Pure and Applied Geophysics

Check for updates

Estimation of Evaporation from Saline-Water with More Efficient Input Variables

SEYED MOSTAFA BIAZAR,[1] AHMAD FAKHERI FARD,[1] VIJAY P. SINGH,[2,3] YAGOB DINPASHOH,[4] and
ABOLFAZL MAJNOONI-HERIS[4]

Abstract—Selection of optimal model inputs is a challenge for non-linear dynamic models. The questions as to which inputs should be used for model development have been a challenge in practice. Despite its importance, the literature on comparison of different methods for choosing inputs for estimating evaporation from saline water is limited. In this study, used three methods namely the Gamma test (GT), entropy theory (EnT), and procrustes analysis (PA) for determining suitable variables for estimating saline water evaporation using non-linear models of artificial neural network (ANN). The weather station near Lake Urmia was used for this experiment. At this station, pans of different concentrations (500 g/L, 300 g/L, 100 g/L, 50 g/L, 20 g/L, 10 g/L, 5 g/L, and drinking water) were prepared. In addition to evaporation data, surface water temperature (measured for each pan separately), air temperature, mean cloudiness, sunshine hours, mean relative humidity, mean wind speed, solar radiation, maximum wind speed, station pressure, mean station vapor pressure, maximum and minimum temperatures, and precipitation were also used. Model results were compared with field measurements and model performance was evaluated by the coefficient of correlation, root mean square error, and Nash–Sutcliffe efficiency coefficient. The most important variables identified by GT were surface water temperature, air temperature, mean relative humidity, mean wind speed, mean station pressure, minimum temperature, precipitation, mean station vapor pressure, and solar radiation. Also, as can be seen the most important variables for evaporation from saline water using the EnT method were water surface temperature, wind speed, and precipitation. The three important variables in the estimation of saline water, evaporation selected by the PA method, were air temperature, sunshine hours, and mean wind speed. According to

results, as the concentration increased, the mean station vapor pressure and temperature variables had the most influence on saline water evaporation. The uncertainty of model output was determined using the 95 percent prediction uncertainty (95PPU or P-factor) and d-factor. Although ANN-GT and ANN-EnT showed better goodness-of-fit metrics, ANN-PA had the lowest uncertainty among the three models in estimating evaporation from saline water. Generally, the PA method was able to demonstrate acceptable performance over the other two methods, with the least number of input variables.

Keywords: Artificial neural network, Evaporation from saline water, Entropy theory, Gamma test, Procrustes analysis.

1. Introduction

Literature and observations show that a significant portion of lake water is lost through evaporation in arid and semi-arid areas (Chattopadhyay et al. 2009; Biazar et al. 2019). Evaporation from the free water surface has been measured by direct and indirect methods. Direct methods include Eddy covariance (EC) method, Bowen ratio energy balance, and measurement of water loss with lysimeters or mas-balance methods (Allen et al. 2011). Indirect methods include experimental and theoretical equations (Meyer 1915; Penman 1948; Harbeck 1955; Priestly and Taylor 1972; Brutsaert and Stricker 1979; De Bruin and Keijman 1979), and artificial intelligence (Piri et al. 2009; Kisi et al. 2015; Vaheddoost and Koack 2019; Ashrafzadeh et al. 2019) and satellite-based methods (Zhao and Gao 2019; Holmes 2019). Direct measurement of evaporation is expensive, cumbersome, and has different sources of error, and becomes even more difficult for evaporation from saline water with different concentrations. Climatic variables, influencing saline water evaporation, not only vary from station to station, but also vary based

[1] Department of Water Engineering, University of Tabriz, 29 Bahman St., 14, 51666-16471 Tabriz, Iran. E-mail: seyed-mostafa.b@gmail.com; SM.Biazar@tabrizu.ac.ir; affard312@yahoo.com

[2] Caroline & William N. Lehrer Distinguished Chair in Water Engineering, Department of Biological and Agricultural Engineering, Texas A&M University, 321 Scoates Hall, 2117 TAMU, College Station, TX 77843-2117, USA. E-mail: vsingh@tamu.edu

[3] Zachry Department of Civil & Environmental Engineering, Texas A&M University, 321 Scoates Hall, 2117 TAMU, College Station, TX 77843-2117, USA.

[4] Department of Water Engineering, University of Tabriz, 29 Bahman St., 14, 51666-16471 Tabriz, Iran. E-mail: din-pazhoh@tabrizu.ac.ir; majnooni@tabrizu.ac.ir

on the degree of salinity. Therefore, different models can be derived for saline water evaporation. The rate of evaporation from saline water surface is usually less than that from fresh water (Asmar and Ergenzinger 1999). Many of the world's saline lakes are shrinking, with the result that bird habitats are undermined, people living around the lakes suffer with economic losses, and dusty winds are being caused which threaten people's health (Wurtsbaugh et al. 2017).

Many investigators have estimated evaporation from lake water surfaces (Gianniou and Antonopoulos 2007; Piri et al. 2009; Hamdani et al. 2018; Wang et al. 2018a, b, Mor et al. 2018; Ashrafzadeh et al. 2019). The mechanisms of evaporation from saline water are somewhat different from those from fresh water. The complexity of the process of evaporation from water bodies (e.g., lakes, ponds, dams, and natural reservoirs) and the lack of sufficient and reliable information are the main obstacles (Sandler 1999; Hamdani et al. 2018). Additionally, the selection of meteorological variables that influence saline water evaporation is a difficult task.

This study attempts to find the main important meteorological variables that affect the rate of daily evaporation from saline water. This task was performed using three models which were (1) procrustes approach (PA) (Dinpashoh et al. 2004), Gamma Test (GT) (Ashrafzadeh et al. 2018), and entropy based method (EnT) (Ahmadi et al. 2009). Once the variables were selected, the artificial neural network (ANN) method was applied to model daily evaporation rate using daily meteorological time series.

Lee (1927) experimentally compared evaporation from pure water and from brine lake in Nevada, USA, for different densities and found that the brine water evaporation in comparison with pure water reduced by 0.01% for a 1% increase in water specific weight. Young (1947) compared evaporation from both saline water (with different concentrations of sodium chloride) and that from fresh water. Results of this study did not differ significantly from the findings of Lee (1927). Kokya and Kokya (2008) analyzed the effect of salinity of water on evaporation rate from a water pan in an experimental study near the Urmia Lake by using Meyer and Harbeck methods, and they developed these methods. The results showed that the

developed methods could estimate evaporation with higher accuracy than classical methods. Al-Khalifat (2008) reported that the average volume of evaporation losses from the Dead Sea was between 2 and 4 billion m^3/year using the mass and energy balance during 1800–2000. He also showed that the decrease in the lake level over time had resulted in reducing evaporation rate. Furthermore, the evaporation rate had decreased due to increasing salinity during the periods with significantly lower lake water level. Piri et al. (2009) investigated the evaporation of free water in hot and dry regions of Iran. In their research, they used an ANN model to estimate evaporation and compared the results with two experimental models (Linacre and Marciano). They used a GT to select the best input combination for the ANN. Wind speed, saturation vapor pressure, and mean relative humidity were selected as the most effective input variables in estimating evaporation. Results showed the ANN-GT model outperformed the experimental models. Moghaddamnia et al. (2009) estimated, sing ANN and adaptive neuro fuzzy inference system (ANFIS), evaporation from the free water surface in Iran. They used data on air temperature, wind speed, saturation vapor pressure, and mean relative humidity. The GT was also used to select the effective input variables. In this study, variables such as wind speed, relative humidity, and saturation vapor pressure were selected as the most effective variables by the GT. They used three empirical models of Hefner, Linkier, and Marciano to compare the model performances. Results showed that the intelligence models performed better than the experimental models. Goyal et al. (2014) modeled tropical evaporation using ANNs, least squares-support vector regression (LS-SVR), fuzzy logic, and ANFIS. They used precipitation data, maximum and minimum temperatures, maximum and minimum relative humidity, and sunshine hours. They also used the GT to select the effective inputs to improve model performance. Results were compared with the experimental models of Hargreaves and Samani and Sphere Stewart, and showed that the intelligence models performed better than the experimental models. Kisi et al. (2015) predicted the Urmia Lake water level with the support vector machine (SVM) model by applying a novel method based on firefly algorithm (FA). The output of the

SVM–FA model was compared to the Genetic Programming (GP) and ANNs models. Results showed that the SVM-FA model was superior to others. Shiri et al. (2016) predicted the water level in the Urmia Lake using the Extreme Learning Machine (ELM) method. In their study the ELM method was compared with ANN and GP. Results showed that the ELM method was better than others. Wang et al. (2018a, b) evaluated different approaches for estimating evaporation in a small high-elevation lake on the Tibetan Plateau (TP). They conducted the evaluation by using EC observation-based reference datasets. They used Bowen ratio energy budget, Penman, Priestley–Taylor, Brutsaert–Stricker and DeBruin–Keijman, Dalton and Ryan–Harleman, Jensen–Haise, and Makkink methods. All methods were significantly improved after parameter optimization, with better efficiency by the former than the latter. The deBruin method yielded the largest error due to the poor relationship between evaporation and the drying power of the air. The good efficiency of the Makkink approach, with no considerable differences before and after optimization, indicated the significance of solar radiation and air temperature in the estimation of lake evaporation. The Makkink method was used for long-term evaporation estimation due to the lack of water temperature observations in lakes on the TP. Wang et al. (2018a, b) carried out research about global lake evaporation accelerated by changes in surface energy allocation in a warmer climate. They reported simulations with a numerical model of lake surface fluxes, with input data based on a high-emission climate change scenario (Representative Concentration Pathway 8.5). In their simulations, the global annual lake evaporation increased by 16% by the end of the century, despite little change in incoming solar radiation at the surface. At the current work, we investigated the effect of variables on evaporation in different concentrations (fresh water to high saline water) separately. Seifi and Riahi (2018) estimated daily reference evapotranspiration using the least square support vector machine (LS-SVM)-GT, ANN-GT and ANFIS-GT models in Iran. Results of GT revealed that three climate variables which were minimum air temperature, maximum air temperature, and wind speed were the most important climatic variables, and

the LSSVM model performed better than did ANFIS and ANN for the same meteorological input variables. Hamdani et al. (2018) assessed seasonal and daily evaporation of the deep and brine Dead Sea. They provided observations in two consecutive years using the eddy covariance system, meteorological stations, and a floating station (which measured the water profile temperature). They found that the peak of evaporation rate occurring in summer was related to solar radiation. The winter peak evaporation was also related to the thermal storage of the lake, and evaporation due to high vapor pressure combined with wind and resulting thermal static. They compared several models of evaporation with direct measurements and showed that the mass transfer model was more reliable than others. Nozari and Azadi (2019) predicted the salinity of drainage and groundwater at various drain depths and spaces using ANN. Results revealed that ANN had a reasonable accuracy in the simulation of temporal shallow groundwater and drainage of water salinities at different drain depths and drain spaces. Guo et al. (2019) conducted research on long-term changes in evaporation over Siling Co Lake on the Tibetan Plateau and its impact on recent rapid lake expansion. In this study, long-term evaporation over Lake Siling Co was simulated using a single-layer lake evaporation model, and simulated results were verified by observation from an EC https://www.science direct.com/topics/earth-and-planetary-sciences/eddy-covariance system in the lake. Results showed that the single-layer lake evaporation model was capable of accurately simulating lake evaporation on a daily scale. Ashrafzadeh et al. (2019) estimated evaporation of free water in northern Iran using intelligent models, ANN and ANN-krill herd optimization algorithm (ANN-KHA). Maximum temperature, maximum and minimum relative humidity, rainfall, wind speed, and sunshine were used. They also used the GT to select the effective inputs and results showed that the ANN-KHA model performed better than did the ANN model.

