CrossMark

A Comparative Study of Various Hybrid Wavelet Feedforward Neural Network Models for Runoff Forecasting

Muhammad Shoaib[1] (ID) · **Asaad Y. Shamseldin**[2] · **Sher Khan**[2] ·
Mudasser Muneer Khan[1] · **Zahid Mahmood Khan**[1] ·
Tahir Sultan[1] · **Bruce W. Melville**[2]

Abstract Considering network topologies and structures of the artificial neural network (ANN) used in the field of hydrology, one can categorize them into two different generic types: feedforward and feedback (recurrent) networks. Different types of feedforward and recurrent ANNs are available, but multilayer perceptron type of feedforward ANN is most commonly used in hydrology for the development of wavelet coupled neural network (WNN) models. This study is conducted to compare performance of the various wavelet based feedforward artificial neural network (ANN) models. The feedforward ANN types used in the study include the multilayer perceptron neural network (MLPNN), generalized feedforward neural network (GFFNN), radial basis function neural network (RBFNN), modular neural network (MNN) and neuro-fuzzy neural network (NFNN) models. The rainfall-runoff data of four catchments located in different hydro-climatic regions of the world is used in the study. The discrete wavelet transformation (DWT) is used in the present study to decompose input rainfall data using db8 wavelet function. A total of 220 models are developed in this study to evaluate the performance of various feedforward neural network models. Performance of the developed WNN models is compared with their counterpart simple models developed without applying wavelet transformation (WT). The results of the study are further compared with - multiple linear regression (MLR) model which suggest that the WNN models outperformed their counterpart simple models. The hybrid wavelet models developed using MLPNN, the GFFNN and the MNN models performed best among the six selected data driven

✉ Muhammad Shoaib
msho127@aucklanduni.ac.nz

Asaad Y. Shamseldin
a.shamseldin@auckland.ac.nz

[1] Bahauddin Zakariya University, Multan, Pakistan

[2] The University of Auckland, Private Bag, Auckland 92019, New Zealand

✎ Springer

models explored in the study. Moreover, performance of the three best models is found to be similar and thus the hybrid wavelet GFFNN and the MNN models can be considered as an alternative to the most commonly used hybrid WNN models developed using MLPNN. The study further reveals that the wavelet coupled models outperformed their counterpart simple models only with the parsimonious input vector.

1 Introduction

The relationship between rainfall and runoff remains complex due to the spatial and temporal variability of watersheds. Since the introduction of the rational method (Mulvany 1850) for determination of peak runoff, hydrologists have proposed numerous empirical and process based models for estimating rainfall-runoff relationships. Process-based models apply physical principles to model various constituents of physical processes of the hydrological cycle. The black-box data-driven models, on the other hand, are primarily based on the measured data and map the input-output relationship without giving consideration to the complex nature of the underlying process. Among data-driven models, the ANNs have appeared as powerful black-box models and received a great attention during last two decades. The merits and shortcomings of using ANNs in hydrology can be found in ASCE Task Committee (200a, b) and Abrahart et al. (2012).

Despite good performance of the ANNs in modelling of non-linear hydrological relationships, yet these models may not be able to cope with non-stationary data if pre-processing of input and/or output data is not performed (Cannas et al. 2006). Application of WT has been found effective in dealing with this issue of non-stationary data (Nason and Von Sachs 1999). The WT is a mathematical tool that improves the performance of hydrological models by simultaneously considering both the spectral and the temporal information contained in the data. It decomposes the main time series data into its sub-components. Thus, the hybrid wavelet data driven models which use multi scale-input data, result in improved performance by capturing useful information concealed within the main time series data. Recently, various hydrological studies successfully applied WT to increase forecasting efficiency of neural network models. Partal (2016), Wei et al. (2013) and Anctil and Tape (2004) applied wavelet based data driven models for streamflow forecasting. Partal (2016) investigated wavelet based multilinear regression (WMLR) model and WMLPNN model and concluded that the WMLR can be considered as an alternative to the WMLPNN models. Wei et al. (2013) examined MLPNN and WMLPNN models and found that the WMLPNN models are superior in performance. Anctil and Tape (2004) compared WMLPNN and WNFNN models and found that the WNFNN model have relatively better performance. Nourani et al. (2009a) found that WMLPNN model can successfully predict short and long term rainfall events. Partal et al. (2015) used WMLPNN model, WRBFNN model and wavelet coupled generalized regression neural network (WGRFNN) model for precipitation forecasting and found WMLPNN model to yield better results. Wang et al. (2009) and Kumar et al. (2015) applied WMLPNN models for reservoir inflow prediction. Kim and Valde's (2003) developed WMLPNN model for drought forecasting whereas Mirbagheri et al. (2010) used WNFNN model for estimating river suspended sediment concentration and Abghari et al. (2012) used WMLPNN model for daily pan evaporation forecast.

Among various types of available feedforward neural networks, selection of a suitable type is very important and difficult task. Network type determines the number of connection weights and the way information flows through the network. Although the application aspect of different feedforward neural network models have been well documented, the selection basis of these models have received very limited attention. Currently there are no scientific basis reported in the literature for the selection of these models. For instance, MLPNN is considered to produce accurate results, yet it does have some disadvantages. For example, it takes more time to train an MLPNN network because of the number of parameters to be determined. The RBFNN, on the other hand, is trained in a fraction of the time as it has fewer parameters to be determined. Similarly, majority of the wavelet coupled neural network studies are also considered to be limited as they either used the MLPNN or the neuro-fuzzy network type only. Different network types may behave differently with increased number of inputs in the case of wavelet coupled models. No study is reported to evaluate the effect of network type on the performance of wavelet coupled neural network models. This study compares the performance of wavelet coupled neural network models by considering numerous neural network types.

Section 1 gives the introduction and review of literature. Section 2 describes the methodology used in this study. Moreover, the development of simple and the hybrid wavelet rainfall-runoff models and the performance indices used to evaluate the models developed in this study is also reported in section 2. The data used in the study is described in Section 3 while results of the different developed models are elaborated in Section 4. The conclusions are provided in Section 5.

2 Methodology

2.1 Wavelet Transformation

Grossmann and Morlet (1984) introduced wavelet transformation (WT) which is capable of providing the time and frequency information simultaneously, hence giving the time-frequency representation of the temporal data. The wavelet in WT refers to the window function that is of finite length (the sinusoids used in the Fourier Transformation are of unlimited duration) and is also oscillatory in nature. The WT is considered to be capable of revealing aspects of the original time series data such as trends, breakdown points, and discontinuities that other signal analysis techniques might miss (Adamowski and Sun 2010; Singh 2012).

There are two types of the wavelet transformations; Continuous Wavelet Transformation (CWT) and Discrete Wavelet Transformation (DWT). The Continuous Wavelet Transform (CWT) of a signal $f(t)$ is defined as follows:

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

where * refers to the complex conjugate and ψ(t) is called the wavelet function or the mother wavelet. The entire range of the signal is analysed by the wavelet function by using the parameters 'a' and 'b'. The parameter 'a' and 'b' are the dilation (scale) and translation (position) parameters, respectively. The calculation of CWT coefficients at each scale 'a' and translation 'b' results in a large amount of data. This problem was resolved in the DWT which

operates scaling (low pass filter) and wavelet (High pass filter) functions. The DWT scales and positions are based on power of two (dyadic scales and positions) and can be defined for a discrete time series $f(t)$, which occurs at any time t as;

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

where the real numbers, j and k are the integers which control the wavelet dilation and translation, respectively. More details on the DWT can be found in many text books including Daubechies and Bates (1993) and Addison (2002).

