JOURNAL OF
ENVIRONMENTAL HYDROLOGY

The Electronic Journal of the International Association for Environmental Hydrology
On the World Wide Web at http://www.hydroweb.com

VOLUME 19 2011

NEURAL NETWORK APPLICATION FOR MONTHLY PRECIPITATION DATA RECONSTRUCTION

Zohre Khorsandi[1]
Mohammad Mahdavi[2]
Ali Salajeghe[2]
Saeid Eslamian[3]

[1]Department of Natural Resources, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2]Department of Natural Resources, Tehran University, Karaj, Iran
[3]Department of Water Engineering, College of Agriculture, Isfahan University of Technology, Iran

The need for precipitation data and the importance of its duration in hydrological and climatic phenomena motivate investigators to develop reconstruction methods. Four methods are artificial neural network, normal ratio, inverse distance weighting, and geographical coordinate. These methods are compared in this study using monthly precipitation data of three stations, Dolat-Abad, Kabootar-Abad and Refinery plant around the city of Esfahan, Iran. Mean absolute error and coefficient of correlation of the results are compared to select the most appropriate method. The neural network approach shows the best performance compared with the other methods.

INTRODUCTION

Precipitation data are critical for hydrological analysis, design of water resources systems and management programs, and investigation of drought periods. Missing data are one of the most important problems in this field. Various methods are developed to solve this problem and reconstruct data interpolation methods such as regression methods, time series and artificial neural networks. McCuen (1998) recommended a simple method to estimate missing rainfall values. In his method, simple mean will be used as the estimated value to perform calculations if the difference of annual precipitation is less than 10% in neighbor stations. Otherwise if the difference is more than 10%, the normal ratio method will be used.

Vicente-Serrano et al. (2010) used values from the nearest neighbor observations to reconstruct a daily precipitation database for northeast Spain with field gaps of 3106 daily precipitation observatories.

In applied hydrology literature, the inverse distance weighted method (IDWM) and normal ratio method are the most common approaches for estimation of missing values in hydrology sciences (ASCE 1996).

Abebe et al. (2000) compared fuzzy methods, neural network methods and statistical methods efficiencies. Their results showed better performance of the neural network method in comparison with the other methods.

Nalder et al. (1998) used different approaches such as nearest neighbor, inverse distance weighted and different types of kriging to interpolate climatic parameters. Their results showed that the best method for calculating temperature and monthly precipitation is the inverse square distance method. They described this method as simple, applicable and more accurate with respect to other methods. Also Dirks et al. (1998) used the mean of stations, inverse distance weighted, Thiessen polygon and kriging methods in order to estimate precipitation, and the inverse distance square method showed the best performance for estimation of precipitation values. MCcuen (1998) proposed the inverse distance weighted method for complete precipitation data and he suggested that the inverse distance can be used instead of geographical coordinates distance.

Goovaerts (2000) used three multivariate geostatistical algorithms incorporating a digital elevation model into the spatial prediction of rainfall, and compared these methods with the ordinary kriging, Thiessen polygon and inverse square distance. The results showed higher error values for the Thiessen method and inverse square distance in prediction of precipitation values, and his results showed that the kriging method has better results than the linear regression method when there is remarkable correlation between elevation and precipitation. Teegavarapu et al. (2005) compared commonly used spatial interpolation methods (IDWM and coefficient of correlation weighted method), nearest neighbor, kriging and artificial neural networks in reconstruction of missing rainfall data. The results showed that the methods of coefficient of correlation weighting, ANN and kriging have better results in comparison with IDWM.

The neural network method is suitable to find a relation among variables that have unknown highly nonlinear relation. Regression models and time series are used commonly to complete missing precipitation data. All these methods require the numerical form of relation variables as a priori and this is the most important limitation of these methods (Teegavarapu et al., 2005). The artificial neural network methods determine this relation function automatically. Therefore this method has found much popularity among investigators. Neural network is used in spatial and temporal interpolation and in prediction of hydrological and climatically phenomena (Tolika et al., 2007; Malek et al., 2009).

Neural network has been used more in reconstruction and completion of time series. Although reconstruction of the monthly and daily time series of stream flows can be performed using statistical methods, this reconstruction and completion can not be carried out carefully by the other two methods because there are great changes in spatial and temporal behavior of precipitation. There are a limited number of studies about using this method to complete the precipitation time series (Coulibaly and Evora, 2007). Researchers that used ANN for reconstructing and interpolating climatic data agree on the accuracy of this method.

MATERIAL AND METHODS

Case study

There are 21 climatological stations in the Esfahan area (Figure 1). As a first step, monthly precipitation data of stations were checked for uniformity, homogeneity and length of statistical period. Fourteen stations were chosen and the period of 1989-2008 was used as the statistical period. Station characteristics are shown in Table 1. Three stations near Esfahan (Kabootar-Abad, Esfahan Refinery plant, and Dolat-Abad) were chosen to reconstruct monthly precipitation data.

Methodology

This study aims to evaluate four methods of Normal ratio (NR), geographical coordinate (GC), Inverse distance weighting (IDW) and neural network (ANN) for reconstructing missing values. Each method is described briefly in following.

