Acta Geophysica
https://doi.org/10.1007/s11600-021-00620-7

RESEARCH ARTICLE - HYDROLOGY

Flood prediction based on climatic signals using wavelet neural network

Nguyen Thi Thuy Linh[1] · Hossein Ruigar[2] · Saeed Golian[2,3] · Getnet Taye Bawoke[4] · Vivek Gupta[5] · Khalil Ur Rahman[6] · Adarsh Sankaran[7] · Quoc Bao Pham[8]

Received: 7 January 2021 / Accepted: 2 June 2021

Abstract

Large-scale climatic circulation modulates the weather patterns around the world. Understanding the teleconnections between large-scale circulation and local hydro-climatological variables has been a major thrust area of hydro-climatology research. The large-scale circulation is often quantified in terms of sea surface temperature (SST) and sea-level pressure (SLP). In this paper, we investigate the potential of wavelet neural network (WNN) hybrid model to predict maximum monthly discharge of the Madarsoo watershed, North of Iran considering two large-scale climatic signals like SST and SLP as inputs. Error measures like root-mean-square error (RMSE), and mean absolute error along with the correlation measures like coefficient of correlation (R), and Nash–Sutcliffe coefficient (CNS) were used to quantify the performance of prediction of maximum monthly discharge of three different hydrometry stations of the watershed. In all the cases, the WNN hybrid machine learning model was found to be giving superior performance consistently against the standalone artificial neural network (ANN) model and multiple linear regression model to predict the flood discharges of March and August months. The prediction of flood for August which is more devastating is found to be slightly better than the prediction of floods of March, in the stations served with smaller drainage area. The RMSE, R and CNS of Tamer hydrometry station in August were found to be 0.68, 0.996, and 0.99 m^3/s, respectively, for the test period by using WNN model against 1.55, 0.989 and 0.95 by ANN model. Moreover, when evaluated for predicting the maximum monthly discharge in March and August between 2012 and 2013, the wavelet-based neural networks performed remarkably well than the ANN.

Keywords Large-scale climatic circulation · Sea surface temperature · Sea-level pressure · Maximum discharge prediction · Artificial neural network

Introduction

The prediction of floods has a long history in hydrology. The studies on flood prediction can be categorized as purely time series models with lagged time-step values as inputs and the cause–effect models like rainfall–runoff model or models considering climatic drivers as inputs. The time-series models and rainfall–runoff models are the most popular approaches for flood forecasting (Dawson and Wilby 1998; Sudheer et al. 2002; Sudheer 2005; Chang et al. 2007). Large-scale climatic circulations and their teleconnection

Communicated by Michael Nones, Ph.D. (CO-EDITOR-IN-CHIEF) / Achilleas G. Samaras, Ph.D. (ASSOCIATE EDITOR).

✉ Quoc Bao Pham
phambaoquoc@tdmu.edu.vn

[1] Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Vietnam

[2] Civil Engineering Department, Shahrood University of Technology, Shahrud, Iran

[3] ICARUS, Department of Geography, Maynooth University, Maynooth, Ireland

[4] Department of Geology, School of Earth Sciences, Bahir Dar University, P.O. Box 79, Bahir Dar, Ethiopia

[5] Department of Hydrology, Indian Institute of Technology, Roorkee, India

[6] State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China

[7] Department of Civil Engineering, TKM College of Engineering, Kollam 691005, India

[8] Institute of Applied Technology, Thu Dau Mot University, Thu Dau Mot City, Binh Duong Province, Vietnam

Published online: 30 June 2021

with global climatic variations play a significant role in modulating variability in the hydroclimatic variables such as precipitation and temperature worldwide and are used to predict climate-related variables (Munoz-Diaz and Rodrigo 2006; Adarsh and Janga Reddy 2019a, b; Gupta et al. 2020). It is well-evident that the rainfall process will be modulated by large-scale climatic oscillations, and accordingly, it is logical to ascertain a link between large-scale circulation patterns and the basin-scale streamflow (Maity and Nagesh Kumar 2008). Even though studies on rainfall-climate teleconnection are abundant in literature, the application for basin scale flood forecasting is relatively less and some of the past studies highlighted the importance of using the climatic drivers for streamflow forecasting at river basin scale (Maity and Nagesh Kumar 2008, 2009).

In recent years, many studies have shown the link between climatic signals and precipitation and discharge of large watersheds worldwide (see, e.g., Biasutti et al. 2008; Soukup et al. 2009; Gamiz-Fortis 2010; Farokhnia et al. 2010; Rezaebanafsheh et al. 2011; Kalra et al. 2012; Oubeidillah et al.2012; Berg et al. 2013, Gupta and Jain 2020; Gupta and Jain 2021). In Iran, indices representing El-Nino Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO) processes have been widely used to evaluate the variability in hydro-climatic variables over some regions of Iran (Tabari et al. 2014; Araghi et al. 2017; Dehghani et al. 2020 and the literature cited their-in). In the dry regions of the world, such as Iran, where flood generation processes often reveal highly nonlinear behavior, flood forecasting is challenging. Iran is located in Western Asia and shares its boundaries with the Caspian Sea, Persian Gulf, and the Gulf of Oman, and hence, the use of sea-level pressure (SLP) and sea surface temperature (SST) indicators seems to be a promising tool to improve hydrometeorological forecasting (Meidani and Araghinejad 2014; Ruigar and Golian 2016). Maity and Nagesh Kumar (2008, 2009) used the climatic drivers like ENSO and Equatorial Indian Ocean Oscillation (EQUINOO) for streamflow prediction of Mahanadi River, India, using simple linear regression and artificial neural networks (ANN). But such hydro-climatic processes being multiscale and nonlinear, a multi-time-scale decomposition-based hybrid modeling is reported to be a more appropriate choice for streamflow predictions while considering climatic drivers as inputs (Adarsh and Janga Reddy 2019b). Meidani and Araghinejad (2014) applied lead-time approach based on singular value decomposition (SVD), to correlate autumn (October–December) SSTs with February–May averaged streamflow value of south western Iran. They have further proven that results showed better capability of SSTs temporal expansion series (first mode) than those of well-known climate indices like ENSO, Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO), NAO and Atlantic Multidecadal Oscillation (AMO).

The use of soft computing tools for flood forecasting is much popular in the literature (see Mousavi et al. 2018) and artificial neural networks (ANN) perhaps the most popular one because of being the first recognized tool in the soft computing family of tools and its ability modeling complex nonlinear relationships between multiple physical processes (Sudheer et al. 2002; Hsu et al. 2002; Kisi and Cigizoglu 2007; Wu and Chau 2011). Some of the other studies have used a diverse set of artificial intelligence (AI) approaches or their integrated variants to predict various hydrometeorological variables worldwide (Coulibaly et al. 2001; Sudheer and Jain 2004; Jain and Kumar 2007). Mukerji et al. (2009) performed the flood forecasting of Jamtara gauging site of the Ajay River Basin in Jharkhand, India, using ANN, adaptive neuro-fuzzy inference system (ANFIS) and adaptive neuro-genetic algorithm (GA) hybrid model. In their study, 20 rainfall–runoff events are selected, of which 15 were considered for training and rest of the data were used for validation. Farokhnia et al. (2010) forecasted the drought of northern Iran using SST and SLP indicators as an input of the ANFIS model. Fallah-Ghalhary (2012) showed that the ANN model, with climatic signals as an input, can accurately predict Khorasan Razavi Province precipitation. Choubin et al. (2016) utilized multiple linear regression (MLR), multilayer perceptron (MLP) network, and ANFIS models to predict the seasonal standard precipitation index of Maharlu–Bakhtaran catchment in Iran using large-scale climate signals. They compared the results of MLR, MLP, and ANFIS models and showed that MLP has a better performance compared to the other two models.

