

## Wavelet-IANN model for predicting flow discharge up to several days and months ahead

Mohamad Javad Alizadeh, Vahid Nourani, Mojtaba Mousavimehr and Mohamad Reza Kavianpour

### ABSTRACT

In this study, an integrated artificial neural network (IANN) model incorporating both observed and predicted time series as input variables conjoined with wavelet transform for flow forecasting with different lead times. The daily model employs forecasts of the tributaries in its input structure in order to predict the daily flow in the main river in the next time steps. The predictive models for the tributaries are those of the conventional wavelet-ANN models in which they comprised only observed time series as input variables. The monthly model updates its input structure by other forecasts of the tributaries and also the predicted time series of the main river in the previous time step. The model is utilized for flow forecasting in the Snoqualmie River basin, Washington State, USA. In the integrated model, the output of each tributary (sub-basins) and also the previous flow time series of the main river are used as input variables. Regarding the results of this study, the daily flow discharge can be successfully estimated for up to several days ahead (4 d) in the main river and tributaries. Moreover, an acceptable prediction of the flow within the next two months can be achieved by applying the proposed model.

**Key words** | ANN, integrated model, multi-scale flow forecasting, wavelet transform

**Mohamad Javad Alizadeh** (corresponding author)  
**Mojtaba Mousavimehr**  
**Mohamad Reza Kavianpour**  
Faculty of Civil Engineering,  
K.N.Toosi University of Technology,  
Tehran,  
Iran  
E-mail: [mjalizadeh@mail.kntu.ac.ir](mailto:mjalizadeh@mail.kntu.ac.ir)

**Vahid Nourani**  
Faculty of Civil Engineering,  
University of Tabriz,  
Tabriz,  
Iran  
and  
Department of Civil Engineering,  
Near East University,  
P.O. Box: 99138, Nicosia,  
North Cyprus Mersin 10,  
Turkey

### INTRODUCTION

Providing a suitable predictive model for short-term and long-term flow discharge in rivers can be helpful for water resources planning and management. However, it is not an easy task due to its nonlinear identity and dependence on a large number of parameters including temporal and spatial variations. Also, some of these parameters, such as rainfall, have such an uncertain and stochastic characteristic which makes it difficult to develop an accurate predictive model for flow discharge. More especially, when the models are applied for forecasting in several days/months in advance, the complexity of the problem increases even more.

Forecasting of flow discharge in rivers using available historical time series is a common task in hydrology. Over the past decades, artificial intelligence (AI) techniques

such as artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) have shown great ability in dealing with non-stationary and nonlinear time series. Several research studies can be addressed which have successfully applied ANN and ANFIS models for time series forecasting of hydrological data (Jeong *et al.* 2012; Chen *et al.* 2015). These models were generally applied to predict target variable one time step ahead. However, some studies investigated more time steps ahead (Babovic *et al.* 2000; Alvisi & Franchini 2011). Campolo *et al.* (1999) applied ANN for river flood forecasting in the Tagliamento River, Italy. They applied the model to forecast flood in the river up to 5 hours in advance. They also showed that the model accuracy decreases as the time horizon increases. Maier & Dandy (2000) carried out a review study on

applications of neural networks for the prediction and forecasting of water resources variables. They addressed different issues including the choice of performance criteria, input selection, data division and pre-processing, network architecture, training and validation procedure of the ANN models. *Vojinovic et al. (2003)* developed a hybrid approach for flow modeling within wastewater pipe networks. They used the model for multi-step ahead forecasting purposes with 5 and 15 min time scales. Also, they analyzed the deterioration of forecast skill with increasing forecast horizon. *Campolo et al. (2003)* and *Alvisi & Franchini (2012)* employed ANN models for water level and river stage forecasting from 1 h to 6 h ahead. Moreover, there are some studies which applied previous observed time series for streamflow forecasting up to several days ahead (e.g., *Adamowski & Sun 2010*; *Alizadeh et al. 2017a*). *Huo et al. (2012)* developed two types of integrated ANN models (IANN) for monthly river flow forecasting and compared the performance of the integrated ANN models with those of the lumped ANN models (ANN). Their study showed that for rivers comprising several tributaries, the IANN models outperform the ANN models. Therefore, applying integrated models for watersheds with several sub-basins can be an efficient way to improve the performance of forecasting models.

Dealing with data-driven models, it was found that wavelet transform improves the ability of a predictive model by extracting useful information on various resolution levels (*Nourani et al. 2011*). The unique feature of wavelet transform which made it a popular tool is in its capability to elucidate simultaneously both spectral and temporal information within the signal (*Nourani et al. 2011*). Therefore, applying a conjunctive model of wavelet and AI techniques such as wavelet-ANN has attracted the attention of many researchers in hydrology and water sciences. It has been demonstrated that linking the wavelet transform to the ANN model can improve the performance of the ANN model significantly (*Nourani et al. 2009*; *Krishna et al. 2011*; *Pramanik et al. 2011*; *Ramana et al. 2013*). However, *Anctil & Tape (2004)* compared the results of neuro-wavelet hybrid model with those of ANN models and concluded that the performance of the neuro-wavelet hybrid forecasting system and the multiple-layer artificial neuron system is very similar.

The existing studies based on AI techniques applied for flow forecasting mostly applied previous observed time series as input variables. Moreover, they were mainly lumped models (a single ANN model) which only use the input variables of the same basin. For monthly time scale and even for those of daily models, the forecasting models were often focused on the flow prediction in the next time step because the performance of the predictive models deteriorates as the time horizon increases. Therefore, development of forecasting models with satisfactory performance in upcoming days and months are of great importance for practical applications.

The main contribution of the current study is incorporating both observed and predicted time series as input variables for flow forecasting in the next time steps. The key assumption explored in this study is that the outputs of the models with the best performance are included in the input structure of the models for longer period forecasting. Moreover, an integrated ANN model (here called IANN) is combined with wavelet transform for the forecasting purpose. The main idea with the integrated model is that the predictive model (IANN) in its input structure uses outputs of the separate ANN models obtained for sub-basins as well as the previous time series of the main river. Therefore, in this study, performance of an ANN model with a new input structure is investigated. The model is applied for flow forecasting up to 4 days, and months ahead in the main river in the Snoqualmie watershed. The remainder of the paper is organized as follows. The main concepts of ANN and wavelet transform along with the characteristics of the study area and the datasets are briefly described immediately below. This is followed by the structure and characteristics of the proposed model. Results are then discussed and, finally, the conclusions are summarized.

## MATERIALS AND METHODS

### Artificial neural network

ANN as a black box model has great capability to reveal the nonlinear relationship between inputs and outputs. A feed forward back propagation neural network, here referred to as ANN, is probably the most commonly used form of

ANN. It comprises three distinct layers of input, hidden, and output. Each layer possesses a set of nodes (neurons) in which is fully connected with the nodes in the following layer. The model has a feed forward phase in which input signals propagate in a forward direction (layer by layer) to reach the output layer and a backward error propagation process which modifies the connection strengths (weights).

The training process can be viewed as a procedure to minimize the differences between observed and computed values of the target variable (output). This procedure happens by finding optimum weights and biases through input to hidden layer connections and also from hidden to output layer connections. In this study, the Levenberg–Marquardt (LM) algorithm is employed to train the ANN models. More details of ANN models and the training algorithms can be found in Hagan & Menhaj (1994) and Ham & Kostanic (2000).

