Gorgan University of Agricultural Sciences and Natural Resources

Environmemtal Resources Research (ERR)

Print ISSN: 2783-4832
Online ISSN: 2783-4670

Simulating rainfall-runoff process with a new combined artificial intelligence

Hamidreza Bababali[1*], Reza Dehghani[2]

[1] Assistant Professor, Department of Civil Engineering, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran
[2] PhD in Water Sciences and Engineering, Department of Soil Conservation and Watershed Management, Lorestan Province Agriculture and Natural Resources Research & Education Center, Agricultural Research, Education & Extension Organization, Khorramabad, Iran

Article Info

Article type:
Research Article

Article history:
Received: September 2023
Accepted: April 2024

Corresponding author:
hr.babaali91@gmail.com

Keywords:
Simulation
Artificial Intelligence
Karkheh
Water Resources Management

Abstract

The rainfall-runoff process is one of the most important and complex hydrological phenomena in the management of surface water resources and in taking appropriate measures in the event of floods and droughts. To simulate this process, a proper understanding of the behavior of the basin saves time and plays important role in model selection. To simulate the runoff process of the Karkheh catchment in Iran, statistical models and artificial intelligence approaches—including Multivariate Linear Regression (MLR), Artificial Neural Network (ANN), Support Vector Regression (SVR), and Support Vector Regression-Wavelet (WSVR)—were applied on a daily time scale over the statistical period from 2010 to 2020. To assess simulation performance, statistical indices such as the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Nash-Sutcliffe Efficiency (NSE), and percentage bias (PBIAS) were utilized. Results indicated that the studied models performed better in composite structures, with artificial intelligence models demonstrating lower error rates and superior performance compared to statistical models. Notably, the Wavelet Support Vector Regression model exhibited greater accuracy and reduced error relative to the other models. Overall, the findings suggest that hybrid artificial intelligence models are effective for modeling the runoff process and can serve as a suitable and efficient solution for water resources management.

Cite this article: Bababali, Hamidreza; Dehghani, Reza. 2024. Simulating the Runoff Precipitation Process with New Combined Artificial Intelligence. *Environmental Resources Research*, 12(1), 95-112.

DOI: 10.22069/IJERR.2024.22134.1424
Publisher: Gorgan University of Agricultural Sciences and Natural Resources

Introduction

Nowadays, various factors such as increasing population size and human activity have amplified water demand globally, inducing extensive changes in hydrological aspects of the catchments and overexploitations of rivers. Threats to humans and financial losses occur due to natural disasters such as floods. Various studies underscore that half of the world's countries are at risk of drought. On the other hand, the dangers of torrential rains have engendered the destruction of infrastructure in many parts of the world. In

these areas, the amount of surface runoff has increased due to increasing rainfall, leading to increased base flow in rivers and streams along with soil erosion. Therefore, to prevent, or at least decrease these events, it is crucial to simulate the river's flow, especially in terms of the rainfall-runoff processes (Wang et al. 2013). Correct or appropriate simulation of rainfall-runoff process is indispensable in water resources management. Due to the temporal and spatial variability of catchment features and precipitation patterns, rainfall-runoff is considered as the most complex nonlinear process in hydrology.

In recent years, different methods have been proposed to simulate the rainfall-runoff process, including numerical models. In this technique, the accuracy and efficiency of model may be reduced as it is very difficult to achieve the number of variables in a catchment. Despite being simple in nature and widely used in predicting hydrological data, statistical models have somehow similar disadvantage (Nayak et al. 2004). Due to the constant and limited recording of nonlinear and non-static hydrological data, they may not have desired accuracy (Parisuj et al. 2017). On the other hand, artificial intelligence (AI) method has recently been used to increase the accuracy of simulations in hydrological processes. Regarding AI models, equations and mathematical relationships may not be provided; moreover, the affecting physical variables cannot be effortlessly estimated. The objective of AI models is to develop valid relationships between the measured variables of the hydrological cycle, which are used to solve a hydrological problem. AI models estimate the desired output upon receiving input and performing a series of mathematical operations. The variables and coefficients associated with these models are estimated based on observational input and output (target) data, thus making them dependent on input and output data in terms of both quantity and quality (Adnan et al. 2021; Alizadeh et al. 2017; Nayak et al. 2013; Nourani et al. 2019a, b; Dehghani and Torabi Poudeh 2021; Dehghani et al. 2020a, b; Kesgin et al. 2020a, b; Ouma et al. 2021; Swathi et al. 2020; Tian et al. 2021; Tikhamarine et al. 2020; Wu and Chau 2013; Xiang et al. 2020; Zare and Koch 2018).

Okkan et al. (2021) used AI models to evaluate the runoff process in the Gediz River in western Turkey. They employed two models, namely artificial neural networks (ANNs) and support vector regression (SVR), as well as monthly barometric and hydrometric station data. The study underlined a great performance in the simulation of runoff process. Indeed, a relative advantage in regard to the accuracy of the SVR model was asserted.

Ridwan et al. (2021) evaluated performance of Bayesian Linear Regression (BLR) models, Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR), and Neural Network Regression (NNR) to simulate the Hulu Terengganu catchment runoff process located in Malaysia. Thence, they analyzed the flow and precipitation variables of a hydrometric station on a daily time scale during the statistical period of 2010–2020. Results emphasized that the BDTR model performs the best among all models.

In general, according to the outcomes of the previous relevant research, it is indispensable to provide a solution and ensure a proper prediction of surface water resources to prevent potential droughts all around the world and in Iran as one of the countries with the highest water challenges. In Iran, Karkheh catchment is of crucially importance in terms of drinking and agriculture, such that it is one of the most vital catchments for agricultural production. For the purpose of growth and development, the products of this plain are primarily fed by surface water. Excessive extraction of groundwater through pumping and deep digging in adjacency to rivers have caused a sharp decline in surface water resources in this catchment in recent years. Therefore, identifying the changes in precipitation and runoff are the sine qua non of predicting and taking management measures to improve the prediction efficacies. Therefore, the purpose of this study is to (a) analyze the rainfall-runoff process using climatic variables, e.g. precipitation, and integrated vector

regression models with wavelet transform (WT) as well as new optimization algorithms, e.g. innovative gunner and black widow spider, and (b) compare different modeling approaches, including hybrid and standalone AI techniques and statistical models.

Materials and Methods

Karkheh catchment in Iran was selected as the study area. It is located between 48° 10′ to 50° 21′ E and 31° 34′ to 34° 7′ N with an area of 59143 km^2. The area extends from the western to the central and southwestern Zagros Mountains in the Persian Gulf region. It is bounded by the Sirvan, Sefid-rud, and Ghareh-chai rivers in the north, the Dez catchment area in the west, and parts of the country's western borders in the south. Average annual rainfall in the Karkheh catchment area ranges from 951 mm in the southern lowlands to 9111 mm in the northern highlands and eastern regions. Vegetation is sparse at lower altitudes but increases with elevation. Rainfall is higher in the northern and eastern regions of the catchment, with an average of 48.8% of the total precipitation occurring in winter, 30.6% in autumn, 20.4% in spring, and only 0.2% in summer. Figure 1 shows the geographical position of the selected four stations in Karkheh catchment area, namely Cham Anjir, Kashkan, Pole-zal, and Jologir. Table 1 presents the geographical data on these stations.

