atmosphere

MDPI

Article

Modeling Pan Evaporation Using Gaussian Process Regression K-Nearest Neighbors Random Forest and Support Vector Machines; Comparative Analysis

Sevda Shabani [1], Saeed Samadianfard [1], Mohammad Taghi Sattari [1,2], Amir Mosavi [3,4,5,6], **Shahaboddin Shamshirband [7,8,*], Tibor Kmet [9] and Annamária R. Várkonyi-Kóczy [3,9]**

[1] Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz 51666, Iran; sevdashabani7364@gmail.com (S.S.); s.samadian@tabrizu.ac.ir (S.S.); mtsattar@gmail.com (M.T.S.)
[2] Department of Farm Structures and Irrigation, Faculty of Agriculture, Ankara University, Ankara, Turkey
[3] Institute of Automation, Kalman Kando Faculty of Electrical Engineering, Obuda University, 1034 Budapest, Hungary; amir.mosavi@kvk.uni-obuda.hu (A.M.); varkonyi-koczy@uni-obuda.hu (A.R.V.-K.)
[4] Institute of Structural Mechanics, Bauhaus University Weimar, D-99423 Weimar, Germany; amir.mosavi@uni-weimar.de
[5] Faculty of Health, Queensland University of Technology, Queensland 4059, Australia; a.mosavi@qut.edu.au
[6] School of the Built Environment, Oxford Brookes University, Oxford OX3 0BP, UK; a.mosavi@brookes.ac.uk
[7] Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[8] Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[9] Department of Mathematics and Informatics, J. Selye University, 94501 Komarno, Slovakia; kmett@ujs.sk or koczya@ujs.sk
* Correspondence: shahaboddin.shamshirband@tdt.edu.vn

check for updates

Abstract: Evaporation is a very important process; it is one of the most critical factors in agricultural, hydrological, and meteorological studies. Due to the interactions of multiple climatic factors, evaporation is considered as a complex and nonlinear phenomenon to model. Thus, machine learning methods have gained popularity in this realm. In the present study, four machine learning methods of Gaussian Process Regression (GPR), K-Nearest Neighbors (KNN), Random Forest (RF) and Support Vector Regression (SVR) were used to predict the pan evaporation (PE). Meteorological data including PE, temperature (T), relative humidity (RH), wind speed (W), and sunny hours (S) collected from 2011 through 2017. The accuracy of the studied methods was determined using the statistical indices of Root Mean Squared Error (RMSE), correlation coefficient (R) and Mean Absolute Error (MAE). Furthermore, the Taylor charts utilized for evaluating the accuracy of the mentioned models. The results of this study showed that at Gonbad-e Kavus, Gorgan and Bandar Torkman stations, GPR with RMSE of 1.521 mm/day, 1.244 mm/day, and 1.254 mm/day, KNN with RMSE of 1.991 mm/day, 1.775 mm/day, and 1.577 mm/day, RF with RMSE of 1.614 mm/day, 1.337 mm/day, and 1.316 mm/day, and SVR with RMSE of 1.55 mm/day, 1.262 mm/day, and 1.275 mm/day had more appropriate performances in estimating PE values. It was found that GPR for Gonbad-e Kavus Station with input parameters of T, W and S and GPR for Gorgan and Bandar Torkmen stations with input parameters of T, RH, W and S had the most accurate predictions and were proposed for precise estimation of PE. The findings of the current study indicated that the PE values may be accurately estimated with few easily measured meteorological parameters.

Keywords: machine learning; meteorological parameters; pan evaporation; advanced statistical analysis; hydrological cycle; big data; hydroinformatics; random forest (RF); support vector regression (SVR)

1. Introduction

Pan evaporation (PE) is usually affected by the thermal energy and the vapor pressure gradient, which depends mainly on meteorological parameters [1]. The evaporation from the soils, lakes, and water reservoirs, is one of the most critical processes in meteorological and hydrological sciences [2–6]. The evaporation from the pan measures the combined weight of the temperature, relative humidity, wind speed, and sunny hours on the evaporation rate from the water surface [7] and this combination is effective on plant evapotranspiration. Hence, there is an excellent correlation between PE and plant evapotranspiration. Therefore, the plant evapotranspiration estimated by applying the coefficients on PE [8,9]. However, there is no possibility to install and maintain a meteorological device at any place, especially in impassable regions. The performance of the pan is affected by instrumental restrictions and operational issues such as human errors, instrumental errors, water turbidity, the animals, and birds interfering with the water in the pan and maintenance problems [10–12]. Therefore, the need for models for accurate estimation of the evaporation felt more than ever before. In other words, the straight measurements of PE are usually limited according to the problems of installing and maintaining measuring devices [13]. In many studies, researchers have been tried to estimate PE using indirect methods utilizing climatological and meteorological parameters, which have resulted empirical models. In the other hand, physical models requirement are various input data; therefore, the applicability of these established empirical and physical models are limited according to the low accessibility of the mentioned data, especially in developing countries. So, in the recent years, machine learning methods have been implemented and applied successfully for PE estimation according to their superior capabilities in learning nonlinear and complex interactions of this phenomenon, which are difficult for empirical models. In this regard, Keskin and Terzi [14] studied the meteorological data of the stations near the lake in western Turkey to determine the daily PE using the neural network model. The results revealed that the best structure of the model obtained with four input data including air temperature, water surface temperature, sunny hours, and air pressure; however, wind speed and relative humidity have a low correlation with the evaporation in the study area. Guven and Kisi [15] studied the ability of linear genetic programming (LGP) in modeling the PE. They compared the results of this method with the results of Radial Basis Function Neural Network, Generalized regression neural network (GRNN), and the Stefanz-Stewart's model. Comparing the results showed that the LGP was more accurate than the other mentioned methods. Traore and Guven [16] studied the ability of Gene Expression Programming (GEP) in Burkina Faso coastal areas to model the evapotranspiration, and they used the combined meteorological data as inputs to the GEP model. They concluded that the GEP model has an excellent ability based on regional data. Gundalia and Dholakia [17] assessed the performance of six empirical models in predicting the daily PE. They found that the Yansen's model, based on the sunny hours, was the most appropriate method for evaluating the daily evaporation at the studied stations in India. Kisi and Zounemat-Kermani [18] compared two methods of neuro-fuzzy and Gene Expression Programming (GEP) to model the daily reference evapotranspiration of the Adana station in Turkey. In this study, the wind speed identified as an effective parameter in modeling. Wang et al. [19,20] evaluated six machine learning algorithms as well as two empirical models for predicting monthly evaporation in different climate regions of China during the years 1961–2000. They found that the multi-layer perceptron model by using the regional input data was better at the studied stations. Malik et al. [21] used the Multi-Layer Perceptron Model, Adaptive Neuro-Fuzzy Inference System, and Radial Basis Function Neural Network to predict evapotranspiration in two regions of Nagar and Ranichari (India). The results showed that the models of Adaptive Neuro-Fuzzy Inference System and multi-layer perceptron neural networks with six meteorological input data were better than other models for estimating the monthly evaporation. Ghorbani et al. [22] implemented a hybrid Multilayer Perceptron-Firefly Algorithm (MLP-FFA) for predicting daily PE values. The obtained results indicated the superiority of developed MLP-FFA model comparing standalone MLP in studied stations. Tao et al. [23] utilized fuzzy model combined with firefly algorithm (ANFIS-FA) for reference evapotranspiration prediction in Burkina Faso region.

Their obtained results revealed high capabilities of implemented firefly algorithm in decreasing the prediction error of standalone ANFIS model in all studied stations. Khosravi et al. [24] examined the potential of five data mining and four ANFIS models for predicting reference evapotranspiration in two stations in Iraq. They stated that for both studied stations, the ANFIS-GA generated the most accurate predictions. Salih et al. [25] investigated the capabilities of co-ANFIS for predicting evaporation from reservoirs using meteorological parameters. The findings of the mentioned study indicated the suitable accuracy of the co-ANFIS model in evaporation estimation. Recently, Feng et al. [26] examined the performance of two solar radiation-based models for the estimation of daily evaporation in different regions of China. They suggested that Stewart's model can be preferred when the meteorological data of sunny hours and air temperature are available. Therefore, it is possible to estimate the evaporation through intrinsically nonlinear models. Qasem et al. [27] examined the applicability of wavelet support vector regression and wavelet artificial neural networks for predicting PE at Tabriz and Antalya stations. Obtained results indicated that artificial neural networks had better performances, and the wavelet transforms did not have significant effects in reducing the prediction errors at both studied stations. Yaseen et al. [28] predicted PE values using four machine learning models in two stations of Iraq. They reported that the SVM indicated the best performance comparing to other studied methods.

