Kuwait J. Sci. **45 (4)** pp 66 - 75, 2018

Improving the accuracy of rainfall prediction using a regionalization approach and neural networks

Mohammad Arab Amiri[1,*], Christian Conoscenti[2], Mohammad Saadi Mesgari[3]
[1]*Dept. of Geographic Information Systems, Faculty of Geodesy and Geomatics Eng.K. N. Toosi University of Technology, Tehran, IRAN*
[2]*Dept. of Earth and Marine Sciences, University of Palermo, Palermo, Italy*
[3]*Dept. of Geographic Information Systems, Faculty of Geodesy and Geomatics Eng., and Center of Excellence in Geospatial Information Technology (CEGIT) K. N. Toosi University of Technology, Tehran, IRAN*
**Corresponding author:* mohamadamiri89@yahoo.com

Abstract

Spatial and temporal analysis of precipitation patterns has become an intense research topic in contemporary climatology. Increasing the accuracy of precipitation prediction can have valuable results for decision-makers in a specific region. Hence, studies about precipitation prediction on a regional scale are of great importance. Artificial Neural Networks (ANN) have been widely used in climatological applications to predict different meteorological parameters. In this study, a method is presented to increase the accuracy of neural networks in precipitation prediction in Chaharmahal and Bakhtiari Province in Iran. For this purpose, monthly precipitation data recorded at 42 rain gauges during 1981-2012 were used. The stations were first clustered into well-defined groupings using Principal Component Analysis (PCA) and Cluster Analysis (CA), and then one separate neural network was applied to each group of stations. Another neural network model was also developed and applied to all the stations in order to measure the accuracy of the proposed model. Statistical results showed that the presented model produced better results in comparison to the second model.

Keywords: Artificial Neural Networks (ANN); Chaharmahal and Bakhtiari Province; Cluster Analysis (CA); precipitation.

1. Introduction

Precipitation is a complex nonlinear phenomenon and its distribution varies in time and space, but there are many studies in the literature showing that precipitation is predictable (Azadi & Sepaskhah, 2012; Arab Amiri *et al.*, 2016). An accurate rainfall prediction is crucial for water resource planning and management. Rainfall variations can also have various impacts on human society, economy, agriculture, ecosystems, and the environment (Alexander *et al.*, 2006). Hence, spatial and temporal modeling of precipitation is of utmost importance.

Several researchers worldwide have attempted to accurately predict the spatial and temporal distribution of rainfall using various techniques such as simple linear regression and Artificial Neural Networks (ANN) (Vamsidhar *et al.*, 2010; Azadi & Sepaskhah, 2012; Deshpande, 2012; Nanda *et al.*, 2013; Dubey, 2015; Mislan *et al.*, 2015; Sharma & Nijhawan, 2015; Arab Amiri *et al.*, 2016). However, the accuracy of prediction obtained by some of these techniques could not achieve a satisfactory level because of the complex and nonlinear nature of precipitation over the region. An artificial neural network algorithm is an inductive, data driven approach that is able to model both linear and non-linear systems without needing to make pre-assumptions. It is the most popular approach for rainfall prediction (Kuligowski & Barros, 1998; Nayak *et al.*, 2013; Arab Amiri *et al.*, 2016).

There are various sources of raw data that can be used for rainfall forecasting, e.g. weather stations, weather radars and satellite sensors (Adamowski *et al.*, 2013; Bai *et al.*, 2014; Devasthale & Norin, 2014). Precipitation time-series have been used in this study. They are recorded by weather stations and composed of the geographic coordinates of static space object and corresponding time series data.

A hybrid model was developed to predict the temporal behavior of rainfall across Chaharmahal and Bakhtiari Province, a principle administrative division in the central part of Iran. For this purpose, the weather stations were grouped into clusters using a spatial regionalization approach which uses Principal Component Analysis (PCA) and Cluster Analysis (CA) (Arab Amiri & Mesgari, 2016). Then the weather stations which were grouped in each cluster were entered into separate ANN models. The prediction accuracy of the developed ANN models was compared with an ANN which was developed for all the stations.

2. Materials and methods

2.1. Study area and data

Chaharmahal and Bakhtiari Province is situated between 31° 09'- 32° 48' N and 49° 28' - 51° 25' E in the central part of Iran (Fig. 1). It is located in the foothills of the Zagros Mountain range (Nabavi *et al.*, 2014), and has a land area of about 16,332 km^2. The total population was over 900,000.

There are four main rivers in the Province, including Zayandeh-rood, Karoon, Karkheh and Dez. Rain and snow melt from the Zagros Mountains flow into these rivers, making the area a significant source of fresh water in Iran (Alahbakhshian Farsani *et al.*, 2013). The mean annual rainfall over the region is shown in Fig. 1.

