OKU Fen Bilimleri Enstitüsü Dergisi
5(3): 1141-1154, 2022

OKU Journal of The Institute of Science and Technology, 5(3): 1141-1154, 2022

Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi

Osmaniye Korkut Ata University Journal of The Institute of Science and Technology

Estimation of Streamflow Using Different Artificial Neural Network Models

Cihangir KOYCEGIZ[1*], Meral BUYUKYILDIZ[2]

[1,2]Konya Technical University Faculty of Engineering and Natural Sciences, Department of Civil Engineering, 42250, Konya

[1]https://orcid.org/0000-0002-0510-1164
[2]https://orcid.org/0000-0003-1426-3314
*Corresponding author: ckoycegiz@ktun.edu.tr

Research Article

Article History:
Received: 16.12.2021
Accepted: 10.03.2022
Published online: 12.12.2022

Keywords:
Artificial neural network
Seyhan basin
Streamflow
Water resources

ABSTRACT

Accurate estimation of streamflow is crucial for water resources planning, design and management, determining of flood and drought management strategies, and minimizing their adverse effects. In this study, the usability of Artificial Neural Network (ANN) models to estimate of monthly streamflow was investigated. For this purpose, monthly data of two stations located in the Seyhan Basin in the south of Turkey were used. The data of Sarız River-Şarköy observation station (No: D18A032) for the streamflow and Sarız meteorology station (No: 17840) for precipitation were used. The precipitation and flow data used belong to the period 1990-2017. Nine input combinations consisting of lags of streamflow and precipitation data were obtained and used in ANN models. We used two ANN techniques, namely Multilayer Perceptron (MLP) and Radial Basis Neural Networks (RBNN) to estimate the monthly streamflow. In the MLP technique, three learning algorithms with gradient descent with momentum and adaptive learning rule backpropagation (GDX), Levenberg-Marquardt (LM) and resilient backpropagation (RBP) were used. The parameters of each different ANN model obtained by using nine input combinations were obtained by trial and error. The success of the models used was evaluated using five different performance metrics. Which of the input combinations used in the streamflow estimation was more successful was decided according to the combination with the highest Nash Sutcliffe efficiency coefficient (NSE) value of the test period. Although similar results were obtained in MLP-GDX, MLP-RBP, MLP-LM and RBNN models, MLP models (except MLP-LM) were slightly more successful than RBNN models. The most successful streamflow estimation model was the MLP-GDX-M6 model. In the MLP-GDX-M6 model, MAE=1.148 m^3/s, RMSE=1.815 m^3/s, R^2=0.724, NSE=0.717, and CA=1.069 were obtained for the testing period. The novelty of the study is that we have examined the credibility of ANN models, including the MLP-GDX, MLP-RBP, MLP-LM and RBNN for predicting the monthly streamflow in natural rivers.

Farklı Yapay Sinir Ağı Modelleri Kullanarak Nehir Akımı Tahmini

Araştırma Makalesi

Makale Tarihçesi:
Geliş tarihi: 16.12.2021
Kabul tarihi: 10.03.2022
Online yayınlanma: 12.12.2022

Anahtar Kelimeler:
Yapay sinir ağları
Seyhan havzası
Nehir akımı

ÖZ

Su kaynaklarının planlanması, tasarımı ve yönetimi, taşkın ve kuraklık yönetim stratejilerinin belirlenmesi ve olumsuz etkilerinin minimize edilebilmesi nedeniyle nehir akımının doğru bir şekilde tahmin edilmesi hayati önem taşımaktadır. Bu çalışmada, Yapay Sinir Ağları (YSA) modellerinin aylık nehir akımı tahmininde kullanılabilirliği araştırılmıştır. Bu amaçla, Türkiye'nin güneyinde Seyhan Havzasında yer alan iki istasyonun aylık verileri kullanılmıştır. Nehir akımı için Sarız Nehri-Şarköy gözlem istasyonu (No: D18A032), yağış için Sarız meteoroloji istasyonu (No:

Su kaynakları

17840) verilerinden faydalanılmıştır. Kullanılan yağış ve akış verileri 1990-2017 periyoduna aittir. Akım ve yağış verilerinin gecikmelerinden oluşan dokuz giriş kombinasyonu elde edilmiş ve YSA modellerinde kullanılmıştır. Aylık nehir akımını tahmin etmek için Çok Katmanlı Algılayıcı (MLP) ve Radyal Temelli Sinir Ağları (RBNN) olmak üzere iki YSA tekniği kullanılmıştır. MLP tekniğinde adaptif öğrenmeli ve momentum özellikli en dik iniş (GDX), esnek geri yayılım (RBP) ve Levenberg-Marquardt (LM) olmak üzere üç adet öğrenme algoritması kullanılmıştır. Farklı giriş kombinasyonları kullanılarak elde edilen her bir farklı YSA modelinin parametreleri deneme yanılma yoluyla belirlenmiştir. Kullanılan modellerin başarısı beş farklı performans ölçütü kullanılarak değerlendirilmiştir. Akarsu tahmininde kullanılan giriş kombinasyonlarından hangisinin daha başarılı olduğuna, test döneminin Nash Sutcliffe verimlilik katsayısı (NSE) değeri en yüksek olan kombinasyona göre karar verilmiştir. MLP-GDX, MLP-RBP, MLP-LM ve RBNN modellerinde benzer sonuçlar elde edilmiş olmasına rağmen MLP modelleri (LM hariç) az da olsa RBNN modellerinden daha başarılı olmuştur. En başarılı akım tahmin modeli MLP-GDX-M6 modeli olmuştur. MLP-GDX-M6 modelinde test periyodu için MAE=1.148 m^3/s, RMSE=1.815 m^3/s, R^2=0.724, NSE=0.717 ve CA=1.069 olarak elde edilmiştir. Çalışmanın yeniliği, doğal nehirlerdeki aylık akış akışını tahmin etmek için MLP-GDX, MLP-RBP, MLP-LM ve RBNN dahil olmak üzere YSA modellerinin güvenilirliğini incelemiş olmamızdır.

