



Trends in
**Applied Sciences
Research**

ISSN 1819-3579



Academic
Journals Inc.

www.academicjournals.com

The Application of Factor Analysis and Artificial Neural Networks in Predicting Spring Precipitation by Means of Climatic Parameters of the Upper Levels of Atmosphere

¹M.H. Nokhandan, ²G.A.F. Ghalhary and ³M. Mousavi-Baygi

¹Climatological Research Institute, Mashhad, Iran

²Isfahan University and Researcher of Climatological Research Institute, Mashhad, Iran

³Department of Water Engineering, Faculty of Agriculture,
Ferdowsi University of Mashhad, Iran

Abstract: This research aims to study the relationship between climatic large-scale synoptic patterns of the upper levels of atmosphere and rainfall in Khorasan-e Razavi Province. Artificial neural networks and factor analysis were used in this study to predict rainfall in the period between April and June in the province. At the first the relationship between average regional rainfall and the changes in synoptic patterns including the temperature of 700 mb level, the thickness between 500 and 1000 mb levels and the relative humidity of 300 mb level were analyzed. In the selection of these regions, we have considered the effect of synoptic patterns in these regions on the rainfall in the northeast region of Iran. Then, artificial neural networks model for the period 1970-1997 were taught. Finally, the rainfall in the period 1998-2007 has been predicted. The results show that artificial neural networks can predict rainfall with reasonable accuracy in all years. The root mean-square error of the model was 5 mm.

Key words: Rainfall prediction, synoptic patterns, artificial neural networks, factor analysis

INTRODUCTION

All aspects of human life are, directly or indirectly, affected by climatic processes. This effect is more noticeable in such fields as agriculture, irrigation, economy, telecommunications, transportation, traffic, air pollution and military industries (Haltiner and Williams, 1980).

Astine (2001) used the multiple regression models to predict seasonal precipitation in Northern Nigeria. The model was tested on the basis of the amount of the required rainfall for the non-irrigated growth of three crops, namely corn, sorgom and millet during the years 1991-1995. The results indicated a strong correlation between the observed and predicted values which were 0.98 and 0.91, respectively for millet and sorgom, which can be grown by dry-farming.

Sen (2003) used the regression model to predict the seasonal monsoon rainfall in Southwest India. The results indicated the efficiency of the regression methods in the prediction of seasonal rainfall. Singhrattna *et al.* (2005) used statistical methods to predict estival monsoon rainfalls over Thailand. Their results suggest that the predictions of the non-parametric regression model in limit (dry or wet) years are more suitable. Ashby *et al.* (2005) used the multi-variable regression model to predict the rainfall in the premature wet season (June to July) and the late wet season (August to October) in the Caribbean Sea basin. The results showed the efficiency of the regression model in predicting seasonal rainfall. The results also indicate a spatial variety in the importance of seasonal predictors.

Gissila *et al.* (2004) studied the seasonal rainfalls of Ethiopia in the period between June and September. Their results indicated that rainfall forecast was very efficient for the western part of the central Ethiopia. Maria *et al.* (2005) used neural networks and regression models to forecast the rainfall in Sao Paulo, Brazil. The results indicated the efficiency of both methods in forecasting the rainfall. Singh *et al.* (2004) used small-scale atmospheric circulation to predict monsoon rainfall in Anas basin, India. Their results indicate that the statistical small-scale model is preferable for rainfall forecast based on daily atmospheric circulation pattern in Indian semi-arid regions. Mooley and Paolino (2006) studied the relationship between fields of geo-potential height at 700 and 500 hpa levels in northern hemisphere and monsoon rainfalls in India (June to September). The findings revealed that the best region for the effect of geo-potential height at 700 hpa level on Indian monsoon rainfall is over the Pacific in the time-span January to April.

Prasad and Singh (2006) applied factor analysis in predicting the monthly rainfall of geo-potential fields of 700 hpa level over India. The results indicated high efficiency of this method in predicting the monthly rainfall of geo-potential fields of 700 mb level. George and Kimber (2007) used such variables as humidity, wind speed, maximum and minimum daily temperature and the degree of hot and cold days to predict the seasonal rainfall in Florida State. Their results were indicative of the efficiency of regression method in rainfall forecast.

Kim *et al.* (2007) analyzed the effects of large-scale climatic signals such as NAO and SOI on the seasonal rainfall of Colorado River basin in US. Their results showed that modeling the dynamic systems of the climate can be useful in developing a long-term prediction model which will be valuable for the management of water resources. Nazemosadat and Cordery (2000) pointed out that the changes in the temperature of water surface in the Persian Gulf have a significant effect on precipitation changes in Southern and Southwestern Iran. Nazemosadat and Cordery (2000) studies indicated that winter precipitation (January to March) in these regions is in inverse proportion to SST of the Persian Gulf.

Due to the significance of rainfall in many decision making processes such as water resources management and agriculture, the present study aims to find out the relationship between large-scale climatic synoptic patterns and regional rainfall using such synoptic patterns as air temperature at 700 mb level, the thickness between 500 and 1000 mb levels and the relative humidity at 300 mb level.

MATERIALS AND METHODS

The Region Under Study

The region studied in this research is Khorasan-e Razavi Province. The time series studied is the average rainfall from April to June during 38 years. The data of spring rainfall for each year includes the rainfall in 38 synoptic, climatology and rain gauge stations provided by the Weather Bureau and the Power Ministry. Of these, 24 stations rain gauge stations of the Power Ministry and the rest belong to the Weather Bureau. Figure 1 represents the map of the studied area and the name of the relevant stations. To compensate for some defects in rainfall data, subtractions and ratios method have used. Run test was also used to test the homogeneity of the data.

Required Data

Except for rainfall data which were obtained from the Weather Bureau, all other required data were adopted from the site NOAA (www.cdc.noaa.gov) in the networks of 2.5×2.5 degrees in the period between 1970 and 2007.

Calculating the Average Regional Rainfall

The final goal of studying spatial changes of rainfall is to simulate the changes in rainfall data in the spatial dimension in order to pave the way for attaining other goals such as forecasting rainfall and

