IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Runoff Prediction Using (Ann) Based Multiple Layer Perceptron (Mlp) And Multi Linear Regression (Mlr) Technique For Meghalaya District

Akhand Pratap Singh, Dr. Vikram Singh, Sonali Kumari

Abstract

Rainfall-runoff modeling is one of the most important topics in water resources planning, development and management on sustainable basis. The Multilayer Perceptron (MLP) and Multiple Linear Regression (MLR). This study was undertaken to develop and evaluate the applicability of the MLP and MLR models by way of training and testing of developed models during monsoon period (June to September) for Meghalaya district of Assam state of India. The daily data of rainfall, runoff (or stream flow), minimum & maximum temperature and wind speed were used in the study for monsoon season. The daily data were split into two sets: a training data set from 2013 to 2020 and a testing data set from 2021 to 2022 for Meghalaya Assam. The NeuroSolution 6.0 software and Microsoft Excel were used in an analysis and the performance evaluation indices for developed models, respectively. The best input combination of rainfall, runoff minimum & maximum temperature and wind speed were identified using the input-output combination for the simulation of Runoff. On the basis input combination, 10 best models for runoff were selected out of 15 models, respectively with different input combinations. The input pairs in the training data set were applied to the network of a selected architecture and training was performed using back propagation algorithm for MLP models. A number of networks were constructed and each of them was trained separately, and the best network was selected based on the accuracy of the predictions in the testing phase. The following statistical indices such as mean squared error (MSE), coefficient of efficiency (CE), coefficient of determination (R^2) and coefficient of correlation (r) were applied to test the performance of the developed MLP and MLR models. The predicted suspended sediment using MLP models were found to be the best performing models for Meghalaya Assam. It was clearly evident that MLR models fit very poorly for the dataset under study. The current day's runoff can be simulated using the data of current day rainfall.

Keyword: Soft computing, MLP, MLR, Runoff prediction

INTRODUCTION

Water is an essential and precious gift by the nature to the creatures on the earth in the form of rain, snow, rivers, seas, oceans etc. Rainfall is one of the most complex and difficult elements of the hydrology cycle to understand and model due to the complexity of atmospheric processes in its formation and occurrence. No one can imagine the existence of life on the planet earth without water. India receives annual precipitation (rain and snow) of about 4000 km^3, out of which, rainfall during monsoon is of the order of 3000 km^3 (75%). Rainfall in India is dependent on the south-west and north-east monsoons based on shallow cyclonic depressions and disturbances, and on the local storms. Most of the rainfall takes place under the influence of south-west monsoon during June to September, except in Tamil Nadu, where it is under the influence of north-east monsoon during October to November. In urban areas, rainfall has a strong influence on sewer system, water logging , traffic and other human-induced activities. With the advancements in computer technology and geographic information system (GIS), it is now possible to improve spatial interpolations of precipitation measurements.

A large number of rainfall-runoff models with different levels of complexity have been developed for scientific and operational applications. However, depending on the specific purpose and study region, a good model performance is not always assured. A rainfall-runoff model on catchment is a mathematical model producing the surface runoff hydrograph as the response (output) to a rainfall hyetograph as input.
The Multi-layer perceptron (MLP) is the most popular (ANN)architecture in use today **(Dawson and Wil by, 1998)**. As a network, MLP was formed by simple neuron called perceptron, which computes a single output from multiple real valued inputs by forming a linear combination according to its input weights and then possibly expressing the output through an online transfer function. The MLP is a Widely used (ANN)configuration that has been frequently applied in the field of hydrological modeling **(Leahy *et al*. 2008; Tabari *et al.* 2010b; Zadeh and Tabari 2012).** Generally, the MLP consists of three layers of neurons: an input layer, output layer and intermediate or hidden layer. Each neuron has a number of inputs and a number of outputs. A neuron computes its output response based on the weighted sum of all its inputs according to an activation function

REVIEW AND LITERATURE

Anurag Malik *et al*. (2018) Radial Basis Neural Network (RBNN), Self- organizing map neural network (SOMNN), and Multiple Linear Regression (MLR) were used for the estimation of pan evaporation at Pantnagar located at the foot hills of Himalayas in the Uttarakhand, India . Daily climatic data & pan evaporation data were used for model calibration & validation. Combination of significant input variables for RBNN, SOMNN & MLR models were deci.ded using Gamma test. Results obtained by models were compared with climate-based Emperical models such as Penman, Stephens-Stewart and Jensen-Burman-Allen models on the basis of Root mean squared Error (RMSE), Coefficient of efficiency (CE) & Correlation Coefficient(r).