Normal ratio method (NR):

In this method, some neighbor stations are used to reconstruct missing values (Chow, 1964; Linsley et al., 1988). Precipitation values in neighbor stations are weighted by normal precipitation values (as average of annual precipitation). This method is given by:

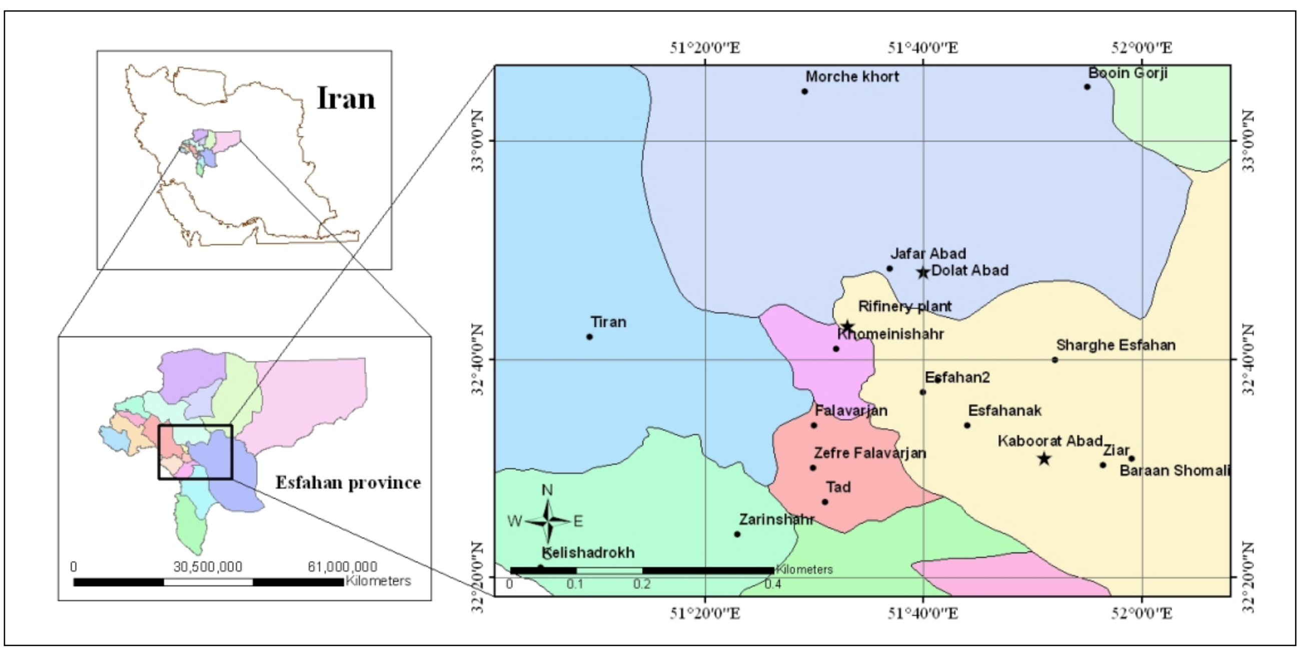

Figure 1. Location of case study.

Table 1. Characteristics of studied meteorological stations.

Stations	Coordinate		Height (m)	Average annual rainfall (mm)
	Longitude	Latitude		
Dolat-Abad	51° 40?	32° 48?	1585	113.67
Khomeini Shahr	51° 32?	32° 41?	1600	123.21
Mahyar 1	51° 47?	32° 17?	1700	139.65
Esfahan 1	51° 40?	32° 37?	1585	131.20
Tiran	51° 9?	32° 42?	1840	172.10
Jafar-Abad	51° 36?	32° 48?	1600	144.85
Zefrh Falavarjan	51° 29?	32° 30?	1605	150.40
Zeyar Braan	51° 56?	32° 30?	1540	119.20
Mourchekhort	51° 29?	33° 4?	1700	123.87
Mahyar 2	51° 48?	32° 16?	1700	148.45
Esfahan 2	51° 41?	32° 38?	1550	125.10
Kabootar-Abad	51° 51?	32° 31?	1545	110.13
Refinery plant	51° 32?	32° 43?	1673	125.71
Sharge Esfahan	51° 52?	32° 40?	1543	103.11

$$Px \quad \frac{1}{3\left(\frac{N_X}{N_A}P_A \quad \frac{N_X}{N_B}P_B \quad \frac{N_X}{N_C}P_C\right)} \tag{1}$$

where P_x is the precipitation of station *x*, *A, B, C* are the nearest neighbor stations, *P* and *N* are amount of precipitation and normal precipitation (Abebe et al., 2000).

Geographical coordinate method (GC):

One of the distance weight methods is to use geographical coordination to determine weight coefficient. After determination of station positions, an incomplete station is considered as the center of coordinate system then distance of any neighbor stations to the center of coordinate system can be calculated. In this method the neighbor stations close to the incomplete station are more important in reconstruction. Therefore, higher weight coefficients will be assigned to them. Weight coefficient and estimation method is given by:

$$W_i = \frac{1}{X_i^2 + Y_i^2} \tag{2}$$

$$P_x = \frac{\sum W_i P_i}{\sum W_i} \tag{3}$$

where W_i: weighted coefficient of neighbor station, *xi* and *yi*: longitude and latitude neighbor stations, P_i: precipitation value in neighbor station and P_x: the reconstructed value (Mahdavi, 1998).

Inverse distance weighting method (IDWM):

Spatial interpolation is another method to reconstruct missing values. Positive spatial relationship is the most important factor in interpolation methods such as inverse distance weighting method and stochastic interpolation method. It means that stations with similar numerical values are closer to each other. Neighbor point values and spatial coefficients are used to estimate missing data. The coefficient is calculated by different equations for different methods. This method is one of the most popular and applied methods in hydrology and geomorphology. The success of inverse distance weighting method depends primarily on the existence of positive spatial autocorrelation (Teegavarapu, 2005). Weighting coefficients and effective neighbor selection are limitations of this approach. The following equation calculates the estimation of missing value in this method.

$$Z = \frac{\sum_{i=1}^{n} Z_i d_{mi}^{-k}}{\sum_{i=1}^{n} d_{mi}^{-k}} \tag{4}$$

where Z is the observation at base station, n is number of stations, Z_i is the observation at station i, d_{mi} is distance from the location of i to the station m, and k is referred to as friction distance that ranges from 1 to 6. The value of 2 is used more commonly for the k coefficient (Teegavarapu, 2005).