Monthly precipitation data were obtained from the Islamic Republic of Iran Meteorological Organization (IRIMO). In this study, the recorded monthly rainfall data were used from 42 weather stations between January 1981 and December 2012. The rainfall data were used from 1981 until 2012 because the precipitation data after 2012 were not available. Quality control was applied to test the homogeneity of the used time series and suspect data were discard. There are no missing data values during the considered period. The homogeneity of the series was also tested using the double-mass curve analysis. If a distinct break is found in a double mass plot, the records before and after the break point are inconsistent and require adjustment. The adjustment consists of adjusting the slope of the curve before the break point to confirm to the slop after it and vice versa. The double-mass

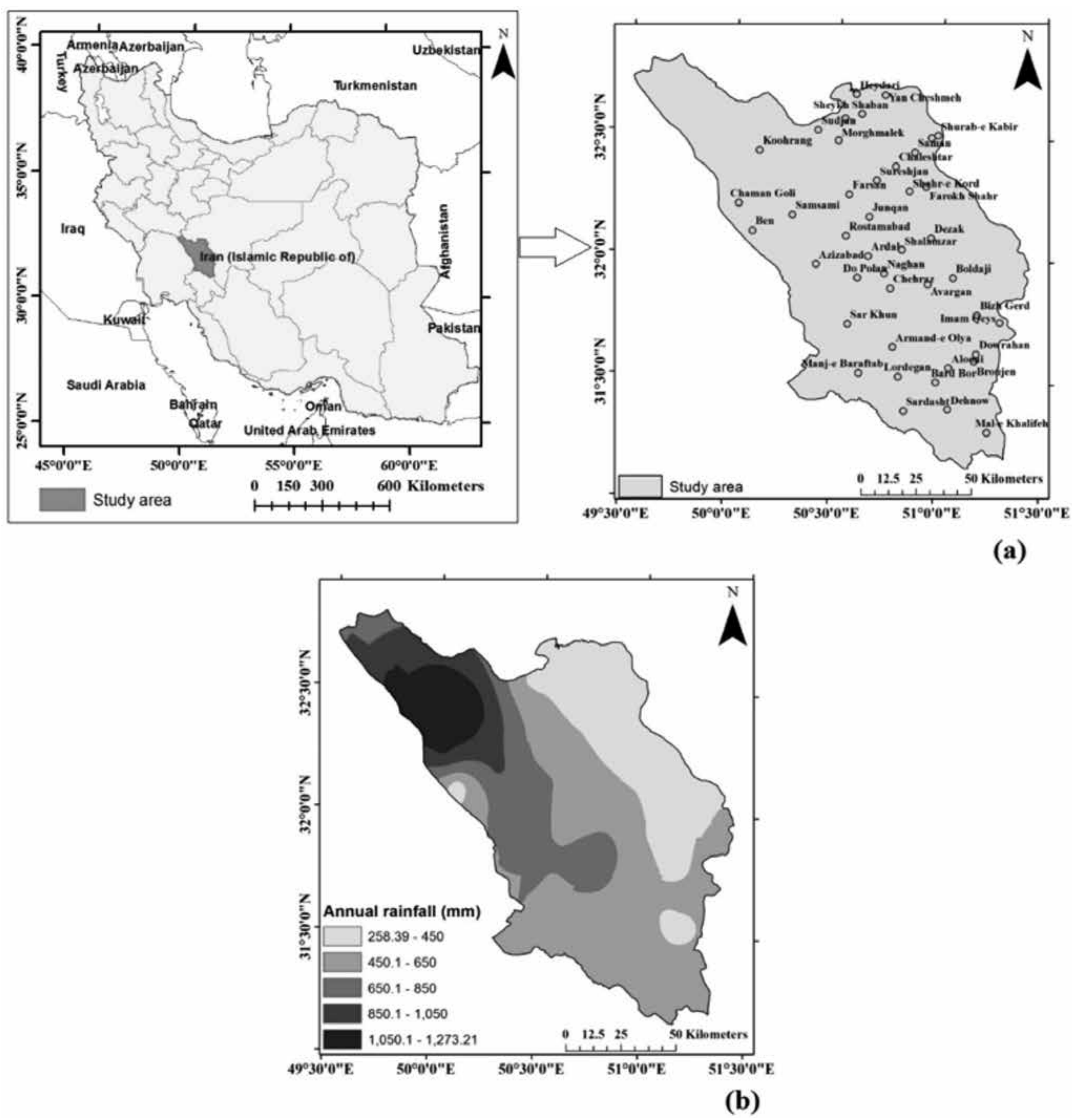

Fig. 1. (a) Study area location, and weather stations in the region, (b) spatial distribution of mean annual rainfall

curve analysis is explained in more details in Kohler (1949). Finally, results showed that the recorded time series by all the stations were sufficiently homogeneous for variability analysis.

2.2. Spatial regionalization of precipitation

A combination of PCA and CA was applied to classify the study area into homogeneous precipitation sub-zones. PCA is a multivariate technique for reducing the dimensionality in a dataset and computing new orthogonal variables called Principal Components (PCs). Six modes of PCA were defined by Richman (1986) for climatological analysis. In this study, the T-mode PCA was used for studying the co-variability of precipitation at given stations across the study region. Hence, the input data matrix was arranged as rows correspond to the stations and columns refer to the observations. The input data consist of annual total precipitation values which were calculated using the original monthly rainfall data.

Applying the Kaiser's criterion for selecting PCs that should be retained, the varimax procedure was used to rotate the retained PCs in order to maximize the correlation between the variables and the components. Then the K-means method, which is the most popular approach for cluster analysis, was used to simplify the spatial structure of the rotated PCs. The optimum number of clusters was determined using the variance ratio criterion (VRC). More information about the VRC index and cluster analysis can be found in Caliński & Harabasz (1974).

2.3. Artificial Neural Networks

ANN is the most widely used approach in the field of rainfall prediction. The method was first developed by Minsky & Papert (1969). An ANN model is comprised of a parallel distributed processor made up of simple processing units called artificial neurons. The basic elements of a neuronal model include a set of connection links, an adder, an activation function and an external bias (Nayak *et al.*, 2013). There are different types of neural networks according to their architecture. The multilayer feedforward neural network is one of the most widely used network architectures. This type has one or more layers of nodes between the input and output nodes. The main advantage of this type of network over the single layer feedforward neural network is the non-linear relationships that is present among its nodes (Hung *et al.*, 2009).

