15th Iranian Hydraulic Conference, 14-15 December 2016
Faculty of Engineering and Technology, Imam Khomeini International University,
Qazvin, Iran

The Time Series Prediction of Meteorological Parameters in the Arid and Semi-Arid Region

ShimaKabiri[1], Masoumeh Alsadat Hashemi Tameh[2*], Reza Ekhlsmand[3]

1- Water Structures, MS. Qazvin Regional Water Authority, Qazvin, Iran
2- Irrigation and Drainage Engineering, Ph.D. Candidate. Faculty of Engineering and Technology, Imam Khomeini International University, Qazvin, Iran
3-Civil Engineering, Ph.D. Candidate. Qazvin Regional Water Authority, Qazvin, Iran

*Hashemitame@yahoo.com

Abstract

Time series prediction of meteorological parameters plays an important role in making decision to decrease the effects of drought and climate change. Temperature is an important factor in planning and making decision of water resources and water balancing. Therefore, precise estimation of temperature is indispensable for every computation in hydrology and other disciplines. There are a lot of methods for estimating time series climate data and all of them can be grouped in (A) statistical methods and (B) intelligent methods. In this study, Artificial Neural Network (ANN) and ANFIS were applied as intelligent methods to estimate the maximum and minimum temperature. The data was spited into two parts (A) 90% data that was used as training and (B) the rest of the data set which was applied as the test set to validate the constructed model. The performance of Multi Layer Perceptron and Neuro-Fuzzy Inference System with the fuzzy c-means clustering (FCM-ANFIS) were investigated using different numbers of neurons in hidden layers and different number of clustering , respectively. Accuracies of the models were evaluated using indices such as R^2, RMSE and MAE.

Keywords: Time series, Artificial neural network, Maximum temperature, Minimum temperature.

15Hydro-2447

1. INTRODUCTION

The accuracy of forecasting climatology data is very important due to the current world climate change. On the other hand, time series modeling is a major tool in planning, operating and decision making of water resources and investigating climatic fluctuations and has been commonly used for data generation, forecasting, estimating missing data and extending hydrologic data records [1, 2, 3].

Therefore, in order to produce efficient accurate results of forecasting weather, a few methods have been developed. Among them, a statistical model has been widely used to make forecasts of climate data. This model works by lowering the equation of the data itself (data driven) such as a simple method regression analysis (SRA), decomposition, exponential smoothing method (ES) and autoregressive integrated moving average (ARIMA). In recent years, several studies have reported that the ANN with its ability to model non-linear relationships may offer a promising alternative for water resource modeling [4, 5, 6, 7]. However, in some cases, rainfall prediction statistical method is also able to produce good and accurate predictions [8].

Along with the development of computing technology, many researchers are trying to make predictions using the ANN method in the field of hydrology. Mislan *et al.*, (2015) applied an Artificial Neural Network (ANN) with the Backpropagation algorithm (BPNN) for forecasting time series of rainfall monthly. The results showed that, BPNN had high accuracy in making predictions of monthly rainfall in the Tenggarong, East Kalimantan – Indonesia [9].

Abhishek et al., (2012) have conducted research with ANN rainfall prediction in Udupi district of Karnataka, India. Other researchers also have predicted the data series of rainfall for 30 years (1977-2006) at the station of Nagpur, India. The results of this research have revealed that by using a Backpropagation in neural networks (BPNN), accurate prediction could be obtained [10]. Kumar Nanda et al., (2013) compared the different ANN models for time series analysis (monthly rainfall), it was found that FLANN gives very close and better prediction result as compared to the ARIMA model with less absolute average percentage error[7]. Kumar and Kumar Jha (2013) concluded the Multilayered Neural Network can be an effective tool

in weather prediction and this type of Network can correctly provide the mapping between the input and the output using historical data [11].
In this study, the ANN and ANFIS were applied to forecasting time series of monthly maximum and minimum temperature in arid and semi-arid region with different combination of their previous values.