## Wavelet transform

Recently, the wavelet approach as a data pre-processing technique has attracted much interest in dealing with non-stationary data in data-driven models. Wavelet transforms are advantageous over Fourier techniques due to their ability to extract information such as time, location, and frequency of signals simultaneously. In this approach, the original time series (input signal) is decomposed to some sub-signals using wavelet function. The mother wavelet (wavelet function) comprises two parameters, translation and dilation. The continuous wavelet transform (CWT) of a signal  $x(t)$  is defined as (Partal 2008):

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

where  $a$  and  $b$  are the dilation and translation parameters, respectively.  $t$  is the time,  $\psi$  is called the mother wavelet, and  $*$  represents the conjugate complex function. The result is a wavelet coefficient ( $W(a,b)$ ) contour map known as a scalogram (Nourani et al. 2011).

For practical applications in hydrology, the discrete wavelet transform (DWT) is preferred to the CWT because DWT requires less computation time and also it is easier

for implementation and development. The DWT can be written as (Grossmann & Morlet 1984):

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} \psi \left( \frac{t - nb_0 a_0^m}{a_0^m} \right) \quad (2)$$

where  $m$  and  $n$  are integers that control the wavelet dilation and transformation, respectively.  $a_0$  is a specified fixed dilation step greater than 1; and  $b_0$  is the location parameter and must be greater than zero. By using DWT, the original time series will be decomposed into various sub-signals of details ( $d_i$ ) and approximations ( $a_i$ ) at different resolution levels. More details about wavelet transform can be found in Mallat (1998) and Labat et al. (2000).

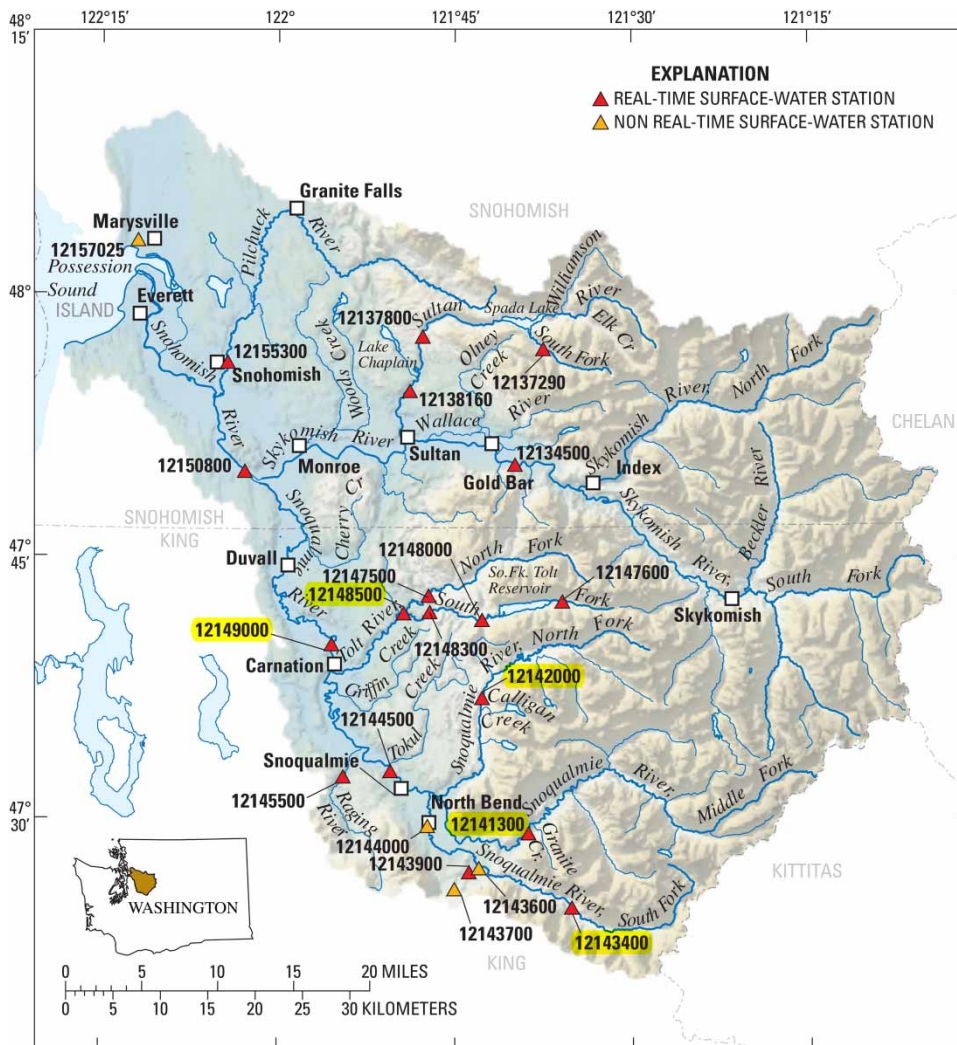
## Study area and data

In this study, a large river basin, Snoqualmie River, was chosen as a case study to develop an integrated flow forecasting model. The main river is 72 km long and is located in King and Snohomish Counties, Washington, USA. The main four tributaries of the Snoqualmie River before Duvall are: South, Middle, and North Forks which join near the city of Snoqualmie and the Tolt River which meets the main river at Carnation. There are different streamgages which measure the flow in these rivers. Therefore, in this study and in the integrated model, flow time series of these four rivers and also the time series of the streamgage near Carnation are applied for predicting flow discharge in the main river. Characteristics related to streamgages and rivers are presented in Table 1. Also, Figure 1 shows the study area and its main tributaries.

The data used in this study include flow time series of the mentioned streamgages recorded at four tributaries and also at the outlet (Carnation) for a 15-year period from 1 January, 1995 to 1 January, 2010. For the monthly time series, the datasets were extended to four months before January, 1995 and four months after January, 2010 to provide data from previous months as input variables. These datasets were downloaded from the USGS web server (<http://waterdata.usgs.gov>). In the model development, the historical time series in daily and monthly time scales have been applied for predicting daily and monthly flow discharge at the outlet and at four tributaries as well.

**Table 1** | Characteristics of the rivers and streamgages in the Snoqualmie River basin

Gage no.	River name	Length (km)	Gage USGS ID	Latitude N	Longitude W
1	South Fork	50	12143400	47.41527	121.58611
2	Middle Fork	66	12141300	47.48611	121.64666
3	North Fork	45	12142000	47.615	121.71222
4	Tolt River	9.17	12148500	47.69583	121.82277
5	Snoqualmie River	72	12149000	47.66611	121.92416



**Figure 1** | Study area and streamgages.

Table 2 gives the statistical analysis of the daily and monthly average flow time series for the five streamgages. These basic statistical measures include minimum ‘Min’, maximum

‘Max’, average ‘Mean’, and coefficient of variation ‘ $C_v$ ’. Regardless of the main river streamgage, here denoted as gage No. 5 (G5), the highest values of the daily and monthly

**Table 2** | Statistical analysis of the flow time series

River	Gage no.	Daily flow (m <sup>3</sup> /s)					Monthly average flow (m <sup>3</sup> /s)				
		Min	Max	Mean	S <sub>d</sub>	C <sub>v</sub>	Min	Max	Mean	S <sub>d</sub>	C <sub>v</sub>
South Fork	1	0.5	164.5	8.4	10.2	1.2	0.7	32.8	8.4	5.9	0.7
Middle Fork	2	3.1	727.7	34.9	42.7	1.2	3.8	128.3	34.6	21.5	0.6
North Fork	3	0.8	373.7	15.1	18.7	1.2	1.1	51.3	14.9	9.2	0.6
Tolt	4	2.8	286	15.8	15.6	0.9	3.1	55.6	15.6	9.5	0.6
Snoqualmie	5	11.8	2,160.5	104	110.6	1.0	13.6	338.8	103.1	61.7	0.5

flows are related to the Middle Fork River recording by gage No. 2 (G2).