To Cite: Koycegiz C., Buyulyildiz M. Estimation of Streamflow Using Different Artificial Neural Network Models. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2022; 5(3): 1141-1154.

1. Introduction

Accurate estimation of streamflow, which is one of the major components controlling the hydrological behavior of basin areas, plays a very important role in making flood warnings, operating reservoirs for flood control, determining the water potential of the river, hydroelectric production in dry periods, distribution of drinking water and irrigation water, and river transportation planning (Mohammadi et al., 2021). Streamflow in a watershed is affected by the physical features of the watershed, such as land use, vegetation, soil types and properties, topography, elevation, size and shape of the basin. In addition to these physical factors, streamflow exhibits a non-linear behavior that is affected by many meteorological factors such as precipitation type, duration of precipitation, intensity of precipitation, distribution of precipitation in the basin, temperature, evapotranspiration, and this complicates its monitoring (Liu et al. 2016).

Physically based models based on the physical process of streamflow formation, which can be revealed through analysis and simulation of hydrological cycles, and data driven models (Latt and Wittenberg, 2014; Cui et al., 2020; Xu et al., 2020) that can apprehend the mathematical relationship (non-linear or linear) between streamflow and its explanatory variables are widely used in flow estimation. Physically based models have the advantage of comprehending the hydrological process as they use the physical properties of the watershed, but require reliable data of the watershed parameters. Due to the limited physical information of most river basins around the world, the inability to comprehend the hydrological behavior of the basin correctly makes it difficult to use physically-based models for flow estimation (Zhang et al., 2015). Data-driven models such as Support Vector Machine (SVM), Artificial Neural Network (ANN) which do not need information about the physical properties of the watershed and are completely based on the characterization of input-output data, are widely used in flow forecasting due to the minimum information requirement, real-time implementation and ease of development (Cui et al., 2020). Adamowski et al. (2012) used Multivariable Adaptive Regression Splines (MARS), wavelet transform-ANN and ANN methods for flow estimation in the Sainji mountain basin where there is not enough data in the Himalayas and compared the results. Hadi and Tombul (2018) used Auto-Regressive (AR), ANN, ANFIS and SVM models to predict streamflow in three basins in Turkey. Consequently, it was obtained that both ANN and ANFIS performed well in streamflow estimation, although ANN outperformed ANFIS for peak values. Liu et al. (2020) used the LSTM network connected Empirical Mode Decomposition (EMD) model for river flow estimation. The performance of the model was evaluated with the Willmott Index (WI) and Legates-McCabe's Index (LMI). The results demonstrated the reliability of this method in flood years and long-term continuous forecasts. Inputs created with monthly flow data yielded close results between forecast and observed values. Latt and Wittenberg (2014) used ANN and multiple linear regression (MLR) methods to estimate of Chindwin River floods using the rainfall and water level data of 1990-2011 periods. In the study by Latifoğlu and Nuralan (2020), monthly river flow data were estimated using Long Short Term Memory (LSTM) networks, which is a Deep Neural Network. The effect of pretreatment applied with Single Spectrum Analysis (SSA) to monthly river flow data on forecast performance was investigated. As a consequence, it was seen that the performance of the SSA-LSTM model was quite good, and the pre-processing of the SSA data significantly increased the model performance. As a result, it has been determined that the SSA-LSTM model can be used as a high-performance tool in river flow estimation studies. ANN, M5 and hybrid wavelet-M5 to model on both daily and monthly scales the rainfall-runoff process at two different basins were used by Nourani et al. (2019). For this purpose, three different data splitting strategies were implemented for the training and testing phases. Firstly, rainfall and runoff time series were decomposed into various sub-time series by applying wavelet transform. The sub-series determined later were used as input to the M5 model. According to the results obtained from the implemented models, the Hybrid Wavelet-M5 model performed better than the original M5 and ANN models.

Xu et al. (2020) used the LSTM network targeting the time series data area for the flow prediction of rivers. The predictions of LSTM are compared with Support Vector Regression (SVR) and Multilayer Perception Models (MLP). In addition, the effect factors of its performance were investigated by carrying out extended experiments on the LSTM model. It was seen that LSTM gave better results in performance results. Cheng et al. (2015) used ANN and SVR models to estimate the monthly flow of the Xinfengjiang Reservoir in China, and found that SVR outperformed ANN, but both models were suitable for the estimation process. Abdullahi et al. (2017) used artificial intelligence (AI) techniques such as ANN, wavelet-ANN (W-ANN), genetic programming (GP) and wavelet-genetic programming (W-GP) to estimate the flow in Iran. For this aim, precipitation data of seventeen meteorological gauge stations for the period 1999-2008 were used. According to the results obtained from the models, the W-ANN model was more performed than the other models. However, it has been determined that the GP model has higher accuracy in estimating peak flow.

