



Input Selection of Wavelet-Coupled Neural Network Models for Rainfall-Runoff Modelling

Muhammad Shoaib^{1,2}  · Asaad Y. Shamseldin² · Sher Khan² · Muhammad Sultan¹ · Fiaz Ahmad¹ · Tahir Sultan³ · Zakir Hussain Dahri^{4,5} · Irfan Ali⁵

Received: 19 March 2018 / Accepted: 13 November 2018 /

Published online: 21 November 2018

© Springer Nature B.V. 2018

Abstract

The use of wavelet-coupled data-driven models is increasing in the field of hydrological modelling. However, wavelet-coupled artificial neural network (ANN) models inherit the disadvantages of containing more complex structure and enhanced simulation time as a result of use of increased multiple input sub-series obtained by the wavelet transformation (WT). So, the identification of dominant wavelet sub-series containing significant information regarding the hydrological system and subsequent use of those dominant sub-series only as input is crucial for the development of wavelet-coupled ANN models. This study is therefore conducted to evaluate various approaches for selection of dominant wavelet sub-series and their effect on other critical issues of suitable wavelet function, decomposition level and input vector for the development of wavelet-coupled rainfall-runoff models. Four different approaches to identify dominant wavelet sub-series, ten different wavelet functions, nine decomposition levels, and five different input vectors are considered in the present study. Out of four tested approaches, the study advocates the use of relative weight analysis (RWA) for the selection of dominant input wavelet sub-series in the development of wavelet-coupled models. The db8 and the dmey (Discrete approximation of Meyer) wavelet functions at level nine were found to provide the best performance with the RWA approach.

Keywords Rainfall-runoff modelling · Artificial neural network · Discrete wavelet transformation · Wavelet sub-series

1 Introduction

The reliable and accurate modelling of the complex phenomenon of rainfall-runoff is vital in the design and management of hydrologic and hydraulic projects such as flood forecasting,

✉ Muhammad Shoaib
msho127@aucklanduni.ac.nz

urban sewer design, drainage system design and catchment management. The transformation of rainfall into runoff is a stochastic process and is reliant on many meteorological parameters and catchment features and therefore many hydrological models have been formulated to simulate this multifarious process. Recently, the ANN has emerged as a powerful black-box model in simulating the process of transformation of rainfall into runoff. Nevertheless, ANN may not be able to deal with non-stationary data without pre-processing of the input and/or output data (Cannas et al. 2006). To overcome this restriction, a hybrid method which involves integration of the WT with the black-box data-driven hydrological models has emerged in hydrology. WT is a mathematical method which enhances the performance of hydrological models by catching the temporal and the spectral information veiled in the data in its raw form. Different hydrological studies used WT in order to enhance accuracy of ANN based hydrological models. A comprehensive appraisal of WT application in hydrology can be found in Nourani et al. (2014).

In spite of good performance of hybrid wavelet ANN models in hydrology, a key limitation related to hybrid wavelet models is the use of a large number of inputs as WT decomposes the input raw data into a number of sub-data series. This not only increases the simulation time but also enhances the computational complexity and makes training of networks harder. Likewise, some of the wavelet transformed data series encompass significant information of the catchment while others may hold no/less information of the system. Consequently, identification of wavelet transformed data sub-series possessing substantial hydrological information is a key issue in development of hybrid wavelet models. This not only affects the accuracy but also computational complexity of the developed hybrid models. One of the most widely-used techniques to identify the information-enriched wavelet data sub-series is the analysis of linear correlation coefficients. Use of linear correlation coefficients for the identification of dominant wavelet sub-series may be further sub-divided into two types. In the first type, a correlation analysis is performed between each decomposed sub-series and the observed output. Only the wavelet sub-series which had a strong correlation with the observed output are considered as the input for the wavelet-coupled ANN models, while those with weak correlation are simply ignored. This approach has been successfully used in some of the hydrological studies. (e.g. Maheswaran and Khosa 2012; Rajae 2011). Likewise, another approach regarding selection of dominant wavelet sub-series based on the cross-correlation analysis has also been employed in the development of ANN and the neuro-fuzzy hydrological models (e.g. Kisi 2011; Kisi and Shiri 2011; Shoaib et al. 2016). In this approach, a correlation analysis is performed between each decomposed sub-series and the observed output. The wavelet sub-series with very weak correlation are ignored and the new input data series is obtained by adding the wavelet sub-series showing strong correlation with the observed output. This new data series is subsequently used as input for the development of wavelet-coupled ANN models. Nevertheless, Nourani et al. (2012) criticize the selection of dominant wavelet sub-series on the basis of linear correlation because a strong non-linear relationship may exist between input and target output despite the presence of weak linear correlation. Likewise, some other studies tested different mathematical methods to select dominant wavelet sub-series in the development of wavelet-coupled models in hydrology.

It is therefore obvious from the above-cited literature that some of the previous hydrological studies used all wavelet sub-series as input, while others identified dominant wavelet sub-series using different methods and subsequently used them as input in the development of wavelet-coupled models. This study therefore aims to compare different methods for the identification of dominant wavelet sub-series for the development of wavelet-based models,

and identify an optimal strategy. The paper is arranged in the following sequence. Section 1 gives the introduction and review of literature. Section 2 describes the methodology used in this study. It contains information about the comprehensive theoretical background of WT, the multilayer perceptron neural network (MLPNN), the development of simple and hybrid wavelet MLPNN rainfall-runoff models, and the performance indices used to evaluate the models developed in this study. The data used in the study is described in Section 3 while results of the different models developed are discussed in Section 4. The conclusions are given in Section 5.

2 Methodology

2.1 Wavelet Transformation

Wavelets are considered as mathematical relations which yield a time-scale depiction of the temporal data and their relationships which are appropriate for examining the data that hold non-stationaries. WT is supposed to have the ability to reveal features of the original data such as trends, breakdown points, and discontinuities that other signal investigation techniques might lack (Singh 2012). As there are many different available references on wavelets, only the main concepts will be introduced here. Continuous Wavelet Transformation (CWT) and Discrete Wavelet Transformation (DWT) are the two types of WT. The CWT of a signal $f(t)$ is defined as follows:

$$W_{a,b}(t) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{a}} \varphi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

where * represents the complex conjugate and $\psi(t)$ is considered as the mother wavelet or wavelet function. In CWT, the whole array of the data is examined by the mother wavelet using the parameters of 'a' and 'b'. The parameters 'a' and 'b' represent the dilation (scale) and translation (position) parameters, respectively. The estimation of CWT coefficients at each scale 'a' and translation 'b' yielded a huge volume of data. This problem was addressed by DWT, which works on low-pass (scaling) and high-pass (wavelet) filter functions. The DWT scales and positions are based on dyadic scales and positions (power of two) and can be defined at any time t for a discrete time series $f(t)$ as;

$$W(a, b)_D = 2^{-j/2} \int_{j=1}^{j=J} \psi^* \left(2^{-j/2-k} \right) f(t) dt \quad (2)$$

where the real numbers, j and k are integers which govern the wavelet dilation and translation respectively. High and low frequencies contained in the temporal data (signal) are examined in DWT by the high- and low-pass filters, respectively.