In the models developed throughout the study, the datasets were divided into three groups, training, validation, and testing periods. Seventy percent of the data were applied in the training stage and the remaining 30% used in the validation and testing stages (15% for validation and 15% for testing stages). Regarding ANN models, it is a common task to apply extreme values of the target in the training dataset. As observed, the extreme values of the streamflow occurred for all the rivers in recent years. Therefore, the training datasets for daily and monthly models are selected from 70% of the end of the time series from 5 July, 1999 to 1 January, 2010 in daily time scale and from May 1999 to April 2010 in monthly time scale. All the datasets have been normalized to bring them in a range of [0, 1]. Normalization helps to transpose the input variables into the data range that the activation functions lie in.

## DESCRIPTION OF THE PROPOSED MODEL

In this study, a forecasting model using wavelet transform and integrated ANN (wavelet-IANN) was developed. The main difference of the wavelet-ANN and the wavelet-IANN is that the wavelet-ANN model is developed for each tributary individually. The wavelet-IANN model assembles the outputs of several wavelet-ANN models from different tributaries and streamgages as input variables to predict the flow in the main river. Prior to developing the wavelet-IANN model, five separate wavelet-ANN models for tributaries and also the main river were considered to predict flow discharge up to four next time steps.

Subsequently, the outputs of the wavelet-ANN models of the sub-basins or tributaries (G1 to G4) were considered as the input variables for the integrated model (wavelet-IANN). It should be mentioned that the preliminary investigations for the study area revealed that the daily/monthly flow is mostly correlated with the precipitation data at the same time. Alizadeh et al. (2017b) showed that incorporating predicted time series of rainfall for flow forecasting two months in advance does not improve the performance of the model. On the other hand, the influence of the current and previous time series of precipitation on flow discharge in the next two months is negligible. Therefore, the data related to precipitation were excluded in the forecasting models because they are not available to be incorporated for the upcoming days/months.

To find the right structure of input parameters of the wavelet-ANN models for the tributaries and the main river, cross-correlation and autocorrelation for different lags were carried out. The results are presented in Tables 3 and 4.

According to Table 3, there is a strong correlation for daily flow discharge in time  $t$  with those of  $t-1$ ,  $t-2$ ,  $t-3$ ,

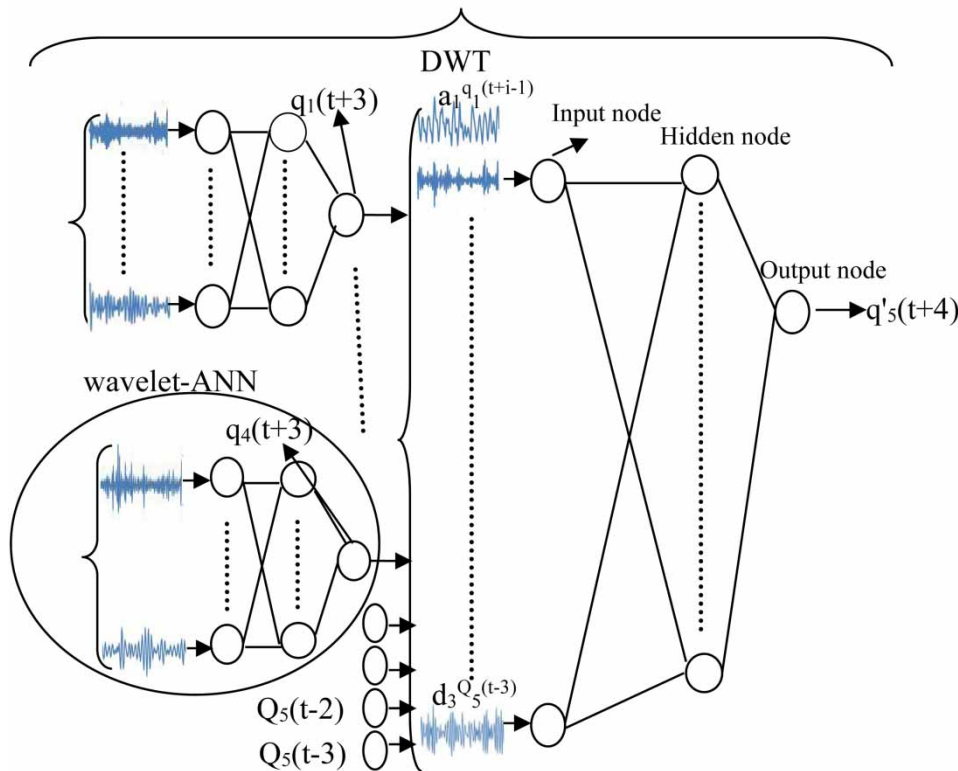
**Table 3** | Autocorrelation for flow in the streamgages with different lags

	Daily				Monthly			
	t-1	t-2	t-3	t-4	t-1	t-2	t-3	t-4
G1	0.75	0.53	0.45	0.39	0.41	0	-0.24	-0.20
G2	0.70	0.43	0.34	0.27	0.32	0.03	-0.26	-0.19
G3	0.66	0.40	0.32	0.26	0.35	0	-0.24	-0.22
G4	0.75	0.55	0.48	0.42	0.48	0.24	-0.05	-0.20
G5	0.79	0.53	0.41	0.35	0.41	0.13	-0.15	-0.20

**Table 4** | Results of cross-correlation for daily and monthly flow time series in the main river

	G1			G2			G3			G4		
	t	t-1	t-2	t	t-1	t-2	t	t-1	t-2	t	t-1	t-2
Daily Q5(t)	0.71	0.77	0.35	0.74	0.81	0.35	0.71	0.83	0.37	0.72	0.82	0.40
Monthly Q5(t)	0.81	0.14	0.00	0.89	0.14	0.00	0.94	0.20	0.02	0.92	0.21	0.05

### The Integrated model (wavelet-IANN) for the main river (G5)

**Figure 2** | Structure of the wavelet-IANN model in  $t+4$ .

and  $t-4$  in the same station. Moreover, for the monthly flow discharge, the correlation is remarkable only between  $t$  and  $t-1$ . Thus, for the ANN and wavelet-ANN models, the daily flow time series up to 4 days ago can be employed as the input variables. For the monthly forecasting models, only the flow time series in the previous time step is important as the input parameter. To select the right input structure of the predictive models for multi-time step ahead forecasting, special care should be taken in order to prevent the overfitting problem. In other words, each dataset should be inserted once in the input structure of the

predictive model throughout the multi-modeling procedure. For the daily ANN/wavelet-ANN models in the tributaries and also in the main river, separate models for each gage are developed. These models are applied to predict daily flow discharge. For each gage, the input structure of the model for predicting flow several days in advance has the same input variables, which are the observed flow discharge within the past 4 days (Equation (3)). In this study, the models with such input structure are called convention ANN/wavelet-ANN models. Figure 2 gives a schematic layout of the components of the integrated model developed

for the daily time scale.

$$Q_j(t), Q_j(t-1), Q_j(t-2), Q_j(t-3) \\ \rightarrow q_j(t+i); i = 1 \dots 4 \text{ for } j = 1 \text{ to } 5 \text{ (G1 to G5)} \quad (3)$$

where indices of  $i$  and  $j$  represent the lead time and stream-gage number, respectively.

