Calibration of SWAT and three data-driven models for monthly stream flow simulation in Maharlu Lake Basin

Fatemeh Moazami Goudarzi, Amirpouya Sarraf, Hassan Ahmadi

ABSTRACT

In this study, the performance of SWAT hydrological model and three computational intelligence methods used to simulate river flow are investigated. After collecting the data required for all models used, the calibration and validation stages were performed. Using the SWAT model and three methods of the Extreme Machine Learning (EML), the Support Vector Regression (SVR), and the Least Squares Support Vector Regression (LSSVR), Maharlu Lake Basin stream flow was simulated and the results obtained at Shiraz station were used for this study. A noise reduction filter was employed to improve the results from the computational intelligence methods, and SUFI-2 algorithm was used to analyze the uncertainty of the SWAT model. Finally, in order to evaluate the models developed and the SWAT model, three statistics (RMSE), (R²), and (NS) coefficient were used. The results indicated that the SWAT model and the machine learning models were generally appropriate tools for daily flow modeling, but the LSSVR model showed less errors in both learning and testing, with the coefficients NS = 0.997 and R² = 0.997 in the calibration stage and NS = 0.994 and R² = 0.994 in the validation stage, which prove their better performance compared to the other methods and the SWAT model.

KEY WORDS: ELM, hydrological modeling, LSSVR, SUFI-2, SVR, SWAT

HIGHLIGHTS

- In this study, we investigated the performance of SWAT hydrological model and three computational intelligence methods for simulating a river flow.
- The calibration and validation stages were performed after collecting the data required for all models.
- The study used the SWAT model and three methods of Extreme Learning Machine (ELM), Support Vector Regression (SVR) and Least Squares Support Vector Regression (LSSVR).

INTRODUCTION

Stream flow is one of the most important variables in the hydrological cycle (Jimeno-Sáez et al. 2018). The estimation of accurate stream flow is required for the estimation of floods, development of agricultural strategies, and planning hydraulic structures (Vilaysane et al. 2015; Jimeno-Sáez et al. 2018).

There are several models for hydrological modeling of catchments and the appropriate model is selected considering the capabilities of such models and the topographic and climatic conditions of the area concerned. A sophisticated mathematical model is the Soil and Water Assessment Tool (SWAT) (Arnold et al. 1998; Jimeno-Sáez et al. 2018). The SWAT model is a semi-distributed, physically-based river basin model that is computationally efficient and operates on a continuous time-step that simulates the processes (water movement, sediment, plant growth, food cycle, etc.) rather than using the regression equations. In this regard, the model will require specific climatic information (soil properties, topography, and type of land vegetation). To facilitate this process, the SWAT-CUP (Calibration and Uncertainty Procedures) model has been developed. The performance of the SWAT Hydrological model was tested on Sher River at Belkheri in Narsimhpur District of Madhya Pradesh, India; the results showed that the performance of the SWAT model in simulating streamflow at Belkheri gauging site can be rated as very good and the calibrated model can be used for the runoff simulation of this agriculture-dominated watershed.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).
On the other hand, the nonlinear property, the lack of inherent certainty of the process, the need for extensive information, and the complexity of these models have led researchers to use common models such as artificial neural networks and new methods such as machine learning methods.

Over the past two decades, the machine learning methods have emerged as a powerful computing system for highly complex and nonlinear systems. These methods provide a relatively flexible and fast technique for modeling and can, thanks to their parallel architecture, make the nonlinear behavior of the system understandable and then estimate the values of the objective function accurately.

The Extreme Machine Learning (EML) Model is a high rate generalization performance learning technique that basically employs the Single hidden-Layer Feedforward Neural (SLFN) network (Huang et al. 2004, 2006). The Extreme Machine Learning model is easy to use and requires no parameter other than a predefined network architecture, so this model has implications such as learning rate, learning period, and local minimum where the slope-based algorithms encountered are far away (Deo et al. 2016).

The EML model not only learns much faster than traditional gradient-based learning algorithms but also avoids many complexities of the algorithms. Therefore, due to the its significant advantages such as simplicity, generalization performance and high implementation rate, many researchers in various research fields have focused on this model (Rezaeenour et al. 2019). Support vector regression (SVR), which is based on the VC-theory (Vapnik 1998), has recently emerged as an effective method for prediction problems. The SVR is a mathematical framework that estimates dependency by learning from finite samples. The SVR minimizes the estimation errors and the complexity of the model, simultaneously. This framework provides a good generalization ability and leads the model to be less prone to over-fitting. Considering these good properties, the SVR has attracted more and more attention in machine condition prognostics in recent years (Qu & Zuo 2012). The type of SVR depends on the loss function used. The standard SVR employs the ε-insensitive loss function which constructs the resultant regression model with selected samples known as Support Vectors. Another SVR called the Least Square SVR (LSSVR) employs the squared loss function to construct the resultant regression model, but it uses the entire samples (Qu & Zuo 2012). The Least Squares Support Vector Regression (LSSVR) is a machine learning technology based on a structure risk minimization (Suykens & Vandewalle 1999; Zhang et al. 2018). The LSSVR transforms the localization problem into the constraint optimization problem of the quadratic programming, which can greatly reduce the computational complexity and improve the computational efficiency. Meanwhile, the LSSVR uses all learning samples as support vectors, which can greatly improve the regression modeling ability of the learning samples (Zhang et al. 2018).

Huang et al. (2014) predicted the monthly flow of the river at Huaxi station using the support vector machine, and their results showed that the support vector machine model has high accuracy in predicting the monthly flow of the river. Ghorbani et al. (2016) used the support vector machine models and the artificial neural network to predict the daily flow of Cypress River in Texas and found that the support vector machine model had a better performance than the neural network model did in terms of the river flow prediction and it also had good accuracy. By examining some of the norm computing methods based on machine learning including the Least Squares Regression Support Vector Machine (LSSVR), the artificial neural network (RBFNN), the Multi-Linear regression (MLR), and the decision tree models such as CART and M5 for predicting Kopili river suspended load in India, Kumar et al. (2016) introduced the LSSVR model as the most accurate method for estimating the above values.

