

Rainfall – Runoff Modeling in the Hemavathi river catchment of Western Ghats of Karnataka using Regression and Artificial Neural Networks

Y. R. Suresh

Assistant Professor, Department of Civil Engineering, Jyothy Institute of Technology, Bengaluru, Karnataka, India

Abstract

Rainfall–Runoff process is a complex phenomenon which depends on various factors like catchment features, intensity and duration of rainfall, soil characteristics, etc. Many researchers have developed models relating rainfall-runoff in various catchments of the world. In the present study, attempt has been made to develop daily monsoon rainfall-runoff models using Multi Linear Regression (MLR) and Artificial Neural Network Feed Forward Multi Layer Perceptron (MLP) Technique for the Hemavathi river catchment in the Western Ghats of Karnataka. Daily rainfall and 10 day Reciprocal Antecedent Precipitation Index (API) are used as the input parameters, daily runoff being the output. Different models have been tried for both calibration and validation data using the above said techniques. The best model is chosen based on the statistical parameters. The models for the calibration period perform relatively better than those for the validation period data sets. An error analysis also has been carried out to compare the results obtained from the models. Hydrographs are drawn to compare the observed and estimated runoff values and the best year has been identified.

Key Words

Antecedent Precipitation Index, Multi Linear Regression, Multi Layer Perceptron, Root Mean Squared Error, Coefficient of Efficiency, Error Analysis

1. Introduction

The use of Artificial Neural Networks (ANN) for developing the relationship between rainfall and runoff has been studied extensively in recent years. Although there have been some efforts to use ANN as a tool to extract knowledge from data (ASCE, 2000), literature on ANN application as black box models has been much more extensive. Despite the enormous work done during the last decade, research in the related fields continue to be popular because of the vastness of the hydrological data that is available and the amount of work that has been done in the field of watershed modeling (Singh V.P, 2010) before the era of soft computing.

The present study has been carried out with the aim of investigating the applicability of ANN to model the relationship between rainfall and runoff in the Hemavathi river catchment area of the Western Ghats in South India and to understand the merits and demerits of ANN relative to the other modeling techniques commonly adopted for hydrological studies. As is the case with studies reported from the

region earlier (Putty Y.R., 2000), the present study also deals with records of daily rainfall and streamflow for the monsoon period. The work has been carried out using the Multi layer feed forward network with 'Back propagation learning', because it is the most popular ANN (Kalteh A.M., 2008) used in Rainfall-Runoff modeling. It is a practice in modeling Rainfall-Runoff using ANN including antecedent rainfall, the rainfall during one or two days prior to the day is considered in simulating daily runoff. However, the runoff relationships studied before the era of soft computing used to consider the Antecedent Precipitation Index (API) as an additional parameter in the model. It has been established through investigations in Karnataka that API is a very useful parameter, because of the nature of the catchments in the study region, where subsurface delayed flow is as dominant as surface flow, which is produced from variable source areas (Putty Y.R., 2000). Delayed flow and the extent of the source areas both depend directly on the wetness of the catchment, which can be conveniently represented by the API. The present study also aims at establishing the relative performance of the models using the two indices of wetness- the antecedent rainfall and the API. A further objective of the study has been to compare the performance of the ANN technique Multi Layer Perceptron (MLP) with that of the statistical technique Multi Linear Regression (MLR), through an error analysis and to understand the extent to which a conceptual model can perform better than the two models developed. The work has been accomplished by using data of Hemavathi river catchment in the region of the Western Ghats of Karnataka.

2. Some Literature Studies

Minns A.W. and Hall M.J. (1996), generated a synthetic rainfall-runoff data using a conceptual model and used one and two hidden layers in their ANN models and found that the additional computational effort of using two hidden layers was not justified giving only marginal improvements in the performance.

Sajikumar N. and Thandaveshwara B.S. (1999), used Temporal Back Propagation NN (TBPNN) to model the monthly R-R processes in Lee River, UK and Thutharpuzha River in Kerala, India. They found that the performance of the TBPNN models was found to be better than Function Series Models (FSM) under a scarce data scenario.

Putty M.R. and Prasad R. (2000), developed a three-component model with runoff consisting of quick flow, lateral flow and groundwater flow for seven catchments in Western Ghats of Karnataka and have shown that groundwater flow forms a dominant component of runoff and the catchment response is strongly dependent on the rainfall magnitude.

