concentration (SSC) in two hydrometric stations. Comparison of the models' results indicated that the ANN model has more ability in predicting SSC in comparison with the MLR and SRC models.

Recent studies on ANN applications in the area of hydrology include rainfall–runoff modeling (Cigizoglu 2003; Wilby *et al.* 2003; Lin & Chen 2004), river stage forecasting (Imrie *et al.* 2000; Lekkas *et al.* 2001; Campolo *et al.* 2003), reservoir operation (Jain *et al.* 1999), flow aeration downstream of outlet gates (Kavianpour & Rajabi 2005), land drainage design (Shukla *et al.* 1996; Yang *et al.* 1998), aquifer parameter estimation (Srinivasa 1998), describing soil water retention curve (Jain *et al.* 2004) and optimization or controlling problems (Wen & Lee 1998; Bhattacharya *et al.* 2003). Diamantopoulou *et al.* (2006) have shown that a three-layer cascade correlation artificial neural network (CCANN) model gave the best results for estimating missing monthly values of water quality parameters in rivers.

Recently, applications of ANN and adaptive neuro-fuzzy system (ANFIS) models in the area of monthly missing data, using data from neighboring sites, was published by Dastorani *et al.* (2010). In their studies, artificial neural networks and the adaptive neuro-fuzzy system (ANFIS), together with two traditionally used methods of the normal ratio method and the correlation method (linear regression method), were employed. According to the results, the ANFIS technique predicted missing data, especially in arid and semiarid areas with variable and heterogeneous data. It is should be mentioned that the linear methods which were used to compare with, are unable to consider the nonlinear characteristics of hydrological data. Therefore, in the present study the results are compared with those obtained from the nonlinear regression method (NLR).

In this paper, the ANN, which is formulated based on the pre-collected measurements, has been applied to evaluate the yearly missing data. The work deals with the reconstruction of missing maximum annual flow discharge (Q_{max}) from the observed one using the ANN model. It is important to realize that ANN has mostly been applied to monthly data, which are most likely different from annual data in terms of their fluctuations and availability. In the previous studies (Dastorani *et al.* 2010), the missing data in each region were reconstructed by one or two stations and thus the conclusions were based on the results of this

station(s), but in the present study, 15 stations were used and thus the accuracy of results should have been improved. The study was performed in mostly humid and mountainous regions with many differences in physiographic, meteorological and hydrological characteristics from arid and semiarid areas studied by Dastorani *et al.* (2010).

METHODS

Artificial neural network

An artificial neural network is defined as a structure composed of a number of interconnected units or artificial neurons. Each unit has input/output (I/O) characteristics and implements a local computation or function. The output of each unit is determined by its I/O, its interconnection to other units and possible external inputs.

Development of an efficient technique for data providing and prediction is getting essential in scientific research and engineering projects. For this purpose, ANN with back propagation (B-P) networks might be one of the most popular techniques in recent years. A B-P network consists of at least three layers of units: an input layer, hidden layer(s) and an output layer. Typically, units are connected in a feed-forward manner with input units completely connected to units in the hidden layer and hidden units completely connected to units in the output layer. When a B-P network runs and is cycled, an input pattern is propagated forward to the output units through the intervening input-to-hidden and hidden-to-output weights. The process of learning starts within a training phase and each input pattern in a training set is applied to the input units and then propagated forward. As the forward processing arrives at the output layer, the forward pattern is then compared with the observed output pattern to calculate an error signal. The error signal for each such target output pattern is then back-propagated from the output layer to the input layer in order to appropriately amend or tune the weights in each layer of the network. After a B-P network has learned the correct classification for a set of inputs, it can be tested on a second set of inputs to see how well it classifies untrained patterns. Thus, an important consideration in applying B-P learning is how well the network generalizes (Chen *et al.* 2007). In this study, a Multi-Layer Perceptron (MLP) network has been used. A back-propagation algorithm has been used by different investigators in the field of hydrology and water resources and resulted in successful results. However, we applied different algorithms with the best results based on the back-propagation algorithm. The schematic structures of this network are shown in Figure 1.

