International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-5, Issue-4, July 2017
DOI:10.21276/ijircst.2017.5.4.5

Evaluation of Performance of Statistical and ANN Approaches for Prediction of Rainfall

N. Vivekanandan

***Abstract*—Prediction of rainfall for a river basin is of utmost importance for planning and design of irrigation and drainage systems as also for command area development. Since the distribution of rainfall varies over space and time, it is required to analyze the data covering long periods and recorded at various locations to arrive at reliable information for decision support. Further, such data need to be analyzed in different ways, depending on the issue under consideration. In the present study, Extreme Value Type-1 (EV1) distribution based on statistical approach and Multi Layer Perceptron (MLP) network based on Artificial Neural Network (ANN) is adopted for prediction of rainfall at Fatehabad and Hansi. The performance of the statistical and ANN approaches used in rainfall predication are evaluated by model performance indicators viz., correlation coefficient, model efficiency and mean absolute percentage error. The study shows the MLP is found to be better suited network for prediction of rainfall at Fatehabad whereas EV1 for Hansi.**

***Index Terms*—Artificial Neural Network, Correlation Coefficient, Extreme Value Type-1, Mean Absolute Percentage Error, Model Efficiency, Multi Layer Perceptron**

I. INTRODUCTION

Prediction of rainfall for a river basin is of utmost importance for planning and design of irrigation and drainage systems as also for command area development. Since the distribution of rainfall varies over space and time, it is required to analyze the data covering long periods and recorded at various locations to arrive at reliable information for decision support [1]. Further, such data need to be analyzed in different ways, depending on the issue under consideration. Out of a number of probability distributions, the family of Extreme Value Distributions (EVDs) includes Generalized Extreme Value, Extreme Value Type-1 (EV1), Extreme Value Type-2 and Extreme Value Type-3 is generally used for rainfall prediction. EVDs arise as limiting distributions for the sample of independent, identically distributed random variables, as the sample size increases. In the group of EVDs, EV1 distribution has no shape parameter as when compared to other distributions and this means that there is no change in the shape of Probability Distribution Function (PDF). Moreover, EV1 distribution has the advantage of having only positive values, since the data series of rainfall are always positive; and therefore EV1 distribution is important in statistics.

Manuscript received 23 July, 2017.

N.Vivekanandan, Scientist-B, Central Water and Power Research Station, Pune 411024, Maharashtra, India,
Phone: +91 20 2410 3367.

Apart from this, with the development of Artificial Intelligence (AI), a number of various AI methods have been developed for prediction of rainfall. The new developed methods include Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System, Fuzzy Logic, Support Vector Machine, Neuro-Fuzzy Network and Evolutionary Optimization Algorithm. Out of these methods, ANN could deal with non-linear and complex problems in terms of classification or forecasting. The ANN models can represent a complex nonlinear relationship and extract the dependence between variables through the training process [2-3]. In the present study, statistical and ANN approaches are adopted for prediction of rainfall for the data under study.

II. METHODOLOGY

A) Extreme Value Type-1 Distribution

The PDF and Cumulative Distribution Function (CDF) of the EV1 distribution are given as:

$$\left.\begin{aligned} &\text{PDF}: f(X)=\frac{e^{-(X-\alpha)/\beta}e^{-e^{-(X-\alpha)/\beta}}}{\beta} \\ &\text{CDF}: F(X)=e^{-e^{-(X-\alpha)/\beta}}, \beta>0\text{, where } (X=X_1, X_2, X_3, \ldots, X_N) \end{aligned}\right\} \quad (1)$$

where, α and β are the location and scale parameters of the distribution [4]. The parameters are computed by Maximum Likelihood Method (MLM) through Equations (2) and (3), and used to estimate the rainfall (X_T) for different return periods from $X_T=\alpha+Y_T\beta$. Here, $Y_T=-\ln(-\ln(1-(1/T)))$ is called as a reduced variate for a given return period T (year).

$$\alpha=-\beta\ln\left[\sum_{i=1}^{N}\exp(-X_i/\beta)\Big/N\right] \quad (2)$$

$$\beta=\overline{X}-\left[\sum_{i=1}^{N}X_i\exp(-X_i/\beta)\Big/\sum_{i=1}^{N}\exp(-X_i/\beta)\right] \quad (3)$$

where X_i is the observed rainfall of i^{th} sample, $\overline{X}$ is the average value of observed rainfall and N is the sample size.