In this study, it is aimed to estimate monthly streamflow with two different ANN techniques, Multi-Layer Perceptron (MLP) and Radial Basis Neural Networks (RBNN), using different input combinations created by utilizing the lags of monthly streamflow and monthly precipitation data. The novelty of the study is that we have examined the credibility of ANN models, including the MLP-GDX, MLP-RBP, MLP-LM and RBNN for predicting the monthly streamflow in natural rivers.

2. Material and Methods

2.1. Multi-Layer Perceptron (MLP)

ANN, which was developed for the mathematical modeling of the learning process, inspired by the working system of the human brain, is known as the most powerful and flexible machine learning methods. ANNs are models with many important features such as learning by using the available data, establishing relationships, classifying, generalizing, and working with an unlimited number of variables (Şen, 2004). Multilayer Perceptron (MLP), which is the most common area of use due to its simple structure and used in our study, can be used in the prediction of nonlinear events (Haykin, 2009). MLP can solve estimation and classification problems with the widely used back propagation algorithm. In MLP networks, neurons are organized in layers. In order for MLP networks to be used in time series estimation, the structure of the network must be determined. The process of determining the network structure includes the number of layers of the network, the number of neurons in the layers, the number of iterations, the learning rate, the momentum coefficient, the activation function, and the determination of the normalization method. By changing parameters such as initial weights, the training of the network can be achieved, and the performance of the network can be measured by testing the trained network. The learning rule of the multilayer network is the generalization of the "Delta Learning Rule" based on the least squares method. For this reason, it is also called the "Generalized Delta Rule". More information on MLP is available in Haykin (2009). In this study, gradient descent with momentum and adaptive learning rule backpropagation (GDX), resilient backpropagation (RBP), and Levenberg-Marquardt (LM) are used as the training algorithm in the MLP technique. GDX is a network training function that updates weight and bias values according to gradient descent momentum and adaptive learning rate. The function traingdx combines adaptive learning rate with momentum training. GDX can train any network as long as its weight, net input, and transfer functions have derivative functions. Backpropagation is used to calculate derivatives of performance based on weight and bias variables (URL-1). MLP-RBP is a network training function that updates weight and bias values. The purpose of the MLP-RBP algorithm is to neutralize the negative conditions of the derivatives of the weights in the iterations. It can train any network as long as its weight, net input, and transfer functions have derivative functions. In RBP, which is a successful training algorithm that manages the direct adaptation of the weight step with local slope information, there is a separate update value (Δij) for each weight. The update value determines the size of the weight update (URL-2). MLP-LM is generally the fastest backpropagation algorithm in the toolbox and is a hybrid of Gauss-Newton and steepest descent approaches to achieve optimal results. This training algorithm generally shows unlimited variations of the correction vector Δp in the inversion of nonlinear problems. Although LM requires more memory than other algorithms, it is highly recommended as a first choice supervised algorithm (URL-3). Detailed information about the GDX, RBP and LM training algorithms are available in literature (URL-1; URL-2; URL-3; Tezel and Buyukyildiz 2016).

2.2. Radial Basis Neural Network (RBNN)

RBNN is an artificial neural network model based on local action and response behaviours seen in neurons in the human nervous system (Broomhead and Lowe, 1988). The training performance of the RBNN model turns into a problem of finding the most suitable surface for the data in the output vector space and thus an interpolation problem. Similar to the general ANN architecture, RBFN models are defined in 3 layers: input layer, hidden layer and output layer. In the RBNN model, radial basis activation functions and nonlinear clustering analysis are used in the transition from the input layer to the hidden layer. There is no parameter learning in RBNN as in MLP and linear adjustment of weights is made for radial bases. This feature provides the advantage of a very fast convergence time without local minimums. Detailed information about the RBNN model is available in Haykin (2009).

2.3. Description of Data

Seyhan Basin, located in the southern part of Turkey, is located in the north of Adana Province in the Eastern Mediterranean Region of Turkey, between 36º 30' and 39º 15' north latitudes and 34º 45' and 37º 00' east longitudes. The Seyhan Basin, with an area of 22035 km^2, extends to the Ceyhan Basin in the east, Konya and the Eastern Mediterranean Basins in the west, Develi Basin and Kulmaç Mountains in the north, and the Mediterranean Sea in the south. Seyhan Basin has a frequent river network. The Seyhan River is formed by the merging of the Zamantı River and the Göksu River. In this study, monthly average streamflow data of Sarız River-Şarköy Station (No: D18A032) and monthly total precipitation data of Sarız Meteorological Station (No: 17840) on the Seyhan Basin were used. The data used belong to the period 1990-2017. The precipitation area of station D18A032 is 752.40 km^2 and is located at an altitude of 1400 m and at 36°19' E - 38°19' N. Sarız meteorological station is located at 36°29' E - 38°29' N and is altitude 1500 m. The location of the used streamflow and precipitation stations in the Seyhan Basin is given in Figure 1.

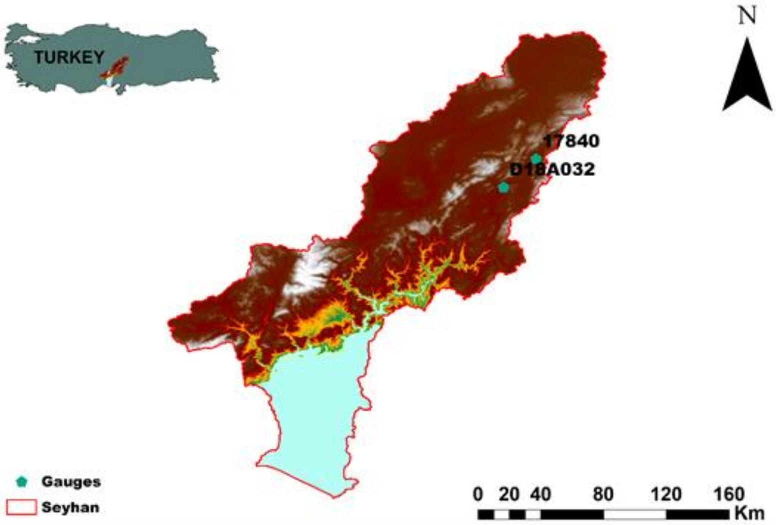


Figure 1. Location of the Seyhan Basin in Turkey, showing the location of the precipitation gauge station (17840) and of the streamflow gauge station (D18A032) in Seyhan Basin.

