Theor Appl Climatol
DOI 10.1007/s00704-015-1392-3

ORIGINAL PAPER

Monthly evaporation forecasting using artificial neural networks and support vector machines

Gulay Tezel · Meral Buyukyildiz

Received: 6 August 2014 / Accepted: 25 January 2015

Abstract Evaporation is one of the most important components of the hydrological cycle, but is relatively difficult to estimate, due to its complexity, as it can be influenced by numerous factors. Estimation of evaporation is important for the design of reservoirs, especially in arid and semi-arid areas. Artificial neural network methods and support vector machines (SVM) are frequently utilized to estimate evaporation and other hydrological variables. In this study, usability of artificial neural networks (ANNs) (multilayer perceptron (MLP) and radial basis function network (RBFN)) and ε-support vector regression (SVR) artificial intelligence methods was investigated to estimate monthly pan evaporation. For this aim, temperature, relative humidity, wind speed, and precipitation data for the period 1972 to 2005 from Beysehir meteorology station were used as input variables while pan evaporation values were used as output. The Romanenko and Meyer method was also considered for the comparison. The results were compared with observed class A pan evaporation data. In MLP method, four different training algorithms, gradient descent with momentum and adaptive learning rule backpropagation (GDX), Levenberg–Marquardt (LVM), scaled conjugate gradient (SCG), and resilient backpropagation (RBP), were used. Also, ε-SVR model was used as SVR model. The models were designed via 10-fold cross-validation (CV); algorithm performance was assessed via mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). According to the performance criteria, the ANN algorithms and ε-SVR had similar results. The ANNs and ε-SVR methods were found to perform better than the Romanenko and Meyer methods. Consequently, the best performance using the test data was obtained using SCG(4,2,2,1) with R^2=0.905.

G. Tezel
Engineering Faculty, Computer Engineering Department, Selcuk University, 42075 Konya, Turkey

M. Buyukyildiz (✉)
Engineering Faculty, Civil Engineering Department, Selcuk University, 42075 Konya, Turkey
e-mail: meralbyildiz@selcuk.edu.tr

1 Introduction

Because of the complexity between the components of land, plant, water surface, and atmosphere system, evaporation is perhaps the most difficult and complicated parameter to estimate among all the components of the hydrological cycle (Singh and Xu 1997). Estimation of evaporation is important in all areas associated with water resources because it affects the reservoir capacity, rainfall-runoff modeling, river basin yield, irrigation water management for the calculation of crop water requirements and its scheduling, etc. In areas with little rainfall, evaporation losses can represent a significant part of the water budget for a lake or reservoir and may contribute significantly to the lowering of the water surface elevation (McCuen 1998).

The actual measurement of evaporation is almost impossible, but evaporation can be estimated using several methods. There are two general approaches to the estimation of evaporation: direct and indirect methods. Direct methods include the application of class A pan measurements, class U pan, lysimeter, and other types of measurements; indirect methods comprise models based on meteorological variables, such as Penman–Monteith, Priestley–Taylor, Blaney–Criddle, and Thornthwaite equations; catchment water budget approach; energy budget method; and mass transfer method (Terzi 2013).

It is impractical to place evaporation pans at every point where there is a planned or existing reservoir and irrigation

Published online: 17 February 2015

project. It is also highly impractical in inaccessible areas, where accurate instruments cannot be established or maintained. A practical means of estimating the amount of pan evaporation where no pans are available is of considerable significance to hydrologists, agriculturists, and meteorologists (Kisi 2009).

The indirect methods of evaporation estimation that were proposed by some researchers often require some forms of input data that are not readily available. Also, the evaporation process is highly nonlinear, and so, simple models cannot satisfactorily describe the nonlinear behavior of the process via the available data (Shiri and Kisi 2011).

The artificial intelligence methods such as artificial neural network (ANN) and support vector machines (SVM) have been successfully utilized to a number of areas of civil engineering (Ahmad et al. 2007; Shahin et al 2004; Yavuz et al. 2014; Goh and Goh 2007; Ulengin and Topcu 2000; Cheng and Cao 2014; Koroglu et al. 2012; Goyal 2014; Kisi 2012, etc). The natural behavior of hydrological processes is appropriate for the application of artificial intelligence methods (McCuen 1998).

In recent years, ANNs and SVM have been extensively used to model numerous nonlinear hydrological processes such as rainfall (Goyal 2014), precipitation (Hamidi et al. 2014), rainfall-runoff modeling (Modarres 2009), evaporation (Terzi 2013; Samui and Dixon 2012; Kisi 2013; Kim et al 2014), temperature (Ertac et al 2014), water quality (Kalin et al 2010), streamflow (Kisi and Cimen 2011), water level (Cimen and Kisi 2009) and suspended sediment (Kisi 2012), etc.

Samui et al. (2011) used support vector machine (SVM) and relevance vector machine (RVM) to predict rainfall of Vellore town in Tamil Nadu State, India. Goyal (2014) investigated the usability of wavelet regression models for monthly rainfall prediction using the rainfall data from 21 stations in Assam, India and to compare them with the performance of artificial neural network models. Consequently, the performances of wavelet regression models are found to be more successful than the ANN models. Hamidi et al. (2014) compared two machine learning techniques, artificial neural network (ANN), and support vector machines (SVM) for modeling monthly precipitation. They used the monthly precipitation data of two synoptic stations in the central area of Hamadan (Airport and Nojeh), the west of Iran. According to the results, the SVM models are found to be better than the ANN models in order to predict monthly precipitation. Modarres (2009) studied an effective ANN model for rainfall-runoff modeling of the Plasjan Basin in the western region of the Zayandehrud watershed, Iran. Ertaç et al. (2014) carried out nonlinear time series analysis and partial least square (PLS) to estimate monthly mean temperature using meteorological variables like temperature, humidity, and rainfall from 54 meteorology observation stations of Istanbul in Turkey and compared with artificial neural networks (ANNs). Kim et al. (2014) used a multilayer perceptron neural networks model (MLP-NNM) and a cascade correlation neural networks model (CCNNM) for predicting daily pan evaporation using five-input models for inland and coastal stations in the Republic of Korea. Terzi (2013) assessed the performance of gene expression programming (GEP) and adaptive neural-based fuzzy inference system (ANFIS) for the estimation daily pan evaporation from Lake Egirdir. Consequently, it was found that the GEP model was better than the ANFIS model. Kisi (2013) examined the accuracy of a least square support vector machines (LSSVM) to estimate reference evapotranspiration (ET_0) and compared empirical models (Priestley–Taylor, Hargreaves, and Ritchie methods) and ANN. It was seen that the LSSVM models performed better than the empirical and ANN models for modeling ET_0. Samui and Dixon (2012) used support vector machine (SVM) and relevance vector machine (RVM) to predict evaporation losses in reservoirs and compared ANN models. The results show that SVM and RVM models can be used as practical tool to predict evaporation. Cimen and Kisi (2009) investigated the usability of SVM and ANN techniques to predict surface water level changes of Van and Egirdir Lakes in Turkey.

