Journal of Computer Applications Research and Development (JCARD), ISSN 2248-9304(Print), ISSN 2248-9312 (Online), Volume 1, Number 1, April-June (2011)
pp. 27-34 © PRJ Publication
http://www.prjpublication.com/JCARD.asp

JCARD

© PRJ PUBLICATION

RAINFALL PREDICTION THROUGH INTEGRATED MA - ANN MODEL

M. Nirmala[1] and S.M. Sundaram[2]
1. Research Scholar, Sathyabama University, Chennai, India
Email Id: monishram5002@gmail.com
2. Department of Mathematics, Sathyabama University, Chennai, India.
Email Id: sundarambhu@rediffmail.com

ABSTRACT

Rainfall prediction has great significance in planning and understanding the rainfall variability that helps the agricultural management in decision-making process. In this research paper, the traditional data pre-processing technique, moving average was coupled with Artificial Neural Network as MA – ANN model to improve the prediction of rainfall in Tamilnadu. The experimental results show that the MA – ANN hybrid model is a better tool than the Moving Average and ANN models, if they applied separately.

Keywords: Moving Average, Artificial Neural Network, Hybrid model, prediction, Mean Absolute Percentage Error

1. INTRODUCTION

Artificial Neural Networks (ANN) have gained significant attention in past two decades and been widely used for hydrological forecasting [6] and [7]. Many studies focused on rainfall prediction have proven that ANN is superior to traditional regression techniques and time series models including Moving Average, Autoregressive Moving Average and Autoregressive Integrated Moving Average [9], [11] and [14]. Neural Networks provide several advantages over regression prediction techniques. These advantages include Neural Networks require almost no assumptions regarding the underlying data to be forecasted and being able to develop models from incomplete or imperfect data. The combination of statistical methods like Moving Average (MA) and the Neural Network model, which are called hybrid models, to solve real-world problems has recently become a new research field in the application of Neural Networks. Hybrid models have been introduced to obtain more accuracy in prediction when compared to the individual models like ARIMA, Multiple Linear Regression, and Artificial Neural Networks etc. [11] and [14]. The objective of this study is to evaluate the effectiveness of the data pre-processing technique MA in the improvement of the ANN model performance. The integrated hybrid models are obtained by coupling the components of MA with ANN. The different models discussed in this research work for prediction of the Annual Rainfall in Tamilnadu are ANN, MA – ANN1and MA – ANN2.

2. RAINFALL DATA

A dataset containing a total of 136 years (1871 - 2006) monthly rainfall totals of Tamilnadu was obtained from Indian Institute of Tropical Meteorology (IITM), Pune, India. The total of all months rainfall in a year is considered as the Annual Rainfall in this analysis and it is calculated for 136 years.

3. METHODS APPLIED

3.1: Moving Average Method

Moving Average rank among the most popular techniques for the pre-processing of time series. The most common Moving Average method is the unweighted Moving Average. They are used to filter random white noise from the data, to make the time series smoother or even to emphasize certain informational components contained in the time series [3]. The Moving Average method smoothes data by replacing each data point with the average of the K neighbouring data points where K is called the length or order of Moving Average. In general, the larger the order of the Moving Average, the greater the smoothing effect is. Given N data points and T, the order of the moving average, MA(T), the forecast F_{T+1} is given by

$$F_{T+1} = \sum_{i=1}^{T} \frac{X_i}{T} \qquad \text{------------------------------- [1]}$$

$$F_{T+2} = \sum_{i=2}^{T+1} \frac{X_i}{T} \qquad \text{------------------------------- [2]}$$

Combining the equations [1] and [2],

$$F_{T+2} = F_{T+1} + \left(\frac{X_{T+1} - X_T}{T} \right) \qquad \text{------------------------------ [3]}$$

From the above equation, it is clear that each new forecast, F_{T+2} is simply an adjustment of the immediately preceeding forecast, F_{T+1} and Moving Average of higher order provide forecasts that do not change very much.