B) Multi-Layer Perceptron Network

ANN modelling procedures adapt to complexity of input-output patterns and accuracy goes on increasing as more and more data become available. Figure 1 shows the architecture of ANN that consists of input layer, hidden layer, and output layer [5]. In turn, these layers have a certain number of neurons or units, so the units are also called input units, hidden units and output units. From ANN structure, it can be easily understood that input units receive data from external sources to the network and send them to the hidden units, in turn, the hidden units send and receive data only from other units in the network, and output units receive and produce data generated by the network, which goes out of the system. In this process, a typical problem is to estimate the output as a function of the input. This unknown function may be approximated by a superposition of certain activation functions such as tangent, sigmoid, polynomial, and sinusoid in ANN. A common threshold function used in ANN is the sigmoid function (f(S)) expressed by Eq. (4), which provides an output in the range of $0 \leq f(S) \leq 1$.

$$f(S) = \left[1 + \exp\left(-S_i\right)\right]^{-1} \text{ and}$$
$$S_i = \sum_{i=1}^{N} I_i W_{ij} + O_i, \quad j=1,2,3,\ldots,M \qquad (4)$$

where S_i is the characteristic function of i^{th} layer, I_i is the input unit of i^{th} layer, O_i is the output unit of i^{th} layer, W_{ij} is the synaptic weights between input and hidden layers, N is the number of observations and M is the number of neurons of hidden layer [6]. The sigmoid function is chosen for mathematical convenience because it resembles a hard-limiting step function for extremely large positive and negative values of the incoming signal and also gives sufficient information about the response of the processing unit to inputs that are close to the threshold value. In ANN, number of training algorithms viz., Multi Layer Perceptron (MLP), Back Propagation, Recurrent, Radial Basis Function and Adaptive Linear Element are used for training the network [7-11]. The objective in training the network is to reduce the global error between the predicted and targeted outputs. In this paper, MLP network is used to predict the rainfall with illustrative example.

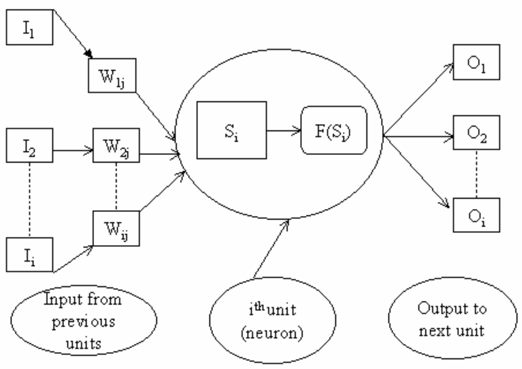

Figure 1: Architecture of ANN

MLP network [12] is the most widely used for rainfall prediction and its architecture with single hidden layer is shown in Figure 1. Gradient descent is the most commonly used supervised training algorithm in MLP in which each input unit of the training data set is passed through the network from the input layer to output layer. The network output is compared with the desired target output and output error (E) is computed using Eq. (5).

$$E = \frac{1}{2}\sum_{i=1}^{N}\left(X_i - X_i^*\right)^2 \qquad (5)$$

where, X_i is the observed rainfall of i^{th} sample and X_i^* is the predicted rainfall for i^{th} sample.

$$\Delta W_{ij}(M) = -\varepsilon \frac{\partial E}{\partial W_{ij}} + \alpha \Delta W_{ij}(M-1) \qquad (6)$$

where, W_{ij} is the synaptic weights between input and hidden layers, $\Delta W_{ij}(M)$ is the weight increments between i^{th} and j^{th} units during M neurons (units) and $\Delta W_{ij}(M-1)$ is the weight increments between i^{th} and j^{th} units during $M-1$ neurons [13]. In MLP, momentum factor (α) is used to speed up training in very flat regions of the error surface to prevent oscillations in the weights and learning rate (ε) is used to increase the chance of avoiding the training process being trapped in local minima instead of global minima [14].

C)Model Performance Analysis

The performance of EV1 distribution and MLP network used in rainfall prediction is evaluated by Model Performance Indicators (MPIs) viz., Correlation Coefficient (CC), Model Efficiency (MEF) and Mean Absolute Percentage Error (MAPE), and are:

$$\left.\begin{aligned}
CC &= \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)\left(X_i^* - \overline{X^*}\right)}{\sqrt{\left(\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2\right)\left(\sum_{i=1}^{N}\left(X_i^* - \overline{X^*}\right)^2\right)}} \\
MEF(\%) &= \left(1 - \frac{\sum_{i=1}^{N}\left(X_i - X_i^*\right)^2}{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}\right) * 100 \\
MAPE &= \frac{1}{N}\sum_{i=1}^{N}\left|\left(\frac{X_i - X_i^*}{X_i}\right)\right| * 100
\end{aligned}\right\} \qquad (7)$$

where $\overline{X}$ is the average value of observed rainfall and $\overline{X^*}$ is the average value of predicted rainfall [15].

