Water Resour Manage (2016) 30:4125–4151
DOI 10.1007/s11269-016-1408-5

Enhancing Long-Term Streamflow Forecasting and Predicting using Periodicity Data Component: Application of Artificial Intelligence

Zaher Mundher Yaseen[1] · **Ozgur Kisi**[2] · **Vahdettin Demir**[2]

Received: 10 April 2016 / Accepted: 13 June 2016 /
Published online: 6 July 2016

Abstract Streamflow forecasting and predicting are significant concern for several applications of water resources and management including flood management, determination of river water potentials, environmental flow analysis, and agriculture and hydro-power generation. Forecasting and predicting of monthly streamflows are investigated by using three heuristic regression techniques, least square support vector regression (LSSVR), multivariate adaptive regression splines (MARS) and M5 Model Tree (M5-Tree). Data from four different stations, Besiri and Malabadi located in Turkey, Hit and Baghdad located in Iraq, are used in the analysis. Cross validation method is employed in the applications. In the first stage of the study, the heuristic regression models are compared with each other and multiple linear regression (MLR) in forecasting one month ahead streamflow of each station, individually. In the second stage, the models are evaluated and compared in predicting streamflow of one station using data of nearby station. The research investigated also the influence of the periodicity component (month number of the year) as an external sub-set in modeling long-term streamflow. In both stages, the comparison results indicate that the LSSVR model generally performs superior to the MARS, M5-Tree and MLR models. In addition, it is seen that adding periodicity as input to the models significantly increase their accuracy in forecasting and predicting monthly streamflows in both stages of the study.

Keywords Streamflow forecasting and prediction · Periodicity component · LSSVR · MARS · M5-Tree · MLR

✉ Zaher Mundher Yaseen
zahermundher@gmail.com

[1] Civil and Structural Engineering Department, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

[2] Department of Civil Engineering, Faculty of Architecture and Engineering, Canik Basari University, 55080 Samsun, Turkey

1 Introduction

Understanding the complicated phenomena of streamflow plays a significant part in water resources management. More specifically, long-term streamflow forecasting (e.g., monthly river flow) is greatly crucial for hydro-power generation, appropriate reservoir operation, effective irrigation management decision and several other hydrological applications. Over the past couple decades, streamflow modeling has received a massive attention by hundreds of researchers. This is due to the fact that, the global climate changes have been influenced the hydrologic cycle that caused numerous of flood and drought events. According to the literature, river flow forecasting has been undertaken based on two main methodologies, physical based models and conceptual based models "e.g., data-driven techniques". Physical models usually required more effort and various hydrological variables to simulate the elemental physical processes of the watershed (Costabile et al. 2012). Whereas, data-driven soft computing approaches have shown the capability to capture the non-linearity relationship between the predictors and predicted without advance knowledge with less inputs hydrological parameters (Ahmed and Sarma 2007; Afan et al. 2014; Singh and Cui 2015; Tigkas et al. 2016).

Classically, black box time series models have been applied for streamflow forecasting since 1970 by (Box and Jenkins). Based on the review researches, those parametric linear models such as Moving Average (MA), Auto Regressive Integrated Moving Average (ARIMA), and Multiple Linear Regression (MLR) have been used in almost all the hydrological variables (Abrahart and See 2000; Maier and Dandy 2000; Abrahart et al. 2010; Abrahart et al. 2012; Yaseen et al. 2015). However, they perform poorly in the conditions of highly non-stationary and non-linear real problems. Since 1990, artificial intelligence methods have been extensively utilized in a wide range of hydrological applications and more specifically for streamflow forecasting, such as artificial neural network (ANN), support vector machine (SVM), adaptive neuro fuzzy inference system (ANFIS), genetic algorithm (GA), and gene expression programming (GEP) (Nourani et al. 2014; Yaseen et al. 2015).

Most recently, three data driven approaches have been gained a remarkable emerging and potential in handling the complex nonlinear problems such as least square support vector regression (LSSVR), multivariate adaptive regression splines (MARS) and M5 Model Tree. Those forgoing approaches have been broadly used in solving hydrologic problems. LSSVR is the modified version of support vector repression (SVR) that can exclude the quadratic programming problems (Suykens and Vandewalle 1999). In addition, it avoids several shortcomings of other data-driven learning processes (e.g., local minima, time consumption and over-fitting) (Ji et al. 2014). LSSVR has received a positive successful application in the engineering field; for instance, bearing raceway prediction (Tao et al. 2008), prediction of effluent parameter of wastewater treatment plant (Huang et al. 2009), airframe wing-box estimation (Deng and Yeh 2010), power system stabilization (Pahasa and Ngamroo 2011), prediction of CO_2 in reservoir (Shokrollahi et al. 2013), oil recovery and economic analysis (Kamari et al. 2014), and oil reservoir viscosity determination (Hemmati-Sarapardeh et al. 2014). In the hydrological context, there are a few studies have been conducted using LSSVR; for example, evapotranspiration prediction (Guo et al. 2011; Kisi 2013), daily water demand estimation (Hwang et al. 2012), sediment transport modeling (Kisi 2012), reservoir inflow modeling (Okkan and Ali Serbes 2013), and water pollution prediction (Kisi and Parmar 2016), authors concluded the outperformance of the LSSVR over the other data-driven used in their researches and recommended its applicability for other hydrological variables.

Multivariate adaptive regression splines is a relatively modern artificial intelligence approach that firstly proposed by (Friedman 1991). The main advantages of this method are the capacity to capture the natural complication of the data mapping in high-dimensional data patterns, quick and flexible model, and perform the forecasting of continuous and binary output variables accurately. In addition, this nonparametric statistical method is a flexible procedure that organize the relationship between the inputs and output variables with less including variable interactions (Leathwick et al. 2006). Previous studies of the MARS algorithm in water resources application include rainfall and temperature forecasting, sediment concentration estimation, water pollution prediction, freshwater distribution system modeling, and drought events river flow simulation (Sarangi and Bhattacharya 2005; Leathwick et al. 2006; Sotomayor 2010; Adamowski et al. 2012; Shortridge et al. 2015). Thus, in the current research, the best knowledge of the authors is to introduce the multivariate adaptive regression splines approach for forecasting and predicting monthly streamflow.

Another new data-driven technique is M5 Model tree. M5 model tree is a data mining approach that splits the data time series into subspace using divide-and-conquer method, which makes it possible to divide the multi-dimensional parameter space and generate the model automatically based on the overall quality criterion (Quinlan 1992). Recently, scholars researched the utility of the M5 model tree in different hydrological applications such as water level optimization (Bhattacharya and Solomatine 2005), precipitation-river flow modeling (Solomatine and Dulal 2003), evapotranspiration prediction (Pal and Deswal 2009), flood events forecasting (Solomatine and Xue 2004), and sedimentation estimation (Sarangi and Bhattacharya 2005). Those are a few studies effectively accomplished in the water resources sector using M5 model tree.

For the best knowledge of the authors, the major objectives of the current research are (i) investigate three different modern heuristic regression approaches (i.e., LSSVR, MARS and M5 model tree) for modeling long-term streamflow, (ii) compare their performance with one classical method such as MLR, (iii) in order to demonstrate the effectiveness, four rivers placed in two different region namely, Batman and Garzan Rivers located in Turkey, Euphrates and Tigris Rivers located in Iraq, have been used to perform the proposed models. In the first phase of the study, streamflow forecasting is demonstrated based on the same river flow data for the same river. Whereas the second phase, streamflow prediction is conducted for specific stream based on the nearby stream. Furthermore, the influence of periodicity on the forecasting and predicting performance was examined.

2 Theoretical Overview

2.1 Least Square Support Vector Regression

LSSVR is the extended version of support vector regression (SVR) model, modified by (Suykens and Vandewalle 1999). Based on the literature, the major drawback of SVR is time consumption that overcame by the improved version of LSSVR via excluding the quadratic programming problem. This enhancement would avoid several limitations (e.g., the local minima, the over-fitting problem). In addition, it may produce a stable solution to crack the quadratic programming problems (Xie et al. 2013; Ji et al. 2014). Statistically, the main

principle knowledge of LSSVR is to accomplish the optimum mapping function between the inputs x and the output y. This process is conducted through non-linear relationship function with high-dimensional feature space. To attain the optimal solution, regression model into the high-dimensional feature space was developed to capture the non-linear regression function. Regression function can be formulated as follows:

$$y(x) = w^{T}\varphi(x) + b \tag{1}$$

where y is the obtained value in terms of x, w is the coefficient vector, φ is the mapping function, b is the bias term achieved by the minimizing the upper bound of the generalization error. According to the standard of minimizing the regularized risk, the regression function of LSSVR (Suykens and Vandewalle 1999) can be well-defined as:

$$min\frac{1}{2}w^{T}w + \frac{1}{2}\gamma\sum_{i=1}^{l}\left(\xi^{2}\right) \tag{2}$$

That subject to the following constraints

$$y = w^{T}\varphi(x_i) + b + \xi_i (i = 1, 2, \ldots, l) \tag{3}$$

Where γ is the regularization parameter which is control the minimization of the forecasting or prediction error and the function smoothness, while ξ is the training error for the inputs (x_i).

At this point, Lagrange Multiplier is utilized to derive solution for w and ξ using formula (2). The objective function obtained by changing the constraint problem into an unconstraint problem. The Lagrange function L written as follows:

$$L(w, b, \xi, \propto) = J(w, \xi) - \sum_{i=1}^{l} a_i\left\{w^{T}\varphi(x_i) + b + \xi_i - y_i\right\} \tag{4}$$

where a_i presents Lagrange Multipliers.

The Lagrangian theorem and Karush-Kuhn-Tucker (KKT) condition permit (Fletcher 1987) to achieve the following function:

$$y(x) = \sum_{i=1}^{l} a_i \; K(x, x_i) + b \tag{5}$$

$K(x)$ denotes the kernel function that satisfies Mercer's conditions; $K(x, x_i) = (\varphi(x) \,.\, \varphi(x_i))$ that eliminate vector dot product operation in some feature space.