2. Research Method and Process

2-1 Artificial Neural Network

The Artificial Neural Network is an engineering concept of knowledge in the field of artificial intelligence designed by adopting the human nervous system [9]. The ANN is a mathematical model that tries to map input data to output data with less error. In the study Multi Layer Perceptron Artificial Neural Network (MLP-ANN) is used for weather prediction.

In general, MLP architecture is arranged in three layers, starting from a first input layer and ending at output layer and among hidden layer between two layers (Fig.1).

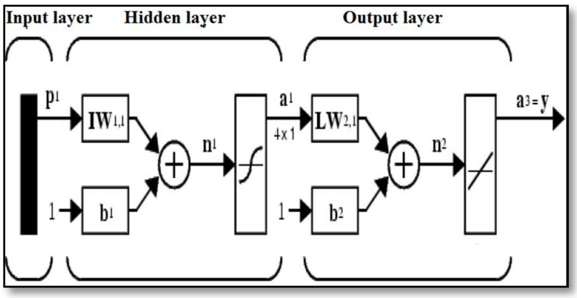

Figure1. Multi perceptron layer

Formally, a one-hidden-layer MLP is a function $f: R^I \rightarrow R^O$, where I is the size of input vector x and O is the size of the output vector H(x) that the mathematical relation is shown in the equation 1.

$$H(x) = G(b^{(2)} + w^{(2)}(F(b^{(1)} + Iw^{(1)}x))) \quad (1)$$

Where, $b^{(1)}\&b^{(2)}$ are the bias vectors, $Iw^{(1)}\&w^{(2)}$ are the weight matrices, G and F are the transfer functions and H(x) is the output vector.

In this experiment, the data was divided into two different groups: training data set and testing data set. The 90% of data were used for training and the rest of them for testing. For getting accurate prediction results the ANN was tested with two hidden layers. The neurons of first hidden layer vary from 4 to 15 and the second hidden layer varies from 5 to 30. Additionally, five transfer functions (Hyperbolic tangent, Log-sigmoid, Hyperbolic tangent sigmoid, Symmetric saturating linear, Linear) for each layer including the hidden layers and the output layer were investigated and the Levenberg-Marqardt was used in learning algorithm. The data with different delay time was selected for input data and the amount of current data was considered for the output data. The best function and the neurons of the hidden layer were selected by a process of trial and error based on the statistical criteria (R2, RMSE and MAE).

2-2 Neuro-fuzzy approach (ANFIS)

ANFIS uses the learning ability of the ANN to define the input-output relationship and construct the fuzzy rules by determining the input structure. The system results were obtained by the thinking and reasoning capability of the fuzzy logic. The detailed algorithm and mathematical background of the model can be found in [12]. There are two types of fuzzy inference system: the Sugeno-Taka. In this study, the Sugeno-Takagi inference system is used for modeling of weather time-series.

The fuzzy system contains three main parts: fuzzification, inference and defizzification. For simplicity, a fuzzy inference system has two inputs x and y and one output is assumed. The fuzzy system was defined with a first-order Sugeno and a common rule set. The rules are as follows:

Rule 1: IF x is A_1 is B_1 THEN $f_1 = p_1x + q_1y + r_1$ (2)

Rule 2: IF x A_2 and is B_2 THEN $f_2 = p_2x + q_2y + r_2$ (3)

Where, x and y are the crisp inputs to the node i, A_i and B_i are the linguistic labels such as low, medium, high, etc., which are characterized by convenient membership functions p_i, q_i and r_i are the consequence parameters. The mechanism of this fuzzy inference system is shown in Figure 2.

According figure 2, the ANFIS has five main layers that they have been briefly explained as follows:
Layer1 (*Input nodes)*: Each node in this layer generates membership grades of the crisp inputs which belong to each of convenient fuzzy sets by using the membership functions. Each node's output O_1^i is computed by:

$$O_1^i = \mu_{A_i}(x) \text{ for } i = 1,2 \quad (4)$$

$$O_1^i = \mu_{B_{i2}}(y) \text{ for } i = 3,4 \quad (5)$$

Where, μ_{A_i} and $\mu_{B_{i2}}$ are the ember function for Ai and Bi respectively. Different membership functions, such as triangular, trapezoidal, Gaussian generalized bell membership function, etc., can be used to determine the membership grades.