Since it is difficult to answer which variables are useful as inputs for modeling (Ahmadi et al. 2009; Isazadeh et al. 2017; Ashrafzadeh et al. 2019), there have been insufficient reports on comparing different techniques for selecting the most effective inputs.

Although much research has been done on evaporation estimation, most of it does not consider effective input selection for evaporation modeling. As a result, the main challenge lies in the evaluation of existing data and assessing the adequacy of data. Although modeling improves performance by adding information, observations show that adding information improves model performance only slightly. Modeling accuracy can be reduced by increasing information. This is because additional information causes the model to be overfitted. The overfitted model also performs well in training but has a very poor performance in testing. Overfitting also occurs when multivariate models have too much input data. It is therefore important to know which inputs are effective in modeling and which are not (Ahmadi et al. 2009). Remesan et al. (2008) used GT in determining the effective inputs for estimating solar radiation using local linear regression (LLR) and ANN in the Brue Catchment in southwest England. Ahmadi et al. (2009) applied GT, Entropy Theory (EnT), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) in the determination of effective inputs for solar radiation estimation in the Brue Catchment in England. They estimated solar radiation with LLR and ANN and found that the GT had the best performance. A limited number of studies have used the procrustes analysis (PA) for the selection of model inputs. Dinpashoh et al. (2004) used this approach in selecting variables in a precipitation climate study in Iran. They found that among the 57 candidate variables, only 12 variables were important in the regionalization of precipitation records. Nam et al. (2015) used the PA in the selection of 33 rainfall-related and geographical variables among the total 42 candidate variables for the delineation of rainfall in South Korea.

In this study, the PA approach was used in the determination of evaporation rate from saline water. Results were compared with those of the GT and EnT schemes. The main objectives therefore were to: (1) identify the best combination of climatic variables for modeling evaporation rate of saline water, (2) identify the best model among the three candidate models which were PA, GT, and EnT using performance criteria.

2. Materials and Methods

2.1. Study Area and Data

Lake Urmia is one of the most significant saline lakes of the world and Iran's largest saline lake, which is at risk of drying up because of excessive agricultural development, climate change, and irrational construction of dams. Various reports have been presented regarding evaporation from the lake surface, with the values generally estimated to be in the range of 890–1360 mm/year (more than 50% difference). However, just one centimeter of error in estimating the evaporation height leads to 30 million cubic meters of error in calculating the lake water balance, considering the average lake area of 3000 km^2. This shows the necessity of accurately estimating the evaporation rate from Lake Urmia using physically-based approaches and accurate observed data.

For modeling the evaporation rate from saline water having different salt concentrations, an experimental study was carried out at a location near the Urmia Lake. Recorded daily data of meteorological variables used in modeling were: (1) surface water temperature, (2) mean air temperature (T_{mean}), (3) mean sky cloudiness, (4) sunshine hours, (5) mean relative humidity (RH_{mean}), (6) mean wind speed, (7) mean air pressure, (8) maximum temperature, (9) minimum temperature, and (10) precipitation. Measurements at daily scale were used for evaporation rate of saline water with different concentration levels. Nine different samples (with salt and water) were made with different concentrations. The sample concentrations were 0.2 g/L (as a fresh water), 5 g/L, 10 g/L, 20 g/L, 50 g/L, 100 g/L, 200 g/L, 300 g/L, and 500 g/L, exposed to free evaporation in the field. This was accomplished using a sensitive electrical conductivity meter (Lide 2004). Also, the total dissolved solids (TDS) was kept unchanged during measurement. The evaporation rate was measured at daily scale from March 1, 2019 to June 30, 2019. In order to keep the water surface clear from algae and other thin films the water surface was cleaned from time to time, since any dust precipitation over the water surface would retard evaporation (El-Dessouky
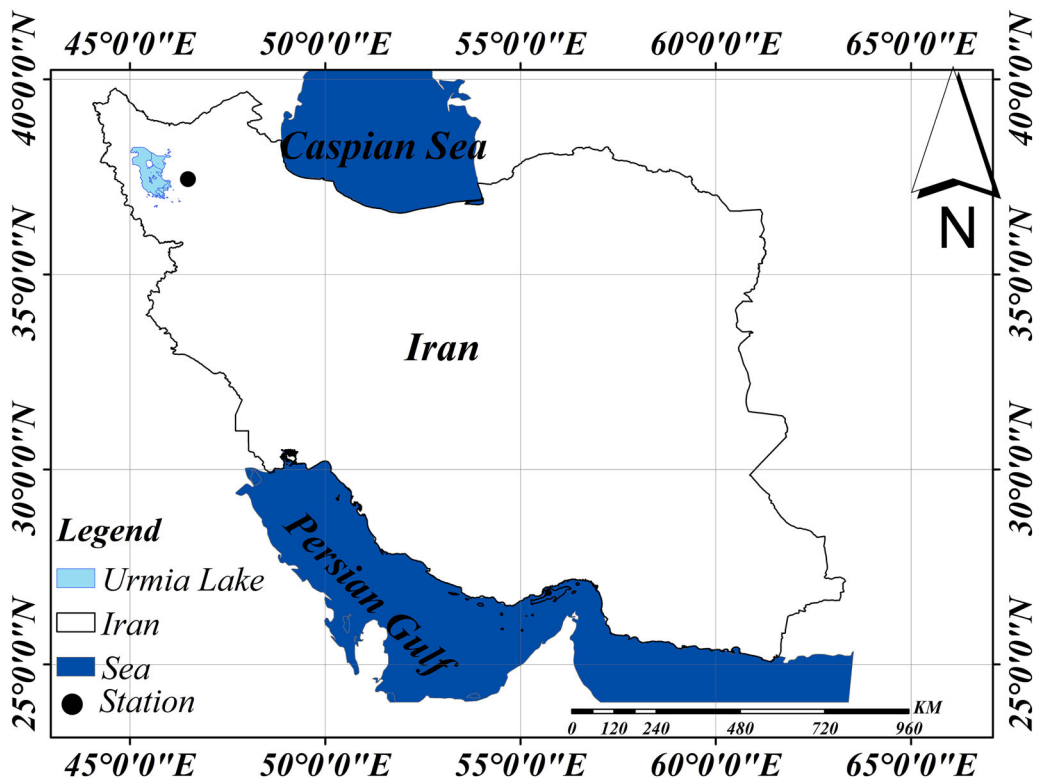
Figure 1
Study area

et al. 2002). The observed evaporation rate was recorded along with the corresponding climatic data.

First, a place closest to Lake Urmia where a station could be built was selected for the study (We used Basmenj weather station). Then, the station was equipped with Pan Evaporation, and waters with different concentrations were monitored to measure the evaporation. Urmia lake salt and distilled water were used for conducting this experiment, because the distilled water concentration was known, and it was possible to easily control the concentration of Pan Evaporation with electrical conductivity meter. The water level in all Pan Evaporation was kept near the Pan Evaporation surface. A separate scale was prepared for each Pan Evaporation, and the numbers on them were read every day. Distilled water was added to the pan every few days to reach the initial level of water and salinity (El-Dessouky et al. 2002;

Lide 2004). After adding water to the Pan Evaporation, they were stirred with separate plastic tubes designed for each Pan Evaporation, and then the tubes were washed using distilled water and the water used to wash them was poured into the Pan Evaporation again, because it was intended to control the Pan Evaporation concentrations (On the days when distilled water was added to the barrels, the scales were read once before adding the water and once after adding the water and stirring).

2.2. Theory of Entropy (EnT)

Entropy theory was developed by Shannon (1948) which states that information reduces uncertainty and vice versa. Entropy can be used as an index of quantification of lack of knowledge of system characteristics. In general, four types of entropy have

been introduced in the quantification of information: marginal entropy, joint entropy, conditional entropy, and information transfer entropy. In order to select the suitable set of inputs for modeling evaporation from saline water, information transfer entropy was used. The marginal entropy, called H(x), based on a discrete random variable, x, can be defined as (Shannon and Weaver 1949):

$$H(x) = -K \sum_{i=1}^{n} P(x_i) \log P(x_i) \qquad (1)$$

where the value of K is equal to 1, if in Eq. (1) the logarithm is based on 2.7182; and n is the number of events with $p(x_i)$; (i = 1, …, N).

The entropy for the sum of the two random independent variables, x and y, can be estimated as (Ahmadi et al. 2009; Saha and Chattopadhyay 2020):

$$H(x, y) = H(x) + H(y) \qquad (2)$$

In the case of two dependent random variables, x and y, the joint entropy is less than the sum of the two distinct entropies (i.e. H(x) + H(y)).

The information transfer entropy specifies the mutual relationship between x and y. The information transfer entropy [denoted by T(x, y)] for the two dependent variables (i.e. x and y) can be obtained as (Ahmadi et al. 2009):

$$T(x, y) = H(x) + H(y) - H(x, y) \qquad (3)$$

which can also be estimated as:

$$T(x, y) = H(x) - H(x|y) = H(y) - H(x|y) \qquad (4)$$

where H(x|y) denotes the conditional entropy. The term H(x|y) expresses the residual uncertainty of x provided that y is known and vice versa.