Zhang B. and Govindaraju S. (2000), reported on the prediction of watershed runoff using Bayesian concepts and modular neural networks (MNN) using monthly rainfall, temperature and runoff data with low, medium and high magnitudes. The results obtained from MNNs were found to be superior to a single fully connected feed-forward network.

Rajurkar M.P., et al (2002), carried out studies on modeling daily flows during monsoon period for Narmada river catchment in India and have shown that a linear multiple- input single- output (MISO)

model coupled with ANN provides a better representation of rainfall-runoff relationship compared with linear and nonlinear MISO models.

Ashu Jain, et al (2003), proved that ANN model outperforms the conventional models of Unit Hydrograph and Regression for the rainfall-runoff data sets of four different storms derived from Salado Creek at Bitters Road, San Antonio.

N J de Vos (2003), in his thesis, has done studies on the Alzette catchment in Luxemburg of North West Europe using ANN to develop rainfall-runoff models by considering rainfall, groundwater flow and evapotranspiration as inputs and stated that ANNs give satisfactory accuracy by pre-processing and post-processing the data.

Senthilkumar A R, et al (2005), studied the performance of Multi Layer Perceptron and Radial Basis Function neural networks to develop rainfall-runoff model for Malaprabha river basin by taking rainfall and runoff as inputs. They have concluded that the generalisation properties of RBF networks are poor compared with those of MLP in rainfall-runoff modeling and the choice of the type of network certainly has an impact on the model prediction accuracy.

Sudheer K.P. (2005), observed that ANN model with back propagation algorithm built on the transformed data of three river flows improves the overall prediction accuracy and in particular prediction of peak flows.

Jain A. and Srinivasulu R. (2006), developed integrated R-R models using conceptual and ANN approaches and found that the performance of integrated model was superior.

Kalteh A M (2007), in his thesis, has considered feed forward multi layer perceptron and self organising map for modelling rainfall-runoff in Caspian Sea watershed and Canadian watershed and found that a supervised SOM model can be considered as a better alternative for FFMLP model.

Oozguur Kisi (2007), in his study, has used four different ANN algorithms for the continuous streamflow data of the North Platte River in US and has proved that the Lavenberg-Marquardt algorithm has given the best model.

Jagan Mohan Reddy A., et al (2008), in their study, revealed that the performance of ANN model using multilayer feed forward with logistic sigmoid transfer function has given the best results by including the residuals derived from simple time series and auto regressive models as additional inputs along with rainfall for three gauging sites of Godavari Basin.

Putty M.R. (2009), in his model, has shown that subsurface flow component forms a very significant part of quick flow in the Western Ghats region of South India on the basis of curve number.

Jani Fathima Jamal and Ashu Jain (2011), have proved that feed forward ANN model performs better than the conceptual model for modeling rainfall-runoff data of Jardine river catchment in Australia by considering daily rainfall, streamflow and evapotranspiration data.

Deepika Yadav, et al. (2011), have done studies on forecasting streamflow using Levenberg-Marquardt algorithm for Sewa Hydroelectric project Stage II, a run-of-the river project in Jammu and Kashmir by taking 18 years of inflow data and proved that the lower proportion of the ratio of 90:5:5 for training, validation and testing gave better results.

Pallavi Mittal, et al (2012), have shown that Dual ANN model was superior to Feed Forward NN model for predicting high flows on hourly basis for the rainfall-runoff data of Kolar Basin, India.

V. Nourani, et.al. (2012), predicted the discharge at multiple discharge stations in Peace-Thampa Bay watershed at Florida State using feed-forward NN (FFNN) with back propagation (BP) training algorithm. They showed that the model performs best by using the upstream discharge records in addition to station discharge records and rainfall of sub basin.

Shwetha C.S. (2013), developed MLR and MLP models for the catchments of Nethravathi and Sita of Western Ghats and showed that MLP model performs better in predicting the runoff.

Chandre Gowda C and Mayya S G (2014), in their study on Nethravathi river basin in India, by considering daily rainfall-runoff data of monsoon period, showed that genetic algorithm coupled with neural network model performs better than back propagation neural network model.

Mahesh B. Shrivastav, et al. (2014), showed that feed-forward networks trained by the Levenberg-Marquardt algorithm gave the best model for the prediction of runoff in the Khodiyar catchment area in Amreli district in Gujarat, India.

Muhammed Shaoib, et al (2014), in their study on the Brosna catchment located in Ireland, concluded that the pre-processing of input rainfall data by the wavelet transformation can significantly increase the performance of the Multilayer Perceptron Neural Network and Radial Basis Function Neural Network rainfall-runoff models.