The literature review showed that the data mining methods had a suitable application in estimating PE in different climates, but to the best of our knowledge, the application of Gaussian process regression not reported for estimating PE. So, in the present study, the ability of four data mining methods, including Gaussian process regression, support vector regression, Nearest-Neighbor, and Random Forest, are studied in estimating PE rates, using different combinations of meteorological parameters. In the following, the results compared. Finally, using performance evaluation indices, the best method is obtained for estimating evaporation in the humid regions of Iran.

2. Study Area

Golestan province is located in the southeastern part of the Caspian Sea with an area of 20,387 km^2 and covers about 1.3% of the total area of Iran (Figure 1). This province has an average annual rainfall of 450 mm in the geographical range of 36°25′ to 38°8′ north latitude and 53°50′ to 56°18′ east longitude. The geographical location and topography of Golestan province influenced by various climatic factors and different climates observed in this province. So that, the semi-arid climate is observed in the international border and the Atrak basin, moderate and semi-humid in the southern and western parts of the province, as well as cold climate in the mountainous regions.
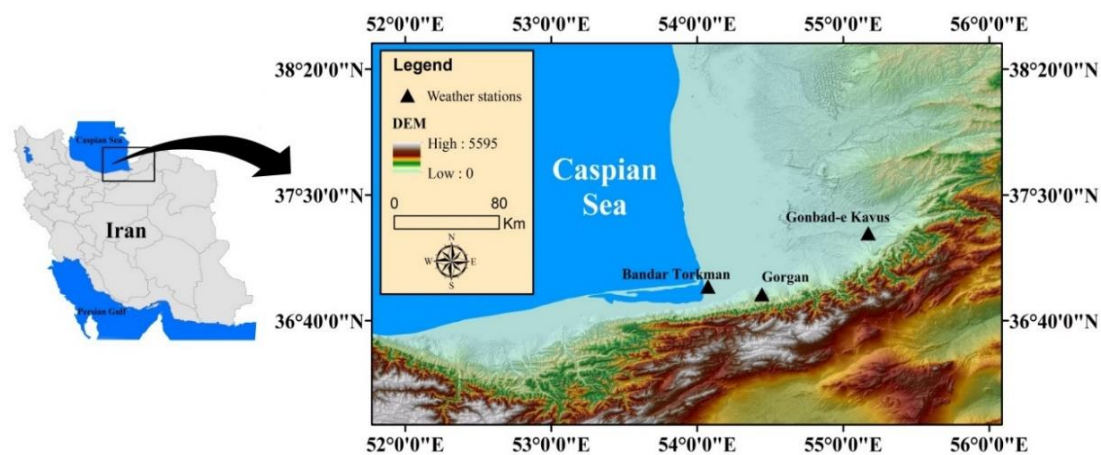
Figure 1. The geographical location of the study area.

Meteorological parameters implemented at the current research are temperature (T), relative humidity (RH), wind speed (W), and sunny hours (S) and PE with the period of 2011 to 2017.

Table 1. Statistical characteristics of the utilized data in three studied stations.

Station	Variable	Mean	Minimum	Maximum	Standard Deviation	Coefficient of Variation	Skewness	Correlation with PE
Gonbad-e Kavus	T (°C)	19.2	−6.8	36.8	8.86	0.46	−0.07	0.85
	RH (%)	66.0	21.5	98.0	14.16	0.21	0.03	−0.63
	W (m/s)	1.7	0.0	7.7	1.07	0.64	0.72	0.24
	S (hr)	6.9	0.0	13.6	4.19	0.61	−0.45	0.46
	PE (mm/day)	3.8	0.0	13.6	3.06	0.81	0.64	1.00

Generally, the class A pan was utilized for the PE measurement due to its international referenceability and acceptability.

Table 1 represents the statistical parameters of all utilized variables in three studied stations, including Gonbad-e Kavus, Gorgan and Bandar Torkman.

Table 1. Statistical characteristics of the utilized data in three studied stations.

Station	Variable	Mean	Minimum	Maximum	Standard Deviation	Coefficient of Variation	Skewness	Correlation with PE
Gonbad-e Kavus	T (°C)	19.2	−6.8	36.8	8.86	0.46	−0.07	0.85
	RH (%)	66.0	21.5	98.0	14.16	0.21	0.03	−0.63
	W (m/s)	1.7	0.0	7.7	1.07	0.64	0.72	0.24
	S (hr)	6.9	0.0	13.6	4.19	0.61	−0.45	0.46
	PE (mm/day)	3.8	0.0	13.6	3.06	0.81	0.64	1.00
Gorgan	T (°C)	18.4	−4.7	35	8.42	0.46	−0.08	0.87
	RH (%)	70.2	20.5	98	12.22	0.17	−0.06	−0.65
	W (m/s)	2.0	0.0	10	1.44	0.71	0.74	0.28
	S (hr)	6.4	0.0	13.1	4.19	0.65	−0.27	0.50
	PE (mm/day)	3.7	0.0	12.8	2.84	0.78	0.59	1.00
Bandar Torkman	T (°C)	18.4	−4.3	34.5	8.08	0.44	−0.11	0.88
	RH (%)	73.6	37.5	98.0	9.73	0.13	−0.12	−0.52
	W (m/s)	3.3	0.0	17.7	1.96	0.6	1.27	0.44
	S (hr)	6.5	0.0	13.3	4.12	0.63	−0.33	0.45
	PE (mm/day)	4.4	0.0	16.0	3.09	0.71	0.52	1.00

As it can be seen from Table 1, W has the highest skewness among implemented parameters. Also, S and PE specify skewed distributions. Nevertheless, T and RH having lower skewness values indicate normal distributions. Additionally, in all studied stations, T and S variables have higher values of correlation with PE. Furthermore, RH has an inverse correlation with PE.

3. Materials and Methods

3.1. Gaussian Process Regression (GPR)

Gaussian processes (GP) defined as a set of random variables, in which a few variables have a multi-variable Gaussian distribution. X and Y are the input and output domains, respectively; n pairs (x_i, y_i) domains are independent and extracted and distributed equally. It assumed that the Gaussian process on X is defined by the average function of μ: Y → Re and the covariance function of k: X * X → Re. The primary assumption of the GPR is that y is determined by y = f (x) + ζ, in which, ζ is Gaussian noise with variance of σ^2. In the Gaussian process regression, there is a random variable f(x) for each input variable x, that is the value of the random function f in that location. In this study, it was assumed that the observation error is independent and has the same distribution as the zero mean value (μ(x) = 0) and the variance (σ^2), and f(x) of the Gaussian process on X (denote with k) (Equation (1)):

$$Y = (Y_1, \ldots, Y_n) \sim N(o, K = \sigma^2 I) \tag{1}$$

where I is the identity matrix and $K_{ij} = k(x_i, x_j)$. Because Y/X ~ N (o, K + σ^2I) is normal, the conditional distribution of the test labels with the condition of the training and testing data $p(Y^*/Y, X, X^*)$; in this condition (Y*/Y, X, X*) ~ N (μ, σ), therefore:

$$\mu = K(X_\diamond, X)(K(X, X) + \sigma^2 I)^{-1} Y \tag{2}$$

$$\sigma = K(X_\diamond, X_\diamond) - \sigma^2 I - K(X_\diamond, X)(K(X, X) + \sigma^2 I)^{-1} K(X, X_\diamond) \tag{3}$$

K(x, x′) is the matrix n × n* of the evaluated covariance in all pairs of test and training data sets that are similar for other values of K(X, X), K(X, X*), and K(X*, X*). Where X and Y are the training vector and training data labels y_i, while X* is the test data. The covariance function specified for creating a

semi-finite positive covariance matrix of k, in which $K_{ij} = k(x_i, x_j)$. By specifying the kernel k and the noise degree σ^2, Equations (3) and (4) are sufficient for a deduction. The selection of the appropriate covariance function and its parameters is essential during the training process of GPR models because the central role in the Gaussian process regression model belongs to the covariance function K(x, x′). This function embeds the geometric structure of the training samples. In other words, for generating precise predictions, the mean and covariance functions should be estimated from utilized data, which are called hyperparameters. [29].

