Hindawi
Journal of Mathematics
Volume 2023, Article ID 4222934, 13 pages
https://doi.org/10.1155/2023/4222934

Research Article

Use of Wavelet and Bootstrap Methods in Streamflow Prediction

Adnan Bashir,[1] **Muhammad Ahmed Shehzad**,[1] **Aamna Khan**,[1] **Ayesha Niaz**,[1] **Muhammad Nabeel Asghar**,[2] **Ramy Aldallal**,[3] **and Mutua Kilai**[4]

[1]*Department of Statistics, Bahauddin Zakariya University, Multan 60000, Pakistan*
[2]*Department of Computer Science, Bahauddin Zakariya University, Multan 60000, Pakistan*
[3]*Department of Accounting, College of Business Administration in Hawtat Bani Tamim, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia*
[4]*Department of Mathematics, Pan African Institute of Basic Science Technology and Innovation Nairobi, Juja, Kenya*

Correspondence should be addressed to Muhammad Ahmed Shehzad; ahmad.shehzad@bzu.edu.pk

Received 3 September 2022; Revised 13 January 2023; Accepted 18 January 2023; Published 17 February 2023

Academic Editor: Xiangfeng Yang

Streamflow prediction is vital to control the effects of floods and mitigation. Physical prediction model often provides satisfactory results, but these models require massive computational work and hydrogeomorphological variables to develop a prediction system. At the same time, data-driven prediction models are quick to apply, easy to handle, and reliable. This study investigates a new hybrid model, the wavelet bootstrap quadratic response surface, for accurate streamflow prediction. Wavelet analysis is a well-known time-frequency joint analysis technique applied in various fields like biological signals, vibration signals, and hydrological signals. The wavelet analysis is used to denoise the time series data. Bootstrap is a nonparametric method for removing uncertainty that uses an intensive resampling methodology with replacement. The authors analyzed the results of the studied models with different statistical metrics, and it has been observed that the wavelet bootstrap quadratic response surface model provides the most efficient results.

1. Introduction

Water is essential for all living things, including plants, animals, and people. Water is an all-purpose solvent and a gift from nature to all living things. The sustainability of modern ecology, human existence on Earth, and the provision of food for an expanding population depend heavily on water availability [1]. Ghafoor and Nawaz [2] explain the worldwide phenomenon of climate change and its variety of implications for of various ecologies. For example, variations in rainfall intensity and duration could cause floods or droughts, and these catastrophes pose a serious risk to regional and global food security. In the last few decades, streamflow prediction plays a significant part in the hydrology field and the administration of water resources. Accurate and reliable streamflow estimation provides the foundation to increase the capability of reservoirs, flood avoidance, water resources, and the design of hydroelectric projects [3]. Thus, for a short period, streamflow prediction can be helpful to enhance the management of water supply. However, accurate and consistent prediction is a tremendous challenge for hydrologists because of the complicated variability in the river system [4].

In the last two decades, artificial intelligence (AI) techniques such as adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM), and neural network (NN), have been employed for river stage prediction. These methods were used because of their skills to model the nonlinear behavior of time series data [5–12]. A deep learning model is used to estimate the uncertainty associated with river flow and determine flood prediction [13]. Wavelet artificial neural network (WANN) model tested by Shafaei and Kisi [14] for river flow predictions. Results depict that the proposed model yields good results than the traditional artificial neural network (ANN) and support vector machine (SVM) models. Delafrouz et al. [15] planned a new hybrid model by conjunction of phase-space reconstruction (PSR) and ANN methods for reliable daily streamflow prediction.

The developed PSR-ANN model was compared with the traditional ANN and gene expression programming (GEP) models. It has been concluded that the proposed model, PSR-ANN produced the best prediction result. Two AI-based techniques named WANN and linear genetic programming (LGP), were introduced by Danandeh Mehr et al. [16] for monthly streamflow forecasting. The results showed that the LGP model is better than the WANN model for monthly streamflow prediction. The SVM model with an adaptive insensitive factors introduced by Gueo et al. [17] for monthly streamflow prediction. The wavelet transform (WT) technique is practiced to reduce noise from data. The phase-space reconstruction technique is applied to determine the structure of the forecasting model. Results claimed that an improved SVM model with adaptive insensitive factor is suitable for processing complex hydrology data.

WT is a prevalent technique to remove noise from nonstationary time series data [18–22]. WANN and wavelet adaptive neuro-fuzzy inference system (WANFIS) models were examined by [23] to check the reliability of the proposed models for maximum lead time. It has been revealed that the WANFIS model is suitable for 1–6-hour prediction and the WANN model is a reliable model for 8–10-hour prediction. Results indicated that WT increases the efficiency of both models. Drisya et al. [24] compared the feedforward neural network (FFNN) and WANN model to analyze streamflow prediction. The WANN model represents more high-quality results than the FFNN model for streamflow prediction. Seo et al. [25] introduced hybrid models: wavelet packet artificial neural network (WPANN), wavelet packet-adaptive neuro-fuzzy inference system (WP-ANFIS), and wavelet packet-support vector regression (WPSVR) by the conjunction of wavelet packet decomposition to the traditional machine learning models like ANN, ANFIS, and support vector regression (SVR). The wavelet packet decomposition (WPD) technique significantly increases the predictive power of the machine learning models. Overall, the WP-ANFIS model produces reliable results for daily river stage prediction. Wavelet-based regression models and wavelet-based NN models are studied by Partal [26] for monthly streamflow prediction. The wavelet transformation has a positive impact on forecasting results when coupled with regression and NN-based models. Khan et al. [27] coupled the wavelet technique with autoregressive integrated moving average (ARIMA) and ANN models for future drought forecasting. Wavelet bootstraps multiple linear regression (WBMLR) model proposed by Sehgal et al. [28] for river stage prediction. They showed that the WBMLR model produced better results than the remaining models used in the study based on ANN and MLR models. Two-hybrid models, wavelet neural network (WNN) and ANN, are attached with block bootstrap sampling (BB) technique by Kasiviswanathan et al. [29]. They concluded that the WNN-BB model consistently yields more performance for flood management than the ANN-BB model. RS method of higher-order polynomial functions is used for flood forecasting. They concluded that the RS model with a fifth-order better forecast river stage [30].

This study aims to determine the performance of the wavelet bootstrap quadratic response surface (WBQRS) model for daily reservoir inflow prediction. In addition, it compares the performance of the developed model with some traditional and hybrid models.

2. Methods

2.1. Wavelet Transform. WT can solve these problems by decomposing one-dimensional signals into two-dimensional time-frequency domains at a similar time. Moreover, wavelets have the property of irregularity and are asymmetric in shape. Due to these properties, the wavelet technique is useful for analyzing signals having harp changes and discontinuities [31]. DWT technique provides a time-scale representation of time-series data and finds a relationship. The DWT technique is useful for nonstationary data to remove noise from data. In addition, the DWT technique is useful when data constitute jumps or shifts and produces a very accurate analysis [32].

Time series analysis is solved by applying numerous techniques, one of which, the most well-known is Fourier transform (FT). Unfortunately, although, FT breaks down time-series data into basic sinusoids of various frequencies while time information is lost during the process of transformation. Therefore, it is recommended that FT is not a suitable choice for analyzing signals with transitory characteristics such as trends, discontinuities, drift, and breakdown points. On the other hand, WT deals with a long interval of time when low-frequency components are required, and it deals with a short interval of time when high-frequency components are the need of time [33].