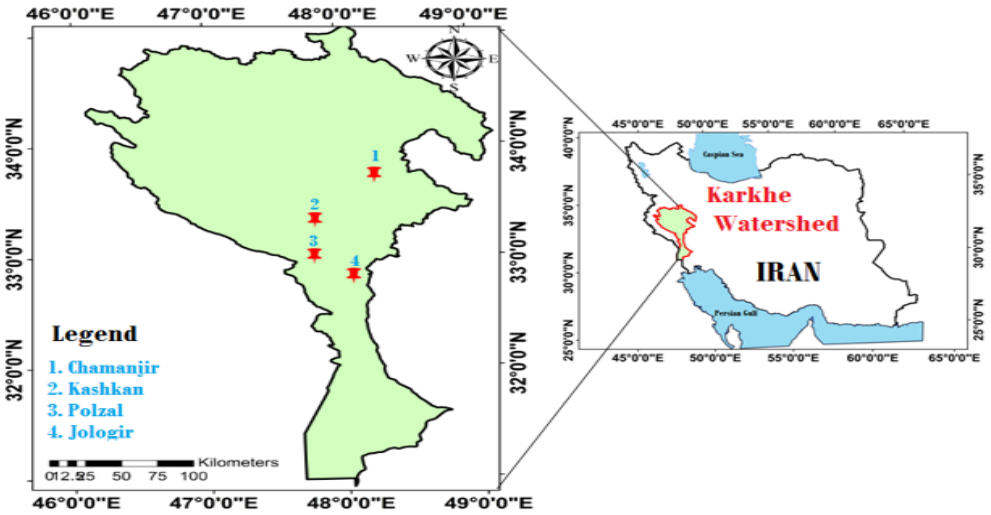


Figure 1. Positions of the selected stations in Karkheh catchment

Table 1. Geographical coordinates of Karkheh stations

Station	Length	Width	Area
Cham Anjir	48º 14′ 38"	32º 26′ 37"	1140
Kashkan	47º 53′ 39"	39º 19′ 52"	820
Pole-zal	48º 9′	32º 25′	90
Jologir	47º 43′	32º 9′	120

Method

To simulate the runoff load process of Karkheh catchment, 4 stations of the basin, constituting the most important hydrometric stations, were selected. Then, different models were defined to simulate the process. The statistical models and artificial intelligence approaches were employed to simulate the runoff precipitation process. In the modeling process, the statistical models were compared with individual models and then, the superior model was converted into a hybrid one with new optimization algorithms. Finally, the performance of the models was measured with different graphs. According to the recent research, different artificial intelligence models are used in the runoff process that are subject to errors. To

reduce the model error, the model setting parameters optimized via meta-heuristic algorithms. Several studies have employed a hybrid model including meta-heuristic algorithms. Here we have employed new algorithms not studied in hydrological or hydrogeological processes before for the purpose of reducing the difficulties and challenges. We have introduced a new algorithm to facilitate the simulation process, predict the runoff in future using dependent parameters, and prevent irreparable damage to Iran's surface water resources. Making effective simulation and predicting of the runoff process is one of the most fundamental and important subjects in Iran's water issues.

Multivariate linear regression

Regression analysis is a statistical method by which the relationship between two or more quantitative variables (independent variables or predictors) is studied to predict the dependent variable (response variable). A multivariate linear regression model is expressed as follows:

$$y = \sum_{i=1}^{N} \beta_i X_i + \varepsilon \quad (1)$$

where β_i is the regression coefficient (regression parameters) X_i the independent variable, ε the eccentricity, and N the number of independent variables. The least squares method is used to estimate the regression coefficients (β_i).

Artificial neural network

Artificial neural networks are widely used in hydrological studies and water resources management (Nourani et al., 2009). The structure of the neural network usually consists of the input layer, the hidden layer, and the output layer. Input layer A is the transmitter layer and a means of providing data. The output layer contains the values predicted by the network and the middle or hidden layer, which consists of processor nodes, is the data processing site. The first viable application of artificial neural networks was detected through the introduction of multilayer perceptron networks. In these networks, it has been proven that among learning algorithms, the error propagation algorithm with feed network structure and three layers is satisfactorily used in solving complex engineering problems and facilitating the simulation and forecasting of hydrological time series (Norani et al., 2011; Tokar and Johnson., 1999). The most common stimulus functions used in reverse diffusion networks are sigmoid and hyperbolic tangent stimulus functions (Hornik, 1988).

Wavelet transform

Wavelet transform is proposed as an alternative method for short-time Fourier transform, and its purpose is to overcome the problems of frequency resolution power in short-time Fourier transform. In wavelet transform, similar to short-time Fourier transform, the signal is divided into windows and the wavelet transform is performed separately on each of these windows (Wang et al., 2000). However, the most important differences are that in wavelet transform, the frequency resolution of the signals or the length of the window changes in accordance with the type of frequency; simultaneously, the width of the window or the frequency scale varies in accordance with the type of frequency. In other words, the role of scale is of higher priority than frequency in wavelet transform. That is, wavelet transform is a form of time-scale transformation. Accordingly, by using wavelet transform, the signal can be expanded on high scales and its details can be analyzed; on low scales, the signal is contracted and the totality of the signal can be examined (Nourani et al, 2009). A wavelet is a small wave, part or window of the main signal whose energy is concentrated in time. By using wavelet transform or analysis, a mother signal or time series can be broken down into wavelets at different resolution levels and scales. Thus, wavelets are either translatable or dilated samples derived from the mother signals that oscillate in a finite length and are strongly damped (Nourani et al., 2018). Based on this important feature of wavelet transform, the timeless and transient time series can be analyzed locally (Shin et al., 2005).

Wavelet transform is defined in two forms of continuous and discrete.
Continuous Wavelet Converter (CWT)
The continuous wavelet transform of the function f (t) is defined through Equations (2) and (3) (Wang et al., 2000).

$$\mathrm{CWT}_f^{\psi}(s,\tau) = \Psi_f^{\psi}(s,\tau) = \frac{1}{\sqrt{|s|}}\int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-\tau}{s}\right)dt = \langle f(t), \psi_{s,\tau}(t)\rangle \quad (2)$$

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \quad (3)$$

Equation (3) is a relation with two variables s and τ denoting the scale parameter (frequency inverse) and the transmission parameter, respectively. The * sign denotes a mixed conjugate. ψ is the function of the mother window or wavelet and 1 / √ (| s |) ψ ((t-τ) / s) represents the wavelets resulting from the transmission and resizing of the mother wavelet (Wang et al., 2000). The word mother is used because all the translated and scaled versions (daughter wavelets) are totally derived from this function. That is, the mother wavelet is a pattern for other windows. The sign 〈…〉 also indicates the multiplication of two functions in the signal space (Nourani et al., 2019a; Karthikeyan and Nagesh Kumar, 2013).

Support Vector Regression

The support vector machine is an efficient learning system based on the theory of constrained optimization which uses the inductive principle of structural error minimization and yields an overall optimal solution (Vapnik, 1995). In the SVR model, a function related to the dependent variable Y, which is itself a function of several independent variables x, is evaluated and measured. Similar to other regression problems, it is assumed that the relationship between independent and dependent variables with an algebraic function like f (x) along with some perturbation (allowable error (ε)) can be determined (Vapnik, 1998) as follows:

$$f(x)=W^T.\emptyset(x)+b \quad (4)$$

$$y=f(x)+\text{noise} \quad (5)$$

where W^T is a transcript of the coefficients, constant *b* belongs to the properties of the regression function, and ∅ is the kernel function based on which the goal is to find a functional form for f (x). This is achieved by training the SVM model through a set of data (training set) (Misra et al., 2009). To calculate W and b, it is necessary to minimize the error function (Equation 6) in the SVM-ε model by considering the conditions (constraints) in Equations (7) and (8) (Hamel, 2009).

$$\frac{1}{2}W^T.W+C\sum_{i=1}^{N}\varepsilon_i+C\sum_{i=1}^{N}\varepsilon_i^* \quad (6)$$

$$W^T.\emptyset(X_i)+b-y_i \le \varepsilon+\varepsilon_i^* \quad (7)$$

$$y_i-W^T.\emptyset(X_i)-b \le \varepsilon+\varepsilon_i, \varepsilon_i^* \ge 0 \quad , \quad i=1,2,\ldots,N \quad (8)$$