A network can find the values of connection weights using the training algorithm. There are different types of training algorithms that differ in how the weights are obtained (Hung *et al.*, 2009). Coulibaly *et al.* (2000) claimed that more than 90% of researchers have used the back propagation algorithm as the training algorithm of ANN models in hydrological applications. The back propagation algorithm is still the most popular and most effective learning algorithm for multilayered networks (Nayak *et al.*, 2013). Hence, the standard back propagation algorithm has been used for training the network. The back propagation algorithm attempts to propagate the errors through the network and adjusting network weights along with this back propagation. In this study, a multi-layered feedforward back propagation algorithm was used. As stated by Nayak *et al.* (2013), multi-layered feedforward back propagation algorithms have been used by most of researchers for rainfall prediction.

Another issue of great importance concerning the implementation of neural networks is specifying the network architecture or the number of nodes and layers in the network. For this purpose, a trial and error procedure was applied to determine the optimum number of clusters. This method begins by applying a small number of nodes, and then the network size is gradually increased. The procedure ends when no significant change in the mean square error at the output layer is achieved.

Two models were developed to precisely predict rainfall over the region. The first model regionalizes the study area into homogeneous precipitation sub-zones and develops a separate neural network for each sub-region. The second model develops one network for all the stations in the study area, and is developed to test the accuracy of the first model. The general framework of the methodology used in this study is shown in Fig. 2.

3. Results and discussion

Annual precipitation data were used to classify the stations into well-defined clusters. Applying the T-mode PCA to the total annual rainfall time-series, three PCs were retained by the Kaiser's criterion. The first three PCs have more than 91% of the total variance. The retained PCs were then

Table 1. Explained variance (%) by the loadings with and without rotation for annual precipitation.

Explained variance (%)	Principal components			Cumulative percentage of total variation
	PC-1	PC-2	PC-3	
Un-rotated (%)	81.45	5.47	4.22	91.14
Varimax rotated (%)	34.82	25.99	30.33	91.14

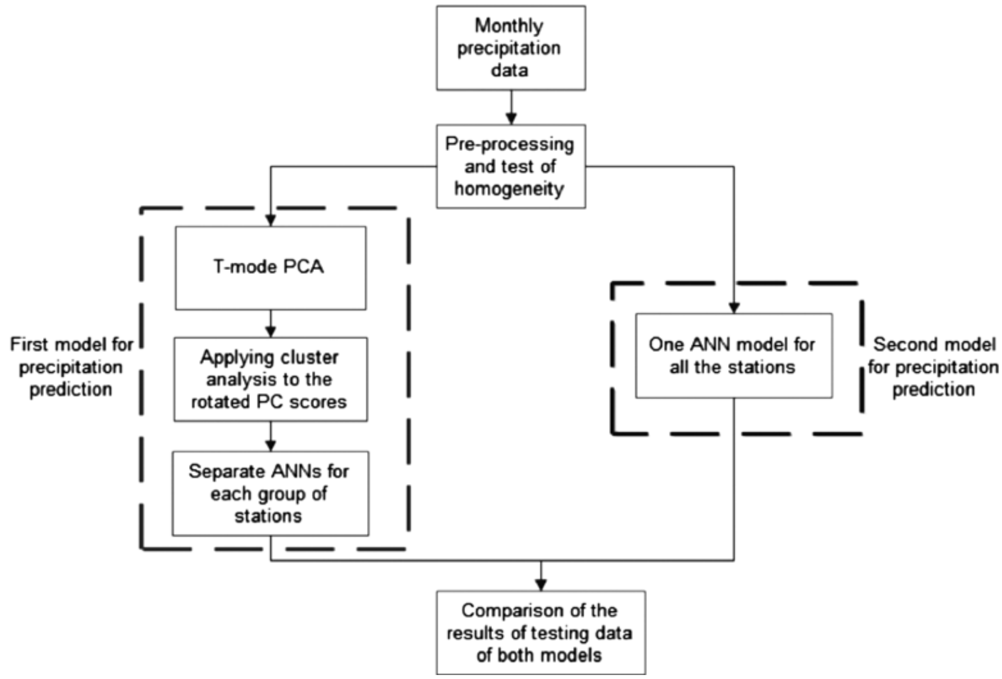

Fig. 2. General methodology framework

rotated by the varimax procedure, which is the most widely used rotation approach (Table 1).

The spatial patterns of the rotated PC scores are depicted in Fig. 3. The three PC scores showed similar patterns, pointing relatively to the same pattern represented by mean annual rainfall (Fig. 1). This indicates that the spatial pattern of mean annual rainfall is shown well by the rotated PC scores.

The rotated PC scores were then subjected to cluster analysis. The K-means clustering was applied to the rotated PC scores in order to classify the weather stations into homogeneous precipitation sub-regions. As stated in the previous section, the optimum number of clusters was determined through the VRC index. The predefined cluster centroids were chosen randomly, and the sum of squared errors between each station and its corresponding centroids were calculated. This sum was minimized by assigning each station to the closest centroid and re-computing new centroids. Having applied the cluster analysis, the weather stations were classified into three distinct groups (Fig. 4). The absolute amounts and percentages of between-classes variance, within-class variance, and total variance of the obtained clusters are presented in Table 2.

Table 2. Variance decomposition of the T-mode PCA using annual precipitation data.