The monthly data of streamflow and precipitation of both stations belong to the duration of 1990–2017 (336 months). Before models' implementation, the data were divided into two phases: approximately 70% (228 months) of the datasets was used for model development (training) phase while the rest 30% (108 months) of the datasets was divided for model evaluation (testing) purposes. The statistical parameters of the streamflow and precipitation data used in the ANN models for training, testing and the whole period are given in Table 1. Figure 2 shows the time series of runoff and precipitation data used in this study for the period 1990-2017.

Table 1. Descriptive statistics for monthly mean streamflow and monthly total precipitation

Parameter	Data set	Minimum	Maximum	Mean	Standard Deviation	Skewness
Streamflow (m^3/s)	Training	0.932	18.700	3.960	2.140	1.933
	Testing	0.677	18.600	3.676	3.411	2.219
	All	0.677	18.700	4.019	3.437	1.997
Precipitation (mm)	Training	0	21.350	4.181	2.788	1.687
	Testing	0	16.700	3.984	2.639	1.461
	All	0	21.350	4.117	2.730	1.618

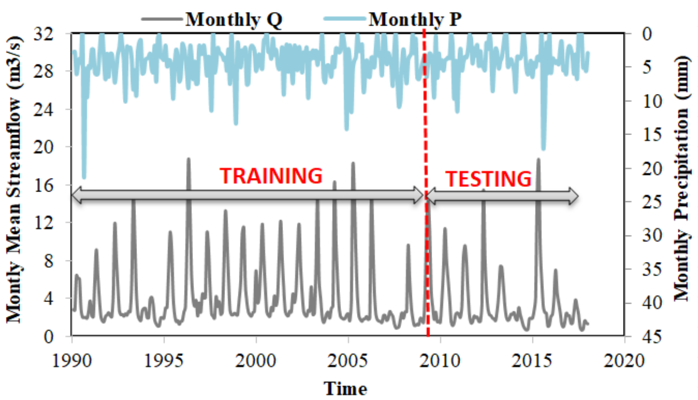


Figure 2. The time series of streamflow and precipitation during 1990-2017 periods

2.4. Model Performance Metrics

The accuracy of the implemented ANN models' was interpreted using Nash Sutcliffe Efficiency Coefficient (NSE), coefficient of determination (R^2), combined accuracy (CA), root mean square error (RMSE) and mean absolute error (MAE) performance metrics. The equations of the performance metrics used are given below.

$$MAE = \frac{\sum_{i=1}^{N} \left| Qi_o - Qi_e \right|}{N} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Qi_o - Qi_e)^2}{N}} \tag{2}$$

$$R^2 = \frac{\left[\sum_{i=1}^{N} (Qi_o - \overline{Q}_o)(Qi_e - \overline{Q}_e) \right]^2}{\sum_{i=1}^{N} (Qi_o - \overline{Q}_o)^2 \sum_{i=1}^{N} (Qi_e - \overline{Q}_e)^2} \tag{3}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N} (Qi_o - Qi_e)^2}{\sum_{i=1}^{N} (Qi_o - \overline{Q}_o)^2} \tag{4}$$

$$CA = 0.33\,(RMSE + MAE + (1 - R^2)) \tag{5}$$

where Q_o and Q_e, are the observed and estimated value of the flow, $\overline{Q}_o$ and $\overline{Q}_e$, are the average of observed and estimated flow data.

3. Results and Discussion

In this study, the usability of two ANN methods, MLP and RBNN, was assessed in estimation of monthly streamflow of Sarız River-Şarköy station using hydro-meteorological inputs. ANN models were created using streamflow and precipitation lags. Nine input combinations were selected based on current time and antecedent precipitation and streamflow values. The input combinations used are given in Table 2. Let us assume that Q_t / P_t represents the streamflow/precipitation at current time (t), in this situation Q_{t-2} / P_{t-2} denotes the streamflow/precipitation two month prior to time t.

Before implementation the ANN models to estimate the monthly streamflow, the streamflow and precipitation data were normalized between 0 and 1 using Equation 6.

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{6}$$

where X_{norm}, X_i, X_{min} and X_{max} represent normalized, observed, minimum and maximum data values, respectively.

