International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume No.3 Issue No.8, pp : 1060-1065 1 Aug 2014

Time Series Data Mining in Rainfall Forecasting Using Artificial Neural Network

Prince Gupta[1], Satanand Mishra[2], S.K.Pandey[3]
[1,3]VNS Group, RGPV, Bhopal, [2]*CSIR-AMPRI, BHOPAL*
prince2010.gupta@gmail.com

Abstract: ***Rainfall is very important parameter in hydrological model. Many techniques and models have been developed for rainfall time series prediction. In this study an artificial neural network (ANN) based model was developed for rainfall time series forecasting. Proposed model used Multilayer perceptron (MLP) network with back propagation algorithm for training. Discharge and rainfall data are took as input parameter for ANN model to predict rainfall time series. Data preprocessing and model's sensitivity analysis were executed. Collected data is divided in three sets for optimal neural network training. The first set is the training set, used for calculate the gradient and updating the network weights and biases. The second set is the validation set. The error on the validation set is follow during the training process. The third set is test set. It is used to compare different models. Different topologies of Neural Networks were created with change in hidden layer, number of processing element and activation function. Mean Absolute error (MAE), Mean Squared error (MSE) and correlation coefficient (CC) are used to evaluate the model performance. On the basis of these evaluation parameter results, it is found that multilayer perceptron (MLP) network predict more accurate than other traditional models.***

Keywords: Data mining, Artificial neural network, Back-propagation, Rainfall-runoff prediction.

1.*Introduction:* The forecasting of variable like runoff, rainfall, precipitation have always been an important issue in hydrology. Rainfall-runoff episode are extremely nonlinear, highly complex and shows a high degree of temporal and spatial variability. In hydrological process rainfall prediction becomes a major issue because it affects the human life in many ways. For rainfall-runoff modelling, Data mining techniques have been widely used. Many researchers have applied different kind of data mining techniques and approaches for rainfall-runoff forecasting. Artificial neural network (ANN) is one of the most widely used data mining technique for rainfall forecasting. There are many characteristics and advantage of artificial neural network which make ANN better than the other traditional techniques. ANN shows mapping ability, perform better generalization, learn by example, advantageous and robust system. Artificial neural network was brought in 1943 by Mcculloch et al. [7]. Later it is advanced with the evolving of Backpropagation algorithm for feed forward algoritm [6]. Artificial neural networks are evaluate the trend form data set and predict the result more accurately than other statistical or mathematical model [1]. Lingsrisawang L. et al. examined Decision tree, support vector machine and artificial neural network prediction model for short-term rainfall forecast. The accuracy of classification of these models was compared. [4]. Behzada et al. compare approach of ANN and SNV to predict one day lead flow runoff. By comparing the forecasting result with Support vector machine concluded that the prediction accuracy of ANN is at least as good as that of other models and in some cases better [2]. Mishra et al (2013) presented the analysis based on data mining technique in hydrological daily discharge time series of the panchratna station in the river Brahmaputra under Brahmaputra and Barak Basin Organization in India. K-means, Dynamic Time Wrapping (DTW), and agglomerative hierarchical clustering are used to cluster and discover the discharge pattern in terms of the modelling. [8, 9, 10]. A performance comparison have been done between Support vector regression and multilayer feed-forward neural network models with respect to their forecasting capabilities. The two models have been designed to estimate the relationship between rainfall and runoff, which describes the most complex phenomenon of hydrological science [3]. A performance comparison of three artificial neural network model has been done. These network models are: the multilayer perceptron neural network (MLPNN), the radial basis function neural network (RBFNN) and the simple neural network (SNN).The result of the study showed that performances of all three combination methods are better than that of the best individual rainfall-runoff model [12]. Artificial neural network was one of most important widely used tool for data processing and hydrological forecasting. In this research feed forward neural network trained with Levenberg-Marquardt back propagation algorithm for rainfall forecasting. The correlation coefficient (R), Root mean square error (RMSE), Mean Absolute error (MAE) is implemented to evaluate the performances [5]. Wang, Z. L. and Sheng, H. H. bought in the application of generalized regression neural network (GRNN) model to forecast annual rainfall. The performance of method is compared with the regression analysis method and back propagation neural network. [13].