In the current research, radial basis function (kernel function) was used to in the regression solution. The formula can be defined as:

$$K(x, x_i) = e^{\frac{\|x - x_i\|^2}{2\sigma^2}} \tag{6}$$

There are two parameters used for tuning LSSVR model, which are γ and σ^2 (Cao et al. 2008). The current state-of-the-art of the authors is the utilization of LSSVR for streamflow forecasting and prediction. This is relying on the robustness of LSSVR model against the chaotic disturbances, complex non-linear and randomness problems. Furthermore, it's utility to reduce the soft computing efforts comparatively to the classical approaches.

2.2 Multivariate Adaptive Regression Splines

MARS is a nonparametric regression model that was initially proposed by (Friedman 1991), which is utilized to forecast continuous numeric outcomes. The main feature of MARS algorithm is the forward and backward stepwise procedure that can controls and explains the complex nonlinear mapping between the inputs and output variables. The advantage of the backward stepwise procedure is to remove the unnecessary input candidates from the previous selected data set in order to enhance the forecasting accuracy. This function forecasts the new output Y according to the input variable X using either of the two basis functions, using a knot or value of variable that defines the inflection point along the inputs range (Sharda et al. 2006):

$$Y = max(0, X - c) \tag{7}$$

$$Y = max(0, c - X) \tag{8}$$

where the c parameter indicates the threshold value. There are two adjacent splines intersect at a knot, in order to maintain the continuity of the basis functions. The function is used in the forward and backward stepwise procedure to each input parameter is to identify the precise location of knots where the function value changes. Great to mention, MARS model is a data-driven process that gained popularity in time series analysis, most recently. In addition, it is even better to explore its capability to enhance river flow forecast models. Authors recommend the following references for the reader to refer for more comprehensive details of MARS model (Friedman 1991; Sharda et al. 2008; Zhang and Goh 2014).

2.3 M5 Model Tree

The complex time series problems can be comprehended by splitting the time space into a number of sub time space and build each category individually using linear regression model. M5 model tree algorithm is one of the new data mining method that divide the data space into smaller sub-spaces using divide and conquer procedure (Quinlan 1992). The fundamental concept of this model is the binary decision tree. The partition procedure follows the idea of a decision tree that has a regression

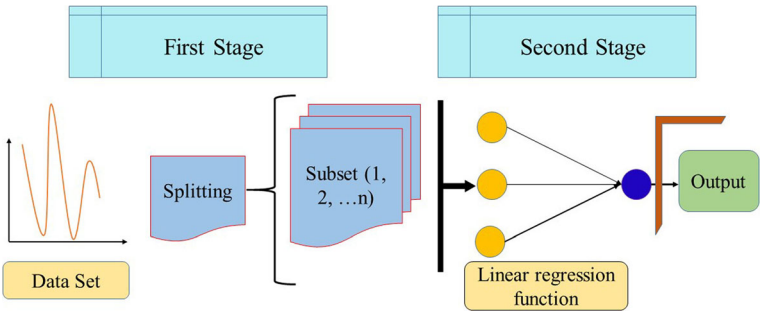

Fig. 1 The two stages of M5 model tree

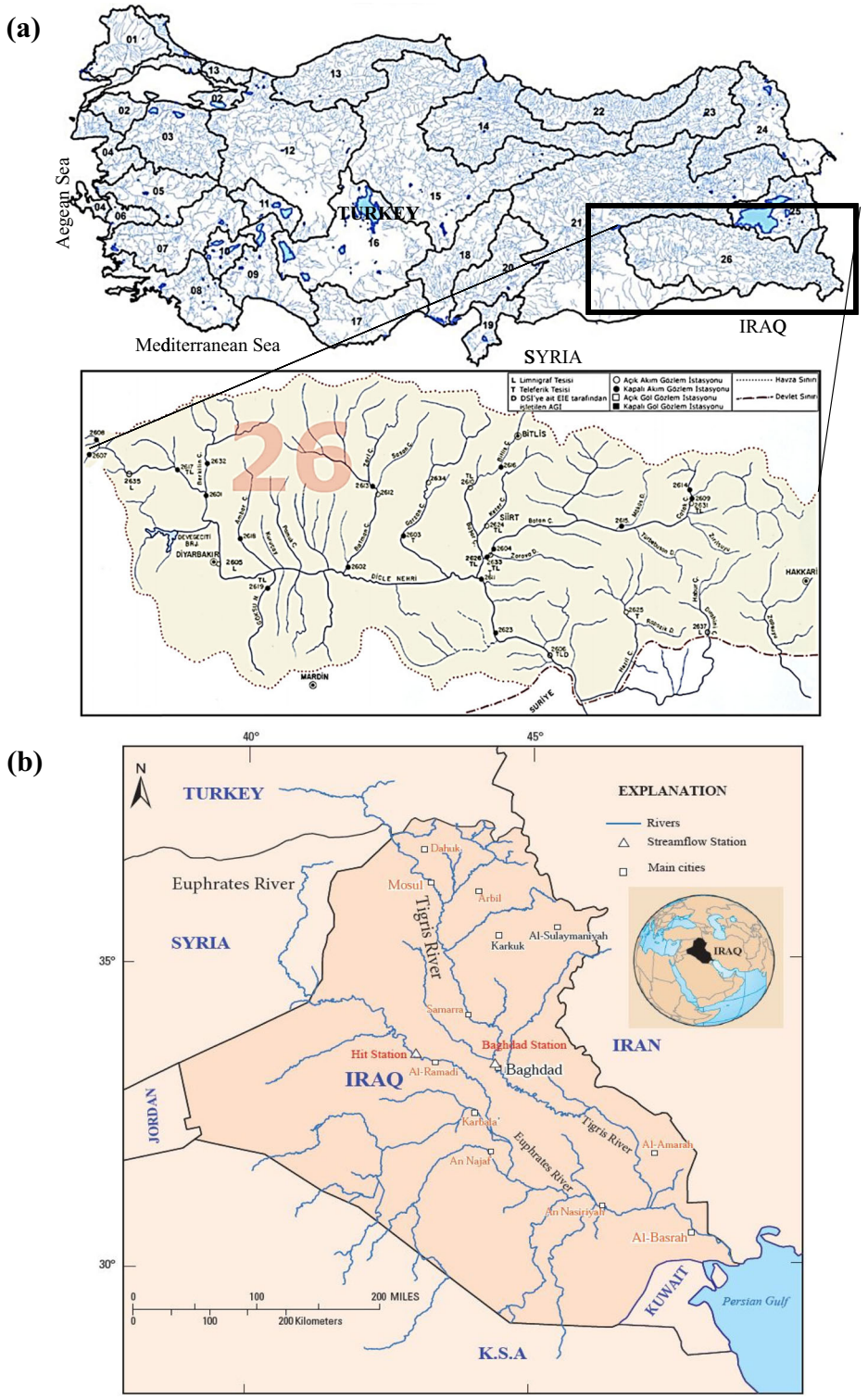
(a)
Aegean Sea
TURKEY
Mediterranean Sea
SYRIA
IRAQ
26
BİTLİS
SİİRT
HAKKARİ
DİYARBAKIR
MARDİN
SURİYE
BİCİLE NEHRİ
(b)
40°
45°
35°
30°
N
TURKEY
SYRIA
IRAQ
IRAN
JORDAN
KUWAIT
K.S.A
Euphrates River
Tigris River
Dahuk
Mosul
Arbil
Karkuk
Al-Sulaymaniyah
Samarra
Hit Station
Al-Ramadi
Baghdad Station
Baghdad
Karbala'
An Najaf
Euphrates River
Tigris River
Al-Amarah
An Nasiriyah
Al-Basrah
Persian Gulf
EXPLANATION
Rivers
Streamflow Station
Main cities
0 100 200 MILES
0 100 200 Kilometers

Fig. 2 **a** The Large basins of Turkey and The Malabadi (2612), Besiri (2603) stations, **b** Hit and Baghdad stations which are located in Iraq region

function, which is able to forecast continuous numerical attribution. As shown in Fig. 1, M5 model tree perform its algorithm based on two stages, at the first stage time series data are divided into subset in order to initiate the decision tree. The splitting criterion for this model is relying on the standard deviation of the class values that reach a node as an amount of error at that node. Then after, computing the expected reduction in this error as a result of testing each attribute at that node (Solomatine and Dulal 2003; Pal and Deswal 2009). Now, the equation that compute the standard deviation reduction (SDR) can be expressed as:

$$\mathrm{SDR} = \mathrm{sd}(\mathrm{K}) - \sum \frac{|\mathrm{Ki}|}{|\mathrm{K}|} \mathrm{sd}(\mathrm{Ki}) \tag{9}$$

The variables of the SDR formula explained are as follows; (i) *sd* represents the standard deviation, (ii) *K* denotes a set of examples that reaches the node, and (iii) the subset of examples that have the ith outcome of the potential set is represented as *Ki*. In the partition procedure, the first generation (child) nodes are less than the origin node in data's standard deviation. As final step in first stage, M5 selects the split that maximizes the envisioned error reduction. Nevertheless, this separation usually produces a large diagram (tree) structure that need to be pruned subtrees using linear regression functions, which is representing the second stage of M5 modeling.

2.4 Multiple Linear Regression

There are several engineering applications involve exploring the relationship between two or more parameters. Regression analysis model is one of the popular statistical approach that is highly recommended for these kind of problems. Throughout the literature, streamflow forecasting has been undertaken using MLR model, due to the fact that this model comprises many regressors to deal with the time series data base. Theoretically, the relationship between the dependent variable (Y) "i.e., one-step-ahead streamflow" and the independent variables (Xi) "i.e., the preceding streamflow records" can be described as followed:

$$\mathrm{Y} = \mathrm{P_o} + \mathrm{P_1X_1} + \mathrm{P_2X_2} + \cdots + \mathrm{P_nX_n} \tag{10}$$

Where Y is the target output, $\mathrm{P_i}$ (i=0,…., n) are the regression coefficients, and Xi (i=0,…., n) are the input variables.

