Ashutosh Sharma and Manish Kumar Goyal*

A Comparison of Three Soft Computing Techniques, Bayesian Regression, Support Vector Regression, and Wavelet Regression, for Monthly Rainfall Forecast

Abstract: Rainfall, being one of the most important components of the hydrological cycle, plays an extremely important role in agriculture-based economies like India. This paper presents a comparison between three soft computing techniques, namely Bayesian regression (BR), support vector regression (SVR), and wavelet regression (WR), for monthly rainfall forecast in Assam, India. A WR model is a combination of discrete wavelet transform and linear regression. Monthly rainfall data for 102 years from 1901 to 2002 at 21 stations were used for this study. The performances of different models were evaluated based on the mean absolute error, root mean square error, correlation coefficient, and Nash-Sutcliffe efficiency coefficient. Based on model statistics, WR was found to be the most accurate followed by SVR and BR. The efficiencies for the BR, SVR, and WR models were found to be 32.8%, 52.9%, and 64.03%, respectively. From the spatial analysis of model performances, it was found that the models performed best for the upper Assam region followed by lower, southern, and middle regions, respectively.

Keywords: Rainfall prediction, support vector regression, wavelet regression, Bayesian regression.

1 Introduction

Rainfall is one of the most important components of the hydrological cycle. It has an importance in many fields, including agriculture, irrigation, hydroelectric production, scientific research, etc. Many catastrophic events like flood and drought are directly linked to rainfall and, hence, it becomes extremely important to have an accurate forecast of rainfall. Rainfall is highly stochastic in nature and is dependent on many local and global atmospheric processes, which makes rainfall prediction one of the most challenging problems around the globe.

Generally, two different approaches are used for rainfall prediction: empirical and dynamical. In the empirical approach (e.g. regression, artificial neural networks, fuzzy logic and other machine learning and data mining techniques), the past records of rainfall, atmospheric parameters, and oceanic parameters are used to capture the trend of rainfall w.r.t. to governing parameters. In the dynamical approach, physical models based on some system of equations are used to predict the rainfall and other atmospheric parameters based on some initial condition provided [27]. The empirical approach has been widely used by researchers for rainfall prediction [19, 22].

For this study, we have used three different regression techniques, viz. Bayesian regression (BR), support vector regression (SVR), and wavelet regression (WR), for rainfall forecasting. BR tries to model the uncertainty in unknown parameters by assigning a probability distribution, and then Bayes' theorem is used to find the posterior distribution. It intends to find a value for unknown parameters that maximizes the posterior distribution. Many applications of Bayesian methods in different fields of science and engineering are available in the

---
*Corresponding author: Manish Kumar Goyal,** Department of Civil Engineering, Indian Institute of Technology, Guwahati 781039, India, Tel.: +91 361 258 3328, e-mail: mkgoyal@iitg.ernet.in
Ashutosh Sharma: Department of Civil Engineering, Indian Institute of Technology, Guwahati 781039, India

literature [4, 5, 8, 15, 23, 28, 32]. A Bayesian framework is used to analyze a daily rainfall series for estimation of extreme rainfall events, and it is shown that use of prior expert information can supplement data and can improve the estimate of extreme events [5]. A Bayesian network is used for modeling the spatial and temporal dependencies among the 100 stations, and the network is combined with the numerical atmospheric prediction to forecast the precipitation [4]. A multivariate relevance vector machine model can be used to forecast releases from a system of multiple reservoirs with good accuracy in learning the input-output patterns [32].

Support vector machine (SVM) was introduced by Vapnik [33] based on the statistical learning theory (SLT). It tries to find a decision rule that provides good generalization by selecting some data points in the training data called support vectors. SVM methods have been widely used in the field of hydrology [1, 2, 7, 20, 25, 30, 38, 39]. Hong and Ping-Feng [12] made an assessment of the support vector regression (SVR) techniques in rainfall prediction, and showed that SVR could be a better alternative for rainfall forecasting. A hybrid support vector technique was used to forecast two hydrological time series [30]. Wei [37] used a coupled wavelet SVMs for hourly rainfall forecast during the tropical cyclones.

Wavelet transform has been used for investigating the variation, periodicity, and trends in data time series in many studies. Discrete wavelet transform (DWT) can effectively be utilized to classify distinct hydroclimatic categories [31]. The wavelet methods are used to model rainfall and runoff for daily and half-hourly sampling rates of karstic springs, and are found able to discriminate the karstic rapid response and recharge, from slower infiltration responses [18]. Chou and Wang [3] applied DWT to decompose unit hydrograph and utilized updated wavelet coefficients of unit hydrograph to accurately predict one-step-ahead runoff. Dai et al. [6] used wavelet analysis to study the decay in summer monsoon in north China over the interdecadal time scale. Zhou et al. [40] have shown that a wavelet predictor-corrector model performed more accurately compared to the ARMA model, seasonal ARIMA model, and BP artificial neural network for the prediction of discharge. Kisi [16] compared the performance of the WR model to the artificial neural network (ANN) and autoregressive (AR) models for streamflow prediction, and found that WR outperformed the ANN and AR models. The WR model was found superior to the ANN model for prediction of daily river stage [17]. A comparison of the WR model and ANN has shown that the WR model provided better monthly rainfall forecast [9]. Many studies have shown SVR models to be better than ANN models [1, 7, 20, 29]. The WR model decomposes the monthly rainfall time series into detail and approximation components using DWT, and a new time series is generated by adding an approximation component and effective detail components. This new time series is given as input to the linear regression (LR) model, keeping the original rainfall time series as output. Some hydrological applications of WR have revealed that WR can perform better than non-linear models like ANN, etc., in terms of accuracy [16, 17]. In the present research work, the accuracy of WR in rainfall prediction is compared with that of SVR and BR.

Assam is a northeastern state of India, where agriculture is the backbone of state economy. The north-eastern region of India is one of the most rainfall-receiving regions in the world. Flood occurs in many parts of the state every year, causing huge loss in terms of infrastructure and human lives. An advance prediction of extreme rainfall events causing floods can help in reducing the impact of floods. The aforementioned three soft computing techniques are used in this paper for rainfall forecast at 21 stations in Assam, India. The BR approach finds the posterior distribution of regression coefficients and averages over a large number of samples. The Bayesian averaging is effective in avoiding the overfitting of the training data. Vapnik's ε-insensitive loss function is used for the SVR, and the Gaussian radial basis function is used for non-linear mapping of input data space. A coupled WR and LR approach is used for rainfall forecast. The present study compares six different models developed based on the aforementioned three soft computing techniques.

2 Study area and data

2.1 Study Area

Assam is a northeastern state of India that is located between latitudes 24°8′N to 28°2′N and longitudes 89°42′E to 96°E. The location of the state on the India map is shown in Figure 1. The state is located at the

foothills of the eastern Himalayas and receives an average annual rainfall of around 300 cm from the southwest monsoon. There are around 51 forests and subforests in the states because of its topographical and climatic features. The state experiences heavy rainfall and floods every year, which cause riverbank erosions and landslides in many parts of the state. Rainfall and floods also affect agriculture, which accounts for more than a third of the state domestic product and is the principal occupation of the rural people, who constitute nearly 90% of the total population [41].