2.2 Data Driven Models

2.2.1 Multiple Linear Regression (MLR)

MLR is a statistical approach to modelling the relationship between a dependent variable y and independent variables $x_1, x_2 \ldots \ldots x_m$ which can be expressed as:

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \tag{3}$$

a, b_1, b_2, b_n are the coefficients which are evaluated in this study using the least square method and the training data set.

2.2.2 Multilayer Perceptron Neural Network (MLPNN)

The MLPNN consists of several neurons (computational elements) arranged in a series of different layers. Typically, it consists of an input layer, hidden layer and an output layer. A layer usually contains a group of neurons each of which has the same pattern of connections to the neurons in other layers. Each layer has a different role in overall operation of the network. Each neuron is connected to the neuron in the next layer through connections called weights. Each neuron receives an array of inputs and produces a single output. The output of a neuron in the input layer will be the input for the neurons in hidden layer. Similarly, the output of the neuron in the hidden layer will be the input for the neurons in the output layer. Each neuron in the hidden and output layer processes its input by a mathematical function known as the neuron transfer function. However, in the case of neurons in the input layer the transfer function is the identity function. The neurons in the input layer have connections with the neuron in the hidden layer while the neuron in the output layer is only connected to the neuron in the hidden layer. There is no direct connection between the neuron in the input layer with those neurons in the output layer. MLPNN is considered as the the most widely used neural network type in various application of hydrology.

2.2.3 Generalized Feed Forward Neural Network (GFFNN)

The GFFNN is a simplification of the MLPNN such that connections can jump over one or more layers. The GFFNN theoretically can solve any problem that a MLPNN can solve. However, in practice, GFFNN can often solve problems much more efficiently. A classic example of this is the two-spiral problem (Hassoun 1995). The advantage of GFFNN is in the

ability to project activities forward by bypassing layers. As a result, the training of the layers closer to the input becomes much more efficient. The architecture of GFFNN, having two hidden layers is illustrated in Fig. 1. The circles represents processing elements (PEs) called neurons arranged in the input, hidden and output layers. The lines represent weighted connections between neurons. By adapting its weights, the network evolves towards an optimal solution. The back propagation algorithm is normally used for training of the GFFNN which propagates the error through the network and allows adaption of the hidden neurons weights.

2.2.4 Radial Basis Function Neural Network (RBFNN)

The RBFNN closely resembles the configuration of MLPNN. It was first proposed by Luo and Unbehauen (1998) as an alternative to the MLPNN. Similar to the MLPNN, the RBFNN has three layers, namely, the input, the hidden and the output layer. The hidden layer uses Gaussian transfer functions, rather than the standard sigmoidal functions employed by the MLPNN. More details on RBFNN can be found in Waszczyszyn (1999).

2.2.5 Modular Neural Network (MNN)

Modular feed forward networks are a special class of MLPNN, such that layers are segmented into independent modules. All modules are functionally integrated and each module operates on separate inputs to accomplish some sub-task of the neural network's global task. This tends to create some structure within the topology (i.e. how the connections are made), which will foster specialization of function in each sub-module. In contrast to MLPNN, the modular networks do not have full interconnectivity between their layers. Therefore, a smaller number of weights are required for the same sized network (i.e. same number of PEs). This tends to speed up training time and reduce the number of required training data samples. A MNN generally trains faster
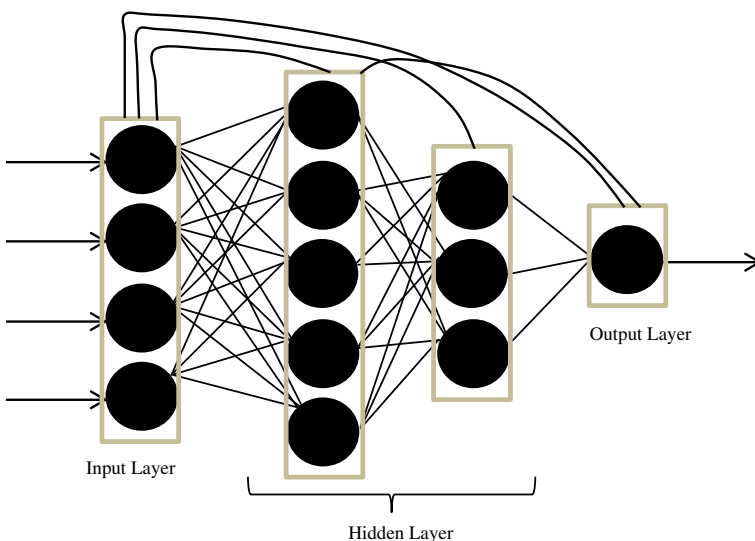

Fig. 1 Schematic diagram of generalized feed forward neural network

than a MLPNN, due to its "short-cut" connections to the output, aiding in the weight adaptation for the hidden and input layers. There are many ways to segment a MNN into modules. There is no strict rule available to select best modular topology based on the data. The schematic diagram of MNN along with topology used in the present study is shown in Fig. 2.

2.2.6 Co-Active Neuro-Fuzzy Inference System (CANFIS)

The CANFIS is considered as the more general class of ANFIS introduced by Maier and Dandy (2000). CANFIS model integrates the modular neural network with the fuzzy inference system (FIS) in the same topology. An FIS is a system that uses the fuzzy set theory to map inputs to outputs. An adaptive network is a feed-forward network which employs a collection of modifiable parameters for determining the output of the network. Like other neural networks, an adaptive network also consists of a set of nodes connected through directional links and each node is a process unit that performs a static node function on its incoming signal to generate the signal output. Unlike other neural networks, the links in an adaptive network only indicate the flow direction of signals between nodes and no weights are associated with these links.

The CANFIS architecture generally uses a neural network learning algorithm for constructing a set of fuzzy-if-then rules with appropriate membership function (MF) from the stipulated input-output pairs. The fundamental component of CANFIS is a fuzzy axon which applies MF to the inputs. The present study uses the well-known fuzzy structure suggested by Takagi and Sugeno (1985) owing to its simplicity. Also, this structure suits well for multi-input and single-input system. More details on CANFIS can be found in Krishnapura and Jutan (1997) and Shoaib et al. (2016).

2.3 Development of Wavelet Coupled Data Driven Models

In this study, wavelet coupled data driven models are developed by integrating them with the DWT.. In order to examine impact of WT on model performance, the models are developed
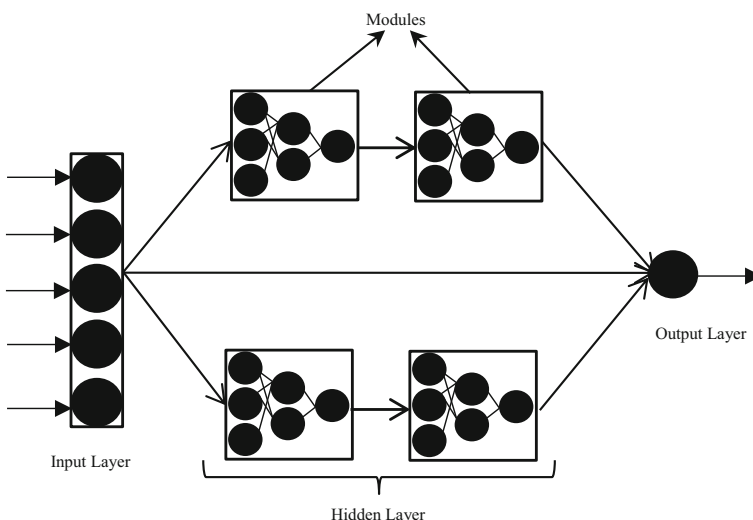

Fig. 2 Schematic diagram of modular neural network

with and without WT. The DWT is used to decompose the daily observed rainfall into approximations and details. The Mallat's algorithm (Mallat 1989) is used in the present study for DWT as it is the most efficient algorithm. The following two options are considered in order to observe the impact of WT on the performance of selected rainfall-runoff data driven rainfall models.