3.2. K-Nearest-Neighbor-IBK

Here the K-Nearest Neighbors or KNN is implemented through the instance-bases learning with parameter k (IBK). In general, this algorithm used for two purposes: 1. estimation of the density function of the distribution of test data and 2. classification of the test data based on the test patterns. The first step in applying this algorithm is to find a method and a relationship to calculate the distance between the test and training data. Euclidean distance is usually used to determine the distance between test and training data

$$d(X, Y) \;=\; \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{4}$$

where X represents the training data with specified parameters (x_n to x_1), and Y represents the training data with the same number of parameters (y_n to y_1).

$$X \;=\; (x_1, x_2, \ldots, x_n) \tag{5}$$

$$Y \;=\; (y_1, y_2, \ldots, y_n) \tag{6}$$

After determining the Euclidean distance between the data, the database samples sorted in ascending order from the least distance (maximal similarity) to the maximum distance (minimum similarity). The next step in this model is to find the number of points (k) of the experiment to estimate the characteristics of the desired database. Determining the number of neighbors (k) is one of the most critical steps, and the efficiency of this method is depended to the selection of the closest (the most similar) samples from the reference database considerably. If k is assumed to be small, the results are sensitive to the unconventional single points of the model, and if k is assumed to be significant, it is possible to place some points from other classes within the desired range. Usually, the best value for k is calculated using cross-validation [30].

3.3. Random Forest (RF)

Random forest (RF) utilizes classification and regression tree (CART) as a learning algorithms of decision trees. The RF is a set of decision trees, which in each one, the space of the variables is divided into smaller subspace so that the data in each region has as uniform as possible. This classification pattern is employed by a structure decision tree. In this tree structure, the branching point to two sub-branches is called node. The first node of the tree is called the root, and the last one is the leaf [31]. In RF, each tree grows with a self-serving sample of the original data, and in order to perform the best division, the number of m variables selected randomly by variables is searched [32]. In the RF method, the data discrepancy are determined in a completely different way from the usual distance functions. The similarity of data in RF is measured based on their placement in the same leaves (the final subspaces). In the RF, the similarity between i and j (s(i, j)) defined as the ratio of the times that the two given data are in the same leaf. The forest similarity matrix is random, symmetric, and positive. This matrix converts the following transformation into a non-similar matrix:

$$d(i, j) \;=\; \sqrt{1} - s(i, j) \tag{7}$$

The formation of the classification tree is not dependent on the values of the variables; hence, the lack of similarity of the random forest applied to the types of variables [33]. The splitting procedure is repeated in each tree until reaching a predefined stop condition.

3.4. Support Vector Regression (SVR)

Support vector machine is one of the learning methods introduced by Bousser et al. [34] on the basis of statistical learning theory. In the following years, they introduced the theory of optimum hyperplane as linear classifiers and introduced nonlinear classifiers by kernel functions [35]. The models of support vector machines are divided into two main groups: (a) the classifier models of support vector machine and (b) a support vector regression model. A support vector machine model is used to solve the classification problems of data in different classes, and the SVR model is used to solve the prediction problems. Regression is meant to obtain a hyperplane that is fitted to the given data. The distance from any point on this hyperplane indicates the error of that particular point. The best method suggested for linear regression is the least-squares method. However, for regression issues, the use of the least-squares estimator in the presence of outlier data may be completely impossible; as a result, the processor will show poor performance. Therefore, a robust estimator that would not be sensitive to small variations in the model should be developed. In fact, a penalty function (ε) is defined as follows [36]:

$$L^{\varepsilon}(x, y, f) \begin{cases} 0 & \text{if } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases} \tag{8}$$

The training data sets are S = {(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)} and the class of the function is as $f(x) = \{w^T x + b, w \in R, b \in R\}$. If the data deviate from the value of ε, a deficiency variable must be defined according to the value of the deviation. In accordance with the penalty function, the minimization is defined as follows:

$$\text{Minimize } \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \text{ Subject to} \begin{cases} f(x) - y_i \leq \varepsilon + \xi_i \\ f(x) - y_i \leq \varepsilon + \xi_i^* & i = 1, 2, 3, \ldots, n \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{9}$$

where $\|w\|^2$ is the norm of the weight vector, and ξ_i and ξ_i^* are auxiliary deficiency variables, and parameter C is the coefficient of equilibrium of complexity between the machine and the number of inseparable points obtained by trial and error.

3.5. Model Development

For developing the studied models of GPR, IBK, RF, SVR different model characteristics were examined using trial and error procedure. For GPR model, radial basis kernel function was implemented with kernel length scale of 3.0 and kernel bias of 1.0. Also, the maximum basis vectors were set to be 100. In the IBK algorithm, k is typically a small, positive and odd integer. So, k was set to be 5. For RF, random forest number of trees was 100, random forest maximum depth was 10 and random forest subset ratio was set to be 0.2. In the SVR, radial basis kernel function was implemented with kernel parameter gamma of 1.0. Moreover, maximum iterations and convergence epsilon were 100,000 and 0.001, respectively.

3.6. Evaluation Parameters

Error-values between computed and observed data are evaluated by Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and correlation coefficient (R) defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right)^2} \quad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - y_i\right| \tag{10}$$

$$R = \frac{\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)}{\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)\left(\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right)} \tag{11}$$

where, x_i and y_i are observed and estimated PE and n is the number of observations. Additionally, Taylor diagrams were employed for inspecting the accuracy of the implemented models. It is outstanding that in the mentioned diagram, measured and some correspondent statistical parameters are presented at the same time. Moreover, different points on a polar plot are used in Taylor diagrams for investigating the differences between observed and estimated values. Also, the CC and normalized standard deviation are indicated by azimuth angle and radial distances from the base point, respectively [37].

4. Results and Discussion

Before implementing the studied GPR, IBK, RF, SVR models in PE estimation, the preliminary statistical analysis was performed for three considered stations and the obtained results are presented in Table 2.

Table 2. Results of homogeneity, trend and outlier tests in three studied stations.

Test Type	Test Name	*p*-Value (Bandar Torkman)	*p*-Value (Gorgan)	*p*-Value (Gonbad-e Kavus)
Homogeneity Test	Pettitt's test	<0.0001	<0.0001	<0.0001
	Buishand's test	<0.0001	<0.0001	<0.0001
	SNHT	<0.0001	<0.0001	<0.0001
	VNR	<0.0001	<0.0001	<0.0001
Trend Test	Mann-Kendall	<0.0001	<0.0001	0.599
Outlier Test	Grubbs test	0.411	0.60	<0.0001
	Dixon test	0.005	0.599	0.791

According to Table 2, the results of trend and data outlier tests confirmed the hypothesis H_o, meaning that our data is free of trend and outlier. Then, in order to evaluate the possibility of using different combinations of meteorological data, seven different scenarios including various meteorological data, were defined for a more accurate estimation of PE (Table 3). It should be noted that these input combinations were established according to the correlation coefficients of meteorological parameters with PE values. So, T with highest correlation coefficient was considered as a necessary parameter for PE estimation. Therefore, T is present in all combinations. Also, different input combinations were examined by presence of T parameter. Then, these input combinations were used in data mining methods to estimate evaporation at three stations of Gonbad-e Kavus, Gorgan and Bandar Torkaman. There is no straightforward guideline for splitting the training and testing data in machine learning modeling [38–46]. For instance, the study of Choubin [47] used a total of 63% of their data for model development, whereas Qasem et al., [48] utilized 67% of data, Asadi et al., [41], Samadianfard et al., [49,50], and Dodangeh et al., [51] used 70%, and Zounemat-Kermani et al., [52] implemented 80% of total data to develop their models. Thus, to develop the studied models for PE estimation, we divided the data into training (70%) and testing (30%). So, data in the time period of 2011–2015 was utilized for training and the residual data from 2016–2017 was used for testing the implemented models. The general results of the computations for the defined scenarios for the above-mentioned methods are presented in Table 4.

