water

MDPI

Article

Estimation of Reference Evapotranspiration in Semi-Arid Region with Limited Climatic Inputs Using Metaheuristic Regression Methods

Saad Sh. Sammen [1], Ozgur Kisi [2,3,*], Ahmed Mohammed Sami Al-Janabi [4], Ahmed Elbeltagi [5] and Mohammad Zounemat-Kermani [6,*]

1 Department of Civil Engineering, College of Engineering, Diyala University, Baquba 32001, Iraq; saad123engineer@yahoo.com
2 Department of Civil Engineering, Technical University of Lübeck, 23562 Lübeck, Germany
3 Department of Civil Engineering, Ilia State University, 0162 Tbilisi, Georgia
4 Department of Civil Engineering, Cihan University-Erbil, Kurdistan Region, Erbil 44001, Iraq; ahmed.mohammedsami@cihanuniversity.edu.iq
5 Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt; ahmedelbeltagy81@mans.edu.eg
6 Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman 93630, Iran
* Correspondence: ozgur.kisi@th-luebeck.de (O.K.); zounemat@uk.ac.ir (M.Z.-K.)

Abstract: Different regression-based machine learning techniques, including support vector machine (SVM), random forest (RF), Bagged trees algorithm (BaT), and Boosting trees algorithm (BoT) were adopted for modeling daily reference evapotranspiration (ET_0) in a semi-arid region (Hemren catchment basin in Iraq). An assessment of the methods with various input combinations of climatic parameters, including solar radiation (SR), wind speed (WS), relative humidity (RH), and maximum and minimum air temperatures (Tmax and Tmin), indicated that the RF method, especially with Tmax, Tmin, Tmean, and SR inputs, provided the best accuracy in estimating daily ET_0 in all stations, while the SVM had the worst accuracy. This work will help water users, developers, and decision makers in water resource planning and management to achieve sustainability.

Keywords: climatic inputs; evapotranspiration; Hemren catchment; machine learning; prediction; semi-arid region

Citation: Sammen, S.S.; Kisi, O.; Al-Janabi, A.M.S.; Elbeltagi, A.; Zounemat-Kermani, M. Estimation of Reference Evapotranspiration in Semi-Arid Region with Limited Climatic Inputs Using Metaheuristic Regression Methods. *Water* **2023**, *15*, 3449. https://doi.org/10.3390/w15193449

Academic Editor: Jianjun Ni

Received: 28 August 2023
Revised: 25 September 2023
Accepted: 27 September 2023
Published: 30 September 2023

1. Introduction

Evapotranspiration is an important factor in the hydrological cycle and is used in water resource management for irrigation [1,2], drought estimation and monitoring [3–5], and in estimation of crop production [6,7]. Evapotranspiration (ET) could be defined as the amount of water that is transferred from the Earth's surface to the atmosphere, and it plays a significant role in the world's ecosystem that is related to water, energy, and carbon cycles [8,9]. ET is affected by different factors, including precipitation, temperature, relative humidity, wind speed, and solar radiation [10]. In addition, there is a common consent that terrestrial ET around the world has been changed as a result of climate change and human activity in the last decades [11–15]. Therefore, in order to calculate terrestrial ET, it is necessary to recognize the roles of water management, the hydrological cycle, and the impact of climate change [16,17].

Reference evapotranspiration (ET_0) can be estimated using different methods and approaches, including statistical or empirical methods, remote-sensing methods, and physical model-based methods [11]. In the first method, ET_0 is estimated using flux tower observations [18,19], while in the second method, ET_0 is calculated using the integration of remote-sensing data with experimental observations [20,21], the surface energy balance equation [22,23], Penman–Monteith or Priestley–Taylor equations [24–26], and data assimilation methods [27–29]. In the third method, ET_0 is estimated using a physical model alone or integrated with data-simulation algorithms [30–32]. Although several studies have used these methods globally, the daily estimation of ET_0 using these methods comes with some uncertainty [33]. Relative variation between the observed and estimated values was found to range from 14% to 44% [34,35]. On the other hand, Lysimeters provide more precise measurements of ET values, and many researchers estimate ET by employing the measurements of Lysimeters [36,37]. Unfortunately, the use of Lysimeters has some drawbacks such as the cost of installation and maintenance, in addition to environmental impacts. Additionally, the restricted number of Lysimeters impedes the observation of ET at specific locations [38]. Based on the above, the development of new adequate methods to estimate reference evapotranspiration (ET_0) with more accuracy and low cost is important and necessary.