As the flow in monthly time scale is mostly correlated with the flow in the previous time step, applying the conventional predictive models for more than one time step forecasting is not efficient. The key assumption of the monthly models developed in this study is to consider the forecasts of the previous time step as the input variable. The models incorporating the predictions of the previous time step are applied for more than one lead time forecasting in monthly time scale. Therefore, the input and output variables can be written in a conditional form as:

$$Q_j(t) \rightarrow q_j(t+1), q_j(t+1|t) \rightarrow q_j(t+2), q_j(t+2|t+1) \\ \rightarrow q_j(t+3), q_j(t+3|t+2) \rightarrow q_j(t+4) \quad (4)$$

where  $q$  denotes the predicted time series.

The main contribution of this study is to explore the efficiency of the integrated models which incorporate the forecasts of the tributaries to predict flow discharge in the main river. In this regard, the cross-correlation between the flow in the main river (G5) and those of the other gages (G1 to G4) is carried out to extract the right structure of the input variables. Table 4 presents the results of the cross-correlation between the main river and the flow in the tributaries up to four lags. As can be observed from Table 4, the flow in the main river (G5) and in time  $t$  is mainly correlated with the flow in the tributaries in  $t-1$  which shows the traveling time of flow from the tributaries to the outlet. In the daily models, the forecasts of the conventional ANN/wavelet-ANN models are imported as the input variables of the integrated models (IANN/wavelet-IANN) to predict the flow discharge in the next time step. For the daily time scale, Equation (5) can be applied to exemplify the structure of the integrated model developed in the main river (G5) in  $t+4$ . A similar expression can be

proposed for the other lead times.

$$Q_5(t), Q_5(t-1), Q_5(t-2), Q_5(t-3), q_1(t+3), \\ q_2(t+3), q_3(t+3), q_4(t+3) \rightarrow q'_5(t+4) \quad (5)$$

Regarding Table 4, it can be seen that the flow in the main river (G5) is mainly correlated with the flow in the tributaries in the same time. Therefore, the forecasts of the monthly models in the tributaries with no lag are included in the integrated model. Moreover, the output of the conventional model in G5 with one lag is incorporated as input variable. Therefore, the input and output of the integrated models (in  $t+3$  as an example) for the monthly time scale can be formulated as follows:

$$q_5(t+2|t+1), q_1(t+3|t+2), q_2(t+3|t+2), \\ q_3(t+3|t+2), q_4(t+3|t+2) \rightarrow q'_5(t+3) \quad (6)$$

where  $q'$  stands for the predicted time series of the integrated model.

In this study, the daily model begins with constructing conventional wavelet-ANN models for each time step and each stream-gage (G1 to G5). Subsequently, the outputs of the conventional models are combined as input variables to develop the integrated models (IANN/wavelet-IANN). For the monthly time scale, the models for the tributaries incorporate observed/predicted time series of the previous time step to predict the flow in 1 month ahead. Finally, the integrated models for the monthly time scale are designed using the forecasts of the tributaries and also the previous time series of the main river as input variables. For the main river (G5), performance of the wavelet-IANN and wavelet-ANN models are compared. Also, separate ANN models with similar original input variables with those of the wavelet-ANN models but not decomposed time series are developed in order to provide more comparisons.

In the developed models, DWT as a data pre-processing tool was employed to decompose the original time series of the input variables. The efficient time series (i.e.,  $D_1$  to  $D_n$  and  $A_n$  in which  $D$ ,  $A$ , and subscript  $n$  represent the detailed, approximation, and decomposition level, respectively) obtained through DWT were entered as

input variables of the ANN models. To achieve the most accurate models, different types of mother wavelets and also different decomposition levels have been investigated. For more lead times ( $t=2, 3, 4$ ), the predicted flow discharge in the precedent time steps ( $t=1 \dots 3$ ) were decomposed by DWT to be imposed as input variables. After trying with different wavelet functions of Daubechies, Haar, Coiflets, and discrete Meyer, it was found that the discrete Meyer wavelet function has a better performance than the others. Also, the 3 and 2 decomposition levels of the wavelet were the most efficient decomposition levels for daily and monthly forecasting models, respectively.

To evaluate the performance of the developed models, the coefficient of determination ( $R^2$ ) and root mean square error (RMSE) were employed. Equations (7) and (8) give a definition for these two indices:

$$R^2 = \frac{(\sum_{i=1}^n (Q_i - \bar{Q})(q_i - \bar{q}))^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2 \sum_{i=1}^n (q_i - \bar{q})^2} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - q_i)^2}{n}} \quad (8)$$

where  $Q_i$ ,  $\bar{Q}$ ,  $q_i$  and  $n$  denote observed flow, mean observed flow, predicted flow and number of data, respectively.

## RESULTS AND DISCUSSION

### Daily forecasting models

In the daily models, five individual ANN and wavelet-ANN models for streamgages (G1 to G5) were developed. The forecasts related to the best model of the tributaries in terms of  $R^2$  and RMSE were included in the IANN and wavelet-IANN models (G5I). The results in terms of  $R^2$  and RMSE during the testing set are presented in Table 5.

Although the performance of the developed models during the testing period is of great interest for researchers, the performance of the models for training and validation periods should be evaluated to find the consistency and reliability of the predictions. In this regard, the performance of the wavelet-ANN models for training and validation datasets are computed and the results are presented in Table 6. Regarding Table 6, the results of the training and validation periods are in good agreement with those computed for the testing set. Therefore, it can be shown that the models have good capability for flow forecasting, in general.

According to Tables 5 and 6, it can be concluded that the proposed models (wavelet-ANN and wavelet-IANN) have great capability for predicting the flow discharge in the Snoqualmie River. The models provide accurate predictions of the flow for different lead times. A comparison between the ANN models and the wavelet-ANN models reveals that the wavelet transform, as a data pre-processing

**Table 5** | Results of the daily ANN and wavelet-ANN models for the testing period

Lead time	G1		G2		G3		G4		G5		G5I	
	$R^2$	RMSE (m <sup>3</sup> /s)	$R^2$	RMSE (m <sup>3</sup> /s)	$R^2$	RMSE (m <sup>3</sup> /s)	$R^2$	RMSE (m <sup>3</sup> /s)	$R^2$	RMSE (m <sup>3</sup> /s)	$R^2$	RMSE (m <sup>3</sup> /s)
<i>ANN models</i>												
t + 1	0.77	4.42	0.66	21.38	0.62	9.32	0.73	6.25	0.72	47.83	0.87	32.46
t + 2	0.54	6.31	0.34	29.70	0.31	12.09	0.50	8.57	0.49	62.74	0.52	60.64
t + 3	0.44	6.95	0.26	31.48	0.22	12.75	0.38	9.37	0.35	70.61	0.39	68.74
t + 4	0.39	7.28	0.20	32.78	0.20	12.93	0.35	9.63	0.32	73.22	0.33	71.68
<i>Wavelet-ANN models</i>												
t + 1	0.99	0.88	0.99	3.90	0.98	1.72	0.99	1.29	0.99	9.29	0.99	8.33
t + 2	0.96	1.87	0.93	9.54	0.93	3.83	0.95	2.64	0.94	21.97	0.95	18.72
t + 3	0.92	2.55	0.88	13.04	0.86	5.52	0.93	3.13	0.93	23.65	0.92	25.38
t + 4	0.86	3.51	0.84	14.67	0.82	6.17	0.88	4.02	0.90	28.94	0.89	31.33