3.2: Artificial Neural Networks

An Artificial Neural Network is a computational model inspired by biological neural networks both structurally and functionally. It consists of a group of interconnected computation units called neurons. Numerous Artificial Neural Network models have been developed for time series prediction [1] and [10]. Among them, the two major types of Artificial Neural Networks are Multi-Layer Perceptron (MLP) networks and Radial Basis Function (RBF) networks. Multi-Layer Perceptron and Radial Basis Function are feed forward networks. Artificial Neural Networks do not have any stationarity constraint on the time series to be learned and predicted. A Multilayer Perceptron is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. The Multilayer feed forward network consist of a set of sensory units that constitute the input layer, one or more hidden layers of computation nodes, and an output layer of computation nodes. The input signal propagates through the network in a forward direction, on a layer-by-layer basis. Mathematically, a multilayer perceptron network is a function consisting of composition of weighted sum of the functions corresponding to the neurons. The feed forward neural network is iteratively adopted according to the recursion formula,

$$w_{k+1} = w_k + \eta d_k \qquad \text{----------------------------- [4]}$$

where w_k denotes the weight matrix at epoch k. The positive constant η is called the learning rate. The direction vector, d_k is negative of the gradient of the output error function,

$$d_k = -\nabla E(w_k) \qquad \text{----------------------------- [5]}$$

4. APPLICATION OF MODELS TO THE ANNUAL RAINFALL DATA

The figure 1 shows the plot of the Annual Rainfall series of Tamilnadu from 1871 to 2006. For developing training through ANN, the Annual Rainfall data set was splitted into a training data (1871 – 1970) consisting of 100 years and a testing data (1971 -2006) of 36 years. In this research work, three predictors for the year Y are used to predict the Annual Rainfall in Tamilnadu in the year (Y + 1). The three predictors are the Tamilnadu Rainfall amount in the Northeast Monsoon, Tamilnadu Rainfall amount in the Southwest Monsoon and the Annual Rainfall in Tamilnadu. To explore the performance of ANNs, two ANN models are generated using the time series methods and for comparison the original ANN was also considered. The prediction models used in this analysis are ANN, MA – ANN1and MA – ANN2.

4.1: Forecasting Accuracy Procedure

Forecasting accuracy was determined using mean absolute percent error as a measure of forecasting accuracy. The Mean Absolute Percentage Error (MAPE) measure is given by

$$MAPE = 100 \ X \ \frac{\sum_{t=1}^{N} \frac{|E_t|}{Y_t}}{N} \qquad \text{------------------------------ [6]}$$

where Y_t and E_t represents inputs and corresponding errors at t = 1,2,...,N respectively and N denotes the number of forecasting periods.

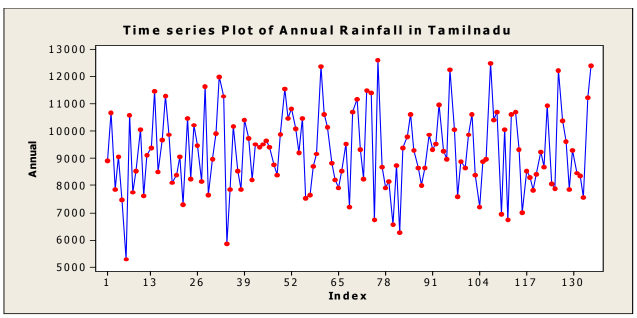
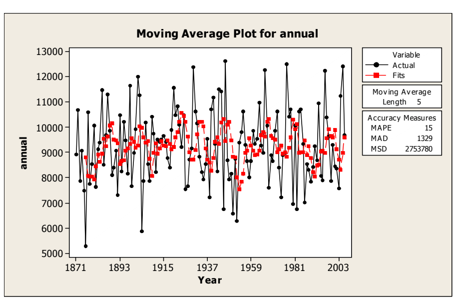
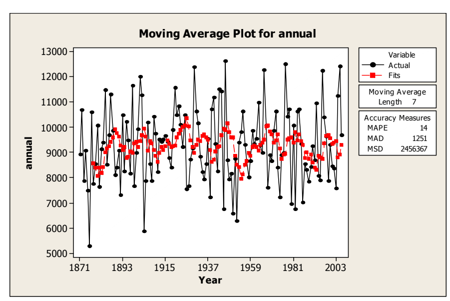


Fig.1: Time series plot of annual rainfall in Tamilnadu (1871 - 2006)

4.2: Moving Average Method

The Moving Average method is applied to smooth the Annual Rainfall data. The Northeast and Southwest Monsoon Rainfall data series are also smoothed by the Moving Average Method.