AnuragMalik*et al*. (2019) In this research work**,** four heuristic methods**,** namely, radial basis neural network(RBNN), self-organizing map neural network(SOMNN), least square support vector regression(LSSVR), and multivariate adaptive regression spline(MARS) were employed for daily SSC modeling at Asthi, Bamini, and Tekra stations located in Godavari River basin, Andhra Pradesh, India. The Gamma Test (GT) was utilized for identifying the most significant input variables for the applied heuristic approaches. The results obtained by RBNN, SOMNN, LSSVR, and MARS models were compared with those of the traditional sediment rating curve (SRC). The results of comparison revealed that the four heuristic methods gave higher accuracy than the SRC model.

Md. AjazAlam and Vikram Singh (2019) In this research work, co-active neuro fuzzy inference system (CANFIS) and Multi Linear Regression (MLR) techniques were used to forecast the daily rainfall. Fifteen CANFIS models and MLR were selected based on the performance evaluation indices during testing period. Results showed that CANFIS model was better as compared to MLR.

TusharRathodetal. (2022) This paper describe the evaluation &application of Co-Active Neuro Fuzzy Interface System(CANFIS) model to simulate rainfall from a watershed of Umargaon area in Nagpur, Maharashtra, India. The Neuro Solution 5.0 software & Microsoft Excel were used in analysis & performance evaluation of developed models, respectively. The result indicated that the CANFIS is suitable for rainfall prediction.

Mirabbasi*etal*.(2023) determined the best input combinations using Gamma test for runoff prediction. As well as gamma test was applied to determine the optimum length of training and testing datasets. Results showed the input combination based on the gamma test method has better performance than other combinations.

MATERIALS AND METHODS

This chapter deals with the study area, data acquisition and methodology adopted suspended sediment simulation using artificial neural networks for Meghalaya Assam at Meghalaya, Assam India, procedure used for calibration and validation of the model and various criteria for evaluating performance of the model are also discussed here.

Study Area

Meghalaya is bounded by Assam on its east, north and north-west and by Bangladesh on the south and south-west. The Khasi Hills and Jaintia Hills which form the central and eastern parts of Meghalaya are an imposing plateau with rolling grassland, hills and river valley. One of the seven sister states of the north-eastern part of the country, Meghalaya has a geographical area of 22,429 km².

In this study 10 year daily rainfall, minimum & maximum temperature, wind speed and runoff data of period of monsoon (June to September) of the year 2013 to 2022 at Meghalaya Assam Catchment from Central Water Commission (CWC) Meghalaya.

