Uncertainty analysis of monthly river flow modeling in consecutive hydrometric stations using integrated data-driven models

Karim Amininia and Seyed Mahdi Saghebian

ABSTRACT

The flow assessment in a river is of vital interest in hydraulic engineering for flood warning and evacuation measures. To operate water structures more efficiently, models that forecast river discharge are desired to be of high precision and certain degree of accuracy. Therefore, in this study, two artificial intelligence models, namely kernel extreme learning machine (KELM) and multivariate adaptive regression splines (MARS), were applied for the monthly river flow (MRF) modeling. For this aim, Mississippi river with three consecutive hydrometric stations was selected as case study. Using the previous MRF values during the period of 1950–2019, several models were developed and tested under two scenarios (i.e. modeling based on station's own data or previous station's data). Wavelet transform (WT) and ensemble empirical mode decomposition (EEMD) as data processing approaches were used for enhancing modeling capability. Obtained results indicated that the integrated models resulted in more accurate outcomes. Data processing enhanced the model's capability up to 25%. It was observed that the previous station's data could be applied successfully for MRF modeling when the station's own data were not available. The best-applied model dependability was assessed via uncertainty analysis, and an allowable degree of uncertainty was found in MRF modeling.

Key words | consecutive stations, EMD, KELM, pre-processing, river discharge

Karim Amininia (corresponding author)
Department of Geography, Ahar Branch,
Islamic Azad University,
Ahar,
Iran
E-mail: *k-amininia@iau-ahar.ac.ir*

Seyed Mahdi Saghebian
Department of Civil Engineering, Ahar Branch,
Islamic Azad University,
Ahar,
Iran

HIGHLIGHTS

- Kernel extreme learning machine (KELM) and multivariate adaptive regression splines (MARS) approaches were used for MRF modeling in three successive hydrometric stations.
- The WT and EEMD as pre-processing methods were used for improving the model's efficiency.
- Monte Carlo uncertainty analysis was applied to investigate the dependability of the applied models.

GRAPHICAL ABSTRACT
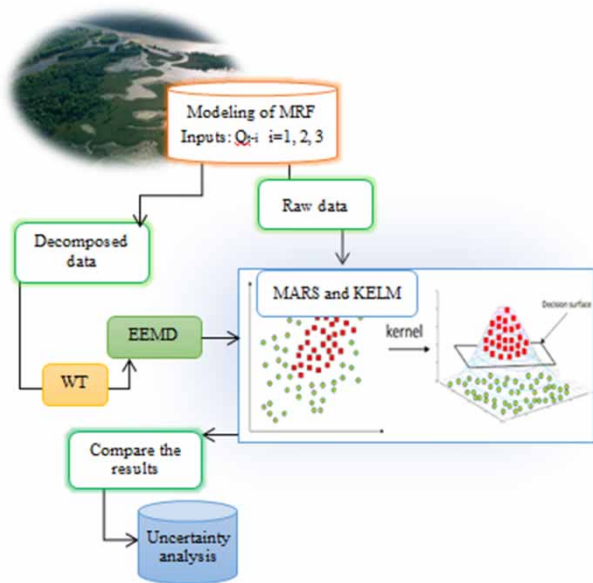
INTRODUCTION

River flow is one of the most important issues in river engineering. Accurate estimation and prediction of river flow is essential to various activities in hydrological and water resources management, such as monitoring pollutant load, calculating sediment transport, controlling flood and drought, determining environmental flows, power generation, reservoir operation and agricultural irrigation, as well as water supply to industry and households. Sufficient flow in a river not only guarantees socioeconomic development but also maintains a healthy river system. River flows, like many other natural processes, are governed by dynamical and nonlinear physical mechanisms or systems which have various different independent variables, primarily precipitation, temperature, evaporation, water use for socioeconomic growth, river channel seepage, as well as forest-vegetation coverage and soil characteristics in the catchment. Therefore, reliable river discharge prediction is particularly important for hydrological operations and hydro-environmental management. A variety of methods have been developed to predict the flow of a river, including time series methods (linear and nonlinear),

conceptual models, physical models, artificial intelligence (AI) models, and models integrating two or more of these approaches (Garg & Jothiprakash 2013). Land use, land cover, topography, soil, precipitation, intensity of rainfall, and other hydrometeorological parameters are required for the conventional models (Talei *et al.* 2010; Yaseen *et al.* 2019). The availability of such dataset for all locations is very difficult. In such a case, the missing parameters are normally assumed for that location or basin, which influences the effectiveness of the model. On the other hand, theoretical models are based on the physical behavior of the water cycle (Evsukoff *et al.* 2012). Several researchers estimate river discharge using the complex relation of several partial differential equations (Koycegiz & Buyukyildiz 2019). These equations are very difficult, and researchers did not find a reliable solution (Solomatine & Dulal 2003). There is high risk of error during solution of these equations. Moreover, for water management, the researcher and scientist look for a quick solution with high precision.

Due to the complexity and nonlinearity of the process and validated models, it is difficult to simulate the accurate

amount of discharge carried by rivers. Therefore, it is necessary to use other methods with higher efficiency. In the past few decades, several AI methods (e.g. artificial neural networks (ANNs), neuro-fuzzy (NF) models, genetic programming (GP), multivariate adaptive regression splines (MARS), support vector machine (SVM), and kernel extreme learning machine (KELM)) have been developed and applied for assessing the complex hydraulic and hydrologic phenomena efficiency. Real-time flood forecasting, longitudinal dispersion coefficients computing in natural streams (Azamathulla & Wu 2011), prediction of rainfall time series (Wu & Chau 2013; Haidar & Verma 2018), forecasting of streamflow discharges (Taormina & Chau 2015), prediction of flow parameters in 90° open-channel (Gholami et al. 2016), side weir discharge coefficient modeling (Azamathulla et al. 2017), predicting bedload in sewer pipes (Roushangar & Ghasempour 2017), daily suspended sediment concentration modeling (Kaveh et al. 2017), scour depth prediction around pier groups (Ebtehaj et al. 2018), flood prediction (Mosavi et al. 2018), flood management (Fotovatikhah et al. 2018), form roughness coefficient modeling in alluvial channels (Saghebian et al. 2020), and predicting Standardized Streamflow Index (SDI; Shamshirband et al. 2020) are some examples.

In AI models, we are looking for a learning machine capable of finding an accurate approximation of a natural phenomenon, as well as expressing it in the form of an interpretable equation. However, this bias toward interpretability creates several new issues. The computer-generated hypotheses should take advantage of the already existing body of knowledge about the domain in question. However, the method by which we express our knowledge and make it available to a learning machine remains rather unclear (Babovic 2009). Machine learning, a branch of AI, deals with the representation and generalization using data learning technique. Representation of data instances and functions evaluated on these instances are part of all machine learning systems. Generalization is the property that the system will perform well on unseen data instances; the conditions under which this can be guaranteed are a key object of study in the subfield of computational learning theory. There is a wide variety of machine learning tasks and successful applications (Mitchell 1997). In general, the task of a machine learning algorithm can be described as

follows: Given a set of input variables and the associated output variable(s), the objective is learning a functional relationship for the input-output variables set. It should be noted that AI models typically do not really represent the physics of a modeled process; they are just devices used to capture relationships between the relevant input and output variables. However, when the interrelationships among the relevant variables are poorly understood, finding the size and shape of the ultimate solution is difficult, and conventional mathematical analysis methods do not (or cannot) provide analytical solutions; these methods can predict the interest variable with more accuracy.

On the other hand, hybrid models based on signal decomposition can be effective in increasing the time series prediction method's efficiency (Pachori et al. 2015). Wavelet analysis is one of the commonly used methods for signal decomposition. The wavelet transform (WT) as signal pre-processing method provides useful information in the temporal and frequency domains for non-stationary signals. Besides the WT, the empirical mode decomposition (EMD) method has been used recently for signal decomposition. This method is suitable for nonlinear and non-stationary time series (Huang et al. 1998). Unlike wavelet decomposition, EMD extracts the data oscillatory mode components without a priori determining the basis functions or level of decomposition (Labate et al. 2013).