Table 3. Parameters involved in defined scenarios of GPR, IBK, RF and SVR models.

Number	Input Parameters
1	T and RH
2	T and W
3	T and S
4	T, RH and W
5	T, RH and S
6	T, W, and S
7	T, RH, W and S

Table 4. Statistical results of the computations for the defined scenarios for GPR, IBK, RF, SVR models.

Model	Gonbad-e Kavus			Gorgan			Bandar Torkman		
	R	MAE (mm/day)	RMSE (mm/day)	R	MAE (mm/day)	RMSE (mm/day)	R	MAE (mm/day)	RMSE (mm/day)
GPR1	0.898	1.173	1.575	0.890	1.014	1.307	0.894	1.023	1.372
GPR2	0.898	1.170	1.560	0.884	1.026	1.337	0.905	0.973	1.299
GPR3	0.894	1.153	1.550	0.890	1.001	1.313	0.900	1.003	1.346
GPR4	0.903	1.148	1.545	0.897	0.980	1.265	0.907	0.972	1.294
GPR5	0.900	1.161	1.561	0.894	0.993	1.289	0.898	1.002	1.344
GPR6	0.899	1.128	1.521	0.897	0.965	1.265	0.912	0.939	1.257
GPR7	0.904	1.134	1.530	0.901	0.958	1.244	0.912	0.946	1.254
IBK1	0.795	1.547	2.069	0.784	1.434	1.895	0.820	1.375	1.840
IBK2	0.784	1.593	2.106	0.772	1.40	1.898	0.823	1.340	1.817
IBK3	0.776	1.585	2.154	0.788	1.393	1.865	0.818	1.391	1.876
IBK4	0.810	1.513	1.991	0.798	1.4	1.827	0.833	1.285	1.737
IBK5	0.789	1.543	2.100	0.804	1.343	1.824	0.835	1.291	1.763
IBK6	0.809	1.507	1.994	0.808	1.340	1.775	0.844	1.289	1.745
IBK7	0.804	1.521	2.028	0.799	1.361	1.841	0.865	1.179	1.577
RF1	0.859	1.322	1.755	0.856	1.149	1.492	0.865	1.139	1.546
RF2	0.844	1.379	1.814	0.832	1.186	1.590	0.876	1.092	1.484
RF3	0.865	1.268	1.703	0.851	1.155	1.522	0.877	1.128	1.502
RF4	0.880	1.239	1.647	0.875	1.059	1.387	0.892	1.023	1.386
RF5	0.877	1.241	1.673	0.870	1.082	1.423	0.887	1.045	1.419
RF6	0.879	1.225	1.621	0.879	1.030	1.374	0.900	1.007	1.349
RF7	0.886	1.199	1.614	0.885	1.011	1.337	0.903	0.980	1.316
SVR1	0.895	1.207	1.629	0.888	1.006	1.317	0.891	1.018	1.389
SVR2	0.896	1.184	1.585	0.883	1.017	1.340	0.904	0.971	1.314
SVR3	0.892	1.154	1.574	0.889	0.984	1.315	0.9	1.003	1.358
SVR4	0.901	1.178	1.590	0.893	0.982	1.284	0.906	0.961	1.298
SVR5	0.894	1.186	1.619	0.894	0.974	1.287	0.895	1.009	1.368
SVR6	0.895	1.129	1.550	0.895	0.962	1.278	0.911	0.943	1.275
SVR7	0.898	1.146	1.572	0.898	0.958	1.262	0.909	0.960	1.278

The presented results in Table 4 showed that for the GPR at Gonbad-e Kavus station, GPR6 with R = 0.899, MAE = 1.128 mm/day, and RMSE = 1.521 mm/day has less error than the other GPR combinations. However, the GPR7 model with R = 0.904, MAE = 1.134 mm/day and the RMSE = 1.530 mm/day requires more meteorological parameters. After GPR6, it presented the most accurate estimations of PE. On the other hand, GPR3 using two parameters of T and S, with R = 0.894, MAE = 1.153 mm/day, and RMSE = 1.550 mm/day have higher accuracy than the other SVR models. Also, due to fewer implemented parameters, GPR3 can be used in the case of data deficiency with an acceptable error and high reliability. Based on the results obtained at Gorgan Station, GPR7 with meteorological data of T, RH, W, and S has the lowest error with RMSE = 1.244 mm/day, MAE = 0.958 mm/day, and R = 0.901 and selected as the most accurate method among the GPR models. After the GPR7, GPR6 with RMSE = 1.265 mm/day, MAE = 0.965 mm/day, and R = 0.897 and GPR4 with RMSE = 1.265 mm/day, MAE = 0.980 mm/day and R = 0.897 with higher error than the GPR7 is in the second order. Based on the results obtained in the Bandar Torkaman station, GPR7 with RMSE = 1.254 mm/day, MAE = 0.946 mm/day, and R = 0.912 is selected as superior GPR model. In the next rank, the GPR6 with RMSE = 1.257 mm/day, MAE = 0.939 mm/day, and R = 0.912 presented precise estimations.

According to the results obtained at Gonbad-e Kavus station in the nearest-neighbor method, the IBK-4 with R = 0.810, MAE = 1.513 mm/day, and RMSE = 1.991 mm/day showed better performance than the other models. This scenario has a higher MAE than the IBK-6, however, because of the low RMSE and also the higher R-values, it can be described as the best nearest-neighbor model in evaporation estimation at Gonbad-e Kavus station. Also, IBK-6 with R = 0.809, MAE = 1.507 mm/day, and RMSE = 1.994 mm/day had an acceptable performance. Besides, at Gorgan Station, IBK6 with RMSE = 1.775 mm/day, MAE = 1.34 mm/day, and R = 0.808, by having the lowest error rate, was selected as the superior IBK model. Furthermore, IBK5 and IBK6 can be used as the second and third ranks, respectively. According to the results obtained in Bandar Torkaman, IBK-7 with RMSE = 1.577 mm/day, MAE = 1.17 mm/day, and R = 0.885 shows the best result among the IBK models. In the next rank, the IBK-4 with RMSE = 1.737 mm/day, MAE = 1.285 mm/day, and R =0.833 presented the acceptable estimations.

According to the results of the Gonbad-e Kavus station on the RF method, the RF7 with the lowest RMSE = 1.614 mm/day, the lowest MAE = 1.999 mm/day, and the highest R = 0.886 showed the best results among the RF models. In the next rank, the RF6 with RMSE = 1.621 mm/day, MAE = 1.225 mm/day, and R = 0.879 presented relatively precise estimations. Based on the results obtained at Gorgan, the RF7 with the least error (R = 0.885, MAE = 1.011 mm/day, and RMSE = 1.337 mm/day) had the best performance as compared to the other RF models. Also, in Bandar Torkaman station, RF7 with RMSE = 1.316 mm/day, MAE = 0.980 mm/day, and R = 0.903 introduced as the superior RF model due to the lowest error rate; however, as mentioned before, the RF7 model requires many meteorological data to develop an accurate relationship between PE and meteorological data. Similar to the GPR method, the RF6 with RMSE = 1.349 mm/day, MAE = 1.007 mm/day, and R = 0.90 presented the best results after RF7 model. Although the RF6 model has slightly higher error than RF7, it described as the optimum RF model due to the use of the low meteorological data.

According to the computations, among the kernel functions in all SVR models, the Pearson function provided the best results. So, for Gonbad-e Kavus station, SVR6 with the least error (R = 0.895, MAE = 1.129 mm/day, and RMSE = 1.55 mm/day) has the best performance in comparison with other SVR models. After that, SVR7 has the least error, but it is not recommended due to the use of more meteorological parameters than the other models. On the other hand, SVR3, with the use of two parameters of T and S, is more precise than the other SVR models (with R = 0.892, MAE = 1.154 mm/day, and RMSE = 1.574 mm/day). Moreover, at Gorgan Station, SVR7 with RMSE = 1.262 mm/day, MAE = 0.958 mm/day and R = 0.898 has the lowest error rate and considered as the superior SVR model; however, as previously mentioned, the SVR7 requires more meteorological parameters to make accurate estimations of PE. Based on the results obtained in the Bandar Torkaman station, SVR6 with the lowest RMSE = 1.275 mm/day and MAE = 0.943 mm/day and the highest R = 0.911 yield the best result among the SVR models. On the other hand, the SVR2 using only two parameters of T and W can be used with acceptable reliability and error in case of data deficiency. These uncertainties in the obtained results were maybe due to data division, input variability and model parameter optimization.