Variance decomposition	Absolute	Percent
Between-classes variance	3910.27	84.45
Within-class variance	720.11	15.55
Total variance	4630.37	100

Regarding precipitation prediction, two models were developed. The first model utilized three different neural networks for each group of stations, while the second model applied one neural network for all the stations. There are some studies in the literature that used one neural network for multiple stations (Weerasinghe *et al.*, 2007; Hung *et al.*, 2009; Azadi & Sepaskhah, 2012).

Table 3. Structure of the neural networks in this study, and errors obtained from cross-validation of each model

Model no.	Sub-region	Input and output variables	Activation function for the hidden layer	Activation function for the output layer	Learning rate	Epoch	The average of RMSE (mm)
	1	P(t)= P(t-1), P(t-2)	Logarithmic sigmoidal	Pure linear	0.01	10	11.01
1	2	P(t)= P(t-1), P(t-2)	Logarithmic sigmoidal	Pure linear	0.01	12	18.27
	3	P(t)= P(t-1), P(t-2)	Logarithmic sigmoidal	Pure linear	0.01	13	30.23
2	-	P(t)= P(t-1), P(t-2)	Logarithmic sigmoidal	Pure linear	0.01	16	35.49

P precipitation, *t* time

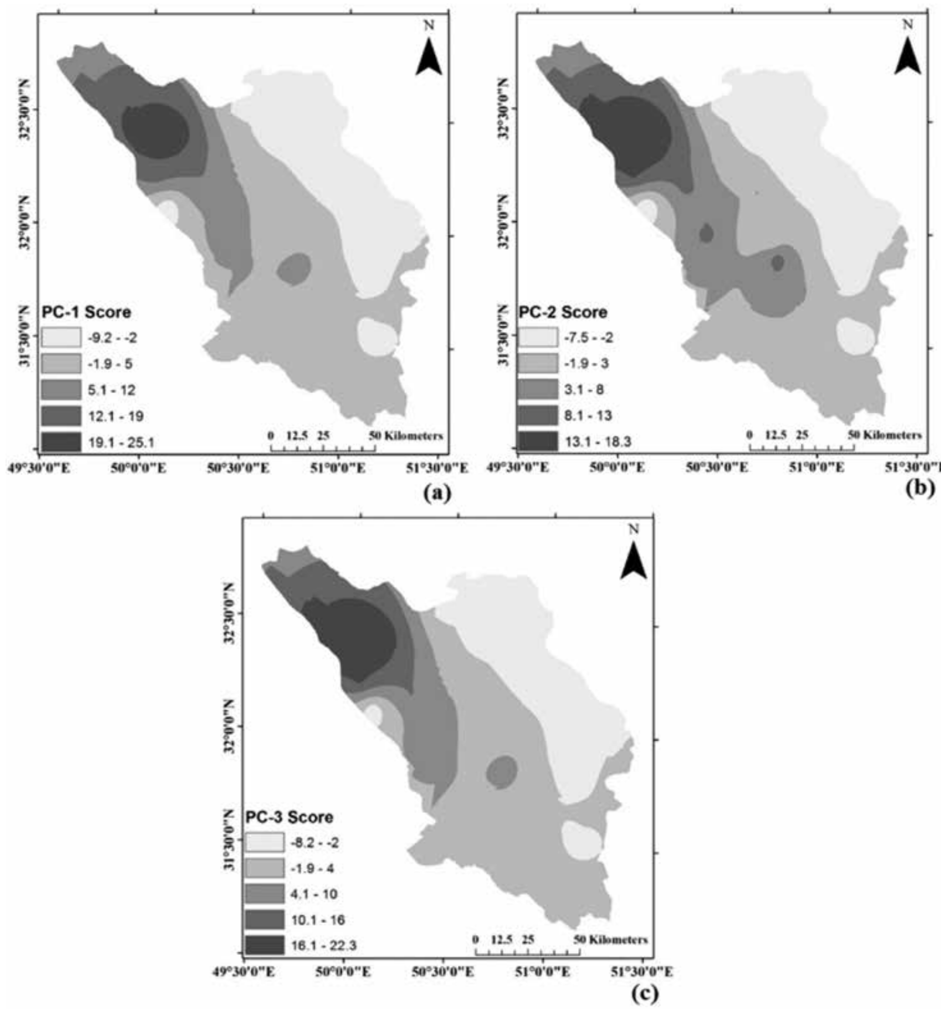

Fig. 3. Spatial distribution of the varimax rotated PC scores: (a) PC-1, (b) PC-2, and (c) PC-3

Since monthly precipitation data of 32 years were used, the input dataset includes 384 samples corresponding to the 12 months of a year for each station. The input data were classified into 3 datasets, namely calibration, validation and testing datasets.

The recorded precipitation data from 1981 to 2008 were used for training the networks. For this purpose, 60% of the training data was selected randomly for calibration of the weights of the neural networks, and the remaining 40% was set aside for cross validation or monitoring the progress of training process. The recorded monthly precipitation data of the years 2009, 2010, 2011 and 2012 were used as testing data to evaluate the performance of the developed models.

A pictorial representation of the used neural network is presented in Fig. 5. Moreover, the architectures of the developed models are presented in Table 3.