The objective of this study was to investigate usability of ε-support vector regression (SVR), radial basis function network (RBFN), and multilayer perceptron (MLP) methods for estimating monthly pan evaporation using meteorological data such as rainfall, wind speed, relative humidity, and temperature. In MLP method, four different training algorithms, gradient descent with momentum and adaptive learning rule backpropagation (GDX), Levenberg–Marquardt (LVM), scaled conjugate gradient (SCG), and resilient backpropagation (RBP) were used.

2 Material and methods

2.1 Case study

Monthly data were obtained from Beysehir meteorology station operated by the General Directorate of State Meteorological Affairs, located near Lake Beysehir in the southwestern part of Turkey. Beysehir observation station is one of the three meteorology stations which make class A pan evaporation observations and represent the evaporation on lake surface area in Beysehir Lake catchment area and its immediate surroundings.

Lake Beysehir is the largest freshwater lake and drinking water reservoir in Turkey and located (37° 40′ 54″ N, 31° 43′ 22″ E) between Beysehir (Konya) and Sarkikaraagac (Isparta) provinces. It has a surface area of approximately 650 km^2, is 45 km long and 20 km wide, with a maximum depth of 10 m, and has a precipitation area of 3095 km^2. It carries the same name as the principal urban center of the region, Beysehir. Lake Beysehir is the most important drinking and irrigation water source for Central Anatolia. The water level in the lake often fluctuates by year and by season. Lake Beysehir is also a national park.

2.2 Data

Monthly data for the period 1972 to 2005 were used for training and testing ANN models. Of the entire data set (227 monthly series), 150 were used for training the ANN and ε-SVR models and 77 monthly values were used for testing. The meteorological data used in this study were divided as input and output for the ANN and ε-SVR to determine the effects of meteorological factors on evaporation. Months for which there were inadequate data were removed. Input data were mean air temperature (T), mean relative humidity (RH), wind speed (WS), and total rainfall (R); output data was class A pan evaporation (E). The monthly statistical parameters of the climatic data are given in Table 1; X_{min}, X_{max}, X_{mean}, S_{x}, C_{v}, and C_{sx} symbolize the minimum, maximum, mean, standard deviation, variation coefficient, and skewness, respectively.

Statistical analysis of the meteorological parameters given in Table 1 indicated that rainfall data demonstrated the highest variation (1.03) and skewness (1.41), whereas the relative humidity showed the lowest variation (0.14) and skewness (0.10); since temperature is the most effective factor for evaporation (correlation coefficient 0.94), evaporation increases as temperature increases; the relative humidity and rainfall exhibited significantly higher negative correlations (−0.70 and −0.56, respectively) with evaporation; the wind speed had a negligible effect on evaporation, with the lowest correlation coefficient of 0.12.

2.3 Artificial neural networks

Artificial neural networks are inspired by biological neuron processing and are utilized in many studies of various disciplines such as mathematics, statistics, computer science, and engineering to realize many tasks—modeling, time series analysis, signal processing, pattern recognition, and classification (Haykin 1999). ANNs are distributed, adaptive, generally nonlinear learning machines that are built from many different nonlinear processing elements called "neurons," each of which receives connections from other neurons or/and itself according to the training algorithm. The signals flowing on the connections are scaled by adjustable parameters known as weights (Principe et al. 2000).

2.3.1 Multilayer perceptron

Gradient descent with momentum and adaptive learning rule backpropagation algorithm The GDX is a backpropagation method that uses a gradient descent technique to train the supervised, layered, feed-forward network. The architecture of the method is designed with one input layer, one or more hidden layers, and one output layer.

The number of hidden layers and the number of neurons on each layer differ between applications (Skapura 1996).

In this method, before training, the training patterns, target output patterns, and other network parameters (learning rate, minimum error, transfer functions of each layer's neurons, the number of hidden neuron, and maximum iteration) are determined according to the application. The number of input- and output neurons is dependent on the application and is automatically assigned. The connection weightings between neurons in the input, hidden, and output layers and biases are initially randomized small values, which are modified as the process of training the network proceeds (Haykin 1999; Fausett 1994).

The training process begins with the presentation of an input pattern to ANN, through which the propagation of the input pattern was performed using neurons until an output value is produced. Finally, each neuron modifies its input connection weights slightly in a direction that reduces the error signal, and the process is repeated for the next pattern (Fausett 1994).

Scaled conjugate gradient algorithm In the traditional backpropagation algorithm, the weights are modified in the steepest descent direction; in an SCG algorithm, a search is conducted along conjugate directions (Principe et al. 2000).

Each of the conjugate gradient algorithms requires a line search in each iteration. These algorithms start with the gradient descent direction and search for the minimum along a line. However, this method is time-consuming because it requires

Table 1 Statistical characteristics of the data used in the study

Data	X_{min}	X_{max}	X_{mean}	S_{x}	C_{v} ($S_{\mathrm{x}}/X_{\mathrm{mean}}$)	C_{sx}	Correlation (with E)
Evaporation, E (mm/month)	4.40	294.70	140.29	59.34	0.42	−0.25	1.00
Wind speed, WS (m/s)	0.00	2.20	0.80	0.47	0.59	0.60	0.12
Relative humidity, RH (%)	36.00	76.00	57.52	8.19	0.14	0.10	−0.70
Temperature, T (°C)	3.90	24.60	16.63	4.79	0.29	−0.55	0.94
Rainfall, R (mm)	0.00	147.20	26.87	27.62	1.03	1.41	−0.56

the network responses to all training inputs to be computed several times for each search (Principe et al. 2000; Moller 1990).

The SCG algorithm has fully automated line search including user-independent parameters and was developed by Moller (1990) to avoid time-consuming line search per learning iteration. The convergence criterion of SCG produces a speed increase of at least one order of magnitude relative to backpropagation (BP). SCG uses second-order information from the neural network but requires only O(N) memory usage. In comparison with BP, SCG involves twice as much calculation work per iteration, since BP has a calculation complexity of O(N) per iteration (Moller 1990).