Table 2. Input combinations used in models

Model Names	Input	Output
M1	Q_{t-1}	
M2	Q_{t-1}, Q_{t-2}	
M3	Q_{t-1}, Q_{t-2}, Q_{t-3}	
M4	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}	
M5	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, P_t, P_{t-1}, P_{t-2}, P_{t-3}, P_{t-4}	Q_t
M6	Q_{t-1}, Q_{t-2}, Q_{t-3}, P_t, P_{t-1}, P_{t-2}, P_{t-3}	
M7	Q_{t-1}, Q_{t-2}, P_t, P_{t-1}, P_{t-2}	
M8	Q_{t-1}, P_t, P_{t-1}	
M9	Q_{t-1}, P_t	

Two different ANN techniques, MLP and RBNN, were used to estimate the monthly average streamflow. Two hidden layers are used in the ANN structures trained with the GDX, RBP and LM training algorithms. Tangent sigmoid activation functions are used in the hidden layers and logarithmic sigmoid activation functions are used in the output layer. Momentum coefficient and learning rate were determined in 0.1 increments between 0.1 and 1 in MLP-GDX models. In the application of both MLP and RBNN models, the number of neurons in the hidden layers was identified in increments of 1 between 1 and 10, and the number of iterations was taken as 1000. In RBNN models, the spread number was obtained in increments of 0.01 between 0.01 and 2. As a result of all these assumptions, the most successful input combination in the streamflow estimation was decided according to the maximum NSE value of the testing period.

The model parameters of the most successful network structures obtained for each input combination in the MLP-GDX, MLP-RBP, MLP-LM and RBNN models are given in Table 3. In Table 3; n and l represent the number of neurons in the first and second hidden layers, while lr and mc represent the learning rate and momentum coefficient, respectively. In the RBNN model, n and σ represent the number of neurons in the hidden layer and the spread number, respectively.

Table 3. Most successful network structures according to NSE

Input Combination	MLP-GDX				MLP-RBP		MLP-LM		RBNN	
	n	**l**	**lr**	**mc**	**n**	**l**	**n**	**l**	**n**	**σ**
M1	1	6	0.9	0.1	2	1	1	2	3	0.58
M2	3	1	0.6	0.9	2	10	2	1	6	0.43
M3	3	2	0.1	0.2	2	4	2	1	9	1.15
M4	7	4	0.2	0.8	2	4	2	1	9	1.08
M5	5	3	0.8	0.1	1	2	1	2	9	0.55
M6	6	3	0.1	0.5	2	6	1	1	10	0.63
M7	10	7	1	0.7	2	2	1	3	9	0.51
M8	4	10	0.6	0.8	4	1	4	1	8	0.20
M9	4	6	0.5	0.2	4	4	2	9	6	0.27

The training and testing statistics of the MLP-GDX, MLP-RBP, MLP-LM and RBNN models are given in Table 4 for the Sarız River-Şarköy station.

Table 4. Comparison of statistical errors for MLP-GDX, MLP-RBP, MLP-LM and RBNN models

Model Names		Performance Metric	M1	M2	M3	M4	M5	M6	M7	M8	M9
MLP-GDX	TRAINING	**MAE** (m^3/s)	**1.683**	1.416	1.371	1.295	1.254	1.354	1.241	1.486	1.512
		RMSE (m^3/s)	**2.644**	2.282	2.271	2.196	1.996	2.093	1.997	2.314	2.400
		$\mathbf{R^2}$	**0.416**	0.565	0.569	0.597	0.668	0.635	0.667	0.553	0.519
		NSE	**0.416**	0.565	0.569	0.597	0.667	0.634	0.667	0.553	0.519
		CA (m^3/s)	**1.621**	1.364	1.344	1.285	1.182	1.258	1.178	1.402	1.450
	TESTING	**MAE** (m^3/s)	**1.508**	1.187	1.127	1.163	1.183	**1.148**	1.108	1.245	1.256
		RMSE (m^3/s)	**2.348**	1.881	1.873	1.888	1.950	**1.815**	1.846	2.118	2.093
		$\mathbf{R^2}$	**0.532**	0.710	0.711	0.703	0.689	**0.724**	0.719	0.628	0.631
		NSE	**0.526**	0.696	0.699	0.694	0.673	**0.717**	0.707	0.614	0.624
		CA (m^3/s)	**1.427**	1.108	1.085	1.105	1.137	**1.069**	1.067	1.233	1.227
MLP-RBP	TRAINING	**MAE** (m^3/s)	**1.701**	1.350	1.338	1.404	1.490	1.246	1.432	1.587	1.465
		RMSE (m^3/s)	**2.668**	2.195	2.247	2.277	2.308	1.962	2.189	2.316	2.262
		$\mathbf{R^2}$	**0.405**	0.598	0.578	0.567	0.555	0.679	0.600	0.553	0.573
		NSE	**0.405**	0.598	0.578	0.567	0.555	0.678	0.600	0.552	0.573
		CA (m^3/s)	**1.638**	1.303	1.322	1.358	1.400	1.165	1.327	1.436	1.371
	TESTING	**MAE** (m^3/s)	**1.528**	1.184	1.197	1.209	1.343	1.315	**1.191**	1.327	1.278
		RMSE (m^3/s)	**2.373**	1.925	1.897	1.925	2.075	2.017	**1.880**	2.192	2.072
		$\mathbf{R^2}$	**0.520**	0.691	0.700	0.694	0.645	0.661	**0.712**	0.589	0.634
		NSE	**0.516**	0.682	0.691	0.682	0.630	0.650	**0.696**	0.587	0.631
		CA (m^3/s)	**1.446**	1.128	1.120	1.135	1.245	1.211	**1.109**	1.297	1.226
MLP-LM	TRAINING	**MAE** (m^3/s)	**1.724**	1.330	1.307	1.353	1.479	1.512	1.483	1.416	1.041
		RMSE (m^3/s)	**2.689**	2.244	2.235	2.252	2.302	2.343	2.339	2.255	1.660
		$\mathbf{R^2}$	**0.396**	0.579	0.583	0.577	0.557	0.541	0.543	0.575	0.770
		NSE	**0.396**	0.579	0.583	0.577	0.557	0.541	0.543	0.575	0.770
		CA (m^3/s)	**1.656**	1.318	1.307	1.329	1.394	1.423	1.412	1.352	0.967
	TESTING	**MAE** (m^3/s)	**1.548**	1.200	**1.179**	1.277	1.345	1.370	1.318	1.420	1.320
		RMSE (m^3/s)	**2.416**	1.946	**1.933**	2.007	2.088	2.063	2.025	2.210	2.162
		$\mathbf{R^2}$	**0.507**	0.686	**0.689**	0.689	0.639	0.651	0.662	0.590	0.599
		NSE	**0.498**	0.674	**0.679**	0.654	0.625	0.634	0.647	0.580	0.598
		CA (m^3/s)	**1.471**	1.142	**1.136**	1.186	1.252	1.248	1.215	1.333	1.282
RBNN	TRAINING	**MAE** (m^3/s)	**1.719**	1.409	1.454	1.435	1.401	1.418	1.300	1.550	1.587
		RMSE (m^3/s)	**2.698**	2.261	2.300	2.292	2.127	2.201	2.121	2.477	2.522
		$\mathbf{R^2}$	**0.392**	0.573	0.558	0.561	0.622	0.595	0.624	0.488	0.469
		NSE	**0.392**	0.573	0.558	0.561	0.622	0.595	0.624	0.488	0.469
		CA (m^3/s)	**1.658**	1.352	1.385	1.375	1.289	1.328	1.253	1.498	1.531
	TESTING	**MAE** (m^3/s)	**1.516**	**1.165**	1.184	1.192	1.248	1.181	1.170	1.322	1.317
		RMSE (m^3/s)	**2.355**	**1.909**	1.923	1.926	2.038	1.933	1.914	2.142	2.180
		$\mathbf{R^2}$	**0.528**	**0.697**	0.691	0.688	0.646	0.684	0.691	0.609	0.605
		NSE	**0.523**	**0.687**	0.682	0.681	0.643	0.679	0.685	0.606	0.592
		CA (m^3/s)	**1.433**	**1.114**	1.128	1.132	1.201	1.132	1.120	1.272	1.284