The input layer consists of two antecedent rainfall values at times t-2 and t-1; and the output layer of the neural network consists of the precipitation at time t. For all the developed neural networks, the logarithmic sigmoidal

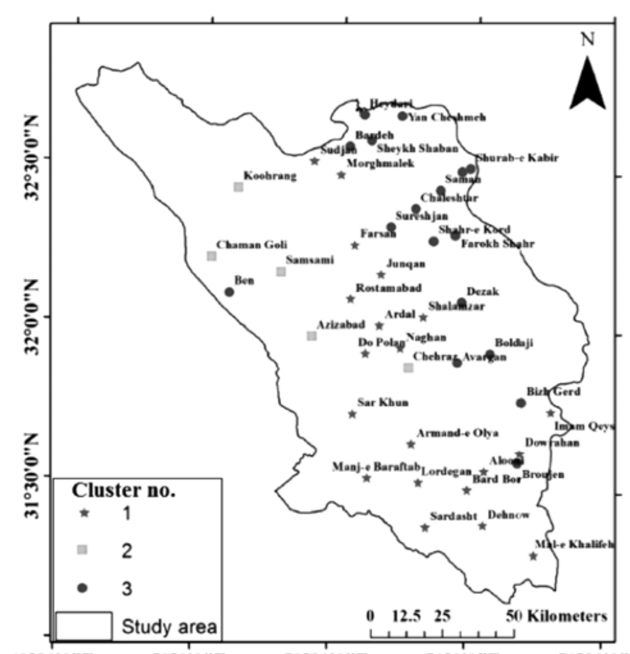

Fig. 4. Homogeneous clusters for which the amount of precipitation (mm) predominates

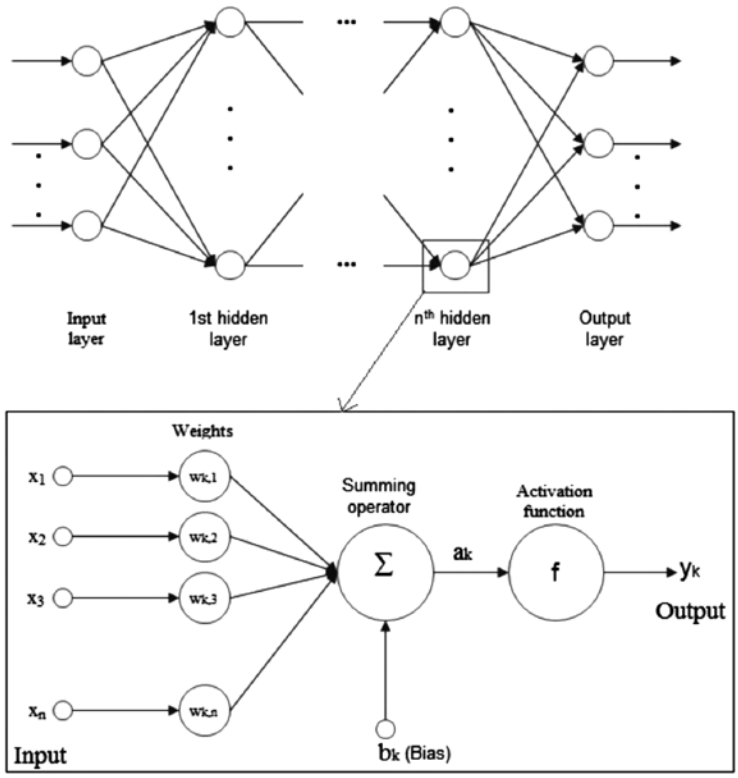

Fig. 5. Pictorial representation of the multilayer feedforward neural network

function and pure linear function were used as the activation function for the hidden layer and output layer, respectively. The average values of Root Mean Square Error (RMSE) indicate that the first model performed better than the second model. In other words, applying one neural network to each homogeneous sub-region showed less RMSE values than applying one network to the whole region.

Applying the developed neural networks to each group of stations, the accuracy of the networks for the testing data was calculated at each station (Table 4). Three statistical