Since river flow's accurate prediction in watersheds has a significant impact on efficient management of water projects, in the current study, the KELM and MARS approaches were used for monthly river flow (MRF) modeling in three successive hydrometric stations. KELM as a kernel-based approach is a relatively new important method based on the different kernels type which is based on statistical learning theory initiated. Such model is capable of adapting itself to predict any variable of interest via sufficient inputs. This method can model nonlinear decision boundaries, and there are many kernels to choose from. A kernel-based approach is also fairly robust against overfitting, especially in high-dimensional space. Also, MARS is a non-parametric method and determines a nonlinear relationship between dependent and independent variables through adjunction of several linear models. The training of this method is fast, has high accuracy, and the probability of occurrence of data overtraining in this method, is less.

Also, the WT and ensemble EMD (EEMD) were used as data processing methods to improve the model's efficiency. In integrated models, the inputs data were decomposed into subseries by WT and EEMD. Then, the most effective subseries were used as inputs in KELM and MARS methods. MRF data of the Mississippi river in the period of 1950–2019 were used for this aim. Various models were developed using only the previous values of river flow time series. Two scenarios were considered in the modeling process. In the first scenario, the intended parameter of each station was estimated using the station's own data, and in the second scenario, the MRF was estimated using the previous station's data. In fact, in the second scenario, the impacts of the existing sub-basins between the consecutive stations on the flow regime of the downstream station were assessed. Also, Monte Carlo uncertainty analysis was applied to investigate the dependability of the applied models.

MATERIALS AND METHODS

Study area

The Mississippi river is the second longest river and chief river of the second-largest drainage system on the North American continent. From its traditional source of Lake Itasca in northern Minnesota, it flows generally south for 3,730 km to the Mississippi River Delta in the Gulf of Mexico. In the current study, river flow data of three consecutive stations, namely station 1 (7010000), station 2 (7020500), and station 3 (7022000), were selected and MRF was investigated under two scenarios. In the first scenario, modeling was done based on each station's data, and in the second scenario, the previous stations' data were used. Figure 1 shows the location of the selected stations and their MRF. To compare the performance of applied models, the total data were divided into three sets: the training, validation, and testing sets. The first 70% of the whole data were used for training the models, and the last 30% of data were used for validating and testing the models (15% for validating and 15% for testing). The training set trains the scheme on the basis of a minimization criterion, and the validation set is used as a stopping criterion for training to avoid overfitting to the data. The testing set is used to evaluate the generated model and assess its generalization capability (Kitsikoudis et al. 2015).

Pre-processing approaches

One of the most popular approaches in time series processing is the WT (Farajzadeh & Alizadeh 2017). The WT uses flexible window function (mother wavelet) in signal processing. The flexible window function can be changed over time according to the signal shape and compactness (Mehr et al. 2013). After using WT, the signal will decompose into two approximation (large-scale or low-frequency component) and detailed (small-scale component) components. The other approach for time series processing is EMD. The EMD method is an effective self-adaptive dyadic filter bank which applied to the white noise. By applying this method, each signal can be decomposed into a number of inherent mode functions (IMFs) which can be used to process nonlinear and non-stationary signals. One of the advantages of this method is the ability to determine the instantaneous frequency of the signal. At each step of the signal decomposition into its frequency components, the high-frequency components are separated first and this process must continue until the component with the lowest frequency remains (see Lei et al. (2009) for more details). EEMD is developed based on EMD. The main benefit of EEMD is solving the mode mixing problem of EMD which determines the true IMF as the mean of an ensemble of trials (Wu & Huang 2009). The EEMD algorithm can be described as: (1) for a given signal $x(t)$, random white noise is added to the signal, (2) the noise-added signal is decomposed using EMD for obtaining IMF series, (3) steps 1 and 2 are repeated until the number of added white noises be greater than or equal to the number of trials, (4) for obtaining the ensemble IMF, the average of the sum of all IMFs is computed ($I_j(t)$), and (5) the original signal is formed as $x(t) = \sum_{j=1}^{n} I_j(t)$.

Data-driven methods

Two data-driven techniques, namely KELM and MARS, were used to predict MRF. In the following, a description of these methods is given.
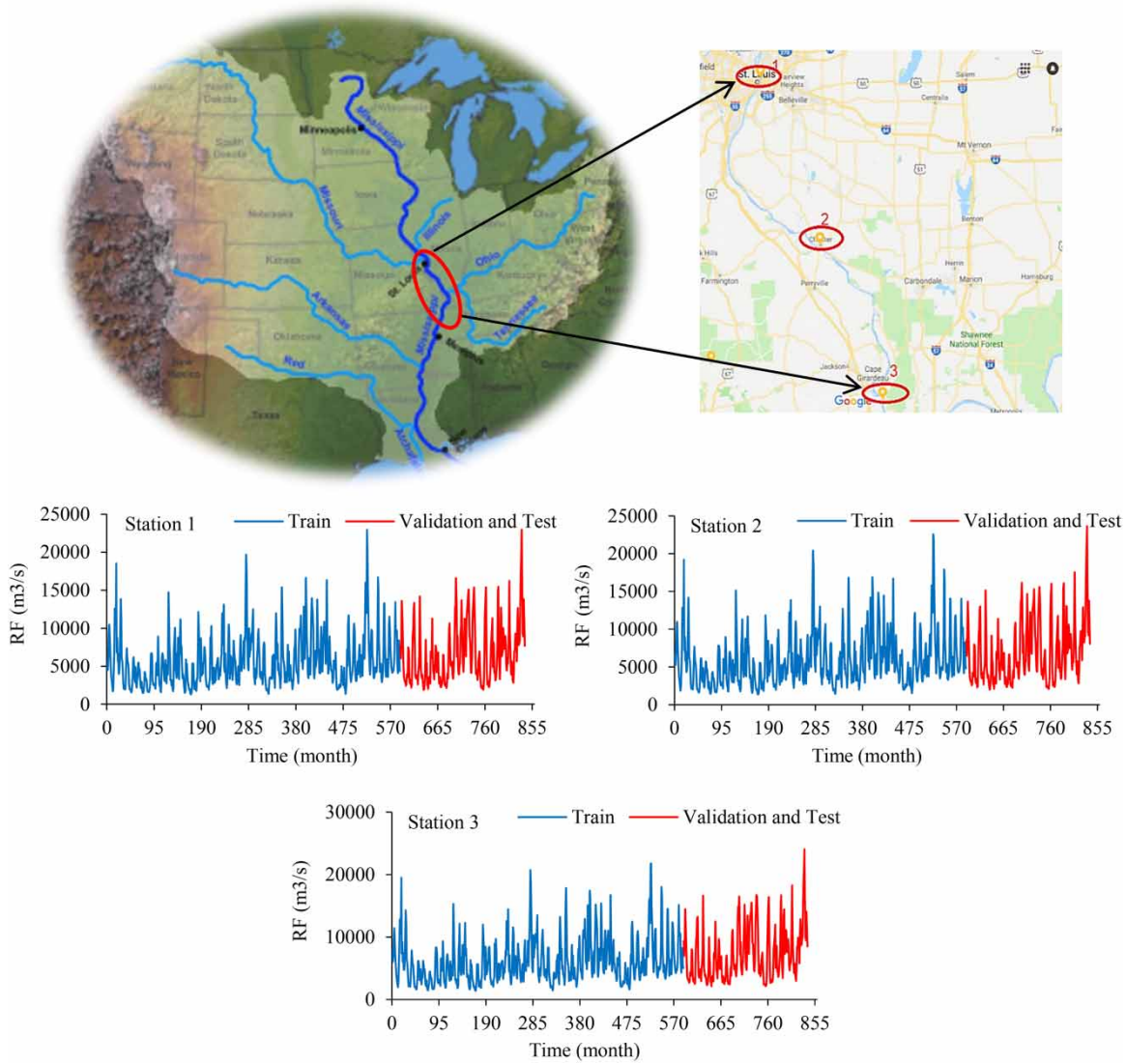
Figure 1 │ The location of selected consecutive stations of Mississippi river and their MRF.

