

Runoff modeling in Kolar river basin using hybrid approach of wavelet with artificial neural network

Deepak Kumar Tiwari , Hari Lal Tiwari  and Raman Nateriya

Department of Civil Engineering, Maulana Azad National Institute of Technology, Bhopal 462003, Madhya Pradesh, India

*Corresponding author. E-mail: deepaktiwarcivil7sept@gmail.com

 DKT, 0000-0002-9253-1153; HLT, 0000-0002-6817-6999

ABSTRACT

In this paper, the Kolar River watershed, Madhya Pradesh is taken as the study area. This study area is located in Narmada River in Central India. The data set consists of monthly rainfall of three meteorological stations, Ichhawar, Brijesh Nagar, and Birpur rainfall stations from 2000 to 2018, runoff data at Birpur and temperature data of Sehore district. In this paper, radial basis function neural network models have been studied for generation of rainfall–runoff modeling along with wavelet input and without wavelet input to the RBF neural network. A total of 15 models were developed in this experiment based on various combinations of inputs and spread constant of RBF model. The evaluation criteria for the best models selected are based on R^2 , AARE, and MSE. The best predicting model among the networks is model 8, which has input of $R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$, and $Q(t-1)$. For the RBFNN model, the maximum value of R^2 is 0.9567 and the lowest values of AARE and MSE are observed. Similarly, for the WRBFNN model, the maximum value of R^2 is 0.9889 and the lowest values of AARE and MSE are observed. WRBF performs better than RBF with any data processing techniques which shows the proposed model possesses better predictive capability.

Key words: ANN, rainfall–runoff modeling, RBF, watershed management, wavelets

HIGHLIGHTS

- Fifteen ANN model used for analysis.
- New data processing technique of wavelets used.
- Analysis of Kolar river basin, main tributary of Narmada in Central India.
- Radial basis function used for modeling.
- Suitable method for data scarce region and semiarid environment.

INTRODUCTION

Rainfall–runoff modeling is a versatile tool for water resource planning and management, city planning, flood, land use, etc. (Ghumman *et al.* 2011). It also helps in minimizing the effect of drought-related issues on water resources. Owing to change in climatic conditions in recent decades because of global warming, the hydrological cycle in India has been adversely affected (Sonali & Nagesh Kumar 2016a) and due to anthropogenic activities there is an increase in global surface temperature which is clear from various evidence (Sonali & Nagesh Kumar 2016b). Many hydrological models have been developed since 1850, ranging from mathematical relations between them, empirical models, conceptual ones in which the physical processes are based on physical laws existing in nature and physical models which are small-scale prototypes of the models (Todini 2007). Conceptual and physical models account for all the physical processes involved in the catchment but they are very data-intensive and time-consuming (Sušanjan *et al.* 2016). These types of models are not suitable for areas which suffer from data scarcity and poorly managed data sites. General time series models such as auto regressive integrated moving average (ARIMA) are popularly used for hydrological time series forecasting (Nourani *et al.* 2009), but Mujumdar & Kumar (1990) suggested that ARIMA should be avoided due to an increase in variance continuously on differencing the series. Also, these models are linear in nature and do not account for the non-stationarities and non-linearities in hydrologic time series data. ARIMA is commonly used for the hydrological time series data which have periodicity over time (Zhang *et al.* 2011). Moreover, seasonal variability is also one of the climatic factors for runoff variation (Bekele *et al.* 2021).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

Nowadays, ANN models are popularly used to develop the rainfall–runoff (RR) relationship and they are black-box models which are data-driven and give the relation between rainfall and runoff without considering the physical explanation of processes involved (Todini 2007). These are mathematical models with repetitive iterations which help in the development of some non-linear relationships between two hydrological phenomena, rainfall and runoff (Poonia & Tiwari 2020). They also do not require prior knowledge of physical processes and morphometry of the basin for prediction. ANN consists of three layers: input layer, hidden layer, and output layer (Figure 1). The input layer consists of the number of nodes equal to input parameters. The hidden layer consists of the mathematical formula for processing of inputs to find which matches best to parameters which are to be predicted. Many times, the input data non-stationarity or dynamic space time variability for data makes it difficult to predict (Sharghi *et al.* 2018). Also, predicting the capability of hybrid models, data processing techniques combined with ANN gives better results (Barman & Bhattacharjya 2020). Among the publications regarding ANN used for runoff prediction, most of them have focused on back propagation neural networks (Asadi *et al.* 2013). Multilayer perceptron (MLP) and radial basis function (RBF) are the two most widely used neural networks in rainfall runoff analysis (Phukoetphim *et al.* 2016; Tayyab *et al.* 2016; Shoaib *et al.* 2018; Poonia & Tiwari 2020), but RBFNN techniques provide better solutions as compared to MLP methods (Kumar & Yadav 2011). Moreover, data quality is key to better predicting capability of the model. In recent years, wavelet transforms have emerged as an excellent tool for predicting (Krishna *et al.* 2011; Badrzadeh *et al.* 2015; Alizadeh *et al.* 2017a). Wavelet transform reduces the data noising in natural data and soothes data range by driving data to multiresolution and on different scales.

Radial basis function models have been used for flow forecasting during the last two decades using Gaussian function for network processing (Dawson *et al.* 2002). Lin & Chen (2004) applied the RBF network for runoff analysis for hourly prediction and successfully applied it for determining the complex relationship between rainfall and runoff. Later, a study on the internal functioning of RBF and its hydrological interpretation was conducted (Achela *et al.* 2009). The results of the study show that a single hidden layered RBFNN is an effective tool to forecast daily flows and that the activations of the hidden layer nodes are far from arbitrary, but appear to represent flow components of the predicted hydrograph. Miaoli *et al.* (2020) used the Levenberg–Marquardt (LM) algorithm for training RBF neural networks which was really an innovative approach to RBF networks. It was found that the LM algorithm makes a network increase in efficiency by carrying out more error value, improves convergence speed, reduces the storage space, and can be successfully applied to problems in various fields. Poonia & Tiwari (2020) also applied the RBF network for runoff simulation. However, with wavelet processed input data, RBFNN models are one of the innovative approaches attempted in this study. Researchers have identified wavelets as excellent data processing techniques to increase the efficiency of any network (Nourani *et al.* 2011).