In the last decades, the use of machine learning (ML) has received more attention in the field of water resource management around the world, including ET_0 estimation. ML has been applied to estimate the parameters of hydrology [39–45], hydraulics [46–49], and water quality by many researchers [42,50–53]. Several previous studies have reported the capability of ML techniques in estimating ET_0 [54]. A comparative study using two ML techniques, namely generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN), in addition to empirical methods for ET_0 estimation in Algeria, is presented in [55]. It concluded that for ET_0 estimation, the results obtained from the use of GRNN are better than those obtained from using the RBF. Another ML model, which is the support vector machine (SVM) model, was developed for ET_0 estimation in [56] using limited climatic data. They used different parameters such as maximum and minimum temperature, wind speed, and solar radiation with several input combinations. The results acknowledged that the SVM is useful for ET_0 estimation with acceptable accuracy. A comparison between an artificial neural network (ANN) with empirical approaches to estimate ET_0 using the daily meterological data has also been conducted [57]. It used two types of ANN with three empirical approaches that included Priestley–Taylor, Makkink, Hargreaves, and mass transfer. In [58], different machine learning techniques were used for ET_0 estimation with more actual and precise limited meteorological variables. The results generalized the relation between the various meteorological parameters. Moreover, the performance of ten ML techniques was evaluated in [59]. It used three statistical indices that included RMSE, R^2, and bias to evaluate the modeling results. In [60], a deep neural network (DNN) model was developed for ET_0 estimation using four meteorological stations in Turkey. In that study, the results of DNN were compared with results of ANN. The study revealed that the output of DNN was more accurate compared to that of ANN. In [61], eight ML techniques were evaluated in estimating ET_0 using temperature data only. Additionally, the results were compared with the Hargreaves–Samani equation (a temperature-based equation). It was concluded that the accuracy of the developed models varied with various scenarios. Five machine learning models to predict daily ET_0 across ten meteorological stations in China were developed in [62]. The results from comparison showed that the CAT model outperformed the other models. The overall findings of the previous studies indicate that the use of soft computing techniques for modeling the evaporation process is very promising, and further studies incorporating these techniques are recommended [63].

In this research, seven scenarios with different climate variables were evaluated by employing four regression-based ML techniques. To the best of the authors' knowledge, these regression-based ML methods have not been previously compared in estimating ET_0 in a semi-arid region.

2. Materials and Methods

2.1. Study Area

The catchment area of the Diyala River is at the eastern border of Iraq towards Iran. The northern part of it (within Iran) is mostly of mountainous character, with about 3000 m height. The Hemren Basin, a large catchment area located in the northeast of Iraq within the Diyala governorate, extends between (33°53′13.00″ to 35°25′41.61″ Northern latitude) and (44°30′47.68″ to 45°48′39.59″ Eastern longitude) inside Iraqi land, and it is located about 120 km northeast of Baghdad, the capital city in Iraq (Figure 1).

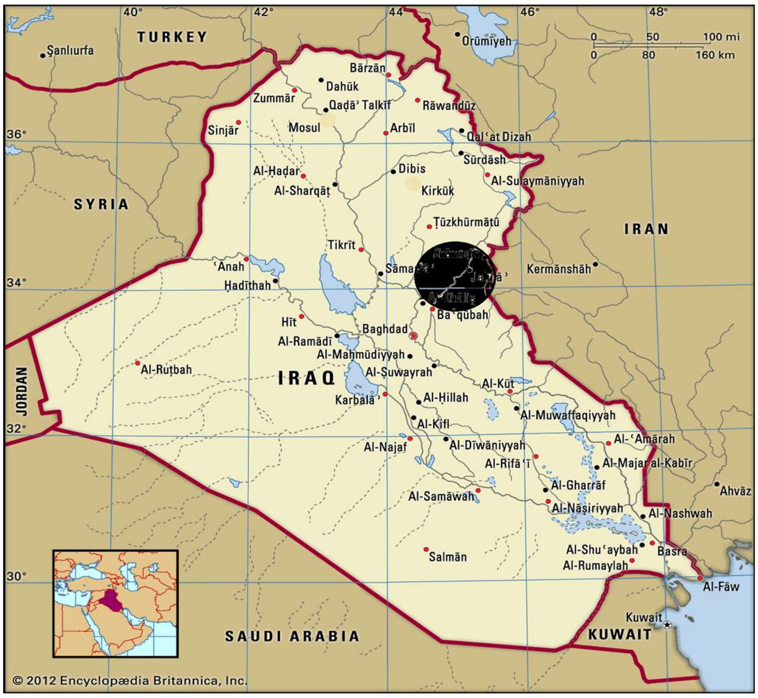


Figure 1. Location of the study area.

The relief of study area is characterized with topographic differences, their elevation ranges vary from 225 to 900 m above M.S.L. Therefore, the area was divided into three main regions. The length of the Diyala River within the catchment area is about 150 km, with an average gradient of 1 m per kilometer. Meanwhile, the Alwand River, which is the main tributary on the left side of the Diyala River, has a gradient of 2 m/km. It drains an area of 3974 km^2, and without the part in Iranian land, the area is 1974 km^2. The Narin River, which is the largest tributary on the right side, has a small gradient with a catchment area of 2344 km^2, and empties into the Diyala River near Hemren Mountains. Moreover, the downstream part of the catchment area, located between Derbendi Khan and Hemren, has lower altitudes and gradients.

2.2. Employed Data

In the present study, the capability of four regression-based machine learning methods, SVM, RF, BoT and BaT, was investigated for ET_0 estimation. Seven input scenarios were considered in this study using six climatic parameters, namely solar radiation (SR), wind speed (WS), relative humidity (RH), and maximum, mean, and minimum air temperatures (T_{max}, T_{min}, T_{mean}) as model inputs. The data were collected daily from five stations in Iraq, namely Mandali, Kalar, Iran–Iraq Border, Qarah-Tapah, and Adhim stations. Table 1 shows the statistical properties of the meteorological stations. The daily climate data during the period of 1979–2014 were collected from the study area and used for model development.

Table 1. Details of meteorological stations used in this study.