Satyaji Rao Y. R., et al (2014), proved that wavelet based neural networks are more effective than ANN models in forecasting daily streamflow in four catchments of Western Ghats of India. They have evaluated the model performance for slow, medium and high flows and found that the selected model was suitable for all range of runoff.

3. The Study Area and the Data used

The Hemavathi river drains east in to the Mysore Plateau. A major portion of the river drains the steep Western slopes of the Western Ghats, and the rest of their catchments lie in the relatively flatter coastal winter land characterised by small hills.

The catchment is characterised by deep and highly pervious soils, dense and evergreen to semi-evergreen vegetation on the slopes and thick grass lands on the crests. Rainfall in the catchment is very high ranging between 2500 mm and 7500 mm, over the study area. The catchment has an area of 600 sq. km. with 6 rain gauge stations.

Studies on catchment response to rainfall using conceptual models have been carried out by earlier workers, using data from many catchments including those of Lakshmanathirtha, Hemavathi (Putty and Prasad, 2000a) and Nehtravathi (Pradeep, 2008). Hence, in order to investigate the performance of the ANN models relative to that of the conceptual models, the same sets of data used earlier have been used in the present study also. In each case, data of the months of S-W monsoon, during which the region receives 90% of its annual rainfall, has been used.

The daily rainfall- runoff data during monsoon period for 5 years (from 1976 to 1980) is used in the present study. The first 3 years data is considered for calibration and the next 2 years data is used for validation.

4. The Models used in the present study

Both the MLR and MLP are very popular models being used in engineering sciences. Hence, an attempt to describe their principles could only be an unwarranted exercise. However, it should be noted that while the MLR model, associated with a specified number of coefficients, is tenable for the physical interpretation, the ANN is a highly involved and non-linear model formation, seldom providing opportunity to comprehend the kind of association that may exist between the variables involved. In fact, ANN is often termed the ultimate black box (Sajikumar et al, 1999). Hence, when the goal of modeling is limited to estimation of the magnitude of the dependent variable from the known values of the independent variables, the ANN provides upstem opportunities for the modeler and hence is the popularity of the technique. The present study, in a way, attempts to understand the extent to which the new computing abilities help improve simulation of runoff from rainfall. This being the aim, the two models are run parallel, in each case the independent variables being the same and the simulations compared mutually, using the commonly adopted tests of performance. The following paragraphs discuss the model formulation and the tests of performance used in the study.

5. The Model Formulation

It has been an established practice (Linsley et al.,1949) in rainfall-runoff modeling investigations to include an index of catchment moisture condition as an influencing variable, in addition to the rainfall causing runoff. While the Antecedent Precipitation Index (API) has been the most popular moisture index in the pre soft computing days, use of rainfall magnitude during the previous two or three days, in addition to that on the day of runoff, has become a popular practice in studies using ANN. The present study uses both these approaches and compares them mutually. Hence, models developed in the present study can be listed as:

$$(i) RO_i = f(RF_i, API) \quad \dots (1)$$

$$(ii) RO_i = f(RF_i, RF_{i-1}, RF_{i-2}, \dots) \quad \dots (2)$$

Where RO_i is the runoff on the day numbered i and API is the Antecedent Precipitation Index on the day. The API used for the present study is the 10 day reciprocal index (Linsley, et al, 1949) defined by

$$API(i) = \frac{1}{\sum_{t=1}^{10} R(t-t)/t} \dots (3)$$

Three cases of Eq. (2) respectively containing one, two and three previous day rainfall were studied. The Regression was carried out on the EXCEL spread sheet and the Feed Forward MLP models were developed using the facility provided by WEKA 3.7.5 (Waikato Environment for Knowledge Analysis), developed by the University of Waikato, Hamilton, New Zealand. In the latter case, the best fitting ANN architecture was arrived at by a trial and error exercise, considering minimization of the Root Mean Squared Error (RMSE) as the goal, keeping AIC and BIC (Aytek, A. et al, 2008) within limits.

6. Tests of Model Performance

The above tests of model performance are applied on daily simulations of runoff. The commonly adopted split record technique of model testing is used in this study also. In this method, the available length of data is split into two parts, the first part is used for calibrating the model, while the second is used for validating the model performance. RMSE values, AIC and BIC scores (Wilby R.L. et al, 2003), CD and CE values have been used for assessing the performance of the models.