Few studies have been conducted aimed at comparing the SWAT and the data-driven models in the past decade. Jajarmizadeh et al. (2015) performed monthly stream flow estimations for southern Iran using the SWAT and the SVM models. Although high accuracy levels were achieved with both models, the SVM model was found to be more successful in the estimation study. Noori & Kalin (2016) employed the SWAT and ANN models in order to perform stream flow estimations at 29 watersheds located around Atlanta during hot and cold seasons. At the end of the study, higher accuracy was attained in hot seasons than in cold seasons (Koycegiz & Buyukyildiz 2019). Alkhamifar & Sarraf (2020) investigated a hybrid wavelet-transform-based extreme learning machine (ELM) model predicting GWL. Also, two other popular models – a wavelet-transform based artificial neural network and a wavelet-transform-based adaptive neuro-fuzzy interference system – were used to evaluate the model. The wavelet-transform-based ELM model outperformed the other two models with a correlation coefficient of 0.983 during a 1-month period. The model was also superior
to the others in terms of training and testing speeds. Donyaii & Sarraf (2020) evolved the Multi-Objective Coronavirus Optimization Algorithm (MOCVOA) to obtain the operational optimum rules of Voshmgir Dam reservoir under climate change conditions. The climatic variables downscaled and predicted by the Bias Correction Spatial Disaggregation (BCSD) method of the MIROC-ESM model were introduced into the Extreme Learning Machine (ELM) model to evaluate the future runoff flowing into the reservoir. Donyaii et al. (2020) re-evaluated the management crisis of Iran's groundwater resources from the perspective of the administrative, social, and legal system with the GRP challenging approach. Then, the best-worst fuzzy multi-criteria decision model in Ling09 software was developed by collecting the opinions of 23 GRP experts, based on three general criteria of change in administrative structure (C1), policy structure (C2) and modification of GRP legal structure (C3), which includes 18 sub-criteria. Donyaii et al. (2021) evaluated and compared the performance of the farmland fertility algorithm (FFA) and that of the Harris hawks optimization (HHO) algorithm with each other to solve the two-objective problem of optimal operation of the Golestan dam reservoir under baseline (April 2006–October 2018) and climate change conditions (April 2021–October 2033) in Golestan Province, Iran and the results showed that when considering a constant value of the reliability index (i.e. 80%), the rate of vulnerability changes in the multi-objective FFA (Moffa) is lower than that of the multi-objective HHO (Mohho) algorithm. Ghanbari et al. (2021) obtained the important socioeconomic parameters of the city of Tehran, Iran in the period of 1991–2013. Important and optimum variables were analyzed and selected using the Pearson correlation analysis, and four variables including income, pop, GDP, and month were selected. In order to find a proper model, three common machine learning (ML) techniques including artificial neural network (ANN), random forest (RF), and multivariate adaptive regression splines (MARS) were used and the results revealed that the MARS model outperformed all the other models. In addition, in the last step, the crow search algorithm (CSA) was applied to the MARS model to increase the accuracy of the selected model. Therefore, the SWAT and the computational intelligence methods have been widely used for stream flow estimation. However, few studies have compared both models for daily stream flow estimation. Koycegiz & Buyukyildiz (2019) investigated the performance of two SWAT and Data-Driven Models (the RBNN and the SVR models) for simulating river runoff in Turkey. The results showed that the data-driven models were more successful than the SWAT model, although the SWAT model is a comprehensive physical model and has a significant problem-solving capability (Eini et al. 2020b).

Therefore, the aim of the present study was to investigate the limitations and strengths of machine learning based models compared the SWAT hydrological model for monthly streamflow modeling in Maharlu Lake Basin. Also, the review of previous studies showed that no study has been conducted in this field so far. So, this study attempted to evaluate the effectiveness of the SWAT hydrological model and three computational intelligence methods.

MATERIALS AND METHODS

Methods

Firstly, the SWAT model was developed regarding the required inputs. The facts that a section of the basin is of karstic type and it has different hydrological characteristics should be considered to increase the accuracy of the basin simulations. It should be noted that for the calibration and the validation of simulation, we used the SUFI2 provided in the software SWAT-CUP (Eini et al. 2020a).

Regarding the widespread use of the machine learning methods, the flexibility as well as the high power of these methods for estimating nonlinear functions, in the next section, the three methods of the Extreme Learning Machine (EML), the Support Vector Regression (SVR); and the Least Squares Support Vector Regression (LSSVR) are used to estimate the streamflow. In this study, the mean-weighted noise reduction filter was used to eliminate the out-of-range points. Eighty percent of the data was used to train the model and the remaining 20% was used to evaluate the accuracy of the models. Finally, the accuracies of the developed models and the SWAT model for the estimation of the streamflow were evaluated and compared using various statistical criteria such as the Root Mean Square Error (RMSE), the Determination Coefficient (R²), and the NSE. A graphical error analysis was also performed by the use of the regression and comparison graphs between the actual values and the estimated values of the models in order to provide a better analysis of the error of the developed models and the semi-distributed SWAT model.
The study area and the input data

To compare the accuracy of the SWAT model and the methods of extreme machine learning, the watershed in Iran was selected as the case study in this work. Figure 1 shows the location map of the watersheds. It is located in the catchment area of the central plateau of Iran and in the basin of Bakhtegan and Maharlu lakes with an area of about 4,270 km² and lies between 29°1′–30°6′ N and 52°12′–53°28′ E. The average annual precipitation and temperature in the region are 368 mm and 17.4 °C, respectively. The precipitation in wet seasons (occurring from November to March) accounts for more than 90% of the annual precipitation (Eini et al. 2020a).

Soil and water assessment tool (SWAT)

The SWAT model is one of the strongest and most sophisticated hydrological models due to its wide range of governing equations. This model can predict the trend and effect of runoff changes under different land use patterns and the conditions of the soil and river basin management in large basins (Eini et al. 2018). For the SWAT modeling purposes, a basin may be partitioned into several sub-basins and then further sub-classified as the Hydrologic Response Units (HRU) with a unique grouping of land use, soil, and slope (Flügel 1995). The calculation of the runoff yield after the basin subdivision can improve the accuracy of the runoff simulation (Liang et al. 2018). The hydrological simulation at each HRU is performed using the water balance equation (Arnold et al. 1998):

\[
SW_t = SW_0 + \sum_{i=1}^{t} \left( R_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw} \right)
\]

where,
- \(SW_t\): Final soil water content (mm).
- \(SW_0\): Initial soil water content (mm).
- \(R_{day}\): The amount of precipitation on day \(i\) (mm).
- \(Q_{surf}\): The amount of surface runoff on day \(i\) (mm).
- \(E_a\): The amount of evapotranspiration on day \(i\) (mm).
- \(W_{seep}\): The amount of water entering the vadose zone from the soil profile on day \(i\) (mm).
- \(Q_{gw}\): The amount of return flow on day \(i\) (mm).

Figure 1 | Map of the study area.
The SWAT model setup and data sets

The model data mainly include two parts: the model input and the calibration verification data. In the SWAT model, it is necessary to build a spatial database of the elevation, the land use, the soil type, and the meteorological data. The elevation data used a Digital Elevation Model (DEM) with a spatial resolution of 15 m (https://gdex.cr.usgs.gov/gdex/) (Figure 2(a)). The DEM and the land use data are from the United States Geographical Survey (USGS). According to the standard, the land use data are divided into five categories (Table 1 and Figure 2(b)). Shrub lands are dominant in this basin, covering 76.60% of the area. The soil data were obtained by downloading the attributes from the FAO global soil data (1:50000) (https://www.fao.org) and include 3 soil categories and 5 sub-categories (Figure 2(c) and Table 2). Also, the dominant soil in this basin is loamy (59.5%).

The historical meteorological period was selected from January 1980 to December 2013 and included the daily precipitation and the maximum and minimum temperatures. The meteorological and hydrological data were extracted from one Synoptic station in Shiraz, two rain gauge stations, and three hydrological stations for model calibration and assessment including Bagh Safa, Chenar Sokhte Azam, and Chenar Sokhte Khoshk stations. For precise simulation, the major crops in the region were modeled according to the comprehensive reports provided by the Iranian Ministry of Energy. These crops include almonds, apples, peaches, tomatoes, olives, winter wheat, and grapes. The agricultural area is 25,840 ha (17% of the total area).