Kernel extreme learning machine

The ELM method is a single-layer feed-forward neural network (SLFFNN) preparing initially introduced by Huang et al. (2006). SLFFNN is a straight framework where information weights linked to hidden neurons and hidden layer biases are haphazardly chosen, while the weights among the hidden nodes are resolved logically. This strategy likewise has preferred execution and adapts progressively over the bygone era learning methods (Huang et al. 2006), in light of the fact that not at all like traditional techniques that involve numerous variables to setup, demonstrating a complex issue utilizing this technique does not need much human intercession to accomplish ideal parameters (Ebtehaj et al. 2018). The standard single-layer neural system with N random information (a_i, b_i), M hidden neurons, and the active function $f(a)$ are shown as follows:

$$a_i = [a_{i1}, a_{i2}, \ldots, a_{in}]^T \in R^n \tag{1}$$

$$b_i = [b_{i1}, b_{i2}, \ldots, b_{in}]^T \in R^m \tag{2}$$

$$\sum_{i=1}^{N} \alpha_i f(w_i \alpha_j + c_i) = X_j \quad j = 1, 2, \ldots, N \tag{3}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector that joins the input layer to the hidden layer, and $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in}]^T$ is the weight vector that joins hidden layer to the target layer. c_i shows the hidden neuron biases. The general SLFFNN network with the M hidden neuron and the activation function $f(a)$ can predict N information with an average zero error $\left(\sum_{i=1}^{N} \|X_j - X_i\| = 0\right)$, in which:

$$\sum_{i=1}^{N} \alpha_i f(w_i \alpha_j + c_i) = Y_j \quad j = 1, 2, \ldots, N \tag{4}$$

Equations (4) can be summarized as:

$$A\alpha = K \tag{5}$$

$$A(w_1, \ldots, w_M, c_1, \ldots, c_M, a_1, \ldots, a_M)$$
$$= \begin{bmatrix} f(w_1 a_1 + c_1) \cdots f(w_N a_1 + c_M) \\ \ldots \\ f(w_1 a_N + c_1) \cdots f(w_M a_N + c_M) \end{bmatrix}_{N \times M} \tag{6}$$

$$K = \begin{bmatrix} p_1^T \\ . \\ . \\ p_N^T \end{bmatrix}_{N \times L}, \quad \alpha = \begin{bmatrix} \alpha_1^T \\ . \\ . \\ \alpha_N^T \end{bmatrix}_{M \times L} \tag{7}$$

The matrix K is identified as the target matrix of the hidden layers of the neural network. Huang *et al.* (2012) also introduced kernel functions in the design of ELM. Now number of kernel functions is used in the design of ELM such as linear, radial basis, normalized polynomial, and polynomial kernel functions. Kernel function-based ELM design is named as kernel extreme learning machine. For more detail about KELM, readers and researchers are referred to Huang *et al.* (2012).

Multivariate adaptive regression splines

MARS as a non-parametric method determines a nonlinear relationship between dependent and independent variables through adjunction of several linear models (Sharda *et al.* 2008). The MARS approach modeling process includes several forward and backward steps. The selection of appropriate independent variables is done at forward step; however, in backward stepwise, unnecessary variables are removed to prevent model over-fitting and enhance the

efficiency of the model (Sharda *et al.* 2008). The MARS approach is based on basis functions in which variable X is projected to variable Y through applying one of the following basis functions:

$$Y = max(0, X - m), \quad Y = max(0, m - X) \tag{8}$$

where m is a threshold value. In the MARS model, the desired basis function numbers are determined at the first step. The selected basis function numbers are applied in the model in forward phase, while in the backward stepwise, the model is simplified by removing the less important basis functions.

Performance criteria

In this study, three criteria were used to assess the applied model's capability including determination coefficient (DC), root-mean-square errors (RMSEs), and mean absolute percentage error (MAPE). These statistical criteria are formulated as:

$$DC = 1 - \frac{\sum_{i=1}^{N} (l_o - l_p)^2}{\sum_{i=1}^{N} (l_o - \overline{l_p})^2}$$

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(l_o - l_p)^2}{N}} \tag{9}$$

$$MAPE = \frac{l}{\overline{l_o}} \frac{\sum_{i=1}^{n} |l_o - l_p|}{N} \times 100$$

where l_o, l_p, $\overline{l_o}$, $\overline{l_p}$ are the observed, estimated, mean observed, mean estimated values, respectively. N is the number of data. The DC measures the fraction of the variance in the data explained by the model. The RMSE describes the average difference between predicted and measured values corresponding to the predicted values, and the MAPE indicates the absolute differences between estimated and observed values as depicted in Equation (9). These measurements are not oversensitive to extreme values (outliers), but are rather sensitive to additive and proportional differences between model predictions and

observations. Therefore, correlation-based measures (e.g. DC) can indicate that a model is a good predictor (Legates & McCabe 1999). Evidently, a high value for DC (up to one) and a small value for RMSE and MAPE indicate a high efficiency of the model. It should be noted that in this study, all input variables were scaled between 0 and 1 for eliminating the input and output variable dimensions.

Uncertainty analysis

The aim of a model uncertainty analysis is to determine the statistical characteristics of the outputs of that model as a function of the uncertainty of the input parameters (Noori *et al.* 2015). Uncertainty is a factor associated with the estimation result which determines the estimation values range. Its value indicates the level of confidence that the actual measured value falls within the specified range (Noori *et al.* 2015). In the current research, the Monte Carlo method proposed by Abbaspour *et al.* (2007) was used to evaluate the uncertainty of the AI models in the MRF series modeling. To verify model results uncertainties, 95% confidence interval (95PPU) and bandwidth factor (*d*-factor) which is the average distance between the upper (XU) and lower (XL) uncertainty bands should be used (Noori *et al.* 2015). In this regard, the considered model should be developed many times (1,000 times in this research), and the empirical cumulative distribution probability of the models should be calculated.

The appropriate confidence limits are mostly from measured data within the width of 95PPU and have a reasonable average width (*d*-factor → 0) (Abbaspour *et al.* 2007). For evaluating the mean width of the confidence band, Abbaspour *et al.* (2007) suggested the below equation:

$$d\text{–factor} = \frac{\overline{dx}}{\sigma x} \qquad (10)$$

where σx and $\overline{dx}$ are the observed data standard deviation and the confidence band's average width, respectively. The percentage of the data within the confidence band of 95% is calculated as:

$$\text{Bracketed by 95PPU} = \frac{1}{k}Cont(j|X_L < X_{reg} < X_U) \qquad (11)$$

where 95PPU shows 95% predicted uncertainty, k shows the number of observed data, and X_{reg} shows the current registered data.

Model development

Appropriate selection of input combination has a significant impact on the accuracy of developed models. River flows can be affected by various different independent variables such as primarily precipitation, temperature, evaporation, river channel seepage, as well as soil characteristics in the catchment. The use of all mentioned dataset for all locations is very difficult. Therefore, according to Table 1, in this research, only monthly previous values of river flow time series over the period of 1950–2019 were used as inputs to predict the next month river flow value. In fact, it was tried to achieve accurate results by developing several simple models. In the modeling process, two states were considered: in the first state, the selected stations' river flow was predicted based on the station's own data, and in the second state, modeling was done based on the data from previous stations. For all defined combinations, t demonstrates the time step (monthly step), and $Q(t-1)$ and $Q(t-2)$ represent river flow values at time $(t-1)$ and $(t-2)$, respectively. The output is the river flow in the current time step (i.e. MRF(t)). The WT and EEMD were

Table 1 | Considered models in the study

Model	Inputs		
Scenario 1			
Station 1, 2, and 3			
I	$Q(t-1)$		
II	$Q(t-1)$, $Qw(t-2)$		
III	$Q(t-1)$, $Q(t-2)$, $Q(t-3)$		
Scenario 2			
Station 2		Station 3	
IV	$Q1(t)$	VII	$Q1(t)$, $Q2(t)$
V	$Q1(t)$, $Q1(t-1)$	VIII	$Q1(t-1)$, $Q2(t-1)$
VI	$Q1(t)$, $Q1(t-1)$, $Q1(t-2)$	IX	$Q2(t)$, $Q2(t-1)$

applied to decompose river flow series. WT is a very useful technique for time series data processing of many aspects (Ercelebi 2004). WT is suitable for analyzing time series data since it is an effective method for time series data reduction. WT can detect sudden signal changes well because it transforms an original time series data into two types of wavelet coefficients: approximation and detail. Approximation wavelet (low-frequency) coefficients capture rough features that estimate the original data, while detail wavelet (high-frequency) coefficients capture detail features that describe frequent movements of the data. WT is useful in supporting multiresolution analysis. In addition to projecting a time series into approximation and detail wavelet coefficients, WT decomposes these coefficients into various scales (Chaovalit et al. 2011).

