RESEARCH ARTICLE

SVR-RSM: a hybrid heuristic method for modeling monthly pan evaporation

Behrooz Keshtegar [1,2] · Salim Heddam [3] · Abderrazek Sebbar [4] · Shun-Peng Zhu [5] · Nguyen-Thoi Trung [1,2]

Abstract

In the present study, a hybrid intelligent model called SVR_RSM, which was extracted using response surface method (RSM) combined by the support vector regression (SVR) approaches was applied for predicting monthly pan evaporation (E_{pan}). This method is established based on two basic calibrating process using RSM and SVR. In the first process, an input data group with two different input variables are used to calibrate the RSM; hence, the calibrating data by RSM in the first process are applied as input database for calibrating the SVR in the second process. Results obtained using the proposed SVR_RSM was compared with those obtained using the RSM, SVR, and the well-known multilayer perceptron neural network (MLPNN) models. Climatic variables including maximum and minimum temperatures (T_{max}, T_{min}), wind speed (U_2), and relative humidity (H%), and the periodicity represented by the month number (α) were selected for predicting the monthly E_{pan} measured with the standard class A evaporation pan. Data was collected at six climatic stations located at the northern East of Algeria. The performances of the proposed models were compared using the RMSE, MAE, modified index of agreement (d), coefficient of correlation (R), and modified Nash and Sutcliffe efficiency (NSE). Using various input combination, the results show that the hybrid SVR_RSM model performed better than all the proposed models. Overall, better accuracy was observed when the model contained the periodicity (α), and it was demonstrated that the best accuracy was obtained using only T_{max} and T_{min}, coupled with the periodicity.

Responsible Editor: Marcus Schulz

✉ Behrooz Keshtegar
beh.keshtegar@tdtu.edu.vn

✉ Salim Heddam
heddamsalim@yahoo.fr

Abderrazek Sebbar
rsebbar@yahoo.fr

Shun-Peng Zhu
zspeng2007@uestc.edu.cn

Nguyen-Thoi Trung
nguyenthoitrung@tdtu.edu.vn

[1] Division of Computational Mathematics and Engineering, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[2] Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[3] Faculty of Science, Agronomy Department, Hydraulics Division, Laboratory of Research in Biodiversity Interaction Ecosystem and Biotechnology, University 20 Août 1955, Route El Hadaik, 26 Skikda, BP, Algeria

[4] Soil and Hydraulics Research Laboratory, Faculty of Engineering Sciences, Hydraulics Department, University Badji-Mokhtar, Annaba, Algeria

[5] School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Introduction

Water loss through evaporation constitutes the main part in the water balance of a catchment, reservoir, and lake (McMahon et al. 2013), and must be estimated as accurately as possible. Exact quantification of the evaporation shall constitute a principal requirement for many applications, among them water balance studies, hydrological modeling, and irrigation scheduling (Cahoon et al. 1991). Over the years, several studies conducted worldwide have shown that direct measurement of evaporation using pan evaporimeter (E_{pan}) is a commonly used approach for estimating the evaporation rate. Undoubtedly, direct measurement is a best and useful method. However, despite its importance, several alternative indirect empirical models were proposed and have been widely studied and they have a proven effectiveness (e.g., the Penman, Stephens-Stewart, and Hargreaves and Samani models). The use of indirect empirical models for evaporation estimation is often associated by the need of a large number of climatic variables. Unfortunately, except the temperature-based methods, that needs generally fewer climatic variables as input, the large amount of input variables were required for applying the empirical and semi empirical models reflects the difficulties which were needed to be overcome by using alternative methods, offered by artificial intelligence (AI), which relate the E_{pan} to several climatic variables.

A number of researchers have been demonstrated the successful application of several AI models for modeling and predicting E_{pan} using several climatic variables as input. Sebbar et al. (2019) proposed a new model for predicting daily E_{pan} using four daily climatic variables from two stations in Algeria, which is based on the extreme learning machine (ELM) approach. They have applied and compared (i) the optimally pruned extreme learning machine (OPELM); and (ii) the online sequential extreme learning machine (OSELM) models. Minimum and maximum air temperatures (T_{min} and T_{max}), wind speed (U_2), and relative humidity (H%) have been employed to predict E_{pan} measured using class A evaporation pan, and the obtained result agrees well with the measured values with a coefficient of correlation (R) between 0.800 and 0.872 using the OSELM, and between 0.808 and 0.853 using the OPELM. Feng et al. (2018) applied three AI and two empirical models in predicting monthly E_{pan} in China. The proposed AI models were (i) ELM, (ii) multilayer perceptron neural network (MLPNN) optimized by particle swarm optimization (MLPNN-PS), (iii) and MLPNN optimized by genetic algorithm (MLPNN-GAANN), and the obtained results were compare to those provided by the Stephens and Stewart (SS) and the Penman empirical models. From the obtained results, they reported that the best accuracy was achieved using the ELM model with average relative root mean square error (RRMSE) and mean absolute error (MAE) of 12.5–15.2% and 11.7–19.9 mm, respectively. In

the same year, Lu et al. (2018) presented a new kind of models based on the tree-based machine learning (TBM) models: (i) M5 model tree (M5Tree), (ii) random forests (RFs), and (iii) gradient boosting decision tree (GBDT). In addition, the authors compared the obtained results using the TBM models with those provided by four empirical equations. All the models were applied and compared using data at daily time step in China. The more accurate prediction by the TBM models was obtained using the GBDT model compared to the M5Tree and RFs with root mean square error (RMSE), mean bias error (MBE), and Nash-Sutcliffe efficiency (NSE) of 0.86 mm, 0.07 mm, and 0.68, respectively. Regarding the empirical model, the authors reported that the Priestley-Taylor model was the most accurate model and the Trabert model performed worst. Another new kind of AI model was introduced by Eray et al. (2018). They applied at the first time an evolving connectionist systems (ECoS) called dynamic evolving neural-fuzzy inference systems named (DENFIS) for modeling monthly E_{pan} using data from two stations in Turkey. Compared to another AI model, the multi-gene genetic programming (MGGP), DENFIS performed best in one station while the MGGP performed best into the second station. More recently, hybrid models evolutionary algorithms have been successfully applied for modeling E_{pan}. For example, Ghorbani et al. (2018) combined the firefly algorithm (FFA) with the standard MLPNN and a hybrid model called MLP-FFA was employed for modeling daily E_{pan} in the arid regions of Iran. From the obtained results, the authors reported that MLP-FFA was more accurate than the MLPNN and support vector machine models (SVM) for both tested stations, with Willmott's index of agreement (WI), NSE, and RMSE ranged from 0.926 to 0.976, 0.791 to 0.922, and 1.007 mm to 1.406 mm, respectively. Shiri (2019) compared adaptive neuro-fuzzy inference systems (ANFIS) and multilayer perceptron neural network (MLPNN) for modeling daily E_{pan} using data from four stations in the USA. Qasem et al. (2019) compared four AI models for predicting monthly E_{pan} using data from two stations in Iran and Turkey. The proposed models were the MLPNN, SVM, and combination of them with wavelet transforms (WSVR and WANN). According to the obtained results, the authors demonstrated that E_{pan} was highly related to the temperature and solar radiation and on the other hand, the wavelet decomposition does not contribute to the overall improvement of the models performances.

Different AI models for E_{pan} can be found in the literature, including radial basis function neural network (Allawi and El-Shafie 2016), support vector machine (Rezaie-Balf et al. 2018), modified response surface method (Keshtegar and Kisi 2016, 2017), least square support vector regression (Wang et al. 2017), co-active neuro-fuzzy inference system (Malik et al. 2017), minimax probability machine regression (Deo et al. 2016), generalized regression neural networks (Kim and Kim 2008), linear genetic programming (Guven

and Kisi 2011), gene-expression programming (Shiri and Kisi 2011), ANFIS (Shiri et al. 2011), evolutionary neural networks (Kisi 2013), fuzzy genetic approach (Kisi and Tombul 2013), and multivariate adaptive regression splines (Kisi 2015). Hence, it is clear that significant efforts to predict daily and monthly pan evaporation using AI models have been carried out in the last two decades. The review of the literature reveals that models based on response surface method were extremely rare. To the best of our knowledge, only the investigations conducted by Keshtegar and Kisi (2016, 2017), no other studies, have reported an application of the response surface method (RSM) for modeling evaporation. The purpose of this paper is to develop a hybrid model that combines the standard support vector regression (SVR) and the RSM in order to develop a hybrid model called SVR_RSM. The present study is based on measured data at monthly time step in Algeria. The suitability of the proposed hybrid method is analyzed in terms of prediction accuracy for estimating monthly E_{pan}, and the obtained results were compared to those obtained using MLPNN, SVR, and RSM models.

Materials and methods

Case study

In this study, six sites were selected to develop the proposed models: Ain Dalia, Beni Haroun, Bouhamden, Chaffia, El Agram, and Zit Emba. The sites are located in northern East of Algeria. The spatial location of these stations is shown in Fig. 1. Table 1 shows the coordinates of these sites with respect to latitude, longitude, time period of record, and the total pattern used for developing the models. Evaporation pan (E_{pan}), which is the predicted variable, was measured using the class A evaporation pan. In addition, data on measured, maximum and minimum temperatures (T_{max}, T_{min}), wind speed (U_2), and relative humidity (H %) were selected as input variables. Table 2 reports the various input combinations for prediction of monthly E_{pan}. The data for all stations were divided into two subset; training and validation subsets, which correspond to 70% and 30%, respectively. Six combinations of input variable were used in the MLPNN, SVR, RSM, and SVR_RSM models to predict the monthly E_{pan} (Table 2). It is clear from Table 2 that the periodicity (α) that corresponds to the month number from 1 to 12 is included in the all combination. Combinations 4 to 6 include only two climatic variables in addition to α, while combination 2 is the unique combination for which the periodicity is not included. A model without wind speed was also considered as one of the combinations (combination 3). The descriptive statistics of the climatic variables and the E_{pan} are presented in Table 3, where X_{mean}, X_{max}, X_{min}, S_x, Cv, and R denote the mean, maximum,

minimum, standard deviation, coefficient of variation, and coefficient of correlation with E_{pan}, respectively.

Modeling methods

Artificial neural network

Artificial neural network (ANN) is a powerful modeling tool which is extended based on biological nervous system (Pathirage et al. 2018). Predicting data, classifying database, or performing pattern can be applied based on ANNs, and also in complex modeling events such as hydrology, reliability analysis, structural design, chemical process, environmental problems, and medicinal patterns. The multilayer perceptron neural network (MLPNN) is a popular well-known ANNs algorithm using train-based optimization methods. The MLPNN is usually structured by an input, one or more hidden and one output layers, that accuracy of prediction using this model strongly depended on the number of neurones in each hidden layer. In the first stage, the input layer is connected using relative weights to the neurones in first hidden layer (φ^1) by the following relation:

$$\varphi_j^1 = f_j^1\left(b_j^1 + \sum_{i=1}^{n} w_{ij}^1 x_i\right) \tag{1}$$

where φ_j^1 represents jth neuron of hidden layer 1, b_j^1 represents the bias of the jth node in hidden layer 1, w_{ij}^1 denotes weights to connect the jth node of hidden layer 1, and ith input node of input layer with n number of input nodes x_i, $i = 1$, 2,…, n. f is the activation function, for which the sigmoid function f is commonly utilized as follows:

$$f_j^1 = \left[1 + \exp\left[-\left(b_j^1 + \sum_{i=1}^{n} w_{ij}^1 x_i\right)\right]\right]^{-1} \tag{2}$$

The second hidden layer is built using the first hidden layer as input data as well as the first hidden layer by the following relation:

$$\varphi_j^2 = f_j^2\left(b_j^2 + \sum_{i=1}^{M} w_{ij}^2 \varphi_i^1\right) \tag{3}$$

where φ_j^2 represents jth neuron of hidden layer 2, b_j^2 and w_{ij}^2 respectively denote the bias of the jth node in hidden layer 2 and weights which connect jth node of hidden layer 2 to ith node of hidden layer 1 with M elements φ_i $i = 1,2,…, M$ and f_j^2 is the activation function in hidden layer 2 as follows:

$$f_j^2 = \left[1 + \exp\left[-\left(b_j^2 + \sum_{i=1}^{n} w_{ij}^2 \varphi_i^1\right)\right]\right]^{-1} \tag{4}$$

Fig. 1 Geographical locations of the six weather stations across Algeria

The predicted function for pan evaporation is applied to connect the input database on hidden layer 2 and output neuron as (E_{pan}) which is expressed as below:

$$\hat{E} = b + \sum_{i=1}^{h} w_i \varphi_i^2 \qquad (5)$$

where h represents number of hidden nodes in layer 2, b is the bias of the output layer, and w_j represent the connection weights between the output node and to jth neuron in hidden layer 2. φ_i^2 is ith hidden node in layer 2 which is computed based on Eq. (2) with sigmoid active function.

