Vol. 47, No. 3, pp 231-238, 2019

Indian Journal of Soil Conservation

Online URL : http://indianjournals.com/ijor.aspx?target=ijor:ijsc&type=home

Rainfall-runoff simulation modelling using artificial neural networks in semi-arid middle Gujarat region

A.K. Koradia, A.D. Bhalala and M.K.Tiwari*

College of Agricultural Engineering and Technology, Anand Agricultural University, Godhra-389001, Gujarat.

*Corresponding author:

E-mail: tiwari.iitkgp@gmail.com (M.K. Tiwari)

ARTICLE INFO

Article history:

Received : March, 2019

Revised : December, 2019

Accepted : December, 2019

Key words:

ANN

GIS

Gujarat

Rainfall

Runoff

Watershed

ABSTRACT

Rainfall-runoff modelling is important for water resources planning, development and management. Water resource managers require information about runoff from a hydrologic catchment area in order to assess runoff potential, reservoir and canal operation, flood and drought management, etc. The present work involves the development of artificial neural network (ANN), principal component analysis (PCA) based ANN (PCA-ANN) and multiple linear regression (MLR) models for establishing rainfall-runoff relationship. In this study, 10 years (2007-2016) of rainfall and runoff data were applied. A robust ANN model was developed by considering different types of training algorithms such as LM, GDX, BFG, CGF, SCG, BR, CGP and RP. The performance of ANN models was also compared with PCA-ANN and MLR models by using statistical indices. It was found in this study that ANN (ANN-1) model with only one lag data at outlet (Santrod gauging station) is suitable to effectively and precisely predict runoff at one day lead time. It was also observed in this study that performance of ANN model is better than PCA-ANN and MLR models for prediction of one day lead discharge at Santrod. Hence, it is recommended to use ANN-1 model to predict runoff for Santrod gauging station of Panam watershed. It will help the water resource managers and field engineers to take suitable decisions related to reservoir and canal operation, flood and discharge management.

1. INTRODUCTION

Understanding of hydrological processes is the backbone for efficient water resources planning and management. The knowledge of catchment runoff gives an idea of water that is available to replenish water bodies in the catchment, and therefore very important in the management of both potable and agricultural water. Furthermore, quantification of runoff, understanding of different hydrological components and water budgeting gives indications about opportunities to harvest rain water (Welderufael *et al.,* 2009; Sharda *et al.,* 2019; Patel *et al.,* 2019), and future plans and management options. A hydrologic model is capable of establishing rainfall–runoff relationship and forecasting future river discharge values that are useful for hydrologic and hydraulic engineering design.

An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks. ANN models have been successfully used for modelling complex nonlinear input / output time series relationships, classification, pattern recognition and other problems in a wide variety of fields (Shiva Prasad *et al.,* 2017; Singh, 2015; Sharma *et al.,* 2015; Shirgure and Rajput, Tiwari and Chaterjee, 2010[a,b]). The high degrees of empiricism and approximation in the analysis of hydrologic systems are highly suitable for the application of ANNs (Hsu *et al.,*1995). As the hydro-climatic data are generally highly correlated, to deal with such issues, principal component analysis (PCA) has been successfully applied in earlier studies (Noori *et al.,* 2010). Sehgal *et al.* (2014) proposed a new hybrid model, the wavelet-bootstrap-multiple linear regression (WBMLR) to explore potential of wavelet analysis and bootstrap resampling techniques for daily discharge forecasting. The results showed that the wavelet bootstrap hybrid models (*i.e.* WBMLR and WBANN) produced significantly better results in comparison to the MLR and ANN models. MLR based hybrid models (WMLR, WBMLR) performed better than the ANN based hybrid models (WBANN, WANN) in the study. Kumar *et al.* (2015) developed a BWANN model for reservoir inflow forecasting by combining ANN model with wavelet and bootstrap analysis, and compared performance of the developed model with wavelet based ANN (WANN), wavelet based MLR (WMLR), bootstrap and wavelet analysis based multiple linear regression models (BWMLR), standard ANN, and standard MLR models for inflow forecasting. Different performance indices indicated better performance of WANN model in comparison with WMLR, ANN and MLR models for inflow forecasting. BWANN model was found better than BWMLR model for uncertainty assessment. Limited research work is carried out using ANN for rainfall runoff simulation in Indian catchments. Makwana and Tiwari (2017) conducted rainfall runoff simulation study of data scarce semi-arid region of Gujarat, India. They found that performance of ANN modelling techniques was better compared to SWAT model.

In this study, ANNs, PCA based ANN (PCA-ANN), and MLR models were developed to predict daily runoff of the catchment area, locted in Panam watershed, middle Gujarat region, India. Performance comparison of different learning algorithms for runoff forecasting and performance evaluation of all the models using different statistical indices are presented and discussed during training and testing period of the data.

The study was taken for the first time to develop a robust ANN model using training, testing and additional cross-validation dataset for better generalization of ANN models, the best ANN models was developed by considering several input combinations along with different training algorithms. Morover for Indian and semi-arid conditions for the first time, a PCA based ANN models were developed and compared. The combination of all the above components was a novel idea applied in this study for rainfall runoff modelling.

2. MATERIALS AND METHODS

Study Area

In the present study, a watershed of Panam river basin in middle Gujarat region was selected. The major river in watershed is Panam river. It has a total length of about 116 km and drainage area of about 2305 km^2.The base map of study area under Panam watershed is show in Fig.1.

For conducting the present study, three runoff gauging stations were considered namely, Rampur, Limkheda, and Santrod, whereas gridded rainfall data represented as A, B, C, and D falling in the study area as presented in Fig. 1, were collected from IMD, Pune.

