

# Scenario-based prediction of short-term river stage–discharge process using wavelet-EEMD-based relevance vector machine

Kiyoumars Roushangar and Farhad Alizadeh

## ABSTRACT

In this study, daily river stage–discharge relationship was predicted using different modeling scenarios. Ensemble empirical mode decomposition (EEMD) algorithm and wavelet transform (WT) were used as hybrid pre-processing approach. In the WT-EEMD approach, first temporal features were decomposed using WT. Furthermore, the decomposed sub-series were further broken down into intrinsic mode functions via EEMD to obtain features with higher stationary properties. Mutual information was used to select dominant sub-series and determine efficient input dataset. Relevance vector machine (RVM) was applied to forecast river discharge. Three scenarios were developed to predict river stage–discharge process. First, a successive-station form of forecasting was proposed by incorporating geomorphological features into the modeling process. Subsequently, an integrated RVM (I-RVM) was trained based on the concept of the cascade of reservoirs and the meta-learning approach. The proposed I-RVM had the semi-distributed characteristics of the river discharge model. Finally, a multivariate RVM was trained to predict discharge for different points of the river. For this reason Westhope station's features were used as input to predict discharge at downstream of the river. Results were compared with rating curve and capability of proposed models were approved in prediction of short-term river stage–discharge.

**Key words** | ensemble empirical decomposition mode, modeling scenarios, mutual information, relevance vector machine, river stage-discharge process, wavelet transform

Kiyoumars Roushangar (corresponding author)

Farhad Alizadeh

Department of Civil Engineering,  
University of Tabriz,  
29 Bahman Ave., Tabriz,  
Iran  
E-mail: [kroshangar@yahoo.com](mailto:kroshangar@yahoo.com)

## INTRODUCTION

Water management policy is a crucial global issue that intends to govern water acquisition and determine proper ways to minimize the misapplication and spoil of attainable water resources. River discharge is very important in the hydrological cycle and hydro-environmental management and plays a basic designation in designing and operating irrigation schemes in a watershed scale. One of the most important problems in analysis and management of river discharge is the modeling/simulation. For this purpose, the complexity and uncertainty of the process is widely modeled using artificial intelligence (AI) by hydrologists and scientists.

The collection of continuous discharge measurements is a very hard task and, therefore, a stage–discharge (S-D) relationship is commonly used to estimate stream discharges as measured stage values. Rating curves are widely used to determine the S-D relationship, although they are not able to provide sufficiently accurate results. A S-D rating curve (RC) is a relationship between stream stage (water level) and discharge for a particular section of a stream. Such a relationship is usually established by a regression-based analysis, and the curves are generally expressed in the form of a power equation (Kisi & Cobenar 2009). However, using the S-D relationship for different river conditions

might not be sufficiently capable. In other words, to capture precise results, various researchers suggested applying AI-based models that proved to provide more accurate outcome in comparison to the S-D relationship (Ghorbani *et al.* 2016).

AI has been successfully applied in a number of diverse fields including water resources (e.g., Coulibaly *et al.* 1999; Anmala *et al.* 2000; Tokar & Markus 2000; Chou 2013; Singh *et al.* 2015), whereas there are many AI-based approaches that are driven by knowledge discovery (Babovic 2009; Meshgi *et al.* 2015). Likewise, machine learning approaches are remarkable AI methods that have been implemented in the river modeling field in recent years (e.g., Yu *et al.* 2004; Azamathulla *et al.* 2009; Chen *et al.* 2010; Nourani *et al.* 2016). Also, some researchers have performed reviews on good state-of-the-art application of AI and some pre-processing approaches to categorize different researches and their advantages (Labat 2005; Abrahart *et al.* 2012; Sang 2013).

The relevance vector machine (RVM), was first proposed by Tipping (2001), and is unique in its form. RVM was designed to enhance some of the shortcomings of the original support vector machine (SVM), including (a) probabilistic prediction issues, (b) solution solving problems, and (c) less tuning parameters (kernel or covariance functions). Similar to the SVM, RVM is a kernel-based method that parameterizes the unknown function like a weighted sum of non-linear basis functions in the feature space. Opposite to the SVM, RVM supposes that the weights are random variables and employs a Bayesian framework to evaluate the posterior distribution of weights using data. A considerable work on the use of RVM is the statistical downscaling of climate model outputs to predict the runoff for several Indian river basins (Ghosh & Mujumdar 2008).

Wavelet transform (WT) and ensemble empirical decomposition mode (EEMD) are pre-processing methods that donate remarkable vision into the physical form of the data by representing information in both time and frequency domains (Roushangar & Alizadeh 2018). EEMD has been proven to be a quite effective method for extracting signals from data generated in noisy non-linear and non-stationary processes (Huang *et al.* 1998). A time series in the WT breaks down into a series of linearly independent detailed signals and one approximation signal by using a

specific wavelet function. Mallat (1998) presented a complete theory for wavelet multi-resolution signal decomposition (also mentioned as a pyramid decomposition algorithm). Research proved that a proper data pre-processing by using the WT and EEMD can lead the models to adequately illustrate the real specifications of the basic system. WT along with EEMD decomposes a non-stationary signal into a given quantity of stationary sub-signals. Then, AI methods can be combined with WT and EEMD to improve preciseness of the prediction. In the present research, feature extraction and learning of historical data are the key to improve forecasting accuracy. For this reason, much of the actual effort is put into the processes of feature extraction, data pre-processing and transformations. Maheswaran & Khosa (2013) also used wavelet technique to improve non-linear model forecasting performances in Krishna River. Bai *et al.* (2015) developed a multiscale deep feature learning method with hybrid models to deal with daily reservoir inflow forecasting. Precise hybrid models were employed in recent years to forecast hydrological and hydrogeological processes (e.g., Partal 2009; Tiwari & Chatterjee 2010; Wang *et al.* 2015).

In the modeling process via AI approaches, some of the input variables might present correlation, noise, or have no significant relationship with target variables and generally are not equally informative. Therefore, one important step is to determine dominant input variables that are independent, informative, and effective. Mutual information (MI) as a non-linear measure of information content can be useful in obviating the selection of effective inputs among huge numbers of WT-based or WT-EEMD sub-series (Babovic & Keijzer 2000; Sannasiraja *et al.* 2004; Sun *et al.* 2010; Roushangar *et al.* 2016). Shannon entropy-based measures are applied in this study to extract dominant inputs of the proposed models for discharge-stage modeling. Since the entropy measures arbitrary dependencies between random variables, it is suitable to be applied to complex classification tasks, where methods based on linear relations are prone to mistakes. Entropy can be regarded as a measure of information, disorder, chaos, and uncertainty (Shannon 1948; Roushangar *et al.* 2016). Meyer *et al.* (2008) presents an original filter approach for effective feature selection in microarray data characterized by a large number of input variables and a few samples.

In most former studies of AI-based hydrological modeling techniques, temporal variables (as historical time series) of the process were proposed as inputs of the model to predict runoff (Anmala *et al.* 2000). On the other hand, in some studies, the AI-based and conceptual models have been combined to estimate the river flow (Tokar & Markus 2000; Ashu & Sanaga 2006; Chen & Adams 2006; Corzo *et al.* 2009; Roushangar *et al.* 2018a). These models, which combine conceptual hydrological models and AI-based model usually include two parts, one is a conceptual model for runoff of every sub-catchment and the other is the AI-based model to ensemble the runoff from each catchment. Nourani *et al.* (2013) developed a geomorphological model to handle Eel River watershed runoff. The quality of results supported the beneficial application of the approach in spatio-temporal modeling of the hydrological processes. Multi-station modeling is a conceptual approach whose capability has been confirmed, and in some cases, imposing geomorphological features on the model caused an improvement in results (Zhang & Govindaraju 2003; Sarangi *et al.* 2005).

By considering the aforementioned issues in prediction of discharge in rivers, the main purpose of the present study is to predict the daily river stage–discharge time series and enhance the capability of modeling scenarios by using hybrid WT-EEMD-MI and RVM for: (1) investigating capability of WT-EEMD-MI based multi-station model along with physical characters to forecast discharge and find improvements; (2) reconstructing integrated RVMs (I-RVMs) by a hybrid use of WT-EEMD-MI and find progress; (3) multivariate prediction of river discharge via multivariate RVM (MRVM) and enhance its capability by the proposed pre-processing approach. Results are going to

be compared with classic RC to investigate the improvements of results.