Generally, back-propagation learning tools-based optimization approaches can provide the suitable connections between input and output data (Kurt and Kayfeci 2009). By using a random weight as initial weights, the optimum weights and biases are generally searched using a mathematical optimization method as gradient, conjugate gradient, or Newton approaches (Dao and Vemuri 2002). In the present study, the learning approach to obtain the nonlinear relation between the input and output variables was applied using Levenberg-Marquardt algorithm (Fun and Hagan 1996). The randomly weights are adjusted after each iteration process. In this study, the number of neuron in the two hidden layers is explored

Table 1 Data set presentation for all stations

Station	Latitude	Longitude	Period of record	Total pattern	Incomplete pattern	Final pattern
Ain Dalia	36° 15′ 49″	7° 51′ 43″	01/01/2003–31/12/2016	168	55	111
Beni Haroun	36° 34′ 00″	6° 16′ 31″	01/01/2003–31/12/2014	144	20	124
Bouhamden	36° 27′ 42″	7° 14′ 16″	01/01/2004–31/12/2016	156	39	117
Chaffia	36° 36′ 33″	8° 02′ 32″	01/01/2000–31/12/2016	204	01	203
El Agram	36° 43′ 36″	5° 49′ 37″	01/01/2003–31/12/2015	156	09	147
Zit Emba	36° 41′ 41″	7° 18′ 07″	01/01/2005–31/12/2016	144	00	144

The six input combinations of different models

Models				Inputs combinations
MLPNN	SVR	RSM	SVR_RSM	
MLPNN1	SVR1	RSM1	SVR_RSM1	T_{max}, T_{min}, H%, U_2, α
MLPNN2	SVR2	RSM2	SVR_RSM2	T_{max}, T_{min}, H%, U_2
MLPNN3	SVR3	RSM3	SVR_RSM3	T_{max}, T_{min}, H%, α
MLPNN4	SVR4	RSM4	SVR_RSM4	T_{max}, T_{min}, α
MLPNN5	SVR5	RSM5	SVR_RSM5	T_{max}, H%, α
MLPNN6	SVR6	RSM6	SVR_RSM6	T_{min}, U_2, α

α: the periodicity (month number)

using mean squared errors (MSE) by trial and error to give the best connection between input layer and output layer with one neuron of pan evaporation. In Fig. 2, a MLPNN with two hidden layers and output layer with one neuron is presented which is applied in the current study.

Response surface method

The response surface method (RSM) is a modeling-based mathematical simple tool with low computational burden to predict the engineering problem. The mathematical relation of this model using second-order polynomial functions is presented as follows (Keshtegar and Heddam 2018; Heddam et al. 2019; Keshtegar et al. 2019b):

$$\hat{E} = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} x_i x_j \qquad (6)$$

where $\hat{E}$ is the predicted pan evaporation using n - input data, and w_i and w_{ij} are connected weights between the polynomial functions and the observed data with bias w_0. In the RSM, the polynomial nodes are directly computed using input data (x) with linear, second order and cross terms. In the modeling process of the RSM, N nodes, i.e., $N = n (n + 1)/2$ is applied using one hidden layer which is computed using the input layer elements. The schematic view of this model is presented in Fig. 3. Commonly, the weights and bias of RSM is computed using last square estimator (Keshtegar et al. 2018; Keshtegar and Seghier 2018).

Support vector regression

The nonparametric modeling approaches can be applied to predict the performances of complex real engineering problems. The learning theory basis the support vector machines is a powerful intelligence tool for regression (SVR) and

classification problems (Brereton and Lloyd 2010). Consequently, the SVR modeling approach as a nonlinear model to provide the suitable relation between the pan evaporation and climatological data can be used to predict these complex environmental problems by the following model:

$$\hat{E} = b + \sum_{i=1}^{N} w_i K(x, x_i) \qquad (7)$$

where b is bias and $K(x, x_i)$ represents the Kernel function which is transferred the n-input database from X- space into N-dimensional feature space. Generally, Gaussian kernel function is used for transferring the input data as follows (Brereton and Lloyd 2010):

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \qquad (8)$$

where σ is the kernel parameter which provides the smoothness of the Kernel function; it is given as $\sigma = 8$ in this study. In Eq. (7), w_i is the weight to connect the predicted pan evaporation and input random data basis in feature space which is computed using two slack variables ξ_i, ξ_i^* by the following optimization problem (Lu 2014):

$$\text{Minimize} \frac{\|w\|^2}{2} + C \sum_{i=1}^{N} \left(\xi_i + \xi_i^*\right)$$
$$\text{Subjected} \quad \text{to} \begin{cases} y_i - < w.K(x, x_i) > -b \leq \varepsilon + \xi_i \\ < w.K(x, x_i) > +b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \qquad (9)$$

In which, factor $C \geq 0$ is the regularization coefficient which is given as $C = 300$ in this study, and ε is insensitive loss function which is given as $\varepsilon = 0.15$. The ε- insensitive loss function is used to neglect the calibrating process-based SVR when differences between the predicted and observed pan evaporation are less than ε. The SVR model is schematically shown in Fig. 4a that the structure-based prediction using nonlinear Kernel function is presented in Fig. 4b with input data set (x). By applying the Karush-Kuhn-Tucker (KKT) condition, the optimum parameters of Eq. (9) can be computed using the Lagrange optimization model in regression process as below (Thissen et al. 2004):

$$\text{Maximize} - \frac{1}{2} \sum_{i,j=1}^{N} \left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right) K(x_i, x_j) - \varepsilon \sum_{i=1}^{N} \left(\alpha_i - \alpha_i^*\right) + \sum_{i=1}^{N} E_i\left(\alpha_i - \alpha_i^*\right)$$
$$\text{Subjected} \quad \text{to} \begin{cases} \sum_{i=1}^{N} \left(\alpha_i - \alpha_i^*\right) = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases}$$
$$(10)$$

where α_i and α_i^* are Lagrange multipliers. Thus, the predicted function-based SVR is given as follows:

Table 3 Statistical parameters of the used data sets for all stations

Station	Data set	Unit	X_{mean}	X_{max}	X_{min}	S_x	C_v	R
Ain Dalia	T_{max}	°C	21.534	35.590	7.679	8.157	0.379	0.876
	T_{min}	°C	11.238	36.497	0.959	6.667	0.593	0.802
	H	%	63.833	83.742	34.871	12.490	0.196	−0.818
	U_2	m/s	3.360	5.681	0.449	1.022	0.304	−0.103
	E_{pan}	mm	6.107	14.287	1.550	2.952	0.483	1.000
Beni Haroun	T_{max}	°C	21.533	36.883	6.408	8.801	0.409	0.901
	T_{min}	°C	8.918	20.174	-1.886	5.961	0.668	0.814
	H	%	0.673	0.941	0.311	0.172	0.256	−0.854
	U_2	m/s	2.500	4.239	0.896	0.652	0.261	−0.237
	E_{pan}	mm	5.265	13.923	0.950	3.597	0.683	1.000
Bouhamden	T_{max}	°C	25.595	39.516	12.818	7.616	0.298	0.936
	T_{min}	°C	11.414	27.455	2.879	5.674	0.497	0.897
	H	%	62.725	80.733	36.032	10.762	0.172	−0.836
	U_2	m/s	1.993	3.222	0.811	0.523	0.262	0.106
	E_{pan}	mm	3.769	10.190	0.780	2.648	0.702	1.000
Chaffia	T_{max}	°C	23.693	35.426	13.059	5.648	0.238	0.916
	T_{min}	°C	12.625	21.374	4.500	4.891	0.387	0.895
	H	%	75.768	86.129	64.145	4.890	0.065	−0.621
	U_2	m/s	3.701	5.693	2.537	0.462	0.125	0.016
	E_{pan}	mm	2.793	7.087	0.820	1.681	0.602	1.000
El Agram	T_{max}	°C	23.488	33.823	13.103	5.628	0.240	0.910
	T_{min}	°C	13.549	22.300	4.103	5.165	0.381	0.872
	H	%	72.984	82.167	58.150	4.855	0.067	−0.576
	U_2	m/s	2.539	4.565	0.658	0.741	0.292	−0.135
	E_{pan}	mm	4.104	9.560	0.970	2.280	0.556	1.000
Zit Emba	T_{max}	°C	15.480	24.781	5.906	5.074	0.328	0.881
	T_{min}	°C	22.710	30.813	8.877	5.123	0.226	0.875
	H	%	70.381	85.742	56.587	4.608	0.065	−0.047
	U_2	m/s	3.413	4.710	0.000	0.604	0.177	−0.489
	E_{pan}	mm	4.973	11.820	0.587	3.038	0.611	1.000

X_{mean} mean, X_{max} maximum, X_{min} minimum, S_x standard deviation, C_v coefficient of variation, R coefficient of correlation with E_{pan}

$$\hat{E} = b + \sum_{i=1}^{N} \left(\alpha_i - \alpha_i^*\right) K(x, x_i) \qquad (11)$$
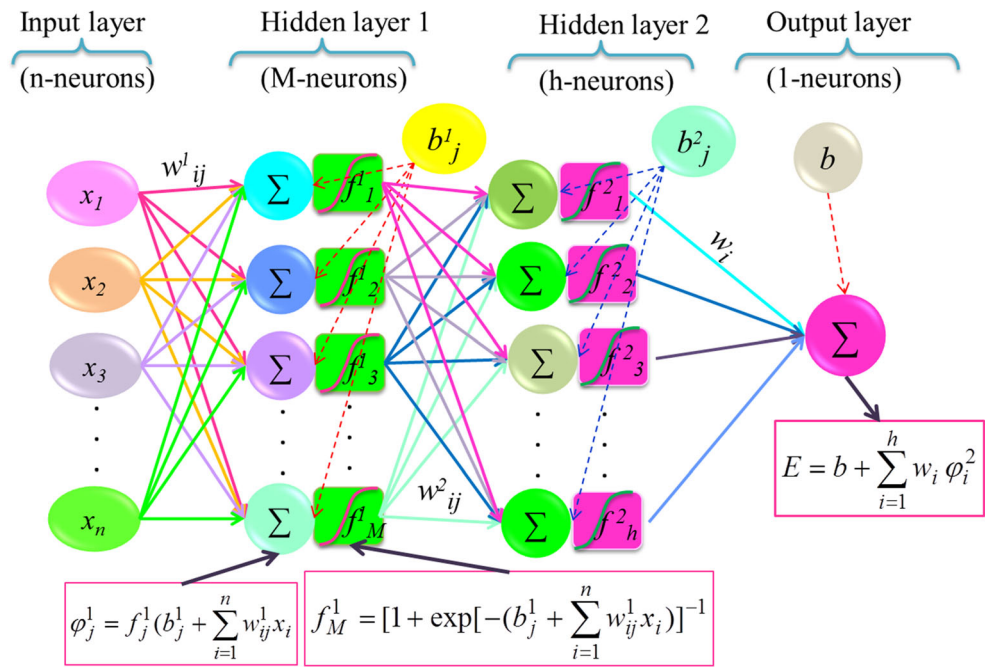
As seen, the weight is computed using the Lagrange multipliers as $w_i = \alpha_i - \alpha_i^*$ which depended on the parameters of the SVR as σ, ε, and C.