Data Applied for Model Development

For ANN, PCA-ANN and MLR models development and simulation, hydro-climatic data of daily discharge and

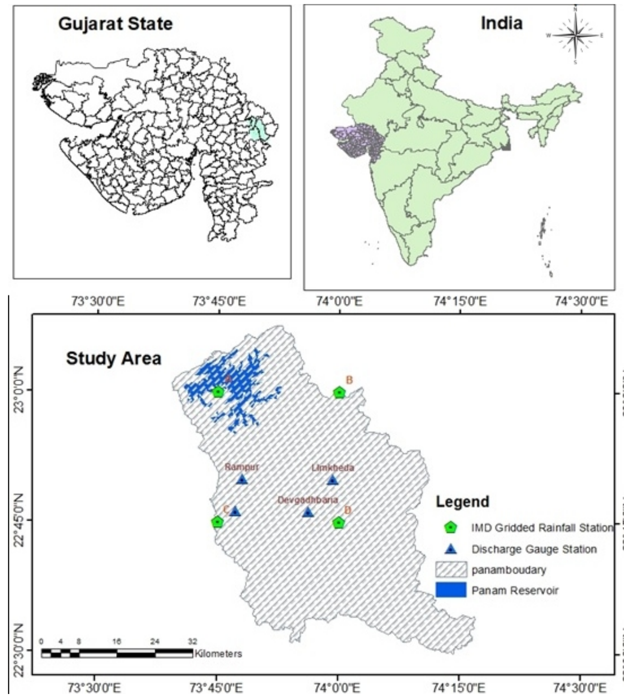


Fig. 1. Base map of study area

rainfall data (duration 2007-2016) for three runoff gauge stations and four IMD gridded locations, respectively, were collected from State Water Data Center (SWDC), Gandhi Nagar and Indian Metrological Department (IMD), respectively, and SRTM-DEM remote sensing data was collected from shuttle radar topography mission (SRTM) United State Geological Survey (USGS). In this study, Arc-GIS 10.0 software was used for watershed delineation and base map preparation. MATLAB R2018a software was applied for development of ANN and PCA-ANN models, and Origin Professional 2019 was applied for MLR modelling.

ANN Architecture

ANN architecture is defined based on the way neurons are connected to each other, which determines how computations proceed. ANN is naturally composed of three different types of layers of neurons, an input layer, one or more hidden layer and an output layer. Back-propagation training algorithm is the most commonly used supervised algorithm for training the multi-layer ANN. In the ANN model, activation function plays an important role in capturing any kind of non-linearity between input and output, and therefore the widely applied sigmoid function is used in this study (Makwana and Tiwari, 2017).

Development of ANN Models

The identification of input and output variables is the first step for developing the ANN models. Cross correlation analysis as presented in Fig's. 2 and 3 was conducted to identify significant input variables, and the procedure was followed by incremental input selection and verification (Tiwari and Chatterjee, 2010[a,b]). In the present study, different models were developed by applying different combinations of lag of daily rainfall and runoff data sets, as presented by the following relationship:

$$Q_t = f\,(R_{t-1}, R_{t-2} \dots R_{t-n} \text{ and } Q_{t-1}, Q_{t-2}, \dots Q_{t-n}) \qquad \dots(1)$$

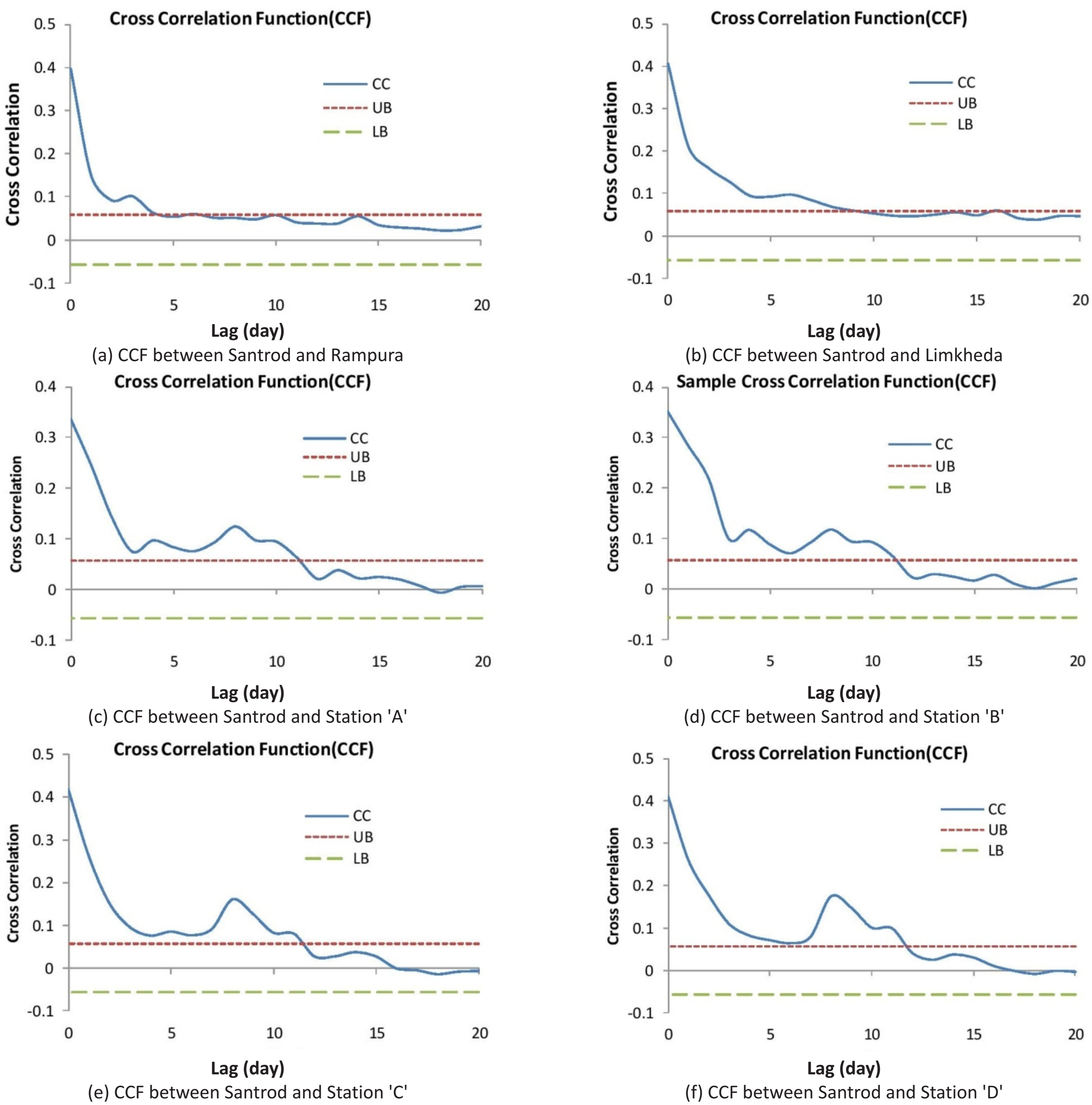
Where, Q_t is daily runoff rate; Q_{t-1} is daily runoff rate at time *t-1*; Q_{t-2} is daily runoff rate at time *t-n*; R_{t-1} is daily rainfall at time *t-1*; R_{t-2} is daily rainfall at time *t-2*; and R_{t-n} is daily rainfall at time *t-n*.