## MATERIAL AND METHODS

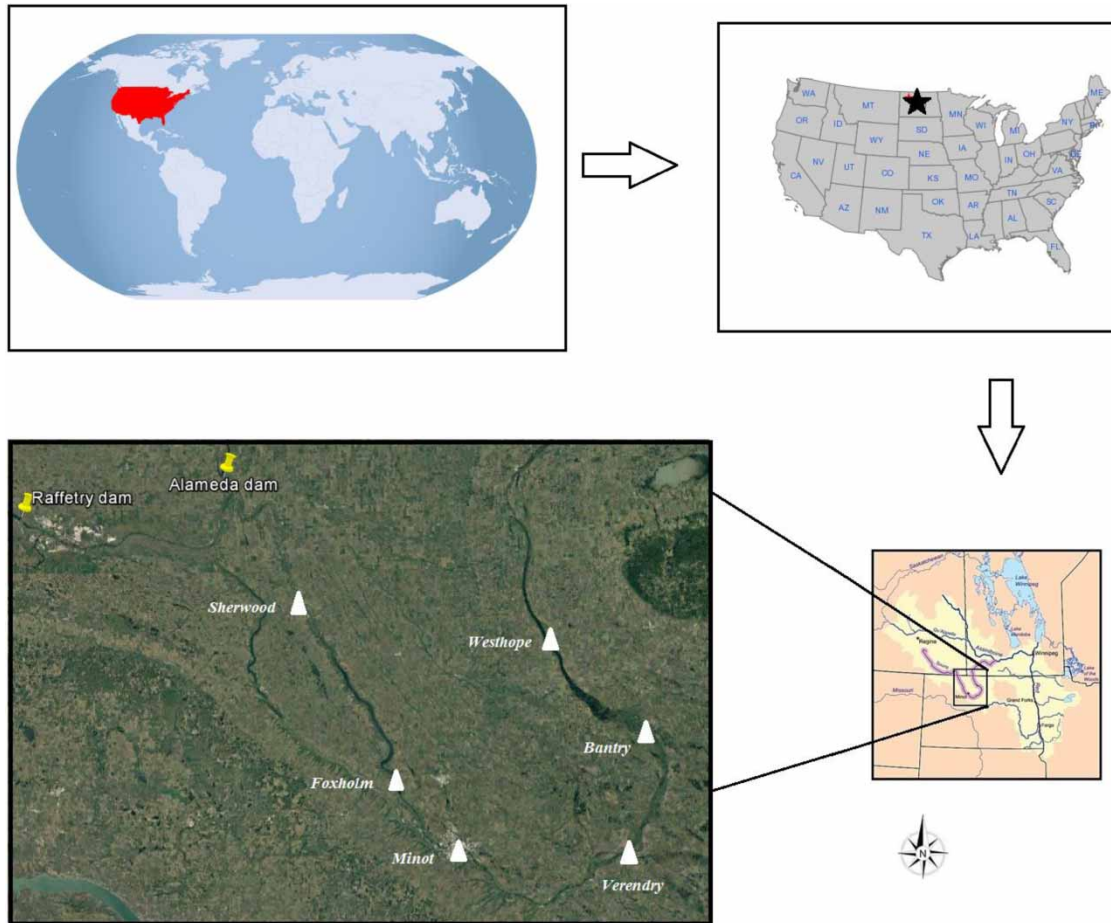
### Study area and used data

The Souris River is a river in central North America. It is about 700 km in length and drains about 61,100 km<sup>2</sup>. It rises in the Yellow Grass Marshes north of Weyburn, Saskatchewan. It wanders south through North Dakota beyond Minot to its most southern point in the city of Velva, and then back north into Manitoba. The river passes through the communities of Melita, Hartney, Souris, and Wawanesa and on to its confluence with the Assiniboine River at Treesbank, about 40 km south east of Brandon. The main tributaries which flow into the Souris in Manitoba are the Antler River, the Gainsborough and Plum Creeks. Much of its drainage basin is fertile silt and clay deposited by former glacial Lake Souris. The channel capacity of the river in Manitoba varies from about 4.2 m<sup>3</sup>/s near the border, to about 40 m<sup>3</sup>/s through Melita, to about 31 m<sup>3</sup>/s near Lauder and 48 m<sup>3</sup>/s near Hartney. North of Hartney the capacity increases to more than 85 m<sup>3</sup>/s. The drop between the border and Hartney is only about 9.5 cm/km. Two large dams in Saskatchewan, Rafferty Dam and Alameda Dam, were built, in part, to reduce flood peaks on the Souris River. Table 1 demonstrates the characteristics of the Souris River and used data.

Selected stations (Sherwood, Foxholm, Minot, Verendrye, Bantry, and Westhope) as shown in Figure 1, are consecutive which is suitable for use in the proposed

**Table 1** | Characteristics of six sub-basins in the Souris River

Hydrometric station	Stations discharge (m <sup>3</sup> /s)	Stage (m) Max.	Max. discharge (m <sup>3</sup> /s)	Min. discharge (m <sup>3</sup> /s)	Area (m <sup>2</sup> )	Distance from next hydrometric station (km)
Sherwood	4.1	1.02	10.8	0.36	23,154.4	–
Foxholm	6.74	2.15	11.44	0	24,527.2	145
Minot	7.28	1.65	12.32	0.07	27,453.9	42
Verendrye	13.74	1.96	14.78	0.31	29,266.9	126
Bantry	14.13	2.28	17.28	0.42	31,856.8	124
Westhope	15.66	3.05	14.8	0.12	43,770.8	121



**Figure 1** | The geographic location of Souris River and location of hydrometric stations.

models. Also, the dataset was separated into two parts: 70% of total data as calibration dataset and the rest was considered as verification dataset. The historical daily discharge data and morphology of the river were available for six stations on the Souris River at the USGS website (<http://co.water.usgs.gov/sediment>).

### Rating-curves

A discharge–height RC technique consists of a graph or equation, relating discharge to stream stage (water level), which can be used to estimate the discharge from the water level record. The S-D RC generally represents a functional relationship of the form given in Equation (1):

$$Q = aH^b \quad (1)$$

where  $Q$  is the stream discharge and  $H$  is the water level. The values of  $a$  and  $b$  for a particular stream are determined from data via a linear regression between  $\log H$  and  $\log Q$ . A major limitation of this approach is that it is not able to consider the hysteresis effect. In this study, the values of  $a$  and  $b$  are computed by using the least squares method.

### Concept of modeling via AI-based approaches

With respect to the reviewed researches, most of the deficiencies about application of AI-based models can be related to its structure. A simple AI model might not be able to handle the natural uncertainty of specific hydrological processes completely. As an instance, an ad hoc AI model is not designed to predict the variable of interest

in an arbitrary point across the river. Therefore, it is required to design models which have semi-distributed models' characteristics. To this end, in this research, different scenarios in river discharge modeling context were proposed. Multi-station modeling via RVM, I-RVMs, and MRVMs are the conceptual models used in this study, which will be discussed. Figure 2 shows the modeling flowchart in this study.

On the other hand, discharge data include a broad domain of values and fluctuations and have to be pre-processed. As is demonstrated in Figure 2, it is proposed to use the WT to decompose the original time series into two components: the approximation components and the detailed components. Secondary decomposition was performed by applying the EEMD to the detailed components obtained from WT into a new sub-series of intrinsic mode functions (IMFs). The goal of this step is to additionally decrease the non-stationarity of the detailed components captured from WT. Although decomposing twice might lead to the required outcome, reproduction of sub-series might defect the RVM performance. For this reason, MI was used to select dominant sub-series. Such a feature selection could increase the performance of RVM.

## Scenarios of modeling via RVM

### Multi-station modeling of river discharge

The multi-station modeling of river discharge used in this study was designed to catch the non-linearity, uncertainty, and complexity of the river discharge. For this purpose, the multi-station model was established according to the observed discharge time series and spatially varying physical parameters like upstream drainage area of sub-basins. In other words, the model inputs consisted of the temporal features and geomorphological characteristics of the sub-basins. A schematic of the proposed model is shown in Figure 3. The physical characters are used as input data of the model since most physical models take advantage of this parameter (e.g., Melesse *et al.* 2003). Likewise, these characteristics are the prime features used by several experimental equations (e.g., SCS and rational equations). Upstream drainage area of the sub-basins is used in most physically based models. Hence, the spatial feature was used along with other features to improve results.

Since the model is in the multi-station form, the input matrix was determined in a way to cover the spatio-temporal variation of the river discharge uncertainties by using

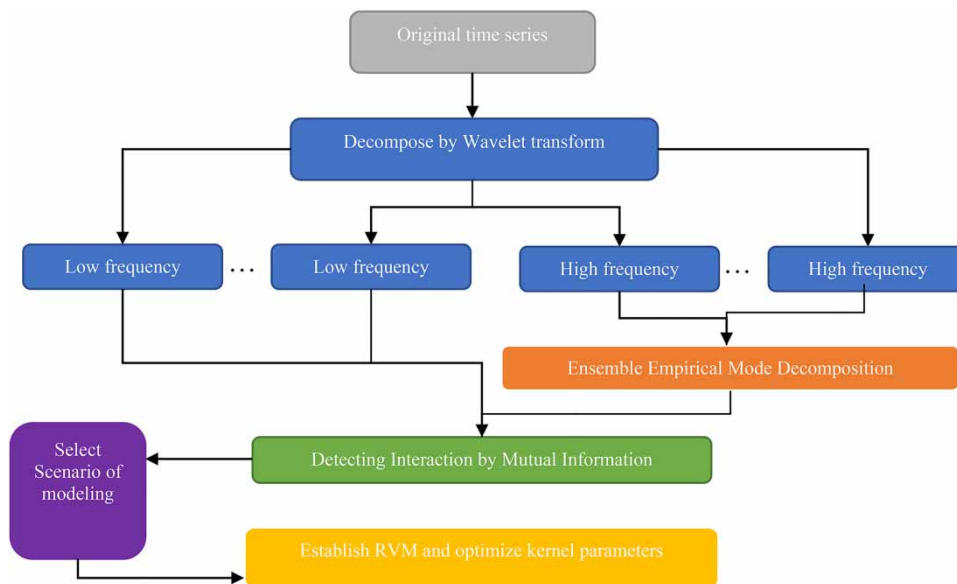
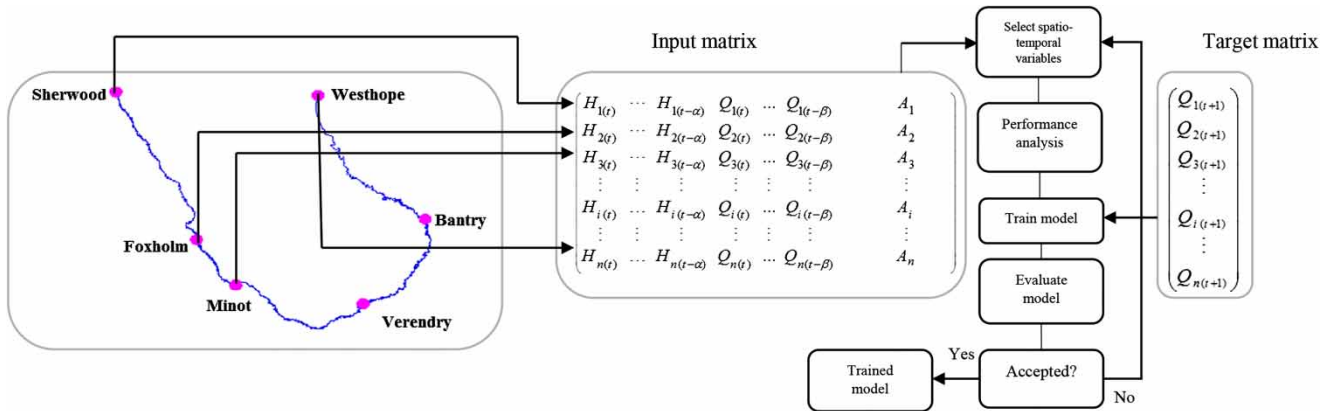


Figure 2 | Flowchart of pre-processing via WT-EEMD-MI in this study.





**Figure 3** | Schematic of predicting via multi-station model. The figure displays the input data selection and process of prediction.