Resilient backpropagation algorithm The RBP is a successful training algorithm that manages a direct adaptation of the weight step with local gradient information. An individual update value (Δij) for each weight determines the size of the weight update. During the learning process, this adaptive update value changes depending on its local sight on the error function E, according to the following learning rule (Riedmiller and Braun 1993):

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^{+} \times \Delta_{ij}^{(t-1)}, & \text{if} \quad \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} \times \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ \eta^{-} \times \Delta_{ij}^{(t-1)}, & \text{if} \quad \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} \times \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ \Delta_{ij}^{(t-1)}, & \text{else} \end{cases} \tag{1}$$

where $0<\eta^{-}<1<\eta^{+}$

In this algorithm, the sign of the partial derivative of error function (E) with respect to the corresponding weight changes (w_{ij}) is used to determine the direction of the weight update. However, the magnitude of the derivative does not affect the weight update (Riedmiller and Braun 1993).

After the update value for each weight is computed, the weight update is realized with a simple rule: if the derivative is positive (increasing error), the weight is decreased by its update value, and if the derivative is negative, the update value is added. Furthermore, the update value remains the same when the derivative is zero (Riedmiller and Braun 1993). If the weight oscillates, the weight change is reduced. Whenever the weight continues to change in the same direction for several iterations, then, the magnitude of the weight change will be increased (Kisi and Uncuoglu 2005).

The Levenberg–Marquardt algorithm The LVM algorithm is observed to be more effective than simple gradient descent and many other conjugate gradient methods in a wide variety of problems. LVM presents local search properties of Gauss–Newton with consistent error decrease using a gradient descent algorithm. The training of a feed-forward network based on LVM is considered as an unconstrained optimization problem. The crucial disadvantages of the LVM algorithm are the requirement of increased memory capacity to calculate the Jacobian matrix of the error function and that it is not guaranteed to find global optimum. If the solution is unacceptable, the whole training process should be restarted (Mukherjee and Routroy 2012).

LVM represents a simplified version of Newton's method which is a well-established numerical optimization technique with quadratic speed of convergence. An obvious problem with Newton's method is the computational requirement involved in calculating the inverse of a Hessian matrix. Even for moderate-sized neural networks, the complexity of the algorithm limits its practical use. LVM offers a viable alternative to Newton's method, but has approximately the same convergence speed and significantly less complexity. To apply LVM, the problem of training the MLP has to be formulated as a nonlinear optimization (Ham and Kostanic 2001).

In the gradient descent algorithm, a momentum is integrated to assist potential overshooting of local minima. By this approach, the local approximation of the cost function is assumed to be a quadratic and expanded second-order Taylor series. The main drawback of this algorithm is the computational complexity of calculating the matrix inversion with several thousand variables (Mukherjee and Routroy 2012).

2.3.2 Radial basis function network

Similar to MLP, RBFN is an example of a nonlinear layered feed-forward network, which can be used for both classification and functional approximation. For classification, the network can determine how close the given input is to the center of the Gaussian kernel by the response of the corresponding hidden unit (Haykin 1999; Fu 1994). The main difference between the RBFN network and the backpropagation network is their transfer functions—the radial basis function deals with only the small regions whereas the sigmoid function adopts non-zero values over an infinitely large region of input space (Fu 1994). The RBFN is principally designed as three layers with entirely different roles: the input layer comprised of source nodes connects the network to its environment, the linear output layer supplies the response of the network to the activation pattern applied to the input neurons, and the only hidden layer between the input and output layers applies a nonlinear transformation from input to hidden (Haykin 1999; ASCE Task Committee 2000). A RBFN network is generally preferred due to its rapid training ability and general applicability (Han and Felker 1997).

The Gaussian function, the most common activation function chosen for RBFN in the hidden layer, produces a localized response to the input. The training process of RBFN can be evaluated in two phases. In the first phase, learning is performed using unsupervised methods such as a *k*-means clustering algorithm in the hidden layer, and the learning process uses supervised methods such as the least mean square (LMS) algorithm for the output layer in the second phase (Fu 1994).

2.4 Support vector machines

The decision support systems which are called machine learning method were put forward to solve classification (SVM) and then regression (SVR) type problems by Vapnik for the first time in 1995 (Vapnik 1995). In both classification and regression operations, the learning problem is represented in the form of an optimization problem with quadratic objective function. SVM is a next-generation learning method oriented under the roof of statistical learning theorem which can be learned from high dimensional and small number of training data (Shen 2005), and the general structure of SVM is seen in Fig. 1. The SVR function is given in Eq. 2.

$$f(x) = \sum_{i=1}^{N_d} \alpha_i^* K(x_i, x) \quad + \quad b \tag{2}$$

where α_i^* is Lagrange multipliers, $K(x_i,x)$ is the kernel function, and b is the bias value. As $\phi\ (x_i)^T$ and $\phi(x)$ feature vectors, the general structure of the kernel function is denoted as in Eq. 3.

$$K\ (x_i, x) = \phi\ (x_i)^T \phi\ (x) \tag{3}$$

The selection of the kernel function to be used and of model parameters plays a vital role to determine the SVR performance. However, there is no any determinant criterion with respect to selection of either kernel function or model parameters (Lin 2006).

Three factors which are effective in SVR performance are ε error term, C configuration factor, and type of kernel function, thereby the parameter of the kernel function (Ekici 2007). The kernel functions commonly used in SVR applications are give in Table 2.

The SVR applications can be performed in two different models such as ε-SVR and υ-SVR. In this study, ε-SVR model and radial basis kernel function commonly preferred in SVR applications in the literature as a kernel function are used.

The program known as ORANGE Software has been used to estimate by the SVR method the evaporation amount in Beysehir Lake, and the applications for ANN are performed by MATLAB 2010b.

2.5 Romanenko method

This equation, developed by Romanenko (1961), is one of the empirical equations used for monthly evaporation amount estimation and is stated as follows:

$$E = 0.0018\ \ (25 + T_a)^2\ \ (100 \quad - \quad RH) \tag{4}$$

where

E Evaporation amount (mm/month)
T_a Air temperature (°C)
RH Relative humidity (%).