2.*Study Area:* Narmada basin located at $77^{o}45'$ E and $80^{o}30'$ E longitude and $21^{o}45'$N and $24^{o}00'$N latitude with an area of 98,796 km^2 and length 1312 km. It runs about 1112 km in the Madhya Pradesh. Monthly time series data of discharge and rainfall is collected for seven sites of Narmada basin in Madhya Pradesh. For this study out of the seven Narmada basin sites Hosahangabad is selected. To forecast the rainfall at hoshangabad site, monthly rainfall and discharge data of previous 10 years from 01 January 2000 to 31 December 2010 was used.

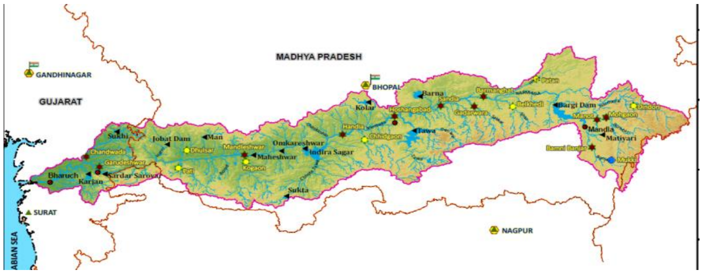


Fig1: Narmada basin in Madhya Pradesh

3.Methodology: Daily basis discharge, rainfall data gathered for Hoshangabad monitoring station located in the Narmada River Basin from Central Water Commission, Ministry of Water Resources and setup the data in compatible format. A machine learning tool Neurosolution 6.2 was used for prediction of rainfall time series. NeuroSolutions is an easy-to-use neural network development tool for Microsoft Windows. It associate advanced learning procedures and genetic optimization with icon based network design interface. Neurosolution make it easy to build customizable neural networks by provide facility of Modify model parameters like hidden layers, the number of processing elements and the learning algorithm which gives best forecasting model.

In this research preprocessing of data was done through normalization. The data was divided in three sets, training, validation and testing sets. Then on normalized data multilayer perceptron neural network with back-propagation algorithm (MLP-BPNN) was implemented by choosing the previous year rainfall and discharge data as an input parameter and next year rainfall data as a desired output parameter. In this study we compare the forecasting accuracy of many kinds of MLP models with linear regression model. Models used in this study for comparison have following naming syntax:
Topology name – Number of hidden layers – learning rule – input projection algorithm.
For example, the model MLP-1-5-B-L is a multilayer perceptron have one hidden layer with 5 neurons and batch method for updating weight is used with Levenberg-Marquardt learning rule.

3.1Artificial Neural Network: A neural network consists of a set of neurons (processing element). Each neuron connected to other neuron by direct link, each link has some weight attached with them. The weight shows knowledge used by network to solve a problem. neurons are arranged in layers: an input layer, an output layer and one or more hidden layers. Information propagates from one layer to other layer by neurons which have input vector and some weight associated with them. At each neuron, the weighted input vector are summed and a threshold value θ_j is added. This added input I_j is then go through a activation function (a non-linear function) $f(.)$ to produce the output of the neuron y_j. The output of one neuron provides the input to the neurons in the next layer. Mathematically it is represented as:

$$I_j=\sum w_{ji}x_i+\theta_j \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$$y_j=f(I_j) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

There are a wide variety of neural network model and learning procedure. In this study, Multilayer perceptron (MLP) neural network was used. MLP is most important neural network. It use linear combination function in input layer to calculate single output from multiple inputs and then apply nonlinear activation function on generated output. The general structure of MLP is given in fig 2. The first layer is called input layer which have input neuron. Each input neuron represents an input data. Second layer is called hidden layer. MLP may have more than one hidden layer. The last layer is called output layer which have output neurons. Output neurons consists the predicted value. Mathematically this is representing as:

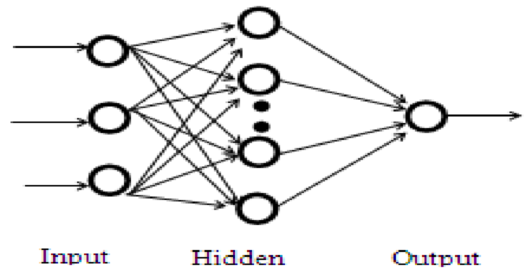


Fig 2: Artificial neural network structure

$$Y = \varphi\left(\sum_{i=1}^{n} W_i X_i + b\right)\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Where W_i denotes the vector of weights, X_i is the vector of inputs (i =1, 2..n), b is the bias, y is the output and φ is the activation function[11]. MLP is trained using back propagation algorithm. Back propagation algorithm can be applied in many ways. At the beginning of training the network weights are initialized with random value. This algorithm works iteratively by updating a network's interconnecting weights such that the overall error is reduced. MLP can be used where you have little knowledge of relationship between inputs and targets.