WT is obtained through the time-frequency components of a signal. WT is divided into two main types: discrete wavelet transforms (DWT) and continuous wavelet transforms (CWTs). The function of CWT is given by

$$W(a,\tau) = \int_{-\infty}^{+\infty} f(t)\psi_{a,\tau}(t)\,\mathrm{d}t, \tag{1}$$

where

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right). \tag{2}$$

Calculating wavelet coefficients at every possible resolution level (scale) creates a massive amount of data. Therefore, the choice of resolution level based on the dyadic scale and wavelet analysis obtained through dyadic scale should be more accurate and efficient. This transformation is called DWT. In equation (1), $f(t)$ is converted into a wavelet coefficient $W(a,\tau)$. Here, τ (real number) is indicated by the translation parameter, whereas a (real and positive number) is symbolized by the dilation parameter. These two parameters are dilating and translating the mother wavelet, $\psi(t)$, respectively. The DWT was obtained by these two parameters and dilation parameter a is discretized while the $a = 2^{-j}$ translation parameter. τ is discretized with $\tau = K2^{-j}$. In the process of discretization, the wavelet function gets the form

$$\psi_{j,k}(t) = 2^{(j/2)}\psi\left(2^j t - K\right), \tag{3}$$

where j and K are the integers represented by scale and translation factors, respectively [34].

Mallat [35] introduced two filters that isolate the signal into two different scales: the high-pass filter and low-pass filter. A high pass filter (wavelet function) represents high frequency and low scale. Wavelet functions are rapidly changing signal features, and these functions are acquired by correlating the original signal with the compressed signal. Other names of high-pass filter (wavelet function) include running differences, detail, and fluctuation. Low-pass filter (scaling function) constitutes low frequency and high scale. Scaling function is also known as an approximation or trend. Figure 1 represents the scaling function and wavelet function of symlet wavelet with vanishing moment 15.

2.2. MLR Model. In regression analysis, intercept and regression coefficients are estimated by employing the method of minimizing the sum of squared residuals. The MLR model is used in hydrological modeling for several decades. Tiwari and Chatterjee [36] and Kisi [37] introduced recent applications in the hydrological literature.

$$Y = \beta^{\circ} + \beta_1 x_1 + \beta_2 x_2 + \varepsilon. \tag{4}$$

The response variable Y is a column vector of order $n \times 1$.

The predictor variable represented by X is an $n \times 2$ matrix.

Regression coefficients β correspond to a 2×1 column vector.

The error term u is an $n \times 1$ column vector.

2.3. Bootstrap Technique. The bootstrap technique resamples the original data set with replacements and trains the model on each resampled data point instead of the original data. The bootstrapping technique applied to develop a single realization of a distribution to generate a set of bootstrap samples. These bootstrap samples furnish a better understanding of the average and variability of the original unknown distribution [36]. Suppose ordinary bootstrap proof is inconsistent with a wider confidence interval. So, there is no need to use the $m = n$ bootstrap sample size. For consistency, the m-out-of-n bootstrap technique is useful in such a situation and proves bootstrap to be consistent [38].

Three steps are involved in applying the bootstrap: sampling for estimation of bootstrap samples with replacement, determining the bootstrap distribution, and final application of the bootstrap distribution. The bootstrap distribution of statistics is indicated as the sampling distribution of statistics based on the resampling technique. The main advantage of bootstrap sampling has the simplicity of performing the complex calculation of mean, standard error, and confidence interval. Therefore, the bootstrap technique could be successfully applied compared to other resampling techniques like Jackknife in estimating unsmoothed parameters and estimating the variance of nonlinear statistics. The vital point to be remembered here is that the parametric bootstrap is used when we have information about the normal distribution. On the other hand, when the distribution is non-normal, it cannot fulfill the assumption of normality. In such a situation, we use a nonparametric bootstrap sample [39].

Let ε denoted by a set of the random sample that is independently and identically (iid) distributed with sample size n. These samples are drawn from an unknown probability distribution $F(X_i \sim \text{iid}\, F)$ while $(1/n)$ assumed to be observed value. Bootstrap samples generated from observed values and empirical distribution F (bootstrap population) developed using uniform sampling with replacement of observed time series data. $X = (x_1, x_2, x_3, \ldots, x_n)$ be a sample taking probability $(1/n)$ on each x_i value. The set of a bootstrap sample can be indicated as $X^1, X^2, X^3, \ldots, X^n$, where B is the total number of bootstrap samples.

2.4. Response Surface Method. Response surface modeling is a procedure to mimic the system of responses because of changes in predictors [40]. RS modeling is a helpful technique to deal with natural events in the field of hydrology. A mathematical equation can express the relationship between regressors and the response variable.

$$Y = f\left(x_1, x_2, x_3, \ldots, x_n\right) + \varepsilon. \tag{5}$$

In equation (5), the response variable is expressed by Y, f represents the response function, $x_1, x_2, \ldots, x_n$ are regressors, and ε denoted by the error term.

The mathematical equation expresses a linear relationship between regressors and the response variables.

$$Y = \beta^{\circ} + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon. \tag{6}$$

FORS model involves the cross-product terms.

The mathematical equation of QRS model is

$$Y = \beta^{\circ} + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon. \tag{7}$$

QRS model comprises additional second-order terms in comparison with the FORS model.

2.5. Performance Indices. The statistical indicators to compare the performance of the developed models are RMSE, MAE, NSE, and CP.

$$\begin{aligned}
\text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{n}\left(O_i - P_i\right)^2}{n}},\\
\text{MAE} &= \frac{1}{n}\left(\sum_{i=1}^{n}\left|O_i - P_i\right|\right),\\
\text{NSE} &= 1 - \frac{\sum_{i=1}^{n}\left(O_i - P_i\right)^2}{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2},\\
\text{CP} &= 1 - \frac{\sum_{i=2}^{n}\left(P_i - O_i\right)^2}{\sum_{i=1}^{N-1}\left(O_{i+1} - O_i\right)^2},
\end{aligned} \tag{8}$$

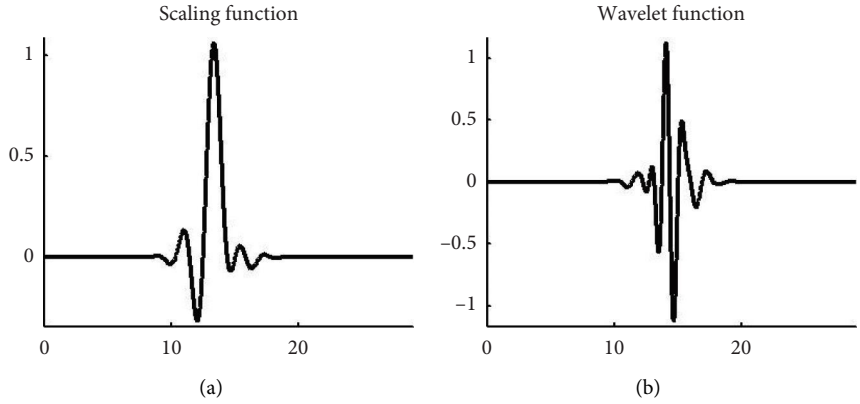

Figure 1: (a) Represents the scaling function of symlet wavelet with vanishing moment 15; (b) denotes the wavelet function of symlet wavelet with vanishing moment 15.

where O_i is represented by the observed flow, P_i is the predicted flow, and the size of the data set indicates n. RMSE determines the variation in error under the independence of sample size. It determines the discrepancy between observed and predicted streamflow values. The value of RMSE lies between zero and 1. The average absolute error is called the MAE. It shows how close the forecast value is to the observed value [41]. NSE is applied in the hydrologic field to appraise the predictive power. The range of NSE is from $-\infty$ to 1. As the value of NSE is close to 1, it means that the model fit is good.