2.1.1 Choice of Wavelet Families

There are different wavelet functions grouped into various wavelet families. These wavelet functions are identified by their individual features including the region of support and the number of vanishing moments. The region of support of a wavelet function is linked with the span length of the wavelet which affects its feature localization properties of a signal. However, the vanishing moment limits the wavelet's capability to denote polynomial

behaviour or information in data. Earlier hydrological studies have employed various mother wavelet functions. The present study is therefore conducted using ten different wavelet functions from five different wavelet families in order to cover a wide range of wavelet functions. The wavelet functions being utilized in the current study include; Haar, db2, db4, db8, Coif2, Coif4, Sym2, Sym4, Sym8 and the Discrete approximation of Meyer (*dmey*) wavelet function. The details of different wavelet families can be found in many text books such as Daubechies (1992) and Addison (2002).

2.2 Artificial Neural Networks (ANN)

2.2.1 Multilayer Perceptron Neural Network (MLPNN)

There are different types of the Artificial Neural Networks, but the Multilayer Perceptron Neural Network (MLPNN) with back propagation is considered the most widely used ANN in hydrology (Principe et al. 2000). The MLPNN consists of a number of neurons (computational elements) arranged in a series of different layers. Each neuron of MLPNN receives an array of inputs and yields a single output. The output generated by the neuron of input layers is fed as input for the neuron in the next hidden layer. Likewise, the output of the hidden layer neuron is input for the neuron of output layer. A mathematical function known as neuron transfer function processes inputs to each neuron in all three layers of MLPNN to yield output. The neurons in the input layer are connected with the neuron in the hidden layer while the neuron in the output layer is connected to the neuron in the hidden layer.

2.3 Relative Predictor Importance

There are four main statistical methods available to determine the relative predictor importance (Van Iddekinge and Ployhart 2008); (i) examination of regression coefficients, (ii) analysis of correlation coefficients, (iii) dominance analysis and (iv) relative weight analysis (RWA). The first method makes use of regression coefficients in the regression equation. The regression coefficient describes the rate of change of dependent variables as a function of predictor variable(s). Thus the importance of the predictor variable can be estimated by the magnitude and sign of its coefficient value. The second method explores the use of the correlation coefficient values to determine the relative importance of each predictor variable. This index gives an indication of how much of the variance in a dependent variable can be attributed to the selected predictor variable. Lebreton et al. (2004) performed a Monte Carlo study and established that, as the mean validity of the predictors, amount of predictor co-linearity, or the number of predictors increased (beyond 3), the interpretability of the beta and correlation coefficients suffered seriously due to coefficient instability. Thus, they warned against using either of these first two methods in isolation, and instead supported the use of Dominance Analysis (Azen and Budescu 2003; Budescu 1993) or a form of Relative Weight Analysis (RWA), the epsilon (ϵ) statistic (Johnson 2000).

The Dominance Analysis (DA) and the Relative Weight Analysis (RWA) are the recently developed strategies for assessing relative importance of predictor variables. However, Johnson and LeBreton (2004) found a weakness associated with the DA in that as the number of predictor increases, it becomes more computationally difficult because of the exponentially increasing number of sub-models involved. The final strategy for identifying relative importance of predictors is RWA which involves variable transformation of the original predictors

into orthogonal (uncorrelated) variables that are related to dependent variables, but not to each other, which are then related back to the original predictors (Van Iddekinge and Ployhart 2008). One of the major weaknesses of the method of RWA was in estimating the orthogonal variables. The Johnson (2000) epsilon (ϵ) statistic addressed this weakness and has become one of the more recommended methodologies of RWA (Johnson and LeBreton 2004; Van Iddekinge and Ployhart 2008). An epsilon statistic is calculated for each predictor variable for its relative importance. Epsilon can also be easily transformed into a statistic that can be interpreted as the percentage of the model R^2 associated with each predictor. Further details on RWA can be found in many readings (e.g. Johnson 2000; Johnson and LeBreton 2004; LeBreton and Tondiandel 2008).

2.4 Development of Hybrid Wavelet MLPNN Models

The MLPNN models are integrated with the DWT in this study to form hybrid WMLPNN rainfall runoff models. The MLPNN models are developed with and without WT. The daily rainfall data is decomposed into approximation and details using the DWT in this study. Ten different mother wavelet functions are selected, and used to decompose the input rainfall data. In the present study, the MLPNN comprises three layers; input, hidden and output layers. For the development of simple MLPNN models, the neurons in the input and output layers are fixed as one. For the development of wavelet-based WMLPNN models, the selected decomposition level determines the number of neurons in the input layer. The choice of the number of neurons in the hidden layer is vital for better ANN performance. The trial and error method is followed in the present study to choose the number of neurons in the hidden layer similar to several other hydrological studies (e.g. Wang and Ding 2003; Singh 2012; Kisi et al. 2013). The neuron transfer function for the neurons of the hidden and output layers is selected as a sigmoid function for both the simple MLPNN and WMLPNN models. The Levenberg-Marquardt algorithm (LMA) is used in the present study for training the simple and wavelet-based models because of its simplicity. The criteria selected for stopping training of the developed models in the present study is either a maximum of 100 epochs, or when the mean squared error (MSE) of the cross-validation testing data set begins to rise. This is a sign that the network has started to over-train. Testing the trained network is the next step, which is done by presenting a changed set of testing data it has not been trained with., If it failed to perform satisfactory during testing stage, the network is re-trained by altering the number of neurons in the hidden layer. Network testing confirms that it has learned the general patterns of the system and has not simply memorized a given set of data.