2.5 Model Performance Indicators

Hydrological applications usually are evaluated based on quantitative indicators. Legates and McCabe (1999) stated in their study that predictive models in the scope of hydrology recommended to be examined using "goodness-of-fit" for example determination coefficient (*R*) and minimum one of absolute error performance criteria (e.g., mean absolute error (MAE) and root mean square error (RMSE)). Thus, the proposed data-driven models were evaluated

with respect to RMSE, MAE and R for each input combination. The statistic measure RMSE and MAE are formulated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{Q_o}-\mathrm{Q_f})^2} \tag{11}$$

$$\mathrm{MAE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{Q_o}-\mathrm{Q_f}|} \tag{12}$$

$$R = \frac{\sum_{t=1}^{n}\left[\left(Q_o-\overline{Q_o}\right)\left(Q_f-\overline{Q_f}\right)\right]}{\sqrt{\sum_{t=1}^{n}\left(Q_o-\overline{Q_o}\right)^2\sum_{t=1}^{n}\left(Q_f-\overline{Q_f}\right)^2}} \tag{13}$$

where N is the number of the raw streamflow data, Q_o is the actual (observed) flow values and Q_f is the model output.

3 Cases Studies and Data Preparation

3.1 Turkey Region

Average monthly intermittent streamflow data of two stations in the East-Anatolia region located in Southeast Turkey were used. The location of the stations was

Table 1 The monthly statistical parameters of data set for Besiri, Malabadi, Hit, Baghdad stations

Stations	Data set	x_{mean} (m^3/s)	Sx (m^3/s)	Csx (m^3/s)	x_{min} (m^3/s)	x_{max} (m^3/s)	r1	r2	r3
Besiri	1991-1999	55.93	68.78	1.66	2.5	306	0.658	0.189	-0.050
	1982-1990	42.39	52.25	1.94	2.0	284.4	0.633	0.170	-0.134
	1973-1981	47.66	63.37	2.27	1.0	354	0.623	0.183	-0.038
	1964-1972	68.68	137.68	4.39	0.1	964	0.745	0.370	0.155
Malabadi	1991-1999	134.18	142.21	1.21	5.1	568.6	0.681	0.253	-0.057
	1982-1990	120.58	127.95	1.17	0.2	519.7	0.712	0.270	-0.080
	1973-1981	126.10	132.49	1.19	1.2	533.9	0.661	0.280	0.022
	1964-1972	136.64	158.52	1.46	0.15	608	0.679	0.1655	-0.190
Hit	1989-1997	1146.42	1137.79	1.87	85.7	5797	0.702	0.270	-0.010
	1980-1988	680.32	546.43	1.85	71.50	3212	0.648	0.249	-0.020
	1970-1989	725.44	291.52	1.35	244.80	1987	0.693	0.414	0.190
	1960-1969	448.08	297.85	1.26	152.30	1442	0.620	0.339	0.121
Baghdad	1996-2005	1085.61	718.76	0.94	334.50	2865	0.756	0.322	-0.060
	1987-1995	816.37	422.55	1.45	292.70	2275	0.776	0.415	0.055
	1977-1986	855.89	446.89	1.91	379	2651	0.832	0.656	0.451
	1968-1976	597.50	196.65	1.05	331.40	1386	0.798	0.600	0.460

Table 2 Regularization constant and width of RBF kernel parameters of the optimal LSSVR models for each combination Besiri, Malabadi, Hit and Baghdad stations

Cross validation	Training data set	Test data set	Input combination		
			(i)	(ii)	(iii)
Besiri					
M1	1964-1981	1991-1999	(3,100)	(100,32)	(10,29)
M2	1964-1972 and 1991-1999	1982-1990	(100,5)	(1,1)	(11,2)
M3	1964-1972 and 1982-1999	1973-1981	(60,5)	(100,6)	(2,2)
M4	1973-1999	1964-1972	(93,3)	(86,10)	(100,4)
Malabadi					
M1	1964-1981	1991-1999	(100,29)	(4,5)	(5,6)
M2	1964-1972 and 1991-1999	1982-1990	(100,3)	(1,2)	(1,2)
M3	1964-1972 and 1982-1999	1973-1981	(76,100)	(100,18)	(100,22)
M4	1973-1999	1964-1972	(72,4)	(100,10)	(100,10)
Hit					
M1	1960-1980	1989-1997	(7,11)	(2,26)	(15,100)
M2	1960-1970 and 1989-1997	1980-1988	(4,100)	(2,89)	(2,100)
M3	1960-1969 and 1980-1997	1970-1989	(2,100)	(3,100)	(3,100)
M4	1970-1997	1960-1969	(49,18)	(100,61)	(2,100)
Baghdad					
M1	1968-1987	1996-2005	(65,24)	(4,11)	(90,100)
M2	1968-1977 and 1996-2005	1987-1995	(16,100)	(5,100)	(5,100)
M3	1968-1976 and 1987-2005	1977-1986	(1,1)	(1,2)	(28,100)
M4	1970-2005	1968-1976	(3,1)	(5,4)	(43,82)

illustrated in Fig. 2a. In this study, the Besiri Station (Station No: 2603) on the Garzan Stream and Malabadi Station (Station No: 2612) on the Batman Stream, in the Firat-Dicle Basin of Turkey were used. The drainage areas at these sites are 2450 km^2 for Besiri and 4105 km^2 for Malabadi. In Turkey, the first largest basin is Firat (basin number 21) with an approximately 127,000 km^2 of land zone. Dicle Basin (basin number 26) is the third largest basin with an almost 57,000 km^2 of land zone. Rely on basin land area, the Firat basin is the largest, with a total yearly flow volume approximately 32 billion m^3. The second one is Dicle Basin, with approximately 25 billion m^3 (Kaygusuz 1999; Demirbas and Bakis 2003). Streamflow forecasting for this region is very important for many of the activities such as flood mitigation, management of water reservoirs, distribution of drinking water and management of water infrastructures and dam planning etc. The observed data are 35 years (420 months) long with an observation period between 1964 and 1999 for mentioned stations. The observed data were obtained from the report of the Turkish General Directorate of Electrical Power Resources Survey and Development Administration.

3.2 Iraq Region

Another two stations were selected to apply in this study which are Hit station on the Euphrates River and Baghdad station on Tigris River in Iraq region, as shown in Fig. 2b. Hit and

Table 3 Comparison of LSSVR models

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
Besiri						
RMSE	M1	1991-1999	109.21	104.78	37.02	**83.67**
	M2	1982-1990	48.32	45.74	43.78	**45.95**
	M3	1973-1981	39.10	33.11	32.67	**34.96**
	M4	1964-1972	49.03	42.96	42.18	**44.72**
		Mean	**61.42**	**56.65**	**38.91**	**52.33**
MAE	M1	1991-1999	50.75	43.21	23.85	**39.27**
	M2	1982-1990	31.15	30.28	27.14	**29.52**
	M3	1973-1981	25.54	20.12	20.36	**22.01**
	M4	1964-1972	32.53	27.91	26.90	**29.11**
		Mean	**34.99**	**30.38**	**24.56**	**29.98**
R	M1	1991-1999	0.702	0.711	0.796	**0.736**
	M2	1982-1990	0.653	0.715	0.738	**0.702**
	M3	1973-1981	0.672	0.784	0.785	**0.747**
	M4	1964-1972	0.698	0.782	0.800	**0.760**
		Mean	**0.681**	**0.748**	**0.780**	**0.736**
Malabadi						
RMSE	M1	1991-1999	111.84	86.85	85.15	**94.61**
	M2	1982-1990	95.45	91.58	91.41	**92.81**
	M3	1973-1981	88.33	74.35	74.11	**78.93**
	M4	1964-1972	100.40	88.10	84.03	**90.84**
		Mean	**99.01**	**85.22**	**83.68**	**89.30**
MAE	M1	1991-1999	77.14	59.54	56.64	**64.44**
	M2	1982-1990	67.93	62.05	62.87	**64.28**
	M3	1973-1981	65.01	51.23	51.20	**55.81**
	M4	1964-1972	74.39	61.67	56.36	**64.14**
		Mean	**71.12**	**58.62**	**56.77**	**62.17**
R	M1	1991-1999	0.710	0.850	0.859	**0.806**
	M2	1982-1990	0.696	0.740	0.731	**0.722**
	M3	1973-1981	0.721	0.814	0.817	**0.784**
	M4	1964-1972	0.704	0.782	0.805	**0.764**
		Mean	**0.708**	**0.796**	**0.803**	**0.769**
Hit						
RMSE	M1	1989-1997	268.79	292.52	296.26	**285.86**
	M2	1980-1988	210.42	211.25	212.42	**211.36**
	M3	1970-1989	415.14	404.69	405.75	**408.53**
	M4	1960-1969	842.02	859.47	906.45	**869.31**
		Mean	**434.09**	**441.98**	**455.22**	**443.77**
MAE	M1	1989-1997	198.06	226.40	227.85	**217.44**
	M2	1980-1988	138.39	142.05	144.84	**141.76**
	M3	1970-1989	286.02	274.95	273.95	**278.31**
	M4	1960-1969	520.63	468.67	509.31	**499.54**

Table 3 (continued)

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
		Mean	**285.78**	**278.02**	**288.99**	**284.26**
R	M1	1989-1997	0.616	0.624	0.628	**0.623**
	M2	1980-1988	0.694	0.699	0.701	**0.698**
	M3	1970-1989	0.649	0.679	0.679	**0.669**
	M4	1960-1969	0.711	0.790	0.803	**0.768**
		Mean	**0.667**	**0.698**	**0.703**	**0.689**
Baghdad						
RMSE	M1	1996-2005	137.14	164.05	173.92	**158.37**
	M2	1987-1995	248.34	253.19	252.92	**251.48**
	M3	1977-1986	264.16	233.92	235.42	**244.50**
	M4	1968-1976	467.32	411.36	414.09	**430.92**
		Mean	**279.24**	**265.63**	**269.09**	**271.32**
MAE	M1	1996-2005	110.21	132.92	143.77	**128.97**
	M2	1987-1995	152.47	168.38	174.29	**165.05**
	M3	1977-1986	199.32	170.68	173.68	**181.23**
	M4	1968-1976	346.36	291.36	282.81	**306.84**
		Mean	**202.09**	**190.84**	**193.64**	**195.52**
R	M1	1996-2005	0.799	0.786	0.785	**0.790**
	M2	1987-1995	0.831	0.828	0.831	**0.830**
	M3	1977-1986	0.782	0.840	0.832	**0.818**
	M4	1968-1976	0.764	0.822	0.822	**0.803**
		Mean	**0.794**	**0.819**	**0.818**	**0.810**

Baghdad stations are covered a drainage area approximately 264,100 km^2 and 134,000 km^2, respectively. The geographic position of the Hit and Baghdad stations areas are stretched between (33° 36' 23") N Latitude and (42° 50' 14") E Longitude, (33° 24' 34") N Latitude and (44° 20' 32") E Longitude. Euphrates and Tigris Rivers are the essential source of fresh water, socioeconomic development and the political stabilization in this region. Developing such accurate forecasting and predicting river flow modeling in particular long-term (e.g., monthly streamflow) are significantly important to provide a considerable economic benefit, improve the irrigation sector, and solve the water shortage problems. The monthly streamflow data records 38 years (456 months) between (1960-1997) for Hit and Baghdad stations between (1968-2005) were used for this application. The hydrological data were obtained from the descriptive research that was conducted by Saleh (2010).

