

Rainfall Runoff Modeling using Gene Expression Programming and Artificial Neural Network

Raviraj Singh, Sunil Ajmera

Abstract: In water resource management and planning the Rainfall-Runoff models play a crucial role and depends mainly on the data available for planning activities. The rainfall-runoff relationship comes under the nonlinear and complex hydrological Event. In the present study two data driven modeling approaches, Artificial Neural Network (ANN) and Gene Expression Programming (GEP) has been used for modeling of rainfall-runoff process as these methods does not consider the physical nature of the process, which is complex to understand. GEP and ANN are used to model rainfall-runoff relationship for Dindori catchment in upper Narmada River Basin. Daily hydro-meteorological data of Dindori gauging station and precipitation of the catchment for a period of eighteen years were used as input in the model design. Various combinations of input variables for training and testing of models were selected based on statistical parameters. The performance of model was evaluated in term of the root mean square error (RMSE), coefficient of determination, RMSE to standard deviation ratio (RSR) and Nash Sutcliffe Efficiency. The results obtained after applying the two techniques were compared. Which indicates that GEP performed better in all performance evaluation parameters (R^2 is 0.92) then ANN (R^2 0.90) and is able to give mathematical relationship for rainfall-runoff modeling.

Keywords - Gene Expression Programming, Artificial Neural Network, Rainfall-Runoff.

I. INTRODUCTION

In response to these water challenges, hydrological phenomenon have been developed to examine, comprehend and develop solutions for water management. Hydrological model is collection of several processes (e.g. precipitation, evapotranspiration, ground water, and runoff) which represent real world system and help in managing and predicting water resources.

The rainfall-runoff process is one of the sophisticated hydrological phenomena and considered as central problem in hydrology. The process is not easily described by simple model due to temporally and spatially distributed watershed characteristics and precipitation patterns. Experts in water assets have utilized information driven displaying approaches, as these have been found to conquer a portion of the troubles related with physical based model. Such approaches have the capacity to show the precipitation overflow process without the point by point comprehension of the complex physical qualities of catchment. In the most recent decade, Progression in the field of Artificial Intelligence (AI) has impacts numerous

science themes just as water resources engineering applications [5]. The use of AI techniques, such as Artificial Neural Network (ANN) and Genetic Programming (GP) has become feasible in water resources engineering. Until now, various works have been accounted in literature with respect to function of ANN, in rainfall-runoff modeling (e.g., Hsu et al., 1995 [13]; Tokar and Markus, 2000 [24]; Rajurkar et al., 2002 [22]; Shrinivasulu and jain, 2005 [23]; Machado et al., 2011 [19]; Chakravarti et al., 2015 [7]; Hussain et al., 2017 [15]; Chankaln et al., 2018 [8]; Hussain et al., 2019) [14].

Gene Expression Programming (GEP) was first presented by Ferreira (2001) [11], as an expansion of Genetic Programming. GEP is a moderately new space free technique for developing PC programs in roughly taking care of issues and produces a 'straightforward' and organized portrayal of the framework being examined. GEP outperforms the GP framework in hundred to thousand times [11].

Therefore the objective of the present study is to evolve rainfall-runoff model, which requires lesser types of data using GEP and ANN techniques. Also to provide the mathematical relationship for the same using GEP technique and compare the accuracy of GEP and ANN model. These technique can be Efficient to perform and are not stochastic in nature.

II. LITERATURE REVIEW

In 2008, Aytek et al. proposed a utilization of two systems of Artificial Intelligence for precipitation runoff demonstrating: the ANN and the transformative calculation (EC). Two different ANN procedures are contrasted and Gene Expression Programming GEP which is another transformative calculation that develops programs. Their outcomes demonstrate that GEP can be proposed as an option to ANN models and performed well in examination. In 2011, Fernando et al. explored the utilization of GEP to create one-day-ahead stream anticipating models for catchments with generally contrasting qualities. Four GEP models were produced for four catchments that demonstrated precise gauges fit.

In 2012, Kisi et al. demonstrated precipitation overflow process for a little catchment in Turkey utilizing Neuro-Fuzzy, GEP and ANN technique. The models were prepared and tried utilizing different mixes of the autonomous factors. The investigation gives proof that GP is fit for displaying precipitation spillover process and is a feasible option in contrast to other applied artificial insight and MLR time-arrangement strategies.