2.5 Performance Parameters

In this study, the performance of the developed rainfall-runoff models is determined by using two statistical parameters, namely, the Root Mean Squared Error (RMSE) and the Nash-Sutcliffe Efficiency (NSE) Nash and Sutcliffe (1970). These are defined by the following equations:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N (Q_{obs} - Q_{mod})^2} \quad (3)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_{obs} - Q_{mod})^2}{\sum_{i=1}^N (Q_{obs} - \overline{Q_{obs}})^2} \quad (4)$$

where Q_{obs} is the observed discharge and Q_{mod} is the modelled discharges while $\overline{Q_{obs}}$ is the mean of observed discharge of training data. These two statistical parameters can satisfactorily be used to assess the performance of hydrological models. (Legates and McCabe 1999).

3 Data

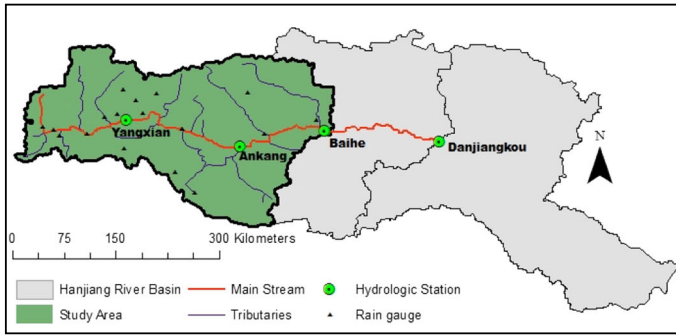
Rainfall-runoff data of two catchments - namely, Baihe and Brosna - situated in various climatological conditions in the world is used in the study. The Baihe catchment is located in north-eastern China and the Brosna catchment located in Ireland. The area of the Baihe catchment is 61,780 km² and it is a sub-basin of Upper Hanjian River basin as shown in Fig. 1a. The topography of this basin is mountainous and it is semi-arid in nature. The daily rainfall data from 19 rain gauge stations and three hydrological stations for a period of 8 years starting from 1st January 1972 onwards is used to obtain daily average rainfall. The corresponding discharge data used in the study is the daily averaged data measured at the Baihe station. The total drainage area of the Brosna catchment is 1207 km² up to the Ferbane flow gauging station. The catchment contains very flat topography except for some undulations caused by glacial deposits. Ten years' daily rainfall data for a period from 1st January 1969 to 31st December 1978 are lumped by averaging the data from four rain gauge stations located in the Brosna basin. The corresponding discharge data used in the study is the daily averaged data measured at the Ferbane station. The first 8 years' data from both catchments is used for training purposes, while the remaining 2 years' data is used for the purpose of testing. The location of the catchment along with rainfall gauging sites is shown in Fig. 1b.

4 Results and Discussion

4.1 Selection of Input Vector

The performance of rainfall-runoff data-driven models is heavily reliant on the use of suitable input vectors, as current responses of hydrological systems are intrinsically dependent on their preceding states. So the use of past data is essential in order to encrypt temporal and spatial features of the data. The present study therefore explores the size of input vector by considering the following five input vectors to predict runoff at the current time;

1. M_1 $r(t-1)$
2. M_2 $r(t-1), r(t-2)$
3. M_3 $r(t-1), r(t-2), r(t-3)$
4. M_4 $r(t-1), r(t-2), r(t-3), r(t-4)$
5. M_5 $r(t-1), r(t-2), r(t-3), r(t-4), r(t-5)$



(a)



(b)

Fig. 1 Location of the study area a Baihe, b Brosna

4.2 Selection of Decomposition Level

The performance of the hybrid wavelet models is very much dependent on the selection of appropriate decomposition level. Since numerous seasonal and cyclic features may be concealed in the hydrological data, a thorough insight into the procedure and attention to the periodicity features might be useful in the choice of a suitable decomposition level for the DWT. The total number of decomposition levels depends on the available data length. A DWT decomposition comprises $\text{Log}_2(N)$ levels at maximum where N is the total number of available data points. As there is only 2-year daily data available for testing in the current study, it therefore can be decomposed from level one to level nine at most. Level one decomposition generated one approximation (a_1) and one detail (d_1). One approximation (a_2) and two details (d_1 and d_2) are produced at level two decomposition, one approximation (a_3) and three details (d_1 , d_2 and d_3) are resulted at level three decomposition, and so on. This results in a total of 180 decomposition levels for the two selected catchments in this study. Regression analysis is then performed with the decomposed input data and the observed output for all 180 decomposition levels. It is found that R^2 value increases with the increase of decomposition level, and maximum value is obtained at a maximum possible decomposition level of nine in the present study for the two catchments with all ten selected wavelet functions.

Consequently, level nine decomposition was selected in the present study. This is also in agreement with numerous preceding studies such as Shoab et al. (2014, 2016). Level nine decomposition resulted in one approximation sub-series (a_9) and nine detail sub-series ($d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8$ and d_9). The detail $d(n)$ characterizes high- frequency, low- scale contents of the data while low- frequency, high- scale contents are represented by the approximation $a(n)$. The sub-series d_1 relates to the time series of 2-day mode, which denotes the characteristics of original raw data discernible at a scale of up to 2 days. The detail sub-series d_2 relates to 4-day mode, d_3 relates to 8-day mode and can catch characteristics of input data on about a weekly basis. The detail d_4 corresponds to 16-day mode, d_5 to 32-day mode (about monthly mode), d_6 to 64-day mode, d_7 to 128-day mode (about 4 months), d_8 to 256-day mode (about eight and half months) and d_9 to 512-day mode (about 17 months). Furthermore, detail sub-series d_8 and d_9 are responsible for catching seasonal and cyclic variations in the input rainfall data on an approximately annual scale. This annual periodicity is considered a very important and leading seasonal cycle in the hydrologic time series data.

4.3 Selection of Appropriate Decomposed Sub-Series

To develop WMLPNN models, daily observed rainfall data is decomposed into approximation (low frequency and large scale) and details (high frequency and small scale) using the DWT and ten selected mother wavelet functions. In order to see the effect of different wavelet sub-series on the performance of the WMLPNN models, the following four options were considered.

Option 1: Starting with 1-day lagged rainfall data $r(t-1)$, followed by 2-day lagged rainfall data $r(t-2)$, and so on up to 6-day lagged rainfall data $r(t-6)$ are transformed using DWT and ten selected wavelet functions. A correlation analysis is then performed between each decomposed sub-series of all lagged rainfall data and the observed discharge for all ten mother wavelet functions and for the two selected catchments used in this study. The correlation results reveal (not shown here) that the d_1 and d_2 sub-signals have a very poor correlation with the observed discharge for all ten selected mother wavelet functions for the two selected catchments. Consequently, d_1 and d_2 , sub-series are eliminated and a new series is formed by adding $d_3, d_4, d_5, d_6, d_7, d_8, d_9$ and a_9 . This approach is commonly used in the implementation of neuro-fuzzy hydrological models (e.g. Partal and Kisi 2007; Shiri and Kisi 2010; Shoab et al. 2016)

Option 2: The WMLPNN models are developed using all the sub-series obtained by wavelet transformation irrespective of their linear correlation with the observed discharge. This is the most common approach used in wavelet-ANN studies (Shoab et al. 2018).