3.3 Data Time Series Preparation

For all presented stations, streamflow data time series were splitted into four training/testing divisions in order to achieve the best effective model formulation. For both of the applications forecasting and predicting, three divisions of the data were utilized to train the models, while the fourth was used to validate (test) the models network. The testing data phase was changed in all application; therefore, four different scenarios were investigated. Table 1

Table 4 Comparison of MARS models

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
Besiri						
RMSE	M1	1991-1999	93.28	86.96	86.26	**88.83**
	M2	1982-1990	50.80	51.49	51.42	**51.24**
	M3	1973-1981	41.20	38.24	33.93	**37.79**
	M4	1964-1972	48.67	40.51	42.82	**44.00**
		Mean	**58.49**	**54.30**	**53.61**	**55.47**
MAE	M1	1991-1999	44.13	38.48	37.83	**40.15**
	M2	1982-1990	32.19	31.81	30.91	**31.64**
	M3	1973-1981	26.46	24.28	20.48	**23.74**
	M4	1964-1972	32.65	26.19	27.41	**28.75**
		Mean	**33.86**	**30.19**	**29.16**	**31.07**
R	M1	1991-1999	0.750	0.778	0.782	**0.770**
	M2	1982-1990	0.643	0.677	0.677	**0.666**
	M3	1973-1981	0.627	0.718	0.775	**0.707**
	M4	1964-1972	0.704	0.808	0.782	**0.764**
		Mean	**0.681**	**0.745**	**0.754**	**0.727**
Malabadi						
RMSE	M1	1991-1999	119.08	91.39	88.67	**99.71**
	M2	1982-1990	95.11	93.84	93.87	**94.27**
	M3	1973-1981	94.03	76.07	77.92	**82.67**
	M4	1964-1972	101.74	90.39	89.83	**93.99**
		Mean	**102.49**	**87.92**	**87.57**	**92.66**
MAE	M1	1991-1999	81.20	61.89	59.59	**67.56**
	M2	1982-1990	67.53	61.41	62.16	**63.70**
	M3	1973-1981	66.65	52.37	49.62	**56.21**
	M4	1964-1972	75.88	63.37	63.01	**67.42**
		Mean	**72.82**	**59.76**	**58.60**	**63.72**
R	M1	1991-1999	0.662	0.817	0.828	**0.769**
	M2	1982-1990	0.700	0.729	0.722	**0.717**
	M3	1973-1981	0.684	0.806	0.798	**0.763**
	M4	1964-1972	0.695	0.769	0.773	**0.746**
		Mean	**0.685**	**0.780**	**0.780**	**0.749**
Hit						
RMSE	M1	1989-1997	284.23	313.96	313.96	**304.05**
	M2	1980-1988	231.63	251.35	251.16	**244.71**
	M3	1970-1989	473.50	450.58	436.41	**453.50**
	M4	1960-1969	862.64	828.14	847.49	**846.09**
		Mean	**463.00**	**461.01**	**462.26**	**462.09**
MAE	M1	1989-1997	186.00	198.59	198.59	**194.39**
	M2	1980-1988	151.04	177.77	177.71	**168.84**
	M3	1970-1989	308.61	294.17	273.88	**292.22**
	M4	1960-1969	546.92	470.54	470.39	**495.95**

Table 4 (continued)

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
		Mean	**298.14**	**285.27**	**280.14**	**287.85**
R	M1	1989-1997	0.643	0.653	0.653	**0.649**
	M2	1980-1988	0.633	0.683	0.683	**0.667**
	M3	1970-1989	0.598	0.653	0.689	**0.647**
	M4	1960-1969	0.685	0.804	0.794	**0.761**
		Mean	**0.640**	**0.698**	**0.705**	**0.681**
Baghdad						
RMSE	M1	1996-2005	141.68	178.06	181.27	**167.00**
	M2	1987-1995	262.21	293.01	289.85	**281.69**
	M3	1977-1986	276.86	259.60	260.68	**265.71**
	M4	1968-1976	475.92	420.44	477.41	**457.92**
		Mean	**289.17**	**287.78**	**302.30**	**293.08**
MAE	M1	1996-2005	109.14	136.025	139.92	**128.36**
	M2	1987-1995	160.54	192.85	187.72	**180.37**
	M3	1977-1986	204.16	180.68	183.51	**189.45**
	M4	1968-1976	351.54	286.66	320.19	**319.46**
		Mean	**24.38**	**23.98**	**24.97**	**204.41**
R	M1	1996-2005	0.805	0.802	0.808	**0.805**
	M2	1987-1995	0.814	0.774	0.781	**0.790**
	M3	1977-1986	0.763	0.826	0.826	**0.805**
	M4	1968-1976	0.756	0.819	0.796	**0.790**
		Mean	**0.784**	**0.806**	**0.803**	**0.798**

indicated the statistical characteristics of each data set used in this study for all stations. Those statistical indicators included over all mean (X_{mean}), standard deviation (Sx), minimum and maximum flow records (X_{min} and X_{max}), skewness (Csx), and the antecedent values of auto-correlation coefficient.

4 Application and Analysis

The effectiveness of the proposed artificial intelligence approaches were examined upon actual streamflow data obtained from official organizations authorized for monitoring such river flows. In the first part of the current study, it was decided to prove the efficiency of the LSSVR, MARS and M5-Tree models to forecast one month ahead streamflow and compare the results with MLR model. In addition, the effect of the periodic time scale on the forecasting results was also explored. Whereas, the second part of the study is to investigate the applicability of the data-driven to predict monthly streamflow using inflow time series data belonging to the nearby river. Different input combinations based on the present and antecedent streamflow were used to model the forecasting and prediction. In other words, Q_t indicates the streamflow at time t, the input variables are; (i) Q_t, (ii) Q_t, Q_{t-1}, (iii) Q_t, Q_{t-1} and Q_{t-2}. This application section provides a comprehensive detailed discussion and analysis of the proposed

Table 5 Comparison of M5-Tree models

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
Besiri						
RMSE	M1	1991-1999	109.14	123.09	124.66	**118.96**
	M2	1982-1990	49.81	54.24	57.70	**53.92**
	M3	1973-1981	42.86	42.97	43.83	**43.22**
	M4	1964-1972	52.77	52.04	52.64	**52.48**
		Mean	**63.65**	**68.09**	**69.71**	**67.15**
MAE	M1	1991-1999	45.18	43.90	45.87	**44.98**
	M2	1982-1990	32.31	29.35	30.14	**30.60**
	M3	1973-1981	28.45	23.88	23.79	**25.37**
	M4	1964-1972	34.73	31.38	31.23	**32.45**
		Mean	**35.17**	**32.13**	**32.76**	**33.35**
R	M1	1991-1999	0.666	0.467	0.462	**0.532**
	M2	1982-1990	0.649	0.729	0.657	**0.678**
	M3	1973-1981	0.620	0.708	0.704	**0.678**
	M4	1964-1972	0.650	0.697	0.688	**0.678**
		Mean	**0.646**	**0.650**	**0.628**	**0.641**
Malabadi						
RMSE	M1	1991-1999	120.55	99.06	111.92	**105.56**
	M2	1982-1990	105.79	103.25	107.65	**100.39**
	M3	1973-1981	95.81	89.40	115.97	**161.87**
	M4	1964-1972	287.70	97.43	100.49	**119.59**
		Mean	**152.46**	**97.29**	**109.01**	**73.57**
MAE	M1	1991-1999	81.16	68.01	71.55	**69.97**
	M2	1982-1990	76.14	67.54	66.24	**66.11**
	M3	1973-1981	69.08	54.97	74.28	**104.81**
	M4	1964-1972	192.03	62.23	60.16	**78.62**
		Mean	**104.60**	**63.19**	**68.06**	**79.87**
R	M1	1991-1999	0.647	0.778	0.706	**0.710**
	M2	1982-1990	0.651	0.705	0.639	**0.665**
	M3	1973-1981	0.677	0.736	0.628	**0.680**
	M4	1964-1972	0.590	0.730	0.716	**0.679**
		Mean	**0.641**	**0.737**	**0.672**	**0.684**
Hit						
RMSE	M1	1989-1997	287.70	412.05	405.92	**368.56**
	M2	1980-1988	319.24	334.48	352.34	**335.35**
	M3	1970-1989	473.50	478.36	538.34	**496.73**
	M4	1960-1969	860.22	829.50	859.90	**742.69**
		Mean	**485.17**	**513.60**	**458.74**	**485.83**
MAE	M1	1989-1997	192.03	257.84	264.82	**238.23**
	M2	1980-1988	233.0	241.88	257.68	**244.19**
	M3	1970-1989	308.61	309.75	325.10	**314.49**
	M4	1960-1969	525.87	477.71	510.56	**343.09**

Table 5 (continued)

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
		Mean	**314.88**	**321.80**	**218.33**	**285.00**
R	M1	1989-1997	0.590	0.513	0.531	**0.545**
	M2	1980-1988	0.462	0.490	0.431	**0.461**
	M3	1970-1989	0.598	0.634	0.574	**0.602**
	M4	1960-1969	0.713	0.788	0.763	**0.769**
		Mean	**0.591**	**0.606**	**0.586**	**0.594**
Baghdad						
RMSE	M1	1996-2005	161.54	208.59	225.09	**198.41**
	M2	1987-1995	279.48	340.39	366.44	**328.77**
	M3	1977-1986	300.78	307.99	333.92	**314.23**
	M4	1968-1976	480.81	459.73	451.24	**463.93**
		Mean	**305.65**	**329.18**	**344.17**	**326.33**
MAE	M1	1996-2005	123.50	156.57	172.98	**151.01**
	M2	1987-1995	177.08	219.42	239.74	**212.08**
	M3	1977-1986	217.63	211.31	237.78	**222.24**
	M4	1968-1976	353.89	329.70	319.32	**334.30**
		Mean	**218.02**	**229.25**	**242.45**	**229.90**
R	M1	1996-2005	0.767	0.697	0.681	**0.715**
	M2	1987-1995	0.800	0.716	0.689	**0.735**
	M3	1977-1986	0.756	0.788	0.763	**0.769**
	M4	1968-1976	0.755	0.783	0.790	**0.776**
		Mean	**0.770**	**0.746**	**0.731**	**0.749**

methods. It should be remarked that the utilized river flow data for all rivers are continuous and do not experience any missing monitoring events data during the examination period.