Artificial neural network is a generalization computational model of central nervous system. Neural network has capability to learn from experience to raise the performance of model. Artificial neural networks (ANN) have strong ability of pattern recognition and classification and forecasting from their learning experience. Artificial neural network (ANN) use the concept of nonlinear mapping which is useful when the data given is incomplete and noisy where rules cannot be defined.

3.2Back Propagation Algorithm: Back propagation algorithm is used to train artificial neural networks. It is a supervised learning method, i.e., a teacher is required to calculate the desired output for any input in the training set. The goal of any supervised learning algorithm is to find a function that best maps a set of inputs to its correct output. As the algorithm's name implies, the errors propagate backwards from the output nodes to the input nodes. Since this algorithm is based on the supervised learning approach, therefore the desired result is already known to the network. We then calculate the error of each neuron, which is computed – desired output. This error is then back propagate to change the weights in such a way that the error will get smaller. In order for the hidden layer to serve any useful function, multilayer networks must have non-linear activation functions for the multiple layers. Back propagation algorithm requires that the activation function used by the artificial neurons be differentiable. This algorithm operates in either of the 2 modes: Incremental mode in which each propagation is followed immediately by the weight adjustment and batch mode in which the weight updates take place after various consecutive propagations

ALGORITHM:

1) Randomly initialize the weights in the network.

2) Apply the input to the network and obtained computed output

3) Calculate the error (e)

e = desired-computed

4) Calculate the Δw_i for all weights in backward pass from hidden layer to output layer.

5) Calculate the ΔW_i for all weights in backward pass from input layer to hidden layer .

6) Update the weights in the network

7) Repeat the step 2 to 6 for every training pattern until all pattern classified correctly.

A flowchart of methodology has been drawn.

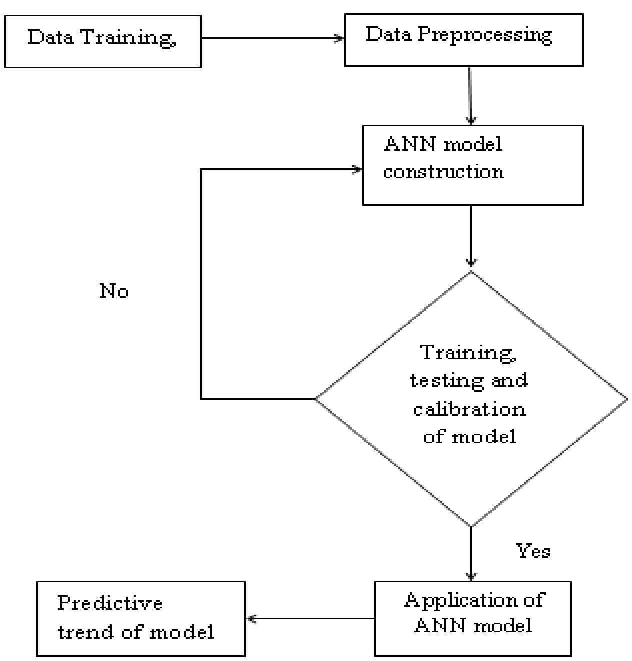


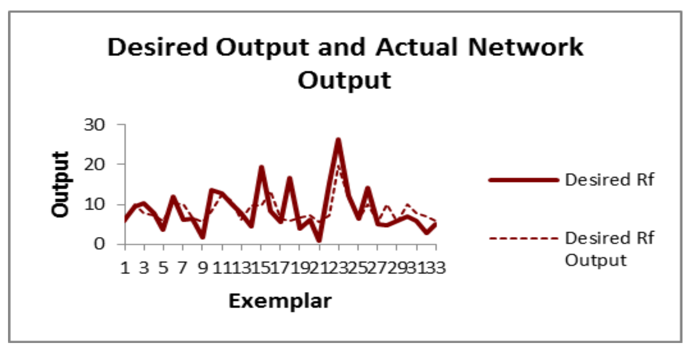
Fig 3 : Flow chart of process

3.3 Design of data set: Hoshangabad is located at 22.75^0N 77.72^0E. An average height from the sea level is 331 meters and average rain fall is 134 cms. Discharge and rainfall data is collected for Hoshangabad site from Central Water Commission, Ministry of Water Resources India. The data set is divided in three subsets : Training, validation and testing sets. Input data contain four months (June to September) rainfall and discharge mean data of year's 2000-2010.The data set contain 43 exemplar