There is significant spatial and temporal variation in rainfall over the state due to topographic and geographic features. Southwest monsoon brings heavy rainfall during the June to August period and causes flood in many parts of the state. The state suffers huge monetary losses every year because of heavy rainfall and floods. The amount of rainfall received by some regions in monsoon months is >800 mm. During the non-monsoon period, the monthly mean rainfall remains much smaller. A box plot showing the annual mean rainfall received by 21 stations in Assam between 1901 and 2002 for every month is presented in Figure 2. Due to this high temporal variability, prediction of rainfall becomes very difficult.

2.2 Data

The monthly mean precipitation data of 21 stations spread over the whole state is used for this study. Data for precipitation were obtained from the India water portal website (http://www.indiawaterportal.org/met_data/). The monthly mean rainfall is available for 102 years from January 1901 to December 2002. The distribution of rainfall is shown in Figure 1. The eastern part of the state receives less rainfall, compared to other parts of the state. The rainfall statistics of 21 stations for 102 years data are presented in Table 1. The
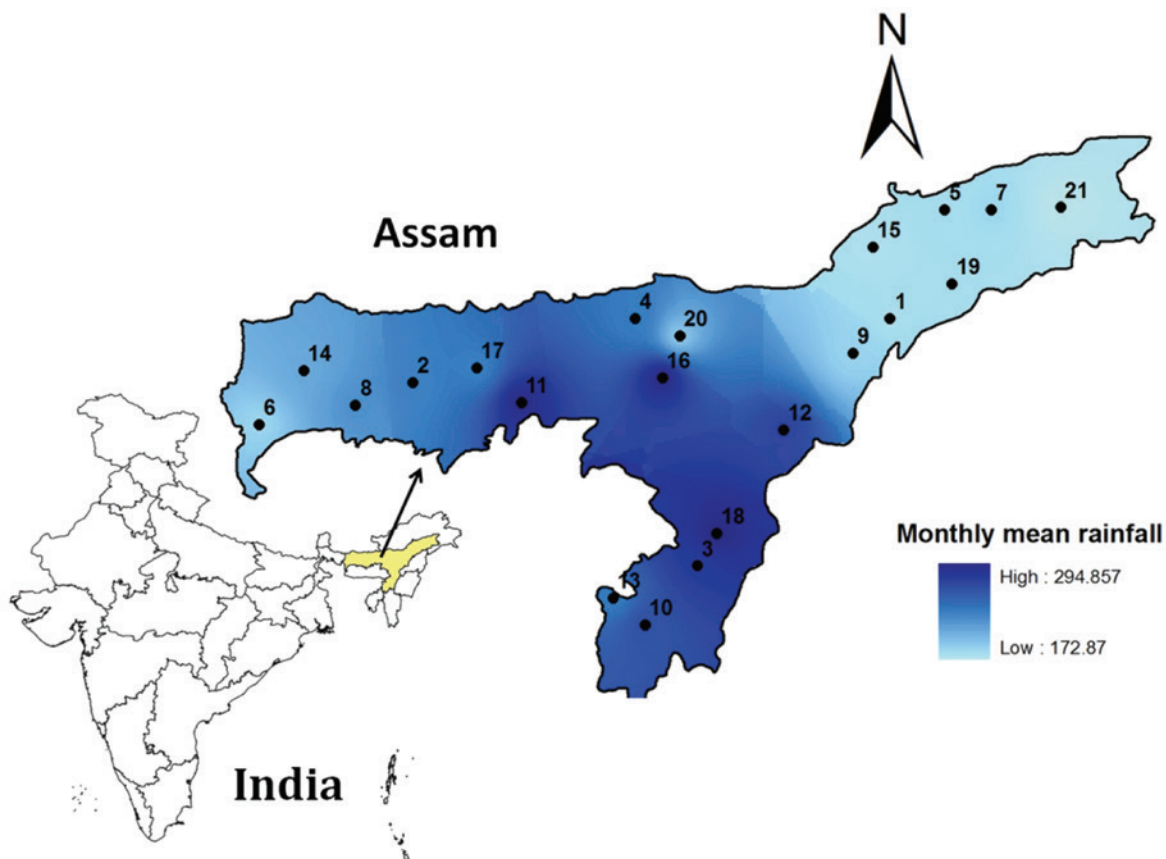

Figure 1: Location of Assam and Stations on India Map, and Distribution of Monthly Mean Rainfall over Assam.
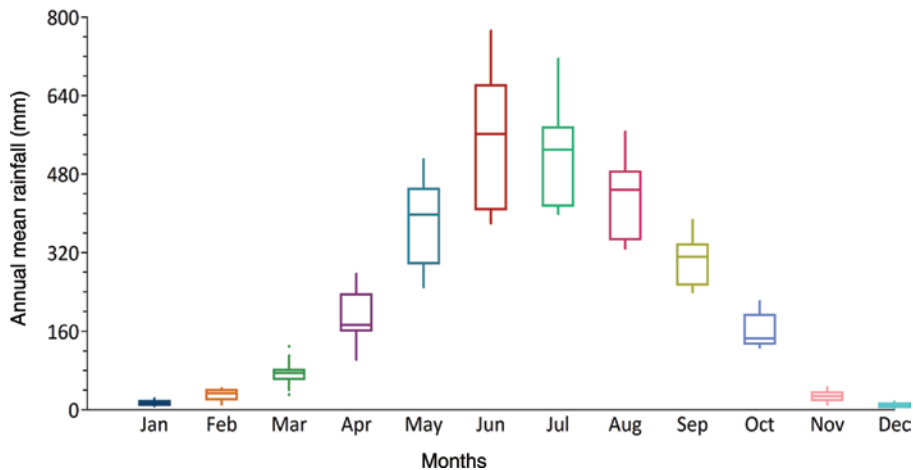The numbers represent the stations.

Figure 2: Box Plot Showing Annual Mean Rainfall at All Stations for Every Month Between 1901 and 2002.

Table 1: Geographical Latitude, Longitude, and Rainfall Statistics of 21 Stations.