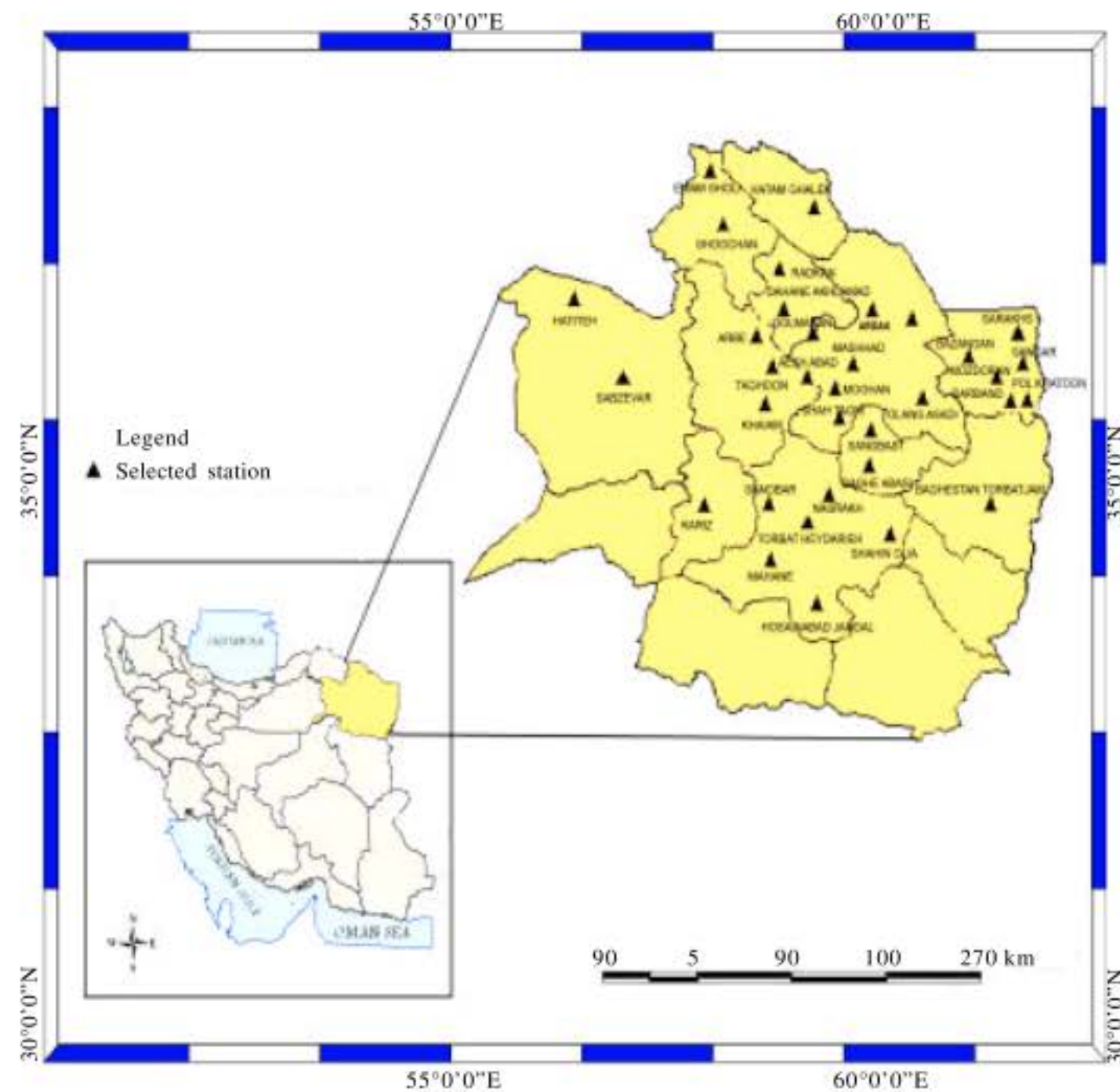


Fig. 1: Map of the region under study and selected stations

getting necessary information for the long-term analysis of rainfall in every region in the area under study. As mentioned before, kriging method used in this study to calculate the average regional rainfall. The following steps were taken to obtain the time series of average regional rainfall:

- Making input files for the Arcmap software
- Obtaining the experimental variogram
- Analyzing and drawing annual spatial changes of rainfall in the region
- Obtaining the values of annual average rainfall in the region under study
- Making time series of rainfall in the region under study

The Relation Between Index Signals and Rainfall

The signals studied in this research are meteorological parameters at upper levels of the atmosphere include:

- Air temperature at 700 mb level
- The thickness between 500 and 1000 mb levels
- Relative humidity of 300 mb level

Determining Seasonal Rainfall and Signals

Seasonal rainfall and signals has been determined using the average of the values of a signal in several successive months in order to predict the amount of seasonal rainfall in the following months. Attempts have been made to make sure that the seasons of the signal do not include months with rainfalls.

Since, we aimed to investigate the relation between meteorological parameters and spring rainfall in this study, the average value of meteorological signals in the period between October and March have used as the time series of the signals and the average rainfall in the period between April and June as the time series of rainfall.

To analyze the parameters of the upper levels of the atmosphere in the present study, two networks of 5×5 and 10×10 degrees have used. The region under study, where meteorological parameters at ground level and upper levels of atmosphere have been factor-analyzed, is located between 0-80 eastern degrees and 10-50 northern degrees in networks of 5×5 degrees and between 0-100 eastern degrees and 0-75 northern degrees in networks of 10×10 degrees. The area includes regions where changes in the pattern of temperature, pressure, humidity and wind speed affect rainfall (Alijani, 2001).

To do the relevant statistical tests and obtain correlation between index signals and regional rainfall, the soft wares EXCELL and JUMP 4 have used.

Research Methodology

We have used factor analysis to analyze the meteorological parameters at upper atmosphere In this statistical method designed to reduce the number of variables, the initial parameters are transformed into independent variables based on their correlation coefficients. These independent variables are called factors. The value of each of the observations in the new factors is calculated as factorial score. Hence, rather than true values of observations, their scores in new components are used as new criteria for clustering. The advantage of this method is that while it reduces the number of variables, it preserves the initial variance of the main data (Alijani, 2001).

The Structure of Artificial Neural Networks

Artificial Neural networks were first introduced in 1943 by McCulloch and Pitts. Later, with the development of back propagation algorithm for feed forward networks, the application of neural network entered a new stage (Mahdizadeh, 2004).

Like natural neural networks, artificial neural networks are made up of parts called neural cells. As in natural neural networks where some cells are responsible for receiving the external stimulus, some for processing and some for the transfer of response to the intended part, in artificial neural networks, too, some cells receive the data of the problem, some process the data and some provide the solution to the problem. Thus, every neural network is made up of the input layer, the hidden layer and the output layer, with the three layers connected by means of connectors of different weights. In all neural networks, there is one input layer, one output layer and several hidden layers (Mahdizadeh, 2004). Figure 2 shows the structure of one kind of such networks (Mohammadi and Misaghi, 2003).

Figure 3 shows the model of a multi-input neuron (Christodoulou and Georgiopoulos, 2001). The three elements of a multi-input neuron are the following:

- The set of synapses each specified with a certain weight. As it is shown in the Fig. 3, the neuron K with the output x_k is connected to the intended neuron j through a proper weight connector called w_{jk} . The effect of the neuron k on the neuron j is calculated through $x_k \cdot w_{jk}$. If the neuron K is active and w_{jk} is positive (excitatory synapse), the neuron K will have a positive effect on the neuron j. On the other hand, if the neuron k is active, but w_{jk} is negative (inhibitory synapse), the