Multi-layer perceptron networks

In a forecasting application, the ANN is iteratively presented with a set of input–output examples known as the training set for the system, from which the network learns the values of the internal parameters of the model. The learning process involves comparing each of the model forecasts with the known correct answer, and adjusting the synaptic weights based on a selected learning rule to minimize the network error. The number of neurons in the input and output layers of the neural network are selected based on the number of input features and the number of outputs defined for the problem. The hidden-layer neurons provide the network with its ability to generalize on new data. Theoretically, a network having one hidden layer with a sufficient number of neurons is capable of approximating any continuous function (Hornik *et al.* 1989). ANNs with one and occasionally two hidden layers are widely used. Since there is no definite rule for fixing the hidden layer's size, a neural network is initially

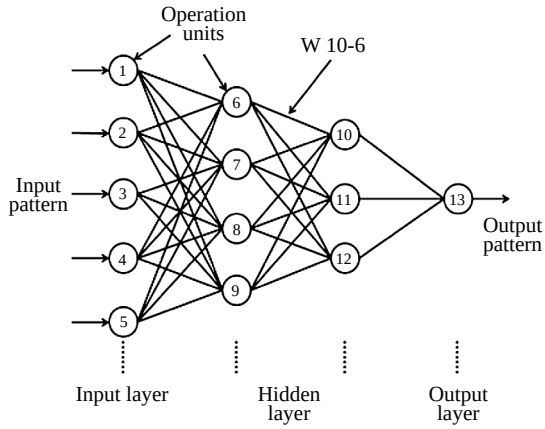

Figure 1 | The ANN network structure.

Table 1 | Details and characteristics of stations and data used in this research

Stations	River	Area above station (km²)	Elevation (m) above sea level	Number of existing data	Number of remained data after outlier tests
Motorkhaneh	Aidoghmosh	1,767	1,060	19	17
Baghkolaye	Almoot taleqan	646	1,350	28	26
Hashtadjoft	Gamishgairoud	1,787	1,450	18	16
Qaregoni	Qezel Ozan	19,401	1,410	18	17
Gilvan	Qezel Ozan	49,068	320	26	24
Ostor	Qezel Ozan	42,022	1,000	23	20
Poldokhtar Mianeh	Qezel Ozan	32,923	1,080	24	20
Kohsalar Mianeh	Shahr chai	1,001	1,500	28	25
Galinak	Shahroud	716	1,770	23	22
Loshan	Shahroud	4,238	300	22	21
Yankikand	Sojasroud	2,494	1,420	19	18
Palti	Tahamchai	173	1,700	16	16
Dehgolan	Talvar	216	1,820	21	20
Salamatabad	Talvar	6,236	1,650	20	20
Sarcham	Zanzanroud	4,546	1,150	16	15
Siahdasht	Almoot	2,300	970	16	15

trained with a minimum number of neurons, and the size is increased until the ANN is able to learn the input–output relationship (Gopakumar *et al.* 2007).

The forward pass of MLP network training using the iterative back-propagation algorithm begins by presenting an input pattern $x(t)$ from the training dataset to the input layer neurons. In the backward pass, the network error is back-propagated through the network and synaptic weights are modified as per the generalized delta rule by considering each of the two adjacent layers starting from the output layer up to the input layer. The rate of change of error, δ_j, with respect to the input to neuron j is computed as

$$\delta_j = \begin{cases} (d_k - o_k)o_k(1 - o_k) \\ y_i(1 - y_i)\sum_k \delta_k W_{jk} \end{cases} \quad (1)$$

where y_i is the output from neuron i and W_{jk} is the weight connecting the output of neuron j to the input of neuron k.

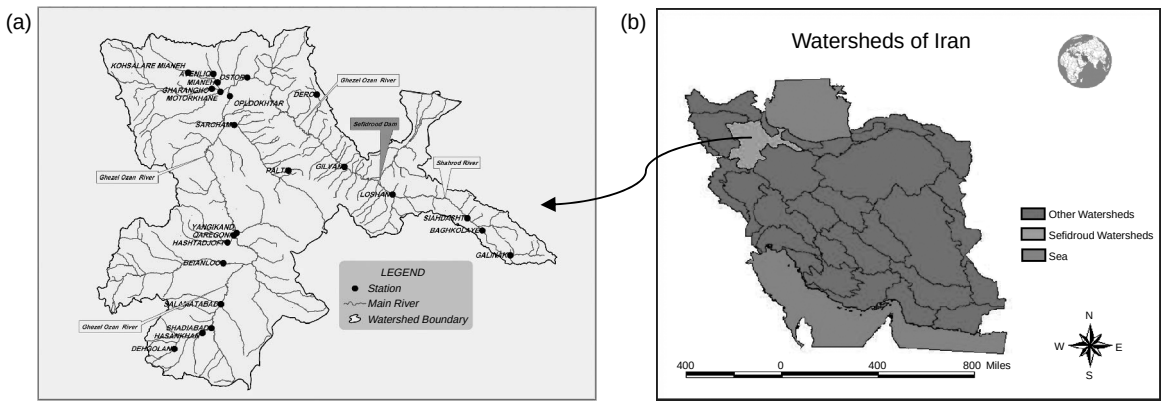

Figure 2 | (a) Hydrometric stations in the Sefidrood watershed. (b) Watersheds of Iran.