4.1 Streamflow Forecasting

As mentioned in the previous section, the first scenario was undertaken to forecast monthly streamflow. For the purpose of how the statistical analysis will generalize an independent data set, each input combination was cross validated by partitioning the time series data into four sets. By recalling the main parameters of LSSVR model, different regularization constant and width of radial basis function kernel were tried to obtain the minimum RMSE indicator. Table 2 displayed the optimal LSSVR parameters models of each input combination for the testing phase. Tables 3, 4, 5, and 6 indicated the testing phase outcomes using LSSVR, MARS, M5 model tree and MLR models for the all stations (Besiri, Malabadi, Hit and Baghdad). According to the mean values of the performance indicators (e.g., RMSE and MAE) of the modeling, there is a remarkable difference can be observed in the results, which are the values of the root mean square error and mean absolute error. The Turkish rivers modeling showed low percentages of RMSE and MAE comparing the Iraq Rivers. This is due to the mean average flow of the rivers, Garzan and Batman Rivers are characterized by mean river flow 53.66 and 129.37 m^3/s, respectively. While Euphrates and Tigris rivers are 750.06 and 838.84 m^3/s, respectively.

Table 6 Comparison of MLR models

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
Besiri						
RMSE	M1	1991-1999	92.68	85.95	82.54	**87.06**
	M2	1982-1990	51.39	50.3	49.81	**50.50**
	M3	1973-1981	42.05	40.48	40.9	**41.14**
	M4	1964-1972	53.58	50.5	49.12	**51.07**
		Mean	**59.93**	**56.81**	**55.59**	**57.44**
MAE	M1	1991-1999	46.08	40.85	41.01	**42.65**
	M2	1982-1990	31.09	29.53	29.2	**29.94**
	M3	1973-1981	25.56	22.6	23.38	**23.85**
	M4	1964-1972	32.58	30.14	30.98	**31.23**
		Mean	**33.83**	**30.78**	**31.14**	**31.92**
R	M1	1991-1999	0.745	0.790	0.805	**0.780**
	M2	1982-1990	0.623	0.672	0.671	**0.655**
	M3	1973-1981	0.633	0.691	0.673	**0.666**
	M4	1964-1972	0.658	0.721	0.727	**0.702**
		Mean	**0.665**	**0.718**	**0.719**	**0.701**
Malabadi						
RMSE	M1	1991-1999	120.37	109.27	107.63	**112.42**
	M2	1982-1990	103.56	103.75	103.53	**103.61**
	M3	1973-1981	93.14	86.57	86.18	**88.63**
	M4	1964-1972	107.95	103.67	103.19	**104.94**
		Mean	**106.26**	**100.82**	**100.13**	**102.40**
MAE	M1	1991-1999	79.76	71.82	70.81	**74.13**
	M2	1982-1990	71.82	67.72	67.3	**68.95**
	M3	1973-1981	63.25	54.85	55.07	**57.72**
	M4	1964-1972	71.36	68.57	69.28	**69.74**
		Mean	**24.38**	**23.98**	**24.97**	**67.63**
R	M1	1991-1999	0.679	0.755	0.758	**0.731**
	M2	1982-1990	0.656	0.691	0.686	**0.678**
	M3	1973-1981	0.712	0.771	0.767	**0.750**
	M4	1964-1972	0.680	0.731	0.726	**0.713**
		Mean	**0.682**	**0.737**	**0.734**	**0.718**
Hit						
RMSE	M1	1989-1997	252.165	264.00	267.60	**250.88**
	M2	1980-1988	240.01	258.76	253.86	**250.87**
	M3	1970-1989	440.96	440.01	438.94	**439.97**
	M4	1960-1969	844.34	829.62	819.43	**831.13**
		Mean	**444.37**	**448.10**	**444.96**	**443.21**
MAE	M1	1989-1997	163.13	178.44	177.19	**172.92**
	M2	1980-1988	161.27	181.30	175.42	**172.66**
	M3	1970-1989	280.20	284.61	280.55	**281.79**
	M4	1960-1969	566.78	548.52	557.18	**557.49**

Table 6 (continued)

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
		Mean	**292.85**	**298.22**	**297.59**	**296.21**
R	M1	1989-1997	0.620	0.621	0.597	**0.613**
	M2	1980-1988	0.693	0.694	0.673	**0.686**
	M3	1970-1989	0.648	0.680	0.668	**0.665**
	M4	1960-1969	0.701	0.716	0.716	**0.711**
		Mean	**0.665**	**0.678**	**0.663**	**0.669**
Baghdad						
RMSE	M1	1996-2005	127.18	139.63	139.23	**135.35**
	M2	1987-1995	257.48	274.15	253.86	**261.83**
	M3	1977-1986	277.86	262.05	260.41	**266.77**
	M4	1968-1976	492.59	462.63	819.43	**591.55**
		Mean	**288.78**	**284.62**	**368.23**	**313.88**
MAE	M1	1996-2005	93.81	102.17	101.15	**99.04**
	M2	1987-1995	150.21	173.30	175.42	**166.31**
	M3	1977-1986	204.08	185.79	182.33	**190.73**
	M4	1968-1976	359.54	322.89	557.18	**413.20**
		Mean	**201.91**	**196.04**	**254.02**	**217.32**
R	M1	1996-2005	0.797	0.788	0.780	**0.788**
	M2	1987-1995	0.832	0.827	0.673	**0.778**
	M3	1977-1986	0.777	0.820	0.815	**0.804**
	M4	1968-1976	0.756	0.794	0.716	**0.755**
		Mean	**0.791**	**0.807**	**0.746**	**0.781**

Based on the mean performance of RMSE and MAE, Tables 3, 4, and 5) exhibited M3 as the best data set to forecast one month ahead for Besiri and Malabadi stations. This might be because M3 data set provides a knowledgeable pattern of flow in the training and testing phases of the models that could perform very well comparing to the other data sets. On the other hand, the worst data set was M1 for LSSVR, MARS and M5 model tree for all the investigated inputs combination. This can be expounded that LSSVR, MARS and M5 model tree could not explore the nature of the streamflow of the M1 data set in the training and testing periods. However, LSSVR results outperformed MARS and M5-Tree models and the outstanding outcome presented for M3 data set period for the input combination (iii). The optimal LSSVR model (M3 data set and input iii) increased the RMSE accuracy of the optimal MARS and M5-Tree models by 3.9 and 31.2 % for the Besiri and by 2.6 and 20.6 % for the Malabadi stations, respectively. It should be noted that there is also a significant difference between MARS and M5-Tree for the both stations. Euphrates and Tigris Rivers modeling were totally different with obvious fluctuation of the best performance results. The consistency of the Iraq rivers region modeling conclusion was diverse, various data sets with different inputs combination performed the remarkable results of the used intelligence approaches. Hit station modeling showed the best accuracy belonging to M2 with one lag time for the LSSVR and MARS models, while M5 model tree demonstrated the best results for the M1 with one lag as well (the input combination (i)). Baghdad station obtained its best application using the first

Table 7 Comparison of the P-LSSVR models

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
Besiri						
RMSE	M1	1991-1999	95.82	96.34	30.27	**74.14**
	M2	1982-1990	34.26	35.24	37.02	**35.51**
	M3	1973-1981	23.18	22.88	22.03	**22.70**
	M4	1964-1972	35.30	35.82	33.65	**34.92**
		Mean	**47.14**	**47.57**	**30.74**	**41.82**
MAE	M1	1991-1999	39.52	38.89	19.74	**32.72**
	M2	1982-1990	22.02	23.23	24.19	**23.15**
	M3	1973-1981	13.46	13.76	13.01	**13.41**
	M4	1964-1972	22.80	23.29	22.31	**22.80**
		Mean	**24.45**	**24.79**	**19.81**	**23.02**
R	M1	1991-1999	0.781	0.768	0.869	**0.806**
	M2	1982-1990	0.846	0.838	0.823	**0.836**
	M3	1973-1981	0.902	0.903	0.909	**0.905**
	M4	1964-1972	0.857	0.852	0.871	**0.860**
		Mean	**0.846**	**0.840**	**0.868**	**0.852**
Malabadi						
RMSE	M1	1991-1999	59.55	77.48	64.07	**67.03**
	M2	1982-1990	69.42	53.11	74.56	**65.70**
	M3	1973-1981	57.16	57.54	59.52	**58.07**
	M4	1964-1972	62.25	63.72	65.32	**63.76**
		Mean	**62.10**	**62.96**	**65.87**	**63.64**
MAE	M1	1991-1999	38.80	49.19	43.51	**43.83**
	M2	1982-1990	44.61	33.12	49.52	**42.42**
	M3	1973-1981	37.81	38.65	41.20	**39.22**
	M4	1964-1972	38.53	38.57	40.18	**39.09**
		Mean	**39.94**	**39.88**	**43.60**	**41.14**
R	M1	1991-1999	0.932	0.888	0.920	**0.913**
	M2	1982-1990	0.863	0.935	0.832	**0.877**
	M3	1973-1981	0.898	0.896	0.888	**0.894**
	M4	1964-1972	0.899	0.894	0.888	**0.894**
		Mean	**0.898**	**0.903**	**0.882**	**0.894**
Hit						
RMSE	M1	1989-1997	325.16	327.19	330.16	**327.50**
	M2	1980-1988	251.96	249.14	255.68	**252.26**
	M3	1970-1989	258.28	371.25	364.35	**331.29**
	M4	1960-1969	739.75	842.85	891.93	**824.84**
		Mean	**393.79**	**447.61**	**460.53**	**433.98**
MAE	M1	1989-1997	234.89	232.38	235.29	**234.19**
	M2	1980-1988	188.51	186.25	193.50	**189.42**
	M3	1970-1989	53.11	249.79	242.39	**181.76**
	M4	1960-1969	425.21	467.19	500.79	**464.40**