Station Name	Station Location		Temperature (°C)		Relative Humidity (%)		Solar Radiation (MJ/m^2)		Wind Speed (m/s)	
	Lon.	Lat.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.
Mandali	45.31	33.88	51.18	−8.76	96	3.2	32.2	0.81	12.7	0.7
Kalar	45.31	34.50	50.31	−4.69	98	3.5	32.65	0.3	10.25	0.72
Iraq–Iran Border	45.63	34.19	48.82	−11.67	96.6	3.62	33.28	0.56	11.13	0.63
Adhim	44.68	34.18	52.23	−3.45	99	3.7	32.05	0.3	10.82	0.66

2.3. Machine Learning Models

This section briefly explains the input combinations and machine leaning methods used in this study. ET mainly depends on temperature and other climate variables as stated by previous studies. The idea in this study was creating some scenarios including temperature as the first variable and then combining it with other variables to select which scenario was the best for predicting ET_0.

Support Vector Machines (SVM) are widely recognized as powerful machine learning (ML) models that yield valuable outcomes in both classification and regression problems. The SVM methodology employs structural risk minimization during the training process, which results in several effective features for simulating complex problems [46]. These features include the sparse presentation of solutions, good generalization ability, and the ability to avoid trapping in local minima. It is worth noting that the term Support Vector Regression (SVR) is commonly used for regression-based problems.

In this method, the input vector x is transformed into a higher-dimensional space through nonlinear mapping, where linear regression is applied to the input vector. Considering a solution space with x as the independent vector and y as the dependent vector variables for the dataset having N number of data pairs, the linear regression function can be written as:

$$f(x) = \sum_{n=1}^{N} w_n \varphi_n(x) + b \tag{1}$$

where $\varphi(x)$ represents a non-linear function that maps the low input space to the high output space; w represents the weights vector, while b denotes the threshold.

Unlike the SVR model, the other three ML models, RF, BaT, and BoT, are based on the concepts of decision and regression tree models that employ ensemble learning techniques. Decision tree learning is a supervised learning approach that is used to solve both classification and regression problems. Examples include Classification and Regression Tree (CART) models. In a decision tree model such as CART, each decision node in the tree represents a test on some input variables. Ensemble learning is a prosperous ML paradigm that merges a group of learners, rather than relying on a solitary learner, to forecast unfamiliar target attributes. It has been proven that using ensemble learning can improve the simulating and predicting results of individual models [64]. In this respect, two types of ensemble learning methodologies, namely bagging (here for the Bat and RF models) and boosting (here for the BoT model), are applied.

As indicated by [65], the BoT model incorporates important advantages of tree-based methods and has unique features. Its performance is based on an ensemble for training new samples. On the other hand, the bagging technique, also known as bootstrap aggregated, is an early ensemble method, which has numerous trees designed to improve the stability and accuracy of models. In the BaT, multiple independent decision trees can be constructed simultaneously on different segments of the training samples by utilizing distinct subsets of accessible characteristics.

Random forests are one of the most popular machine learning algorithms. They are so successful because they provide in general good predictive performance, low over-fitting, and easy interpretability. This interpretability is given by the fact that it is straightforward to derive the importance of each variable on tree decision. In other words, it is easy to compute how much each variable is contributing to the decision. The RF model acts similar to the BaT by constructing different decision trees. However, it uses a classification methodology for combining multiple trees to arrive at a conclusive outcome using the voting technique. Consequently, the RF classifier exhibits a robust ability to generalize. The RF can be considered as a specified type of the Bootstrap model. In each stage, the system has two subdivisions as unconnected segments to reduce the mean squared error values.

Feature selection through the random forest (RF) is categorized as an embedded method. Embedded methods encompass the advantages of both filter and wrapper techniques, as they rely on algorithms with built-in feature selection capabilities. Embedded methods offer several advantages, including:

High accuracy: They yield precise results.

Improved generalization: They enhance the model's ability to apply learned patterns to new data.

Interpretability: They provide insights into the significance of selected features.

Random forest comprises multiple decision trees, each constructed using a random subset of observations and a random subset of features from the dataset. This means that not every tree processes all the features or observations. This design ensures that the trees are uncorrelated, reducing the risk of overfitting. Each tree is essentially a sequence of binary questions based on individual or combined features. At each node (corresponding to each question), the tree partitions the dataset into two groups, each containing observations that are more similar to each other and dissimilar to those in the other group. Consequently, the importance of each feature is determined by how "pure" each of these partitioned groups becomes.

2.4. Evaluation of Models' Performance

The most important step in using machine-learning models is evaluating their accuracy. Performance evaluation of the four soft computing models was conducted based on regression analysis using four statistical indices, namely mean absolute error (MAE), root mean square error (RMSE), mean square error (MSE), and coefficient of determination (R^2).

3. Results

The utilized input scenarios for daily ET_0 estimation are presented in Table 2. From the table, the first Scenario (M1) used full variables as inputs, while the seventh Scenario (M7) had only two variables, T_{mean} and SR. The performance metrics of the employed methods in estimating daily ET_0 of the five stations are presented in the following sections.

Table 2. Scenarios of input combinations used for daily ET_0 estimation at study stations.