Artificial neural network (ANN):

The artificial neural network is extensively parallel processers composed of simple units of processor that have a natural tendency to preserve experimental knowledge and restore it as required. A neural network is similar to brain because of two reasons:

1- Network learns knowledge from the environment through a training process.

2- Strength of connections between neurons is used to store information (Haykin, 1999).

Neural networks have different structures but their components are similar. A neural network is defined by its components, including neurons, type of neuron connections and neural network training algorithm.

The neuron, the basic component of neural network, is a data processing unit whose diagram is shown in Figure 2. The components of a neuron are as follows:

- A set of connections or synapses, each of which has a specific weight coefficient or strength. The transmitting signal will be multiplied by the weight of synapse. Unlike real synapses in the brain, the neural network weight coefficient can be negative.

- A function to evaluate the sum of the total input signal.

- An activation function is used to limit the output range and non-linearization of neuron output.

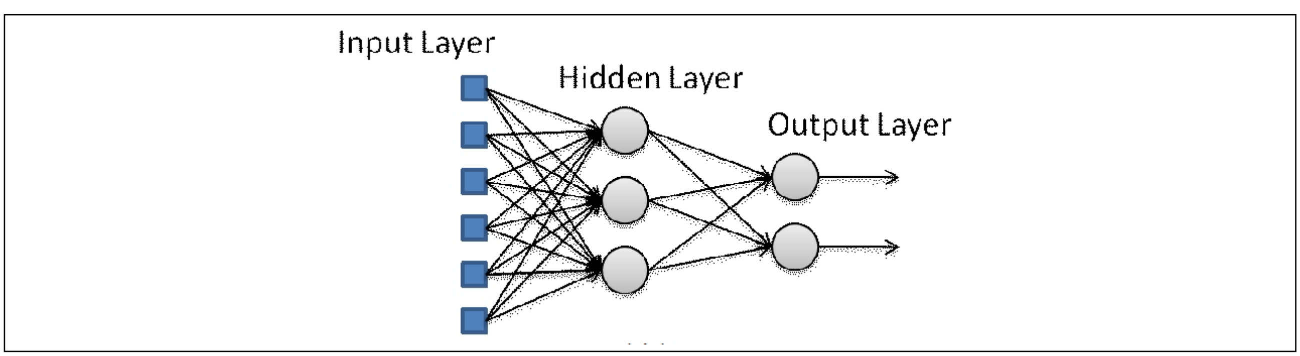

Figure 2. Multi-layer Perceptron.

Neural network types:

Neural networks differ in type of neurons, connection types and the learning algorithms. Neural network structure is shown by graphs for simplicity. In this case, a neuron is shown as a solid circle and synapses by arrows. The arrows represent direction of information flow. Common neural networks are: single-layer perceptron, multi-layer perceptron, general feed forward neural network, recursive neural networks and RBF neural networks. The multi-layer perceptron which is used in this study is described in the following.

Multi-layer perceptron:

This type of neural network consists of a number of neurons located in different layers. Connections are forward and neuron values will be calculated layer by layer from input layer to output layer to calculate outputs of the network. This kind of neural network has been used in various engineering applications.

Training Neural Network

The most important feature of neural networks is learning from their environment to increase their performance. Different processes are called "learning" processes and there are different views about the definition of the learning. Mendel and McClaren defined the learning process as:

"Learning is a process in which the free parameters of neural network will be adjusted through the network environment stimulation. The behaviour of adjustment of free parameters defines the type of learning" (Mendel, 1970).

The collection of accurate and specific rules to solve the problem of learning is called the training algorithm. So far various algorithms for training of neural networks are presented, including the Levenberg–Marquardt algorithm, competitive adaptive learning, and learning. Coulibay et al. (2007) compared different neural networks for reconstruction of daily data and suggested the multi-layer perceptron procedure as the most appropriate method for reconstruction of rainfall data. In this study the suggested method is used to reconstruct the data.

Application

In all of the methods used in this study, selection of the neighbor stations is one of the most important tasks. To select neighbor stations, first of all, the Thiessen Polygon method was used to determine range of effect of each station then neighbor stations were determined (Figure 3). Then according to the length of statistical period and accuracy of the recorded data a number of stations were selected (Table 2).

Monthly precipitation data and average annual rainfall from nearby stations were used in the NR. In the GC, the coordinates and Equations (2 and 3) are used, and in the IDWM, the distance between stations and the Equation (4) are used to calculate monthly precipitation values for the stations with missing data.

ANN input data are monthly rainfall of neighbor stations and the outputs are monthly precipitation of missing station. Multi-layer perceptron and Levenberg–Marquardt algorithm are used in this study. In order to select the best neural network structure for the reconstruction of precipitation data, different structures considering the number of inputs were evaluated, results of this evaluation are presented in Table 3.