Station no.	Station name	Latitude	Longitude	Monthly mean rainfall				
				Minimum	Maximum	Mean	Standard deviation	Skewness
1	Jorhat	26°45′N	94°13′E	0.19	677.06	179.26	162.86	0.7
2	Barpeta	26°19′N	91°00′E	0	1033.69	230.73	239.38	0.83
3	Cachar	25°05′N	92°55′E	0	1293.43	275.71	265.96	0.87
4	Darrang	26°45′N	92°30′E	0	1069.2	234.79	237.25	0.83
5	Dhemaji	27°29′N	94°35′E	0.2	685.51	180.44	160.98	0.69
6	Dhubri	26°02′N	89°58′E	0	1020.58	185.85	198.97	0.88
7	Dibrugarh	27°29′N	94°54′E	0.2	703.83	183.69	164.32	0.73
8	Goalpara	26°10′N	90°37′E	0	1108.47	222.13	231.72	0.83
9	Golaghat	26°31′N	93°58′E	0.06	705.11	186.69	170.58	0.69
10	Hailakandi	24°41′N	92°34′E	0	1323.74	261.82	256.93	0.92
11	Kamrup	26°11′N	91°44′E	0	1594.27	292.04	301.43	0.89
12	Karbi Anglong	26°00′N	93°30′E	0.03	1234.26	270.23	265.14	0.87
13	Karimganj	24°52′N	92°21′E	0	1197.53	233.98	230.07	0.91
14	Kokrajhar	26°24′N	90°16′E	0	997.85	208.74	221.32	0.83
15	Lakhimpur	27°14′N	94°6′E	0.13	674.85	178.1	162.08	0.67
16	Nagaon	26°21′N	92°41′E	0	1525.96	293.95	294.85	0.89
17	Nalbari	26°25′N	91°26′E	0	1055.19	231.85	239.49	0.85
18	North Cachar Hills	25.18°N	93.03°E	0	1233.48	294.86	280.44	0.8
19	Sibsagar	26°59′N	94°38′E	0.19	694.85	179.75	161.6	0.71
20	Sonitpur	26°38′N	92°48′E	0	784.22	192.8	185.47	0.71
21	Tinsukia	27°30′N	95°22′E	0.19	722.23	172.87	155.35	0.78

maximum recorded monthly rainfall is 1594.27 mm at Kamrup station, which occurred in July 1974. The mean monthly rainfall varied between 172 and 294 mm. The standard deviation varied between 155.35 mm at station no. 21 (Tinsukia) and 301.43 mm at station no. 11 (Kamrup). The skewness varied between 0.67 at Lakhimpur station and 0.92 at Hailakandi station. Skewness is positive at all the stations, which shows that the frequency distribution of rainfall is asymmetric and it lies on the right side of the mean. The distribution of the data is less scattered as the values of skewness become lower. Out of 102 years data, the first 60 years data (from January 1901 to December 1960) were used for calibrating the models and the remaining 42 years data (from January 1961 to December 2002) were used for validation.

3 Methodology

A general LR problem can be described in the form of a target vector (t), a set of input vectors (x_n), and a functional relationship (f) between t and x, which also includes some additive noise components. This can be represented as

$$t_n = f(x_n, \beta) + \epsilon_n, \tag{1}$$

where ϵ is an additive noise process in which ϵ_n are i.i.d. (independent identical distributions) and β is the vector of adjustable parameters. The main objective of all regression techniques is to set a selection criterion for the function f such that it fits the maximum data points in the training data set and gives a good generalization for new data points. The widely used least square technique is based on error function minimization, in which it tries to minimize the sum-of-squares error and set β^* for the least sum-of-square errors. Error function minimization techniques suffer from overfitting of training data and poor generalization. Limiting the model complexity can help in avoiding the overfitting; however, the model can still suffer from poor generalization in case it is not flexible enough to capture the trend of the data. In the maximum likelihood technique, a normal distribution with zero mean and variance σ^2 is assumed for the additive terms (ϵ_n) in Eq. (1) so that $p(\epsilon|\sigma^2) = N(\epsilon|0, \sigma^2)$ and we select the value of β^*, which maximizes the likelihood function.

3.1 Bayesian LR

In BR, the uncertainty in β is represented by a probability distribution, $p(\beta)$. Bayes' theorem is used to express the posterior distribution for β as the product of prior distribution and likelihood function:

$$p(\beta|t, \alpha, \sigma^2) \propto p(\beta|\alpha)p(t|\beta, \sigma^2). \tag{2}$$

The value of β is chosen, which maximizes the posterior distribution. We integrate over the distribution of β to make predictions. The integration over the distribution overcomes the problem of overfitting, as we are averaging over a large number of possible solutions. The maximization of the posterior distribution in Eq. (2) is equivalent to the minimization of the negative logarithm of the distribution. Different samples of the solution are generated using the mean and variance of the distribution. Solving the minimization problem, the posterior can be expressed in terms of mean and variance.

For this study, we have assumed a Gaussian prior on β (regression coefficients) with mean zero and covariance matrix Λ^{-1} so that $\beta \sim N(0, \Lambda^{-1})$, where Λ is the inverse covariance or precision matrix that is used to make computation easier. The posterior distribution on β can be represented as

$$\beta = N(\mu_\beta, \Sigma_\beta), \tag{3}$$

where

Posterior mean, $\mu_\beta = (X^TX + \sigma^2\Lambda)^{-1}X^Ty$,

Covariance matrix, $\Sigma_\beta = \sigma^2(X^TX + \sigma^2\Lambda)^{-1}$.

This model works in two steps. In the first step, the mean and covariance matrix for posterior distribution of β is calculated using Eq. (3). In the second step, β is approximated from this normal distribution by averaging over a large number of samples generated by using multivariate normal distribution:

$$\beta \sim \text{MVN}(\mu_\beta, \Sigma_\beta). \tag{4}$$

3.2 Support Vector Regression

SVM is a method of supervised learning that is based on SLT. It was developed by Vapnik and his team, and is generally used for classification and regression analysis [33, 34]. The most important advantage of SVM is
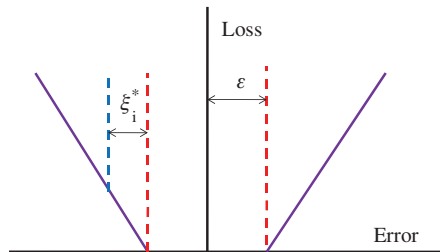
Figure 3: Vapnik's ε-Insensitive Loss Function.

the principle of structural risk minimization, which is shown to be better than traditional empirical risk minimization [11, 24]. It tries to minimize the upper-bound generalization errors instead of minimizing the local training errors. It tries to find a decision rule that provides better generalization by selecting some subset of training data called support vectors. To take into account the non-linearity present in the data, the kernel functions are used for non-linearly mapping the input space to a higher-dimensional feature space.