Performance	Desired Rf
MSE	17.87293619
NMSE	1.014895788
MAE	3.359021882
Min Abs Error	0.20057731
Max Abs Error	7.18416321
r	0.456888021

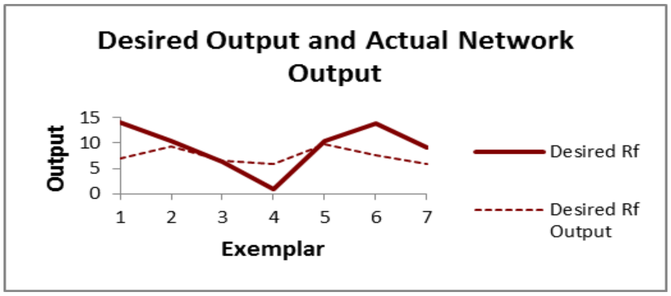
for the year's 2001 to 2010. In this study, 75% of data is used for training, 15% of data is used for validation and 10% data is used for testing.

4.Results: In this study, discharge and rainfall data of year 2000-2010 were used to trained the network. Rainfall and discharge data as an input parameter and next year rainfall data as a desired output parameter. Various experiments were perform on network's structure and algorithm by change in number of hidden layer, number of neurons, learning algorithm and activation function. After performing all experiments, the model with 2 hidden layer along with online update method for updating weight is used with momentum learning rule, found optimum for this research work. Fig4 shows the result obtained from training, cross validation and testing phase of the model.



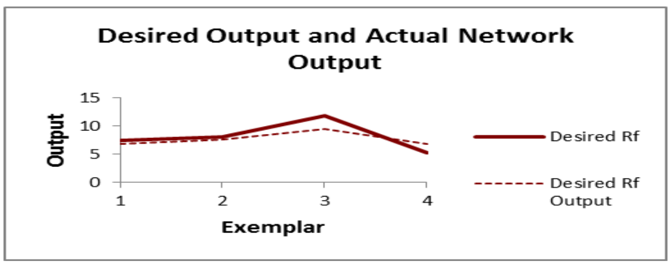
Performance	Desired Rf
MSE	16.479802
NMSE	0.583784492
MAE	2.973255089
Min Abs Error	0.020170507
Max Abs Error	10.67431173
r	0.653898533

Fig4(a). Desired versus actual network output at training stage and error measures



Performance	Desired Rf
MSE	2.08340233
NMSE	0.381095659
MAE	1.238887023
Min Abs Error	0.422597333
Max Abs Error	2.27064122
r	0.936043821

Fig4(b). Desired versus actual network output at cross validation stage and error measures



Performance	Desired Rf
MSE	2.08340233
NMSE	0.381095659
MAE	1.238887023
Min Abs Error	0.422597333
Max Abs Error	2.27064122
r	0.936043821

Fig4(c) Desired versus actual network output at testing stage and error measures.

5.performance Evaluataion and Comparision: To measure the performance of this research, statistics of correlation coefficient, Mean absolute error (MAE), Mean Square Error (MSE) are used. If value of Correlation coefficient is more, than it shows strong correlation between the parameter. This study compared four MLP neural network model with two linear regression models to find the best model for rainfall forecasting. The best model has minimum MSE and MAE. MLP-2-O-M (multilayer perceptron) found the best model among others model. It has minimum MSE, MAE and maximum correlation coefficient value. Table 1 shows the summary of all network performance.

6.Conclusion and Future Work: Current results using MLP with a Backpropagation yield a low mean square error and mean absolute error as compare to linear regression for rainfall forecast in table 1. This is significant improvement over the current forecasts and yields a good model for producing future forecasts. This result clearly indicating that artificial neural network approach is a more convincing and relatively predicting more accurate simulation results than linear regression.

our most promising line of future work is to apply our methods on other Narmada Basin Sites in Madhya Pradesh.