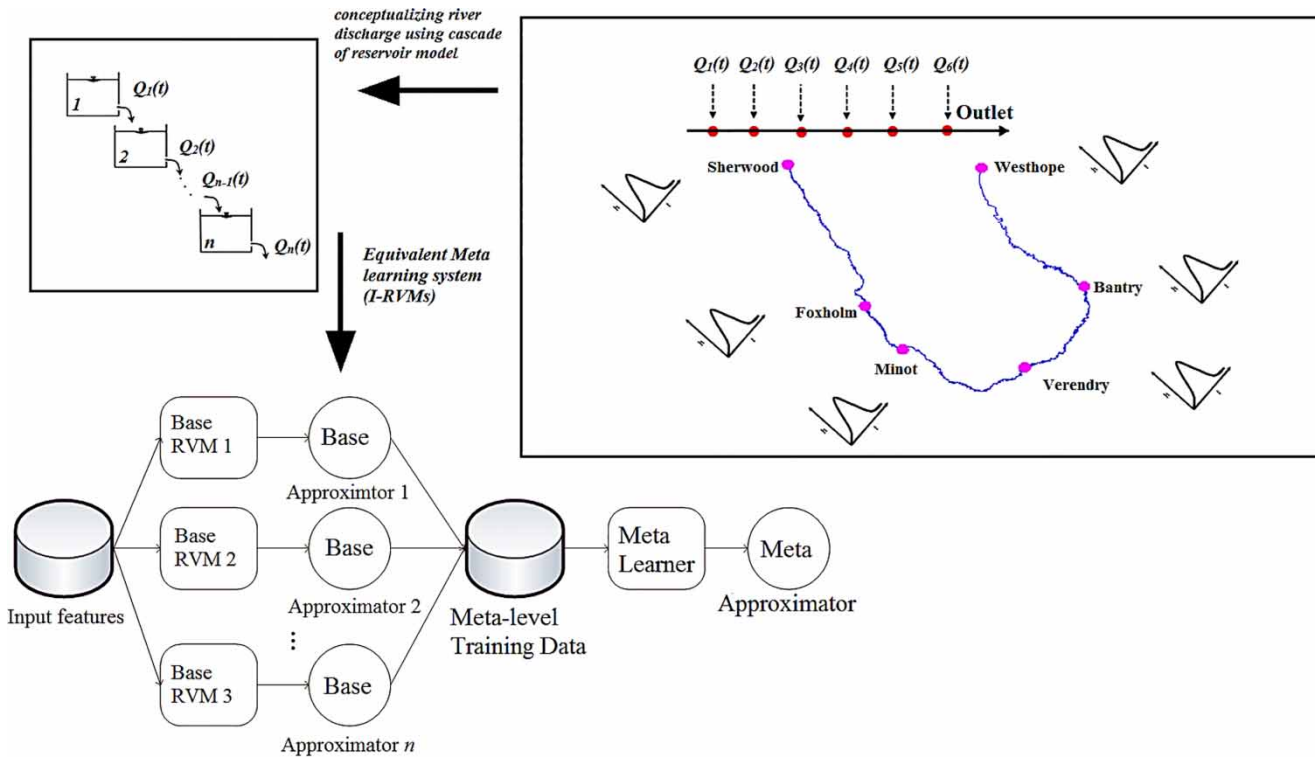
temporal features and geomorphological characteristics of the hydrometric stations. Therefore, multiple inputs were set in a way that all temporal and spatial features could establish a unit matrix. The input variables comprised different sets of antecedent and current gauge S-D values and upstream drainage area of the stations to forecast the discharge values ( $Q_i(t)$ ,  $i = 1, 2, 3 \dots n$ , where  $n$  is the number of hydrometric stations). The multi-station model presents a single model which is capable enough to be employed instead of several models within the watershed. The proposed model has spatio-temporal characteristics which can be applied to forecast entire river discharge using only one model. Selection of an appropriate input combination for the model is the most important step in the modeling process (Nourani *et al.* 2013). In the multi-station model, the input variables of the model are selected among the temporal and spatial features of river to forecast the discharge values for each station ( $Q_i(t+1)$ ) by assuming the Markovian property of the process. In addition, the proposed model is able to predict runoff in stations with lack of data or any desired point within the river due to employing the spatial and temporal features of the sub-basins as the model inputs. A MATLAB code was developed to address the model.

### Integrated RVMs

Spatial variations in a river or watershed heterogeneity is common in nature and can be a conclusive factor affecting the modeling performance. Generally, the river discharge

can be predicted using a simple AI model by employing time series of discharge and possibly other relevant features (Huo *et al.* 2012). Nevertheless, such a model cannot be trained using other stations' features which might lead to a weak outcome. To overcome this issue, this study planned a conceptual model inspired by the cascade of reservoirs model by the integrated form of the RVMs. The I-RVMs strategy used herein utilizes the concept of representing river as a cascade of reservoirs with different runoff capacity (Nash & Sutcliffe 1970) as presented in Figure 4. Similarly, for  $n$  reservoirs placed as a cascade, representing each of the stations, the outflow from  $n$ th reservoir can be specified based on the channel drainage network. In the proposed model, a watershed is divided into different sub-basins, on the basis of approximately the same interval lines passing through the main channel reach joints, as shown in Figure 4. Then, each sub-basin, accompanied by its main channel, is represented by a cascade of linear reservoirs and a linear channel. On the other hand, to transform the hydrologic model to AI-based model, the I-RVMs were designed to take advantage of the meta-learning system. Meta-learning is a privileged technique to unify or aggregate the results of multiple learners (Liao & Feng 2014).

Meta-learning is a general technique to aggregate the results of multiple learners, and it can be loosely defined as learning from information generated by a learner(s) (Chan & Stolfo 1993). Meta-RVM model is achieved by a basic configuration that has several base-RVMs and one meta-learner. The meta-learner learns from the outputs of base-RVMs, as presented in Figure 4. The meta-RVM



**Figure 4** | Scheme of integrated RVMs (I-RVMs). The concept is based on cascade of reservoirs and meta-learning system.

trains the base-RVMs on disjoint subsets of a dataset and a ‘top’ RVM on the whole dataset where hierarchical architecture forms.

### Multivariate prediction of river discharge

The proposed methodology was designed as a multi-input multi-output method in which the MRVM was trained based on an input dataset consisting of temporal features (S-D) of Westhope station for prediction of river discharge for the stations located at upstream of Westhope station (based on cross-correlation values captured from MI values). Therefore, by training only one MRVM for Souris River, the discharge values for all hydrometric stations could be predicted based on the S-D dataset of the Westhope station. The great difference between multi-station model and multivariate model is that multi-station model predicts the discharge time series as a united time series, whereas the proposed MRVM model predicts the discharge values of all stations separately. In such multi-output models, the interaction between the discharge values of all hydrometric

stations on the river must be considered through the training process of the MRVM and, therefore, the trained system can be employed to fill the missing, or to predict the future discharge values of all hydrometric stations on the river using the hydrological data observed in the Westhope station.

Selecting an appropriate input for MRVM is a very substantial process due to multi-output structure. Therefore, employing the WT-EEMD-MI could increase capability of the MRVM in the modeling of non-linear and non-stationary features of discharge. Also, MI was used to select dominant subcomponents in order to improve MRVM performance.

### Relevance vector machine

RVM has good generalization performance, as attaining sparsity in the exposure. The RVM is a kind of Bayesian regression structure, in which the weights of each input feature are governed by a set of hyperparameters. These hyperparameters demonstrate the posterior diffusion of the weights which are estimated frequently during training. Most hyperparameters near infinity, causing the posterior

diffusions and effectively setting the corresponding weights to zero. The remaining features with non-zero weights are called *relevance vectors*.

Tipping's formulation (Tipping 2001) was designed to allow regression from a multivariate input to a univariate output feature. In multiple modeling, one solution is to use a single RVM for each output dimension. This solution leaves the problem that several RVMs must be trained and might face some issues. MRVM (Thayanathan *et al.* 2006) extended the RVM structure to multivariate outputs, establishing a general regression tool. This formulation allows designation of the same set of patterns for all output targets. In addition, MRVM is optimized based on the weights without the use of hyperparameters. A data likelihood is acquired as a function of weight variables and hyperparameters. The weight variables are then analytically consolidated to obtain a marginal likelihood as a function of the hyperparameters. An optimal set of hyperparameters is obtained by maximizing the marginal likelihood over the hyperparameters using a version of the fast marginal likelihood maximization algorithm. The optimal weight matrix is obtained using the optimal set of hyperparameters.

The MRVM framework was designed to handle multivariate outputs and minimize the cost function according to Equation (2) and learn the parameters of mapping functions,  $\mathbf{W}^k$  and  $\mathbf{S}^k$ .

$$L^k = \sum_{n=1}^N \log N(\hat{x}_k^{(n)} | \mathbf{W}^k \hat{\phi}_k(\mathbf{Z}^{(n)}), \mathbf{S}^k) \quad (2)$$

Where  $\hat{x}_k^{(n)} = \sqrt{C_k^{(n)}} x^{(n)}$  and  $\hat{\phi}_k(\mathbf{Z}^{(n)}) = \sqrt{C_k^{(n)}} \phi(\mathbf{Z}^{(n)})$

The weight matrix must be determined to elude overfitting. According to Tipping's relevance vector approach (Tipping 2001) and assuming a Gaussian prior for the weights of each basis function, let  $A = \text{diag}(\alpha_1^{-2}, \dots, \alpha_p^{-2})$ , where each element  $\alpha_j$  is a hyperparameter that determines the *relevance* of the associated basis function. The prior distribution over the weights is then

$$\rho(\mathbf{W}^k | A^k) = \prod_{r=1}^M \prod_{j=1}^p N(W_{rj}^k | 0, \alpha_j^{-2}) \quad (3)$$

where  $W_{rj}^k$  is the element at  $(r, j)$  of the weight matrix  $\mathbf{W}^k$ .

