World Academy of Science, Engineering and Technology
International Journal of Geological and Environmental Engineering
Vol:13, No:7, 2019

River Stage-Discharge Forecasting Based on Multiple-Gauge Strategy Using EEMD-DWT-LSSVM Approach

Farhad Alizadeh, Alireza Faregh Gharamaleki, Mojtaba Jalilzadeh, Houshang Gholami, Ali Akhoundzadeh

Abstract**—**This study presented hybrid pre-processing approach along with a conceptual model to enhance the accuracy of river discharge prediction. In order to achieve this goal, Ensemble Empirical Mode Decomposition algorithm (EEMD), Discrete Wavelet Transform (DWT) and Mutual Information (MI) were employed as a hybrid pre-processing approach conjugated to Least Square Support Vector Machine (LSSVM). A conceptual strategy namely multi-station model was developed to forecast the Souris River discharge more accurately. The strategy used herein was capable of covering uncertainties and complexities of river discharge modeling. DWT and EEMD was coupled, and the feature selection was performed for decomposed sub-series using MI to be employed in multi-station model. In the proposed feature selection method, some useless sub-series were omitted to achieve better performance. Results approved efficiency of the proposed DWT-EEMD-MI approach to improve accuracy of multi-station modeling strategies.

Keywords**—**River stage-discharge process, LSSVM, discrete wavelet transform (DWT), ensemble empirical decomposition mode (EEMD), multi-station modeling.

I. Introduction

THE collection of continuous discharge measurements is very hard mission, and therefore, a stage-discharge (Q-H) relationship is commonly used to estimate stream discharges as measured stage values. Rating curves are widely used to determine the Q-H relationship, although they are not able to provide sufficiently accurate results. A Q-H rating curve is a relationship between stream stage (water level) and discharge for a particular section of a stream. Usually regression-based power equations are used to analysis Q-H relation [1]. However, using Q-H relationship for different river conditions might not be capable enough. In the other words, to capture precise results, various researchers suggested to apply AI-based models that proved to provide more accurate outcome in comparison to Q-H relationship [2].

In modeling process based on 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. Shannon entropy-based measures are applied in this study to extract dominant inputs of the proposed models for discharge-stage modeling [3], [5].

The present study is aimed to predict the daily river stage-discharge process and enhance the capability of modeling scenarios by using Hybrid discreet wavelet transform (DWT)-EEMD-mutual information (MI) and least square support vector machine (LSSVM). To achieve this goal, capability of WT-EEMD-MI based multi-station model was investigated and improvements were studied, on the other hand, results were compared with classic rating curve (RC).

II. Material and Methods

A. Study Area and Used Data

Souris River is a river in central North America. It is about 700 km in length and drains about 61,100 km^2. The Souris River flows through the Melita, Hartney, Souris and Wawanesa tributies and on to its meeting with the Assiniboine River at Treesbank. Discharge of the river in varies from 4.2 m^3/s to about 85 m^3/s in downstream. There are two large dams in upstream of the river (Saskatchewan) namely Rafferty Dam and Alameda Dam, which were constructed, to reduce flood peaks on the Souris River. Table I demonstrates the characteristics of the Souris River and used data.

Selected stations (Sherwood, Foxholm, Minot, Verendry, Bantry and Westhope) as shown in Fig. 1, are in continued form which is suitable to be used in the proposed models. Also, dataset was separated into two parts; first division including 70 percent of total data as calibration dataset, and the rest was considered as verification dataset.

TABLE I
Characteristics of Six Sub-Basins in the Souris River

Hydrometric station	Stations Discharge (m^3/s)	Stage (m) Max.	Max. discharge (m^3/s)	Min. discharge (m^3/s)	Area (m^2)
Sherwood	4.1	1.02	10.8	0.36	23154.4
Foxholm	6.74	2.15	11.44	0	24527.2
Minot	7.28	1.65	12.32	0.07	27453.9
Verendrye	13.74	1.96	14.78	0.31	29266.9
Bantry	14.13	2.28	17.28	0.42	31.856.8
Westhope	15.66	3.05	14.8	0.12	43770.8

Farhad Alizadeh (M. Sc.) is with the East Azerbaijan Regional Water Company, Azerbaijan (e-mail: f.alizadeh.ce@gmail.com).
Alireza Faregh Gharamaleki (Ph.D.) and Ali Akhoundzadeh (Ph.D.) are with the East Azerbaijan Regional Water Company, Azerbaijan.
Mojtaba Jalilzadeh (M. Sc.) is with the Ministry of Energy of Iran, Iran.
Houshang Gholami (Ph.D. Candidate) is with the East Azerbaijan Regional Water Company, Azerbaijan.