The streamflow time-series records are highly nonlinear and non-stationary, and it is still difficult to using tools like ANN to simulate such streamflow measurements (Wei et al. 2013). To circumvent this short coming, the preprocessing using a multiscale decomposition tool is identified as the solution by the researchers because of its ability to capture both spectral and temporal information in a signal (Nourani et al. 2009, 2014). These capabilities eventually make the modeling of a hydro-climatic variable by a machine learning model more efficient. The wavelet transform is a powerful tool to deal with non-stationary time series, as it is unique in detecting unusual events through local time and frequency analysis (Adamowski and Prokoph 2013; Nourani et al. 2014; Rathinasamy et al. 2014; Tiwari and Adamowski 2015; Araghi et al. 2015). In the past two decades, integrated wavelet-artificial neural network (WNN) models are increasingly becoming popular due to their superiority over simple neural network models or other data-based models for accurately fitting and predicting the variable of interest (Kim and Valdes 2003; Wei et al. 2013; Nourani et al. 2014). Wei et al (2013) developed the WNN and ANN modeling approach for 48-month-ahead monthly river discharge prediction in the Weihe River in China. The WNN hybrid model was

found to have significant potential to improve the prediction relative to those of the single ANN model. Wavelet-based approaches have also been used to model other hydro-meteorological variables such as rainfall, discharge, wind, and suspended sediment transported (Wei et al. 2013).

On examining the literature, it is evident that even though numerous studies were reported on flood forecasting, the prediction for longer lead times is still not satisfactory. Moreover, recent review studies highlighted the importance of long-lead prediction of flood, and the importance of the experimentation on the use of machine learning models for flood forecasting (Mosavi et al. 2018). Although the teleconnection between climate indicators and hydrological variables is evident and significant, the direct use of such variables for forecasting of hydrologic extremes like floods is relatively less appeared in the literature, in particular using machine learning-based hybrid models. Also, a multi-time scale decomposition based hybrid modeling is a preferable choice for the prediction of multiscale processes like floods. The hybrid modeling practices using decomposition methods were found to be much efficient in annual runoff forecasting, where the data length is often limited, say less than 60 (Wang et al. 2013, 2015a, b; Zhao et al. 2017; Song et al. 2020; Zhang et al. 2020). Moreover, in some of the comparison studies, the superior performance of WT-based hybrid modeling over empirical mode decomposition (EMD)-based modeling was reported (Karthikeyan and Nagesh Kumar 2013; Zhu et al. 2016). Even though the correlation between SST and SLP signals and extreme precipitation over Madarsoo watershed at the upstream of Golestan dam, Iran, was analyzed (Ruigar and Golian 2016), such analysis for basin-scale flood forecasting is not reported, which is of paramount importance in reducing flooding risks of the basin. Additionally, use of hybrid machine learning models is believed to improve the performance of flood predictions at river basin scale. This, in turn, leads to the development of robust and efficient models that can efficiently solve complex flooding scenarios in an adaptive condition. Therefore, this integrated application of machine learning models in the area is unique and can be useful to solve flood-related vulnerabilities. In this context, this study uses WNN for the flood forecasting of three stations of Madarsoo watershed, Iran, considering the climatic drivers like SLP and SST as inputs.

Study area and dataset

The Madarsoo watershed is located upstream of Golestan dam (Iran) with a 5155 km^2 areal coverage approximately. Data of three hydrometric stations, namely Tangrah, Tamer, and Galikesh, were utilized to find the correlation between climatic indices and maximum monthly discharge of the catchment. Figure 1 shows the geographic location of the watershed and gauging stations.

In the past, many researchers used datasets of 20–55 years long, centering around 40 years for flood or annual runoff forecasting using machine learning models or their hybrid variants (Mukerji et al. 2009; Wang and Ding 2003; Wang et al. 2015a, b; Ruigar and Golian 2016; Zhou et al. 2017; Song et al. 2020; Zhang et al. 2020). This leads to the inference that the annual flood/runoff prediction exercises, even though the number of samples are less, the data diversity will also be less (unlike the case of daily data where the diversity is quite high, ranging from very low to very high) and within a specific range. The hybrid decomposition models could capture such pattern very well which might have led to improved predictions. The maximum monthly discharge values between the years 1974–2011 for March and August of all the three stations were collected and used in this study. March has the highest inter-annual flow values, while the most devastating historical floods have been observed in August. The statistical properties of the datasets are presented in Table 1. This also shows that the diversity is more for the data of August months than that of March month.

The selection of predictor variables is one of the crucial steps in machine learning-based modeling of hydrological variables. Many of the past studies identified SST and SLP as two most influential drivers affecting Iranian climatology (Farokhnia et al. 2010; Meidani and Araghinejad 2014). In a physical sense, by incorporating their influence, an improved forecasting results can be anticipated, as Iran is surrounded with seas on north and south. As the response time of oceans is slower than atmosphere, some efforts have been focused on investigating the links of SST and SLP over oceans with atmospheric changes for climate monitoring or use as potential hydroclimatic predictors (Meidani and Araghinejad 2014). Nazemosadat et al. (1995) showed that the SST variations in the Persian Gulf have significant effects on precipitation variability in the south-western Iran. They attributed the occurrence of localized heavy winter rainfall in the region with movement of the Sudan current toward Iran. Fallah-Ghalhary (2012) analyzed 38 years of rainfall data in Khorasan-e-Razavi province, northeastern Iran and found a strong link between SLP, and SLP difference (ΔSLP) changes with the rainfall of the region. Rezaebanafsheh et al. (2011) reported that the cooler condition of the sea in autumn has led to wetter winter in western Iran, because of its relation with seasonal anomalies of Mediterranean SST. Moreover, the effect of hydro-climatic teleconnections will be influential at coarser spatiotemporal resolution (Maity and Nagesh Kumar 2008). Considering the above feedbacks, the monthly data of climate signals, i.e., SST and SLP, for the period 1974–2011, were acquired from (www.esrl.noaa.gov). The spatial resolution for SST