III. APPLICATION

In this paper, a study on prediction of rainfall at Fatehabad and Hansi rain-gauge stations using EV1 distribution and MLP network was carried out. The series of Annual Maximum 1-Day Rainfall (AMDR) was extracted from the daily rainfall data recorded at these rain gauge stations was used. For Fatehabad and Hansi, the rainfall data for the period 1954 to 1996 was used for training the network whereas data for the period 1997 to 2011 was used for testing the network.

IV. RESULTS AND DISCUSSIONS

Statistical software, namely, 'Hydrognomon' was used to determine the parameters of EV1 distribution whereas 'SPSS Neural Connection' was used to train the network data with different combinations of parameters to determine optimum network architecture of MLP. The determined optimum network architecture with model parameters was used for prediction of rainfall at Fatehabad and Hansi.

A) Prediction of Rainfall using EV1 Distribution

By using the annual maximum series of rainfall recorded data Fatehabad and Hansi stations, the location and scale parameters of EV1 were determined by MLM and presented in Table 1. For Fatehabad and Hansi, the parameters obtained from EV1 were used to determine the predicted value of rainfall at consecutive years based on the probabilities of observed value of rainfall and presented in Figures 2 and 3.

Table 1: Parameters of EV1 Distribution

Rain-gauge station	Location	Scale
Fatehabad	48.004	24.038
Hansi	39.453	39.811

International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-5, Issue-4, July 2017
DOI:10.21276/ijircst.2017.5.4.5

B) Prediction of Rainfall using MLP Network

For Fatehabad, the momentum factor (α) and learning rate (ε) were fixed as 0.6 and 0.04 while optimizing the network architecture of MLP. Similarly, the values viz., α =0.7 and ε = 0.06 were used for determination of optimum network architecture of MLP for Hansi. The network data was trained with optimum MLP networks, viz., 1-12-1 for Fatehabad and 1-15-1 for Hansi. The networks were also tested with model parameters for prediction of rainfall.

C) Performance Analysis of EV1and MLP using MPIs

The model performance of EV1 distribution and MLP network used in rainfall prediction was evaluated by MPIs and given in Tables 2 and 3.

Table 2: Values of MPIs for Fatehabad

MPIs	EV1		MLP	
	Training	Testing	Training	Testing
CC	0.985	0.975	0.997	0.999
MEF (%)	95.2	96.0	99.2	99.0
MAPE (%)	10.4	21.3	3.7	6.7

Table 3: Values of MPIs for Hansi

MPIs	EV1		MLP	
	Training	Testing	Training	Testing
CC	0.996	0.998	0.982	0.949
MEF (%)	99.0	99.0	96.1	97.2
MAPE (%)	7.4	11.3	15.0	17.4

From Table 2, it may be noted that: (i) The MAPE obtained from MLP is comparatively less than the corresponding values of EV1 while training the network data and therefore the MLP is found to be better suited for prediction of rainfall at Fatehabad; (ii) The percentage of MEF is computed as about 99% while training and testing the network data with MLP; and (iii) There is generally a good correlation between the observed and predicted rainfall using EV1 and MLP, with CC values vary from 0.975 to 0.999.

From Table 3, it may be noted that: (i) The MAPE obtained from EV1 is comparatively less than the corresponding values of MLP while training the network data and therefore the EV1 is found to be better suited for prediction of rainfall at Hansi; (ii) The percentage of MEF in prediction of rainfall using EV1 during training and testing the network data is computed as 99%; and (iii) There is generally a good correlation between the observed and predicted rainfall using EV1 and MLP, with CC values vary from 0.949 to 0.998. Figures 2 and 3 give the plots of observed and predicted rainfall (using EV1 and MLP) for Fatehabad and Hansi respectively.

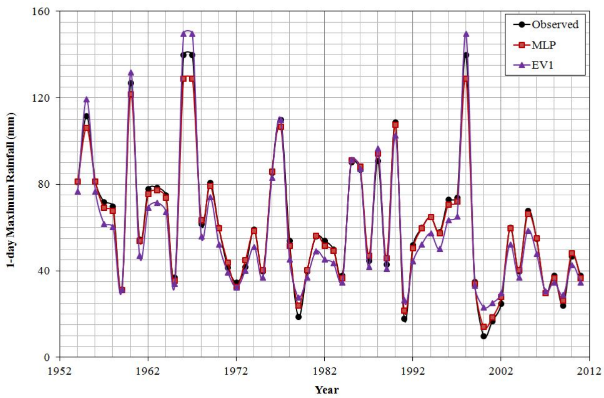

Figure 2: Plots of observed and predicted rainfall using EV1 distribution and MLP network for Fatehabad