Hybrid model using SVR and RSM

In this study, the parameters of the SVR are selected as the constant values, while optimization processes can be applied to search the acceptable parameters for improving the accuracy of the predicted pan evaporation. Applying the optimization process for modeling using the SVR is a good strategy but it is a time-consuming algorithm with more computational burden. Consequently, the main effort in this study is to improve the predictions of SVR for calibrating pan evaporation with constant SVR parameters in order to reduce the computational burden of modeling process. The accuracy of the predicted model can be improved based on the applied input database as wavelet functions. Nevertheless, the effect of the input data can be controlled by filtering them using a mathematical relation based on the RSM. Generally, the AI-based data-driven using two-step calibration is a strategy to improve accuracy of prediction. This strategy is recently applied to enhance the RSM as modified RSM (Keshtegar and Kisi 2017) and multi-layer RSM (Kowsar et al. 2019). Using two calibrating process for improving the performance of the RSM approach, Keshtegar and Seghier (2018) applied the hybrid model for the burst pressure of corroded pipes and for solving hydrological problems (Keshtegar and Kisi 2017). This

Fig. 2 Schematic view of the MLPNN model for calibrating pan evaporation with structure n-M-h-1

$$\varphi_j^1 = f_j^1 \left(b_j^1 + \sum_{i=1}^{n} w_{ij}^1 x_i \right) \qquad f_M^1 = \left[1 + \exp\left[-\left(b_j^1 + \sum_{i=1}^{n} w_{ij}^1 x_i \right) \right] \right]^{-1}$$

$$E = b + \sum_{i=1}^{h} w_i \, \varphi_i^2$$

strategy is used to modify the prediction of SVR with hybrid modeling approach using RSM in first calibrating process and SVR in second calibrating process. Consequently, the nonlinear relations-based two modeling approaches may improve the accuracy predictions of pan evaporation.

These modeling methods SVR and RSM were combined using three basic layers in SVR_RSM model as the input data nodes in first layer, modeling nodes using RSM in second layer, and predicted pan evaporation using SVR in third layer. In the second and third layers, two modeling approaches of

Fig. 3 Schematic view of the RSM model

$$\hat{E} = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} w_{ij} x_i \, x_j$$
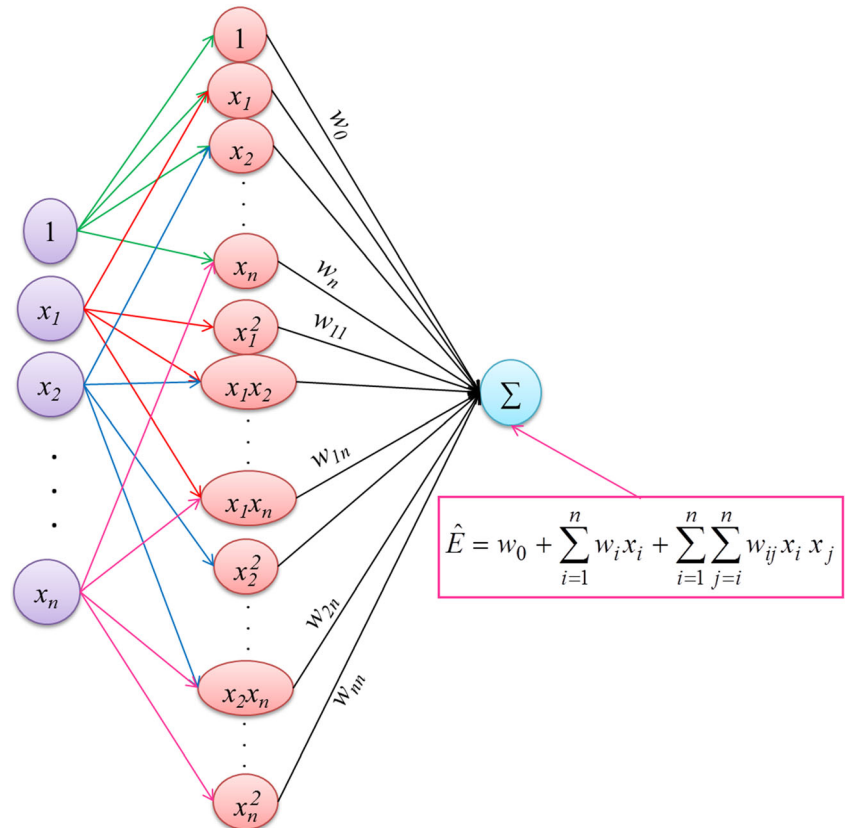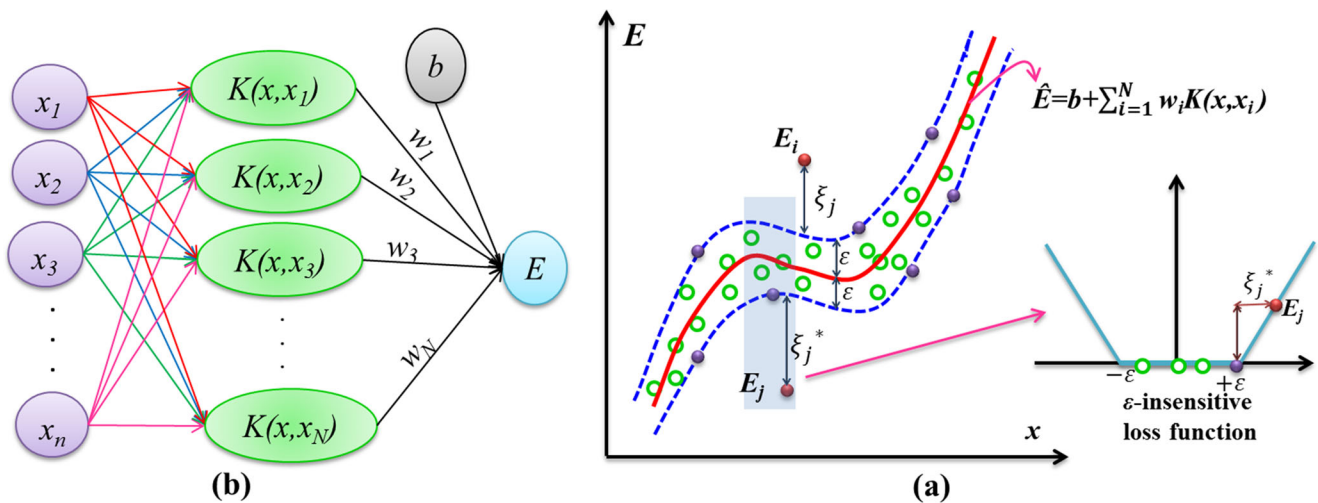
Fig. 4 Schematic view of SVR model: **a** structure of SVR and **b** calibrating data with ε-insensitive loss function

RSM and SVR are used to calibrate the pan evaporation. Generally, the nodes in the second calibrating process are computed using the predicted results of RSM with two individual input variables which are given from input database. Therefore, the RSM with cross terms is used in the first calibrating database on second layer as below relation:

$$\varphi_m = w_0 + w_1 x_{n-1} + w_2 x_n + w_3 x_{n-1}^2 + w_4 x_n^2$$
$$+ w_5 x_{n-1} x_n \tag{12}$$

where w_0-w_5 are weights which represent the connection of the input databases x_{n-1} and x_n with the data of node m. The total number of nodes in the second layer which are obtained using Eq. (12) with n input data is $m = \frac{n!}{2! \times (n-2)!}$, where (!) represents the factorial operator. Thus, it can be provided m predicted nodes-based RSM that this dataset is used to calibrate SVR in third layer by the following relation:

$$\hat{E} = b + \sum_{i=1}^{N} w_i K(\varphi, \varphi_i) \tag{13}$$

where b is bias and $K(\varphi, \varphi_i)$ represents the Kernel function which transferred the data-based predicted RSM from the real-space into feature-space. Commonly, the Gaussian Kernel function as well as the original SVR with parameter of $\sigma = 2$ while the other SVR in this hybrid modeling approach are given as $\varepsilon = 0.1$ and $C = 300$. The structure of the SVR_RSM is presented in Fig. 5. As seen from Fig. 5, the SVR_RSM was structured using two nonlinear models of RSM combined by SVR which is calibrated using input data handing by RSM predictions. The SVR predictions depended on the input dataset which are provided using RSM. Consequently, the accuracy prediction can be affected on agreement and accuracy of SVR. The cross-linear correlation of input variables is considered in first calibrating process using RSM, while the nonlinear correlation between the input

data and the observed pan evaporation are given using SVR in the second calibrating process by using m predicted nodes of RSM database. This calibrating model-based two phases may improve the prediction of monthly pan evaporation based on the following steps:

Step 1: Input dataset as train and test data points
Step 2: Separate different two data sets form training database
Step 3: Calibrate the nodes of first stage using the RSM with 2-set original input variables which are given from step 2
Step 4: Give the input database from data provided by RSM in step 3 for SVR as the input database in the training phase
Step 5: Set the parameters of SVR
Step 6: Calibrate SVR using the parameters in step 5 and input data in step 4
Step 7: Predict the data using hybrid intelligent model by using steps 3–6 for test data point

The SVR_RSM model using two calibrating processes may provide highly correlated between the observed and predicted pan evaporations. Consequently, it can provide the accurate predictions and is defined as a robust modeling approach.

Comparative statistics

Four models such as MLPNN, RSM, SVR, and SVR_RSM are used for nonlinear modeling of pan evaporation that their predictions were compared using five comparative statistics which are used to compare the accuracy and the agreement between the predicated and observed datasets. The statistics

Fig. 5 Schematic view of SVR_RSM model
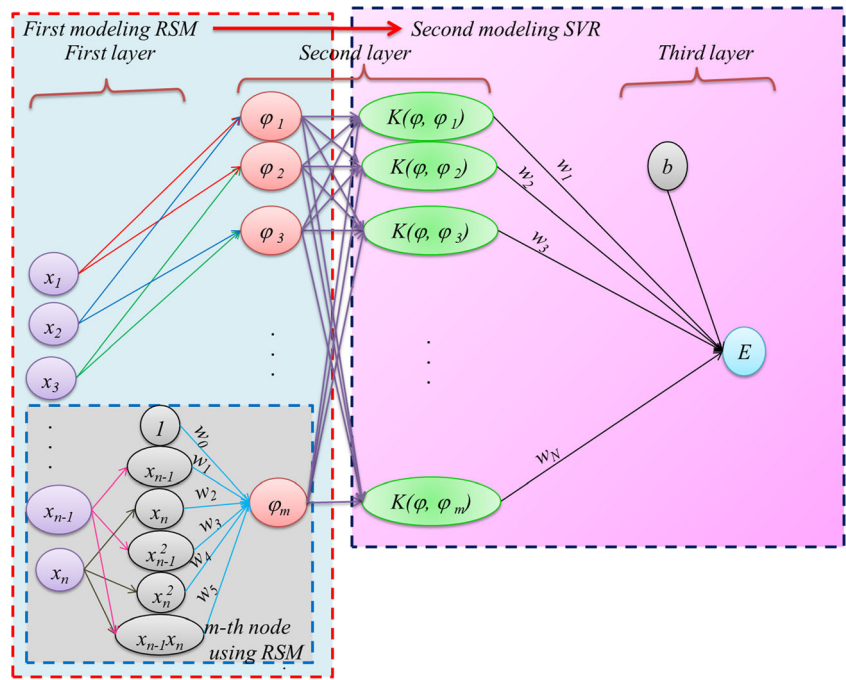
are root mean square error (RMSE), mean absolute error (MAE), modified index of agreement (*d*), coefficient of correlation (*R*), and modified Nash and Sutcliffe efficiency (NSE) with below relations (Keshtegar et al. 2019a, b):

$$R = \left[\frac{\frac{1}{N}\sum(E_i - E_m)(P_i - P_m)}{\sqrt{\frac{1}{N}\sum_{i=1}^{n}(E_i - E_m)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{n}(P_i - P_m)^2}} \right] \quad (14)$$

$$d = 1 - \frac{\sum_{i=1}^{N}(E_i - P_i)}{\sum_{i=1}^{N}(|O_i - E_m| + |P_i - E_m|)} \quad (15)$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}[E_i - P_i]}{\sum_{i=1}^{N}[E_i - E_m]} \quad (16)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(E_i - P_i)^2} \quad (17)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|E_i - P_i| \quad (18)$$

Where N is number of observed datasets E_i and P_i are observed and predicted pan evaporation for ith data point, respectively. O_i and P_i are the mean of observed and calculated data. The nonlinear model is better than the other modeling techniques when its comparative statistics are computed for

MAE, RMSE values tended to zero and other agreement indexes as R, d, and NSE tended to 1.