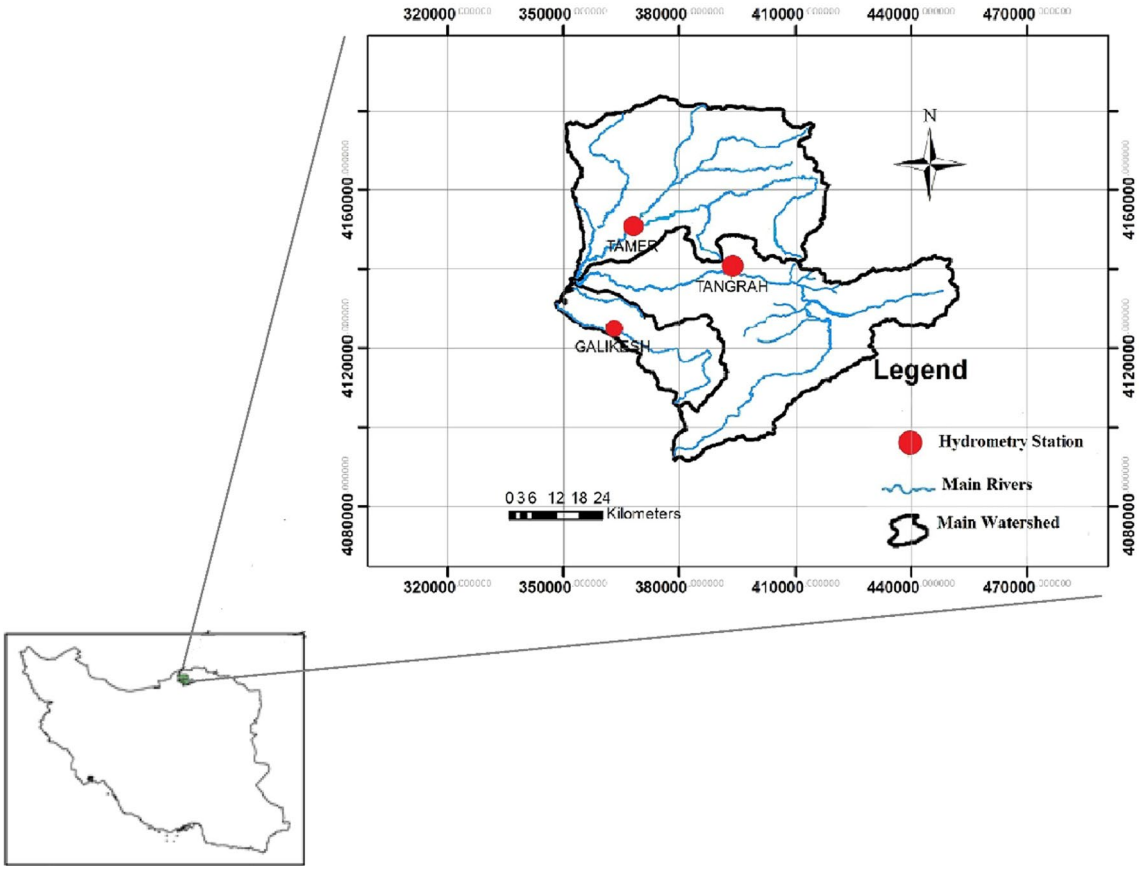


Fig. 1 Location map of hydrometric stations in Madarsoo watershed

Table 1 Statistical properties of maximum monthly discharge data of March and August

Month	Station	Max (m^3/s)	Mean (m^3/s)	Standard deviation	Skewness
March	Tangrah	19.93	6.96	4.93	1.26
	Tamer	37.45	6.82	7.01	2.96
	Galikesh	29.1	10.3	5.84	1.08
August	Tangrah	777	37.2	142.07	4.49
	Tamer	225	18.62	41.05	3.82
	Galikesh	339.3	17.28	57.6	5.1

and SLP networks is 2°. To investigate the relation between SST and SLP with maximum monthly discharge of the three hydrometric stations of Madarsoo watershed, 19 points were selected as indicator points. Fallah-Ghalhary (2012) performed a detailed study in identification of indicator points affecting Iranian climatology employing the factor analysis, Kriging and the statistical correlation. Moreover, the same indicator points were considered by Ruigar and Golian (2015) in the prediction of monthly maximum precipitation of Madarsoo watershed in one of the past study. Table 2 contains the indicator point's information, including their coordinates and their associated sea/ocean. Figure 2 depicts the position of the indicator points.

Methodology

Multiple linear regression

Multiple linear regression (MLR) is a simple method that predicts the values of a dependent variable, Y, based on values of independent variables ($x_1, x_2 \ldots x_n$) (Quraishi and Mouazen 2013). In this study, MLR was used for predicting maximum monthly discharge values based on climatic signals as predictors. MLR equation is as follows:

$$Y = a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n, \quad (1)$$

Table 2 Name and coordinates of points used in the analysis of relationship between SST, SLP and maximum monthly discharge over the study area (Ruigar and Golian 2016)

Points	Latitude (°)	Longitude (°)	Points	Latitude (°)	Longitude (°)
Adiabatic Sea (ADI)	43	15	Caspian sea (CAS)	45	50
Aden Gulf(ADE)	12.5	45	Northern Persian Gulf (NPG)	27.5	50
Arabian Sea (ARA)	20	65	Northern Red Sea (NRS)	25	35
Azores Sea (AZE)	40	330	Northern sea (NS)	55	5
East Mediterranean (EM)	35	30	Oman sea (OS)	22.5	65
Black sea (BS)	45	35	Southern Red Sea (SRS)	15	41
Atlantic Ocean (ATL)	30	320	West Mediterranean Sea (WM)	40	5
Greenland (GR)	35	30	Labrador Sea (LS)	60	310
Indian Ocean (IO)	10	65	Sorannetwork (SN)	20	25
Southern Persian Gulf (SPG)	25	55			

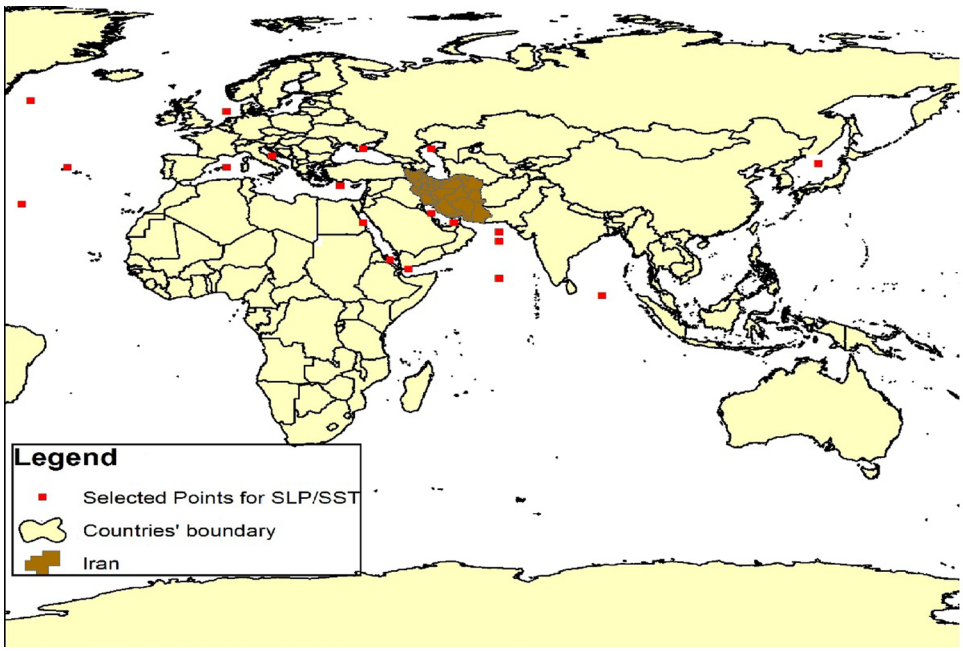


Fig. 2 Schematic view of the 19 indicator points which their SST and SLP data were used in the present study (Ruigar and Golian 2016)

where b_i is the parameter of ith variable, and a is the constant value. In the present research, 80% of the maximum monthly discharge data were used to estimate regression parameters, while the remaining 20% of the data were used for the validation phase.

Artificial neural networks

Neural networks are computational models that can determine the complex relationship between inputs and outputs of a physical system (Ruigar and Golian 2016). An MLP with one hidden layer and sigmoid and linear activation function at hidden and output layers, respectively, is capable of approximating most of the functions with acceptable precision ANN depending on several neurons of the hidden layer (Hecht-Nielson 1987). Figure 3 shows the schematic view of MLP. For each neuron, the input is the weighted sum of the output from the previous layer, and the outputs are calculated as follows:

$$\text{net}_i^n = \sum_{i=1}^{m} W_{ji}^n \times O_j^{n-1}, \tag{2}$$

where net_j^n is the input value of i-th neuron in n-th layer and W_{ji}^n is the connection weight between ith neuron in nth layer and O_j^{n-1} is the output of j-th neuron in the$(n-1)$th layer; m is the number of neurons in the $(n-1)$th layer.