The different combinations of input-output relationships of data sets are given in the Table 1 for the ANN models. Different training variables applied for ANN model development are presented in Table 2.

Principal Components Analysis (PCA)

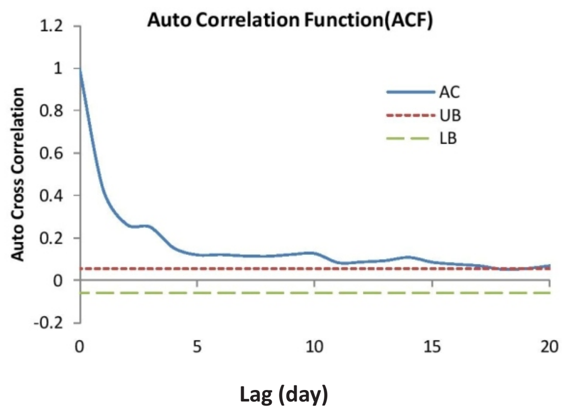
In PCA, a reduced set of *m* components or factors are extracted from a set of *p* variables that accounts for most of the variance in the *p* variables. In other words, a set of *p* variables are reduced to a set of *m* underlying super ordinate dimensions.

These underlying factors are inferred from the correlations among the *p* variables. Each factor can be

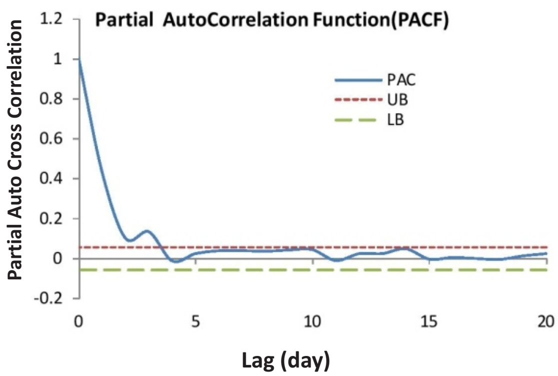


(a) CCF between Santrod and Rampura
(b) CCF between Santrod and Limkheda
(c) CCF between Santrod and Station 'A'
(d) CCF between Santrod and Station 'B'
(e) CCF between Santrod and Station 'C'
(f) CCF between Santrod and Station 'D'

Fig. 2. Cross correlation statistics for input variables selection



(a) ACFs of Santrod gauging station



(b) PACFs of Santrod gauging station

Fig. 3. (a) Auto-correlation functions (ACFs) and (b) Partial auto-correlation functions (PACFs) statistics for input variables selection

Table: 1
Input-output combinations for ANN models

Model no.	Model	Input-output combinations
1	ANN-1	Qt=f(sant(t-1))
2	ANN-2	Qt =f(sant(t-1)ram(t-1))
3	ANN-3	Qt =f(sant(t-1)ram(t-1),lim(t-1))
4	ANN-4	Qt=f(sant(t-1)ram(t-1),lim(t-1),raind(t-1) raind(t-2))
5	ANN-5	Qt =f(sant(t-1)ram(t-1),lim(t-1),raind(t-1) raind(t-2)raina(t-1))
6	ANN-6	Qt =f(sant(t-1)ram(t-1),lim(t-1),raind(t-1) raind(t-2)raina(t-1)rainb(t-1))
7	ANN-7	Qt =f(sant(t-1)ram(t-1),lim(t-1),raind(t-1) raind(t-2)raina(t-1)rainb(t-1)rainc(t-1))

represented as a weighted sum of the p variables. The i^{th} factor is represented as:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + K + W_{ip}X_p \qquad ...(2)$$

In this way, each of the p variables is a linear function of the m factors and can be represented as:

$$X_j = A_{ij} + A_{2j}F_1 + A_{2j}F_2 + K + A_{mj}F_m + U_j \qquad ...(3)$$

Where, U_j is the unique variance to variable j that cannot be explained by any of the common factors.