Table 2 | Results of applying an MLP network and NLR method for estimating missing data at Baghkolaye station

	Testing data				Training data			
Source stations	R^2 **(MLP)**	R^2 **(NLR)**	**RMSE (MLP)**	**RMSE (NLR)**	R^2 **(MLP)**	R^2 **(NLR)**	**RMSE (MLP)**	**RMSE (NLR)**
Hashtadjoft	0.82	0.23	30.47	63.03	0.84	0.14	22.93	53.30
Qaregoni	0.64	0.17	23.70	35.87	0.83	−0.07	24.11	60.86
Gilvan	0.40	0.13	43.92	52.88	0.74	0.26	29.21	48.79
Ostor	0.65	0.55	31.49	35.66	0.71	−0.04	32.41	61.01
Kohsalar Mianeh	0.26	0.17	42.79	45.27	0.41	0.01	45.23	58.34
Galinak	0.50	0.23	28.92	35.96	0.96	0.52	11.20	38.80
Loshan	0.79	0.46	32.36	51.74	0.81	0.38	21.33	38.79
Yankikand	0.36	0.26	57.51	61.75	0.77	0.04	25.11	51.53
Palti	0.54	0.40	21.43	24.59	0.83	−0.07	19.82	49.39
Dehgolan	0.28	0.09	25.11	28.22	0.57	0.02	33.63	51.10
Salamatabad	0.25	0.13	35.50	38.15	0.84	0.10	20.88	48.97
Siahdasht	0.77	0.61	21.83	28.78	0.93	0.18	12.33	42.48

o_k is the final model output and d_k is the desired outputs. The weight update rule is

$$W_{ij}(t+1) = W_{ij}(t) + \alpha[W_{ij}(t-1)] + \eta\delta_j(t)y_i(t) \quad (2)$$

where $W_{ji}(t+1)$, $W_{ji}(t)$ and $W_{ji}(\mathrm{t}-1)$ are the weights connecting neurons j and i at iterations $(t+1)$, t and $(t-1)$, respectively. η is the learning rate parameter and α is the momentum constant. The learning rate parameter controls the rate of network learning, whereas the momentum term helps to improve the learning accuracy and to avoid local minima. The forward and backward passes of computations are repeated on the complete set of training data, passed many times to the ANN. With the progress of training, the randomly initialized synaptic weights of the network gradually change to their optimal value. The network training with the back-propagation algorithm is the search for the global minimum of the error function.

A target error is used as the criteria used to stop the training of the network. Performance of the trained ANN is assessed by observing the model errors on; (a) the entire set of training data and (b) on separate validation datasets (Gopakumar *et al.* 2007).

In this study the sigmoid function, bounded to the range of [0, 1], was used to have better normalized mean square error (NMSE) results in all the dataset. The ANN would require an extremely small weighting factor causing computational inaccuracies due to floating point calculation, sluggish training and the near-zero gradient of the sigmoid function at extreme values (Dawson & Wilby 1998). Therefore, in the present study, the input values were standardized with respect to the maximum and minimum values in the range, which provided better model predictions than other approaches of standardization:

$$\bar{X}_i = \frac{(X_i - X_{i(\min)})}{X_{i(\max)} - X_{i(\min)}} \quad (3)$$

where $\bar{X}_i$ is the respective standardized value for the node i; X_i is the actual value of node i; $X_{i(\min)}$ and $X_{i(\max)}$ are the minimum and the maximum of all values applied to the nodes, respectively.

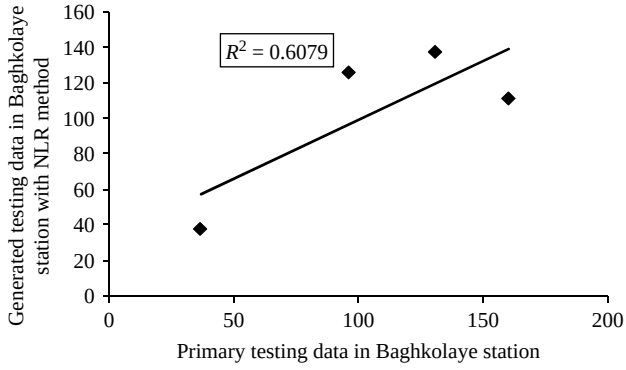

Figure 3 | The R^2 index in Baghkolaye station based on NLR method.