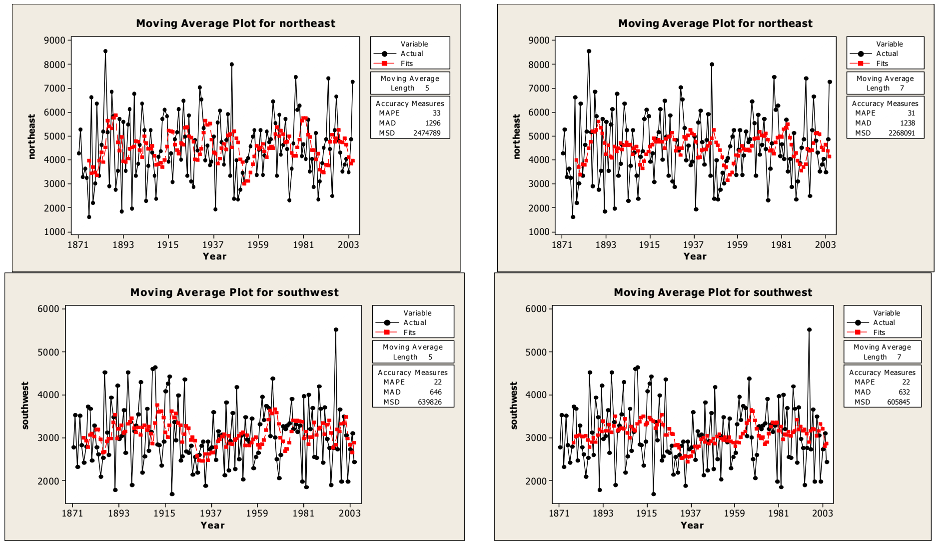


Fig. 2: Moving average plots for annual and monsoons of Tamilnadu rainfall of length 5 and 7.

The order was determined by varying K from 1 to 10 and the best order was identified by using the error measure MAPE, while fitting. Thus, the best K values were found to be 5 and 7 for these Rainfall series. The fig. 2 shows the plots of observed and predicted obtained by Moving Average method of order 5 and of order 7 for the Annual Rainfall series and also for the predictors, Northeast Monsoon and Southwest Monsoon Rainfall series.

4.3: Artificial Neural Networks

For developing an Artificial Neural Network model, the whole dataset is divided into training set consisting of 100 years (1871 - 1970) and testing set consisting of 36 years (1971 - 2006). The two algorithms that are considered for training the Annual Rainfall series are Gradient Descent Algorithm (GDA) and Scaled Conjugate Algorithm (SCG). The original procedure used in the Gradient Descent Algorithm is to adjust the weights towards convergence using the gradient. A search is made along conjugate directions rather in steepest descent directions for faster convergence in Conjugate Gradient Algorithm. The online learning is applied in both the algorithms of MLP in this analysis. A three-layered architecture constructed with three units in input and in the hidden layers and one unit in the output layer in both the MLP and RBF algorithms.