Since a large number of models are being tested and compared mutually using different tests of performance, the quantity of results obtained is huge and hence only sample results with significance are presented. In the case of ANN, the final architecture yielding the best results needs to be arrived at by trial and error. Hence, the model performance referred to in the following discussion pertains only to the most suitable architecture, arrived at by experimentation.

7. The Best Performing Model

The performance of the various models, which include different indices of wetness, is presented in **Table 1**. Since this part of the work intends just to select the best performing model, the values of CE only are presented. It is evident from the table that both in regression and ANN, the three variable model relating daily runoff to rainfall on the day and API is the best performing. Hence, this model is chosen for further discussions. Experiments on ANN also indicate the architecture yielding the best results obtained in each case. The best performing architecture for the Hemavathi catchment is (2,2,1).i.e., 2 input nodes (Daily rainfall and API), 2 hidden nodes and 1 output node (Daily Runoff). The discussions below pertain to the comparison of the performance of MLP with that of MLR.

The results from the performance tests are presented in **Table 2** for daily data. It shows the values of CE separately for the calibration and validation periods. **Table 3** gives the values of CD and CE for each year of data used. **Fig. 1(a), 1(b)** and **Fig. 2(a), 2(b)** show the scatter diagrams and regression lines for the calibration data and validation data respectively by Multi Linear Regression and Multi-Layer Perceptron Techniques. **Fig. 3** and **Fig. 4** show the hydrographs of observed and estimated daily runoff obtained for the calibration data and validation data by both MLR and MLP techniques respectively.

8. Error Analysis

Analysis of errors is often a method to improve the structure of a model. It can also be a means to compare the simulations of different models. The error analysis carried out in the present study comprised of the following.

- (i) A study of the frequency distribution of the errors. (**Table 4**).
- (ii) A comparison between the magnitudes of individual errors in the MLP with the corresponding errors in MLR.

9. Conclusions

Among the models developed, the one with RF and API as the input parameters performs the best in both MLR and MLP techniques with higher values of CE for the calibration data. The best model architecture obtained for the Hemavathi catchment was (2,2,1). These models have been obtained for a learning rate of 0.3 and epoch 500. The models for the calibration period perform relatively better than those for the validation period data sets in the catchment for the daily data.

A careful study of the results presented in **Table 3, Figs. 3 and 4** indicate that the performance of both the models is fairly reliable. However, the difference between the two models is evident from their performance during the validation period.

It is observed that the frequency of errors in the range of -10 to 10 is the highest for the catchment, which indicates that both MLR and MLP models fit reasonably well during calibration period as well as validation period.

After obtaining the weights connecting the inputs, hidden nodes and the output for the best model using MLP, the relative importance of each input on the prediction of output was calculated by Garson Algorithm. The weights obtained for the Hemavathi catchment for calibration and validation periods are as shown. The best model has 2 hidden nodes in a hidden layer. The weights and the relative importance of each input are tabulated in **Tables 4 and 5** for calibration and validation period data respectively.

It is observed that both daily rainfall and 10 day reciprocal API are significant in predicting the daily runoff. The relative importance is shown graphically in **Fig. 5**.

Table 1. Performance of models with different independent variables – Coefficient of Efficiency for the calibration period

Model formulation	Multi Linear Regression (MLR)	Multi-Layer Perceptron (MLP)
$RO_i=f(RF_i)$	0.55	0.54
$RO_i=f(RF_i,RF_{i-1})$	0.65	0.64
$RO_i=f(RF_i,RF_{i-1},RF_{i-2})$	0.66	0.66
$RO_i=f(RF_i, API_i)$	0.69	0.70

Table 2. Performance of the MLR and MLP models for Hemavathi catchment

Period	Data	Coefficient of Efficiency (CE) for daily data	
		MLR	MLP
1976-1978	Calibration	0.69	0.70
1979-1980	Validation	0.74	0.74

Table 3. Values of CD and CE for ROE and ROO for each year – Best year is highlighted

Catchment	Period	Year	Multi	Linear	Multi	Layer
			Regression		Perceptron	
			CD	CE	CD	CE
Hemavathi	Calibration	1	0.75	0.75	0.80	0.80
		2	0.75	0.75	0.73	0.73
		3	0.68	0.68	0.71	0.71
	Validation	1	0.80	0.78	0.82	0.82
		2	0.69	0.62	0.73	0.55