The calculation of entropy for the M distinct variables can be generalized in a similar manner (Harmancioglu and Alpaslan 1992; Singh 2011). In this case the total entropy for the M independent variables X_m (m = 1, 2, …M) can be obtained as:

$$H(x_1, x_2, etc., x_M) = \sum_{m=1}^{M} H(x_m) \qquad (5)$$

For dependent variables, the joint entropy can be computed as (Harmancioglu and Alpaslan 1992; Singh 2011; Chattopadhyay et al. 2018):

Table 1

Characteristics of station sued in this study

Concentrations (g/L)	Evaporation (mm/day)	Surface water temperature (°C)	Air temperature (°C)	mean cloudiness (Octa)	Sunshine (Hr)	Mean relative humidity (%)	Mean wind speed (M/s)	Station pressure (Pa)	Maximum temperature (°C)	Minimum temperature (°C)	Precipitation (mm/day)	Mean station vapor pressure (pa)	Solar radiation (j/cm²/day)	Maximum wind speed (m/s)
Fresh water	7.4	18.5	14.32	4.1	8.6	52.8	3.4	863.73	20.1	8.85	1.3	8	2254.17	7.23
5	7.14	18.2												
10	7.1	18												
20	7.12	17.7												
50	6.8	17.5												
100	6.44	17.5												
300	5.47	17.4												
500	4.82	17.4												

$$H(x_1, x_2, etc., x_M) = H(x_1) + \sum_{m=1}^{M} H(x_m|x_1, etc., x_{m-1})$$

$$(6)$$

In order to select the most suitable variables, the value of T(x, y) was calculated for all the 14 variables. Then, one of the variables was removed from the set of x, and T(x, y) was evaluated for this reduced set. This process was repeated for the other 13 variables. The computed values of T(x, y) were tabulated for different concentrations of saline water (i.e. y). For a distinct concentration (for example, 5 mg/L), those variables, for which T(x, y) was more than that of the corresponding T(x, y) obtained using the total 14 variables, the redundant variables were selected and then were removed from the set. Thus, the most appropriate variables were selected in modeling saline water evaporation.

2.3. Gamma Test (GT)

GT is a nonlinear approach which supposes that the observation set is defined by the following relationship (Durrant 2001):

$$\{(x_i - y_i)\}, 1 \le i \le M\} \qquad (7)$$

where x_i is the input observation vector (here, surface water temperature, air temperature, …), and y_i is the output of the GT (here, evaporation rate from a distinct concentration), and M is the total number of observations (here, total days for which evaporation was measured). The relationship between the inputs and output can be expressed as:

$$y = f(x) + r \qquad (8)$$

where f(x) is the smooth variable and r is the error term. It is assumed that the mean of the distribution fitted to r is zero and its variance is limited. The gamma statistic denoted by (Γ) expresses the variance of those observations, which the model is incapable in its determination. For a distinct input vector xi, the set N[i, k], for ($1 \le k \le p$), is called the set of closest neighbors.

The gamma test is based on this set (i.e. N [i,k]). The term X $_{[I,k]}$ is the closest neighbor for ith x (i.e. x_i), such that ($1 \le k \le p$), and($1 \le i \le m$). Also p is the maximum number of neighbors that are usually

assumed to be between 10 and 50. In order to estimate (Γ) the values of ($\delta_M(k)$) should be calculated according to the input data (Evans and Jones 2002).

$$\delta_M(k) = \frac{1}{M} \sum_{i=1}^{M} |x_{N[i,k]} - x_i|^2, \quad 1 \le k \le p \qquad (9)$$

where $|\ldots|$ denotes the Euclidean distance between vector $x_{N[i,k]}$ and its neighbors. Also, the value of $\gamma_M(k)$ can be estimated using the output data from the following relationship:

$$\gamma_M(k) = \frac{1}{2M} \sum_{i=1}^{M} |y_{N[i,k]} - y_i|^2, \quad 1 \le k \le p \qquad (10)$$

where $y_{N[i,k]}$ is the value of output corresponding to the kth neighborhood of x_i vector. In this manner the p values for $\delta_M(k)$ and the p values for $\gamma_M(k)$ can be calculated. Then a relationship between $\{\delta_M(k), \gamma_M(k)\}$ would exist as:

$$\gamma = A\ \delta + \Gamma \qquad (11)$$

The value of Γ statistic is indeed the intercept of the above mentioned regression model. Also, A is the slope of the line which shows the complexity of model derived from data (Isazadeh et al. 2017; Ashrafzadeh et al. 2018).

2.4. Principal Components Analysis (PCA)

PCA is one of the oldest multivariate methods that was introduced by Pearson in 1901 and modified later by Hotelling in 1933 (Hotelling 1933; Jolliffe 1986). The main idea in PCA is the reduction of dimensions of data that feature a relatively high correlation (in terms of modulus). The reduction of dimensions by converting data into a few new variables, namely PCs. These new variables are independent of each other. The first PC can be known as the first independent variable which has the largest variance compared with the other PCs. The second PC incorporates the largest variance following the PC1 among others. In the same role the third PC has a large variance after PC1 and PC2 among others. The sum of variances of all the PCs is 100%. In PCA the first few PCS usually incorporated a large variance for data which led to easy interpretation of data.

As stated by Johnson and Hanson (1995), instead of directly using the input variables, which are usually dependent on each other, the variables can be converted to the few independent PCs that can be subsequently applied for data interpretation. In this method, only negligible information would be lost after converting the original variables into the PCS (Johnson and Hanson 1995).

2.5. Procrustes Analysis (PA)

Assuming p to be the number of total variables, and k the number of the dominant PCs in PCA, the PA selects a set of important variables which show the feature of nearly the whole number of PCs. At the same time the number of selected variables (q) is much less than p (total candidate variables) and is less than k (selected PCs). This property of PA which at last selects less variables (say q) helps with interpretation of the whole data. The retained q variables incorporate the feature of nearly all candidate variables. This method was introduced by Krzanowski (1987) and then applied in hydrology by Dinpashoh et al. (2004).

In this method, it is necessary to select q as a number of important variables to be much less than total candidate variables (i.e. p), that is, q < < p. By the implementation of PCA to original variables the number of important PCs was selected to be k, in which k < p. In PA, the selection of q variables was done by the minimization of the objective function denoted by M^2. This function can be defined as follows:

$$M^2 = Trace\{YY' + ZZ' - 2ZQ'Y'\} \qquad (12)$$

Figure 2 illustrates the general stages of procrustes analysis.

In Fig. 2, X is the matrix of the standardized data and $\underline{X}$ is the reduced matrix (standardized), i.e., after removing a most redundant variable from X. Y is the output of PCA conducted on X. and Z is the output of PCA carried out for $\underline{X}$. Y is the real feature of original data, but Z is called the approximate feature of data. PA evaluates the differences between the two matrices (Y and $\underline{X}$) by means of differences between the sums of squares of the corresponding points in the two mentioned arrangements. If the variables were selected precisely, the similarity between the two
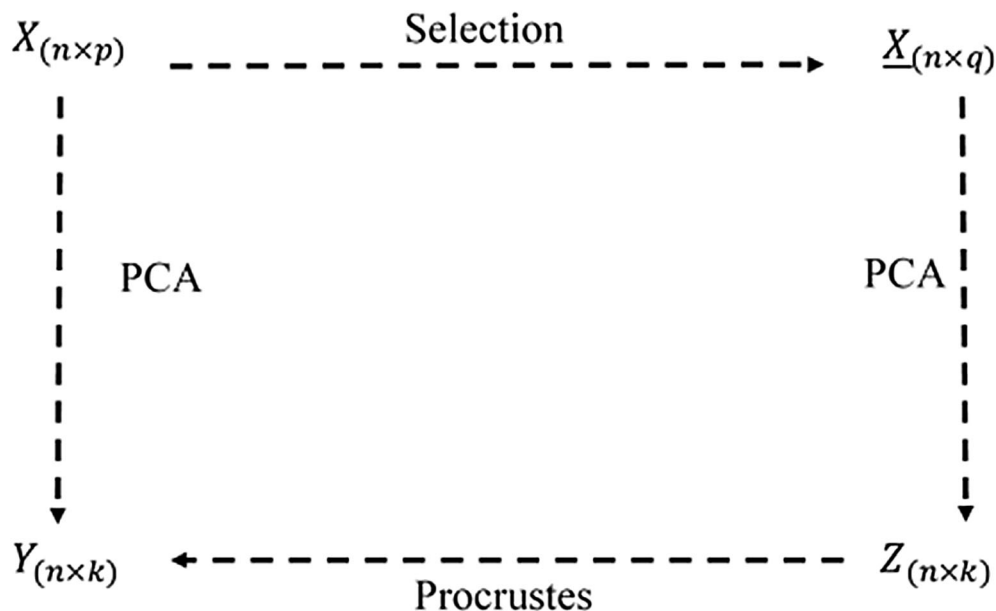
Figure 2
The cyclone of variable selection using PA method (Dinpashoh et al. 2004)

arrangements would be large, therefore, the differences between the approximate and real configuration would be less. In Eq. (12), Q is obtained as:

$$Q = VU' \qquad (13)$$

where matrices U and V can be obtained using the singular value decomposition (SVD) of the matrix $Z'Y$ having dimension of k × k. The SVD relationship is as follows:

$$Z'Y = U \sum V' \qquad (14)$$

where $UU' = I_k,$ $V'V = VV' = I_k,$ and $\sum = diag(\sigma_1, \sigma_2, \ldots, \sigma_k).$ I_k is the square identity matrix with dimensions k × k.

M^2 was determined for any of the subsets pertaining to the set of variables with at least q members. The best subset of the variables' reference set was selected in a repeated manner by which at last the lowest amount of M^2 was obtained and the retained variables were defined to be the most important variables (Krzanowski 1987).