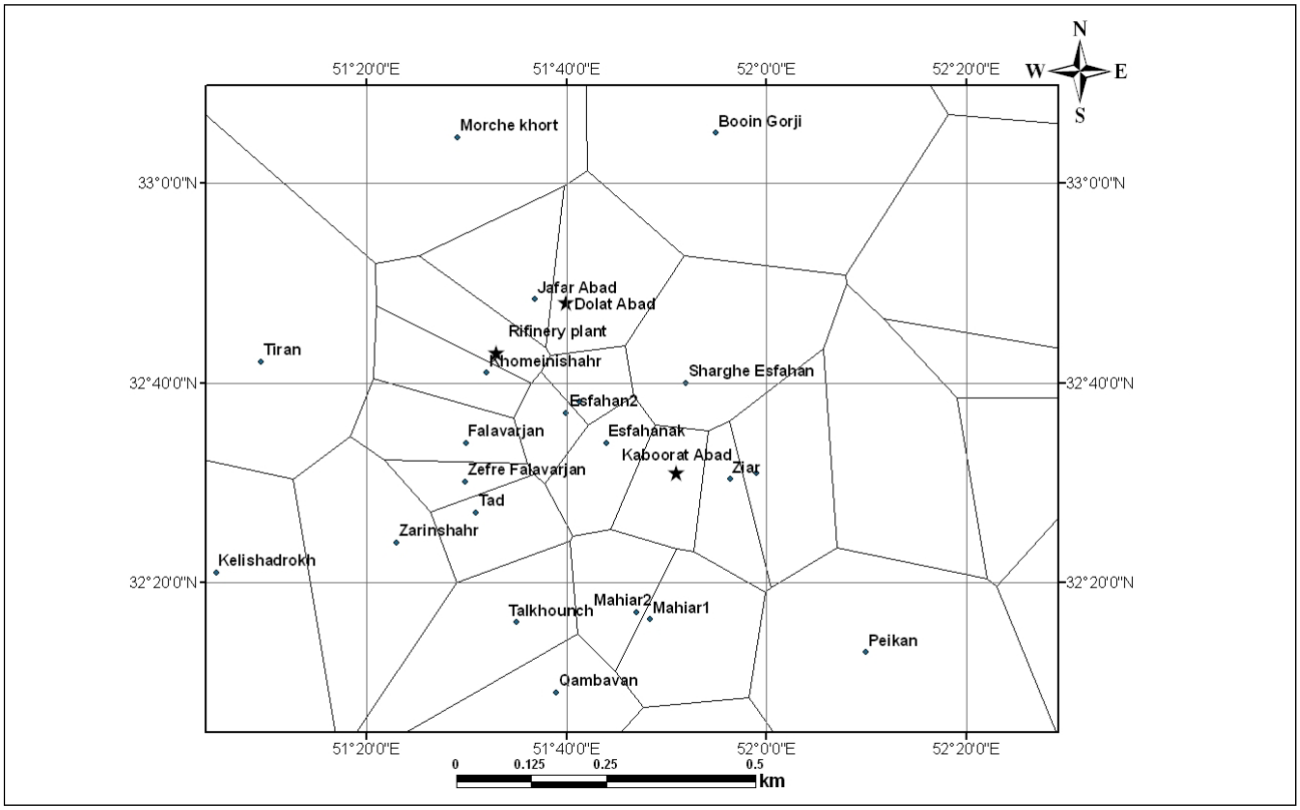

Figure 3. Position of each of the stations Thiessen Polygon method.

Table 2. Stations used in the precipitation data reconstruction.

Incomplete station	Adjacent station	Average annual rainfall (mm)	Distance incomplete station (km)
Refinery plant	Jafar-abad	144.9	11.6
	Mourchekhort	123.9	40.3
	Esfahan 1	131.2	15.8
	Esfahan 2	125.1	15.6
	Khomeini Shahr	123.2	4
	Tiran	172.1	36.8
Kabootar-Abad	Mahyar 2	148.5	26.6
	Sharge Esfahan	103.1	16.7
	Esfahan 2	125.1	20.5
	Mahyar 1	139.7	27.4
	Zeyar Braan	119.2	8.5
	Esfahan 1	131.2	20.1
Dolat-Abad	Sharge Esfahan	103.1	23.9
	Esfahan 1	131.2	18.3
	Jafar-abad	144.9	4.9
	Mourchekhort	123.9	34.9

Reconstruction

Two factors were considered to select the best structure for the neural network: the correlation coefficient function of training and the correlation coefficient function of evaluation.

For Esfahan Refinery plant station, a network with a hidden layer with 12 neurons had highest correlation of training data and a network with two hidden layers with 5 and 3 neurons had highest correlation of evaluation data. Considering the importance of evaluation, the second structure is

Table 3. The structure survey results.

Station	Number of neurons in the hidden layer		Training		Evaluation	
	First Layer	Second Layer	R	SD	R	SD
Refinery plant	5	3	0.93	0.03	0.76	0.07
	3	2	0.83	0.06	0.71	0.11
	5	2	0.93	0.02	0.69	0.14
	12	0	0.97	0.01	0.64	0.12
Kabootar-Abad	10	0	0.88	0.02	0.44	0.21
	12	0	0.9	0.02	0.42	0.18
	5	3	0.84	0.05	0.61	0.14
	5	2	0.89	0.04	0.51	0.03
	3	2	0.69	0.27	0.68	0.1
	5	4	0.84	0.04	0.63	0.13
Dolat-Abad	3	2	0.85	0.08	0.8	0.06
	3	1	0.69	0.3	0.72	0.25
	8	0	0.91	0.02	0.69	0.06
	10	0	0.94	0.02	0.77	0.12
	20	0	0.95	0.02	0.63	0.15

used to reconstruct data. For the other two stations, a network with two hidden layers with 3 and 2 neurons had the highest correlation of evaluation values. Finally the best methods for reconstructing monthly precipitation data were chosen by correlation coefficient and mean absolute error.

$$MAE = \frac{\sum_{i=1}^{n} |\overline{p} - p_i|}{n} \tag{5}$$

where $\overline{p}$ is reconstructed precipitation, p_i is the measured precipitation and n number of data.

RESULTS AND DISCUSSIONS

The described methods are used to reconstruct monthly precipitation data of Esfahan refinery plant, Kabootar-Abad and Dolat-Abad. Figures 4-9 compare the results of NR, ANN, IDW and GC methods with observed values.