Revised Manuscript Received on February 5, 2020.

Raviraj Singh, M.E student, civil engineering department, Shri. G.S. Institute of Technology & Science, Indore, Madhya Pradesh, India. singhravigsits@gmail.com.

Dr. Sunil Ajmera, Professor in civil engineering department, Shri. G.S. Institute of Technology & Science, Indore, Madhya Pradesh, India. sbajmera@gmail.com.

In 2014 Hasmi et al. Utilized GEP to perform emblematic relapse for building up a parametric plan of stream span bend (FDC) regionalization, to relate chose FDC qualities to catchment attributes. Study demonstrated that utilization of this man-made consciousness system facilitates the choice of a lot of the most applicable autonomous factors out of an enormous set, in light of the fact that these are consequently chosen through the GEP procedure.

In 2016 Phukoetphim et al. contrasted the presentation of GEP and diverse neural system mix techniques when utilized in the improvement of multimodal frameworks. The strategies were utilized to join the outcomes from various sorts of precipitation overflow models to test the multimodal. Examination of the outcomes uncovered that the GEP performed superior to neural system strategies on account of the catchment situated in New Zealand.

In 2017 Abdollahi et al. analyzed and thought about the presentation of four man-made reasoning procedures including ANN, cross breed wavelet-ANN, GEP, and half and half wavelet-GEP for every day mean stream expectation. Their examination indicated that in spite of the fact that the GEP model was the most exact in anticipating top streams, yet in general among the four referenced models in both lasting and non-enduring waterways, WANN had the best execution.

In 2018 Ercan et al. estimated the flow under influence of different meteorological parameters using GEP, ANN and Regression analysis program. During the analysis, precipitation, humidity and temperature were used as input parameters and discharge was used as output parameter. Their study demonstrated a good capturing skill of GEP driven flow estimations relative to observation data and model results.

III. MATERIAL AND METHODS

A. Study Area and data used

The Narmada River is located in central India with stretch over Madhya Pradesh and Gujrat state having a catchment area of 98,796 km² (38,145.3 sq mile), circumscribed by the Satpura and Vindhya Mountain Ranges. It lies between longitudes 72°32' and 81°45' east and latitude 21°20' to 23°45' north. Dindori catchment area at upper most mountainous region of Narmada River, located in Dindori district of Madhya Pradesh has been taken. The catchment area of Narmada River up to Dindori gauging site from its origin (Amarkantak hills) is 2292 km². Gauging site is located at 22°57' north; 81°04' east in dindori district. The daily rainfall Gridded data of

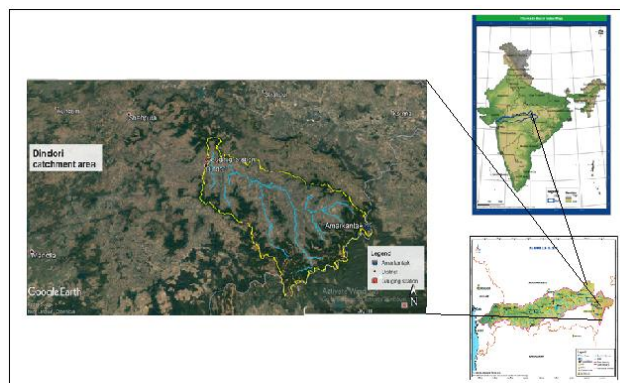


Fig 1: Location map of study area (Dindori catchment)

CFSR reanalysis product for the study area is extracted from website of Global weather data for SWAT for 18 years. The daily discharge data in cumec for dindori gauging site is obtained from the government authorized body. The whole data were separated into training and testing data set for 12 and 8 years individually. So as to choose the best possible composition of precipitation (P) and discharge (Q), the Auto-correlation of discharge with one day lag and cross correlation between the P and Q is resolved and coefficients of the functions are shown in Table I. The Table I and Fig 2 and 3 shows that the rainfall runoff process with lag of two time series of P and Q data will fulfill the process.