In this study, a multilayer perceptron network (MLP) is used to predict maximum monthly discharge for Madarsoo watershed stations, and all necessary steps have been programmed in MATLAB software. The number of neurons in the hidden layer was obtained by trial and error between 1

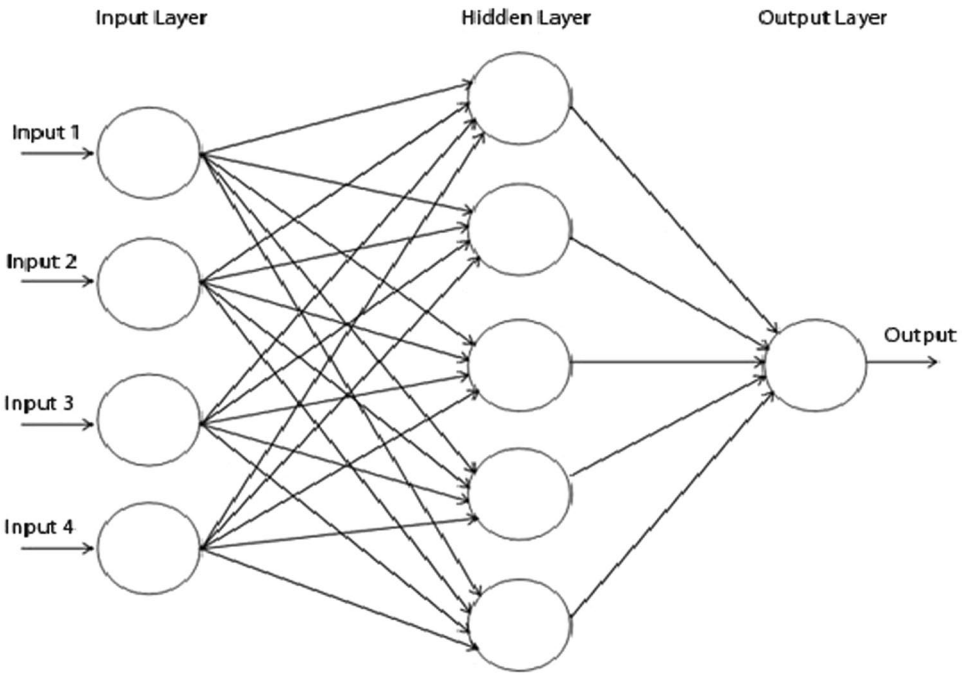


Fig. 3 Schematic view of MLP neural network

and 20 neurons. Also, the Levenberg–Marquardt (LM) algorithm was used for training the network with activation functions as sigmoid and linear transfer functions in the hidden and output layers, respectively. As stated before, large-scale climate signals with the highest correlation values with maximum monthly discharge and maximum monthly discharge itself formed the input–output datasets.

Wavelet transform

The WA is an advancement over short-time Fourier transform used to discover time characteristics in data (Daubechies 1990). For signal processing, the wavelet transform depends on the scale (or frequency) and time. The most common two wavelets transform approaches are continuous (CWT) and discrete (DWT) wavelet transform. The CWT is preferred for working with functions defined over the whole real axis.

Continuous wavelets transform (CWT).

CWT is a tool that converts the one-dimensional function to a two-dimensional function. The new space basis is the wavelet functions. CWT can be expressed by the following equation:

$$\mathrm{CWT}_x^{\psi}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} \int X(t)\psi_{a,b}^{*}(t)\mathrm{d}t. \tag{3}$$

This relationship is a function of two variables of (a) and (b). The results of this transformation are the wavelet coefficients of CWT. By multiplying each coefficient in the wavelet, the wavelet component is obtained (Polikar 1996).

Discrete wavelets transform (DWT).

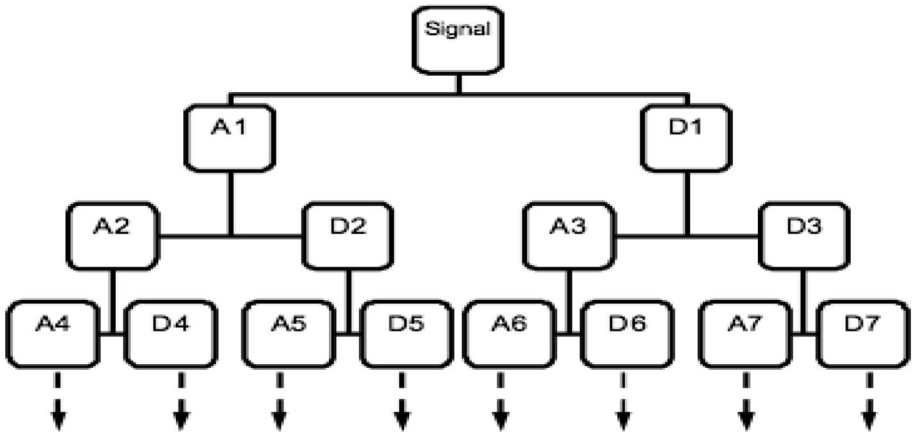
The DWT can be equated using the following equation:

$$\mathrm{DWT}_{f(t)} = \sum_{j,k} a_{j,k}\left(2^{\frac{j}{2}}\psi - k\right). \tag{4}$$

Like the CWT and DWT are also adjustable. Based on the coefficients of the discrete wavelet, the original wave can be reconstructed under some situations. Recently, DWT is used for extracting information about non-stationary signals (Daubechies 1990).

In this study, first, the DWT method is used for data decomposition because CWT requires a significant amount of data and computation time (Partal 2009). Also, one of the most widely used Daubechies wavelet 5 (db5) was employed as the mother wavelet to decompose the series (Wei et al. 2013). The schematic view of the multistage decomposition of the wave is shown in Fig. 4. The general rule for an appropriate scale of decomposition degree is its wavelet that the greatest degree of decomposition should be less than the size of test data in prediction (Wu et al. 2006). In this study, the testing sizes of datasets are 7 months (7 years at August and March). Therefore, the input datasets are decomposed into various details (D5) and an approximation (A5) at two resolution levels (21–22) using the db5 decomposition level. Next, the fitted DWs are selected as the inputs of ANNs for the estimation of maximum monthly discharge. This phase is the most significant and effective part of the ANN estimation.

Fig. 4 Wavelet decomposition



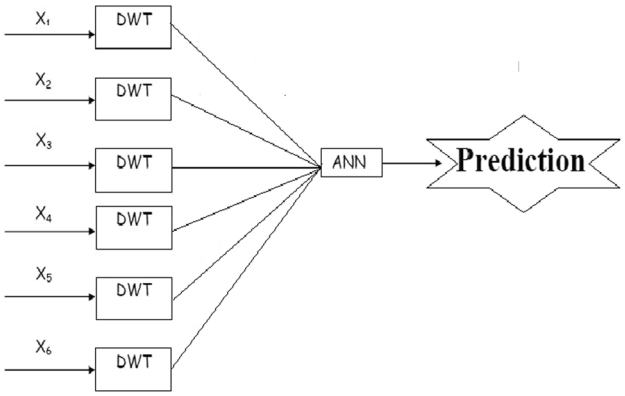


Fig. 5 Schematic view of wavelet neural network method (Partal and Cigioghlu 2009)

Figure 5 shows a schematic view of prediction using the wavelet neural network (WNN) method.

Model Performance Evaluation

All input and output data values are standardized to range between zero and one using the following equation:

$$X_n = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \tag{5}$$

where X_n is standardized data; i, max, and min indices show i-th, maximum and minimum of original data, respectively.

To determine the performance of models for discharge prediction, root-mean-square error (RMSE), correlation coefficient (R), mean absolute error (MAE), and Nash–Sutcliffe coefficient (CNS) were used in this study (Dibike and Solomatine 2001; Baratti et al. 2003). These indices are defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (Q_s - Q_o)^2}{n}} \tag{6}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} |Q_s - Q_o|}{n} \tag{7}$$

$$\text{CNS} = 1 - \left(\frac{\sum_{i=1}^{n} (Q_o - Q_s)^2}{\sum_{i=1}^{n} (Q_o - \overline{Q})^2} \right) \tag{8}$$

$$R = \left[\frac{\sum_{i=1}^{n} (Q_o - \bar{Q}_o)(Q_s - \bar{Q}_s)}{\sqrt{\sum_{i=1}^{n} (Q_o - \bar{Q}_o)^2} \sqrt{\sum_{i=1}^{n} (Q_s - \bar{Q}_s)^2}} \right], \tag{9}$$

where Q_o is the observed discharge; Q_s is the simulated value and $\overline{Q}_o$ and $\overline{Q}_s$ are the average of observed and simulated discharge values, respectively.

To compare the prediction performance of the three data-driven methods, a classical split sample test was used in the terminology of Klemes (1986). The input data is divided into two samples. Data from 1974 to 2004 (31 years), which includes 80% of data, were used to train the data-driven methods, whereas during the years 2005 to 2011 (7 years), which includes 20% of data, used to validate the trained algorithms.