3. Study Area and Data Set

River Chenab is structured by joining the Chandra and Bhaga rivers at Tandi in Himachal Pradesh state, India. Then, it enters into the plains downward from the uplands to the vast alluvial lowlands of Punjab and Pakistan. The Chenab river basin is positioned in eastern Punjab and flows through the southwestern direction in Punjab province. Marala, Khanki, and Qadrabad gauging stations play an important role in the river canal link system in Punjab and Pakistan. Marala gauging station was built in district Sialkot in 1968 and had a length of 1.366 kilometers. It is situated between $32^{\circ}40'24''$ N to $74^{\circ}27'50''$ E and has a discharge capacity of 1.1 million cusec water. Khanki gauging station is situated on the river Chenab when it enters in Gujranwala district. It was constructed in 1982 and had a maximum discharge capacity of 0.8 million cusecs. The next gauging station on river Chenab is Qadrabad at district Mandi Bahauddin. The total length of the Chenab river basin is around 974 km, and several irrigation canals are fed by it. Table 1 indicates the summary statistics of the three selected gauging stations for the analysis of this study while Figure 2 represents the Chenab river basin with its adjoining rivers. Where for simplicity, authors use abbreviations of gauging stations Marala = Mar, Khanki = Kha, Qadrabad = Qad, Trimmu = Tri, and Punjab = Punj.

Data used for this study are the daily water discharge of Chenab river basin from three gauging stations during 2005–2010 (1 July–30 September). The data constitutes into two parts. The first part consists of training data from 2005–2009 while the testing data constitute of the 2010 year.

Table 1: Summary of gauging stations of the Chenab river basin.

Location	Marala	Khanki	Qadrabad
Latitude (N)	32.78°	32.40°	32.32°
Longitude (E)	73.88°	73.98°	73.70°
Altitude (m)	234	214	210
The area above sea level (m)	250	250	250

4. Model Development

The prime step in hydrological modeling is selecting appropriate input variables for the models [29, 42]. However, there is no exact method for the selection of the input variables of prediction models. Therefore, different researchers use different techniques for the selection of significant variables. In the present study, the correlation technique is applied for the selection of input variables and considers the geographic location of gauging stations. Figure 1 represents the Chenab river basin with its adjoining rivers.

Correlations of gauging stations are shown in Table 2 to select the relevant input variables for streamflow prediction. Pearson correlation coefficient applies to calculate correlation among gauging stations (variables) on complete data to select suitable input variables. Its mathematical form is given as

$$r = \frac{n\sum \mathrm{XY} - \sum X \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\left[n\sum Y^2 - \left(\sum Y\right)^2\right]\right]}}. \quad (9)$$

Table 2 clearly shows that the Mar gauging station strongly correlates with Kha and Qad gauging station, but it has a week correlation with Tri and Punj gauging stations. Kha gauging station has a strong correlation with Mar and Qad gauging stations. Tri and Punj have a strong correlation with each other. If we look at the geographic location of the Chenab river basin, it is observed that the Jehlum river joins river Chenab at Tri Gauging station. Therefore, Qad gauging

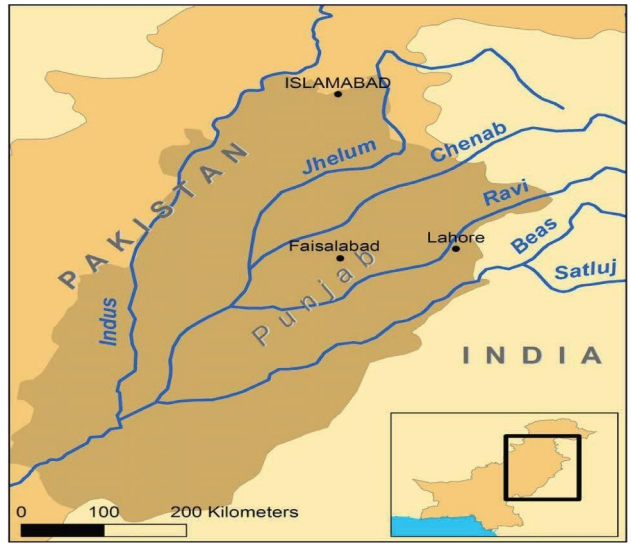

Figure 2: Chenab river basin and its adjoining river map.

Table 2: Linear correlation analysis among variables.

Combinations	Correlations
Mar and Kha	0.8690
Mar and Qad	0.7780
Mar and Tri	0.4980
Mar and Punj	0.4040
Kha and Qad	0.9370
Kha and Tri	0.4910
Kha and Punj	0.3830
Qad and Tri	0.4700
Qad and Punj	0.3470
Tri and Punj	0.8360

station is our forecast site and is used as a response variable, while Mar and Kha used it as predictors.

The main idea behind the correlation analysis is to determine the relationship between two variables. "Correlation analysis is the study of the relationship between variables" [43] p. 375). The correlation coefficient between Mar and Kha is 0.8690 (Table 2), which is the sign of high multicollinearity among both predictors. The principal component analysis technique was applied to remove the multicollinearity problem. By using this operation on the predictors, principal components are formed. These components are orthogonal among each other and have an independent linear relationship. Results obtained from these components produced principal component scores. Principle scores are applied as input to the predictors to eliminate multicollinearity among predictors [44].

DWT method is applied to decompose the data of river streamflow into wavelet components. Choice of the mother wavelet is another important step to perform wavelet analysis because the results obtained from time-series data greatly depend upon selecting the mother wavelet. Maheswaran and Khosa [45] explain that the choice of a suitable mother wavelet is based on the properties and application of the wavelet function, that is, the vanishing moments and region of support. The support region of the wavelet function explains the characteristics of localization and vanishing moments dealing with the polynomial behavior of time series data [46]. Four different wavelet families with different vanishing moments are tested on a traditional FORS-based model for 1-d ahead prediction to select the best one wavelet function. The wavelet families to be tested are Haar, symlets (sym2, sym6, sym12, sym15, and sym20), coiflets (coif1, coif2, and coif4), and Daubechies (db4, db12, db15, and db20). The symlet mother wavelet function with vanishing moment 15 is selected to decompose the time series data because its performance is good on all performance indices compared to other mother wavelet functions, as represented in Table 3.

The best performance in the wavelet domain is also based on selecting the optimal level of decomposition. However, the decomposition level of WT can be determined using the mathematical formula as given in the following [47, 48]:

$$L = \text{int}[\log(N)]. \tag{10}$$

In this empirical formula, L indicates the length of data; it is the function to convert the decimal parts to an integer, and N represents the total length of data. Three levels of decomposition are chosen by using the empirical formula of the equation. First, the data decomposes into three wavelet components: detail (D_1, D_2, and D_3) and approximation (A_3). Then, the effective wavelet components (D_3 and A_3) are chosen to provide input to the wavelet-based models. Figure 3 indicates the time series data with wavelet decomposition at level third.

4.1. Wavelet-Based Model Development. DWT technique coupled with the RS-based model (FORS and QRS) to produce new hybrid models: WFORS and WQRS. The algorithm of WFORS and WQRS models consists of two steps.

(1) At the first step, we decompose the original streamflow data into wavelet components and use

Table 3: Performance of various WFORS models with different mother wavelet functions.