The minimum decomposition level (L) in the WT method can be obtained as following:

$$L = \text{int } [\log N] \tag{12}$$

In this study, $L = 3$ was used for data decomposition. Also, the EEMD method was used for the additional decomposing of the high-frequency time series decomposed by WT due to the nonlinearity and non-stationarity of MRF series. Besides the WT, EEMD is another efficient noise reduction technique, which enables adaptive decomposition based only on signal characteristics. EEMD and WT donated incredible vision in time and frequency domain of data to capture the nonlinear and seasonal properties. The EMD represents an adaptive method to recognize oscillations from the signal. Analogously to Discrete Wavelet Transform (DWT), EEMD describes signal via IMFs as details (high-frequency oscillations), and residual as approximation (low-frequency oscillations) from the last decomposition level. This new method is based on the insight from recent studies of the statistical properties of white noise, which showed that the EMD method is an effective self-adaptive dyadic filter bank when applied to the white noise (Wu & Huang 2004). On the other hand, studies demonstrated that noise could help data analysis in the EMD method. All these investigations promote the advent of the EEMD method. An attempt was done to choose the most dominant subseries

based on their correlation with original time series. Figure 2 shows the considered modeling process.

RESULTS AND DISCUSSION

The results of kernel-based approaches

The capability of kernel-based approaches was assessed in predicting the river flow using the main data (i.e. without processing). It should be noted that each AI method has its own parameters which for achieving the desired results, the optimized amount of these parameters should be determined. For example, in designing the kernel-based approaches, the selection of appropriate type of kernel function is needed. A number of kernels are discussed in the literature (Gill et al. 2006). In this research, river flow was predicted using different kernel types. Therefore, the model (III) was run via KELM and according to Figure 3, the radial basis functions (RBFs) kernel was found as the best kernel function. According to Noori et al. (2011), the RBF kernel is very desirable to be used in prediction of hydraulic and hydrological features since: (1) unlike the linear kernel, the RBF kernel can handle the case when the relation between class labels and attributes is nonlinear.
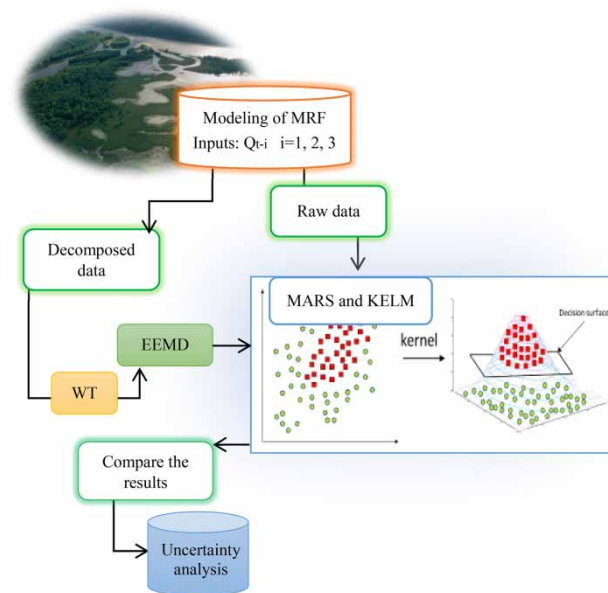

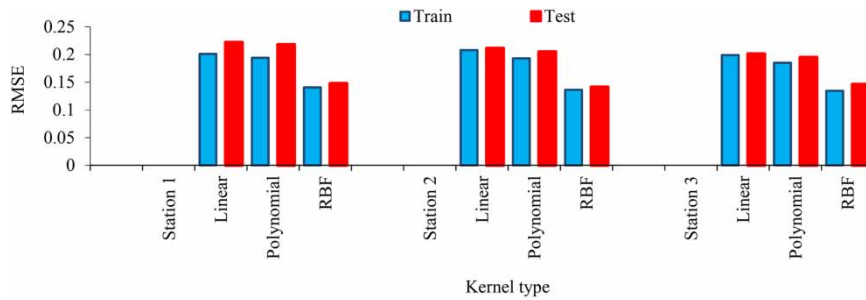**Figure 2** | Considered modeling process in the study.

Figure 3 | Performance of the KELM method with different kernel functions for the model (III).

(2) The RBF kernels tend to give better performance under general smoothness assumptions. (3) It has fewer tuning parameters than the polynomial and the sigmoid kernels. Therefore, RBF kernel was used as the core tool of KELM. Also, in the MARS model, maximum number of basis function was set as 25, and the best model performance is obtained with 10 basis functions.

It should be noted that monthly river flows can be affected by various hydrometeorological parameters. However, the use of all them for all cases is not possible. In this research, it was assumed that by using only previous values of river flow time series, an accurate prediction of MRF could be obtained. Therefore, several simple models were developed based on only monthly previous values of river flow time series. The selection of input combinations was done based on train and error process, and the most accurate combinations were selected. The obtained results from data-driven single models are listed in Table 2. As it could be seen, the used models slightly showed similar results in the selected stations. Among the used models, the KELM model yielded to better predictions. According to Li *et al.* (2020), the simulation performance can be affected by the size and distribution characteristics of the streamflow. As the distribution range becomes wider, the prediction performance gradually becomes unstable. Considering the obtained statistical indicators and according to Figure 4, it could be stated that the MARS and KELM models did not lead to acceptable efficiency in modeling the maximum and minimum amounts of river flow. From the obtained results, it could be stated that in the first state, the model (III) with input parameters of $Q(t-1)$, $Q(t-2)$, $Q(t-3)$ performed better than the others. Based on the results, it could be seen that adding $Q(t-2)$, $Q(t-3)$ parameters to input

combinations increased the model's accuracy. In the second state, for the station 2, the model (VI) with parameters of $Q1(t)$, $Q1(t-1)$, $Q1(t-2)$ was selected as the superior model and for the station 3, the model (VII) with parameters of $Q1(t)$, $Q2(t)$ led to better predictions. Figure 4 illustrates the scatter plots between observed versus predicted river flow for the KELM model.

Data processing impacts on model's efficiency

In this section, the effect of pre-processing of time series on increasing of model's accuracy was investigated. Therefore, the time series were decomposed using the WT method. To decompose the time series by WT, a mother wavelet which is more similar to the signal should be selected. There are different types of mother wavelets such as Daubechies (db), Haar, Morlet, Symlets (sym), Coifets, and Meyer. According to Roushangar *et al.* (2020), in hydrological processes, the db and sym mother wavelets perform better than others. In this study, the daubechies (db2, db4) and Symlets (sym2, sym4) mother wavelets were trained for decomposition of $Q(t-1)$ in station 1. According to Figure 5(a), it was found that the db4 mother wavelet led to better outcomes. Therefore, db4 mother wavelet and decomposition level 3 were used for time series decomposition. Four subseries (one approximation and three detailed) were obtained for each time series. In the second step, two first detail subseries (i.e. details 1 and 2) were further decomposed via EEMD. The principle of EEMD is decomposition of signal to different IMFs and one residual signal. The sum of these signals will be the same original signal. The formation of IMFs is based on subtracting the basic function from the original signal. This process