The mean absolute error is used to compare results of methods (Table 4).

Mean absolute error of Dolat-Abad station was 6.73 for IDW method, 6.62 for GC, 5.56 for NR method, and 3.75 for ANN. Therefore the ANN method has the best performance for reconstruction of monthly precipitation data in comparison to other three methods. In Esfahan Refinery plant station mean absolute error has a value of 5.22 for IDW method, 5.23 for GC method, 6.31 for NR method and 4.52 for ANN. The ANN method has the lowest value of mean absolute error for the results of Kabootar-Abad also. The neural network coefficients of correlation for all the stations are more than values of the coefficient of correlation of the other methods. Therefore, according to the statistical comparison in these stations, ANN has the best performance.

It should be noted that efficiency of methods differs for the different stations. This difference is obvious in results of Kabootar-Abad station. Mean distances of neighbor stations in the three case studies are about 20 km, but distance changes in Dolat-Abad and refinery stations (12 and 14 km) are more than Kabootar-Abad station (6.9 km). Also mean of monthly precipitation of Kabootar-Abad neighbor stations has more difference with the Kabootar Abad station (17.17 mm)

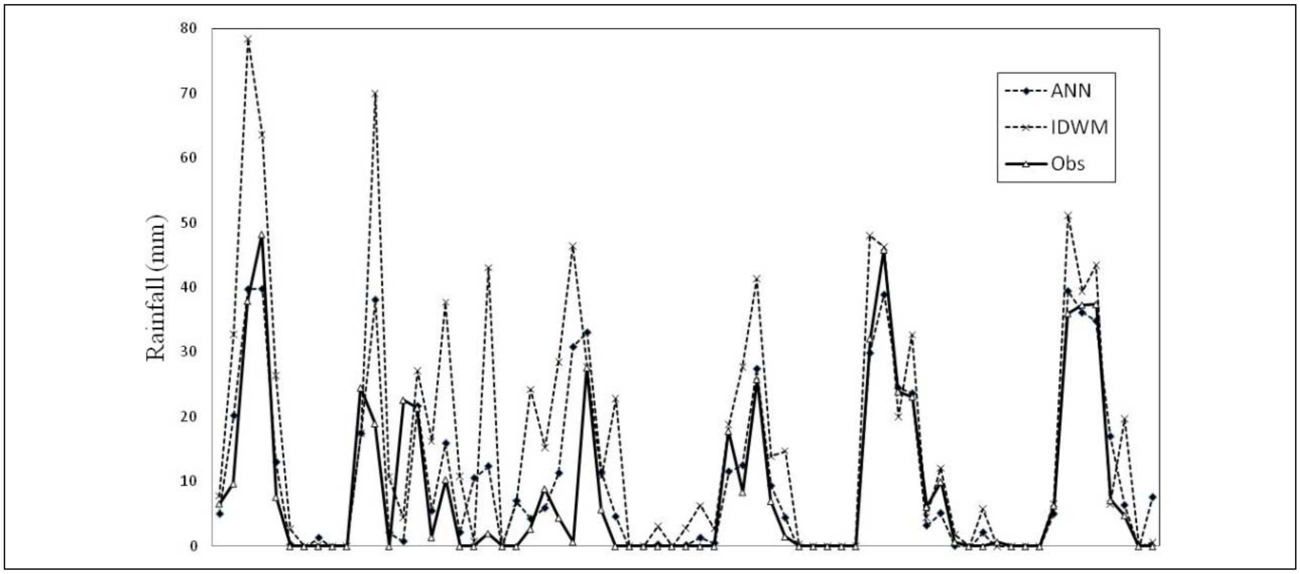

Figure 4. ANN, IDWM and observed monthly precipitation amounts at Dolat-abad station.

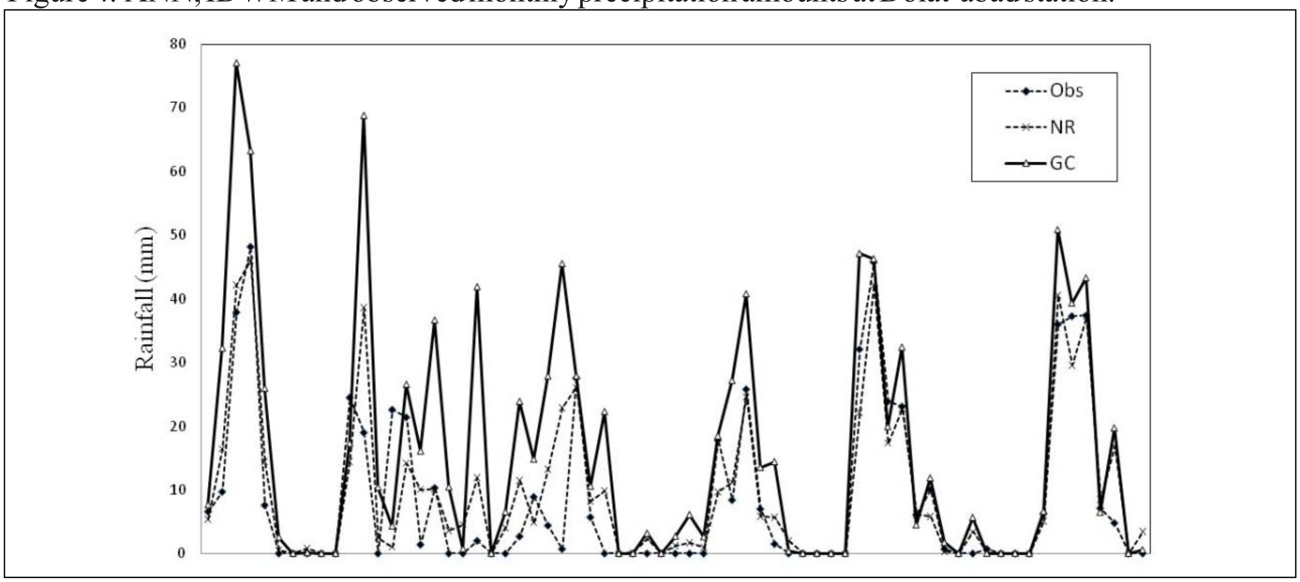

Figure 5. NR, GC methods and observed monthly precipitation amounts at Dolat-abad station.