Wavelet	Vanishing moment	NSE	RMSE (m^3/s)	MAE (m^3/s)	CP
Haar	1	0.9560	0.3850	0.2688	0.9596
Coif1	1	0.9283	0.4600	0.2627	0.9288
Coif2	2	0.9582	0.3261	0.2127	0.9389
Coif4	4	0.9634	0.2764	0.1818	0.9314
Db4	4	0.9476	0.3770	0.2519	0.9418
Db10	10	0.9469	0.3225	0.2172	0.9275
Db12	12	0.9641	0.2700	0.1926	0.9514
Db15	15	0.9617	0.2726	0.2114	0.9616
Db20	20	0.9578	0.3250	0.2291	0.9465
Sym2	2	0.9764	0.2533	0.1903	0.9725
Sym6	6	0.9770	0.2516	0.1767	0.9612
Sym12	12	0.9687	0.2576	0.1520	0.9378
Sym15	15	0.9776	0.2250	0.1469	0.9521
Sym20	20	0.9681	0.2390	0.1538	0.9467

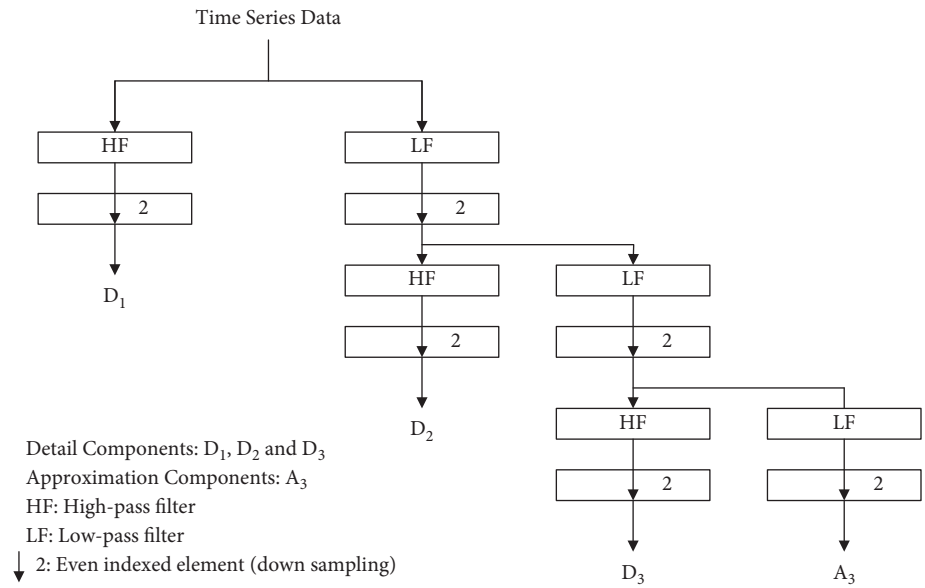

Figure 3: Shows time series data with wavelet components (D_1, D_2, and D_3) and approximation component (A_3).

the DWT technique after selecting the optimal decomposition level.

(2) In the second phase, the effective discrete wavelet component provides input to the FORS and QRS models.

4.2. Wavelet-Bootstrap-Based Model Development. Finally, we develop new hybrid models using wavelet and bootstrap methods on the FORS model to establish the WBFORS model. These two methods are coupled with the QRS model to develop the WBQRS model. To develop the wavelet-bootstrap-based models, we operate the bootstrap technique on the effective wavelet components and provide this bootstrap resampled data to the FORS and QRS models. Flowchart in Figure 4 shows the process of the hybrid models.

Figure 4 shows the algorithm of the developed models in this study. This article studies the relationship between Qad gauging station (predicting variable, Y), Mar gauging station, and Kha gauging station (predictors: $X1$ and $X2$). For that matter, the traditional models employed by using the least square method are

$$Y = \beta^{\circ} + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \text{ (MLR)},$$
$$Y = \beta^{\circ} + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon, \text{ (FORS)},$$
$$Y = \beta^{\circ} + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon, \text{ (QRS)}, \tag{11}$$

Authors collect data on the monsoon season for the river Chenab (1 July–30 September) for all five gauging stations in Pakistan. Then, the correlation technique is applied to select suitable input variables. So, after applying the correlation technique, the selected variables are Mar, Kha, and Qad.

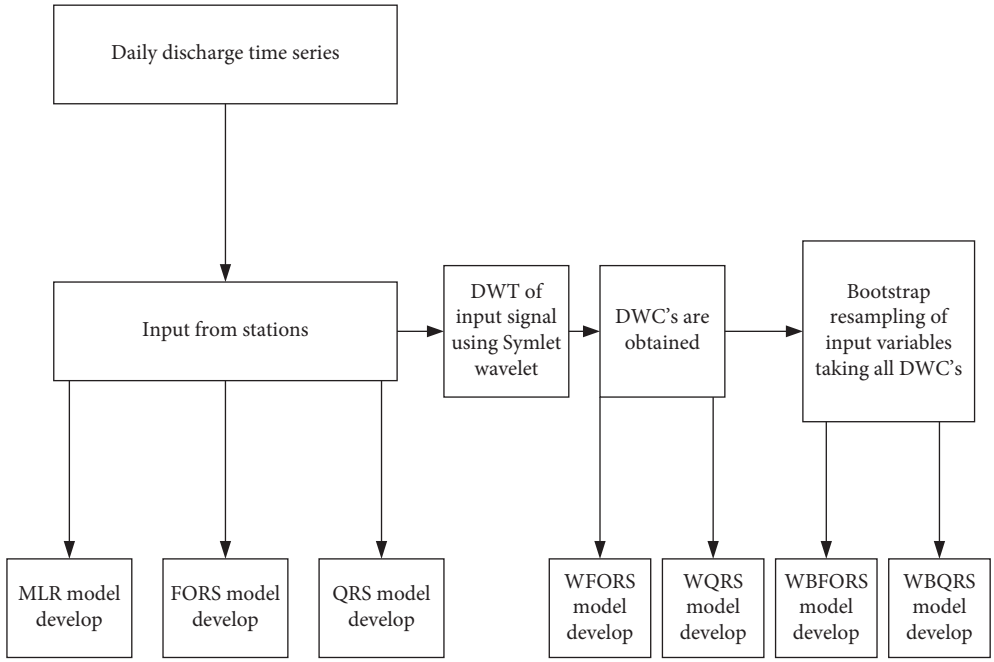

Figure 4: The flowchart illustrates the proposed model development.

Table 4: Statistical indicator for MLR, FORS, QRS, WFORS, WQRS, WBFORS, and WBQRS models.

Models	Days	RMSE (m^3/s)	MAE (m^3/s)	NSE	CP
MLR	1 day	0.4010	0.2082	0.8426	0.7398
	2 day	0.4010	0.2064	0.8429	0.7399
	3 day	0.4010	0.2059	0.8430	0.7402
FORS	1 day	0.3780	0.1941	0.8621	0.7719
	2 day	0.3770	0.1923	0.8622	0.7719
	3 day	0.3770	0.1922	0.8622	0.7719
WFORS	1 day	0.2250	0.1470	0.9761	0.9521
	2 day	0.2230	0.1437	0.9779	0.9526
	3 day	0.2220	0.1415	0.9782	0.9529
WBFORS	1 day	0.0686	0.0466	0.9711	0.9816
	2 day	0.0313	0.0234	0.9925	0.9977
	3 day	0.0115	0.0086	0.9965	0.9984
QRS	1 day	0.3390	0.1952	0.8913	0.8201
	2 day	0.3390	0.1942	0.8913	0.8202
	3 day	0.3390	0.1936	0.8914	0.8202
WQRS	1 day	0.2170	0.1354	0.9796	0.9563
	2 day	0.2160	0.1328	0.9799	0.9565
	3 day	0.2150	0.1316	0.9800	0.9567
WBQRS	1 day	0.0356	0.0283	0.9877	0.9957
	2 day	0.0165	0.0106	0.9965	0.9974
	3 day	0.0147	0.0108	0.9954	0.9970

According to its geographical location, the Qad gauging station is the response variable and the remaining two gauging stations (Mar and Kha) are the independent variables. Next, the performance of all studied models (Figure 4) is checked on the performance indices: RMSE, MAE, CP, and NSE. Finally, the results are presented based on the testing data set in Table 4. Then in the second step, the authors applied the wavelet technique to the time series data by using symlet mother wavelet function with vanishing moment fifteen and at level third. Next, data from effective wavelet components A_3 and D_3 are applied to get WFORS and WQRS models. After that, the bootstrap technique was applied to the DWCs to get WBFORS and WBQRS models. The bootstrap technique removes uncertainty from data because hydrological data have nonstationary trends. When authors hybrid the bootstrapping technique with the wavelet method, the model gets outstanding prediction results, as represented in Table 4. For attaining model consistency and removing the effect of randomness, the authors use the m-out-of-n technique. In this method, m represents the