Table 2 │ The results obtained for the scenario 1 in the MRF modeling

Method	Model	Validation			Test			Validation			Test		
		DC	RMSE	MAPE	DC	RMSE	MAPE	DC	RMSE	MAPE	DC	RMSE	MAPE
State 1													
		Station 1						Station 2					
MARS	I	0.512	0.161	20.15	0.413	0.173	21.37	0.484	0.150	18.5	0.420	0.159	19.62
	II	0.557	0.155	19.51	0.446	0.169	20.87	0.512	0.146	18.125	0.445	0.153	18.87
	III	0.576	0.148	18.37	0.498	0.164	20.25	0.539	0.143	17.75	0.469	0.150	18.5
KELM	I	0.523	0.153	18.87	0.415	0.166	20.54	0.494	0.147	18.05	0.423	0.155	19.42
	II	0.568	0.146	17.6	0.449	0.162	20.05	0.523	0.138	17.125	0.448	0.147	18.25
	III	**0.588**	**0.140**	**15.8**	**0.521**	**0.158**	**19.5**	**0.550**	**0.135**	**16.75**	**0.481**	**0.142**	**17.62**
		Station 3											
MARS	I	0.537	0.148	18.37	0.462	0.166	20.52						
	II	0.569	0.145	18	0.489	0.162	20.01						
	III	0.599	0.141	17.5	0.513	0.158	19.5						
KELM	I	0.554	0.138	17.12	0.476	0.157	19.37						
	II	0.587	0.135	16.75	0.503	0.152	18.75						
	III	**0.618**	**0.132**	**16.37**	**0.530**	**0.141**	**17.5**						
State 2													
		Station 2						Station 3					
MARS	IV	0.720	0.090	11.12	0.703	0.097	12	VII	0.748	0.071	8.75	0.705	0.080	9.29
	V	0.743	0.079	9.75	0.722	0.086	10.62	VIII	0.715	0.119	14.75	0.616	0.135	17.76
	VI	0.741	0.082	10.12	0.729	0.084	10.37	IX	0.510	0.147	18.25	0.288	0.173	22.11
KELM	IV	0.782	0.071	8.75	0.753	0.074	9.12	**VII**	**0.758**	**0.062**	**7.62**	**0.735**	**0.064**	**9.41**
	V	0.788	0.061	7.59	0.767	0.069	8.54	VIII	0.744	0.119	14.75	0.647	0.126	16.7
	VI	**0.797**	**0.062**	**7.62**	**0.785**	**0.066**	**8.12**	IX	0.531	0.137	17.08	0.308	0.162	20.82

Bold values show the superior model in each state.

continues until the residual signal remains almost constant. In this study, details 1 and 2 subseries were decomposed into nine IMFs and one residual signal. Further decomposition using EEMD leads to more stationary and with less noise subseries. In Figure 5(b), the $Q(t-1)$ decomposed subseries by db4-EEMD was shown for station 1. Since the number of input data increased after time series decomposition, the correlations of subseries with original signal were calculated. Figure 5(c) shows the correlations among the IMFs subseries and MRF series. Each subseries which its correlation was more than the average value of correlations of all subseries was taken as significant out of others. The selected subseries were used as inputs in data-driven methods to predict the MRF.

The results of the integrated pre-processing models are listed in Table 3. According to the results presented in Tables 2 and 3, it could be induced that data pre-processing significantly improved the results' accuracy and integrated models were more accurate than data-driven methods. In fact, the use of WT and further decomposition of the detailed series led to an improvement in the outcomes. For example, for station 3 in scenario 1, the RMSE values of the MARS and KELM models for the model (III) were 0.158 and 0.141, respectively; the values of RMSE for the WT-EEMD-MARS and WT-EEMD-KELM models decreased to 0.091 and 0.087, respectively. It could be seen that using the integrated models, the river flow modeling was done with higher accuracy and the applied methods
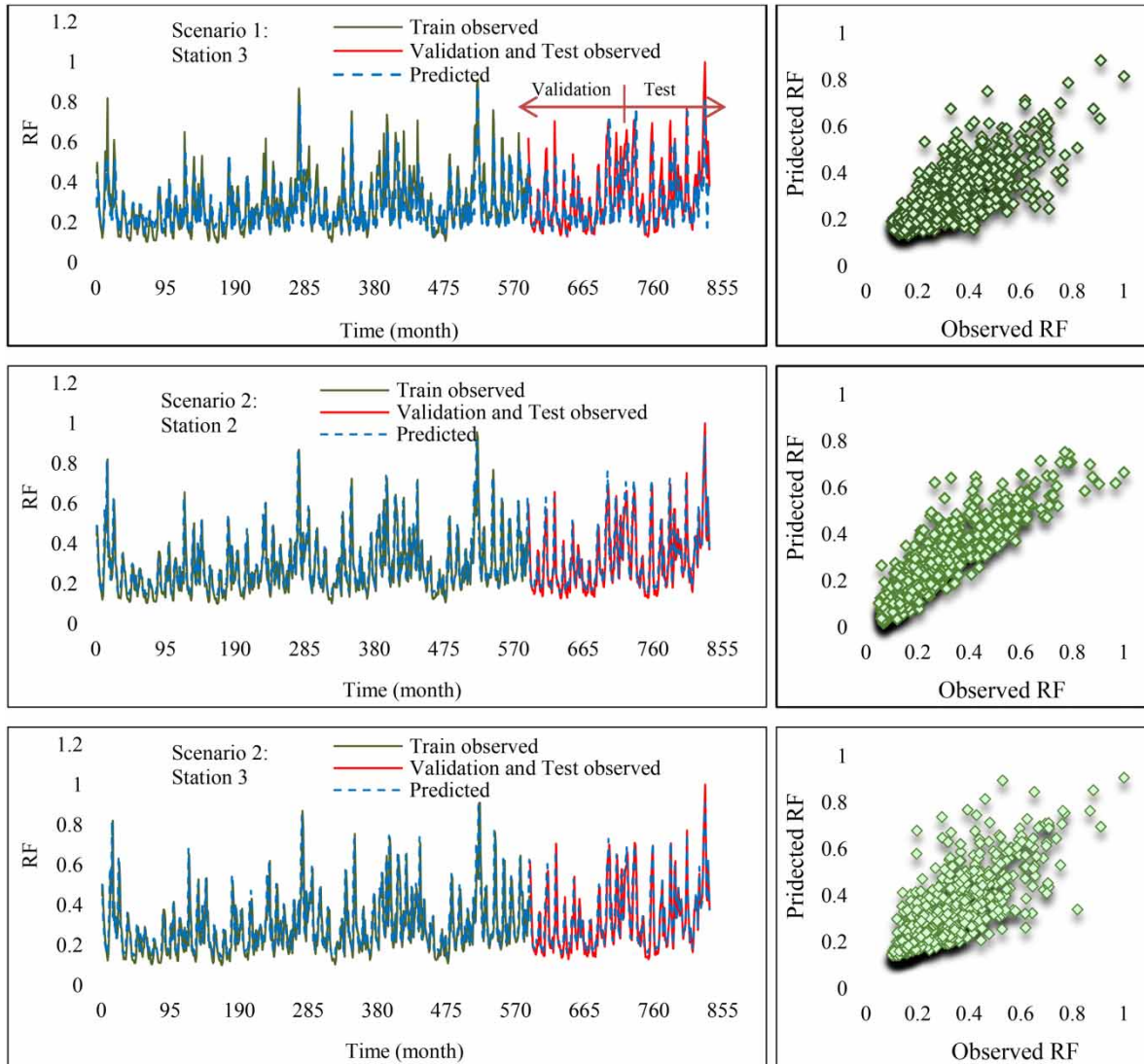
Figure 4 | Scatterplot of MRF modeling via KELM superior models.

were successful in the modeling process. Alizadeh *et al.* (2018) tried to design, construct, and evaluate the efficiency of the wavelet-ANN model for flow time series forecasting. They showed that the integrated models (wavelet-ANN) provide acceptable predictions of the monthly flow discharge. It was found that the WT is a powerful tool which has a great ability to extract useful information from time series. Consequently, it increases the ANN models' performances significantly. However, in this study, the MRF time series decomposed to various periodicity scales with WT and further decomposition of subseries using EEMD to obtain more stationary subseries. In general, the data pre-

processing by WT-EEMD increased the model's accuracy approximately up to 25%.