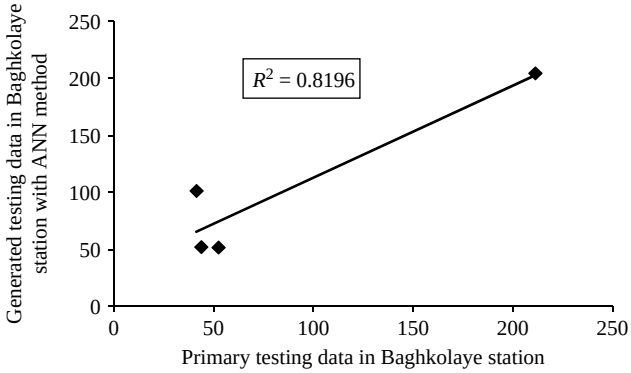

Figure 4 | The R^2 index in Baghkolaye station based on ANN method.

Nonlinear regression method

Since the relation between hydrometric data is not truly linear, we must make use of nonlinear techniques, one of which is the nonlinear regression that is widely applied in hydrology. By this method, the relationship between the target and the related parameters will be in the following form:

$$Q = aX_1^{\theta_1}X_2^{\theta_2}X_3^{\theta_3}\ldots X_n^{\theta_n} \tag{4}$$

where Q is the target parameter, θ_i is the ith model parameter, a is the multiplicative error term, X_i is the related parameter and n is the number of related parameters (Shu & Ouarda 2008). Solving Equation (4) using linear regression techniques generally requires linearizing the power form by a logarithmic transformation. However, the estimation of the linearized model is theoretically unbiased in the logarithmic domain (McCuen *et al.* 1990). Using nonlinear regression (NLR) methods, parameters can be directly estimated by minimizing the error in the actual domain. Nonlinear regression with a properly selected objective function can generally provide more accurate estimates than linear regression (Pandey & Nguyen 1999; Grover *et al.* 2002). In this paper, the NLR method is selected to be compared with the proposed ANN approach.

Evaluation methods

To assess the performance of each presenting methods, some indices must be applied. In this study, we use the maximum coefficient of determination (R^2) and the RMSE to compare ANN and NLR methods. These indices are computed according to the flowing equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{n}} \tag{6}$$

Table 3 | Results of missing data estimation by MLP network and NLR method using Baghkolaye station

	Testing data				Training data			
Related station	**R^2 (MLP)**	**R^2 (NLR)**	**RMSE (MLP)**	**RMSE (NLR)**	**R^2 (MLP)**	**R^2 (NLR)**	**RMSE (MLP)**	**RMSE (NLR)**
Motorkhaneh	0.41	0.41	47.04	47.22	0.62	−0.03	43.80	71.62
Hashtadjoft	0.64	0.24	25.05	36.20	0.68	0.11	21.56	36.06
Qaregoni	0.54	0.32	169.29	207.25	0.57	−0.09	142.18	225.61
Gilvan	0.59	0.56	184.54	191.63	0.76	0.30	186.45	316.68
Ostor	0.70	0.47	170.28	225.34	0.87	0.06	93.96	250.61
Poldokhtar Mianeh	0.62	0.28	63.00	86.60	0.84	0.47	76.42	139.84
Kohsalar Mianeh	0.38	−0.01	38.51	49.25	0.62	−0.06	35.31	59.09
Galinak	0.41	0.10	43.89	54.09	0.88	0.64	13.45	23.21
Loshan	0.60	0.42	76.54	91.58	0.71	0.13	63.29	109.32
Yankikand	0.37	−0.11	34.23	45.16	0.70	−0.04	29.40	54.71
Palti	0.16	−0.04	18.32	20.36	0.61	−0.12	10.11	17.14
Dehgolan	0.43	−0.08	8.90	12.27	0.67	−0.07	6.41	11.54
Salamatabad	0.46	−0.02	82.97	113.89	0.63	−0.11	65.75	114.65
Sarcham	0.24	−0.13	86.31	105.50	0.51	−0.12	29.26	44.06
Siahdasht	0.61	0.52	41.12	45.76	0.88	0.34	51.02	118.93