Results and discussion

In this section, the MLPNN, SVR, RSM, and SVR_RSM algorithms were validated on the six dataset described in the previous section. The correspondence between the measured and model estimates of E_{pan} are reported in Tables 4, 5, 6, 7, 8, and 9, in terms of RMSE, MAE, R, NSE, and d, during the training and validations phases. Hereafter, we focused our discussion on the results obtained during the validation phase. At Ain Dalia station (Table 4), the estimation of monthly E_{pan} was robust, having RMSE ranging from 1.523 to 2.218 mm, MAE ranging from 1.094 to 1.651 mm, and correlation coefficient (R) ranged from 0.774 to 0.891, based on the six MLPNN models. Estimates based on NSE and d indexes were almost equally accurate results with NSE ranged from 0.508 to 0.768, and d ranged from 0.780 to 0.925. Table 4 shows that high variability of models performances has been observed between the first five MLPNN models (MLPNN1 to MLPNN5) and the sixth model (MLPNN6). These differences in accuracy of models could possibly be due to the exclusion of the relative humidity (H %) from the input variables of the MLPNN6 model. Among the six MLPNN models (Table 4), the MLPNN1 has the highest accuracy with lower RMSE and MAE, and high NSE and d values. The MLPNN1 has R value of 0.889, NSE of 0.768, and d of 0.922, when evaluated using the validation data. Predictor variables selected by this model

Table 4 Performances of different models in modeling monthly E_{pan} at Ain Dalia Station

Models	Training					Validation				
	MAE	RMSE	R	d	NSE	MAE	RMSE	R	d	NSE
MLPNN1	0.784	1.029	0.925	0.958	0.853	1.094	1.523	0.889	0.922	0.768
MLPNN2	0.826	1.091	0.916	0.953	0.835	1.150	1.542	0.876	0.925	0.762
MLPNN3	0.917	1.145	0.905	0.947	0.819	1.228	1.583	0.874	0.914	0.749
MLPNN4	0.899	1.179	0.900	0.947	0.808	1.153	1.574	0.870	0.922	0.752
MLPNN5	0.776	1.050	0.922	0.958	0.848	1.108	1.558	0.891	0.919	0.757
MLPNN6	1.289	1.684	0.836	0.841	0.608	1.651	2.218	0.774	0.780	0.508
RSM1	0.611	0.814	0.953	0.976	0.908	1.112	1.550	0.880	0.925	0.760
RSM2	0.774	1.009	0.927	0.961	0.859	1.114	1.548	0.874	0.924	0.760
RSM3	0.729	0.993	0.929	0.962	0.864	1.133	1.541	0.886	0.923	0.762
RSM4	0.800	1.049	0.921	0.957	0.848	1.123	1.514	0.885	0.926	0.771
RSM5	0.822	1.064	0.918	0.956	0.844	1.163	1.544	0.881	0.922	0.762
RSM6	0.918	1.216	0.892	0.941	0.795	1.455	1.846	0.813	0.890	0.659
SVR1	0.525	0.868	0.948	0.973	0.896	1.096	1.400	0.899	0.939	0.804
SVR2	0.680	1.049	0.923	0.958	0.848	1.057	1.476	0.889	0.931	0.782
SVR3	0.658	1.006	0.929	0.963	0.860	1.125	1.513	0.881	0.927	0.771
SVR4	0.750	1.062	0.921	0.958	0.844	1.119	1.492	0.884	0.930	0.777
SVR5	0.683	1.022	0.929	0.962	0.855	1.078	1.424	0.895	0.937	0.797
SVR6	0.761	1.115	0.911	0.953	0.828	1.212	1.632	0.859	0.917	0.734
SVR_RSM1	0.459	0.701	0.966	0.983	0.932	0.935	1.270	0.920	0.951	0.839
SVR_RSM2	0.612	0.925	0.940	0.967	0.882	0.977	1.374	0.907	0.941	0.811
SVR_RSM3	0.458	0.831	0.951	0.974	0.904	1.064	1.399	0.899	0.939	0.804
SVR_RSM4	0.723	0.997	0.930	0.964	0.863	1.001	1.394	0.903	0.940	0.806
SVR_RSM5	0.586	0.931	0.939	0.968	0.880	1.064	1.395	0.903	0.938	0.805
SVR_RSM6	0.704	1.046	0.923	0.960	0.849	1.190	1.616	0.861	0.923	0.739

included the four original climatic variables (T_{max}, T_{min}, H%, U_2) coupled with the periodicity (the month number: α). Finally, the R, NSE, and d values were further improved to 0.891, 0.757, and 0.919 when H% was used (MLPNN5) instead of U_2 (MLPNN6), and model errors as measured using RMSE and MAE were reduced by 29.75% and 32.89%, respectively. Table 4 indicates better error statistics in terms of MAE (1.455 mm) and RMSE (1.846 mm), and higher R (0.813), NSE (0.659), and d (0.890) values, using the RSM6 model against (R = 0.774), (NSE = 0.508), and (d = 0.780) obtained suing the MLPNN6 model. In addition, it is clear from Table 4 that using fewer input variables, the RSM model was able to provide high accuracy compared to the MLPNN model. Results clearly indicated that the MAE and RMSE of the RSM4 was decreased from 1.574 to 1.514 (2.6%) and from 1.153 to 1.123 (3.8%), respectively, compared to the MLPNN4. Overall, when comparing the RSM and the MLPNN models, it is clear that there was a strong relationship between the number of input variables and the accuracy of the models. The relationship varied, however, significantly for the MLPNN model, which means that RSM is more suitable for building robust models using only fewer inputs. The RSM6

with only T_{min}, U_2, and α as input variables improved the R, NSE, and d of the MLPNN6 by 3.9%, 11%, and 15.1%, respectively, and decreasing the values of the RMSE and MAE by 11.87% and 16.77%, respectively. According to Table 4, high accuracy was obtained using SVR models compared to the MLPNN and RSM models, for all the six input combination. The E_{pan} showed good correlation between measured and calculated values using the SVR models. This class of models was characterized by R, NSE, and d values varying from 0.859 to 0.899, 0.734 to 0.804, and from 0.917 to 0.939, higher than the values obtained using the MLPNN and the RSM models. SVR1 exhibited a decrease in RMSE and MAE value by 8.78% and 0% compared to the MLPNN1, and by 9.67% and 1.46% compared the RSM1, respectively. The good accuracy of the SVR models is obvious, especially when comparing the models having a fewer input variables. SVR6 decreased the RMSE and the MAE of the MLPNN6 by 26.42% and 26.95%, respectively, and by 11.59% and 16.70% compared to the RSM6, respectively. The proposed hybrid method of SVR_RSM produces high nonlinear mapping of E_{pan} which are substantially higher than the three other models, as stated in Table 4. This can be nearly always

Table 5 Performances of different models in modeling monthly E_{pan} at Beni Haroun Station

Models	Training					Validation				
	MAE	RMSE	R	d	NSE	MAE	RMSE	R	d	NSE
MLPNN1	0.676	1.013	0.959	0.979	0.915	1.083	1.458	0.935	0.959	0.835
MLPNN2	0.688	1.061	0.953	0.975	0.907	1.028	1.357	0.931	0.959	0.857
MLPNN3	0.827	1.286	0.933	0.964	0.863	0.976	1.308	0.948	0.964	0.867
MLPNN4	0.768	1.191	0.940	0.969	0.883	0.921	1.224	0.944	0.968	0.884
MLPNN5	1.194	1.493	0.908	0.944	0.816	1.022	1.342	0.937	0.959	0.860
MLPNN6	0.829	1.118	0.950	0.974	0.897	1.124	1.373	0.932	0.962	0.854
RSM1	0.709	0.998	0.958	0.978	0.918	0.957	1.323	0.942	0.964	0.864
RSM2	0.752	1.107	0.948	0.973	0.899	0.964	1.364	0.943	0.959	0.856
RSM3	0.702	1.051	0.953	0.976	0.909	0.925	1.309	0.942	0.964	0.867
RSM4	0.800	1.163	0.943	0.970	0.888	0.949	1.175	0.950	0.970	0.893
RSM5	0.781	1.135	0.945	0.971	0.894	0.949	1.272	0.942	0.966	0.875
RSM6	0.951	1.273	0.931	0.963	0.866	1.044	1.401	0.925	0.959	0.848
SVR1	0.724	1.207	0.939	0.968	0.880	0.779	1.130	0.952	0.973	0.901
SVR2	0.731	1.054	0.954	0.975	0.908	1.078	1.357	0.939	0.960	0.857
SVR3	0.739	1.201	0.939	0.968	0.881	0.808	1.134	0.951	0.973	0.900
SVR4	0.744	1.006	0.957	0.978	0.917	0.976	1.193	0.948	0.970	0.890
SVR5	0.789	1.145	0.945	0.970	0.892	0.944	1.189	0.945	0.969	0.890
SVR6	0.889	1.373	0.919	0.957	0.845	1.019	1.393	0.923	0.959	0.850
SVR_RSM1	0.675	0.943	0.963	0.980	0.927	0.745	1.095	0.954	0.975	0.907
SVR_RSM2	0.523	0.873	0.969	0.983	0.937	0.798	1.179	0.948	0.971	0.892
SVR_RSM3	0.517	0.869	0.969	0.983	0.938	0.841	1.114	0.953	0.974	0.904
SVR_RSM4	0.518	0.872	0.969	0.983	0.937	0.762	1.071	0.957	0.977	0.911
SVR_RSM5	0.585	1.023	0.957	0.977	0.914	0.776	1.129	0.952	0.974	0.901
SVR_RSM6	0.685	0.932	0.964	0.980	0.928	0.938	1.232	0.944	0.969	0.882

advantageous, especially in the situation where the input variables were selected smaller. As can be seen, using only the T_{max}, H%, and the periodicity (α), an R, NSE, and d values of 0.903, 0.805, and 0.923 were achieved, which were not exhibited by any of the other models. An RMSE of 1.395 mm was achieved using the SVR_RSM5 model less than all the values provided by the MLPNN5, RSM5, and SVR5 models. Similarly, the hybrid SVR_RSM1 model generated with all input variables has produced an R of 0.920, NSE of 0.839, and d of 0.951 (Table 4) which are higher than the values provided by the MLPNN1, RSM1, and SVR1, models. Both SVR_RSM4 and SVR_RSM5 yielded similar accuracy in terms of all the five statistical indexes, and slightly higher than the SVR_RSM3. Performance characteristics also differ between SVR_RSM1 and SVR_RSM2, with and without periodicity (α). Generally, the SVR_RSM2 without periodicity (α) yielded higher RMSE (1.374 mm) compared to (1.270 mm) achieved using the SVR_RSM1. Of the six input combination, the SVR_RSM6 had the largest RMSE value (1.616 mm). Moreover, the SVR_RSM6 generally yields the highest MAE (1.190 mm) and the lowest R, NSE, and d values.

