Daily Rainfall Forecasting for Mashhad Synoptic Station using Artificial Neural Networks

2011 International Conference on Environmental and Computer Science
IPCBEE vol.19(2011) © (2011) IACSIT Press, Singapore

Daily Rainfall Forecasting for Mashhad Synoptic Station using Artificial Neural Networks

Najmeh Khalili [1+], Saeed Reza Khodashenas [2], Kamran Davary[3] and Fatemeh Karimaldini [4]

[1+] Ph.D Student, Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Iran

[2] Assistant Professor, Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Iran

[3] Associated Professor, Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Iran

[4] M.Sc Student, Department of Agriculture and Biological, Faculty engineering, University Putra Malaysia, Malaysia

Abstract. In this paper, we have utilized ANN (Artificial Neural Network) modeling for daily rainfall forecasting in Mashhad synoptic station. To achieve such a model, we have used daily rainfall data of March as a month with high humidity and May and December as months with medium humidity from 1986 to 2010 for this synoptic station. First, the Hurst rescaled range statistical (R/S) analysis is used to evaluate the predictability of the collected data. Then, to extract the precipitation dynamic of this station using ANN modeling, a new approach of three-layer feed-forward perceptron network with back propagation algorithm is proposed. Using this ANN model as a black box model, we have realized the hidden dynamics of rainfall through the past information of the system. The approach employs the gradient decent algorithm to train the network. Trying different parameters, some structures including GS_{531} and GS_{651} for March, GS_{521} and GS_{681} for May and GS_{571} and GS_{631} for December, have been selected which give the best estimation performance. Performance statistical analysis of the obtained models shows that in the best chosen model of daily forecasting, the correlation coefficient (R), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are 0.89, 0.14(mm) and 1.15(mm) for March, 0.85, 0.14(mm) and 1.16(mm) for May and 0.86, 0.15(mm) and 1.17(mm) for December, respectively which presents the effectiveness of the proposed models.

Keywords: Artificial Neural Networks, Daily Rainfall Forecasting, Feed-Forward Perceptron, Mashhad.

1. Introduction

The rainfall as meteorological parameters is very complex nonlinear phenomena and varies along with time and place. Nevertheless, literatures show that rainfall is predictable [1], [2], [3].

In the last years, many models of ANNs have been developed for rainfall prediction. Luk, (2001) predicted rainfall in catchment's upper Parramatta River in Australia using Multi Layer Feed-forward Neural Network (MLFN)[2]. Their results showed that MLFN has more accuracy in rainfall modelling in comparison to Time Delay Neural Network (TDNN) and Recurrent Neural Network (RNN). While TDNN anticipated to RNN and MLFN through rainfall prediction using large scale continental signals in west of Iran. Ramirez et al. (2005) also used a Multi Layer Feed-forward Perceptron (MLFP) neural network for daily rainfall prediction in the region of Sao Paolo State, Brazil. In their research, potential temperature, vertical component of the wind, specific humidity, air temperature, perceptible water, relative vortices and moisture divergence flux are used as input data for training of networks. Results of ANN were superior to the ones obtained by the linear regression model, which is revealing a great potential for suitable performance

---

+ Corresponding author. Tel.: + (98-511-8795612); fax: +(98-511- 8787430).
E-mail address: (najmehkhalili@gmail.com).

[4]. In a similar study, Saplioglu et al (2010) employed a three layer feed-forward neural network for daily rainfall prediction in the meteorological stations of Burdur, Egirdir, and Isparta cities in Turkey [5]. The results indicate ANN models are superior to the commonly used weighted average and harmonic average methods in this study. Generally, there isn't any fixed ANN model as a suitable network for all of the problems, instead any ANN model using different criterion should be tested until one obtain the appropriate model for desirable purpose [6]. However, according to the literature, it seems that for rainfall prediction, Multi Layer Feed-forward Perceptron (MLFP) has more reasonable outputs in comparison to other ANN types.

In this paper, we have used ANNs to obtain a forecasting model for the daily rainfall of Mashhad's synoptic station in Iran. This is due to the intelligent capability of the neural networks in the extraction of the features of the systems, even in the cases that there is not much information about the system dynamics. In these cases, we will suppose to use a black box model for the system and capture system's dynamics through its memory. Having more information about the affective factors on the system, we can use a gray box model and utilize the additional information in our modelling. However, Hurst's Rescaled range statistical (R/S) analysis shows that the collected data in Mashhad station is predictable using only the past information of system and only a black box modelling of the system could capture its dynamic.

Therefore, a new approach has been used for rainfall forecasting in Mashhad's synoptic station in Iran. Some black box structures of MLFP were used and in these MLFP structures only previous daily rainfall data are exploited for precipitation prediction. A test model that is applied to specify the predictability of data and artificial neural network method is introduced. The utilized ANN structures and obtained results are illustrated.