In the above equations, C is an integer and a positive number, which determines the penalty when an error occurs in model training. The kernel function is N (the number of instances and the two properties ε_i and ε_i^* are deficient variables). Finally, the SVM regression function can be rewritten as follows:

$$f(x)=\sum_{i=1}^{N}\bar{\alpha}_i\,\emptyset(x_i)^T.\emptyset(x)+b \quad (9)$$

In Equation 11, $\bar{\alpha}_i$ is the mean of the Lagrangian coefficients. Calculating ∅ (x) in its characteristic space can be very complex (Yoon et al., 2011). To solve this problem, the usual procedure in the SVM regression model is to select a kernel function as follows:

$$K(X_J,X)=\emptyset(X_i)^T\sqrt{b^2-4ac} \quad (10)$$

Different kernel functions can be used to build different types of SVM-ε. The types of kernel functions applicable to the SVM regression model include polynomial kernel 1, Radial Base Function (RBF) kernel, and linear kernel, which are respectively calculated by the equations given below. Figure 2 shows the structure of the backup vector machine model. Given that the most widely used kernel functions are radial, linear, and polynomial (Vapnik and Chervonenkis, 1991), this study used these three kernel functions to constitute the vector regression model. The parameters of

the studied kernels C, t, and d converged to the optimal values using new ultra-exploration algorithms and then, the hybrid model was investigated. It is noteworthy that the process of backup vector machine calculations was implemented based on coding in MATLAB environment and the kernel functions were optimized through trial and error.

$$k(x,x_j)=(t+x_i.x_j)^d \quad (11)$$

$$K(x,x_i)=\exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right) \quad (12)$$

$$k(x,x_j)=x_i.x_j \quad (13)$$

Evaluation Criteria and Comparison of Models

In any project, certain criteria are employed to evaluate the efficiency of modeling. In the present study, different statistical criteria were employed to evaluate the efficiency of the models including the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Nash-Sutcliffe Efficiency (NSE) coefficient, and Bias (Chai & Draxler, 2014; Legates & McCabe, 1999). The value of R^2 is in the range of [0-1] and the closer it gets to one, the higher the prediction ability of the model will be, and vice versa. Zero suggests that the model does not define the variations of the response data around the mean value, while one implies that it defines all of them around the mean (Nagelkerke, 1991). Nash-Sutcliffe Efficiency (NSE) coefficient is a normalized statistic that defines the relative value of residual variance in comparison with the variance of the measured data (Nash & Sutcliffe, 1970; Moriasi et al., 2007). The NSE ranges between $-\infty < NSE < 1$ and the higher its value approaches one, the more optimized the answer will be. The values between zero and one are generally accepted as the acceptable performance ratings, while $NSE \ll 0$ suggests that the mean observational values point to higher predictive power than the estimated values, implying unacceptable performance of the model. This criterion is recommended by ASCE (1993) and its use is very common because it provides a vast array of information regarding the reported values (ASCE, 1993). The use of this criterion has been highly welcomed in different scientific fields and numerous researchers throughout the world are benefiting from it (Sevat & Dezetter, 1991; Kesgin et al., 2020). Percentage of Bias (PBIAS) measures the orientation of computational (simulated) data to their smaller or larger observational counterparts (Dabanli & Sen, 2018). The PBIAS value can be positive, negative, or zero. Zero suggests the optimal value and low-magnitude values point to the precision of the model in the simulation process. Positive and negative values denote the underestimation and overestimation of the model, respectively (Gupta et al., 1999). This criterion is favored more by the scholars in this field and applied by nearly all of them. It is also prevalent in most hydrological, water resource management, and geological studies (Musi et al., 2019; Pengxin et al., 2019). The above-mentioned criteria are derived through Equations (15-19) presented here.

Besides the aforementioned statistical criteria, this study applies box plot ad Taylor's diagram criteria, which are among the prevalent graphic approaches to visual comparison of the performance of models. Taylor's diagram is presented in two different forms, namely a semi-circle displaying positive and negative correlations and a quadrant displaying only positive correlation. In both cases, the correlation coefficient values are in the form of the circle's radius on its arch and the standard deviation values are displayed in the form of two concentric circles toward the center of the circle (Taylor, 2001; Pincus et al., 2008; Wehner, 2013). Boxplot was introduced by John Tokey in 1969 and it is one of the most widespread types of diagrams, displaying many descriptive statistics among the data. In other words, this plot facilitates visual comparison of different groups (Conti et al., 2014).

$$R^2=\left[\frac{\sum_{i=1}^{n}(M_{oi}-\bar{M}_0)(M_{ei}-\bar{M}_e)}{\sqrt{\sum_{i=1}^{n}(M_{oi}-\bar{M}_0)^2.\sum_{i=1}^{n}(M_{ei}-\bar{M}_e)^2}}\right]^2, 0\le R^2\le 1 \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_{ei} - M_{oi})^2}\ 0 \leq RMSE \leq +\infty \quad (16)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|M_{ei} - M_{oi}|\ , 0 \leq MAE \leq +\infty \quad (17)$$

$$\text{NSE}=1-\frac{\sum_{i=1}^{n}(M_{ei}-M_{oi})^2}{(M_{ei}-\bar{M}_e)^2}, -\infty < NSE < 1 \quad (18)$$

$$PBIAS = \frac{\sum_{i=1}^{n}(M_{oi}-M_{ei})}{\sum_{i=1}^{n}M_{ei}} \times 100\ , -100 \leq PBIAS \leq 100 \quad (19)$$

Discussion and results

One of the most important steps in modeling is choosing the right combination of input variables. Therefore, first, the correlation between input and output variables was calculated and the input parameters were selected in order to achieve the optimal model for predicting the flow of rivers in the Karkheh catchment, as presented in Table 2. In this table, P (t) rainfall and Q (t) runoff were considered at times t-1, t-2, t-3, t-4. Considering that in the present study, the effect of the flow sequence specific to previous days was considered in predicting the daily flow rate, only the normalized flow rate data with a return sequence of up to 4 days were used as training data in different combinations according to the table. For convenience, they are called patterns. For this purpose, the data obtained from Cham-e-Fig, Kashkan, Pol-e-Zal, and Zalgir hydrometric stations located in Karkheh catchment with 3650 records recorded during the period (2020-2010) on a daily time scale were used. Finally, 2920 records were selected for training and the remaining 730 records for validation of the studied models. It should be noted that 80% of the data for for training and the remaining 20% for testing were randomly selected to cover a wide range of data types (Kisi et al., 2006; Nagy et al., 2002).

Table 2. Combinations of input variables for the best model selection

Number	Input	Output
1	P(t)	Q(t)
2	P(t), P(t-1)	Q(t)
3	P(t), P(t-1), P(t-2)	Q(t)
4	P(t), P(t-1), P(t-2), P(t-3)	Q(t)
5	P(t), P(t-1), P(t-2), P(t-3),Q(t-1)	Q(t)
6	P(t), P(t-1), P(t-2), P(t-3),Q(t-1), Q(t-2)	Q(t)
7	P(t), P(t-1), P(t-2), P(t-3),Q(t-1), Q(t-2), Q(t-3)	Q(t)
8	P(t), P(t-1), P(t-2), P(t-3),Q(t-1), Q(t-2), Q(t-3), Q(t-4)	Q(t)

The models and algorithms used in this research were evaluated with respect to experimental datasets and the highest efficiency was selected for further simulation and analysis. This stage has eight main phases, which are described in Table (3). In simpler terms, these phases are the best input combinations that were selected based on the correlation coefficient. Also, for each model, all eight combinations were used at the training and testing stages. Researchers generally base their evaluation of models on either R2 or RMSE. The main objective of artificial intelligence-based systems is to reduce the estimated error rate; therefore, the criterion for superiority of models is RMSE in this research. It can generally be said that the addition of variables with high CC in determining the output of the model increases the point prediction power. Table (3) shows the selection of the optimal input combination based on RMSE.