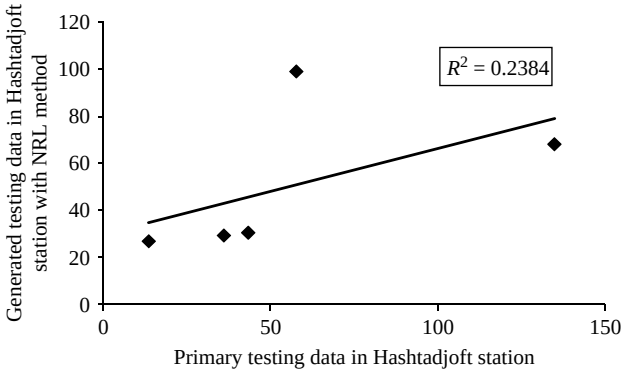

Figure 5 | The R^2 index in Hashtadjoft station based on NLR method.

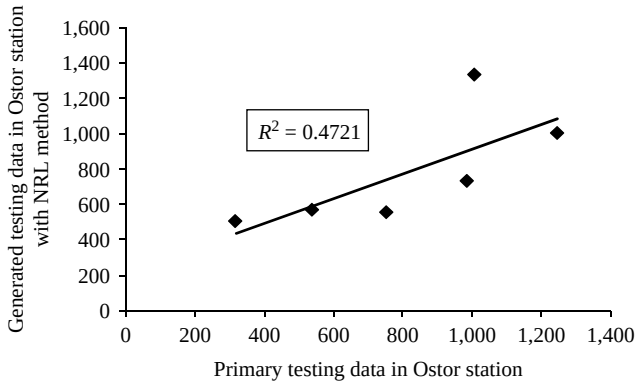

Figure 7 | The R^2 index in Ostor station based on NLR method.

where n is the number of data, X_i is the real observation data, Y_i is the predicted data and $\bar{X}$ is the average of observation data.

Description of datasets

The ANN approach has been applied to the hydrometric station networks of the Sefidroud watershed in the north of Iran. The catchment area is located between 35°N and 38°N and its area is about 60,496 km^2. There are 58 hydrometric stations in this watershed. In this study, only 16 stations in the catchment area which consist of 15 or more recorded data were selected. Table 1 shows the characteristics of these selected stations and Figure 2 shows the watersheds of Iran and the Sefidroud watershed. For this study, the data of maximum annual discharge (Q_{max}), which were recorded from 1973 to 2002, were used. In the beginning, the outlier data in the datasets were found by using the US Water Resources Council (USWRC 1981) and Grubs tests (Graphpad 2009). The USWRC test was used to determine the minimum level of the dataset, while the Grubs test was used to determine the maximum level of the dataset.

RESULTS AND DISCUSSION

Removing the outlier data was followed by emerging gaps in datasets of a few stations. Thus, the station with the maximum remaining data was used to estimate the missing data in other stations. Baghkolaye station, which had the maximum natural data, was selected as the source station for completing the data of other stations. To complete the Baghkolaye station's data, the stations with a complete data in those missing data periods for Baghkolaye station were used. Finally, the data of 12 stations were selected and applied to a Multi-Layer Perceptron (MLP) network and

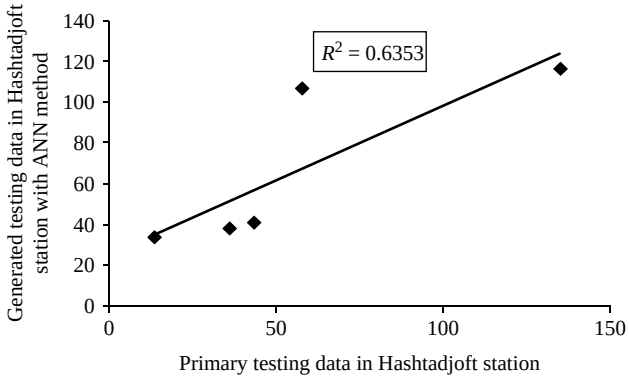

Figure 6 | The R^2 index in Hashtadjoft station based on ANN method.

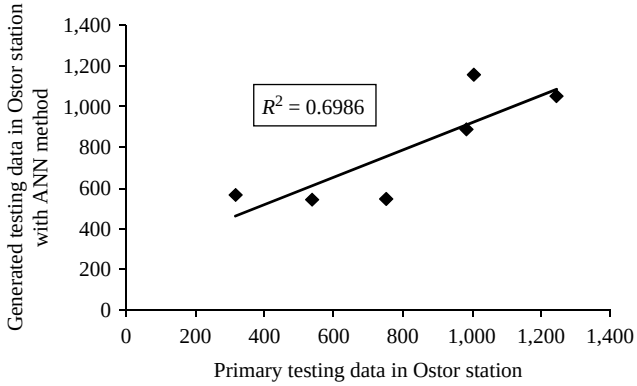

Figure 8 | The R^2 index in Ostor station based on ANN method.