Layer2 *(Rule nodes)*: In this layer, the AND/OR operator is applied to produce one output of all the incoming signals. Each node output represents the results of the antecedent for a fuzzy rule, i.e. firing strength. The outputs of the second layer, called firing strengths O_1^i, are the products of the corresponding degrees obtained from the first layer as follows:

$$O_2^i = w_i = \mu_{B_i}(y) \times \mu_{A_i}(x) \qquad \text{for } i=1, \quad (6)$$

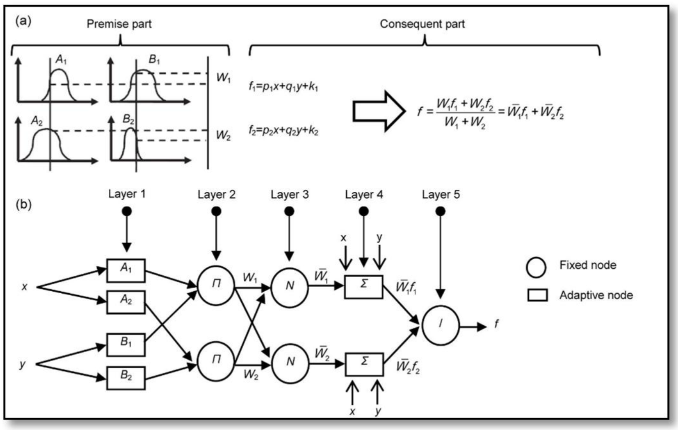

Figure 2. (a) Sugeno's fuzzy if then rule and fuzzy reasoning mechanism; (b) equivalent ANFIS architecture.

Layer3 (*Average nodes)*: The main target is to compute the ratio of the firing strength of each *i*th rule to the sum of the firing strength of all rules. The outputs of this layer are called normalized firing strengths that are calculated as follows:

$$O_3^i = \bar{W}_i = \frac{w_i}{\sum_{i=1}^{n} w_i} \quad (7)$$

Layer4 (*Consequent nodes)*: Every node in this layer is an adaptive node that is calculated by:

$$O_4^i = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \qquad \text{for } i=1,2 \quad (8)$$

Where, $\overline{w}_i$ is the ith node output from the previous layer as demonstrated in the third layer. The q_i, p_i and r_i is the parameter set in the consequence function and also the coefficients of linear combination in the Sugeno inference system.

Layer5 (*Output nodes)*: This layer is called the output node, in which a single node computes the overall output by summing all input signals. The output of the system is computed by:

$$O_5^i = f(x, y) = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{9}$$

In this study, the ANFIS was generated using fuzzy c-means (FCM) clustering. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters**.** The divided data in last step was used for finding best FIS.

For getting accurate prediction results the ANFIS was tested with the different number of clusters to be generated by FCM. The number of clusters varies from 4 to 22. The data with different delay time was used for input data and the amount of the current month was considered for the output data. The best Model was selected based on the statistical criteria (R2, RMSE and MAE).

2-3 Study Area

Evapotranspiration is an important factor for water resource management. Nazari (2015) showed that Blaney-Criddle is the best equation for estimating evapotranspiration in Qazvin [13]. Also, maximum and minimum temperatures are crucial components of the equation. On other hand, maximum and minimum temperatures are important components of the equation. Hence, Qazvin province was selected as study area. The study area (Fig.3) is in the north-west of Iran with an area of 3918.79 km^2. The region has a cold but dry climate with the annual mean temperature of 13°C and the annual precipitation is about 320mm. For this study, the monthly averages of maximum and minimum temperature were used from year 1961 to 2015.

