

Research Article

Improving Accuracy of River Flow Forecasting Using LSSVR with Gravitational Search Algorithm

Rana Muhammad Adnan,¹ Xiaohui Yuan,¹ Ozgur Kisi,² and Rabia Anam³

¹*School of Hydropower and Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China*

²*Center for Interdisciplinary Research, International Black Sea University, Tbilisi, Georgia*

³*Faculty of Agricultural Engineering & Technology, Department of Farm Machinery & Power, University of Agriculture, Faisalabad, Pakistan*

Correspondence should be addressed to Xiaohui Yuan; yxh71@hust.edu.cn

Received 1 December 2016; Revised 15 December 2016; Accepted 18 December 2016; Published 19 January 2017

Academic Editor: James Cleverly

Copyright © 2017 Rana Muhammad Adnan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

River flow prediction is essential in many applications of water resources planning and management. In this paper, the accuracy of multivariate adaptive regression splines (MARS), model 5 regression tree (M5RT), and conventional multiple linear regression (CMLR) is compared with a hybrid least square support vector regression-gravitational search algorithm (HLGSA) in predicting monthly river flows. In the first part of the study, all three regression methods were compared with each other in predicting river flows of each basin. It was found that the HLGSA method performed better than the MARS, M5RT, and CMLR in river flow prediction. The effect of log transformation on prediction accuracy of the regression methods was also examined in the second part of the study. Log transformation of the river flow data significantly increased the prediction accuracy of all regression methods. It was also found that log HLGSA (LHLGSA) performed better than the other regression methods. In the third part of the study, the accuracy of the LHLGSA and HLGSA methods was examined in river flow estimation using nearby river flow data. On the basis of results of all applications, it was found that LHLGSA and HLGSA could be successfully used in prediction and estimation of river flow.

1. Introduction

River flow forecasting plays a vital role in planning of water projects, irrigation systems, hydropower system, and optimized utilization of water resources [1]. Due to continuous increase of population growth, industrial uses, and irrigation needs, the river flow forecasting has received great attentions of researchers for operational river management [2]. Forecasting of river flow provides alerts of approaching floods and also assists in controlling the outflows of reservoir during low flows days of river. Floods affect countless lives, infrastructure, and property and cause limitless damage more than any other natural disaster. Due to no assessment of flood magnitude, a flood resulted in a loss of thousand lives and damage of agriculture land of million dollars in Pakistan in 2010 [3]. It is not possible to provide complete safety from flood, but high amounts of money and many lives

can be saved by providing accurate flood predictions, flood magnitude, and flood duration [4]. The importance of water measurement compelled researchers to apply various types of forecasting methods to estimate and forecast river flows.

From the last three decades of the previous century, the statistical methods were applied successfully in the field of hydrology including the river flow forecasting. Statistical methods try to find inherent relationships within the actual data. The autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) methods are the most popular in the statistical methods category and have been extensively used to model different variables in the field of hydrology [5–15]. Ahlert and Mehta [5] and Kurunç et al. [7] used ARIMA statistical models for modeling river flows data. Ahmad et al. [6] and Mirzavand and Ghazavi [8] applied ARIMA statistical methods to analyse water quality and groundwater data,

respectively. Otok and Suhartono [11], Rabenja et al. [13], and Valipour [14] forecasted runoff data in Indonesia and USA, respectively, by applying SARIMA model and compared with ARIMA. Psilovikos and Elhag [12] and Yang et al. [15] applied ARIMA models successfully in modeling different processes of evaporation data. Mishra and Desai [9] and Modarres [10] used SARIMA method efficiently for drought forecasting in India and Iran, respectively. In the previous two decades, the artificial neural networks (ANN) have been replaced with the statistical methods in solving different problems due to their flexible nature and capturing the nonlinearity in the data. In the literature, many researchers compared the ANN with statistical methods in solving many problems of hydrology and reported that ANN outperformed the statistical methods [16–21]. Huang et al. [18] used the ANN method to forecast the river flows of Apalachicola River, USA, using the previous rainfall and river flow data. They compared the quarterly and yearly river flow forecasts results with the ARIMA method's results and found that the ANN performed better than the ARIMA method in prediction of river flow. The detailed discussion of all ANN applications in comparison with statistical methods to forecast different variables in hydrology is not possible in this paper. However, ANN also have some major weakness, that is, overfitting, falling into local minimum, slowing convergence speed, and requiring large number of training data. Thus, in the last decade, the support vector regression (SVR) took priority over ANN due to its parallel distributed processing, self-learning features, avoiding the overfitting problems, and providing globally optimal solutions [22–28]. Ahmad et al. [22] applied SVR model to forecast runoff of Bakhtiyari Basin, Iran, and results explored that SVR showed better accuracy than the ANN methods for daily runoff forecasting especially in case of prediction of higher values of river flows. However, SVM faces computationally difficulties in determining optimal solution due to use of quadratic programming with nonlinear equation. This procedure is time consuming.

Recently least square support vector regression (LSSVR), the improved version of SVR, received much attention in the field of prediction methods due to use of linear squares principle for the loss function instead of the quadratic programming in the SVR method and fast computational speed [29–38]. Shabri and Suhartono [37] and Kisi [34] applied LSSVR successfully to forecast river flows. Shabri and Suhartono [37] compared the prediction accuracy of LSSVR with ANN and multivariate linear regression methods whereas Kisi [34] compared it with adaptive neurofuzzy embedded fuzzy *c*-means clustering (ANFIS-FCM) method and they both found that LSSVR performed better than the other methods. Kisi [33] and Goyal et al. [30] forecasted the reference evapotranspiration and pan evaporation by using LSSVR method. Kisi [33] compared its results with feed forward ANN whereas Goyal et al. [30] compared it with ANN and ANFIS methods and they found that LSSVR performed better than the other methods. Okkan and Serbes [36] and Bhagwat and Maity [29] successfully used LSSVR method to forecast runoff data by using the previous meteorological and river flows data. Kisi [32] estimated the suspended sediment by using the river flow data through

LSSVR method and reported that the LSSVR gave better estimates in comparison with ANN and sediment rating curve (SRC) methods. Hwang et al. [31] predicted the daily water demand of the Seoul City, Korea, and daily mean inflow of Cheng-ju Dam by using LSSVR model. He compared the predicted results of LSSVR with the conventional multiple linear regression (CMLR) and back propagation neural network methods in both cases and found that LSSVR showed superiority in prediction accuracies. Wu et al. [38] and Mellit et al. [35] applied LSSVR method to predict the different meteorological variables and found that LSSVR performed better than ANN method. Motivated by these successful applications of the LSSVR, the LSSVR method was selected as a forecasting method in this research. LSSVR method has parameters which play a vital role in determining the prediction accuracy of the method. Determining suitable value of these parameters will produce better river flow prediction results. Still, there is no specific way to determine optimal parameters for LSSVR method in the literature of river flow forecasting. Thus, the novelty of this study is to generate a hybrid LSSVR-gravitational search algorithm (HLGSA) river flow forecasting method. In this study, GSA is used to find the optimal values of LSSVR method to increase the prediction accuracy of the method. GSA was preferred in this study over other heuristic algorithms such as simulated annealing algorithm, genetic algorithm, memetic algorithm, differential evolution, and particle swarm optimization due to their premature convergence, parameter sensitivity, and consuming too much time to obtain global optimal solution. Instead of these heuristic algorithms, GSA improves the global search ability and optimization speed by using the principle of gravity and motion. To the best knowledge of the authors, there is not any published work in the literature that predicts the river flow using hybrid LSSVR-gravitational search algorithm (HLGSA) method. Recently, researchers preferred hybrid methods for solving different problems in the field of hydrology.

In addition to the HLGSA method, multivariate adaptive regression splines (MARS) is another popular regression method used to model the complex nonlinear relationships among the variables. MARS is a nonparametric regression method and it has been applied extensively nowadays in the field of hydrology to predict different variables [39–45]. To determine the benefits of using MARS over other conventional regression methods, MARS method was compared with CMLR and model 5 regression tree (M5RT) methods in this study. Cross validation (CV) technique was used to better see the prediction accuracy of all applied methods. Log transform function was also utilized in this study to see its effect on the prediction accuracy of these methods.

2. Hybrid LSSVR-Gravitation Search Algorithm (HLGSA) Method for River Flow Prediction

2.1. LSSVR. LSSVR introduced by Suykens and Vandewalle [46] is a modification version of SVR and has advantage on SVR due to reduction in complexity of optimization process

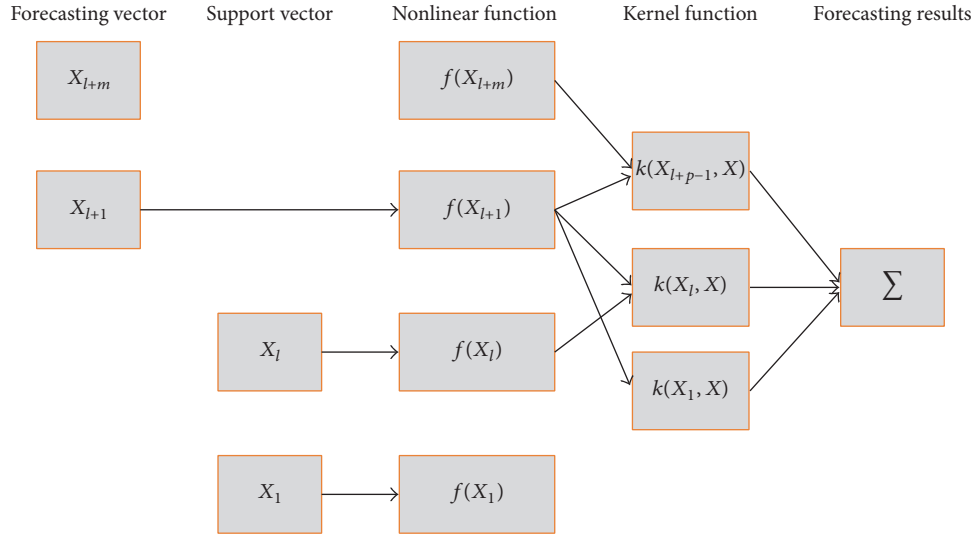


FIGURE 1: LSSVR structure for river flow forecasting.

by using linear equation instead of quadratic equations [47]. Figure 1 demonstrates the process of LSSVR algorithm. By using time series inputs x_i (lagged river flows) and output y_i (predicted river flow), the function of nonlinear LSSVR is given as

$$y(x) = \langle \omega^T, \varphi(x) \rangle + b, \quad (1)$$

where $\langle \cdot, \cdot \rangle$ represents dot product, $\varphi(x)$ is a nonlinear function that employs regression, and ω and b are the weight vector and bias term, respectively [48]. The cost function (C) of LSSVR can be minimized as

$$\text{Min } C = \frac{1}{2} \omega^T \omega + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \quad (2)$$

$$\text{Subject to } y_i = \langle \omega, \varphi(x) \rangle + b + e_i \quad (3)$$

$$(i = 1, 2, \dots, n),$$

where γ, e_i represent the regularization constant and the training error for x_i , respectively.

To solve (2), the Lagrange function is used to find the solutions of ω and e . The Lagrange function can be calculated as [49]

$$L_{\text{LSSVR}} = C - \sum_{i=1}^N \alpha_i \{ \langle \omega, \varphi(x) \rangle + b + e_i - y_i \}, \quad (4)$$

where $\alpha_i \in R$ is the Lagrange multipliers.

The solution of above equation can be achieved by determining the partial differential of Lagrange function and applying the kernel function (KF) to satisfy Mercer's condition. To solve regression problems, there are many types of KF including polynomial, radial basis, Gaussian, sigmoid, Mexican hat, Meyer, and Morlet. The KF type plays a vital role in constructing high accurate LSSVR model [50]. This study used the radial basis KF (RBKF) due to its effectiveness for

the nonlinear regression problems [51]. The performance of the RBKF with other KFs is shown in Section 7. RBKF can be expressed as

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right). \quad (5)$$

After selecting the RBKF for the LSSVR method, finding proper values for penalty factor parameter that is γ and RBKF parameter that is σ^2 is necessary. There is no specific way to obtain the optimal values of parameters. Due to these reasons, GSA is adopted in the study to calculate the suitable parameter values.

2.2. Gravitational Search Algorithm (GSA). GSA is one of the effective optimization algorithms compared with other evolutionary algorithms. It is based on the law of gravity and motion and first proposed by Rashedi et al. [52]. In GSA, each agent has four parameters: position, velocity, inertial mass, and gravitational mass. The location of the agent corresponds to a solution of the problem whereas its gravitational and inertia masses are obtained utilizing a fitness function [53]. The location of particle can be expressed as

$$X_i = (x_i^1, \dots, x_i^k, \dots, x_i^s) \quad i = 1, 2, \dots, np, \quad (6)$$

where x_i^k represents the location of the i th agent in the k th dimension. The mass of each agent is computed after calculating the fitness of current population as [52, 54]

$$m_i(t) = \frac{\text{fit}_i(t) - \text{worst}(t)}{\text{best}(t) - \text{worst}(t)}, \quad (7)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^N m_j(t)},$$

where $\text{fit}_i(t)$ and $M_i(t)$ represent the fitness value and mass of the i th agent at time t , respectively, whereas $\text{best}(t)$ and

worst(t) represent the minimum fitness value and maximum fitness value, respectively.

To calculate the gravitational acceleration of the agent i , firstly the force exerted by heavy agents on this agent should be computed as

$$F_{ij}^k(t) = G_c(t) \frac{M_i(t) \cdot M_j(t)}{\|x_i(t), x_j(t)\|_2 + \varepsilon} \cdot (x_i^k(t) - x_j^k(t)), \quad (8)$$

where $M_i(t)$ and $M_j(t)$ are the passive and active gravitational mass, respectively, corresponding to agents i and j at the t generation, $G_c(t)$ and ε are the gravitational and small constant, $x_i^k(t)$ and $x_j^k(t)$ indicate position of k th dimension of agents i and j at the t generation, and $\|x_i(t), x_j(t)\|$ is Euclidean distance between agents i and j .