Option 3: The WMLPNN models are developed using all the wavelet sub-series except the ones having very weak correlation with the observed discharge. These are d_1 and d_2 in the present study.

Option 4: The input wavelet sub series are selected based on the results of RWA for the development of WMLPNN models.

At first, a feed forward MLPNN with back propagation training algorithm is developed without any WT to model the rainfall-runoff transformation process of the two selected catchments. For this, the observed daily rainfall and daily discharge is used as input and target output respectively. The MLPNN developed without WT are referred to as Simple MLPNN

models in this study. Afterwards, the WMLPNN models are developed with DWT using ten selected mother wavelet functions for the first three options stated above in order to evaluate the effect of different wavelet functions with different input vectors. In order to compare performance of WMLPNN models relative to their counterpart simple MLPNN models, the relative percentage increase or decrease in the NSE and the RMSE values are calculated for both selected catchments with ten different wavelet functions and the results are shown in Tables 1 and 2, respectively.

Examination of Table 1, which presents the relative percentage change in NSE for the Baihe catchment, reveals that with input vector M_1 , the best results (shown in bold) are obtained with the db2, db8, sym2, coif2 and the dmey wavelet functions which enhance the NSE values by about 16–18% relative to their counterpart simple MLPNN model. The wavelet functions db8, coif4 and dmey wavelet functions performed best and enhances the relative NSE (%) values by about 50% with input vector M_2 which corresponds to the lead time of 2 days. The WMLPNN models developed with the input vector M_3 , which corresponds to the lead time of 3 days, yielded best results with the db4, db8 and dmey wavelet functions in option 1, but these results were poor compared with the M_2 input vector. Furthermore, the negative change in NSE percentage with input vector M_4 and M_5 with almost all the WMLPNN models shows a poor performance of the WMLPNN models relative to their respective simple MLPNN models for option 1. The performance of the WMLPNN models relative to their respective simple MLPNN models is best with input vector M_2 . A similar trend of relative percentage change in the RMSE values of the developed WMLPNN models for option 1 can be found in Table 2. The improvement in the RMSE values (negative values) shows the WMLPNN model's ability to capture the high flow trails of the observed hydrograph. The best results (shown in bold) for WMLPNN models are achieved with the input vector M_2 with wavelet functions db8 and dmey which improves the relative RMSE values by about 30% as shown in Table 2 for Baihe catchment.

For option 2, the WMLPNN models are developed by considering all the sub-series obtained by WT. Contrasting wavelet functions performed differently and an examination of Tables 1 and 2 shows that the best results are again obtained with input vector M_2 with wavelet functions db8 and dmey. The percentage change in NSE value is found to be about 85% for WMLPNN models developed with db8 and dmey wavelet functions relative to their counterpart simple MLPNN models. Similarly, the relative percentage change in RMSE values for the best models with M_2 input vector is found to be in the range of about 60% for the best models, as shown bold in Table 2 for Baihe catchment. It can be further revealed that for option 2 with all the sub-series used in the development of WMLPNN, considerable improvement of WMLPNN models relative to their counterpart simple MLPNN models is found with all the five input vectors considered in the present study. This is in contrast to option 1 where improvement in the results is found only with input vector M_1 and M_2 .

Option 3 used all sub-series obtained by the WT except d_1 and d_2 , which both had weak correlation with the observed discharge for the development of WMLPNN models. The best results are shown in bold in Tables 1 and 2. The tables show the WMLPNN models developed with the M_2 with the wavelet functions db8 and dmey. A relative improvement of about 70 and 47% is found in the NSE and RMSE values respectively for the WMLPNN models developed with the db8 and the dmey wavelet functions as shown (bold) in Tables 1 and 2. Similar to option 2, the performance of best WMLPNN models relative to their counterpart simple MLPNN models is better for all the five input vectors considered in this study.

Table 1 Relative percent change in NSE for Baihe catchment

Model	Haar		db2		db4		db8		sym2		
	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	
Option 1											
M1	11.42	10.68	19.79	17.60	12.99	11.07	18.66	14.78	19.79	17.23	
M2	38.27	40.74	29.63	35.78	46.99	37.68	55.17	52.06	33.02	25.42	
M3	-4.99	-18.90	-16.15	-6.31	29.00	29.66	36.06	28.41	-34.47	-13.30	
M4	-27.83	-55.09	-66.36	-84.32	-14.10	-12.80	12.19	-8.34	-35.85	-53.42	
M5	-56.70	-49.28	-72.09	-2.99	-38.17	11.72	-7.98	7.88	-38.68	-24.83	
Option 2											
M1	35.34	38.36	52.79	45.84	33.97	31.16	48.86	45.44	47.27	38.21	
M2	59.15	52.78	76.15	70.58	62.21	62.77	85.61	79.92	68.57	61.80	
M3	31.79	28.87	53.17	49.15	63.50	55.15	72.43	68.93	53.11	49.16	
M4	40.93	14.59	36.51	16.40	29.79	20.55	62.69	43.76	35.89	17.77	
M5	27.03	26.02	29.97	-0.22	36.39	6.89	36.60	28.34	11.79	-52.57	
Option 3											
M1	41.65	42.20	33.53	28.65	29.96	27.27	56.20	51.35	32.66	27.97	
M2	40.80	43.59	53.89	50.96	59.17	58.75	72.66	66.46	53.89	50.96	
M3	-2.36	11.51	6.03	15.45	28.36	40.33	47.29	49.73	6.61	14.43	
M4	-21.43	-31.98	-30.99	-38.52	21.71	-1.64	25.14	27.04	-30.99	-38.52	
M5	-35.06	-27.91	-59.07	-19.09	-11.25	24.29	7.72	7.75	-57.57	-56.30	
Option 4											
Model											
		sym4		sym8		coef2		coef4		dmev	
		Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.
Option 1											
M1	14.42	14.14	16.31	14.74	16.99	16.47	13.24	12.50	16.95	16.15	
M2	42.94	46.81	49.79	46.99	47.45	39.17	51.32	54.35	53.58	55.90	
M3	9.12	19.70	35.30	42.21	24.88	21.00	27.99	37.61	41.19	40.94	
M4	-10.55	-35.18	15.38	-0.76	-22.59	-56.84	4.67	-10.77	16.67	-10.20	
M5	-4.81	26.20	0.01	1.37	-29.57	1.01	-6.82	-7.61	-7.76	-1.15	
Option 2											