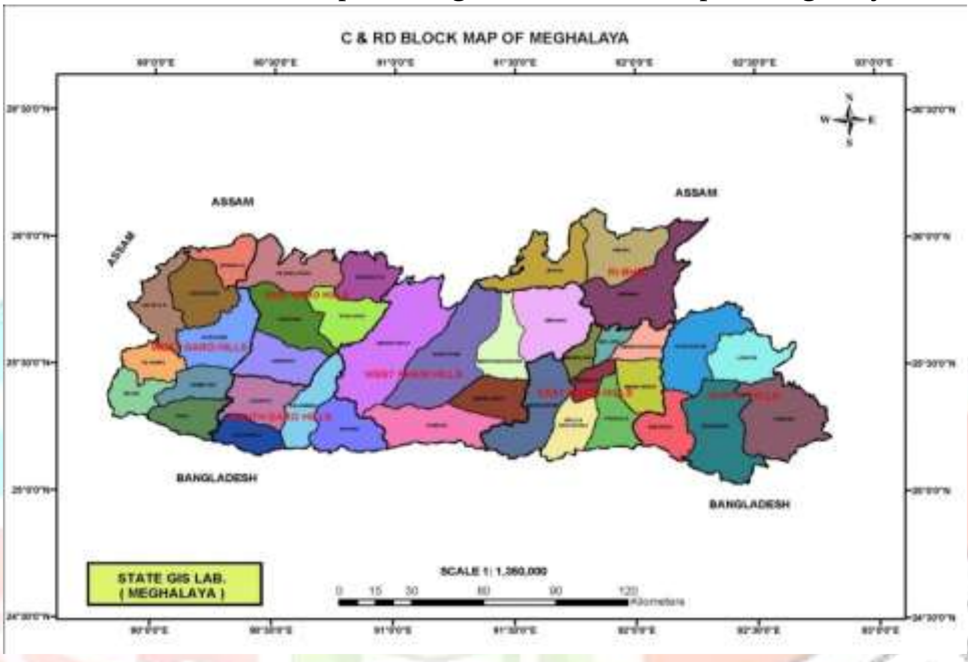
Climate and topography

The climate in Meghalaya Assam was the wet season is oppressive and mostly cloudy, the dry season is mostly clear, and it is hot year.The climate of Meghalaya is generally very humid. It is directly influenced by the south west monsoon and the north east winter winds

Land use pattern

Land Use Statistics, Ministry of Agriculture, GOI, 2008-09. The recorded forest area of the state is 9,496 km² which is 42.34% of its geographical area. The Reserved Forests constitute 11.72%, Protected Forests 0.13% and Unclassed Forests 88.15%.which involves indigenous, low yielding and high value crops.

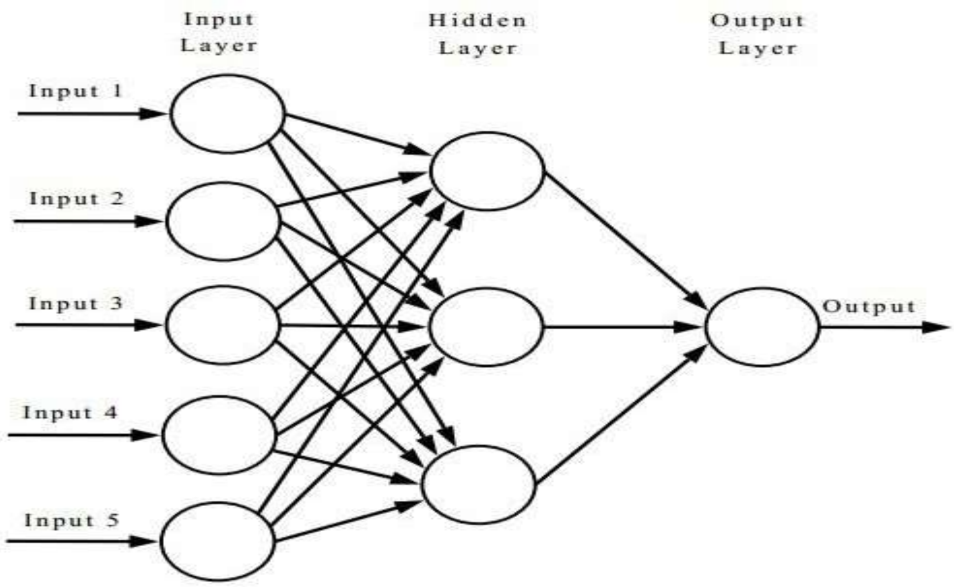
Rice is the most important food crop and occupies about 44% of the total agricultural land. Maize is the next important agricultural food crop of Meghalaya



Location map of Meghalaya, Assam

Learning in MLP

An important step in developing an MLP model is the determination of its weight matrix through training. There are two types of learning, supervised and unsupervised. In supervised learning, a set of input pattern and its known output pattern is used to train the network. An external teacher finds the error between computed output and desired output during the training and this error is used to make adjustments in the weights to minimize the error. On the other hand, there is no teacher present to train the patterns in unsupervised learning. Here the system learns of its own detecting regularities in the input space through correlation, without direct feedback from the teacher. Supervised learning is used in the applications involving classification, functional mapping. Whereas unsupervised learning is employed in clustering type of applications. As the present work involves functional mapping, it is mandatory to use the supervised learning.