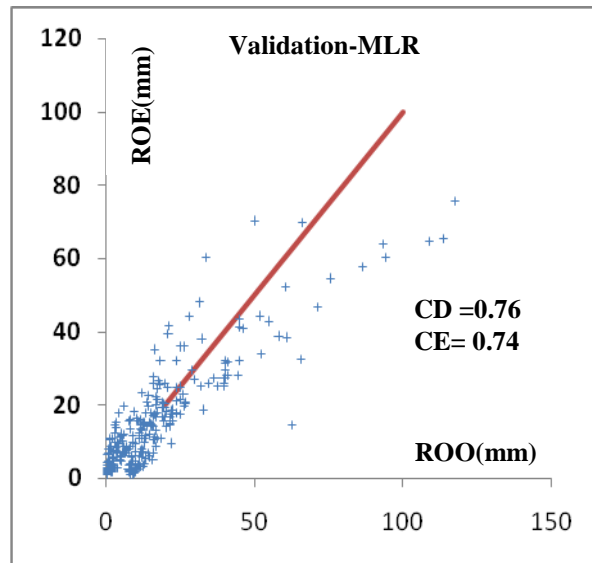
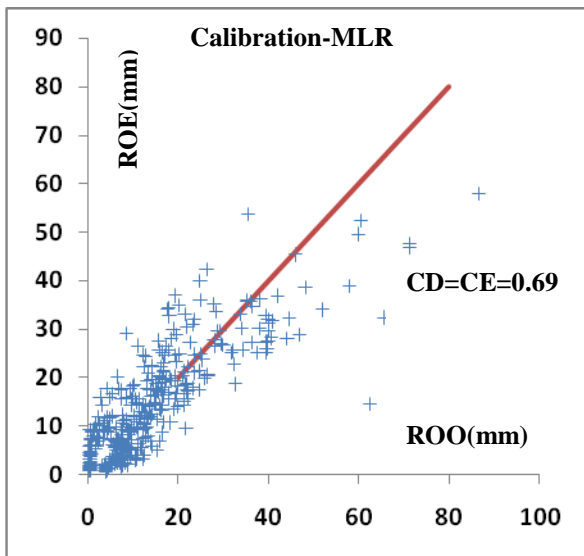


Fig. 1(a). Regression Line for calibration data (Multi Linear Regression)

Fig. 1(b). Regression Line for validation data (Multi Linear Regression)

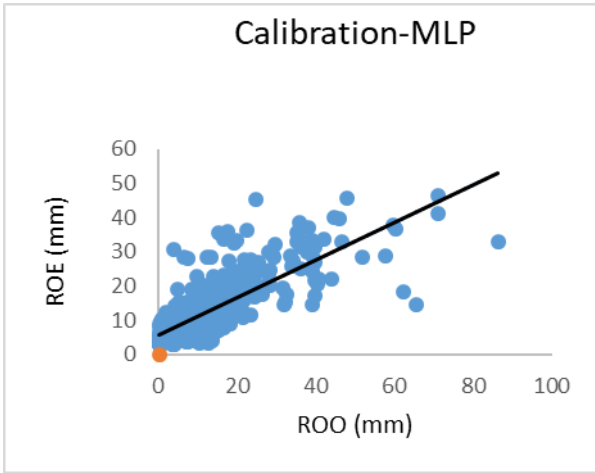


Fig. 2(a). Regression Line for calibration data (Multi-Layer Perceptron)

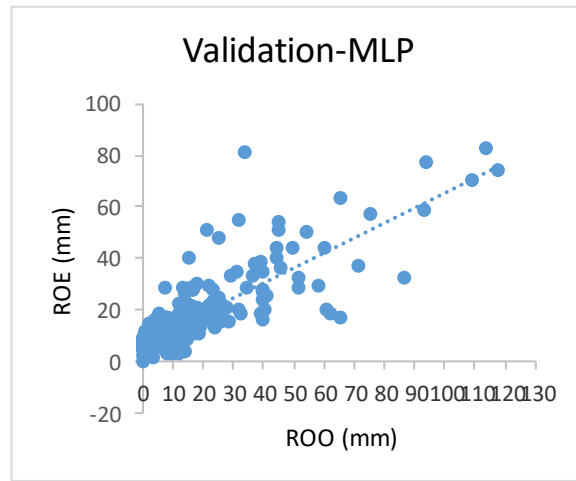


Fig. 2(b). Regression Line for validation data (Multi-Layer Perceptron)