In SVR, a loss function ($L_\varepsilon(t, f(x))$) is used, which describes the deviation between the function $f(x_i)$ outcomes and the actual target values t_i, and allows some deviation between the target values t_i and the function values $f(x_i)$. Vapnik's ε-insensitive loss function is used, which has some algorithmic advantages like providing sparse decomposition and rendering the computation more amenable [35]. Figure 3 shows Vapnik's ε-insensitive loss function. It is defined as

$$L_\varepsilon(t, f(x)) = \begin{cases} 0 & \text{if } |t - f(x)| \leq \varepsilon \\ |t - f(x)| - \varepsilon & \text{otherwise} \end{cases}. \tag{5}$$

The points that lie outside the ε-insensitive zone (outside red dotted lines) add to the cost. The deviations outside the ε-insensitive zone are penalized in a linear manner. The associated convex quadratic programming problem is

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*). \tag{6}$$

$$\text{Subject to} \quad \begin{aligned} t_i - (w.x_i + b) &\leq \varepsilon + \xi_i \\ (w.x_i + b) - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i &\geq 0 \\ \xi_i^* &\geq 0 \qquad i = 1, 2, \ldots, n, \end{aligned} \tag{7}$$

where ξ_i and ξ_i^* are slack variables. These variables represent the deviations for the training data points lying outside the ε-insensitive zone. The slack variables have zero value for all the data points lying inside the ε-insensitive zone and increase progressively for the data points lying outside the zone. C is the regularization constant that assigns a penalty when a training error occurs. The current optimization problem can be easily optimized in the dual form.

A dual set of variables, $\underline{\alpha}_i$ and $\overline{\alpha}_i$, is introduced to build the Lagrange function from the primal objective function [Eq. (6)] and the corresponding constraints [Eq. (7)]. The dual formation of the optimization problem is

$$\text{Maximize} \quad -\frac{1}{2}\sum_{i,j=1}^{n} \underline{\alpha}_i - \overline{\alpha}_i(\underline{\alpha}_j - \overline{\alpha}_j)(x_i.x_j) - \varepsilon\sum_{i=1}^{n}(\underline{\alpha}_i + \overline{\alpha}_i) + \sum_{i=1}^{n} y_i(\underline{\alpha}_i - \overline{\alpha}_i) \tag{8}$$

$$\text{Subject to} \quad \begin{aligned} \sum_{i=1}^{n}(\underline{\alpha}_i - \overline{\alpha}_i) &= 0 \\ 0 \leq \underline{\alpha}_i \leq C, &\qquad i = 1, 2, \ldots, n. \\ 0 \leq \overline{\alpha}_i \leq C, &\qquad i = 1, 2, \ldots, n \end{aligned}$$

After determining $\underline{\alpha}_i$ and $\overline{\alpha}_i$, the Karush-Kuhn-Tucker method is utilized to find the parameters w and b. Prediction in an LR function is expressed as

$$f(x) = \sum_{i=1}^{n} \alpha_i - \overline{\alpha}_i \langle x_i . x_j \rangle + b. \tag{9}$$

To make SVR deal with non-linear cases, a function ϕ is used to map the decision space variable (X) to a higher-dimensional space. The dual problem, after applying this transformation, becomes

$$\text{Maximize} \quad -\frac{1}{2} \sum_{i,j=1}^{n} \underline{\alpha}_i - \overline{\alpha}_i (\underline{\alpha}_i - \overline{\alpha}_j) \langle \phi(x_i).\phi(x_j) \rangle - \varepsilon \sum_{i=1}^{n} (\underline{\alpha}_i + \overline{\alpha}_i) + \sum_{i=1}^{n} y_i (\underline{\alpha}_i - \alpha_i). \tag{10}$$

$$\text{Subject to} \quad \sum_{i=1}^{n} (\underline{\alpha}_i - \overline{\alpha}_i) = 0$$
$$0 \leq \underline{\alpha}_i \leq C, \qquad i = 1, 2, \ldots, n,$$
$$0 \leq \overline{\alpha}_i \leq C, \qquad i = 1, 2, \ldots, n \tag{11}$$

where $k(x_i, x_j) = (\phi(x_i).\phi(x_j))$ is called the kernel function. Any function can be used as kernel function if it satisfies Mercer's theorem [34]. Gaussian radial basis function (GRBF) is used for this study, which can be defined as $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$.

The hyperparameters such as the SVR constant ε, regularization constant C, and σ (for radial basis function kernels) influence the performance of SVR. Parameter C allows a trade-off between training error and model complexity. If a model is not sufficiently complex, then it may fail to capture the underlying trend of the data and hence cause underfitting. On the other hand, if the model is too complex, then it may capture the noise of the data and, hence, can suffer from overfitting.

The magnitude of training error increases for the small values of C and hence tends to underfit the training data. However, a higher value of C will tend to a behavior as that of a hard-margin SVM and hence overfits the training data [13]. The parameter ε influences the number of support vectors by deciding the thickness of the ε-insensitive zone. Hence, its value affects both generalization capability and the complexity of the approximation function. A lower value of ε leads to higher number of support vectors and increases the complexity. A higher value of ε leads to smaller number of support vectors and results in more flat estimates of the regression function. The performance of SVR is sensitive to these parameters; hence, it is important to find appropriate values for C and ε [14]. Determining the appropriate value of these parameters is often a heuristic trial-and-error process [25]. It is shown that an exponentially growing sequence of parameters works better [29].

3.3 Wavelet Regression

Wavelet analysis is an advanced signal processing technique that has been extensively used in communications, data compression, acoustics, signal processing, image processing, etc. It was introduced by Grossmann and Morlet [10] in the early 1980s. The signal or function is expressed as the combination of "little waves" called wavelets. Unlike Fourier transform, where the signal is divided into smooth sinusoids of unlimited duration, the wavelet transform splits the signal into wavelets of finite duration and zero mean [16]. Wavelets are localized both in frequency and time domains. This property of wavelets is called time-frequency localization. Wavelet analysis features the analysis of temporal variations at different time scales, and can work on non-stationary signals as well. Fourier transform assumes infinite length signals and concentrates mainly on their frequencies only while wavelet transform can be applied to any kind and any size of the signal.

The wavelet function $\psi(t)$ is defined as $\int\limits_{-\infty}^{+\infty}\psi(t)dt=0$. The term $\psi_{a,b}(t)$ is obtained by compression and expansion of $\psi(t)$:

$$\psi_{a,b}(t)=|a|^{\frac{-1}{2}}\,\psi\left(\frac{t-b}{a}\right)b\in R,\,a\in R,\,a\neq 0, \tag{12}$$

where R represents real numbers, $\psi_{a,b}(t)$ is the successive wavelet, and a and b are the scale of frequency factor and time factor. When the term $\psi_{a,b}(t)$ satisfies Eq. (12), for a finite energy signal [16, 26, 40], the successive wavelet transform of $f(t)$ is

$$W_{\psi}f(a,\,b)=|a|^{\frac{-1}{2}}\int_{R}f(t)\overline{\psi}\left(\frac{t-b}{a}\right)dt, \tag{13}$$

where $\overline{\psi(t)}$ represents the complex conjugate function of $\psi(t)$. $f(t)$ is decomposed at a different resolution level (scale) in Eq. (13). In real applications, the successive wavelets are generally discrete. Let $a=a_0^j$, $b=kb_0a_0^j$, $a_0>1$, $b_0\in R$, and k,j are integers. The DWT of function $f(t)$ is defined as

$$W_{\psi}f(j,\,k)=a_0^{-j/2}\int_{R}f(t)\overline{\psi}(a_0^{-j}t-kb_0)dt. \tag{14}$$