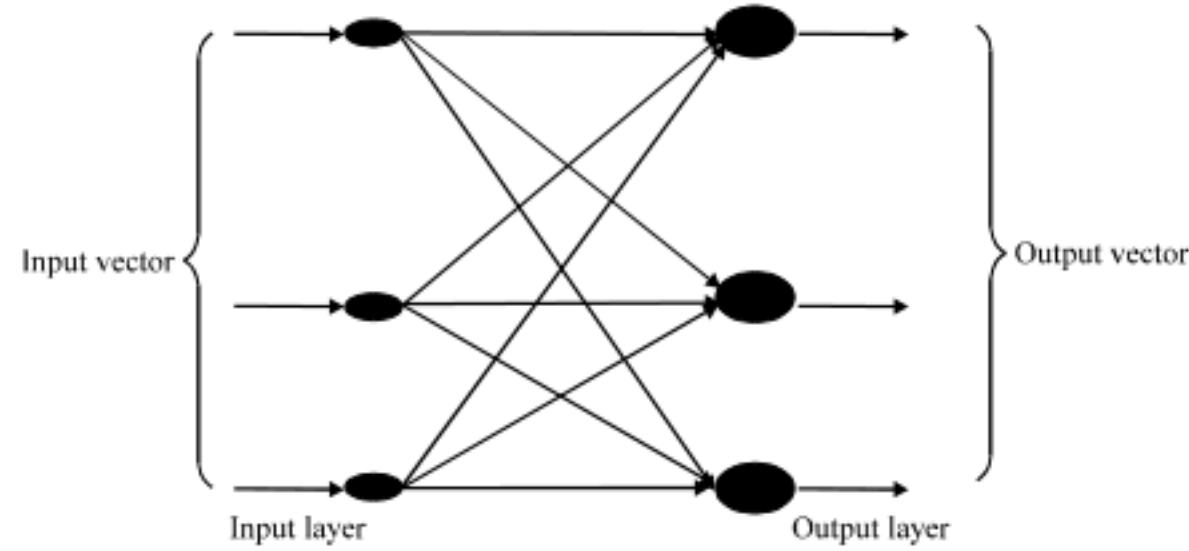


Fig. 2: The overall structure of feed forward monolayer neural networks

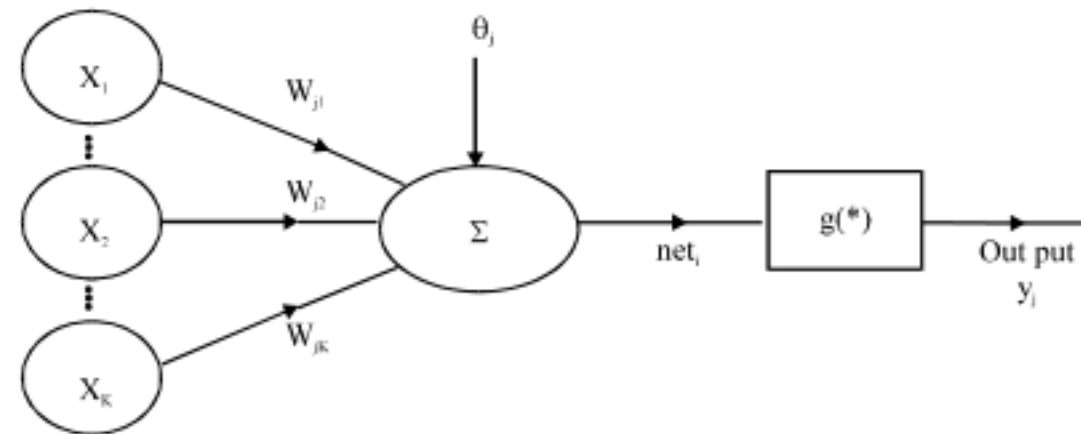


Fig. 3: A model of a multi-input neuron

neuron K will have a negative effect on the neuron j. Special attention should be paid to the written form and the subtitle of the weight of the synapse w_{jk} . The first subtitle belongs to the target neuron and the second to the source neuron of the intended synapse

- A capacitor for collecting the incoming signals which are weighted by the synapses of the neuron. The accumulating effect of all neurons which are connected to the intended neuron (neuron j) is calculated by adding up all the effects of individual neurons on the neuron j
- A function of activity is used to limit the output range of the neuron. Activity function is considered as a constraining function in which the eligible changes of the range of output signals is restricted to some finite values

The net input and the output y are calculated by Eq. 1 and 2:

$$\text{net } j = \sum_{k=0}^K w_{jk} x_k + b \tag{1}$$

$$y_i = g(\text{net}_j) \tag{2}$$

In the formulae x_1, x_2, \dots, x_k represent the incoming signals, $w_{j1}, w_{j2}, \dots, w_{jk}$ stand for synaptic weights accumulating in a neuron, net is the accumulated effect of all the neurons connected to the neuron j and the internal threshold of the neuron j.

G is activity function and y_j is the output signal of the neuron.

To assess the accuracy of the model, the index of Root Mean Square Error (RMSE) has used which is calculated by the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (o_i - e_i)^2}{n}} \quad (3)$$

In the Eq. 3, RMSE is root mean square error, o_i and e_i are the observed and predicted value of the variable respectively in the point I and n number of observations.

RESULTS AND DISCUSSION

Results concerning meteorological parameters changes at upper levels of the atmosphere and changes in spring rainfall.

In this study, we used factor analysis to detect index regions for each of the factors in the time-span October to March. We, then, calculated the correlation between changes in the rainfall in April to June next year and temperature changes at 700 mb level, the thickness between 500 and 1000 mb level and relative humidity changes at 300 mb level in index regions.

For the time-span October to March, the time series of the data was prepared in the network of 5×5 and 10×10 grade in the statistical period 1970 to 2007. SPSS software was, then, used to factor-analyze the data. By variables in this process, we mean the values of the time series of the data of the temperature at 700 mb level, the thickness between 500 and 1000 mb level and relative humidity at 300 mb level in each of the grades of the network. The result of the factor analysis of those factors covering a considerable portion of the data was reviewed. Each factor, also, includes variables that have a correlation of 0.6 with the factor. Finally, in the aforementioned time span, some regions, called index regions, were detected for each of the extracted factors. In order to study the relationship between the precipitation in the next season (April to June) and the data of the temperature at 700 mb, the thickness between 500 and 1000 mb level and relative humidity at 300 mb level by means of the results of factor analysis method, we, first, calculated the factorial scores of the aforementioned data in the specified index regions in every year and then draw its changes with the rainfall in the next season. In this regard, the correlation coefficient between factorial scores in selected index regions and the rainfall in the next season at 5% level was subjected to statistical analysis.

Table 1 shows the observations about the relation between changes in parameters of upper atmosphere and the spring precipitation of the region under study. As we can shown in Table 1, for the temperature of 700 mb level in networks of 5×5 grade, the index regions of factors 2 and 3 and for networks of 10×10 grade, the index regions of factor 5 show strong correlation with rainfall in the region under study. These regions can be used in rainfall forecast models. These regions have shown in Fig. 4 and 5. As far as the thickness between 500 and 1000 mb levels is concerned, in networks of 5×5 grade, the index regions of factor 1 and for networks of 10×10 grade, the index factors of 1 and 7 manifest strong correlation with the precipitation in the region under study and can be used in rainfall forecast models. These regions have shown in Fig. 6 and 7. As for the relative humidity of 300 mb

Table 1: Summary of the observation of the effects of the parameters of upper atmosphere in the index regions selected on the basis of the rainfall of the region under study

Variable name	Season	Factor	Used network
Temperature at 700 hpa	October to March	Factor 2	5×5
		Factor 3	
		Factor 5	10×10
The thickness between 500 and 1000 hPa	October to March	Factor 1	5×5
		Factor 1	10×10
		Factor 7	
Relative humidity at 300 hpa	October to March	Factor 2	5×5
		Factor 4	
		Factor 2	10×10
		Factor 4	