Option 1: Develop the models with observed rainfall as input and observed discharge as output without applying WT on input rainfall data. These models are referred to as Simple models in this study

Option 2: Develop the coupled wavelet rainfall-runoff models with wavelet transformed rainfall data as the input and the observed discharge as the output. These models are referred to as wavelet models.

2.4 Flow Duration Curves (FDCs)

Flow Duration Curves (FDCs) are used to test ability of the developed models to capture low, medium and high flows of the observed hydrograph. The FDCs illustrate the percentage of time a given flow was equalled or exceeded during a specified period of time. From the FDC, the 10 percentile flow, the flow that is equalled or exceeded 10% of the period of record, can be considered as high flow percentile. Likewise, 11 to 89 percentile flow is considered as medium flow percentile while 90 percentile flows are considered as low flow percentile. The medium flow percentile can be further divided in to high medium flow percentile and low medium flow percentile from 11 to 49 percentile and 50 to 89 percentile flows, respectively.

2.5 Performance Indicators

Statistical measures of goodness of fit can be used to evaluate performance of the developed models. Two statistical measures, namely, the Root Mean Squared Error (RMSE) and the Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe 1970) are used in the present study to measure performance of the models. These two indicators are defined by the following equations:

$$RMSE = \sqrt{\frac{1}{2}\sum_{i=1}^{N}(Q_{obs}-Q_{mod})^2} \qquad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(Q_{obs}-Q_{mod})^2}{\sum_{i=1}^{N}(Q_{obs}-\overline{Q_{obs}})^2} \qquad (5)$$

where N is the total number of observations, Q_{obs} and Q_{mod} are the observed and modelled discharges respectively, and $\overline{Q_{obs}}$ is the mean of observed discharge of training/calibration data. Eq. (5) is multiplied by 100 to express NSE in terms of percentage.

The RMSE is used to measure the estimated output accuracy. A RMSE value of zero indicates perfect match between the estimated and the observed outputs. Similarly, a higher value of RMSE suggests that there is no match between the estimated and the observed output.

Karunanithi et al. (1994) suggested that the RMSE value is a good measure for indicating the goodness of fit at high flows. Likewise, the NSE provides a measure of the capacity of the model to predict observed values. In general, high values of NSE (up to 100%) and small values for RMSE indicate a good model. These two statistical measures can be used to evaluate performance of hydrological models satisfactorily (Legates and McCabe 1999).

3 Study Area and Data

The daily rainfall runoff data of four catchments located in different hydrological conditions in the world is used in the present study. The four selected catchments are the Baihe and Yanbian catchments located in China, the Brosna catchment located in Ireland and the Nan catchment located in Thailand.

The Baihe catchment is a sub-basin of Upper Hanjian River basin. The Hanjian River is the largest tributary of the Yangtze River in China, covering a total drainage area of approximately 151,000 km^2 (34^0 30′- 30^0 49′ N, 106^0 14′- 114^0 56′E) with a total length of 1577 km. The altitude of the basin decreases from 3500 m in the northwest to 88 m at the Danjiangkou reservoir in the southeast (Sun et al. 2014). The Hanjian River is about 737 km long from northwest to Baihe station in the southeast. The average width of the main stream of Hanjiang is about 200–300 m with an average slope of 6% and surrounded by high and steep mountains, narrow valleys and swift torrents. Normally, evaporation is greater than rainfall from November to April and in August, whereas rainfall is less in the other months resulted in low evaporation as well. The daily rainfall data for a period of eight years starting from 1st January1972 onwards from nineteen rain gauge stations and three hydrological stations (having rainfall data recording facility) are used to obtain lumped daily average rainfall. The concurrent discharge data used in the study is the daily averaged data measured at the Baihe station. The Yanbian catchment is also located in the monsoon climatic region of China. The rainfall occurs mainly during the summer when the rainfall is greater than the evaporation. The river flow rate is also high during this period. The peaks of the rainfall usually occur between June and September while discharge peaks mainly occur between July and September. The flood hydrographs have narrow and high peaks. Most flooding happens during the wet season that lasts from June to October, and the flow rate gradually recedes to reach critically low levels from November to May until the rains start in June. The summers are generally hot while winters are mild to cold in this humid climatic regions. The Brosna catchment located in Ireland has a very flat topography except for some undulations caused by glacial deposits. There is no noticeable evidence of substantial groundwater movement across the topographical boundary of the catchment. Daily rainfall data for a period of ten years (1st January1969 to 31st December 1978) are lumped by averaging the data from four rain gauge stations. The concurrent discharge data used in the study is the daily averaged data measured at the Ferbane station. The Nan river basin is located in Thailand and topographically it is a plain area surrounded by mountains. The overall climate of the area is tropical monsoon and categorized by winter, summer and rainy seasons. The area is affected by the northeast and southwest monsoons. The rainy season, accomplished by the southwest monsoon, persists from mid-May until the end of October. July and August are usually months of strong rainfall. During the winter season, the weather is cold and dry because of the northeast monsoons, commencing in November and ending in February. From February until mid-May the weather is slightly warm. More than 80 % of the rainfall is concentrated in the wet season. Both Yanbian and Nan catchments are located in humid/monsoon climatic zones. Humid

climate is associated with an excess of moisture from the precipitation. A brief detail of these catchments is presented in Table 1 and 2. The data is divided into two parts, first part for training/calibration and the second part for testing/validation purposes. About three quarters of the data was used for training purposes while the remaining one quarter was used for testing purposes in the present study.

Sample size is very much important for a data driven models to make a reasonable accurate estimate. A larger sized sample is normally preferred as it helps to better comprehend structure and configuration of the data. However, a long time series data does not inevitably supposed to have additional information, since there can be several repetitions of similar type of information in the time series data (Wagener et al. 2003). This may result only in increasing the computational time and numerous over-fit series rather than refining the predicting efficacy (Gaweda et al. 2001; Fernando et al. 2009). Normally, it is suggested that the sample size must not be lesser than the number of parameters used in the model (Hyndman and Kostenko 2007). Therefore, a sample size of eight years (2918 data points) for Brosna catchment and six years (2190 data points) for Baihe, Yanbian and Nan catchments is used in present study for training of the developed models.

4 Results and Discussion

4.1 Selection of Input Vector

The performance of rainfall-runoff data driven models is dependent on the selection of appropriate input vector. The common approach for input selection of data driven models is the selection of input vector comprising of sequential time series data which starts from containing only 1-day lagged time series data in the input vector and then modifying the external input vector by successively adding one more lag time series into input vector and this continues up to a specific lag time (e.g. Furundzic 1998; Tokar and Johnson 1999; Riad et al. 2004; Chua et al. 2008; Moosavi et al. 2013).

The present study investigates the size of input vector by considering the following five input vectors.