The widely used values for parameter a_0 is 2 and for b_0 is 1 time step. When a_0 is 2 and b_0 is 1, Eq. (14) changes into binary wavelet transform [16] as

$$W_{\psi}f(j,\,k)=2^{-j/2}\int_{R}f(t)\overline{\psi}(2^{-j}t-k)dt. \tag{15}$$

$W_{\psi}f(a,\,b)$ or $W_{\psi}f(j,\,k)$ reflects the frequency characteristics (a or j) and time characteristics (b or k) of the original time series. The parameter a decreases when the temporal resolution is high and the frequency resolution of the wavelet transform is low. On the other hand, a increases when the temporal resolution is low and the frequency resolution is high [36]. For discrete time series $f(t)$ for different time steps (integral), the DWT is defined as [42]

$$W_{\psi}f(j,\,k)=2^{-j/2}\sum_{t=0}^{N-1}f(t)\overline{\psi}(2^{-j}t-k), \tag{16}$$

where $W_{\psi}f(j,\,k)$ is the wavelet coefficient for the discrete wavelet of scale $a=2^j$ and $b=2^jk$.

The input signal (monthly rainfall time series) is decomposed into a certain number of sub-time series components (Ds) using the Mallat DWT algorithm [21]. Every component has an important role in the original time series, and all have distinct behavior. Thus, the contribution of different subcomponents to the original time series varies from each other [36]. The decomposition process consists of a number of successive filtering steps. The original signal is first decomposed into approximation and detail components. It is broken down into many lower-resolution components. The details are low-scale high-frequency components of the signal, while the approximations are the high-scale, low-frequency components (Figure 4). The higher scales consist of the extended version of a wavelet, and the coefficients refer to the slowly-changing coarse features of low-frequency components. The lower scales present the condensed wavelet and follow the rapidly changing details (high-frequency components) of the signal [21].

The WR model used in the present study is a combination of DWT and LR. DWT is utilized for preprocessing of the original monthly rainfall series before use in the LR model. The decomposed subcomponents (detail and approximation) are utilized as inputs for LR, keeping the original rainfall series as input. The rainfall time series is processed using DWT separately for both calibration and validation period.
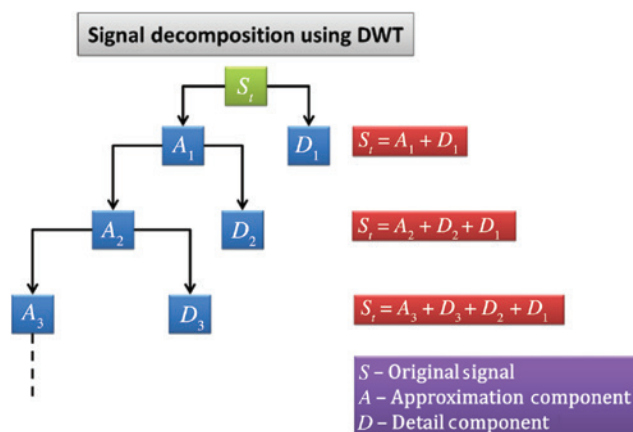
Signal decomposition using DWT

$S_t = A_1 + D_1$

$S_t = A_2 + D_2 + D_1$

$S_t = A_3 + D_3 + D_2 + D_1$

S – Original signal
A – Approximation component
D – Detail component

Figure 4: Signal Decomposition by DWT for Three Resolution Levels.
S_t denotes the original signal, A_i denotes the approximation components, and D_i denotes the detail components.

3.4 Models

Using these three techniques and two different sets of input variables, six different models were developed. The details of these models are provided in Table 2. All the models were calibrated and validated for the same data, as specified in Section 2.2.

4 Results and Discussion

Monthly rainfall at 21 stations in Assam is forecasted using three different techniques, namely BR, SVR, and WR. Out of 102 years of available data, the first 60 years data were used to calibrate the models and the remaining data were used for validation. For SVR, Gaussian radial basis function kernel was used. The first 60 years data were used for optimizing the value of SVM parameters, ε, C, and γ. A grid search was used for optimizing the values of parameters using 10-fold cross-validation. The values for the parameters have been varied in an exponentially growing sequence with $C = 2^0$, 2^2, …, 2^{22}, $\gamma = 2^{-15}$, 2^{-13}, …, 2^3 and epsilon $= 2^{-1}$, 2^0, …, 2^4. The values of these parameters were selected separately for all the stations. The optimized values of these three parameters for different stations are presented in Table 3.

For WR, rainfall series was pre-processed using DWT for calibration and validation periods separately. In this study, Haar wavelet (db1) and five levels of decompositions were used. The WR model structure is shown in Figure 5. Using DWT, a lagged monthly rainfall time series was decomposed into subcomponents, i.e. approximation (A_i) and detail (D_i) components. A new series was generated by adding the effective D_i components and the fifth approximation component (A_5). The effective D_i were chosen based on their correlations

Table 2: Different Models and Input Variables.

Model	Approach used	Input variables
WR1	Wavelet regression	
SVR1	Support vector regression	$R(t) = f(R[t-1])$
BR1	Bayesian regression	
WR2	Wavelet regression	
SVR2	Support vector regression	$R(t) = f(R[t-1], R[t-2])$
BR2	Bayesian regression	

Table 3: Optimized Values of SVM Model Parameters for Different Stations.

| Station	SVR1			SVR2		
	$\varepsilon^{(2x)}$	$C^{(2x)}$	$\gamma^{(2x)}$	$\varepsilon^{(2x)}$	$C^{(2x)}$	$\gamma^{(2x)}$
1	−1	2	−13	1	2	−15
2	4	8	−13	1	6	−15
3	−1	10	−15	−1	4	−15
4	0	10	−13	0	4	−15
5	0	2	−13	−1	4	−13
6	4	10	−15	−1	6	−15
7	1	4	−11	1	6	−11
8	4	6	−9	0	6	−15
9	−1	2	−13	−1	2	−15
10	−1	10	−15	1	6	−15
11	−1	8	−15	0	8	−15
12	1	6	−9	−1	4	−15
13	−1	16	−15	0	4	−15
14	3	6	−15	1	6	−15
15	1	2	−13	−1	4	−13
16	2	4	−13	−1	4	−15
17	4	6	−13	0	4	−15
18	−1	6	−11	1	4	−15
19	0	2	−13	−1	4	−13
20	3	4	−13	3	4	−13
21	−1	2	−11	2	4	−13

with the original time series of rainfall. The newly generated time series was used as input to the LR model, and the original rainfall time series was used as output.