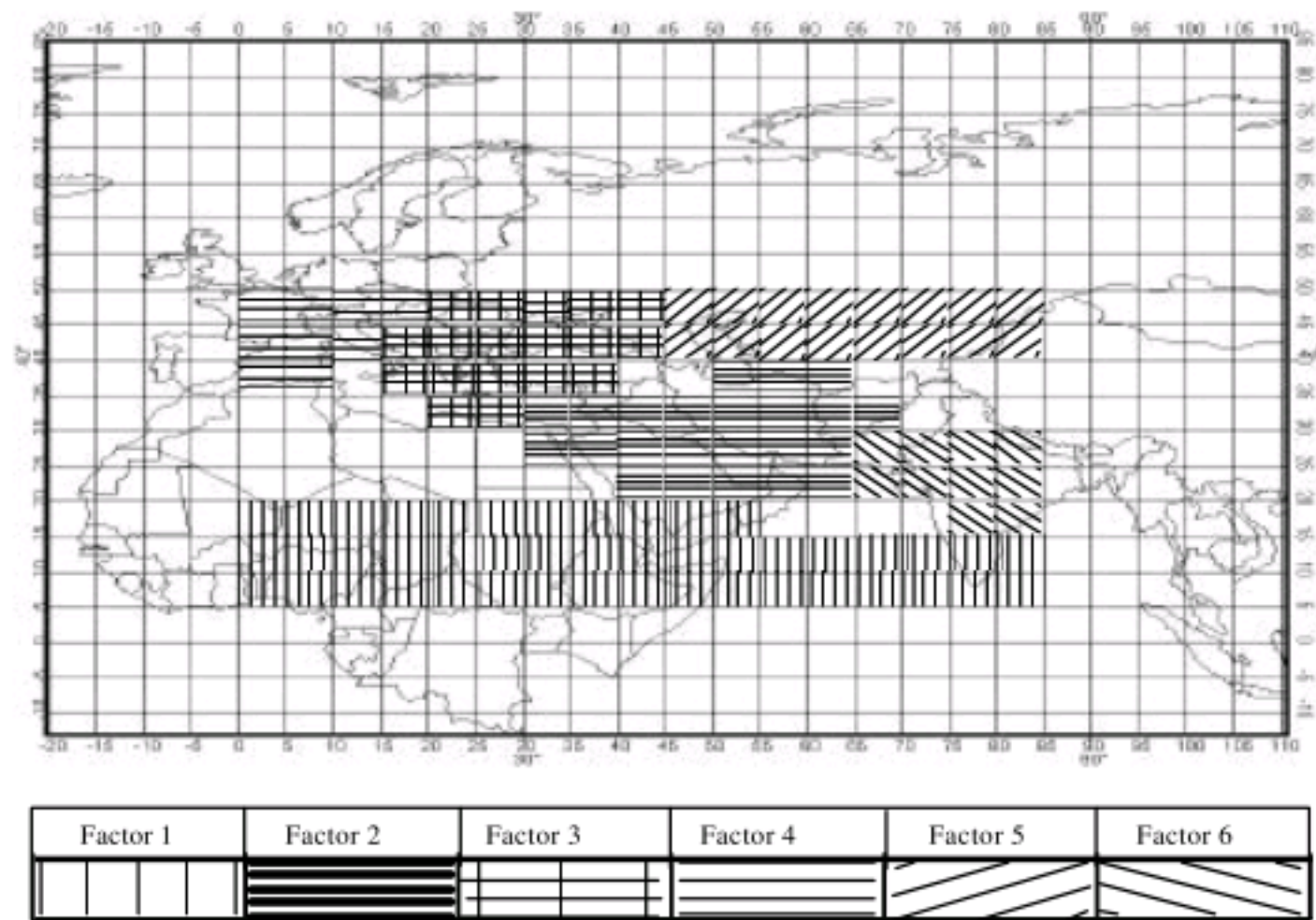


Fig. 4: The index areas detected by factor analysis at the temperature of 700 mb level in the period October to March in networks of 5x5 degree

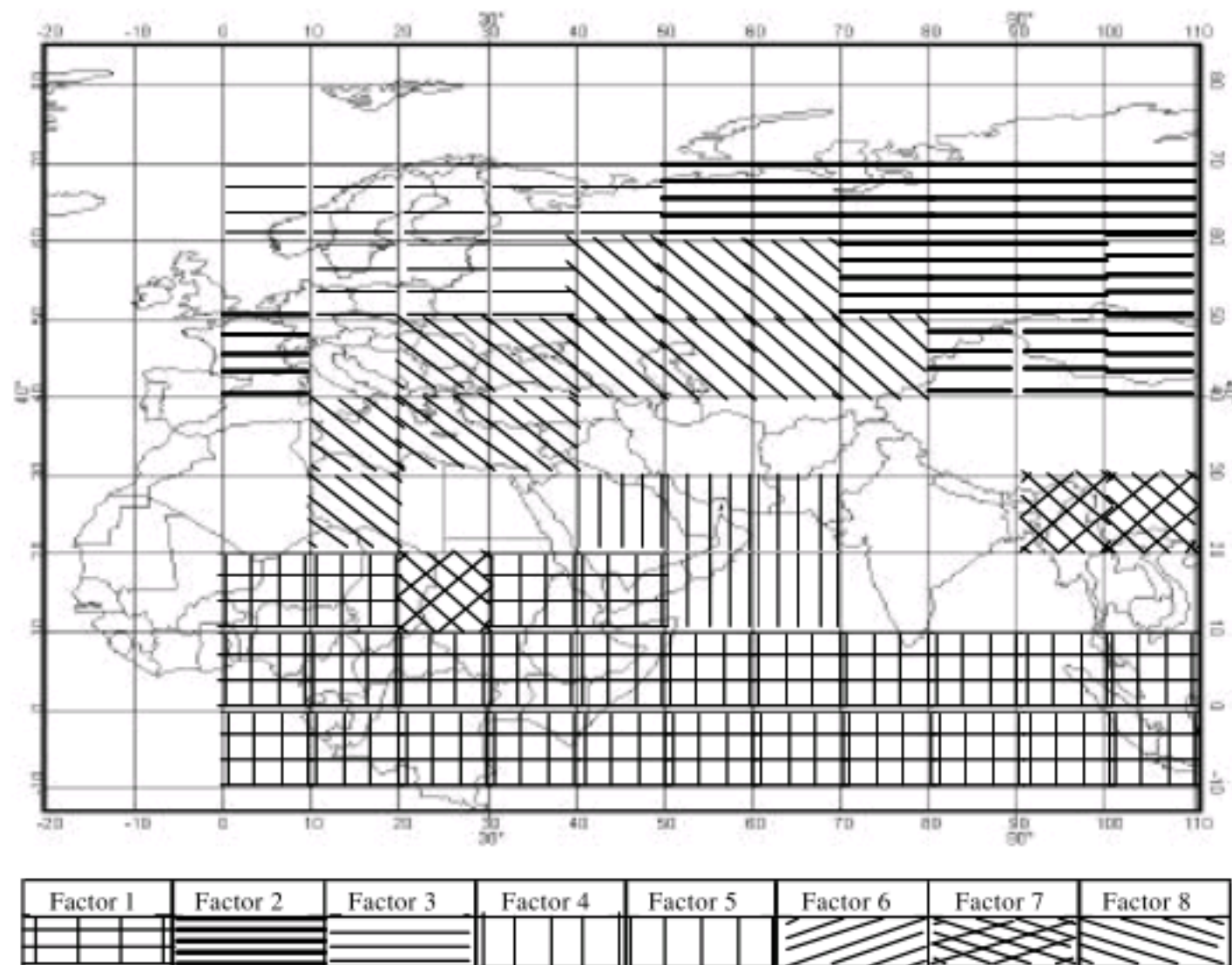


Fig. 5: The index areas detected by factor analysis at the temperature of 700 mb level in the period October to March in networks of 10x10 degree