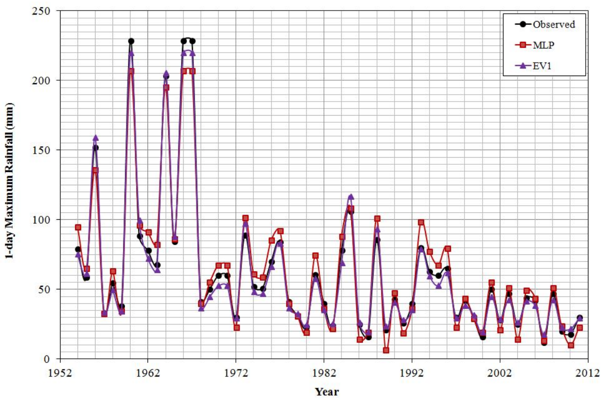

Figure 3: Plots of observed and predicted rainfall using EV1 distribution and MLP network for Hansi

D)Analysis Based on Descriptive Statistics

In addition to MPIs, the overall performance of EV1 distribution and MLP network used in prediction of rainfall was analyzed through the descriptive statistics. For Fatehabad and Hansi, the statistical parameters such as Average, Standard Deviation, Coefficient of Variation, Coefficient of Skewness and Coefficient of Kurtosis for the observed and predicted rainfall (using EV1 and MLP) were computed and given in Tables 4 and 5.

Table 4: Descriptive statistics of observed and predicted rainfall for Fatehabad

Statistical Parameters	Observed rainfall		Predicted rainfall			
			EV1		MLP	
	Training	Testing	Training	Testing	Training	Testing
Average (mm)	67.4	46.7	64.2	46.3	66.5	46.4
Standard Deviation (mm)	29.5	31.6	31.7	31.2	27.4	28.5
Coefficient of Variation (%)	43.8	67.6	49.3	67.3	41.2	61.3
Coefficient of Skewness	0.756	1.929	1.255	2.928	0.635	1.840
Coefficient of Kurtosis	0.315	5.014	1.042	9.693	-0.153	4.482

For Fatehabad, it may be noted that the percentage of variation on the average predicted rainfall using MLP, with reference to average observed rainfall, is 1.3% during training period and 0.6% during testing period. From Table 3, it may be noted that these values were computed as 4.7% and 0.9% while training and testing the rainfall data with

EV1. For Hansi, it was found that the percentage of variation on the average predicted rainfall using MLP, with reference to average observed rainfall, is 3.2% during training period and 2.5% during testing period. From Table 4, it may be noted that these values were computed as 1.9% and 1.2% while training and testing the rainfall data with EV1. Based on the results obtained from model performance analysis and descriptive statistics, the MLP network is considered as better suited for prediction of rainfall at Fatehabad whereas EV1 distribution for Hansi.

Table 5: Descriptive statistics of observed and predicted rainfall for Hansi

Statistical Parameters	Observed rainfall		Predicted rainfall			
			EV1		MLP	
	Training	Testing	Training	Testing	Training	Testing
Average (mm)	73.0	32.1	71.6	31.7	75.3	31.3
Standard Deviation (mm)	55.1	12.5	54.7	9.2	51.5	15.8
Coefficient of Variation (%)	75.5	39.0	76.4	28.9	68.4	50.5
Coefficient of Skewness	1.906	-0.057	1.810	-0.027	1.288	0.279
Coefficient of Kurtosis	3.143	-1.366	2.575	-1.421	1.552	-1.638

V. CONCLUSIONS

The paper described the procedures involved in prediction of rainfall using statistical and ANN approaches. From the results of data analysis, the following conclusions were drawn from the study:

[1] Optimum MLP network architectures viz., 1-12-1 for Fatehabad and 1-15-1 for Hansi were used for training the network data.

[2] For Fatehabad, the values of CC, MEF and MAPE between the observed and predicted rainfall (using MLP network) were computed as 0.999, 99% and 6.7% respectively while testing the network data.

[3] For Hansi, the values of CC, MEF and MAPE between the observed and predicted rainfall (using EV1 distribution) were computed as 0.998, 99% and 11.3% respectively.

[4] For Fatehabad, the overall MAPE on the predicted rainfall using MLP and EV1, with reference to recorded rainfall, was found to be 4.5% and 13.2% respectively. For Hansi, the overall MAPE was computed as 15.6% for MLP whereas 8.4% for EV1.