Sensitivity analysis: calibration and validation

The accuracy of the SWAT simulations is highly dependent on the calibration and the validation processes (Zhang et al. 2019). The observed stream flow was obtained from three hydrologic stations. The observed data

![Figure 2](https://example.com/f2.png)

**Figure 2** | The spatial input details used in the SWAT model, (a) the DEM (15 m resolution) ranges from 1,426 to 2,991 m, (b) the GLCC classification, (c) the soil classification.
were divided into two parts before running the SWAT simulations: one was used for the model calibration and the other for the model validation.

The SWAT-CUP app is available to calibrate and analyze the uncertainty of the SWAT model. This application is capable of performing calibration and uncertainty analyses using Sequential Uncertainty Fitting (SUFI2), Particle Swarm Optimization (PSO), Generalized Likelihood Uncertainty Estimation (GLUE), the Parameter Solution (ParaSol) and Markov Chain Monte Carlo (MCMC) algorithms (Abbaspour 2007). In this study, the SUFI-2 algorithm, which is widely used to calibrate the SWAT model, was used for the calibration and uncertainty analyses. The SUFI-2 algorithm is known as an inverse optimization method based on the Bayesian framework, in which all sources of uncertainty, including the uncertainties of the parameters, the input variables like precipitation, the model conceptualization, and the measured data in the domain introduced for each parameter are considered. The initial range of the parameters after each iteration is substituted using a new range of parameters obtained from the model, and a limited uncertainty domain of the parameters is calculated. This procedure continues until the two quantifying uncertainty indices and the R-Factor-P-Factor reach a proper limit (Abbaspour 2007). In the SUFI-2, there were 11 statistical indices, including the product of the RMSE, the sum of the RMSE, the interpreting coefficient, the Nash-Sutcliffe coefficient, and the PBIAS method.

**Table 1** | The land use layer in the map 2012

| Land use type                  | SWAT code | Area (%) |
|-------------------------------|-----------|----------|
| Shrub land                    | SHRB      | 76.60    |
| Crop land                     | CRWO      | 8.90     |
| Crop land                     | CRDY      | 8.74     |
| Crop land                     | CRGR      | 0.51     |
| Grass land                    | GRAS      | 3.08     |
| Crop land                     | CRIR      | 0.62     |
| Urban construction land       | URMD      | 1.45     |
| Water area                    | WATR      | 0.04     |

*GLCC organization (The Global Land Cover Characterization (GLCC) is a series of the global land cover classification datasets that are primarily based on the unsupervised classification of 1-km AVHRR (Advanced Very High Resolution Radiometer) 10-day NDVI (Normalized Difference Vegetation Index) composites. The AVHRR source imagery dates from April 1992 through March 1993. The Ancillary data sources included the digital elevation data, the eco-regions interpretation, the country- or regional-level vegetation, and the land cover maps. [https://lta.cr.usgs.gov/GLCC](https://lta.cr.usgs.gov/GLCC)*

**Table 2** | The soil types used in modeling

| FAO soil layer* (1 km² cells) | Soil type (FAO classification) | Partial | Aggregated |
|-------------------------------|--------------------------------|---------|------------|
| I-Rc-Yk-c-3508                | LOAM                           | 29.67   | 59.56      |
| Rc37-3c-3532                  | LOAM                           | 29.89   |            |
| Xk32-3a-3573                  | SANDY-CLAY                     | 21.37   | 21.37      |
| Xk33-3a-3574                  | CLAY-LOAM                      | 18.89   | 19.05      |
| Xk5-2-3a-3578                 | CLAY-LOAM                      | 0.16    |            |

*The Food and Agriculture Organization (FAO) is a specialized agency of the United Nations that leads international efforts to defeat hunger. The goal is to achieve food security for all and make sure that people have regular access to enough high-quality food to lead active, healthy lives. With over 194 member states, FAO works in over 130 countries worldwide. They believe that everyone can play a part in ending hunger. [http://www.fao.org/soils-portal/soil-survey](http://www.fao.org/soils-portal/soil-survey)*

**Machine learning models**

**Extreme machine learning (EML)**

Extreme Machine Learning (EML) has three main layers consisting of the input layer, the hidden layer, and the output layer. As with any general neural network, the input data is first presented to the network and then each of these input data enters the neurons in the hidden layer. In the hidden layer, the characteristics are selected by mapping the input data to the N-dimensional space, where the most important and relevant characteristics are
extracted effectively (Roul et al. 2015). The output of the hidden layer is then processed by the output layer neurons for classification or regression. The distinctive characteristic of the EML compared to the conventional neural network methods is that the network training process is performed only by adjusting the output layer values, while the weights of the hidden layer neurons are randomly initialized and remain unchanged in the training process. This is a unique advantage of fast network training, which reduces the computational volume.

An EML network is shown in Figure 3. At the runtime, each sample in input dataset \( L \) enters the network as a pair. Each pair contains an input characteristic vector \( X_P \) and its corresponding surface label \( t_P \), so that \( X_P \in \mathbb{R}^n, \forall P = 1, 2, \ldots, L \). The network feedback output can be calculated using Equation (2), where \( t'_i \) represents the predicted output of the unit \( i \) in \( M \). \( M \) is the total number of the units in the output layer with \( K_i \) activation functions for \( \forall i = 1, 2, \ldots, M \). Also, \( N \) denotes the number of hidden units with \( f_j \) active functions for \( \forall j = 1, 2, \ldots, N \).

\[
  t'_i = K_i \left( \sum_{j=1}^{N} \beta_j f_j(X, b) \right)
\]

The results of \( \beta_j \) output layer weight vector from the Moore-Penrose generalized inverse multiplication of the hidden layer output matrix \( H \) are expressed by the training vector \( T \) and as Equation (3). The inverse of matrix \( H \) can be obtained using the orthogonal projection method where \( H^{-1} = (H^T H)^{-1} H^T \). As a result, the output weight matrix \( \beta \) can be obtained from Equation (4) (Zyarah & Kudithipudi 2017).

\[
  \beta = H^{-1} T \quad \text{(3)}
\]

\[
  \beta = T H^T (H^T H)^{-1} \quad \text{(4)}
\]

**Support vector regression (SVR)**

The Support Vector Machine (SVM) is a type of the supervised learning system established by Vapnik and Chernovenkis, and has gained popularity since then (Cortes & Vapnik 1995). The SVMs are used both for classification and estimation of the data fitting functions in regression problems so that the least error occurs in grouping data or fitting function. The support vector machine is called Support Vector Classification (SVC) and, on the one hand, the SVMs used for modeling and prediction are called support vector regression (SVR) (Burges 1998; Smola & Schölkopf 2004). If the output \( y \) is a function of \( n \) input variables \( x \), we assume that a training dataset of length \( N \) is \( T = \{(x_1, y_1), (x_2, y_2), \ldots (x_N, y_N)\} \), where \( x_k \in \mathbb{R}^n \) and \( y_k \in \mathbb{R}, K = 1, 2, \ldots, N \). In fact, \( x_k \)s are \( n \) dimensional vectors that take the values of each input in step \( k \) and \( y_k \)s are the output variable values in step \( k \). The basic SVR formula for a linear model is as follows:

\[
  y(x) = (w, x) + b
\]

where, \( y \) is the estimated output of the model, \( w \) is the weight vector and/or an element of the feature space of the problem, and \( b \) is the bias value. However, most real-world problems cannot be modeled using linear shapes. For nonlinear modeling, the SVR method uses nonlinear functions called Kernel functions. These nonlinear functions are responsible for mapping data from the input space to the feature space. So, the previous equation
can be rewritten as follows:

\[ y(x) = (w, \theta(x)) + b \]  