**Table 6** | Results of the daily wavelet-ANN models during training and validation periods

Lead time	G1		G2		G3		G4		G5		G5I	
	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)
<i>Training period</i>												
t + 1	0.99	0.89	0.99	3.58	0.99	2.01	0.99	1.57	0.99	8.87	0.99	6.74
t + 2	0.96	2.01	0.94	10.25	0.95	4.94	0.96	2.74	0.97	17.79	0.98	16.77
t + 3	0.95	2.11	0.91	12.73	0.88	6.26	0.95	3.32	0.97	17.83	0.97	18.92
t + 4	0.86	3.66	0.89	13.89	0.84	7.29	0.90	4.62	0.93	28.05	0.94	27.95
<i>Validation period</i>												
t + 1	0.98	1.76	0.98	7.40	0.98	3.02	0.99	2.19	0.98	17.53	0.99	17.25
t + 2	0.92	3.72	0.91	15.99	0.90	7.48	0.95	5.24	0.95	32.04	0.96	29.44
t + 3	0.90	3.93	0.87	19.64	0.85	9.08	0.91	6.34	0.95	33.72	0.94	35.28
t + 4	0.84	5.17	0.82	22.78	0.81	10.44	0.88	7.76	0.87	52.60	0.90	45.19

technique, improves the performance of the models significantly. Therefore, the wavelet-ANN models (whether usual or integrated) are superior to the ANN models without wavelet transform. The ANN models can only provide acceptable predictions for flow discharge in 1 day ahead. For more lead times, their efficiency decreases remarkably. Among the ANN models, the models applied for flow forecasting in streamgauge No. 5 (at Carnation) have the best performance. Also, the IANN model for gage No. 5 outperforms the ANN model in gage No. 5. Regarding Table 6, the developed wavelet-ANN models have roughly similar performance during training and validation periods, which demonstrates the models are trained sufficiently. Therefore, the overfitting problem is not expected to happen for these models. The wavelet-ANN models proposed by this study can be successfully used for flow forecasting in the Snoqualmie River and its tributaries up to 4 days ahead. It is observed from Table 4 that the wavelet-ANN models show excellent performance when applied for flow predictions in 1 and 2 days ahead. Moreover, the proposed models give acceptable predictions of the flow time series for more lead times (3 and 4 days). For gages No. 1 to 5, R<sup>2</sup> changes from 0.98 to 0.82 when the conventional wavelet-ANN models are employed for flow simulation from 1 to 4 days in advance. The results are compatible with those of Alizadeh et al. (2017a), who developed conventional wavelet-ANN models (only observed time series as input variables) and demonstrated the models can be successfully

applied for flow discharge in several days ahead. The R<sup>2</sup> of the models changes from 0.99 to 0.88 during the testing period when they are applied for 1 to 4 lead times. Also, Adamowski & Sun (2010) reported that R<sup>2</sup> values for the testing period changes from 0.97 to 0.76 for the Kargotis River and 0.78 to 0.42 for the Xeros River when they are predicting flow discharge in 1 and 3 days ahead. Campolo et al. (2003) showed that the percentage error ranges from 7 to 15% from 1 hour ahead to 6 hour ahead predictions.

For time series forecasting in gage No. 5 (the main river), two kinds of the wavelet-ANN model including conventional and integrated models were developed. The conventional model uses only the precedent observed time series of gage No. 5, while the integrated model also includes the predicted time series of the wavelet-ANN models of the tributaries.

The difference between the performance of the ANN and the wavelet-ANN models for 1 day ahead forecasting is reasonable. It is in good agreement with those of the cited references with R<sup>2</sup> between 0.64 and 0.75 for ANN models and R<sup>2</sup> up to 0.92 for wavelet-ANN models during the verification/testing period (Nourani et al. 2009). On the other hand, Ancil & Tape (2004) showed that the ANN and neuro-wavelet models perform similarly to each other which is different from the results obtained in this study. The difference may originate from differences in the input variables, type of wavelet function, type of algorithms in the training stage, characteristics of the basins, etc. In their study, precipitation and evapotranspiration were also

included in the input structure of the models, which improved the performance of the models. They applied Bayesian regularization technique in the model training and the Morlet wavelet, while in this study, Levenberg–Marquardt algorithm and other types of wavelet have been employed. Wavelet transform as a pre-processing technique is usually employed to improve performance of the existing models, while the ANN model developed by Anctil & Tape (2004) performs satisfactorily ( $R^2 > 0.9$ ) alone. In this study, as the data were noisy and the performance of the ANN models were not as efficient as desired, the wavelet technique was employed to improve the performance of the existing ANN models. However, for predictions with multiple time steps ahead, the performance of the ANN models decreases rapidly, while performance variations in the wavelet-ANN models are gradual.

Generally, the integrated models outperform the conventional models whether combined with wavelet or not.

However, the superiority of the wavelet-IANN models over the wavelet-ANN models is not remarkable. On the other hand, applying the integrated models requires more computational time. The conventional wavelet-ANN models can be applied adequately to provide sound predictions of the flow in the main river up to 4 days in advance.

The predicted time series and the observed time series of the usual and integrated wavelet-ANN models developed for the main channel during testing period are presented in Figure 3.

According to Figure 3, the wavelet-ANN and the wavelet-IANN models illustrate acceptable predictions of the flow discharge for all lead times and for low to high flows. Both of the predictive models give accurate predictions of flow for 1 and 2 days ahead. Moreover, the models have a relatively high accuracy for flow time series forecasting with 3 and 4 days ahead. To provide more discussion on the capability of the predictive models in the main river,