Table 1 (continued)

Model	sym4		sym8		coif2		coif4		dmey	
	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.
M1	38.16	35.88	54.70	49.55	76.10	39.15	75.81	47.49	40.78	38.19
M2	72.71	74.36	71.77	73.79	79.45	70.91	75.75	69.44	85.34	83.34
M3	74.56	74.84	64.00	60.87	71.70	64.78	63.35	60.99	67.27	74.65
M4	64.46	41.08	55.61	37.89	51.24	35.31	47.06	26.59	52.88	55.75
M5	39.36	3.90	24.21	2.00	-5.24	4.16	-14.41	-2.25	40.93	23.17
Option 3										
M1	56.75	50.33	46.99	47.10	33.41	32.96	38.14	40.07	34.96	32.75
M2	72.95	74.30	76.86	75.32	60.78	63.25	70.86	68.04	76.15	73.16
M3	55.16	61.12	54.79	55.70	41.62	43.44	44.81	46.12	53.99	52.28
M4	16.96	15.09	8.61	1.95	15.14	11.34	16.77	0.14	32.01	12.76
M5	3.94	49.84	11.32	-26.62	7.97	14.60	-1.78	-24.21	7.25	10.28

Tr Training data set, Ts Testing data set

Table 2 Relative percent change in RMSE for Baihe catchment

Model	Haar		db2		db4		db8		sym2	
	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.
Option 1										
M1	-19.67	-21.49	-31.36	-26.41	-6.81	-5.70	-9.90	-7.68	-10.53	-9.02
M2	-36.08	-29.05	-51.16	-42.49	-27.20	-21.05	-33.05	-30.76	-18.16	-13.64
M3	-17.41	-11.20	-31.57	-20.52	-15.74	-16.13	-20.04	-15.39	15.96	6.44
M4	-23.14	-7.58	-20.32	-8.57	6.82	6.21	-6.29	4.09	16.56	23.86
M5	-14.58	-13.99	-16.32	-13.89	17.55	16.50	3.91	11.82	17.76	24.93
Option 2										
M1	-19.67	-21.49	-31.36	-26.41	-18.82	-17.03	-28.56	-26.13	-27.46	-21.39
M2	-36.08	-29.05	-51.16	-42.49	-38.53	-36.20	-62.07	-51.24	-43.94	-35.46
M3	-17.41	-11.20	-31.57	-20.52	-39.59	-23.63	-47.49	-31.66	-31.53	-20.53
M4	-23.14	-7.58	-20.32	-8.57	-16.21	-10.87	-38.92	-25.01	-19.93	-9.32
M5	-14.58	-13.99	-16.32	-13.89	-20.24	-16.91	-20.38	-29.66	-6.08	-13.12
Option 3										
M1	-23.68	-23.97	-18.55	-15.53	-16.39	-14.72	-33.88	-30.25	-18.02	-15.13
M2	-23.06	-24.89	-32.10	-29.97	-36.10	-35.77	-47.71	-42.09	-32.10	-29.97
M3	1.17	-5.93	-3.06	-8.05	-15.36	-22.75	-27.40	-29.10	-3.36	-7.49
M4	10.19	14.88	14.45	17.69	-11.52	0.81	-13.48	-14.58	14.45	17.69
M5	16.22	13.10	26.12	23.42	5.48	7.39	-3.94	3.15	25.53	28.96
Model										
Model	sym4		sym8		coef2		coef4		dmey	
	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.
Option 1										
M1	-7.58	-7.34	-8.61	-7.66	-8.98	-8.60	-6.94	-6.46	-8.95	-8.43
M2	-24.46	-27.07	-29.14	-27.19	-27.51	-22.01	-30.23	-32.43	-31.87	-26.07
M3	-4.67	-10.39	-19.57	-23.98	-13.33	-11.12	-15.14	-21.01	-23.32	-23.15
M4	5.14	16.27	-8.01	0.38	10.72	25.24	-2.36	5.25	-8.71	4.98
M5	2.38	7.32	-0.01	6.59	13.83	6.05	3.36	10.01	3.81	10.64

Table 2 (continued)

Model	sym4		sym8		coif2		coif4		dmey	
	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.	Tr.	Ts.
M1	-21.44	-19.92	-32.76	-28.97	-22.54	-21.99	-29.05	-27.54	-23.12	-21.38
M2	-47.76	-69.45	-46.86	-45.31	-45.97	-42.77	-45.64	-41.52	-61.71	-54.95
M3	-49.56	-35.66	-40.00	-26.79	-37.77	-29.08	-32.39	-26.86	-42.79	-35.52
M4	-40.39	-23.24	-33.37	-21.19	-34.65	-19.57	-25.63	-14.32	-31.36	-33.48
M5	-22.13	-14.83	-12.94	-15.68	-21.96	-17.46	-18.69	-16.53	-23.14	-26.84
Option 3										
M1	-34.30	-29.53	-27.26	-27.27	-37.38	-39.38	-46.02	-43.47	-19.43	-17.99
M2	-47.99	-49.25	-51.90	-50.32	-23.59	-24.79	-25.71	-26.60	-51.16	-48.19
M3	-33.04	-37.64	-32.76	-33.44	-7.88	-5.84	-8.77	-0.07	-32.17	-30.92
M4	-8.87	-7.85	-4.40	-0.98	-4.07	-5.03	0.89	5.85	-17.54	-6.60
M5	-1.99	-8.67	-5.83	2.77	0.00	0.00	0.00	0.00	-3.69	0.26

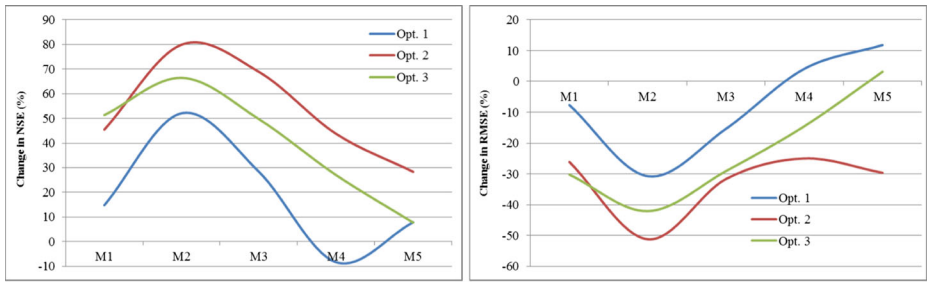
Tr: Training data set, Ts: Testing data set