Scenario Code	Inputs	Technique			
		RF	SVM	BoT	BaT
M1	T_{max}, T_{min}, T_{mean}, SR, WS, and RH	×	×	×	×
M2	T_{max}, T_{min}, T_{mean}, and SR	×	×	×	×
M3	T_{mean}, SR, and WS	×	×	×	×
M4	SR, WS, RH, andT_{min}	×	×	×	×
M5	RH, WS, and T_{max}	×	×	×	×
M6	T_{max}, RH, and T_{min}	×	×	×	×
M7	T_{mean} and SR	×	×	×	×

3.1. Qarah-Tapah Station

From Figure 2 and Table 3, the performance metrics of the employed methods in estimating the daily ET_0 of the Qarah-Tapah station shows that the RF method had R^2, MSE, RMSE, and MAE ranges from 0.86 (RF-5) to 1 (RF-1), from 0.05 (RF-2) to 0.414 (RF-5), from 0.074 (RF-2) to 0.643 (RF-5), and from 0.055 (RF-2) to 0.487 (RF-5), respectively. It is evident from the metrics' ranges that the RF method was generally more successful in estimating the daily ET_0 of the Qarah-Tapah station. Another finding is that for the RF method, there was a small difference between the first and second scenarios, and the second one produced the best accuracy. The other methods behaved differently, for example, the first and second input combinations provided the same accuracy for the BoT method. The M1 scenario had slightly better accuracy than the M2 for the SVM method, while the first scenario performed worse compared to latter for the BaT. This difference can be explained by the different working principles of the four implemented methods. The best estimates belonged to the RF method, followed by the BaT methods, while the SVM generally produced the worst ET_0 estimates.

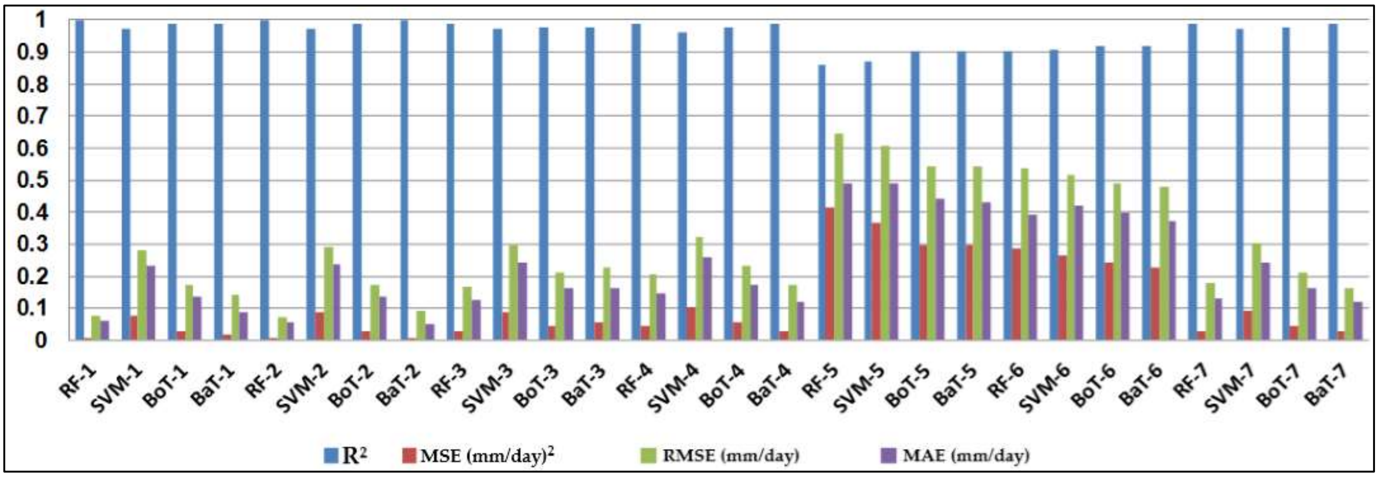


Figure 2. The effectiveness of the applied models in daily ET_0 estimation at the Qarah-Tapah station.

Table 3. The effectiveness of the applied models in daily ET_0 estimation at the Qarah-Tapah station during the testing period.

Model	ML Algorithm	Performance Metric			
		R^2	MSE (mm/day)^2	RMSE (mm/day)	MAE (mm/day)
M1	RF-1	1.00	0.006	0.077	0.061
	SVM-1	0.97	0.077	0.280	0.230
	BoT-1	0.99	0.029	0.172	0.135
	BaT-1	0.99	0.019	0.141	0.085
M2	RF-2	0.998	0.005	0.074	0.055
	SVM-2	0.97	0.086	0.293	0.239
	BoT-2	0.99	0.029	0.172	0.135
	BaT-2	1.00	0.008	0.091	0.052
M3	RF-3	0.99	0.029	0.170	0.125
	SVM-3	0.97	0.089	0.299	0.241
	BoT-3	0.98	0.044	0.211	0.161
	BaT-3	0.98	0.053	0.229	0.163
M4	RF-4	0.99	0.043	0.208	0.149
	SVM-4	0.96	0.105	0.325	0.258
	BoT-4	0.98	0.054	0.232	0.174
	BaT-4	0.99	0.029	0.171	0.122
M5	RF-5	0.86	0.414	0.643	0.487
	SVM-5	0.87	0.367	0.606	0.488
	BoT-5	0.90	0.297	0.545	0.439
	BaT-5	0.90	0.297	0.545	0.428
M6	RF-6	0.90	0.288	0.536	0.395
	SVM-6	0.91	0.267	0.516	0.421
	BoT-6	0.92	0.241	0.490	0.398
	BaT-6	0.92	0.229	0.478	0.372
M7	RF-7	0.99	0.031	0.176	0.129
	SVM-7	0.97	0.092	0.303	0.245
	BoT-7	0.98	0.045	0.212	0.161
	BaT-7	0.99	0.026	0.162	0.119

3.2. Mandali Station

The test performances of the RF, SVM, BoT, and BaT methods in estimating ET_0 of the Mandali Station are reported in Figure 3 and Table 4. Here, it was also clear that the RF-2 model had the lowest MSE (0.024), RMSE (0.156), and MAE (0.059) followed by those of the RF-1 and BaT-2 models, while SVM produced the worst estimates, similar to the previous station.