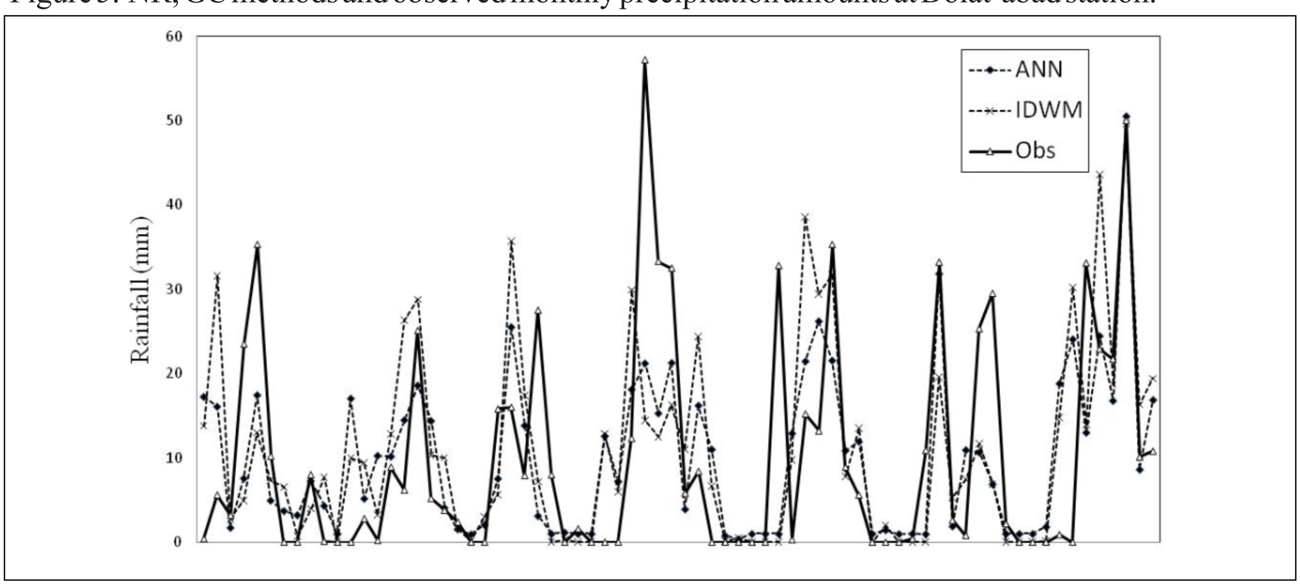

Figure 6. ANN, IDWM and observed monthly precipitation amounts at Kabootar-abad station.

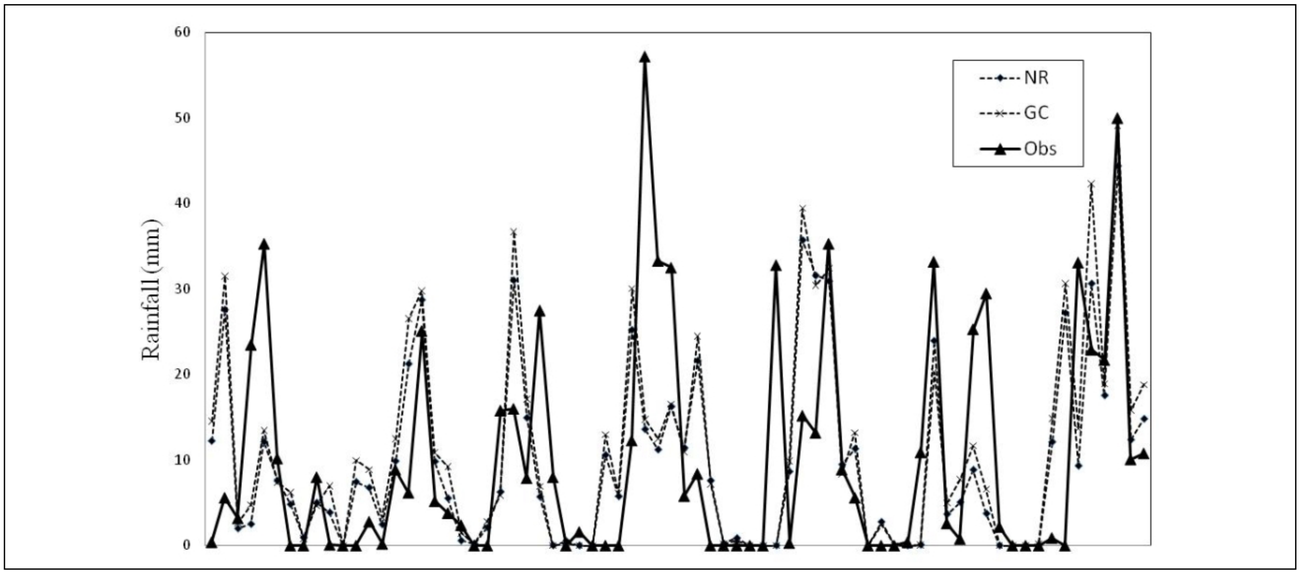

Figure 7. NR, GC methods and observed monthly precipitation amounts at Kabootar-abad station.