Table 7 (continued)

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
		Mean	**225.43**	**283.90**	**292.99**	**267.44**
R	M1	1989-1997	0.521	0.525	0.519	**0.522**
	M2	1980-1988	0.594	0.622	0.604	**0.607**
	M3	1970-1989	0.736	0.736	0.746	**0.739**
	M4	1960-1969	0.828	0.796	0.807	**0.810**
		Mean	**0.670**	**0.670**	**0.669**	**0.669**
Baghdad						
RMSE	M1	1996-2005	228.54	191.43	193.39	**204.45**
	M2	1987-1995	284.63	287.86	288.77	**287.09**
	M3	1977-1986	207.19	210.57	212.20	**209.99**
	M4	1968-1976	256.43	361.95	370.49	**329.62**
		Mean	**244.20**	**262.95**	**266.21**	**257.79**
MAE	M1	1996-2005	181.10	157.63	158.75	**165.83**
	M2	1987-1995	205.54	209.49	209.78	**208.27**
	M3	1977-1986	147.86	151.64	149.68	**149.73**
	M4	1968-1976	240.33	241.16	249.23	**171.57**
		Mean	**139.71**	**189.98**	**191.86**	**173.85**
R	M1	1996-2005	0.605	0.733	0.726	**0.688**
	M2	1987-1995	0.773	0.766	0.764	**0.768**
	M3	1977-1986	0.881	0.877	0.879	**0.879**
	M4	1968-1976	0.860	0.865	0.858	**0.861**
		Mean	**0.780**	**0.810**	**0.807**	**0.799**

data set (M1) with one antecedent value of flow to forecast one-month-ahead. The variance of the best results here is because of the phenomena that characterized Iraq climatology which is highly nonstationary and each approach dealt with the data base with different consistency. Here, the lowest standard indicators appeared for the fourth data set (M4) with respect to the all inputs combination. In general, it could be noticed that LSSVR provides the admirable forecasting modeling of streamflow over the other methods. The RMSE performance of the best MARS and M5-Tree models was increased using the best LSSVR model by 10.1 and 36.7 % for the Hit and by 3.3 and 17.8 % for the Baghdad stations, respectively. Similar to the previous application, here also a considerable difference exists between MARS and M5-Tree models.

Traditionally, MLR models were examined for the same data sets and the remarkable goodness in term of RMSE and MAE were selected for comparison purpose. MLR results presented in Table 6 for all the stations. There is an outstanding harmony with gained results regarding the data sets and the preceding input vectors comparing with LSSVR, MARS and M5 model methods. What is worth to be observed? There is a noteworthy enhancement in the application of LSSVR, MARS and M5-Tree model methods comparatively with MLR method. In order to describe this improvement in rational way, the percentages of the accuracy increment for the performance criteria have been calculated. The mean RMSE and MAE accuracies of the MLR model successfully

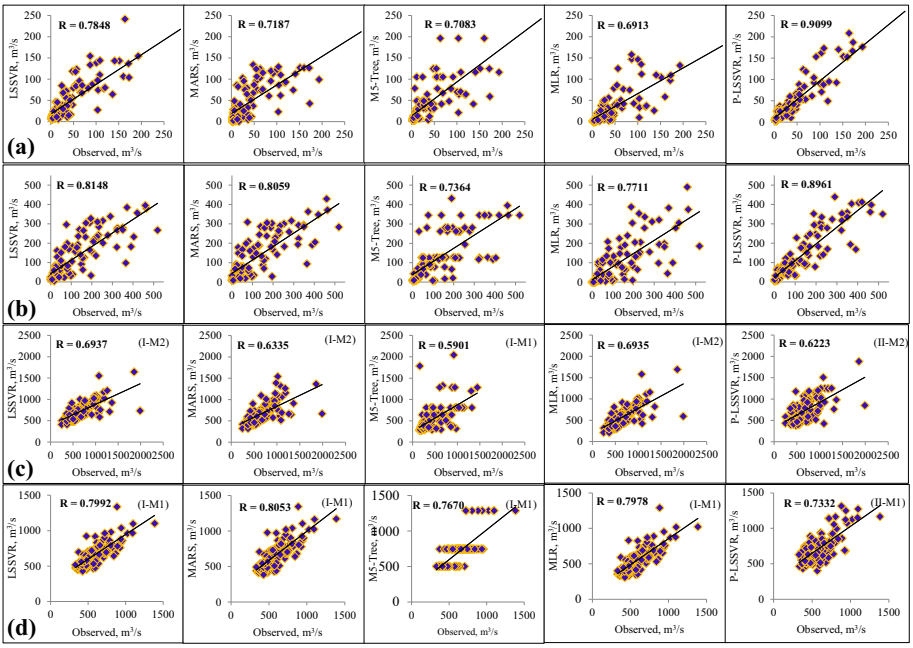

Fig. 3 The observed and forecasted streamflows scatterplot by the LSSVR, MARS, M5-Tree, MLR and P-LSSVR, **a** the M3 data set-Beşiri station, **b** the M3 data set-Malabadi station **c** the M1 and M2 data sets–Hit station, and **d** the M1 data set - Baghdad station

increased using LSSVR model by 8.95-4.19 %, 12.8-8.08 %, -0.12-4.03 % and 13.56-10.03 % for Besiri, Malabadi, Hit and Baghdad stations, respectively.

The periodicity data component was also examined and evaluated for the forecasting modeling section. In fact, the main idea behind including this periodic sub data which is one year to forecast one month ahead, is to supply the modeling an external pattern of flow that might give a comprehensive knowledge and better accuracy of results. Table 7 displayed the results of the testing phase for periodic LSSVR model. Obviously, adding the periodicity component has increased the average LSSVR model performance accuracy in term of the RMSE and MAE by 20-23.21 %, 28.73-33.82 %, 2.20-5.91 % and 4.98-11.08 % for Besiri, Malabadi, Hit and Baghdad stations, respectively. By comparing Table 7 with 3, the periodic LSSVR indicates the same consistency of modeling accuracy with LSSVR for Besiri and Malabadi stations which are M3 the best model and M1 the worst model. In addition, Hit station gives the same combination of results M2 the best model and M4 the worst model.

Table 8 The optimal parameters of the LSSVR models in cross application

Cross validation	Test data set	Input combinations		
		(i)	(ii)	(iii)
M1	1991-1999	(25,1)	(2,1)	(1,1)
M2	1982-1990	(100,3)	(7,1)	(32,2)
M3	1973-1981	(100,4)	(17,2)	(100,5)
M4	1964-1972	(16,1)	(24,1)	(91,3)

Table 9 Comparison of the LSSVR and MARS models in predicting monthly streamflow's of the Malabadi Station by using the data of Besiri station

Model	Statistics	Cross validation	Test data set	Input combinations			
				(i)	(ii)	(iii)	Mean
LSSVR	RMSE	M1	1991-1999	125.54	104.86	108.27	**112.89**
		M2	1982-1990	97.76	85.20	84.25	**89.07**
		M3	1973-1981	92.02	75.77	74.47	**80.75**
		M4	1964-1972	97.36	82.968	83.48	**87.94**
			Mean	**103.17**	**87.20**	**87.62**	**92.66**
	MAE	M1	1991-1999	84.01	68.42	71.02	**74.48**
		M2	1982-1990	71.91	59.44	58.11	**63.15**
		M3	1973-1981	66.30	55.04	52.08	**57.81**
		M4	1964-1972	73.00	57.64	57.94	**62.86**
			Mean	**73.81**	**60.14**	**59.79**	**64.58**
	R	M1	1991-1999	0.610	0.750	0.737	**0.699**
		M2	1982-1990	0.672	0.771	0.780	**0.741**
		M3	1973-1981	0.691	0.804	0.812	**0.769**
		M4	1964-1972	0.729	0.811	0.808	**0.783**
			Mean	**0.676**	**0.784**	**0.784**	**0.748**
MARS							
	RMSE	M1	1991-1999	166.28	136.72	137.73	**146.91**
		M2	1982-1990	98.89	86.60	90.63	**92.04**
		M3	1973-1981	93.04	78.41	79.49	**83.65**
		M4	1964-1972	106.79	84.15	88.25	**93.06**
			Mean	**116.25**	**96.47**	**99.03**	**103.92**
	MAE	M1	1991-1999	100.74	80.41	81.12	**87.42**
		M2	1982-1990	70.06	61.58	67.38	**66.34**
		M3	1973-1981	66.98	56.29	58.45	**60.57**
		M4	1964-1972	77.98	60.36	63.96	**67.43**
			Mean	**78.94**	**64.66**	**67.73**	**70.44**
	R	M1	1991-1999	0.525	0.661	0.656	**0.614**
		M2	1982-1990	0.663	0.762	0.731	**0.719**
		M3	1973-1981	0.683	0.790	0.783	**0.752**
		M4	1964-1972	0.656	0.806	0.782	**0.748**
			Mean	**0.632**	**0.755**	**0.738**	**0.708**

Whereas, Baghdad station presents different outcome the best testing data set was 1977-1986 (M3) and the worest testing data set was similar to the previous application od the LSSVR, 1968-1976 (M4).