A Basic overview of MLP

Back –propagation of training algorithm

Feed forward back propagation (FFBP) was selected for rainfall-runoff modeling as it is the most commonly used (ANN)approach in hydrological predictions and in approximating nonlinear functions. The FFBP is a supervised learning technique used for training set of input-output data, the most common learning rule for multi-layer perception is the back-propagation algorithm (BPA). This involves two phases: a feed-forward phase in which the external input information at the input nodes is propagated forward to compute the output information signal at the output unit, and a backward phase in which modifications to the connection strengths are made based on the differences between the computed and observed information signals at the output units **(Eberhart and Dobbins, 1990)**. The neural network structure in this study possessed a three-layer learning network which consists of an input layer, a hidden layer and an output layer. In NeuroSolution 6.0 the Neural Builder offer 6 learning rules, such as step, Momentum, Delta-Bar-Delta (DBD),

Description of fifteen MLP model and corresponding output and input parameters for prediction of daily runoff

Model No.	Output-Input Variables*
MLP-1	$Q_t=f\,(T_{max})$
MLP-2	$Q_t=f\,(T_{min})$
MLP-3	$Q_t=f\,(WS)$
MLP-4	$Q_t=f\,(R_t)$
MLP-5	$Q_t=f(T_{max},T_{min})$
MLP-6	$Q_t=f\,(T_{max},WS)$
MLP-7	$Q_t=f\,(T_{max},R_t)$
MLP-8	$Q_t=f(T_{min},WS)$
MLP-9	$Q_t=f(T_{min},R_t)$
MLP-10	$Q_t=f\,(WS,R_t)$
MLP-11	$Q_t=f(T_{max},T_{min},WS)$
MLP-12	$Q_t=f(T_{max},T_{min},R_t)$
MLP-13	$Q_t=f\,(T_{max},WS,R_t)$
MLP-14	$Q_t=f\,(T_{min},WS,R_t)$
MLP-15	$Q_t=f\,(T_{max},T_{min}, WS,R_t)$

Output-Input combinations for MLR models for runoff simulation

Model No.	Output-Input Variables*
MLR-1	$Q_t=e_1+ g_1T_{max}$
MLR-2	$Q_t =e_2+ h_1T_{min}$
MLR-3	$Q_t=e_3 + k_1WS$
MLR-4	$Q_t= e_4 + l_1R_t$
MLR-5	$Q_t =e_5+g_2T_{max}+h_2T_{min}$
MLR-6	$Q_t=e_6 + g_3T_{max}+k_2WS$
MLR-7	$Q_t=e_7+g_4T_{max}+ l_2R_t$
MLR-8	$Q_t=e_8+h_3T_{min}+ k_3WS$
MLR-9	$Q_t=e_9+h_4T_{min}+ l_3R_t$
MLR-10	$Q_t=e_{10} + k_4WS +l_4R_t$
MLR-11	$Q_t=e_{11}+ g_5T_{max}+ h_5T_{min}+k_5WS$
MLR-12	$Q_t =e_{12}+g_6T_{max}+h_6T_{min}+l_5R_t$
MLR-13	$Q_t=e_{13} +g_7T_{max}+k_6WS+ l_6R_t$
MLR-14	$Q_t=e_{14}+h_7T_{min}+k_7WS+l_7R_t$
MLR-15	$Q_t=e_8+ g_8T_{max}+ h_8T_{min}+ k_8WS+ l_8R_t$

*e_i,g_i,g',h_i and h_i'are regression coefficients (i=1,2,….,15)

Training and testing of MLP and MLR models

The daily data of rainfall, runoff (or discharge or streamflow) minimum & maximum temperature and wind speed were split into two sets: a training data set from 2013 to 2020and a testing data set from 2021 to 2023 for Meghalaya Assam. The input pairs in the training data set were applied to the network of a selected architecture and training was performed using back propagation algorithm for MLP models

Statistical Indices

Correlation coefficient(r)

The correlation coefficient (r) is an indicator of degree of closeness between observed and predicted values. If observed and predicted values are completely independent, the correlation coefficient will be zero (Mutreja, 1992). The correlation coefficient is estimated by the following equation:

$$r=\frac{\sum_{j=1}^{n}\left\{\left(Y_j-\bar{Y}\right)\left(Y_{ej}-\bar{Y}_{ej}\right)\right\}}{\sum_{j=1}^{n}\left(Y_j-\bar{Y}\right)^2\sum_{j=1}^{n}\left(Y_{ej}-\bar{Y}_{ej}\right)^2}$$

Y= mean of observed values and Y_{ej}= mean of predicted values

Y_j= observed values and Y_{ej}= predicted values

n= number of observations

Coefficient of efficiency (CE)