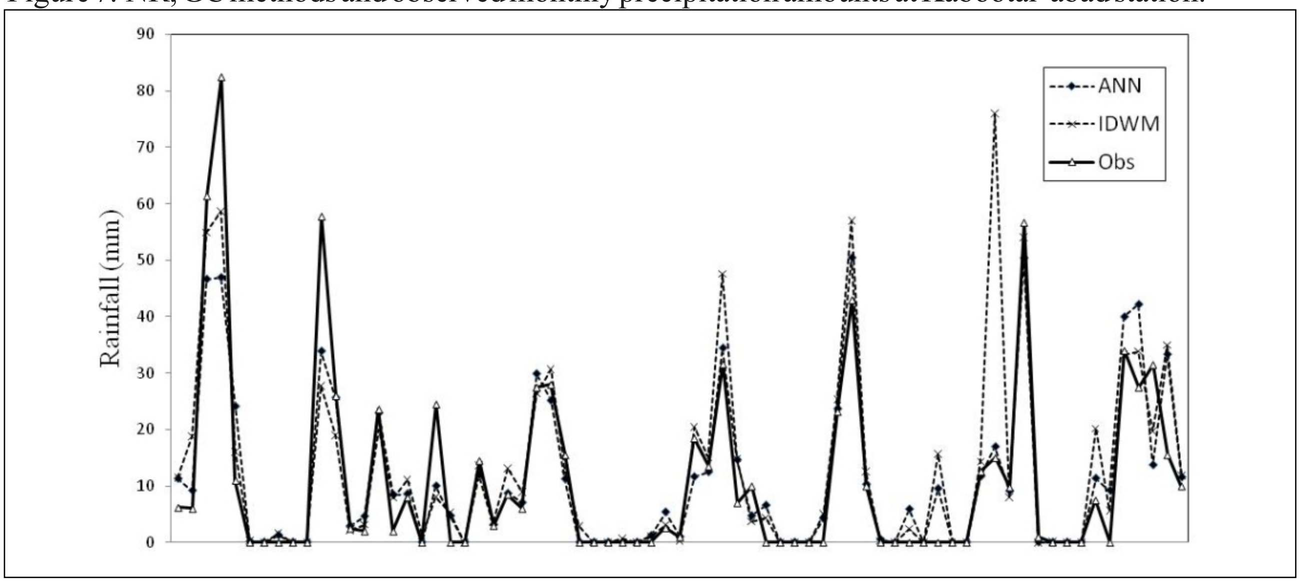

Figure 8. ANN, IDWM and observed monthly precipitation amounts at Refinery plant station.

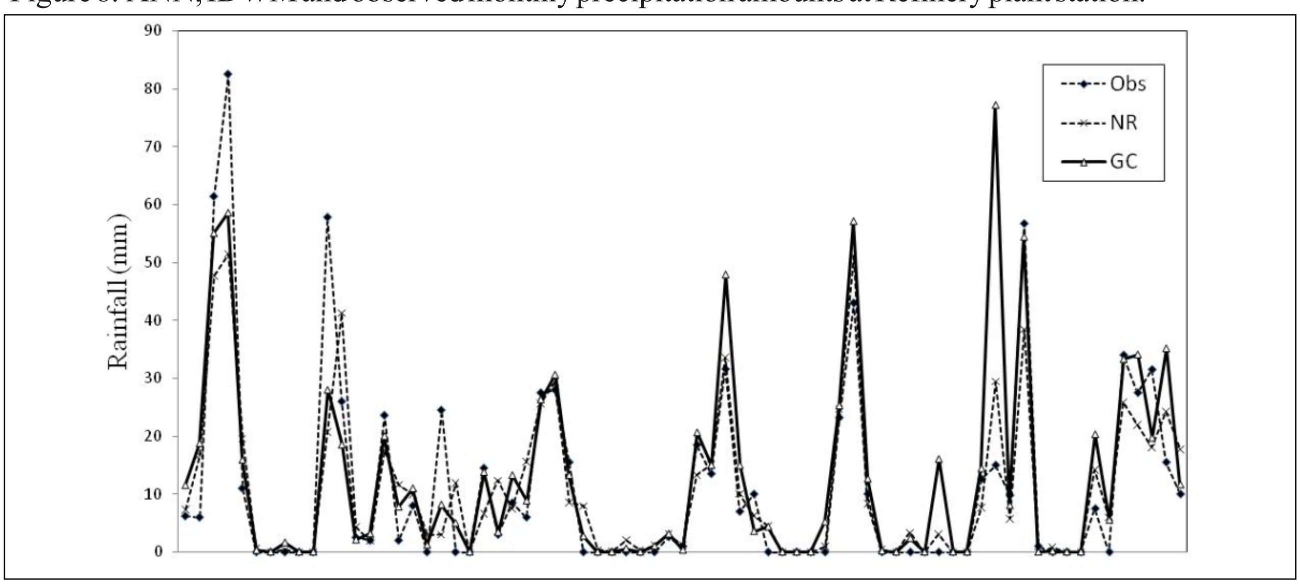

Figure 9. NR, GC methods and observed monthly precipitation amounts at Refinery plant station.

Table 4. Statistical evaluation of methods used in the precipitation data reconstruction.