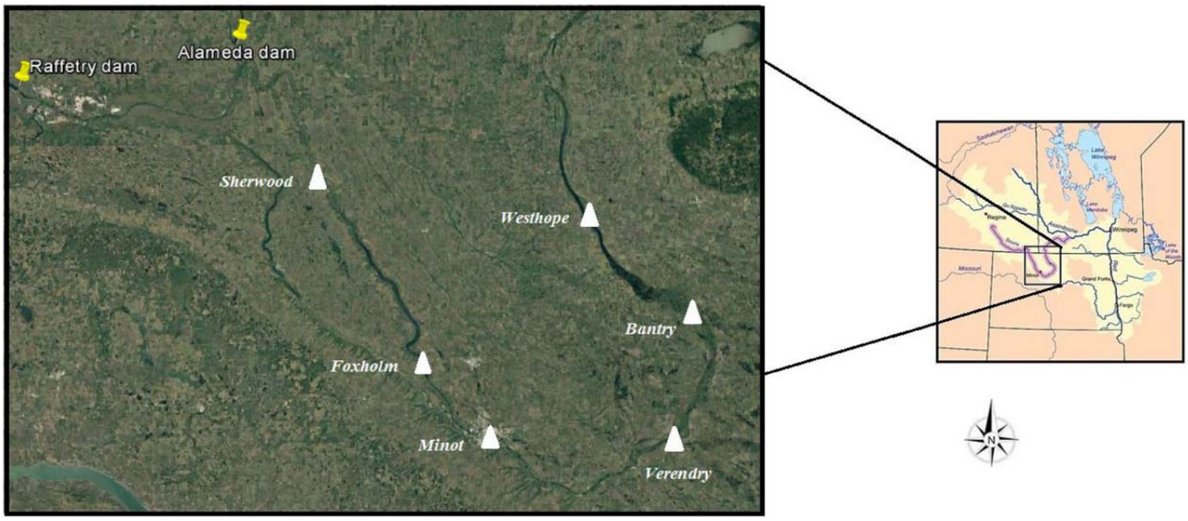

Fig. 1 The geographic location of Souris River and location of hydrometric stations

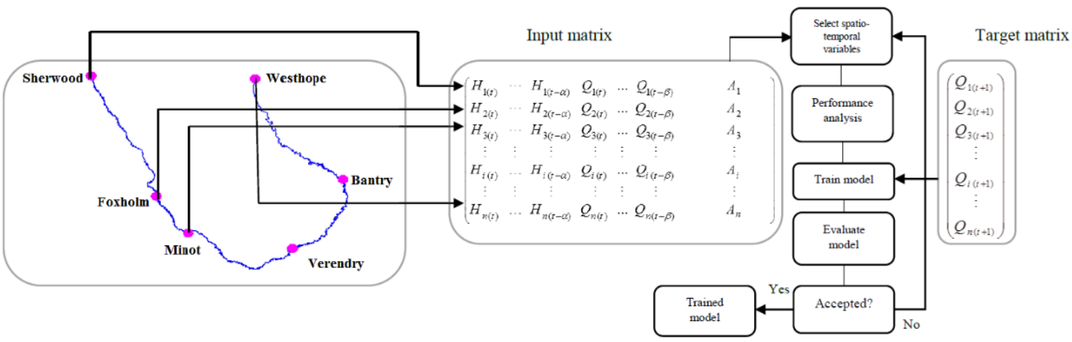

Fig. 2 Schematic of predicting via multi-station model. The figure displays the input data selection and process of prediction

B. Proposed and Applied Methodologies

Since discharge data include broad domain of values, they have to be pre-processed. It is proposed to use the DWT to decompose the original time series into the approximation and the detail components. Secondary decompose was performed by applying the EEMD to the detail components obtained from DWT into a new subseries of Intrinsic Mode Functions (IMFs). The goal of this step is to additionally decrease of the non-stationarity of the detail components captured from DWT. Although decomposing twice might led to required outcome, reproduction of subseries might defect the LSSVM performance. For this reason, mutual information was used to select dominant subseries. Such a feature selection could increase the performance of LSSVM.