3. Artificial Neural Network (ANN)

ANNs are considered as one of the data processing methods, including input/output processing and generally one or more hidden layers as their main components. Even though there are various neural network architectures, about 90% of them are the feed-forward type (Coulibaly et al. 2000; Aghelpour et al. 2019). A feed-forward ANN can have one or more hidden layers which their nodes are called as hidden units or nodes. The ANN can solve high ordered statistical problems by including hidden layers. The ANN can have several outputs, however, in the presented study one output was utilized. The perceptron, the very basic form of an artificial neural network, is a binary classifier and can be described using the following equation:

$$g(z) = \begin{cases} 1 & if \ z = w \times X + b > 0 \\ 0 & otherwise \end{cases} \qquad (15)$$

where $g\ (z)$ is the Heaviside step function (i.e., a limited activation function), w is the weight vector, b is the perceptron parameter deviation, and X is the input vector. Multilayer perceptron (MLP) networks learn to simulate the treatment of a wrapped, nonlinear system through learning algorithms and observed data. There are different kinds of learning algorithms, the most famous of which is back-propagation (Rumelhart et al. 1988), delta-bar-delta (Jacobs 1988), quick-prop (Fombellida and Destiné 1992), conjugate gradient (Charalambous 1992), and Levenberg–Marquardt (Hagan and Menhaj 1994; Deo et al. 2018; Ashrafzadeh et al. 2019). These learning algorithms are generally used to detect the optimal set of MLP model parameters. This study used the MLP model along with learning algorithm Levenberg–Marquardt learning algorithm. 1 to 20 neurons in the hidden layer are employed to evaluate the effects of network structure on its performance in RS simulation. The "Trial and Error" method has been used to obtain the optimal neuron. The sigmoid tangent function is applied to map information from the input layer to the hidden layer and from the hidden layer to the output layer (Lagos-Avid and Bonilla 2017; Naganna et al. 2019; Ashrafzadeh et al. 2020). Moreover, this study used 70% of the data as the training set and the remaining 30% as the test set.

3.1. Model Performances

Three metrics were applied to assess the model performances, including the (1) coefficient of correlation (CC), (2) root mean square error (RMSE), and (3) Nash–Sutcliffe model efficiency coefficient (NS) as you know One of the most popular evaluation indices is the Nash Sutcliffe Index, whose range varies from 1 to negative infinity. The intervals of 0.75 1, 0.36–0.75 and less than 0.36 for this index in a simulation, respectively, show very good, satisfactory, and poor performance (Nash and Sutcliffe 1970; Isazadeh et al. 2017; Ashrafzadeh et al. 2018).

$$CC = \sqrt{\frac{(\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})^2)}{(\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2)}} \qquad (16)$$

$$RMSE = \sqrt{\frac{(\sum_{i=1}^{N}(x_i - \bar{y})^2)}{N}} \qquad (17)$$

$$NS = 1 - \sqrt{\frac{(\sum_{i=1}^{N}(x_i - \bar{y})^2)}{(\sum_{i=1}^{N}(x_i - \bar{x})^2}} \qquad (18)$$

where x_i is the ith observed value, $\bar{x}$ is the mean of observations, y_i is the ith value estimated from the model, $\bar{y}$ is the mean of the estimated values, and N is the number of observations.

3.2. Uncertainty Analysis

In this study we used an approach presented by Abbaspour et al. (2007) to analyze uncertainty in the predict evaporation from saline water. In this method, the percentage of measured data bracketed by 95 percent of predicted uncertainties (95PPU) calculated by the 2.5th (Xl) and 97.5th (Xu) percentiles of normal distribution function obtained from n (in this study 1000) times of the simulation results as follows

$$\text{Bracketed by 95PPU} = \frac{1}{k}\text{count}\left(j|X_L^l \leq X_{reg}^l \leq X_U^l\right) \times 100$$

(19)

where l is the item number from one to k, X_{reg}^l is the observed value on day l, and j is the numerator parameter of the number of observed values placed on the 95% prediction uncertainties (95PPU) band. If all values are within the confidence band of uncertainty, they are then bracketed by 95PPU = 100 (Abbaspour et al. 2007; Isazadeh et al. 2017; Biazar et al. 2019, 2020).

Also, 95 percent prediction uncertainties (95PPU or P-factor) and d-factor coefficients were presented to quantify the authority of calibration and uncertainty analyses (Abbaspour et al. 2007). Equation (20) was used to determine the average width of the band (d-factor) index:

$$d - factor = \frac{\overline{d}x}{\sigma x}$$

(20)

where σx is the normal deflection of the observed data and $\overline{d}x$ is the average width of the confidence interval that can be achieved by

$$\overline{d}x = \frac{1}{k}\sum_{t=1}^{K}(X_U - X_L)$$

(21)

where t = 1,…, K is the number of observed data, X_U is the 97.5th percentile of model output, X_L is the

2.5th percentile of model output, and K is the number of observation data.

4. Results

The experiment was conducted in an environment close to Lake Urmia and the environmental conditions were the same for all pans. In this study, three methods were used separately to determine the input variable or select the appropriate input composition and to estimate evaporation using an ANN. Finally, results were compared. To estimate the evaporation by the ANN model, the input variables were selected based on the performance characteristics of each method. The performance and specificity of each of the preprocessing methods are described in Sect. 2. This means that each of the three methods, depending on their performance characteristics, had determined the input variables or appropriate input composition for estimating evaporation at different concentrations. Therefore, the variables in the input composition selected by each of the methods were different for each concentration in the same way as Asmar and Ergenzinger (1999).

4.1. Entropy Theory (EnT)

In the entropy section, the effective input variables were used to estimate evaporation for each pan separately. Thus, the entropy value was first calculated for all selected input variables (row 1 in Table 2). Then, among all input variables one variable was eliminated and the Entropy Test was conducted for the remainder of variables. In this process all input variables were examined by removing the variables one by one. Then, the output of the entropy method was included (row 2 to 15 in Table 2). The first combination (row 1) containing all the variables was designated as the base, meaning that the subsequent combinations were compared with the first combination containing the most information (all variables). If the entropy value of the next compound (2–15) was lower than the first compound, the omitted variable contained information that was effective in estimating the evaporation,

Table 2

Summary of entropy values obtained for different inputs of the model in estimating evaporation from saline water with different concentrations

No.	Eliminated parameters	Trans-information concentration (g/L)							
		5	10	20	50	100	300	500	Fresh water
1	–	1.099	2.194	1.893	1.034	1.954	2.014	3.001	2.985
2	Surface water temperature	**0.996**	**1.032**	**0.781**	**0.781**	**0.987**	**1.458**	**2.502**	**1.006**
3	Air temperature	1.770	**2.097**	2.246	2.246	3.001	3.241	**1.942**	3.127
4	Cloudiness	2.628	3.528	2.421	2.421	2.789	2.945	3.832	3.432
5	Sunshine	3.625	3.625	2.471	2.471	2.471	2.871	3.195	2.994
6	Mean relative humidity	**0.708**	2.807	2.997	2.997	3.104	**1.145**	3.731	3.010
7	Mean wind speed	**0.785**	**0.986**	1.407	1.407	**1.007**	**1.975**	**1.035**	**2.325**
8	Station pressure	3.754	**1.052**	2.123	**2.123**	2.578	2.418	**1.855**	3.014
9	Maximum temperature	2.711	3.502	3.001	3.001	2.897	2.145	3.483	3.049
10	Minimum temperature	**0.859**	1.958	2.914	**2.914**	3.492	**1.921**	**1.192**	3.442
11	precipitation	**0.885**	**0.995**	**0.874**	**0.874**	**0.997**	**0.878**	**1.154**	**2.139**
12	Mean station vapor pressure	3.754	3.543	2.784	**2.784**	**1.014**	2.244	**0.548**	2.996
13	Solar radiation	**0.911**	2.809	3.129	3.129	3.012	3.173	**2.476**	2.973
14	Maximum wind speed	**0.859**	1.984	**1.454**	**1.454**	**0.994**	2.098	**2.723**	**2.173**
15	SQRT ($T_{maximum} - T_{minimum}$)	3.285	2.587	3.015	3.015	3.173	3.413	4.000	3.571

The bold values indicate important variables in EnT

thus it was known as the effective variable (bold variable).

The entropy values of the information transfer for different variables in line with estimating water evaporation with different salinities are presented in Table 2. In this table, the input data were: water surface temperature, air temperature, degree of cloudiness, hours of solar radiation, relative humidity, wind speed, station pressure, maximum temperature, minimum temperature, precipitation, air vapor pressure, solar radiation, maximum wind speed, and the square root of maximum and minimum temperature (SQRT ($T_{maximum} - T_{minimum}$)) in Table 2, As can be seen the three most important variables for evaporation from saline water using the entropy method were water surface temperature, wind speed and rainfall. Air temperature, solar radiation, and relative humidity were important for only two different concentrations identified as 10 and 500 g/L for air temperature, 5 and 500 for solar radiation, and 5 and 300 for relative humidity). On the other hand, the variables of cloudiness degree, solar radiation, maximum temperature, and the square root of maximum and minimum temperature differences, at any of the concentrations, were not selected as effective input variables by the

information transfer entropy method. Other variables were also identified as effective input variables in 3 or 4 compositions by the entropy method.

In this section, the effect of salinity can be generally investigated by initially selecting water surface temperature, wind speed, and precipitation as the primary variables for low concentrations according to the entropy method. As the concentration increased, the station pressure and air temperature variables were added. At the concentration of 50 g/L, the mean station vapor pressure variable was also added to the initial variables according to entropy in the same way as Anderson (1936).

4.2. Gamma Test (GT)

The GT for each pan was used to determine the effective input variables to estimate evaporation. Therefore, first, the GT was performed for all data (row 1) and then this test was performed based on the elimination of each of the variables from all available variables (row 2–15), which is shown in Table 3. In the first-row GT method, in which all the variables in the composition were assigned as bases and the next row or combinations were compared with the first combination, if the gamma value of the next

Table 3

Summary of the output of GT for estimating daily evaporation using the normalized option

No.	Eliminated parameter in GT	The values of Gamma concentration (g/L)							
		5	10	20	50	100	300	500	Fresh water
1	–	0.030	0.033	0.035	0.036	0.029	0.033	0.025	0.043
2	Surface water temperature	**0.036**	**0.035**	0.034	**0.040**	**0.035**	**0.034**	**0.030**	0.042
3	Air temperature	**0.034**	**0.036**	0.034	**0.040**	**0.035**	0.030	**0.031**	0.042
4	mean cloudiness	**0.030**	0.027	0.031	0.033	0.025	0.030	**0.032**	0.037
5	Sunshine	0.029	**0.034**	**0.036**	0.035	**0.033**	0.031	**0.026**	0.042
6	Mean relative humidity	**0.036**	**0.040**	**0.039**	**0.039**	**0.037**	**0.036**	**0.032**	**0.044**
7	Mean wind speed	0.028	0.032	**0.042**	**0.044**	**0.045**	**0.048**	**0.034**	**0.050**
8	Station pressure	0.025	**0.033**	**0.035**	**0.043**	**0.032**	**0.035**	**0.042**	0.039
9	Maximum temperature	**0.032**	0.032	0.034	**0.041**	**0.033**	**0.036**	**0.030**	0.042
10	Minimum temperature	**0.037**	**0.042**	**0.041**	**0.043**	**0.038**	**0.037**	**0.035**	**0.047**
11	Precipitation	**0.032**	**0.038**	**0.040**	**0.046**	**0.035**	0.032	0.034	0.047
12	Mean station vapor pressure	**0.035**	**0.035**	0.032	**0.043**	**0.038**	**0.042**	**0.031**	0.041
13	Solar radiation	**0.031**	**0.036**	**0.036**	0.036	**0.032**	**0.033**	**0.025**	0.039
14	Maximum wind speed	0.026	0.030	**0.035**	0.033	0.026	0.030	**0.029**	0.035
15	SQRT $(T_{maximum} - T_{minimum})$	0.024	0.027	0.032	0.035	**0.031**	**0.041**	**0.029**	0.037

The bold values indicate the important variables in GT

compound (2–15) was greater than the first compound, the omitted variable was known as the effective variable (bold variable in the table).