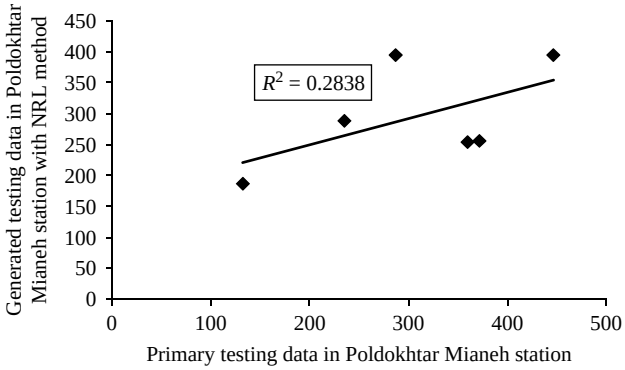

Figure 9 | The R^2 index in Poldokhtar Mianeh station based on NLR method.

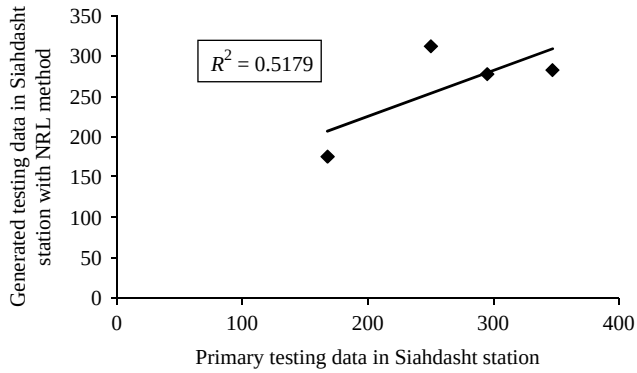

Figure 11 | The R^2 index in Siahdasht station based on NLR method.

a NLR model to estimate the missing data in this station. The network used had an input layer with one neuron, a hidden layer with 10 neurons and an output layer with one neuron. The learning rate of 0.05 with the ratios of 1.05 to increase the learning rate and 0.7 to decrease the learning rate were used in this model. To increase the convergence, the number of epochs of 10,000 was selected by trial and error; an index of the performance goal of 0.001 was also used with improved results.

The hyperbolic tangent sigmoid function was used as the transfer and output functions in this network. The number of layers and neurons in each layer and the type of transfer and output functions were selected based on which applied to find out the best structure of MLP network. In this stage, 70% of input and target data were used for training and the remaining 30% for validation and testing the network. The missing data in Baghkolaye station was first reconstructed using data from other stations, such as Hashtadjoft station. To generate the missing data in Baghkolaye, the network was trained and tested independently based on Baghkolaye and every other source (column 1, Table 2). Then, with using every source station, the missing data in Baghkolaye station was reconstructed. To obtain the best results, the R^2 and RMSE indices had been used and applied between the main data and the generated data.

According to Table 2, the best stations for estimating the missing data were Hashtadjoft and Siahdasht in ANN and NLR, respectively. Figures 3 and 4 compare the results of the R^2 index between primary testing data and those generated in Baghkolaye station based on NLR and ANN methods, respectively.

After the estimation of missing data in Baghkolaye station, the data at this station was used to estimate the missing data in other stations. Therefore, the MLP network with one hidden layer was applied. The structure of this network was similar to the previous architecture. Therefore, the best structure for this MLP using these numbers of

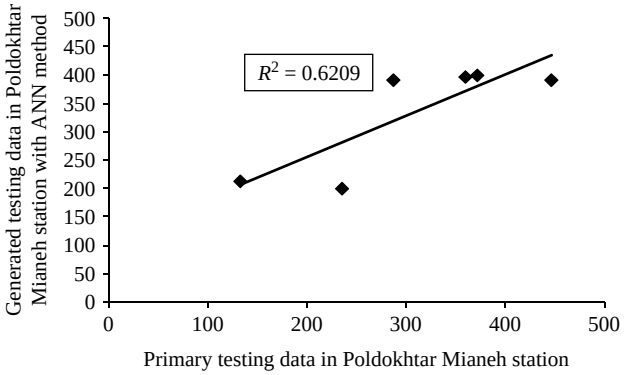

Figure 10 | The R^2 index in Poldokhtar Mianeh based on ANN method.