As shown in Table (3), according to different structures of each model, the composition of the optimal input variable varies in the models. Also, for each model, RMSE values were calculated for both testing and training sections. In this section, the lowest amount of RMSE was selected in the testing section to comment on the accuracy of the models. As can be seen from Table (3), for all models in all four stations, the eighth model combination

(input combination (8)) had the best performance because this combination had the lowest RMSE. Also, given that the compositional structure or pattern number 8 includes a number of more effective parameters or variables, the error is reduced by the same amount and, therefore, the combination number (8) will be preferable to other combinations.

Table 3. Selection of the optimal input combination based on RMSE

Chamanjir										
Model	Evaluation criterion	Phase	1	2	3	4	5	6	7	8
MLR	RMSE(m^3/s)	training	0.524	0.506	0.488	0.461	0.442	0.425	0.406	0.386
		testing	0.412	0.408	0.397	0.383	0.372	0.366	0.354	0.312
ANN	RMSE(m^3/s)	training	0.321	0.307	0.296	0.281	0.276	0.262	0.249	0.234
		testing	0.252	0.244	0.237	0.225	0.217	0.208	0.198	0.188
SVR	RMSE(m^3/s)	training	0.236	0.228	0.217	0.208	0.196	0.184	0.172	0.168
		testing	0.188	0.172	0.164	0.155	0.142	0.131	0.126	0.115
WSVR	RMSE(m^3/s)	training	0.161	0.152	0.146	0.133	0.121	0.111	0.101	0.092
		testing	0.085	0.074	0.066	0.058	0.047	0.035	0.028	0.016

Kashkan										
Model	Evaluation criterion	Phase	1	2	3	4	5	6	7	8
MLR	RMSE(m^3/s)	training	0.476	0.462	0.451	0.446	0.432	0.418	0.397	0.372
		testing	0.354	0.344	0.337	0.324	0.316	0.308	0.296	0.284
ANN	RMSE(m^3/s)	training	0.347	0.335	0.326	0.314	0.306	0.291	0.284	0.276
		testing	0.268	0.254	0.242	0.236	0.222	0.214	0.204	0.187
SVR	RMSE(m^3/s)	training	0.242	0.236	0.224	0.218	0.207	0.198	0.186	0.175
		testing	0.179	0.168	0.155	0.142	0.135	0.127	0.116	0.108
WSVR	RMSE(m^3/s)	training	0.168	0.152	0.145	0.131	0.123	0.115	0.106	0.096
		testing	0.081	0.075	0.068	0.054	0.048	0.037	0.029	0.018

Polzal										
Model	Evaluation criterion	Phase	1	2	3	4	5	6	7	8
MLR	RMSE(m^3/s)	training	0.482	0.476	0.462	0.448	0.437	0.422	0.408	0.386
		testing	0.375	0.365	0.356	0.342	0.336	0.322	0.309	0.294
ANN	RMSE(m^3/s)	training	0.356	0.345	0.334	0.327	0.314	0.307	0.291	0.288
		testing	0.276	0.261	0.255	0.242	0.236	0.225	0.211	0.198
SVR	RMSE(m^3/s)	training	0.254	0.246	0.232	0.225	0.214	0.206	0.192	0.186
		testing	0.188	0.175	0.166	0.152	0.141	0.132	0.128	0.111
WSVR	RMSE(m^3/s)	training	0.174	0.162	0.155	0.141	0.132	0.124	0.115	0.101
		testing	0.097	0.086	0.075	0.064	0.058	0.043	0.033	0.025

Jologir										
Model	Evaluation criterion	Phase	1	2	3	4	5	6	7	8
MLR	RMSE(m^3/s)	training	0.472	0.464	0.452	0.441	0.432	0.421	0.401	0.384
		testing	0.366	0.654	0.342	0.337	0.324	0.317	0.308	0.291
ANN	RMSE(m^3/s)	training	0.342	0.336	0.327	0.315	0.309	0.297	0.286	0.274
		testing	0.268	0.254	0.246	0.237	0.226	0.214	0.203	0.193
SVR	RMSE(m^3/s)	training	0.245	0.233	0.228	0.215	0.208	0.196	0.187	0.172
		testing	0.178	0.166	0.157	0.142	0.134	0.121	0.111	0.103
WSVR	RMSE(m^3/s)	training	0.168	0.152	0.143	0.131	0.125	0.114	0.106	0.096
		testing	0.086	0.075	0.064	0.055	0.042	0.037	0.025	0.014

Model performance evaluation

In this study, for the four stations and for all models studied (MLR, ANN, SVR, WSVR), scatter plots and graphs of time changes concerning the observational and computational data associated with the four stations were respectively given in Figure 4. The results demonstrated that for all the stations studied, the WSVR model had the highest accuracy and the lowest error. Also, upon comparing the WSVR model with the SVR model, it was found that the new optimizer model outperformed the SVR model with good accuracy. After selecting the best input combination for each model and plotting the scatter plots and graphs of time changes concerning the observational and computational data for the four stations, the performance statistics of the studied station models for the test data are shown in Table (4). Therefore, in summary, after selecting the best input combination for each model, predictive models for flow simulation were studied in four stations. The performance statistics of WSVR model in all the studied stations at the validation stage demonstrated that this hybrid model enjoyed the following evaluation criteria (PBAIS = 0.001, NSE = 0.977-0.983, MAE = 0.008-0.009, RMSE = 0.014-0.016, R2 = 0.968- 0.973). Also, the BIAS value was positive for all stations, meaning that that the model was underestimated.

Figure 4 shows the time series variation diagram and the distribution of observational and computational values. This figure illustrates that WSVR and SVR models experienced less errors maximum and minimum estimations. In addition, SVR had favorable accuracy in estimation of median and minimum values, while the statistical model of MLR and ANN had poor performance.

Box diagrams were used for visual analysis and evaluation of the models used in the research. The closer the predicted value to the observational value is in terms of correlation coefficient and standard deviation, the higher the predictability will be (Sigaroodi et al., 2014). The advantage of a box chart is that it can demonstrate how a model predicts maximum, median, and square values.

The diagram of the runoff precipitation process box in Figure 5 shows that the WSVR model is of suitable fit with the maximum observed runoff. Also, ANN and MLR have the least compatibility. The same result was observed upon predicting the minimum observed runoff. These two results show that although SVR is one of the smart and accurate models, it cannot predict maximum values well. However, when combined with hybrid algorithm such as W, SVR's performance in predicting maximum values is greatly enhanced. The WSVR models predict the values seen in the third quarter better than the other models and the SVR model predicts the values in the first quarter better.