(6)

Radial Basis Functions (RBF) are among the most commonly used nonlinear Kernel functions for the SVR models. In the SVR algorithm, the aim is to minimize norm \( \|w\| \) vector \( w \) for each data point \( i = 1, 2, \ldots, N \), which are defined as:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \|w\|^2 \\
\text{subject to} & \quad y_i - (w, \phi(x_i)) - b \leq \varepsilon \\
& \quad (w, \phi(x_i)) + b - y_i \leq \varepsilon
\end{align*}
\]

(7)

A new error function is used to apply the support vector machine for regression problems; this function considers the error of the data exit from a specified range. The performance of this linear function is shown in Figure 3. The indicated deviation is defined as Equation (8):

\[
|\xi_i| = \begin{cases} 
0, & \text{if } |\xi| \leq \varepsilon \\
|\xi| - \varepsilon, & \text{otherwise}
\end{cases}
\]

(8)

Finally, based on the principle of the structural error minimization, the optimization problem is defined as Equation (9) where the aim is to find the values of \( W \) and \( b \), so that the value of the cost function is minimized (Ma et al. 2019).

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} W^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \\
\text{subject to} & \quad y_i - (w, \phi(x_i)) - b \leq \xi_i \\
& \quad (w, \phi(x_i)) + b - y_i - \xi_i \leq \xi_i^* \\
& \quad \xi_i, \xi_i^* \geq 0
\end{align*}
\]

(9)

where slack variables \( \xi_i \) and \( \xi_i^* \)\((\xi_i > 0, \xi_i^* > 0, i = 1, 2, \ldots, n)\) are adopted to measure the deviations outside the \( \varepsilon \)-insensitive zone, and parameter \( C \) is a positive constant called the regularization parameter, which determines a compromise between the model complexity and the amount up to which deviations larger than \( \varepsilon \) should be tolerated (Qu & Zuo 2012). This optimization problem is often solved in its dual form, in which the problem constraints are applied using Lagrange multipliers, and Lagrange \( (L) \) is minimized as follows:

\[
\begin{align*}
L & = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) - \sum_{i=1}^{N} \beta_i (y_i - (w, \phi(x_i)) + b) - \sum_{i=1}^{N} \beta_i^* (y_i - (w, \phi(x_i)) + b) - \sum_{i=1}^{N} \eta_i \xi_i \\
& \quad - \eta_i^* \xi_i^*
\end{align*}
\]

(10)

where, \( \beta, \beta^* \), \( \eta \) and \( \eta^* \) are Lagrange multipliers. The above statement, together with the Kuhn-Tucker complementary terms, produces a quadratic programming problem as follows:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_{ij} (\beta_i - \beta_j) (\beta_i - \beta_j^*) + \varepsilon \sum_{i=1}^{N} (\beta_i + \beta_i^*) - \sum_{i=1}^{N} y_i (\beta_i - \beta_i^*) \\
\text{subject to} & \quad \sum_{i=1}^{N} (\beta_i - \beta_i^*) \\
& \quad \beta_i \leq 0, \beta_i^* \leq C
\end{align*}
\]

(11)

where

\[
K_{ij} = K(x_i, x_j) = \phi(x_i)^T \phi(x_j)
\]

while \( K_{ij} \) is a Kernel function, which handles the data transfer from the input space to the feature space. The solution of the above Quadratic Programming (QP) problem is obtained from the optimization of the parameters.
\( \beta_i \) and \( \beta_i^* \) for \( i = 1, 2, \ldots, N \) such that for each training dataset, there is a pair of \( \beta_i \) and \( \beta_i^* \). Achieving a global optimum is guaranteed by the use of a QP problem. Finally, the model form in the dual space can be written as follows:

\[
y(x) = \sum_{i=1}^{N} (\beta_i - \beta_i^*) k(x, x_i) + b
\]

Also, the Kernel function used in this study is the radial basis function (Gaussian function) which is defined as follows:

\[
K_{ij} = k(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{2\sigma^2}\right)
\]

This function has a single parameter \( \sigma \) that calculates the width of the function (Kavaklioglu 2011).

**Least square SVR (LSSVR)**

Compared to the standard SVR, the LSSVR is based on a squared loss function and is optimal for prediction of the data with Gaussian additive noise (Suykens et al. 2002; Qu & Zuo 2012). The LSSVR model is formulated as:

\[
\min J = \frac{1}{2} \|W\|^2 + \frac{1}{2} \sum_{i=1}^{N} e_i^2
\]

subject to \( y_i = \omega^T \phi(x_i) + b + e_i \) \( (i = 1, 2, \ldots, n) \)

where a mapping \( \phi(.) \) is directly adopted to allow the LSSVR to work in a linear space, as for the regular SVR. Parameter \( C \) is the regularization parameter, which needs to be pre-specified. Equation (14) has equality constraints resulting from the squared loss function. Using the Lagrange multiplier method for Equation (14) yields an un-constrained optimization problem (Qu & Zuo 2012):

Maximize \( L(\omega, b, e; \alpha) = J(\omega, e) - \sum_{i=1}^{N} \alpha_i (\omega^T \phi(x_i) + b + e_i - y_i) \)

where \( \alpha_i \)'s are the Lagrange coefficients. The optimization conditions are as follows:

\[
\frac{\delta L}{\delta \omega} = 0 \rightarrow \omega = \sum_{i=1}^{N} \alpha_i \phi(x_i)
\]

\[
\frac{\delta L}{\delta b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0
\]

\[
\frac{\delta L}{\delta e_i} = 0 \rightarrow \alpha_i = \lambda e_i \quad (i = 1, 2, \ldots, n)
\]

\[
\frac{\delta L}{\delta e_i} = 0 \rightarrow \alpha_i = \lambda e_i \quad (i = 1, 2, \ldots, n)
\]

These conditions are similar to those of the standard SVR except for the condition \( \alpha_i = C e_i \). It can be seen that \( \alpha_i \) is dependent on \( e_i \), which is usually non-zero for all the data points, so for the LSSVR all the data points are involved in constructing its regression model, and this is different from the regular SVR (Qu & Zuo 2012). After removing \( e_i \) and \( \omega_i \), the optimal values of \( \alpha_i \) and \( b \) can be obtained by solving the linear equation as follows:

\[
\begin{pmatrix}
0_{1 \times l} & I_l^T \\
1_{1 \times l} & \Omega + \frac{1}{\lambda} I_l
\end{pmatrix}
\begin{bmatrix}
b \\
\alpha
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
y
\end{bmatrix}
\]