In this study, RBF, a branch of ANN, has been employed to predict the RR relationship for Kolar River, Madhya Pradesh, India. Also, to counter the non-stationarity, wavelet preprocessing is performed to input data and compared with simple RBF models. A total of 15 models based on input combinations of runoff, rainfall, and temperature have been developed for RBF models and wavelet RBF (WRBF) and compared based on three evaluation criteria, namely, coefficient of determination (R^2), mean square error (MSE), and root mean square error (RMSE).

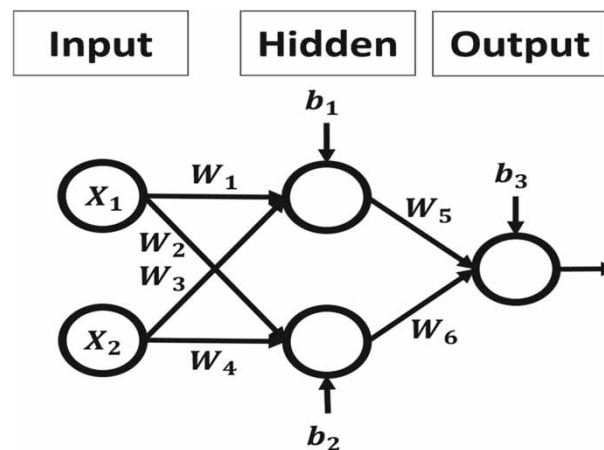


Figure 1 | A general schematic view of the three-layered artificial neural network.

STUDY AREA AND DATA SET

The study area is Kolar river catchment, which rises in the Vindhya range of Sehore district and flows in a south-westerly direction to meet the Narmada near Nasrullahganj in the Raisen district of Madhya Pradesh. The Kolar river basin has an upper elevation of 600 m, elevation of 432 m at the dam and downwards until Narmada, and 350 m at the downstream. The main purpose of the Kolar dam is to supply water for Bhopal city, irrigation for Jholiapur area of Raisen district, and fishing activities. It is also a tourist attraction for people nearby. Kolar catchment is situated in the Sehore district in the state of Madhya Pradesh, India. The salient features for the catchment are given in Table 1, and Figure 2 presents the Kolar catchment geographical map.

Data collection

The present study uses the daily data of rainfall for 30 years from 1988 to 2018 for three sites, namely, Birpur, Brijesh Nagar, and Ichhawar, in millimeters. Also, average daily data of temperature is used for the same period in degrees Celsius. Data of average daily discharge in cubic meters per second were available from October 1999 to October 2018 (19 years and 3 months) and provided by Kolar dam authority, a state government body associated with the management of the dam. Figure 3 shows the daily rainfall measurements of three rainfall stations, Birpur, Brijesh Nagar, and Ichhawar, in millimeters. Here, places with high spatiotemporal rainfall variability, such as mountain regions as in Kolar region, have a variety of slopes from very steep to mild, thus, input data can be a large source of uncertainty in hydrological modeling. Modeling of mountain regions is highly uncertain as it contains data of spatially distributed rainfall and also depends on the density of rain gauge stations. For the temperature data set, the nearest daily data were available at Birpur site. These temperature data were validated with the nearest available data at Bhopal airport, which is about 30 km from the site.

METHODOLOGY

In this study, the RBFNN model is used for RR modeling of Kolar river basin, located around 20 km from Bhopal, the capital city of Madhya Pradesh state, India. First, daily data for rainfall, runoff, and temperature were converted into monthly scale and then normalized in the range of 0.1 to 0.9 to overcome any abnormalities due to data range (Seo *et al.* 2015). After normalization of input, the RBFNN model is applied with 15 different combinations of inputs. Again, the input data are post-processed with wavelet transformation application and the RBFNN model is used for RR modeling. This network is termed the WRBFNN model. These models are less complex, with a lower number of input nodes, single hidden layers of ten neurons, and output layer with a single neuron. Data required for the model are divided into 70, 15, and 15% for training, testing, and validation stages as standard percentages given in MATLAB. Other researchers have used different percentages of 60, 20, and 20% also (Nourani & Komasi 2013).

Normalization

Normalization of input and output data safeguards enhanced computation efficiency, data integrity, and redundancy elimination (Nawi *et al.* 2013). Before applying rainfall, discharge, and temperature data to the models, data were normalized between 0.1 and 0.9 to remove any error caused due to null value of normalized data as given by Equation (1):

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X_n is the normalized value of the data set, X_i is the actual value of the data set, and X_{min} and X_{max} are the minimum and maximum values of the data set, respectively.