Table: 2
Training variables and their assigned values for ANN models

Training variables	Assigned value
Neural Network	Feed-forward back propagation
Hidden layer	1
No. of neurons	1 to 15
Learning function	LEARNGDM (Levenberg Marquardt)
Transfer function	Tansig
Performance function	MSE (Mean Squared Error)
Training function	TRAINLM
Epoch	500
Training threshold	0.001

Training Variables and their Assigned Values for ANN Models

Performance indices

The statistical parameters namely, coefficient of determination (R^2), Nash-Sutcliffe Efficiency (NSE), root mean square error (RMSE), percentage deviation (P_{dv}), and mean absolute error (MAE) were used to evaluate the performance of developed models for comparison between observed and predicted runoff values.

3. RESULTS AND DISCUSSION

Statistical Analysis of Raw Data

The daily rainfall and runoff data of the Santrod gauging station of Panam watershed for ten years period from 2007 to 2016 was used. First six years (2007 to 2012) data was used for training purpose, two years (2015 to 2016) data for cross-validation, and remaining two years data (2013-2014) were used for testing the developed rainfall-runoff models. A brief statistical analysis of the applied data is presented in Table 3.

Model Structure Identification

The current study used a statistical approach suggested by Sudheer *et al.* (2002) to identify the appropriate input vectors. The method presents the identification of significant input variables using statistical analysis of the data series such as cross-correlation functions (CCFs), auto-correlation functions (ACFs) and partial auto-correlation functions (PACFs) between the variables. This process was also applied by Tiwari and Chatterjee (2010[a]) to select significant inputs from the seven discharge gauging stations for daily river flow forecasting. CCFs were developed between discharge at Santrod gauging station and different variables at different discharge (Rampura and Limkheda) and rainfall gauging (A, B, C, and D) stations located upstream and nearby. Through cross correlation analysis, it was identified that the different variables have good correlation with discharge at outlet, varying from 1 day lag to 4 days lag (Fig. 2). By using ACFs and PACFs (Fig. 3) it was identified that only 1 day lag is significant for surface runoff simulation. This prior knowledge was also used to further select significant inputs.

Table: 3
A brief statistics of data sets applied for modeling

Data and duration	Discharge (m^3sec^{-1})								
	Rampur			Limkheda			Panam at Santrod		
	Max	Min	Std	Max	Min	Std	Max	Min	Std
Training (2007-2012)	538.9	0.0	21.7	728.1	0.0	35.4	1704.8	0.0	91.0
Cross-validation (2015-2016)	17.9	0.0	1.8	124.0	0.0	22.3	480.3	0.0	44.8
Testing (2013-2014)	57.2	0.0	7.1	455.1	0.0	34.8	519.4	0.0	65.3

Performance of ANN Models during Training and Testing

Performance of different ANN models for training and testing period is presented in Table 4. In terms of different performance indices, it can be observed that ANN-1 model performs best. It can be observed from the table that in terms of different performance indicators, the best input variable for runoff simulation is sant (t-1) to simulate the runoff sant(t) at the Santrod. The performance of ANN-1 model using LM training algorithm in terms of different performance indices such as R^2, NSE, RMSE, P_{dv}, and MAE was found as 0.83, 55.59%, 29.88 m^3sec^{-1}, 35.67%, and 19.68 m^3sec^{-1}, respectively.

It can be further observed from the table that performance of ANN models are better in testing period compared to training period. The reason behind this is that the variability in the training dataset is more with maximum value being 1704.8 m^3sec^{-1}, whereas in testing period the maximum value is 519.4 m^3sec^{-1} at Panam at Santrod gauging station. It makes training difficult to generalize with such a high variability in the training dataset, but once the model is trained, it performs better for the testing dataset with less variability within and being within the training dataset's limit. It may be due to the reason that variability in the training dataset is more compared to testing dataset.The hydrograph and scatter plot between observed and predicted runoff during training year 2007-2012 and during testing year 2013-2014 using ANN-1 model are shown in Fig. 4. It can be observed from the figure during the training period that most of the observed predicted value are showing general behaviour as that of observed value. It can be observed from the figure during the testing period that predicted values are more closer to the observed values, and performance can be considered as satisfactory.

Performance of ANN Models during Training and Testing Using Different Learning Algorithms

Performance of the best found ANN-1 model was further evaluated using 9 different learning algorithms, and the results are presented in Table 5. The performance ANN-1 model using CGB training algorithm in terms of different performance indices such as RMSE, P_{dv}, NSE, R^2 and MAE were found as 29.45 m^3sec^{-1}, 45.52%, 55.59%, 0.82 and 14.71 m^3sec^{-1}, respectively. The performance of CGB training algorithm was found best followed by LM, GDX, BFG, CGF, SCG, BR, CGP and RP.

Performance of ANN-1 model in terms of hydrograph and scatter plot using CGB training algorithm for the training (2007-2012) and testing period (2013-2014) is presented in Fig. 5. The general behavior between observed and predicted values can be easily identified using these graphs.