Figure 2 compares the variations of the predicted evaporation for the superior models (GPR, IBK, RF, and SVR) with the observed evaporation at one year of the testing period. Also, the distribution patterns of the methods mentioned above shown in Figure 3.
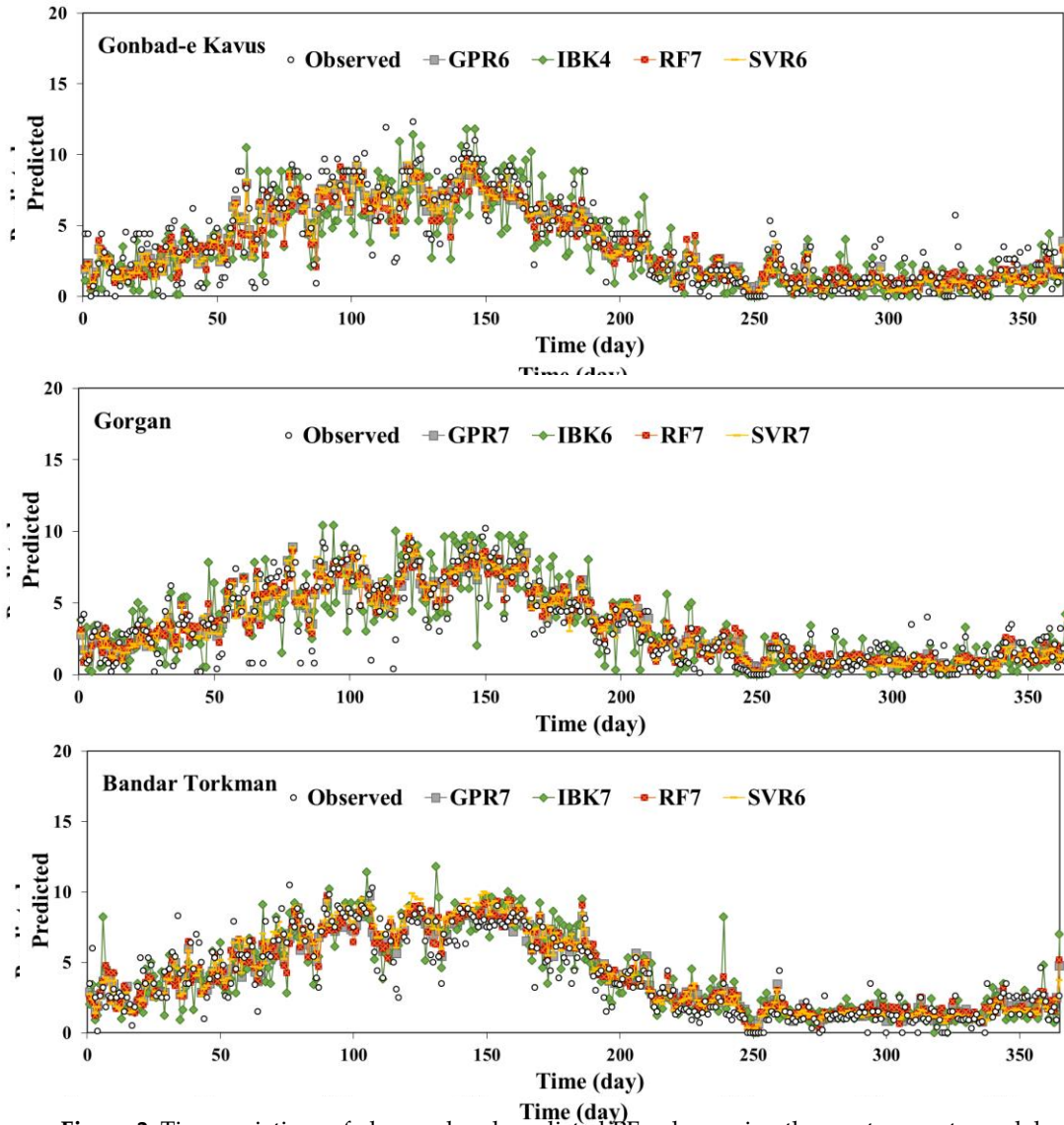
Figure 2. Time variations of observed and predicted PE values using the most accurate models.
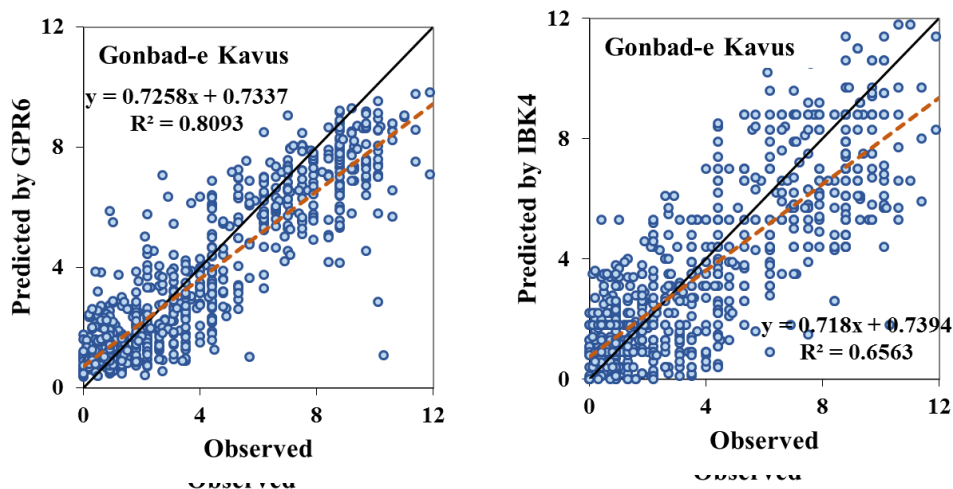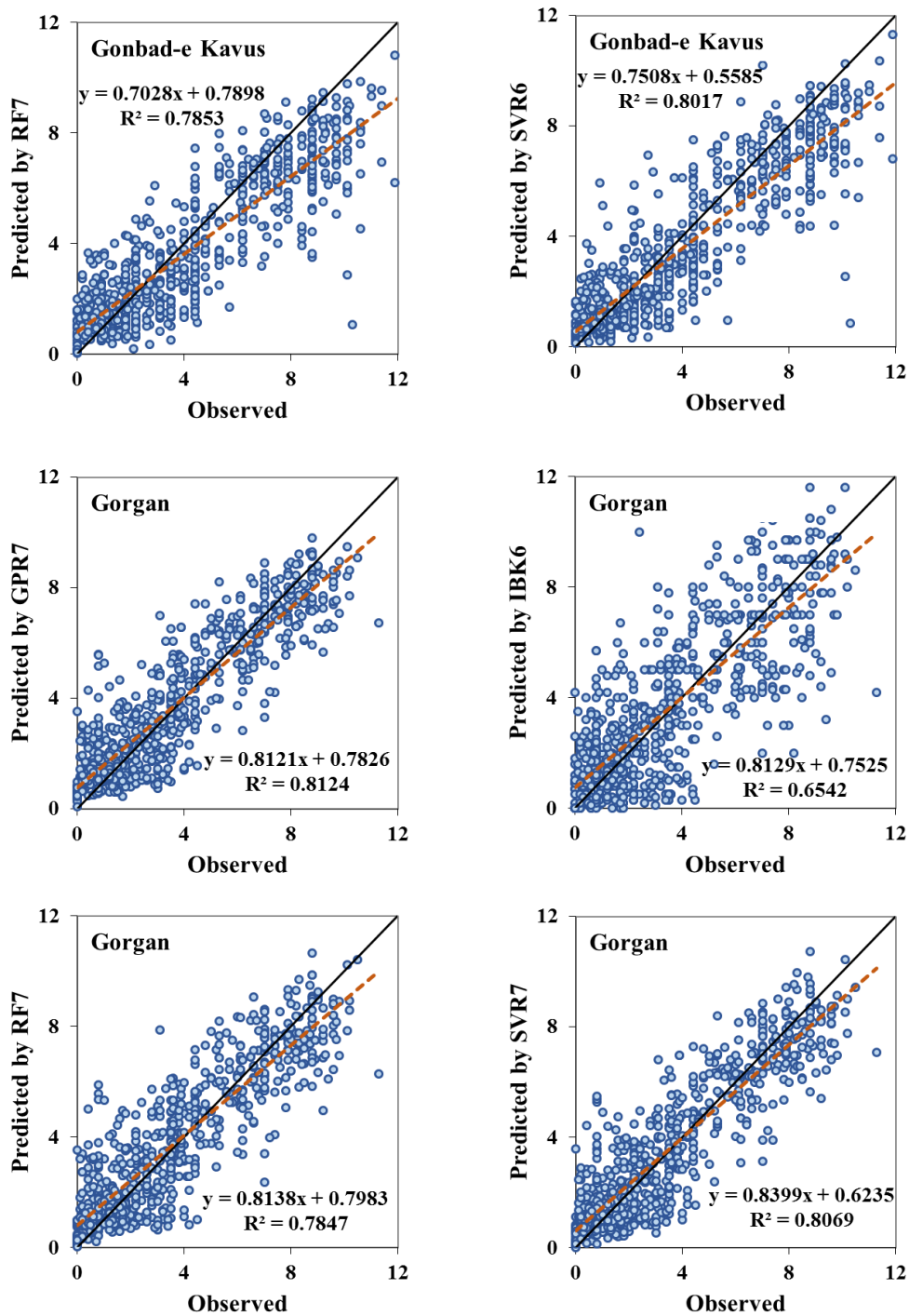
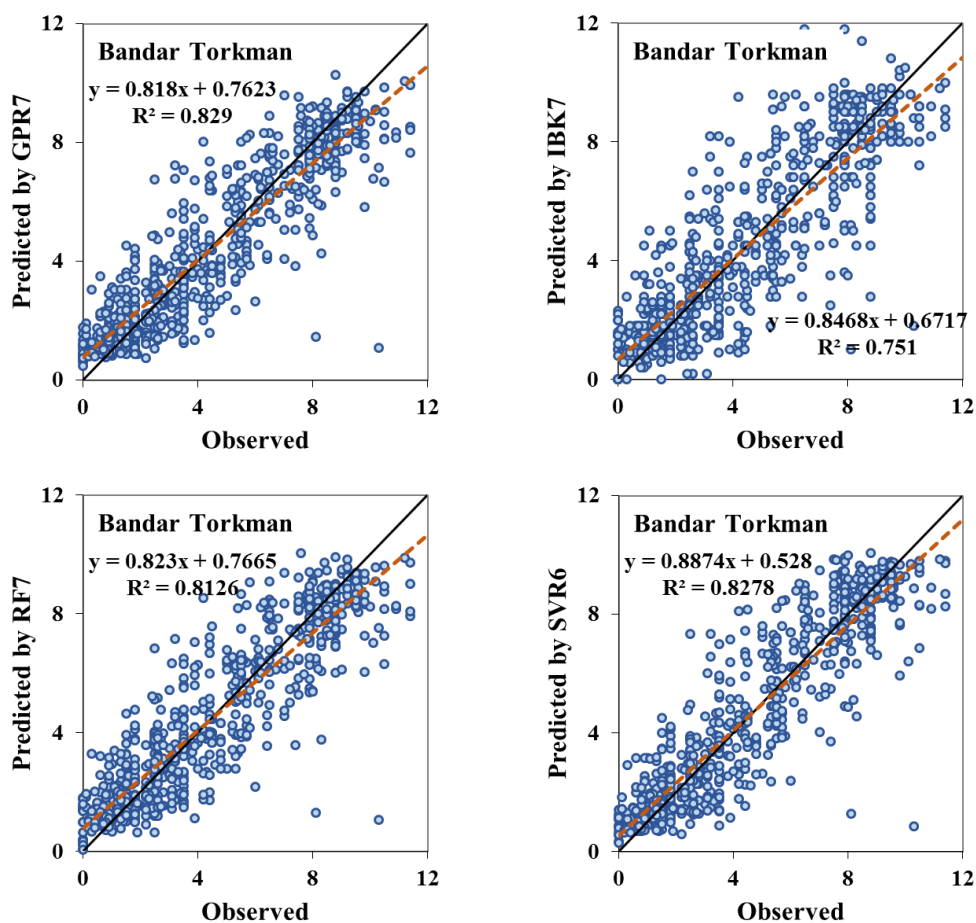**Figure 3.** *Cont*.