The percentage change in the NSE and the RMSE values of the developed WMLPNN models relative to their counterpart simple MLPNN models for the Brosna catchment is calculated next for all the first three options considered (results not shown here). The results revealed that percentage improvement in the NSE values is found only with input vectors M_1 , M_2 and M_3 and is in the range of about 6 to 10% for the best WMLPNN models. However, the relative improvement in the percentage change in RMSE values for all the best models is very marginal in the range of about 2 to 5%. For option 2 and 3, the best WMLPNN models are developed with the wavelet functions db8 and dmev. The above analysis suggests that different wavelet functions behaved differently for each input vector. But of all ten wavelet functions tested, the db8 and the dmev wavelet functions performed best for all the input vectors and for the two selected catchments.

To further elaborate the effects of DWT on different input vectors, the performance of the WMLPNN models with the db8 and the dmev wavelet functions relative to their respective simple MLPNN models for each input vector is measured, and the testing results are shown in Fig. 2 for Baihe catchment. Examination of Fig. 2 for the Baihe catchment reveals that the performance of WMLPNN models developed under options 2 and 3 was better for all input vectors M_1 to M_5 with the two wavelet functions. However, WMLPNN models developed under option 2 performed poorly with M_4 input vector, in that they produced negative NSE and RMSE relative percentage values. The best results are obtained with option 2 with input vector M_2 . It is further shown in Fig. 2 that option 1 produced the poorest results among the three options considered. The results for the Brosna (not shown here) catchment for the two best wavelet functions db8 and dmev revealed that the performance of the WMLPNN models under option 1 with both db8 and dmev wavelet functions are very poor compared with the WMLPNN models developed under option 2 and 3. The performance of the WMLPNN models developed for options 2 and 3 is almost similar with the best performance obtained with input vector M_3 . On the basis of the above analysis, it can be concluded that options 2 and 3 give best results for both catchments. However, WMLPNN models give best results with input vectors M_2 for Baihe catchment, and M_3 for Brosna catchment. This may be due to different time of focus and lag-times for both selected catchments.

It can also be concluded on the basis of above analysis that the selection of input wavelet sub-series based on the results of a correlation coefficient between the selected input wavelet series and the observed discharge did not provide satisfactory results. Therefore, option 4 is tested next. Only the sub-series obtained by the WT using the best identified wavelet functions db8 and dmev are used for Option 4. RWA of the input wavelet sub-series was conducted using RWA-Web (Tonidandel and LeBreton 2011) for both selected catchments in order to find the relative importance of all sub-series obtained by the WT. Individual relative weights confidence intervals (Johnson 2004) and all relevant significance tests were based on bootstrapping with 10,000 replications as suggested by Tonidandel et al. (2009). As suggested by Tonidandel et al. (2009), bias-corrected and accelerated-confidence intervals were used because of their higher coverage accuracy. Furthermore, 95% confidence intervals (CI) were used in all cases. Typical results of RWA for Baihe catchment for lagged 2-day wavelet-transformed rainfall data series are shown in Table 3. These results show that a weighted linear combination of our ten transformed sub-series explained 44% of the variance in the observed discharge. The raw weight provides estimates of variable importance. These weights denote an additive decomposition of the total model R^2 and can be interpreted as the proportion of variance in observed discharge that is appropriately attributed to each sub-series obtained by the WT. By summing the raw weights of all the predictors, the total model R^2 can be

(a)



(b)

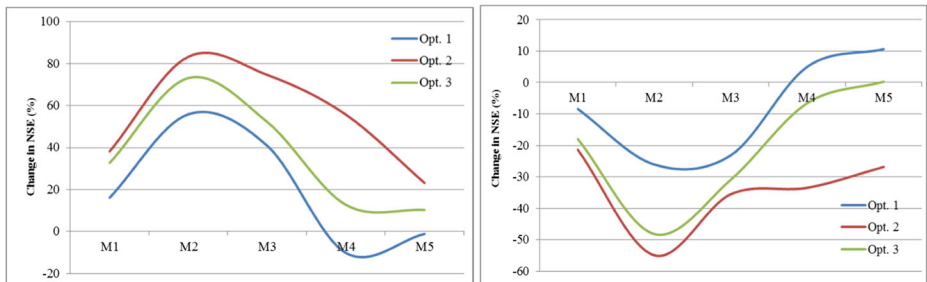


Fig. 2 Effect of input vector on the relative performance of DWTMLPNN models for Baihe catchment **a** db8 **b** dmey wavelet functon

determined. Each relative weight is divided by the model R^2 to find the rescaled weights. These rescaled weights provide assessments of relative importance using the metric of percentage of predicted variance attributed to each variable. For example, the d_4 sub-series explains about 10% of the predicted variance in the observed discharge. The next column in the table provides confidence intervals (CIs) around the raw weights. These CIs are useful for describing the precision of the raw relative weights. Larger values indicate less precision and smaller values indicate greater precision. The confidence interval tests of significance is given in the last two columns of the Table 3, which provide information about the statistical significance of the raw relative weights obtained by calculating bias-corrected and accelerated

Table 3 Relative weight analysis results

$R^2 = 0.442$			CI around raw weight		CI test of significance	
Variable	Raw weight	Rescaled weight	Lower CI	Upper CI	Lower CI	Upper CI
d_1	0.001	0.152	0.000	0.006	0.012	0.023
d_2	0.000	0.011	0.000	0.000	-0.004	0.006
d_3	0.042	9.564	0.015	0.082	-0.007	0.001
d_4	0.048	10.859	0.025	0.081	0.015	0.082
d_5	0.077	17.334	0.056	0.100	0.026	0.080
d_6	0.080	18.004	0.056	0.107	0.055	0.099
d_7	0.062	14.119	0.039	0.086	0.055	0.106
d_8	0.107	24.202	0.088	0.128	0.038	0.086
d_9	0.007	1.599	0.004	0.011	0.087	0.127
a_9	0.018	4.156	0.013	0.024	0.002	0.011

CIs as described by Tonidandel et al. (2009). Examination of the relative weights revealed that all ten predictor variables are found to be statistically significance as none of the 95% CIs for the tests of significance contained zero.