Performance of ANN Models during Training and Testing Using MLR

The performance of the developed MLR models was evaluated using five statistical indices namely, RMSE, $P_{dv,}$ NSE, R^2, MAE as 71.01 m^3sec^{-1}, 51.06%, 26.48%, 0.26 and19.68 m^3sec^{-1}, respectively, during the testing period. It was observed in this study that performance of MLR models is inferior compared to ANN models. From the hydrograph and scatter plots between observed and predicted values (Fig. 6) it can be verified that the predicted values are deviating much from the observed values.

Performance of ANN Models during Training and Testing Using PCA

The performance of the developed PCA-ANN models during the testing period was evaluated using five statistical indices namely, RMSE, $P_{dv,}$ NSE, R^2, MAE, and were found

Table: 4
Performance of different ANN models using different input variables

S.No.	Model	Optimum No of HN	Performance Indicator for Training					Performance Indicator for Testing				
			R^2	NSE (%)	RMSE (m^3sec^{-1})	P_{dv} (%)	MAE (m^3sec^{-1})	R^2	NSE (%)	RMSE (m^3sec^{-1})	P_{dv} (%)	MAE (m^3sec^{-1})
1	ANN-1	15	0.26	25.64	77.84	69.83	25.27	0.83	55.59	29.88	35.67	19.68
2	ANN-2	4	0.22	21.65	79.90	81.47	23.93	0.71	43.15	33.81	60.90	19.15
3	ANN-3	7	0.28	27.01	77.11	81.70	26.17	0.73	42.22	34.08	50.15	21.75
4	ANN-4	9	0.24	24.03	78.67	74.51	23.46	0.77	46.66	32.87	54.27	21.31
5	ANN-5	2	0.23	23.28	78.93	69.72	22.26	0.72	45.88	33.11	62.52	18.22
6	ANN-6	4	0.23	22.54	79.00	85.93	23.56	0.71	46.22	33.00	56.89	18.10
7	ANN-7	1	0.22	21.62	79.47	88.03	23.50	0.70	43.93	33.70	57.62	18.76

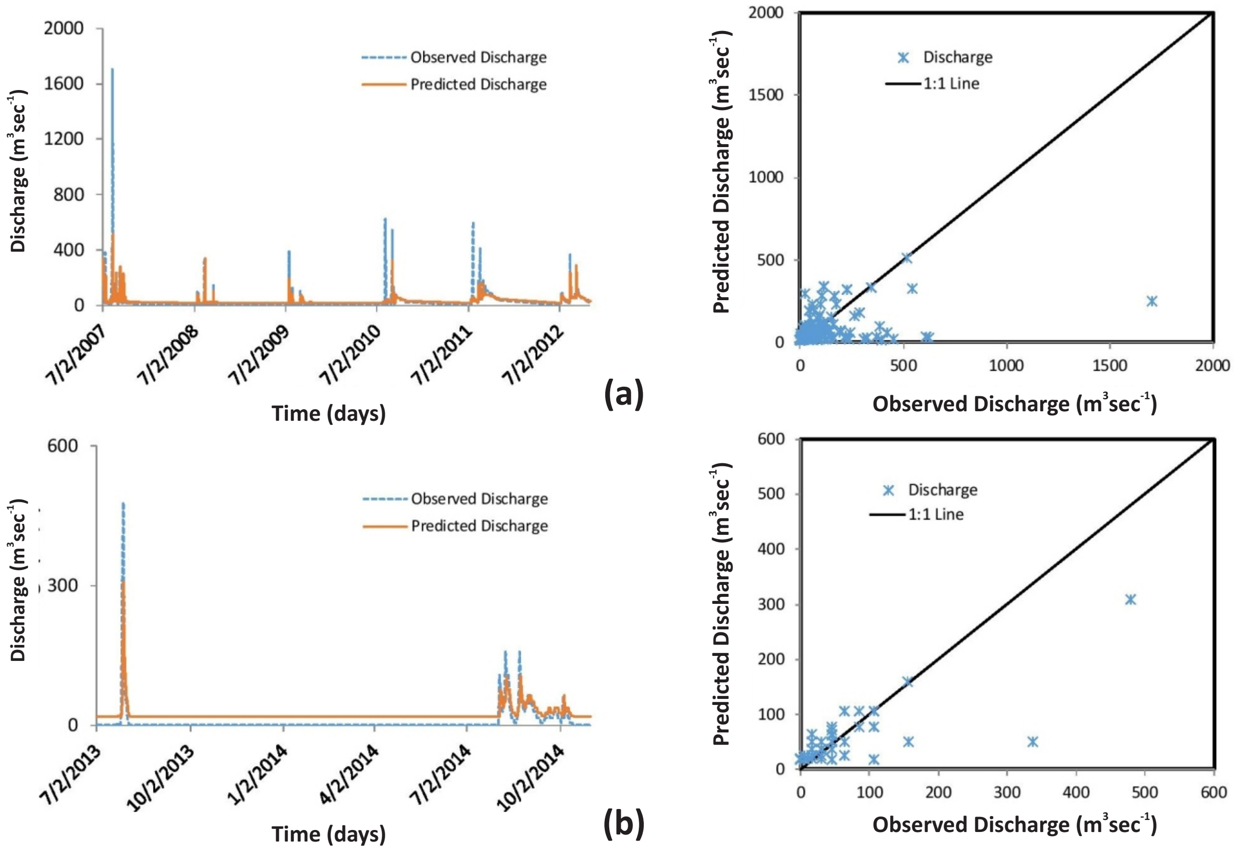


Fig. 4. Hydrograph and scatter plot of observed and predicted runoff using ANN-1 for (a) training and (b) testing