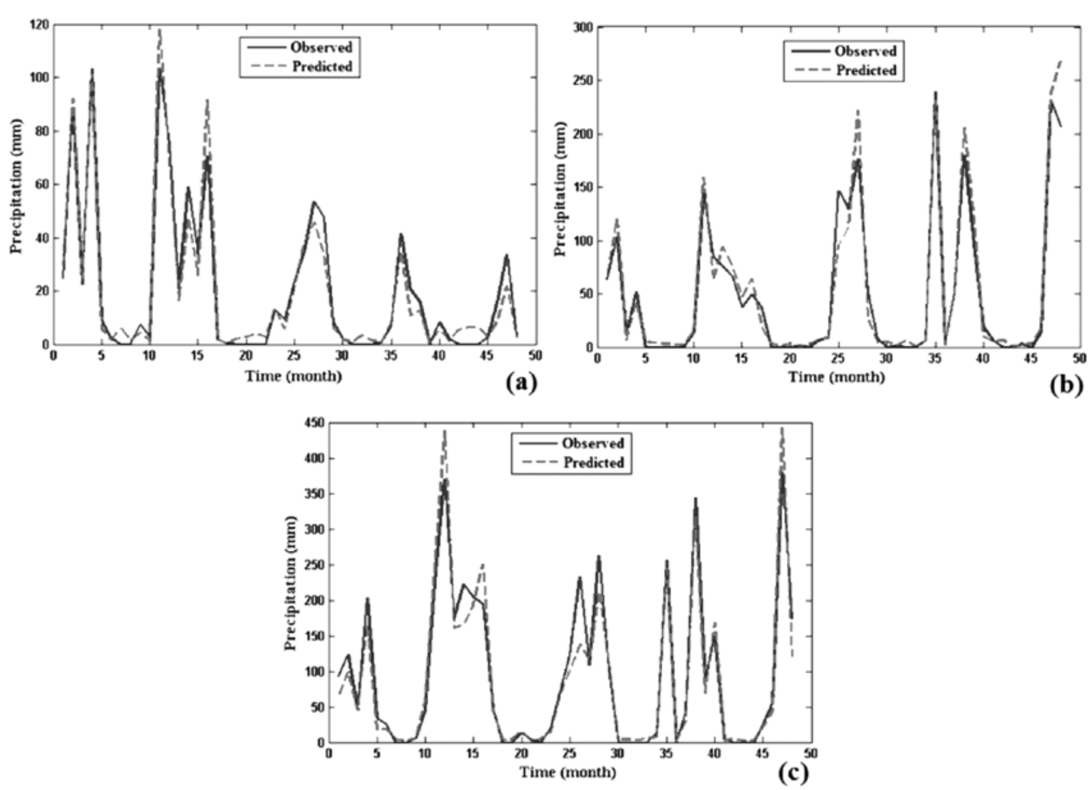

Fig. 6. The predicted and observed time series of monthly precipitation at (a) Shahr-e Kord, (b) Mal-e Khalifeh, and (c) Chehraz stations, during the testing period (2010-2012)

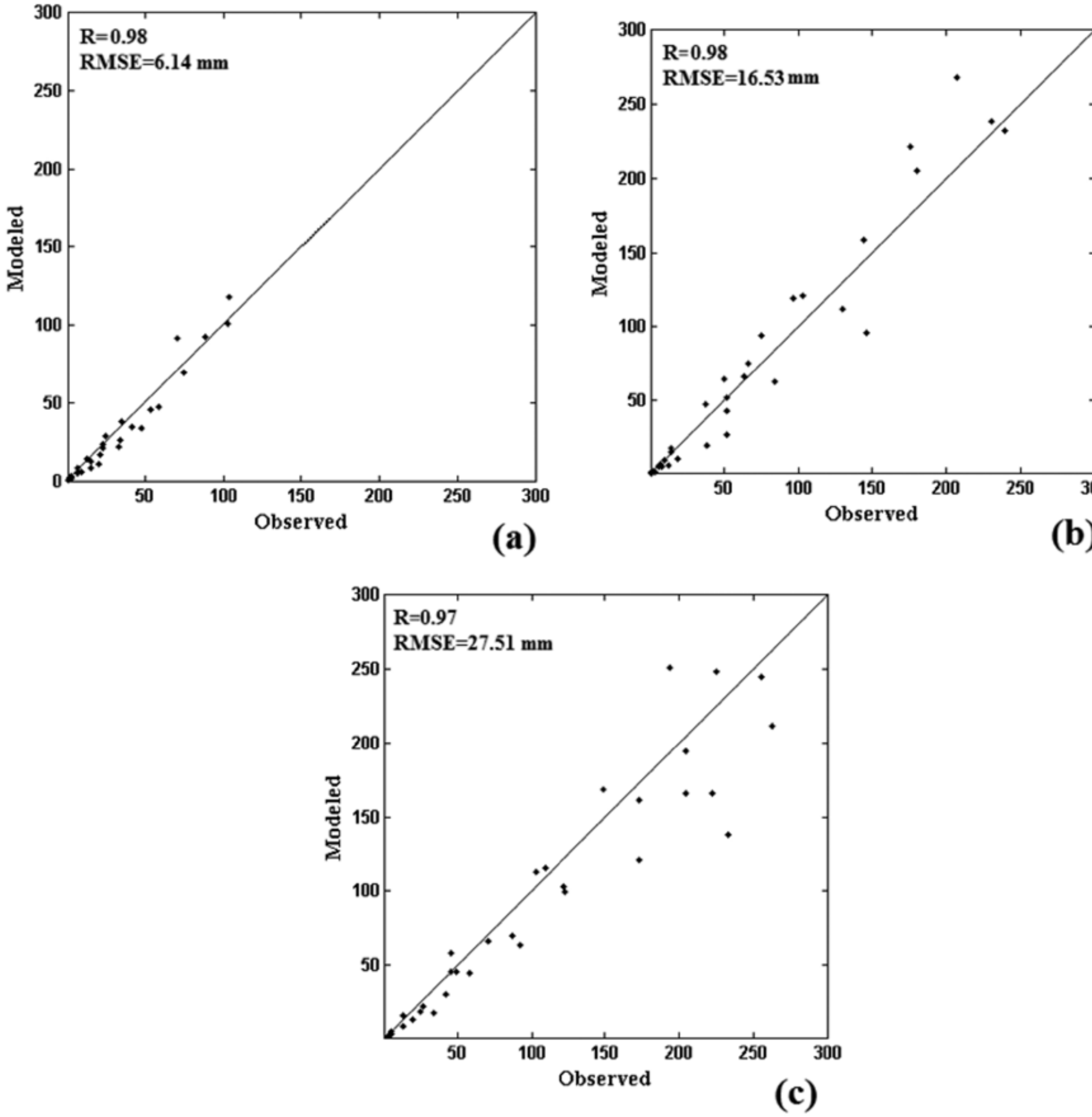

Fig. 7. Scatter plots between the observed and modeled monthly precipitation at (a) Shahr-e Kord, (b) Mal-e Khalifeh, and (c) Chehraz stations, during the testing period

indices were used to evaluate the performance of the neural networks, namely Mean Error (ME), RMSE and Correlation Coefficient (R). The magnitude of ME and RMSE are relatively small in comparison with the magnitude of precipitation at each station. The values of R are also close to one at all the stations, which means that the developed networks truly predicted our set of data.

One station in each cluster with the best performance was selected as a sample station in order to undertake a time series comparison. Fig. 6 depicts the observed and predicted values of monthly precipitation which were plotted against each other in a graph for each selected station during the testing period (i.e. 2010 to 2012). It can be concluded that there are small over-predictions and under-predictions at all the three stations during the testing period.