Results and discussion

Relationship between climate signals and maximum monthly discharges

The Pearson correlation coefficient between each input variable (SST and SLP) and maximum monthly discharge values for each month and various lag times (from 1 to 12 months) was calculated at significance levels of 1% and 5%. Then, variables with a significant correlation with discharge data were used as input variables for the data-driven method to predict discharges. The correlation coefficients and lag times for the different indicator points and hydrometric stations are in Table 3. For the Tangrah station, indicators with a correlation coefficient of 0.735 for March and 0.909 for August are found as the highest values. For the Tamer and the Galikesh stations, the correlation coefficient for March and August are 0.644, 0.447 and 0.782, 0.389, respectively. This may be related to the size of the sub-catchments. Tangrah catchment is the largest among the three, and floods in large catchments tend to be more influenced by large-scale weather patterns, which can be linked to SST and SLP. Hence, a higher correlation between SST and SLP with discharges in a larger catchment is expected, whereas in smaller catchments, the flood process may be mostly due to local conditions.

Prediction of maximum monthly discharges

As discussed before, the training model includes SST and SLP signals that have the best correlation coefficient with maximum monthly discharge. Table 4 shows the performance evaluation of the multiple linear regression (MLR) method for the training and test datasets for March and August. For example, in August, the value of RMSE, *R*, and CNS for Tangrah hydrometric station are 3.39, 0.985, and 0.938 m^3/s, respectively, for the training phase and 8.7, 0.593 and 0.421 m^3/s for the test phase. It can be seen that the training performance of the model is satisfactory, i.e., the model is flexible enough to reproduce the main characteristics of the dataset. However, it is important to notice that the performances for the validation phase are weaker. For example, CNS for August at Tangrah stations drops to 0.421, which is about 50% lower compared to the training phase. A

Table 3 Correlation coefficients and lag times for different indicator points and hydrometric stations used for peak discharge prediction

	Tangrah				Tamer				Galikesh			
M	Point	*R*	M	*L*	Point	*R*	M	*L*	Point	*R*	M	*L*
March	BS SST	0.735**	May	10	SPG SST	0.511**	Sep	6	GR SST	0.782**	Feb	1
	GR SST	0.733*	Jan	2	EM SST	0.644**	Sep	6	GR SST	0.632*	Jan	2
	GR SST	0.565**	Feb	1	CAS SST	0.591*	Sep	6	EM SST	0.768**	May	10
	CAS SST	0.713*	May	10	CAS SST	0.591**	Aug	7	BS SST	0.746**	May	10
	EM SST	0.698**	May	10	BS SST	0.538*	July	8	–	–		–
	SPG SST	0.679**	May	10	BS SST	0.578**	May	10	–	–		–
August	EM SST	0.909**	March	5	GR SST	0.327**	Oct	10	ATL SST	0.358*	Jan	7
	NPG SLP	0.404*	May	3	GR S ST	0.347*	Dec	8	AZE SLP	0.344**	Nov	9
	OS SLP	0.342*	May	3	AZE SLP	0.425*	Feb	6	CAS SLP	0.343*	July	1
	SPG SLP	0.393**	May	3	AZE SLP	0.311**	May	3	NPG SLP	0.389*	May	3
	ADG SLP	0.338*	Nov	9	AZE SLP	0.447*	Nov	9	OS SLP	0.342**	May	3
	IO SLP	0.408**	Nov	9	CAS SLP	0.319*	Apr	4	SPG SLP	0.376**	May	3

M, month; L, lag

*Correlation is significant at the 0.05 level

**Correlation is significant at the 0.01 level

Table 4 Performance of the multiple linear regression model for flood prediction of the three stations of Madarsoo watershed

Station	Month	Train				Test			
		RMSE (m^3/s)	MAE (m^3/s)	*R*	CNS	RMSE (m^3/s)	MAE (m^3/s)	*R*	CNS
Tangrah	March	6.88	3.97	0.894	0.784	15.44	14.83	0.374	0.276
	August	3.39	2.62	0.986	0.938	8.70	7.78	0.593	0.421
Tamer	March	8.02	5.66	0.822	0.675	15.67	14.49	0.298	0.227
	August	25.43	11.70	0.761	0.544	84.60	60.53	0.153	0.087
Galikesh	March	1.96	1.44	0.961	0.907	4.92	4.66	0.698	0.615
	August	5.36	3.30	0.721	0.655	0.161	0.128	0.571	0.410

similar decrease in performance from training to the validation phase is found for the other months and stations.

Such high performance for training and poor performance for the testing period indicate the possibility of overfitting. This shows that the MLR approach is not reliable enough in predicting the maximum monthly discharges. Bergström (1991), for example, states that "If the model performance is significantly lower for the independent period used for validation than it was for the calibration period, the modeler should seriously consider if there are problems of overparameterization. The model may simply have too many degrees of freedom for the information contained in the observed records". Although this statement refers to classical rainfall–runoff modeling, it also holds true in the case of data-driven forecasting. Hence, the large drop in the performance from training to validation phase questions the capability of the MLR approach to successfully predict the future discharge series.

Next, the ANN is invoked to develop the flood prediction model, using the same inputs considered for the development MLR model. The performance indices for the ANN model are given in Table 5. Similar to the MLR method, the use of ANN also resulted in very good performance for the training phase. For example, the CNS for all three stations is about 0.85 or larger, while the R values for the training phase are larger than 0.92 for all stations and months. However, unlike the MLR, the decline in the performance of the ANN model for the validation phase is much smaller, e.g., for Tangrah Station, the CNS for March decreases from 0.903 to 0.772. This indicates that ANN is a much more stable approach for predicting the maximum monthly discharge in this case study. Further it is noted that the use of ANN results in better performance for flood prediction of stations with smaller catchments. Tamer and Galikesh than the Tangrah station. Also it is noted that the flood prediction for August month is better than that of March in the stations of Tamer and Galikesh (*R* values of 0.989 and 0.941 in August against 0.948 and 0.878 in March), while a contrasting behavior is noted in the flood prediction of Tangrah station. This could also be connected with faster accumulation of flow in smaller catchments leading to the devastating floods in August. For the Tangrah station with larger drainage area, the dampening effect will be more leading to more lag time in flow accumulation and subsequent lower inter-annual flood values. Moreover, the smaller difference in performance between the training and testing data further reinforces that the ANN model is not overfitted and well-trained and exhibit better generalization capability (Ghosh and Mujumdar 2007).

Using WNN to predict discharges results in excellent performance for the training phase (Table 6). For all three stations, CNS is always higher than 0.97, and *R* is larger than 0.99. This means that there is a nearly perfect consistency between the observed and predicted discharges. Also, the performance measures for the validation phase are very high, with a CNS larger than 0.79 and *R* larger than 0.87. In the case of WNN model also, the performance of prediction of Tamer and Galikersh is slightly better than

Table 5 Performance of ANN model for flood prediction of the three stations of Madarsoo watershed

Station	Month	Train				Test			
		RMSE (m^3/s)	MAE (m^3/s)	*R*	CNS	RMSE (m^3/s)	MAE (m^3/s)	*R*	CNS
Tangrah	March	6.12	4.28	0.951	0.903	2.76	1.06	0.887	0.772
	August	0.08	0.08	0.999	0.999	3.54	2.62	0.793	0.710
Tamer	March	5.81	3.60	0.948	0.882	3.68	3.02	0.922	0.849
	August	1.58	0.68	0.998	0.996	1.55	0.98	0.989	0.948
Galikesh	March	3.11	2.63	0.961	0.853	3.40	3.07	0.878	0.728
	August	0.23	0.16	0.999	0.999	2.47	1.27	0.941	0.873

Table 6 Performance of WNN model for flood prediction of the three stations of Madarsoo watershed

Station	Month	Train				Test			
		RMSE (m^3/s)	MAE (m^3/s)	*R*	CNS	RMSE (m^3/s)	MAE (m^3/s)	*R*	CNS
Tangrah	March	0.54	0.46	0.999	0.998	0.69	0.46	0.943	0.935
	August	0.08	0.08	0.999	0.999	2.39	1.39	0.866	0.791
Tamer	March	1.99	1.32	0.99	0.979	1.25	0.96	0.953	0.892
	August	1.35	0.68	0.998	0.996	0.68	0.75	0.996	0.990
Galikesh	March	0.41	0.30	0.998	0.996	0.41	0.37	0.936	0.793
	August	0.19	0.12	0.999	0.999	0.39	0.27	0.978	0.914

that of Tangrah station. Also, the predictability of August flood is more reliable in these two stations than the August floods of Tangrah station.