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

Table 1 Summary description of test catchments of data used

Catchment	Area Km^2	Calibration Period (years)	Verification Period (years)	Data starting date	Climate	Topography
Baihe (China)	61780	6	2	1/01/1972	Semi-arid	Mountainous
Brosna (Ireland)	1207	8	2	1/01/1969	Temperate	Flat
Nan (Thailand)	4609	6	3	1/04/1978	Humid	Flat
Yanbian (China)	2350	6	2	1/01/1978	Humid	Flat to Mountainous

Table 2 Statistical summary of data of test catchments

Catchment	Data type	Calibration Period			Validation Period		
		Max. (mm/d)	Avg. (mm/d)	C_v	Max. (mm/d)	Avg. (mm/d)	C_v
Baihe	Rainfall	47.08	2.59	2.17	79.98	2.48	2.53
	Evaporation	12.80	2.89	0.80	8.10	2.53	0.71
	Discharge	28.25	1.04	1.89	22.66	0.78	2.23
Brosna	Rainfall	32.67	2.20	1.63	27.56	2.47	1.51
	Evaporation	9.80	1.31	1.04	6.90	1.32	1.04
	Discharge	6.94	0.98	0.83	6.62	1.22	0.86
Nan	Rainfall	128.01	3.89	2.39	113.83	4.05	2.33
	Evaporation	4.70	3.33	0.20	4.70	3.33	0.20
	Discharge	41.20	1.82	1.68	25.78	1.82	1.68
Yanbian	Rainfall	70.53	3.28	2.40	72.07	3.36	2.42
	Evaporation	17.70	5.79	0.58	15.60	6.08	0.55
	Discharge	29.56	2.55	1.36	22.43	2.65	1.32

The first input vector M_1 contains only 1-day lagged rainfall r(i-1) in input vector. The input vector M_2 is obtained by adding lagged-2 day rainfall r(i-2) in M_1. Likewise, M_3 and M_4 are obtained by adding lagged-3 day rainfall series r(i-3) in M_2 and r(i-4) in M_3 respectively. Finally, M_5 is obtained by adding lagged-5 day r(i-5) rainfall series in M_4.

4.2 Selection of Mother Wavelet Function

The performance of different wavelet coupled based models is very sensitive to the selection of the mother wavelet function to be used for transformation of data using DWT. A number of wavelet families are available, each having different members. Examples of these families include the most popular Daubechies db wavelet family containing db2, db3, db4, db5, db6, db7, db8, db9 and db10 members and the Coiflet wavelet family having Coif1, Coif2, Coif3, Coif4, Coif 5 as members (Daubechies and Bates 1993). These different wavelet functions are characterized by their distinctive features including the region of support and the number of vanishing moments. The region of support of a wavelet is associated with the span length of the wavelet which affects its feature localization properties of a signal and the vanishing moment limits the wavelet's ability to represent polynomial behaviour or information in a signal. The details of different wavelet families can be found in many text books such as Daubechies and Bates (1993) and Addison (2002). The present study employed the db8 wavelet function to decompose the input rainfall data. The db8 wavelet function with eight vanishing moments has the ability to best describe the temporal and the spectral information in the input rainfall data, (Shoaib et al. 2014, 2015, 2016).

4.3 Selection of Decomposition Level

The choice of suitable decomposition level in the development of wavelet coupled data driven models is vital. A DWT decomposition consists of Log_2N levels/stages at most. Aussem et al. (1998) and Nourani et al. (2009a, b) used the following equation to calculate suitable decomposition level:

$$L = int(log (N)) \tag{5}$$

where L is the level and N is the total data points in data. This equation was derived for fully autoregressive data by only considering the length of data without giving attention to the seasonal signature of the hydrologic process (Nourani et al. 2011). With $N = 730$ testing data points in the present study, the L value is calculated as three by using Eq. (5) which contains one approximation at level three 'a_3' and three details 'd_1' and 'd_2' and 'd_3'. Hence, using decomposition at level 3, only small temporal and spectral variation of input rainfall data may be considered. In the present study, the daily observed rainfall data of two selected catchments is transformed by DWT using the db8 mother wavelet functions. The input rainfall data is decomposed up to maximum decomposition level nine as suggested by Shoaib et al. (2014). The decomposition of data at level nine contains one large scale (lower frequency) sub-signal approximation (a_9) and nine small scale (higher frequency) sub-signals detail (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8 and d_9) in order to get the temporal and the spectral information contained in the original time series rainfall data. The detail sub-series d_1 corresponds to time series of 2-day mode which represents the features of original data discernible at a scale of up to two days. The d_2 sub-series correspond to a 4-day mode, which represents features detectable at a scale of 2–4 days. Likewise, d_3 corresponds to 8-day mode, d_4 to 16-day mode, d_5 to 32-day mode, d_6 to 64-day mode, d_7 to 128-day mode, d_8 to 256-day mode and d_9 to 512-day mode. All these ten data series are used as input for developing wavelet coupled data driven models in this study, except for the wavelet coupled CANFIS model because of the exponential relationship between the number of inputs and the number of internal processing elements in case of neuro-fuzzy modelling.

4.4 Simple and Wavelet Coupled Data Driven Models

Firstly, simple neural network (SNN) models including the MLR, MLPNN, GFFNN, MNN, RBFNN and CANFIS models were developed without applying any data pre-processing (wavelet transformation) on the rainfall data to simulate the rainfall-runoff transformation process of the two selected catchments. The observed daily discharge is taken as the target (output) and the observed rainfall is used as the input. The selected simple models of MLPNN, GFFNN, MNN, RBFNN and CANFIS are comprised of three layers; input, hidden and output layer. The input layer of all SNN models contain one neuron with input vector M_1 to five neurons with input vector M_5.

The wavelet coupled neural network (WNN) models are developed next using the wavelet transformed rainfall data as input and the observed discharge as the target output. The input layers of all WNN models except CANFIS contain ten neurons with input vector M_1 to fifty neurons with input vector M_5. As an exponential relationship exists between the number of input variables and the number of internal processing elements in case of the CANFIS model, all wavelet transformed data series cannot be used. To avoid this, a special procedure is adopted as suggested by Krishnapura and Jutan (1997) and Nourani et al. (2011). For this, firstly a correlation analysis is performed between each wavelet sub-time series and the observed discharge. It is revealed from the correlation analysis that some of the sub-signal/coefficient has a very poor correlation with the observed discharge. Based on the correlation analysis, the poorly correlated coefficients are eliminated and a new series is formed by adding the strongly correlated coefficients.

The number of neurons in the output layer is fixed as one for all the SNN and WNN models developed in the present study. The selection of the number of neurons in the hidden layer is important for better enhancing the SNN and WNN models performance. The

selection of the appropriate number of neurons in the hidden layer is done by trial and error procedure in the present study. This is accomplished by training the network and evaluating its performance over a range of different increasing values of number of hidden layer neurons in order to obtain near maximum efficiency with as few neurons as necessary (Hammerstrom 1993). Similar to the number of neurons in the hidden layer of MLPNN, GFFNN and MNN models, there is also no basic rule to find the number of MF's in the CANFIS model and these are determined in the present study iteratively. However, large number of MF's should be avoided to reduce the computational time and effort. The number of MF's in the present study is selected by trial and error. The sigmoid activation function is used for the neurons of the hidden and the output layer for the SNN and WNN models except for RBFNN and the CANFIS models. The RBFNN models use Gaussian transfer function in the RBF/hidden layer. For simple and wavelet coupled CANFIS models, the generalized bell shaped function is used as the membership function (Jang 1993) in hidden layer, as it is considered to be more flexible because it has three parameters to modify, versus two parameters for the Gaussian function. Training/calibration is a process of adjusting connection weights in the network so that the network's response best matches desired response (Muttil and Chau 2006). The simple and the wavelet coupled models used in the study are trained in a process called supervised learning. In supervised learning, the input and output are repeatedly fed into the neural network. With each presentation of input data, the model output is matched with the given target output and an error is calculated. This error is back propagated through the network to adjust the weights with the goal of minimizing the error and achieving simulation closer and closer to the desired target output. Learning in the RBFNN is conducted in two stages: first for the hidden layer, and then for the output layer. The centres and widths of the gaussians are set by unsupervised learning rules which uses a competitive rule with full conscience. The Levenberg-Marquardt algorithm (LMA) (Levenberg 1944; Marquardt 1963) is used in the current study to train all selected SNN and WNN models because it is considered to be fast, accurate and reliable (Adamowski and Sun 2010). It is an iterative algorithm that locates the minimum function value which is expressed as the sum of squares of nonlinear functions. LMA is a combination of nonlinear optimization techniques using the steepest descent combined with the Gauss-Newton method to solve a set of nonlinear equations using least-squares. The stopping criteria for training of all the developed models in the present study is either a maximum of 1000 epochs or training is set to terminate when the mean squared error (MSE) of the cross validation testing data set begins to increase. This is an indication that the network has begun to overtrain. Overtraining is when the model simply memorizes the training set and is unable to generalize the problem. The trained models are then tested by presenting different sets of two years rainfall-runoff data of the selected catchments it has not been trained with. The models are re-trained by changing the number of neurons in the hidden layer, if it failed to perform satisfactorily during testing phase. Testing of the network ensures that it has learned the general patterns of the system and has not simply memorized a given set of data.