2. Materials and methods

2.1. Data and Location of the study

Daily rainfall data has collected from Mashhad's synoptic weather station. Mashhad's synoptic weather station is located in the north-east of Iran at $36^{o}16'$ Northern longitude, $59^{o}38'$ Eastern latitude and 999.2 meter elevation. In Mashhad's special geographical situation, interfacing different air masses make it a region with a special continental climate. This location affected by Polar continental, Maritime Tropical and Sudanian air masses. Overall, its climate is as mid dry and cold with dry-hot summers and wet-cold winters. The maximum annual temperature is about +35 and the minimum is about -15. The annual average precipitation is about 253 mm in Mashhad.

Using the collected data for the mentioned case, some tests such as Mackus and Run Test [7] were carried out for determination of data sufficiency and homogenous. In addition, Hurst's Rescaled range statistical (R/S) analysis test [8] was used for assessment of data predictability. Indeed, this statistical index captures the existence of memory effect in the given data. Finally, rainfall was predicted using MLFP that is a suitable type of ANNs for meteorological predictions [1], [3].

The main objective of the research is to develop an artificial neural network for the purpose of daily rainfall forecasting. To achieve this objective, we used the daily precipitation data from 1986 to 2010 of March as a month with high humidity and May and December as months with medium humidity.

2.2. Artificial Neural Networks (ANNs)

Figure 1 shows an example for ANN networks, with three layers. In the ANN model, the relation between neurons inputs (x_i) and their outputs (y_i) is specified by synapse weight parameters (w), bias (b), the activation functions in hidden layer ($f(x)$), and output function in the layer ($g(x)$). This relation is explained mathematically as described in Eq. (1):

$$y = g[[\sum_{j=1}^{n} w_{kj} f(\sum_{i=1}^{n} (w_{ji}x_i + b)) + b]] \tag{1}$$

Where, $x_1, x_2, \ldots, x_i$ are inputs, $w_{j1}, w_{j2}, \ldots, w_{jn}$ are synapse weights towards the hidden layer, $w_{k1}, w_{k2}, \ldots, w_{kn}$ are synapse weight towards the output layer, b is the bias or external threshold, $f(x)$ is the activation function in the hidden layer, $g(x)$ is the activation function in output layer, and y is the output.

It is assumed that the network does not have any a priori knowledge about the problem. Therefore, the network must be firstly trained with repeated sets of input patterns. During the training process, the error between the desired output and the calculated output is propagated back through the network, hence is called "back propagation algorithm" [9].

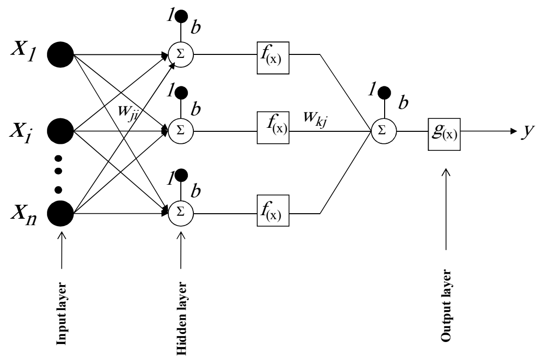

Fig. 1: The artificial neural network structure.

2.3. Structure of utilized ANNs

After the assessment of the predictability of available data (R/S analysis), different types of neural network models were tested and their performance were evaluated. Based on our results and also based on the literature, Multi Layer Feed-forward Perceptron (MLFP) neural network with back propagation training method can provide us with proper results [1], [3]. The daily data for 23 years from 1986 to 2010 were arranged as a time series with length of the 713 daily data. Then, we have constructed a matrix with m rows and n columns in which, each row corresponds to one of data to be given to the ANN model. Each row is an augmented data in which the columns 1 to n-1 are the information that are required to be used to predict the target data in the last column. In other words, columns 1 to n-1 are the inputs of the ANN and the last column is the data to be estimated (target data).

Among these collected data samples, 580 data samples were used for the training phase and the rest were used for the validation phase. The training data could be selected in order; however, more comprehensive approach of training is random selection of augmented data in the constructed matrix. In this case, the ANN model has more chance to capture the dynamics of the system and later on, the trained model is able to respond to the given data in any order. For validation phase, we have used a similar procedure and instead of validating the system with the ordered data, we gave the validation data to the trained ANN model in a randomly selected way. Therefore, after training, we expect that the trained ANN model to be able to predict the precipitation value for a desired day by providing the necessary information about the history of that particular day. In the above mentioned procedure, after trying different structures of MLFP models, the number of hidden layer neurons was chosen to achieve the best possible output. It should also be noted that the number of epochs was selected as 1000 and η was held at constant value of 0.5. The selected activation functions in hidden and output layer of the networks were sigmoid and linear, respectively.