Results showed that the MLR is not stable, and hence MLR is not suggested to predict future discharges in our case study. This might be due to the nonlinear relationship between the predictor and predicted variables that cannot be captured by the MLR method. In other words, the ANN model can predict discharge with better performance due to its more flexibility and capability to model nonlinear relationships (Tushar and Bhatt 2013). ANN and WNN approaches show satisfactory performance measures, while WNN slightly outperforms the ANN approach in this case study. Notably, the decrease in performance between the training and validation phase is marginal, indicating that WNN is a more stable approach to predict discharges.

Interestingly, the performance indicators for ANN and WNN are similar for predicting discharge values in March and August, although the variability in the runoff in August is higher than runoff in March. The coefficients of variation (CV) in the runoff in March for the Tangrah, Tamer, and Galikesh stations are 0.71, 1.03, 0.57, respectively (Table 1), while CV increases to 3.82, 2.20, and 3.33 in August, respectively. The teleconnection signals SLP and SST seem to have a similar strong link to the runoff generation processes in March and August, although the variability of discharges is different. Using the WNN approach, the predicted discharges for Tangrah, Tamer, and Galikesh stations are plotted against the observed discharges (Fig. 6).

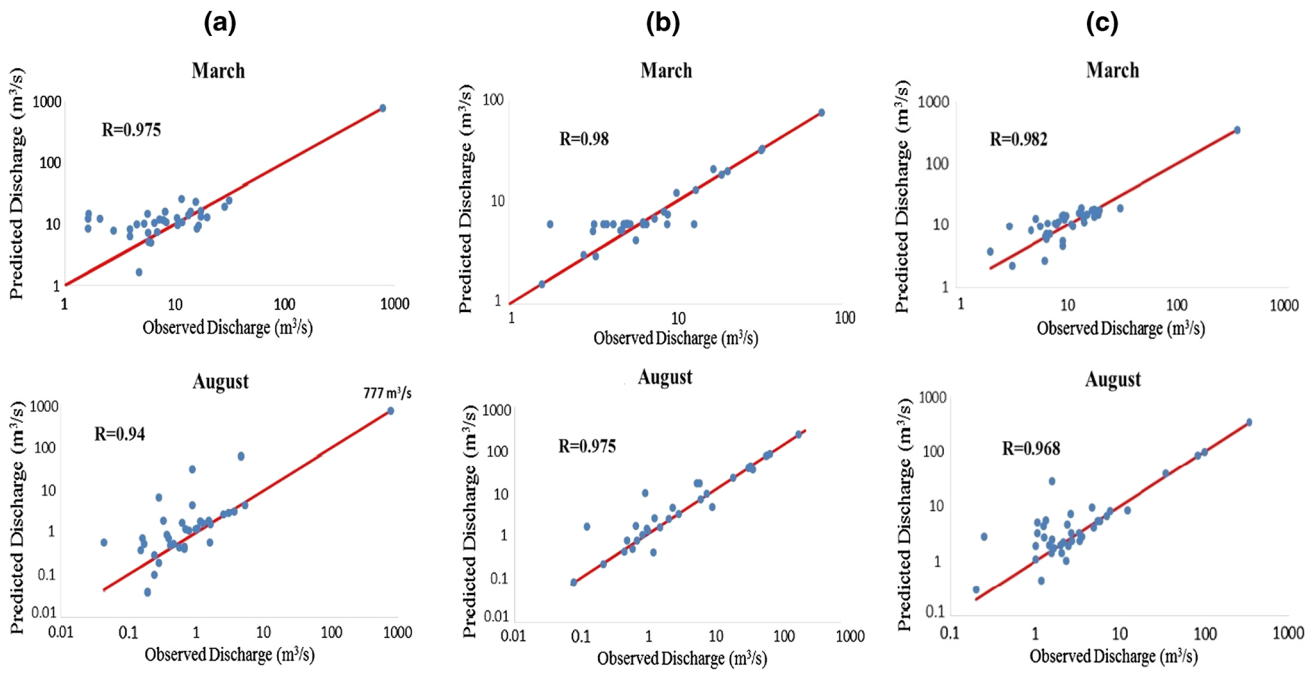


Fig. 6 Scatter plots of observed versus simulated maximum monthly discharge values for March and August months using the best WNN model for **a** Tangrah and **b** Tamer **c** Galikesh hydrometric stations. Upper panels show the results for the month of March and lower panel shows the results of month of August

Due to the skewed distribution of discharges, the values are plotted in Log–Log axes, and the 1:1 line is given to show over- and underestimations better. There is a slight tendency to overestimate smaller discharges in March, most visible at the Tangrah station. However, both over and underestimation errors tend to be equally distributed across all discharge ranges in August. Also, the large and devastating floods that occurred in August 2001 and 2004 in Tangrah station were predicted reasonably well. As shown in Fig. 6, the WNN is capable of predicting large flood events, e.g., the flood of August 2001 at Tangrah station with a peak discharge of 777 m^3/s, is accurately predicted. Comparing R calculated with and without this flood event (777 m^3/s) showed that the correlation coefficient changes from 0.94 to 0.88 after removing this event from the dataset. It can be concluded that the omission of this value has influenced the interpretation of the results.

Although the maximum correlation of the teleconnection signals and discharges (Table 4) are much higher for the Tangrah station compared to the other two stations, i.e., Tamer and Galikesh, the performance of the discharge predictions is not much different. For example, CNSs for the validation phase of Tangrah station are 0.935 and 0.791. The CNSs for Tamer and Galikesh stations are 0.892, 0.990, 0.793, and 0.914. This indicates that the WNN is not too sensitive to the selection of the indicator station. Even SST and SLP signals from stations with a lower correlation of teleconnection signals and discharge seem to contain enough information of runoff generation to predict maximum discharges at a reasonable performance.

To test the capabilities of the developed data-driven models, we used peak floods of 2012, 2013, and 2014, which are not included in the training, or validation of the models in the previous stages. Tables 7 and 8 represent predicted and observed discharge values in March and August using ANN and WNN models. Similar to the previous results, ANN and WNN predicted the discharges reasonably well, while WNN exhibited better performance than other models.

Table 7 shows that in all the cases except one, the WNN predictions gives much closer estimates to the observed values. The difference of WNN estimates from observed discharge ranges between 1.44 and 7.5 m^3/s in different cases, while it ranges between 1.1 and 18.6 m^3/s in different cases. The slightly better performance or ANN over WNN is noted for March 2012 flood of Galikesh, where the peak discharge was anomalously high. In this case also, the estimate by WNN was satisfactory (22.2 by WNN and 20.4 by ANN against the observed value of 19.3) and within acceptable range of 3%. Table 8 shows that in the prediction of August floods of both of the years, the predictions by WNN was much impressive than that by ANN. All the ANN predictions were found to be over estimations, with a mean difference of 5 m^3/s (difference in predictions ranging between 2.3 and 8.67 m^3/s). The difference predictions by WNN were less than 2.1 m^3/s in all the cases.