The performance of the developed SNN and WNN models are determined in terms of performance indicators of NSE (%) and the RMSE (m^3/s). Variation of NSE for Baihe, Brosna, Nan and Yanbian catchments is presented in Tables 3, 4, 5 and 6 respectively while Tables 7, 8, 9 and 10 shows corresponding RMSE variation for all the input vectors considered in this study. Examination of Table 3 reveals that for all SNN models including MLR, MLPNN, GFNN, MNN, RBFNN and CANFIS, the value of NSE (%) is found minimum for input vector M_1 and maximum for input vector M_5. The input vector M_1 contains only 1-day lagged

Table 3 Variation of NSE (%) for SNN and WNN Models (Baihe catchment)

	Simple					Wavelet				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
MLR										
Training	4.54	20.43	48.86	60.92	65.87					
Testing	3.40	21.26	55.55	64.34	69.91					
MLPNN										
Training	3.60	12.30	28.72	70.72	78.10	64.69	84.28	88.88	86.78	88.85
Testing	3.60	10.97	27.31	69.31	76.48	60.75	84.45	85.26	85.57	88.64
GFNN										
Training	4.19	20.19	56.89	73.55	88.57	62.74	83.73	84.01	93.13	83.11
Testing	2.09	18.54	53.69	70.17	84.82	56.63	82.35	83.04	87.10	82.00
MNN										
Training	6.15	24.00	63.04	75.37	82.59	61.83	88.76	87.15	87.83	86.70
Testing	6.62	23.29	61.90	74.41	80.35	57.79	83.48	86.51	86.11	83.18
RBFNN										
Training	3.96	11.81	12.36	17.76	43.70	22.46	34.54	35.54	49.79	60.78
Testing	2.33	10.55	11.11	16.43	43.31	21.92	30.02	31.70	45.69	57.67
CANFIS										
Training	4.41	21.91	56.66	68.91	78.13	24.17	64.96	73.44	73.03	78.76
Testing	4.45	17.29	56.11	66.70	74.30	23.40	60.56	63.41	73.59	70.69

rainfall data series while input vector M_5 contains all rainfall data series from 1-day up to 5-day lagged. It can also be seen from Table 3 that the performance of all WNN models is found superior relative to their counterpart SNN models for all the input vectors M_1 to M_5. Furthermore, Table 3 suggested that the performance of simple and wavelet coupled MLPNN, GFNN and MNN is found better relative to their counterpart MLR, RBFNN and

Table 4 Variation of NSE (%) for SNN and WNN Models (Brosna catchment)

	Simple					Wavelet				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
MLR										
Training	1.67	9.60	17.61	20.83	23.31					
Testing	2.24	12.09	21.08	25.14	27.92					
MLPNN										
Training	4.30	11.92	19.52	24.03	28.06	68.01	69.05	68.01	68.80	60.26
Testing	4.74	12.50	21.54	26.60	31.39	48.08	63.47	63.09	65.59	52.52
GFNN										
Training	2.32	11.34	19.06	22.43	28.07	51.73	65.29	64.08	63.38	51.97
Testing	7.39	13.67	23.54	28.24	32.50	50.28	61.96	63.87	63.07	45.78
MNN										
Training	4.58	11.86	18.19	24.53	28.27	52.28	69.05	68.01	68.80	60.26
Testing	4.44	13.13	23.51	27.66	31.69	48.08	63.47	63.09	65.59	52.52
RBFNN										
Training	1.46	10.80	14.22	19.13	20.12	18.17	23.35	31.02	32.68	27.30
Testing	2.19	12.44	16.03	22.80	25.81	18.74	25.02	26.04	29.82	26.95
CANFIS										
Training	3.04	9.96	16.52	19.21	23.18	9.08	17.11	21.08	22.82	24.77
Testing	6.05	12.24	19.65	21.99	29.87	12.04	21.13	23.79	28.99	29.66

Table 5 Variation of NSE (%) for SNN and WNN Models (Nan catchment)

	Simple					Wavelet				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
MLR										
Training	2.83	5.64	8.22	10.70	15.33					
Testing	3.86	6.29	8.60	10.78	13.84					
MLPNN										
Training	7.17	9.43	10.46	13.97	16.66	51.99	68.28	80.51	78.81	79.27
Testing	5.14	6.29	8.56	10.35	13.98	49.90	64.02	71.03	72.72	72.72
GFNN										
Training	5.97	7.02	9.50	12.94	16.89	47.51	65.06	67.93	70.44	76.84
Testing	5.37	6.59	8.82	10.44	13.36	43.80	61.59	68.02	70.41	67.80
MNN										
Training	5.89	7.54	9.89	12.38	15.96	48.58	72.36	77.70	79.52	82.46
Testing	4.86	6.73	8.93	10.72	13.58	40.62	67.00	67.71	75.39	76.15
RBFNN										
Training	4.11	6.50	8.47	11.59	13.99	26.15	34.57	42.77	61.00	60.89
Testing	4.43	6.53	8.68	10.28	13.48	27.35	34.30	37.40	56.73	59.04
CANFIS										
Training	5.70	5.44	7.18	7.72	10.33	36.14	37.45	38.60	38.17	40.18
Testing	4.82	5.68	7.18	8.11	10.37	36.06	36.66	35.74	34.48	34.66

CANFIS models. Table 7 presents RMSE values (cumecs) for the all the SNN and WNN models for Baihe catchment with all the input vectors considered in the study. It is apparent from Table 7 that all the SNN models yielded best results with input vector M_5 by having minimum value of RMSE and worst results with input vector M_1 by having maximum value. It is also evident that all the WNN models gives best results relative to their counterpart SNN

Table 6 Variation of NSE (%) for SNN and WNN Models (Yanbian catchment)

	Simple					Wavelet				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
MLR										
Training	14.28	38.81	49.82	56.65	60.90					
Testing	13.36	36.30	48.16	54.51	57.71					
MLPNN										
Training	14.28	38.81	49.82	56.65	60.90	54.44	83.34	83.82	84.93	84.10
Testing	13.36	36.19	48.16	54.51	57.71	50.50	80.50	81.99	81.81	81.33
GFNN										
Training	9.28	42.71	51.96	58.45	61.75	57.83	79.96	83.02	89.18	85.33
Testing	5.70	42.69	50.48	57.05	58.83	54.40	73.43	81.71	85.23	82.50
MNN										
Training	10.51	42.31	52.48	58.18	62.53	56.89	84.08	90.37	86.71	86.10
Testing	8.62	42.35	51.35	57.24	60.72	52.77	80.91	85.27	82.13	82.17
RBFNN										
Training	10.49	34.91	44.51	50.42	54.06	20.25	42.12	42.82	45.25	54.90
Testing	11.57	33.08	44.90	50.04	53.17	19.99	39.63	39.69	39.48	50.11
CANFIS										
Training	14.31	32.10	44.16	52.98	55.99	48.93	51.37	57.18	60.64	60.59
Testing	14.22	30.52	42.60	49.70	51.90	45.81	51.12	52.61	55.04	60.03

Table 7 Variation of RMSE (cumecs) for SNN and WNN Models (Baihe catchment)