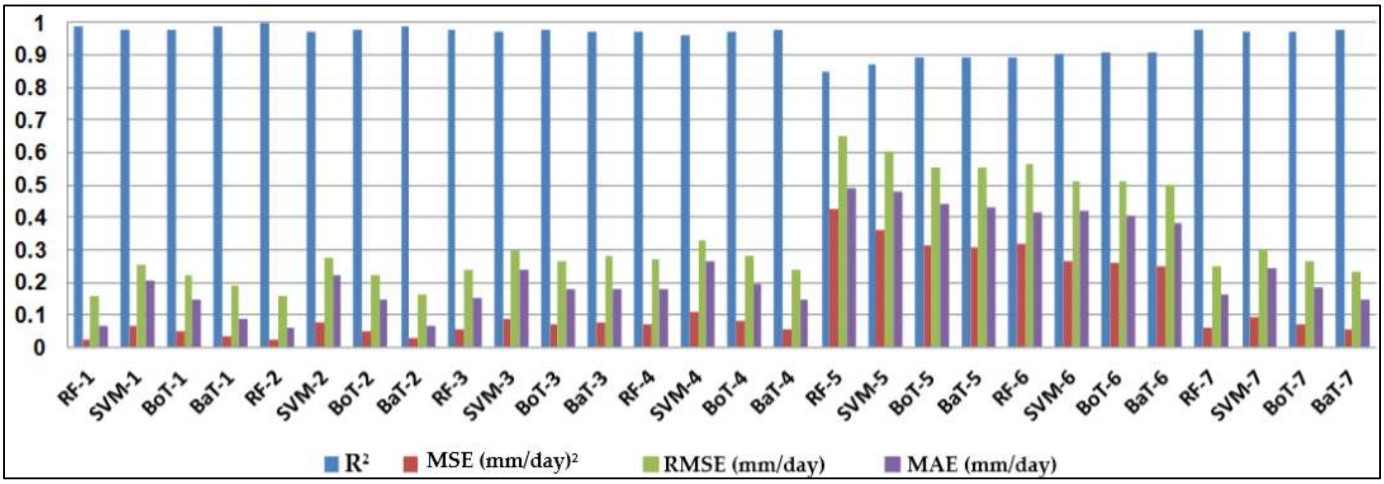


Figure 3. The effectiveness of the applied models in daily ET_0 estimation at the Mandali station.

Table 4. The effectiveness of the applied models in daily ET_0 estimation at the Mandali station during the testing period.

Model	ML Algorithm	Performance Metric			
		R^2	MSE $(mm/day)^2$	RMSE (mm/day)	MAE (mm/day)
M1	RF-1	0.99	0.025	0.157	0.067
	SVM-1	0.98	0.065	0.255	0.206
	BoT-1	0.98	0.048	0.220	0.146
	BaT-1	0.99	0.036	0.189	0.090
M2	RF-2	0.998	0.024	0.156	0.059
	SVM-2	0.97	0.075	0.274	0.223
	BoT-2	0.98	0.048	0.220	0.146
	BaT-2	0.99	0.026	0.161	0.066
M3	RF-3	0.98	0.056	0.237	0.150
	SVM-3	0.97	0.088	0.296	0.237
	BoT-3	0.98	0.069	0.263	0.180
	BaT-3	0.97	0.077	0.278	0.181
M4	RF-4	0.97	0.074	0.272	0.176
	SVM-4	0.96	0.109	0.330	0.262
	BoT-4	0.97	0.080	0.283	0.195
	BaT-4	0.98	0.055	0.236	0.147
M5	RF-5	0.85	0.425	0.652	0.492
	SVM-5	0.87	0.360	0.600	0.481
	BoT-5	0.89	0.310	0.556	0.442
	BaT-5	0.89	0.309	0.556	0.431
M6	RF-6	0.89	0.316	0.562	0.414
	SVM-6	0.90	0.263	0.513	0.418
	BoT-6	0.91	0.259	0.509	0.406
	BaT-6	0.91	0.250	0.500	0.385
M7	RF-7	0.98	0.063	0.251	0.162
	SVM-7	0.97	0.093	0.304	0.244
	BoT-7	0.97	0.071	0.266	0.182
	BaT-7	0.98	0.054	0.234	0.148

3.3. Kalar Station

Figure 4 and Table 5 present the test performances of the implemented four methods in estimating ET_0 of the Kalar Station. Similarly to Qarah-Tapah and Mandali stations, the RF-2 model provides the best performance with the lowest MSE (0.022), RMSE (0.148), and MAE (0.056) and the highest R^2 (0.998), followed by those of RF-1 and BaT-2 models. In this station, the BaT and BoT ranked second and third, while the SVM was the least accurate model.