Table I: Cross- Correlation and Partial Auto-correlation coefficient values between discharge and Precipitation data

Function	Time series	Coefficient
Cross correlation	$P_{t-1} - Q_t$	0.541
	$P_{t-2} - Q_t$	0.550
	$P_{t-3} - Q_t$	0.396
Partial Auto-correlation	$Q_{t-1} - Q_t$	0.810
	$Q_{t-2} - Q_t$	0.027
	$Q_{t-3} - Q_t$	0.139

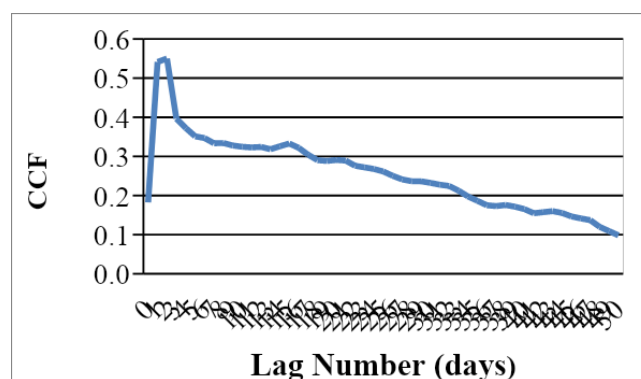


Fig 2: Cross-Correlation function (CCF) of rainfall with discharge (lag in days)

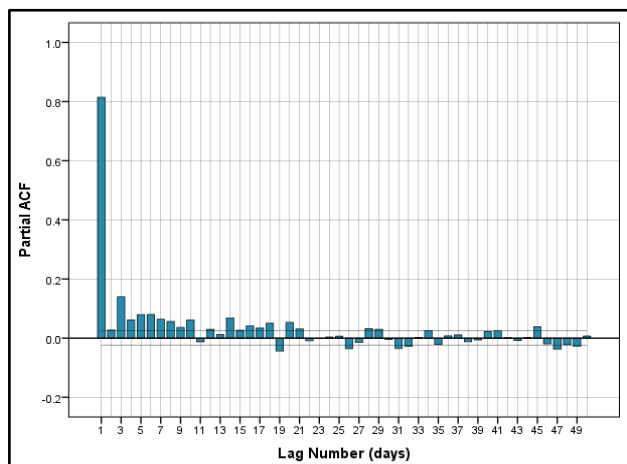


Fig 3: Partial Auto-Correlation function (PACF) of discharge (lag in days)

B. Artificial Neural Network

ANN is the computational methodology roused from natural sensory system. Neural system acts like human mind that shows the capacity to review, take in and sum up from training patterns. Preparing design. ANN models comprise of a puddle of straightforward handling units called neurons which convey by sending sign to each other over weighted associations. The Back propagation algorithm (BPA) is most widely recognized learning rule for Multi-layer perceptrons, which is utilized in this examination and the network framed through this algorithm is known as Feed Forward Back Propagation Neural Network (FFBPN). In FFBPN error in first iteration is calculated and is back propagated to the network so that error can be minimized in next iteration. Fig 4 shows the network of FFBPN used in the study for rainfall runoff relationship. Itemized application of ANN in hydrology can be found in ASCE (2000) [1].

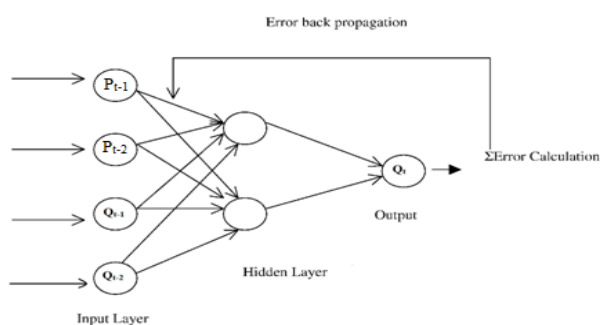


Fig 4: Back propagation neural network

The architecture of network is represented as (l-m-n), where l is number of input layer, m is number of hidden layer and n is number of output layer. The input layer consist of input nodes representing input variables rainfall and runoff with time series lag (Pt-1..., Qt-1, Qt-2...), one neuron on output layer as runoff (Qt) and various combination of neuron in hidden layer are employed to model the process which are connected through nodes. The activation function used in the study for the hidden node and output nodes to determine the weightage of variables is Tan sigmoid.