2.6 Meyer method

Meyer method, used for the estimation of evaporation amount from water surface and developed by Meyer (1915), is given

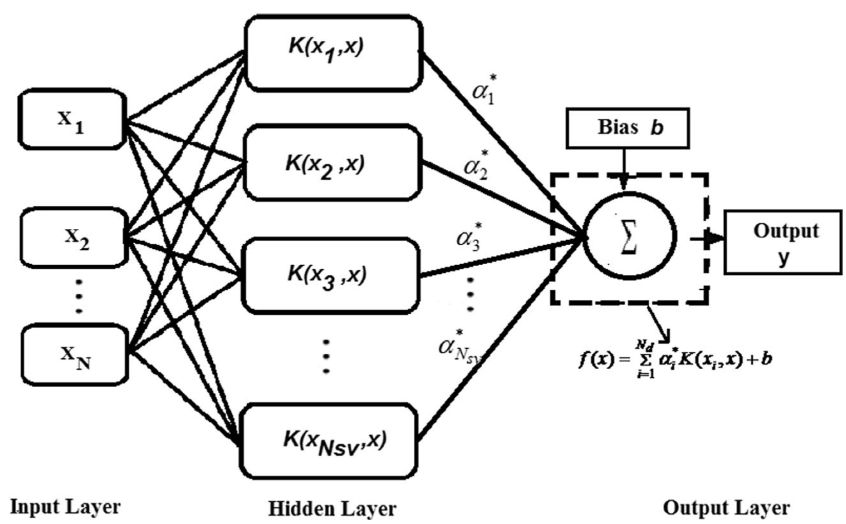


Fig. 1 General structure of support vector machines

Table 2 Commonly used kernel functions

Kernel type	Kernel function
Linear	$K(x_i,x)=x_i^T x_j$
Polynomial	$K(x_i,x)=(\gamma x_i^T x_j+r)^d$
Radial basis function	$K(x_i,x)=\exp(-\gamma\|x_i-x_j\|^2)$
Sigmoid	$K(x_i,x)=\tanh(\gamma x_i^T x_j+r)$

below.

$$E = 11\cdot(e_s - e)\left(1+\frac{\mathrm{WS}}{16}\right) \quad (5)$$

where

E Evaporation amount (mm/month)
e_s Saturated vapor pressure at (mmHg)
e Vapor pressure (mmHg)
WS wind speed (km/h).

RH, e_s parameter used in Eq. 5 as relative humidity, is calculated with equations below.

$$RH = \frac{e}{e_s} \quad (6)$$

$$e_s = 6.11\exp\left[17.3T\Big/(237.3+T)\right] \quad (7)$$

2.7 *k*-fold cross-validation

In the *k*-fold cross-validation method, data are separated into *k* partitions/folds (approximately equal), and each of the partitions in the turn is used for test while the rest is used for training. Thus, *k*-1 partitions of the *k* parts are worked for training and the remaining one (the *k*th hold-out part) is used for testing.

This procedure is repeated *k* times, so that every sample is exercised strictly once during the test process. The average error of all *k* periods is then calculated. Thus, the confidence of classifier and data was demonstrated using cross-validation (CV) (Lehmann et al. 2007).

2.8 Performance criteria

The performance criteria of root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were used to evaluate the results. Here, the strength of the relationship between two variables was measured with R^2; various types of information about the estimation capabilities of the model were provided by RMSE and MAE, and since MAE yields a more balanced perspective of the goodness-of-fit at moderate evaporation, the goodness-of-fit related to high evaporation values was measured by RMSE (Karunanithi et al. 1994). RMSE and MAE are defined as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|Ei_{\mathrm{observed}} - Ei_{\mathrm{estimate}}\right| \quad (8)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Ei_{\mathrm{observed}} - Ei_{\mathrm{estimate}}\right)^2} \quad (9)$$

in which N is the number of data set and Ei is the monthly evaporation.

3 Results and discussion

In this study, before applying ANNs and SVM to data, the input and output values were normalized in the range of [0, 1] using Eq. 10 in order to remove disparities related to the differing units used to represent the parameters.

$$X_{\mathrm{norm}} = \frac{X - X_{\mathrm{min}}}{X_{\mathrm{max}} - X_{\mathrm{min}}} \quad (10)$$

where X_{norm}, X, X_{min}, and X_{max} are the normalized value, observed value, minimum value, and maximum value, respectively.

Initially, many trials were carried out to determine the most appropriate network models, by using different numbers of hidden layer neurons and iterations for all the training ANN methods and, additionally, by changing the learning rate and momentum coefficient for MLP.

Furthermore, many trials were made for various transfer functions to determine the most appropriate network structure for the learning algorithms except the RBFN method. As a result of trials, the appropriate transfer functions were determined as hyperbolic tangent function and logarithmic sigmoid function for the hidden layer neurons and output layer neurons, respectively. Using different iterations, trials between 500 and 10,000 were conducted for LVM, GDX, and RBP, and between 25 and 1000 for SCG algorithms. The most appropriate values were determined as 2000, 2000, 50, and 1000 for the LVM, GDX, SCG, and RBP algorithms, respectively.

Table 3 MAE and RMSE values of the most successful models for all methods used

Methods	Models	Training		Testing	
		MAE (mm/month)	RMSE (mm/month)	MAE (mm/month)	RMSE (mm/month)
ANN	SCG	12.340	16.052	27.512	31.770
	RBP	11.692	15.258	25.745	30.308
	LVM	*10.721*	*14.621*	25.845	30.754
	GDX	11.501	14.894	22.754	26.996
	RBFN	11.652	15.401	24.546	28.652
ε-SVR (C, ε, γ)	ε-SVR (512, 0.05, 0.5)	12.320	16.968	*17.485*	*22.647*
Meyer		28.739	33.525	60.367	24.497
Romanenko		17.512	66.131	31.720	37.394

Italic values indicate the best results

According to the trials, the performance criteria for LVM increased when a greater number of iterations were used, and the best results were obtained with 10,000 iterations. When these results were compared with those obtained using 2000 iterations, the R^2 values remained the same. An improvement of 0.1 was obtained for other performance criteria. However, this improvement was disregarded because of the increased processing time required. In conclusion, it was determined that 2000 iterations were more appropriate for LVM.

Additionally, two important parameters for determining the optimum network topology of MLP, learning ratio (lr) and momentum coefficient (mc), were tried for different combinations within the range of 0.1 and 1, at an interval of 0.01. The most appropriate values of lr and mc were acquired as 0.4 and 0.2, respectively. For RBFN, the parameter of spread was scanned within the range of 0.01 and 3, at an interval of 0.01. The maximum number of neurons was scanned within the range of 1 and 25. The most appropriate values of spread and the number of neurons were 2.58 and 14, respectively.

The best results were obtained in topologies with two hidden layers for all ANN algorithms except RBFN. As a result of trials, which examined the number of neurons in hidden layers, iteration number, and values of lr and mc, the most successful network topologies were determined as LVM(4,1,5,1), GDX(4,6,20,1), RBP(4,2,7,1), and SCG(4,2,2,1).