[5] The overall MEF in rainfall prediction using MLP for Fatehabad and EV1 distribution for Hansi was found to be about 99%.

[6] From the values of CC, it was found that there is generally perfect line of agreement between the recorded and predicted rainfall for Fatehabad and Hansi.

[7] Results of model performance analysis and descriptive statistics indicated the MLP is better suited for prediction of rainfall at Fatehabad whereas EV1 for Hansi.

[8] The results presented in the paper would be helpful to the stakeholders for planning, design and management of hydraulic and civil structures in the vicinity of Fatehabad and Hansi rain-gauge stations.

ACKNOWLEDGEMENTS

The author is grateful to the Director, Central Water and Power Research Station, Pune, for providing research facilities to carry out the study. The author is thankful to India Meteorological Department, Pune, for supply of rainfall data used in the study.

REFERENCES

[1] V.K. Somvanshi, O.P. Pandey, P.K. Agrawal, N.V. Kalanker, M. Ravi Prakash, and Ramesh Chand, "Modelling and prediction of rainfall using artificial neural network and ARIMA techniques", *Journal of Indian and Geophysical Union*, vol. 10, no. 2, 2006, pp. 141-151.

[2] X.K. Wang, W.Z. Lu, S.Y. Cao, and D. Fang, "Using time delay neural network combined with genetic algorithm to predict runoff level of Linsham watershed, Sinchuan, China", *Journal of Hydrologic Engineering*, vol. 12, no. 2, 2007, pp. 231-236.

[3] K.P. Sudheer, K. Srinivasan, T.R. Neelakantan, and V. Srinivas, "A nonlinear data-driven model for synthetic generation of annual streamflows", *Hydrological Processes*, vol. 22, no. 12, 2008, pp. 1831-1845

[4] E.J. Gumbel, "Statistic of Extremes", Columbia Univ. Press, New York, 1960.

[5] S. Tokar, and M. Markus, "Precipitation runoff modelling using artificial neural network and conceptual models", *Journal of Hydrologic Engineering*, vol. 5, no. 2, 2000, pp. 156-161

[6] M. Kaltech, "Rainfall-runoff modelling using artificial neural networks: modelling and understanding", *Journal of Environmental Sciences*, vol. 6, no. 1, 2008, pp. 53-58.

[7] M.P. Rajurkar, U.C. Kothyari, and U.C. Chaube, "Modelling of the daily rainfall- runoff relationship with artificial neural network", *Journal of Hydrology*, vol. 285, nos. 1-4, 2004, pp. 96-113.

[8] C. Sarangi, A. Madramootoo, P. Enright, S.O. Prashar, and R.M. Patel, "Performance evaluation of ANN and geomorphology-based models for runoff and sediment yield prediction for a Canadian watershed", *Current Science*, vol. 89, no. 12, 2005, pp. 2022-2033.

[9] L.M. See, A. Jain, R.J. Abrahart, and C.W. Dawson, "Visualization of hidden neuron behaviour in a neural network rainfall-runoff model, Practical Hydroinformatics: Computational intelligence and technological developments in water applications", Springer-Verlag, 2008, pp. 87-99

[10] E. Vamsidhar, K.V.S.R.P. Varma, P. Sankara Rao, and R. Satapati, "Prediction of rainfall using back propagation neural network model", *International*

ISSN: 2347-5552, Volume-5, Issue-4, July 2017
DOI:10.21276/ijircst.2017.5.4.5

Journal on Computer Science and Engineering, vol. 2, no. 4, 2010, pp. 1119-1121.

[11] H.E. Amr, A. El-Shafie, G.E. Hasan, A. Shehata, and M.R. Taha, "Artificial neural network technique for rainfall forecasting applied to Alexandria", *International Journal of the Physical Sciences*, vol. 6, no. 6, 2011, pp. 1306-1316.

[12] R.R. Deshpandey, "On the rainfall time series prediction using multilayer perceptron artificial neural network", *International Journal of emerging technology and advanced engineering*, vol. 2, no. 1, 2012, pp. 2250-2459.

[13] N. Vivekanandan, "Comparison of MLP, CCL and CGT networks for prediction of rainfall", *International Journal of Computer and Communication Engineering Research*, vol. 1, no. 2, 2014, pp. 36-40.

[14] L. Ma, S.Y. Luan, C.W. Jiang, H.L. Liu, and Y. Zhang, "A Review on the forecasting of wind speed and generated power", *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, 2009, pp. 915-920.

[15] Chen, and B.J. Adams, "Integration of artificial neural networks with conceptual models in rainfall-runoff modelling", *Journal of Hydrology*, vol. 318, no. 1-4, 2006, pp. 232-249.