Results at Beni Haroun station are reported in Table 5. The statistics indexes given in Table 5 show how well E_{pan} can be estimated from the climatic variables using the proposed models. The results obtained using the four models showed generally strong relationships between measured and calculated E_{pan} values (Table 5). The low RMSE and MAE values indicate small variations between measured and estimated E_{pan} for the training and validation data set. Using all the four climatic variables in addition to the periodicity, the hybrid model SVR_RSM1 provided slightly better prediction results ($R = 0.954$ and NSE $= 0.907$) and the lowest errors indexes (RMSE $= 1.095$ mm and MAE $= 0.745$ mm). Overall accuracy results were high for all the proposed models and the SVR_RSM1 produced the highest overall accuracy, followed by SVR1, and the RSM1 performed best compared to the MLPNN1. In summary, the statistics performance calculated between measured and predicted E_{pan} values showed that both SVR1 and SVR_RSM1 methods similarly perform with accuracy slightly higher than the RSM1 model, and largely higher than the MLPNN1 model. The performance differences between SVR1 and SVR_RSM1 methods, however,

Table 6 Performances of different models in modeling monthly E_{pan} at Bouhamden Station

Models	Training					Validation				
	MAE	RMSE	R	d	NSE	MAE	RMSE	R	d	NSE
MLPNN1	0.469	0.691	0.966	0.981	0.931	0.588	0.826	0.936	0.964	0.874
MLPNN2	0.423	0.677	0.969	0.982	0.934	0.653	0.842	0.940	0.963	0.869
MLPNN3	0.450	0.698	0.965	0.982	0.930	0.576	0.798	0.943	0.970	0.882
MLPNN4	0.527	0.743	0.961	0.978	0.920	0.652	0.869	0.932	0.960	0.860
MLPNN5	0.430	0.661	0.969	0.983	0.937	0.685	0.834	0.936	0.965	0.872
MLPNN6	0.460	0.624	0.972	0.986	0.944	0.684	0.898	0.927	0.958	0.851
RSM1	0.436	0.658	0.968	0.984	0.938	0.574	0.798	0.941	0.968	0.882
RSM2	0.442	0.669	0.967	0.983	0.935	0.578	0.793	0.942	0.968	0.884
RSM3	0.486	0.692	0.965	0.982	0.931	0.634	0.828	0.936	0.965	0.873
RSM4	0.500	0.716	0.962	0.980	0.926	0.549	0.746	0.949	0.971	0.897
RSM5	0.488	0.720	0.962	0.980	0.925	0.569	0.760	0.945	0.971	0.893
RSM6	0.654	0.879	0.943	0.970	0.889	0.753	0.991	0.910	0.943	0.818
SVR1	0.582	0.682	0.969	0.982	0.933	0.681	0.822	0.936	0.966	0.875
SVR2	0.585	0.715	0.968	0.979	0.926	0.729	0.903	0.924	0.956	0.849
SVR3	0.584	0.687	0.968	0.982	0.932	0.669	0.814	0.938	0.967	0.877
SVR4	0.559	0.707	0.965	0.982	0.928	0.618	0.826	0.935	0.966	0.874
SVR5	0.607	0.718	0.964	0.981	0.926	0.557	0.690	0.960	0.977	0.912
SVR6	0.586	0.730	0.962	0.979	0.923	0.735	0.931	0.918	0.954	0.840
SVR_RSM1	0.477	0.584	0.980	0.986	0.951	0.479	0.677	0.961	0.976	0.915
SVR_RSM2	0.463	0.605	0.976	0.985	0.947	0.521	0.708	0.956	0.974	0.907
SVR_RSM3	0.478	0.585	0.980	0.986	0.951	0.475	0.677	0.961	0.976	0.915
SVR_RSM4	0.453	0.638	0.972	0.984	0.941	0.466	0.687	0.962	0.975	0.913
SVR_RSM5	0.523	0.672	0.970	0.982	0.935	0.471	0.634	0.963	0.980	0.926
SVR_RSM6	0.417	0.599	0.975	0.986	0.948	0.626	0.830	0.946	0.963	0.873

are not large and therefore, E_{pan} based on climatic variables can be predicted very well by the models. Using only fewer input variables, the SVR_RSM6 showed the best performances among the other models, with overall R and NSE of 0.944 and 0.882, respectively, RMSE of 1.232 mm, and MAE of 0.938 mm. For comparison, the RSM6 and SVR6 performed with equal accuracy, while MLPNN6 modeled E_{pan} reasonably and slightly less than the SVR_RSM6 (Table 5). When the periodicity (α) was excluded from the input variables, the RMSE and MAE of the SVR_RSM2 decreased by 13.11% and 25.97% compared to the SVR2, by 13.56% and 17.22% compared to the RSM2, and by 13.11% and 22.337% compared to the MLPNN2, respectively. In summary, MLPNN2 and SVR2 performed similarly in comparison with RSM2 and SVR_RSM2. RSM2 produces relatively large RMSE error (1.364 mm) slightly higher than the values calculated using MLPNN2 and SVR2 (RMSE = 1.357 mm). SVR_RSM2 shows improved performance estimates compared to RSM2, showing the best performance among the three models, indicated by the model performance statistics in Table 5. Nonetheless, the use of the data-driven models without hybridization did not result in substantial differences

with respect to the statistical indexes, and only the hybrid model was characterized by the strongest accuracy.

Results at Bouhamden Station are reported in Table 6. According to Table 6, the models performance varied significantly with R values ranging from 0.927 to 0.943, NSE values ranging from 0.851 to 0.882, and the d values ranging from 0.958 to 0.970, for the MLPNN models. The SVR_RSM1 model was able to predict E_{pan} much better than the SVR1 (R = 0.961 versus R2 = 0.936). The RSM1 model performs better than the MLPNN1 (R = 0.941 versus R = 0.936) and better than SVR1 (R = 0.941 versus R = 0.936). Overall, except the high performances exhibited using the hybrid SVR_RSM1 model (R = 0.941, NSE = 0.915, and d = 0.976), the difference in R among the various approaches was not large and did not considerably affect the correlation between measured and estimated monthly evaporation (E_{pan}) and no greatest variability was observed between the models performances. Although slightly more accurate than the SVR_RSM1, estimated statistical indexes values using the SVR_RSM5 (R = 0.963, NSE = 0.926, and d = 0.980) were only weakly higher than those provided by the SVR_RSM1 (R = 0.961, NSE = 0.915, and d = 0.976). Consequently, the

Table 7 Performances of different models in modeling monthly E_{pan} at Chaffia Station

Models	Training					Validation				
	MAE	RMSE	R	d	NSE	MAE	RMSE	R	d	NSE
MLPNN1	0.376	0.519	0.957	0.972	0.906	0.369	0.509	0.950	0.969	0.893
MLPNN2	0.392	0.549	0.946	0.971	0.894	0.362	0.466	0.955	0.976	0.910
MLPNN3	0.309	0.416	0.969	0.984	0.939	0.366	0.494	0.949	0.974	0.899
MLPNN4	0.331	0.450	0.964	0.981	0.929	0.354	0.504	0.947	0.972	0.895
MLPNN5	0.347	0.456	0.964	0.980	0.927	0.395	0.510	0.949	0.970	0.892
MLPNN6	0.359	0.504	0.955	0.977	0.911	0.467	0.652	0.924	0.958	0.824
RSM1	0.296	0.412	0.970	0.984	0.941	0.328	0.463	0.955	0.976	0.911
RSM2	0.396	0.529	0.950	0.974	0.902	0.322	0.456	0.958	0.976	0.914
RSM3	0.303	0.416	0.969	0.984	0.939	0.320	0.457	0.956	0.977	0.914
RSM4	0.337	0.455	0.963	0.981	0.927	0.333	0.472	0.953	0.975	0.908
RSM5	0.326	0.439	0.966	0.982	0.932	0.324	0.448	0.959	0.977	0.917
RSM6	0.406	0.553	0.945	0.971	0.893	0.445	0.611	0.923	0.960	0.846
SVR1	0.296	0.372	0.976	0.987	0.951	0.394	0.501	0.947	0.972	0.896
SVR2	0.355	0.474	0.960	0.979	0.921	0.373	0.489	0.950	0.973	0.901
SVR3	0.295	0.377	0.976	0.987	0.950	0.372	0.495	0.948	0.973	0.899
SVR4	0.332	0.431	0.967	0.983	0.935	0.376	0.496	0.948	0.972	0.898
SVR5	0.309	0.398	0.972	0.986	0.945	0.354	0.470	0.954	0.976	0.909
SVR6	0.386	0.533	0.949	0.973	0.900	0.436	0.598	0.927	0.962	0.852
SVR_RSM1	0.346	0.396	0.977	0.984	0.945	0.329	0.442	0.964	0.977	0.919
SVR_RSM2	0.378	0.444	0.970	0.981	0.931	0.289	0.423	0.965	0.979	0.926
SVR_RSM3	0.233	0.347	0.979	0.989	0.958	0.289	0.428	0.964	0.979	0.924
SVR_RSM4	0.335	0.426	0.968	0.984	0.936	0.329	0.458	0.957	0.976	0.913
SVR_RSM5	0.306	0.389	0.973	0.986	0.947	0.289	0.408	0.968	0.981	0.931
SVR_RSM6	0.313	0.466	0.961	0.980	0.924	0.397	0.564	0.937	0.967	0.869

hybrid SVR_RSM models are usually characterized by fairly high capabilities for providing high models accuracy using only fewer input variables and are more highly suitable for modeling nonlinear process such as pan evaporation. The different input combinations had a greater impact on the statistical errors index (RMSE and MAE). Errors were consistently lower when fewer variables were used. Excluding the U_2 from the input variables of the MLPNN2 model resulted in smaller gains in accuracies compared to the MLPNN3: (RMSE = 0.576 mm versus R = 0.653 mm) and (MAE = 0.576 mm versus MAE = 0.653 mm). Similarly, using only T_{max} and T_{min} as input of the RSM4 instead of the all four climatic variables (RSM2) slightly improves its performances, despite the smaller gains in accuracies: (RMSE = 0.746 mm versus R = 0.798 mm) and (MAE = 0.549 mm versus MAE = 0.574 mm). Using only the T_{min} and U_2% coupled with the periodicity (α), the difference in RMSE and MAE terms is maximized between RSM6 and RSM2, permitting an improved estimation of E_{pan}: the RMSE and MAE were decreased by 19.98% and 23.24%, respectively. Finally, when comparing the performances of the two hybrid models: SVR_RSM5 and SVR_RSM2 approaches, results were not significantly different. RMSE and MAE do not change more than 4.16% and 1.55% between these two approaches.

The comparisons between calculated E_{pan} using the proposed models and measured values at Chaffia station are presented in Table 7. For this station, it is clear that all the models performed very well, and the statistical indexes indicate an important finding: inclusion of the periodicity coupled with the four climatic variables do not help to improving the performances of the models; on the contrary, the accuracy of the models was decreased. From Table 7, when the periodicity (α) is excluded from the inputs variable, we obtained the best performances among the compared four models. Globally, exclusion of the α improves the performance of the MLPNN2 model by slightly increasing the values of the R, NSE, and d by 0.5%, 0.7%, and 1.7%, and decreasing the values of the RMSE and MAE by 8.44% and 1.9%, respectively, (Table 7, models MLPNN1 and MLPNN2). In addition, the RSM2 model (without α) improves the accuracy of the RSM1 model (with α) by increasing the R and d values by 0.3%, and decreasing the RMSE and MAE values by 1.51% and 1.82%, respectively. Nevertheless, by comparing the performances of the SVR2 model with the performances of the

Table 8 Performances of different models in modeling monthly E_{pan} at El Agram Station