4.1 Comparison of BR, SVR, and WR

Four different statistics, namely mean absolute error (MAE), root mean square error (RMSE), correlation coefficient (R), and Nash-Sutcliffe efficiency coefficient (NS) [43], were used to evaluate the performance of the three techniques. These statistics are defined as follows:
1. The MAE between the observed and the predicted outputs can be defined as

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i|.$$
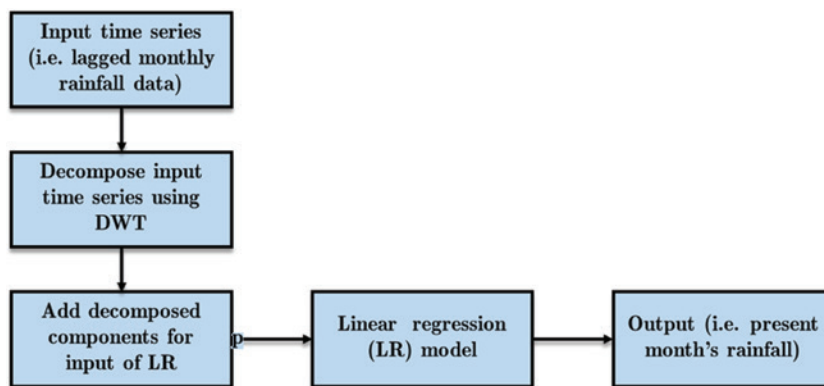
Figure 5: Wavelet Regression Model Structure.

2. The RMSE between the observed and the predicted outcomes can be defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2}.$$

3. *R* is defined as

$$R = \frac{\sum_{i=1}^{n}(y_i - \overline{y}_i)(f_i - \overline{f}_i)}{\left[\sum_{i=1}^{n}(y_i - \overline{y}_i)^2 \cdot \sum_{i=1}^{n}(f_i - \overline{f}_i)^2\right]^{\frac{1}{2}}}.$$

4. NS is defined as

$$\mathrm{NS} = 1 - \frac{\sum_{i=1}^{n}(f_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2},$$

where f_i are the estimated values, $\overline{f}_i$ is the average of estimated values, y_i are the observed values, and $\overline{y}_i$ is the average of observed values.

Figure 6A–D shows the comparison of the performance of six models during the validation period. As observed from the figure, the WR2 model outperformed all other models for all 21 stations. The MAE for WR2 was observed to lie between 62.89 mm at station no. 19 and 119.813 mm at station no. 11. The RMSE for WR2 was found to lie between 93.06 mm (for station no. 15) and 185.79 mm (for station no. 11), while the RMSE for SVR2 model was observed between 106.15 and 214.08 mm. The average RMSEs for BR1, SVR1, WR1, BR2, SVR2, and WR2 were 172.84, 151.24, 142.58, 173.89, 145.42, and 127.06 mm, respectively. The WR2 model had the least RMSE compared to all other models.

Figure 6C shows the *R* for different models at all stations during the validation period. The value of *R* for WR2 was found to be between 0.794 at station no. 21 and 0.855 at station no. 8. The average values of *R* for BR1, SVR1, WR1, BR2, SVR2, and WR2 were 0.667, 0.702, 0.750, 0.643, 0.729, and 0.828, respectively. Figure 6D shows the Nash-Sutcliffe efficiency index (NSE) for different models at all stations during the validation period. The maximum and minimum values of NSE for WR2 were 0.694 at station 8 and 0.578 at station 21. The average values of NSE for BR1, SVR1, WR1, BR2, SVR2, and WR2 were 0.334, 0.491, 0.547, 0.327, 0.529, and 0.640, respectively. Considering *R* and NSE, it is clear that the WR2 model performed better than all other models.

Rainfall is one of the most complex and challenging hydrological processes to understand and to model because of the complexities involved with the associated atmospheric processes that generate rainfall. The rainfall time series is affected by many complex factors having different frequency components. Considering only one resolution component, the model is unable to capture the internal mechanism behind the rainfall process. Using DWT, the complex rainfall time series were decomposed into several simple time series, and the WR model was developed utilizing appropriate subcomponents. The performance of WR model was better than the other models due to the fact that it uses different resolution components for LR. The ineffective frequency components were removed from the rainfall time series, which effectively improved the performance of the LR model. WR is a simpler model compared to the non-linear SVR model; however, WR has shown better accuracy in rainfall prediction. This signifies the role of wavelet preprocessing of the rainfall time series, which de-noised the inputs and resulted in improved prediction ability of the model.

To assess the spatial variations in the performance of the forecasting models, the state was divided into four zones: lower Assam, middle Assam, upper Assam, and southern Assam (Figure 7). Table 4 presents
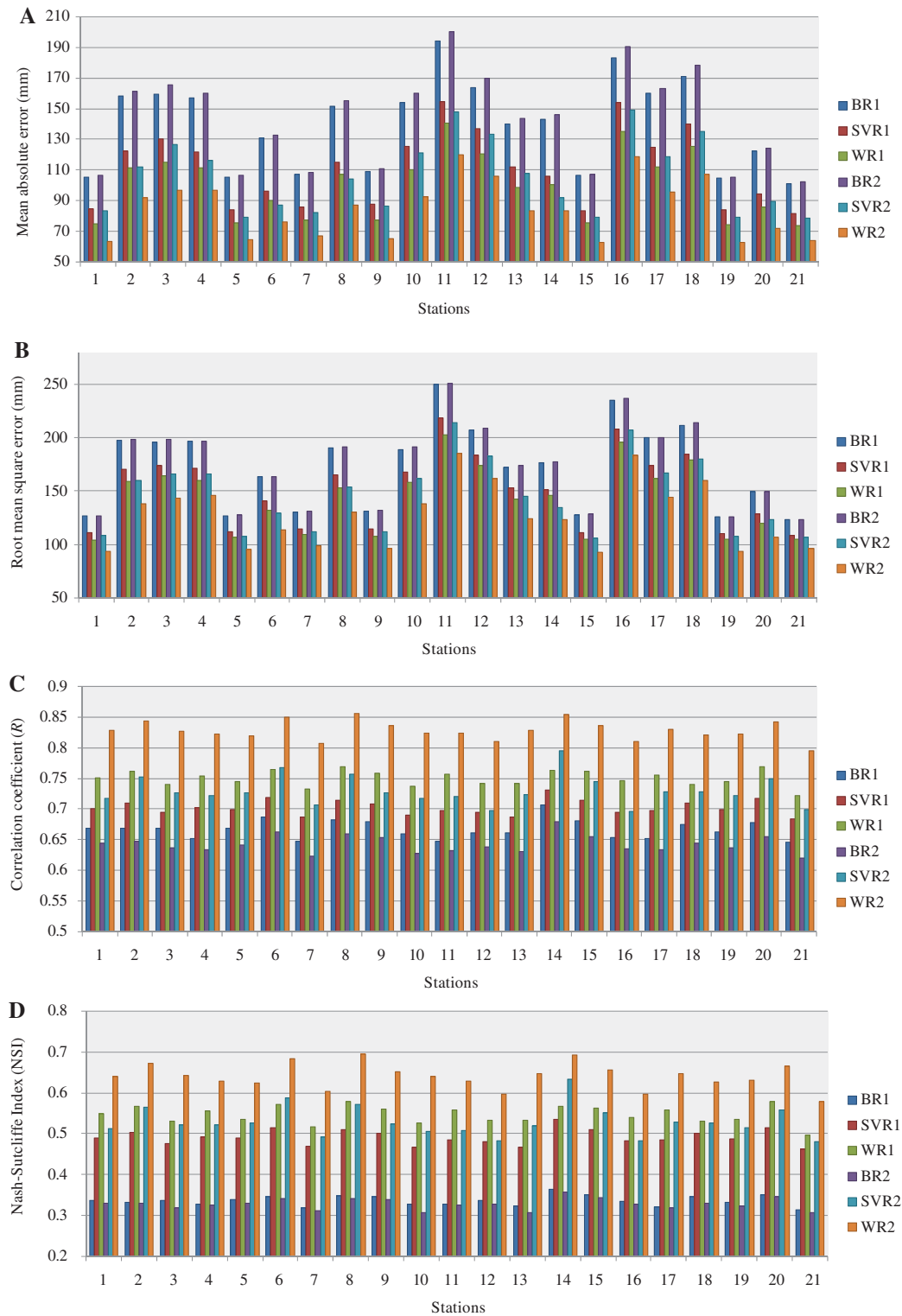
Figure 6: (A) MAE for Different Models at 21 Stations. (B) RMSE for Different Models at 21 Stations. (C) Correlation Coefficient (*R*) for Different Models at 21 Stations. (D) Nash-Sutcliffe Index (NS) for Different Models at 21 Stations.