Table: 5
Performance of ANN-1 model using different algorithms

Model No.	Learning Algorithm	No. of Hidden Neurons	Performance Indices				
			R^2	NSE (%)	RMSE (m^3sec^{-1})	P_{dv} (%)	MAE (m^3sec^{-1})
1	CGB	15	0.82	56.91	29.43	45.52	14.71
2	LM	15	0.83	55.59	29.88	35.67	19.68
3	GDX	15	0.83	54.59	30.22	37.37	20.42
4	BFG	15	0.82	53.86	30.46	48.81	17.98
5	CGF	12	0.82	53.48	30.58	41.35	19.19
6	SCG	15	0.82	52.25	30.98	46.71	19.78
7	BR	8	0.77	50.33	31.60	57.06	17.22
8	CGP	15	0.72	49.53	31.86	59.44	15.23
9	RP	14	0.75	46.95	32.66	57.16	18.71

as 33.05 m^3sec^{-1}, 61.20%, 45.67%, 0.72 and 18.67 m^3sec^{-1}, respectively. Whereas performance of best found ANN-1 model using CGB training algorithms in terms of RMSE, P_{dv}, NSE, R^2 and MAE was found as 29.45 m^3sec^{-1}, 45.52%, 55.59% , 0.82 and 14.71 m^3sec^{-1}, respectively. Therefore, it can be observed that PCA-ANN model does not perform well as its performance is far inferior than ANN model.

It can be observed from the Fig. 7 that performance of PCA-ANN model is not better than the results shown by ANN-1 model with CGB training algorithms owing to poor closer association between observed and predicted values in the hydrograph. Moreover, the scatter plot shows that the predicted vales are not much close to 1:1 line. But when performance of PCA model is compared with MLR model, it has better performance as it shows closer association between observed and predicted values in the hydrograph.

4. CONCLUSIONS

Based on the performance indices among ANN, ANN-1 model performed better than other ANN models in daily runoff forecasting for Santrod gauging station of Panam watershed, and the results show that ANN models are more efficient than PCA-ANN and MLR models in the rainfall-runoff modeling for Santrod gauging station of Panam watershed. Further, performance of CGB training algorithm

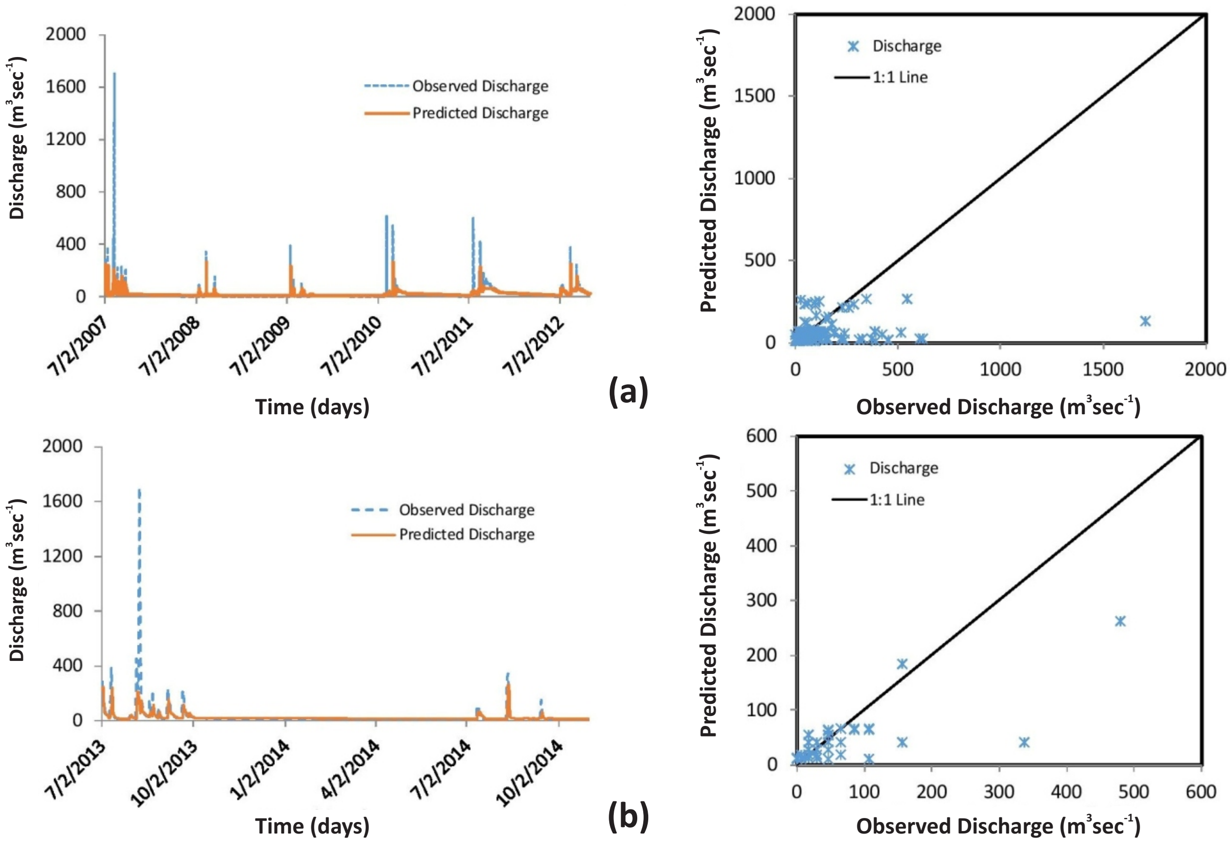


Fig. 5. Hydrograph and scatter plot of observed and predicted runoff using ANN-1 with CGB training algorithm for (a) training and (b) testing dataset