Table 1 | Characteristics of the three weather stations

Station	Latitude	Longitude	Elevation (m)	Avg Tmax (c)	Avg Tmin (c)	Rainfall (mm)	Climate
Birpur	22°58'	77°20'	441	31.88	18.76	940	Semi-arid
Brijesh Nagar	22°57'	77°08'	505	30.66	16.45	1,370	Humid
Ichhawar	23°01'	77°01'	515	32.24	17.8	1,032	Dry – Sub humid

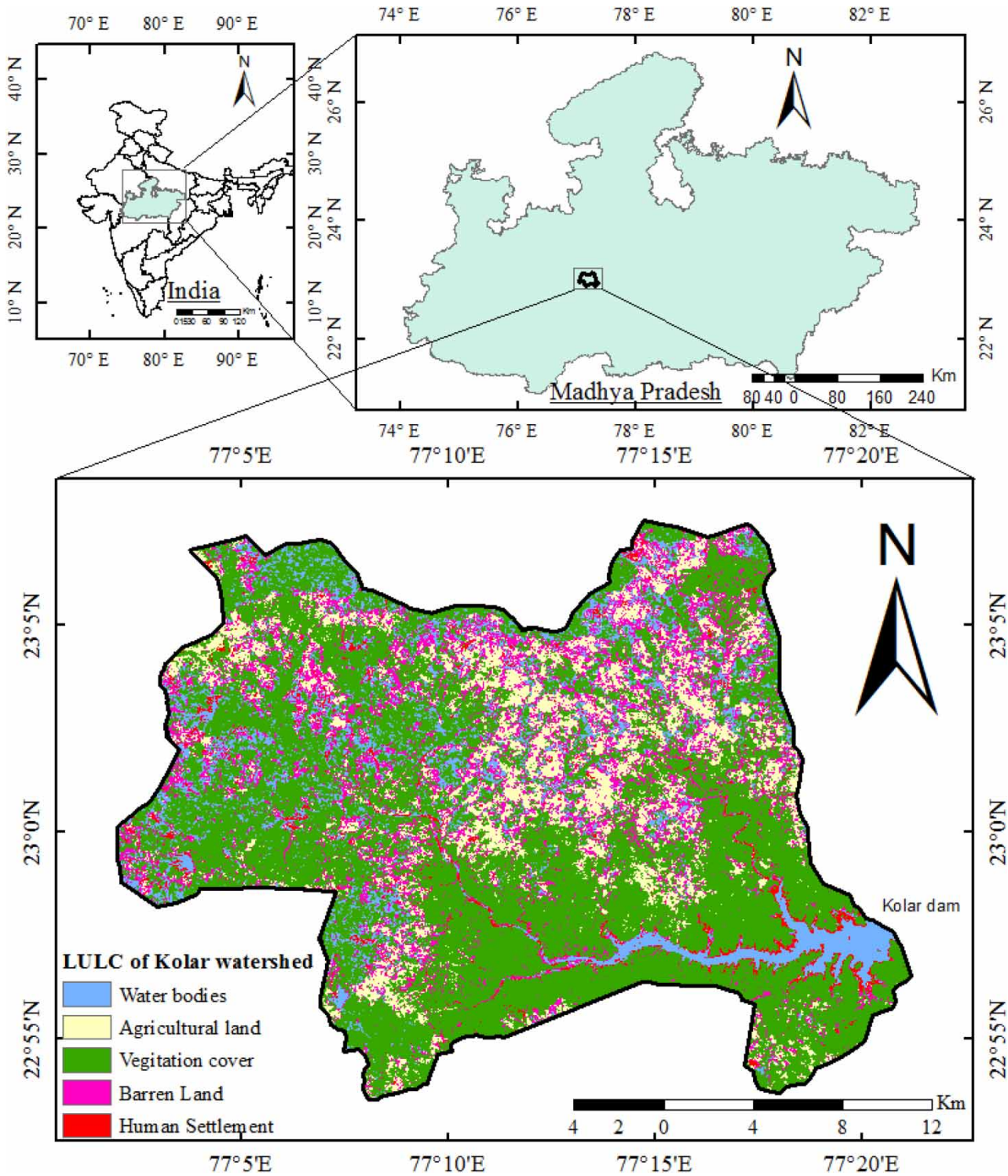


Figure 2 | Kolar River watershed map.

Radial basis function NN (RBFNN)

The radial basis function network (RBFN) has been used for non-linear pattern recognition because of its simple structural arrangement and its ability to predict how learning proceeds in an explicit manner (Shamseldin 2009). RBFN was initially introduced to solve the real multi-variable interpolation problem (Powell 1987). Broomhead & Lowe (1988) proposed the