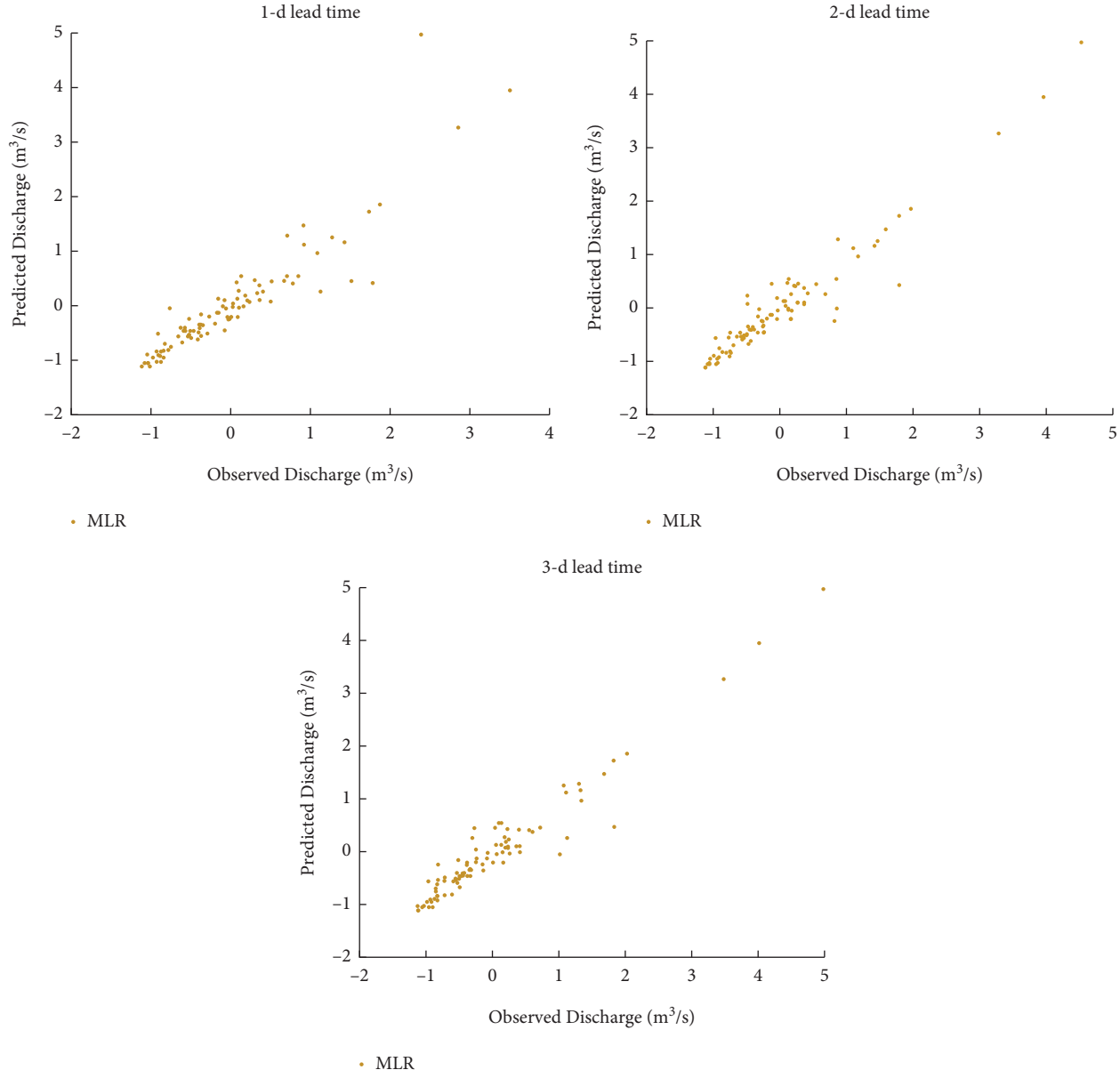

Figure 5: Scatter diagram represents the observed discharge on the horizontal axis means the data recorded on the Qad gauging station (represented by Y in the model). On the vertical coordinate, predicted discharge means the data obtained after fitting the MLR model.

bootstrap sample size, and n is the number of repetitions. So, carefully selecting the bootstrap sample size and the number of bootstrap repetitions to ensure the model gets consistent results. The m-out-of-n bootstrap technique is useful for consistency and proves bootstrap models get consistent results with repetition [38].

5. Results and Discussion

To determine the predictive ability of the determined models, the performance indices to be used are RMSE, MAE, NSE, and CP. The performance of the MLR, FORS, WFORS, WBFORS, QRS, WQRS, and WBQRS models is presented in Table 4. In general, the QRS model attains the optimum performance in this study for river inflow prediction than other traditional models, MLR and FORS.

This study compares the wavelet-based RS models: WFORS and WQRS with the traditional models: QRS, FORS, and MLR. It has been observed from Table 3 that the predictive efficiency of the wavelet-based models (WFORS and WQRS) is much better than the traditional models MLR, FORS, and QRS concerning statistical techniques: RMSE, MAE, NSE, and CP in Chenab river basin. The results of MLR, FORS, and QRS models are poor compared to the WFORS and WQRS models. The reason behind the week performance of the MLR, FORS, and QRS models is that these models use raw streamflow data for modeling. In contrast, such raw streamflow data comprises various frequency components. River streamflow data do not present actual characteristics of the time series data when it utilizes one resolution component for prediction purposes [14]. Therefore, wavelet-based models (WFORS and WQRS) use different frequency resolution levels for streamflow data.