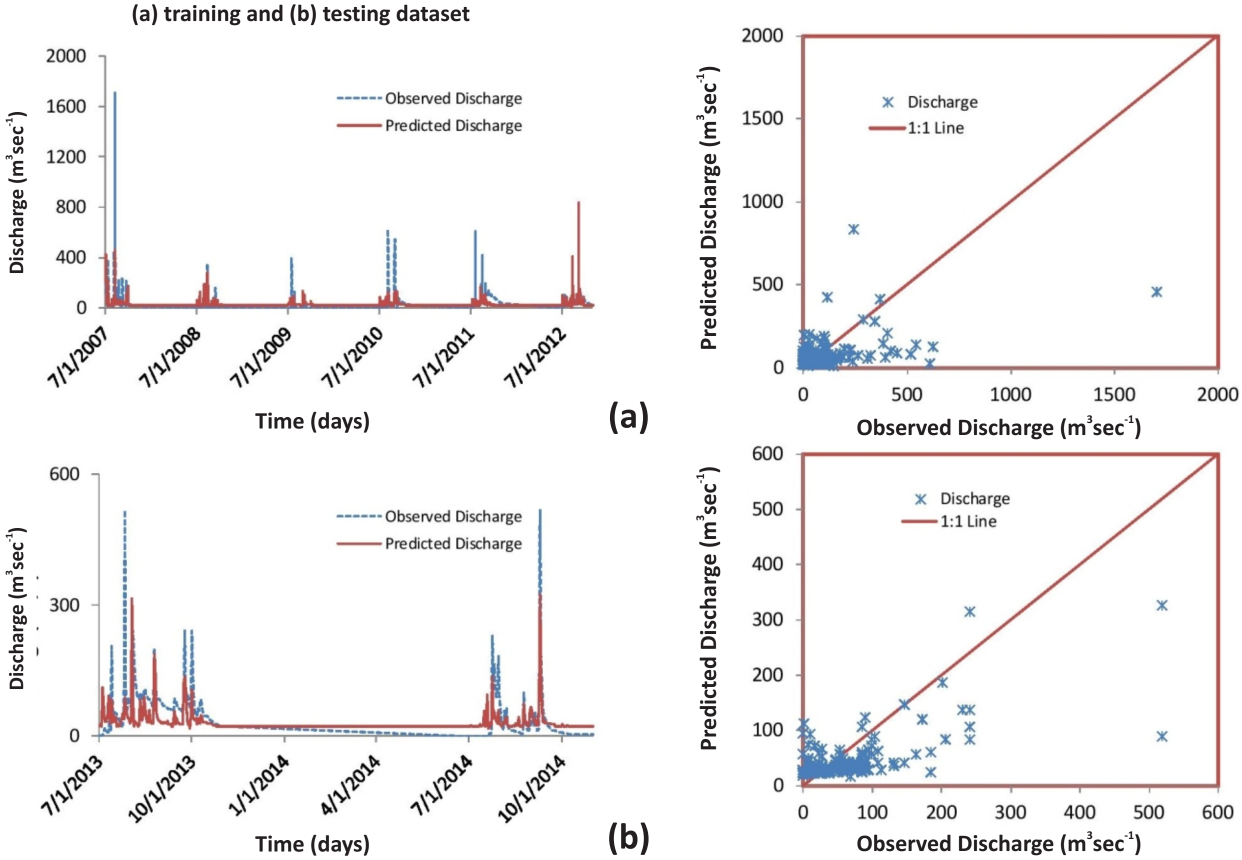


Fig. 6. Hydrograph and scatter plot of observed and predicted runoff using MLR for (a) training and (b) testing