the average RMSEs for four regions, obtained using different models. A comparison of model performances reveals that the upper Assam part has the highest accuracy (least RMSEs), whereas middle Assam has the least accuracy (highest RMSEs). The average standard deviations of monthly rainfall for upper, middle, lower,
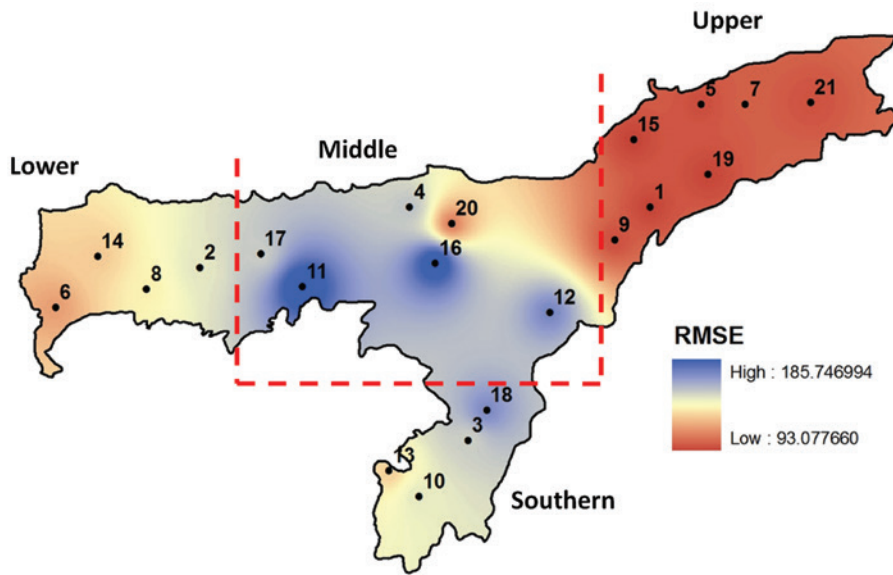
Figure 7: Map Showing the Distribution of RMSEs for the WR2 Model over Assam.

Table 4: RMSEs Obtained for Different Parts of Assam Using Six Models.

Region	RMSE (mm)					
	BR1	SVR1	WR1	BR2	SVR2	WR2
Upper	127.42	111.81	106.19	128.12	108.71	95.31
Middle	206.65	180.82	169.10	207.37	176.85	154.33
Lower	182.18	157.02	147.59	182.88	144.55	126.26
Southern	192.27	170.08	161.48	194.80	163.41	141.25

and southern Assam were 168, 254, 222, and 259 mm, respectively. The upper Assam region had the minimum average standard deviation, which could be the reason for the better performance of models in this region, as variability in station rainfalls for this region was least. Figure 7 shows the performance of the WR2 model (best model) in the form of a map made using inverse distance weighted interpolation of RMSEs at different stations. The average RMSEs for lower, middle, upper and southern Assam were 126.26, 154.33, 95.31, and 141.25 mm, respectively. Upper Assam had the lowest RMSE followed by lower, southern, and middle Assam, respectively. The WR2 model performed best for the stations in the upper Assam zone. Stations in the middle Assam had the highest RMSEs, which imply that the forecasting model performed badly at these stations. Of 21 stations, station no. 15 (Lakhimpur) had the least RMSE for the WR2 model, whereas station no. 11 (Kamrup) had the highest RMSE. It is already shown in Table 1 that the Kamrup station had the highest values of standard deviation and skewness.

5 Conclusions

A comparison of three different forecasting techniques, BR, SVR, and WR, for monthly rainfall data at 21 stations in Assam, India, is presented in this paper. The GRBF kernel was used for SVR, and the SVR parameters were optimized utilizing a grid search technique. For the WR model, the rainfall time series was decomposed into subcomponents using DWT and a new time series was generated by adding effective subcomponents, which were used as input for the LR model. The monthly rainfall data of 102 years were used for the study,

which were divided into 60:40 ratio for calibration and validation, respectively. Based on the MAE, RMSE, R, and NSE, WR was found to be the most efficient and BR was found the least efficient for rainfall prediction. The efficiency index for BR, SVR, and WR was found to be 32.8%, 52.9%, and 64.03%, respectively. The WR model used in this study was a linear model; however, SVR is an advanced non-linear model. WR is simpler than SVR but was able to perform better than SVR. Hence, it can be concluded that wavelet processing of the input time series resulted in better performance of the WR model. Spatial assessment of model performance reveals that model performance was found best for stations in the upper part of the state, followed by the lower, southern, and middle parts. The study can be useful for better management of water resources in the state.