From the results, it could be seen that the integrated methods resulted in desirable accuracy in both scenarios and the use of previous station's data could be applied as a reliable scenario to predict river flow values for the stations with the lack of the observational data. Also, it could be seen that the use of upstream flow models caused a significant improvement in the performance levels. This issue indicated that the existing sub-basins between the consecutive stations may have noticeable physical impacts (i.e. increasing drainage area) on the flow regime of the
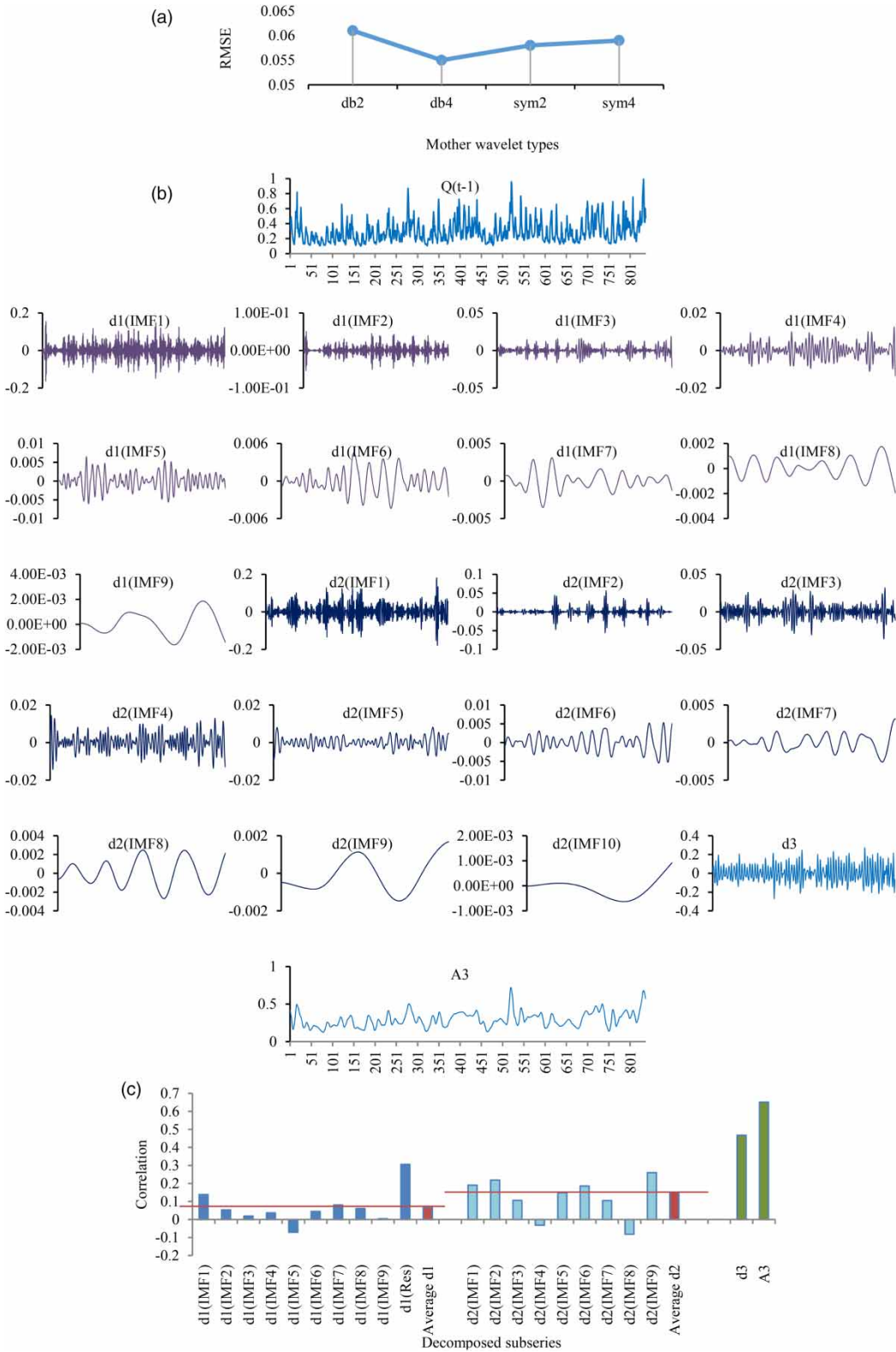
Figure 5 | (a) The RMSE values for decomposed $Q(t-1)$ using different mother wavelets for the station 1, (b) decomposed $Q(t-1)$ time series using db4-EEMD for the station 2, and (c) correlation between detail subseries and original MRF input series.

Table 3 | The results obtained for the scenario 2 in the MRF modeling

Method	Model	Validation			Test			Validation			Test		
		DC	RMSE	MAPE	DC	RMSE	MAPE	DC	RMSE	MAPE	DC	RMSE	MAPE
State 1													
		Station 1						Station 2					
MARS	I	0.743	0.088	10.87	0.708	0.102	12.625	0.729	0.081	10	0.697	0.089	11
	II	0.808	0.085	10.51	0.743	0.097	12	0.772	0.080	9.87	0.718	0.085	10.52
	III	0.834	0.081	10	0.831	0.094	11.62	0.812	0.078	9.62	0.749	0.083	10.21
KELM	I	0.773	0.083	10.2	0.730	0.095	11.75	0.758	0.077	9.5	0.724	0.085	10.53
	II	0.841	0.080	9.87	0.763	0.093	11.5	0.803	0.075	9.25	0.745	0.081	10
	III	**0.873**	**0.076**	**9.35**	**0.854**	**0.090**	**11.12**	**0.845**	**0.074**	**9.12**	**0.814**	**0.078**	**9.62**
		Station 3											
MARS	I	0.817	0.079	9.75	0.708	0.099	14.25						
	II	0.865	0.078	9.62	0.745	0.097	13.08						
	III	0.910	0.076	9.37	0.791	0.091	12.92						
KELM	I	0.859	0.074	9.12	0.739	0.093	11.5						
	II	0.897	0.072	8.87	0.786	0.090	12.14						
	III	**0.913**	**0.071**	**8.75**	**0.833**	**0.087**	**10.87**						
State 2													
		Station 2						Station 3					
MARS	IV	0.809	0.078	9.625	0.790	0.084	10.37	VII	0.885	0.057	7.12	0.835	0.063	8.85
	V	0.834	0.069	8.51	0.810	0.075	9.25	VIII	0.846	0.097	12.84	0.729	0.106	15.08
	VI	0.832	0.071	8.75	0.819	0.074	9.12	IX	0.603	0.120	14.88	0.502	0.135	19.14
KELM	IV	0.881	0.062	7.62	0.859	0.065	8.08	**VII**	**0.897**	**0.050**	**6.25**	**0.890**	**0.050**	**7.143**
	V	0.906	0.051	6.37	0.881	0.063	7.75	VIII	0.881	0.094	12.02	0.764	0.101	14.28
	VI	**0.902**	**0.053**	**6.62**	**0.893**	**0.060**	**7.37**	IX	0.628	0.112	13.87	0.518	0.126	17.85

Bold values show the superior model in each state.

downstream station. According to Figure 6, it could be seen that the extreme values in the river flow series calculated more correctly by the data processing methods.