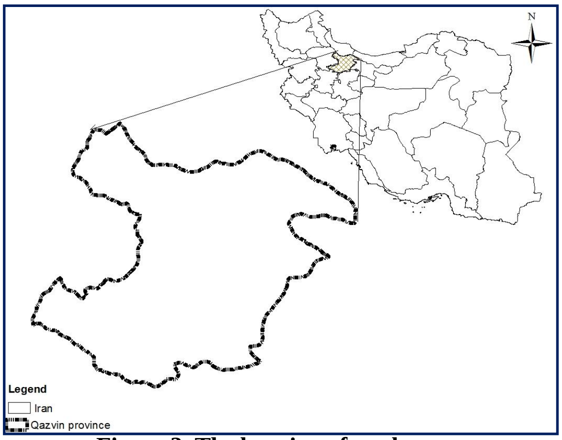

Figure 3. The location of study area

2-4 Statistical Analysis

Statistical analysis was used to calculate the accuracy of models. In this study, these indices have been calculated for determining accuracy as follows:

1- Mean Absolute Error (MAE):

$$MAE = \frac{\sum_{i=1}^{N} |(P_{real} - P_{sim})|}{N} \tag{10}$$

Where, P_{sim} is the amount of simulated climate parameter, P_{real} is the amount of observed climate parameter and N is the number of observed climate parameter.

2- R Square

$$R^2 = \frac{\sum_{i=1}^{N}(P_{sim}-\bar{P}_{sim})^2(P_{real}-\bar{P}_{real})^2}{\sum_{i=1}^{N}(P_{sim}-\bar{P}_{sim})^2\sum_{i=1}^{N}(P_{real}-\bar{P}_{real})^2} \quad (11)$$

Where, $\bar{P}_{sim}$ is the mean amount of simulated climate parameter and $\bar{P}_{real}$ is the mean amount of observed climate parameter.

3- Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(P_{real}-P_{sim})}{N}} \quad (12)$$

3. Result and Discussion

ANN and ANFIS models are introduced into hydrology science as a powerful, flexible, and statistical modeling technique to address complex pattern recognition problems. One of the most important tasks in developing a satisfactory ANN and ANFIS forecasting model is the selection of the input variables, but it is not clear how long the time-delay should be continued to achieve the proper structure. Therefore, the statistical indices between initial data should be considered first [14]. In this study, different combinations of input data were explored to assess their influence on the minimum and maximum temperature estimation modeling.

3-1 Result of Multi Layer Perceptron Artificial Neural Network (MLP-ANN)

The testing result showed that Multi Layer Perceptron Artificial Neural Network's structures with two hidden layers, six inputs, Levenberg-Marqardt learning algorithm, hyperbolic tangent transfer function for hidden layers and linear transfer function for output layer had the best result for simulation minimum and maximum temperature. The networks' structure6- 6-18-1 and 6-5-23-1 had best performance for prediction minimum and maximum temperature respectively. The best fit models structure are determined according to criteria of performance that is shown in table 1and 2.

Moreover, with liner regression, the relationship between the predicted values by neural network and the observed value has been plotted. The plotted line has been compared with 1:1 line. Finally, with a linear regression model without intercept, the relationship between the simulated weather data by MLP-ANN and the observed values has been calculated for the testing and training data. Results have been shown in figure 4a, b. The equations of the liner regression are as follows:

$$T_{min\ observed} = 1.01 \times T_{min\ Simulated} \quad (13)$$

$$T_{max\ observed} = 1.02 \times T_{max\ Simulated} \quad (14)$$

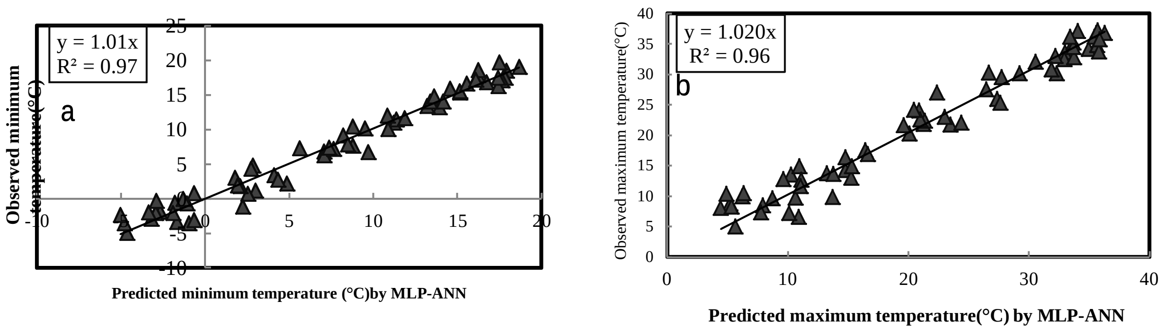

Figure 4. Linear regression between the observed and predicted values of a. minimum temperature and b. maximum temperature using MLP-ANN

Table 1. Statistical indices for different methods (ANN and ANFIS) for prediction minimum temperature

Minimum temperature					
Type of model	**Input structure numbers of the variable**	**Best structure**	**R^2**	**RMSE**	**MAE**
ANN1	$T_{min}(t-1)$	1-19-16-1	0.77	3.7	3
ANN2	$[T_{min}(t-1), T_{min}(t-3)]$	2-5-30-1	0.94	1.62	1.3
ANN3	$[T_{min}(t-1), T_{min}(t-3), T_{min}(t-6), T_{min}(t-9), T_{min}(t-12), T_{min}(t-36)]$	6-6-18-1	0.97	1.28	1.02
ANFIS1	$T_{min}(t-1)$	16	0.74	3.38	3.34
ANFIS2	$[T_{min}(t-1), T_{min}(t-3)]$	7	0.97	1.12	2.37
ANFIS3	$[T_{min}(t-1), T_{min}(t-3), T_{min}(t-6), T_{min}(t-9), T_{min}(t-12), T_{min}(t-36)]$	6	0.98	1.05	1.8

3-2 Result of Neuro-fuzzy approach (ANFIS)

The structure parameters, which were selected for ANFIS are shown in table 3.The best fit models structure was determined due to the statistical indices that is presented in table 1and 2. Result showed that the ANFIS had best performance with six inputs and 6 clusters for the minimum and maximum temperature, respectively. A linear regression model has been fitted to the relationship between the simulated amounts of climate data by ANFIS and the observed values.

Table2. Statistical indices for different methods (ANN and ANFIS) for prediction maximum temperature

Maximum temperature					
Type of model	**Input structure numbers of the variable**	**Best structure**	**R^2**	**RMSE**	**MAE**
ANN1	$T_{max}(t-1)$	1-13-8-1	0.74	5.2	4.26
ANN2	$[T_{max}(t-1), T_{max}(t-3)]$	2-5-20-1	0.95	2.09	1.63
ANN3	$[T_{max}(t-1), T_{max}(t-3), T_{max}(t-6), T_{max}(t-9), T_{max}(t-12), T_{max}(t-36)]$	6-5-23-1	0.96	2.12	1.4
ANFIS1	$T_{max}(t-1)$	14	0.73	5.37	4.4
ANFIS2	$[T_{max}(t-1), T_{max}(t-3)]$	5	0.95	4.5	1.69
ANFIS3	$[T_{max}(t-1), T_{max}(t-3), T_{max}(t-6), T_{max}(t-9), T_{max}(t-12), T_{max}(t-36)]$	6	0.97	1.5	3.2

Table 3. The training parameters of the ANFIS model

types of fuzzy inference system	Sugeno
Epoch	150
AND method	Prod
Implication method	Minimum
Aggregation method	sum
Defuzzification method	Wtaver
Initial Step size	0.01
Step size decrease rate	0.9
Step size increase rate	1.1
Number of clusters	Vary from 4 to 22