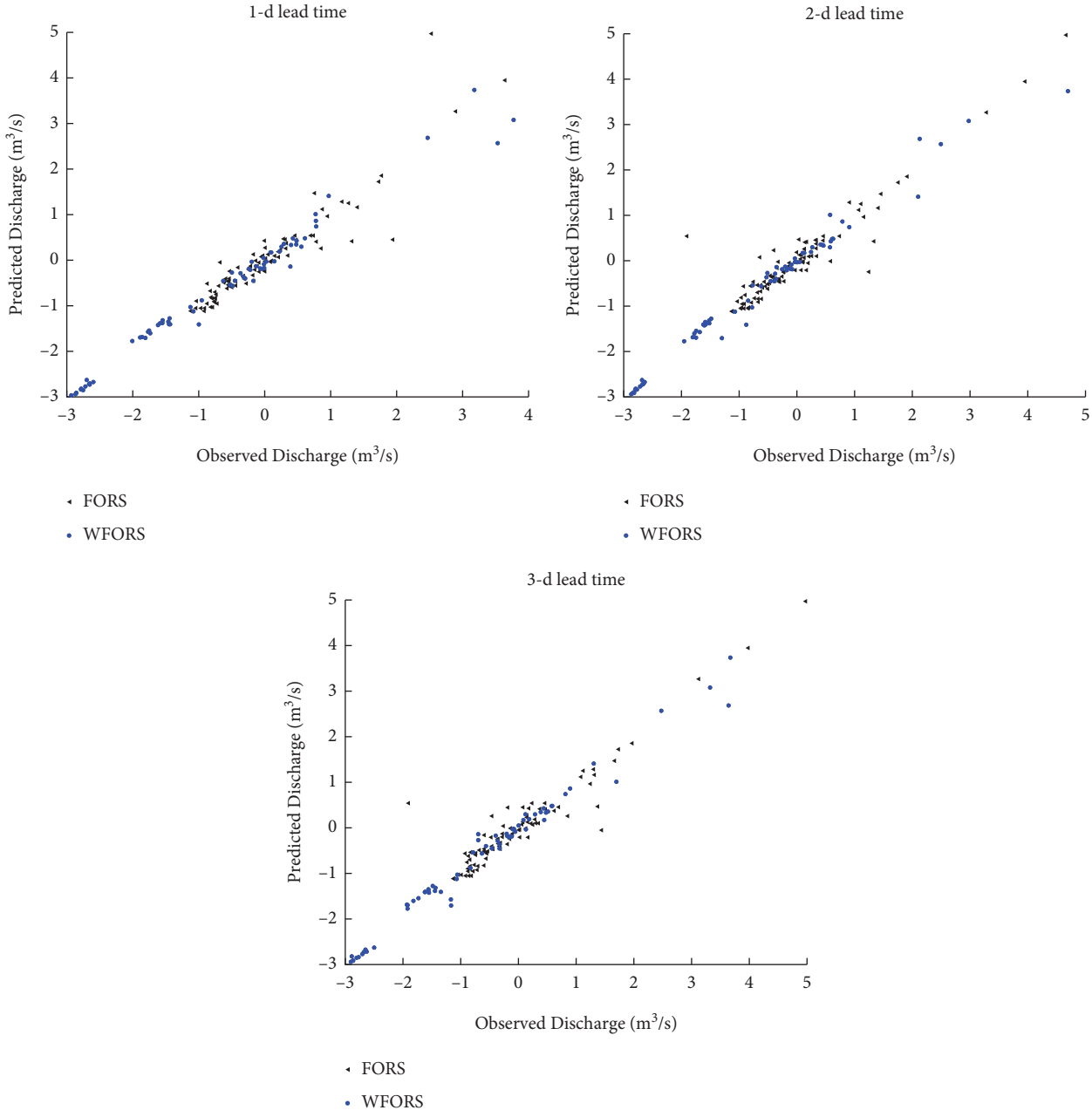

Figure 6: Represents scatter diagram of the observed reservoir flow against the predicted reservoir flow of FORS and WFORS models.

Wavelet transform is the most suitable choice for reservoir inflow prediction rather than the traditional models. On the basis of performance indices (RMSE, MAE, NSE, and CP), the WQRS (RMSE = 0.217 m^3/s, MAE = 0.1354 m^3/s, NSE = 0.9796, CP = 0.9563) and WFORS (RMSE = 0.2250 m^3/s, MAE = 0.1470 m^3/s, NSE = 0.9761, CP = 0.9521) models has much better results for 1-d ahead prediction of testing data set than the traditional models MLR (RMSE = 0.4010 m^3/s, MAE = 0.2082 m^3/s, NSE = 0.8426, CP = 0.7398), FORS (RMSE = 0.3780 m^3/s, MAE = 0.1941 m^3/s, NSE = 0.8621, CP = 0.7719), and QRS (RMSE = 0.3390 m^3/s, MAE = 0.1952 m^3/s, NSE = 0.8913, CP = 0.8201).

Values of the performance indices RMSE and MAE are higher for models MLR, FORS, QRS, WFORS, and WQRS than the models WBFORS and WBQRS, while the values of performance indices NSE and CP are closer to the one for WBFORS and WBQRS models compared to the remaining models for 1–3 ahead prediction. Thus, this study reveals

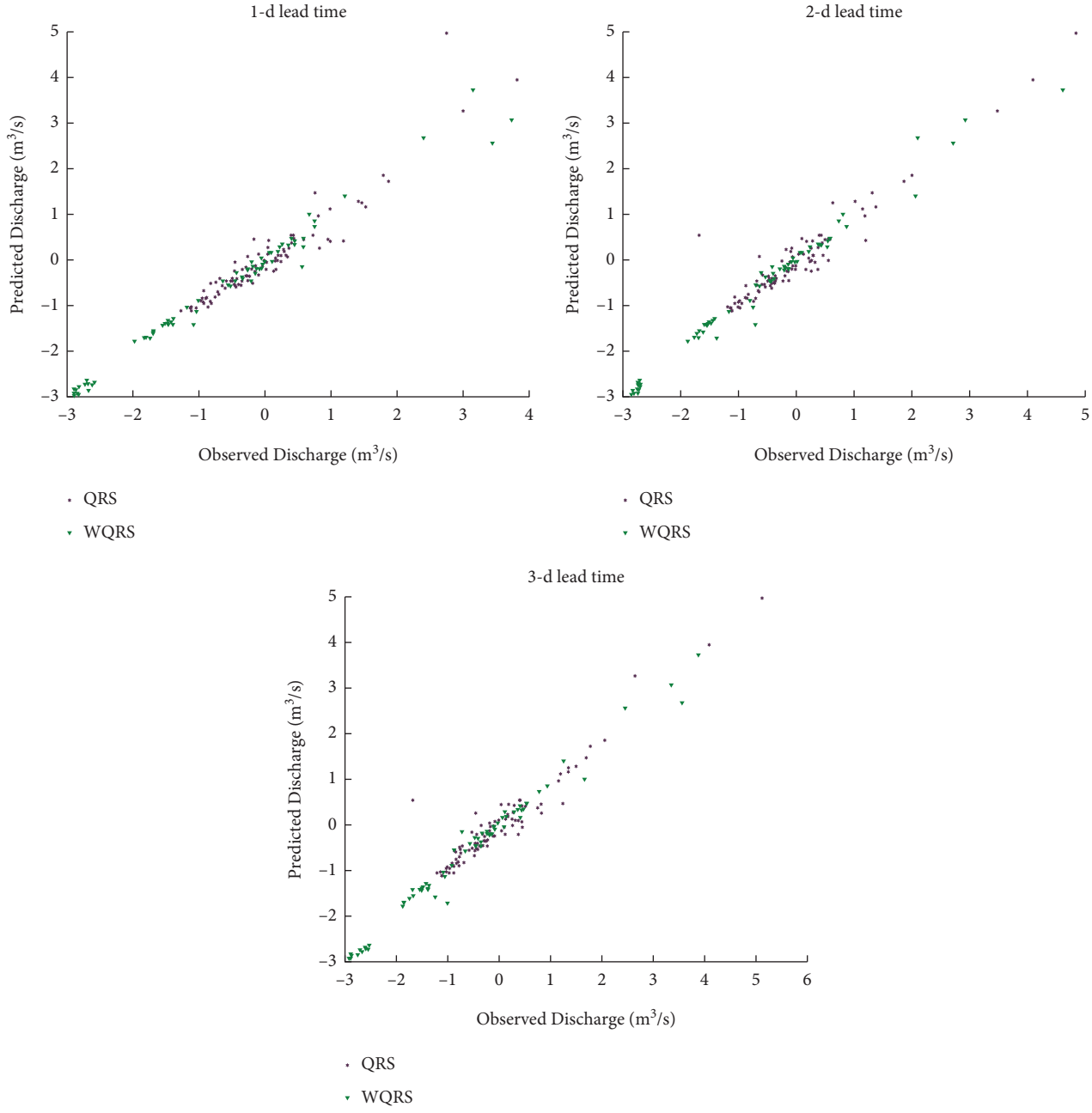

Figure 7: Represents scatter diagram of the observed reservoir flow against the predicted reservoir flow of QRS and WQRS models.

that both models: WBFORS and WBQRS yield better performance in terms of prediction, but the performance of the WBQRS model is excellent.

Scatter plots for 1 d and 3 d ahead prediction of MLR, FORS, WFORS, WBFORS, QRS, WQRS, and WBQRS model are presented in Figures 5–8, respectively. According to the NSE criterion, the WBQRS model shows that the straight line is good and in better agreement with the observed streamflow. Therefore, the WBQRS model indicates good predictive power with 1–3 days ahead prediction compared to the other models.