Models	Training					Validation				
	MAE	RMSE	R	d	NSE	MAE	RMSE	R	d	NSE
MLPNN1	0.390	0.562	0.965	0.981	0.929	0.425	0.578	0.978	0.985	0.946
MLPNN2	0.600	0.752	0.935	0.965	0.873	0.666	0.840	0.949	0.968	0.887
MLPNN3	0.455	0.607	0.959	0.978	0.917	0.432	0.550	0.977	0.987	0.951
MLPNN4	0.439	0.570	0.964	0.981	0.927	0.410	0.520	0.981	0.988	0.957
MLPNN5	0.534	0.653	0.952	0.973	0.904	0.563	0.714	0.972	0.975	0.918
MLPNN6	0.669	0.836	0.919	0.954	0.843	0.719	0.901	0.945	0.959	0.870
RSM1	0.377	0.489	0.973	0.986	0.946	0.466	0.634	0.970	0.982	0.935
RSM2	0.602	0.756	0.934	0.965	0.872	0.717	0.884	0.940	0.965	0.875
RSM3	0.399	0.523	0.969	0.984	0.939	0.420	0.590	0.976	0.985	0.944
RSM4	0.416	0.534	0.968	0.983	0.936	0.376	0.515	0.981	0.988	0.957
RSM5	0.462	0.586	0.961	0.980	0.923	0.452	0.643	0.972	0.982	0.934
RSM6	0.583	0.747	0.935	0.966	0.875	0.554	0.776	0.962	0.971	0.903
SVR1	0.301	0.453	0.977	0.988	0.954	0.574	0.866	0.944	0.965	0.880
SVR2	0.496	0.701	0.945	0.969	0.890	0.654	0.971	0.927	0.954	0.849
SVR3	0.377	0.497	0.972	0.985	0.945	0.442	0.682	0.971	0.978	0.925
SVR4	0.411	0.532	0.968	0.983	0.937	0.380	0.511	0.983	0.988	0.958
SVR5	0.431	0.560	0.965	0.981	0.930	0.500	0.771	0.961	0.972	0.904
SVR6	0.581	0.743	0.937	0.965	0.876	0.569	0.778	0.965	0.970	0.903
SVR_RSM1	0.288	0.390	0.983	0.991	0.966	0.358	0.457	0.985	0.991	0.966
SVR_RSM2	0.473	0.657	0.951	0.974	0.903	0.698	0.839	0.950	0.966	0.887
SVR_RSM3	0.324	0.433	0.979	0.989	0.958	0.390	0.502	0.985	0.989	0.960
SVR_RSM4	0.400	0.465	0.976	0.988	0.952	0.359	0.436	0.989	0.992	0.970
SVR_RSM5	0.427	0.526	0.969	0.984	0.938	0.372	0.512	0.984	0.988	0.958
SVR_RSM6	0.510	0.680	0.947	0.971	0.896	0.503	0.714	0.973	0.975	0.918

SVR1 model, we concluded that the periodicity (α) had only a marginal effect and the two models have similar overall performances; SVR2 model, improves the accuracy of the SVR1 model by increasing the R, NSE, and d values by 0.3%, 0.1%, and 0.1%, respectively, and decreasing the RMSE and MAE values by 2.39% and 5.33%, respectively. Finally, the hybrid SVR_RSM2 model is a particular case in which the obtained results reveal significant improvement of the models performances. While the R, NSE, and d statistical indexes appear to be relatively equal, the SVR_RSM2 model outperforms the SVR_RSM1 when looking to the MAE value: the hybrid SVR_RSM2 decreased the MAE of the SVR_RSM1 by 12.15%. Overall, among all the proposed models, SVR_RSM5 was found to produce the highest correlation between calculated and measured E_{pan} ($R = 0.968$, NSE = 0.931, $d = 0.981$).

Results reported in Table 8 deals with the accuracy of the proposed models at El Agram station, and it is clear that the four proposed models were found to have a high performance in estimating E_{pan}. By comparison, the hybrid SVR_RSM model has better performances, both during training or validation phase. Also, it is found that models with T_{max}, T_{min}, and the periodicity (combination 4, Table 8) were more powerful than the others in terms of prediction accuracy. Results of the accuracy assessment for the MLPNN models showed that MLPNN4 produced the best accurate estimation of E_{pan} as shown by all statistical indexes. For comparison, MLPNN4 decreased the RMSE and MAE of the MLPNN1 by 10.03% and 3.53%, respectively. Also, it is clear from Table 8 that when T_{min} was used as input variable for the MLPNN4 instead of U_2 for model MLPNN6, the performances was increased, by increasing the R, NSE, and d values, and decreasing the RMSE and MAE values. MLPNN4 improves the accuracy of the MLPNN6 model by increasing the R, NSE, and d values by 3.6%, 2.9%, and 8.7%, respectively, and significantly decreasing the RMSE and MAE by 42.28% and 42.97%, respectively. Regarding the RSM models, RSM4 produced the best accurate estimation of E_{pan} and contributed to the overall decrease in RMSE and MAE of the RSM1 by 18.76% and 19.31%, respectively. In addition, it is clear from Table 8 that RSM4 is less affected by removing H% (RSM5) than U_2 (RSM6): the RMSE and MAE of the RSM4 was decreased by 19.90% and 16.81% compared to the RSM5, and by 33.63% and 32.13% compared to RSM6. A comparison of

Table 9 Performances of different models in modeling monthly E_{pan} at Zit Emba Station

Models	Training					Validation				
	MAE	RMSE	R	d	NSE	MAE	RMSE	R	d	NSE
MLPNN1	0.971	1.215	0.915	0.955	0.830	0.913	1.159	0.933	0.964	0.854
MLPNN2	0.932	1.176	0.919	0.957	0.840	1.138	1.403	0.901	0.940	0.786
MLPNN3	0.875	1.127	0.925	0.959	0.854	0.969	1.172	0.928	0.961	0.851
MLPNN4	0.838	1.083	0.933	0.964	0.865	0.869	1.101	0.936	0.967	0.868
MLPNN5	0.857	1.181	0.917	0.956	0.839	0.922	1.214	0.921	0.957	0.840
MLPNN6	0.785	1.016	0.943	0.967	0.881	0.869	1.064	0.938	0.965	0.877
RSM1	0.779	0.968	0.944	0.971	0.892	0.896	1.165	0.939	0.963	0.853
RSM2	0.978	1.222	0.910	0.951	0.828	1.310	1.547	0.890	0.926	0.740
RSM3	0.770	0.986	0.942	0.969	0.888	0.919	1.205	0.932	0.961	0.842
RSM4	0.815	1.042	0.935	0.965	0.875	0.914	1.150	0.940	0.965	0.856
RSM5	0.894	1.165	0.918	0.956	0.843	0.872	1.101	0.938	0.963	0.868
RSM6	0.914	1.130	0.923	0.959	0.853	0.952	1.149	0.937	0.962	0.856
SVR1	0.643	0.908	0.952	0.975	0.905	0.918	1.171	0.931	0.963	0.851
SVR2	0.886	1.165	0.919	0.957	0.843	1.202	1.467	0.889	0.931	0.766
SVR3	0.651	0.923	0.950	0.974	0.902	0.937	1.243	0.923	0.959	0.832
SVR4	0.751	1.089	0.930	0.963	0.863	1.072	1.269	0.933	0.961	0.825
SVR5	0.817	1.148	0.922	0.958	0.848	0.865	1.108	0.932	0.964	0.867
SVR6	0.782	1.055	0.934	0.964	0.872	0.825	0.986	0.947	0.972	0.894
SVR_RSM1	0.511	0.742	0.968	0.984	0.937	0.781	0.991	0.950	0.972	0.893
SVR_RSM2	0.752	1.035	0.937	0.967	0.876	1.161	1.371	0.911	0.942	0.796
SVR_RSM3	0.542	0.838	0.959	0.979	0.919	0.807	1.037	0.944	0.970	0.883
SVR_RSM4	0.671	0.992	0.942	0.969	0.886	0.832	1.007	0.947	0.972	0.890
SVR_RSM5	0.761	1.104	0.928	0.962	0.859	0.838	1.094	0.938	0.967	0.870
SVR_RSM6	0.655	0.978	0.946	0.972	0.890	0.814	1.047	0.949	0.972	0.881

results obtained using the SVR models shows that E_{pan} estimates based on the T_{max}, T_{min}, and α variable (SVR4, Table 8) are better than those from the other models. The performances of the SVR4 model are improved by an increase of R from 0.944 to 0.983, NSE from 0.880 to 0.958, and d from 0.965 to 0.988 respectively, compared to the SVR1 with all the four climatic variables. The SVR1 method shows slightly less error than the SVR2 (without α) in terms of RMSE (10.81%) and MAE (12.23) statistics, while it performs worse than the SVR3: the SVR3 decreased the RMSE and MAE of the SVR1 by 21.24% and 22.98%, respectively. Finally, it can be seen that the hybrid SVR_RSM4 models shows the best performance for the all six input combinations, owing to its smallest RMSE and MAE values and largest values of R, NSE, and d. The SVR_RSM4 model reduce the RMSE and MAE to nearly 0.436 mm and 0.359 mm, less than any values calculated using the other models, and the peak values of the R (0.989), NSE (0.970), and d (0.992) reached during the validation phase clearly underlines the superiority and the robustness of the hybrid approach. The SVR_RSM1 model had a comparable performance with SVR_RSM3, with slight superiority in favor to the SVR_RSM1. While results obtained

from the SVR_RSM1 show an increasing performances compared to the SVR_RSM2, for which the periodicity is removed with an increasing in the R, NSE, and d by 3.5%, 2.5%, and 7.5%, respectively. Similarly, SVR_RSM5 performs well compared to the SVR_RSM2, but relatively poor compared to the SVR_RSM1 and SVR_RSM4.