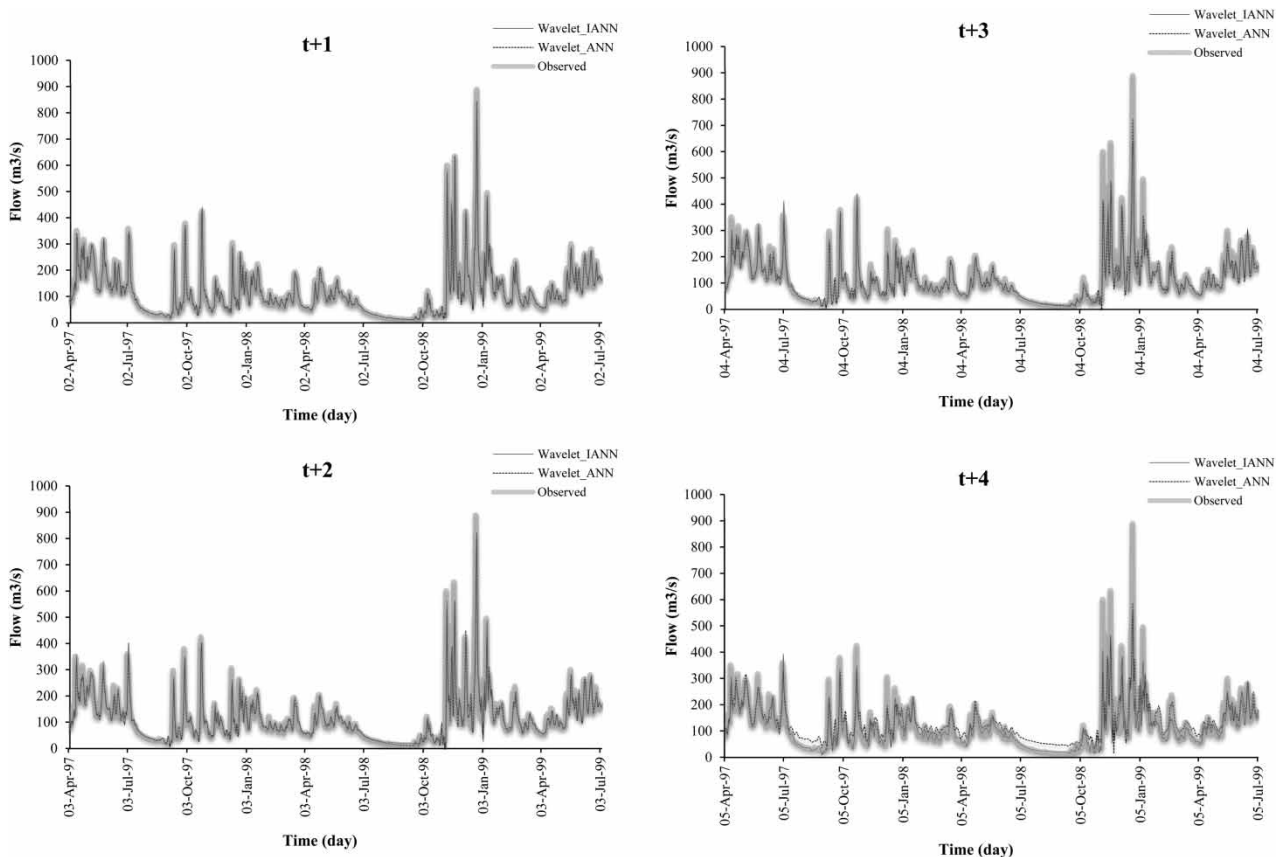


Figure 3 | Daily time series of the measured and predicted flow in the Snoqualmie River.

the 10% of the highest flows were selected as peak flows and the  $R^2$  and RMSE have been computed for them during the testing period. The results are given in Table 7. Generally, both models provide acceptable predictions of high flows in daily time scale. As can be obtained from Table 7, the integrated model outperforms the conventional model when it is applied for the next 2 days. However, for 1, 3, and 4 days in advance forecasting, the difference in performance of the models is negligible. As can be observed, the efficiency of the predictive models for high flows decreases rapidly as the time horizon increases. This conclusion is expected because the correlation between input and output decreases.

### Monthly forecasting models

Similar to the daily models, separate ANN and wavelet-ANN models were developed to evaluate the efficiency of

**Table 7** | Performance of the daily wavelet-ANN and wavelet-IANN models for high flow conditions

Lead time	Wavelet-ANN		Wavelet-IANN (Integrated)	
	$R^2$	RMSE ( $m^3/s$ )	$R^2$	RMSE ( $m^3/s$ )
t + 1	0.97	18.89	0.97	18.88
t + 2	0.79	51.54	0.86	42.83
t + 3	0.70	63.04	0.68	65.96
t + 4	0.71	71.34	0.73	70.34

the proposed models for monthly flow forecasting. The results obtained through the testing period are presented in Table 8. Also, the results for training and validation periods are presented in Table 9. A comparison among the results of training, validation, and testing periods reveal that the consistency and reliability of the predictive models decreases as the time horizon increases.

Results of the monthly predictive models show that the ANN models have low values of  $R^2$  and high values of RMSE. The best ANN model has  $R^2$  of about 0.4 obtained for the integrated ANN in the main river. The value is close to the results of the best support vector regression models (without wavelet) developed by Kisi & Cimen (2011) to forecast monthly stream flow in two different rivers with  $R^2$  about 0.35 and 0.28 for the test set. The ANN models' performance decreases significantly in the way that the predictions of the ANN models are not reliable and contain low accuracy and a great amount of uncertainty. The main reason why the performance of the model for longer periods (2 to 4 months ahead) decreases rapidly is due to using the predictions of the preceding time steps which are not accurate and differ significantly to those of the observed ones. On the other hand, the wavelet-ANN models give relatively high accurate predictions, which do not differ from the actual values significantly and, consequently, applying them in the next time step forecasting can help the models' performance significantly.

**Table 8** | Results of the monthly ANN and wavelet-ANN models

Lead time	G1		G2		G3		G4		G5		G5I	
	$R^2$	RMSE ( $m^3/s$ )	$R^2$	RMSE ( $m^3/s$ )	$R^2$	RMSE ( $m^3/s$ )	$R^2$	RMSE ( $m^3/s$ )	$R^2$	RMSE ( $m^3/s$ )	$R^2$	RMSE ( $m^3/s$ )
<i>ANN models</i>												
t + 1	0.39	4.80	0.23	16.26	0.26	7.40	0.39	7.97	0.35	46.82	0.40	38.21
t + 2	0.15	6.10	0.13	20.75	0.24	8.28	0.28	8.75	0.22	54.22	0.26	49.20
t + 3	0.14	6.14	0.12	21.48	0.23	8.30	0.29	8.70	0.14	61.46	0.22	56.42
t + 4	0.12	6.3	0.11	22.21	0.17	8.69	0.28	8.60	0.20	59.10	0.22	56.17
<i>Wavelet-ANN models</i>												
t + 1	0.98	0.97	0.95	4.65	0.94	2.04	0.97	1.70	0.96	11.85	0.98	10.16
t + 2	0.81	2.57	0.73	14.49	0.77	4.25	0.81	4.82	0.72	33.64	0.86	27.92
t + 3	0.70	4.30	0.54	12.26	0.61	5.77	0.65	5.66	0.57	38.49	0.67	38.18
t + 4	0.48	5.38	0.37	20.37	0.36	6.96	0.40	6.81	0.31	53.14	0.50	49.27