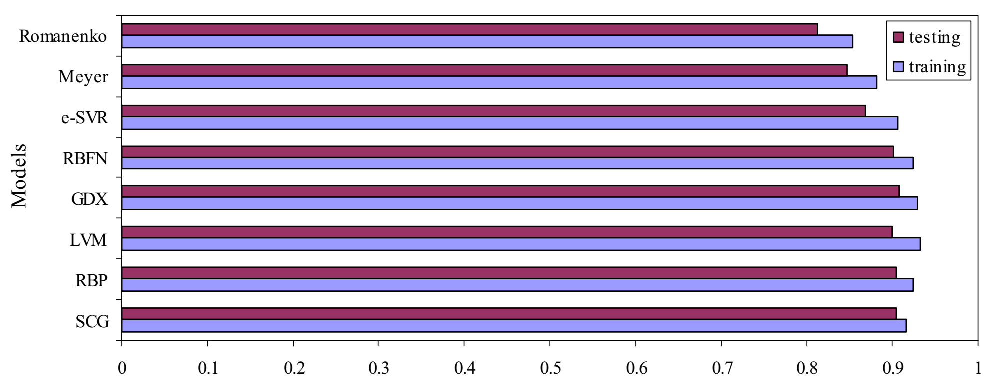
During the application of ε-SVR model, the most successful ε-SVR (C, ε, γ) architecture has been determined as ε-SVR (512, 0.05, 0.5) along with the values such as 0.05 for error term (ε), 512 for configuration factor (C), and 0.5 for γ parameter of the radial basis kernel function, as a result that the data set used is subjected to 10-fold cross-validation.

The performance criteria for these most successful network topologies of ANN and ε-SVR models, which were designed to estimate evaporation, are shown in Table 3 which indicates that the algorithms had similar performance.

Evaporation data obtained with both Romanenko and Meyer empirical equations and ANN and SVR were compared with the evaporation data taken into consideration and observed with the dividing of training and test periods. MAE and RMSE values of all methods used are shown in Table 3, while R^2 values are shown in Fig. 2. According to Table 3, the best estimation achievement was reached with LVM algorithm, MAE=10.721 mm/month and RMSE=14.621 mm/month values in training data and with ε-SVR method MAE=17.485 mm/month and RMSE=22.647 mm/month values in test data.

If an evaluation is made according to the R^2 values in Fig. 2, when all methods used are compared, ANN models have been more successful than both ε-SVR method and Meyer and Romanenko equations with higher R^2 values. Among ANN models, LVM algorithm with R^2=0.932 value in training data and GDX algorithm with R^2=0.908 value in test data have reached the highest achievement. According to

Fig. 2 Determination coefficients (R^2) values of the most successful models for all methods used



	SCG	RBP	LVM	GDX	RBFN	e-SVR	Meyer	Romanenko
testing	0.905	0.904	0.9	0.908	0.902	0.868	0.847	0.813
training	0.916	0.925	0.932	0.929	0.924	0.907	0.881	0.853

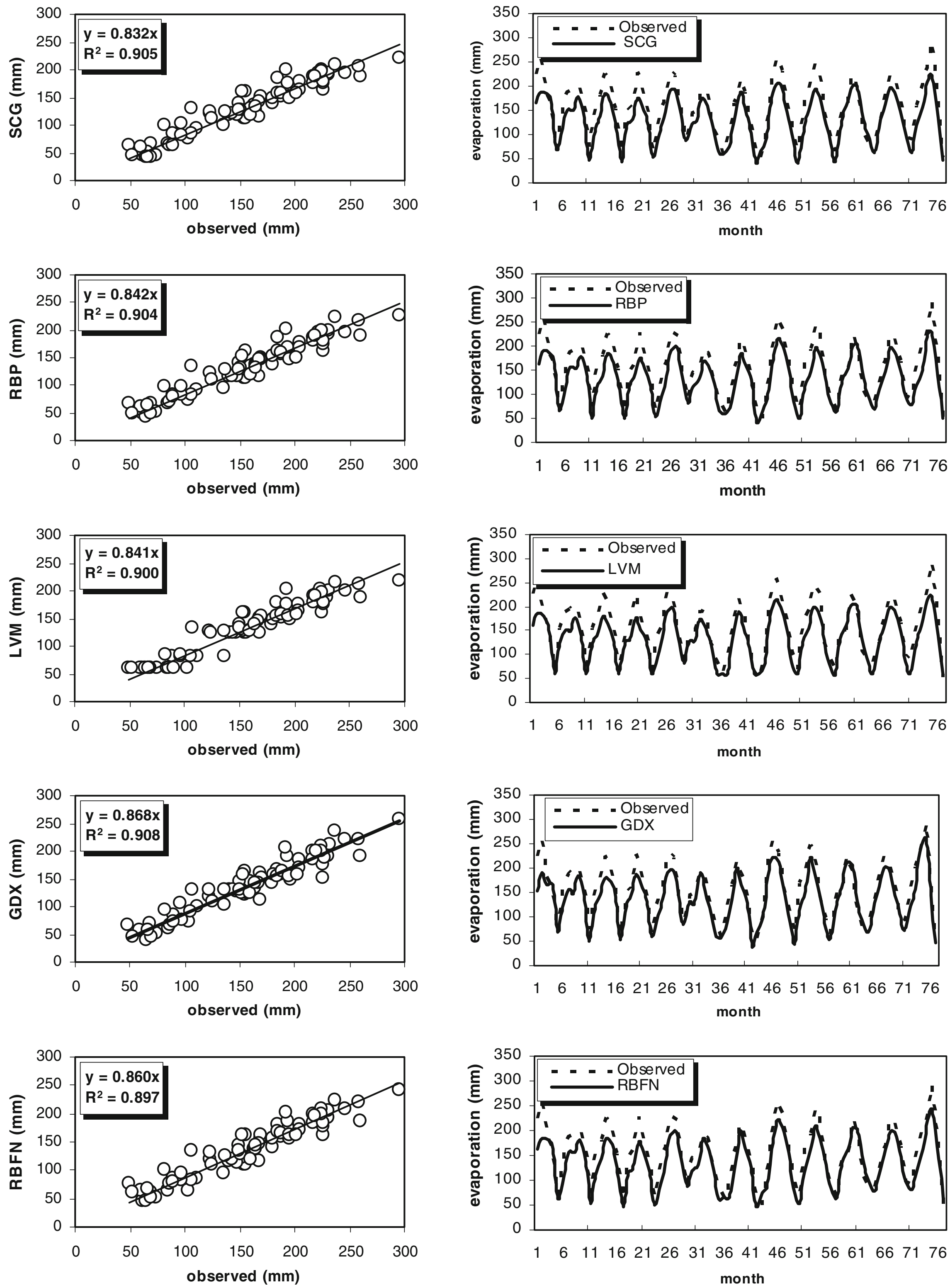


Fig. 3 The scatter diagrams and time series of ANNs, ε-SVR, Meyer, and Romanenko models in test period