Table 7 Predicted and observed flood discharge in March

Year	Station	Observed discharge (m^3/s)	ANN predicted discharge (m^3/s)	WNN predicted discharge (m^3/s)
2012	Tangrah	14.8	9.11	17.04
	Tamer	7.47	20.01	3.75
	Galikesh	19.3	20.42	22.19
2013	Tangrah	4.18	19.91	11.67
	Tamer	6.02	24.65	8.32
	Galikesh	12.9	17.11	14.34

Table 8 Predicted and observed flood discharge in August

Year	Station	Observed discharge(m^3/s)	ANN predicted discharge (m^3/s)	WNN predicted discharge (m^3/s)
2012	Tangrah	6.71	11.5	5.06
	Tamer	4.18	10.8	3.37
	Galikesh	5.65	7.95	6.78
2013	Tangrah	0.246	4.52	1.51
	Tamer	0.19	8.86	2.25
	Galikesh	1.08	4.56	2.59

The study clearly established that the MLR method is suffering from overfitting problem and the a pure linear model is model is not enough for flood modeling, which is quite anticipated because of nonlinear processes. The ANN model can predict the flood with reasonable degree of accuracy, while wavelet-based processing can enhance the performance of ANN in flood modeling, similar to the observations made in the studies pertaining to several other hydrological variables (Jian and Adeli 2005; Adamowski and Chan 2011; Adamoswski et al. 2012; Wei et al. 2012; Belayneh et al. 2014). The improved accuracy using the WNN approach will immensely help the policymakers to predict the floods with sufficient lag times and manage the water resources such that the negative impact of floods on society can be reduced to as much extent as possible. Based on the developed models, flood hazard mapping and flood early warning system developments should be done for the basin with an objective to reduce the flooding risks. Similar studies are to be encouraged in watersheds of diverse size and with datasets of different length and property diversity. Such experimentations may give potential insights to the capabilities of decomposition-based hybrid modeling in flood modeling or annual runoff predictions, which are often constrained by data length. Further, the use of climatic signals as potential inputs for modeling floods needs further

investigation. Such studies can be performed at diverse spatial scales by considering more oscillations and interrelationships among them, using multiscale methods like the popular wavelet coherence or time-dependent intrinsic correlations (Adarsh and Janga Reddy 2019a, b) in different parts of the world. Moreover, the hybrid variants can be developed using other artificial intelligence (AI) methods, to identify the best models to deal with the uncertainty in performance in long lead prediction of floods.

Conclusion

This study investigated the potential of WNN for the flood prediction of three hydrometry stations of Madarsoo watershed, Iran, considering large-scale climatic signals SLP and SST as inputs. The predictions of maximum discharges were made for the months of August and March using WNN and the performance comparisons were made with the predictions by the ANN and MLR. The nonlinear relationships between flood discharge and climatic drivers was captured well by both the ANN and its wavelet hybridization, while MLR was not found to be successful in this exercise. The generalization capabilities of WNN were found to be better than that of ANN despite the smaller sample size, in the flood modeling of Madarsoo watershed as confirmed by a set of diverse statistical correlation and error measures. The best performance was noted in the prediction of August flood of Tamer hydrometry station by WNN (RMSE, *R* and CNS of 0.68, 0.996, and 0.99 m^3/s, respectively, against 1.55, 0.989 and 0.95 by ANN model) for the test period of 2005–2011. The flood predictions of August month, which is characterized with more diversity in data, was found to be better for the stations served by smaller drainage area by both ANN and WNN methods. Moreover, the performance in flood forecasts of August and March of 2012 and 2013 years were made by WNN and ANN methods and WNN found to be performing remarkably well than the ANN model. This study further shows the potential of teleconnected climate signals in transferring the climate–rainfall linkages to catchment-scale processes to predict maximum monthly discharge and hydrologic extremes like flood.

Declarations

Conflict of interest The authors declared that they have no conflict of interest.

References

Adamowski J, Chan HF (2011) A wavelet neural network conjunction model for groundwater level forecasting. J Hydrol 407(1–4):28–40. https://doi.org/10.1016/j.jhydrol.2011.06.013

Adamowski J, Chan HF, Prasher SO, Sharda VN (2012) Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. J Hydroinfor 14(3):731–744

Adamowski J, Prokoph A, Adamowski K (2013) A spectral analysis based methodology to detect climatological influences on daily urban water demand. Math Geosci 45:49–68

Adarsh S, Janga Reddy M (2019a) Links between global climate teleconnections and indian monsoon rainfall. In: Venkataraman C, Mishra T, Ghosh S, Karmakar S (eds) Climate change signals and response. Springer, New York, pp 61–72. https://doi.org/10.1007/978-981-13-0280-0

Adarsh S, Janga Reddy M (2019b) Multiscale characterization and prediction of reservoir inflows using MEMD-SLR coupled approach. J Hydrol Eng ASCE. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001732

Akansu AN, Serdijn WA, Selesnick IW (2010) Wavelet transforms in signal processing: a review of emerging applications. Phys Commun 3:1–18

Araghi A, Adamowski J, Nalley D, Malard J (2015) Using wavelet transforms to estimate surface temperature trends and dominant periodicities in Iran based on gridded reanalysis data. Atmos Res 155:52–72

Araghi A, Mousavi-Baygi M, Adamowski J, Martinez C (2017) Association between three prominent climatic teleconnections and precipitation in Iran using wavelet coherence. Int J Climatol 37(6):2809–2830

Baratti R, Cannas B, Fanni A, Pintus M, Sechi GM, Toreno N (2003) River discharge forecast for reservoir management through neural networks. Neurocomputing 55:421–437

Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. J Hydrol 508:418–429

Berg N, Hall A, Capps SB, Hughes M (2013) El Niño-Southern Oscillation impacts on winter winds over Southern California. Clim Dyn 40:109–121

Bergström S (1991) Principles and confidence in hydrological modelling. Hydrol Res 22:123–136

Biasutti M, Held IM, Sobel AH, Giannini A (2008) SST forcings and Sahel rainfall variability in simulations of the twentieth and twenty-first centuries. J Climate 21:3471–3486

Chang FJ, Chiang Y-M, Chang L-C (2007) Multi-step-ahead neural networks for flood forecasting. Hydrol Sci J 52(1):114–130

Choubin B, Khalighi-Sigaroodi S, Malekian A, Kişi O (2016) Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. Hydrol Sci J. https://doi.org/10.1080/02626667.2014.966721

Coulibaly P, Anctil F, Bobée B (2000) Daily reservoir indischarge forecasting using artificial neural networks with stopped training approach. J Hydrol 230:244–257

Coulibaly P, Bobée B, Anctil F (2001) Improving extreme hydrologic events forecasting using a new criterion for artificial neural network selection. Hydrol Process 15(8):1533–1536

Daubechies I (1990) The wavelet transform time-frequency localization and signal analysis. IEEE Trans Inf Theory 36:961–1005

Dawson CW, Wilby R (1998) An artificial neural network approach to rainfall-runoff modeling. Hydrol Sci J 43(1):47–66

Dehghani M, Salehi S, Mosavi A, Nabipour N, Shamshirband S, Ghamisi P (2020) Spatial analysis of seasonal precipitation over Iran: co-variation with climate indices. ISPRS Int J Geo-Inf 9(2):73. https://doi.org/10.3390/ijgi9020073

Dibike YB, Solomatine DP (2001) River discharge forecasting using artificial neural networks. Phys Chem Earth Part B 26:1–7

Fallah-Ghalhary GH (2012) Rainfall prediction using teleconnection patterns through the application of artificial neural networks. Mod Climatol 1:362–386

Farokhnia A, Morid S, Byun HR (2010) Application of global SST and SLP data for drought forecasting on Tehran plain using data mining and ANFIS techniques. Theoret Appl Climatol 104:71–81

Gamiz-Fortis SR, Esteban-Parra MJ, Trigo RM, Castro-Diez Y (2010) Potential predictability of an Iberian river discharge based on its relationship with previos winter global SST. J Hydrol 385:143–149