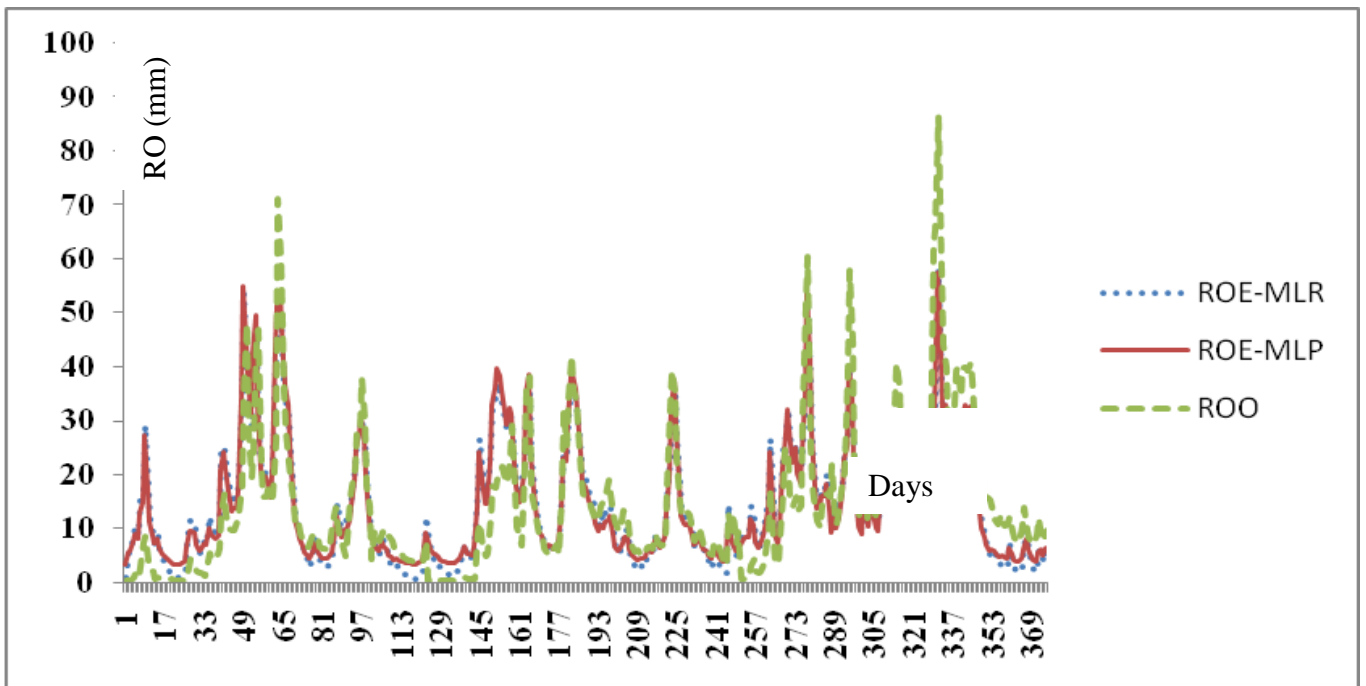


Fig. 3. Hydrographs for Calibration data

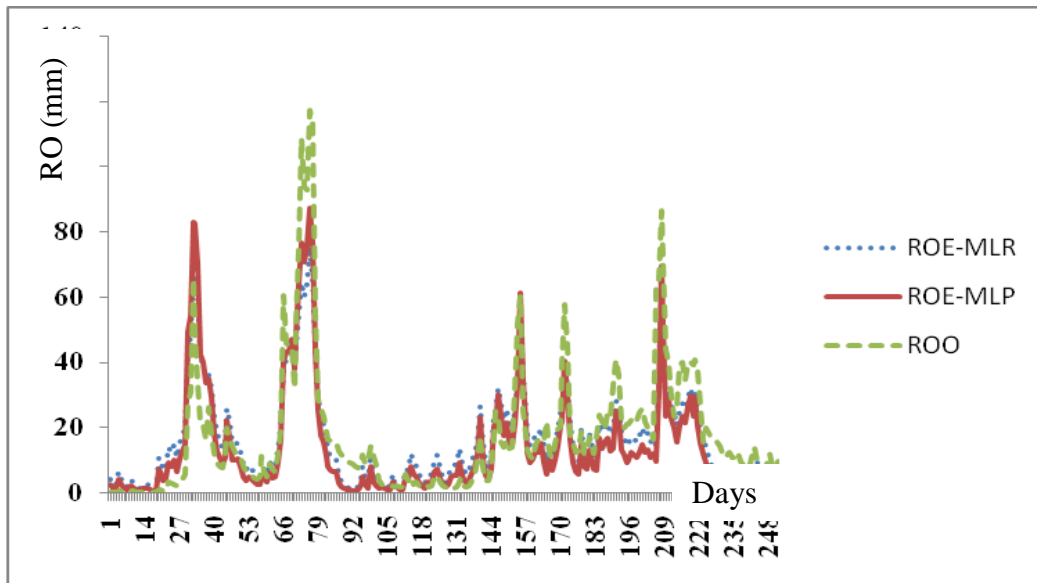


Fig. 4. Hydrographs for Validation data

Table 4. Error frequency analysis

Calibration period			Validation period		
Range	MLR	MLP	Range	MLR	MLP
< -50	0	0	< -50	4	0
-50 to -40	1	1	-50 to -40	1	0
-40 to -30	1	1	-40 to -30	5	0
-30 to -20	3	1	-30 to -20	4	0
-20 to -10	16	16	-20 to -10	15	0
-10 to 0	171	183	-10 to 0	92	0
0 to 10	160	151	0 to 10	105	146
10 to 20	22	21	10 to 20	19	50
20 to 30	1	1	20 to 30	5	26
30 to 40	0	0	30 to 40	2	8
>40	0	0	>40	0	22

Table 4. Weights obtained for the best model for calibration data:

Hidden Layer	RF	10 day reciprocal API	Output (RO)
Hidden node 1	-1.574	-1.212	-1.523
Hidden node 2	0.117	-1.984	-1.18
Relative Importance	54.2	45.8	

Table 5. Weights obtained for the best model for validation data:

Hidden Layer	RF	10 day reciprocal API	Output (RO)
Hidden node 1	1.102	-2.229	-1.537
Hidden node 2	-1.783	-0.762	-1.716
Relative Importance	51.6	48.4	

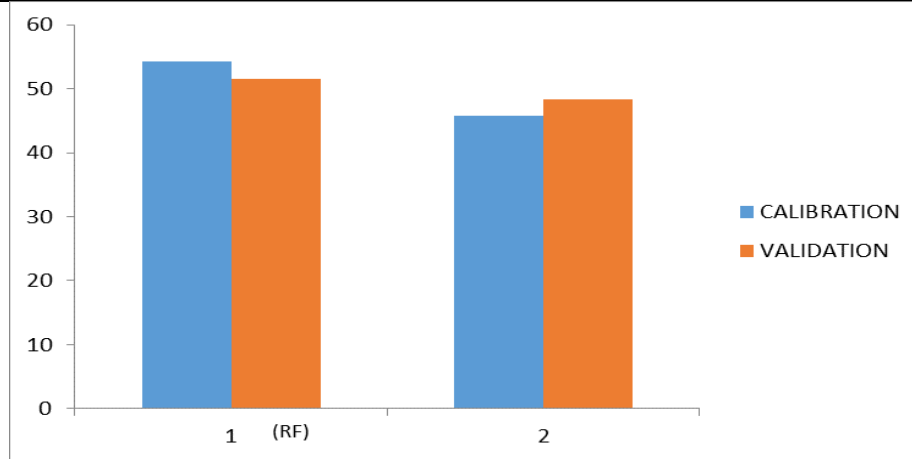


Fig. 5. Relative Importance of inputs in the prediction of RO (Garson Algorithm)
(1 – Daily rainfall, 2- 10 day reciprocal API)

References

1. Aman Mohammad Kalteh (2007), “Rainfall-Runoff Modeling Using Artificial Neural Networks (ANNs)”, Doctoral Thesis, Lund Institute of Technology, Lund University.