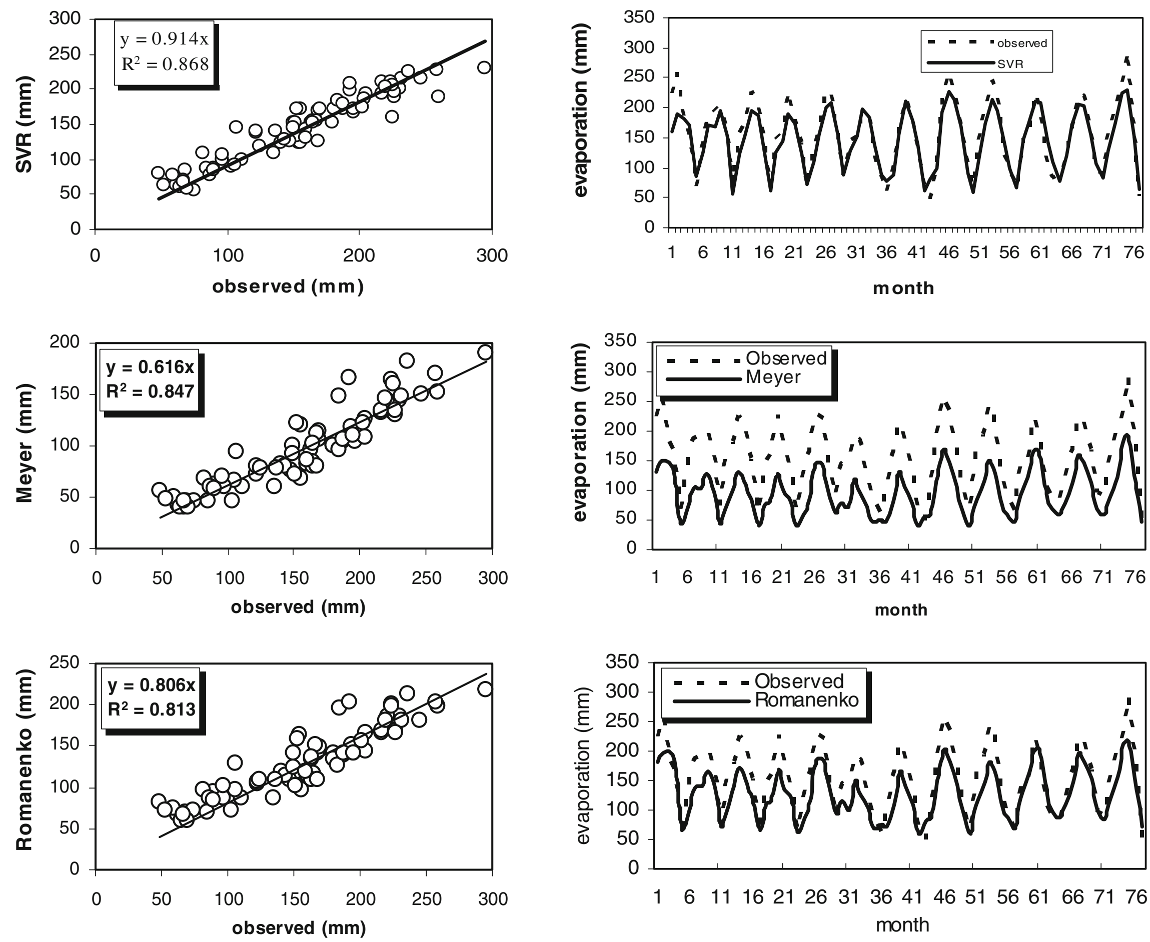


Fig. 3 (continued)

R^2 values, Meyer method with its R^2=0.847 value obtained for test data is more successful than Romanenko method by which R^2=0.813 value was obtained. However, both Romanenko and Meyer methods underperformed compared with ANNs and ε-SVR methods in estimating evaporation.

Figure 3 shows the scatter diagrams (a) where the monthly evaporation estimated via ANNs, ε-SVR, Meyer, and Romanenko models are compared with observed evaporation and time series (b). According to the scatter diagrams, the equations obtained from all algorithms present similar gradients, and the time series reveal nearly the same tendency for observed and estimated data. While the proposed models generated lower

Table 4 Performance criteria obtained via 10-fold cross-validation

ANN models	Training			Testing		
	MAE (mm/month)	RMSE (mm/month)	R^2	MAE (mm/month)	RMSE (mm/month)	R^2
SCG	14.840	2.034	0.896	*0.516*	*2.419*	*0.913*
RBP	13.219	1.833	0.912	0.671	3.147	*0.914*
LVM	13.191	2.007	0.911	0.635	2.978	0.907
GDX	*12.820*	*1.793*	*0.915*	0.582	2.730	0.893
RBFN	13.285	2.268	0.910	0.607	2.847	0.913

Italic values indicate the best results

Table 5 Test performance criteria for each fold for GDX and SCG algorithms

Fold number	GDX			SCG		
	MAE	RMSE	R^2	MAE	RMSE	R^2
1	*1.567*	*7.349*	*0.814*	*1.702*	*7.984*	*0.868*
2	0.177	0.830	0.875	0.959	4.500	0.950
3	0.830	3.895	0.929	0.080	0.375	0.937
4	0.689	3.229	0.909	0.453	2.125	0.912
5	0.016	0.077	0.896	0.024	0.111	0.917
6	0.067	0.315	0.921	0.492	2.308	0.936
7	*0.840*	*3.941*	*0.795*	*0.053*	*0.250*	*0.833*
8	0.410	1.922	0.959	0.281	1.318	0.964
9	0.652	3.059	0.953	0.277	1.297	0.958
10	0.572	2.685	0.881	*0.836*	*3.923*	*0.858*
Mean	0.582	2.730	0.893	0.516	2.419	0.913

Italic values indicate the best results

values than the maximum peaks of the observed data, they became more successful in obtaining minimum peaks.

In addition, 10-fold CV was used to prevent overfitting and to check the reliability of the ANN algorithms. The network architectures which obtained optimum results (Table 3) were implemented while applying the CV method. At first, the data set was divided into 10 partitions in order to use CV.

One of these partitions was used as a test, and the others for training at every step. Therefore, each data set was used both as training data and test data. Then, the average of the performance criteria obtained in all folds was calculated for each ANN algorithm (see Table 4).

Investigating the training results in Table 4, GDX gave the best results according to all performance criteria. While the SCG algorithm was the most successful in terms of MAE and RMSE in testing, RBP was the best one regarding R^2.