According to the values given in Table 4, the lowest successful input combination in MLP-GDX, MLP-RBP, MLP-LM and RBNN models was the M1 model, in which Qt-1 data were used in both the training and test periods. Higher MAE, RMSE and CA, lower R^2 and NSE values were obtained in M1 input combination compared to other input combinations.

In MLP-GDX models for testing period, the MAE (=1.108 m^3/s) and CA (=1.067 m^3/s) values of the M7 input combination are lower than the MAE and CA values of all input combinations. However, the input combination with the lowest RMSE (=1.815 m^3/s), the highest R^2 (=0.724) and NSE (=0.717) values in MLP-GDX technique was obtained as M6. For this reason, the most successful input combination in MLP-GDX technique was accepted as the M6 model, in which the parameters Q_{t-1}, Q_{t-2}, Q_{t-3}, P_t, P_{t-1}, P_{t-2}, P_{t-3} were used. In the MLP-RBP models for testing period, the most successful input combination was the M7 model, in which the Q_{t-1}, Q_{t-2}, P_t, P_{t-1}, P_{t-2} parameters were used. The MLP-RBP-M7 model has lower MAE (=1.191 m^3/s), RMSE (=1.880 m^3/s), CA (=1.109 m^3/s) values and higher R^2 (=0.712) and NSE (=0.696) values than the other MLP-RBP models.

In the MLP-LM models for testing period, the most successful input combination was the M3 model, in which the Q_{t-1}, Q_{t-2}, Q_{t-3} parameters were used. The MLP-LM-M3 model has lower MAE (=1.179 m^3/s), RMSE (=1.933 m^3/s), CA (=1.136 m^3/s) values and higher R^2 (=0.689) and NSE (=0.679) values than the other MLP-RBP models.

In the RBNN models for testing period, the most successful input combination was the M2 model, in which the Q_{t-1} and Q_{t-2} parameters were used. The RBNN-M2 model has lower MAE (=1.165 m^3/s), RMSE (=1.909 m^3/s), CA (=1.114 m^3/s) values and higher R^2 (=0.697) and NSE (=0.687) values than the other RBNN models. On the other hand, according to the performance criteria, the results of the M7 input combination in RBNN models are very similar to the results of the M2 input combination.

When comparing the MLP-GDX, MLP-RBP, MLP-LM and RBNN models, the MLP-GDX models with the lowest MAE, RMSE, CA and highest R^2, NSE values outperformed the RBNN, MLP-LM and MLP-RBP models for flow prediction at all input combinations.