Station	Inverse square distance		Geographical coordinate		Normal ratio		Neural network	
	MAE	R	MAE	R	MAE	R	MAE	R
Dolat-Abad	6.73	0.75	6.62	0.76	5.56	0.79	3.75	0.87
Kabootar-Abad	6.02	0.70	5.97	0.71	5.53	0.72	4.88	0.76
Refinery plant	5.22	0.79	5.23	0.79	6.31	0.79	4.52	0.88

in comparison to the other stations (Dolat-Abad 12 mm and refinery plant 11 mm). The amount of rainfall standard deviation of Kabootar-Abad station is smaller than the value for the two other stations. This higher difference indicates the higher differences between precipitation of Kabootar-Abad station and its nearby stations, and this difference is smaller in the two other stations. Therefore, the reasons of different efficiencies in different stations are the distances of neighbor stations to the missing stations and differences in their mean precipitation changes. Therefore, the lower efficiency of the neural network can be explained based on distance between the stations and their differences in precipitation behavior.

CONCLUSIONS

Methods based on distance weights are commonly used in missing data calculations and in spatial interpolation of climatic data. Major limitations in these approaches is that these methods can not be used if there is no positive spatial correlation between stations. Determination and selection of missing station neighbors is very important in the normal ratio method. In ANN methods, neighbor stations should also be selected correctly and input parameters that are monthly precipitation data affect the accuracy of the neural network. On the other hand, the type of network, numbers of layers and neurons, and training algorithms are important in the efficiency of the method. There are many studies about reconstruction of climatic and hydrological data. These studies propose multi-layer perceptron as a suitable structure to reconstruct data. Different investigators suggest different criteria such as training and evaluation correlation coefficients to select an appropriate structure of the network. According to these parameters the suitable structure is selected by the method presented here, three stations data are reconstructed and the results show that ANN is most efficient in comparison with the other three methods. The NR method is second in measure of efficiency. Although the NR is simple, it is more accurate in comparison with the GC and IDW methods. Two spatial interpolation methods used have similar results and occupy the third place in efficiency of reconstruction of monthly precipitation data.

ACKNOWLEDGMENTS

The author wishes to thank the following two reviewers of the paper: Dr. Saeid Soltani: Department of Natural Resources, Isfahan University of Technology, Iran, and Dr. Jamal Ghodusi, Department of Natural Resources, Science and Research Branch, Islamic Azad University, Iran.

REFERENCES

Abebe, J., D.P. Solomatine, and R.G.W. Venneker. 2000. Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. Hydrological Sciences-Journal-des Sciences Hydrologiques, Vol.45(3), pp. 425-436.

ASCE. 1996. Hydrology Handbook, second ed. American Society of Civil Engineers (ASCE); New York.

Boulanger, JP., F. Martinez, O. Penalba, and E.C. Segura. 2007. Neural Network based daily precipitation generator (NNGEN-P). Climate Dynamics 28, pp. 307–324.

Chow, V.T. 1964. Handbook of Applied Hydrology. McGraw-Hill; New York.

Coulibaly, P., and N.D. Evora. 2007. Comparison of neural network methods for infilling missing daily weather records. Journal of Hydrology, Vol. 341, pp. 27–41.

Dirks, K.N., J.E. Hay, C.D. Stow, and D. Harris. 1998. High resolution studies of rainfall on Norfolk Island Part II: interpolation of rainfall data. J. Hydrol., Vol. 208(3-4), pp. 187-193.

Goovaerts, P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. hydrol, Vol. 228(2000), pp. 113-129.

Haykin, S. 1999. Neural Networks, A Comprehensive Foundation, Prentice Hall Int; Englewood Cliffs, NJ.

Nalder I.A., and R.W. Wein. 1998. Spatial interpolation of climatic Normal's: test of a new method in the Canadian boreal forest. Agricultural and Forest Meteorology, 92, pp. 211-225.

Malek, M. A., S. Harun , S.M. Shamsuddin, and I. Mohammad. 2009. Reconstruction of Missing Daily Rainfall Data Using Unsupervised Artificial Neural Network. International Journal of Computer Systems Science and Engineering, vol. 4(2), pp. 79-84.

Mahdavi, M. 1998. Applied Hydrology, vol 1, Tehran university publications; Tehran.

McCuen, R.H. 1998. Hydrologic Analysis and Design. Prentice-Hall; Englewood Cliffs, NJ.

Mendel, J.M., and R.W. McClarn. 1970. Reinforcement-Learning Control and Pattern Recognition System. In Adaptive, Learning, and Pattern Recognition Systems: Theory and applications, 66.

Linsley, R.K., M.A. Kohler, and J.L.H. Paulhus. 1988. Hydrology for Engineers. McGraw-Hill Co.; Singapore

Dolling, O.R., and E.A. Varas. 2002. Artificial neural networks for stream flow prediction, Journal of Hydraulic Research.Vol. 40(5), pp. 547-554.

Lucio, P.S., F.C. Conde, I.F.A. Cavalcanti, A.I. Serrano, A.M. Ramos, and A.O. Cardoso. 2007. Spatiotemporal monthly rainfall reconstruction via artificial neural network – case study: south of Brazil Adv. Geosci., Vol. 10, pp. 67–76.

Teegavarapu, R.S.V., and V. Chandramouli. 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. Journal of Hydrology, Vol. 312, pp. 191–206.

Tolika, K., P. Maheras, M. Vafiadis, H.A. Flocas, and A. Arseni-Papadimitriou. 2007. Simulation of seasonal precipitation and rain days over Greece: a statistical downscaling technique based on artificial neural networks (ANNs). International Journal of Climatology Vol. 27(7), pp. 861–881.

Vicente-Serrano, S.M., S. Beguería, J.I. López-Moreno, M.A. García-Vera, Petr Stepanek. 2010. A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. International Journal of Climatology 30:8, 1146-1163

ADDRESS FOR CORRESPONDENCE
Zohre Khorsandi
Department of Natural Resources
Science and Research Branch
Islamic Azad University
Tehran, Iran

Email: khorsandi_zohra@yahoo.com