## Proposed pre-processing approach

### Discrete wavelet transform

The WT is a popular method and a very precise method for time series processing (Aussem *et al.* 1998; Nourani *et al.* 2013; Farajzadeh & Alizadeh 2018; Roushangar *et al.* 2018a). While the general theory behind WT is quite analogous to that of the short time Fourier transform, WT allows for a completely flexible window function (called the mother wavelet), which can be changed over time based on the shape and compactness of the signal. Given this property, WT can be used to analyze the time-frequency characteristics of any kind of time series. In recent years, WT has been widely used for the analysis of many hydro-meteorological time series (Nourani *et al.* 2013; Roushangar *et al.* 2018a, 2018c). As the mother wavelet moves across the time series during the WT process, it generates several coefficients that represent the similarity between the time series and the mother wavelet (at any specific scale). There are two main types of WT: continuous and discrete. Use of the continuous wavelet transform can generate a large number of (often unnecessary) coefficients, making its use and interpretation more complicated. On the other hand, the discrete wavelet transform (DWT) method simplifies the transformation process while still providing a very effective and precise analysis, since DWT is normally based on the dyadic calculation. DWT coefficients can be calculated by the following equation:

$$W_\psi(a, b) = \frac{1}{(2)^{a/2}} \sum_{t=0}^{N-1} x(t) \psi\left(\frac{t}{2^a} - b\right) \quad (4)$$

where  $2^a$  represents the dyadic scale of the DWT. Applying DWT to a time series decomposes that time series into two ancillary time series shape components, called the approximation (A) and detailed (D) components. Component A comprises the large-scale, low-frequency component of the time series, while component D represents the small-scale, high-frequency component.

### Ensemble empirical mode decomposition

EEMD was proposed to solve the mode mixing issue of empirical mode decomposition (EMD) which specifies the



true IMF as the mean of an ensemble of trials (Wu & Huang 2009). Each trial consists of the decomposition results of the signal plus a white noise of finite amplitude. This new method is based on the insight from recent studies of the statistical properties of white noise (Wu & Huang 2004), which showed that the EMD method is an effective self-adaptive dyadic filter bank when applied to the white noise. 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.

The principle of the EEMD algorithm is as follows: the added white noise will populate the whole time–frequency space uniformly with the constituting components of different scales. When a signal is added to this uniformly distributed white noise background, the components in different scales of the signal are automatically projected onto proper scales of reference established by the white noise in the background. Because each of the noise-added decompositions consists of the signal and the added white noise, each individual trial may certainly produce very noisy results. However, the noise in each trial is different in separate trials (Wang *et al.* 2013; Roushangar & Alizadeh 2018). Thus, it can be decreased or even completely canceled out in the ensemble mean of enough trials. The ensemble mean is treated as the true answer because the only persistent part is the signal as more and more trials are added in the ensemble.

### Shannon entropy and MI and selecting dominant sub-series

Shannon formulated entropy ( $H$ ), now called the Shannon entropy or information content, for a discrete random variable of  $X$  with sample size of  $N$  (bin number), which takes values  $x_1; x_2; \dots; x_N$  with probabilities of  $p_1, p_2, \dots, p_N$ , respectively, as (Shannon 1948):

$$H(x) = H(p) = - \sum_{i=1}^N p(x_i) [p(x_i)] \quad (5)$$

Feature selection is more appropriate by MI than other linear correlation measuring tools since sometimes there may be a weak linear but strong non-linear relationship between input(s) and output(s). Such a usage of MI along

with RVM is favorable since both tools are beneficial in finding non-linear properties of features. MI between two random variables  $X$  and  $Y$  is defined as (Yang *et al.* 2000):

$$MI(A, B) = H(A) + H(B) - H(A, B) \quad (6)$$

with  $H(A)$  and  $H(B)$  being the entropy of  $A$  and  $B$ , respectively, and  $H(A, B)$  their joint entropy as:

$$H(A, B) = - \sum_{a \in A} \sum_{b \in B} P_{AB}(a, b) \log_{AB}(a, b) \quad (7)$$

WT was conjugated to the proposed scenarios. Application of such hybrid models in which spectral and temporal analysis are altogether can improve performance of RVM. The WT was applied to decompose stage–discharge time series. For this end, daubechies 4 (db4) mother wavelet was used since previous studies approved its capability to be applied in hydrology (Roushangar *et al.* 2016). Using a trial and error process to obtain the optimum decomposition level ended at level 5 in six sub-series (one approximation and five detailed). As was studied by Aussem *et al.* (1998), on finding a correct decomposition level, initially the following formula which offers a minimum level of decomposition was introduced:

$$L = \text{int}[\log N] \quad (8)$$

where  $L$  is decomposition level and  $N$  is the number of time series features (Nourani *et al.* 2013; Farajzadeh & Alizadeh 2018; Roushangar *et al.* 2018b). In the present study,  $L$  is equal to 3. This experimental equation was derived for fully autoregressive signals and only considering the time series length without considering the seasonal efficacy of the process. In order to handle the seasonality of the process, decomposition level up to 5 ( $2^5$ ; approximately monthly mode) was selected as the optimum level. Therefore, the seasonality of the process up to one month could be handled by the model. Due to the proportional relationship between the amount of discharge and river stage, these signals were supposed to have the same seasonality level and both time series were decomposed at the same level (i.e., level 5). On the other hand, since daily time series include various properties, non-linearity and non-stationarity, EEMD was applied to further breakdown previously decomposed time series (by WT).

In order to prevent probable issues from overtraining or other problems, it was tried to select the dominant sub-series from the decomposed time series prior to modeling. In this study, feature selection was performed using MI approach. MI is an unsupervised measuring tool computed with respect to input and target values. Input matrix was determined by considering general ranking of sub-series according to MI.

### Efficiency criteria

In this study, the total data were separated into calibration and verification sets. Three different criteria were selected to measure the revenue of the proposed forecasting methods; the root mean square error (RMSE), mean absolute error (MAE), and the determination coefficient (DC). The RMSE, MAE, and DC were applied to exhibit discrepancies between predicted and observed values (Adamowski & Chan 2011). The RMSE and MAE were used to quantify the modeling accuracy by generating a positive value. The RMSE and MAE grow from zero for perfect forecasts through large positive values as the differences between forecasts and observations become increasingly large. Obviously, a high value for DC (up to 1) and small value for RMSE indicate high efficiency of the model. Legates & McCabe (1999) indicated that a hydrological model can be sufficiently evaluated by these statistics. These measures are not oversensitive to extreme values (outliers) and are sensitive to additive and proportional differences between model predictions and observations. Therefore, correlation-based measures (e.g., DC statistic) can indicate that a model is a good predictor (Legates & McCabe 1999).

Input and output variables are normalized by scaling between 0 and 1 to eliminate their dimensions and to ensure that all variables receive equal attention during calibration of the model. Therefore, RMSE values are to be transformed between 0 and 1 (Nourani *et al.* 2016).

## RESULTS AND DISCUSSION

The RC is an empirical model which extracts information from recorded stage values. The RC was applied for all six

hydrometric stations. As an instance, Figures 5(a) and 5(b) demonstrate the RC for Sherwood and Minot stations. Results of RC using least square method led to the following results:

$$Q_{(Sherwood)} = 2.7422H^{2.8127}$$

$$Q_{(Minot)} = 0.0281H^{11.313}$$

Results of modeling are demonstrated in Figure 5(c). Based on the results it was observed that results need to be strengthened in terms of DC, RMSE, and MAE. We discuss the results of modeling using RVM in the next section.

### Optimizing of Gamma values

In this study, RVM was applied to predict discharge time series for the Souris River. An accurate use of RVM is to select an appropriate kernel function and tune related hyperparameters. A number of kernels were discussed by researchers, but studies suggested the effectiveness of radial basis kernel (RBF) function in the case of machine learning approaches in the majority of civil engineering applications (Gill *et al.* 2006; Goel & Pal 2009; Nourani *et al.* 2016). For a fair comparison of results, a RBF, where  $\gamma$  (kernel width) is a kernel specific, was used by the RVM algorithm in this study.

### Results of multi-station model

The river flow process in any cross section of river system can be characterized as the function of various variables such as discharge, catchment, and river physical characteristics. The relationship between river discharge and influential variables can be expressed by:

$$Q(t+1) = f(x_{t+1}) + \varepsilon_{t+1} \quad (9)$$

where  $Q(t+1)$  denotes the river discharge in any cross section of river system,  $\varepsilon_{t+1}$  is the random error,  $x(t+1)$  is the input vector, which may include many variables such as antecedent flow data and other relevant variables at various time lags, but in the present application, it consists of

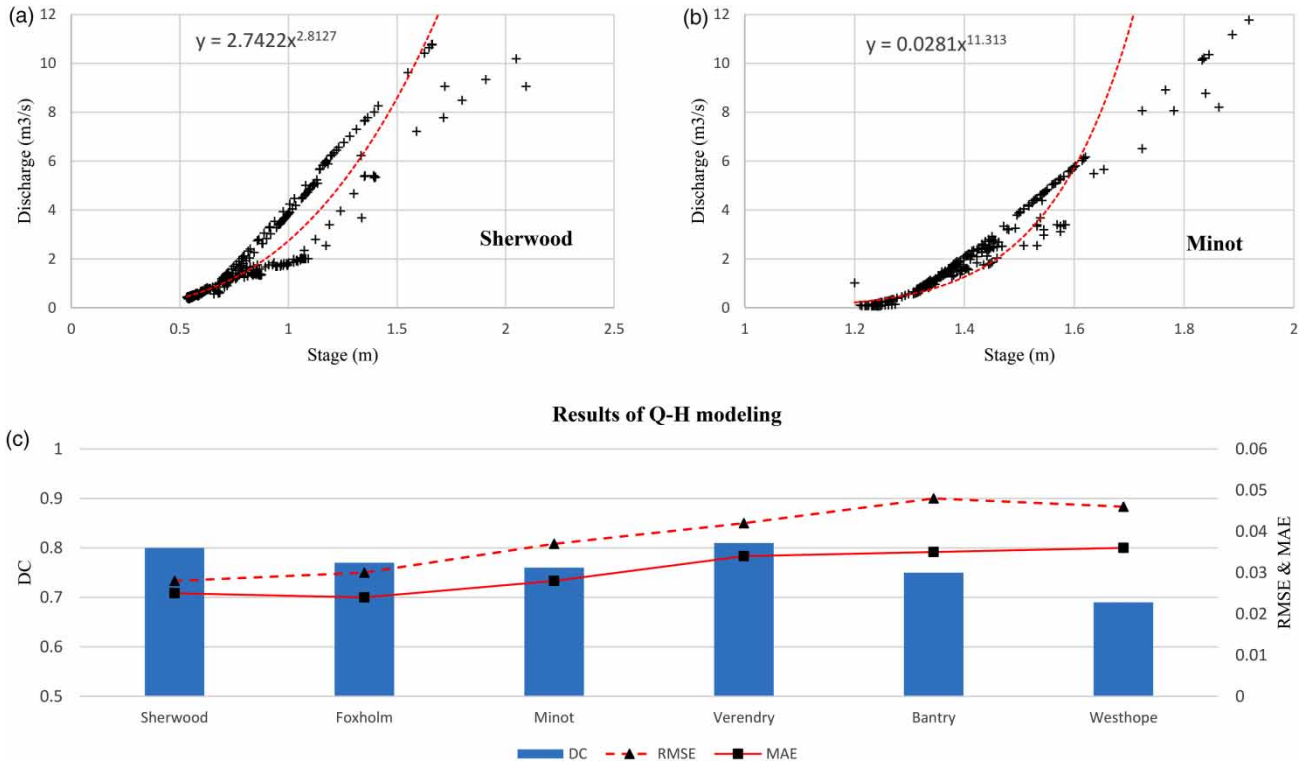


Figure 5 | (a) Results of modeling via RC for Sherwood station; (b) results of modeling via RC for Minot station; (c) overall results of RC for all stations.

combinations of antecedent daily S-D records. It is evident that the training datasets should cover all the characteristics of the problem in order to capture effective estimation. One of the most important steps in developing a satisfactory forecasting model is the selection of the input variables. Hence, different combinations of the antecedent S-D are used to construct the appropriate input structure for all scenarios.