Coefficient of efficiency (CE) has been recommended by many researchers in the field of hydrology for evaluating the model performance (Mutreja, 1992 and Basu, 1993). They reported that the value of CE should be at least 80% or so for a model to be accepted. A negative efficiency represents that the predicted values are less than the observed mean. The value of CE is determined by the following equation:

$$CE = \left[1 - \frac{\sum_{j=1}^{n} (Y_j - Y_{ej})^2}{\sum_{j=1}^{n} \left(Y_j - \bar{Y} \right)^2} \right] \times 100\%$$

Mean square error (MSE)

The mean square error (MSE) is determined to measure the prediction accuracy. It always produces positive values by squaring the errors. The MSE is zero for perfect fit and increased values indicate higher discrepancies between predicted and observed values (Wilks, 1995). The Mean Square Error between observed and predicted values is determined by the following equation:

$$MSE = \frac{\sum_{j=1}^{n} (Y_j - Y_{ej})^2}{n}$$

RESULTS AND DISCUSSION

This chapter deals with the development and application of (ANN)and MLR model to simulate rainfall-runoff of Meghalaya Assam at Meghalaya Catchment, Assam. The subsequent use in training and testing during year 2013-2020 for the training or calibration of the model, where as the data of year 2021-2022 were used for verification or testing of the development model. Out of 1170 datasets, 926 (80%) were used for training and 244 (20%) were used for testing in (ANN)based software.

The performance of the development model were evaluated qualitatively and quantitatively by the visual observation, and based on various statistical and hydrological indices such as correlation coefficient (r), coefficient of efficiency (CE) and mean squared error (MSE). The thirty one model having high values of rand CE and lower values of MSE is considered as the better fit model.

Statistical indices for MLP models for runoff simulation during training & testing phase for Meghalaya Assam

Sr. No.	Model No	Structure	Testing			
			CE	r	R2	MSE
1	MLR4	1-10-1	0.9631	0.9950	0.9951	0.0313
2	MLR 10	2-10-1	0.9625	0.9726	0.9726	0.0321
3	MLR 3	1-10-1	0.9616	0.9955	0.9955	0.0326
4	MLR 1	1-10-1	0.9607	0.9965	0.9965	0.0328
5	MLR 12	3-4-1	0.9600	0.9674	0.9672	0.0328
6	MLR 6	2-10-1	0.9599	0.9913	0.9911	0.0342
7	MLR 7	2-10-1	0.9599	0.9794	0.9933	0.0345
8	MLR 2	1-10-1	0.9531	0.9958	0.9958	0.0345
9	MLR 5	2-10-1	0.9525	0.9941	0.9940	0.0347
10	MLR 15	4-10-1	0.9524	0.9760	0.9760	0.0352

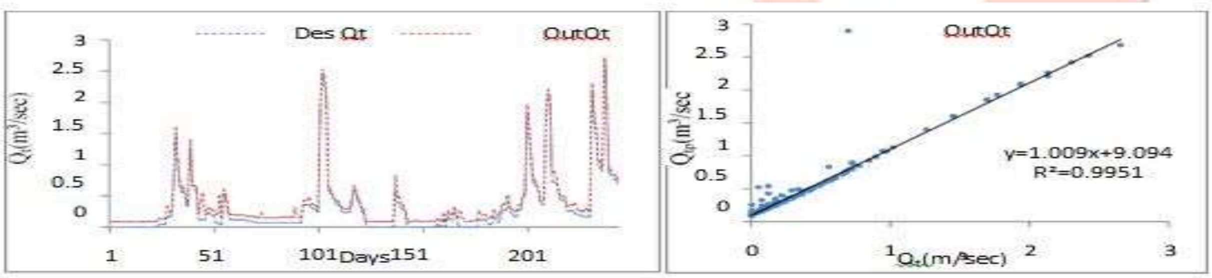


Figure 4.1 Comparison of observed (Q_{to}) and predicted (Q_{tp}) runoff and corresponding scatter plot in testing MLP-4 (4-10-1) model for Meghalaya Assam.