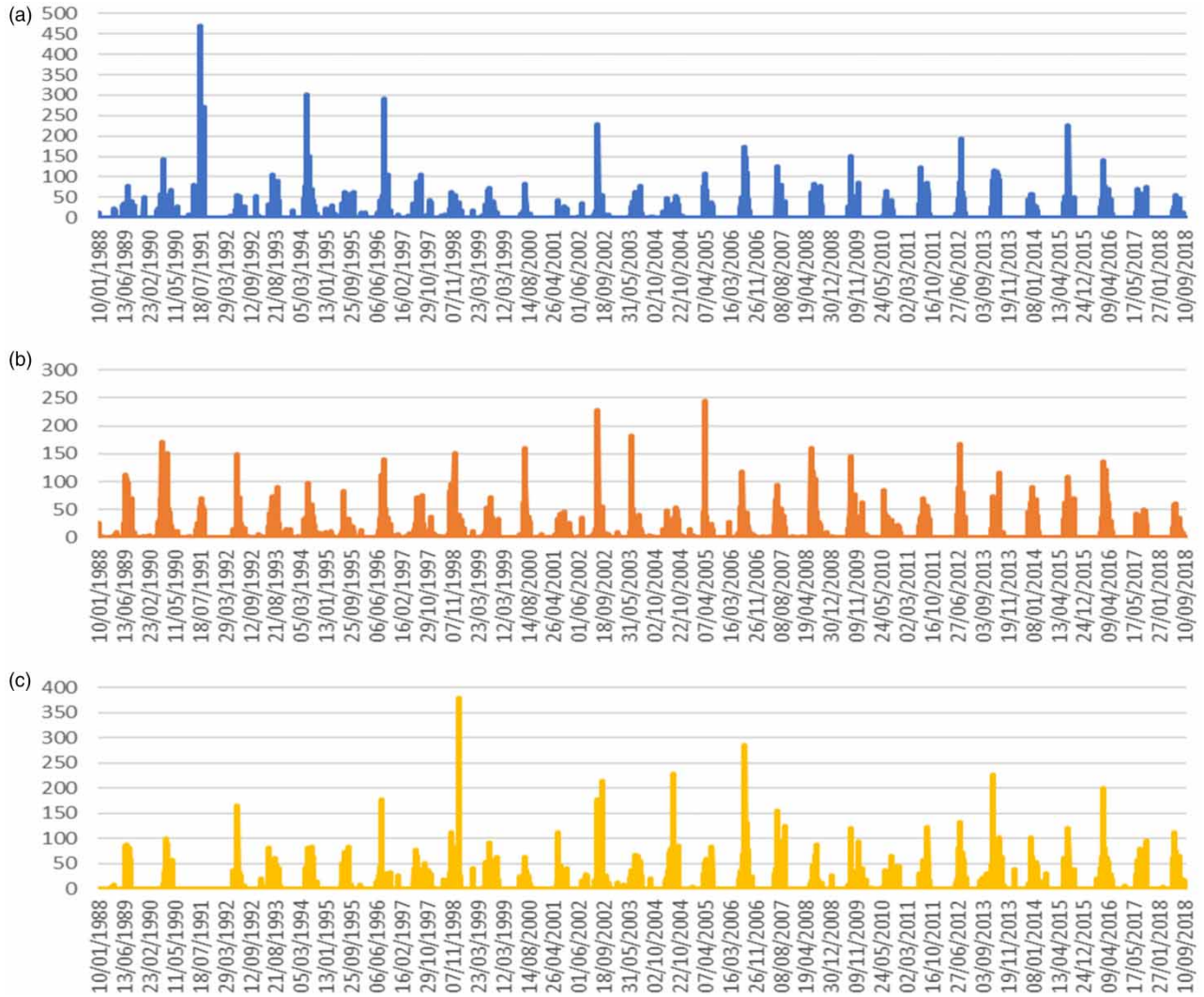


Figure 3 | Daily rainfall data of (a) Birpur, (b) Brijsh Nagar, and (c) Ichhawar rain gauge stations.

RBF network which consists of three layers: an input layer, a hidden layer, and an output layer. A general RBF model is shown in Figure 4. In this network, the quantity of input and output nodes depends upon the number of input and output variables. X is the input variable and Y is the output variable. φ is radial basis function and c is the center of each hidden node, corresponding to observed input data. Mostly, sigmoidal and Gaussian kernel functions are used in RBF systems. This is represented by Equation (2):

$$Y = \sum_{i=1}^n W_i \varphi(\|X - C_i\|) \tag{2}$$

where W is weights associated with each input parameter, φ is radial basis function, c is the center of each hidden node, and n is number range of input variable. This RBF Gaussian function is given as Equation (3):

$$f(y) = e^{-y^2} \tag{3}$$

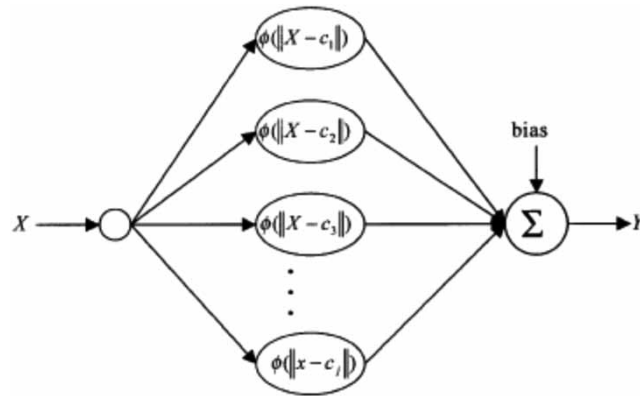


Figure 4 | General structure of RBFNN (Lin & Chen 2004).

Wavelet transforms

Wavelet decomposition is an advanced tool in data processing techniques. Its use has increased nowadays in data handling, signal processing, image processing, and optical engineering (Seo *et al.* 2015). Wavelet transform is used to discover, denoise, and smooth time series. Wavelet transform breaks the original time series into various time series which have sinusoidal wave patterns output and these outputs are used as input for the neural network. The end result will be a collection of time series frequencies which represent the original signal with different scales and resolution. The main advantage of decomposing the time series is that the hidden collective frequency in the time series will be released and various parameters such as variations in mode and its temporal change can be easily identified (Kim *et al.* 2017). It will help tremendously in data synthetization and analysis. Also, here, the time period of data is on a monthly basis and discrete in nature. The wavelet transform of a time series $f(t)$ is defined in Equation (4) as:

$$f(x, y) = \frac{1}{\sqrt{x}} \int_{-\infty}^{\infty} f(t) t \Psi\left(\frac{t-y}{x}\right) dt \quad (4)$$

where $\Psi(t)$ is the basic wavelet having time series length t . The variables are x and y , where x is the scale or dilation factor that determines the characteristic frequency so that its variation gives rise to a 'spectrum'; and y is the translation in time so that its variation represents the 'sliding' of the wavelet over $f(t)$ (Krishna *et al.* 2011). The original signal $X(n)$ passes through two complementary filters (low pass and high pass filters) and emerges as two signals as approximations (A) and details (D). The approximations are the high-scale, low frequency components of the signal. The details are the low-scale, high frequency components. Normally, the low frequency content of the signal (approximation, A) is the most important part. It demonstrates the signal identity (Figure 5). To find out the best decomposition level for optimum performance of the model, levels at 2, 3, and 4 decomposition level were tested and the best performance was found for decomposition level of 4. This is in accordance with Equation (5) (Nourani *et al.* 2014):

$$L = \text{int}\{\log(n)\} \quad (5)$$

where L and n denote decomposition level and data number (time series length), respectively.