Ghosh S, Mujumdar PP (2007) Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment. Water Resour Res 43:W07405. https://doi.org/10.1029/2006WR005351

Gupta V, Jain MK (2020) Impact of ENSO, global warming, and land surface elevation on extreme precipitation in India. J Hydrol Engng 25(1):05019032

Gupta V, Jain MK (2021) Unravelling the teleconnections between ENSO and dry/wet conditions over India using nonlinear Granger causality. Atmos Res 247:105168

Gupta V, Singh V, Jain MK (2020) Assessment of precipitation extremes in India during the 21st century under SSP1-1.9 mitigation scenarios of CMIP6 GCMs. J Hydrol 590:125422

Haykin S (1994) Neural networks: a comprehensive foundation. MacMillan Publishing Company, New York

Hecht-Nielson R (1987). Kolmogorov's mapping neural network existence theorem. In: IEEE international conference on neural Net 3. IEEE Press, New York

Hsu KL, Gupta HV, Gao X, Sorooshian S, Imam B (2002) Self-organizing linear output map (SOLO): an artificial neural network suitable for hydrologic modeling and analysis. Water Resour Res 38(12):38–41

Jain A, Kumar AM (2007) Hybrid neural network models for hydrologic time series forecasting. Appl Soft Comput 7(2):585–592

Jiang X, Adeli H (2005) Dynamic wavelet neural network model for traffic flow forecasting. J Trans Eng 131(10):771–779

Kalra A, Miller WP, Lamb KW, Ahmad S, Piechota T (2012) Using large-scale climatic patterns for improving long lead time stream discharge forecasts for Gunnison and San Juan River Basins. Hydrol Process 27:1543–1559

Karthikeyan L, Nagesh Kumar D (2013) Predictability of non-stationary time series using wavelet and EMD based ARMA models. J Hydrol 502:103–119

Kim TW, Valdes JB (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural network. J Hydrol Eng 8:319–328

Kisi O, Cigizoglu HK (2007) Comparison of different ANN techniques in river discharge prediction. Civil Eng Envrion Syst 24:211–231

Klemeš V (1986) Operational testing of hydrological simulation models. Hydrol Sci J 31:13–24

Kumar DN, Reddy MJ, Maity R (2007) Regional discharge forecasting using large scale climate teleconnections and artificial intelligence techniques. Int J Intell Syst 16:307–322

Maity R, Nagesh Kumar D (2008) Basin-scale stream-flow forecasting using the information of large-scale atmospheric circulation phenomena. Hydrol Process 22(5):643–650

Maity R, Nagesh Kumar D (2009) Hydroclimatic influence of large-scale circulation on the variability of reservoir inflow. Hydrol Process 23(6):934–942

Meidani E, Araghinejad S (2014) Long-lead stream discharge forecasting in the southwest of Iran by sea surface temperature of the mediterranean sea. J Hydrol Eng 1–10

Mizanur R, Rafiuddin MM (2013) Seasonal forecasting of Bangladesh summer monsoon rainfall using simple multiple regression model. J Earth Syst Sci 122:551–558

Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: literature review. Water 10(11):1536. https://doi.org/10.3390/w10111536

Mukerji A, Chatterjee C, Raghuwanshi NS (2009) Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. J Hydrol Eng. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000040

Munoz-Diaz FS, Rodrigo FS (2006) Effects of the North Atlantic oscillation on the probability for climatic categories of local monthly discharge in Southern Spain. Int J Climatol 23:381–397

Nazemosadat MJ, Cordey I, Eslamian S (1995) The impact of the Persian Gulf Sea surface temperature on Iranian rainfall. In: Proceedings of the Iranian water resource management conference, Esfahan, Iran, pp 809–819

Nourani V, Alami MT, Aminfar MH (2009) A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng Appl Artif Intell 22:466–472

Nourani V, Baghanam AH, Adamowski J, Kisi O (2014) Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. J Hydrol 514:358–377

Oubeidillah AA, Tootle G, Anderson SR (2012) Atlantic Ocean sea-surface temperatures and regional stream discharge variability in the Adour-Garonne basin, France. Hydrol Sci J 57:496–506

Partal T, Cigoghlu K (2009) Prediction of daily precipitation using wavelet—neural networks. Hydrol Sci J 54:234–246

Polikar R (1996) Fundamental concepts and an overview of the wavelet theory, 2nd edn. Rowan University College of Engineering Web Servers, Glassboro

Quraishi MZ, Mouazen AM (2013) Development of a methodology for in situassessment of topsoil dry bulk density. Soil till Res 126:229–237

Rathinasamy M, Adamowski J, Khosha R (2014) Multiscale streamflow forecasting using a new Bayesian model average based ensemble multi-wavelet Volterra nonlinear method. J Hydrol 507:186–200

Rezaebanafsheh M, Jahanbakhsh S, Bayati M, Zeynali B (2011) Forecas autumn and winter precipitation of west of Iran applying Mediterranean SSTs in summer and autumn. Phys Geogr Res Quart 74:47–62 (**In Persian**)

Ruigar H, Golian S (2016) Prediction of precipitation in Golestan dam watershed using climate signals. Theoret Appl Climatol 123:671–682

Song P, Liu W, Sun J, Wang C, Kong L, Nong Z, Lei X, Wang H (2020) Annual runoff forecasting based on multi-model information fusion and residual error correction in the Ganjiang River Basin. Water 12:2086. https://doi.org/10.3390/w12082086

Soukup TL, AzizOA TGA, Piechota TC, Wulff S (2009) Long lead-timestreamdischarge forecasting of the North Platte River incorporating ocean-atmospheric climate variability. J Hydrol 368:131–142

Sudheer KP (2005) Knowledge extraction from trained neural network river flow models. J Hydrol Eng 10(4):264–269

Sudheer KP, Jain A (2004) Explaining the internal behaviour of artificial neural network river flow models. Hydrol Process 18(4):833–844

Sudheer KP, Gosain AK, Ramasastri KS (2002) A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrol Process 16(6):1325–1330

Tabari H, Abghari H, Talaee PH (2014) Impact of the North Atlantic Oscillation on streamdischarge in Western Iran. Hydrol Process 28(15):4411–4418

Tiwari M, Adamowski J (2015) Medium-term urban water demand forecasting with limited data using an ensemble wavelet-bootstrap machine-learning approach. J Water Resour Plan Manag 141:2

Tushar M, Patel BMN (2013) ANN and MLR model for shear stress prediction of Eicher 11.10 chassis frame. Int J Mech Sci 4:216–223

Wang W, Ding J (2003) Wavelet network model and its application to the prediction of hydrology. Nat Sci 1(1):67–71

Wang WC, Xu DM, Chau KW, Chen S (2013) Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. J Hydroinform 15(4):1377–2139

Wang WC, Chau KW, Xu DM, Chen XY (2015a) Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. Water Resour Manage 29(8):2655–2675. https://doi.org/10.1007/s11269-015-0962-6

Wang WC, Chau KW, Qiu L, Chen YB (2015b) Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. Environ Res 139:46–54. https://doi.org/10.1016/j.envres.2015.02.002

Wei S, Song J, Khan NI (2012) Simulating and predicting river discharge time series using a wavelet-neural network hybrid modelling approach. Hydrol Process 26(2):281–296

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

Wu CL, Chau KW (2011) Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. J Hydrol 399(3–4):394–409

Wu A, Hsieh WW, Tang B (2006) Neural network forecasts of the tropical Pacific sea surface temperatures. Neural Netw 19:145–154

Zhang J, Li H, Bin S, Fang H (2020) Multi-time scale co-integration forecast of annual runoff in the source area of the Yellow River. J Wat Clim Change 12(1):101–115

Zhao Z, Chen X, Xu Y, Xi D, Zhang Y, Zheng Z (2017) An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting. Water 9:153. https://doi.org/10.3390/w9030153

Zhu S, Zhou J, Ye L et al (2016) Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of Yangtze River, China. Environ Earth Sci 75:531. https://doi.org/10.1007/s12665-016-5337-7