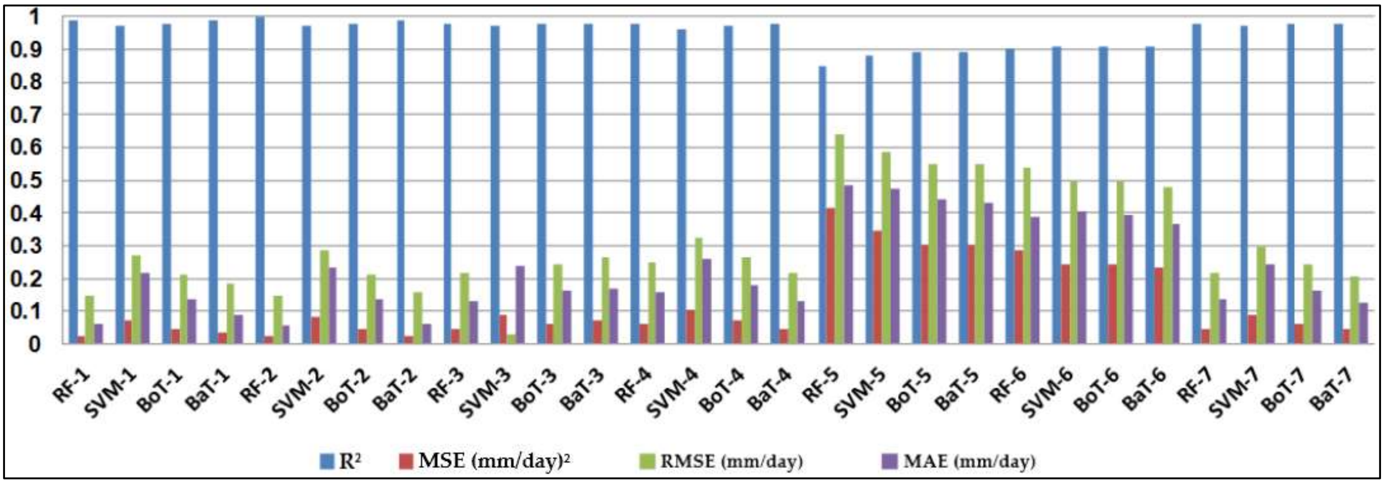


Figure 4. The effectiveness of the applied models in daily ET_0 estimation at the Kalar station.

Table 5. The effectiveness of the applied models in daily ET_0 estimation at the Kalar station during the testing period.

Model	ML Algorithm	Performance Metric			
		R^2	MSE (mm/day)^2	RMSE (mm/day)	MAE (mm/day)
M1	RF-1	0.99	0.022	0.149	0.062
	SVM-1	0.97	0.071	0.268	0.216
	BoT-1	0.98	0.044	0.210	0.138
	BaT-1	0.99	0.034	0.186	0.087
M2	RF-2	0.998	0.022	0.148	0.056
	SVM-2	0.97	0.080	0.284	0.231
	BoT-2	0.98	0.043	0.209	0.137
	BaT-2	0.99	0.024	0.155	0.060
M3	RF-3	0.98	0.046	0.215	0.132
	SVM-3	0.97	0.087	0.029	0.238
	BoT-3	0.98	0.058	0.242	0.162
	BaT-3	0.98	0.069	0.263	0.169
M4	RF-4	0.98	0.061	0.248	0.157
	SVM-4	0.96	0.104	0.322	0.257
	BoT-4	0.97	0.069	0.263	0.178
	BaT-4	0.98	0.047	0.217	0.131
M5	RF-5	0.85	0.412	0.641	0.484
	SVM-5	0.88	0.343	0.586	0.476
	BoT-5	0.89	0.302	0.549	0.439
	BaT-5	0.89	0.300	0.548	0.429
M6	RF-6	0.90	0.288	0.536	0.390
	SVM-6	0.91	0.243	0.493	0.403
	BoT-6	0.91	0.244	0.494	0.395
	BaT-6	0.91	0.230	0.480	0.368
M7	RF-7	0.98	0.047	0.217	0.134
	SVM-7	0.97	0.089	0.299	0.242
	BoT-7	0.98	0.058	0.242	0.163
	BaT-7	0.98	0.042	0.206	0.124

3.4. Iraq–Iran Station

Figure 5 and Table 6 give the performance metrics of the four methods in the test period of the Iraq–Iran Border station. In this station, RF-2 also ranked first by providing the lowest MSE (0.020), RMSE (0.143), and MAE (0.055) and the highest R^2 (0.998), followed by those of the RF-1 and BaT-2 models. Here as well BoT and BaT performed superior to the SVM method. Again, the fifth scenario provided the worst results, while the first and second input scenarios had the best ET_0 estimates.