Input data preparation

The daily data of rainfall, runoff, and temperature are available from 2000 to 2018. Rainfall, runoff, and temperature data are utilized for developing the RBFNN network. First, 15 ANN models were developed based on input combinations and results were obtained. After that, input variables were normalized and wavelet decomposed at level 3 and these wavelet transforms were used to feed as input signals in the WRBFNN network. This normalization is done due to large variation in values of rainfall and runoff. Moreover, logistic functions (Equation (5)) vary between 0 and 1, due to which this normalization is justified. Data of runoff and rainfall during rainy seasons are only considered for network architecture for better results and used

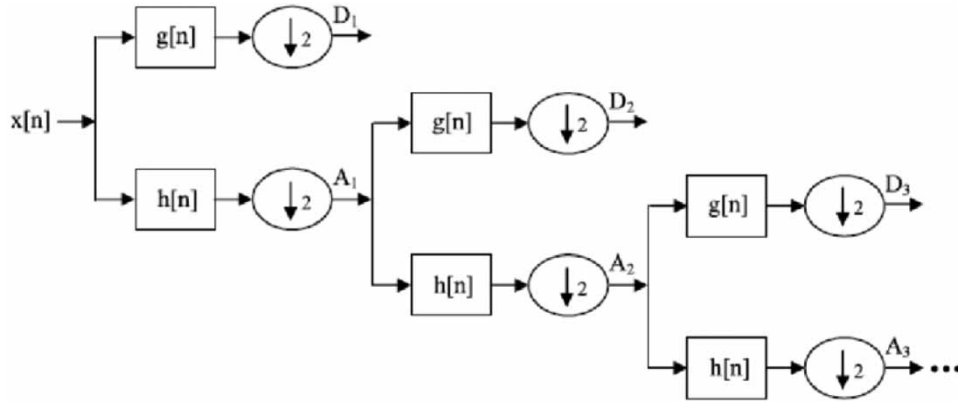


Figure 5 | Diagram of wavelet decomposition of high and low filter pass.

for training, testing, and validation of the neural network, that is, from 15 June to 15 October, which are monsoon months in India.

ANN model architecture

In the present study, 15 models are developed using several input combinations of daily antecedent precipitation of the last 4 days, i.e., R , $R(t-1)$, $R(t-2)$, $R(t-3)$, and $R(t-4)$, where these are precipitation on the same day, previous day, 2 days before, 3 days before, and 4 days before the present day and $Q(t-1)$, $Q(t-2)$, and $Q(t-3)$ are precipitation on the previous day, 2 days before, and 3 days before the present day. Also mean monthly temperature, $T(t-1)$, which is the previous day’s temperature, is also used for the network. In the case of RBFNN network models, input combinations are used just after normalization, but in the case of WRBFNN, after normalization, discrete wavelet decompositions of input data up to level 2 are correlated with output and used for feeding as input signals. Table 2 shows the various input combinations for both the RBFNN and WRBFNN networks.

In Table 2, based on various combinations of input variables, a total of 15 models are generated. These models are used for analysis for both the networks. These network variations are based on the trial-and-error method. Other combinations can also be used, but due to time constraints only 15 models are selected.

Table 2 | Model variants for radial basis function neural network (RBFNN) and wavelet radial basis function neural network (WRBFNN)

Model No.	Input combinations	Spread constant	Output layer
M1	$R(t-1)$	1	$Q(t)$
M2	$R(t-1)$, $R(t-2)$	1	$Q(t)$
M3	$R(t-1)$, $R(t-2)$, $R(t-3)$	1	$Q(t)$
M4	$R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$	1	$Q(t)$
M5	$R(t-1)$, $Q(t-1)$	1	$Q(t)$
M6	$R(t-1)$, $R(t-2)$, $Q(t-1)$	1	$Q(t)$
M7	$R(t-1)$, $R(t-2)$, $R(t-3)$, $Q(t-1)$	1	$Q(t)$
M8	$R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$, $Q(t-1)$	1	$Q(t)$
M9	$R(t-1)$, $Q(t-1)$, $Q(t-2)$	1	$Q(t)$
M10	$R(t-1)$, $R(t-2)$, $Q(t-1)$, $Q(t-2)$	1	$Q(t)$
M11	$R(t-1)$, $R(t-2)$, $R(t-3)$, $Q(t-1)$, $Q(t-2)$	1	$Q(t)$
M12	$R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$, $Q(t-1)$, $Q(t-2)$	1	$Q(t)$
M13	$R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$, $Q(t-1)$, $T(t-1)$	1	$Q(t)$
M14	$R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$, $Q(t-1)$, $Q(t-2)$, $T(t-1)$	1	$Q(t)$
M15	$R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$, $T(t-1)$	1	$Q(t)$

Evaluation criteria

Performance of the models is evaluated based on three criteria, namely, coefficient of determination (R^2), average absolute residual error (AARE), and mean square error (MSE), which are given by Equations (6), (7), and (8), respectively:

$$R^2 = \frac{n(\sum Q_{obs} * Q_{sim}) - (\sum Q_{obs}) * (\sum Q_{sim})}{\sqrt{[n(\sum Q_{obs}^2) - (\sum Q_{obs})^2] * [n(\sum Q_{sim}^2) - (\sum Q_{sim})^2]}} \quad (6)$$

$$AARE = \sum_{i=1}^n \frac{|Q_{sim} - Q_{obs}|}{n} \quad (7)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Q_{sim} - Q_{obs})^2 \quad (8)$$

where Q_{obs} is observed value of discharge and Q_{sim} is simulated value of discharge; n is number of observations. The range of R^2 lies between 0 and 1, which represents no correlation and perfect correlation between observed and simulated value. A lower value of AARE represents greater predictive capability of the model. MSE is calculated at the output layer, and if MSE is within acceptable limits the training of the NN model is then assumed to be completed, otherwise the feed backward pass is continued for updating of synaptic weights between hidden and output layers by using the learning algorithm. The synaptic weights between input and hidden layers can also be updated in an analogous method.