7.Acknowelegement: The authors would like to thank the Central Water Commission, Ministry of Water Resources, India for providing Water Level and Discharge data.

References:

i. Abhishek, K., *Kumar, A., Ranjan, R., and Kumar, S. (2012), A Rainfall Prediction Model using Artificial Neural Network, IEEE*

ii. *Behzad M., Asghari K., Eazi M. and Palhang M., Generalized performance of SVM and NN in runoff modelling, 2009, ELSEVIER SCIENCES, Expert System with Application, Vol. 36, Issue 4, pp 7624-7629*

iii. *Burbridge Robert and Buxton Bernard, An Introduction to Support Vector Machines for Data Mining, UCL, Gower Street, WC1E 6BT, UK.*

iv. *Ingsrisawang L., Ingsrisawang S., Somchit S., Aungsuratana P. and Khantiyanan W. (2008), Machine Learning techniques For Short-Term rain forecasting System in the northeastern part of Thailand, World Academy of Science, Engineering and Technology, vol:2*

v. *Kumar, R., and Yadav, G. S. (2013), Forecasting of Rain Fall in Varanasi District, Uttar Pradesh Using Artificial Neural Network, JECET.*

vi. *Mahdizadeh, M.B., (2004), Artificial Neural Networks and their Application in Civil Engineering, !st Edn., Ebady publication , iran, ISBN: 964-6531-35-0,pp:192*

vii. *McCullagh, J., Bluff and E. Ebert, (1995), A neural network model for rainfall estimation, artificial neural network and expert system, proceeding of 2nd New Zealand International two stream conference on artificial neural networks and expert system.*

viii. *Mishra S., Majumder S., and Dwivedi V.K.* **,** *pattern discovery in hydrological time series data mining in a Sustainable Water resources Management And Climate Change Adaptation, Vol.-II, pp.107-115, February 17-19, 2011, NIT Durgapur.*

ix. *Mishra S., Dwivedi V.K. , Sarvanan C. and Pathak K. K., Pattern Discovery in Hydrological Time Series Data Mining during the Monsoon Period of the High Flood Years in Brahmaputra River Basin, IJCA(0975-8887), doi-10.5120/11397-6698, Vol. 67, No.6, pp.7-14,April, 2013.*

x. *Mishra S., Sarvanan C., Dwivedi V.K. , and Pathak K. K., Discovering Flood Rising Pattern in Hydrological Time Series Data Mining during the Pre Monsoon Period, Indian. Journal of Geo-Marine Science, Accepted on 12/01/2014.*

xi. *Mutlu, E., Chaubey, I., Hexmoor, H., and Bajwa, S. G. (2008), Comparison of artificial neural network models for hydrologic* *predictions at multiple gauging stations in an agricultural watershed, Hydrological Process.*

xii. Shamseldin, A. Y., O'CONNOR, K. M., and Nasr, A.E. (2007), A comparative study of three neural network forecast combination methods for simulated river flows of different rainfall–runoff models, Hydrological Sciences Journal.

xiii. Wang, Z. L., and Sheng, H. H. (2010), Rainfall Prediction Using Generalized Regression Neural Network: Case study Zhengzhou,IEEE

Table 1: Summary of all networks

Summary of All Networks

Performance Metrics

	Training			Cross Validation			Testing		
Model Name	**MSE**	**r**	**MAE**	**MSE**	**r**	**MAE**	**MSE**	**r**	**MAE**
MLP-1-O-M (Multilayer Perceptron)	16.48854	0.66652	3.033368	19.4035	0.458292	3.639182	3.672059	0.74303	1.700497
LR-0-B-M (Linear Regression)	30.33534	0.075114	4.305287	14.72496	0.530509	2.934238	4.586494	0.508656	1.828986
LR-0-B-L (Linear Regression)	23.69843	0.400626	3.700809	23.84557	-0.11621	4.229682	10.53255	-0.50956	2.418188
MLP-1-B-M (Multilayer Perceptron)	30.26885	0.059556	4.341953	14.949	0.507207	2.933167	4.900828	0.529443	1.736907
MLP-2-O-M (Multilayer Perceptron)	16.4798	0.653899	2.973255	17.87294	0.456888	3.359022	2.083402	0.936044	1.238887
MLP-2-B-M (Multilayer Perceptron)	27.04729	0.246192	4.213447	20.22151	-0.33849	3.65608	8.171495	-0.50813	2.459326