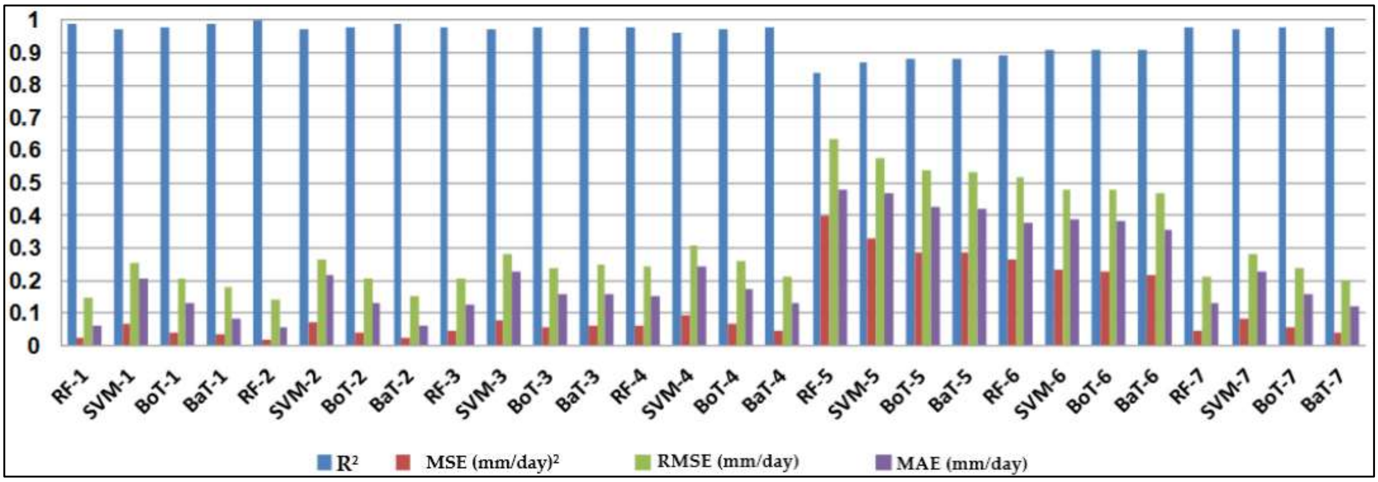


Figure 5. The effectiveness of the applied models in daily ET_0 estimation at the Iraq–Iran Border station.

Table 6. The effectiveness of the applied models in daily ET_0 estimation at the Iraq–Iran Border station during the testing period.

Model	ML Algorithm	Performance Metric			
		R^2	MSE $(mm/day)^2$	RMSE (mm/day)	MAE (mm/day)
M1	RF-1	0.99	0.021	0.145	0.062
	SVM-1	0.97	0.064	0.253	0.204
	BoT-1	0.98	0.041	0.203	0.133
	BaT-1	0.99	0.032	0.178	0.084
M2	RF-2	0.998	0.020	0.143	0.055
	SVM-2	0.97	0.071	0.266	0.216
	BoT-2	0.98	0.041	0.203	0.133
	BaT-2	0.99	0.022	0.151	0.059
M3	RF-3	0.98	0.042	0.205	0.125
	SVM-3	0.97	0.078	0.279	0.225
	BoT-3	0.98	0.055	0.237	0.157
	BaT-3	0.98	0.060	0.246	0.156
M4	RF-4	0.98	0.058	0.241	0.154
	SVM-4	0.96	0.093	0.306	0.242
	BoT-4	0.97	0.067	0.259	0.175
	BaT-4	0.98	0.045	0.213	0.130
M5	RF-5	0.84	0.401	0.633	0.478
	SVM-5	0.87	0.329	0.574	0.466
	BoT-5	0.88	0.287	0.536	0.427
	BaT-5	0.88	0.286	0.534	0.418
M6	RF-6	0.89	0.264	0.514	0.375
	SVM-6	0.91	0.230	0.480	0.390
	BoT-6	0.91	0.228	0.477	0.383
	BaT-6	0.91	0.217	0.466	0.358
M7	RF-7	0.98	0.045	0.213	0.132
	SVM-7	0.97	0.080	0.283	0.228
	BoT-7	0.98	0.056	0.237	0.158
	BaT-7	0.98	0.041	0.202	0.121

3.5. Adhim Station

The performance measures of the RF, SVM, BoT, and BaT methods in estimation at the Adhim Station are presented in Figure 6 and Table 7. The RF-2 model had the lowest MSE (0.006), RMSE (0.078), and MAE (0.058), followed by those of the RF-1 and BaT-2 models. SVM generally produced the worst estimates, while BaT and BoT methods followed the RF in accuracy for estimating the daily ET_0. Models including the first and/or scond scenarios perform the best, while models with the fifth combination had the worst results. In this station, BoT-1 had better accuracy than BoT-2 did; however, this difference was marginal.

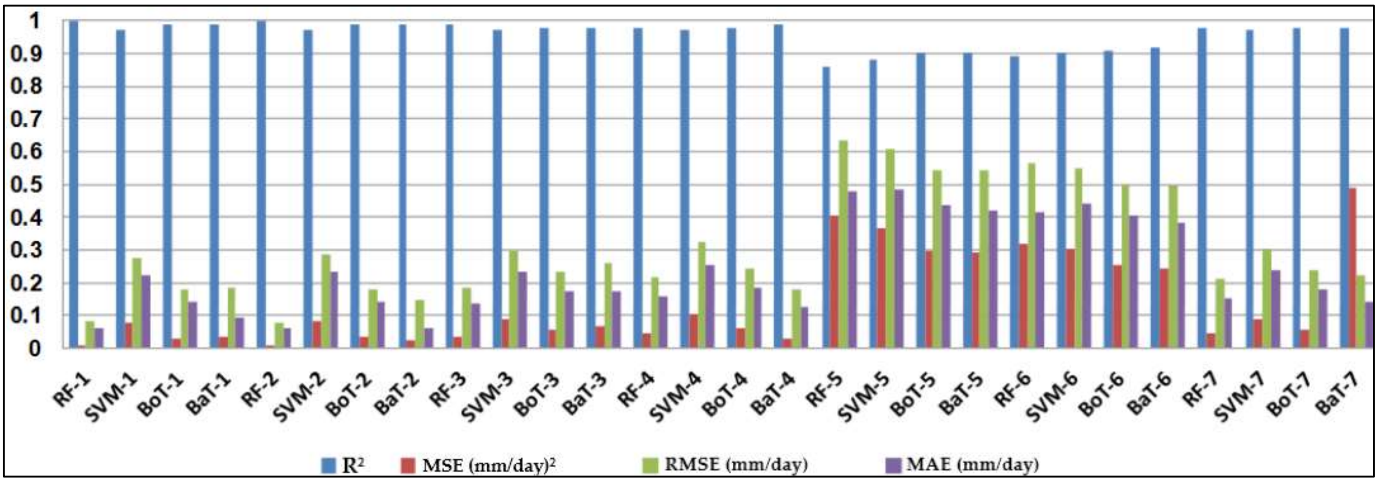


Figure 6. The effectiveness of the applied models in daily ET_0 estimation at the Adhim station.