RESULTS AND DISCUSSION

The selection of variables used for input inspires the performance of the rainfall-runoff model. These variables applied for selection in the rainfall-runoff model or network depend on either number of time lagged input variables (Tokar & Markus 2000; Riad & Mania 2004; Wang *et al.* 2017) or input variables based on most correlated variables in lag time series data (Sudheer *et al.* 2002; Ali *et al.* 2010). The authors have applied the first approach to the study due to there being a large number of modeling networks. The daily observed rainfall data of Kolar catchment is transformed by the DWT and applied at level 4. With $N = 11,202$ testing/validation rainfall data points in the current study, the input rainfall data can be decomposed into approximations and details. The simulated values of the runoff components, averaged over the length of the data, are shown in Table 3. In the paper, two models, RBFNN and WRBFNN, were compared based on

Table 3 | RBFNN and WRBFNN models' simulation results for given period

Model No.	RBFNN			WRBFNN		
	R^2	AARE	MSE	R^2	AARE	MSE
M1	0.0312	0.0843	0.01501	0.0856	0.0813	0.01463
M2	0.0414	0.0798	0.01311	0.1567	0.0768	0.01273
M3	0.353	0.0753	0.01245	0.4516	0.0723	0.01207
M4	0.4315	0.0716	0.01059	0.5149	0.0686	0.01021
M5	0.0563	0.0811	0.01487	0.2156	0.0781	0.01449
M6	0.1127	0.0804	0.01456	0.3895	0.0774	0.01418
M7	0.6784	0.0489	0.00741	0.7916	0.0459	0.00703
M8	0.9567	0.0356	0.00088	0.9889	0.0097	0.00051
M9	0.5423	0.0512	0.00959	0.6846	0.0482	0.00922
M10	0.6458	0.0498	0.00815	0.7829	0.0468	0.00777
M11	0.7856	0.0417	0.00453	0.8357	0.0387	0.00415
M12	0.9483	0.0388	0.00186	0.9786	0.0114	0.00148
M13	0.8856	0.0401	0.00316	0.9046	0.0371	0.00278
M14	0.8623	0.0395	0.00245	0.8938	0.0365	0.00207
M15	0.3827	0.0738	0.01122	0.5109	0.0708	0.01084

15 model combinations. They are tested at the spread constant of 1 to observe the correlation between output and input parameters. These values are also better as antecedent rainfall. The trend increases in R^2 value until model 4 for both the cases of RBFNN and WRBFNN. For model 5, the value of R^2 decreases for the previous day's antecedent rainfall and previous day's discharge. Again, it started increasing for model 6 onwards until model 8, which has highest R^2 for both the models. However, the value of R^2 in the case of the WRBFNN model is greater than the RBFNN model which shows that use of wavelet transforms of the input in neural networks makes the network more efficient.

In the last three models of the NN network, when temperature is used as input for the best performing model in the first 12 models, the value of coefficient of determination decreased significantly. Therefore, it may be concluded that for this Kolar river watershed, temperature is not a significant factor for prediction of runoff.

This best performing model for both the networks is used for prediction for one-year ahead runoff. Figures 6 and 7 show the regression for predicted value and observed value of both the models.

Figure 6 shows the regression plot of the RBFNN network of model 8. Coefficient of determination between observed and predicted value for this model shows better predicting capability of the model. However, during regression analysis of WRBFNN model 8, better predicting capability has been seen, as is evident from training, testing, and validation of the model. Based on results of both the models, it can be concluded that RBF is an efficient neural network for the rainfall-runoff modeling and the efficiency can be further increased with use of data processing technique of wavelet transforms.

These results were compared with almost similar research work in the recent past, which shows that the present work is also novel and efficient in the field of rainfall-runoff modeling. Alizadeh *et al.* (2017) used the ANN network alone and along

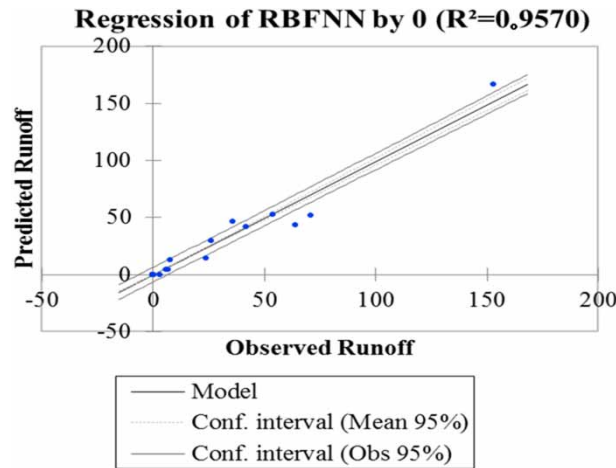


Figure 6 | The runoff prediction for 1 day ahead for the best performing network model of RBFNN scatter plot (runoff in m^3/sec).