Bibliography

[1] M. Behzad, K. Asghari, M. Eazi and M. Palhang, Generalization performance of support vector machines and neural networks in runoff modeling, *Expert Syst. Appl.* **36** (2009), 7624–7629.
[2] M. Bray and D. Han, Identification of support vector machines for runoff modelling, *J. Hydroinform.* **6** (2004), 265–280.
[3] C. Chou and R. Wang, On-line estimation of unit hydrographs using the wavelet-based LMS algorithm, *Hydrol. Sci. J.* **47** (2003), 721–738.
[4] A. S. Cofino, R. Cano, C. Sordo and J. M. Gutierrez, Bayesian networks for probabilistic weather prediction, *Proc. 15th Eur. Conf. Artif. Intell.* **700** (2002), 695–700.
[5] S. Coles and J. A. Tawn, A Bayesian analysis of extreme analysis rainfall data process, *J. Roy. Stat. Soc. Ser. C* **45** (2011), 463–478.
[6] X. Dai, P. Wang and J. Chou, Multiscale characteristics of the rainy season rainfall and interdecadal decaying of summer monsoon in North China, *Chin. Sci. Bull.* **48** (2003), 2730–2734.
[7] Y. B. Dibike, S. Velickov, D. Solomatine and M. B. Abbott, Model induction with support vector machines: introduction and applications, *J. Comput. Civ. Eng.* **15** (2001), 208–216.
[8] L. Frechette and A. Khan, Bayesian regression-based urban traffic models, *Transp. Res. Rec.* **1644** (1998), 157–165.
[9] M. K. Goyal, Monthly rainfall prediction using wavelet regression and neural network: an analysis of 1901–2002 data, Assam, India, *Theor. Appl. Climatol.* **118** (2014), 25–34.
[10] A. Grossmann and J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape, *SIAM J. Math. Anal.* **15** (1984), 723–736.
[11] S. R. Gunn, Support vector machines for classification and regression, *Technical Report*, 1998.
[12] W. C. Hong and P. Ping-Feng, Potential assessment of the support vector regression technique in rainfall forecasting, *Water Resour. Manag.* **21** (2007), 495–513.
[13] T. Joachims, Optimizing search engines using clickthrough data, in: *Proc. Eighth ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. KDD' 02*, pp. 133–142, 2002.
[14] K. Kim, Financial time series forecasting using support vector machines, *Neurocomputing* **55** (2003), 307–319.
[15] S. U. Kim and K. S. Lee, Regional low flow frequency analysis using Bayesian regression and prediction at ungauged catchment in Korea, *KSCE J. Civ. Eng.* **14** (2009), 87–98.
[16] O. Kisi, Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting, *Hydrol. Process.* **23** (2009), 3583–3597.
[17] O. Kisi, Wavelet regression model as an alternative to neural networks for river stage forecasting, *Water Resour. Manag.* **25** (2011), 579–600.
[18] D. Labat, R. Ababou and A. Mangin, Rainfall-runoff relations for karstic springs, part II: Continuous wavelet and discrete orthogonal multiresolution analyses, *J. Hydrol.* **238** (2000), 149–178.
[19] S. Lee, S. Cho and P. M. Wong, Rainfall prediction using artificial neural networks, *J. Geogr. Inf. Decis. Anal.* **2** (1998), 233–242.
[20] J. Y. Lin, C. T. Cheng and K. W. Chau, Using support vector machines for long-term discharge prediction, *Hydrol. Sci. J.* **51** (2006), 599–612.
[21] S. G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (1989), 674–693.
[22] H. D. Navone and H. A. Ceccatto, Predicting Indian monsoon rainfall: a neural network approach, *Clim. Dyn.* **10** (1994), 305–312.
[23] V. B. Nikam and B. B. Meshram, *Modeling Rainfall Prediction Using Data Mining Method: A Bayesian Approach,* pp. 132–136, IEEE, 2013.
[24] E. Osuna and R. Freund, An improved training algorithm for support vector machines, in: *Neural Networks Signal Process VII – Proc. IEEE*, pp. 276–285, 1997.

[25] S. Raghavendra and P. C. Deka, Support vector machine applications in the field of hydrology: a review, *Appl. Soft Comput. J.* **19** (2014), 372–386.

[26] O. A. Rosso, A. Figliola, S. Blanco and P. M. Jacovkis, Signal separation with almost periodic components: a wavelets based method, *Rev. Mex. Fis.* **50** (2004), 179–186.

[27] N. Sethi and K. Garg, Exploiting data mining technique for rainfall prediction, *Int. J. Comput. Sci. Inf. Technol.* **5** (2014), 3982–3984.

[28] A. Sharma and M. K. Goyal, Bayesian network model for monthly rainfall forecast, in: *2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India*, pp. 241–246, IEEE, 2015.

[29] N. Sharma, M. Zakaullah, H. Tiwari and D. Kumar, Runoff and sediment yield modeling using ANN and support vector machines: a case study from Nepal watershed, *Model Earth Syst. Environ.* **1** (2015), 23.

[30] C. Sivapragasam, S. Liong and M. Pasha, Rainfall and runoff forecasting with SSA-SVM approach, *J. Hydroinform.* **3** (2001), 141–152.

[31] L. Smith, D. Turcotte and B. Isacks, Stream flow characterization and feature detection using a discrete wavelet transform, *Hydrol. Process.* **12** (1998), 233–249.

[32] A. M. Ticlavilca and M. McKee, Multivariate Bayesian regression approach to forecast releases from a system of multiple reservoirs, *Water Resour. Manag.* **25** (2011), 523–543.

[33] V. N. Vapnik, *The nature of statistical learning theory*, Springer, Berlin, 1995.

[34] V. N. Vapnik, An overview of statistical learning theory, *IEEE Trans. Neural Netw.* **10** (1999), 988–999.

[35] V. Vapnik, S. E. Golowich and A. Smola, Support vector method for function approximation, regression estimation, and signal processing, in: *Advances in Neural Information Processing Systems*, vol. 9, San Mateo, CA, 1996.

[36] W. Wang and J. Ding, Wavelet network model and its application to the prediction of hydrology, *Nat. Sci. Sleep*, **1** (2003), 67–71.

[37] C. C. Wei, Wavelet support vector machines for forecasting precipitation in tropical cyclones: comparisons with GSVM, regression, and MM5, *Weather Forecast* **27** (2012), 438–450.

[38] J. P. Liu, M. Q. Chang, X. Y. Ma, Groundwater quality assessment based on support vector machine. In: *HAIHE River Basin Research and Planning Approach-Proceedings of 2009 International Symposium of HAIHE Basin Integrated Water and Environment Management*, Beijing, China, 2009, pp. 173–178.

[39] Z. Zakaria and A. Shabri, Streamflow forecasting at ungaged sites using support vector machines, *Appl. Math. Sci.* **6** (2012), 3003–3014.

[40] H. C. Zhou, Y. Peng and G. H. Liang, The research of monthly discharge predictor-corrector model based on wavelet decomposition, *Water Resour. Manag.* **22** (2008), 217–227.

[41] A. Mahanta, D. Jhingran, C. K. Das and A. Partie. *Assam Human Development Report 2003*, 2003, Available at http://hdr.undp.org/en/reports/nationalreports/asiathepacific/india/name,3268,en.html. Accessed 20 July, 2013.

[42] G. Urcid and G. X. Ritter, Advances in knowledge-based and intelligent information and engineering systems, in: *Advances in Knowledge-Based and Intelligent Information and Engineering Systems*, M. Graña, C. Toro, J. Posada, R. J. Howlett and L. C. Jain, eds., pp. 2140–2149, IOS Press, 2012.

[43] J. E. Nash and J. V. Sutcliffe, River flow forecasting through conceptual models, part I – a discussion of principles, *J. Hydrol.* **10** (1970), 282–290.