Based on results of RWA approach, wavelet sub-series which had the least importance in explaining the variance of the observed discharge are identified. The new WMLPNN models with db8 and dmey wavelet functions are then developed by excluding those wavelet sub-series with less relative importance (up to 5% rescaled weigh value). The percentage improvement in the NSE and the RMSE values relative to the respective simple MLPNN models are then calculated. The results of WMLPNN models developed using all wavelet sub-series (All) and with wavelet sub-series identified based on the RWA are presented in Fig. 3a for Baihe and 3b for Brosna catchments during testing. It can be seen from Fig. 3a that WMLPNN models developed using all wavelet sub-series obtained by the dmey wavelet function performed poorly in terms of percentage change in NSE and RMSE among all the models developed. Furthermore, the other three models developed performed similarly. The best results in terms of percentage change in NSE and the RMSE are achieved with WMLPNN models using RWA wavelet series obtained by the dmey wavelet function. Figure 3b shows the performance in terms of percentage change in NSE and the RMSE for the Brosna catchment. The figure shows that all the developed models, including those developed using all wavelet sub-series, and those developed using wavelet sub-series identified on the basis of RWA, performed similarly. It can now be concluded on the basis of the above analysis that the WMLPNN models developed using wavelet sub-series identified on the basis of RWA yielded results that are comparable with the models developed using all wavelet sub-series. RWA can be used for identification of wavelet-series with least importance and these can be neglected without loss of performance of the WMLPNN models.

In order to ascertain the ability of modelled hydrographs to trail low-, medium- and high-flow features of the hydrograph, flow duration curves (FDCs) of the best selected simple and

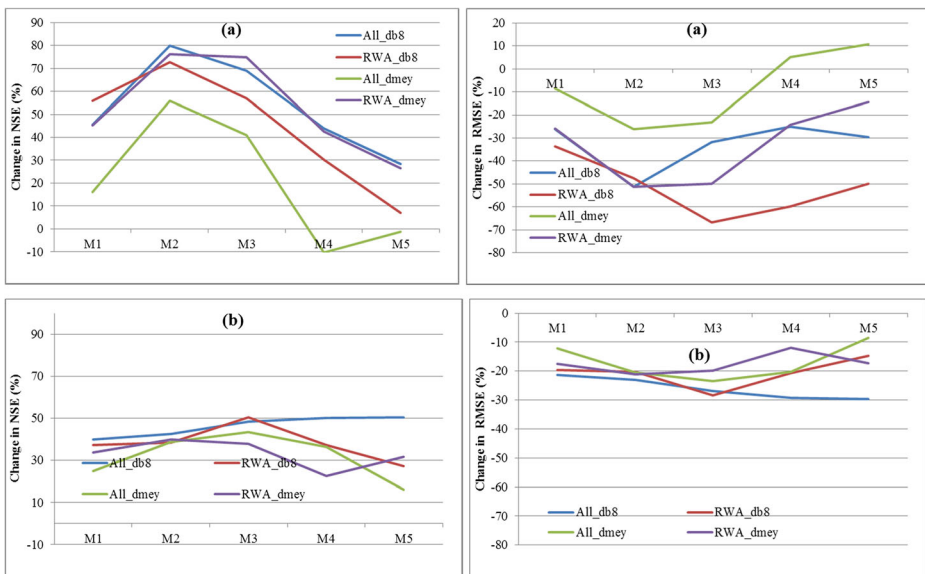


Fig. 3 Comparison of DWTMLPNN models developed using all wavelet sub-series and using selected sub-series **a** Baihe **b** Brosna catchments

wavelet-coupled models for both the selected catchments are shown in Fig. 4. FDC shows the percentage of time a given flow was equalled or exceeded for a given time period.. From the FDCs, the 10th percentile flow, the flow that is equalled or exceeded 10% of the period of record, can be considered as high-flow percentile. Similarly, the 11th to 89th percentile flow is considered as a medium-flow percentile, while 90th percentile flows are considered as low-flow percentile. The medium-flow percentile can be further divided into high-medium flow and low-medium flow from percentiles 11 to 49 and 50 to 89 respectively. It is obvious from flow duration curves that modelled hydrographs generated by the WMLPNN models using all wavelet sub-series, and the ones developed with wavelet sub-series based on the RWA, are equally good. For Baihe catchment, both simulated hydrographs captured well the high-, high-medium, low-medium and low-flow trails of the observed hydrograph. The FDCs generated by the simple ANN model underestimate the observed flow during medium-flow range, while they overestimate the observed flow during low-flow trails of the observed hydrograph. It is further evident from Fig. 4b that the performance of the both simulated hydrographs is equal in capturing low-, medium- and high-flow trails of the observed hydrograph. However, both the simulated hydrographs overestimated the flows only during the low-medium flow trails of the observed hydrograph. The FDC generated by simple ANN model yield underestimated flows during the high- and medium-flow trails and overestimated flows during the low-flow trails of the observed hydrographs. It can therefore be concluded that the hybrid wavelet ANN models out-performed their counterpart simple ANN models for the two selected catchments and modelled hydrographs have the ability to trail well the low-, medium- and high-flow features of the observed hydrograph.

5 Conclusions

The following conclusions are hereby drawn from the study:

- The study favoured the use of dominant wavelet sub-series only as input for the wavelet-coupled models instead of using all sub-series of data obtained by the WT. This considerably reduces the computational complexity and the simulation time as well. The dominant wavelet sub-series identified on the basis of RWA performed better than those identified on the basis of correlation analysis. This may be due to the fact that cross-correlation reveals a linear relationship embedded in the data which is not shown by the non-linear rainfall-runoff transformation process.

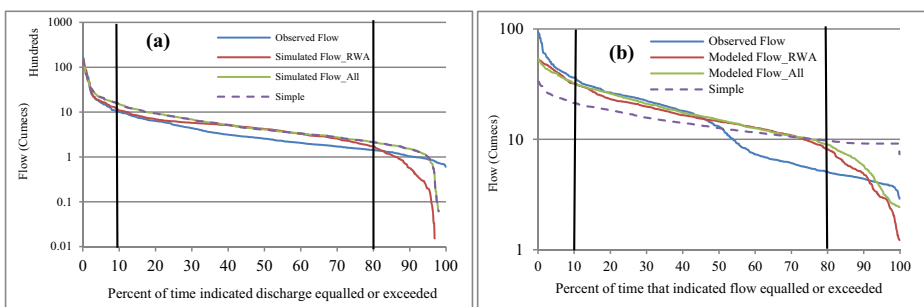


Fig. 4 Comparison of Flow duration curves **a** Baihe **b** Brosna catchments

- Hybrid wavelet models are found to outperform their counterpart simple models with parsimonious input vectors. The best results are found with input vector M_2 (containing up to 2-day lagged rainfall data series) and M_3 (containing up to 3-day lagged rainfall data series) for the Baihe and Brosna catchments respectively. This may be due to different times of concentration and the antecedent moisture conditions of the selected catchments.
- Out of the ten wavelet functions tested in this study, the db8 and dmev wavelet functions performed best in both selected catchments for all five input vectors considered. This may be due to good time-frequency localization properties of the db8 and the dmev wavelet functions.
- The study advocated the use of level nine decomposition of the input data for the development of hybrid wavelet models. The level nine decomposition yielded the best results as it is able to capture the significant and leading annual seasonal cycle of the hydrological time series data.