Further assessment for the effectiveness of the utilized data-driven models, it seems reasonable to investigate the linear relationship between the observed and forecasted time series for the testing period. Scatter plots are illustrated in Figs. 3a, b belonging to Besiri and Malabadi stations, respectively. Those figures demonstrated the best models of LSSVR, MARS, M5 model tree, periodic LSSVR (P-LSSVR) and MLR models for M3 input combination. P-LSSVR has been found the best model displayed closed to the fit line

Table 10 Comparison of the M5-Tree and MLR models in predicting monthly streamflow's of the Malabadi Station by using the data of Besiri station

Model	Statistics	Cross validation	Test data set	Input combinations			
				(i)	(ii)	(iii)	Mean
M5-Tree	RMSE	M1	1991-1999	134.28	116.83	121.29	**124.13**
		M2	1982-1990	100.92	101.20	104.19	**102.10**
		M3	1973-1981	100.08	109.58	101.85	**103.84**
		M4	1964-1972	111.77	93.58	83.33	**96.23**
			Mean	**111.76**	**105.30**	**102.67**	**106.58**
	MAE	M1	1991-1999	86.06	76.06	73.66	**78.59**
		M2	1982-1990	72.08	61.14	68.65	**67.29**
		M3	1973-1981	71.06	69.59	65.00	**68.55**
		M4	1964-1972	79.99	65.10	58.43	**67.84**
			Mean	**77.30**	**67.97**	**66.44**	**70.57**
	R	M1	1991-1999	0.570	0.683	0.663	**0.639**
		M2	1982-1990	0.650	0.716	0.713	**0.693**
		M3	1973-1981	0.623	0.664	0.696	**0.661**
		M4	1964-1972	0.622	0.754	0.811	**0.729**
			Mean	**0.616**	**0.704**	**0.721**	**0.680**
MLR							
	RMSE	M1	1991-1999	230.94	228.73	227.95	**229.21**
		M2	1982-1990	129.86	127.84	126.56	**128.09**
		M3	1973-1981	121.19	118.64	118.30	**119.38**
		M4	1964-1972	125.55	122.67	121.39	**123.20**
			Mean	**151.89**	**149.47**	**148.55**	**149.97**
	MAE	M1	1991-1999	115.60	111.42	113.38	**113.47**
		M2	1982-1990	82.05	81.22	79.92	**81.06**
		M3	1973-1981	77.20	75.21	74.34	**75.58**
		M4	1964-1972	79.93	78.57	77.75	**78.75**
			Mean	**88.70**	**86.61**	**86.35**	**87.22**
	R	M1	1991-1999	0.483	0.530	0.522	**0.512**
		M2	1982-1990	0.594	0.642	0.638	**0.625**
		M3	1973-1981	0.651	0.704	0.687	**0.681**
		M4	1964-1972	0.676	0.725	0.720	**0.707**
			Mean	**0.601**	**0.650**	**0.642**	**0.631**

comparing to the other models. Similarly, Fig. 3c, d showed the best fit line regression indicator regarding Hit and Baghdad stations. Hit station performed the best value of R for LSSVR model with M2 data set and input combination (i). However, it is evident based on Fig. 3c that there is a slight deviation between LSSVR model and MLR. Fig. 3d displayed the best fit line all the models for M1 and combination (i), except MLR method with combination (ii), for Baghdad station.

Overall, LSSVR and MARS generally performed superior to M5-Tree and MLR models. The reason behind this may be the fact that the linear structure of the M5-Tree and MLR models prevents them from accurately modeling highly nonlinear streamflow process. Wang

et al. 2009 compared the ability of autoregressive moving-average ARMA, ANN, ANFIS, genetic programming (GP) and SVM methods in forecasting monthly discharge time series and they obtained R of 0.786, 0.786, 0.801, 0.815 and 0.823 for the ARMA, ANN, ANFIS, GP and SVM, respectively. Rezaeian-Zadeh et al. 2013 predicted monthly discharges in a semi-arid region using ANN with different training algorithms and they found that the best ANN model trained with scaled conjugate gradient algorithm provided a correlation 0.78. Turan and Yurdusev 2014 used ANFIS and genetic fuzzy system (GFS) in predicting monthly river flows of Gediz Basin in Turkey and they obtained R of 0.84 and 0.85 for the best ANFIS and GFS models. It is clear from the presented tables "performance metrics" that the LSSVR and MARS models provided accurate results in forecasting monthly streamflow from the R^2 viewpoint.

4.2 Streamflow Predicting

In this section, streamflow's prediction has been conducted using the LSSVR, MARS, M5 model tree, P-LSSVR and MLR based on nearby streamflow data for particular station. The significant of this kind of modeling is for the cases of missing river flow or the poor quality of discharge monitoring (e.g., upstream or downstream stations). For this kind of problem, streamflow prediction using nearby station can be highly useful to predict the missing data. In this study, the prediction was undertaken for the Turkish streams. This is for the reason that Garzan and Batman rivers have the same drainage hydrological features; so that, the prediction will be implemented in homogenous physical characteristics. Here also, the data base was cross-validated and divided into four divisions. With similar to the previous sub section application procedure, Table 8 expresses the optimal parameters of LSSVR model. For the scenario of predicting streamflow at Malabadi station (Batman River) using river flow data of Besiri station (Garzan River), Table 9 and 10 provided the modeling evaluators of LSSVR, MARS and M5 Tree models, respectively. According to the mean RMSE and MAE indicators, the highest score given by LSSVR and MARS models for M3 and input combination (iii) and (ii); in that order, while M5 Tree model score the best accuracy of M4 data set and two lagged times. Negatively, the three models gave the lowest accuracy scores for M1 data set. The best LSSVR model (M3 data set and input iii) increased the RMSE performance of the best MARS (M3 data set and input ii) and M5-Tree (M4 data set and input iii) models by 5.3 and 11.9 %, respectively. Comparison of the best explored model which is using LSSVR approach with MLR model (table 10), there were a positive improvement in the prediction scenario accuracies in term of mean RMSE and MAE by 37.04-29.95 %, respectively.

Table 11 The optimal parameters of P-LSSVR models in cross application

Cross validation	Test data set	Input combinations		
		(i)	(ii)	(iii)
M1	1991-1999	(9,1)	(4,2)	(63,7)
M2	1982-1990	(100,15)	(1,2)	(1,2)
M3	1973-1981	(17,2)	(35,3)	(32,3)
M4	1964-1972	(1,5)	(8,3)	(5,2)

Table 12 Comparison of the P-LSSVR models in predicting monthly streamflow's of the Malabadi Station by using the data of Besiri station

Statistics	Cross validation	Test data set	Input combinations			
			(i)	(ii)	(iii)	Mean
RMSE	M1	1991-1999	82.91	83.34	86.57	**84.27**
	M2	1982-1990	81.45	85.63	85.81	**84.30**
	M3	1973-1981	58.15	57.72	58.22	**58.03**
	M4	1964-1972	68.79	65.81	64.89	**66.50**
		Mean	**72.83**	**73.13**	**73.87**	**73.27**
MAE	M1	1991-1999	49.16	50.50	50.96	**50.21**
	M2	1982-1990	54.94	53.37	53.21	**53.84**
	M3	1973-1981	39.57	39.49	39.68	**39.58**
	M4	1964-1972	47.86	41.80	41.22	**43.63**
		Mean	**47.88**	**46.29**	**46.27**	**46.81**
R	M1	1991-1999	0.861	0.858	0.848	**0.856**
	M2	1982-1990	0.797	0.784	0.784	**0.788**
	M3	1973-1981	0.891	0.893	0.891	**0.892**
	M4	1964-1972	0.875	0.885	0.888	**0.883**
		Mean	**0.856**	**0.855**	**0.853**	**0.854**

The effect of embedding the periodicity feature was tested for prediction phase. This was conducted for the best accurate model has been obtained in the forgoing applications, which is least square support vector regression model. Again, the ideal regularization constant and RBF kernel values are visualized in Table 11. The test results of P-LSSVR is exhibited in Table 12; however, the best average performances accuracies of P-LSSVR were gained from M3 data

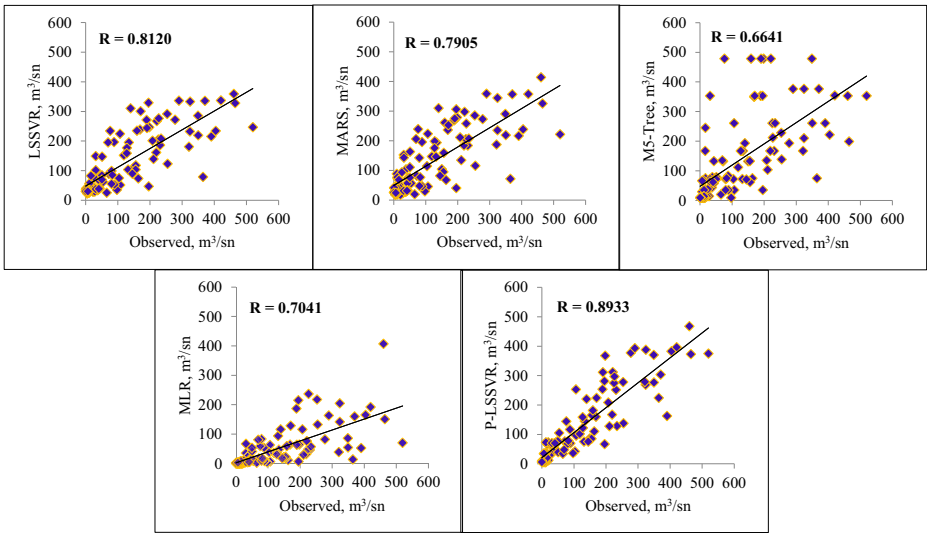

Fig. 4 The streamflow prediction of the Malabadi Station by LSSVR, MARS, M5-Tree, MLR and P-LSSVR using M3 data sets of Beşiri Station

set, whereas the worst model from M1 and M2 with slight variation. To further visualize the effect of including the periodic component, the percentages of the prediction development between LSSVR and P-LSSVR in term of the mean RMSE and MAE were 22.50-24.17 %, respectively. Finally, the actual and predicted river flow for LSSVR, MARS, M5 model tree, MLR and P-LSSVR are illustrated in Fig. 4 of the best sophisticated data set. Clearly, it was found that the closet prediction model is P-LSSVR with R value 0.89.