2. Ankit Chakravarti et al. (2015), “Rainfall Runoff Analysis Using Artificial Neural Network”, Indian Journal of Science and Technology, Vol 8(14), 54370, July, 1-7.
3. ASCE Task Committee (2000). “ Application of Artificial Neural Networks in Hydrology”, Journal of Hydrologic Engineering, April 2000, 115-123.
4. Ashu Jain, S.K.V. Prasad Indurthy (2003), “Comparitive Analysis of Event-based Rainfall-runoff Modeling Techniques – Deterministic, Statistical and Artificial Neural Networks”. Journal of Hydrologic Engineering, March-April, 93-98.
5. Aytek, A. and Alp, M. (2008). “An application of artificial intelligence for rainfall– runoff modeling”. J. Earth Syst. Sci., 117(2), 145-155.
6. Chandre Gowda C., S.G. Mayya (2014), “Comparison of Back Propagation Neural Network and Genetic Algorithm Neural Network for Stream Flow Prediction”, Journal of Computational Environmental Sciences, Volume 2014, 1-6.
7. Dawson, C.W., Wilby, R. (1998). “ An artificial neural network approach to rainfall- runoff modeling”. Hydrological Sciences Journal ,43, 47–66.
8. Deepika Yadav, et al (2011), “Stream flow forecasting using Levenberg-Marquardt algorithm approach”, International Journal of Water Resources and Environmental Engineering, Vol.3(1), January, 30-40.
9. Haykin, S. (2005). “Neural Networks—A Comprehensive Foundation”. Prentice-Hall.