The total gravitational acceleration of the i th agent can be calculated using the law of motion as follows:

$$a_i^k(t) = \frac{\sum_{j=1, j \neq i}^N \text{rand} \cdot F_{ij}^k(t)}{M_i(t)}, \quad (9)$$

where $a_i^k(t)$ represents the gravitational acceleration of the agent i in the k th dimension and rand indicates a random variable with uniform distribution in the interval $[0, 1]$. With the help of (9), the total gravitational force exerted on the agent i in the k th dimension can be calculated as

$$F_i^k(t) = M_i(t) \times a_i^k(t) = \sum_{j=1, j \neq i}^N \text{rand} \cdot F_{ij}^k(t). \quad (10)$$

Then the speed and location of the agent are updated as follows:

$$\begin{aligned} v_i^k(t+1) &= \text{rand} \cdot v_i^k(t) + a_i^k(t), \\ x_i^k(t+1) &= x_i^k(t) + v_i^k(t+1). \end{aligned} \quad (11)$$

It is clearly seen from the brief description of the GSA that it utilizes the gravitational force as the direct form to communicate the agents' cooperation. The heavy agents in GSA are processed, infer good solutions, and move more gradually than lighter ones, which guarantee the algorithm's exploitation step. In other words, the GSA searches for the ideal solution by appropriately calibrating the inertia and gravitational masses of agents where every agent provides a solution. As time progresses, the heaviest agent will exhibit an ideal solution in the search space [55].

2.3. HLGSA (Hybrid LSSVR-GSA). The process of constructing the river flow prediction model HLGSA by using the hybrid of LSSVR and GSA methods is described in this section and shown in Figure 2. The process is as follows:

- (i) Firstly, divide all river flow data sets into training and test parts.
- (ii) Select the RBF kernel function and initial parameters for the HLGSA method to build the initial LSSVR model. The initial value of the parameters is set as

follows: the range of penalty factor γ is 0.1 to 2000, the range of RBF parameter σ^2 is 0.001 to 20, number of iterations is 15, the number of particles can be set up to 40, and constant alpha is found to be better in range of 16 to 20, whereas initial gravitational constant G_0 is found to be better in range from 105 to 115.

- (iii) Compute the particle fitness value of each agent. In this paper, RF_{RMSE} is selected as the fitness function. The fitness function for this method can be defined as

$$\begin{aligned} \text{fitness function} &= \text{RF}_{\text{RMSE}} \\ &= \sqrt{\frac{1}{N_{\text{RF}}} \sum_{t=1}^{N_{\text{RF}}} (\text{RF}_o - \text{RF}_f)^2}. \end{aligned} \quad (12)$$

- (iv) Choose the best parameters combination through GSA to obtain the optimal values of the LSSVR parameters.
- (v) If it does not meet the stopping criterion, then utilize the new combination of parameters to reconstruct the LSSVR. Compute the fitness until it suits the stopping criterion.
- (vi) The ideal parameter values are achieved to build the optimal LSSVR model for forecasting river flow. Now, the testing values are used for the optimal LSSVR model to get river flow prediction results.

3. Regression Methods

In this study, the performance accuracy of a hybrid nonlinear optimized regression method (HLGSA) was compared with a nonlinear, nonparametric regression method (MARS), with a piecewise linear regression method (M5RT), and with a conventional linear regression method (CMLR) in forecasting monthly river flow.

3.1. Multivariate Adaptive Regression Splines. MARS is a flexible method which finds relationships that are nearly additive or involve interactions with fewer parameters. The general MARS method is introduced by Friedman [56] and is expressed by the following equation:

$$\text{RF}_f = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^K b_{km}(x_{(k,m)}), \quad (13)$$

where RF_f is the forecasted river flow by the MARS that is dependent variable, a_0 is a constant, a_m are the model coefficients calibrated to provide the best fit to the used data, M is the quantity of basis functions (BFs), k is the "splits" quantity that generates the m th BFs, and b_{km} gets values of 1 or -1 and represents the (right/left) sense of the associated step function. $x_{(k,m)}$ is the independent variable's label [57].

Two-step MARS provides optimal MARS model. MARS develops a huge number of BFs chosen to overfit the data at first step, where variables are permitted to enter—as continuous, categorical, or ordinal—the formal system by

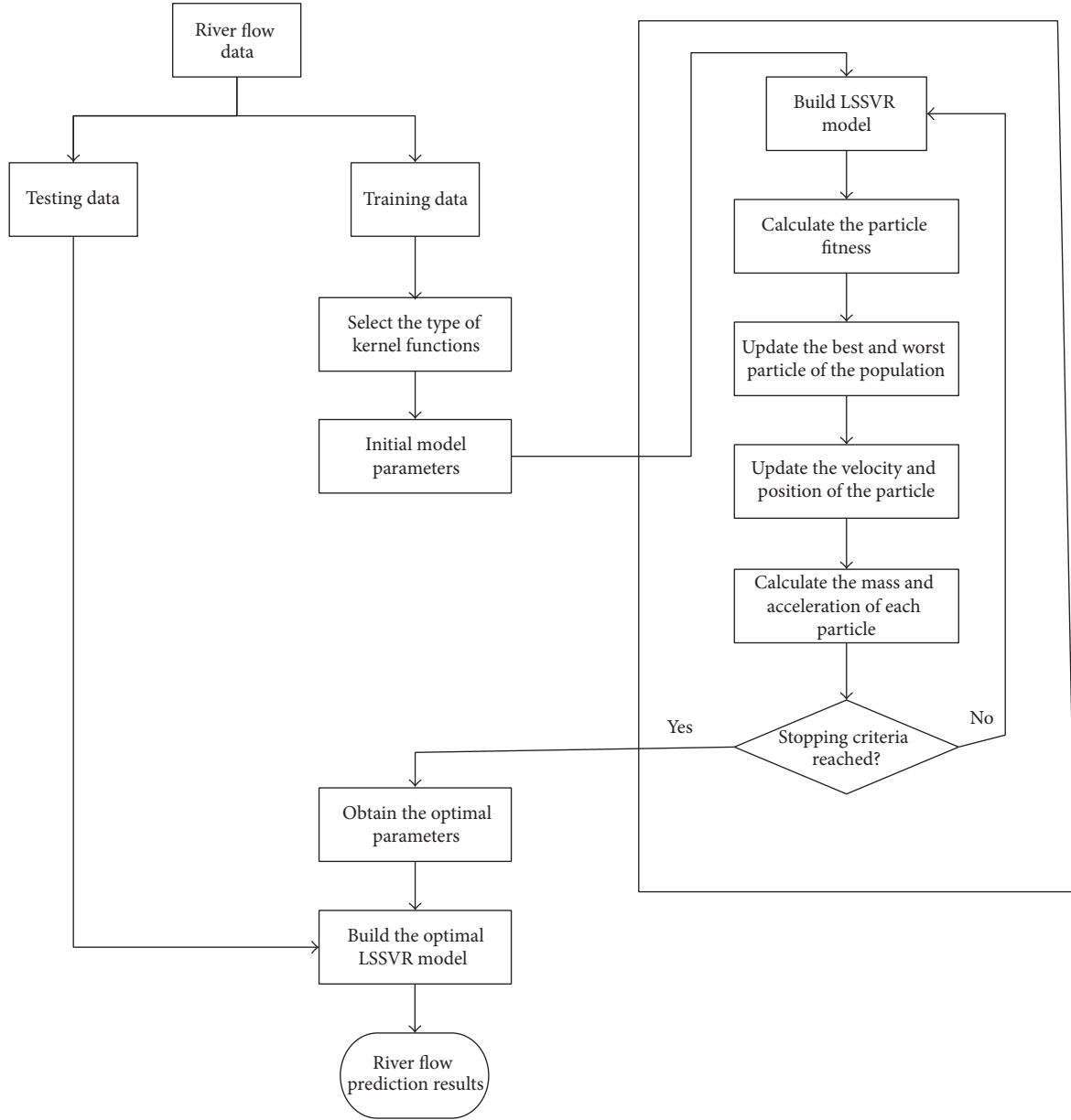


FIGURE 2: Flow chart of HLGSA model for river flow forecasting.

which variable ranges are characterized, and they can interact with each other or be restricted to enter only as additive components. In the second step, BFs are erased in the order of minimum effect utilizing the generalized cross validation criterion (GCV). A measure of variable significance can then be evaluated by watching the decrement in the computed GCV when a parameter is excluded from the model. This procedure proceeds until the rest of the BFs all satisfy the predecided necessities. The GCV can be computed as [58]:

$$GCV(M) = \frac{(1/n) \sum_{i=1}^n (RF_o - RF_f)^2}{(1 - (C(M)/n))^2}, \quad (14)$$

where RF_o and RF_f are the actual and predicted river flow values and $C(M)$ is a complexity penalty function.

After building the MARS model, the relative importance of a variable in terms of its contribution to the fit of the model can be estimated. MARS is capable of tracking very complex data structures, so selected in this study for modeling river flow time series.

3.2. *Model 5 Regression Regression Tree (M5RT)*. In decision tree (DT), each branch node indicates a choice between a number of alternatives and a decision is made in every leaf node [59]. Regression trees (RT) are applied to solve those forecasting problems having numeric response variable. They are different from the DT only in that they involve a numeric value rather than a class label combined with the leaves [60]. The M5RT method combines the features of DT and RT methods because the construction of the M5RT is similar to

the DT but, instead of the class labels, it has linear regression functions at the leaves. The M5RT is a piecewise linear method that was introduced by Quinlan [61] and has many successful applications in the field of water resources [41, 62–68] that compelled the authors to use M5RT method in this paper for river flow prediction.

The division criteria for the M5RT method are based on reducing the standard deviation of the class values that reach a node as an error measure and computing the estimated reduction in this error as a consequence of testing each attribute at that node. The standard deviation reduction (SDR) is computed by

$$\text{SDR} = \text{sd}(T) - \sum \frac{|T_i|}{|T|} \text{sd}(T_i), \quad (15)$$

where T stands for set of samples that enters the node, T_i indicates the subset of samples that have the i th output of the potential set, and sd is the standard deviation [69].

3.3. Conventional Multiple Linear Regression (CMLR). The multiple linear regression methods forecast values of a dependent variable Y based on independent variables (X_1, X_2, \dots, X_n). Two main advantages of the CMLR are that it has simple structure and it is included in lots of statistical packages [70]. In this study, after determining the independent lagged river flow values for dependent river flows of both basins, the CMLR can be constructed as follows:

$$\text{RF}_{(tn)} = b_0 + b_1 \text{RF}_{(tn-1)} + b_2 \text{RF}_{(tn-2)} + \dots + b_n \text{RF}_{(tn)}, \quad (16)$$

where $\text{RF}_{(tn)}$ is the dependent variable, $b_0 - b_n$ are the equation parameters for the linear relation, and $\text{RF}_{(tn-1)}$, $\text{RF}_{(tn)}$ are the independent lagged river flow value used to forecast river flow. However, CMLRs have some disadvantages in predicting nonlinear situations because of their linear structure [71].

4. Study Sites and Data Preprocessing

The study used the river flows data from two catchments, Astore and Shyok, on the Upper Indus Basin of Pakistan. Figure 3 shows the location map of the catchments. The geographical location of Astore Basin is approximately between longitudes $74^\circ, 24'$ and $75^\circ, 14'E$ and between latitudes $34^\circ, 45'$ and $35^\circ, 38'N$. The river covers a catchment area of about 3750 km^2 . Water and power development authority (WAPDA), Pakistan, has one flow gauging station, that is, Doyian in this area for flow record under Surface Water Hydrology Project (SWHP). The elevation of this gauging station is 1583 masl and its geographical location in the basin is $35^\circ, 33'N$ latitude and $74^\circ, 42'E$ longitude. The Shyok Basin covers drainage area of $68,458 \text{ km}^2$ with average basin elevation of 4940 m. WAPDA also installed one flow gauging station hydrometric station in this area for flow record at Yogo with an elevation of 2469 m and its geographical location in the catchment is $35^\circ, 11'N$ latitude and $76^\circ, 06'E$ longitude. The recorded monthly data of river flows of both catchments were collected through WAPDA for the duration of 1975–2006 and the total time span of this duration is 384 months.

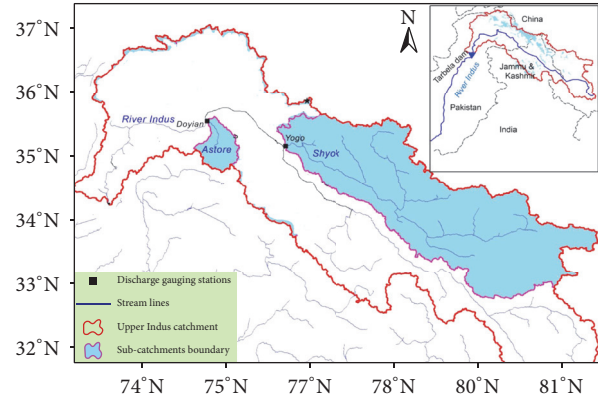


FIGURE 3: The location map of the Astore and Shyok catchments.

According to the WAPDA, the mean annual river flow 32 yr (1975–2006) flow record is $142 \text{ m}^3/\text{s}$ for Astore catchment whereas, for Shyok catchment, it is $457 \text{ m}^3/\text{s}$.

In this research, cross validation (CV) technique was applied to better see the prediction accuracy of the applied methods. In CV technique, the whole data is divided into n equal data sets (DS), then the $n-1$ DS is used to train, and the other one DS is used to test the accuracy of the method. This process is repeated n times till every DS of the data is used to test the applied method. This CV technique is preferred over k -fold cross validation techniques due to usage of every data set for testing and $n > k$ which makes it closer to the real world problem [69, 72]. Similarly, in this research, the whole river flow data was divided into four equal DS. In all the applications, the three DS were used to train and remaining one DS was adopted to test the method. This procedure was repeated four times till every DS of data was used to test the method. The monthly river flow time series statistics of Astore and Shyok catchments are reported in Table 1. Here, the DS1, DS2, DS3, and DS4 represent four equal data sets of whole data for CV analysis whereas RF_{mean} , RF_{sd} , RF_{sc} , RF_{min} , and RF_{max} represent mean, standard deviation, skewness coefficient, minimum, and maximum river flows, respectively. The recorded monthly river flows data show similarly high positive skewed distribution for Astore and Shyok catchments ($\text{RF}_{\text{sc}} = 1.51$ and 1.88). However, the range of the flow data of Shyok Basin ($36.7\text{--}2080.7 \text{ m}^3/\text{s}$) is much higher than that of the Astore Basin ($19.3\text{--}654.9 \text{ m}^3/\text{s}$). The lagged values of river flows show low persistence (e.g., Lag 1 = 0.735, Lag 2 = 0.266, and Lag 3 = -0.152). However, the lagged values of Astore Basin give a little better persistence than those of the Shyok Basin.