Figure 3. *Cont.*

Figure 3. Scatterplots of observed and predicted PE values using the most accurate models.

It can be comprehended from Figure 2 that the estimations of GPR6 at Gonbad-e Kavus station and GPR7 at both Gorgan and Bandar Torkman stations are in better agreement with observed PE. Similarly, it indicates from Figure 3 that the estimates of GPR6 and GPR7 are less scattered through the bisection line, and they preferred in correspondent stations. Moreover, an individual assessment of observed and estimated PE values by the best GPR, IBK, RF, and SVR models accomplished for each studied stations (Figure 4). Taylor diagrams, presented in Figure 4, are practical tools for better understanding the different potential of studied models. In the Taylor diagram, the most accurate model is explained by the point with the lower RMSE and higher R values. So, the Taylor diagrams proved that GPR6 in Gonbad-e Kavus and GPR7 in both Gorgan and Bandar Torkman stations indicated the best performances and presented the most accurate predictions of PE.
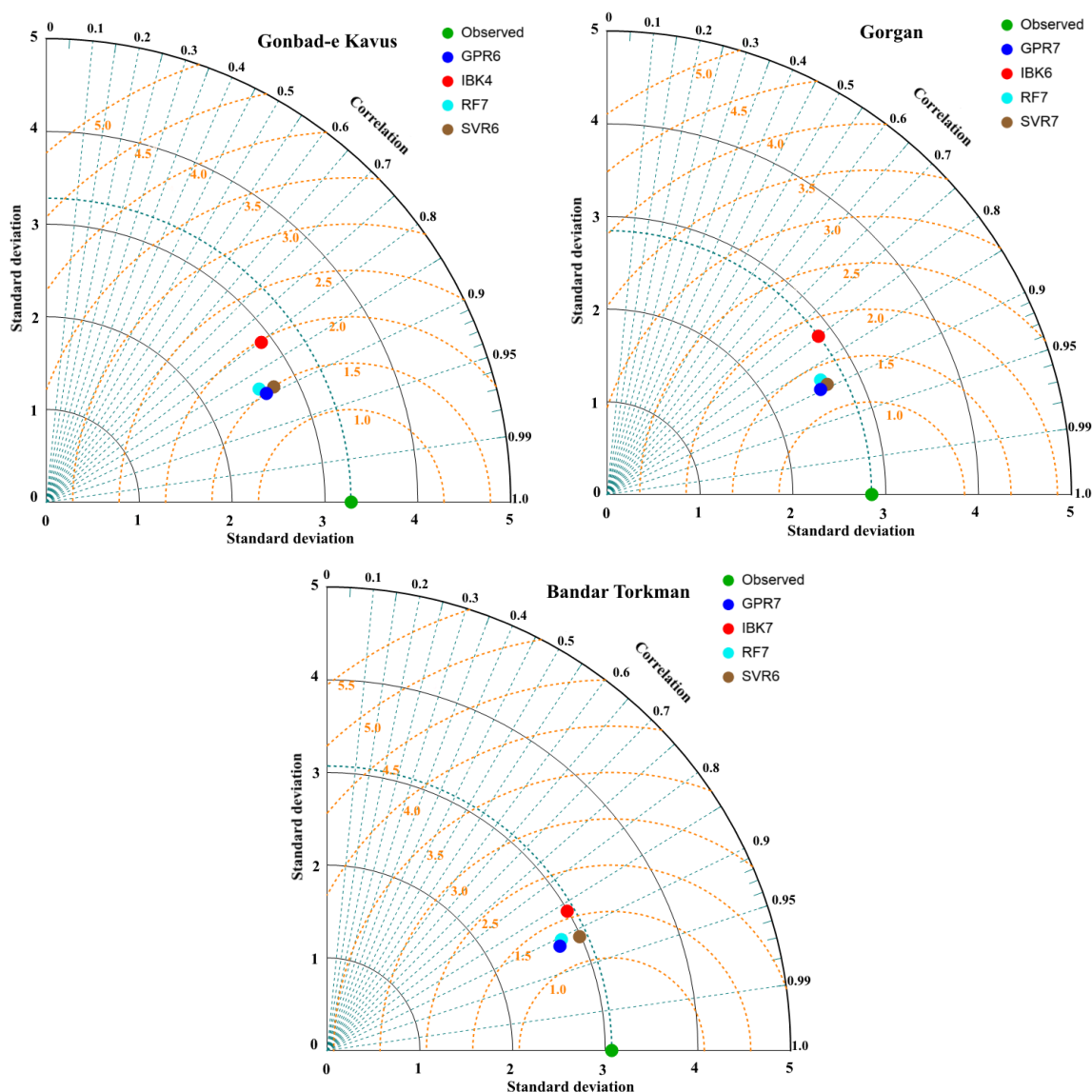
Figure 4. Taylor diagrams of estimated PE values using the most accurate models.