Multi-station modeling of river discharge used in this study, was designed to catch the nonlinearity of the river discharge. For this purpose, the multi-station model was established according to the observed Q-H time series. The schematic of proposed model is shown in Fig. 2. LSSVM is the least squares formulation of a SVM. It was first proposed by Suykens and Vandewalle as a classi-fier in 1999. Unlike the inequality constrains in the SVM, LSSVM proposed equality constrains in the formulation [5].

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 temporal features of the hydrometric stations. Therefore, multiple inputs were set in a way that all temporal could establish a unit matrix. The input variables were comprised with different sets of antecedent and current Q-H values of the all stations to forecast the discharge values (Qi(t), i = 1, 2, 3… 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 [7].

The DWT is a popular method and very precise method for time series processing [6], [8]. While the general theory behind DWT is quite analogous to that of the short time Fourier transform (STFT), DWT 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, DWT can be used to analyze the time-frequency characteristics of any kind of time series. In recent years, DWT has been widely used for the analysis of many hydro-meteorological time series [7]. As the mother wavelet moves across the time series during the DWT process, it generates several coefficients that represent the similarity between the time series and the mother wavelet (at any specific scale). A time series is decomposed into details (D) and approximations (A) when using DWT.

EEMD was proposed to solve the mode mixing issue of

Empirical Mode Decomposition (EMD) which specifies the true Intrinsic Mode Function (IMF) as the mean of an ensemble of trials (Wu and Huang, 2009). Each trial consists of the decomposition results of the signal plus a white noise of finite amplitude. Recently developed approach is captured from outcome of recent studies which have proved capability of white noise [9], 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.

In the process of EEMD, a white noise is added to capture a uniform time–frequency space based on its components with various scales. By the time, signal is decomposed via EEMD, due to its uniformly distributed white noise; the components in various scales are transformed to appropriate scales of reference created by the white noise in the background. On the other hand, produced sub-series can be very noisy. However, due to its nature and difference of the noise in each it can be decreased or even completely removed in the ensemble mean of enough trails [4], [10].

In this study, the total data were separated into calibration and verification sets. Three different criteria were selected to meter the revenue of the proposed forecasting methods; the Root Mean Square Error (RMSE) and the determination coefficient (DC). The RMSE and DC were applied to exhibit discrepancies between predicted and observed values [11].

III. Results and Discussion

The rating curve (RC) is an empirical model, which extracts information from recorded stage values. The RC was applied for all 6 hydrometric stations. As an instance, Figs. 3 (A) and (B) demonstrate the RC for Sherwood and Minot stations. Results of RC using least square method led to the following results:

Q(Sherwood)=2.7422H2.8127
Q(Minot)=0.0281H11.313

Results of modeling are demonstrated in Fig. 3 (C). Based on the results it was observed that results needs to be strengthen in terms of DC and RMSE.