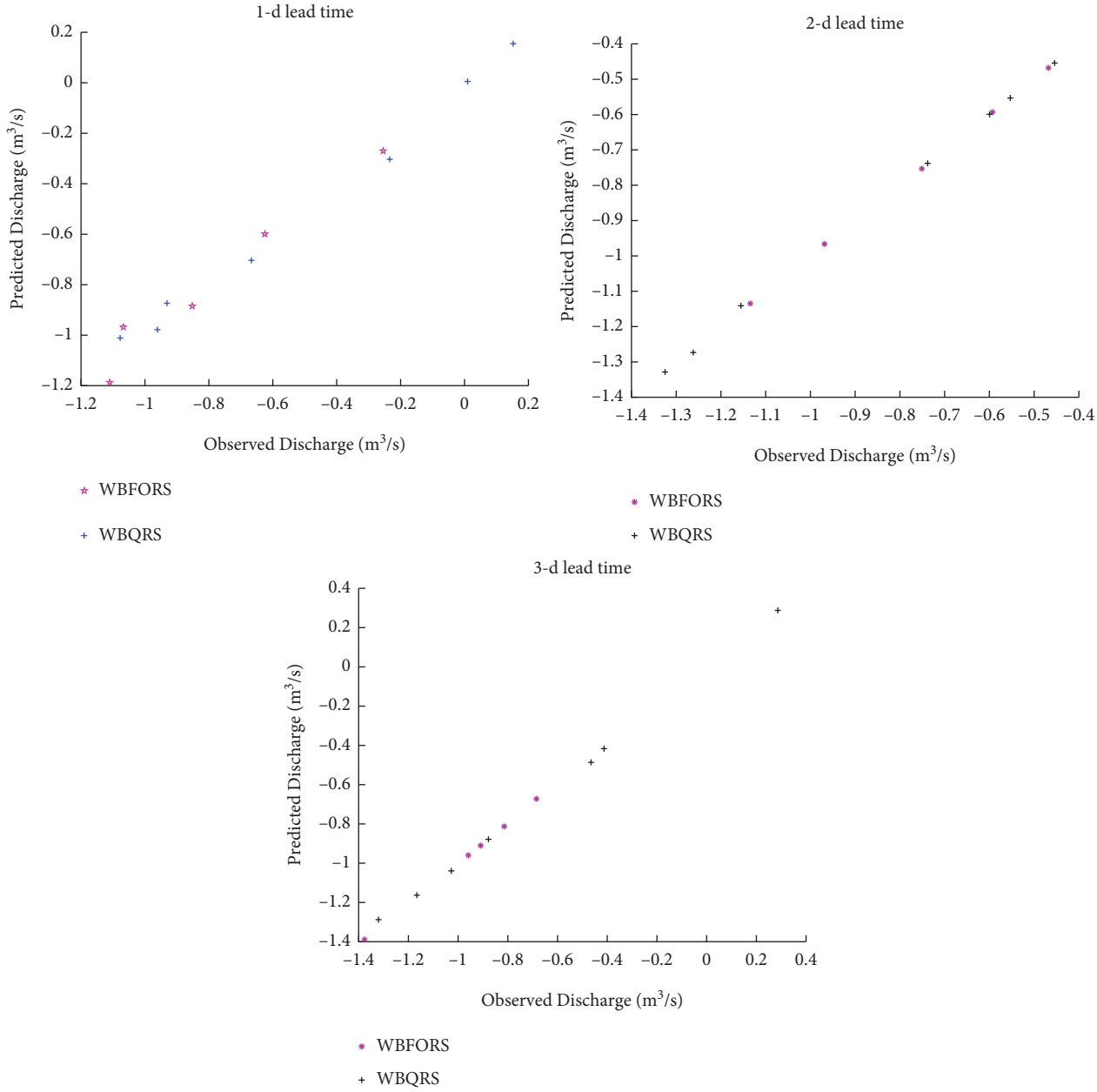

Figure 8: Represents scatter diagram of the observed reservoir flow against the predicted reservoir flow of WBFORS and WBQRS models.

6. Conclusion

Streamflow prediction plays a vital role in hydrology for assessing water patterns. This study explains the water flow prediction using wavelet and bootstrap techniques in the RS model. In this paper, the authors introduce a new hybrid model, WBQRS, for reservoir inflow prediction. To check the validity, the proposed WBQRS model was compared with the MLR, FORS, QRS, WFORS, WQRS, and WBFORS models on different statistical criteria: RMSE, MAE, NSE, and CP. The observed data is decomposed into different frequency components by applying the DWT technique at level third. The mother wavelet function used in this whole scenario is symlet with vanishing moment 15 (sym15). In comparison with other models, WBQRS model depicts the best prediction results with different performance indices for 1-d ahead predictions are RMSE = 0.0356 m^3/s, MAE = 0.0283 m^3/s, NSE = 0.9877, and CP = 0.9957. The results depict that the WBQRS model is a better choice for prediction than the remaining models in the study. In all cases, the models show reliable prediction for 1-3-day prediction. This study will be helpful for authorities to make an early prediction about floods and activate the flood warning response system. This research depicts that the WBQRS model is the most suitable choice for short-term streamflow prediction.

Data Availability

Data are available upon request in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported via funding from Prince Sattam Bin Abdulaziz University project number (PSAU/2023/R/1444).

References

[1] A. Ali, "Assessment of groundwater quality using geographical information system: a case study of faisalabad, Pakistan," *Journal of Environmental and Agricultural Sciences*, vol. 23, no. 2, pp. 30–35, 2021.

[2] A. Ghafoor and R. Nawaz, "Changes in climatic parameters in metropolitan city of lahore, Pakistan," *Journal of Environmental and Agricultural Sciences*, vol. 22, no. 3, pp. 23–33, 2020.

[3] H. Y. Dalkiliç and S. A. Hashimi, "Prediction of daily streamflow using artificial neural networks (ANNs), wavelet neural networks (WNNs), and adaptive neuro-fuzzy inference system (ANFIS) models," *Water Supply*, vol. 20, no. 4, pp. 1396–1408, 2020.

[4] J. Tripura, P. Roy, and A. Barbhuiya, "Simultaneous streamflow forecasting based on hybridized neuro-fuzzy method for a river system," *Neural Computing & Applications*, vol. 33, pp. 3221–3233, 2020.

[5] Ö. Kisi, "Multi-layer perceptrons with Levenberg-Marquardt training algorithm for suspended sediment concentration prediction and estimation/Prévision et estimation de la concentration en matières en suspension avec des perceptrons multi-couches et l'algorithme d'apprentissage de Levenberg-Marquardt," *Hydrological Sciences Journal*, vol. 49, no. 6, Article ID 55720, 2004.

[6] P. S. Yu, S. T. Chen, and I. F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 4, pp. 704–716, 2006.

[7] C. Wu, K. W. Chau, and Y. S. Li, "Methods to improve neural network performance in daily flows prediction," *Journal of Hydrology*, vol. 372, no. 4, pp. 80–93, 2009.

[8] R. R. Sahay and V. Sehgal, "Wavelet-ANFIS models for forecasting monsoon flows: case study for the Gandak River (India)," *Water Resources*, vol. 41, no. 5, pp. 574–582, 2014.

[9] J. Yu, X. Qin, O. Larsen, and L. H. C. Chua, "Comparison between response surface models and artificial neural networks in hydrologic forecasting," *Journal of Hydrologic Engineering*, vol. 19, no. 3, pp. 473–481, 2014.

[10] M. G. Orimi, A. Farid, R. Amiri, and K. Imani, "Cprecip parameter for checking snow entry for forecasting weekly discharge of the Haraz River flow by artificial neural network," *Water Resources*, vol. 42, no. 5, pp. 607–615, 2015.

[11] M. R. Yazdani and A. A. Zolfaghari, "Monthly River forecasting using instance-based learning methods and climatic parameters," *Journal of Hydrologic Engineering*, vol. 22, no. 6, Article ID 4017002, 2017.

[12] M. A. Elganainy and A. E. Eldwer, "Stochastic forecasting models of the monthly streamflow for the blue nile at eldiem station," *Water Resources*, vol. 45, no. 3, pp. 326–337, 2018.