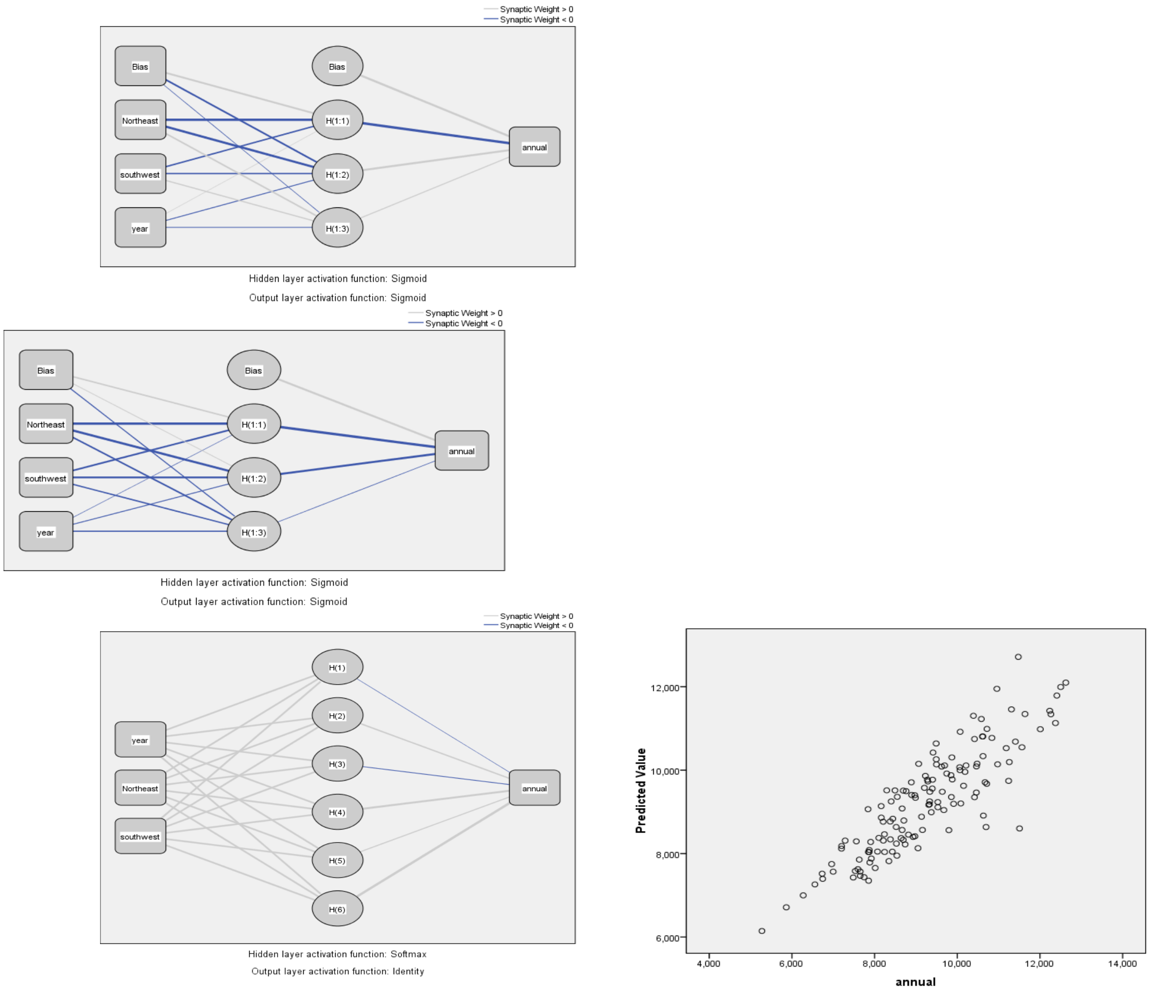


Fig.3: Architecture of ANN model using GDA, SCG AND RBF for Tamilnadu Annual Rainfall series and the graph showing the predicted values of Annual Rainfall in Tamilnadu using RBF

The data was trained upto 1000 epochs. In this analysis, the sigmoidal function is chosen in the hidden and the output layers as the activation function for all algorithms since it is nonlinear and continuously differentiable which are required for network learning. The sigmoid function is given by

$$\sigma(y) = \frac{1}{1 + e^{-y}} \qquad \text{-------------------------- [7]}$$

The various parameters used when applying the algorithms to the Annual Rainfall series is tabulated in table 1. The training was tested using the test data set. The architecture of the algorithms of ANN is shown in figure 3 together with the plot of observed versus predicted values using RBF algorithm. The results of the error measure, MAPE obtained in testing are tabulated in table 2.

TABLE 1: Various Parameters used in the models

	ANN			MA – ANN1			MA – ANN2		
Parameters	**SCG**	**GDA**	**RBF**	**SCG**	**GDA**	**RBF**	**SCG**	**GDA**	**RBF**
Number of Layers	3	3	3	3	3	3	3	3	3
Number of Input Nodes	3	3	3	1	1	1	1	1	1
Number of Output Nodes	1	1	1	2	2	2	2	2	10
Number of Hidden Nodes	3	3	6	1	1	1	1	1	1
Activation function in the Hidden Layer	Sigmoid	Sigmoid	Softmax	Sigm oid	Sigm oid	Softmax	Sigm oid	Sigm oid	Softm ax
Activation function in the Output Layer	Sigmoid	Sigmoid	Identity	Sigm oid	Sigm oid	Identity	Sigm oid	Sigm oid	Identi ty

4.4: Hybrid Models

4.4.1: MA – ANN1 model

In this model, first both the predictors and the dependent rainfall data series are smoothed through Moving Average method of order 5 and are trained through Gradient Descent Algorithm. The number of hidden layer is one and the number of nodes is two. The activation function used in both hidden layer and output layer in GDA is sigmoidal function. The error measure MAPE obtained during training and testing were tabulated in table 2. Next, the data was trained through the SCG algorithm where the training was online. The number of hidden layer is one and the number of nodes in the hidden layer is two. Finally, the data was trained through RBF algorithm. Here also the number of hidden layer is one, which was chosen after training with two hidden layers, and the best MAPE value was found in the case with one hidden layer.