This linear regression has been compared with a 1:1 line, and the comparison between the fitted line and 1:1 line at 5% level of probability has been investigated. Result showed that the relationship between the simulated weather data by ANFIS and the observed values with a linear regression model without intercept was statistically significant (Fig. 5a, b).The equations of liner regression without intercept of data are as follows:

$$T_{min\ observed} = 1.01 \times T_{min\ Simulated} \tag{15}$$

$$T_{max\ observed} = 1.01 \times T_{max\ Simulated} \tag{16}$$

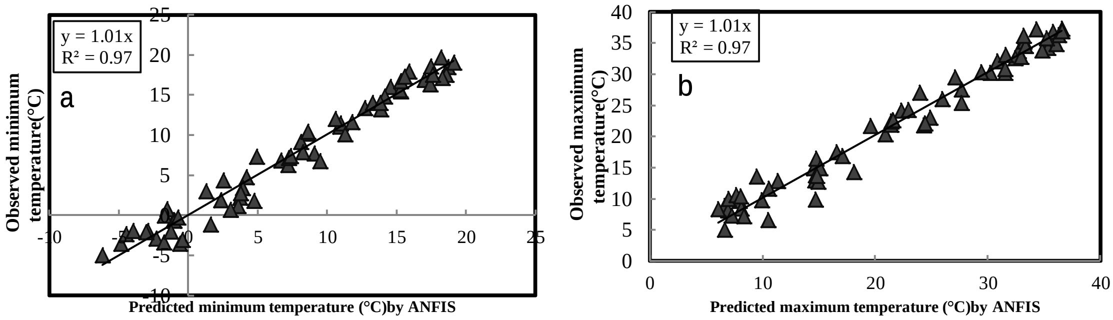

Figure 5. Linear regression between the observed and predicted values of a. minimum temperature and b. maximum temperature using ANFIS

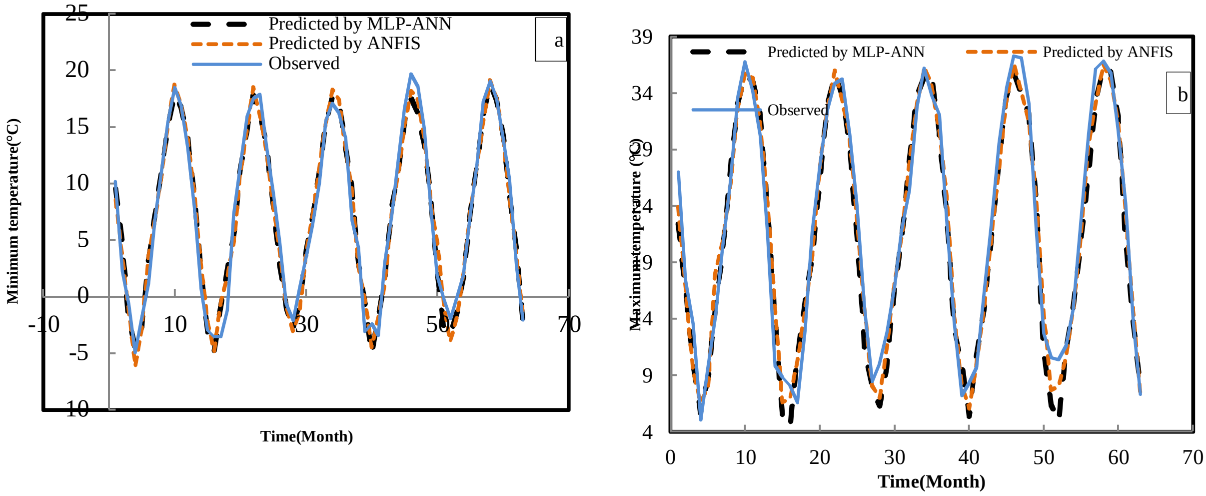

Figure 1. Comparison between the observed and simulated the Maximum and minimum temperature by MLP-ANN and ANFIS as a function of time (month) a. **Minimum temperature and b. Maximum temperature**

4. CONCLUSION

In this study, the potential of multi-layer perceptron artificial neural network and ANFIS were investigated in arid and semi- arid region. MLP-ANN with Levenberg-Marqardt learning algorithm, hyperbolic tangent transfer function for hidden layers and linear transfer function for output layer had the best output for both weather data. The networks' structure 6- 6-18-1 and 6-5-23-1 had best efficiency for simulation minimum and maximum temperature, respectively. Moreover, the performance of the ANFIS Neuro-Fuzzy Inference System with the fuzzy c-means clustering (FCM-ANFIS), six inputs and six clusters was the best. Result indicated that time series of minimum and maximum temperature could be successfully predicted in semi-arid region by MLP-ANN and ANFIS methods.