5. Input Combination Selection and Performance Evaluation Criteria

Input combinations (IC) selection is an important step in model development and it plays a key role in increasing the accuracy of the model. To see the correlation effect of the lagged values, the autocorrelation function is generally used to determine the number of affective lagged input values. For river flow forecasting, previous lagged river flows values

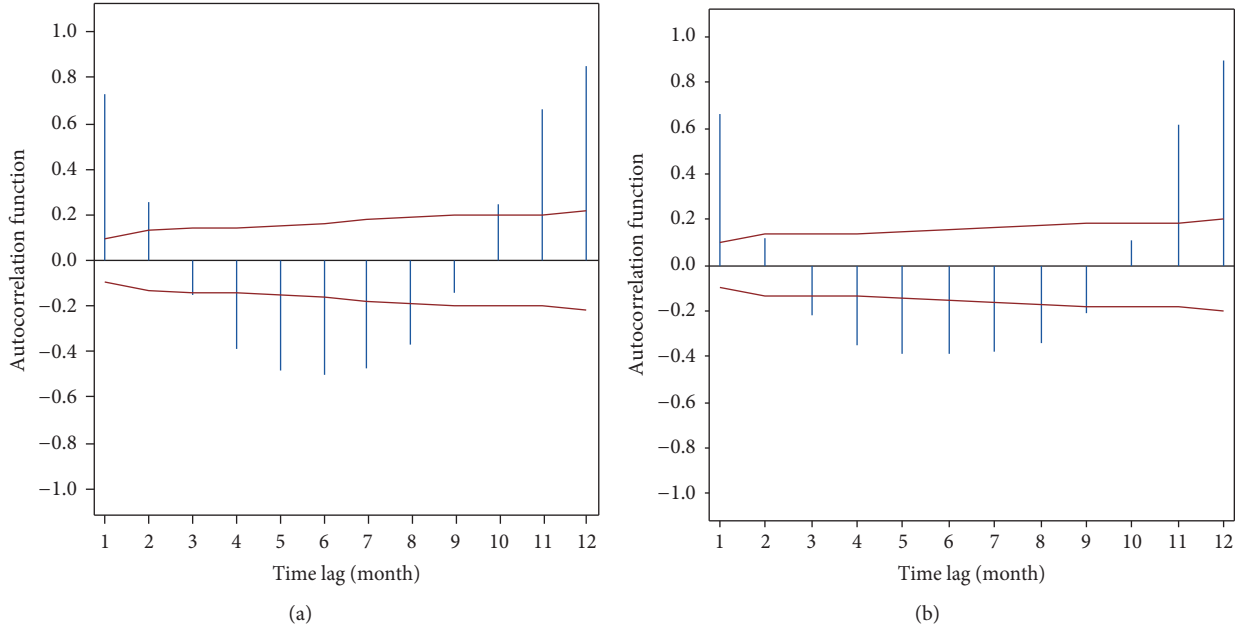


FIGURE 4: Autocorrelation function of river flow series of (a) Astore catchment and (b) Shyok catchment.

TABLE 1: The monthly statistical parameters of river flow data for Astore and Shyok catchments.

Station	Data set	RF _{min}	RF _{max}	RF _{mean}	RF _{SD}	RF _{SC}	Lag 1	Lag 2	Lag 3
Astore	DS1 (1999–2006)	26.7	612.3	138.2	1371	1.40	0.735	0.266	−0.156
	DS2 (1991–1998)	26.4	654.9	155.3	162.2	1.51	0.728	0.229	−0.178
	DS3 (1983–1990)	19.3	558.8	137.9	149.1	1.31	0.724	0.234	−0.185
	DS4 (1975–1982)	20.1	447.3	119.7	123.1	1.21	0.727	0.237	−0.202
Shyok	DS1 (1999–2006)	36.9	1855.2	406.9	527.6	1.50	0.669	0.119	−0.238
	DS2 (1991–1998)	39.0	1995.6	361.1	492.1	1.71	0.665	0.109	−0.242
	DS3 (1983–1990)	37.4	1954.8	327.6	437.0	1.72	0.660	0.118	−0.237
	DS4 (1975–1982)	36.7	2080.7	345.6	492.5	1.88	0.615	0.065	−0.233

are generally taken as IC in many researches of river flow forecasting [37, 73, 74]. In this research, the effect of lagged river flows values of both catchments was examined through autocorrelation function and was reported in Figure 4. According to the analysis, the following three IC were selected as inputs on the basis of most significant lagged river flows values for both basins, that is, IC1 = RF_{t−1}, RF_{t−2}; IC2 = RF_{t−1}, RF_{t−2}, RF_{t−11}; and IC3 = RF_{t−1}, RF_{t−2}, RF_{t−11}, RF_{t−12}.

In this paper, two error indices were selected to evaluate the performance of the models in prediction of monthly river flows including the root mean square error (RF_{RMSE}) and mean absolute error (RF_{MAE}). The similarity between the observed value and the forecasted value of river flow is measured by using the determination coefficient (RF_{DC}) index. These three indexes have been extensively applied in many problems of water resources for evaluating the model performance [19, 26, 75]. All the performance evaluation indexes can be calculated as

$$RF_{RMSE} = \sqrt{\frac{1}{N_{RF}} \sum_{t=1}^{N_{RF}} (RF_o - RF_f)^2},$$

$$RF_{MAE} = \frac{1}{N_{RF}} \sum_{t=1}^{N_{RF}} |RF_o - \overline{RF_o}|,$$

$$RF_{DC} = \left[\frac{\sum_{t=1}^{N_{RF}} (RF_o - \overline{RF_o})(RF_f - \overline{RF_f})}{\sqrt{\sum_{t=1}^{N_{RF}} (RF_o - \overline{RF_o})^2 (RF_f - \overline{RF_f})^2}} \right]^2, \quad (17)$$

where N_{RF} is the total size of observations of river flow time series, RF_o is observed river flow, RF_f is forecasted river flow, $\overline{RF_o}$ is average of river flows, and $\overline{RF_f}$ is average forecasted river flow.

6. River Flow Prediction Using Soft Computing Methods

In the first part of the paper, the performance of the proposed method hybrid LSSVR-GSA (HLGSA) was compared with other regression methods in predicting monthly river flows

TABLE 2: Comparison of HLGSA, MARS, M5RT, and CMLR methods: Astore catchment.

Indexes	Cross validation data sets	Input combinations	Methods			
			HLGSA	MARS	M5RT	CMLR
RF _{RMSE}	DS1	IC1	56.80	57.75	71.96	83.44
		IC2	44.80	53.48	56.66	59.75
		IC3	51.01	55.28	80.52	55.66
	DS2	IC1	72.58	73.83	83.09	97.06
		IC2	65.42	67.06	68.28	69.44
		IC3	66.97	68.07	101.99	68.08
	DS3	IC1	58.27	65.16	66.36	91.31
		IC2	47.38	51.59	59.57	69.20
		IC3	47.61	56.20	67.22	65.93
	DS4	IC1	46.93	57.75	61.29	74.79
		IC2	40.09	43.46	57.26	61.40
		IC3	42.57	48.14	59.20	58.08
RF _{MAE}	DS1	IC1	34.32	38.34	42.05	51.91
		IC2	27.03	32.30	33.45	26.36
		IC3	31.67	33.75	46.80	25.05
	DS2	IC1	44.24	49.12	52.21	57.67
		IC2	37.86	37.96	39.12	42.11
		IC3	38.27	38.92	53.77	40.25
	DS3	IC1	33.80	35.79	40.08	52.33
		IC2	27.14	29.23	32.47	34.33
		IC3	28.81	30.84	42.56	31.06
	DS4	IC1	32.20	37.19	38.46	46.80
		IC2	25.05	28.36	29.83	39.48
		IC3	26.36	30.22	33.74	32.70
RF _{DC}	DS1	IC1	0.831	0.822	0.732	0.693
		IC2	0.889	0.851	0.835	0.815
		IC3	0.852	0.849	0.745	0.837
	DS2	IC1	0.826	0.820	0.729	0.685
		IC2	0.850	0.846	0.822	0.811
		IC3	0.848	0.834	0.630	0.824
	DS3	IC1	0.849	0.830	0.773	0.694
		IC2	0.901	0.880	0.851	0.831
		IC3	0.898	0.861	0.810	0.844
	DS4	IC1	0.892	0.882	0.813	0.699
		IC2	0.916	0.900	0.888	0.853
		IC3	0.906	0.893	0.874	0.882

of the Astore and Shyok catchments, separately by using the three input combinations comprising antecedent river flows. CV technique was used for each applied method by dividing river flows data into four equal DS. Test statistics of HLGSA, MARS, M5RT, and CMLR methods for the Astore catchment in the test duration is compared in Table 2. It is clear from the table that all four applied methods provide different prediction results for different DS and input combinations. In case of input combinations, IC1 comprising the two consecutive previous months' river flow values provides the worst prediction results for the HLGSA, MARS, and CMLR methods. IC2 comprising the two consecutive antecedent months' river flow values including antecedent eleventh

month's flow value gives the best prediction results for the HLGSA, MARS, and M5RT methods whereas, for the CMLR method, IC3 comprising the two consecutive antecedent months' river flow values and antecedent eleventh and twelfth months' river flow value provides better performance than the other two input combinations. In the case of data sets, it is clear from the table that DS2 gives the worst forecasts results for all the regressions models including proposed HLGSA method. The reason of this is the maximum river flow value of Astore Basin's test set; DS2 ($RF_{\max} = 654.9 \text{ m}^3/\text{s}$) is higher than the corresponding extreme value of the training DS value (see Table 1). This indicates that all trained methods encounter problems in constructing extrapolation in higher

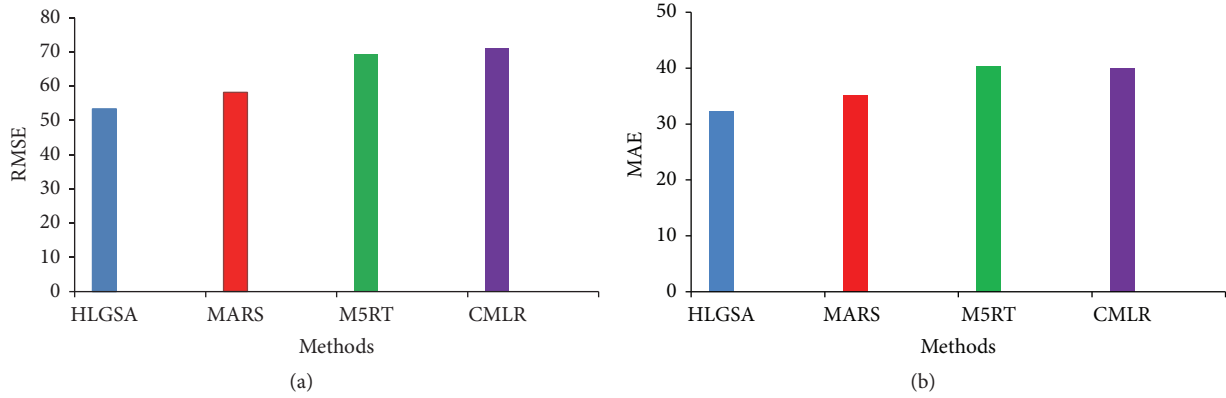


FIGURE 5: Overall errors of forecasted river flow using all normal models for Astore catchment.

value of DS2. The higher values of RF_{sd} and RF_{sc} parameters for the DS2 data set in comparison with other data sets can also be another reason for the worst results. It is evident that all methods provide good forecasts for the DS4 under all input combination scenarios. The best model structures for HLGSA, MARS, and M5RT methods were found for the DS4 and IC2. However, in case of CMLR method, the best model structure was found for the DS4 and IC3. The best RF_{RMSE} ($40.09 \text{ m}^3/\text{s}$) for the HLGSA is better as compared to MARS, M5RT, and CMLR methods (43.46 , 57.26 , and $58.08 \text{ m}^3/\text{s}$), respectively. This is also true for RF_{MAE} values where the best RF_{MAE} for the HLGSA is $25.05 \text{ m}^3/\text{s}$ compared to MARS, M5RT, and CMLR methods (28.36 , 29.83 , and $32.70 \text{ m}^3/\text{s}$), respectively. Table 2 clearly shows that the HLGSA method provides better prediction results than the other regression methods under all data sets and input combinations scenarios. MARS is ranked as the second best and performs better than the M5RT and CMLR whereas M5RT performs better than the CMLR under all data sets. In case of IC3, however, CMLR gives better forecasts than the M5RT for all data sets by having lower values of error indexes (RF_{RMSE} and RF_{MAE}) and higher value of correlation index (RF_{DC}). The best prediction results of HLGSA, MARS, and M5RT for IC2 indicate that the river flow of the two preceding months and eleventh month highly affects the current month river flow. The results of IC3 justify this statement by adding the river flow data of twelfth month that negatively affect the method performance. However, the CMLR showed positive dependence on the input combinations by showing the best prediction results for IC3.

For the sake of simplicity, the performance accuracy of all methods was evaluated by comparing the overall mean errors representing the mean error of all data sets and input combinations. Figures 5(a)-5(b) show the overall average errors statistics of all methods. Mean errors statistics of RF_{RMSE} and RF_{MAE} clearly indicate that the HLGSA method performs better by having relatively less value of error indexes than the other methods in prediction of Astore catchment's river flows. M5RT and CMLR give almost same mean errors statistics and provide the worst accuracy prediction in comparison to HLGSA and MARS due to having higher values of both errors

indexes. This indicates the nonlinearity of the investigated phenomenon because both M5RT and CMLR have linear structures. HLGSA decreases the overall mean RF_{RMSE} of the MARS, M5RT, and CMLR by 8.22%, 23.15%, and 24.49%, respectively. The observed and forecasted monthly river flows of Astore Basin by all the methods using their best model structures are reported in Figures 6(a)–6(d). The figure clearly explores that the HLGSA method is in good fit with the original river flows data in comparison to the other methods. The HLGSA gives higher value of correlation index (RF_{DC}) than the other methods. From Figure 6, it can be clearly seen that the RF_{DC} value of the HLGSA method is 0.916, which is higher than RF_{DC} of MARS, M5RT, and CMLR methods (which are 0.900, 0.888, and 0.882).