**Table 9** | Results of the monthly wavelet-ANN models during training and validation periods

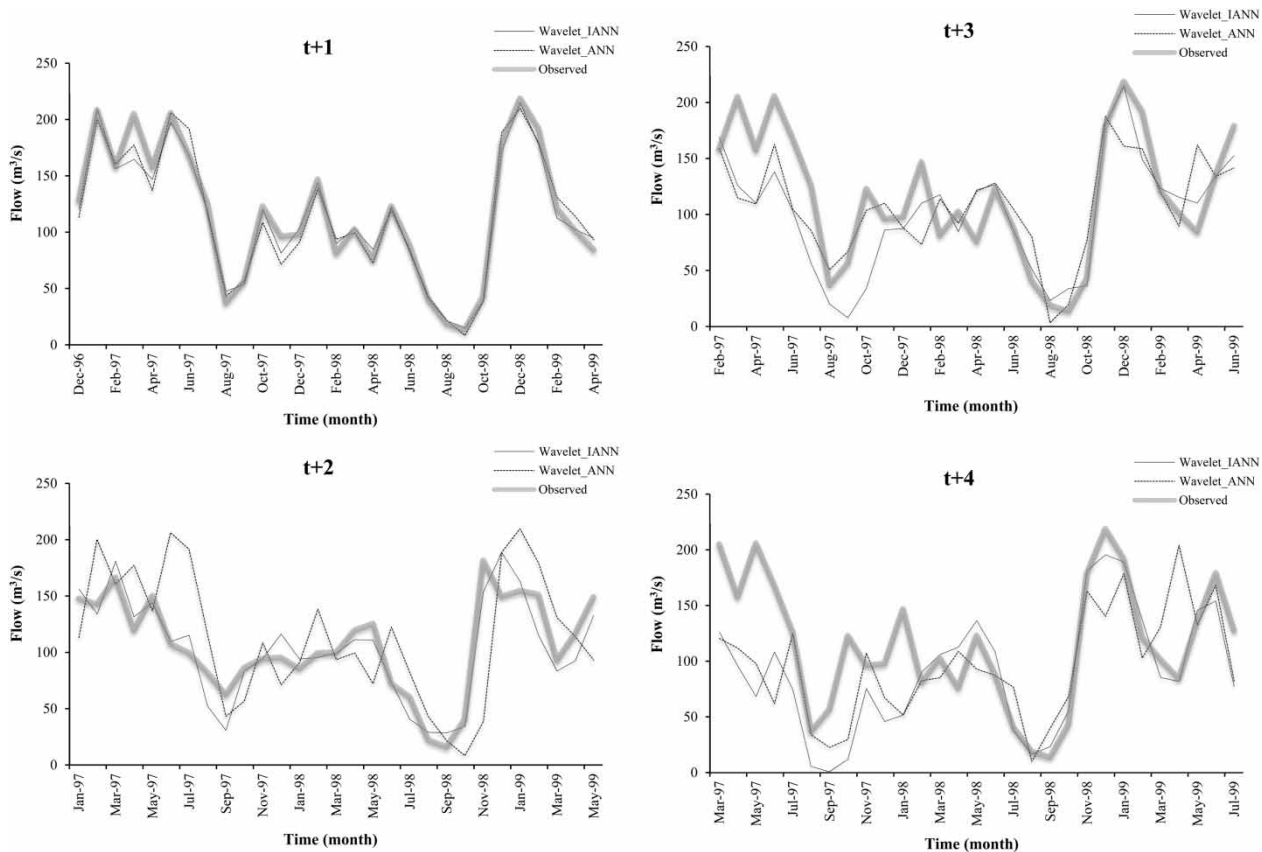
Lead time	G1		G2		G3		G4		G5		G5I	
	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)	R <sup>2</sup>	RMSE (m <sup>3</sup> /s)
<i>Training period</i>												
t + 1	0.98	0.82	0.97	3.63	0.98	1.25	0.98	1.15	0.98	8.66	0.98	7.36
t + 2	0.87	2.50	0.75	10.65	0.87	3.30	0.89	2.98	0.89	18.98	0.92	17.11
t + 3	0.77	2.73	0.74	10.71	0.68	5.49	0.81	4.11	0.73	31.48	0.87	21.49
t + 4	0.71	3.06	0.52	14.47	0.49	6.68	0.54	6.14	0.70	32.53	0.86	22.97
<i>Validation period</i>												
t + 1	0.88	2.41	0.88	8.66	0.92	3.19	0.92	3.36	0.93	23.01	0.91	24.94
t + 2	0.54	4.64	0.54	17.37	0.66	6.40	0.72	6.54	0.69	46.28	0.75	41.67
t + 3	0.38	6.10	0.38	19.83	0.51	8.00	0.69	7.47	0.61	54.60	0.61	54.38
t + 4	0.29	6.33	0.28	22.59	0.21	10.18	0.51	9.68	0.30	69.83	0.50	63.01

Table 8 indicates that the wavelet-ANN models can be successfully used to predict the flow in the tributaries one month ahead. A relatively high R<sup>2</sup> (more than 0.9) and low RMSE confirm the efficiency of the proposed models. For longer predictions, the R<sup>2</sup> of the models during the testing period decreases rapidly. Similarly, a remarkable rise in the RMSE can be found with increasing the lead times. However, acceptable predictions for tributaries and also the main channel up to two months ahead can be achieved using wavelet-ANN models. Based on the results presented in Table 9, the monthly wavelet-ANN models for flow forecasting more than two lead times may suffer from overfitting problem due to the significant difference between performance evaluation criteria of training and validation datasets. Therefore, more action should be taken in consideration of improving performance of such models, especially when they are used for more than two lead times forecasting.

From Table 8, it can be concluded that an accurate prediction of the flow time series in gage No. 5 in the next month can be achieved when wavelet-ANN and wavelet-IANN models are applied. For more lead times, the model's performance decreases rapidly. Generally, the wavelet-IANN model gives acceptable predictions for the flow discharge in all lead times. However, the results for more than two months ahead may be susceptible regarding reliability and overfitting problem. It is deduced that the wavelet-IANN model outperforms the wavelet-ANN model,

especially when the forecasting lead times increase. For example, in  $t + 4$ , the wavelet-ANN model has R<sup>2</sup> of 0.31 and RMSE of 53 m<sup>3</sup>/s, while the corresponding values for the wavelet-IANN model (the integrated model) are computed as about 0.5 and 49 m<sup>3</sup>/s, respectively. These models satisfy the model performance evaluating criteria of R<sup>2</sup> < 0.5 and a low RMSE. Also, the wavelet-IANN models developed for discharge prediction in one and two months ahead show a relatively high efficiency. As there is a high correlation between the flow in the main river with its upstream tributaries, the predicted values of flow with different lead times in the main river can be helpful to draw some useful information for its upstream tributaries. The observed and predicted time series of the wavelet-ANN model for gage No. 5 and for one to four lead times are presented in Figure 4.

Figure 4 indicates that the wavelet-ANN model provides a relatively high accurate prediction of the flow for one month ahead. As the time horizon grows, the model's efficiency decreases. Considering the depicted time series from one to four months ahead against the measured values, the wavelet-ANN model includes the four levels of relatively high, acceptable, relatively low, and low degree of accuracy for  $t + 1$ ,  $t + 2$ ,  $t + 3$ , and  $t + 4$ , respectively. The predictive model for  $t + 1$  has a relatively high coefficient of determination and low RMSE while it becomes worse with more lead times. For the next month, the model has a reliable output even for extreme flows.



**Figure 4** | Monthly predicted and observed time series for gage No. 5 during the testing period.

Moreover, an approximately complete fit between measured and predicted time series can be obtained for relatively low and medium flows.

In the integrated model denoted as wavelet-IANN (Figure 4), there is a high correlation between the observed and predicted time series for  $t+1$ . Moreover, the model shows an acceptable performance for time series forecasting in two months in advance. It can be observed that the model does not have a reliable prediction for time step  $t+4$ . Similar to the daily models, the integrated model (wavelet-IANN) developed for monthly time series forecasting outperforms the wavelet-ANN model. However, the daily predictions are frequently more accurate, especially for more than one lead time.

Generally, it can be concluded from the figures that the integrated model has a better performance than conventional models for predicting high flows. For the next month ( $t+1$ ), the wavelet-ANN model predicts high flows of 218.88,

208.68, and 205.95  $\text{m}^3/\text{s}$  as 209.83, 200.39, and 206.35  $\text{m}^3/\text{s}$ , respectively. Similarly, their forecasting for  $t+2$  are about 149, 147, and 150  $\text{m}^3/\text{s}$ . On the other hand, the predictions of high flows obtained by the wavelet-IANN models for the same three high flows in  $t+1$  are about 214, 209, and 197  $\text{m}^3/\text{s}$  and for  $t+2$  are about 188, 156, and 144  $\text{m}^3/\text{s}$ . Therefore, as the time horizon increases, the performance of the forecasting models (both the wavelet-ANN and Wavelet-IANN) for extreme values deteriorates rapidly.

Finally, the proposed model improves the performance of the daily ANN model mainly because of conjoining with wavelet transform whereas the contribution of the forecasts as input variables is of less importance. For the monthly model, both wavelet transform and incorporating predicted time series as input variables significantly contribute to increase the model's efficiency. For example, the monthly model for the main river in  $t+2$  and  $t+3$  and

for the test period have  $R^2$  of about 0.72 and 0.53, without including the predicted time series in the wavelet-ANN input structure. These values are significantly lower than the corresponding values obtained by the proposed model (Table 8).