5. REFERENCES

1. Delleur, JW. Tao, PC. and Kavass, ML. (1976). An Evaluation of Practicality and Complexity of Some Rainfall and Runoff Time Series Models, *Water Resources Research*, 12, 953–970.

2. Salas, JD and Fernandez, B. (1993). Models for Data Generation in Hydrology: Univariates Techniques, In Stochastic Hydrology and Its Use in Water Resources Systems Simulation and Optimization, Macro JB, Harboe R, Salas JD (eds), NATO Advanced Study Institute, Kluwer Academic Publishers: Boston, 47–73.

3. Hipel, KW. And McLeod, AE. (1994). Time Series Modeling of Water Resources and Environmental Systems. *Elsevier*, Amsterdam.

4. Tabari, H. Hosseinzadeh Talaeeb, P. and Willems, P. (2015). Short-Term Forecasting of Soil Temperature Using Artificial Neural Network, *Meteorological applications*, DOI: 10.1002/met.1489

5. Haviluddin. and Alfred, R. (2014). Forecasting Network Activities Using ARIMA Method, *Journal of Advances in Computer Networks*, 2, 173-179.

6. Shrivastava, G. Karmakar, S. and Kowar, M.K. (2012). Application of Artificial Neural Networks in Weather Forecasting: A ComprehensiveLiterature Review, *International Journal of Computer Applications*,51, p. 17-29.

7. Kumar Nanda, S. Prasad Tripathy, D. Kumar Nayak, S. and Mohapatra, S. (2013). Prediction of Rainfall in India Using Artificial Neural Network (ANN) Models, *International Journal of Intelligent Systems and Applications*, DOI: 10.5815/ijisa.2013.12.01

8. Farajzadeh, J. Fard A.F. and Lotfi, S. (2014). Modeling of Monthly Rainfall and Runoff Of Urmia Lake Basin Using Feed-Forward Neural Network and Time Series Analysis Model, *Water Resources and Industry*,7(8) , 38–48.

9. Mislan. Haviluddin. Sigit Hardwinarto. Sumaryono and Marlon Aipassa. (2015). Rainfall Monthly Prediction Based on Artificial Neural Network: a Case Study in Tenggarong Station, East Kalimantan – Indonesia, *International Conference on Computer Science and Computational Intelligence (ICCSCI 2015). Procedia Computer Science*, 59 , 142 – 151.

10. Abhishek, K . Kumar,A. Ranjan, R . and Kumar, S. (2012). A Rainfall Prediction Model using Artificial Neural Network, IEEE Control and System Graduate Research Colloquium (ICSGRC 2012).

11. Kumar, N. and Kumar Jha, G. (2013). A Time Series ANN Approach for Weather Forecasting, *International Journal of Control Theory and Computer Modeling (IJCTCM)*, 3(1).

12. Jang, JSR. Sun, CT. and Mizutani, E. (1997). Neuro-Fuzzy and Soft Computing: Computational Approach to Learning and Machine Intelligence. Prentice Hall: Upper Saddle River, NJ.

13. Nazari, R.(2015). Estimation of Actual Evapotranspiration of Crops and Orchards in Qazvin Using Satellite Imagery and METRIC Algorithm, *M.Sc Thesis , Imam Khomeini International University*.(in Persian)

14. Chen, SH and Lin, YH .(2006). The Strategy of Building a Flood Forecast Model by Neuro-Fuzzy Network. *Hydrological Processes*, 20(7), 1525-1540.