Table 3 reports the results of three performance evaluation statistical indexes for the HLGSA, MARS, M5RT, and CMLR methods in forecasting river flows of the Shyok catchment. In case of input combinations, here also IC1 gives the worst forecast results compared to the IC1 and IC3. However, in contrast to Astore application, here IC3 shows better accuracy than the IC2 for all the applied methods. In the case of data sets, it is evident from the table that DS4 gives the worst prediction results for all the methods. Similar to Astore Basin, here also the maximum river flow value of the Shyok Basin's test set, DS4 ($RF_{max} = 2080.7 \text{ m}^3/\text{s}$), is higher than that of the training value (see Table 1). Another reason of this may be the fact that the DS2 has low correlations with the preceding river flow input data in comparison to the other data sets. It is obvious from the table that all methods give good prediction results for the DS3 among all input combination scenarios. The best model structures for all four applied methods in case of Shyok catchment are found for the DS3 and IC3. Similar to the Astore Basin, here also Table 3 clearly explores that the HLGSA method outperforms the other methods from RF_{RMSE} , RF_{MAE} , and RF_{DC} viewpoints for all data sets and input combination scenarios. Here, also the MARS gives better accuracy than the M5RT and CMLR methods for all data sets and input combinations. However in contrast to Astore Basin, here M5RT provides better prediction results than the CMLR method. The best RF_{RMSE} for the proposed method is better as $109.20 \text{ m}^3/\text{s}$ compared

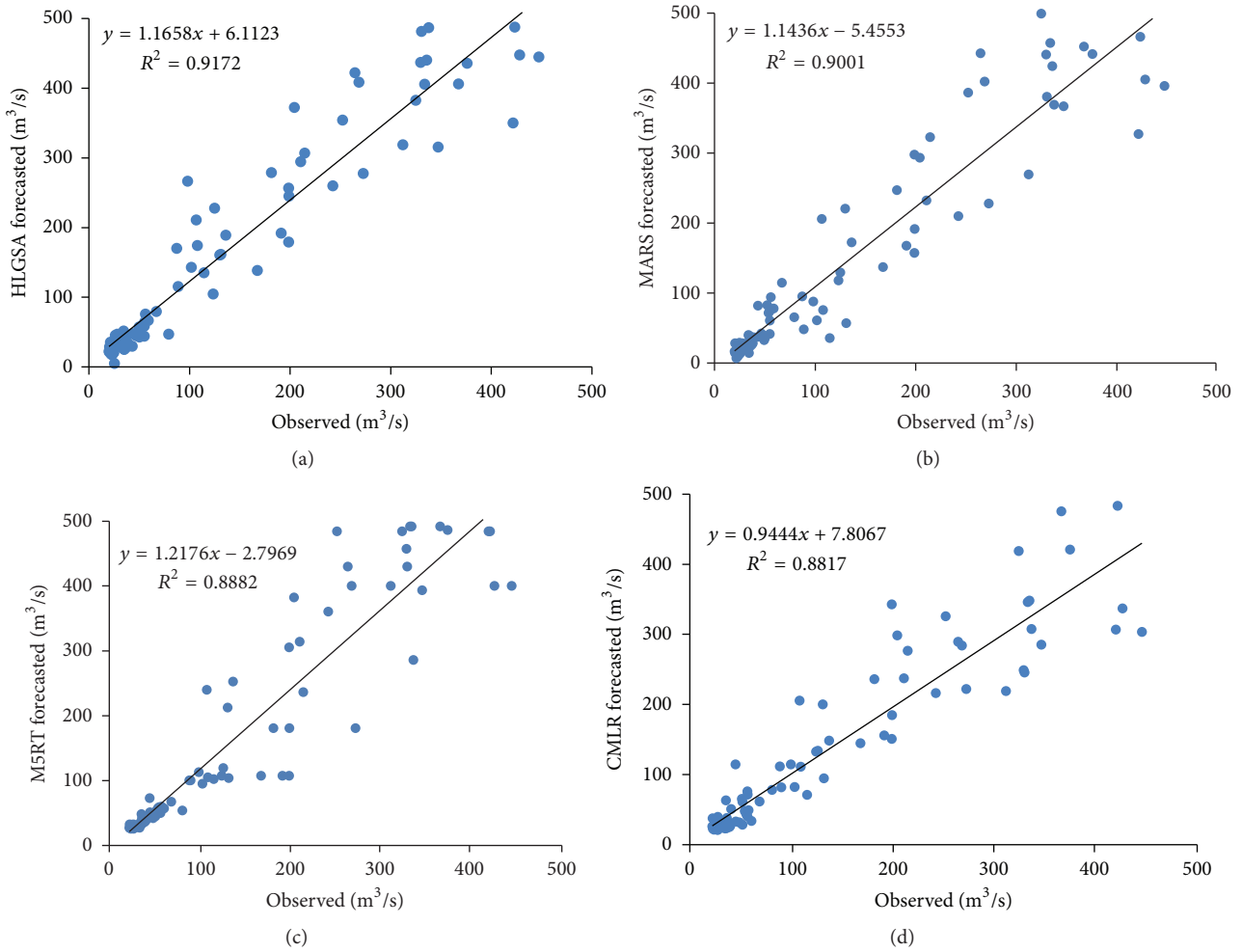


FIGURE 6: The observed and forecasted river flow by the HLGSA, MARS, M5RT, and CMLR models using DS4 data set: Astore catchment.

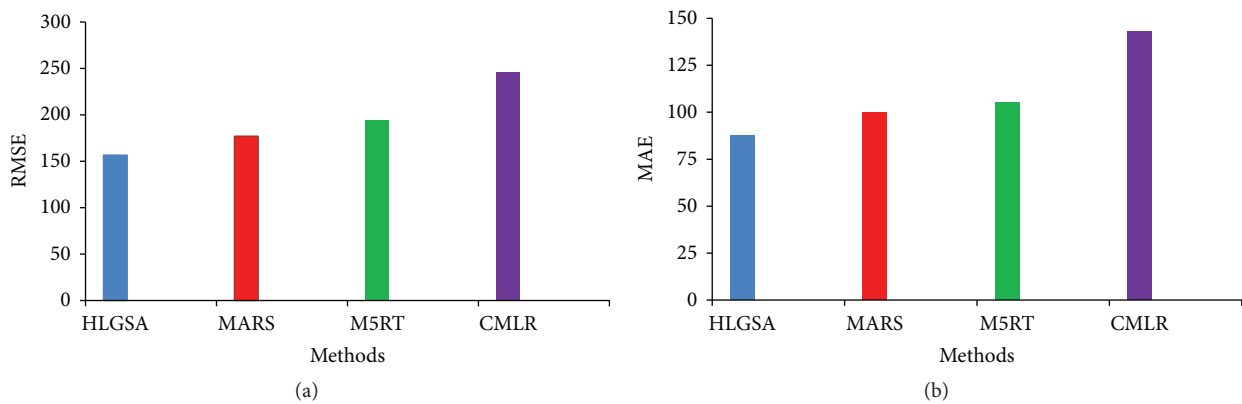


FIGURE 7: Overall errors of forecasted river flow using all normal models for Shyok catchment.

to MARS, M5RT, and CMLR methods (128.38, 141.13, and 151.21 m³/s), respectively. This is also true for RF_{MAE} values where the best RF_{MAE} for the proposed method is 55.75 m³/s compared to MARS, M5RT, and CMLR methods (69.52, 77.27, and 86.70 m³/s), respectively.

Figures 7(a)-7(b) compare the overall average errors statistics of all methods in prediction river flows of Shyok Basin. Similar to Astore, here the figure clearly shows that RF_{RMSE} and RF_{MAE} indexes of the HLGSA method give lower values in comparison to the MARS, M5RT, and CMLR

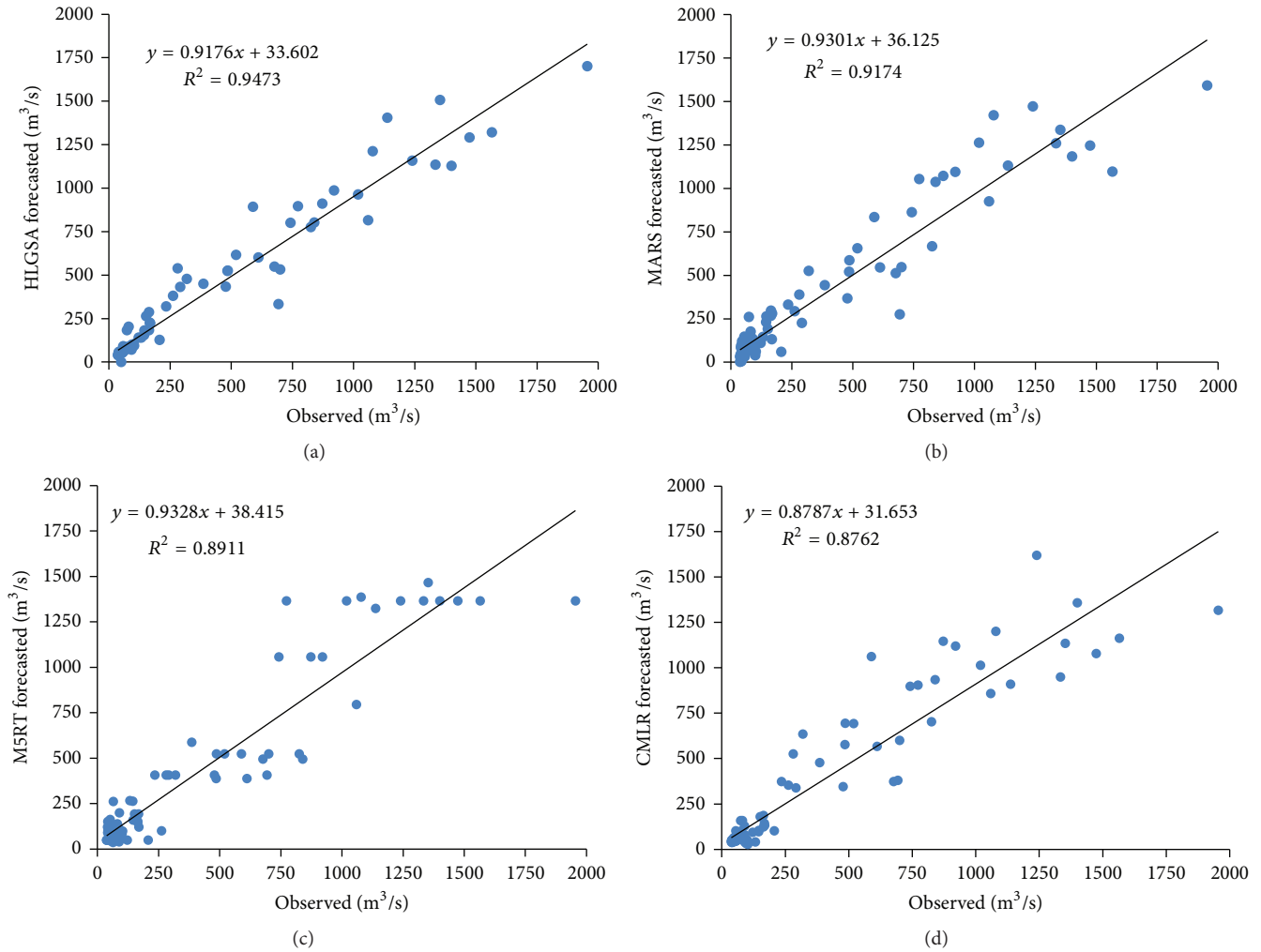


FIGURE 8: The observed and forecasted river flow by the HLGSA, MARS, M5RT, and CMLR models using DS3 data set: Shyok catchment.

methods and report that the HLGSA performs better than the other methods. However, in contrast to Astore, here the MARS and M5RT give almost same accuracy results whereas CMLR provides the worst performance in comparison to the other methods due to having higher values of both error indexes. In case of Shyok catchment, HLGSA decreases the overall mean RF_{RMSE} of the MARS, M5RT, and CMLR by 11.48%, 19.55%, and 36.19%, respectively. The scatterplots of the original and predicted monthly river flows of Shyok Basin by all the methods using their best model structures are reported in Figures 8(a)–8(d). The figure clearly shows the superior accuracy of the HLGSA method over the MARS, M5RT, and CMLR methods. It is evident from the figure that the HLGSA method has a good fit to the observed river flows data in comparison to the other methods by having the higher value of RF_{DC} . From Figure 8, it can be clearly seen that the RF_{DC} value of the HLGSA method is 0.947, which is higher than RF_{DC} of MARS, M5RT, and CMLR methods (which are 0.917, 0.891, and 0.876).

Selection of the proper kernel function (KF) is very important in obtaining highly accurate HLGSA method. The

accuracy of the HLGSA method produced by different types of KF is different. In this research, the RBKF was utilized for determining the prediction results of river flows of both basins. However, the KF has many other types, and mostly common KF types are polynomial KF, sigmoid KF, Gaussian KF, Morlet KF, Mexican hat KF, and Meyer KF. To evaluate the performance of the applied KF in this study, six different kernel functions were compared in prediction of river flows of both catchments by using the best HLGSA model structures and were reported in Table 4. Table 4 clearly proves the superiority of the RBKF over the other kernel functions due to having smaller values of both errors indexes ($\text{RF}_{\text{RMSE}} = 40.09$, $\text{RF}_{\text{MAE}} = 25.05$ for Astore and $\text{RF}_{\text{RMSE}} = 109.20$, $\text{RF}_{\text{MAE}} = 55.75$ for Shyok) and higher value of RF_{DC} ($\text{RF}_{\text{DC}} = 0.916$ for Astore and $\text{RF}_{\text{DC}} = 0.947$ for Shyok) for both basins.

7. Effect of Log Transform on the Prediction Accuracy of the Soft Computing Methods

In this section of the paper, the effect of log transform on the prediction accuracy of the applied methods was investigated

TABLE 3: Comparison of the HLGSA, MARS, M5RT, and CMLR methods: Shyok catchment.