Using GT, various combinations of input variables were identified for different Pans for improving the evaporation estimation. The important variables identified by GT were surface water temperature, air temperature, mean relative humidity, mean wind speed, mean station pressure, minimum temperature, precipitation, mean station vapor pressure, and solar radiation. This result was obtained in the case of six (or more) out of the eight concentrations (input parameter) used in this study (Table 3). The use of GT led to the elimination of mean wind speed for all the densities (except 20 (g/L) and 500 (g/L) concentrations). The parameter denoted by the square root of maximum and minimum temperature (SQRT ($T_{maximum} - T_{minimum}$)) eliminated by GT for densities of 10, 20, 50 (g/L) and fresh water. Results showed that maximum temperature had been eliminated just for three densities, namely 10, 20 (g/L) and fresh water (Table 3).

As can be seen, the main variables at low concentrations included the mean relative humidity, mean wind speed, and air temperature which were almost the same as Piri et al. (2009). As concentration increased, surface water temperature, sunshine hours, station pressure, mean station vapor pressure, and

solar radiation were added to the initial variables. Wind speed, temperature, and relative humidity were among the factors that contributed to evaporation (Moghaddamnia et al. 2009; Biazar et al. 2019). As the concentration of the solution increased, the salt in the solution absorbed the sun's radiation energy and broke its bonds. Therefore, solar radiation, sunshine hours, and surface water temperature were added to the initial variables. Also, at higher concentration, the soluble salts decreased the free energy of the water molecules and hence the lowered the saturated vapor pressure above the saline surface. As the concentration increased toward supersaturated concentration, the maximum temperature parameter and the SQRT ($T_{maximum} - T_{minimum}$) were added to the input variables. This indicated that as the concentration increased, the role of temperature increased.

4.3. Procrustes Analysis (PA)

The PA method was used for detecting the most effective variables in saline water evaporation. The PCA model should be applied for data in an iterative manner. The number of dominant components was detected by eigenvalue calculation. Those eigenvalues, whose values were more than one, were selected as the most dominant PCs. Results showed that the first three components had eigenvalues more than 1.

Table 4

Optimal input variables selected, based on PA for each saline water concentration

Concentration (g/L)	The main effective variables in evaporation		
5	Air temperature	Mean wind speed	Solar radiation
10	Sunshine hours	Mean wind speed	Maximum temperature
20	Air temperature	Sunshine hours	Mean wind speed
50	Air temperature	Sunshine hours	Mean wind speed
100	Air temperature	Sunshine hours	Mean wind speed
300	Air temperature	Sunshine hours	Mean wind speed
500	Air temperature	Sunshine hours	Mean wind speed
Fresh water	Air temperature	Sunshine hours	Mean wind speed

Therefore, the first three components were selected to be the most dominant components here. Also, the number of most important variables in the estimation of saline water evaporation was chosen to be three. Bearing in mind that the number of dominant effective PCs and the number of effective variables were three, the PA analysis was carried out for data and the M^2 index was calculated for the selection of important variables. Table 3 shows the most important variables which were detected by PA to be effective in saline water evaporation with different concentrations.

From Table 4, it can be seen that, for example, in the 20 (g/L) concentration, the three important variables in the estimation of saline water were air temperature, sunshine hours, and mean wind speed same as Wang et al. (2018a, b), respectively. This result is valid for the other five concentrations, namely 50, 100, 300, 500 g/L, and "Fresh Water" too. Similarly, sunshine hours, mean wind speed, and maximum temperature were the effective variables in evaporation from water with concentration of 10 (g/ L). The best combinations of variables for the estimation of saline water evaporation with different concentrations are represented in Table 4.

Depending on the dataset used, the PA method first classifies the data and then selects the variable that contains the most variance in each data set. Therefore, the selected variable from each category is most similar to its members, thus using the selected variable can represent other variables in each category.

As can be seen from the results, the wind speed variable was recognized as the effective input

variable by almost all three methods, the same as Mor et al. (2018). This may be due to the fact that the station studied was near the lake and humidity was high in the area and because of high wind speed in the area, this variable was able to show its effect well. The physical effect of wind speed is to reduce the moisture content in the region and increase the evaporation rate, so the wind speed is also a good representative of maximum wind speed, relative humidity, and saturated vapor pressure. Another variable that can be considered as an effective input variable in evaporation modeling is the air temperature and surface water temperature which were selected by the GT and EnT at most concentrations. However, the PA' method identified air temperature as an effective input variable rather than surface water temperature. The reason is that depending on the physical conditions and the variance of data, air temperature can be representative of surface water temperature. It should be noted that the most important energy source for warming the environment is solar radiation (Dinpashoh et al. 2019; Biazar and Ferdosi 2020), and in fact, the physical effect of solar radiation is the increase in ambient temperature. The temperature itself is also an effective factor in evaporation. Thus, the temperature can somehow reflect the physical effect of solar radiation and represent variable solar radiation. Sunshine hours can also be a good representative of the variable cloudy and rainy weather. In fact, the longer the sunshine hours, the lower the cloudiness, and the lower cloudiness degree will also bring about a decrease in humidity.

In saline water, stratification is carried out at high concentrations and the lower part has the highest
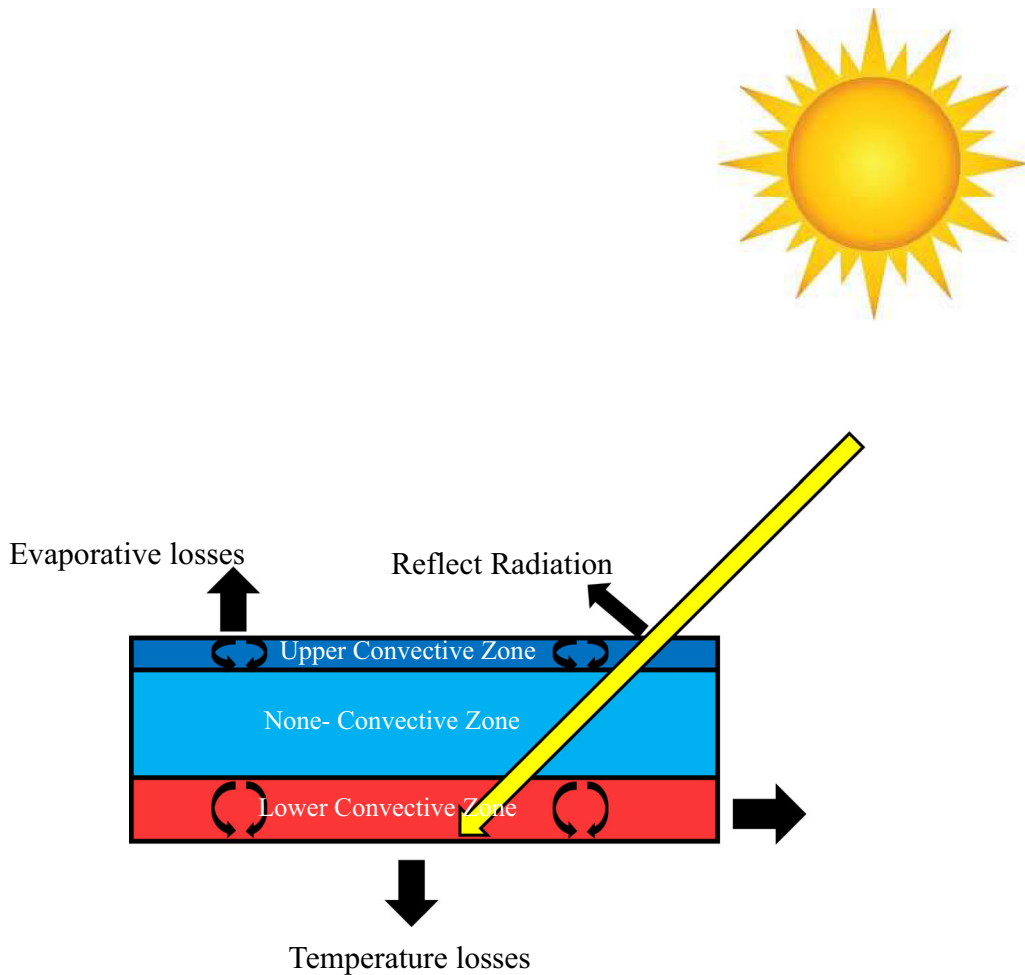
Figure 3
Where Pans were divided into three parts: UCZ (Upper Convective Zone), NCZ (No-Convective Zone) and LCZ (Low Convective Zone)

concentration possible. It should be noted that the sunlight penetrating the pan will pass through the first and second layers and fall into the third layer at high concentrations, which is why the temperature was very high in the pan with high concentrations-the same as Rabl and Nielsen (1975), Hull et al. (1988), and El-Sebaii et al. (2011). In fact, salts or high concentrations of the solution absorb and store the sun's energy. The second layer acts as an insulator for the third layer, which reduces the water surface temperature (the first layer) and increases the temperature at the bottom of the pan (the third layer) at high concentration, the same as Kurt et al. (2000), Suárez et al. (2010), and Ruskowitz et al. (2014).