In order to determine input variables for AI-based models, the input data were extracted based on the principle underlying the RC, in which different model structures were selected by a combination of stage and/or discharge variables including:  $H_t, H_{t-1}, H_{t-2}, \dots$  and  $H_{t-7}$  representing stage values at times  $(t), (t-1), (t-2), \dots$  and  $(t-7)$  and  $Q_{t-1}, Q_{t-2}, \dots$  and  $Q_{t-7}$  representing discharge values at times  $(t-1), (t-2), \dots$  and  $(t-7)$ . Models 1–3 extract information from stage values alone; Models 4–6 extract information from discharge values alone with a structure similar to autocorrelation; and Models 7–10 extract information from both stage and discharge values. While it is clear that the RC is not the only technique, Table 2 shows

Table 2 | Input dataset applied in various models

Input number	Input definition
1	$H_{(t)}$
2	$H_{(t)}, H_{(t-1)}$
3	$H_{(t)}, H_{(t-1)}, H_{(t-2)}$
4	$Q_{(t)}$
5	$Q_{(t)}, Q_{(t-1)}$
6	$Q_{(t)}, Q_{(t-1)}, Q_{(t-2)}$
7	$Q_{(t)}, H_{(t)}$
8	$Q_{(t)}, Q_{(t-1)}, H_{(t)}$
9	$Q_{(t)}, Q_{(t-1)}, H_{(t)}, H_{(t-1)}$
10	$Q_{(t)}, Q_{(t-1)}, Q_{(t-2)}, H_{(t)}, H_{(t-1)}$

the possible different choices of dependent variables for each modeling technique but these are required to have optimized structures by avoiding over-fitting. In the present study, only the best structures are used to report the performance of the proposed models. The performance measures outlined in ‘Performance criteria’ were used to

select the particular model structure but other techniques will be used to assist the choice, e.g., scatter diagram, or plotting difference between individual model performances.

The proposed multi-station model, in addition to S-D time series, took advantage of spatially varying parameters as model inputs. The geomorphological variables used in the multi-station model had the issue of incompatibility in dimension. In order to avoid possible conflicts with incorrect dimensionality of obtained formulations, it was suggested to use the dimensionless values. This is a prevalent scientific application where units of measurements are effectively omitted through the introduction of dimensionless ratios (Nourani *et al.* 2013). In this way, the dimensionless geomorphological parameters were applied in the model. Therefore, the model construction for each sub-basin could be represented as:

$$Q_i(t+1) = f\left[H_i(t), H_i(t-1), Q_i(t), Q_i(t-1), \frac{A_i}{A_t}\right] \quad (10)$$

In Equation (10),  $i$  refers to the sub-basin number ( $i = 1, 2, 3 \dots 6$ ).  $H_i(t-\alpha)$  is river stage values with 1-day lag time at  $i$ th sub-basin.  $A_i$  is the drainage area for each sub-basin and  $A_t$  is the total watershed area. In order to develop the multi-station model for the entire watershed, the data of all six stations were imposed onto the RVM framework. Since the S-D values are sorely affected by recent conditions of watershed at daily time scale, only values with 2-day lag times (i.e.,  $\alpha = 2$ ) was used in the multi-station model. By considering the aforementioned issues for the RVM modeling, the multi-station model was calibrated and verified.

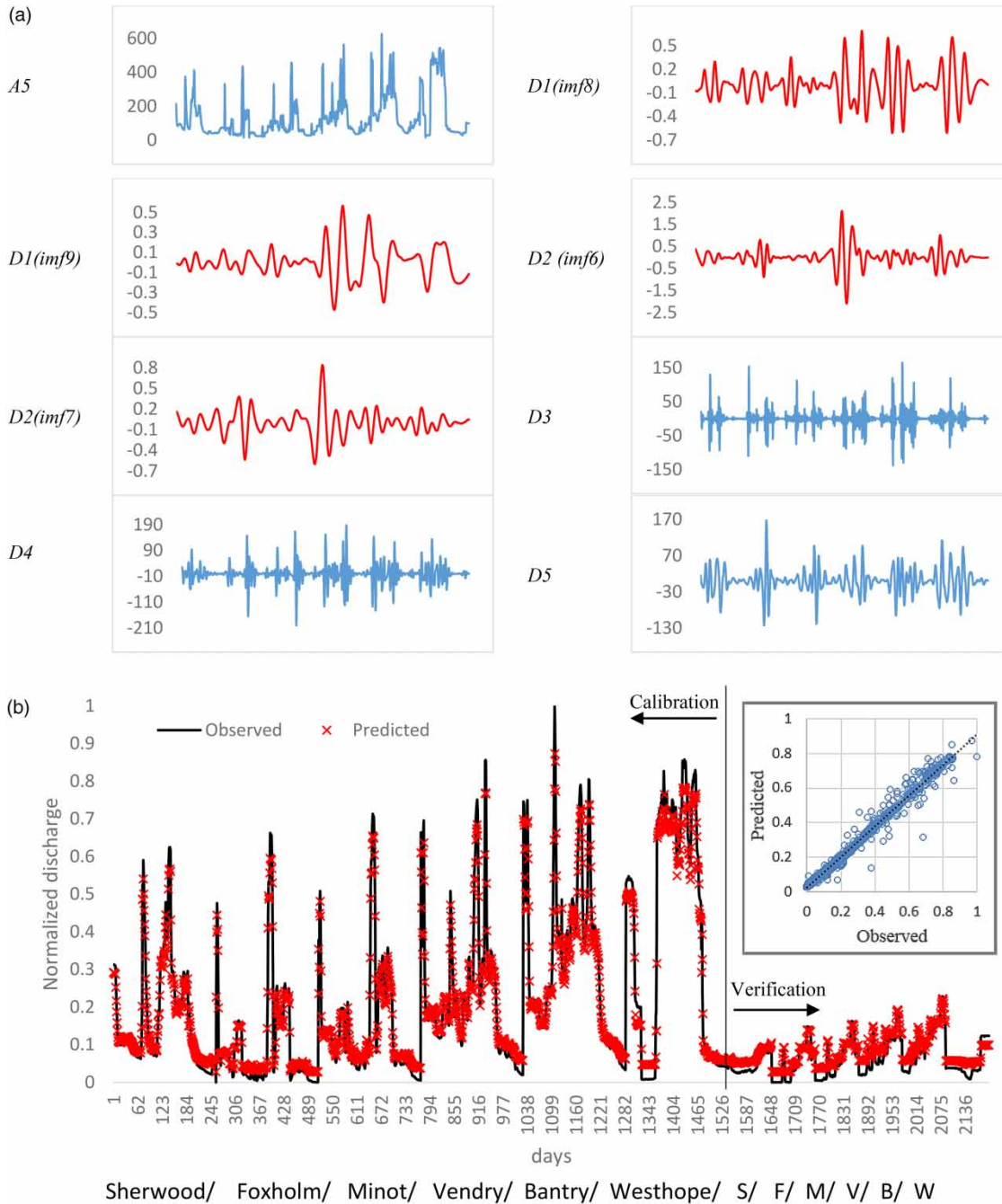
In order to find the efficient structure of the proposed multi-station model, sensitivity analysis was performed. For this end, the six best input combinations were considered to create the input matrix (Table 3). Table 3 shows the multi-station model's performance for different input combinations. In order to explore the efficiency of physical parameters, Comb. (1) was designed using only temporal variables (i.e., discharge-gauge stage data).

According to the obtained results, imposing dimensionless physical parameters along with temporal dataset caused an increase in modeling accuracy (Combs. (2)–(6)). Generally, Combs. (2)–(6) provided higher efficiency in comparison to Comb. (1). Moreover, replacing spatially varying parameters (i.e.,  $A_i$ ) instead of temporal variables in the input combination increased precision of outcome. Conjugating WT to temporal features in Comb. 5 gave an incredible insight into temporal features which caused Comb. 5 to have a better outcome. By considering improvement in Comb. 5, WT-EEMD-MI pre-processing approach was applied to find more dominant sub-series. For this reason, *detail 1* and *2* of *db(4,5)* was further decomposed using EEMD. MI was used to select appropriate IMFs from decomposed sub-series. According to Comb. 6, *imf 8* and *9* from *detail subseries 1* and *imf 6* and *7* from *detail subseries 2* was selected via the proposed pre-processing approach. Figure 6 presents decomposed time series of the discharge values by *db(4,5)*-EEMD-MI for the multiple station format. By comparing the obtained results for Combs. (1)–(6) (Table 3), it could be inferred that Comb. (6) was more efficient than others. The input structure of this combination consisted of two temporal variables and

**Table 3** | Results of multi-station modeling

Combination no.	Input matrix variables	Calibration			Verification		
		DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>
1	$Q(t), Q(t-1), H(t)$	0.82	0.024	0.015	0.7	0.031	0.021
2	$Q(t), Q(t-1), H(t), A$	0.84	0.021	0.014	0.74	0.029	0.19
3	$Q(t), H(t), A$	0.83	0.022	0.014	0.74	0.03	0.019
4	$Q(t), Q(t-1), H(t), H(t-1), A$	0.85	0.02	0.013	0.79	0.027	0.17
5	$Qd3, Qd4, Qd5, Qa, A$	0.88	0.019	0.012	0.82	0.025	0.016
6	$Qd1(imf\ 8,9), Qd2(imf\ 6,7), Qd3, Qd4, Qd5, Qa, Ha, A$	0.90	0.017	0.011	0.85	0.021	0.014

<sup>a</sup>RMSE and MAE values are dimensionless due to normalization.