Indexes	Cross validation data sets	Input combinations	Methods			
			HLGSA	MARS	M5RT	CMLR
RF _{RMSE}	DS1	IC1	168.82	181.90	208.64	352.90
		IC2	145.96	156.39	171.05	209.44
		IC3	123.69	152.65	155.46	173.55
	DS2	IC1	176.70	214.12	251.98	328.46
		IC2	151.87	178.95	209.81	224.69
		IC3	135.67	176.82	177.60	182.70
	DS3	IC1	163.33	178.43	200.20	300.36
		IC2	138.85	149.03	149.96	201.21
		IC3	109.20	128.38	141.13	151.21
	DS4	IC1	209.95	232.83	264.04	371.26
		IC2	185.74	193.75	218.38	257.42
		IC3	170.68	181.39	189.40	195.10
RF _{MAE}	DS1	IC1	102.17	110.63	109.72	220.14
		IC2	78.63	92.31	107.78	135.56
		IC3	67.52	87.96	100.71	94.33
	DS2	IC1	106.96	128.49	113.52	187.27
		IC2	81.68	94.76	100.12	121.25
		IC3	87.98	93.26	93.68	96.48
	DS3	IC1	90.56	105.66	105.78	181.46
		IC2	66.59	74.29	97.34	116.65
		IC3	55.75	69.52	77.27	86.70
	DS4	IC1	120.19	132.56	137.95	224.07
		IC2	104.50	107.79	112.15	147.17
		IC3	88.99	102.18	103.53	105.04
RF _{DC}	DS1	IC1	0.861	0.840	0.810	0.597
		IC2	0.904	0.878	0.841	0.786
		IC3	0.909	0.895	0.888	0.864
	DS2	IC1	0.872	0.816	0.757	0.616
		IC2	0.886	0.869	0.851	0.794
		IC3	0.905	0.879	0.865	0.846
	DS3	IC1	0.920	0.859	0.816	0.621
		IC2	0.929	0.890	0.876	0.841
		IC3	0.947	0.917	0.891	0.876
	DS4	IC1	0.811	0.781	0.719	0.524
		IC2	0.858	0.867	0.802	0.726
		IC3	0.920	0.879	0.861	0.820

TABLE 4: Comparison of HLGSA performance for different kernel functions using best data set of Astore and Shyok catchments.

Catchment	Performance indexes	Kernel functions						
		RBF	Polynomial	Sigmoid	Gaussian	Morlet	Mexican hat	Mayer
Astore	RF _{RMSE}	40.09	70.55	67.12	66.88	66.51	65.57	57.90
	RF _{MAE}	25.05	42.25	41.48	39.55	39.53	39.38	34.27
	RF _{DC}	0.916	0.878	0.892	0.889	0.887	0.901	0.903
Shyok	RF _{RMSE}	109.20	156.53	155.45	138.47	137.72	137.95	144.92
	RF _{MAE}	55.75	83.72	82.40	72.52	70.93	76.30	83.21
	RF _{DC}	0.947	0.873	0.875	0.901	0.901	0.908	0.903

TABLE 5: Comparison of the LHLGSA, LMARS, LM5RT, and LCMLR methods: Astore catchment.

Indexes	Cross validation data sets	Input combinations	Methods			
			LHLGSA	LMARS	LM5RT	LCMLR
RF _{RMSE}	DS1	IC1	54.73	62.76	70.22	123.24
		IC2	43.31	50.88	69.52	65.05
		IC3	52.54	54.65	71.86	60.76
	DS2	IC1	72.25	78.42	80.68	135.75
		IC2	56.78	59.43	74.33	65.48
		IC3	58.11	64.52	83.89	68.60
	DS3	IC1	56.65	72.06	64.28	129.02
		IC2	45.75	48.49	57.39	61.24
		IC3	48.69	53.97	64.91	54.83
	DS4	IC1	43.36	59.96	61.91	96.04
		IC2	35.33	38.56	46.18	40.11
		IC3	36.72	38.78	51.49	39.02
RF _{MAE}	DS1	IC1	33.28	43.89	40.52	60.16
		IC2	26.91	27.05	38.73	31.79
		IC3	28.51	29.98	40.32	30.29
	DS2	IC1	42.04	54.60	47.71	66.67
		IC2	29.12	30.29	41.03	32.68
		IC3	30.62	36.75	45.64	31.15
	DS3	IC1	32.83	38.91	34.88	60.72
		IC2	27.27	26.81	28.27	27.54
		IC3	29.37	28.06	32.86	27.07
	DS4	IC1	20.29	39.92	34.89	52.10
		IC2	13.08	20.88	24.80	21.42
		IC3	17.58	21.06	27.02	21.08
RF _{DC}	DS1	IC1	0.842	0.806	0.741	0.714
		IC2	0.894	0.877	0.789	0.824
		IC3	0.882	0.859	0.763	0.846
	DS2	IC1	0.822	0.785	0.760	0.695
		IC2	0.882	0.876	0.775	0.820
		IC3	0.866	0.845	0.729	0.836
	DS3	IC1	0.856	0.798	0.815	0.736
		IC2	0.905	0.890	0.817	0.836
		IC3	0.896	0.893	0.779	0.855
	DS4	IC1	0.894	0.854	0.843	0.746
		IC2	0.921	0.906	0.899	0.896
		IC3	0.914	0.896	0.888	0.901

by applying log transform on the time series data of both basins before applying these methods. Here also the CV technique was applied and log transform data was divided into four equal data sets. Table 5 reports the test results of the LHLGSA, LMARS, LM5RT, and LCMLR methods for the Astore Basin. Here, also the IC2 generally provides the best forecast of LHLGSA, LMARS, and LM5RT methods whereas IC3 generally gives better results for the LCMLR method. However, IC3 generally provides worse results for LM5RT method in comparison to the IC1 and IC2. LCMLR gives better performance than the LM5RT method in case of IC2 and IC3 scenarios due to having lower values of RF_{RMSE} and RF_{MAE} and higher values of RF_{DC}. Similar to previous Astore

application, here also the DS2 gives the worst results whereas DS4 performs the best among all data sets. The reason of the worst results of DS2 was already mentioned before. In case of Astore Basin, the best RF_{RMSE}, RF_{MAE}, and RF_{DC} values (35.33 m³/s, 13.08 m³/s, and 0.921) of the LHLGSA are better than those of the HLGSA (40.09 m³/s, 25.05 m³/s, and 0.916), respectively. This is also true for the LMARS, LM5RT, and LCMLR methods where the best RF_{RMSE}, RF_{MAE}, and RF_{DC} values of the LMARS, LM5RT, and LCMLR methods, respectively, are 38.78 m³/s, 20.88 m³/s, and 0.906, 46.18 m³/s, 24.80 m³/s, and 0.899, and 39.02 m³/s, 21.08 m³/s, and 0.901, which are better in comparison to those of the MARS, M5RT, and CMLR methods (43.46 m³/s, 28.36 m³/s, and 0.900,

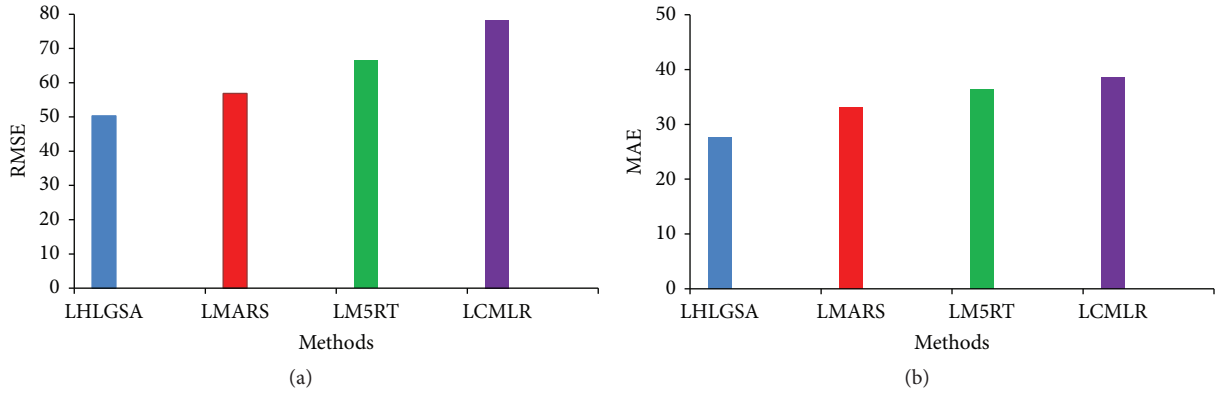


FIGURE 9: Overall errors of forecasted river flow using all log models for Astore catchment.

57.26 m³/s, 29.83 m³/s, and 0.888, and 58.08 m³/s, 32.70 m³/s, and 0.888, resp.).

The mean errors statistics of all log models is illustrated in Figures 9(a)–9(b). The figure clearly explores the performance dominance of LHLGSA over the other corresponding methods. According to Table 5, LCMLR generally performs better than LM5RT method in mostly cases of input combinations. However, on the mean errors values basis, the LCMLR performs worse than the LM5RT method due to inaccurate forecast results of ICI ($DS1_{LCMLR} = 123.24 > DS1_{LM5RT} = 70.22$, $DS2_{LCMLR} = 135.75 > DS2_{LM5RT} = 80.68$, $DS3_{LCMLR} = 129.02 > DS3_{LM5RT} = 64.28$, $DS4_{LCMLR} = 96.04 > DS4_{LM5RT} = 61.91$ according to RF_{RMSE} viewpoint) which affects the mean errors values. Figures 10(a)–10(d) illustrate the observed and forecasted river flows of the Astore Basin by using the log transform methods. The figure clearly shows that the LHLGSA method is in good fit with the original data. According to comparison of scatter plots of log and normal methods (Figures 6 and 10), it is evident that the log methods have better fits with the original data in comparison to normal methods. On the basis of fit line equation, the log methods are closer to the exact line than the normal methods (see a_0 and a_1 coefficients in Figures 6 and 10). LHLGSA decreases the overall mean RF_{RMSE} of the LMARS, LM5RT, and LCMLR by 11.46%, 24.16%, and 35.65%, respectively.

Test statistics of the LHLGSA, LMARS, LM5RT, and LCMLR methods for the Shyok Basin are reported in Table 6. Here, in contrast to previous Shyok application, IC2 gives better forecast results for the LHLGSA and M5RT methods whereas IC3 performs better for the LMARS and LCMLR methods. However, IC3 performs slightly better for the LM5RT, whereas IC2 performs better for the LMARS in the case of DS3 data set. Similar to the previous Shyok application, here also DS3 performs the best whereas DS4 performs worst among all data sets. To check the best log method among all log methods, the mean errors indexes (RF_{RMSE} and RF_{MAE}) of all log models are plotted in Figures 11(a)–11(b) in the form of bar graphs. The bar graphs of mean error indexes clearly show that the LHLGSA performs better than the other log methods due to having lower values of error indexes. Scatter plots of the observed and predicted river flows of Shyok catchment for all log methods by using

their best model structures are shown in Figures 12(a)–12(d). The figure clearly shows that the LHLGSA method gives less scatter estimates with a higher value of RF_{DC} in comparison to the LMARS, LM5RT, and LCMLR methods. The figure also reveals that all log methods are closer to exact line than the normal methods (compare Figures 8 and 12). On the basis of comparison, it is obvious that the normal methods give more scattered forecasts than the log methods. In case of Shyok catchment, LHLGSA decreases the overall mean RF_{RMSE} of the LMARS, LM5RT, and LCMLR by 14.49%, 21.73%, and 43.84%, respectively. The log transform does not equally affect all input combinations and data sets but it can be observed that this effect is more prominent when more inputs are used in the case of CMLR methods (see Tables 5 and 6). In contrast to Astore Basin application, for the Shyok Basin, the best RF_{RMSE} and RF_{MAE} values for the HLGSA (109.20 m³/s and 55.75 m³/s) are a little better compared to LHLGSA (109.41 m³/s and 59.01 m³/s), respectively. However in case of similarity index (RF_{DC}), the best RF_{DC} value for LHLGSA (0.959) is better than that of the HLGSA (0.947) method. In case of LMARS, LM5RT, and CMLR method, the best RF_{RMSE} , RF_{MAE} , and RF_{DC} values are better in comparison with MARS, M5RT, and CMLR methods, similar to Astore Basin application.

To evaluate the overall effect of log transform function on all applied methods, the overall mean error indexes (RF_{RMSE} and RF_{MAE}) for both basins with and without log methods are compared in Figures 13(a), 13(b), 14(a), and 14(b). Both graphs clearly explore that the log methods give better accuracy than the normal methods except the LCMLR which gives worse accuracy compared to the CMLR with respect to RF_{RMSE} error index. However, the LCMLR gives better accuracy than the CMLR from the RF_{MAE} index viewpoint for both basins. The reason of the LCMLR's bad performance is due to its worse results in case of ICI that affects the overall RF_{RMSE} value. The bar graphs also prove that the proposed HLGSA method shows better accuracy than the other models in both cases of normal and logarithm transformed time series data. According to the bar graphs, the LHLGSA reduces the overall RF_{RMSE} of the HLGSA by 5.66% and 4.87% for the Astore and Shyok Basins, respectively. The LMARS reduces the overall RF_{RMSE} of the MARS by 2.20% and 1.52% for the Astore and

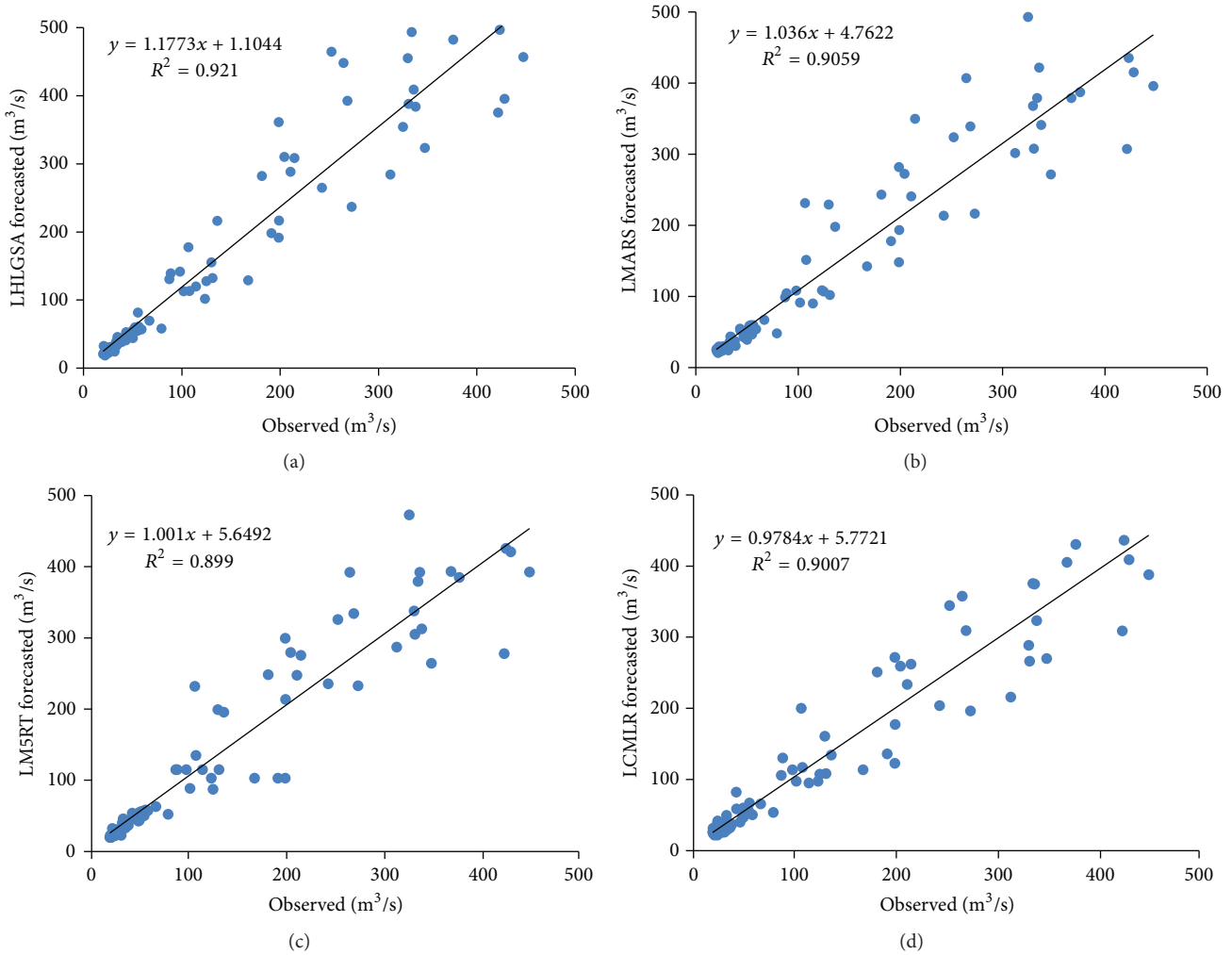


FIGURE 10: The observed and forecasted river flow by the LHLGSA, LMARS, LM5RT, and LCMLR models using DS4 data set: Astore catchment.