This will reduce the release of water molecules into the saline water. In saline waters, the first layer has the lowest concentration compared to the other two layers, but the soluble salts in this layer decrease the free energy of water molecules. According to the second law of thermodynamics, the increase in ionic activity due to the presence of ions in the solvent and the chemical potential of the solvent will cause the evaporation of saline waters, especially the saline waters, not following the usual models of evaporation. As a result, it reduces the rate of conversion of water molecules from liquid to gas, and reduces the vapor pressure, as is known, evaporation takes place whenever there is a deficit between a water surface

Table 5

Test results and uncertainty analysis of ANNs models

Models	Concentration (g/L)	R	RMSE (mm/day)	NS	d-factor	p-factor	Model structure[a]
ANN-GT	500	0.88	1.64	0.75	1.88	0.86	(14-10-1)
	300	0.88	1.83	0.68	1.37	0.64	(9-5-1)
	100	0.85	2.62	0.71	1.24	0.64	(12-5-1)
	50	0.95	2.01	0.85	0.90	0.55	(9-2-1)
	20	0.93	2.04	0.86	1.42	0.82	(8-8-1)
	10	0.93	1.97	0.86	1.34	0.82	(9-8-1)
	5	0.98	1.23	0.95	1.23	0.91	(9-17-1)
	Fresh water	0.94	1.81	0.88	0.91	0.68	(4-10-1)
ANN-EnT	500	0.88	1.74	0.72	1.16	0.68	(9-2-1)
	300	0.87	1.68	0.73	1.16	0.64	(5-7-1)
	100	0.87	2.62	0.71	1.24	0.55	(5-9-1)
	50	0.92	2.28	0.81	1.07	0.50	(7-3-1)
	20	0.94	1.92	0.87	0.76	0.64	(4-2-1)
	10	0.94	1.86	0.87	1.23	0.77	(5-8-1)
	5	0.96	1.60	0.92	0.74	0.68	(7-4-1)
	Fresh water	0.92	2.06	0.84	1.04	0.68	(4-13-1)
ANN-PA	500	0.88	1.69	0.71	1.10	0.59	(3-8-1)
	300	0.83	1.85	0.67	0.99	0.55	(3-3-1)
	100	0.86	2.53	0.73	1.12	0.55	(3-7-1)
	50	0.92	2.24	0.82	0.66	0.50	(3-2-1)
	20	0.96	1.71	0.90	1.43	0.77	(3-13-1)
	10	0.92	2.03	0.85	0.85	0.59	(3-8-1)
	5	0.95	1.80	0.90	0.92	0.77	(3-9-1)
	Fresh water	0.93	1.91	0.86	0.70	0.50	(3-4-1)

[a]Model structure for example (14-10-1), It means: 14 input, 10 hidden and one output nodes or neuron

and the overlying atmosphere and sufficient energy is available. This will be neutralized as the concentration decreases, and the energy from sunlight is being distributed more evenly throughout the pan. Also, the concentration in the first layer decreases from super salty to freshwater, respectively, which increases the temperature in the surface layer of water, thereby increasing evaporation from saline to freshwater. As mentioned earlier, the climate was the same for all pans. Therefore, according to the above discussion, the physical effect of concentration of the solutions was mainly on the temperature changes, so that the surface temperature of the saline water was observed to be lower than the freshwater surface temperature for the above-mentioned reasons. Evaporation measurements in this study were performed for each concentration separately so that the effect of the concentration factor was considered on the volatility values themselves.

4.4. Modeling Results

The best input combinations were identified by GT, EnT, and PA, and then ANN was used for evaporation estimation. Table 5 shows the results of performance measures of the intelligent models with different scenarios, namely, ANN-GT, ANN-EnT and ANN-PA. As seen from Table 5, the ANN-GT model showed better performance in the estimation of daily saline water evaporation as compared with the other two models. This result was obtained, based on the performance measures (RMSE, CC, and NS), which is valid for all salt concentrations [except 20 (g/L)]. One of the uncertainties of statistical and intelligent models is the use of iterations in calculations. The coefficients such as the d-factor and p-factor can be calculated based on the results of each iteration, each representing part of the uncertainty of each model. Therefore, to determine the uncertainty band after a thousand times iteration in each model, a 95% probability band was obtained for each pan. Although the model ANN-GT revealed better performance than
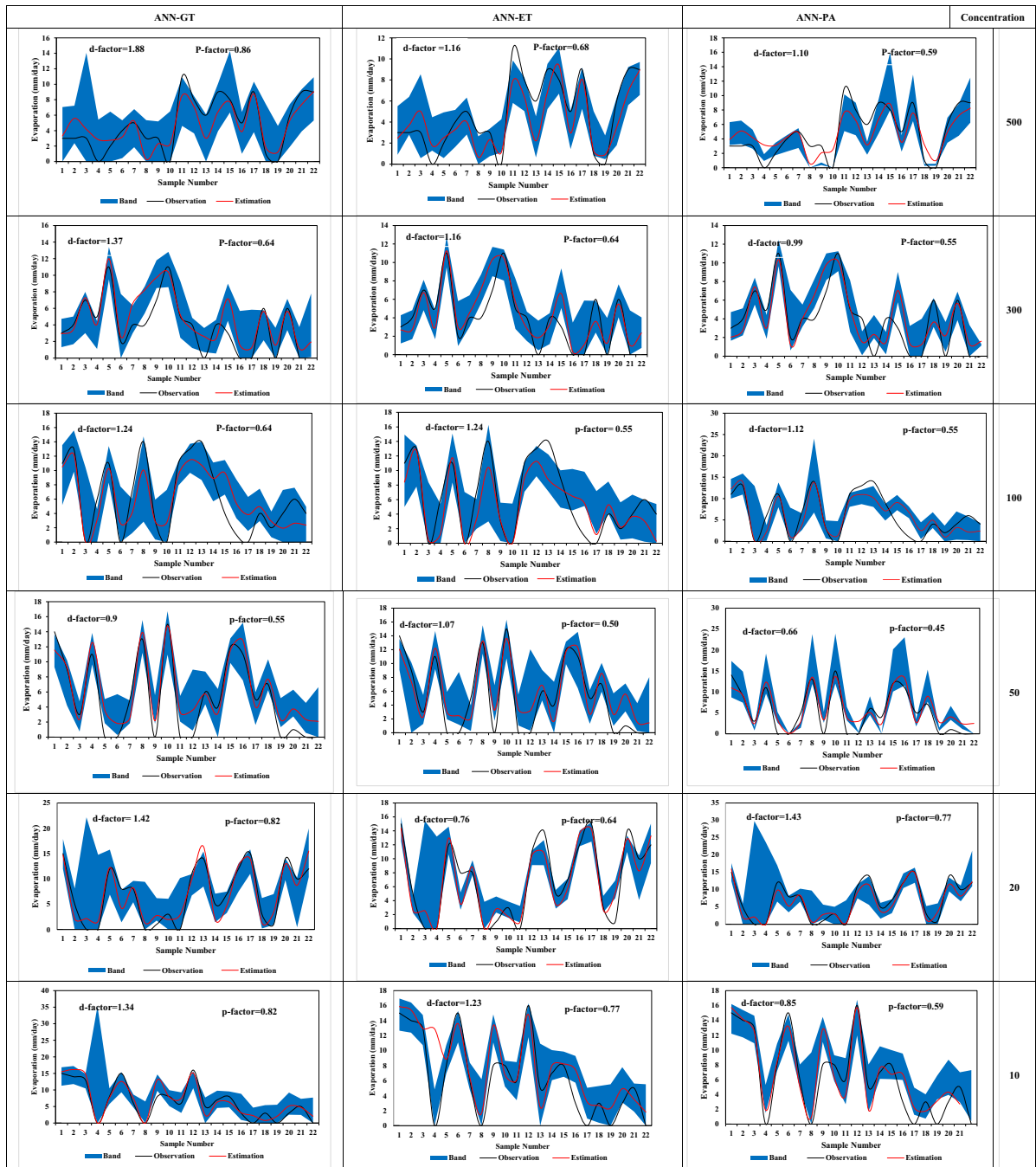
Figure 4
Uncertainty of ANN-GT, ANN-EnT and ANN-PA in estimation of evaporation from saline water
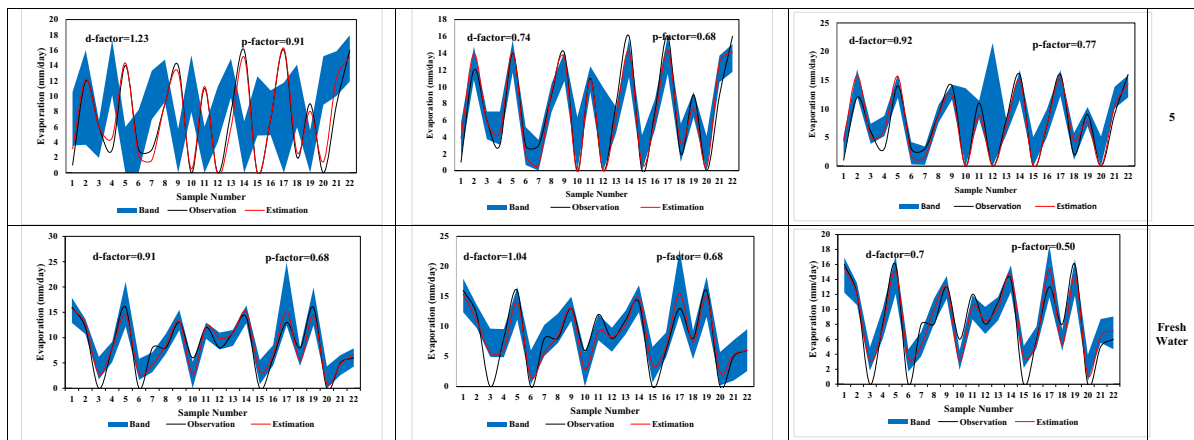
Figure 4
continued

ANN-EnT and ANN-PA, however, ANN-PA showed lower uncertainty than the other two, because the values of the d-factor measure obtained for ANN-GT and ANN-EnT were larger than ANN-PA. However, as can be seen from Table 5 and Fig. 4, the values obtained here for the d-factor and 95% PPU (i.e. p-factor) indicated the lower uncertainty of ANN-PA model in all the tested concentrations [except 20 and 5 (g/L)] (Table 5 and Fig. 4). Results of uncertainty analysis for different salt concentrations are shown in Fig. 4. As can be seen from Fig. 4, ANN-GT and ANN-EnT had a wider 95% PPU band than ANN-PA. This implies that the obtained values of d-factor for the ANN-GT and ANN-EnT models were greater than for ANN-PA. According to Fig. 4, almost all models showed that the higher the bandwidth of 95 PPU, the greater the d-factor, and the lower the bandwidth, the lower the d-factor, the same as Abbaspour et al. (2007), and Noori et al. (2011). The lowest bandwidth also belonged to the ANN-PA model for concentrations of 50 g/L with a P-factor of 0.45 and a d-factor of 0.66. The highest d-factor belonged to the ANN-GT model for a concentration of 500 g/L with a d-factor of 1.88 and a p-factor of 0.86. One reason for the high uncertainty in this model may be due to the number of input variables selected by the gamma test-the same as Isazadeh et al. (2017). As mentioned in the GT section, for the concentration of 500 g/L all variables except the precipitation variable were selected as inputs. Comparative diagrams of calculated and observed values are given in Fig. 4, which shows the proper overlap between the estimated and observed diagrams. Besides, according to the NS index, given the high values of this index, which in most models were in a very good range, it can be concluded that the performance of the models was acceptable. Generally, results with lower uncertainty are more reliable in the estimation of hydrologic variables (Khaledian et al. 2020; Isazadeh et al. 2017; Ghorbani et al. 2016; Noori et al. 2011; Yang et al. 2008; Abbaspour et al. 2007).