[13] V. Gude, S. Corns, and S. Long, "Flood prediction and uncertainty estimation using deep learning," *Water*, vol. 12, no. 3, p. 884, 2020.

[14] M. Shafaei and O. Kisi, "Predicting river daily flow using wavelet-artificial neural networks based on regression analyses in comparison with artificial neural networks and support vector machine models," *Neural Computing & Applications*, vol. 28, no. S1, pp. 15–28, 2017.

[15] H. Delafrouz, A. Ghaheri, and M. A. Ghorbani, "A novel hybrid neural network based on phase space reconstruction technique for daily river flow prediction," *Soft Computing*, vol. 22, no. 7, pp. 2205–2215, 2018.

[16] A. Danandeh Mehr, E. Kahya, and E. Olyaie, "Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique," *Journal of Hydrology*, vol. 505, pp. 240–249, 2013.

[17] J. Guo, J. Zhou, H. Qin, Q. Zou, and Q. Li, "Monthly streamflow forecasting based on improved support vector machine model," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13073–13081, 2011.

[18] S. N. Londhe and S. Narkhede, "Forecasting stream flow using hybrid neuro-wavelet technique," *ISH Journal of Hydraulic Engineering*, vol. 24, no. 3, pp. 275–284, 2018.

[19] K. F. Roushangar and F. Alizadeh, "Scenario-based prediction of short-term river stage–discharge process using wavelet-EEMD-based relevance vector machine," *Journal of Hydroinformatics*, vol. 21, no. 1, pp. 56–76, 2019.

[20] F. Alizadeh, A. F. Gharamaleki, A. F. Gharamaleki, M. Jalilzadeh, and A. Akhoundzadeh, "Prediction of river stage-discharge process based on a conceptual model using EEMD-WT-LSSVM approach," *Water Resources*, vol. 47, no. 1, pp. 41–53, 2020.

[21] A. Solgi, A. Pourhaghi, R. Bahmani, and H. Zarei, "Preprocessing data using wavelet transform and PCA based on support vector regression and gene expression programming for river flow simulation," *Journal of Earth System Science*, vol. 126, no. 5, p. 65, 2017.

[22] A. Bashir, M. A. Shehzad, A. Khan et al., "Response surface models using the wavelet technique for reservoir inflow prediction," *Mathematical Problems in Engineering*, vol. 2022, Article ID 5171969, 10 pages, 2022.

[23] Y. Seo, S. Kim, and V. P. Singh, "Multistep-ahead flood forecasting using wavelet and data-driven methods," *KSCE Journal of Civil Engineering*, vol. 19, no. 2, pp. 401–417, 2015.

[24] J. Drisya, D. S. Kumar, and T. Roshni, "Hydrological drought assessment through streamflow forecasting using wavelet enabled artificial neural networks," *Environment, Development and Sustainability: A Multidisciplinary Approach to the Theory Practice of Sustainable Development*, vol. 23, pp. 3653–3672, 2020.

[25] Y. Seo, S. Kim, O. Kisi, V. P. Singh, and K. Parasuraman, "River stage forecasting using wavelet packet decomposition and machine learning models," *Water Resources Management*, vol. 30, no. 11, pp. 4011–4035, 2016.

[26] T. Partal, "Wavelet regression and wavelet neural network models for forecasting monthly streamflow," *Journal of Water and Climate Change*, vol. 8, no. 1, pp. 48–61, 2017.

[27] M. M. H. Khan, N. S. Muhammad, and A. El-Shafie, "Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting," *Journal of Hydrology*, vol. 590, Article ID 125380, 2020.

[28] V. Sehgal, M. K. Tiwari, and C. Chatterjee, "Wavelet bootstrap multiple linear regression based hybrid modeling for daily river discharge forecasting," *Water Resources Management*, vol. 28, no. 10, pp. 2793–2811, 2014.

[29] K. Kasiviswanathan, J. He, K. Sudheer, and J. H. Tay, "Potential application of wavelet neural network ensemble to

forecast streamflow for flood management," *Journal of Hydrology*, vol. 536, pp. 161–173, 2016.
[30] B. Keshtegar, M. F. Allawi, H. A. Afan, and A. El-Shafie, "Optimized river streamflow forecasting model utilizing high-order response surface method," *Water Resources Management*, vol. 30, no. 11, pp. 3899–3914, 2016.
[31] D. Nalley, J. Adamowski, and B. Khalil, "Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008)," *Journal of Hydrology*, vol. 475, pp. 204–228, 2012.
[32] Z. Liu, P. Zhou, G. Chen, and L. Guo, "Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting," *Journal of Hydrology*, vol. 519, pp. 2822–2831, 2014.
[33] V. Nourani, A. Hosseini Baghanam, J. Adamowski, and O. Kisi, "Applications of hybrid wavelet–artificial intelligence models in hydrology: a review," *Journal of Hydrology*, vol. 514, pp. 358–377, 2014.
[34] B. Kim, H. Jeong, H. Kim, and B. Han, "Exploring wavelet applications in civil engineering," *KSCE Journal of Civil Engineering*, vol. 21, no. 4, pp. 1076–1086, 2017.
[35] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
[36] M. K. Tiwari and C. Chatterjee, "Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach," *Journal of Hydrology*, vol. 394, no. 4, pp. 458–470, 2010.
[37] O. Kisi, "Wavelet regression model as an alternative to neural networks for river stage forecasting," *Water Resources Management*, vol. 25, no. 2, pp. 579–600, 2011.
[38] M. R. Chernick and R. A. LaBudde, *An Introduction to Bootstrap Methods with Applications to R*, John Wiley & Sons, Hoboken, NJ, USA, 2014.
[39] A. Zhang, H. Shi, T. Li, and X. Fu, "Analysis of the influence of rainfall spatial uncertainty on hydrological simulations using the bootstrap method," *Atmosphere*, vol. 9, no. 2, p. 71, 2018.
[40] L. T. Huang and M. M. Gromiha, "Analysis and prediction of protein folding rates using quadratic response surface models," *Journal of Computational Chemistry*, vol. 29, no. 10, pp. 1675–1683, 2008.
[41] S. Mouatadid, J. F. Adamowski, M. K. Tiwari, and J. M. Quilty, "Coupling the maximum overlap discrete wavelet transform and long short-term memory networks for irrigation flow forecasting," *Agricultural Water Management*, vol. 219, pp. 72–85, 2019.
[42] N. Pramanik, R. K. Panda, and A. Singh, "Daily river flow forecasting using wavelet ANN hybrid models," *Journal of Hydroinformatics*, vol. 13, no. 1, pp. 49–63, 2011.
[43] D. A. Lind, W. G. Marchal, and S. A. Wathen, "Basic statistics for business and econometrics," *Mcgraw-Hill*, 2006.
[44] S. Sharma, *Applied Multivariate Techniques*, John Wiley & Sons, Hoboken, NJ, USA, 1996.
[45] R. Maheswaran and R. Khosa, "Comparative study of different wavelets for hydrologic forecasting," *Computers & Geosciences*, vol. 46, pp. 284–295, 2012.
[46] A. A. Alexander, S. G. Thampi, and C. Nr, "Development of hybrid wavelet-ANN model for hourly flood stage forecasting," *ISH Journal of Hydraulic Engineering*, vol. 24, no. 2, pp. 266–274, 2018.
[47] V. Nourani, M. Komasi, and A. Mano, "A multivariate ANN-wavelet approach for rainfall–runoff modeling," *Water Resources Management*, vol. 23, no. 14, pp. 2877–2894, 2009.
[48] A. Bashir, M. A. Shehzad, I. Hussain, M. I. A. Rehmani, and S. H. Bhatti, "Reservoir inflow prediction by ensembling wavelet and bootstrap techniques to multiple linear regression model," *Water Resources Management*, vol. 33, no. 15, pp. 5121–5136, 2019.