5 Conclusion

As a matter of fact, streamflow modeling is a challenging task for the hydrology researchers. This is due to the chaotic disturbances, complex non-linear dynamics and randomness phenomena of this hydrological variable. In the current research, the potential of three heuristic regression models namely; LSSVR, MARS and M5 model tree were investigated in forecasting and predicting long-term streamflow. The application and analysis were numerically conducted based on four rivers flow, Batman and Garzan Rivers located in Turkey, Euphrates and Tigris Rivers located in Iraq. However, the findings are enumerated as follows.

(i) LSSVR, MARS and M5 tree models outperformed the classical MLR method in both scenarios forecasting and predicting.
(ii) In general, LSSVR indicated better forecasted and predicted accuracies for one-month-ahead over MARS and M5 model tree. Indeed, this is due to the capability of the novel application of least square support vector regression which is developed version of support vector regression via excluding the quadratic programming problem in addition to the skill to capture the complicated non-linear relationship.
(iii) The periodic component feature was embedded and considered within the input combinations of the modeling, the results illustrated that adding this component data was remarkably helpful to provide a detailed intuition into the process of the forecasted and predicted monthly streamflow and improves the accuracy modeling for all the examined rivers.

References

Abrahart RJ, See L (2000) Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrol Process 14:2157–2172. doi:10.1002/1099-1085(20000815/30)14:11/12<2157::AID-HYP57>3.0.CO;2-S

Abrahart RJ, See LM, Dawson CW, et al (2010) Nearly two decades of neural network hydrologic modeling. Adv Data-Based Approaches Hydrol Model Forecast NJ World Sci Publ 267–346.

Abrahart RJ, Anctil F, Coulibaly P, et al. (2012) Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Prog Phys Geogr 36:480–513. doi:10.1177/0309133312444943

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 Hydroinf 14:731. doi:10.2166/hydro.2011.044

Afan HA, El-Shafie A, Yaseen ZM, et al. (2014) ANN Based Sediment Prediction Model Utilizing Different Input Scenarios. Water Resour Manag 29:1231–1245. doi:10.1007/s11269-014-0870-1

Ahmed JA, Sarma AK (2007) Artificial neural network model for synthetic streamflow generation. Water Resour Manag 21:1015–1029. doi:10.1007/s11269-006-9070-y

Bhattacharya B, Solomatine DP (2005) Neural networks and M5 model trees in modelling water level–discharge relationship. Neurocomputing 63:381–396. doi:10.1016/j.neucom.2004.04.016
Box GEP, Jenkins GM (1970) Time Series Analysis, Forecasting and Control, 1st editio. Holden-Day, San Francisco, CA
Cao SG, Liu YB, Wang YP (2008) A forecasting and forewarning model for methane hazard in working face of coal mine based on LS-SVM. J China Univ Min Technol 18:172–176. doi:10.1016/S1006-1266(08)60037-1
Costabile P, Costanzo C, Macchione F, Mercogliano P (2012) Two-dimensional model for overland flow simulations: A case study. Eur Water 38:13–23
Demirbas A, Bakis R (2003) Turkey's water resources and hydropower potential. Energy Explor Exploit 21:405–414
Deng S, Yeh T-H (2010) Applying least squares support vector machines to the airframe wing-box structural design cost estimation. Expert Syst Appl 37:8417–8423
Fletcher R (1987) Practical methods of optimization. John Wiley & Sons.
Friedman JH (1991) Multivariate Adaptive Regression Splines. Ann Stat 19:1–67. doi:10.1214/aos/1176347963
Guo X, Sun X, Ma J (2011) Prediction of daily crop reference evapotranspiration (ET0) values through a least-squares support vector machine model. Hydrol Res 42:268
Hemmati-Sarapardeh A, Shokrollahi A, Tatar A, et al. (2014) Reservoir oil viscosity determination using a rigorous approach. Fuel 116:39–48. doi:10.1016/j.fuel.2013.07.072
Huang Z, Luo J, Li X, Zhou Y (2009) Prediction of effluent parameters of wastewater treatment plant based on improved least square support vector machine with PSO. In: Information Science and Engineering (ICISE), 2009 1st International Conference on. IEEE, pp 4058–4061
Hwang SH, Ham DH, Kim JH (2012) Forecasting performance of LS-SVM for nonlinear hydrological time series. KSCE J Civ Eng 16:870–882. doi:10.1007/s12205-012-1519-3
Ji Z, Wang B, Deng S, You Z (2014) Predicting dynamic deformation of retaining structure by LSSVR-based time series method. Neurocomputing 137:165–172. doi:10.1016/j.neucom.2013.03.073
Kamari A, Nikookar M, Sahranavard L, Mohammadi AH (2014) Efficient screening of enhanced oil recovery methods and predictive economic analysis. Neural Comput & Applic 25:815–824. doi:10.1007/s00521-014-1553-9
Kaygusuz K (1999) Hydropower potential in Turkey. Energy Sources 21:581–588
Kisi O (2012) Modeling discharge-suspended sediment relationship using least square support vector machine. J Hydrol 456-457:110–120. doi:10.1016/j.jhydrol.2012.06.019
Kisi O (2013) Least squares support vector machine for modeling daily reference evapotranspiration. Irrig Sci 31: 611–619
Kisi O, Parmar KS (2016) Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. J Hydrol 534:104–112. doi:10.1016/j.jhydrol.2015.12.014
Leathwick JR, Elith J, Hastie T (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecol Model 199:188–196. doi:10.1016/j.ecolmodel.2006.05.022
Legates DR, McCabe GJ Jr (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 35:233–241
Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. Environ Model Softw 15:101–124. doi:10.1016/S1364-8152(99)00007-9
Nourani V, Hosseini Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. J Hydrol 514:358–377. doi:10.1016/j.jhydrol.2014.03.057
Okkan U, Ali Serbes Z (2013) The combined use of wavelet transform and black box models in reservoir inflow modeling. J Hydrol Hydromechanics 61:112–119. doi:10.2478/johh-2013-0015
Pahasa J, Ngamroo I (2011) A heuristic training-based least squares support vector machines for power system stabilization by SMES. Expert Syst Appl 38:13987–13993
Pal M, Deswal S (2009) M5 model tree based modelling of reference evapotranspiration. Hydrol Process 23: 1437–1443. doi:10.1002/Hyp.7266
Quinlan JR (1992) Learning with continuous classes. In Proceedings of the 5th Australian joint Conference on Artificial Intelligence (Vol. 92, pp. 343–348). http://sci2s.ugr.es/keel/pdf/algorithm/congreso/1992-Quinlan-AI.pdf
Rezaeian-Zadeh M, Tabari H, Abghari H (2013) Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions. Arab J Geosci 6:2529–2537. doi:10.1007/s12517-011-0517-y
Saleh DK (2010) Stream gage descriptions and streamflow statistics for sites in the Tigris River and Euphrates River basins, Iraq. US Department of the Interior, US Geological Survey

Sarangi A, Bhattacharya AK (2005) Comparison of Artificial Neural Network and regression models for sediment loss prediction from Banha watershed in India. Agric Water Manag 78:195–208. doi:10.1016/j.agwat.2005.02.001

Sharda VN, Patel RM, Prasher SO, et al. (2006) Modeling runoff from middle Himalayan watersheds employing artificial intelligence techniques. Agric Water Manag 83:233–242. doi:10.1016/j.agwat.2006.01.003

Sharda VN, Prasher SO, Patel RM, et al. (2008) Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data. Hydrol Sci J-J Des Sci Hydrol 53:1165–1175. doi:10.1623/hysj.53.6.1165

Shokrollahi A, Arabloo M, Gharagheizi F, Mohammadi AH (2013) Intelligent model for prediction of CO2 - Reservoir oil minimum miscibility pressure. Fuel 112:375–384. doi:10.1016/j.fuel.2013.04.036

Shortridge JE, Guikema SD, Zaitchik BF (2015) Empirical streamflow simulation for water resource management in data-scarce seasonal watersheds. Hydrol Earth Syst Sci Discuss 12:11083–11127. doi:10.5194/hessd-12-11083-2015

Singh VP, Cui H (2015) Entropy Theory for Streamflow Forecasting. Environ Process 2:449–460. doi:10.1007/s40710-015-0080-8

Solomatine DP, Dulal KN (2003) Model trees as an alternative to neural networks in rainfall—runoff modelling. Hydrol Sci J 48:399–411. doi:10.1623/hysj.48.3.399.45291

Solomatine DP, Xue Y (2004) M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China. J Hydrol Eng 9:491–501. doi:10.1061/(ASCE)1084-0699(2004)9:6(491)

Sotomayor KAL (2010) Comparison of adaptive methods using multivariate regression splines (MARS) and artificial neural networks backpropagation (ANNB) for the forecast of rain and temperatures in the Mantaro river basin. 58–68.

Suykens JA, Vandewalle J (1999) Least Squares Support Vector Machine Classifiers. Neural Process Lett 9:293–300. doi:10.1023/A

Tao B, Xu W, Pang G, Ma N (2008) Prediction of bearing raceways superfinishing based on least squares support vector machines. In: Natural Computation, 2008. ICNC'08. Fourth International Conference on. IEEE, pp 125–129

Tigkas D, Christelis V, Tsakiris G (2016) Comparative Study of Evolutionary Algorithms for the Automatic Calibration of the Medbasin-D Conceptual Hydrological Model. Environ Process. doi:10.1007/s40710-016-0147-1

Turan ME, Yurdusev MA (2014) Predicting Monthly River Flows by Genetic Fuzzy Systems. Water Resour Manag 28:4685–4697. doi:10.1007/s11269-014-0767-z

Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol 374:294–306. doi:10.1016/j.jhydrol.2009.06.019

Xie G, Wang S, Zhao Y, Lai KK (2013) Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study. Appl Soft Comput 13:2232–2241. doi:10.1016/j.asoc.2013.02.002

Yaseen ZM, El-shafie A, Jaafar O, et al. (2015) Artificial intelligence based models for stream-flow forecasting: 2000–2015. J Hydrol 530:829–844. doi:10.1016/j.jhydrol.2015.10.038

Zhang W, Goh ATC (2014) Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geosci Front 7:45–52. doi:10.1016/j.gsf.2014.10.003