	Simple					Wavelet				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
MLR										
Training	1368	1249	1001	875	818					
Testing	1236	1116	838	751	690					
MLPNN										
Training	1375	1311	1182	758	655	832	555	467	509	468
Testing	1235	1187	1072	697	610	788	496	483	478	424
GFNN										
Training	1370	1251	919	720	473	855	565	560	367	575
Testing	1244	1135	856	687	490	828	528	518	452	534
MNN										
Training	1356	1221	851	695	584	865	469	502	488	511
Testing	1215	1102	776	636	557	817	511	462	469	516
RBFNN										
Training	1372	1315	1311	1270	1050	1233	1133	1124	992	877
Testing	1243	1189	1186	1150	947	1111	1052	1039	927	818
CANFIS										
Training	1369	1237	922	781	655	1219	829	721	727	645
Testing	1229	1144	833	726	638	1101	790	761	646	681

models by yielding smaller RMSE values. Moreover, performance of simple wavelet coupled MLPNN, GFNN and MNN models in terms of RMSE values is found higher (smaller value of RMSE) than their counterpart MLR, RBFNN and CANFIS models. Tables 4 and 8 demonstrate the NSE (%) and RMSE (cumecs) values respectively, for Brosna catchment. Examination of Tables 4 and 8 reveals results which are quite similar to Baihe catchment.

Table 8 Variation of RMSE (cumecs) for SNN and WNN Models (Brosna catchment)

	Simple					Wavelet				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
MLR										
Training	11.24	10.78	10.29	10.09	9.93					
Testing	14.81	14.05	13.31	12.96	12.72					
MLPNN										
Training	11.09	10.64	10.17	9.88	9.61	6.41	6.31	6.41	6.33	7.15
Testing	14.41	13.79	13.05	12.84	12.41	9.10	8.79	9.10	8.79	10.32
GFNN										
Training	11.20	10.67	10.20	9.98	9.61	7.88	6.68	6.79	6.86	7.86
Testing	14.42	13.92	13.10	12.69	12.31	10.56	9.24	9.01	9.10	11.03
MNN										
Training	11.07	10.64	10.25	9.85	9.60	7.83	6.31	6.41	6.33	7.15
Testing	14.65	13.96	13.10	12.74	12.38	10.80	9.05	9.10	8.79	10.32
RBFNN										
Training	11.25	10.71	10.50	10.19	10.13	10.25	9.92	9.41	9.30	9.67
Testing	14.82	14.02	13.73	13.16	12.91	13.51	12.97	12.88	12.55	12.81
CANFIS										
Training	11.16	10.76	10.36	10.19	9.93	10.81	10.32	10.07	9.96	9.83
Testing	14.52	14.04	13.43	13.23	12.55	14.05	13.31	13.08	12.63	12.57

Table 9 Variation of RMSE (cumecs) for SNN and WNN Models (Nan catchment)

	Simple					Wavelet				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
MLR										
Training	160.93	158.59	156.40	154.28	150.22					
Testing	160.35	158.31	156.35	154.47	151.80					
MLPNN										
Training	157.30	155.37	154.49	151.43	149.04	113.11	91.93	72.07	75.14	74.32
Testing	159.33	158.32	156.44	154.89	151.74	115.79	98.12	88.06	85.44	87.41
GFNN										
Training	158.31	157.42	155.31	152.33	148.84	118.27	96.48	92.44	88.75	78.55
Testing	159.09	158.06	156.16	154.77	152.22	122.63	101.39	92.52	88.98	92.83
MNN										
Training	158.38	156.99	154.97	152.82	149.67	117.05	85.81	77.09	73.87	68.36
Testing	159.51	157.94	156.07	154.53	152.03	126.02	93.94	92.93	81.13	79.86
RBFNN										
Training	159.87	157.87	156.19	153.51	151.41	140.27	132.04	123.49	101.94	102.08
Testing	159.88	158.11	156.28	154.91	152.11	139.39	132.56	129.39	107.58	104.66
CANFIS										
Training	158.54	158.76	157.29	156.83	154.60	130.44	129.10	127.90	128.35	126.25
Testing	159.55	158.83	157.56	156.76	154.83	130.81	130.19	131.14	132.42	132.23

All the SNN models yielded best results (higher values of NSE (%) and lower values of RMSE (m^3/s)) with input vector M_5. Moreover, all WNN models performed good compared to their respective simple models for all input vector M_1 to M_5. Likewise, simple and wavelet coupled MLPNN, GFNN and MNN models are found to perform well compared to their respective MLR, RBFNN and CANFIS models for Brosna catchment. It is also apparent from Table 4

Table 10 Variation of RMSE (cumecs) for SNN and WNN Models (Yanbian catchment)

	Simple					Wavelet				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
MLR										
Training	87.38	73.83	66.86	62.14	59.02					
Testing	88.58	76.02	68.52	64.19	61.88					
MLPNN										
Training	87.38	73.83	66.86	62.14	59.02	63.71	38.52	37.97	36.64	37.64
Testing	88.58	76.02	68.52	64.19	61.88	66.96	42.02	40.39	40.59	41.12
GFNN										
Training	89.90	71.44	65.42	60.84	58.37	61.29	42.25	38.89	31.04	36.15
Testing	92.41	72.04	66.97	62.37	61.06	64.26	49.05	40.69	36.57	39.81
MNN										
Training	89.28	71.69	65.06	61.03	57.77	61.97	37.65	29.28	34.41	35.18
Testing	90.97	72.25	66.37	62.23	59.65	65.40	41.58	36.53	40.23	40.18
RBFNN										
Training	84.29	71.80	71.37	69.84	63.39	84.29	71.80	71.37	69.84	63.39
Testing	85.12	73.94	73.90	74.03	67.22	85.12	73.94	73.90	74.03	67.22
CANFIS										
Training	87.37	77.77	70.53	64.72	62.61	67.45	65.82	61.76	59.21	59.25
Testing	88.14	79.33	72.10	67.49	66.00	70.06	66.54	65.51	63.81	60.17

that NSE (%) values for the SNN models are found slightly higher during testing than during training. This may be due to the overfitting. Overfitting is when the model performance in the training period is markedly inferior to that of the testing period. This may be because of over-parameterization of the model as the model has too many degrees of freedom for the information carrying capacity of the observed data (Viney et al. 2009). The model learns to reproduce noise of data or the data pairs itself rather than the general trends in the data set as a whole in case of over-fitting.

The results for the Nan catchment are presented in Tables 5 and 9 while Tables 6 and 10 demonstrates the results for the Yanbian catchment. Examination of these tables reveals that the WNN models yielded better results compared with their corresponding SNN models.

4.4.1 Effect of Size of Input Vector

The effect of size of input vector on the performance of best simple and wavelet coupled models, namely, MLPNN, GFFNN and MNN is investigated next and the results are presented in Fig. 3 for all four selected catchments. Examination of the figure reveals that the wavelet coupled models outperformed their counterpart simple models for all five input vectors considered in the study for the selected catchments. However, Fig. 3a reveals that as the size of input vector becomes dense by adding more and more lagged time rainfall series in the input vector, the gap in the NSE (%) values between both the simple and the wavelet coupled models constantly decreases. The performance of the simple and wavelet coupled models becomes almost similar for Baihe catchment with input vector M_5 as shown in Fig. 3a. A similar trend of NSE (%) values can be found for the Brosna, Nan and Yanbian catchments as shown in Fig. 3c, e and g, respectively. The performance of the simple models keeps on increasing as the size of input vectors increases by adding more and more raw lagged rainfall data series in the input vector. However, the performance of the wavelet coupled models increases from about 50% with input vector M_1 to about 80% with input vector M_2 and then it remains constant with M_3 and M_4 for the Baihe, Brosna and Yanbian catchments. The NSE (%) value for the Nan catchment is found maximum with input vector M_4 and then it remains same afterwards for input vector M_5. Like NSE(%) values, the RMSE values also show a similar trend for all four selected catchments as shown in Fig. 3b, d, f and h.