(17)
where \( y = \{y_1; y_2; \ldots; y_n\} \), \( \alpha = \{\alpha_1; \ldots; \alpha_n\} \), and \( 0_{1 \times l} \) is a unit vector of dimensions \( m \times n \), \( 0_{l \times n} \) is a matrix with components 0 of dimensions \( m \times n \), and \( \Omega \) is a matrix with elements \( (x_i)^T \varphi (x_j) \), \( i = 1, \ldots, n \). The optimal value of bias \( \hat{b} \) and Lagrange coefficients \( \alpha_i \) are obtained by solving a linear equation. Then, the optimal regression function for inputs \( x_0 \) is written as follows (Yuan & Lee 2015; Moqaddasi Amiri et al. 2019):

\[
y(x_0) = \sum_{i=1}^{N} \alpha_i K(x_0, x_i) + \hat{b}
\]

(18)

Method of modeling machine learning models

In order to develop the models studied in this paper, one part of the collected data should be randomly separated for the model training and another for the model performance evaluation. For this purpose, after improving the data quality through the pre-processing process, the data from 1980 to 1996 were used to train the models and the data from 1997 to 1999 were used for testing and validation. Also, in order to remove the effect of the variable interval on the modeling performance, all variables were normalized in a range between zero and one. The modeling process by the methods studied in this paper is illustrated as a flowchart in Figure 4, through which the generalizations of these methods can be well understood.

One of the most important components affecting the SVR and LSSVR models is the type of Kernel function used. Three Gaussian, linear, and polynomial functions are the functions commonly used as the Kernel functions in support vector machine models. In this study, the Gaussian Kernel function was used to develop the two models SVR and LSSVR. The training process of these two models is similar to that of the EML method, the training model is first produced by the training dataset and is finally evaluated and validated by the test dataset.

Evaluation criteria for model comparison

In the present paper, the performance of the models used for estimating the streamflows were evaluated by the use of the RMSE, the Determination Coefficient (\( R^2 \)), and the Nash-Sutcliffe efficiency (NSE), which are most widely used in hydrology studies. These statistics are defined in Table 3.

\( R^2 \) describes the degree of the collinearity between the simulated and the measured data and the proportion of the variance in the measured data explained by the model (Santhi et al. 2001; Moriasi et al. 2007). The Nash-Sutcliffe efficiency (NSE) is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance (Nash & Sutcliffe 1970). The NSE indicates how well the plot of the observed versus the simulated data fits the 1:1 line.

The Root-Mean-Square Deviation Error is the difference between the value predicted by the model and the actual value. The closer it is to zero, the less it is.

RESULTS AND DISCUSSION

Results of the SWAT model

The first step in the calibration and validation of the model is to identify the parameters that affect the variables. The effects of the parameters on the variables varies depending on the hydrological conditions in different basins. Each parameter in the SWAT model has a specific impact on the streamflow or on the other components of the water balance. Since each of these parameters can have different effects alone or together, the sensitivity analysis of the input parameters to the model was performed using the SWAT-CUP program. The sensitivity analysis was performed using two t-stat and p-value parameters that determine the effect of each parameter on the variables. The higher the absolute value of t-stat is \([\pm \infty]\) and the closer the p-value is to zero, the greater the effect of the parameter changes on the variable is (Abbaspour et al. 2015).

In the SWAT Model, crack flow is defined as a process during which water directly infiltrates into the aquifer via cracks and gaps. To control crack flow, it is essential to calibrate the parameters linked to the cracks. The \( CN2 \) parameters showed the highest impacts on the streamflow (Eini et al. 2018). The basic flow parameters associated with the groundwater also exhibited a high sensitivity to the model. The soil parameters such as

1 The number of curves in the average moisture conditions of the basin.
Sol_BD² and Sol_AWC³ along with the parameters such as the surface water time delay, the sub-basin average slope, and the hydraulic conductivity of waterways, were also sensitive. Table 4 presents 16 parameters with higher sensitivity, which were selected to calibrate and verify the model. The initial values and the range of the

² Moist bulk density.
³ Available water capacity of the soil layer.

Figure 4 | The flowchart of the methods studied to estimate the flow rate.
parameters can refer to the existing research, which can save time for parameter adjustment and improve efficiency.

Also, the performance criteria for the SWAT-Crack flow model are shown in Table 5.

The period 1980–1982 was used as the warm up period of the model, 1982–1997 was used as the calibration period of the model, and 1997–1999 was used as the validation period of the model. The observed and the simulated streamflow time series in the baseline period (1980–1999) is shown in Figure 5.

### Table 3 | Performance metrics

| Equation | Rang |
|----------|------|
| \( R^2 = \frac{\sum_{i=1}^{n} (Y_{obs}^i - \bar{Y}_{mean}^i)(Y_{sim}^i - \bar{Y}_{mean}^i)}{\sum_{i=1}^{n} (Y_{obs}^i - \bar{Y}_{mean}^i)^2} \) | \([0, 1]\) |
| \( NSE = 1 - \frac{\sum_{i=1}^{n} (Y_{obs}^i - Y_{sim}^i)^2}{\sum_{i=1}^{n} (Y_{obs}^i - \bar{Y}_{mean}^i)^2} \) | \([-\infty, 1]\) |
| \( RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_{sim}^i - \bar{Y}_{mean}^i)^2}{n}} \) | \([0, \infty]\) |

\( Y_{obs}^i \) is the \( i \)th observation for the constituent being evaluated, \( Y_{sim}^i \) is the \( i \)th simulated value for the constituent being evaluated, \( \bar{Y}_{mean}^i \) is the mean of the observed data for the constituent being evaluated, and \( n \) is the total number of the observations.

### Table 4 | The sensitivity analysis of the parameters affecting on the surface streamflow

| Rank | Parameter name | Parameter name | Initial values | Final values | t-stat | p-value |
|------|----------------|----------------|---------------|--------------|--------|---------|
| 1    | ALPHA_BF.gw(v) | Base flow alpha factor (days) | 1 0 | 0.0037 | −10.35 | 0 |
| 2    | CH_K2.rte(v)   | Effective hydraulic conductivity in main channel alluvium. | 500 0.01 | 352.12 | 4.15 | 0 |
| 3    | OV-N.hru(v)    | Manning’s ‘n’ value for overland flow. | 30 0.01 | 12.41 | 2.41 | 0.016 |
| 4    | SOL_AWC.sol(r) | Soil available water storage capacity | 0 0 | 0.214 | 1.83 | 0.067 |
| 5    | GW_REVAP.gw(v) | Groundwater ‘revap’ coefficient. | 0.2 0.02 | 0.095 | −1.73 | 0.085 |
| 6    | CANMX.hru(v)   | Maximum canopy storage. | 100 0 | 53.74 | −1.63 | 0.104 |
| 7    | CH_S2.rte(v)   | Average slope of main channel. | 10 0.001 | 6.7 | 1.59 | 0.112 |
| 8    | GW_DELAY.gw(v) | Groundwater delay (days). | 500 0 | 108.7 | −1.48 | 0.14 |
| 9    | REVPAMN.gw(v)  | Threshold depth of water in the shallow aquifer for ‘revap’ to occur (mm). | 500 0 | 407.1 | −1.38 | 0.168 |
| 10   | GWQMN.gw(v)    | Threshold depth of water in the shallow aquifer required for return flow to occur (mm). | 5,000 0 | 2,275.8 | −1 | 0.317 |
| 11   | SLSUBBSN.hru(v) | Average slope length. | 150 10 | 72.98 | 0.79 | 0.425 |
| 12   | CH_N2.rte(v)   | Manning’s ‘n’ value for the main channel. | 0.3 −0.01 | 0.1 | −0.68 | 0.498 |
| 13   | CN2.mgt(r)     | SCS runoff curve number | 0.2 −0.2 | −0.013 | −0.67 | 0.505 |
| 14   | SOL_BD.sol(r)  | Moist bulk density. | 0.6 −0.5 | 0.302 | −0.66 | 0.507 |
| 15   | ESCO.hru(v)    | Soil evaporation compensation factor. | 1 0 | 0.59 | 0.6 | 0.544 |
| 16   | SOL_ALB.sol(r) | Moist soil albedo. | 0.5 −0.5 | −0.036 | −0.55 | 0.579 |
| 17   | EPCO.hru(v)    | Plant uptake compensation factor. | 1 0 | 0.56 | −0.14 | 0.888 |