Uncertainty analysis results

In this part of the study, the uncertainty analysis was done in order to find the uncertainty of the integrated superior model (i.e. WT-EEMD-KELM). In the proper confidence level, two important indices should be considered. First, the 95PPU band brackets most of the observations, and second, the d-factor should be smaller than the standard deviation of the observed data. The two mentioned indices were applied for accounting input uncertainties. The

obtained results for test series are listed in Table 4 and shown in Figure 7. Based on the values obtained for the d-factor and 95PPU, it could be indicated that for the integrated model, most of the observed and predicted values were within the 95PPU band for MRF predictions. For example, in station 1, the amount of d-factor was 0.125 which was smaller than the standard deviation of the observed data (i.e. 0.142). Also, for verifying the uncertainty analysis results, the lower bound and upper bound estimation (LUBE) method was used. The upper bound and the lower bound are calculated according to the forecasting value and the confidence level. The accuracy of the point forecasting has played a key in the obtained results accuracy. The LUBE tries to directly approximate upper and lower
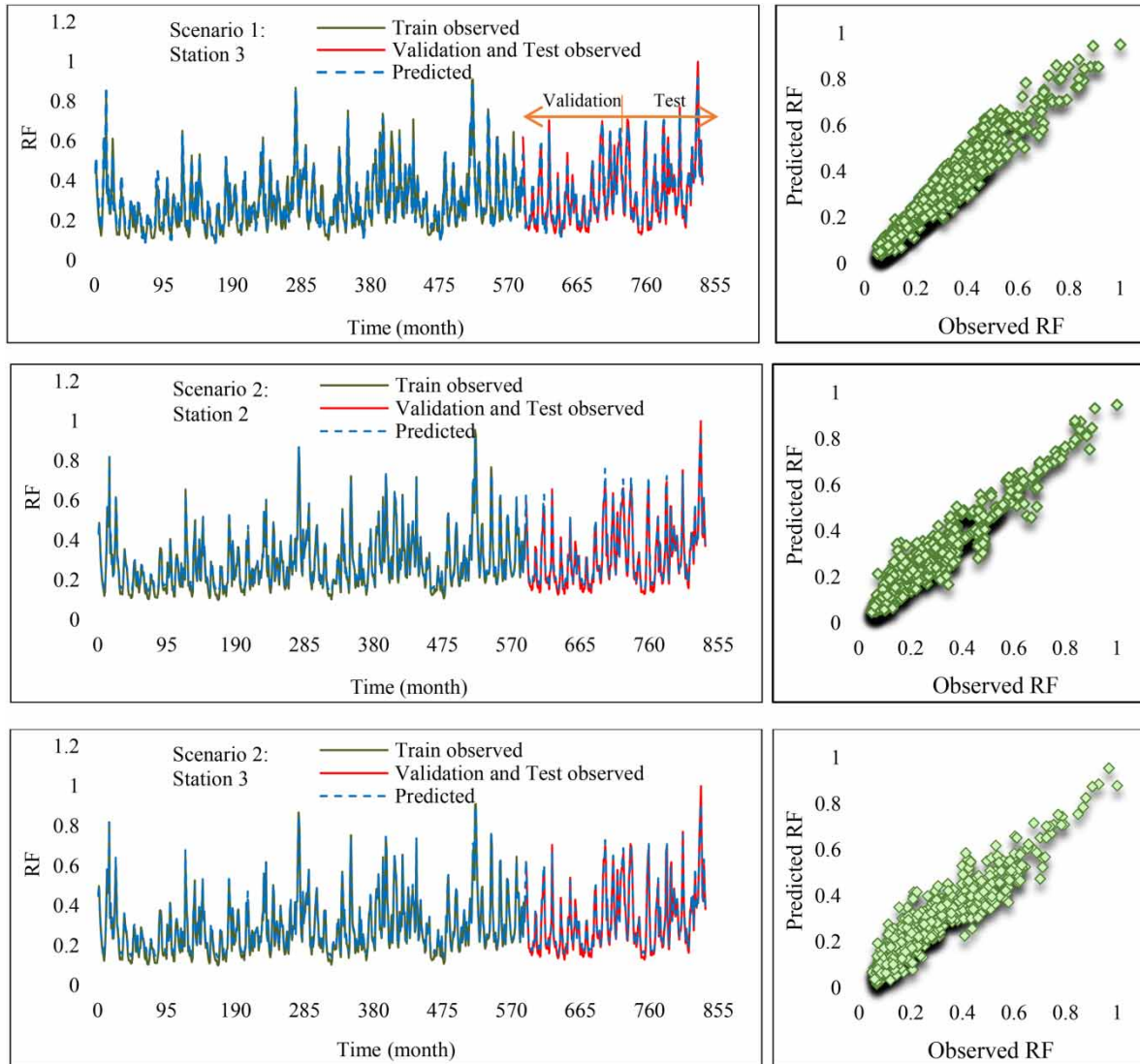
Figure 6 | Scatterplot of MRF modeling via WT-EEMD-KELM superior models.

Table 4 | Uncertainty indices of the WT-EEMD-KELM model

Station	Performance criteria		Station	Performance criteria	
	95PPU	d-factor		95PPU	d-factor
Monte Carlo method					
1	77%	0.125	3	89%	0.165
2	84%	0.162			
LUBE method					
1	75%	0.127	3	86%	0.169
2	81%	0.169			

bounds of prediction intervals (PIs) which is prevalent measure for the assessment of prediction/forecast uncertainty. PIs deal with the uncertainty in the prediction of a future realization of a random variable and account for more sources of the uncertainty (model misspecification and noise variance). As it is seen from Table 4, for all cases, most of the observed values were within the 95PPU band and d-factor values were smaller than the standard deviation of the observed data. In general and according to the results, it could be induced that the MRF modeling via WT-EEMD-KELM model led to an allowable degree of uncertainty.
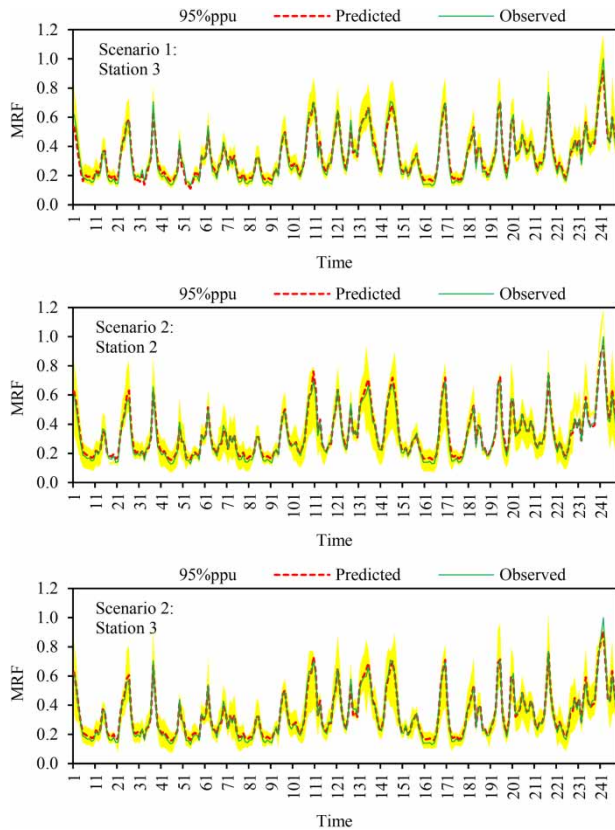
Figure 7 │ Comparative plots of observed and predicted the MRF by WT-EEMD-KELM for test series via Monte Carlo method.

CONCLUSION

Prediction of river flow plays a crucial role in various fields of water resource projects. Therefore, in this study, the MRF was investigated via AI and data pre-processing methods. In this regard, under two scenarios, different models were developed based on station's own data and previous stations' data. The efficiency of developed models was tested via single and integrated AI methods. In the integrated models, time series were decomposed to several subseries using WT and further decomposition was performed via EEMD and most effective subseries were used as inputs for MARS and KELM models. The results showed that the single methods led to poor predictions based on the obtained statistical indicators. It was observed that data pre-processing improved the model's efficiency up to 25%. The maximum and minimum values of river flow were well predicted using the integrated models. The

integrated methods resulted in desirable accuracy in both considered scenarios and it was proved that the previous stations' data could be used as a reliable scenario in river flow prediction for stations without data. Also, the superior applied model dependability was investigated using uncertainty analysis. The results showed that the KELM model had an allowable degree of uncertainty in MRF modeling. Therefore, the integration of data-driven models with data processing methods could be useful for accurate prediction of the MRF time series as a more suitable and reliable method. It should, however, be noted that the KELM and MARS are data-driven models and the KELM- and MARS-based models are data sensitive, so further studies using data ranges out of this study or using other AI techniques should be carried out in the future studies. Also, the use of other input combinations based on precipitation, temperature, evaporation, and river channel seepage is suggested to find the merits of the intelligence models in MRF modeling and investigate the impact of the mentioned independent variables on the modeling process.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories (https://waterdata.usgs.gov).