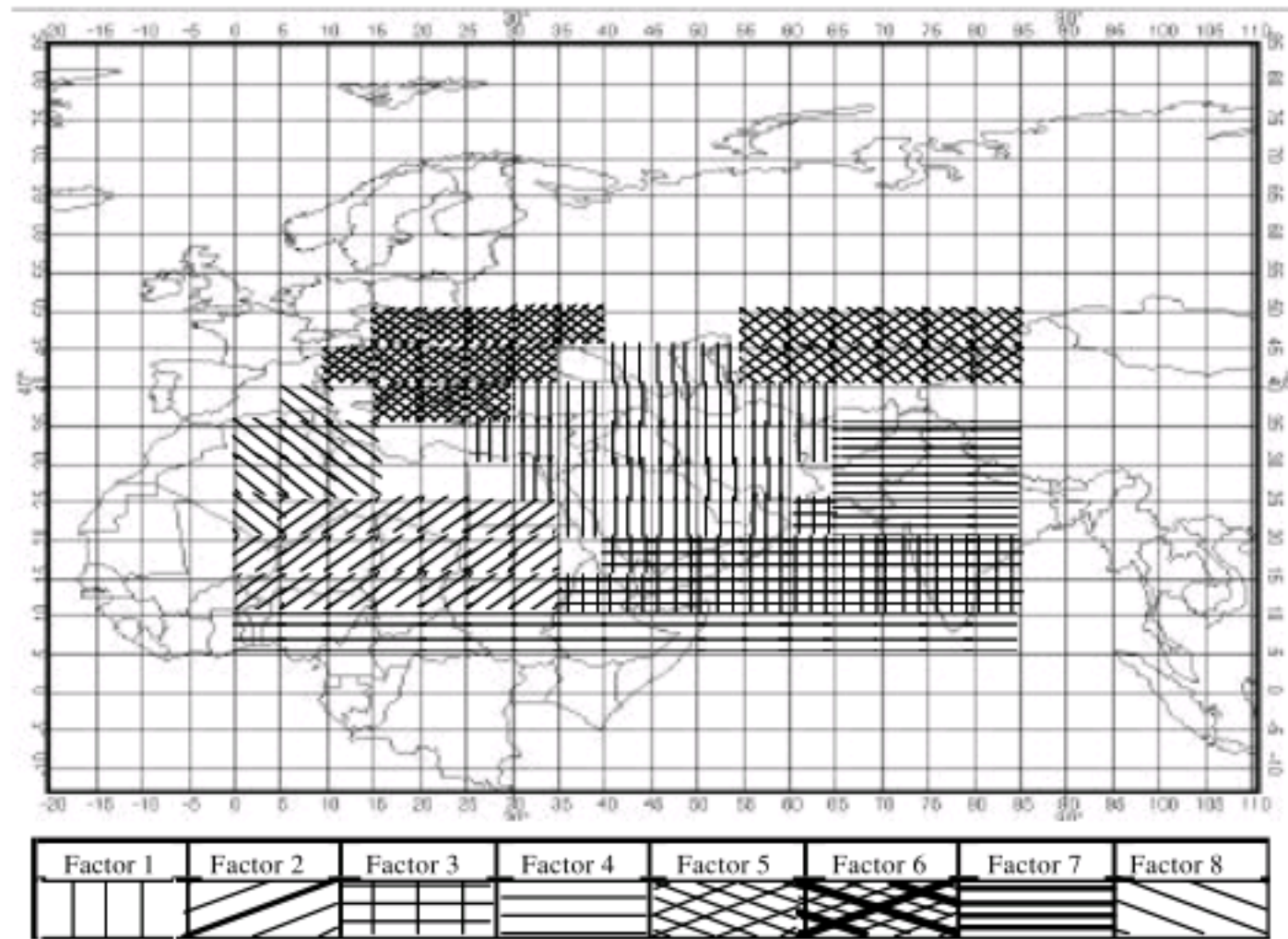


Fig. 6: The index areas detected by factor analysis at the thickness between 500 and 1000 mb level in the period October to March in networks of 5x5 degree

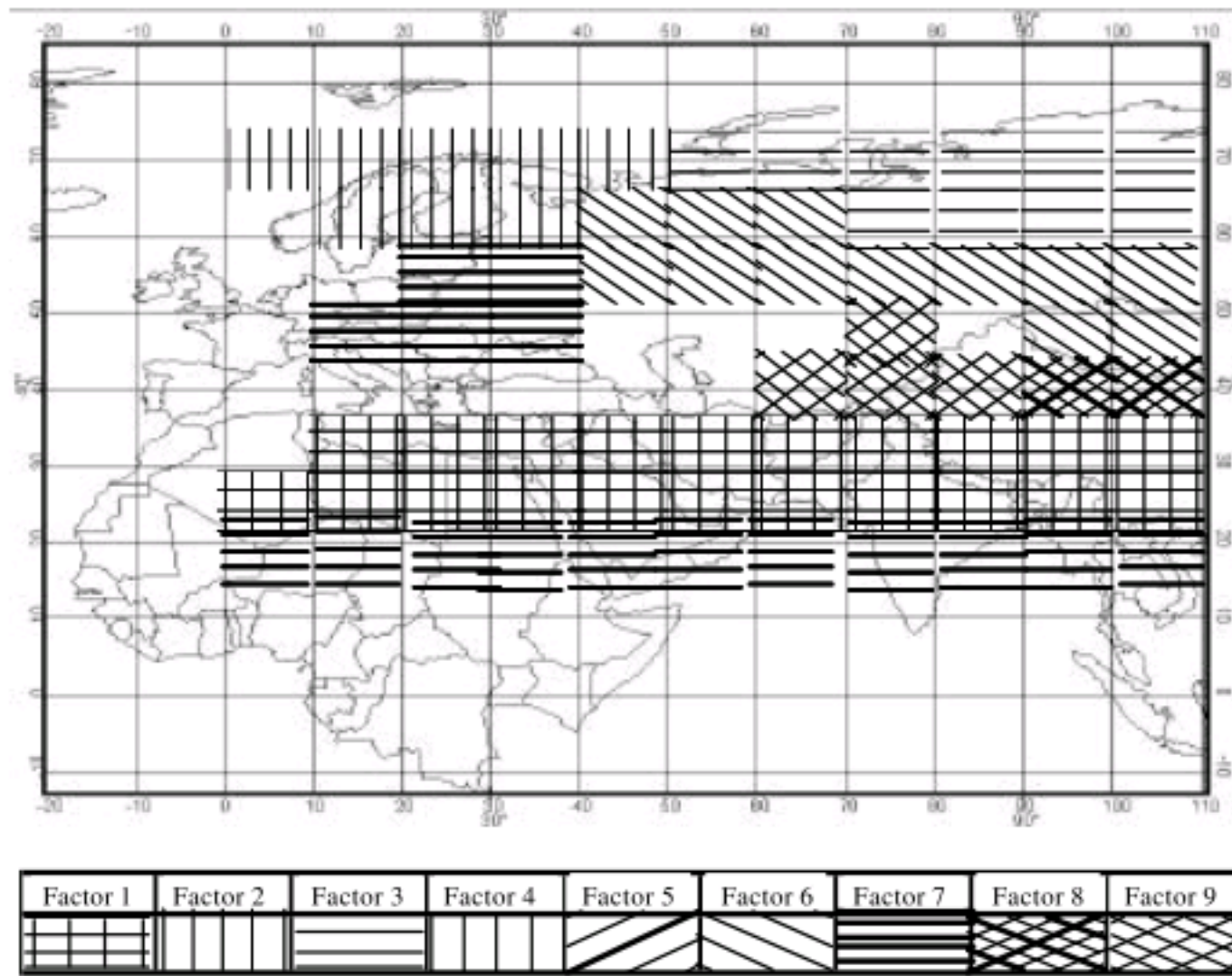


Fig. 7: The index areas detected by factor analysis at the thickness between 500 and 1000 mb level in the period October to March in networks of 10x10 degree