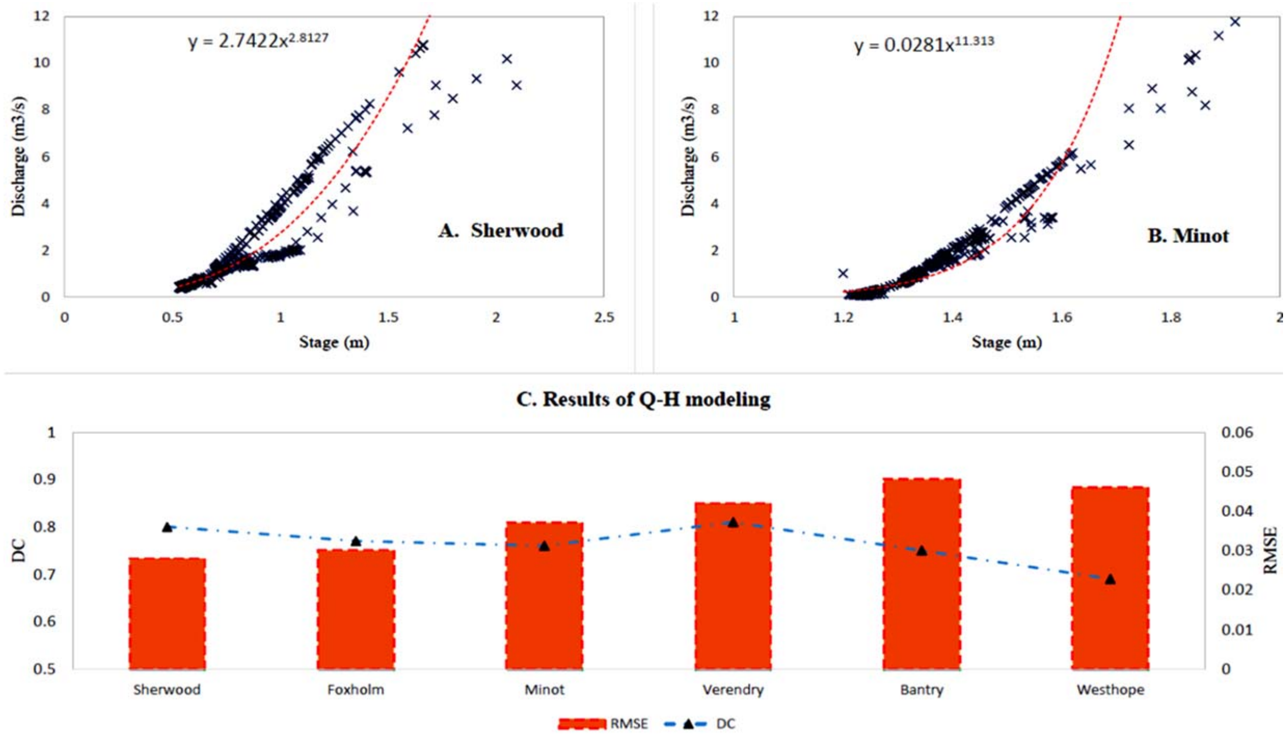

Fig. 3 (A) Results of modeling via RC for Sherwood station, (B) Results of modeling via RC for Minot station, C Overall results of RC for all stations

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 [7]. Therefore, the model construction for each sub-basin could be represented as:

$$Qi(t+1) = f[H_i(t), H_i(t-1), Q_i(t), Q_i(t-1)] \quad (1)$$

In (1), i refers to the sub-basin number (*i = 1, 2, 3 . . . 6*). $H_i(t\text{-}\alpha)$ is river stage values with day lag time at ith sub-basin. In order to develop the multi-station model for the entire watershed, the data of all six stations were imposed to the LSSVM framework. Since the *Q-H* values are sorely effected by recent conditions of watershed at daily time scale, only values with two days lag times (i.e., $\alpha = 2$) were used in the multi-station model.

In order to find the efficient structure of proposed multi-station model, sensitivity analysis was performed. For this end, six best input combinations were considered to create the input matrix (Table II). Table II shows the multi-station model's performance for different input combinations. According to the obtained results, imposing pre-processed temporal dataset caused an increase in modeling accuracy (Combs. (5) and (6)). Conjugating DWT-EEMD to temporal features in Comb.6 donated an incredible insight to temporal features which caused Comb. 6 to have the best outcome. By comparing the obtained results for Combs. (1)–(6) (Table II), it could be inferred that the Comb. (6) was more efficient than others. The input structure of this combination was consisted of two temporal variables. In Table II, 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 is alike.

IV. Concluding Remarks

In this research, capability of LSSVM by using conceptual modeling scenarios were verified to discover their quantitative and qualitative aspects in prediction of river Q-H process. EEMD and WT donated incredible vision in time and frequency domain of data to capture the non-linear and seasonal properties. By capturing more correlated first and second detailed subseries decompositions by MI and imposing it to LSSVM, an increase in performance was observed. Especially in multivariate modeling, because of its sensitive structure, EEMD was applied for all detailed subseries and obtained results showed a good agreement with observed time series.

The present study took advantage multi-station model which was designed to predict the river discharge in multiple-station form by training only one LSSVM model. This strategy was capable of forecasting at the point of interest in a river or watershed by including temporal properties of the case study.