**Figure 6** | (a) Decomposed time series using db4-EEMD-MI (8 sub-series); (b) multi-station forecasting of Souris River discharge. In verification step, first letter of each station represents the station's name.

one spatial variable (i.e.,  $A_i$ ). In Table 3,  $Qd1$  means detail subseries 1 and  $Qd2(imf 6, 7)$  means decomposed detail subseries 2 via EEMD and selected  $imf 6$  and  $7$  via MI and the rest are alike.

It must be mentioned that the drainage area ( $A$ ) is probably one of the most important physical characteristics of the watershed. Drainage area reflects the volume of water that can be generated from rainfall. Considering this fact,



the volume of available water for runoff would be the product of rainfall depth and the drainage area. Results approved the importance of using such a variable in structure of the proposed multi-station model.

### Integrated RVMS

For assessing the overall performance of I-RVMs in forecasting monthly discharge for the Westhope station, first temporal features of each station were pre-processed by WT-EEMD to improve the ability of base-RVMs. Then, each base-RVM was used to forecast target discharge values. Outcome of these models was aggregated to form a single time series. In the second stage, decomposed aggregated discharge time series were pre-processed by WT-EEMD-MI and then imposed into meta-learner to forecast Westhope's discharge values. Such a modeling strategy has the advantage of enjoying all datasets so that meta-learner can handle characteristics of the variable of interest (i.e.,  $Q(t+1)$  for Westhope station). One of the great differences between multi-station modeling and I-RVMs is using spatial data of the watershed. I-RVMs did not use spatial data, where the multi-station model was designed based on employing spatial data. However, I-RVMs were designed to explore spatial variation in S-D process and without requiring physical characteristics of the watershed. This model can operate as a non-linear time-space regression tool rather than a fully lumped model. Table 4 represents the input features of each base-RVM and meta-learner. According to obtained results from Table 4, I-RVMs performed well and in comparison to the multi-station RVM, outperformed it relatively. However, still both these strategies are unique and can be applied for different spatial and temporal conditions. In Table 4, for each variable,  $i$  represents the stations used in the modeling process.

Scatterplots of predicted values of discharge via I-RVM are demonstrated in Figure 7.

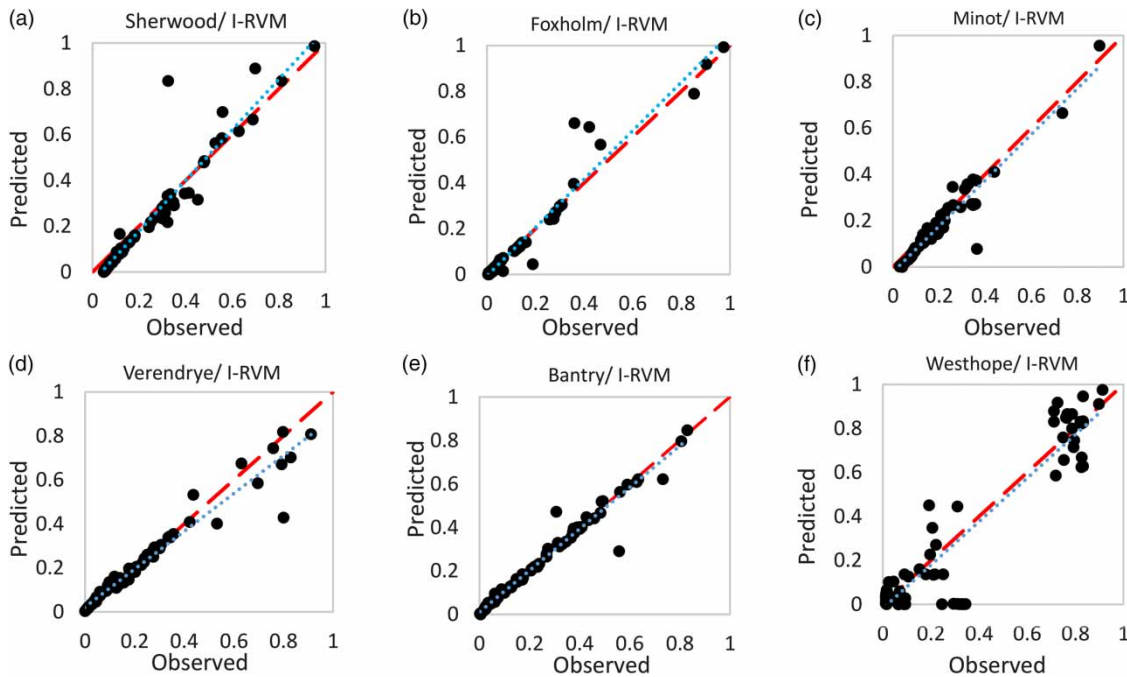
### Multivariate prediction of river discharge

MRVM, which was first proposed by Thayanathan *et al.* (2006), was designed to model multivariate processes. Unlike I-RVM modeling process, MRVM was capable to be extended according to the river discharging system. It was concluded that using Westhope station's dataset was suitable to be used as input to handle other stations' discharge values and forecast them so the outputs lead to appropriate results. According to MI, the highest correlation was detected among Westhope and other stations time series (according to the cross MI values (not reported here)), accordingly, it was selected as input of MRVM. Since such an application of MRVM might be very sensitive to input combinations, pre-processing was carried out to capture the seasonality and possible non-stationarity issues. Similar to other scenarios, WT-EEMD pre-processing was performed and MI was used to select the most dominant sub-series. Table 5 shows results of three different input combinations via sensitivity analysis and the outcome was presented in terms of RMSE and DC. Also, it is noteworthy to say that the MRVM was designed in a way to find its optimized structure without requiring pre-defined parameters (Thayanathan *et al.* 2006). Comb. 1 was designed based on the best classic input structure (Table 2) which led to moderate results. Utilizing WT-EEMD-MI by employing  $db(4,5)$  mother wavelet and further decomposing of detail 1 and 2, caused improvement in results. A recovery was also obtained by further decomposing of all detailed signals obtained from  $db(4,5)$ . In other words, all detailed sub-series of WT were imposed to EEMD and different numbers of IMFs were obtained. MI was then applied to select the

Table 4 | Results of I-RVMs

Input Comb. (base learner)	Input Comb. (meta learner)	Calibration			Verification		
		DC	RMSE	MAE <sup>a</sup>	DC	RMSE	MAE <sup>a</sup>
$Qid1(imf\ 8,9), Qid2(imf\ 6,7), Qid3, Qid4, Qid5, Qia, Hia$	$\Sigma Qid1(imf\ 8,9), Qid2(imf\ 6,7), Qid3, Qid4, Qd5, Qia, Hia$	0.88	0.032	0.021	0.83	0.036	0.024

<sup>a</sup>RMSE and MAE values are dimensionless due to normalization.



**Figure 7** | Scatterplot of modeling via I-RVM model for all stations on Souris River.

dominant IMFs and determined input dataset was fed into MRVM. Such a decomposition method, which was used for the first time in this research, caused improvement in outcome (20–40%). Outcomes of modeling for Minot and Bantry stations are demonstrated in [Figure 8](#).

The proposed pre-processing approach comprised WT, EEMD, and MI methods. WT decomposed the time series into sub-series grown in a dyadic form. Further breakdown was performed via EMMD and caused sub-series to be obtained with more stationary properties and less noise. Since decomposing into two levels led to a number of sub-series (about 40 to 50 sub-series), feeding them into RVM might have caused over-training and other issues. Therefore, MI was used to decrease the number of sub-series and omit noise and select dominant sub-series. Using WT-EEMD-WT caused improvement in results of about 2–20% in average in comparison other models.

With regard to the importance of river discharge predictions in hydro-environmental management, in the present study, the ability of the proposed models was also examined for multi-step-ahead predictions in daily scale. Three- and seven-days (weekly) ahead predictions were performed by taking advantage of each model's best input dataset. The multi-station model did not provide accurate

predictions and sometimes results were very poor. Considering the outcome, lead time forecast for the multi-station model was not reported here. The obtained results for I-RVMs (for Westhope station) and MRVM (all stations except Westhope) are demonstrated in [Table 6](#). Results showed that employing three-step-ahead prediction led to an acceptable outcome by the means of RMSE, MAE, and DC values. Despite the obtained results for 3-days-ahead prediction, outcome of prediction for 7-days-ahead weakened relatively. It is indicated from [Table 6](#) that in multi-days-ahead predictions, selected input led to relatively poor performance in comparison to one-step-ahead modeling and missed its precision by increasing the prediction horizon. According to the results from [Table 6](#), results of prediction using the best structure were relatively weak for some stations. The reason is because in multi-step-ahead prediction, Markovian property weakens by the time step increase.