Table 9 summarizes the results obtained using the six different input combinations for the four different models. In all input combination without exception, all the statistical indexes values show that the hybrid SVR_RSM is the best model. The comparison between measured E_{pan} and the calculated values are reported as well. During the validation phase, the statistical indexes for different models shows that the RMSE and MAE are lowest for the SVR_RSM model (Table 9), and the best accuracy with lowest values was obtained using SVR_RSM1 (RMSE = 0.991 mm and MAE = 0.781 mm). In general, the agreement between measured and calculated values of E_{pan} is good: the R, NSE, and d values are generally larger than 0.90, 0.88, and 0.94, respectively for most of the models, and the RMSE are reasonably low. When using only the T_{max} and T_{min} as the unique input variables coupled with the periodicity (SVR_RSM4), the R values is often close to 0.947 and the
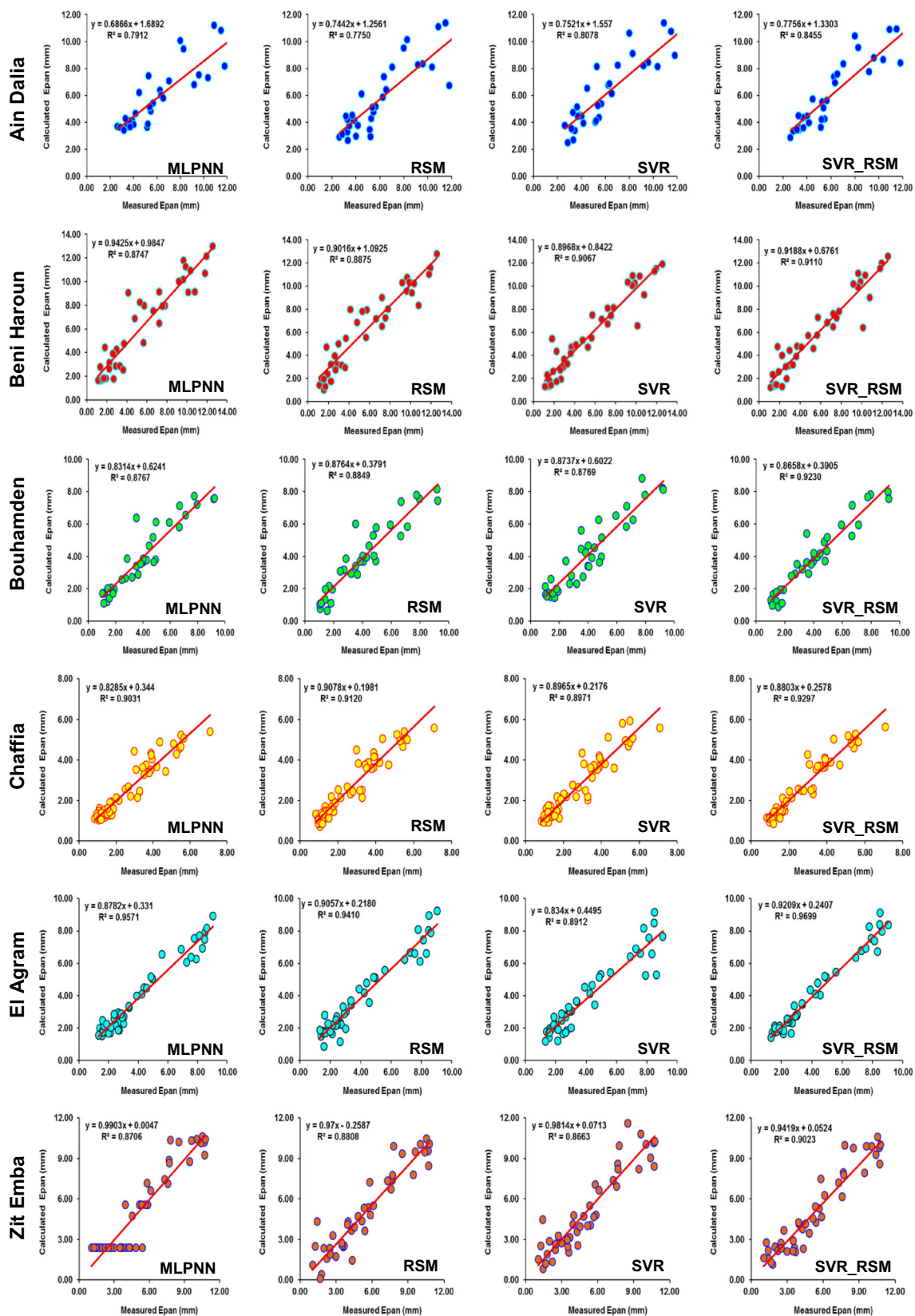
Fig. 6 Scatterplots showing the relation between the measured and calculated values of monthly evaporation (E_{pan}) in the validation phase for all stations
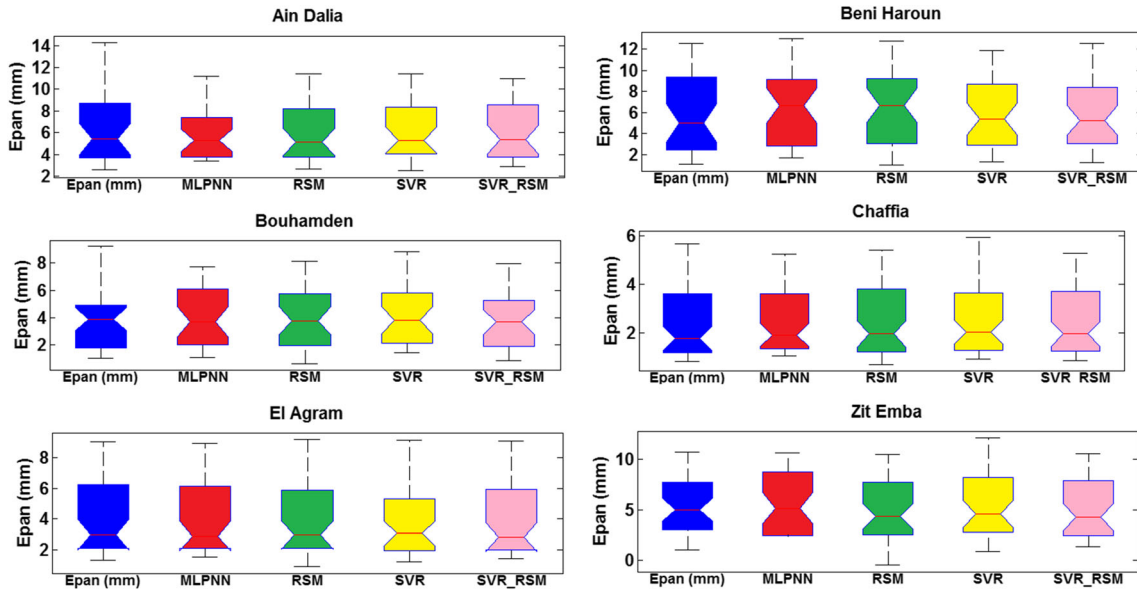
Fig. 7. Boxplots of measured and calculated values of monthly pan evaporation (E_{pan}) in the validation phase of all stations. The central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers correspond to the most extreme data points

RMSE is very low (1.007 mm). Compared to the other models, SVR_RSM1 is more accurate, followed by SVR_RSM4 (RMSE = 1.007 mm and MAE = 0.832 mm) and SVR_RSM6 (RMSE = 1.047 mm and MAE = 0.814 mm) with relatively similar accuracy and a favor to the SVR_RSM4 when looking to the NSE value, and finally the SVR_RSM2 is ranked in the last place with the highest RMSE (1.371 mm) and MAE (1.161 mm) values, respectively. From Table 9, it is clear that using all the four climatic variables

without periodicity reveal important finding: MLPNN2, RSM2, SVR2, and SVR_RSM2 were the worst models, for which the R, NSE, and d values have been decreased consistently with an increasing in the RMSE and MAE values. Removing the periodicity increased the MAE and RMSE of the MLPNN1, RSM1, SVR1, and the hybrid SVR_RSM1 model by (19.77% and 17.39%), (31.6% and 24.69%), (23.63% and 20.18%), and (32.73% and 27.72%), respectively; consequently, an important point should be distinguished:
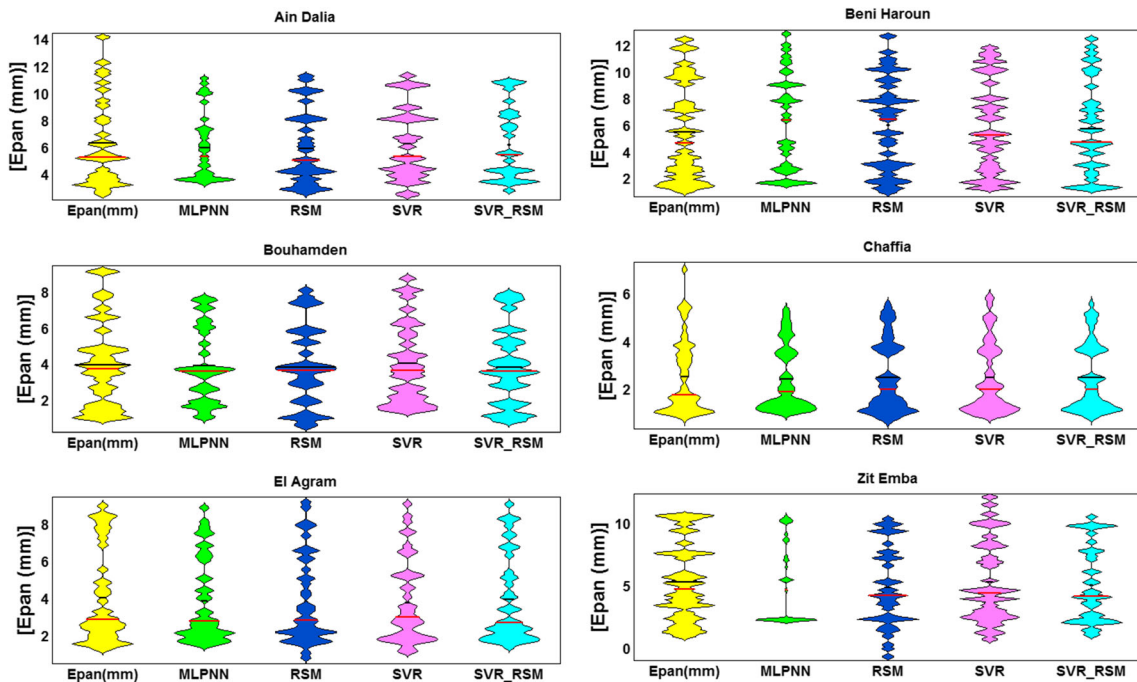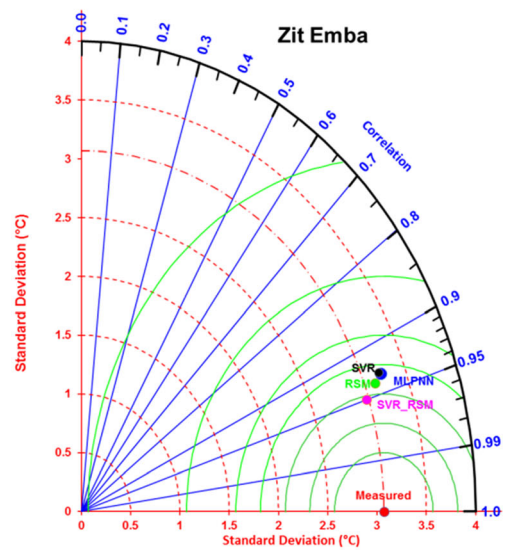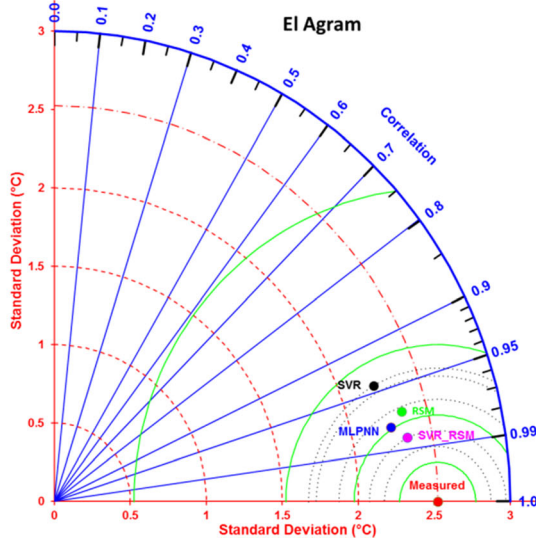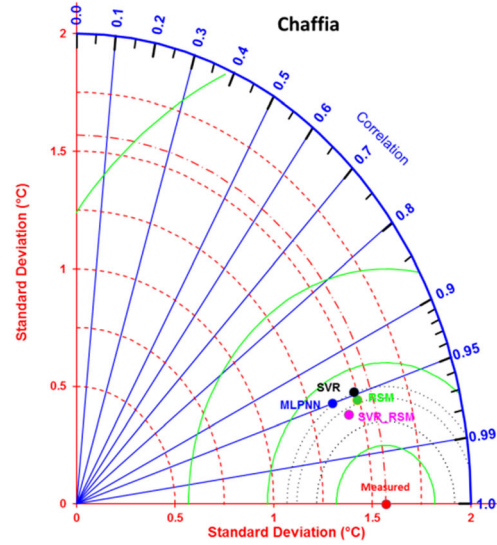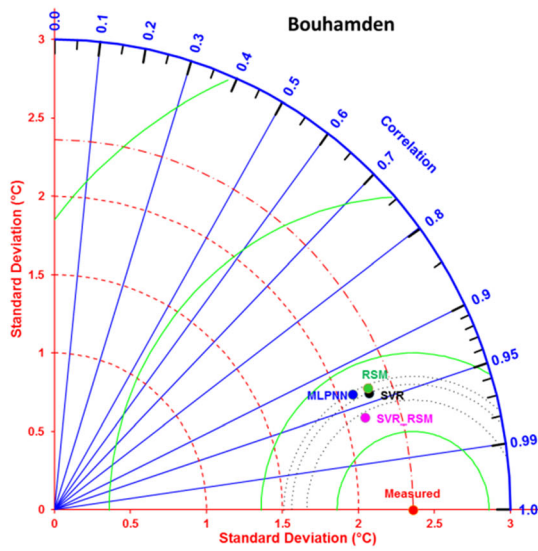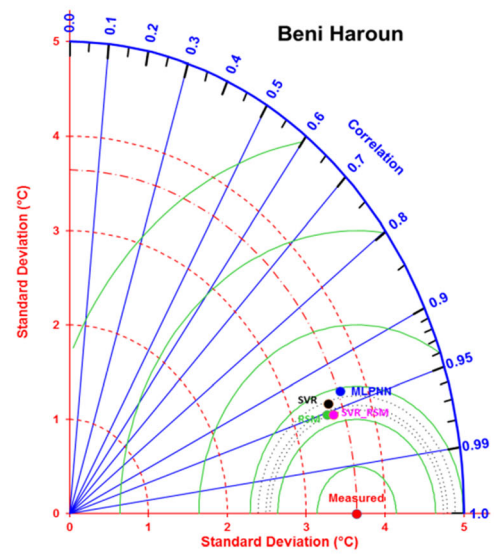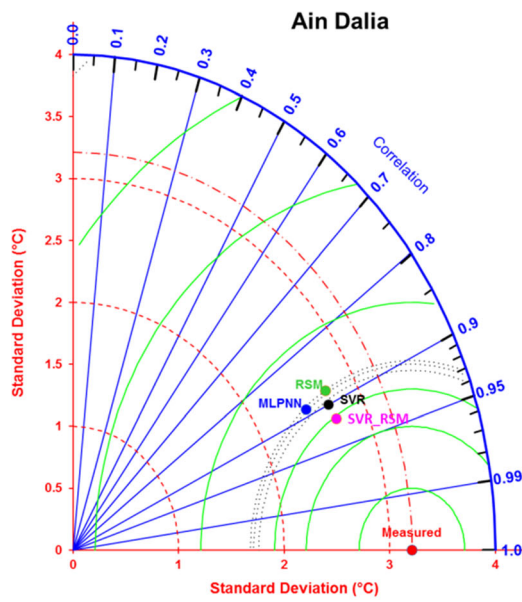