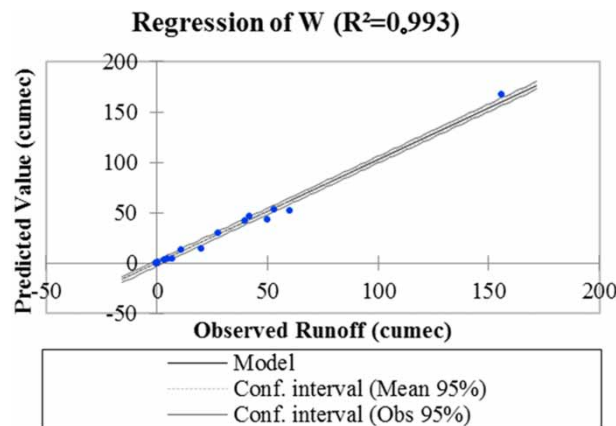


Figure 7 | The runoff prediction for 1 day ahead for the best performing network model of WRBFNN scatter plot (runoff in m^3/sec).

with wavelets (WANN) for rainfall–runoff prediction for Tolt River, Washington, USA. The best performing model in this work has an R^2 value of 0.71 for ANN and 0.97 for WANN. These models were applied on a monthly scale and showed better predicting capability in the multi-model approach. The number of models in the present study is also greater as compared to them. The present models performed better than this model. Poonia & Tiwari (2020) also used feed forward back propagation (FFBP) network and RBF simultaneously and found that RBF performs better, which also supports the results of the findings of the present study.

CONCLUSION

The rainfall–runoff correlation rests on climatic besides physical factors comprising daily variations in precipitation, catchment slope, elevation, land cover, soil humidity, underground water storage, etc. Owing to being dependent on multivariables, the RR relationship shows a complicated non-linear relationship. Many models are available for simulation of relationships but are highly catchment dependent. Owing to its ability to model complicated non-linear relations without any need for a high number of parameters, the RBF model is proposed for Kolar catchment and, to remove bias in data signals, wavelets are proposed for data processing of inputs. In this paper, a new prediction model for rainfall–runoff modeling is proposed which has the advantages of a self-learning dynamic neural network. The main aim of this modeling is to predict the runoff generated in the Kolar river basin. Two methods were applied for the simulation of the network. The first method, RBFNN, is applied with radial basis function neural network for 15 models differing in input combination for the rainfall–runoff modeling. In the second model, WRBFNN, data processing techniques of wavelet which remove the bias in data signals and give better predicting capability were applied for the prediction of runoff. For the first time, RBFNN is applied with wavelets for the central region of India for Narmada basin and for 15 models in one study. The best predicting model in these two networks is model 8, which has inputs of $R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$, and $Q(t-1)$. For the RBFNN model, the maximum value of R^2 is 0.9567 and the least value of AARE and MSE is observed. Similarly, for the WRBFNN model, the maximum value of R^2 is 0.9889 and least values of AARE and MSE are observed. It shows the importance of discharge and time lagged rainfall data for predicting the runoff generated during the hydrological process. These models were applied for one-year ahead data for the next year and gave promising predicting values which shows the model to be useful for prediction for the next year. Also, the use of temperature data in prediction shows no significant contribution and, on the contrary, leads to a decrease in coefficient of determination when introduced with the best performing models. This model has the potential to be improved further with a greater number of input combinations and data from different river basins. This result cannot be made generic to all the basins but further research is required. Most of the models have performed better than the same models with wavelet transformed inputs. This method can be applied to data-scarce regions where data are available for shorter duration and manually recorded which invites human errors. For the network to perform better, data with high accuracy are required to train the network better. Other data, like evapotranspiration, ground water recharge, morphometry of basin, solar radiation, wind velocity, etc. are also required to be taken into account for the runoff analysis. Models like SWAT, PCSWMM, etc. can also be used for better prediction.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Achela, D., Fernando, K. & Shamseldin, A. Y. 2009 Investigation of internal functioning of the radial-basis-function neural network river flow forecasting models. *Journal of Hydrologic Engineering* **14** (3), 286–292.
- Ali, S., Ghosh, N. C. & Singh, R. 2010 Rainfall–runoff simulation using a normalized antecedent precipitation index. *Hydrological Sciences Journal* **55** (2), 266–274.
- Alizadeh, M. J., Kavianpour, M. R., Kisi, O. & Nourani, V. 2017 A new approach for simulating and forecasting the rainfall-runoff process within the next two months. *Journal of Hydrology* **548**, 588–597.
- Asadi, S., Shahrabi, J., Abbaszadeh, P. & Tabanmehr, S. 2013 A new hybrid artificial neural networks for rainfall-runoff process modeling. *Neurocomputing* **121**, 470–480.
- Badrzadeh, H., Sarukkalige, R. & Jayawardena, A. W. 2015 Hourly runoff forecasting for flood risk management: application of various computational intelligence models. *Journal of Hydrology* **529**, 1633–1643.