5. Conclusion

Model input selection is a complicated process, especially for non-linear dynamic models. This study introduced a new input selection method, PA, for the estimation of daily evaporation from saline water with different concentrations. This work is conducted for daily estimation of saline water evaporation using three intelligent models (with different scenarios) which are ANN-GT, ANN-EnT and ANN-PA. The outputs of these models were compared with the corresponding measured evaporation. The amount of evaporation was measured separately for each pan (each pan had different concentrations). As in the

Kokya and Kokya ([2008](#)) study, the concentration was kept constant throughout the experiment. It should be noted that the amount of concentration affected the amount of evaporation and evaporation increased from the high-concentration pan toward the low-concentration pan (Al-Khlaifat [2008](#)). The effect of concentration is included in the evaporation values themselves. And the purpose of this study was to determine the best input variables. Moreover, the selection of effective Inputs for saline water evaporation using PA, GT and EnT can be regarded to be the novelty of this study. The most important variables identified by the GT were surface water temperature, air temperature, mean relative humidity, mean wind speed, mean station pressure, minimum temperature, precipitation, mean station vapor pressure, and solar radiation. Also, as can be seen the most important variables for evaporation from saline water using the EnT method were water surface temperature, wind speed, and precipitation. The three important variables in the estimation of saline water selected by the PA method were air temperature, sunshine hours, and mean wind speed. As the concentration increased, the mean station vapor pressure and temperature variables had the most influence on saline water evaporation. The performances of models were evaluated by using the RMSE, CC, and NS measures. The limitations of these measures are in their inability to assess uncertainty. Therefore, uncertainty of the models used was evaluated using the p-factor (95% PPU) and d-factor. Although ANN-GT was found to be a powerful model in the estimation of evaporation, however, it failed to select the optimal input combinations with lowest uncertainties (Noori et al. [2011](#)). Having lower uncertainty is reported to be the primary criterion in the selection of most suitable model in hydrologic practical applications. This study employed the mentioned two measures (p-factor and d-factor) in uncertainty analysis. Results of this study are interpreted, based on the evaluation of uncertainty in addition to model performance measures. Although there are numerous questions that need to be addressed yet, this work will hopefully stimulate more investigation in input selection methods in model development for hydrological processes, climate simulation, and evapotranspiration estimation. The authors will suggest to develop different scenarios for

ANN models to remove bias and increase sufficiency in other researches. Generally, according to the results the PA method was able to demonstrate acceptable performance over the other two methods, with the least number of input variables, in addition to reducing the uncertainty of the ANN model. It can be concluded that by reducing the number of variables as well as selecting the input combination that is a good representative of the effective variables in the evaporation estimation, the model complexity can be reduced and a good estimate of saline water evaporation can be made.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., et al. (2007). Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology, 333*(2–4), 413–430. https://doi.org/10.1016/j.jhydrol.2006.09.014.

Aghelpour, P., Mohammadi, B., & Biazar, S. M. (2019). Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. *Theoretical and Applied Climatology, 138*(3–4), 1471–1480. https://doi.org/10.1007/s00704-019-02905-w.

Ahmadi, A., Han, D., Karamouz, M., & Remesan, R. (2009). Input data selection for solar radiation estimation. *Hydrological Processes: An International Journal, 23*(19), 2754–2764. https://doi.org/10.1002/hyp.7372.

Al-Khlaifat, A. (2008). Dead Sea rate of evaporation. *American Journal of Applied Sciences, 5*(8), 934–942.

Allen, R. G., Pereira, L. S., Howell, T. A., & Jensen, M. E. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management, 98*(6), 899–920. https://doi.org/10.1016/j.agwat.2010.12.015.

Anderson, D. B. (1936). Relative humidity or vapor pressure deficit. *Ecology, 17*(2), 277–282. https://doi.org/10.2307/1931468.

Ashrafzadeh, A., Ghorbani, M. A., Biazar, S. M., & Yaseen, Z. M. (2019). Evaporation process modelling over northern Iran:

Application of an integrative data-intelligence model with the krill herd optimization algorithm. *Hydrological Sciences Journal, 64*(15), 1843–1856. https://doi.org/10.1080/02626667.2019.1676428.

Ashrafzadeh, A., Kişi, O., Aghelpour, P., Biazar, S. M., & Masouleh, M. A. (2020). Comparative study of time series models, support vector machines, and GMDH in forecasting long-term evapotranspiration rates in northern Iran. *Journal of Irrigation and Drainage Engineering, 146*(6), 04020010.

Ashrafzadeh, A., Malik, A., Jothiprakash, V., Ghorbani, M. A., & Biazar, S. M. (2018). Estimation of daily pan evaporation using neural networks and meta-heuristic approaches. *ISH Journal of Hydraulic Engineering*. https://doi.org/10.1080/09715010.2018.1498754.

Asmar, B. N., & Ergenzinger, P. (1999). Estimation of evaporation from the Dead Sea. *Hydrological Processes, 13*(17), 2743–2750. https://doi.org/10.1002/(SICI)1099-1085(19991215)13:17%3C2743:AID-HYP845%3E3.0.CO;2-U.

Biazar, S. M., Dinpashoh, Y., & Singh, V. P. (2019). Sensitivity analysis of the reference crop evapotranspiration in a humid region. *Environmental Science and Pollution Research, 26*(31), 32517–32544. https://doi.org/10.1007/s11356-019-06419-w.

Biazar, S. M., & Ferdosi, F. B. (2020). An investigation on spatial and temporal trends in frost indices in Northern Iran. *Theoretical and Applied Climatology*. https://doi.org/10.1007/s00704-020-03248-7.

Biazar, S. M., Rahmani, V., Isazadeh, M., Kisi, O., & Dinpashoh, Y. (2020). New input selection procedure for machine learning methods in estimating daily global solar radiation. *Arabian Journal of Geosciences, 13*, 431.

Brutsaert, W., & Stricker, H. (1979). An advection-aridity approach to estimate actual regional evapotranspiration. *Water Resources Research, 15*(2), 443–450. https://doi.org/10.1029/WR015i002p00443.

Charalambous, C. (1992). Conjugate gradient algorithm for efficient training of artificial neural networks. *IEE Proceedings G (Circuits, Devices and Systems), 139*(3), 301–310. https://doi.org/10.1049/ip-g-2.1992.0050.

Chattopadhyay, S., Chattopadhyay, G., & Midya, S. K. (2018). Shannon entropy maximization supplemented by neurocomputing to study the consequences of a severe weather phenomenon on some surface parameters. *Natural Hazards, 93*(1), 237–247.

Chattopadhyay, S., Jain, R., & Chattopadhyay, G. (2009). Estimating potential evapotranspiration from limited weather data over Gangetic West Bengal, India: A neurocomputing approach. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling, 16*(3), 403–411.

Coulibaly, P., Anctil, F., & Bobée, B. (2000). Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology, 230*(3–4), 244–257. https://doi.org/10.1016/S0022-1694(00)00214-6.

De Bruin, H. A. R., & Keijman, J. Q. (1979). The Priestley–Taylor evaporation model applied to a large, shallow lake in the Netherlands. *Journal of Applied Meteorology, 18*(7), 898–903. https://doi.org/10.1175/1520-0450(1979)018%3C0898:TPTEMA%3E2.0.CO;2.

Deo, R. C., Ghorbani, M. A., Samadianfard, S., Maraseni, T., Bilgili, M., & Biazar, M. (2018). Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for wind-speed prediction of target site using a limited set of neighboring reference station data. *Renewable Energy, 116*, 309–323. https://doi.org/10.1016/j.renene.2017.09.078.

Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S., & Mirnia, M. (2004). Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology, 297*(1–4), 109–123. https://doi.org/10.1016/j.jhydrol.2004.04.009.

Dinpashoh, Y., Singh, V. P., Biazar, S. M., & Kavehkar, S. (2019). Impact of climate change on streamflow timing (case study: Guilan Province). *Theoretical and Applied Climatology, 138*(1–2), 65–76.

Durrant, P. J. (2001). winGamma: A non-linear data analysis and modelling tool with applications to flood prediction. Unpublished Ph.D. thesis, Department of Computer Science, Cardiff University, Wales, UK.

El-Dessouky, H. T., Ettouney, H. M., Alatiqi, I. M., & Al-Shamari, M. A. (2002). Evaporation rates from fresh and saline water in moving air. *Industrial & Engineering Chemistry Research, 41*(3), 642–650. https://doi.org/10.1021/ie010327o.

El-Sebaii, A. A., Ramadan, M. R. I., Aboul-Enein, S., & Khallaf, A. M. (2011). History of the solar ponds: A review study. *Renewable and Sustainable Energy Reviews, 15*(6), 3319–3325. https://doi.org/10.1016/j.rser.2011.04.008.

Evans, D., & Jones, A. J. (2002). A proof of the gamma test. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 458*(2027), 2759–2799. https://doi.org/10.1098/rspa.2002.1010.

Fombellida, M., & Destiné, J. (1992). The extended quickprop. In Artificial Neural Networks (pp. 973–977). North-Holland. https://doi.org/10.1016/B978-0-444-89488-5.50032-4

Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., & Terzi, O. (2016). A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences, 75*(6), 476. https://doi.org/10.1007/s12665-015-5096-x.

Gianniou, S. K., & Antonopoulos, V. Z. (2007). Evaporation and energy budget in Lake Vegoritis, Greece. *Journal of Hydrology, 345*(3–4), 212–223. https://doi.org/10.1016/j.jhydrol.2007.08.007.