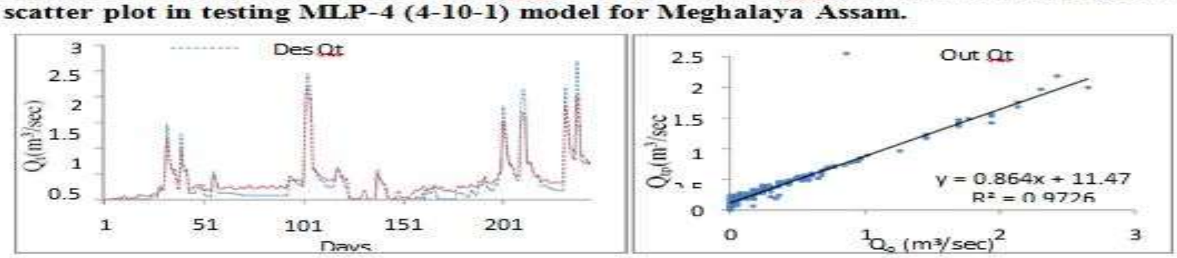


Figure 4.2 Comparison of observed (Q_{to}) and predicted (Q_{tp}) runoff and corresponding scatter plot in testing MLP-10 (10-10-1) model for Meghalaya Assam.

Runoff Modeling using MLR

Multiple linear regression (MLR) models were also used to simulate runoff of current day for study area. The combinations of input variables in MLR models (Table 3.5) consist of the same meteorological variables as used for MLP models. Table 4.2 gives the values of performance evaluation indices for ten MLR models selected in terms of MSE, CE, R^2 and r values during the period of 2013 to 2022. For the selected models (MLR-3, MLR-10, MLR-4, MLR-1,MLR-6,MLR-8,MLR-13,MLR-7,MLR-15andMLR-12), the MSE varied from 18138.13 to 19960.70, CE Varied from 0.0306 to 0.0867; R^2 varied from 0.0328 to 0.0904 and r varied from 0.1810 to 0.3007.

Statistical indices for selected MLR runoff models during testing phase for Meghalaya Assam

Model No.	Regression equation	Statistical index			
		MSE	CE	r	R^2
MLR3	Q_t = 331.2874 + (-4.1500 * Tmax) + (-6.5762* Tmin)+ (1.5260 * WS)+ (0.7464 * Rt)	432.025	-733.569	-0.2990	0.089
MLR 10	Q_t= 325.5376+(-4.2257* Tmax)+ (-6.0549* Tmin) + (0.7511 * Rt)	1764.552	-2999.253	0.134	0.018
MLR 4	Q_t= 236.7774 +(-5.7944* Tmax) + (-2.6071 * WS)+(0.6809 * Rt)	1965.456	-3340.850	0.187	0.035
MLR 1	Q_t= 232.9875 +(-5.9133* Tmax) + (0.6593 * Rt)	2237.494	-3803.393	-0.344	0.118
MLR 6	Q_t =263.7965 +(-6.7291 * T_{max})	2353.608	-4000.821	-0.187	0.035
MLR 8	Q_t= 264.9888 +(-6.7062* Tmax) + (-0.6480*WS)	3032.913	-5156.904	-0.461	0.213
MLR 13	Q_t= 336.3303 +(-5.5640* Tmax) + (-4.8255* Tmin)+ (2.5232 *WS)	3070.815	-5220.283	0.091	0.008
MLR 7	Q_t=326.8621 +(-5.7045* Tmax) + (-3.9416*Tmin)	3168.455	-5386.300	0.106	0.011
MLR 15	Q_t = 375.9319 +(-14.2515* Tmin) + (4.3274* WS) +(0.9935 *Rt)	3538.991	-6016.319	0.049	0.002
MLR 12	Q_t = 66.2292 +(-7.8604* WS) + (1.1240 * Rt)	3568.954	-6067.265	0.040	0.002

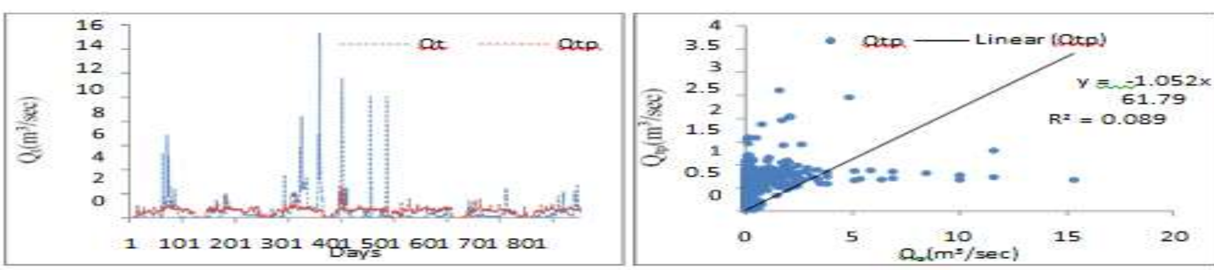


Figure4.11 Comparison of observed (Q_{to}) and predicted (Q_{tp}) runoff and corresponding scatter plot by MLR-3 model for Meghalaya Assam