- Barman, S. & Bhattacharjya, R. K. 2020 ANN-SCS-based hybrid model in conjunction with GCM to evaluate the impact of climate change on the flow scenario of the River Subansiri Swapnali Barman and Rajib Kumar Bhattacharjya. *Journal of Water and Climate Change* **11** (4), 1150–1164.
- Bekele, W. T., Haile, A. T. & Rientjes, T. 2021 Impact of climate change on the streamflow of the Ardjo-Didessa catchment under RCP scenarios. *Journal of Water and Climate Change* **12**, 2325–2337.
- Broomhead, D. S. & Lowe, D. 1988 Multivariable functional interpolation and adaptive networks. *Complex Systems* **2**, 321–355.
- Dawson, C. W., Harpham, C., Wilby, R. L. & Chen, Y. 2002 Evaluation of artificial neural network techniques for flow forecasting in the River Yangtze, China. *Hydrology and Earth System Sciences* **6** (4), 619–626.
- Ghumman, A. R., Ghazaw, Y. M., Sohail, A. R. & Watanabe, K. 2011 Runoff forecasting by artificial neural network and conventional model. *Alexandria Engineering Journal* **50** (4), 345–350.
- Kim, B., Jeong, H., Kim, H. & Han, B. 2017 Exploring wavelet applications in civil engineering. *KSCE Journal of Civil Engineering* **21** (4), 1076–1086.
- Krishna, B., Rao, Y. R. S. & Nayak, P. C. 2011 Time series modeling of river flow using wavelet neural networks. *Journal of Water Resource and Protection* **3** (1), 50–59.
- Kumar, M. & Yadav, N. 2011 Multilayer perceptrons and radial basis function neural network methods for the solution of differential equations: a survey. *Computers and Mathematics with Applications* **62** (10), 3796–3811.
- Lin, G. F. & Chen, L. H. 2004 A non-linear rainfall-runoff model using radial basis function network. *Journal of Hydrology* **289** (1–4), 1–8.
- Miaoli, M., Xiaolong, W. & Honggui, H. 2020 Accelerated Levenberg-Marquardt algorithm for radial basis function neural network. In: *Proceedings - 2020 Chinese Automation Congress, CAC 2020*. Institute of Electrical and Electronics Engineers Inc., pp. 6804–6809.
- Mujumdar, P. P. & Kumar, D. N. 1990 Stochastic models of streamflow: some case studies. *Hydrological Sciences Journal* **35**, 395–410.
- Nawi, N. M., Atomi, W. H. & Rehman, M. Z. 2013 The effect of data pre-processing on optimized training of artificial neural networks. *Procedia Technology* **11**, 32–39.
- Nourani, V. & Komasi, M. 2013 A geomorphology-based ANFIS model for multi-station modeling of rainfall-runoff process. *Journal of Hydrology* **490**, 41–55.
- Nourani, V., Komasi, M. & Mano, A. 2009 A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resources Management* **23** (14), 2877–2894.
- Nourani, V., Kisi, Ö. & Komasi, M. 2011 Two hybrid Artificial Intelligence approaches for modeling rainfall-runoff process. *Journal of Hydrology* **402** (1–2), 41–59.
- Nourani, V., Baghanam, A. H., Adamowski, J. & Kisi, O. 2014 Applications of hybrid Wavelet-Artificial Intelligence models in hydrology. A review. *Journal of Hydrology* **514**, 358–377.
- 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.
- Poonia, V. & Tiwari, H. L. 2020 Rainfall-runoff modeling for the Hoshangabad Basin of Narmada River using artificial neural network. *Arabian Journal of Geosciences* **13** (944), 1–10.
- Powell, W. W. 1987 Hybrid organizational arrangements: new form or transitional development? *California Management Review* **30**, 67–87.
- Riad, S. & Mania, J. 2004 Rainfall-runoff model using an artificial neural network approach. *Mathematical and Computer Modelling* **40**, 839–846.
- Seo, Y., Kim, S. & Singh, V. P. 2015 Multistep-ahead flood forecasting using wavelet and data-driven methods. *KSCE Journal of Civil Engineering* **19** (2), 401–417.
- Shamseldin, A. Y. 2009 Investigation of internal functioning of the radial-basis-function neural network river flow. *Journal of Hydrologic Engineering* **14** (3), 286–292.
- Sharghi, E., Nourani, V., Najafi, H. & Molajou, A. 2018 Emotional ANN (EANN) and wavelet-ANN (WANN) approaches for Markovian and seasonal based modeling of rainfall-runoff process. *Water Resources Management* **32** (10), 3441–3456.
- Shoaib, M., Shamseldin, A. Y., Khan, S., Khan, M. M., Khan, Z. M., Sultan, T. & Melville, B. W. 2018 A comparative study of various hybrid wavelet feedforward neural network models for runoff forecasting. *Water Resources Management* **32** (1), 83–103.
- Sonali, P. & Nagesh Kumar, D. 2016a Spatio-temporal variability of temperature and potential evapotranspiration over India. *Journal of Water and Climate Change* **7** (4), 810–822.
- Sonali, P. & Nagesh Kumar, D. 2016b Detection and attribution of seasonal temperature changes in India with climate models in the CMIP5 archive. *Journal of Water and Climate Change* **7** (1), 83–102.
- Sudheer, K. P., Gosain, A. K. & Ramasastri, K. S. 2002 A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological Processes* **16** (6), 1325–1330.
- Sušanj, I., Ožanić, N. & Marović, I. 2016 Methodology for developing hydrological models based on an artificial neural network to establish an early warning system in small catchments. *Advances in Meteorology* **2016**, 9125219.
- Tayyab, M., Zhou, J., Zeng, X. & Adnan, R. 2016 Discharge forecasting by applying artificial neural networks at the Jinsha River Basin, China. *European Scientific Journal* **12** (9), 108.
- Todini, E. 2007 Hydrological catchment modelling: past, present and future. *Hydrology and Earth System Sciences* **11** (1), 468–482.
- Tokar, A. S. & Markus, M. 2000 Precipitation-runoff modeling using artificial neural networks and conceptual models. *Journal of Hydrologic Engineering* **5** (2), 156–161.

- Wang, W., Chau, K., Xu, D., Qiu, L. & Liu, C. 2017 The annual maximum flood peak discharge forecasting using hermite projection pursuit regression with SSO and LS method. *Water Resources Management* **31** (1), 461–477.
- Zhang, Q., Bin, B. W., Yong, H. & Ren, M. 2011 Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting. *Water Resources Management* **25**, 2683–2703.

First received 16 June 2021; accepted in revised form 26 November 2021. Available online 30 December 2021