The best developed models performed consistently during both training/calibration and testing/validation periods for both selected catchments located in different hydro-climatic regions of the world. However, the results of the study should be validated on some more catchments with bigger data sets.

Compliance with Ethical Standards

Conflict of Interest None.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Addison PS (2002) The illustrated wavelet transform handbook. Institute of Physics Publishing, London
- Azen R, Budescu DV (2003) The dominance analysis approach for comparing predictors in multiple regression. *Psychol Methods* 8(2):129
- Budescu DV (1993) Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. *Psychol Bull* 114(3):542
- Cannas B, Fanni A, See L, Sias G (2006) Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Physics and Chemistry of the Earth, Parts A/B/C* 31(18):1164–1171. <https://doi.org/10.1016/j.pce.2006.03.020>
- Daubechies I (1992) Ten lectures on wavelets (CBMS-NSF regional conference series in applied mathematics), vol 61. Society for Industrial and Applied mathematics, Philadelphia
- Johnson JW (2000) A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivar Behav Res* 35(1):1–19
- Johnson JW (2004) Factors affecting relative weights: the influence of sampling and measurement error. *Organ Res Methods* 7(3):283–299
- Johnson JW, LeBreton JM (2004) History and use of relative importance indices in organizational research. *Organ Res Methods* 7(3):238–257
- Kisi O (2011) Wavelet regression model as an alternative to neural networks for river stage forecasting. *Water Resour Manag* 25(2):579–600. <https://doi.org/10.1007/s11269-010-9715-8>
- Kisi O, Shiri J (2011) Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. *Water Resour Manag* 25(13):3135–3152

- Kisi O, Shiri J, Tombul M (2013) Modeling rainfall-runoff process using soft computing techniques. *Comput Geosci* 51(0):108–117. <https://doi.org/10.1016/j.cageo.2012.07.001>
- Lebreton JM, Ployhart RE, Ladd RT (2004) A Monte Carlo comparison of relative importance methodologies. *Organ Res Methods* 7(3):258–282
- LeBreton JM, Tonidandel S (2008) Multivariate relative importance: extending relative weight analysis to multivariate criterion spaces. *J Appl Psychol* 93(2):329–345
- Legates DR, McCabe GJ (1999) Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35(1):233–241
- Maheswaran R, Khosa R (2012) Comparative study of different wavelets for hydrologic forecasting. *Comput Geosci* 46(0):284–295. <https://doi.org/10.1016/j.cageo.2011.12.015>
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models. Part 1: a discussion of principles. *J Hydrol* 10(3):282–290
- Nourani V, Komasi M, Alami MT (2012) Hybrid wavelet–genetic programming approach to optimize ANN modeling of rainfall–runoff process. *J Hydrol Eng* 17(6):724–741
- Nourani V, Hosseini Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. *J Hydrol* 514:358–377
- Partal T, Kişi Ö (2007) Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *J Hydrol* 342(1–2):199–212. <https://doi.org/10.1016/j.jhydrol.2007.05.026>
- Principe JC, Euliano NR, Lefebvre WC (2000) *Neural and adaptive systems*. Wiley, New York
- Rajae T (2011) Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. *Sci Total Environ* 409(15):2917–2928. <https://doi.org/10.1016/j.scitotenv.2010.11.028>
- Shiri J, Kisi O (2010) Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. *J Hydrol* 394(3):486–493
- Shoaib M, Shamseldin AY, Melville BW (2014) Comparative study of different wavelet based neural network models for rainfall–runoff modeling. *J Hydrol* 515:47–58
- Shoaib M, Shamseldin AY, Melville BW, Khan MM (2016) Hybrid wavelet neuro-fuzzy approach for rainfall-runoff modeling. *J Comput Civ Eng* 30(1):04014125. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000457](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000457)
- Shoaib M, Shamseldin AY, Khan S, Khan MM, Khan ZM, Sultan T, Melville BW (2018) A comparative study of various hybrid wavelet feedforward neural network models for runoff forecasting. *Water Resour Manag* 32(1):83–103. <https://doi.org/10.1007/s11269-017-1796-1>
- Singh R (2012) Wavelet-ANN model for flood events. In: Deep K, Nagar A, Pant M, Bansal JC (eds) *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011)* December 20–22, 2011, vol 131. Springer Berlin, Heidelberg, pp 165–175
- Tonidandel S, LeBreton JM (2011) Relative importance analysis: a useful supplement to regression analysis. *J Bus Psychol* 26(1):1–9
- Tonidandel S, LeBreton JM, Johnson JW (2009) Determining the statistical significance of relative weights. *Psychol Methods* 14(4):387
- Van Iddekinge CH, Ployhart RE (2008) Developments in the criterion-related validation of selection procedures: a critical review and recommendations for practice. *Pers Psychol* 61(4):871–925
- Wang W, Ding J (2003) Wavelet network model and its application to the prediction of hydrology. *Nat Sci* 1(1):67–71

Affiliations

Muhammad Shoaib^{1,2} · Asaad Y. Shamseldin² · Sher Khan² · Muhammad Sultan¹ · Fiaz Ahmad¹ · Tahir Sultan³ · Zakir Hussain Dahri^{4,5} · Irfan Ali⁵

¹ Department of Agricultural Engineering, Bahauddin Zakariaya University, Bosan Road, Multan 60800, Pakistan

² Department of Civil and Environmental Engineering, The University of Auckland, Auckland 1010, New Zealand

³ Department of Civil Engineering, Bahauddin Zakariaya University, Bosan Road, Multan 60800, Pakistan

⁴ Water Systems and Global Change, Wageningen University & Research, Wageningen, Netherlands

⁵ Pakistan Agricultural Research Council, Islamabad, Pakistan