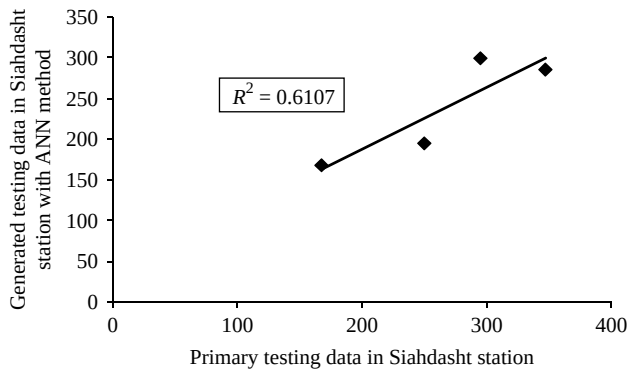

Figure 12 | The R^2 index in Siahdasht station based on ANN method.

layers, neurons, transfer function and output function was reached. In this stage, 70% of input and target data were also used for training and the remaining 30% for validation and testing the network. In addition, the NLR method was used as the second technique for estimating the missing data in other stations. To obtain the best results, the R^2 and RMSE indices were used. The results are shown in Table 3. Figures 5 and 6 provide the R^2 indexes for Hashtadjoft station using the ANN and NLR methods, respectively. Similar results for Ostor station can be observed in Figures 7 and 8, for Poldokhtar Mianeh station in Figures 9 and 10, and for Siahdasht station in Figures 11 and 12.

Considering the results obtained from Table 3 and the above figures, it is concluded that the ANN method with a Multi-Layer Perceptron network provided better results than the nonlinear regression method in estimating the missing data for hydrometrics datasets. In addition, this study illustrated that the ANN method will provide accurate estimations even in the cases of limited or insufficient data.

CONCLUSIONS AND FUTURE WORK

This research was conducted to evaluate the applicability of ANN for the reconstruction of hydrometric data. The comparison between this method and the NLR method showed that in all cases the results from ANN were more accurate than those produced by NLR methods. The study also showed that ANN will certainly be useful in the case with only a few numbers of data, if the structure of the network is correctly selected. In the previous studies, the missing data in each region were reconstructed by one or two stations, but in the present study, 15 stations were used and thus the accuracy is expected to be improved. The obtained results confirmed the main hypothesis of the research, which was preparing another way of application of ANN for reconstructing the annual missing data, in humid and semi-humid regions. In order to improve the current research, wavelet and neuro-fuzzy conjunction, multivariate ANN-wavelet approach and wavelet models for missing hydrometric data estimating for various regions from arid to humid are suggested.

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to the anonymous reviewers, whose comments improved the original manuscript significantly.

REFERENCES

Bhattacharya, B., Lobbrecht, A. H. & Solomatine, D. P. 2003 Neural networks and reinforcement learning in control of water systems. *J. Water Res. Plan. Manag.* **129**, 458–465.

Bin, H. E. & Takase, K. 2006 Application of the artificial neural network method to estimate the missing hydrologic data. *Hydrol. Water Res.* **19** (4), 249–257.

Bustami, R., Bessaih, N., Bong, C. & Suhaili, S. 2007 Artificial neural network for precipitation and water level predictions of Bedup river. *IAENG Int. J. Comput. Sci.* **34** (2), 228–233.

Campolo, M., Soldati, A. & Andreussi, P. 2003 Artificial neural network approach to flood forecasting in the River Arno. *Hydrol. Sci. J.* **48**, 381–398.

Chen, B. F., Wang, H. D. & Chu, C. C. 2007 Wavelet and artificial neural network analyses of tide forecasting and supplement of tides around Taiwan and South China Sea. *Ocean Eng.* **34**, 2161–2175.

Cigizoglu, H. K. 2003 Estimation, forecasting and extrapolation of river flows by artificial neural networks. *Hydrol. Sci. J.* **48**, 349–361.

Dastorani, M. T., Moghadamnia, A., Piri, J. & Rico-Ramirez, M. 2010 Application of ANN and ANFIS models for reconstructing missing flow data. *Environ. Monit. Assess.* **166**, 421–434.

Dawson, C. W. & Wilby, R. 1998 An artificial neural network approach to rainfall-runoff modelling. *Hydrol. Sci. J.* **43**, 47–66.