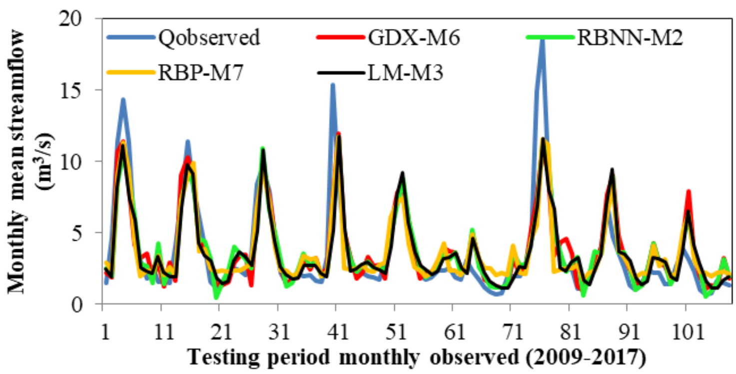
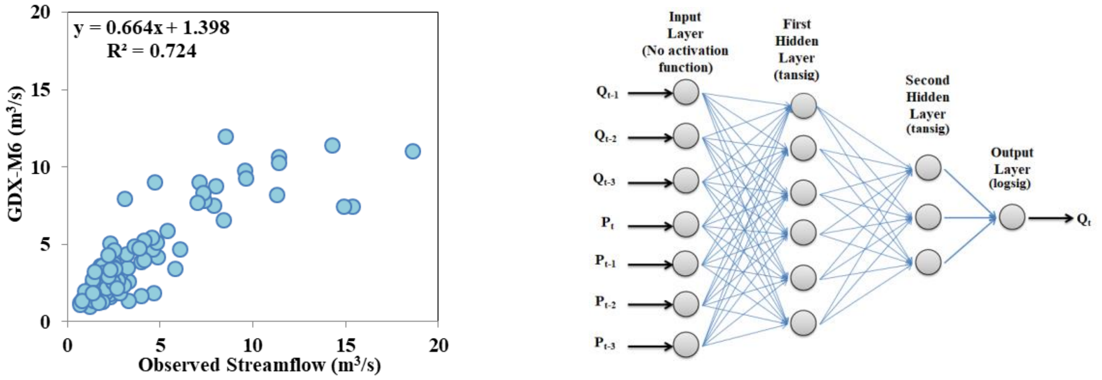
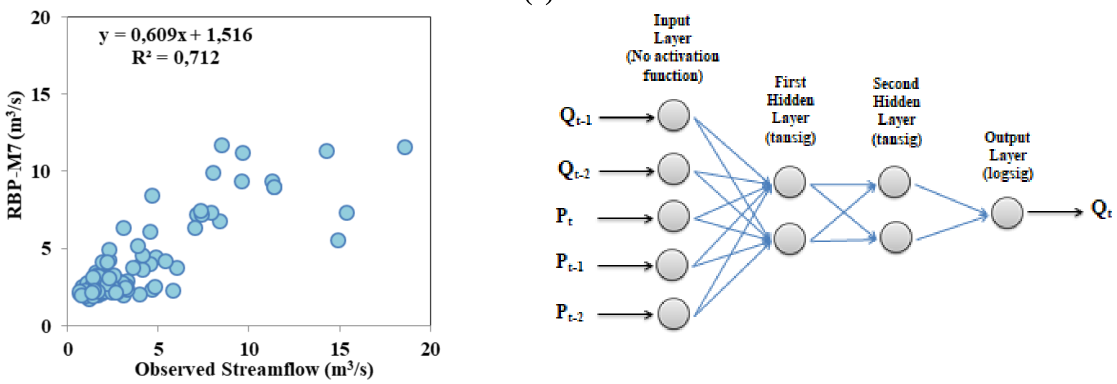


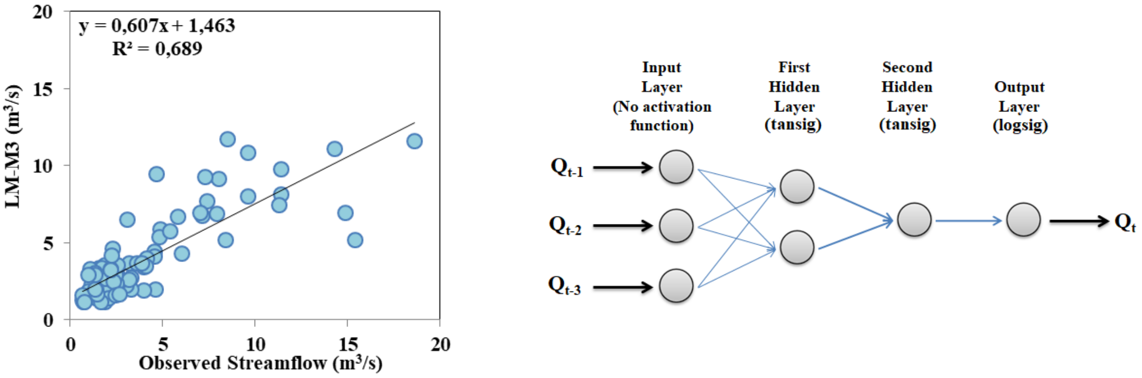
Figure 3. Optimal models and observed monthly mean streamflow for the testing period (2009–2017) using GDX-M6, RBP-M7, LM-M3 and RBNN-M2



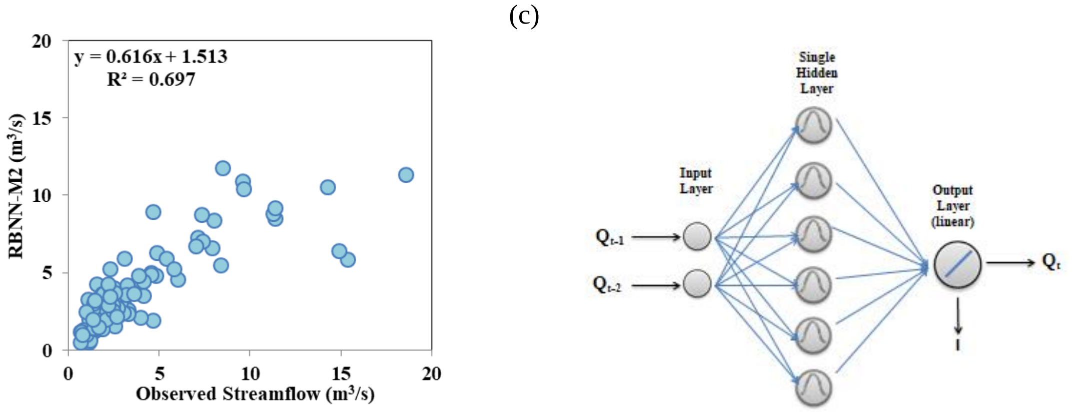
(a)