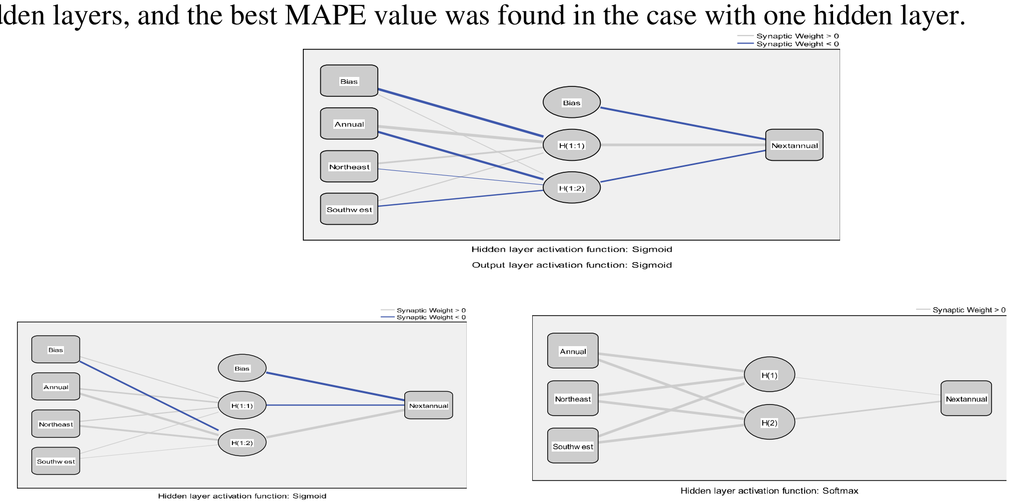


Fig.4: Architecture of MA-ANN1 model using GDA, SCG and RBF for Tamilnadu Annual Rainfall series

The figure 4 shows the architecture of the learning algorithms of ANN when the hybrid model MA – ANN1 was applied to the rainfall series. The various parameters used for training the data through the various learning algorithms in this MA – ANN1 hybrid model are tabulated in the table 1. The MAPE obtained during training and testing is tabulated in the table 2.

4.4.2: MA – ANN2 model

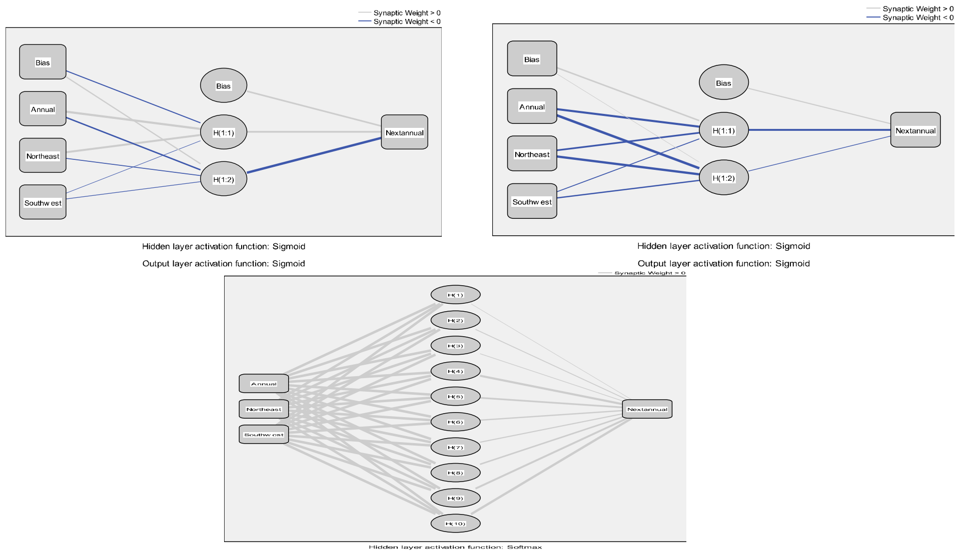


Fig. 5: Architecture of MA-ANN2 model using GDA, SCG and RBF for Tamilnadu Annual Rainfall series