Diamantopoulou, M. J., Antonopoulos, V. Z. & Papamichail, D. M. 2006 Cascade correlation artificial neural networks for estimating missing monthly values of water quality parameters in rivers. *Water Res. Manag.* **21**, 649–662.

Elshorbagy, A., Panu, U. S. & Simonovic, S. P. 2000 Group-based estimation of missing hydrological data: I. Approach and general methodology. *Des sciences hydrologiques* **46** (6), 849–866.

Gopakumar, R., Takara, K. & James, E. J. 2007 Hydrologic data exploration and river flow forecasting of a humid tropical river basin using artificial neural networks. *Water Res. Manag.* **21**, 1915–1940.

Graphpad 2009 How Grubbs' test works. Available at: http://www.graphpad.com/library/BiostatsSpecial/article_39.htm

Grover, P. L., Burn, D. H. & Cunderlik, J. M. 2002 A comparison of index flood estimation procedures for ungauged catchments. *Can. J. Civil Eng.* **29**, 734–741.

Hornik, K., Stinchcombe, M. & White, H. 1989 Multilayer feedforward networks are universal approximators. *Neural Netw.* **2** (5), 359–366.

Hsu, K.-L., Gupta, H. V. & Sorooshian, S. 1995 Artificial neural network modeling of the rainfall-runoff process. *Water Resour. Res.* **31**, 2517–2530.

Imrie, C. E., Durucan, S. & Korre, A. 2000 River flow prediction using artificial neural networks: generalization beyond the calibration range. *J. Hydrol.* **233**, 138–153.

Jain, S. K., Das, A. & Srivastava, D. K. 1999 Application of ANN for reservoir inflow prediction and operation. *J. Water Res. Plan. Manag.* **125**, 263–271.

Jain, S. K., Singh, V. P. & van Genuchten, M. T. h. 2004 Analysis of soil water retention data using artificial neural networks. *J. Hydrol. Eng.* **9**, 415–420.

Kavianpour, M. R. & Rajabi, E. 2005 Application of neural network for flow aeration downstream of bottom outlet leaf gates. *Iran. J. Water Res. Manag.* **3** (1), 96–103.

Lekkas, D. F., Imrie, C. E. & Lees, M. J. 2001 Improved non-linear transfer function and neural network methods of flow routing for real-time forecasting. *J. Hydroinf.* **3** (3), 153–164.

Lin, G. F. & Chen, L. H. 2004 A non-linear rainfall-runoff model using radial basis function network. *J. Hydrol.* **289**, 1–8.

Linsley, R. K., Kohler, M. A. & Paulhus, J. L. H. 1988 *Hydrology for Engineers*. McGraw-Hill, New York.

McCuen, R. H., Leahy, R. B. & Johnson, P. A. 1990 Problems with logarithmic transformations in regression. *J. Hydraul. Eng.* **116** (3), 414–428.

Pandey, G. R. & Nguyen, V.-T.-V. 1999 A comparative study of regression based methods in regional flood frequency analysis. *J. Hydrol.* **225**, 92–101.

Rajaee, T., Mirbagheri, S. A., Zounemat-Kermani, M. & Nourani, V. 2009 Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Sci. Total Environ.* **407**, 4916–4927.

Shu, C. & Ouarda, T. B. M. J. 2008 Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *J. Hydrol.* **349**, 31–43.

Shukla, M. B., Kok, R., Prasher, S. O., Clark, G. & Larcroix, R. 1996 Use of artificial neural network in transient drainage design. *Trans. ASAE* **39** (1), 119–124.

Srinivasa, L. 1998 Aquifer parameter estimation using genetic algorithm and neural networks. *Civ. Environ. Eng. Syst.* **16**, 37–50.

US Water Resources Council 1981 *Guidelines for Determining Flood Flow Frequencies*. Bulletin 17B (revised), Hydrology Committee, Water Resources Research Council, Washington, DC.

Wen, C. G. & Lee, C. S. 1998 A neural network approach to multiobjective optimization for water quality management in a river basin. *Water Resour. Res.* **34**, 427–436.

Wilby, R. L., Abrahart, R. J. & Dawson, C. W. 2003 Detection of conceptual model rainfall-runoff processes inside an artificial neural network. *Hydrol. Sci. J.* **48**, 163–181.

Yang, C. C., Larcroix, R. & Prasher, S. O. 1998 The use of back-propagation neural networks for the simulation and analysis of time-series data in subsurface drainage system. *Trans. ASAE* **41**, 1181–1187.

First received 2 September 2009; accepted in revised form 25 January 2010. Available online 24 June 2010