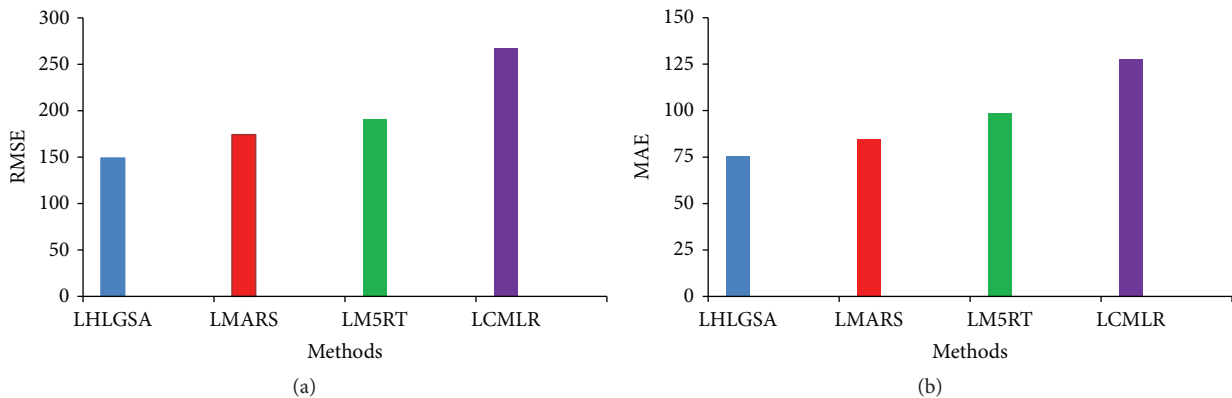


FIGURE 11: Overall errors of forecasted river flow using all log models for Shyok catchment.

Shyok Basins, respectively. The LM5RT reduced the overall RF_{RMSE} of the M5RT by 4.41% and 2.22% for the Astore and Shyok Basins, respectively. However, in the case of multiple linear regression method, CMLR reduced the overall RF_{RMSE}

of the LCMLR by 9.05% and 7.45% for the Astore and Shyok Basins, respectively, while, in the case of RF_{MAE} error index, the LCMLR reduced the overall RF_{MAE} of the CMLR by 3.60% and 10.80%, respectively.

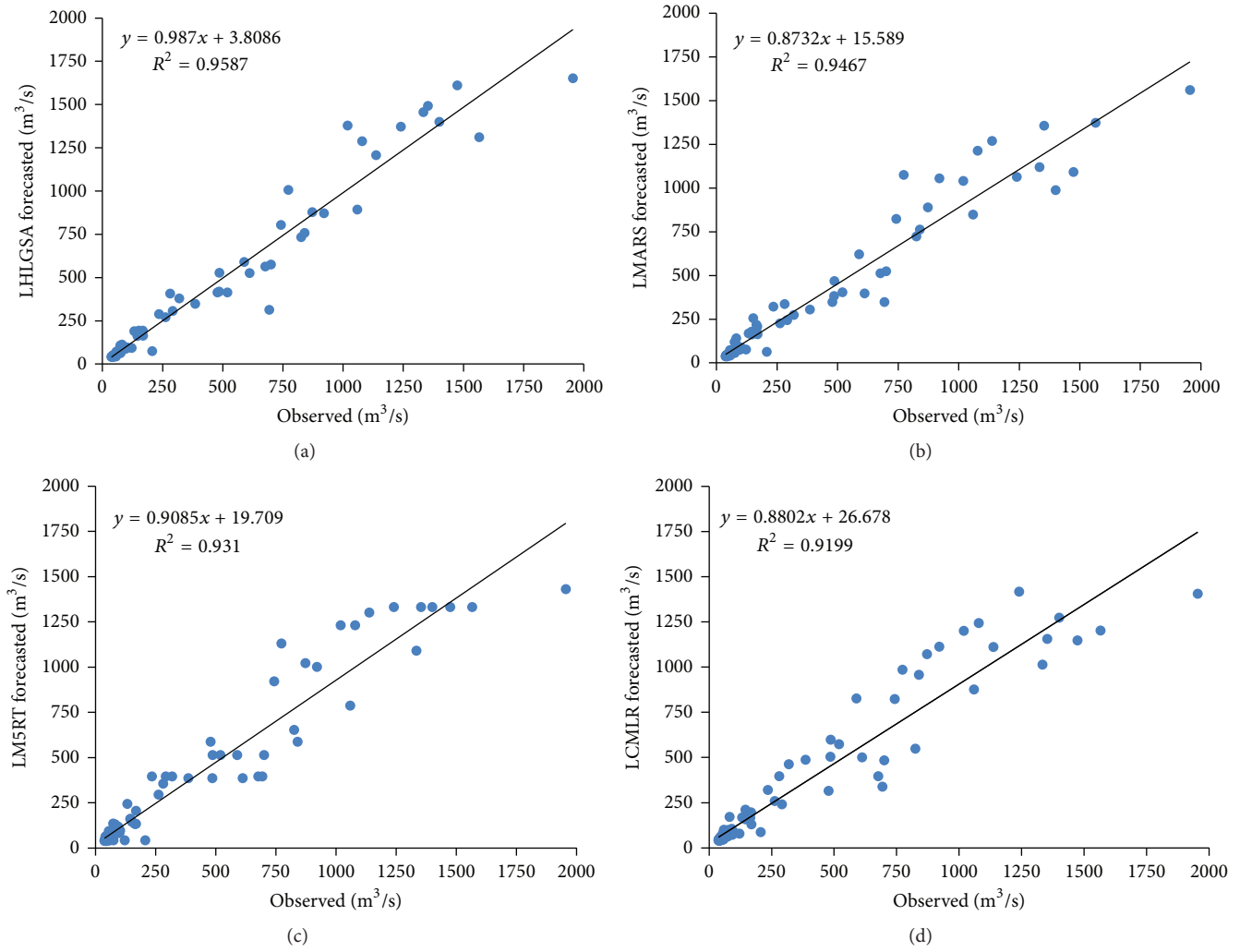


FIGURE 12: The observed and forecasted river flow by the LHLGSA, LMARS, LM5RT, and LCMLR models using DS3 data set: Shyok catchment.

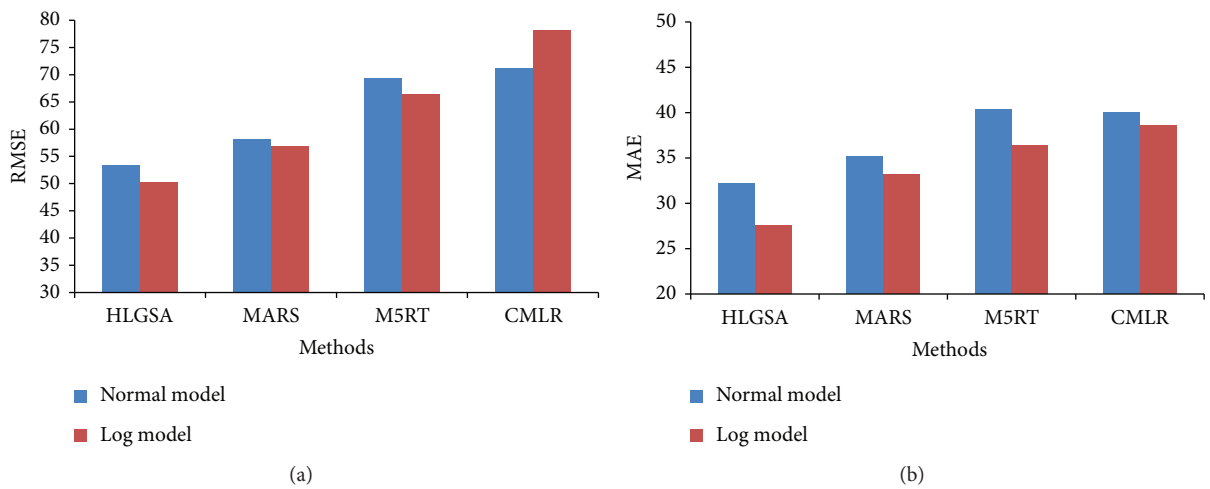


FIGURE 13: Comparison of overall errors of all log and normal models for Astore catchment.

TABLE 6: Comparison of the LHLGSA, LMARS, LM5RT, and LCMLR methods: Shyok catchment.

Indexes	Cross validation data sets	Input combinations	Methods			
			LHLGSA	LMARS	LM5RT	LCMLR
RF _{RMSE}	DS1	IC1	165.03	224.25	240.12	395.55
		IC2	142.12	160.35	163.1	195.03
		IC3	153.55	157.09	171.48	167.49
	DS2	IC1	154.19	210.33	224.47	470.32
		IC2	133.53	172.82	183.1	201.85
		IC3	134.43	155.31	174.73	171.76
	DS3	IC1	146.79	186.3	207.49	370.18
		IC2	109.41	120.66	141.03	162.35
		IC3	117.29	122.36	124.84	138.19
	DS4	IC1	207.08	243.15	259.96	512.25
		IC2	161.58	171.43	185.77	208.59
		IC3	163.91	168.28	209.68	191.93
RF _{MAE}	DS1	IC1	77.85	101.03	131.89	212.74
		IC2	65.08	73.75	78.02	85.02
		IC3	66.44	66.09	102.13	66.82
	DS2	IC1	83.19	99.06	111.59	230.22
		IC2	78.05	82.59	94.03	88.69
		IC3	78.8	80.38	92.82	86.81
	DS3	IC1	75.99	90.21	101.15	190.72
		IC2	59.01	64.28	72.03	76.79
		IC3	65.47	69.9	71.92	74.58
	DS4	IC1	102.66	124.59	138.56	240.1
		IC2	76.67	82.41	84.61	93.64
		IC3	76.89	80.77	105.49	84.59
RF _{DC}	DS1	IC1	0.845	0.721	0.738	0.677
		IC2	0.896	0.884	0.879	0.869
		IC3	0.89	0.888	0.842	0.873
	DS2	IC1	0.896	0.819	0.818	0.694
		IC2	0.91	0.898	0.861	0.856
		IC3	0.904	0.901	0.884	0.896
	DS3	IC1	0.921	0.782	0.899	0.721
		IC2	0.959	0.947	0.923	0.911
		IC3	0.953	0.938	0.930	0.920
	DS4	IC1	0.823	0.752	0.736	0.625
		IC2	0.89	0.887	0.88	0.832
		IC3	0.873	0.895	0.782	0.855

8. Comparison of LHLGSA and HLGSA Methods in Estimating River Flows Using Nearby River Flows Data

In this last section of the research, the performance of the HLGSA and LHLGSA is evaluated in river flow estimation of a basin using flow data from a nearby basin. The river flows estimation using nearby basin's flow data is a vital issue because Pakistan is a developing country and many basins have long duration of missing flows data due to financial problems in the maintenance of the hydraulic gauging stations at higher altitudes. Since river flows play a key role in planning and designing of hydropower projects and for the

flood mitigation, it is necessary to find a suitable way to fill these missing flows data. In this paper, the river flows data of the Astore Basin is used to estimate the flow data of the Shyok Basin. In this application, also the CV technique is applied to better see the accuracy of both methods in estimating river flows. Table 7 shows the RF_{RMSE}, RF_{MAE}, and RF_{DC} values of both methods in estimating monthly river flows of Shyok Basin. The best input combination for the LHLGSA and HLGSA methods is IC3 while IC2 generally gives worse results for both methods. DS4 gives better accuracy than the other DS whereas DS1 provides the worst estimates for both methods. Table 7 clearly shows the superiority of the LHLGSA over HLGSA in case of accuracy under all data sets

TABLE 7: Comparison of the LHLGSA and HLGSA methods in estimating river flow of the Shyok catchment by using the data of Astore catchment.