With a general look at the results and considering the above interpretations, it concludes that the meteorological parameters of T, W, and S have the most crucial role in increasing the accuracy of the PE estimation. Finally, for Gonbad-e Kavus station, GPR6 with input parameters of T, W and S, and for Gorgan and Bandar Torkman stations, GPR7 with input parameters of T, RH, W, and S have the best performance, and they considered as the most accurate models for estimating PE. It should be noted that the obtained results are based on the meteorological parameters of the studied stations and at a certain period, and the results changed in different climate zones. Comparing the obtained results with findings of Ghorbani et al. [22] showed that the accuracy of the GPR7 at Gorgan station with RMSE of 1.244 mm/day, as the most precise model, was better than developed MLP-FFA model by Ghorbani et al. [22] at Manjil station and lower than the mentioned MLP-FFA model at Talesh station, located at north of Iran.

As a concluding remark, the obtained results indicated that the accuracy of machine learning methods were satisfactory while using meteorological parameters of T, W and S. Furthermore, although the accuracies of different machine learning methods vary in three studied stations, but the performance of GPR and SVR were better than other examined models. Moreover, it should be noted that the accuracy and performance of these machine learning methods are not constant in different climates and regions. So, for increasing the applicability of machine learning methods in PE

regions. So, for increasing the applicability of machine learning methods in PE estimation, developing a general model for a homogeneous climatic region is recommended for future researches.

5. Conclusions

Evaporation is of particular importance in agriculture, hydrology, water and soil conservation studies. In this study, GPR, IBK, RF, and SVR machine learning methods were used to simulate daily PE in three stations of Gonbad-e Kavus, Gorgan, and Bandar Torkaman, located in Golestan province (Iran). The results of this study showed that in Gonbad-e Kavus station, the GPR6 and in the Gorgan and Bandar Torkaman stations, GPR7 have the lowest estimation errors and showed higher accuracy than other studied models. In other words, GPR can estimate PE with high accuracy using the meteorological parameters of (T, W, and S) in Gonbad-e Kavus station and the meteorological parameters of T, RH, W and S in Gorgan and Bander Turkmen stations. As a conclusive, overall result proved the superiority of the GPR method in PE estimation. The GPR recommended for PE estimation with a high degree of reliability.

The outcome indicates that the optimum state of Gonbad-e Kavus, Gorgan and Bandar Torkman stations, GPR with the error values of 1.521, 1.244, and 1.254, the KNN with error values of 1.991, 1.775, and 1.577, RF with error values of 1.614, 1.337, and 1.316, and SVR with error values of 1.55, 1.262, and 1.275, respectively, have more appropriate performances in estimating PE. It found that GPR for Gonbad-e Kavus Station with input parameters of T, W and S and GPR for Gorgan and Bandar Torkmen stations with input parameters of T, RH, W, and S had the most accurate performances and proposed for precise estimation of PE. So, the findings of the current study indicated that the PE values might be estimated using few easily measured meteorological parameters and with suitable accuracy in similar climates. Thus, the obtained conclusive remarks should be verified in different climates for evaluating the accuracy of the considered models.

References

1. Yang, H.B.; Yang, D.W. Climatic factors influencing changing pan evaporation across China from 1961 to 2001. *J. Hydrol.* **2012**, *414*, 184–193. [CrossRef]
2. Fan, J.; Chen, B.; Wu, L.; Zhang, F.; Lu, X.; Xiang, Y. Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. *Energy* **2018**, *144*, 903–914. [CrossRef]
3. Fan, J.; Wang, X.; Wu, L.; Zhang, F.; Bai, H.; Lu, X.; Xiang, Y. New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: A case study in South China. *Energy Convers. Manag.* **2018**, *156*, 618–625. [CrossRef]
4. Wang, R.-Y.; Yang, X.-G.; Zhang, J.-L.; Wang, D.-M.; Liang, D.-S.; Zhang, L.-G. A Study of Soil Water and Land Surface Evaporation and Climate on Loess Plateau in the Eastern Gansu Province. *Adv. Earth Sci.* **2007**, *22*, 625.

5.   Monteith, J.L. Evaporation and environment. *Symp. Soc. Exp. Biol.* **1965**, *19*, 205–234. [PubMed]

6.   Jones, F.E. *Evaporation of Water with Emphasis on Applications and Measurements*; CRC Press: Boca Raton, FL, USA, 2018.

7.   Roderick, M.L.; Rotstayn, L.D.; Farquhar, G.D.; Hobbins, M.T. On the attribution of changing pan evaporation. *Geophys. Res. Lett.* **2007**, *34*, L17403. [CrossRef]

8.   Pereira, L.S.; Allen, R.G.; Smith, M.; Raes, D. Crop evapotranspiration estimation with FAO56: Past and future. *Agric. Water Manag.* **2015**, *147*, 4–20. [CrossRef]

9.   Saadi, S.; Todorovic, M.; Tanasijevic, L.; Pereira, L.S.; Pizzigalli, C.; Lionello, P. Climate change and Mediterranean agriculture: Impacts on winter wheat and tomato crop evapotranspiration, irrigation requirements and yield. *Agric. Water Manag.* **2015**, *147*, 103–115. [CrossRef]

10. Smith, M.; Allen, R.; Monteith, J.; Perrier, A.; Pereira, L.; Segeren, A. *Report on the Expert Consultation on Procedures for Revision of FAO Guidelines for Prediction of Crop Water Requirements*; FAO: Rome, Italy, 1991.

11. Abdelhadi, A.; Hata, T.; Tanakamaru, H.; Tada, A.; Tariq, M. Estimation of crop water requirements in arid region using Penman–Monteith equation with derived crop coefficients: A case study on Acala cotton in Sudan Gezira irrigated scheme. *Agric. Water Manag.* **2000**, *45*, 203–214. [CrossRef]

12. Smith, M.; Allen, R.; Pereira, L. *Revised FAO Methodology for Crop-Water Requirements*; International Atomic Energy Agency (IAEA): Vienna, Austria, 1998.

13. Wang, L.; Kisi, O.; Zounemat-Kermani, M.; Li, H. Pan evaporation modeling using six different heuristic computing methods in different climates of China. *J. Hydrol.* **2017**, *544*, 407–427. [CrossRef]

14. Keskin, M.E.; Terzi, Ö. Artificial neural network models of daily pan evaporation. *J. Hydrol. Eng.* **2006**, *11*, 65–70. [CrossRef]

15. Guven, A.; Kişi, Ö. Daily pan evaporation modeling using linear genetic programming technique. *Irrig. Sci.* **2011**, *29*, 135–145. [CrossRef]

16. Traore, S.; Guven, A. Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel. *Water Resour. Manag.* **2012**, *26*, 4367–4380. [CrossRef]

17. Gundalia, M.J.; Dholakia, M. Estimation of pan evaporation using mean air temperature and radiation for monsoon season in Junagadh region. *Int. J. Eng. Res. Appl.* **2013**, *3*, 64–70.