(b)



(c)



(d)

Figure 4. ANN architecture of the selected models, and scatter-diagrams of observed and estimated monthly mean streamflow GDX-M6, b) RBP-M7, c) LM-M3 and d) RBNN-M2 model

The time series for the testing period of the MLP-GDX-M6, MLP-RBP-M7, MLP-LM-M3 and RBNN-M2 models, in which the most successful results were obtained in the streamflow estimation, are shown in Figure 3. ANN architecture of the selected models and the scatter diagrams are shown in Figure 4.

4. Conclusions

In this study, monthly streamflow estimation was made using ANN-based modeling approach. Nine different input combinations consisting of the lags of precipitation and streamflow data of the study area were used in the ANN models. MLP and RBNN models were used for streamflow estimation. In the MLP technique, models were created with GDX, RBP and LM training algorithms. While MLP models showed the highest success in almost every input combination, LM models were the model that showed the lowest prediction success. All four models used were also found to significantly overestimate/underestimate low/high streamflow values in some times of the test period.

There are some limitations that affect the success of the used models in this study. These limitations include the small size of the training and test data, the input variables used, the structure of the models used, and the selection of model parameters. Although the results obtained from the ANN models used in this study are promising, it may be possible to achieve higher streamflow prediction success with different applications. Because, obtaining streamflow forecasting models with high forecasting success will contribute to missing data completion, flood modeling studies, and modeling of other hydrological variables. Streamflow data is under the influence of many meteorological parameters such as temperature, evaporation, snowmelt, humidity as well as precipitation and has a stochastic and non-linear structure. Therefore, the performance of streamflow prediction models can be improved by using more meteorological variables. In addition, using decomposition techniques such as wavelet and empirical mode decomposition, weakening the non-stationary and non-linearity of the streamflow data, using longer-term data, and improving the convergence rate by using more robust algorithms can increase the success of the models. These mentioned points will shed light on future streamflow estimation and similar hydrological studies.

Statement of Conflict of Interest

Authors have declared no conflict of interest.

Author's Contributions

The contribution of the authors is equal.

References

Abdollahi S., Raeisi J., Khalilianpour M., Ahmadi F., Kisi O. Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques. Water Resources Management 2017; 31: 4855–4874.

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

Broomhead D., Lowe D. Multivariable functional interpolation and adaptive networks. Complex Systems 1988; 2: 321-355.

Cheng CT., Feng ZK., Niu WJ., Liao SL. Heuristic methods for reservoir monthly inflow forecasting: A case study of Xinfengjiang Reservoir in Pearl River, China. Water 2015; 7: 4477-4495.

Cui F., Salih SQ., Choubin B., Bhagat SK., Samui P., Yaseen ZM. Newly explored machine learning model for river flow time series forecasting at Mary River, Australia. Environmental Monitoring and Assessment 2020; 192: 761.

Hadi SJ., Tombul M. Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination. Journal of Hydrology 2018; 516: 674–687.

Haykin S. Neural networks and learning machines. Pearson Education Inc., Upper Saddle River, New Jersey, USA, 2009.

Latifoğlu L., Nuralan KB. Tekil spektrum analizi ve uzun-kısa süreli bellek ağları ile nehir akım tahmini. Avrupa Bilim ve Teknoloji Dergisi 2020; 376-381.

Latt ZZ., Wittenberg H. Improving flood forecasting in a developing country: a comparative study of stepwise multiple linear regression and artificial neural network. Water Resources Management 2014; 28(8): 2109-2128.

Liu D., Jiang W., Mu L., Wang S. Streamflow prediction using deep learning neural network: Case study of Yangtze River. IEEE Access, 2020; 8: 90069-90086.

Liu Y., Sang YF., Li X., Hu J., Liang K. Long-term streamflow forecasting based on relevance vector machine model. Water 2016; 9(1): 9.

Mohammadi B., Moazenzadeh R., Christian K., Duan Z. Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models. Environmental Science and Pollution Research 2021; 28, 65752–65768.

Nourani V., Davanlou Tajbakhsh A., Molajou A., Gokcekus H. Hybrid wavelet-M5 model tree for rainfall-runoff modeling. Journal of Hydrologic Engineering 2019; 24(5): 04019012.

Şen Z. Yapay sinir ağları ilkeleri. Su Vakfı Yayınları,183 p., İstanbul.2004.

Tezel G., Buyukyildiz M. Monthly evaporation forecasting using artifcial neural networks and support vector machines. Teor. Appl. Climatol. 2016; 124: 69-80.

URL-1, https://www.mathworks.com/help/deeplearning/ref/traingdx.html (17.01.2022)

URL-2, https://www.mathworks.com/help/deeplearning/ref/trainrp.html (17.01.2022)

URL-3, https://www.mathworks.com/help/deeplearning/ref/trainlm.html (17.01.2022)

Xu W., Jiang Y., Zhang X., Li Y., Zhang R., Fu G. Using long short-term memory networks for river flow prediction. Hydrology Research 2020; 51(6): 1358-1376.

Zhang X., Peng Y., Zhang C., Wang B. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. Journal of Hydrology 2015; 530: 137-152.