10. Jagan Mohan Reddy A. et al. (2008), "Rainfall-Runoff Modeling: Combination of Simple Time Series, Linear Autoregressive and Artificial Neural Network Models", Int. Conf. on Water Resources, Hydraulics and Hydrology, University of Cambridge, UK, Feb 23-25, 32-39.
11. Kalteh, A.M. (2008). "Rainfall-Runoff modeling using artificial neural networks (ANNs): Modeling and Understanding". Caspian J. Env. Sci., Vol. 6, No.1, 53-58.
12. Konda Thirumalaiah, Makarand C. Deo (2000), "Hydrological Forecasting Using Neural Networks", J. of Hydrol. Eng., April, 5(2), 180-189.
13. Linsley Jr., R.K., Kohler, M.A., Paulhus, J.L.H. (1949). "Applied Hydrology". McGraw-Hill, New York.
15. Mahesh B. Shrivastav et al. (2014), "Prediction of Runoff using Artificial Neural Networks (A Case Study of Khodiyar Catchment Area)", International Journal for Scientific Research and Development, Vol. 2, Issue 2, 1-4.
16. Minns, A.W., Hall, M.J. (1996). "Artificial Neural Networks as rainfall runoff models". Hydrologic Sci. J., 41(3), June, 399-418.
17. Muhammad Shoaib et al. (2014), "Comparative study of different Wavelet based neural network models for rainfall-runoff modeling", Journal of Hydrology, 515, 47-58.
18. Mysooru R. Yadupathi Putty (2009), "Curve-Number-Based Watershed Model Incorporating Quick Subsurface Runoff, with Applications in the Western Ghats, South India", J. of Hydrol. Eng., 14(8), August, 876-881.
19. N.J. de Vos (2003), "Rainfall-Runoff Modeling Using Artificial Neural Networks", M.Sc. Thesis Report, Delft University of Technology, Delft, Netherlands.
20. Nourani V., Masoud Mehrvand (2012), "An ANN-Based Data Driven Approach for Multi-Station Rainfall-Runoff Modeling", International Journal of Computer Science and Management Research, Vol. 1, Issue 4, November, 771-780.
21. Oozguur Kisi (2007), "Streamflow Forecasting Using Different Artificial Neural Network Algorithms", J. of Hydrol. Eng., September-October, 12(5), 532-539.
22. Pallavi Mittal et al. (2012), "Dual Artificial Neural Network for Rainfall-Runoff Forecasting", Journal of Water Resource and Protection, 4, 1024-1028.
23. Putty, Y.R. (2011). "Principles of Hydrology". I.K. International Publishing House Pvt. Ltd.
24. Putty, Y.R., Prasad. R. (2000). "Understanding runoff processes using a watershed model – a case study in the Western Ghats in South India". Journal of Hydrology, 228, 215-227.
25. Rajurkar M.P. et al. (2002), "Artificial neural networks for daily rainfall-runoff modeling", Hydrological Sciences, 47(6), December, 865-877.
26. Sajikumar and Thandaveswara. (1999). "A nonlinear rainfall – runoff model using Artificial Neural Network". Journal of Hydrology, Vol. 216, 32-55.
27. Senthilkumar A.R., et al. (2005), "Rainfall-runoff modeling using artificial neural networks: comparison of network types". Hydrol. Process. 19, 1277-1291.
28. Singh V.P. et al. (2010). "Daily outflow prediction by Multi Layer Perceptron with Logistic Sigmoid and Tangent Sigmoid Activation Functions". Water Resour Manage, Springer, 1-16.
29. Sudheer K.P. (2005), "Advances in Neural Network Modeling: Application to River Flow Forecasting", 1-6.
30. WEKA 3-7-5 Manual. University of Waikato, Hamilton, New Zealand, 2011.



31. Wilby R.L. et al (2003), “Detection of conceptual model rainfall-runoff processes inside an artificial neural network”, Hydrological Sciences Journal, 48(2), April, 163-181.