Indexes	Cross validation data sets	Input combinations	Methods	
			LHLGSA	HLGSA
RF_{RMSE}	DS1	IC1	267.33	254.43
		IC2	237.07	234.53
		IC3	232.12	233.22
	DS2	IC1	214.19	218.81
		IC2	192.56	199.51
		IC3	188.45	195.61
	DS3	IC1	217.30	241.62
		IC2	193.67	214.29
		IC3	188.67	204.49
	DS4	IC1	192.60	216.17
		IC2	182.55	189.68
		IC3	177.62	189.03
RF_{MAE}	DS1	IC1	141.17	135.81
		IC2	130.29	128.41
		IC3	115.68	122.55
	DS2	IC1	115.05	119.37
		IC2	105.80	107.56
		IC3	96.79	99.75
	DS3	IC1	104.98	116.19
		IC2	91.91	107.76
		IC3	88.02	102.42
	DS4	IC1	95.95	112.59
		IC2	85.95	105.30
		IC3	80.21	99.66
RF_{DC}	DS1	IC1	0.775	0.786
		IC2	0.847	0.857
		IC3	0.866	0.862
	DS2	IC1	0.772	0.766
		IC2	0.802	0.773
		IC3	0.814	0.776
	DS3	IC1	0.808	0.797
		IC2	0.816	0.801
		IC3	0.833	0.809
	DS4	IC1	0.846	0.814
		IC2	0.878	0.870
		IC3	0.882	0.871

and input combinations. It can also be seen that LHLGSA gave lower values of error indexes in case of the best data set with the best input combination (RF_{RMSE} of LHLGSA = 177.62 < RF_{RMSE} of HLGSA = 189.03 and RF_{MAE} of LHLGSA = 80.21 < RF_{MAE} of HLGSA = 99.66). The original and estimated river flows by the LHLGSA and HLGSA methods using their best model structures (DS4 with IC3) are illustrated in Figure 15. The figure clearly shows that the LHLGSA has higher value of RF_{DC} (0.882 > 0.871) representing better estimation than the HLGSA method. Figure 16 compares the mean error indexes of both methods in estimation river flows of Shyok Basin. It can be seen from the figure that the LHLGSA shows better accuracy than the HLGSA by

having lower values of errors indexes. LHLGSA decreases the overall mean RF_{RMSE} of the HLGSA method by 4.10% in the estimation of Shyok Basin flows using the river flows of Astore Basin.

9. Discussion

On the basis of above results, the key findings can be summarized as follows.

- (1) The Astore Basin reported lower values of RF_{RMSE} and RF_{MAE} in comparison with the Shyok Basin. This is due to the mean river flow of the basins; Astore Basin

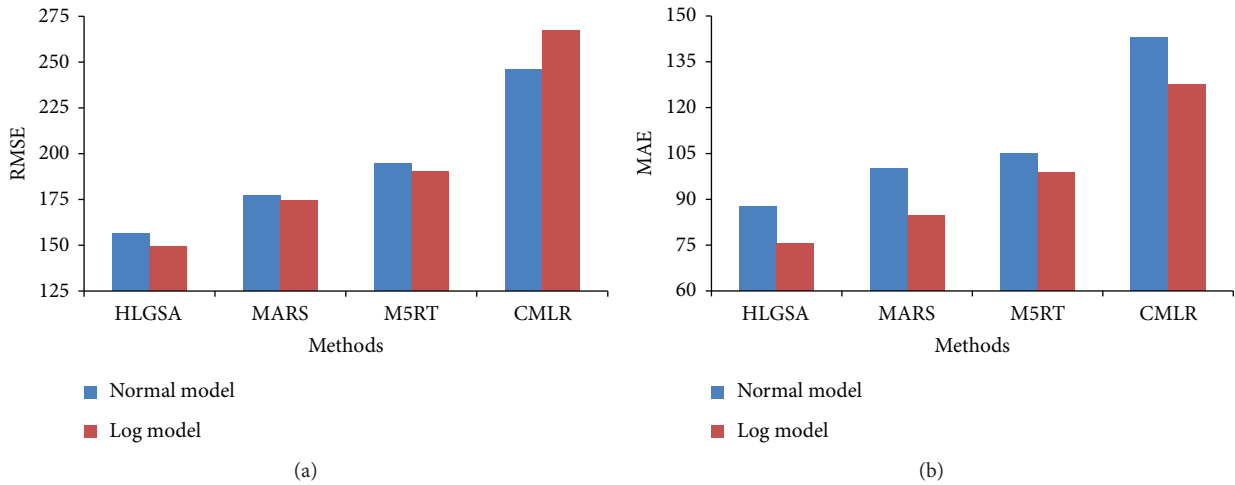


FIGURE 14: Comparison of overall errors of all log and normal models for Shyok catchment.

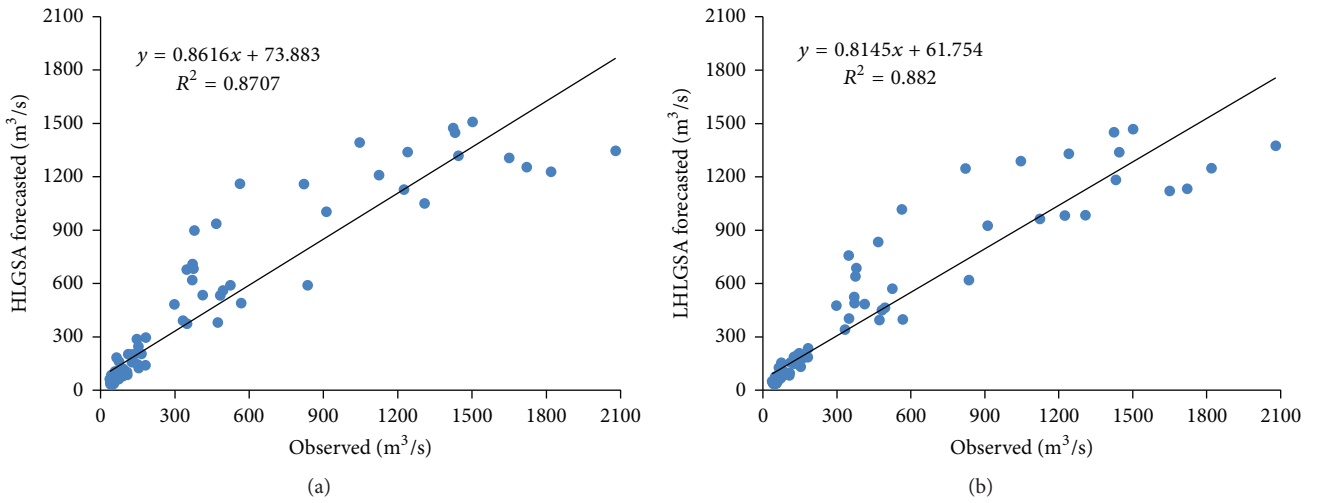


FIGURE 15: The river flow estimates of the Shyok Basin by LHLGSA and HLGSA models using the data of Astore Basin: DS4 data set.

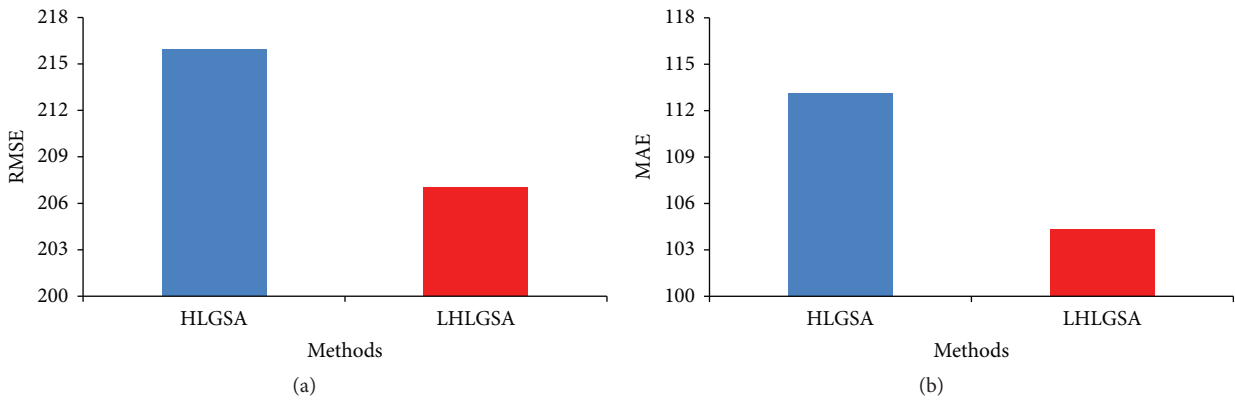


FIGURE 16: Comparison of overall errors of LHLGSA and HLGSA models for estimating river flow of Shyok catchment using Astore catchment's river flow.

is characterized by mean river flow of $142 \text{ m}^3/\text{s}$ while the mean river flow of Shyok Basin is $457 \text{ m}^3/\text{s}$.

- (2) It was also observed that the prediction accuracy of all methods including proposed method was mostly improved with increasing in input numbers which indicated that all input combinations have positive effects on predicting river flow especially in case of CMLR method.
- (3) It was also found that the higher value of testing data set's maximum river flow in comparison with other training data set's maximum river flow values caused the extrapolation difficulties and produced worst prediction results for that data set.
- (4) Overall, the HLGSA and LHLGSA methods outperformed the MARS, M5RT, CMLR and LMARS, LM5RT, and LCMLR methods, respectively. Moreover, the comparison between Figures 13 and 14 indicates that the prediction results with log transform function are better on the mean basis than the prediction results without log transform function using all regression methods including the proposed method which means the log transform function is suitable for denoising the river flow data.
- (5) In literature, many studies reported that MARS performed better or equally in comparison with the LSSVR methods [43, 44, 76–78]. However, in this study, the hybrid LSSVR method with gravitational search algorithm performed better than MARS method. The main reason behind this may be the LSSVR's strong generalization capability and nonlinear fitting ability and the second reason may be the selection of optimal LSSVR control parameters (γ and σ^2) through GSA that directly affects and improves the accuracy of the method. The powerful global search ability of GSA helps to find the optimal and suitable values for the LSSVR control parameters in a shorter time in comparison to other algorithms. We can conclude that, in application of LSSVM, the control parameters should be adequately optimized by using global optimization techniques. This will decrease the uncertainties in obtaining optimal LSSVM models.
- (6) In general, the benchmark regression methods can be ranked according to their prediction accuracy as MARS, M5RT, and CMLR. The reason behind the worst results of M5RT and CMLR methods can be the linear structure of these models.
- (7) The RF_{RMSE} , RF_{MAE} , and RF_{DC} results validate that the HLGSA and LHLGSA methods can be effectively applied for the prediction and estimation of river flow.

10. Conclusions

In the current study, river flow data of Astore and Shyok rivers was used to determine the forecasting capability of HLGSA, MARS, M5RT, and CMLR methods by using the

antecedent river flow values as inputs. Two error indexes (RF_{RMSE} and RF_{MAE}) and one similarity index (RF_{DC}) were used for comparing the prediction accuracy of these methods. CV technique was used in all the applications to better see the prediction accuracy of the data sets. In the first part of the study, among four regression methods, HLGSA provided better results than the other methods in prediction of the monthly river flow data of both catchments. HLGSA improved the prediction accuracy of the MARS, M5RT, and CMLR by 8.22%, 23.15%, and 24.49% in Astore Basin, respectively, whereas, for Shyok Basin, HLGSA improved the prediction accuracy of the MARS, M5RT, and CMLR methods by 11.48%, 19.55%, and 36.19%, respectively. In the current study, radial basis kernel function was selected for HLGSA model due to its better prediction accuracy. In the second part of the study, the effect of logarithm transform function on prediction performance of all regression methods was also investigated. Results reported that after applying logarithm function on river flow time series data, all the regression methods provided better prediction accuracy for both basins. Prediction results also exposed that the HLGSA method outperformed the other methods. LHLGSA decreased the overall mean RF_{RMSE} of the LMARS, LM5RT, and LCMLR by 11.46%, 24.16%, and 35.65%, respectively, for the Astore Basin, whereas, for Shyok Basin, LHLGSA decreases the overall mean RF_{RMSE} of the LMARS, LM5RT, and LCMLR by 14.49%, 21.73%, and 43.84%, respectively. On the comparison of log transformed methods and normal methods, the LHLGSA reduced the overall RF_{RMSE} of the HLGSA by 5.66% and 4.87% for the Astore and Shyok Basins, respectively. LMARS reduced the overall RF_{RMSE} of the MARS by 2.20% and 1.52% for the Astore and Shyok Basins, respectively. The LM5RT reduced the overall RF_{RMSE} of the M5RT by 4.41% and 2.22% for the Astore and Shyok Basins, respectively. The third part of the study evaluated the prediction performance of the HLGSA and LHLGSA methods in river flow estimation using river flow data of the nearby basin. The test results revealed that the LHLGSA performed better than the HLGSA in estimating river flows of Shyok Basin by using Astore Basin data.

In this study we forecasted river flows with only previous river flow values as inputs. These prediction accuracies of the applied methods could be improved if more input variables were available. Further studies may be conducted by including more inputs such as rainfall, snowpack, and temperature or/and building prediction models using more advanced modeling methods at these study sites. The proposed data driven methods may be applied for other regions with similar or different climates. In this case, however, the methods should be properly calibrated by using high number of river flow data.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by National Natural Science Foundation of China (no. 51379080 and no. 41571514) and Hubei

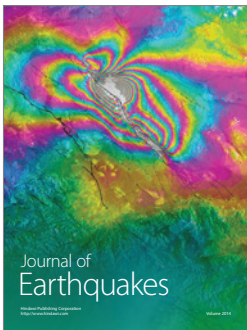
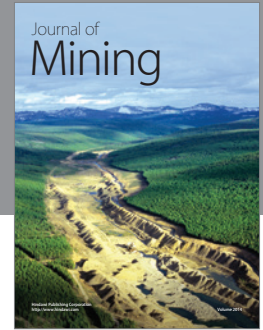
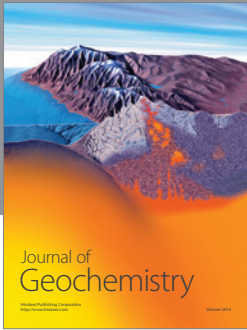
Provincial Collaborative Innovation Center for New Energy Microgrid, China Three Gorges University. The authors thank the staff of WAPDA for providing river flows data of both basins.