However, since there was no significant difference between SCG and RBP in terms of R^2, SCG was accepted to be the most successful algorithm according to the test results.

When the results in Table 3 and the results obtained via application of CV to the data set (Table 4) were compared, the optimum results were obtained in different ANN algorithms for both testing and training. When evaluated according to R^2 values, all of the algorithms performed similarly. This result demonstrated the usability of all the utilized ANN algorithms in estimating evaporation data for the Beysehir observation station.

While the most successful algorithm was GDX (Table 3) before the application of CV via the hold-out method, SCG was determined as the most successful algorithm after the application of CV method (Table 4). Therefore, the results of the test performance criteria of each fold, obtained as a result of 10-fold CV application, are given in Table 5 only for the SCG and MLP models.

As seen in Table 5, both MLP and SCG algorithms performed differently for each fold. Especially in fold 1, both algorithms resulted in failure in terms of MAE and RMSE. In addition, it is observed that the lowest R^2 value was obtained in the seventh fold for both algorithms. The high error value and low R^2 values obtained via CV negatively affect the average performance of the model. GDX gave an R^2 value less than 0.9 in five folds, compared with only three folds for SCG. This resulted in low average performance of R^2 in the MLP model. As a result, the SCG algorithm gave the best average performance in estimating evaporation using 10-fold CV.

Figure 4a shows the evaporation data obtained by combining the test values at the end of each fold for the SCG(4,2,2,1)

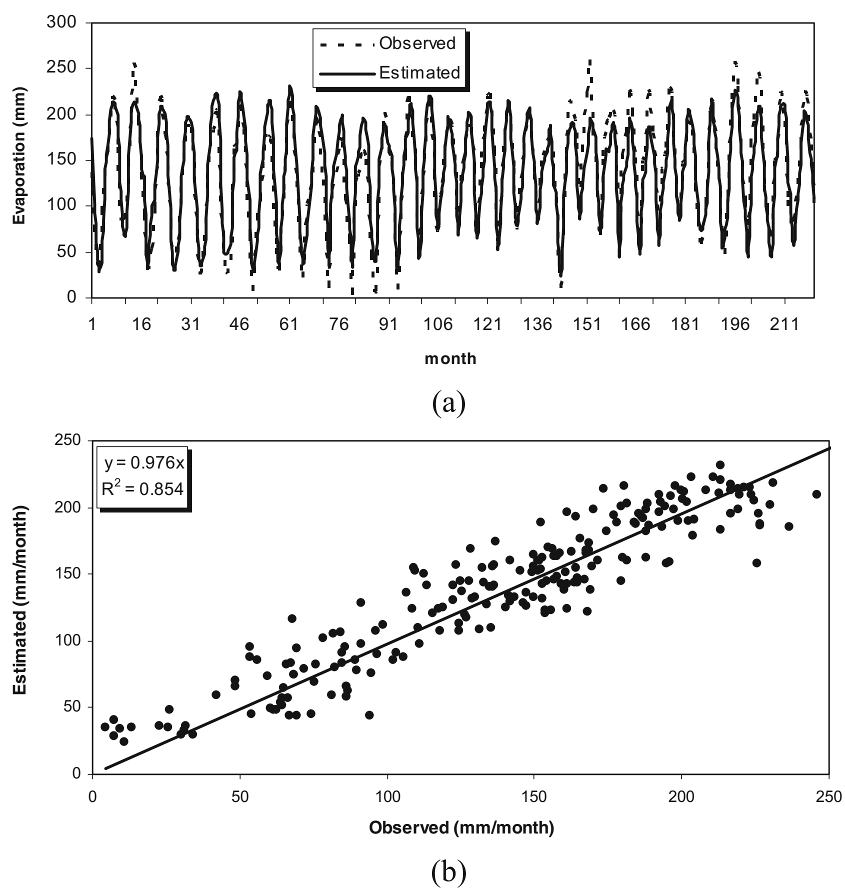


Fig. 4 **a** The time series of all data generated by the SCG(4,2,2,1) model established. **b** The scatter diagram

model; a corresponding scatter diagram is shown in Fig. 4b. As seen in Fig. 4a, the estimated and observed evaporation values converged, with the exception of some peak values.

This is also observed in the scatter diagram in Fig. 4b. The coefficient of determination for all test data was calculated as 0.854. This value is an adequate achievement in terms of usability of the proposed model in situations where evaporation cannot be determined via direct or indirect methods.

4 Conclusions

In this study, usability of ANNs (MLP and RBFN) and ε-SVR artificial intelligence methods in the estimation of evaporation amount has been studied. In the formation of ANN models, four different ANN algorithms consisting of GDX, LVM, SCG, and RBP were used in the MLP method. The obtained results were compared with the results of empirical Meyer and Romanenko equations used in the estimation of evaporation amount. With this purpose, temperature, relative humidity, wind speed, and precipitation parameters of factors affecting evaporation were used as inputs and class A pan evaporation values were used as outputs. Data used belongs to Beysehir meteorology observation station.

Beysehir observation station is one of the three meteorology stations which make class A pan evaporation observations and represent the evaporation on lake surface area in Beysehir Lake catchment area and its immediate surroundings. As a result of the comparison of all obtained model performances, it was observed that all ANN models are more successful than ε-SVR and empirical Meyer and Romanenko methods. Nonetheless, both in ε-SVR and Meyer and Romanenko methods, an achievement of over 80 % was reached compared to R^2.

The best performance among utilized ANN algorithms was displayed with R^2=0.905 value in test period by the model which had SCG(4,2,2,1) network structure. Model results obtained by using the data belonging to Beysehir observation station have shown that artificial intelligence methods such as ANN and SVR have achieved considerable success in the estimation of air evaporation amount. The success of Meyer and Romanenko methods which estimate evaporation by using less parameter is also very close to that of ANN and SVR methods.

As the estimation of evaporation amount from water surface is quite difficult due to the complexity of evaporation mechanism and the abundance of factors affecting evaporation, the success of models obtained from such artificial intelligence methods is significant.

When the obtained results are evaluated, with models to be developed by using the data belonging to other two observation stations which are situated in Beysehir Lake Basin along with Beysehir observation station and which represent the evaporation on water surface, estimation of evaporation amount of Beysehir Lake, which holds great importance in providing both drinking water and irrigation water, shall be possible. Therefore, in terms of water potential of the lake reservoir, the capacity to estimate the evaporation amount is of considerable significance to take precautions against water shortages and develop strategies beforehand. Moreover, with these models to be formed, it shall be possible to make various hydrologic analyses and, hence, conduct studies with the purpose of improving water resources.