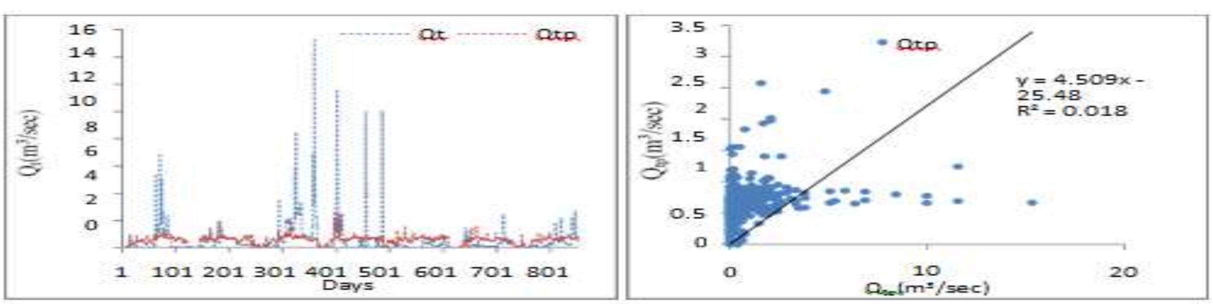


Figure4.12 Comparison of observed (Q_{to}) and predicted (Q_{tp}) runoff and corresponding scatter plot by MLR-10 model for Meghalaya Assam

CONCLUSIONS

The different performance evaluation indices were applied to test the performance of the developed models. The performance evaluations of the models was made on the basis of visual observation through runoff and scatter graphs between observed and predicted runoff, followed by statistical indices such as mean square error (MSE), coefficient of efficiency(CE) and coefficient of correlation (r).

The following conclusions were drawn from the results of this study:

- Runoff can be prediction by using MLP model with input parameters as maximum temperature, minimum temperature, wind speed and rainfall.
- On the basis of lower MSE value and higher CE and r values, MLP-4model was found to be the best model followed by MLP-6 model. According to the MLP-4 model, current day's runoff depends on current day's rainfall.
- The runoff graphs and scatter graphs indicate that the MLP models over-predict the lower values of runoff.
- It was clearly evident that MLR model fits very poorly for the data set understudy. The runoff graphs and scatter graphs indicate that the MLR models under-predict the peak values of runoff.
- It can be concluded that MLP technique is suitable for runoff prediction in river catchment with the use of minimum parameters.

LITERATURE CITED

Dawson and Wil by, 1998. As a network, MLP was formed by simple neuron called perceptron, which computes a single output from multiple real valued inputs by forming a linear combination according to its input weights and then possibly expressing the output through an online transfer function.

(Leahy *et al*. 2008; Tabari *et al.* 2010b; Zadeh and Tabari 2012). The MLP is a Widely used (ANN)configuration that has been frequently applied in the field of hydrological modeling.

Eberhart and Dobbins, 1990. and a backward phase in which modifications to the connection strengths are made based on the differences between the computed and observed information signals at the output units

Abdelazim M. Negm, M.M. Elfiky, T.M. OwaisandM.H. Nassar (2003). Prediction of suspended sediment concentration in river flow using artificial neural networks Proc. of 6thInt.River Engineering Conf., 28-30 Shahid-Chamran Univ., Ahvaz, Iran.

Alam Md. Ajaz and Singh Vikram 2019 Rainfall simulation using co-active neurofuzzy inferencesystem (CANFIS) and multi linear regression method (MLR) International Research Journal of Engineering and Technology (IRJET) 2019: 06(07): 2237-2239.

Alp M, Cigizoglu H K. (2007). Suspended sediment load simulation by two artificial neural network methods using hydro meteorological data. Journal of Environmental Modeling & Software, 22: 2–13.

Arti R. Naik and Prof. S.K.Pathan (2012). Weather Classification and Forecasting using Back Propagation Feed-forward Neural Network International Journal of Scientific and Research Publications, Volume 2, Issue 12.