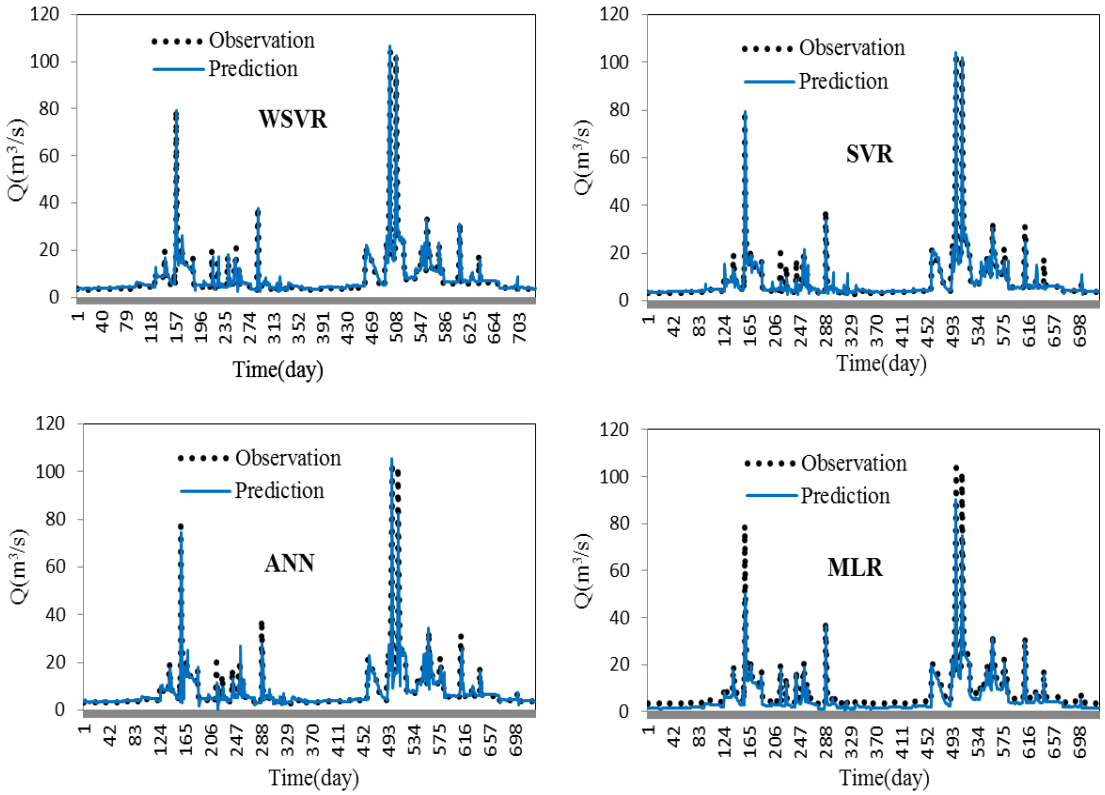
Table 4. Performance evaluation of the studied models for Karkheh catchment stations

Chamanjir										
Model	Training					testing				
	R^2	RMSE	MAE	NSE	PBIAS	R^2	RMSE	MAE	NSE	PBIAS
MLR	0.83	0.386	0.242	0.87	0.004	0.865	0.312	0.168	0.898	0.003
ANN	0.871	0.234	0.125	0.908	0.003	0.91	0.188	0.144	0.922	0.003
SVR	0.92	0.168	0.134	0.93	0.003	0.952	0.115	0.086	0.963	0.002
WSVR	0.956	0.092	0.045	0.96	0.002	0.968	0.016	0.008	0.977	0.001

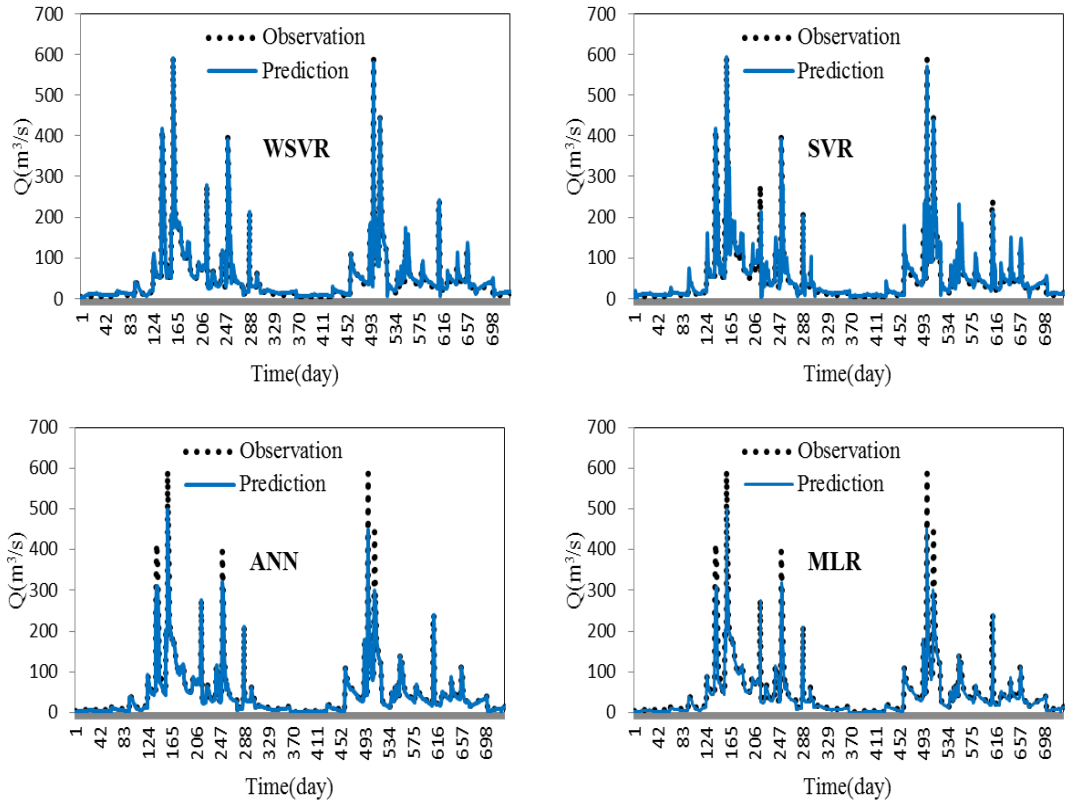
Kashkan										
Model	Training					testing				
	R^2	RMSE	MAE	NSE	PBIAS	R^2	RMSE	MAE	NSE	PBIAS
MLR	0.842	0.372	0.16	0.887	0.004	0.882	0.284	0.142	0.908	0.003
ANN	0.884	0.276	0.134	0.917	0.003	0.921	0.187	0.095	0.936	0.003
SVR	0.927	0.175	0.086	0.938	0.003	0.957	0.108	0.057	0.968	0.002
WSVR	0.96	0.096	0.041	0.964	0.002	0.974	0.018	0.007	0.984	0.001

Polzal										
Model	Training					testing				
	R^2	RMSE	MAE	NSE	PBIAS	R^2	RMSE	MAE	NSE	PBIAS
MLR	0.836	0.386	0.163	0.886	0.004	0.877	0.294	0.143	0.901	0.003
ANN	0.872	0.288	0.144	0.914	0.003	0.918	0.198	0.098	0.928	0.003
SVR	0.923	0.186	0.095	0.936	0.003	0.955	0.111	0.061	0.966	0.002
WSVR	0.955	0.101	0.052	0.961	0.002	0.972	0.025	0.012	0.981	0.001

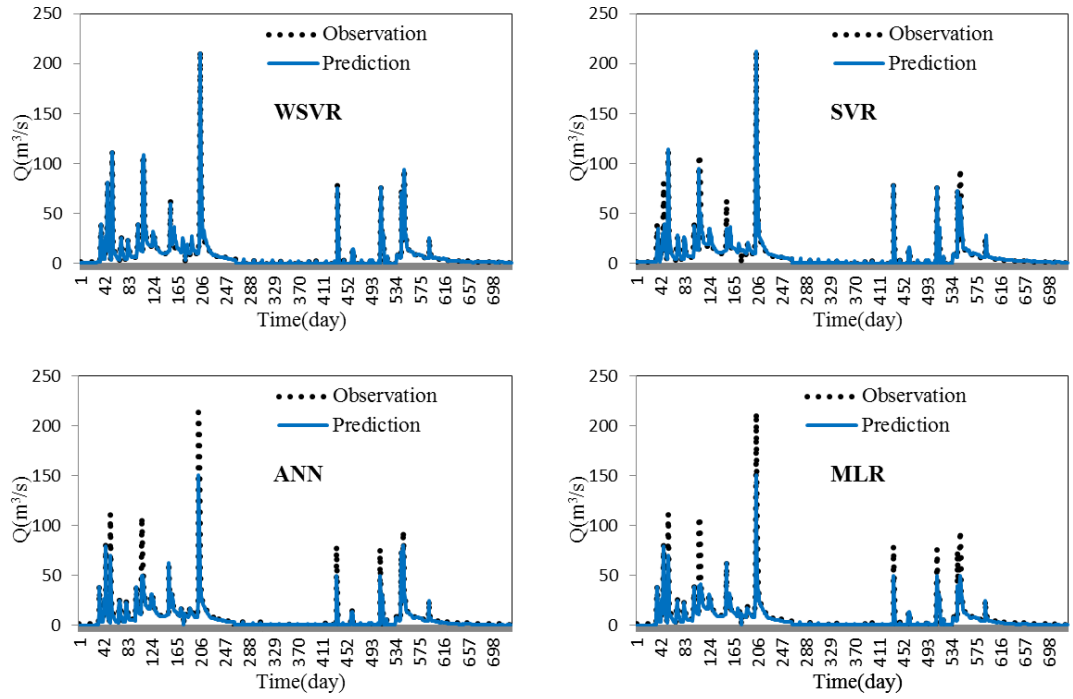
Jologir										
Model	Training					testing				
	R^2	RMSE	MAE	NSE	PBIAS	R^2	RMSE	MAE	NSE	PBIAS
MLR	0.84	0.384	0.161	0.887	0.004	0.88	0.291	0.142	0.905	0.003
ANN	0.88	0.274	0.138	0.915	0.003	0.92	0.193	0.096	0.93	0.003
SVR	0.925	0.172	0.088	0.937	0.003	0.956	0.103	0.058	0.967	0.002
WSVR	0.957	0.096	0.045	0.963	0.002	0.973	0.014	0.009	0.983	0.001