### Table 5 | The performance criteria for the SWAT-crack flow model

| Method  | Set       | RMSE | \( R^2 \) | NS    |
|---------|-----------|------|----------|-------|
| SWAT-CRK | Training  | 1.07 | 0.75     | 0.70  |
|         | Testing   | 1.02 | 0.88     | 0.86  |
The results of machine learning models

The daily meteorological data of Shiraz Synoptic Station were used for the data collection of the machine learning models, covering the period between 1980 and 1999. The parameters used include 20 independent variables (input variables) and one dependent variable (target variable), whose statistical information are presented in Table 6. The initial data could not be used for the modeling because of the large number of out-of-range values and had to be pre-processed before modeling. Finally, after removing incomplete and out-of-range samples, 2689 samples remained, which were used for modeling the discharge value.

Table 6 | Statistical information of Shiraz Station

| No. | Parameter                        | Min    | Max    |
|-----|----------------------------------|--------|--------|
| 1   | Year                             | 1980   | 1999   |
| 2   | Month                            | 1      | 12     |
| 3   | Day                              | 1      | 31     |
| 4   | Maximum wind speed               | 0      | 14     |
| 5   | Maximum wind direction           | 0      | 360    |
| 6   | Mean wind speed                  | 0      | 5.375  |
| 7   | Maximum temperature              | 1.5    | 43.2   |
| 8   | Minimum temperature              | −9.6   | 28.6   |
| 9   | 24-h rainfall                    | 0      | 99     |
| 10  | Mean wind speed                  | 0      | 5.375  |
| 11  | Mean temperature                 | −3.5   | 32.2   |
| 12  | Maximum daily cloudiness         | 19     | 100    |
| 13  | Minimum relative moisture        | 2      | 59     |
| 14  | Mean relative moisture           | 10.75  | 82.62  |
| 15  | Mean soil temperature            | −15    | 24     |
| 16  | Minimum sea level pressure       | 994.81 | 1,027.20 |
| 17  | Maximum sea level pressure       | 998.15 | 1,033.92 |
| 18  | Mean sea level pressure          | 996.4  | 1,030.23 |
| 19  | Maximum station level pressure   | 846.30 | 863    |
| 20  | Minimum station level pressure   | 846.30 | 859.5  |
| 21  | Mean station level pressure      | 845.1  | 861.07 |
| 22  | Discharge                        | 0      | 10     |
According to Table 7, it seems that the input variables including maximum temperature, minimum temperature, rainfall, and mean soil temperature have higher values of linear correlation with the target variable (discharge). Also, the other variables presented in the above table have lower values of linear correlation due to the non-linear relationship with the discharge variable.

The noise in the data is an alternative problem in the hydrological time series that affects the model performance adversely. Also, the obtained flow measurements are often subject to many errors. The presence of noise in the hydrological data is so serious that the noise level in these data has been proven to be between 5 and 15%, which is sufficient to reduce the model performance and produce unreliable results (Sivakumar et al. 1999). One of the fundamental problems of the data mining is the data noise reduction. The noise in the data is an inevitable problem that affects the performance of the model adversely. In this study, after removing the out-of-range points, a noise reduction filter was used to remove the data noise. The filter is based on creating a window of a constant length and then calculating the mean value per window, thereby reducing the data noise. In Figure 6, the flow rates of the two measured and the noise-reduced states are compared.

In order to develop the EML model, selecting the number of neurons in the hidden layer is one of the basic steps before starting the modeling process by this method. As stated in the theory section of this method, the EML model removed the computation of the number of hidden layers, making it much more efficient than the MLP neural networks. Also, in this method, adjusting the weights between the input layer and the hidden layer is not involved in the training process and the main focus is on neuronal computation. In this study, the trial-and-error approach with the Mean Squared Error (MSE) was used to select the number of the neurons and the type of the activation functions. Different number of neurons and activation functions were studied to obtain the Mean Squared Error. Finally, it was found that the EML model with a structure consisting of a hidden layer with 400 neurons containing the sigmoid activator function provided the best performance in the

Table 7 | Profile of the collected parameters to estimate the flow

| No. | Parameter                        | Min   | Max   | Correlation with the target variable (R) |
|-----|----------------------------------|-------|-------|----------------------------------------|
| 1   | Year                             | 1980  | 1999  | 0.002                                  |
| 2   | Month                            | 1     | 12    | -0.193                                 |
| 3   | Day                              | 1     | 31    | -0.011                                 |
| 4   | Maximum wind speed               | 0     | 14    | -0.076                                 |
| 5   | Maximum wind direction           | 0     | 360   | 0.053                                  |
| 6   | Mean wind speed                  | 0     | 5.375 | 0.091                                  |
| 7   | Maximum temperature              | 1.5   | 43.2  | -0.83                                  |
| 8   | Minimum temperature              | -9.6  | 28.6  | 0.42                                   |
| 9   | 24-h rainfall                    | 0     | 99    | 0.67                                   |
| 10  | Mean wind speed                  | 0     | 5.375 | 0.27                                   |
| 11  | Mean temperature                 | -3.5  | 32.2  | 0.201                                  |
| 12  | Maximum daily cloudiness         | 19    | 100   | 0.247                                  |
| 13  | Minimum relative moisture        | 2     | 59    | -0.26                                  |
| 14  | Mean relative moisture           | 10.75 | 82.62 | 0.193                                  |
| 15  | Mean soil temperature            | -15   | 24    | -0.45                                  |
| 16  | Minimum sea level pressure       | 994.81| 1,027.20| 0.193                  |
| 17  | Maximum sea level pressure       | 998.15| 1,033.92| 0.167                  |
| 18  | Mean sea level pressure          | 996.4 | 1,030.23| 0.185                  |
| 19  | Maximum station level pressure   | 846.30| 863   | -0.024                                |
| 20  | Minimum station level pressure   | 846.30| 859.5 | -0.009                                |
| 21  | Mean station level pressure      | 845.1 | 861.07| -0.012                                |
| 22  | Discharge                        | 0     | 10    | 1                                     |
discharge approximation. After training the EML model by the training dataset (the data from 1980 to 1996), the model was evaluated by the test dataset (the data between 1997 and 1999).