## CONCLUSIONS

In the present study, an attempt was made to design, construct, and evaluate the efficiency of an integrated wavelet-ANN model (here called wavelet-IANN) for daily and monthly flow time series forecasting. The model in the main river updates its input structure via the predicted time series of the tributaries and historical time series of itself to predict the flow in the main river for different lead times. The model was applied for a large river basin called the Snoqualmie River basin which comprises four main tributaries (South Fork, Middle Fork, North Fork, and Tolt River) before Duvall. Also, the efficiency of the wavelet-ANN models for future time series forecasting in the tributaries have been taken into consideration. For the daily time scale, the flow time series of each gage up to 4 days ago were considered as input variables for the future time series prediction. For the monthly models, the previous month time series for flow forecasting in  $t + 1$  and the predicted time series were incorporated for other lead times forecasting. To provide more comparisons, separate ANN and IANN models without being connected to wavelet transform were developed.

It was found through this study that the wavelet transform is a powerful tool which has a great ability to extract useful information from time series. Consequently, it increases the ANN models' performances significantly. In the integrated models, the DWT has been employed twice in a predictive model when the target variable is predicted for more than one lead time. First, it was applied to decompose the original time series and next, it was employed for the predicted time series to be added as input variables. Therefore, the DWT as a data pre-processing technique is a key component towards achieving accurate flow forecasting models, especially when monthly forecasting is desired.

The results obtained from the daily flow forecasting models indicate that the proposed model (wavelet-ANN)

can strongly predict the flow discharge in the tributaries and also the main river up to the following 4 days. A very high correlation between the predicted flow and the measured time series was achieved. With increasing time horizon, the models' performance decreases gradually. The daily developed models yield a very high  $R^2$  (higher than 0.8) for all the streamgages and all the lead times under investigation. Furthermore, their RMSE for the testing period is acceptable. However, the performance of the predictive models for high flow conditions deteriorates rapidly as the time horizon increases.

From the results of the monthly models it can be derived that the wavelet-IANN is a suitable model for flow forecasting within the next two months. The results of the wavelet-ANN models represent good predictions of flow in the tributaries and also in the main river only for the next one and two months. The wavelet-IANN model provides an acceptable prediction of the flow up to three months ahead ( $R^2 > 0.65$  and relatively low RMSE). However, the models' efficiency in terms of determination coefficient decreases for flow modeling in the fourth month. Moreover, the forecasting model for the third and fourth months are susceptible to overfitting problems which should be considered carefully.

Generally, the integrated model slightly outperforms the conventional model in daily time scale although it requires more computational time. A comparison between the performance of the wavelet-IANN models and those of the wavelet-ANN models indicates that the integrated models outperform the usual models (wavelet-ANN models) for the monthly forecasting and also for higher lead times. For example, in the monthly model in  $t + 4$ , the wavelet-ANN model has an  $R^2$  of 0.31 and an RMSE of  $53 \text{ m}^3/\text{s}$ , while for the integrated model (wavelet-IANN) it is computed as 0.5 and  $49 \text{ m}^3/\text{s}$ , respectively. Therefore, it can be shown that knowing about the flow discharge in the tributaries can improve the performance of the predictive model in the main river.

The findings of this study demonstrate that accurate predictions of the flow up to several days ahead (here up to 4 d) can be achieved by applying the wavelet-ANN models. Moreover, the integrated models (wavelet-IANN) provide acceptable predictions of the monthly flow discharge in the Snoqualmie River for two months in advance. The

results of this study are promising in order to examine the efficiency of such a model for other hydrological time series forecasting with different lead times.

## REFERENCES

- 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. *Journal of Hydrology* **390** (1), 85–91.
- Alizadeh, M. J., Rajaei, T. & Motahari, M. 2017a Flow forecasting models using hydrologic and hydrometric data. *Proceedings of the Institution of Civil Engineers – Water Management* **170** (3), 150–162.
- Alizadeh, M. J., Kavianpour, M. R., Kisi, O. & Nourani, V. 2017b A new approach for simulating and forecasting the rainfall-runoff process within the next two months. *Journal of Hydrology* **548**, 588–597.
- Alvisi, S. & Franchini, M. 2011 Fuzzy neural networks for water level and discharge forecasting with uncertainty. *Environmental Modelling & Software* **26** (4), 523–537.
- Alvisi, S. & Franchini, M. 2012 Grey neural networks for river stage forecasting with uncertainty. *Physics and Chemistry of the Earth, Parts A/B/C* **42**, 108–118.
- Ancil, F. & Tape, D. G. 2004 An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. *Journal of Environmental Engineering and Science* **3** (S1), S121–S128.
- Babovic, V., Keijzer, M. & Bundzel, M. 2000 From global to local modelling: a case study in error correction of deterministic models. In: *Proceedings of the Fourth International Conference on Hydro Informatics*, Iowa City, USA.
- Campolo, M., Andreussi, P. & Soldati, A. 1999 River flood forecasting with a neural network model. *Water Resources Research* **35** (4), 1191–1197.
- Campolo, M., Soldati, A. & Andreussi, P. 2003 Artificial neural network approach to flood forecasting in the River Arno. *Hydrological Sciences Journal* **48** (3), 381–398.
- Chen, X.-Y., Chau, K.-W. & Wang, W.-C. 2015 A novel hybrid neural network based on continuity equation and fuzzy pattern-recognition for downstream daily river discharge forecasting. *Journal of Hydroinformatics* **17** (5), 733–744.
- Grossmann, A. & Morlet, J. 1984 Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis* **15** (4), 723–736.
- Hagan, M. T. & Menhaj, M. B. 1994 Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks* **5** (6), 989–993.
- Ham, F. M. & Kostanic, I. 2000 *Principles of Neurocomputing for Science and Engineering*. McGraw-Hill Higher Education, New York, USA.
- Huo, Z., Feng, S., Kang, S., Huang, G., Wang, F. & Guo, P. 2012 Integrated neural networks for monthly river flow estimation in arid inland basin of Northwest China. *Journal of Hydrology* **420**, 159–170.
- Jeong, C., Shin, J.-Y., Kim, T. & Heo, J.-H. 2012 Monthly precipitation forecasting with a neuro-fuzzy model. *Water Resources Management* **26** (15), 4467–4483.
- Kisi, O. & Cimen, M. 2011 A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology* **399** (1), 132–140.
- Krishna, B., Satyaji Rao, Y. R. & Nayak, P. C. 2011 Time series modeling of river flow using wavelet neural networks. *Journal of Water Resource and Protection* **3** (1), 50–59.
- Labat, D., Ababou, R. & Mangin, A. 2000 Rainfall-runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. *Journal of Hydrology* **238** (3), 149–178.
- Maier, H. R. & Dandy, G. C. 2000 Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software* **15** (1), 101–124.
- Mallat, S. 1998 *A Wavelet Tour of Signal Processing*. Academic Press, New York, USA.
- Nourani, V., Komasi, M. & Mano, A. 2009 A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resources Management* **23** (14), 2877–2894.
- Nourani, V., Kisi, Ö. & Komasi, M. 2011 Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *Journal of Hydrology* **402** (1), 41–59.
- Partal, T. 2008 River flow forecasting using different artificial neural network algorithms and wavelet transform. *Canadian Journal of Civil Engineering* **36** (1), 26–38.
- Pramanik, N., Panda, R. & Singh, A. 2011 Daily river flow forecasting using wavelet ANN hybrid models. *Journal of Hydroinformatics* **13** (1), 49–63.
- Ramana, R. V., Krishna, B., Kumar, S. & Pandey, N. 2013 Monthly rainfall prediction using wavelet neural network analysis. *Water Resources Management* **27** (10), 3697–3711.
- Vojinovic, Z., Kecman, V. & Babovic, V. 2003 Hybrid approach for modeling wet weather response in wastewater systems. *Journal of Water Resources Planning and Management* **129** (6), 511–521.

First received 23 November 2016; accepted in revised form 1 August 2017. Available online 14 September 2017