The scatter plots between the observed and modeled monthly precipitation for the selected stations are also shown in Fig. 7. The scatter plots showed that the network which was developed for the first sub-region performed slightly better at Shahr-e Kord station than the other two networks at Mal-e Khalifeh and Chehraz stations. However, the difference is negligible and all the developed networks of the proposed model performed well for rainfall prediction.

4. Conclusions

This research aimed to present a method for improving the accuracy of precipitation prediction using neural networks over a region. For this purpose, monthly precipitation data from 1981-2012 in Chaharmahal and Bakhtiari Province, an important watershed area, were used. The proposed model first grouped the stations into well-defined clusters and then one separate neural network was developed for each of the clusters. Monthly precipitation values at two antecedent time lags were used as inputs of the networks. To assess the efficiency of the proposed method, a neural network was also

Table 4. Accuracy of fit statistics of the testing data sets of the developed model

Sub-model no.	Station name	ME (mm)	RMSE (mm)	R
1	Avargan	-3.473	14.946	0.956
	Boldaji	-1.556	11.384	0.963
	Dezak	0.002	10.855	0.976
	Pol-e Zamankhan	-1.776	10.003	0.963
	Broujen	-1.019	9.730	0.938
	Farokh Shahr	0.576	7.661	0.973
	Saman	-3.139	14.027	0.917
	Shahr-e Kord	-0.458	6.139	0.977
	Bardeh	0.454	9.460	0.957
	Bizh Gerd	-2.703	10.591	0.962
	Ben	-0.666	9.995	0.954
	Chaleshtar	-2.953	13.778	0.929
	Heydari	-0.349	10.568	0.956
	Sheykh Shaban	-2.102	12.684	0.960
	Shurab-e Kabir	-2.256	9.089	0.948
	Sureshjan	-2.944	16.504	0.928
	Yan Cheshmeh	-1.175	9.836	0.961
2	Imam Qeys	-5.115	20.750	0.948
	Ardal	-5.137	16.749	0.947
	Farsan	-0.694	12.581	0.967
	Lordegan	-2.489	17.161	0.954
	Alooni	-3.878	15.488	0.949
	Armand-e Olya	-5.624	25.318	0.943
	Bard Bor	-2.601	19.709	0.936
	Dehnow	-6.516	19.222	0.947
	Do Polan	-4.143	17.528	0.949
	Dowrahan	-4.590	20.082	0.939
	Junqan	-4.990	20.904	0.944
	Manj-e Baraftab	-2.697	17.579	0.948
	Morghmalek	-1.841	14.693	0.935
	Naghan	-7.271	21.235	0.949
	Rostamabad	-1.829	21.624	0.943
	Sardasht	-1.320	22.536	0.943
	Sar Khun	-4.504	16.295	0.969
	Shalamzar	-0.642	13.372	0.973
	Sudjan	-5.183	16.026	0.950
3	Koohrang	-12.360	35.612	0.938
	Azizabad	1.389	23.162	0.963
	Chehraz	-4.239	27.506	0.970
	Chaman Goli	-8.053	34.925	0.934
	Samsami	-6.936	29.933	0.933

applied to all the stations. Results pointed out that regionalizing the Province and then applying one separate neural network to each homogeneous group of stations more accurately predicted monthly rainfall values over the study area. This is contrary to a research study by Azadi & Sepaskhah (2012). This investigation of precipitation prediction in west, southwest, and south provinces of Iran, the researchers divided the region into homogeneous precipitation sub-zones, but this did not noticeably increase the accuracy of the neural network predictions.

The antecedent precipitation values at two time lags were used as inputs to the neural networks because only the recorded data of rain gauge stations were used. Using other weather data such as temperature, wind speed, and dew point as inputs are intriguing topics for future studies that may increase the accuracy of rainfall prediction in the region.

References

Adamowski, J., Adamowski, K. & Prokoph, A. (2013). Quantifying the spatial temporal variability of annual streamflow and meteorological changes in eastern Ontario and southwestern Quebec using wavelet analysis and GIS. Journal of Hydrology, **499**: 27-40.

Alahbakhshian Farsani, P., Habibnejad Roshan, M., Vahbzade, G. & Solaimani, K. (2013). Investigation of Trend of Precipitation Variation Using Non-Parametric Methods in Charmahal O Bakhtiari Province. International Journal of Advanced Biological and Biomedical Research, **1**(5): 547-555.

Alexander, L., Zhang, X., Peterson, T., Caesar, J., Gleason, B., *et al.* (2006). Global observed changes in daily climate extremes of temperature and precipitation. Journal of Geophysical Research: Atmospheres, **111**(D5): 1-22.

Arab Amiri, M., Amerian, Y. & Mesgari, M.S. (2016). Spatial and temporal monthly precipitation forecasting using wavelet transform and neural networks, Qara-Qum catchment, Iran. Arabian Journal of Geosciences, **9**(5): 421.

Arab Amiri, M. & Mesgari, M.S. (2016). Spatial variability analysis of precipitation in northwest Iran. Arabian Journal of Geosciences, **9**(11): 578.

Azadi, S. & Sepaskhah, A.R. (2012). Annual precipitation forecast for west, southwest, and south provinces of Iran using artificial neural networks. Theoretical and Applied Climatology, **109**(1-2): 175-189.

Bai, X., Wang, Y., Jiang, H., Liao, Z., Xiong, Y., *et al.* (2014). Mining high-temperature event space-time regions in geo-referenced temperature series data. 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). IEEE, 671-676.