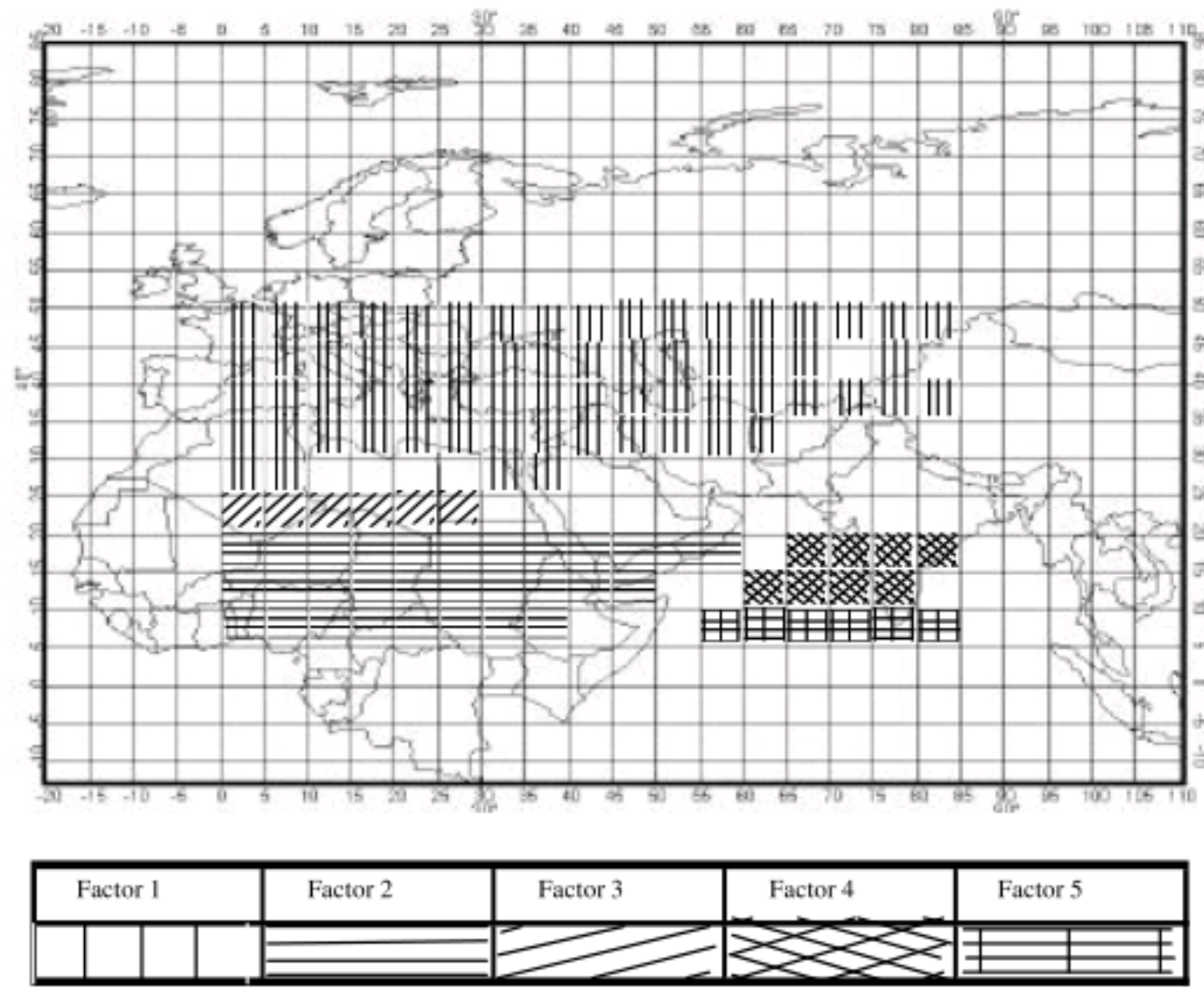


Fig. 8: The detected areas of relative humidity at 300 mb level in networks of 5×5 degree