References

Ahmad I, El Naggar H, Kahn AN (2007) Artificial neural network application to estimate kinematic soil pile interaction response parameters. Soil Dyn Earthq Eng 27(9):892–905

ASCE Task Committee (2000) Artificial neural networks in hydrology, II: hydrological applications. ASCE J Hydrological Eng 5:124–137

Cheng MY, Cao MT (2014) Evolutionary multivariate adaptive regression splines for estimating shear strength in reinforced-concrete deep beams. Eng Appl Artif Intell 28:86–96

Cimen M, Kisi O (2009) Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. J Hydrol 378: 253–262

Ekici S (2007) Elektrik Güç Sistemlerinde Akıllı Sistemler Yardımıyla Arıza Tipi ve Yerinin Belirlenmesi. PhD Thesis. Fırat Üniversitesi. Fen Bilimleri Enstitüsü. (in Turkish)

Ertac M, Firuzan E, Solum S (2014) Forecasting Istanbul monthly temperature by multivariate partial least square. Theor Appl Climatol 117:1–13

Fausett L (1994) Fundamentals of Neural Networks: Architectures, Algorithms and Applications. Prentice Hall, Inc. A Simon&Schuster Company

Fu L (1994) Neural Networks in Computer Intelligence. McGraw-Hill, Singapore

Goh AT, Goh SH (2007) Support vector machines: their use in geotechnical engineering as illustrated using seismic liquefaction data. Comput Geotech 34(5):410–421

Goyal MK (2014) Monthly rainfall prediction using wavelet regression and neural network: an analysis of 1901–2002 data, Assam, India. Theor Appl Climatol 118:25–34

Ham FM, Kostanic I (2001) Principles of neurocomputing for science and engineering. Mc-Graw Hill, New York

Hamidi OH, Poorolajal J, Sadeghifar M, Abbasi H, Maryanaji Z, Faridi HR, Tapak L (2014) A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. Theor Appl Climatol. doi:10.1007/s00704-014-1141-z

Han H, Felker P (1997) Estimation of daily soil water evaporation using an artificial neural network. J Arid Environ 37:251–260

Haykin S (1999) Neural networks: a comprehensive foundation, 2nd edn. Prentice Hall, Upper Saddle River

Kalin L, Isik S, Schoonover JE, Lockaby BG (2010) Predicting water quality in unmonitored watersheds using artificial neural networks. J Environ Qual 39:1429–1440

Karunanithi N, Grenney WJ, Whitley D, Bovee K (1994) Neural networks for river flow prediction. ASCE J Comput Civ Eng 8:201–220

Kim S, Singh VP, Seo Y (2014) Evaluation of pan evaporation modeling with two different neural networks and weather station data. Theor Appl Climatol 117:1–13

Kisi O (2009) Daily pan evaporation modelling using multi-layer perceptrons and radial basis neural networks. Hydrol Process 23: 213–223

Kisi O (2012) Modeling discharge-suspended sediment relationship using least square support vector machine. J Hydrol 456:110–120

Kisi O (2013) Least squares support vector machine for modeling daily reference evapotranspiration. Irrig Sci 31(4):611–619

Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J Hydrol 399(1–2):132–140

Kisi O, Uncuoglu E (2005) Comparison of three back-propagation training algorithms for two case studies. Indian J Eng Mater Sci 12:434–442

Koroglu MA, Ceylan M, Arslan MH, Ilki A (2012) Estimation of flexural capacity of quadrilateral FRP-confined RC columns using combined artificial neural network. Eng Struct 42:23–32

Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund L, Dodge Y, Dierks T (2007) Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). J Neurosci Methods 161:342–350

Lin H (2006) Support vector machines for regression and its application for prediction of machine degradation based on vibration signals. Thesis Master of Science. University of Alberta. Edmonton Alberta

McCuen RH (1998) Hydrologic analysis and design. Prentice Hall, EnglewoodCliffs, NewJersey

Meyer A (1915) Evaporation of surface water bodies. A compendium of Water Resources Technology. McGraw-Hill Book Company

Modarres R (2009) Multi-criteria validation of artificial neural network rainfall-runoff modeling. Hydrol Earth Syst Sci 13:411–421

Moller MF (1990) A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. Computer Science Department, University of Aarhus, Denmark

Mukherjee I, Routroy S (2012) Comparing the performance of neural networks developed by using Levenberg–Marquardt and Quasi-Newton with the gradient descent algorithm for modelling a multiple response grinding process. Expert Syst Appl 39:2397–2407

Principe JC, Euliano NR, Lefebvre WC (2000) Neural and adaptive systems: fundamentals through simulations. John Wiley& Sons Inc., New York

Riedmiller M, Braun H (1993) A direct adaptive method for faster backpropagation learning: the RPROP algorithm. Proceedings of the IEEE International Conference on Neural Networks. 586–591

Romanenko VA (1961) Computation of the autumn soil moisture using a universal relationship for a large area. Proc. Ukrainian Hydrometeorological Research Institute: 3. Kiev

Samui P, Dixon B (2012) Application of support vector machine and relevance vector machine to determine evaporative losses in reservoirs. Hydrol Process 26:1361–1369

Samui P, Mandla VR, Krishna A, Teja T (2011) Prediction of rainfall using support vector machine and relevance vector machine. Earth Sci India 4(4):188–200

Shahin MA, Maier HR, Jaksa MB (2004) Data division for developing neural networks applied to geotechnical engineering. J Comput Civ Eng 18:105–114

Shen J (2005) Fusing Support Vector Machines and Soft Computing for Pattern Recognition and Regression. Kansas State University Manhattan, Kansas, PhD Thesis

Shiri J, Kisi O (2011) Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data in the Khozestan Province (South Western Iran). J Irrig Drain Eng ASCE 137:412–425

Singh VP, Xu CY (1997) Evaluation and generalization of 13 mass transfer equations for determining free water evaporation. Hydrol Process 11:311–324

Skapura DM (1996) Building neural network. ACMPress, Addision-Wesley

Terzi O (2013) Daily pan evaporation estimation using gene expression programming and adaptive neural-based fuzzy inference system. Neural Comput Appli 23:1035–1044

Ulengin F, Topcu YI (2000) Knowledge-based decision support systems techniques and their application in transportation planning systems. Knowl-Based Syst 4:1403–1429

Vapnik V (1995) The nature of statistical learning theory. Springer, New York

Yavuz G, Arslan MH, Baykan OK (2014) Shear strength predicting of FRP-strengthened RC beams by using artificial neural networks. Sci Eng Composite Mat 21(2):239–255