C. Gene Expression Programming Modeling

GEP is the learning technique that realizes explicitly about connections among sets of variable and afterward frame models to state these connections. Computer program of various shapes and size encoded in straight chromosomes or genome of rigid length [11]. The chromosomes includes multiple gene, every gene encoding a littler subprogram. The auxiliary and utilitarian association of the linear chromosomes permits the activity of significant hereditary administrators, for example, transformation, recombination and transposition. The GEP is genotype/phenotype in nature .The quality of the GEP approach is because of its hereditary administrators which work at chromosomes level that make the formation of hereditary decent variety very simplified. The quality of GEP is its extraordinary, multigenic nature which permits the development of progressively complex projects made out of a few subprograms. Therefore GEP perform superior to the GP system [18].

GEP utilizes a similar sort of diagram portrayal of GP, however the substances delivered by GEP (Expression tree) are the statement of chromosomes or genome. In GEP the phenotype limit was crossed, giving new and proficient answers for developmental calculation. GEP resembles GAs and GP as it utilizes population of individual, chooses them as per fitness, and presents genetic variety utilizing at least one genetic operator.

The major steps for symbolic regression or function finding as in preparing a model by using gene expression programming are:

- **Fitness function Selection**
- **Electing terminals and functions**
- **Choosing the linking function and chromosomal architecture**
- **Selecting operators**

D. Performance Evaluation

Four statistical measures were used to determine and compare the model performance and analyze the result, which are as follows:

Root mean square error (RMSE), coefficient of determination (R^2), RSR (RMSE-SD Ratio) and Nash Sutcliffe Efficiency (NSE).

- 1) Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_{im} - Q_{io})^2}{N}} \quad (1)$$

- 2) The coefficient of determination (R^2)

$$R^2 = \left[\frac{\sum (Q_o - Q_{om})(Q_m - Q_{mm})}{\sqrt{\sum (Q_o - Q_{om})^2 \sum (Q_m - Q_{mm})^2}} \right]^2 \quad (2)$$

- 3) RSR (RMSE-SD Ratio)

$$RSR = \frac{RMSE}{SD} \quad (3)$$

4) Nash Sutcliffe Efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_{io} - Q_{im})^2}{\sum_{i=1}^N (Q_{io} - Q_{om})^2} \quad (4)$$

Where Q_{im} and Q_{io} denote the modeled and observed runoff values, Q_{mm} and Q_{om} denotes the mean modeled and observed runoff values, respectively. N denotes number of samples.

IV. RESULTS AND DISCUSSION

A. Rainfall Runoff modeling using ANN

Table II: Testing statistics of ANN model

Input combinations	RMSE (m ³ /sec)		R ²		RSR		NSE	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
P _{t-1}	44.43	37.76	0.552	0.576	0.69	0.59	0.55	0.56
P _{t-1} , P _{t-2}	39.29	33.44	0.654	0.667	0.62	0.52	0.64	0.65
P _{t-1} , P _{t-2} , P _{t-3}	40.84	34.67	0.642	0.651	0.64	0.54	0.63	0.64
P_{t-1}, P_{t-2}, Q_{t-1}	22.67	19.84	0.864	0.901	0.36	0.32	0.85	0.87
P _{t-1} , P _{t-2} , Q _{t-1} , Q _{t-2}	23.93	20.96	0.855	0.882	0.38	0.36	0.84	0.86

B. Rainfall Runoff modeling using GEP

For constructing the structure of expression tree for GEP models, Default function set are applied. Some of them are +, -, x, /, √, x², x³, ln, ∛, sin(x), cos(x), log(x), Min and Max with RMSE as fitness function. Optimal Evolution strategy was assigned with set of genetic operator for optimizing the run. The various genetic operator with their values assigned for training are shown in Table III. For modeling rainfall-runoff process through GEP, GeneXpro tool has been used in present study.

Optimal values for GEP Model for training and testing period through statistical indexes for different input combinations are shown in Table IV. The results indicate that combination P_{t-1}, P_{t-2}, Q_{t-1} is adequate for modeling with GEP architecture having 30 chromosomes, 3 genes and head size 8.