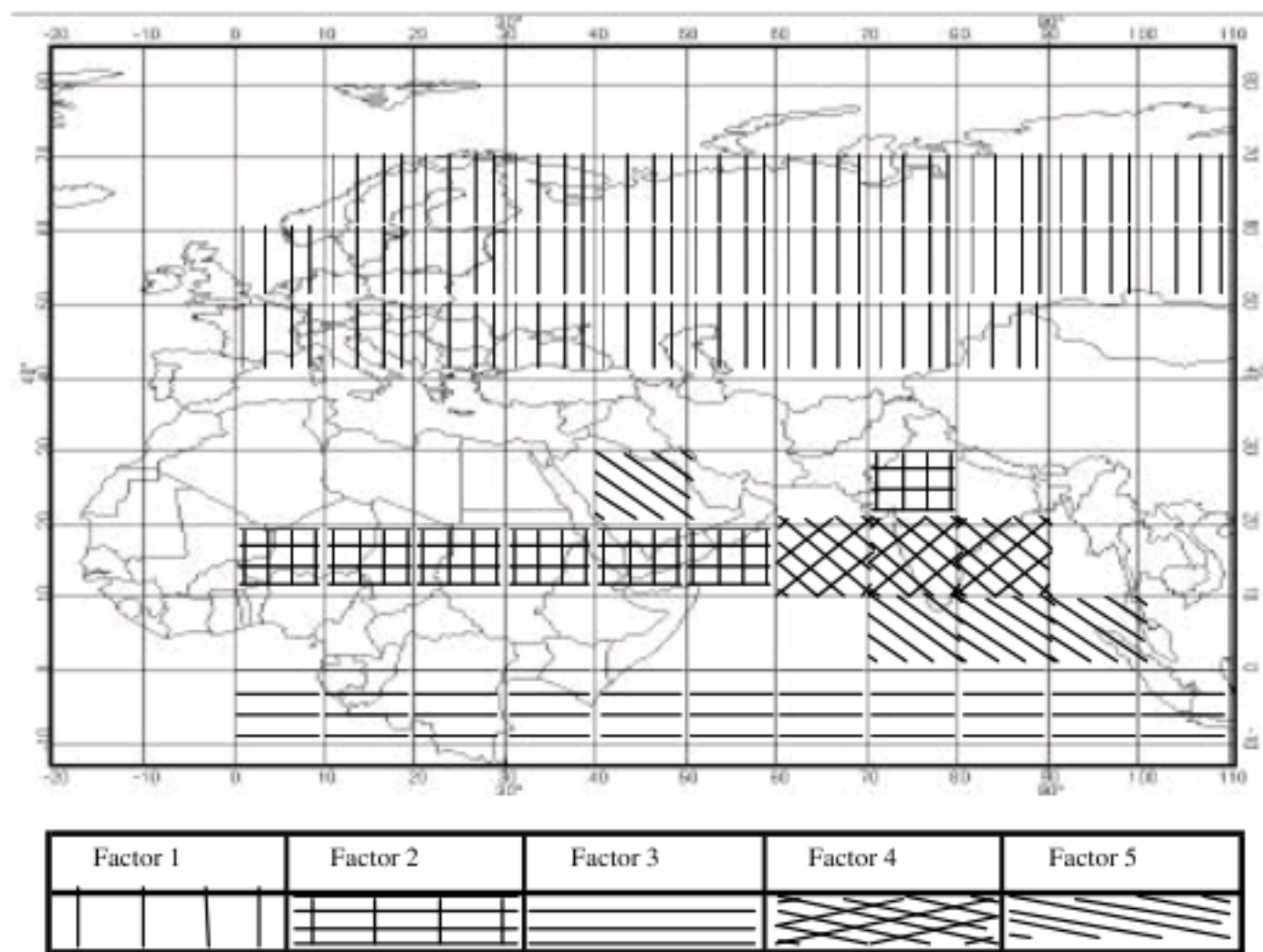


Fig. 9: The detected areas of relative humidity at 300 mb level in networks of 10×10 degree

level, in networks of 5×5 grade the index region of factors 2 and 4 and in networks of 10×10 grade, too, the index region of factors 2 and 4 have strong correlation with the precipitation in the region under study and can be used in rainfall forecast models as well as in the prediction of dry and wet periods in the region being studied. These regions have shown in Fig. 8 and 9. From the above discussion, we can infer that the index signals specified in this research have been able to justify the dispersion and distribution patterns of spring precipitation in the years under study. These signals can be successfully used as predictor together with other climatic parameters to predict above normal, below normal and normal periods in the region. A comparison of the results obtained in this study with those of other researchers such as Nazemosadat (2001) indicate a significant relationship between such signals as temperature, humidity and thickness and the rainfall in the northeast of the country. It is concluded that we can use these signals to predict the rainfall in Northeastern Iran. For example, according to the results obtained by Nazemosadat (2001), autumn rainfall in Southern and Southwestern Iran is in inverse proportion to SST of the Persian Gulf. Likewise, the results of Fallah Ghalhary *et al.* (2007) and Mousavi *et al.* (2008) indicate a significant correlation between signals of pressure and temperature and annual precipitation (December to May) in Grand Khorasan Region including three provinces: Northern, Razavi and Eastern Khorasans. The findings of this study, too, indicate a significant relationship between such signals as temperature of 700 mb level, relative humidity of 300 mb level and the thickness between 500 and 1000 mb levels and the rainfall in the northeast of the country. Therefore, with reference to the results of this study, we can use these climatic signals in rainfall forecast models as well as in predicting dry and wet periods in the region.

Predicting Spring Rainfall by Means of Artificial Neural Networks

In this study, Pearson correlation method has been used to obtain meteorological signals which affect regional rainfall. Thus, all the signals which have shown a correlation with 5% level of significance in the period between October and March have been used as predictors in the structure of the rainfall forecast model. After numerous checking, it became clear that the optimum effect of signals is when the period between October and March is used. Therefore, the mentioned signals in Table 1 in the period between October and March were used as predictors in rainfall forecast models.

The above model divides the data into three different sections, namely, training data, validation data and testing data. The data belonging to 38 years were in turn divided into 19 years of training data, 9 years of validation data and 10 years of testing data. In other words, from the whole set of historical data, two-thirds were considered as calibration data and one-third as testing data.

After conducting various tests to test the network and the number of neurons of the hidden layer and different functions of activity in the hidden and output layers, eventually found out that the final model with its one input layer, one hidden layer and one output layer (average spring rainfall), had the least error, so in this research, used it as the main model. The numbers of the neurons in the input, hidden and output layers is ten, four and one, respectively (10-4-1). The hidden layer activity function is a function of the hyperbolic tangent and the output layer activity function is a function of the linear sigmoid.

Figure 10 and Table 2 present the results of the calibration period of the rainfall forecast model. As is shown, the minimum mean-square error in the repetition of 999 is 0.056. The maximum mean-square error is again 0.056. In other words, at this stage in the repetition of 999, the network shows the maximum error. Also, the minimum mean-square error of the validation epoch in the repetition of 310 is 0.154. The results of the prediction model are shown in Table 3 and Fig. 11. It is to be noted that these results were presented for the testing epoch of the model.

As Table 4 shows, the mean square error is 25.34 and the normalized mean square error is 0.46. Also, the mean absolute error for this model was calculated to be 3.75 mm. The minimum absolute error is 0.73 mm and the maximum absolute error is 11.56 mm. Also, the correlation coefficient between

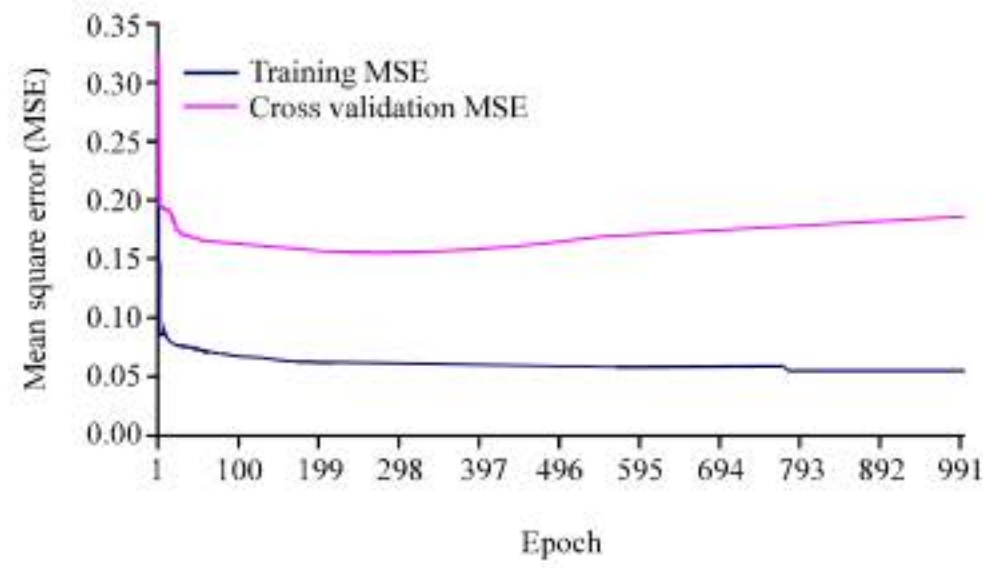


Fig. 10: The diagram of mean-square error in the training and validation epochs in different repetitions. In the figure, the upper curve is the MSE of the validation Epoch