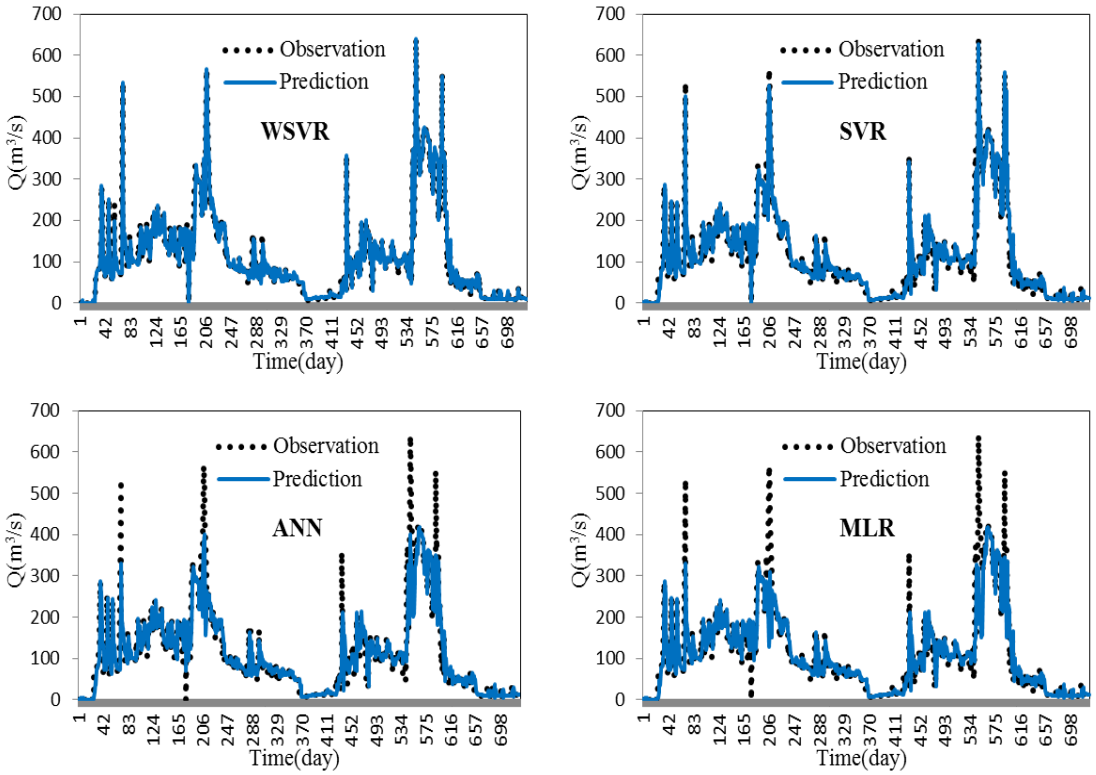
a) Chamanjir



b) Kashkan



c) Polzal



d) Joligir

Figure 4. Scatter diagram and time changes for observational and computational data for the four hydrometric stations under study a) Chamanjir b) Kashkan c) Polzal d) Jologir

According to Figure 5, at Chamanjir station, the SVR models in the first quarter corresponded well to the observational values, and the ANN and WSVR models corresponded to the values in the third quarter, while in the middle values of all models except, the MLR model worked well.

At Kashkan station, in the first quarter, WSVR model corresponded well to the observational values; SVRs functioned properly, while MLR model performed poorly in the case of the mentioned station.

At Polzal station, in the first quarter, the WSVR model corresponded well to the observational values; however, in the third quarter, only the WSVR and SVR models performed well, while the MLR model performed poorly in the case of this station.

At Jologir station, in the first quarter, the results obtained by WSVR models were in good agreement with the observational values; however, in the third quarter, only the WSVR, SVR, and ANN models performed well. while the MLR model performed poorly in this station.

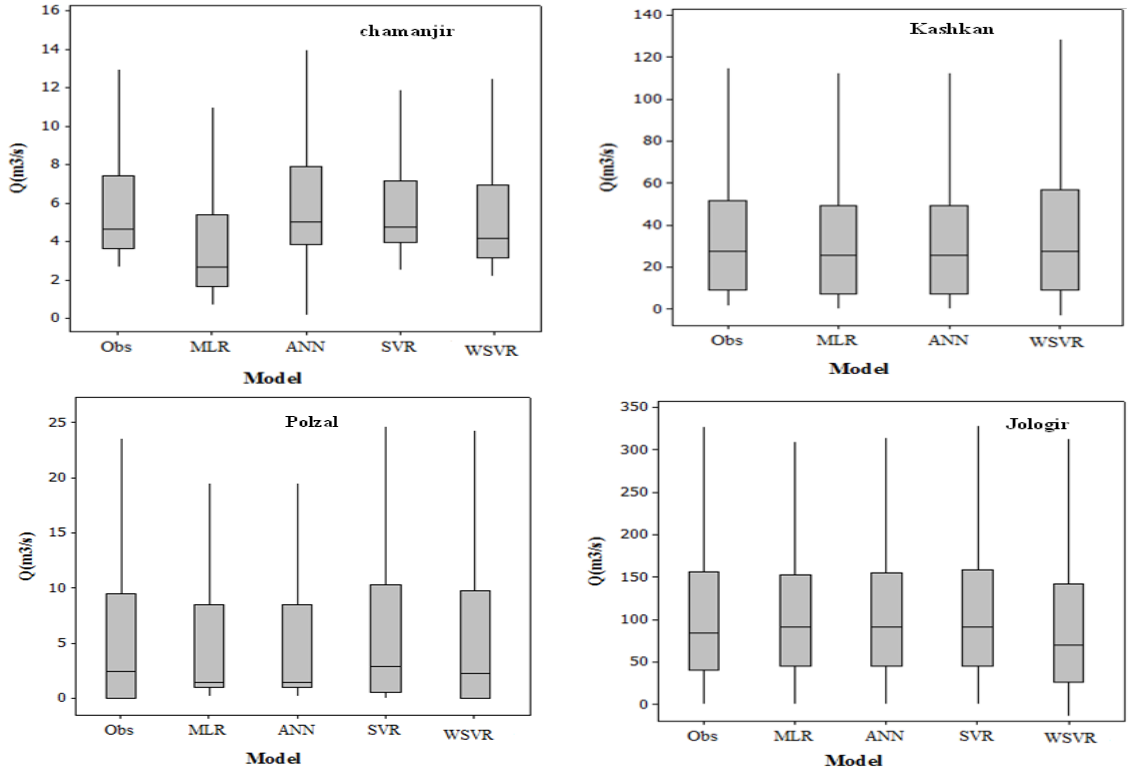


Figure 5. Box plot for the measured and predicted values

Conclusion

In general, the models developed for all the studied hydrometric stations achieved satisfactory prediction results. However, the comparison of model performance across the four stations in the Karkheh catchment revealed that the models for the Kashkan station outperformed those at the other stations. This comparative advantage is attributed to less skewed data, accurate measurement of observational parameters, operator precision, and high data quality. Additionally, the results indicated that the more effective the dependent variables were, the better the network performance. Furthermore, increasing the input to the network corresponded with improved efficiency and accuracy of the model. Also, according to the evaluation criteria, it was concluded that all the six models could estimate the runoff precipitation process with relatively high accuracy. Meanwhile, WSVR model exhibited greater accuracy and less error than MLR, ANN and SVR models.