18. Kisi, O.; Zounemat-Kermani, M. Comparison of two different adaptive neuro-fuzzy inference systems in modelling daily reference evapotranspiration. *Water Resour. Manag.* **2014**, *28*, 2655–2675. [CrossRef]

19. Wang, L.; Kisi, O.; Hu, B.; Bilal, M.; Zounemat-Kermani, M.; Li, H. Evaporation modelling using different machine learning techniques. *Int. J. Climatol.* **2017**, *37*, 1076–1092. [CrossRef]

20. Wang, L.; Niu, Z.; Kisi, O.; Li, C.A.; Yu, D. Pan evaporation modeling using four different heuristic approaches. *Comput. Electron. Agric.* **2017**, *140*, 203–213. [CrossRef]

21. Malik, A.; Kumar, A.; Kisi, O. Monthly pan-evaporation estimation in Indian central Himalayas using different heuristic approaches and climate based models. *Comput. Electron. Agric.* **2017**, *143*, 302–313. [CrossRef]

22. Ghorbani, M.A.; Deo, R.C.; Yaseen, Z.M.; Kashani, M.H.; Mohammadi, B. Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: Case study in North Iran. *Theor. Appl. Climatol.* **2018**, *133*, 1119–1131. [CrossRef]

23. Tao, H.; Diop, L.; Bodian, A.; Djaman, K.; Ndiaye, P.M.; Yaseen, Z.M. Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. *Agric. Water Manag.* **2018**, *208*, 140–151. [CrossRef]

24. Khosravi, K.; Daggupati, P.; Alami, M.T.; Awadh, S.M.; Ghareb, M.I.; Panahi, M.; Pham, B.T.; Rezaie, F.; Qi, C.; Yaseen, Z.M. Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. *Comput. Electron. Agric.* **2019**, *167*, 105041. [CrossRef]

25. Salih, S.Q.; Allawi, M.F.; Yousif, A.A.; Armanuos, A.M.; Saggi, M.K.; Ali, M.; Shahid, S.; Al-Ansari, N.; Yaseen, Z.M.; Chau, K.W. Viability of the advanced adaptive neuro-fuzzy inference system model on reservoir evaporation process simulation: Case study of Nasser Lake in Egypt. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 878–891. [CrossRef]

26. Feng, Y.; Jia, Y.; Zhang, Q.; Gong, D.; Cui, N. National-scale assessment of pan evaporation models across different climatic zones of China. *J. Hydrol.* **2018**, *564*, 314–328. [CrossRef]

27. Qasem, S.N.; Samadianfard, S.; Kheshtgar, S.; Jarhan, S.; Kisi, O.; Shamshirband, S.; Chau, K.-W. Modeling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 177–187. [CrossRef]

28. Yaseen, Z.M.; Al-Juboori, A.M.; Beyaztas, U.; Al-Ansari, N.; Chau, K.W.; Qi, C.; Ali, M.; Salih, S.Q.; Shahid, S. Prediction of evaporation in arid and semi-arid regions: A comparative study using different machine learning models. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 70–89. [CrossRef]

29. Pasolli, L.; Melgani, F.; Blanzieri, E. Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 464–468. [CrossRef]

30. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Philip, S.Y. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]

31. Hastie, T.; Tibshirani, R.; Friedman, J.; Franklin, J. The elements of statistical learning: Data mining, inference and prediction. *Math. Intell.* **2005**, *27*, 83–85.

32. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

33. Shi, T.; Horvath, S. Unsupervised learning with random forest predictors. *J. Comput. Graph. Stat.* **2006**, *15*, 118–138. [CrossRef]

34. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Association for Computing Machinery, New York, NY, USA, July 1992; pp. 144–152.

35. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin, Germany, 2013.

36. Kisi, O.; Cimen, M. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *J. Hydrol.* **2011**, *399*, 132–140. [CrossRef]

37. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. [CrossRef]

38. Mosavi, A.; Ozturk, O.; Chau, K. Flood prediction using machine learning models: Literature review. *Water* **2018**, *10*, 1536. [CrossRef]

39. Mosavi, A.; Salimi, M.; Faizollahzadeh Ardabili, S.; Rabczuk, T.; Shamshirband, S.; Varkonyi-Koczy, A.R. State of the Art of Machine Learning Models in Energy Systems, a Systematic Review. *Energies* **2019**, *12*, 1301. [CrossRef]

40. Ouaer, H.; Hosseini, A.H.; Amar, M.N.; Seghier, M.E.A.B.; Ghriga, M.A.; Nabipour, N.; Andersen, P.Ø. Rigorous Connectionist Models to Predict Carbon Dioxide Solubility in Various Ionic Liquids. *Appl. Sci.* **2020**, *10*, 304. [CrossRef]

41. Asadi, E.; Isazadeh, M.; Samadianfard, S.; Ramli, M.F.; Mosavi, A.; Nabipour, N.; Shamshirband, S.; Hajnal, E.; Chau, K.-W. Groundwater Quality Assessment for Sustainable Drinking and Irrigation. *Sustainability* **2020**, *12*, 177. [CrossRef]

42. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F. An Ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci. Total Environ.* **2019**, *651*, 2087–2096. [CrossRef]

43. Nabipour, N.; Mosavi, A.; Baghban, A.; Shamshirband, S.; Felde, I. Extreme Learning Machine-Based Model for Solubility Estimation of Hydrocarbon Gases in Electrolyte Solutions. *Preprints* **2020**, *2020*, 010010. [CrossRef]

44. Shamshirband, S.; Hadipoor, M.; Baghban, A.; Mosavi, A.; Bukor, J.; Várkonyi-Kóczy, A.R. Developing an ANFIS-PSO Model to Predict Mercury Emissions in Combustion Flue Gases. *Mathematics* **2019**, *7*, 965. [CrossRef]

45. Bemani, A.; Baghban, A.; Shamshirband, S.; Mosavi, A.; Csiba, P.; Varkonyi-Koczy, A.R. Applying ANN, ANFIS, and LSSVM Models for Estimation of Acid Solvent Solubility in Supercritical CO_2. *arXiv* **2019**, arXiv:1912.05612. [CrossRef]

46. Riahi-Madvar, H.; Dehghani, M.; Seifi, A.; Salwana, E. Comparative analysis of soft computing techniques RBF, MLP, and ANFIS with MLR and MNLR for predicting grade-control scour hole geometry. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 529–550. [CrossRef]

47. Choubin, B. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Sci. Total Environ.* **2020**, *701*, 134474. [CrossRef] [PubMed]

48. Qasem, S.N.; Samadianfard, S.; Nahand, H.S.; Mosavi, A.; Shamshirband, S.; Chau, K.W. Estimating daily dew point temperature using machine learning algorithms. *Water* **2019**, *11*, 582. [CrossRef]

49. Samadianfard, S.; Jarhan, S.; Salwana, E.; Mosavi, A.; Shamshirband, S.; Akib, S. Support Vector Regression Integrated with Fruit Fly Optimization Algorithm for River Flow Forecasting in Lake Urmia Basin. *Water* **2019**, *11*, 1934. [CrossRef]

50. Samadianfard, S.; Panahi, S. Estimating Daily Reference Evapotranspiration using Data Mining Methods of Support Vector Regression and M5 Model Tree. *J. Watershed Manag. Res.* **2019**, *9*, 157–167.

51. Dodangeh, E.; Choubin, B.; Eigdir, A.N. Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Sci. Total Environ.* **2019**, *9*, 135983. [CrossRef]

52. Zounemat-Kermani, M.; Seo, Y.; Kim, S.; Ghorbani, M.A.; Samadianfard, S.; Naghshara, S.; Kim, N.W.; Singh, V.P. Can Decomposition Approaches Always Enhance Soft Computing Models? Predicting the Dissolved Oxygen Concentration in the St. Johns River, Florida. *Appl. Sci.* **2019**, *9*, 2534. [CrossRef]