Fig. 8 Violin plots of measured and calculated monthly pan evaporation (E_{pan}) in the validation phase of all stations. The two lines with black and red color display the mean and the median values of E_{pan}

Ain Dalia

Beni Haroun

Bouhamden

Chaffia

El Agram

Zit Emba

Fig. 9 Taylor diagram displaying a statistical comparison of the four developed models with measured values of monthly pan evaporation E_{pan} (mm). The green circles correspond to circumferences of equal centered normalized root-mean-square (NRMS) difference between measured and calculated E_{pan}, the blue lines correspond to lines of equal correlation coefficients, and doted red circles correspond to circumferences of equal standard deviations

the periodicity is more important for the hybrid model compared to the other.

Finally, Scatterplot comparison between measured and calculated values of the E_{pan} was plotted in Fig. 6. Figures 7 and 8 show the boxplots and the violin plots for measured and calculated E_{pan} (mm) values using the SVR, RSM, MLPNN, and SVR_RSM models for all stations. For the violin plot, the two lines with black and red color display the mean and the median values of E_{pan}. Figure 9 presents a Taylor diagram plot calculated against the measured data at the six climatic stations. In the diagram, the accuracy of the four best models was compared during the validation phase using the correlation coefficient and the standard deviation. In the diagram, the measured values are plotted using a red circle along the X-axis. At all stations, SVR_RSM performs better than the three other models, clear evidence that proposed hybrid model improves the accuracy of the standard SVR model. The statistics for SVR, MLPNN, and RSM models varied generally from one climatic station to another, though one of the models has a lower standard deviation or has a higher coefficient of correlation. In each station, the differences between the three models (RSM, SVR, and MLPNN) are much smaller than the differences between the three and the SVR_RSM. Consequently, the improvement in SVR_RSM model throughout the use of the hybrid approach is further demonstrated in Fig. 9.

Conclusion

This study investigated the utility of using the response surface model (RSM) for calibrating the original support vector regression (SVR) in order to obtain a new hybrid model (SVR_RSM), applied for modeling monthly pan evaporation. Different scenarios based on several input combinations have demonstrated that the hybrid SVR_RSM model performs overwhelmingly better than the RSM, SVR, and the MLPNN models. One of the most important conclusions of the present study is that, the proposed hybrid model provides an effective and practical approach to estimate E_{pan} using only fewer input variables. For example, the SVR_RSM model with only the T_{max} and T_{min}, coupled with the periodicity exhibited high accuracy with R, NSE, and d values of 0.989, 0.970, and 0.992 during the validation phase. Superiority of the SVR_RSM for predicting E_{pan} was evident at the all six climatic stations. In addition, the MLPNN, RSM, and SVR

models provided relatively different results, and the accuracy varied from one station to another depending on the input variables. Furthermore, inclusions of the periodicity to the input variables contribute to an improvement of the models performances, except at Chaffia station for which the performances was decreased. The results of the study were encouraging, but need to be extended to other stations.

References

Allawi MF, El-Shafie A (2016) Utilizing RBF-NN and ANFIS methods for multi-lead ahead prediction model of evaporation from reservoir. Water Resour Manag 30(13):4773–4788. https://doi.org/10.1007/s11269-016-1452-1

Brereton RG, Lloyd GR (2010) Support vector machines for classification and regression. Analyst 135:230–267. https://doi.org/10.1039/B918972F

Cahoon JE, Costello TA, Ferguson JA (1991) Estimating pan evaporation using limited meteorological observations. Agric For Meteorol 55(3-4):181–190. https://doi.org/10.1016/0168-1923(91)90061-T

Dao VN, Vemuri VR (2002) A performance comparison of different back propagation neural networks methods in computer network intrusion detection. Diff Equ Dyn Sys 10(1&2):201–214

Deo RC, Samui P, Kim D (2016) Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. Stoch Env Res Risk A 30(6):1769–1784. https://doi.org/10.1007/s00477-015-1153-y

Eray O, Mert C, Kisi O (2018) Comparison of multi-gene genetic programming and dynamic evolving neural-fuzzy inference system in modeling pan evaporation. Hydrol Res. https://doi.org/10.2166/nh.2017.076

Feng Y, Jia Y, Zhang Q, Gong D, Cui N (2018) National-scale assessment of pan evaporation models across different climatic zones of China. J Hydrol 564:314–328. https://doi.org/10.1016/j.jhydrol.2018.07.013

Fun MH, Hagan MT (1996) Levenberg-Marquardt training for modular networks, Neural Networks. IEEE International Conference on. IEEE, pp. 468-473. https://doi.org/10.1109/ICNN.1996.548938.

Ghorbani MA, Deo RC, Yaseen ZM, Kashani MH, Mohammadi B (2018) Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. Theor Appl Climatol 133(3-4):1119–1131. https://doi.org/10.1007/s00704-017-2244-0

Guven A, Kisi O (2011) Daily pan evaporation modeling using linear genetic programming technique. Irrig Sci 29:135–145. https://doi.org/10.1007/s00271-010-0225-5

Heddam S, Keshtegar B, Kisi O (2019) Predicting total dissolved gas concentration on a daily scale using kriging interpolation method (KIM), response surface method (RSM) and artificial neural network (ANN): case study of Columbia River Basin Dams, USA. Nat Resour Res. https://doi.org/10.1007/s11053-019-09524-2

Keshtegar B, Heddam S (2018) Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. Neural Comput & Applic 30(10):2995–3006. https://doi.org/10.1007/s00521-017-2917-8

Keshtegar B, Kisi O (2016) A nonlinear modelling-based high-order response surface method for predicting monthly pan evaporations. Hydrol Earth Syst Sci Discuss. https://doi.org/10.5194/hess-2016-191

Keshtegar B, Kisi O (2017) Modified response-surface method: new approach for modeling pan evaporation. ASCE J Hydro Engin

22(10):04017045. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001541

Keshtegar B, Mert C, Kisi O (2018) Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree. Renewable and Sustainable Energy Reviews 81:330–341

Keshtegar B, Seghier MEAB (2018) Modified response surface method basis harmony search to predict the burst pressure of corroded pipelines. Eng Fail Anal 89:177–199. https://doi.org/10.1016/j.engfailanal.2018.02.016

Keshtegar B, Bagheri M, Yaseen ZM (2019a) Shear strength of steel fiber-unconfined reinforced concrete beam simulation: Application of novel intelligent model. Compos Struct 212:230–242. https://doi.org/10.1016/j.compstruct.2019.01.004

Keshtegar B, Heddam S, Kisi O, Zhu SP (2019b) Modelling total dissolved gas (TDG) concentration at Columbia River Basin dams: high-order response surface method (H-RSM) vs. M5Tree, LSSVM and MARS. Arab J Geosci 12:544. https://doi.org/10.1007/s12517-019-4687-3

Kim S, Kim HS (2008) Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. J Hydrol 351(3-4):299–317. https://doi.org/10.1016/j.jhydrol.2007.12.014

Kisi O (2013) Evolutionary neural networks for monthly pan evaporation modeling. J Hydrol 498:36–45. https://doi.org/10.1016/j.jhydrol.2013.06.011

Kisi O (2015) Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. J Hydrol 528:312–320. https://doi.org/10.1016/j.jhydrol.2015.06.052

Kisi O, Tombul M (2013) Modeling monthly pan evaporations using fuzzy genetic approach. J Hydrol 477:203–212. https://doi.org/10.1016/j.jhydrol.2012.11.030

Kowsar R, Keshtegar B, Miyamoto A (2019) Understanding the hidden relations between pro-and anti-inflammatory cytokine genes in bovine oviduct epithelium using a multilayer response surface method. Sci Rep 9(1):3189. https://doi.org/10.1038/s41598-019-39081-w

Kurt H, Kayfeci M (2009) Prediction of thermal conductivity of ethylene glycol-water solutions by using artificial neural networks. Appl Energy 86:2244–2248. https://doi.org/10.1016/j.apenergy.2008.12.020

Lu CJ (2014) Sales forecasting of computer products based on variable selection scheme and support vector regression. Neurocomputing 128:491–499. https://doi.org/10.1016/j.neucom.2013.08.012

Lu X, Ju Y, Wu L, Fan J, Zhang F, Li Z (2018) Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models. J Hydrol 566:668–684. https://doi.org/10.1016/j.jhydrol.2018.09.055

Malik A, Kumar A, Kisi O (2017) Monthly pan-evaporation estimation in Indian central Himalayas using different heuristic approaches and climate based models. Comput Electron Agric 143:302–313. https://doi.org/10.1016/j.compag.2017.11.008

McMahon TA, Peel MC, Lowe L, Srikanthan R, McVicar TR (2013) Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis. Hydrol Earth Syst Sci 17(4):1331–1363. https://doi.org/10.5194/hess-17-1331-2013

Pathirage CSN, Li J, Li L, Hao H, Liu W, Ni P (2018) Structural damage identification based on autoencoder neural networks and deep learning. Eng Struct 172:13–28. https://doi.org/10.1016/j.engstruct.2018.05.109

Qasem SN, Samadianfard S, Kheshtgar S, Jarhan S, Kisi O, Shamshirband S, Chau KW (2019) Modeling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates. Engin Applic Comput Fluid Mech 13(1):177–187. https://doi.org/10.1080/19942060.2018.1564702

Rezaie-Balf M, Kisi O, Chua LH (2018) Application of ensemble empirical mode decomposition based on machine learning methodologies in forecasting monthly pan evaporation. Hydrol Res. https://doi.org/10.2166/nh.2018.050

Sebbar A., Heddam S., Djemili L. (2019). Predicting daily pan evaporation (E_{pan}) from dams reservoirs in the Mediterranean regions of Algeria: OPELM vs OSELM. Environmental Process. https://doi.org/10.1007/s40710-019-00353-2.

Shiri J (2019) Evaluation of a neuro-fuzzy technique in estimating pan evaporation values in low-altitude locations. Meteorol Appl. https://doi.org/10.1002/met.1753

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). ASCE J Irrig Drain Eng 137:412–425. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000315

Shiri J, Dierickx W, Pour-Ali BA, Neamati S, Ghorbani MA (2011) Estimating daily pan evaporation from climatic data of the State of Illinois, USA using adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN). Hydrol Res 42(6):491–502. https://doi.org/10.2166/nh.2011.020

Thissen U, Pepers M, Üstün B, Melssen WJ, Buydens LMC (2004) Comparing support vector machines to PLS for spectral regression applications. Chemom Intell Lab Syst 73(2):169–179. https://doi.org/10.1016/j.chemolab.2004.01.002

Wang L, Niu Z, Kisi O, Li C, Yu D (2017) Pan evaporation modeling using four different heuristic approaches. Comput Electron Agric 140:203–213. https://doi.org/10.1016/j.compag.2017.05.036

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