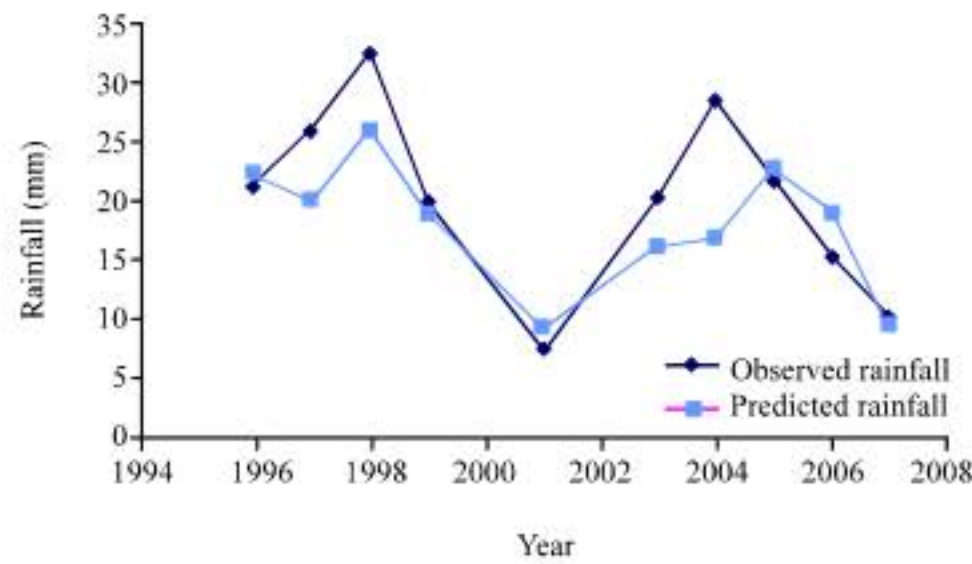


Fig. 11: Comparison of the observed and predicted rainfall in the region under study by means of artificial neural network model

Table 2: Minimum and maximum error in training and validation Epochs

Best networks	Training	Cross validation
Epoch	999.000	310.000
Minimum mean squared error	0.056	0.154
Final mean squared error	0.056	0.186

Table 3: Rainfall prediction in the region under study by means of neural network model

Year	Observed rainfall	Predicted rainfall
1996	21	22.0
1997	25	19.8
1998	32	25.5
1999	19	18.6
2001	7	9.0
2002	20	16.0
2004	28	16.8
2005	21	22.5
2006	15	18.7
2007	10	9.3

recorded and predicted rainfall for the model is 0.8. The root mean square error for this model was calculated to be 5 mm which indicates the high efficiency of the model in predicting rainfall in the area under study.

In sum, the analysis of the results of the model indicates that the difference between the observed and predicted rainfall is within a reasonable range and that the model has been able to predict rainfall

Table 4: The features of artificial neural network model

Performance	Values
Mean squared error	25.34
Normalized mean squared error (MSE/variance desired output)	0.46
Mean absolute error	3.75
Minimum absolute error	0.73
Maximum absolute error	11.56
Linear correlation coefficient	0.80

in all years with an acceptable error. The root mean-square error for this model was 5 mm, which is very small, indicating the accuracy of the model in predicting rainfall. We can realize from the above discussion that the variables used in the model have been able to detect the distribution pattern of rainfall in the region with great ease and accuracy and that the model can be successfully used for predicting rainfall in spring. This plays a vital role in the management and planning of water resources for drinking and agriculture. Considering these predictions, we can design the future policies with the aim of optimal spending of the budget and maximum exploitation of the resources.

ACKNOWLEDGMENTS

This study presents part of the findings of the research project predicting spring rainfall in Khorasan Razavi Province based on meteorological signals by means of fuzzy logic, artificial neural networks and adaptive neuron-fuzzy networks by the same author. The author greatly appreciates the cooperation of the directors and managers of the Climatology Research Center who provided the possibilities required for completing this project.

REFERENCES

- Alijani, B., 2001. Synoptic Climatology. 1st Edn., SAMT Publication, USA., ISBN: 964-459-609-9.
- Ashby, S.A., M.A. Taylor and A. Chen, 2005. Statistical models for predicting rainfall in the Caribbean. *J. Theor. Applied Climatol.*, 82: 65-80.
- Astine, O.N., 2001. Forecasting seasonal rainfall for agricultural decision-making in Northern Nigeria. *Agric. For. Meteorol.*, 107: 193-205.
- Christodoulou, C. and M. Georgiopoulos, 2001. Applications of Neural Networks in Electromagnetics (Artech House Antennas and Propagation Library). Artech House Publishers, USA., ISBN-13: 978-0890068809, pp: 512.
- Fallah Ghalhary, G.A., M.M. Baygi and M.H. Nokhandan, 2007. Seasonal rainfall forecasting based on synoptic patterns of sea level pressure and sea level pressure gradient by means of statistical models. *J. Agric. Sci. Technol.*, 21: 95-101.
- George, W. and J. Kimber, 2007. A statistical model for predicting average rainfall in the state of Florida. Proceedings of the 5th International Conference on Dynamic Systems and Applications, May 30, Morehouse College, Atlanta, Georgia, USA., pp: 1-1.
- Gissila, T., E. Black, D.I.F. Grimes and J.M. Slingo, 2004. Seasonal forecasting of the Ethiopian summer rains. *Int. J. Climatol.*, 24: 1345-1358.
- Haltiner, G.J. and R.T. Williams, 1980. Numerical Prediction and Dynamic Meteorology. 2nd Edn., Wiley and Sons, New York.
- Kim, W.T., C. Yoo and A.J. Hyun, 2007. Influence of climate variation on seasonal precipitation in the Colorado River Basin. *J. Stochastic Environ. Res. Risk Asses.*, 22: 411-420.
- Mahdizadeh, M.B., 2004. Artificial Neural Networks and Their Application in Civil Engineering. 1st Edn., Ebady Publication, Iran, ISBN: 964-6531-35-0, pp: 192 (In Farsi).

- Maria, C., F. Harold and N. Ferreira, 2005. Artificial neural network technique for rainfall forecasting applied to the São Paulo region. *J. Hydrol.*, 301: 146-162.
- Mohammadi, K. and F. Misaghi, 2003. *Artificial Neural Networks*. 1st Edn., Tarbiat Modarres University Press, Iran.
- Mooley, D.A. and D.A. Paolino, 2006. Relationship of the Indian monsoon rainfall to the northern hemispheric 700 mb height tendency. *Int. J. Climatol.*, 8: 499-509.
- Mousavi, B.M., G.A.F. Ghalhari and M.H. Nokhandan, 2008. Assessment of the relation between the large scale climatic signals with rainfall in the Khorassan. *J. Agric. Sci. Natur. Resour.*, 15: 217-224.
- Nazemosadat, M.J. and I. Cordery, 2000. On the relationship between ENSO and autumn rainfall in Iran. *J. Climate.*, 1: 47-62.
- Nazemosadat, M.J., 2001. *Will it Rain? Drought and Rainfall in Iran and their Relation with ENSO*. 1st Edn., Shiraz University Press, Iran, ISBN: 964-455-621-6.
- Prasad, K.D. and S.V. Singh, 2006. Exploring the possibility of forecasting monthly-700 hPa geopotential fields over India. *Int. J. Climatol.*, 14: 371-378.
- Sen, N., 2003. New forecast models for Indian Southwest monsoon season rainfall. *Curr. Sci. India*, 84: 1290-1292.
- Singhrattna, N., B. Rajagopalan, M. Clark and K. Kumar, 2005. Seasonal forecasting of thailand summer monsoon rainfall. *Int. J. Climatol.*, 25: 649-664.