Also, according to the evaluation criteria, it was concluded that all the six models could estimate the runoff precipitation process with a relatively high accuracy. Meanwhile, WSVR model had exhibited great accuracy and less error than MLR, ANN and SVR models.

Overall, the results of this study point to the superiority of the Wavelet model to other algorithms (based on correlation and RMSE criteria). This model provided the best possible response in the case of all the studied stations and had the best accuracy and the highest predictive power. This superiority is rooted in the powerful internal structure of this algorithm and the use of primary and secondary parameters, cost reduction function, and time saving in achieving an optimal solution and a more effective convergence, which has made the weights the most optimal convergence value.

Also, due to the powerful structure of the Wavelet algorithm, it is possible to ensure convergence to the optimal answer and local minima. The performance of these secondary parameters along with other factors reduces the search amplitude, resulting in a better and faster convergence,

because the more limited the search amplitude is, the faster and more accurate the search and convergence will be. Overall, this study shows that the use of WSVR model can be effective in the runoff process. Also, this model can be useful in facilitating the development and implementation of surface water management strategies. This attempt will be quite fruitful in management decisions in order to increase the volume and quantity of surface water resources.

Acknowledgments
The authors thank Lorestan Regional Water Company, Iran for participating in the collection of data needed to do the job.

Conflict of Interest
The authors declare that they have no conflict of interest.

Funding
The University of Lorestan Khorramabad Iran supported our research work (Grant No. 1).

Author's Contribution
The authors include Dr. Reza Dehghani consistently participated in the preparation of this article.

Availability of data and material
The data and material availability will be made available to researchers after receiving the email
Code availability
The code availability will be made available to researchers after receiving the email

Ethics approval
This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate
This study is exempt from Lorestan Regional Water Company. Based on the fact that this type of study is a type of inhuman research and the need for informed consent was waived.

Consent for publication
The author agrees to publish the article in the above publication

References

Adnan, R.M., Petroselli, A., Heddam, S., Guimarães Santos, C.A., and Kisi, O. 2021. Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model. Stochastic Environmental Research and Risk Assessment. 35,597–616.

Adnan, R.M., Petroselli, A., Heddam, S., Guimarães Santos, C.A., and Kisi, O. 2021. Comparison of different methodologies for rainfall–runoff modeling: machine learning vs conceptual approach. Natural Hazards. 105, 2987–3011.

Alizadeh, M.J., Kavianpour, M.R., Kisi, O., and Nourani, V. 2017. A new approach for simulating and forecasting the rainfall-runoff process within the next two months. Journal of Hydrology. 548,588–597

ASCE. 1993. Criteria for evaluation of watershed models. Journal of Irrigation and Drainage Engineering. 119(3), 429-442.

Chai, T., and Draxler, R.R. 2014. Root mean square error (RMSE) or mean absolute error (MAE) -Arguments against avoiding RMSE in the literature. Geoscientific Model Development. 7, 1247–1250.

Chou, C.M. 2014. Random modelling of daily rainfall and runoff using a seasonal model and wavelet denoisin. Mathematical Problems in Engineering.

Dabanlı, İ., and Şen, Z. 2018. Precipitation projections under GCMs perspective and Turkish Water Foundation (TWF) statistical downscaling model procedures. Theoretical and Applied Climatology.132 (1), 153-166.

Danandeh Mehr, A., Nourani, V., Hrnjica, B., and Molajou, A. 2017. A binary genetic programing model for teleconnection identification between global sea surface temperature and local maximum monthly rainfall events. Journal of Hydrology. 555,397-406

Dehghani, R., and Torabi Poudeh, H. 2021. Applying hybrid artificial algorithms to the estimation of river flow: a case study of Karkheh catchment area. Arabian Journal of Geosciences.14, 768.

Dehghani, R., Torabi Poudeh, H., Younesi, H., and Shahinejad, B. 2020. Daily Streamflow Prediction Using Support Vector Machine-Artificial Flora (SVM-AF) Hybrid Model. Acta Geophysica. 68(6), 51-66

Dehghani, R., Torabi Poudeh, H., Younesi, H., and Shahinejad, B. 2020. Forecasting Daily River Flow Using an Artificial Flora–Support Vector Machine Hybrid Modeling Approach (Case Study: Karkheh Catchment, Iran), Air, Soil, and Water.14, 22-35

Deo, R.C., and Sahin, M. 2016. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environmental Monitoring and Assessment. 188(2), 90-108

Duie Tien, B., Khosravi, K., Tiefenbacher, J., Nguyen, H., and Kazakis, N.2020. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Journal of Science of the Total Environment. 721,136612.

Gocić, M., Motamedi, S., Shamshirband, S., Petković, D., Ch, S., Hashim, R., and Arif, M. 2015. Soft computing approaches for forecasting reference evapotranspiration. Computers and Electronics in Agriculture. 113(4), 164–173.

Gupta, H.V., Sorooshian, S., and Yapo, P.O. 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. Journal of Hydrologic Engineering. 4(2), 135-143.

Hamel, L. 2009. Knowledge Discovery with Support Vector Machines, Hoboken, N.J. John Wiley.

Hayyolalam, V., and Pourhaji Kazem, A.A. 2020. Black Widow Optimization Algorithm: A novel meta-heuristic approach for solving engineering optimization problems. Engineering Applications of Artificial Intelligence, 87(4).

Hornik, K. 1988. Multilayer feed-forward networks are universal approximators. Neural Networks. 2 (5), 359–366.

Hossain, S., Hewa, G. A., and Wella-Hewage, S. 2019. A Comparison of continuous and event-based rainfall–runoff (RR) modelling using EPA-SWMM. Water. 11(3), 611.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., and Lou, Z. 2018. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water. 10, 1543.

Karthikeyan, L., Nagesh Kumar, D. 2013. Predictability of nonstationary time series using wavelet and EMD based ARMA models. Journal of Hydrology. 502, 103–119.

Kesgin, E., Agaccioglu, H., and Dogan, A. 2020. Experimental and numerical investigation of drainage mechanisms at sports fields under simulated rainfall. Journal of Hydrology.580, 124251.

Kesgin, E., Agaccioglu, H., and Dogan, A. 2020. Experimental and numerical investigation of drainage mechanisms at sports fields under simulated rainfall. Journal of Hydrology, 580,

Khosravi, K., Nohani, E., Maroufinia, E., and Pourghasemi, H.R. 2016. A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights of evidence bivariate statistical models with multi-criteria method. Natural Hazards. 83(2),1-41

Kisi, O., Karahan, M., and Sen, Z. 2006. River suspended sediment modeling using fuzzy logic approach. Hydrology Process. 20, 2. 4351-4362.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. 2018. Rainfall-Runoff modelling using Long-Short-Term-Memory (LSTM) networks. Hydrology and Earth System Sciences. 22, 6005–6022.

Legates, D.R., and McCabe, G.J. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research. 35, 233–241.