Caliński, T. & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics-theory and Methods, **3**(1): 1-27.

Coulibaly, P., Anctil, F. & Bobee, B. (2000). Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. Journal of Hydrology, **230**(3): 244-257.

Deshpande, R.R. (2012). On the rainfall time series prediction using Multilayer Perceptron Artificial Neural Network. International Journal of Emerging Technology and Advanced Engineering, **2**(1): 2250-2459.

Devasthale, A. & Norin, L. (2014). The large-scale spatio-temporal variability of precipitation over Sweden observed from the weather radar network. Atmospheric Measurement Techniques, **7**(6): 1605-1617.

Dubey, A.D. (2015). Artificial neural network models for rainfall prediction in Pondicherry. International Journal of Computer Applications, **120**(3): 30-35.

Hung, N.Q., Babel, M.S., Weesakul, S. & Tripathi, N. (2009). An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrology and Earth System Sciences, **13**(8): 1413-1425.

Kohler, M.A. (1949). Double-mass analysis for testing the consistency of records and for making adjustments. Bulletin of the American Meteorological Society, **30**: 188-189.

Kuligowski, R.J. & Barros, A.P. (1998). Experiments in short-term precipitation forecasting using artificial neural networks. Monthly Weather Review, **126**(2): 470-482.

Minsky, M. & Papert, S. (1969). Perceptrons. An introduction to computational geometry. MIT Press, Cambridge, MA. Pp. 258.

Mislan, Haviluddin, Hardwinarto, S., Sumaryono & Aipassa, M. (2015). Rainfall monthly prediction based on Artificial Neural Network: A case study in Tenggarong Station, East Kalimantan, Indonesia. Procedia Computer Science, **59**: 142-151.

Nabavi, R., Naeini, K.M., Zebardast, N. & Hashemi, H. (2014). Epidemiological study of gastrointestinal helminthes of canids in Chaharmahal and Bakhtiari Province of Iran. Iranian Journal of Parasitology, **9**(2): 276-281.

Nanda, S.K., Tripathy, D.P., Nayak, S.K. & Mohapatra, S. (2013). Prediction of rainfall in India using Artificial Neural Network (ANN) models. International Journal of Intelligent Systems and Applications, **5**(12): 1-22.

Nayak, D.R., Mahapatra, A. & Mishra, P. (2013). A survey on rainfall prediction using artificial neural network. International Journal of Computer Applications, **72**(16).

Richman, M.B. (1986). Rotation of principal components. Journal of Climatology, **6**: 293-335.

Sharma, A. & Nijhawan, G. (2015). Rainfall prediction using neural network. International Journal of Computer Science Trends and Technology, **3**(3): 65-69.

Vamsidhar, E., Varma, K., Rao, P.S. & Satapati, R. (2010). Prediction of rainfall using backpropagation neural network model. International Journal on Computer Science and Engineering, **2**(4): 1119-1121.

Weerasinghe, H., Premaratne, H. & Sonnadara, D. (2007). A neural network based rainfall forecasting system using multiple stations. 25th National IT Conference, Colombo, Sri Lanka. 82-86.

Submission : 08/01/2017
Revision : 10/04/2017
Acceptance : 02/05/2017

تحسين الدقة في التنبؤ بمعدل سقوط الأمطار باستخدام نهج الأقلمة والشبكات العصبية

محمد عرب أميري[1،*]، كريستيان كونوسنتي[2]، محمد سعدي مسجاري[3]

[1] قسم نظم المعلومات الجغرافية، كلية هندسة الجيوديسيا وجيوماتكس، جامعة K. N. Toosi للتكنولوجيا، طهران، إيران
[2] قسم علوم الأرض والعلوم البحرية، جامعة باليرمو، باليرمو، إيطاليا
3 قسم نظم المعلومات الجغرافية، كلية هندسة الجيوديسيا وجيوماتكس، ومركز التميز في تكنولوجيا المعلومات الجغرافية المكانية (CEGIT)، جامعة K. N. Toosi للتكنولوجيا، طهران، إيران

* mohamadamiri89@yahoo.com

الملخص

أصبح التحليل الزمني والمكاني لأنماط هطول الأمطار من المواضيع الساخنة في علم المناخ المعاصر. يمكن أن يكون للزيادة في دقة التنبؤ بهطول الأمطار نتائج قيمة لمتخذي القرار في مناطق معينة. لذلك، فإن الدراسات حول التنبؤ بهطول الأمطار على المستوى الإقليمي لها أهمية كبيرة. تم استخدام الشبكات العصبية الصناعية بشكل واسع في التطبيقات المناخية للتنبؤ بمعلمات المناخ المختلفة. نقدم في هذه الدراسة طريقة لزيادة درجة الدقة للشبكات العصبية للتنبؤ بهطول الأمطار في مقاطعات شاهرماهال وبختياري في إيران. ولهذا الغرض، تم استخدام بيانات شهرية عن الأمطار في 42 محطة خلال الفترة 1981 – 2012. في البداية تمت عنقدة المحطات في عناقيد معرفة بشكل كامل باستخدام تحليل المكون الرئيسي وتحليل العنقدة، ومن ثم تم تطبيق شبكة عصبية منفصلة لكل مجموعة من المحطات. تم أيضاً تطوير وتطبيق شبكة عصبية أخرى لجميع المحطات لقياس دقة النموذج المقترح. أظهرت النتائج الإحصائية أن النموذج المقدم أنتج نتائج أفضل بالمقارنة مع النموذج الثاني.