REFERENCES

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J. & Srinivasan, R. 2007 Modelling hydrology and water quality in the prealpine/alpine Thur watershed using SWAT. *Journal of Hydrology* **333** (2), 413–430.

Alizadeh, M. J., Nourani, V., Mousavimehr, M. & Kavianpour, M. R. 2018 Wavelet-IANN model for predicting flow discharge up to several days and months ahead. *Journal of Hydroinformatics* **20** (1), 134–148.

Azamathulla, H. M. & Wu, F. C. 2011 Support vector machine approach for longitudinal dispersion coefficients in natural streams. *Applied Soft Computing* **11**, 2902–2905.

Azamathulla, H. M., Haghiabi, A. H. & Parsaie, A. 2017 Prediction of side weir discharge coefficient by support vector machine technique. *Water Science and Technology: Water Supply* **16** (4), 1002–1016.

Babovic, V. 2009 Introducing knowledge into learning based on genetic programming. *Journal of Hydroinformatics* **11** (3–4), 181–193.

Chaovalit, P., Gangopadhyay, A., Karabatis, G. & Chen, Z. 2011 Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys (CSUR)* **43** (2), 1–37.

Ebtehaj, I., Bonakdari, H., Moradi, F., Gharabaghi, B. & Khozani, Z. S. 2018 An integrated framework of extreme learning machines for predicting scour at pile groups in clear water condition. *Coastal Engineering* **135**, 1–15.

Ercelebi, E. 2004 Electrocardiogram signals de-noising using lifting-based discrete wavelet transform. *Computers in Biology and Medicine* **34**, 479–493.

Evsukoff, A. G., Cataldi, M. & Lima, B. 2012 A multi-model approach for long-term runoff modeling using rainfall forecasts. *Expert Systems with Applications* **39**, 4938–4946.

Farajzadeh, J. & Alizadeh, F. 2017 A hybrid linear-nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model. *Hydroinformatics* **20** (1), 246–262.

Fotovatikhah, F., Herrera, M., Shamshirband, S., Chau, K. W., Faizollahzadeh Ardabili, S. & Piran, M. J. 2018 Survey of computational intelligence as basis to big flood management: challenges, research directions and future work. *Engineering Applications of Computational Fluid Mechanics* **12** (1), 411–437.

Garg, V. & Jothiprakash, V. 2013 Evaluation of reservoir sedimentation using data driven techniques. *Applied Soft Computing* **13** (8), 3567–3581.

Gholami, A., Bonakdari, H., Zaji, A. H., Ajeel Fenjan, S. & Akhtari, A. A. 2016 Design of modified structure multi-layer perceptron networks based on decision trees for the prediction of flow parameters in 90 open-channel bends. *Engineering Applications of Computational Fluid Mechanics* **10** (1), 193–208.

Gill, M. K., Asefa, T., Kemblowski, M. W. & Makee, M. 2006 Soil moisture prediction using support vector machines. *JAWRA Journal of the American Water Resources Association* **42** (4), 1033–1046.

Haidar, A. & Verma, B. A. 2018 Novel approach for optimizing climate features and network parameters in rainfall forecasting. *Soft Computing* **22** (24), 8119–8130.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C. & Liu, H. H. 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **454**, 903–995.

Huang, G. B., Zhu, Q. Y. & Siew, C. K. 2006 Extreme learning machine: theory and applications. *Neurocomputing* **70** (1–3), 489–501.

Huang, G. B., Zhou, H., Ding, X. & Zhang, R. 2012 Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **42** (2), 513–529.

Kaveh, K., Bui, M. D. & Rutschmann, P. 2017 A comparative study of three different learning algorithms applied to ANFIS for predicting daily suspended sediment concentration. *International Journal of Sediment Research* **32** (3), 340–350.

Kitsikoudis, V., Sidiropoulos, E. & Hrissanthou, V. 2015 Assessment of sediment transport approaches for sand-bed rivers by means of machine learning. *Hydrological Sciences* **60** (9), 1566–1586.

Koycegiz, C. & Buyukyildiz, M. 2019 Calibration of SWAT and two data-driven models for a data-scarce mountainous headwater in semi-arid Konya closed basin. *Water* **11** (1), 147.

Labate, D., La Foresta, F., Occhiuto, G., Morabito, F. C., Lay-Ekuakille, A. & Vergallo, P. 2013 Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: a comparison. *IEEE Sensors Journal* **13** (7), 2666–2674.

Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of 'goodness of fit' measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35** (1), 233–241.

Lei, Y., He, Z. & Zi, Y. 2009 Application of the EEMD method to rotor fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing* **23** (4), 1327–1338.

Li, Y., Liang, Z., Hu, Y., Li, B., Xu, B. & Wang, D. 2020 A multi-model integration method for monthly streamflow prediction: modified stacking ensemble strategy. *Journal of Hydroinformatics* **22** (2), 310–326.

Mehr, A. D., Kahya, E. & Olyaei, E. 2013 Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *Hydrology* **505**, 240–249.

Mitchell, T. M. 1997 *Machine Learning*. McGraw-Hill, New York.

Mosavi, A., Ozturk, P. & Chau, K. W. 2018 Flood prediction using machine learning models: literature review. *Water* **10** (11), 1536.

Noori, R., Karbassi, A. R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M. H., Farokhnia, A. & Gousheh, M. G. 2011 Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology* **401**, 177–189.

Noori, R., Deng, Z., Kiaghadi, A. & Kachoosangi, F. T. 2015 How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers? *Hydraulic Engineering* **142** (1), 04015039.

Pachori, R. B., Avinash, P., Shashank, K., Sharma, R. & Acharya, U. R. 2015 Application of empirical mode decomposition for analysis of normal and diabetic RR-interval signals. *Expert Systems with Applications* **42** (9), 4567–4581.

Roushangar, K. & Ghasempour, R. 2017 Prediction of non-cohesive sediment transport in circular channels in deposition and limit of deposition states using SVM. *Water Science and Technology: Water Supply* **17** (2), 537–551.

Roushangar, K., Alizadeh, F., Adamowski, J. & Saghebian, S. M. 2020 Exploring the multiscale changeability of precipitation

using the entropy concept and self-organizing maps. *Journal of Water and Climate Change* **11** (3), 655–676.

Saghebian, S. M., Roushangar, K., Ozgur Kirca, V. S. & Ghasempour, R. 2020 Modeling total resistance and form resistance of movable bed channels via experimental data and a kernel-based approach. *Journal of Hydroinformatics* **22** (3), 528–540.

Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S., Kargar, K., Mosavi Nabipour, A. & Chau, K. W. 2020 Predicting standardized streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics* **14** (1), 339–350.

Sharda, V. N., Prasher, S. O., Patel, R. M., Ojasvi, P. R. & Prakash, C. 2008 Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data. *Hydrological Sciences Journal* **53** (6), 1165–1175.

Solomatine, D. P. & Dulal, K. N. 2003 Model trees as an alternative to neural networks in rainfall-runoff modelling. *Hydrological Sciences Journal* **48** (3), 399–311.

Talei, A., Chua, L. H. C. & Quek, C. 2010 A novel application of a neuro-fuzzy computational technique in event-based rainfall–runoff modelling. *Expert Systems with Applications* **37**, 7456–7468.

Taormina, R. & Chau, K. W. 2015 ANN-based interval forecasting of streamflow discharges using the LUBE method and MOFIPS. *Engineering Applications of Artificial Intelligence* **45**, 429–440.

Wu, C. L. & Chau, K. W. 2013 Prediction of rainfall time series using modular soft computing methods. *Engineering Applications of Artificial Intelligence* **26** (3), 997–1007.

Wu, Z. H. & Huang, N. E. 2004 A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **460** (2046), 1597–1611.

Wu, Z. H. & Huang, N. E. 2009 Ensemble empirical mode decomposition: a noise assisted data analysis method. *Advances in Adaptive Data Analysis* **1**, 1–41.

Yaseen, Z. M., Sulaiman, S. O., Deo, R. C. & Chau, K. W. 2019 An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology* **569**, 387–408.