### Comparison of results

Previously used AI-based models or physically based models were different from the proposed models used herein and the results are discussed here in terms of quality and

Table 5 | Results of multivariate forecasting by MRVM

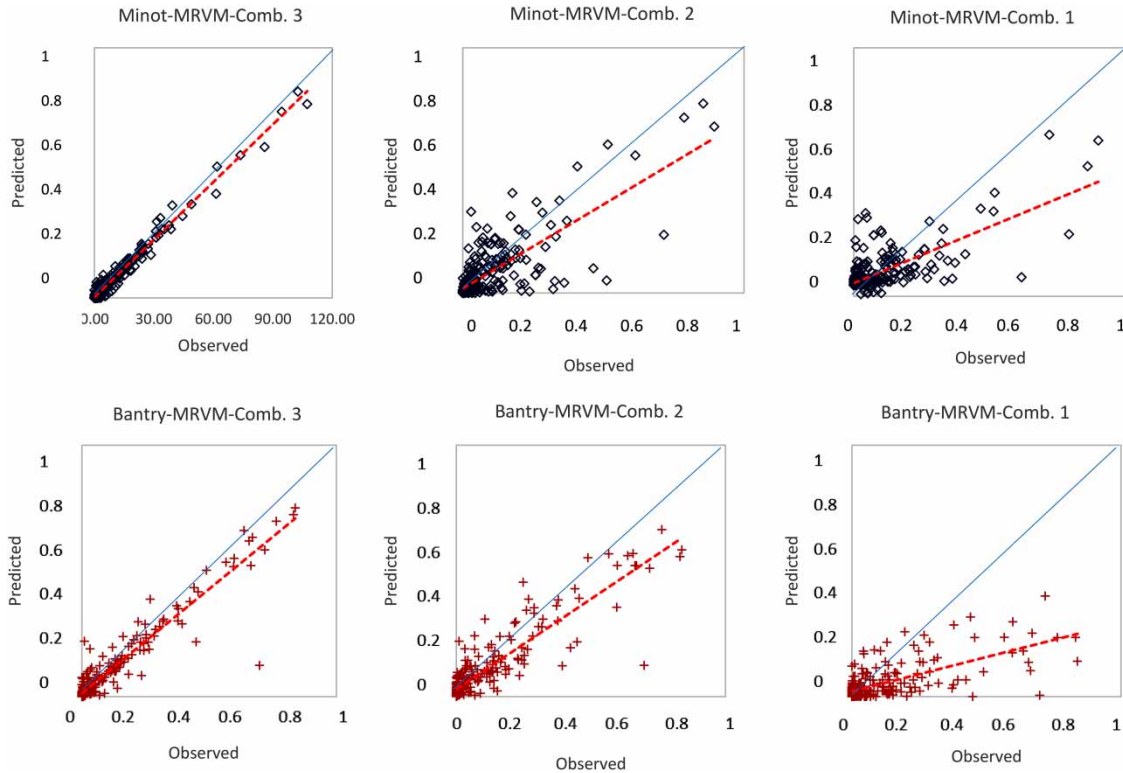
Station	Comb. 3						Comb. 2						Comb. 1					
	Calibration			Verification			Calibration			Verification			Calibration			Verification		
	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>
	Qd1(imf 8.9), Qd2(imf 6.7), Qd3(imf 6.7), Qd4(imf 6.7), Qd5(imf 6.7), Qa, Ha																	
Sherwood	0.90	0.017	0.012	0.82	0.024	0.016	0.86	0.019	0.012	0.75	0.028	0.018	0.71	0.034	0.023	0.66	0.44	0.32
Foxholm	0.86	0.02	0.014	0.80	0.026	0.017	0.82	0.025	0.017	0.73	0.03	0.02	0.7	0.037	0.026	0.63	0.048	0.034
Minot	0.90	0.021	0.014	0.85	0.024	0.016	0.84	0.033	0.21	0.75	0.039	0.26	0.72	0.038	0.027	0.66	0.047	0.33
Vendry	0.90	0.024	0.016	0.86	0.036	0.024	0.83	0.034	0.22	0.76	0.041	0.29	0.69	0.049	0.035	0.62	0.06	0.04
Bantry	0.88	0.025	0.17	0.81	0.034	0.22	0.81	0.042	0.29	0.69	0.046	0.33	0.67	0.052	0.4	0.59	0.068	0.051
Westhope	0.91	0.03	0.02	0.84	0.034	0.22	0.83	0.039	0.26	0.7	0.044	0.31	0.7	0.053	0.41	0.65	0.067	0.05

<sup>a</sup>RMSE and MAE values are dimensionless due to normalization.

quantity. First, we tried to explain the capability of proposed models under different conditions.

In order to demonstrate the capability of the proposed models, it is necessary to investigate in two separate parts: proposed hydrological models and proposed pre-processing approach. Three different strategies were proposed to study Q-H relation of the river, namely, multi-station model, I-RVMs and MRVM. All proposed models have unique capabilities and some deficiencies to be discussed below.

- **Rating curve:** RC is a physically based model which is used to extract information from river stage dataset. The best outcome was captured for Verendry station (DC = 0.81, RMSE = 0.042, and MAE = 0.034), whereas the worst result was captured for Westhope station (DC = 0.69, RMSE = 0.046, and MAE = 0.036).
- **Multi-station model:** The proposed model is in continuous and multi-station form and since this model took advantage of watershed morphology data (upstream drainage area of hydrometric stations) it behaved like a spatio-temporal model that could identify and predict the discharge values more accurately when employed with the proposed WT-EEMD-MI pre-processing approach. The multi-station model could handle the uncertainty and non-linearity of the Souris River by exploring the interaction of stations by using the spatial and temporal variables and it is capable of being applied to predict discharge values at an arbitrary point of the watershed because of its structure (input data, form of prediction) (DC = 0.83, RMSE = 0.32, MAE = 0.21, in verification period).
- **Integrated RVMs:** The integrated form of RVMs was designed to predict the discharge time series of downstream (Westhope station) without using its discharge values which is an advantage in this model. In the proposed model, there are no spatial features and temporal features were pre-processed by the WT-EEMD-MI approach. The mathematical formulation of I-RVMs was inspired from Nash's cascade of linear reservoirs' model and was transformed into AI-based model by using meta-learner system (DC = 0.85, RMSE = 0.21, MAE = 0.014 in verification period (best scenario based on results)).
- **MRMV:** The MRVM took advantage of MRVM's multi input-multi output structure to predict the river discharge via one input dataset. The main difference between the



**Figure 8** | Scatterplot of modeling via MRVM model for the Minot and Bantry stations.

**Table 6** | Results of modeling for multi-days-ahead for I-RVM and MRVM

Station	Three-days-ahead						Seven-days-ahead					
	Calibration			Verification			Calibration			Verification		
	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>	DC	RMSE <sup>a</sup>	MAE <sup>a</sup>
Sherwood	0.84	0.02	0.013	0.72	0.032	0.024	0.7	0.035	0.028	0.65	0.046	0.036
Foxholm	0.80	0.028	0.019	0.71	0.033	0.026	0.67	0.04	0.032	0.58	0.059	0.46
Minot	0.77	0.034	0.024	0.72	0.037	0.029	0.69	0.043	0.034	0.57	0.062	0.049
Vendry	0.79	0.039	0.029	0.71	0.045	0.034	0.64	0.05	0.039	0.53	0.068	0.052
Bantry	0.74	0.04	0.031	0.69	0.046	0.035	0.65	0.05	0.038	0.55	0.073	0.058
Westhope	0.83	0.039	0.027	0.7	0.044	0.032	0.7	0.053	0.04	0.65	0.067	0.053

<sup>a</sup>RMSE and MAE values are dimensionless due to normalization.

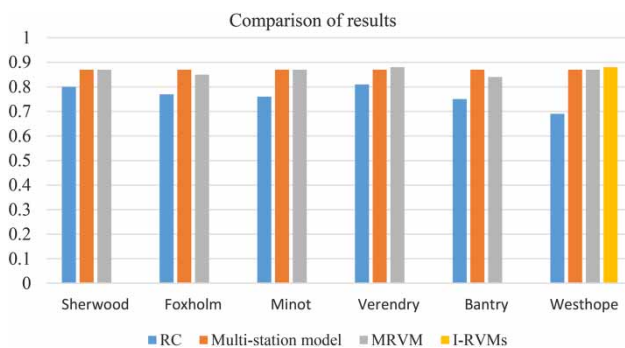
multi-station model and MRVM is the form of prediction. In the multi-station model, the discharge values of various stations are predicted as a united time series, whereas MRVM predicts the values separately for each station. The obtained results proved to be more accurate when they are applied in conjugation with WT-EEMD-MI.

- **WT-EEMD-MI:** The proposed pre-processing approach consisted of WT, EEMD, and MI approaches. WT decomposes the time series into sub-series which grow in a dyadic form. The sub-series carry the seasonal properties which strengthen any AI-based model to predict the time series more accurately. Further breakdown via EEMD led to obtaining sub-series with more

**Table 7** | Properties of applied models

Strategy of modeling	Characteristics	Advantages and disadvantages	Pre-processing method	Multistep-ahead-predictions
Multi-station	Multi-station form Spatio-temporal properties	High accuracy Fill the missing values Predict entire river discharge by training only one RVM Using geomorphological features as input	EEMD-WT-MI	No
I-RVMs	Integrated form of AI Meta-learning system Using geomorphological unit hydrograph based on cascade of linear reservoirs Combined form GUHCR and meta-learning system Semi distributed AI model	High accuracy Take advantage of semi-distributed specification to predict precisely Fill the missing values of target station Unable to predict or fill missing values of other stations in integrated form	EEMD-WT-MI	Yes
MRVM	Multi input multi output	Fill the missing values of target stations Unable to predict or fill missing values of other stations	EEMD-WT-MI	Yes

stationary properties and less noise in the data. Since decomposing in two levels led to a number of sub-series (about 40 to 50 sub-series), feeding them into RVM might lead to over-training or other issues. Therefore, MI was used to select the dominant sub-series and decrease the number of sub-series and omit noise. Using the WT-EEMD-MI caused improvement in results (about 2–20% in comparison to other models. [Table 7](#) demonstrates the properties of the proposed models. Quantitative comparison of overall results are demonstrated in [Figure 9](#).

**Figure 9** | Comparison of results for all scenarios and RC model for all stations.

## CONCLUDING REMARKS

In this research, capabilities of RVM by using conceptual modeling scenarios were verified to discover their quantitative and qualitative aspects in prediction of river S-D process. The reason for selecting RVM and MRVM was due to their structure and least parameters required to be tuned. Captured results could be separated into different categories. Pre-processing and proposed scenarios are the other subjects which are inferred from this research.

EEMD and WT gave incredible vision in the time and frequency domain of data to capture the non-linear and seasonal properties. WT was first applied to decompose both S-D time series. Since first and second detailed sub-series had a weak correlation with target data, EEMD was used to decompose the sub-series obtained from DWT (details and approximations) again into IMFs. By capturing more correlated first and second detailed sub-series decompositions by MI and imposing them on RVM, an increase in performance was observed. Especially in multivariate modeling, because of its sensitive structure, EEMD was applied for all detailed sub-series and obtained results showed a good agreement with observed time series.



Due to issues in hydrological predictions by using simple AI-based models, the present study took advantage of conceptual models. Multi-station model was designed to model river discharge in multiple-station form by training only one RVM. This strategy was capable of forecasting at the point of interest in a river or watershed by including physical properties of the case study. I-RVMs designed based on Nash's cascade of reservoir and meta-learning RVMs system were able to predict the leading point of the river (outlet of watershed) precisely. MRVM, which was the opposite of I-RVMs, used Westhope station's (last station located on the river) temporal features to predict other stations time series. Incorporating WT-EEMD-MI increased its accuracy and led to accurate results.