Goyal, M. K., Bharti, B., Quilty, J., Adamowski, J., & Pandey, A. (2014). Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. *Expert Systems with Applications, 41*(11), 5267–5276. https://doi.org/10.1016/j.eswa.2014.02.047.

Guo, Y., Zhang, Y., Ma, N., Xu, J., & Zhang, T. (2019). Long-term changes in evaporation over Siling Co Lake on the Tibetan Plateau and its impact on recent rapid lake expansion. *Atmospheric Research, 216*, 141–150. https://doi.org/10.1016/j.atmosres.2018.10.006.

Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks, 5*(6), 989–993. https://doi.org/10.1109/72.329697.

Hamdani, I., Assouline, S., Tanny, J., Lensky, I. M., Gertman, I., Mor, Z., et al. (2018). Seasonal and diurnal evaporation from a deep hypersaline lake: The Dead Sea as a case study. *Journal of Hydrology, 562*, 155–167. https://doi.org/10.1016/j.jhydrol.2018.04.057.

Harbeck G. E. (1955). The effect of salinity on evaporation. United States, Geological Survey, Professional Papers 272-A.

Harmancioglu, N. B., & Alpaslan, N. (1992). Water quality monitoring network design: a problem of multi-objective decision making 1. *JAWRA Journal of the American Water Resources Association, 28*(1), 179–192. https://doi.org/10.1111/j.1752-1688.1992.tb03163.x.

Holmes, T. R. (2019). Remote sensing techniques for estimating evaporation. In Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment (pp. 129–143). Elsevier.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(6), 417–441.

Hull, J. R., Nielsen, J., & Golding, P. (1988). Salinity gradient solar ponds. *Advances in Solar Enery.* https://doi.org/10.1007/978-1-4613-9945-2_6.

Isazadeh, M., Biazar, S. M., & Ashrafzadeh, A. (2017). Support vector machines and feed-forward neural networks for spatial modeling of groundwater qualitative parameters. *Environmental Earth Sciences, 76*(17), 610. https://doi.org/10.1007/s12665-017-6938-5.

Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks, 1*(4), 295–307. https://doi.org/10.1016/0893-6080(88)90003-2.

Johnson, G. L., & Hanson, C. L. (1995). Topographic and atmospheric influences on precipitation variability over a mountainous watershed. *Journal of Applied Meteorology, 34*(1), 68–87. https://doi.org/10.1175/1520-0450-34.1.68.

Jolliffe, I. T. (1986). Principal components in regression analysis. In Principal component analysis (pp. 129–155). Springer, New York, NY.

Khaledian, M. R., Isazadeh, M., Biazar, S. M., & Pham, Q. B. (2020). Simulating Caspian Sea surface water level by artificial neural network and support vector machine models. *Acta Geophysica.* https://doi.org/10.1007/s11600-020-00419-y.

Kisi, O., Shiri, J., Karimi, S., Shamshirband, S., Motamedi, S., Petković, D., et al. (2015). A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. *Applied Mathematics and Computation, 270,* 731–743. https://doi.org/10.1016/j.amc.2015.08.085.

Kokya, B. A., & Kokya, T. A. (2008). Proposing a formula for evaporation measurement from salt water resources. *Hydrological Processes: An International Journal, 22*(12), 2005–2012. https://doi.org/10.1002/hyp.6785.

Krzanowski, W. J. (1987). Selection of variables to preserve multivariate data structure, using principal components. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 36*(1), 22–33. https://doi.org/10.2307/2347842.

Kurt, H., Halici, F., & Binark, A. K. (2000). Solar pond conception—experimental and theoretical studies. *Energy Conversion and Management, 41*(9), 939–951. https://doi.org/10.1016/S0196-8904(99)00147-8.

Lagos-Avid, M. P., & Bonilla, C. A. (2017). Predicting the particle size distribution of eroded sediment using artificial neural networks. *Science of The Total Environment, 581,* 833–839.

Lee, C. H. (1927). Discussion of evaporation on reclamation projects. *American Society of Civil Engineers Transactions, 90,* 340–343.

Lide, D. R. (Ed.). (2004). *CRC handbook of chemistry and physics (vol. 85).* Boca Raton: CRC Press.

Meyer, A. F. (1915). Computing run-off from rainfall and other physical data. *Transactions of the American Society of Civil Engineers, 78*(2), 1056–1155.

Moghaddamnia, A., Gousheh, M. G., Piri, J., Amin, S., & Han, D. (2009). Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *Advances in Water Resources, 32*(1), 88–97. https://doi.org/10.1016/j.advwatres.2008.10.005.

Mor, Z., Assouline, S., Tanny, J., Lensky, I. M., & Lensky, N. G. (2018). Effect of water surface salinity on evaporation: The case of a diluted buoyant plume over the Dead Sea. *Water Resources Research, 54*(3), 1460–1475. https://doi.org/10.1002/2017WR021995.

Naganna, S. R., Deka, P. C., Ghorbani, M. A., Biazar, S. M., Al-Ansari, N., & Yaseen, Z. M. (2019). Dew point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms. *Water, 11*(4), 742. https://doi.org/10.3390/w11040742.

Nam, W., Shin, H., Jung, Y., Joo, K., & Heo, J. H. (2015). Delineation of the climatic rainfall regions of South Korea based on a multivariate analysis and regional rainfall frequency analyses. *International Journal of Climatology, 35*(5), 777–793. https://doi.org/10.1002/joc.4182.

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. https://doi.org/10.1016/0022-1694(70)90255-6.

Noori, R., Karbassi, A. R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M. H., Farokhnia, A., et al. (2011). Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology, 401*(3–4), 177–189. https://doi.org/10.1016/j.jhydrol.2011.02.021.

Nozari, H., & Azadi, S. (2019). Experimental evaluation of artificial neural network for predicting drainage water and groundwater salinity at various drain depths and spacing. *Neural Computing and Applications, 31*(4), 1227–1236. https://doi.org/10.1007/s00521-017-3155-9.

Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A Mathematical and Physical Sciences, 193*(1032), 120–145. https://doi.org/10.1098/rspa.1948.0037.

Piri, J., Amin, S., Moghaddamnia, A., Keshavarz, A., Han, D., & Remesan, R. (2009). Daily pan evaporation modeling in a hot and dry climate. *Journal of Hydrologic Engineering, 14*(8), 803–811. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000056.

Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review, 100*(2), 81–92. https://doi.org/10.1175/1520-0493(1972)100%3C0081:OTAOSH%3E2.3.CO;2.

Rabl, A., & Nielsen, C. E. (1975). Solar ponds for space heating. *Solar Energy, 17*(1), 1–12. https://doi.org/10.1016/0038-092X(75)90011-0.

Remesan, R., Shamim, M. A., & Han, D. (2008). Model data selection using gamma test for daily solar radiation estimation. *Hydrological Processes, 22*(21), 4301–4309. https://doi.org/10.1002/hyp.7044.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Learning representations by back-propagating errors. *Cognitive Modeling, 5*(3), 1.

Ruskowitz, J. A., Suárez, F., Tyler, S. W., & Childress, A. E. (2014). Evaporation suppression and solar energy collection in a

salt-gradient solar pond. *Solar Energy, 99*, 36–46. https://doi.org/10.1016/j.solener.2013.10.035.

Saha, S., & Chattopadhyay, S. (2020). Exploring of the summer monsoon rainfall around the Himalayas in time domain through maximization of Shannon entropy. *Theoretical and Applied Climatology*. https://doi.org/10.1007/s00704-020-03186-4.

Sandler, S. I. (1999). *Chemical and engineering thermodynamics (Wiley series in Chemical Engineering)*. New York: Wiley.

Seifi, A., & Riahi, H. (2018). Estimating daily reference evapotranspiration using hybrid gamma test-least square support vector machine, gamma test-ANN, and gamma test-ANFIS models in an arid area of Iran. *Journal of Water and Climate Change*. https://doi.org/10.2166/wcc.2018.003.

Shannon, C. E., & Weaver, W. (1949). *Urban*. Champaign: University of Illinois Press.

Shiri, J., Shamshirband, S., Kisi, O., Karimi, S., Bateni, S. M., Nezhad, S. H. H., et al. (2016). Prediction of water-level in the Urmia Lake using the extreme learning machine approach. *Water Resources Management, 30*(14), 5217–5229. https://doi.org/10.1007/s11269-016-1480-x.

Singh, V. P. (2011). Hydrologic synthesis using entropy theory. *Journal of Hydrologic Engineering, 16*(5), 421–433. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000332.

Suárez, F., Tyler, S. W., & Childress, A. E. (2010). A fully coupled, transient double-diffusive convective model for salt-gradient solar ponds. *International Journal of Heat and Mass Transfer, 53*(9–10), 1718–1730. https://doi.org/10.1016/j.ijheatmasstransfer.2010.01.017.

Vaheddoost, B., & Kocak, K. (2019). Temporal dynamics of monthly evaporation in Lake Urmia. *Theoretical and Applied Climatology, 137*(3–4), 2451–2462. https://doi.org/10.1007/s00704-018-2747-3.

Wang, B., Ma, Y., Ma, W., Su, B., & Dong, X. (2018a). Evaluation of ten methods for estimating evaporation in a small high-elevation lake on the Tibetan Plateau. *Theoretical and Applied Climatology*. https://doi.org/10.1007/s00704-018-2539-9.

Wang, W., Lee, X., Xiao, W., Liu, S., Schultz, N., Wang, Y., et al. (2018b). Global lake evaporation accelerated by changes in surface energy allocation in a warmer climate. *Nature Geoscience, 11*(6), 410. https://doi.org/10.1038/s41561-018-0114-8.

Wurtsbaugh, W. A., Miller, C., Null, S. E., DeRose, R. J., Wilcock, P., Hahnenberger, M., et al. (2017). Decline of the world's saline lakes. *Nature Geoscience, 10*(11), 816. https://doi.org/10.1038/NGEO3052.

Yang, J., Reichert, P., Abbaspour, K. C., Xia, J., & Yang, H. (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *Journal of Hydrology, 358*(1–2), 1–23. https://doi.org/10.1016/j.jhydrol.2008.05.012.

Young, A. A. (1947). Some recent evaporation investigations. *Eos, Transactions American Geophysical Union, 28*(2), 279–284. https://doi.org/10.1029/TR028i002p00279.

Zhao, G., & Gao, H. (2019). Estimating reservoir evaporation losses for the United States: Fusing remote sensing and modeling approaches. *Remote Sensing of Environment, 226*, 109–124. https://doi.org/10.1016/j.rse.2019.03.015.