The Annual Rainfall of Tamilnadu together with its predictors was smoothed by the Moving Average method of order 7. Then the smoothed data were trained through the algorithms GDA, SCG and RBF. The architecture of the learning algorithms of MLP and RBF were chosen using the least error measure. The best error measure was obtained with 10 nodes in the hidden layer of RBF whereas for GDA and SCG it was two, which was shown in figure 5. The various parameters used during training processes are given the table 1. The mean absolute percentage error measure obtained during training and testing when the rainfall series were trained is tabulated in the table 2.

TABLE 2: The MAPE values obtained for testing data for the original ANN Model and the hybrid models

	GDA	SCG	RBF
ANN	6.6061	6.5063	0.3926
MA – ANN1	3.8965	4.0176	4.1987
MA – ANN2	2.629948	2.59588	2.29024

5. CONCLUSION

In this study, the conventional ANN model was coupled with the statistical technique Moving Average. As a result, two new hybrid prediction models MA-ANN1and MA-ANN2 including the original ANN model, are proposed to predict the Annual Rainfall in Tamilnadu. The Moving Average of length 5 and of length 7 predicted values for the Annual Rainfall series of Tamilnadu were found for applying these hybrid models. From the error measure MAPE, it can be concluded that the original ANN when compared to the coupled models gave accurate less predicted values.

REFERENCE

[1]. Abraham A, Philip NS and Joseph KB, (2001), Will we have wet summer? Soft Computing models for long – term rainfall Forecasting, Proceedings of 15th European Simulation Conference ESM2001, Prague.

[2]. Anderson J.A. (2003), An Introduction to Neural Networks, Prentice Hall.

[3]. Damodar N. Gujarati and Sangeetha (2007). Basic Econometrics, 4th edition,the Tata Mcgraw – Hill Publishing Company Limited.

[4]. Gadgil, S., Rajeevan, M. & Nanjundiah, R., (2005) Monsoon prediction – Why yet another failure?, Current Science, 88, 1389-1499.

[5]. George E. P. Box, Gwilym M. Jenkins and Gregory C. Reinsel (1994). Time Series Analysis – Forecasting and Control, 3rd Edition, Pearson Education, Inc.

[6]. Hung N Q, Babel M S, Thailand, Weesakul S, and Tripathi N. K, (2008), An Artificial Neural Network model for rainfall forecasting in Bangkok, Hydrology and Earth System Sciences Discussions, 5, 183–218.

[7]. Khaing Win Mar and Thinn Thu Naing , (2008), Optimum Neural Network architecture for Precipitation Prediction of Myanmar, Proceedings of World Academy of Science, Engineering and Technology, 48,130 – 134.

[8]. Moller A F (1993), A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, 6, 525 – 533.

[9]. Nirmala M and Sundaram S.M. (2010), Comparison of Learning Algorithms with the ARIMA Model for the Forecasting of Annual Rainfall in Tamilnadu, International Journal of Information Sciences and Computing, 4(1), 64 – 71.

[10]. Philip N S and Joseph K B, (2001), On the Predictability of Rainfall in Kerala: An Application of ABF Neural Network, Proceedings of Workshop on Intelligent Systems Design and Applications, International Conference on Computational Sciences, San Francisco, 400 – 408.

[11]. Renji Remesan, Muhammad Ali Shamim, Dawei Han, Jimson Mathew (2009), Runoff prediction using an Integrated hybrid modelling scheme, Journal of Hydrology, 372, 48 – 60.

[12]. Spyros Makridakis, Steven C. Wheelwright, Victor E. McGee. Forecasting – Methods and Applications , 2nd edition, 1983.

[13]. Sunyoung Lee, Sungzoon Cho and Patrick M. Wong, (1998), Rainfall Prediction Using Artificial Neural Networks, Journal of Geographic Information and Decision Analysis, 2 (2), 233 – 242.

[14]. Wu C.L., Chau K.W., Li Y.S., (2009), Methods to improve neural network performance in daily flows prediction, Journal of Hydrology, 372, 80 – 93.

[15]. Yegnanarayana, B (2000) Artificial Neural Networks, Prentice-Hall of India Pvt Ltd, India.