## REFERENCES

- Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E. & Wilby, R. L. 2012 Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Progress in Physical Geography* **36** (4), 480–513.
- Adamowski, J. & Chan, H. F. 2011 A wavelet neural network conjunction model for groundwater level forecasting. *Journal of Hydrology* **407**, 28–40.
- Anmala, J., Zhang, B. & Govindaraju, R. 2000 Comparison of ANNs and empirical approaches for predicting watershed runoff. *Journal of Water Resources Planning and Management* **126**, 156–166.
- Ashu, J. & Sanaga, S. 2006 Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. *Journal of Hydrology* **317**, 291–306.
- Aussem, A., Campbell, J. & Murtagh, F. 1998 Wavelet-based feature extraction and decomposition strategies for financial forecasting. *Journal of Computational Finance* **6** (2), 5–12.
- Azamathulla, H. M., Chang, C. K., Ab Ghani, A., Ariffin, J., Zakaria, N. A. & Abu-Hasan, Z. 2009 An ANFIS-based approach for predicting the bed load for moderately sized rivers. *Journal of Hydro-Environmental Research* **3** (1), 35–44.
- Babovic, V. 2009 Introducing knowledge into learning based on genetic programming. *Journal of Hydroinformatics* **11** (3–4), 181–193.
- Babovic, V. & Keijzer, M. 2000 Forecasting of river discharges in the presence of chaos and noise. In: *Coping with Floods* (J. Marsalek, ed.). NATO Science Series 2 Environmental Security Vol. 71, Springer, Dordrecht.
- Bai, Y., Wang, P., Xie, J. J., Li, J. T. & Li, C. 2015 An additive model for monthly reservoir inflow forecast. *Journal of Hydrologic Engineering* **20** (7), HE.1943-5584.0001101.
- Chan, P. K. & Stolfo, S. J. 1993 Experiments on multistrategy learning by meta-learning. In: *CIKM '93 Proceedings of the Second International Conference on Information and Knowledge Management*, Washington, DC, pp. 314–323.
- Chen, J. & Adams, B. J. 2006 Integration of artificial neural networks with conceptual models in rainfall–runoff modeling. *Journal of Hydrology* **318**, 232–249.
- Chen, S. T., Yu, P. S. & Tang, Y. H. 2010 Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology* **385**, 13–22.
- Chou, C.-M. 2013 Enhanced accuracy of rainfall–runoff modeling with wavelet transform. *Journal of Hydroinformatics* **15** (2), 392–404.
- Corzo, G. A., Solomatine, D. P., Hidayat, H., de Wit, M., Werner, M., Uhlenbrook, S. & Price, R. K. 2009 Combining semi-distributed process-based and data-driven models in flow simulation: a case study of the Meuse river basin. *Hydrology and Earth System Sciences* **13**, 1619–1634.
- Coulibaly, P., Anctil, F. & Bobee, B. 1999 Hydrological forecasting with artificial neural networks: the state of the art. *Canadian Journal of Civil Engineering* **26**, 293–304.
- Farajzadeh, J. & Alizadeh, F. 2018 A hybrid linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model. *Journal of Hydroinformatics* **20** (1), 246–262.
- Ghorbani, M. A., Khatibi, R., Goel, A., FazeliFard, M. H. & Azani, A. 2016 Modeling river discharge time series using support vector machine and artificial neural networks. *Environmental Earth Sciences* **75**, 685.
- Ghosh, S. & Mujumdar, P. P. 2008 Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources* **31**, 132–146.
- Gill, M. K., Asefa, T., Kemblowski, M. W. & Makee, M. 2006 Soil moisture prediction using Support Vector Machines. *Journal of the American Water Resources Association* **42** (4), 1033–1046.
- Goel, A. & Pal, M. 2009 Application of support vector machines in scour prediction on grade-control structures. *Engineering Application of Artificial Intelligence* **22** (2), 216–223.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., 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 Mathematical, Physical, and Engineering Sciences* **454**, 903–995.
- Huo, Z., Feng, S., Kang, K., Huang, G., Wang, F. W. & Guo, P. 2012 Integrated neural networks for monthly river flow estimation in arid inland basin of Northwest China. *Journal of Hydrology* **420–421**, 159–170.
- Kisi, O. & Cobenar, M. 2009 Modeling river stage-discharge relationships using different neural network computing techniques. *Clean* **37** (2), 160–169.

- Labat, D. 2005 Recent advances in wavelet analyses: Part 1. A review of concepts. *Journal of Hydrology* **314**, 275–288.
- 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**, 233–241.
- Liao, S. & Feng, C. 2014 Meta-ELM: ELM with ELM hidden nodes. *Neurocomputing* **128**, 81–87.
- Maheswaran, R. & Khosa, R. 2013 Wavelets-based non-linear model for real-time daily flow forecasting in Krishna River. *Journal of Hydroinformatics* **15** (3), 1022–1041.
- Mallat, S. G. 1998 *A Wavelet Tour of Signal Processing*, 2nd edn. Academic Press, San Diego, CA.
- Mellesse, A. M., Graham, W. D. & Jordan, J. D. 2003 Spatially distributed watershed mapping and modeling: GIS based storm runoff response and hydrograph analysis: Part 2. *Journal of Spatial Hydrology* **3** (2), 2–28.
- Meshgi, A., Schmitter, P., Chui, T. F. M. & Babovic, V. 2015 Development of a modular streamflow model to quantify runoff contributions from different land uses in tropical urban environments using genetic programming. *Journal of Hydrology* **525**, 711–723.
- Meyer, P. E., Schretter, C. & Bontempi, G. 2008 Information-theoretic feature selection in microarray data using variable complementarity. *IEEE Journal of Selected Topics in Signal Processing* **2**, 261–274.
- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models. I: a discussion of principles. *Journal of Hydrology* **10** (3), 282–290.
- Nourani, V., Komasi, M. & Alami, M. T. 2013 Geomorphology-based genetic programming approach for rainfall–runoff modeling. *Journal of Hydroinformatics* **15** (2), 427–445.
- Nourani, V., Alizadeh, F. & Roushangar, K. 2016 Evaluation of a two-stage SVM and spatial statistics methods for modeling monthly river suspended sediment load. *Water Resources Management* **30**, 393–407.
- Partal, T. 2009 Modelling evapotranspiration using discrete wavelet transform and neural networks. *Hydrological Processes* **23**, 3545–3555.
- Roushangar, K. & Alizadeh, F. 2018 Entropy-based analysis and regionalization of annual precipitation variation in Iran during 1960–2010 using ensemble empirical mode decomposition. *Journal of Hydroinformatics* **20** (2), 468–485.
- Roushangar, K., Garekhani, S. & Alizadeh, F. 2016 Forecasting daily seepage discharge of an earth dam using wavelet–mutual information–Gaussian process regression approaches. *Geotechnical and Geological Engineering* **34** (5), 1313–1326.
- Roushangar, K., Alizadeh, F. & Adamowski, J. 2018a Exploring the effects of climatic variables on monthly precipitation variation using a continuous wavelet-based multiscale entropy approach. *Environmental Research* **165**, 176–192.
- Roushangar, K., Alizadeh, F. & Nourani, V. 2018b Improving capability of conceptual modeling of watershed rainfall–runoff using hybrid wavelet-extreme learning machine approach. *Journal of Hydroinformatics* **20** (1), 69–87.
- Roushangar, K., Nourani, V. & Alizadeh, F. 2018c A multiscale time-space approach to analyze and categorize the precipitation fluctuation based on the wavelet transform and information theory concept. *Hydrology Research* **49** (3), 724–743.
- Sang, Y. F. 2013 A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric Research* **122**, 8–15.
- Sannasiraja, S. A., Zhang, H., Babovic, V. & Chand, E. S. 2004 Enhancing tidal prediction accuracy in a deterministic model using chaos theory. *Advances in Water Resources* **27** (7), 761–772.
- Sarangi, A., Madramootoo, C. A., Enright, P., Prasher, S. O. & Patel, R. M. 2005 Performance evaluation of ANN and geomorphology-based models for runoff and sediment yield prediction for a Canadian watershed. *Current Science* **89** (12), 2022–2033.
- Shannon, C. E. 1948 The mathematical theory of communications, I and II. *Bell System Technical Journal* **27**, 379–423.
- Singh, G., Panda, R. K. & Lamers, M. 2015 Modeling of daily runoff from a small agricultural watershed using artificial neural network with resampling techniques. *Journal of Hydroinformatics* **17** (1), 56–74.
- Sun, Y., Babovic, V. & Chan, E. S. 2010 Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory. *Journal of Hydrology* **395** (1–2), 109–116.
- Thayanathan, A., Navaratnam, R., Stenger, B., Philip, H. S., Torr, P. H. S. & Cipolla, R. 2006 Multivariate Relevance Vector Machines for Tracking. Computer Vision – ECCV. In: *Proceedings of the 9th European Conference on Computer Vision*, Graz, Austria, May 7–13, 2006, Part III. Springer-Verlag, Berlin, Heidelberg, pp. 124–138.
- Tipping, M. E. 2001 Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* **1**, 211–244.
- Tiwari, M. K. & Chatterjee, C. 2010 Development of an accurate and reliable hourly flood forecasting model using waveletbootstrap-ANN (WBANN) hybrid approach. *Journal of Hydrology* **394**, 458–470.
- Tokar, A. S. & Markus, M. 2000 Precipitation–runoff modeling using artificial neural networks and conceptual models. *Journal of Hydrologic Engineering* **5** (2), 156–161.
- Wang, W.-C., Xu, D.-M., Chau, K. W. & Chen, S. 2013 Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *Journal of Hydroinformatics* **15** (4), 1377–1390.
- Wang, W. C., Chau, K. W., Xu, D. M. & Chen, X. Y. 2015 Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management* **29** (8), 2655–2675.
- 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 and Physical Sciences* **460A**, 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.
- Yang, H. H., Vuuren, S. V., Sharma, S. & Hermansky, H. 2000 Relevance of time frequency features for phonetic and speaker-channel classification. *Speech Communication* **31**, 35–50.
- Yu, X., Liong, S. Y. & Babovic, V. 2004 EC-SVM approach for real-time hydrologic forecasting. *Journal of Hydroinformatics* **6** (3), 209–223.
- Zhang, B. & Govindaraju, R. S. 2003 Geomorphology-based artificial neural networks (GANNs) for estimation of direct runoff over watersheds. *Journal of Hydrology* **273**, 18–34.

First received 14 March 2018; accepted in revised form 24 October 2018. Available online 23 November 2018