Figures 7–9 show the results of the point diagrams and the fit lines on these points, which can be used to analyze the model error in both training and testing sections as well as the correlation between the actual and the estimated values. In these diagrams, the closer those points are to the first-quarter bisector, the better the prediction will be, and vice versa, the farther away the points are, the worse the data match is. In Figures 10 and 11, the actual and the estimated values of the three EML, SVR, and LSSVR models are compared for the two training and test sections, respectively. Also in Figure 12, the RMSE values of the three developed models are compared. According to Figure 12 it can be said that the results of the LSSVR model have the least error for estimating the streamflow rate and are more accurate than those of the other models.

Comparison of the SWAT and the computational intelligence methods
In order to evaluate the error of the developed models and the SWAT model, different statistical parameters were used. As shown in Table 8, the LSSVR model with the least error shows the best performance compared to the
different models as well as the SWAT model. The results from evaluating the efficiency of the computational intelligence methods show that the models had less error at the validation stage than at the calibration stage because the model learning was performed by the training data. The EML model takes less computation than the MLP neural network method does. The reduced computation in this method has caused the overfitting problem not to occur during the model training process and the model evaluation results to be accurate in spite of the large increase in the hidden layer units. Also, given that one of the important cases for developing the SVR and the LSSVR models is to determine the Kernel function, the results showed that the Gaussian Kernel (radial basis) provides a very good accuracy for estimating the streamflow. However, among the models studied in this paper, the best results were obtained from the LSSVR method. The RMSE values obtained from the LSSVR method for the training and test datasets were 0.038 and 0.057, respectively, which presented less error than the two EML and SVR methods do.

**Figure 8** | Regression of LSSVR model training and test stages at Bagh Safa station.

**Figure 9** | Regression of SVR model training and test stages at Bagh Safa station.
SUMMARY AND CONCLUSION

Determining the flow of rivers is one of the most important and effective components in the management of water resources in watersheds. In this study, the performances of the semi-distributed SWAT model and three computational intelligence methods were evaluated for monthly simulation of Maharlu Lake Basin stream flow. The methods used for modeling and estimating the flow include three methods of the Extreme Machine Learning (EML), the Least Squares Support Vector Regression (LSSVR), and the Support Vector Regression (SVR) and the process of developing these models is performed by the software MATLAB 2018. In order to enhance the quality level of the data, the blank lines of the data and the out-of-range data were first removed. Then, the data noise was removed by a mean weight-based noise reduction filter. The study of the diagram comparing

![Figure 10](image1.png)  
*Figure 10* | Comparison of the measured streamflow noise reduction with the modeled values for the model training section at Bagh Safa station.

![Figure 11](image2.png)  
*Figure 11* | Comparison of the measured streamflow noise reduction with the modeled values for the models section at Bagh Safa station.
the measured values and the noise reduction values indicated by the filter showed that this filter can very well remove the noise data by preserving the original signal form. According to the results of the correlation between the input data and the stream flow variable, it was proved that the four variables of maximum temperature, minimum temperature, 24-h rainfall, and mean soil temperature have a good correlation with the objective variable, so can be effective variables for estimating the watershed flow.

In addition, after preparing the required data, such as DEM, land use, and soil maps, the SWAT model was calibrated using the SUFI-2 algorithm as a semi-auto calibration procedure. Studies show that the models perform well in simulating the watershed monthly flows, but the support vector machine least squares methodology had higher accuracy in estimating monthly flows. The structure of this method is based on the solution of the linear optimization problem of the support vector and has equivalents whose opposite Lagrangian coefficients are nonzero. Here, the data are defined as the support vectors.

The study results show that the support vector machine method is used in cases where complex systems cannot be defined by mathematical equations or problems of data disturbance or model recognition as well as in areas with defects of statistics. These models with the minimum input parameters show acceptable results and given that these models are based on the principle of the structural error minimization, so for simulation, using the learning method that supervises the radial basis functions makes parameter estimations have higher rate and lower error than other Kernels do, and this is one of the distinctive characteristics of radial basis functions. On the one hand, the SWAT model is a semi-distributed-conceptual model with a very high accuracy in solving problems. The reason for using this model is that it simulates all hydrological phenomena in watersheds and is not limited to large volumes of information. This model is especially useful when different areas of the watershed have different soil or land use so that the heterogeneity and difference can affect the hydrology of the watershed. Given that the implementation of this model requires a large amount of catchment information and usually due to the complexity and large volume of the data, it requires computers with high processing ability and more time to analyze the uncertainty. Therefore, in studies where we are faced with the lack of information and time, it is

Table 8 | Results of the estimation accuracy of the models developed by different statistical criteria

| Method    | Set   | RMSE     | $R^2$   | NS   |
|-----------|-------|----------|---------|------|
| ELM       | Training | 0.172   | 0.954   | 0.954 |
|           | Testing | 0.299   | 0.854   | 0.85  |
| LSSVRL    | Training | 0.038   | 0.997   | 0.997 |
|           | Testing | 0.057   | 0.994   | 0.994 |
| SVR       | Training | 0.161   | 0.960   | 0.9601|
|           | Testing | 0.199   | 0.935   | 0.9355|
| SWAT-CRK  | Training | 1.07    | 0.75    | 0.70  |
|           | Testing | 1.02    | 0.88    | 0.86  |
justifiable to use models such as vector machine that can be implemented with the least available data and provide acceptable results. In addition, in studies in which simulation is used for a comprehensive investigation of hydrological events and their effects on each other, using hydrological models such as SWAT is recommended.

**DATA AVAILABILITY STATEMENT**

All relevant data are included in the paper or its Supplementary Information.

**REFERENCES**

Abbaspour, K. 2007 User manual for SWAT-CUP, SWAT calibration and uncertainty analysis programs. In: *Swiss Federal Institute of Aquatic Science and Technology*. Eawag, Duedendorf, Switzerland, p. 93.

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H. & Klöve, B. 2015 A continental-scale hydrology and water quality model for Europe: calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of Hydrology* 524, 733–752.

Afkmifar, S. & Sarraf, A. P. 2020 Comparative study of groundwater level forecasts using hybrid neural network models. *Proceedings of the Institution of Civil Engineers – Water Management*. https://doi.org/10.1080/iwama.20.00062.

Arnold, J. G., Srinivasan, R., Muttiliah, R. S. & Williams, J. R. 1998 Large area hydrologic modeling and assessment part I: model development 1. *JAWRA Journal of the American Water Resources Association* 34 (1), 73–89.

Burges, C. J. 1998 A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2 (2), 121–167.

Cortes, C. & Vapnik, V. 1995 *Support-Vector Networks*. *Machine Learning* Vol. 20. Kluwer Academic Publisher, Boston, MA.

Deo, R. C., Samui, P. & Kim, D. 2016 Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. *Stochastic Environmental Research and Risk Assessment* 30 (6), 1769–1784.