Analysis of result by statistical measures for different combination for ANN model is shown in Table II through the RMSE, R², RSR and NSE statistical indexes. The performance is checked for different input combination for training and testing period. From the Table II it can be concluded that combination (P_{t-1}, P_{t-2}, and Q_{t-1}) with architecture (3-5-1) gave best results for rainfall-runoff modeling by ANN, as it shows better statistical indexes values in comparison to other combination with different architecture of neural network.

In addition to that GEP model gives Mathematical expression between the input and output variables in form of Expression tree. Expression of the GEP model for best combination (P_{t-1}, P_{t-2}, Q_{t-1}) is decoded from three sub-ET.

$$Q_t = ((\text{Max}((\text{Log}(Q_{t-1}) * P_{t-1}), (P_{t-2} - 6.71)) + (1.0 - (6.59 - Q_{t-1}))/2.0) + \text{Max}(((Q_{t-1} + 5.45)/2.0) + \text{Log}(Q_{t-1}), (\text{Atan}(P_{t-2}) * (-9.68) + P_{t-1}))) + (\text{power}(((P_{t-1} + P_{t-2}) * Q_{t-1}), (1/3)) + (\text{Log}(Q_{t-1}) * (Q_{t-1} / (-9.58)))) \quad (5)$$

Here Min, Max and Atan represent the minimum, maximum and inverse tangent function respectively. Q_t denotes simulated discharge.

Table III: Genetic operator employed for training of the GEP model

Genetic operators			
Number of chromosomes	30	Inversion rate	0.054
Head size	8	1-point recombination rate	0.075
Number of genes	3	2- point recombination rate	0.027
Linking function	Addition	Gene recombination rate	0.027
Fitness function	RMSE	Gene transposition rate	0.054
Mutation rate	0.013	Insertion sequence transposition rate	0.054

Table IV: Testing statistics of GEP model

Input combinations	RMSE (m ³ /sec)		R ²		RSR		NSE	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
P _{t-1}	42.07	34.78	0.577	0.596	0.66	0.55	0.57	0.58

P_{t-1}, P_{t-2}	37.26	31.25	0.679	0.688	0.58	0.49	0.66	0.68
$P_{t-1}, P_{t-2}, P_{t-3}$	39.58	32.85	0.664	0.677	0.62	0.52	0.65	0.67
$P_{t-1}, P_{t-2}, Q_{t-1}$	18.78	16.37	0.887	0.921	0.29	0.25	0.87	0.91
$P_{t-1}, P_{t-2}, Q_{t-1}, Q_{t-2}$	19.32	17.01	0.875	0.905	0.30	0.26	0.87	0.89

The comparison of results for optimal input combination during testing period for GEP and ANN model are shown in Table V. Fig 5 shows the observed and simulated discharge runoff value for GEP and ANN model through the scatterplots.

From Table V and Fig 5 it is clear that GEP model gave better results as compared to ANN for rainfall-runoff modeling, as GEP shows higher R^2 (0.92) and NSE (0.91) values with lesser error as RMSE (16.37 cumec) for testing period.

Table V Testing statistics of optimal GEP and ANN model

Period	GEP				ANN			
	RMSE	R^2	RSR	NSE	RMSE	R^2	RSR	NSE
Training	18.78	0.89	0.29	0.87	22.67	0.86	0.36	0.85
Testing	16.37	0.92	0.25	0.91	19.84	0.90	0.32	0.87