References

- [1] Q. Zhang, B.-D. Wang, B. He, Y. Peng, and M.-L. Ren, "Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting," *Water Resources Management*, vol. 25, no. 11, pp. 2683–2703, 2011.
- [2] J. Xu, Y. Chen, W. Li, Q. Nie, C. Song, and C. Wei, "Integrating wavelet analysis and BPANN to simulate the annual runoff with regional climate change: a case study of Yarkand River, Northwest China," *Water Resources Management*, vol. 28, no. 9, pp. 2523–2537, 2014.
- [3] A. R. Ghumman, Y. M. Ghazaw, A. R. Sohail, and K. Watanabe, "Runoff forecasting by artificial neural network and conventional model," *Alexandria Engineering Journal*, vol. 50, no. 4, pp. 345–350, 2011.
- [4] M. Tayyab, J. Zhou, X. Zeng, L. Chen, and L. Ye, "Optimal application of conceptual rainfall-runoff hydrological models in the Jinshajiang River basin," *Proceedings of the International Association of Hydrological Sciences*, vol. 368, pp. 227–232, 2015.
- [5] R. C. Ahlert and B. M. Mehta, "Stochastic analyses and transfer functions for flows of the upper Delaware River," *Ecological Modelling*, vol. 14, no. 1-2, pp. 59–78, 1981.
- [6] S. Ahmad, I. H. Khan, and B. P. Parida, "Performance of stochastic approaches for forecasting river water quality," *Water Research*, vol. 35, no. 18, pp. 4261–4266, 2001.
- [7] A. Kurunç, K. Yürekli, and O. Çevik, "Performance of two stochastic approaches for forecasting water quality and streamflow data from Yeşilirmak River, Turkey," *Environmental Modelling and Software*, vol. 20, no. 9, pp. 1195–1200, 2005.
- [8] M. Mirzavand and R. Ghazavi, "A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods," *Water Resources Management*, vol. 29, no. 4, pp. 1315–1328, 2015.
- [9] A. K. Mishra and V. R. Desai, "Drought forecasting using stochastic models," *Stochastic Environmental Research and Risk Assessment*, vol. 19, no. 5, pp. 326–339, 2005.
- [10] R. Modarres, "Streamflow drought time series forecasting," *Stochastic Environmental Research and Risk Assessment*, vol. 21, no. 3, pp. 223–233, 2007.
- [11] B. W. Otok and Suhartono, "Development of rainfall forecasting model in Indonesia by using ASTAR, transfer function, and ARIMA methods," *European Journal of Scientific Research*, vol. 38, no. 3, pp. 386–395, 2009.
- [12] A. Psilovikos and M. Elhag, "Forecasting of remotely sensed daily evapotranspiration data over Nile Delta region, Egypt," *Water Resources Management*, vol. 27, no. 12, pp. 4115–4130, 2013.
- [13] A. T. Rabenja, A. Ratiarison, and J. M. Rabeharisoa, "Forecasting of the rainfall and the discharge of the Namorona River in Vohiparara and FFT analyses of these data," in *Proceedings of the 4th International Madagascar Conference in High-Energy Physics (HEP-MAD '09)*, pp. 1–12, August 2009.
- [14] M. Valipour, "Long-term runoff study using SARIMA and ARIMA models in the United States," *Meteorological Applications*, vol. 22, no. 3, pp. 592–598, 2015.
- [15] S. Yang, S. Liu, X. Li, Y. Zhong, X. He, and C. Wu, "The short-term forecasting of evaporation duct height (EDH) based on ARIMA model," *Multimedia Tools and Applications*, pp. 1–14, 2016.
- [16] A. Altunkaynak, "Forecasting surface water level fluctuations of lake van by artificial neural networks," *Water Resources Management*, vol. 21, no. 2, pp. 399–408, 2007.
- [17] D. Ömer Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 586–594, 2010.
- [18] W. Huang, B. Xu, and A. Chan-Hilton, "Forecasting flows in Apalachicola River using neural networks," *Hydrological Processes*, vol. 18, no. 13, pp. 2545–2564, 2004.
- [19] G. Landeras, A. Ortiz-Barredo, and J. J. López, "Forecasting weekly evapotranspiration with ARIMA and artificial neural network models," *Journal of Irrigation and Drainage Engineering*, vol. 135, no. 3, pp. 323–334, 2009.
- [20] A. K. Mishra, V. R. Desai, and V. P. Singh, "Drought forecasting using a hybrid stochastic and neural network model," *Journal of Hydrologic Engineering*, vol. 12, no. 6, pp. 626–638, 2007.
- [21] C. L. Wu and K.-W. Chau, "Data-driven models for monthly streamflow time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 8, pp. 1350–1367, 2010.
- [22] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: a machine learning approach," *Advances in Water Resources*, vol. 33, no. 1, pp. 69–80, 2010.
- [23] M. Behzad, K. Asghari, M. Eazi, and M. Palhang, "Generalization performance of support vector machines and neural networks in runoff modeling," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7624–7629, 2009.
- [24] Z. He, X. Wen, H. Liu, and J. Du, "A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region," *Journal of Hydrology*, vol. 509, pp. 379–386, 2014.
- [25] S. K. Jain, "Modeling river stage-discharge-sediment rating relation using support vector regression," *Hydrology Research*, vol. 43, no. 6, pp. 851–861, 2012.
- [26] A. M. Kalteh, "Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform," *Computers & Geosciences*, vol. 54, pp. 1–8, 2013.
- [27] J.-Y. Lin, C.-T. Cheng, and K.-W. Chau, "Using support vector machines for long-term discharge prediction," *Hydrological Sciences Journal*, vol. 51, no. 4, pp. 599–612, 2006.
- [28] H. Yoon, S.-C. Jun, Y. Hyun, G.-O. Bae, and K.-K. Lee, "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer," *Journal of Hydrology*, vol. 396, no. 1-2, pp. 128–138, 2011.
- [29] P. P. Bhagwat and R. Maity, "Hydroclimatic streamflow prediction using least square-support vector regression," *ISH Journal of Hydraulic Engineering*, vol. 19, no. 3, pp. 320–328, 2013.
- [30] M. K. Goyal, B. Bharti, J. Quilty, J. Adamowski, and A. Pandey, "Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5267–5276, 2014.
- [31] S. H. Hwang, D. H. Ham, and J. H. Kim, "Forecasting performance of LS-SVM for nonlinear hydrological time series," *KSCIE Journal of Civil Engineering*, vol. 16, no. 5, pp. 870–882, 2012.
- [32] O. Kisi, "Modeling discharge-suspended sediment relationship using least square support vector machine," *Journal of Hydrology*, vol. 456–457, pp. 110–120, 2012.

- [33] O. Kisi, "Least squares support vector machine for modeling daily reference evapotranspiration," *Irrigation Science*, vol. 31, no. 4, pp. 611–619, 2013.
- [34] O. Kisi, "Streamflow forecasting and estimation using least square support vector regression and adaptive Neuro-Fuzzy embedded fuzzy c-means clustering," *Water Resources Management*, vol. 29, no. 14, pp. 5109–5127, 2015.
- [35] A. Mellit, A. M. Pavan, and M. Benganem, "Least squares support vector machine for short-term prediction of meteorological time series," *Theoretical and Applied Climatology*, vol. 111, no. 1-2, pp. 297–307, 2013.
- [36] U. Okkan and Z. A. Serbes, "Rainfall-runoff modeling using least squares support vector machines," *Environmetrics*, vol. 23, no. 6, pp. 549–564, 2012.
- [37] A. Shabri and Suhartono, "Streamflow forecasting using least-squares support vector machines," *Hydrological Sciences Journal*, vol. 57, no. 7, pp. 1275–1293, 2012.
- [38] J. Wu, M. Liu, and L. Jin, "Least square support vector machine ensemble for daily rainfall forecasting based on linear and nonlinear regression," in *Advances in Neural Network Research and Applications*, pp. 55–64, Springer, Berlin, Germany, 2010.
- [39] J. Adamowski, H. F. Chan, S. O. Prasher, and V. N. Sharda, "Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data," *Journal of Hydroinformatics*, vol. 14, no. 3, pp. 731–744, 2012.
- [40] P. Coulibaly and C. K. Baldwin, "Nonstationary hydrological time series forecasting using nonlinear dynamic methods," *Journal of Hydrology*, vol. 307, no. 1-4, pp. 164–174, 2005.
- [41] R. C. Deo, O. Kisi, and V. P. Singh, "Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model," *Atmospheric Research*, vol. 184, pp. 149–175, 2017.
- [42] R. C. Deo, P. Samui, and D. Kim, "Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models," *Stochastic Environmental Research and Risk Assessment*, vol. 30, no. 6, pp. 1769–1784, 2016.
- [43] O. Kisi, "Modeling reference evapotranspiration using three different heuristic regression approaches," *Agricultural Water Management*, vol. 169, pp. 162–172, 2016.
- [44] O. Kisi and K. S. Parmar, "Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution," *Journal of Hydrology*, vol. 534, pp. 104–112, 2016.
- [45] V. N. Sharda, S. O. Prasher, R. M. Patel, P. R. Ojasvi, and C. Prakash, "Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data," *Hydrological Sciences Journal*, vol. 53, no. 6, pp. 1165–1175, 2008.
- [46] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [47] X. Guo and X. Ma, "Mine water discharge prediction based on least squares support vector machines," *Mining Science and Technology*, vol. 20, no. 5, pp. 738–742, 2010.
- [48] M. Arabloo, H. Ziaee, M. Lee, and A. Bahadori, "Prediction of the properties of brines using least squares support vector machine (LS-SVM) computational strategy," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 50, pp. 123–130, 2015.
- [49] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparse approximation," *Neurocomputing*, vol. 48, no. 1–4, pp. 85–105, 2002.
- [50] X. Yuan, C. Chen, Y. Yuan, Y. Huang, and Q. Tan, "Short-term wind power prediction based on LSSVM-GSA model," *Energy Conversion and Management*, vol. 101, pp. 393–401, 2015.
- [51] D.-Y. Shi, J. Lu, and L.-J. Lu, "A judge model of the impact of lane closure incident on individual vehicles on freeways based on RFID technology and FOA-GRNN method," *Journal of Wuhan University of Technology*, vol. 34, no. 3, pp. 63–68, 2012.
- [52] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "GSA: a gravitational search algorithm," *Information Sciences*, vol. 179, no. 13, pp. 2232–2248, 2009.
- [53] R. K. Sahu, S. Panda, and S. Padhan, "A novel hybrid gravitational search and pattern search algorithm for load frequency control of nonlinear power system," *Applied Soft Computing*, vol. 29, pp. 310–327, 2015.
- [54] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "Filter modeling using gravitational search algorithm," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 117–122, 2011.
- [55] F.-Y. Ju and W.-C. Hong, "Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting," *Applied Mathematical Modelling. Simulation and Computation for Engineering and Environmental Systems*, vol. 37, no. 23, pp. 9643–9651, 2013.
- [56] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.
- [57] S.-M. Chou, T.-S. Lee, Y. E. Shao, and I.-F. Chen, "Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines," *Expert Systems with Applications*, vol. 27, no. 1, pp. 133–142, 2004.
- [58] C.-J. Lu, T.-S. Lee, and C.-M. Lian, "Sales forecasting for computer wholesalers: a comparison of multivariate adaptive regression splines and artificial neural networks," *Decision Support Systems*, vol. 54, no. 1, pp. 584–596, 2012.
- [59] D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*, MIT Press, 2001.
- [60] A. Etemad-Shahidi and J. Mahjoobi, "Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior," *Ocean Engineering*, vol. 36, no. 15-16, pp. 1175–1181, 2009.
- [61] J. R. Quinlan, "Learning with continuous classes," in *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, pp. 343–348, Singapore, 1992.
- [62] B. Bhattacharya and D. P. Solomatine, "Neural networks and M5 model trees in modelling water level-discharge relationship," *Neurocomputing*, vol. 63, pp. 381–396, 2005.
- [63] S. Londhe and P. Dixit, "Stream flow forecasting using model trees," *International Journal of Earth Sciences and Engineering*, vol. 4, no. 6, pp. 282–285, 2011.
- [64] M. Janga Reddy and B. N. S. Ghimire, "Use of model tree and gene expression programming to predict the suspended sediment Load in rivers," *Journal of Intelligent Systems*, vol. 18, no. 3, pp. 211–227, 2009.
- [65] M. T. Sattari, M. Pal, H. Apaydin, and F. Ozturk, "M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey," *Water Resources*, vol. 40, no. 3, pp. 233–242, 2013.

- [66] K. K. Singh, M. Pal, and V. P. Singh, "Estimation of mean annual flood in indian catchments using backpropagation neural network and M5 model tree," *Water Resources Management*, vol. 24, no. 10, pp. 2007–2019, 2010.
- [67] D. P. Solomatine and Y. Xue, "M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China," *Journal of Hydrologic Engineering*, vol. 9, no. 6, pp. 491–501, 2004.
- [68] A. Zahiri and H. M. Azamathulla, "Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels," *Neural Computing and Applications*, vol. 24, no. 2, pp. 413–420, 2014.
- [69] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2005.
- [70] A. Heiat, "Comparison of artificial neural network and regression models for estimating software development effort," *Information and Software Technology*, vol. 44, no. 15, pp. 911–922, 2002.
- [71] M. Zounemat-Kermani, Ö. Kişi, J. Adamowski, and A. Ramezani-Charmahineh, "Evaluation of data driven models for river suspended sediment concentration modeling," *Journal of Hydrology*, vol. 535, pp. 457–472, 2016.
- [72] I. Myrtveit, E. Stensrud, and M. Shepperd, "Reliability and validity in comparative studies software prediction models," *IEEE Transactions on Software Engineering*, vol. 31, no. 5, pp. 380–391, 2005.
- [73] B. Li and C. Cheng, "Monthly discharge forecasting using wavelet neural networks with extreme learning machine," *Science China Technological Sciences*, vol. 57, no. 12, pp. 2441–2452, 2014.
- [74] A. G. Yilmaz and N. Muttil, "Runoff estimation by machine learning methods and application to the Euphrates Basin in Turkey," *Journal of Hydrologic Engineering*, vol. 19, no. 5, pp. 1015–1025, 2014.
- [75] C. Santhi, J. G. Arnold, J. R. Williams, W. A. Dugas, R. Srinivasan, and L. M. Hauck, "Validation of the swat model on a large RWER basin with point and nonpoint sources," *Journal of the American Water Resources Association*, vol. 37, no. 5, pp. 1169–1188, 2001.
- [76] J. Alreja, S. Parab, S. Mathur, and P. Samui, "Estimating hysteretic energy demand in steel moment resisting frames using multivariate adaptive regression spline and least square support vector machine," *Ain Shams Engineering Journal*, vol. 6, no. 2, pp. 449–455, 2015.
- [77] A. Kibekbaev and E. Duman, "Benchmarking regression algorithms for income prediction modeling," *Information Systems*, vol. 61, pp. 40–52, 2016.
- [78] P. Samui and D. Kim, "Least square support vector machine and multivariate adaptive regression spline for modeling lateral load capacity of piles," *Neural Computing and Applications*, vol. 23, no. 3–4, pp. 1123–1127, 2013.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