Donyaii, A. R. & Sarraf, A. P. 2020 Optimization of reservoir operation using a bioinspired metaheuristic based on the COVID-19 propagation model. *Journal of Numerical Methods in Civil Engineering*. *NMCE* 5 (1), 15–28. http://nmce.kntu.ac.ir/article-1-289-en.html.

Donyaii, A. R., Sarraf, A. P. & Mardanifar, M. 2020 Evaluation of crisis resolution strategies for groundwater revival plan using fuzzy best – worst multi-criteria decision model. *Journal of Hydraulic Structures* 6 (3), 21–44. doi: 10.22055/jhs.2020.35415.1149.

Donyaii, A. R., Sarraf, A. P. & Ahmadi, H. 2021 Comparison of meta-heuristic algorithms in optimum operation of a single-reservoir dam system. *Proceedings of the Institution of Civil Engineers – Engineering Sustainability*. https://doi.org/10.1680/jensu.20.00062.

Eini, M. R., Javadi, S. & Delavar, M. 2018 Development of comprehensive karstic watershed model in order to estimate and precise the components of the water balance. *Iran-Water Resources Research* 14, 133–145.

Eini, M. R., Javadi, S., Delavar, M., Gassman, P. W. & Jarilhani, B. 2020a Development of alternative SWAT-based models for simulating water budget components and streamflow for a karst-influenced watershed. *Catena* 195, 104801.

Eini, M. R., Javadi, S., Hashemy Shahdany, M. & Kisi, O. 2020b Comprehensive assessment and scenario simulation for future of the hydrological processes in Dez river basin, Iran. *Water Supply* 21 (3), 1157–1176.

Flügel, W. A. 1995 Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrological Processes* 9 (3–4), 423–436.

Ghanbari, F., Kamalan, H. & Sarraf, A. P. 2021 An evolutionary machine learning approach for municipal solid waste generation estimation utilizing socioeconomic components. *Arabian Journal of Geosciences* 14 (2021), 92. https://doi.org/10.1007/s12517-020-06348-w.

Ghorbani, M. A., Khatibi, R., Goel, A., FazeliFard, M. H. & Azani, A. 2016 Modeling river discharge time series using support vector machine and artificial neural networks. *Environmental Earth Sciences* 75 (8), 675–685.

Huang, G.-B., Zhu, Q.-Y. & Siew, C.-K. 2004 Extreme learning machine: a new learning scheme of feedforward neural networks. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, IEEE, pp. 985–990.

Huang, G.-B., Zhu, Q.-Y. & Siew, C.-K. 2006 Extreme learning machine: theory and applications. *Neurocomputing* 70 (1–3), 489–501.

Huang, S., Chang, J., Huang, Q. & Chen, Y. 2014 Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology* 511, 764–775.

Jajarmizadeh, M., Ladjani, E. K., Harun, S. & Ahmadi, A. 2015 Application of SVM and SWAT models for monthly streamflow prediction, a case study in South of Iran. *KSCE Journal of Civil Engineering* 19 (1), 345–357.

Jimeno-Sáez, P., Senent-Aparicio, J., Pérez-Sánchez, J. & Pulido-Velazquez, D. 2018 A comparison of SWAT and ANN models for daily runoff simulation in different climatic zones of Peninsular Spain. *Water* 10 (2), 192.

Kavaklioglu, K. 2011 Modeling and prediction of Turkey’s electricity consumption using Support Vector Regression. *Applied Energy* 88 (1), 368–375.

Koycegiz, C. & Buyukyildiz, M. 2019 Calibration of SWAT and two data-driven models for a data-scarce mountainous headwater in semi-arid Konya closed basin. *Water* 11 (1), 147.
Kumar, D., Pandey, A., Sharma, N. & Flügel, W.-A. 2016 Daily suspended sediment simulation using machine learning approach. *Catena* 138, 77–90.

Liang, Z., Tang, T., Li, B., Liu, T., Wang, J. & Hu, Y. 2018 Long-term streamflow forecasting using SWAT through the integration of the random forests precipitation generator: case study of Danjiangkou Reservoir. *Hydrology Research* 49 (5), 1513–1527.

Ma, Z., Ye, C. & Ma, W. 2019 Support vector regression for predicting building energy consumption in southern China. *Energy Procedia* 158, 3433–3438.

Moqaddasi Amiri, M., Tapak, L. & Faradmal, J. 2019 A mixed-effects least square support vector regression model for three-level count data. *Journal of Statistical Computation and Simulation* 89 (15), 2801–2812.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50 (3), 885–900.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – a discussion of principles. *Journal of Hydrology* 10 (3), 282–290.

Noori, N. & Kalin, L. 2016 Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology* 533, 141–151.

Qu, J. & Zuo, M. J. 2012 An LSSVR-based algorithm for online system condition prognostics. *Expert Systems with Applications* 39 (5), 6089–6102.

Rezaeenour, J., Roozbahani, E., Roozbahani, M. & Yari, M. 2019 Optimization ELM neural network in prediction problem: case study forecasting demand steel in Iran. *Journal of Development Evolution Management* 2019 (37), 25–34.

Roul, R. K., Nanda, A., Patel, V. & Sahay, S. K. 2015 Extreme learning machines in the field of text classification. In: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE, pp. 1–7.

Santhi, C., Arnold, J. G., Williams, J. R., Dugas, W. A., Srinivasan, R. & Hauck, L. M. 2001 Validation of the swat model on a large river basin with point and nonpoint sources 1. *JAWRA Journal of the American Water Resources Association* 37 (5), 1169–1188.

Sivakumar, B., Phoon, K.-K., Liong, S.-Y. & Liaw, C.-Y. 1999 A systematic approach to noise reduction in chaotic hydrological time series. *Journal of Hydrology* 219 (3–4), 103–135.

Smola, A. J. & Schölkopf, B. 2004 A tutorial on support vector regression. *Statistics and Computing* 14 (3), 199–222.

Suykens, J. A. & Vandewalle, J. 1999 Least squares support vector machine classifiers. *Neural Processing Letters* 9 (3), 293–300.

Suykens, J. A., De Brabanter, J., Lukas, L. & Vandewalle, J. 2002 Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing* 48 (1–4), 85–105.

Vapnik, V. 1998 *Statistical Learning Theory*. John Wiley and Sons, Inc, New York.

Vilaysane, B., Takara, K., Luo, P., Akkharath, I. & Duan, W. 2015 Hydrological stream flow modelling for calibration and uncertainty analysis using SWAT model in the Xedone river basin, Lao PDR. *Procedia Environmental Sciences* 28, 380–390.

Yuan, F.-C. & Lee, C.-H. 2015 Using least square support vector regression with genetic algorithm to forecast beta systematic risk. *Journal of Computational Science* 11, 26–33.

Zhang, L., Wang, R., He, J. & Wang, P. 2018 Mobile node localization method based on KF-LSSVR algorithm. *EURASIP Journal on Wireless Communications and Networking* 2018 (1), 1–8.

Zhang, S., Li, Z., Lin, X. & Zhang, C. 2019 Assessment of climate change and associated vegetation cover change on watershed-scale runoff and sediment yield. *Water* 11 (7), 1373.

Zyarah, A. M. & Kudithipudi, D. (2017). Extreme learning machine as a generalizable classification engine. In: 2017 *International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp. 3371–3376.