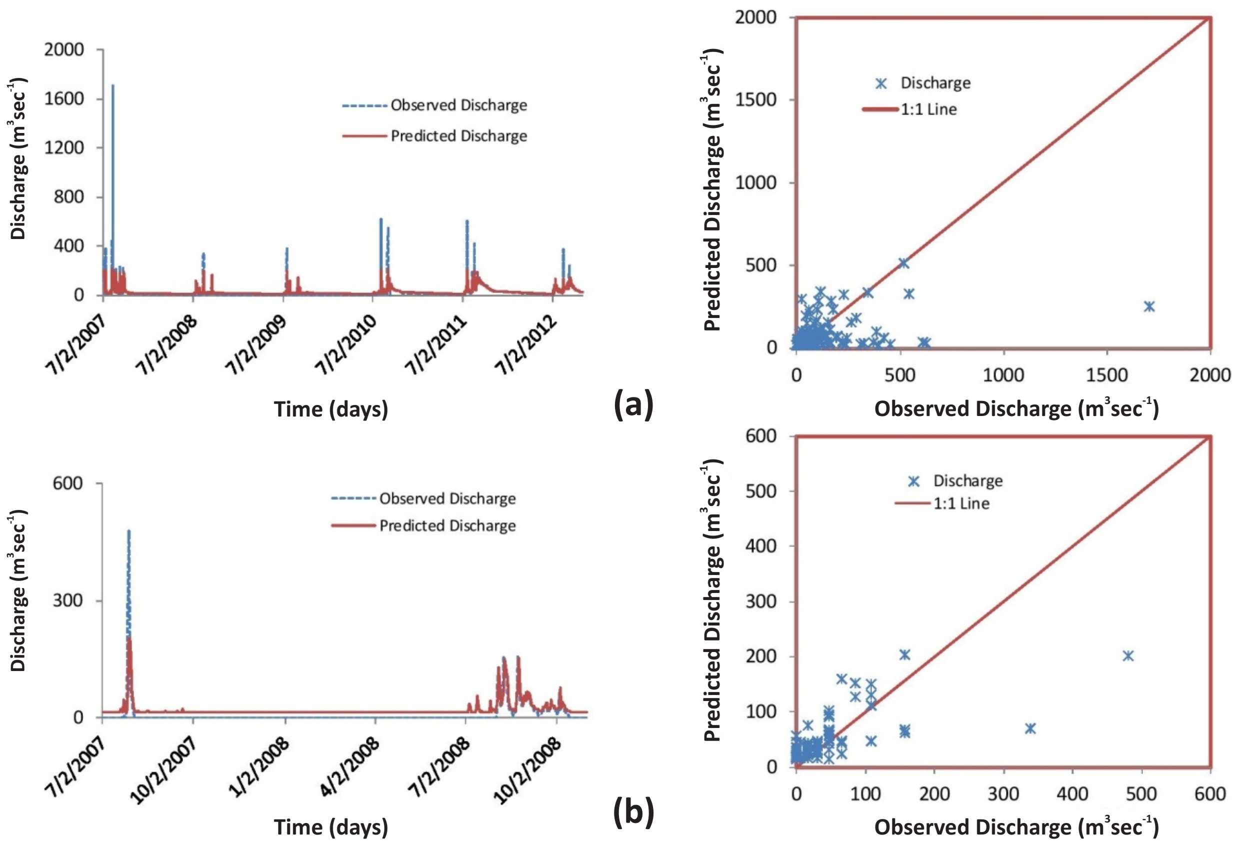


Fig. 7. Hydrograph and scatter plot of observed and predicted runoff using PCA for (a) training and (b) testing

was found best followed by LM, GDX, BFG, CGF, SCG, BR, CGP and RP. Hence, it is recommended to use ANN-1 model to predict runoff for Santrod gauging station of Panam watershed. It will help the water resource managers to operate the Panam watershed properly in the case of extreme events such as flooding.

ACKNOWLEDGEMENTS

The authors convey sincere thanks to the State Water Data Center (SWDC), Gandhi Nagar; and Indian Meteorological Department (IMD), Pune for providing the data; and ITRA, Digital India Corporation, Ministry of Electronics and Information Technology, Government of India for providing the fund for conducting the study.

REFERENCES

Hsu, K.L., Gupta, H.V. and Sorooshian, S. 1995. Artificial neural network modeling of the rainfall-runoff process. *Water Resour. Res.*, 31(10): 2517-2530.

Kumar, S., Tiwari, M.K., Chatterjee, C. and Mishra, A. 2015. Reservoir Inflow Forecasting Using Ensemble Models Based on Neural Networks, Wavelet Analysis and Bootstrap Method. *Water Resour. Manage.*, 29 (13): 4863-4883.

Makwana, J. and Tiwari, M.K. 2017. Hydrological stream flow modelling using soil and water assessment tool (SWAT) and neural networks (NNs) for the Limkheda watershed, Gujarat, India. *Model. Earth Syst. Environ.*, 3(2): 635-645.

Noori, R., Khakpour, A., Omidvar, B. and Farokhnia, A. 2010. Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. *Exp. Syst. Appl.*, 37(8): 5856-5862.

Patil, M., Kothari, M., Gorantiwar, S.D. and Singh, P.K. 2019. Runoff simulation using the SWAT model and SUFI-2 algorithm in Ghod catchment of upper Bhima river basin. *Indian J. Soil Cons.*, 47(1): 7- 13.

Sehgal, V., Tiwari, M.K. and Chatterjee, C. 2014. Wavelet bootstrap multiple linear regression based hybrid modeling for daily river discharge forecasting. *Water Resour. Manage.*, 28(10): 2793-2811.

Sharda, V.N., Mandal, D. and Dogra, P. 2019. Assessment of cost of soil erosion and energy saving value of soil conservation measures in India. *Indian J. Soil Cons.*, 47(1): 1-6.

Sharma, S.K., Hardaha, M.K. and Ranade, D.H. 2015. Estimation of soil loss using artificial neural networks for Kalidevi watershed of Dhar, Madhya Pradesh, India. *Indian J. Soil Cons.*, 43(2): 135-141.

Shirgure, P.S. and Rajput, G.S. 2014. Modeling daily pan evaporation using artificial neural networks. *Indian J. Soil Cons.*, 42(2): 119-129.

Singh, A. 2015. Optimization of neural network structure for radial basis function network for simulation of hydrological processes. *Indian J. Soil Cons.*, 43(3): 250-254.

Shiva Prasad, H.J. and Danu, U.S. 2017. Rainfall-runoff modeling of Ram Ganga river basin using Artificial Neural Network (ANN) and multiple linear regression (MLR) techniques, www.krishikosh.egranth.ac.in/handle/1/5810056215.

Sudheer, K.P., Gosain, A.K. and Ramasastri, K.S. 2002. A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrol. Proces.*, 16(6): 1325-1330.

Tiwari, M.K. and Chatterjee, C. 2010[a]. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). *J. Hydrol.*, 382: 20-33.

Tiwari, M.K. and Chatterjee, C. 2010[b]. Development of an accurate and reliable hourly flood forecasting model using Wavelet-Bootstrap-ANN hybrid approach. *J. Hydrol.*, 394: 458-470.

Welderufael, W.A., Le Roux, P.A.L. and Hensley, M. 2009. Quantifying rainfall-runoff relationships on the Melkassa Hypo Calcic Regosolecotope in Ethiopia. *Water Sa*, 35(5).