In the next section, we have described different MLFP structures that we have tested for this case study. In particular, we will discuss the design of the input layers which are related to rainfall data of the previous time period and the design of the output layers which give the predicted daily rainfall.

3. Results and discussions

Hurst exponent of daily data in our work was estimated 0.9422. This value implies that the daily set of given time series is predictable. Then, using MATLAB software, several different structures of MLFP were designed with different numbers of neurons in the input and hidden layers. Each structure is introduced in the form of M_{ijk} , in which the indices i, j, and k stand for number of neurons in the input layer, the hidden layer, and the output layer, respectively. Among different structures, we found GS_{531} and GS_{651} for March, GS_{521} and GS_{681} for May and GS_{571} and GS_{631} for December as the models that give the best performance.

3.1. Structure of GS_{531}, GS_{521} and GS_{571}

The input layer of this network consists of 5 neurons for the last 5 bi-daily rainfall moving-averages. For this structure, 3, 2 and 7 neurons in the hidden layer yields the best prediction performance for March, May and December, respectively. The structure of GS_{531} (March) is shown in Figure 2.

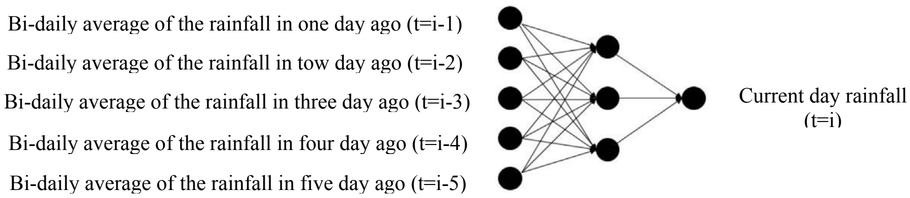

Fig. 2: GS_{531} neural network structure.

3.2. Structure of GS_{651}, GS_{681} and GS_{631}

In this structure, 6 neurons were used in the input layer. These inputs are last year rainfall for the estimated day, and the last 5 bi-daily rainfall moving-averages. After several trials and errors, seeking optimal number of neurons in the hidden layer, 5, 8 and 3 neurons were selected for this layer for March, May and December, respectively. Moreover, for this structure, all input data were initially normalized. Obviously, the network output should be de-normalized correspondingly. Figure 3 shows the topology of this network.

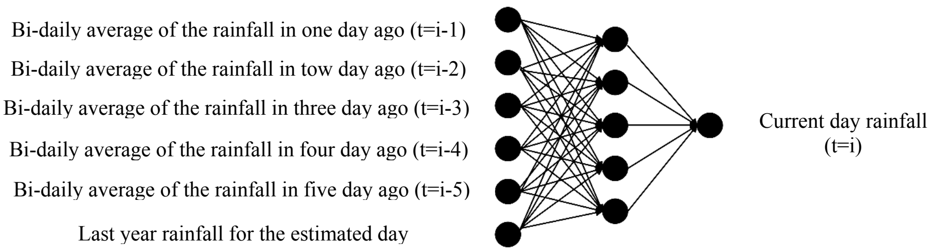

Fig. 3: GS_{651} neural network structure.

The obtained results of validation phase are shown in table 1 that include Correlation Coefficient (*R*), Root Mean Square Error (*RMSE*) and Mean Absolute Error (*MAE*) for daily prediction using GS_{531} and GS_{651}.

Table 1: results of used superior ANNs structures

ANN structures	*R*	*RMSE* (mm)	*MAE*
GS_{531} (March)	0.83	0.17	1.20
GS_{651}(March)	0.89	0.14	1.15
GS_{521}(May)	0.82	0.19	1.22
GS_{681}(May)	0.85	0.14	1.16
GS_{571}(December)	0.82	0.20	1.24
GS_{631}(December)	0.86	0.15	1.17

Table 1 shows a correlation ($R^2 = 0.69$) and indicate that the GS_{531} model provided good predictions of daily rainfall values for Mashhad's synoptic station. However, there is a slight tendency to underestimation of precipitation. Figure 4.a shows the actual and predicted rainfall in validation phase for GS_{531} model. The results do not show significant improvement by changing the structure of the hidden layer or even increasing the number of neurons in the input or hidden layer. Moreover, considering the number of neurons in the whole network, the training data is sufficient and the problem cannot comes from data insufficiency. It should be highlighted that in this MLFP structure, only bi-daily rainfall moving-averages are used in the input layer and this may not capture the dynamics of the system. To leverage the prediction performance, we used the GS_{531} structure in which a richer input data is used to feed the network.