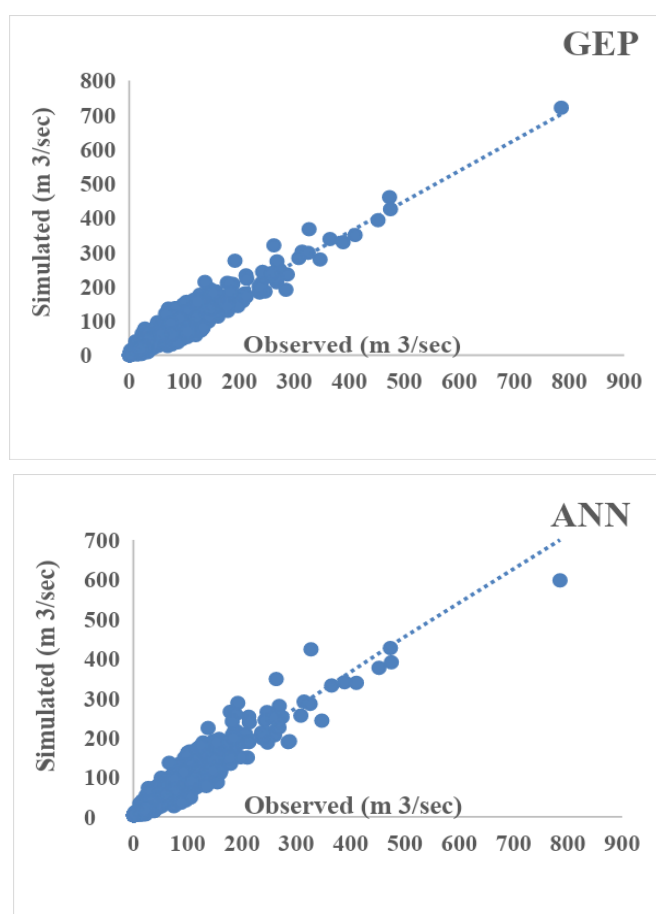


Fig 5: Scatter plots for GEP and ANN model in testing period

V. CONCLUSION

In present paper two data driven techniques ANN and GEP are accounted for developing Rainfall-Runoff model. The objective is to develop the techniques which requires lesser data for rainfall runoff modeling and are easy to perform.

The objective seems to be fulfilled as the proposed methodology based on GEP and ANN are found to perform satisfactorily with the available data for gauged catchment of Narmada River on Dindori gauging site.

The comparison of the results demonstrated that the GEP model is better than ANN model in present case. As GEP can give the expression among the variable of examined procedure while the ANN model is not able to show the interrelationships among the variables, GEP outperformed the ANN technique.

Other relevant optimization technique such as Particle Swarm can also be utilized, other data driven technique such as ANFIS and hybrid of the technique can also be used and accuracies of their results may be evaluated and compared with those of GEP.

REFERENCES

1. ASCE Task Committee on application of ANN in hydrology (2000), "Artificial Neural Networks in Hydrology II: Hydrologic Applications.
2. Abdollahi, S., Raeisi, J., Khalilianpour, M., Ahmadi, F., & Kisi, O. (2017). "Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques". *Water resources management*, 31(15), 4855-4874.
3. Antar, M. A., Ellassiouti, I., & Allam, M. N. (2006). "Rainfall-runoff modeling using artificial neural networks technique: a Blue Nile catchment case study". *Hydrological Processes: An International Journal*, 20(5), 1201-1216.
4. ASCE Task Committee on application of ANN in hydrology (2000), "Artificial Neural Networks in Hydrology II: Hydrologic Applications.
5. Aytek, A., & Alp, M. (2008). "An application of artificial intelligence for rainfall-runoff modeling". *Journal of Earth System Science*, 117(2), 145-155.
6. Babovic v., Maarten Keijzer, (2002) "Rainfall Runoff Modelling Based on Genetic Programming". *Hydrology Research* 1 October. 33 (5): 331-346.
7. Chakravarti, A., Joshi, N., & Panjari, H. (2015). Rainfall Runoff Analysis Using the Artificial Neural Network. *Indian Journal of Science and Technology*, 8(14), 1-7.
8. Chanklan, R., Kaoungku, N., Suksut, K., Kerdprasop, K., & Kerdprasop, N. (2018). Runoff Prediction with a Combined Artificial Neural Network and Support Vector Regression. *International Journal of Machine Learning and Computing*, 8(1).
9. Ercan, B., Yagci, A.E., Yilmaz, A.S., Yuce, M.I. and Unsal, M., (2018). "Annual Runoff Modelling Of Kizilirmak Basin by Artificial Intelligent Techniques". *Fresenius Environmental Bulletin* Volume 28 ± No. 9/2019 pages 6651-6660.
10. Fernando, A. K., Shamseldin, A. Y., & Abrahart, R. J. (2011). Use of gene expression programming for multimodel combination of rainfall-runoff models. *Journal of Hydrologic Engineering*, 17(9), 975-985.
11. Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027*.
12. Hashmi, M. Z., & Shamseldin, A. Y. (2014). Use of gene expression programming in regionalization of flow duration curve. *Advances in water resources*, 68, 1-12.
13. Hsu, K. L., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, 31(10), 2517-2530.
14. Hussain, A., Singh, J. K., Kumar, A. S., & Harne, K. R. (2019). Rainfall-Runoff Modeling of Sutlej River Basin (India) Using Soft Computing Techniques. *International Journal of Agricultural and Environmental Information Systems (IAEIS)*, 10(2), 1-20.
15. Hussain, D., Usmani, A., Verma, D. K., Jamal, F., & Khan, M. A. (2017). Rainfall Runoff Modelling Using Artificial Neural Network. *International Journal of Advance Research, Ideas and Innovations in Technology*, 3(6), 1528-1533.
16. Jayawardena, A. W., Muttill, N., & Fernando, T. M. K. G. (2005). Rainfall-runoff modelling using genetic programming. In *International Congress on Modelling and Simulation, Modelling and Simulation, MODSIM05 Proceedings. Modelling and Simulation Society of Australia & New Zealand*.
17. Karimi, S., Shiri, J., Kisi, O., & Xu, T. (2018). Forecasting daily streamflow values: assessing heuristic models. *Hydrology Research*, 49(3), 658-669.
18. Kisi, O., Shiri, J., & Tombul, M. (2012). Modeling rainfall-runoff process using soft computing techniques. *Computers & Geosciences*, 51, 108-117.
19. Machado, Miriam Mine, Eloy Kaviski & Heinz Fill (2011) Monthly rainfall-runoff modelling using artificial neural networks, *Hydrological Sciences Journal*, 56:3, 349-36