It can be inferred on the basis of the above analysis that the NSE (%) value for all simple models steadily increases from input vector M_1 to M_5. The addition of more and more rainfall data in the input vector results in more and more hydrological information of the catchment being fed to the simple models utilizing raw rainfall data without transformation. Furthermore, the performance of the wavelet coupled models follows an increasing trend from input vector M_1 to M_2 and then stays at a constant value with the remaining input vectors M_3, M_4 and M_5 for all the four selected catchments. This may be due to the reason that the wavelet transformed lagged one and two rainfall data series contained the maximum temporal and the spectral information, thus yielding best results in terms of NSE (%) and RMSE values. The addition of further transformed lagged rainfall data series in the input vector results in no significant improvement as all or most of the hydrologic information of the catchment has already been utilized by the model.

Flow Duration Curves (FDCs) are developed and presented next in Fig. 4 to test the ability of the best developed models to capture low, medium and high flows of the
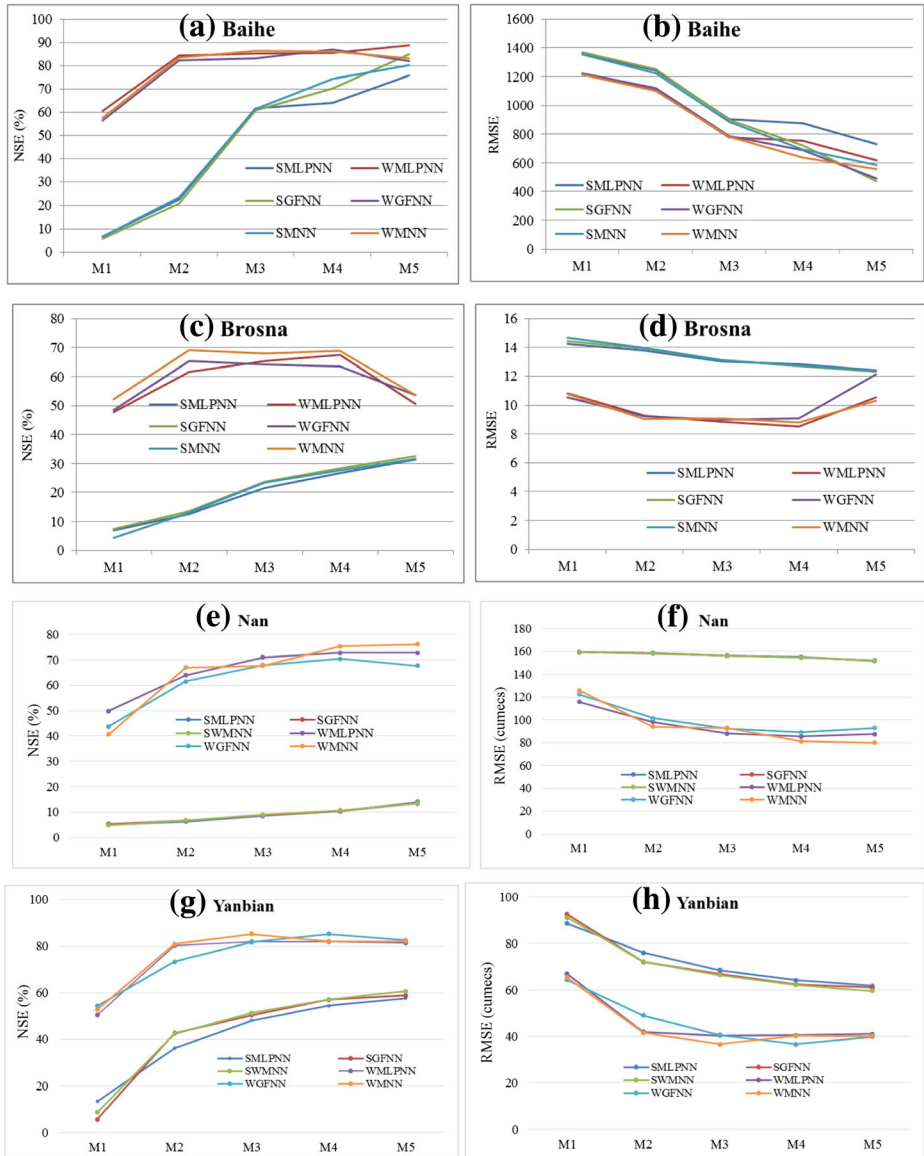
Fig. 3 Impact of input vector

observed hydrograph. The FDCs of the observed discharge and the predicted discharges of the wavelet coupled MLPNN model (WMLPNN), GFFNN (WGFFNN) and MNN (MNN) are presented for input vector M_2 for all four selected catchments. Examination of the figures illustrates that for high, medium and low flow percentile flows, FDCs of the WMLPNN, WGFFNN and WMNN models for all the four selected catchments provide good match with the FDCs of the observed flow except for the Brosna catchment. For Brosna catchment, the modelled discharge line falls below the observed
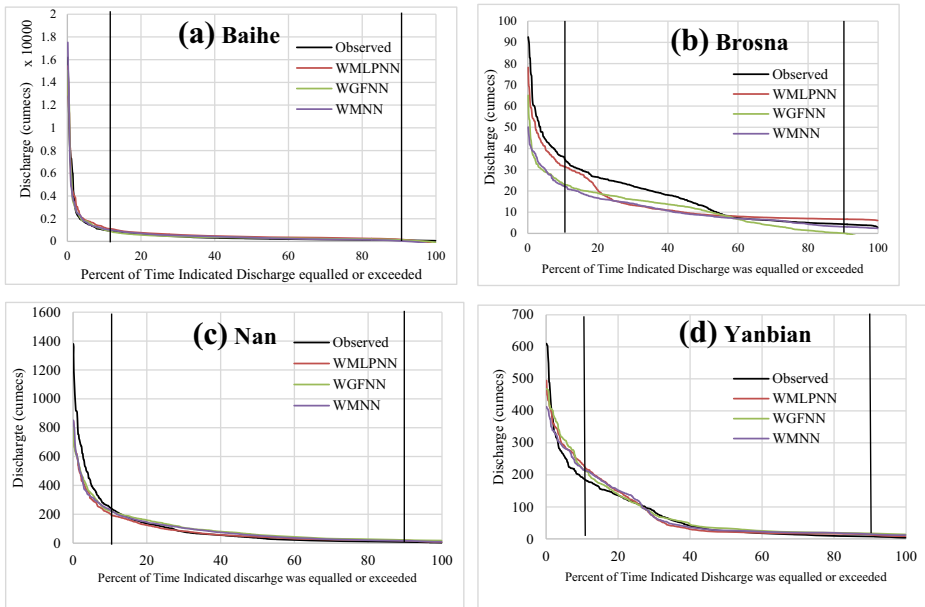
Fig. 4 Flow duration curves of the best models

discharge line for high and medium high flow percentile while it falls above the observed discharge line for low medium flow and low flow percentiles. On the basis of the above analysis of the FDCs, it can be inferred that the WMLPNN, WGFFNN and WMNN models are found to track most of the variation of the observed hydrograph.