A stronger correlation (R^2 = 0.79) is shown in table 1 between predicted and actual rainfall values is observed for validation phase of GS_{651} in comparison to GS_{531}. As it can be seen in Figure 4, GS_{651} give more satisfactory performance. Using input neurons including last year rainfall for the estimated day beside the last 5 bi-daily rainfall moving-averages in GS_{651} provides richer input data which it has been lead to better results in GS_{651} comparing to GS_{531}.

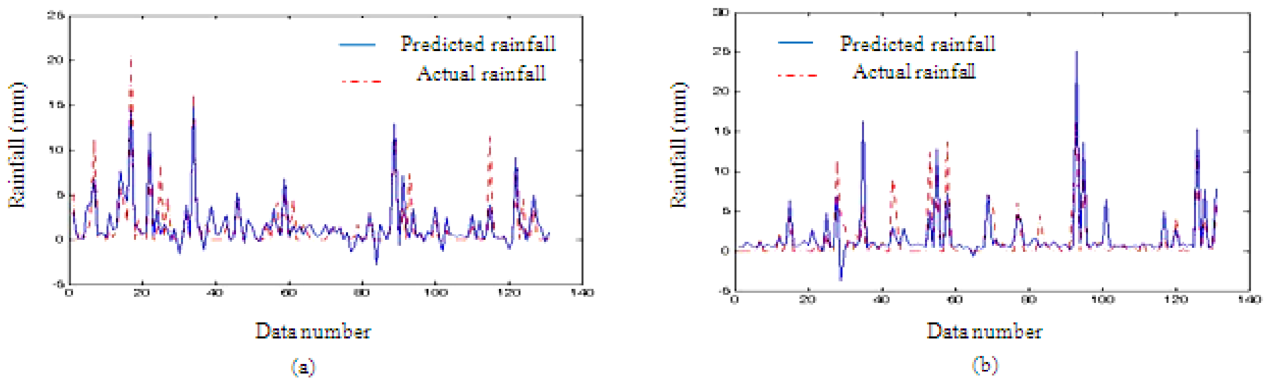

Fig. 3: Accordance between predicted and actual daily rainfall validation phase. (a) GS531 and (b) GS651

4. Conclusion

The proposed method and discussed results on the use of MLFP networks for the precipitation prediction can be recapped as follows:

- There are different ANN model structures, training algorithms, activation functions, and number of epochs. This makes the find of proper model for a particular problem difficult. The selection of appropriate structure could be specified only by designer experience and with a trial and error procedure.
- Using the rainfall data for the corresponding day in the previous year and the last five bi-daily rainfall moving-averages, results in the much better prediction performance.
- The achieved ANN model gave satisfactory prediction performance.

As future work, we are going to use a gray box model instead of the black box. In fact, although this paper shows that the black box model is capable of predicting the rainfall, it is reasonable to employ the prior information in our rainfall model in the form of a gray box ANN model to improve the prediction performance.

5. Acknowledgment

The authors would gratefully acknowledge the support of Meteorology Organization of Iran for providing the rainfall data for this research.

6. References

[1] S. Hastenrath, and L. Greischar. (1999). Prediction of the summer rainfall over South Africa. *Journal of Climate*. 1999, **8**: 1511-1518.

[2] K. C. Luk. An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer Modeling Journal*. 2001, **33**: 883-699.

[3] S. P. Ninan, S. P, and J. Babo. A neural network tool for analyzing trends in rainfall. *Computers & Geosciences Journal*. 2003, **29**: 215–223.

[4] M. C. V. Ramirez , N. J. Ferreira, and H. F. Velho. Artificial neural network technique for rainfall forecasting applied to the Sa˜o Paulo region. *Journal of Hydrology*. 2005, **301**: 146–162.

[5] K. Saplioglu , M. Cimeny, and B. Akman. Daily Precipitation Prediction in Isparta Station by Artificial Neural Network. Proc. of the 4th International Scientific Conference on Water Observation and Information System for Decision Support (BALWOIS). Ohrid, Republic of Macedonia. 2010.

[6] N. Khalili. Rainfall prediction using artificial neural networks. *Thesis of Master degree of Irrigation and drainage engineering*. Faculty of Agriculture, Ferdowsi university of Mashhad (In Persian).2008.

[7] H. Beker, and F. Piper. The Principles of Cryptography, Part IV: Stream Ciphers, Section A: Randomness." *Secure Speech Communications.* Academic Press. 1985, Chapter **3**: 104-109.

[8] B. M. Gammel. Hurst's rescaled range statistical analysis for pseudorandom number generators used in physical simulations. *The American physical society Journal*.1998, **58**(2): 2586-2597.

[9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In: D. E. Rumelhart and J. L. McClelland, (eds.). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* MIT Press, Bradfords Books, Cambridge, MA. 1986, Vol. 1, 318-362.