TABLE II
Results of Multi-Station Modeling

Combination	Input matrix	Calibration		Verification	
		DC	RMSE*	DC	RMSE*
1	*Q(t), Q(t-1), H(t)*	0.84	0.022	0.72	0.029
2	*Q(t), Q(t-1), H(t)*	0.86	0.019	0.76	0.026
3	*Q(t), H(t)*	0.86	0.02	0.75	0.027
4	*Q(t), Q(t-1), H(t), H(t-1)*	0.87	0.022	0.82	0.025
5	*Qd3, Qd4, Qd5, Qa*	0.90	0.017	0.85	0.022
6	*Qd1(imf 8,9), Qd2(imf 6,7), Qd3, Qd4, Qd5, Qa, Ha*	0.93	0.014	0.88	0.019

* *RMSE and MAE values are dimensionless due to normalization.*

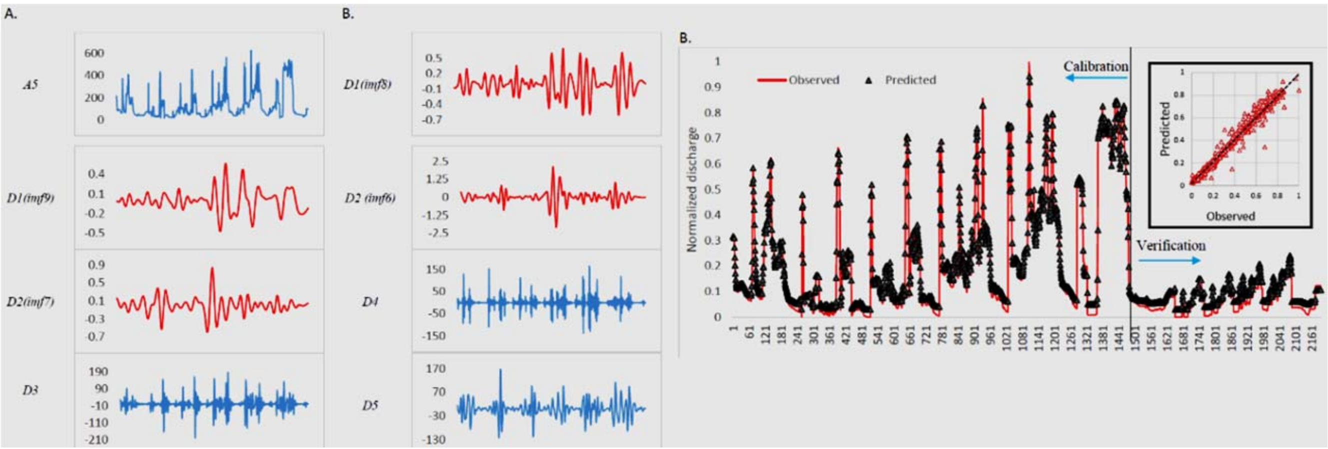

Fig. 4 (A) Decomposed time series using *db4-EEMD-MI* (8 sub-series), (B) multi-station forecasting of Souris River discharge

References

[1] Kisi, O. and Cobenar, M. (2009). "Modeling River Stage-Discharge Relationships Using Different Neural Network Computing Techniques." Clean, Vol. 37 No. 2, PP. 160 – 169.

[2] Ghorbani, M. A., Khatibi, R., Goel, A., Fazeli Fard, M. H. and Azani, A. (2016) "Modeling river discharge time series using support vector machine and artificial neural networks." Environmental Earth Sciences, Vol. 75, PP. 685.

[3] Shannon, C.E. (1948). "The mathematical theory of communications." I and II. Bell System Technical Journal, Vol. 27, PP. 379–423.

[4] Roushangar, K. and 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, Vol. 20 No. 2, PP. 468-485.

[5] Noori, R., Karbassi, A. R., Moghaddamnia, A. et al. 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, Vol. 401, PP. 177-189.

[6] Aussem, A., Campbell, J. and Murtagh, F. (1998). "Wavelet-based feature extraction and decomposition strategies for financial forecasting." Journal of Computational Finance, Vol 6, No. 2, PP. 5-12.

[7] Nourani, V., Komasi, M. and Alami, M. T. (2013). "Geomorphology-based genetic programming approach for rainfall–runoff modeling."

Journal of Hydroinformatics, Vol. 15 No. 2 PP. 427-445.

[8] Farajzadeh, J. and 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, Vol 20, No. 1, PP. 246-262.

[9] Wu, Z. H. and Huang, N. E. (2009). "Ensemble empirical mode decomposition: A noise assisted data analysis method." Advances in Adaptive Data Analysis, Vol. 1, PP.1–14.

[10] Wang, W-C., Xu, D-M., Chau, K. W. and Chen, S. (2013). "Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD." Journal of Hydroinformatics, Vol. 15 No. 4 PP. 1377-1390.

[11] Adamowski, J. and Chan, H.F. (2011). "A wavelet neural network conjunction model for groundwater level forecasting." Journal of Hydrology, Vol. 407, PP. 28–40.