5 Conclusions

The study was conducted to evaluate the effect of feedforward neural network type on the performance of wavelet coupled neural network models. The rainfall-runoff data from four catchments located in two hydrologically different conditions is utilized in the study. The six data driven models, namely, MLR, MLPNN, GFFNN, MNN, RBFNN and CANFIS were used in the study. The input rainfall data transformed using DWT and db8 wavelet function was employed. Decomposition at level nine was selected to decompose the input rainfall data. The study found that the wavelet coupled data driven models outperformed their counterpart simple models which do not use DWT. The MLPNN, GFFNN and MNN models have superior performance among the six selected data driven models tested in the current study. Moreover, the performance of the three superior models resembled one another and thus GFFNN and MNN models can be considered as an alternative to the most commonly used MLPNN models. It is also found that the performance of the SNN and WNN models is dependent on the selection of input vector. The SNN models are found to perform superior with input vector containing more and more lagged rainfall data while the wavelet coupled models outperformed only with the parsimonious input vector. The enhanced performance of wavelet coupled models with parsimonious input vector may be attributed to the fact that the WT provided the information regarding the hydrological signature of the watershed.

References

Abghari H, Ahmadi H, Besharat S, Rezaverdinejad V (2012) Prediction of daily pan evaporation using wavelet neural networks. Water Resour Manag 26(12):3639–3652. https://doi.org/10.1007/s11269-012-0096-z

Abrahart RJ et al (2012) Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Prog Phys Geogr 36(4):480–513

Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J Hydrol 390(1–2):85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033

Addison PS (2002) The illustrated wavelet transform handbook. Institute of Physics Publishing, London

Anctil F, Tape DG (2004) An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. J Environ Eng Sci 3(S1):S121–S128. https://doi.org/10.1139/s03-071

Aussem A, Campbell J, Murtagh F (1998) Wavelet-based feature extraction and decomposition strategies for financial forecasting. Journal of Computational intelligence in Finance 6(2):5–12

Cannas B, Fanni A, See L, Sias G (2006) Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Phys Chem Earth, Parts A/B/C 31(18):1164–1171. https://doi.org/10.1016/j.pce.2006.03.020

Chua LHC, Wong TSW, Sriramula LK (2008) Comparison between kinematic wave and artificial neural network models in event-based runoff simulation for a n overland plane. J Hydrol 357:337–348

Committee AT (2000a) Artificial neural networks in hydrology-I: preliminary concepts. J Hydrol Eng ASCE 5(2):115–123

Committee AT (2000b) Artificial neural networks in hydrology-II: hydrologic applications. J Hydrol Eng ASCE 5(2):124–137

Daubechies I, Bates BJ (1993) Ten lectures on wavelets. J Acoust Soc Am 93(3):1671–1671

Fernando T, Maier H, Dandy G (2009) Selection of input variables for data driven models: an average shifted histogram partial mutual information estimator approach. J Hydrol 367(3):165–176

Furundzic D (1998) Application example of neural netwroks for time series analysis: rainfall-runoff modeling. Signal Process 64:383–396

Gaweda AE, Zurada JM, Setiono R (2001) Input selection in data-driven fuzzy modeling, fuzzy systems, 2001. The 10th IEEE international conference on. IEEE, pp 1251–1254

Grossmann A, Morlet J (1984) Decomposition of hardy functions into square integrable wavelets of constant shape. SIAM J Math Anal 15(4):723–736

Hammerstrom D (1993) Working with neural networks. Spectr IEEE 30(7):46–53

Hassoun MH (1995) Fundamentals of artificial neural networks. MIT press, Cambridge

Hyndman, R.J., Kostenko, A.V., 2007. Minimum sample size requirements for seasonal forecasting models. Foresight, 6(Spring): 12–15

Jang JSR (1993) ANFIS: adaptive-netowrk-based fuzzy inference system. IEE Transactions on Systems, Man and Cybernetics 23(3):665–685

Karunanithi N, Grenney WJ, Whitley D, Bovee K (1994) Neural networks for river flow prediction. J Comput Civ Eng 8(2):201–220

Kim T-W, Valde's JB (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. J Hydrol Eng 8(6):319. https://doi.org/10.1061/(ASCE)1084-0699(2003)8:6(319)

Krishnapura VG, Jutan A (1997) ARMA neuron networks for modeling nonlinear dynamical systems. Can J Chem Eng 75(3):574–582

Kumar S, Tiwari MK, Chatterjee C, Mishra A (2015) Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and bootstrap method. Water Resour Manag 29(13):4863–4883

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

Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. Q Appl Math 2(2):164–168

Luo, F.-L., Unbehauen, R., 1998. Applied neural networks for signal processing. Cambridge University Press

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ Model Softw 15(1):101–124

Mallat GS (1989) A theory for multiresolution signal decomposition: the wavelet Representaiton. IEEE Trans Pattern Anal Mach Intell 11(7):674–693

Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 11(2):431–441

Mirbagheri SA, Nourani V, Rajaee T, Alikhani A (2010) Neuro-fuzzy models employing wavelet analysis for suspended sediment concentration prediction in rivers. Hydrol Sci J 55(7):1175–1189. https://doi.org/10.1080/02626667.2010.508871

Moosavi V, Vafakhah M, Shirmohammadi B, Behnia N (2013) A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. Water Resour Manag 4:1–21

Mulvany TJ (1851) On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment. Proceedings of the institution of Civil Engineers of Ireland 4(2):18–33

Muttil N, Chau KW (2006) Neural network and genetic programming for modelling coastal algal blooms. Int J Environ Pollut 28(3–4):223–238

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models. Part 1: a discussion of principles. J Hydrol 10(3):282–290

Nason GP, Von Sachs R (1999) Wavelets in time-series analysis. Philosophical transactions of the Royal Society of London a: mathematical. Phys Eng Sci 357(1760):2511–2526

Nourani V, Alami MT, Aminfar MH (2009a) A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng Appl Artif Intell 22(3):466–472. https://doi.org/10.1016/j.engappai.2008.09.003

Nourani V, Komasi M, Mano A (2009b) A multivariate ANN-wavelet approach for rainfall–runoff modeling. Water Resour Manag 23(14):2877–2894. https://doi.org/10.1007/s11269-009-9414-5

Nourani V, Kisi Ö, Komasi M (2011) Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. J Hydrol 402:41–59

Partal T (2016) Wavelet regression and wavelet neural network models for forecasting monthly streamflow. J Water Clim Chang 8(1):48–61

Partal T, Cigizoglu HK, Kahya E (2015) Daily precipitation predictions using three different wavelet neural network algorithms by meteorological data. Stoch Env Res Risk A 29(5):1317–1329. https://doi.org/10.1007/s00477-015-1061-1

Riad S, Mania J, Bouchaou L, Najjar, Y. (2004) Rainfall-runoff model usingan artificial neural network approach. Math Comput Model 40(7):839–846

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 (2015) Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach. J Hydrol 527:326–344

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

Singh RM (2012) Wavelet-ANN model for flood events, Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011. Springer, India. p 165–175

Sun Y, Tian F, Yang L, Hu H (2014) Exploring the spatial variability of contributions from climate variation and change in catchment properties to streamflow decrease in a mesoscale basin by three different methods. J Hydrol 508:170–180

Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE.Transactions on systems, man, and cybernetics, SMC-15. 1:116–132

Tokar AS, Johnson PA (1999) Rainfall-runoff modeling using artificial neural networks. J Hydrol Eng 4(3):232–239

Viney NR, Bormann H, Breuer L, Bronstert A, Croke BF, Frede H, Kite GW (2009) Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions. Adv Water Resour 32(2):147–158

Wagener T, McIntyre N, Lees M, Wheater H, Gupta H (2003) Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol Process 17(2):455–476

Wang W, Jin J, Li Y (2009) Prediction of inflow at three gorges dam in Yangtze River with wavelet network model. Water Resour Manag 23(13):2791–2803. https://doi.org/10.1007/s11269-009-9409-2

Waszczyszyn Z (1999) Fundamentals of artificial neural networks. Springer, Berlin

Wei S, Yang H, Song J, Abbaspour K, Xu Z (2013) A wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows. Hydrol Sci J 58(2):374–389. https://doi.org/10.1080/02626667.2012.754102