20. Nayak, P. C., Venkatesh, B., Krishna, B., & Jain, S. K. (2013). Rainfall-runoff modeling using conceptual, data driven, and wavelet based computing approach. *Journal of Hydrology*, 493, 57-67.
21. Phukoetphim, P., Shamseldin, A. Y., & Adams, K. (2016). Multimodel approach using neural networks and symbolic regression to combine the estimated discharges of rainfall-runoff models. *Journal of Hydrologic Engineering*, 21(8), 04016022.
22. Rajurkar, M. P., Kothari, U. C., & Chaube, U. C. (2002). Artificial neural networks for daily rainfall-runoff modelling. *Hydrological Sciences Journal*, 47(6), 865-877.
23. Srinivasulu, S., & Jain, A. (2006). A comparative analysis of training methods for artificial neural network rainfall-runoff models. *Applied Soft Computing*, 6(3), 295-306.
24. Tokar, A. S., & Markus, M. (2000). Precipitation-runoff modeling using artificial neural networks and conceptual models. *Journal of Hydrologic Engineering*, 5(2), 156-161.

AUTHORS PROFILE



Mr. Raviraj Singh, I have completed Bachelor degree in Civil Engineering from Rajiv Gandhi proudyogiki vishwavidyalaya Bhopal and presently a student of post-graduation in Water resources engineering from The Shri. G.S institute of technology and science, Indore. Having one year of teaching experience with GATE qualified. Area of interest is water management and conservation. Completed training from "Construction Division No.1, CPA, Bhopal" on Building estimate construction during under graduation project. Student member of Indian Water Resources Society (IWRS), student chapter Indore. Student member of Indian Water Works Association, student chapter Bhopal. Participated and worked as coordinator in various Technical college level events like paper presentation, Think Quest.



Dr. Sunil Ajmera, Presently working as Professor in Civil Engineering and Applied Mechanics Department of Shri. G.S. Institute of Technology & Science, Indore, Madhya Pradesh, India. Teaching UG Students of Civil Engineering and PG students of Water Resources Engineering specialization and Supervising Thesis Dissertation and Project work of the students, having total teaching experience of more than 33 years. Member of many professional societies and actively involved in various professional societies in different capacity. Presently working as faculty Advisor of Indian Water Resources Society Student Chapter at SGSITS Indore. Area of Research and Consultancy is Water Resources Engineering and Construction Technology. Published 09 research papers in national and International Journals.