Lima, A.R., Cannon, A.J., and Hsieh, W.W. 2016. Forecasting daily streamflow using online sequential extreme learning machines. Journal of Hydrology. 537,431–443

Liu, Z., and Todini, E. 2002. Towards a comprehensive physically-based rainfall-runoff model. Hydrology and Earth System Sciences Discussions. 6(5), 859- 881

Lo Conti, F., Hsu, K.L., Noto, L.V., and Sorooshian, S. 2014. Evaluation and comparison of satellite precipitation estimates with reference to a local area in the Mediterranean Sea. Atmospheric Research. 138, 189-204.

Misra, D., Oommen, T., Agarwal, A., Mishra, S.K., and Thompson, A.M. 2009. Application and analysis of support vector machine based simulation for runoff and sediment yield. Biosystems Engineering. 103(3), 527–535

Moriasi, D.N., Arnold, G.J., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith, T.L. 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, American Society of Agricultural and Biological Engineers. 50(3), 885−900

Musie, M., Sen, S., and Srivastava, P. 2019. Comparison and evaluation of gridded precipitation datasets for streamflow simulation in data scarce watersheds of Ethiopia, Journal of Hydrology. 579.

Nagelkerke, N.J.D. 1991. A note on a general definition of the coefficient of determination. Biometrika. https://doi.org/10.1093/biomet/78.3.691.

Nagy, H., Watanabe, K., and Hirano, M. 2002. Prediction of sediment load concentration in rivers using artificial neural network model. Journal of Hydraulics Engineering. 128: 3. 558-559.

Nash, J.E., and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models: Part 1. A discussion of principles Journal of Hydrology. 10(3), 282-290.

Nayak, P.C., Sudheer, K.P., Rangan, D.M., and Ramasastri, K.S. 2004. A neuro-fuzzy computing technique for modeling hydrological time series. Journal of Hydrology. 291(1-2), 52-66

Nayak, P.C., Venkatesh, B., Krishna, B., and Jain, S.K. 2013. Rainfall-runoff modeling using conceptual, data driven, and wavelet based computing approach. Journal of Hydrology. 493, 57–67

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

Nourani, V., Fard, A. F., Niazi, F., Gupta, H. V., Goodrich, D. C., and Kamran, K. V. 2015. Implication of remotely sensed data to incorporate land cover effect into a linear reservoir-based rainfall–runoff model. Journal of Hydrology. 529, 94–105.

Nourani, V., Kisi, Ö., and Komasi, M. 2011. Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. Journal of Hydrology.402 (1–2), 41–59.

Nourani, V., Komasi, M., and Mano, A. 2009. A multivariate ANN-wavelet approach for rainfall-runoff modeling. Water Resources Management. 23, https://doi.org/ 10.1007/s11269-009-9414-

Nourani, V., Molajou, A., Tajbakhsh, A.D., and Najafi, H. 2019a. A wavelet based data mining technique for suspended sediment load modeling. Water Resources Management. 33, 1769-1784.

Nourani, V., Tajbakhsh, A.D., and Molajou, A. 2018. Data mining based on wavelet and decision tree for rainfall-runoff simulation. Hydrology Research. 50, 75–84.

Nourani,V., Alami, M, T., and Aminfar, M.H. 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Engineering Applications of Artificial Intelligence. 22(2), 466–472.

Okkan, U., BerilErsoy, Z., Kumanlioglu, A., and Fistikoglu, O. 2021. Embedding machine learning techniques into a conceptual model to improve monthly runoff simulation: A nested hybrid rainfall-runoff modeling. Journal of Hydrology.598.

Ouma, Y.O., Cheruyot, R. and Wachera, A.N. 2021. Rainfall and runoff time-series trend analysis using LSTM recurrent neural network and wavelet neural network with satellite-based meteorological data: case study of Nzoia hydrologic basin. Complex & Intelligent Systems. https://doi.org/10.1007/s40747-021-00365-2

Parisuj, P., Goharnejad, H., and Moazami, S. 2017. Rainfallrunoff hydrologic simulation using adjusted satellite rainfall algorithms, a case study: Voshmgir Dam Basin, Golestan. Iran-Water Resources Research. 14(3), 140-159

Pengxin, D., Zhang, M., Bing, J., Jia, J., and Zhang, D. 2019. Evaluation of the GSMaP_Gauge products using rain gauge observations and SWAT model in the Upper Hanjiang River Basin, Atmospheric Research. 219,153-165.

Ramana, R.V., Krishna, B., Kumar, S.R., and Pandey, N.G. 2013. Monthly rainfall prediction using wavelet neural network analysis. Water Resources Managment. 27, 3697 3711.

Ridwan, W., Sapitang, M., Aziz, A., FaizalKushiar, k., NajahAhmed, A., and El-Shafie, A. 2021. Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. Ain Shams Engineering Journal. 12(2), 1651-1663,

FSevat, E., and Dezetter, A. 1991. Selection of calibration objective functions in the context of rainfall-runoff modeling in a Sudanese savannah area. Hydrological Sciences Journal. 36(4), 307-330.

Shin, S., Kyung, D., Lee, S., Taik Kim, J., and Hyun, J. 2005. An application of support vector machines in bankruptcy prediction model. Expert Systems with Applications. 28(4), 127-135.

Sigaroodi, S.K., Chen, Q., Ebrahimi, S., Nazari, A., and Choobin, B. 2014. Long-term precipitation forecast for drought relief using atmospheric circulation factors: a study on the Maharloo Basin in Iran. Hydrology and Earth System Sciences. 18, 1995–2006

Swathi, V., Raju, K.S., and Varma, M.R.R. 2020. Addition of overland runoff and flow routing methods to SWMM—model application to Hyderabad, India. Environmental Monitoring and Assessment. 192, 643-655.

Tian, D., He, X., Srivastava, P., and Kalin, L. 2021. A hybrid framework for forecasting monthly reservoir inflow based on machine learning techniques with dynamic climate forecasts, satellite-based data, and climate phenomenon information. Stochastic Environmental Research and Risk Assessment.

Tikhamarine, Y., Souag-Gamane, D., NajahAhmed, A., Sammen, S., Kis,i O., FengHuang, Y., and El-Shafie, A. 2020. Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization. Journal of Hydrology.589,

Tokar, A.S., and Johnson, P.A. 1999. Rainfall- Runoff modeling using artificial neural

Vapnik, V., and Chervonenkis, A. 1991. The necessary and sufficient conditions for consistency in the empirical risk minimization method. Pattern Recognition and Image Analysis. 1(3), 283-305.

Vapnik, V.N. 1995. The nature of statistical learning theory. Springer, New York, Pp: 250-320.

Vapnik, V.N. 1998. Statistical learning theory. Wiley, New York, Pp: 250-320.

Wang, D., Safavi, A.A., and Romagnoli, J.A. 2000. Wavelet-based adaptive robust M-estimator for non-linear system identification. AIChE Journal. 46(4), 1607-1615.

Wang, W.C., Xu, D.M., Chau, K.W., and Chen, S. 2013. Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. Journal of Hydroinformatics. 15(4), 1377-1390

Wehner, M.F. 2013. Very extreme seasonal precipitation in the NARCCAP ensemble: model performance and projections, Climate Dynamics. 40(1-2), 59-80.

Worland, S.C, Farmer, W.H., and Kiang, J.E. 2018. Improving predictions of hydrological low-flow indices in ungaged basins using machine learning. Environmental Modelling and Software. 101,169–182

Wu, C.L., and Chau, K.W. 2013. Prediction of rainfall time series using modular soft computingmethods. Engineering Applications of Artificial Intelligence. 26(3), 997–1007

Xiang, Z., Yan, J., and Demir, I. 2020. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resources Research. 56(1), 1–17.

Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O., and Lee, K.K. 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. Journal of Hydrology. 396(4), 128–138

Zare, M., and Koch, M. 2018. Groundwater level fluctuations simulation and prediction by ANFIS- and hybrid wavelet-ANFIS/fuzzy C-means (FCM) clustering models: Application to the Miandarband plain. Journal of Hydro-environment Research. 18, 63–76.