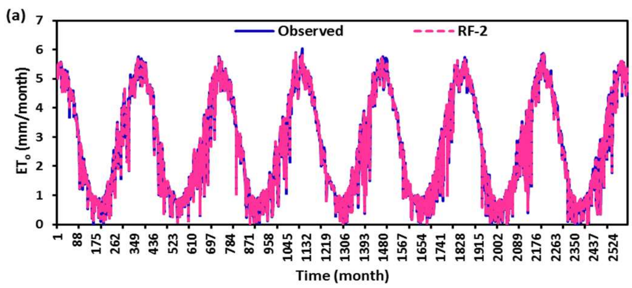
Table 7. The effectiveness of the applied models in daily ET_0 estimation at the Adhim station during the testing period.

Model	ML Algorithm	Performance Metric			
		R^2	MSE (mm/day)^2	RMSE (mm/day)	MAE (mm/day)
M1	RF-1	1.00	0.006	0.082	0.061
	SVM-1	0.97	0.075	0.275	0.222
	BoT-1	0.99	0.031	0.178	0.140
	BaT-1	0.99	0.033	0.184	0.093
M2	RF-2	0.998	0.006	0.078	0.058
	SVM-2	0.97	0.082	0.287	0.234
	BoT-2	0.99	0.032	0.179	0.140
	BaT-2	0.99	0.022	0.148	0.059
M3	RF-3	0.99	0.034	0.185	0.134
	SVM-3	0.97	0.086	0.294	0.234
	BoT-3	0.98	0.053	0.231	0.173
	BaT-3	0.98	0.068	0.261	0.173
M4	RF-4	0.98	0.047	0.218	0.156
	SVM-4	0.97	0.103	0.321	0.253
	BoT-4	0.98	0.058	0.241	0.182
	BaT-4	0.99	0.031	0.177	0.127
M5	RF-5	0.86	0.405	0.636	0.481
	SVM-5	0.88	0.367	0.606	0.485
	BoT-5	0.90	0.296	0.544	0.434
	BaT-5	0.90	0.292	0.541	0.422
M6	RF-6	0.89	0.319	0.565	0.415
	SVM-6	0.90	0.301	0.548	0.443
	BoT-6	0.91	0.253	0.503	0.404
	BaT-6	0.92	0.243	0.493	0.381
M7	RF-7	0.98	0.044	0.211	0.152
	SVM-7	0.97	0.090	0.300	0.239
	BoT-7	0.98	0.057	0.239	0.178
	BaT-7	0.98	0.491	0.221	0.143

Overall, the RF method, especially with the T_{max}, T_{min}, T_{mean}, and SR inputs, provided the best accuracy in estimating the daily ET_0 of all stations. Its accuracy was followed by that of the BaT and BoT methods, while SVM had the worst accuracy. In most cases, the second input scenario provided the best accuracy in estimating the daily ET_0. It is also worth noting that the seventh input scenario, having only T_{mean} and SR inputs, performed superior to the fourth, fifth, and sixth input scenarios. These results are contrary to those of [56], where a support vector machine (SVM) model was developed for ET_0 estimation using limited climatic data (i.e., maximum and minimum temperature, wind speed, and solar radiation). The results in that study acknowledged that the SVM was useful for ET_0 estimation with acceptable accuracy.

On the other hand, [66] used two machine-learning methods, random vector functional link (RVFL) and relevance vector machine (RVM), in modeling ET_0 using limited climatic data, T_{max}, T_{min}, and extraterrestrial radiation with various input combinations and three data split scenarios. The study indicated that using only the temperature input (Tmin, Tmax) provided the worst ET_0 estimations. Other studies also acknowledged similar results [67,68]. Another study found that temperature-based models involving temperature and Ra inputs offer promising results [61].

Figures 7 and 8 illustrate the time variation and scatter plots of the best model (RF-2) estimates. It is clear from Figure 7 that the ET_0 estimates by RF-2 were closely following the observed values. The models' ranks from the best to worst are presented in Table 8. From Table 8, we can say that the first or second input scenarios, inlcuding T_{max}, T_{min}, T_{mean}, SR, WS, and RH; and T_{max}, T_{min}, T_{mean} and SR variables, respectively, generally provided the best estimates, while the fifth scenario, involving RH, WS, and T_{max}, gave the worst ET_0 estimates. The main reason of this might be the fact that the SR input is very effective for ET_0, and not having it in this combination (fifth) and involving the WS parameter may worsen the estimation accuracy. Adding some input variables may negatively affect the variance and cause worse model accuracy in machine learning modeling. Here, adding WS might deteriorate the model's performance, as this can be observed from the first and second input cases. From Table 8 it can also clearly be seen that for the Kalar Station, the first and second scenarios produced the best estimates, whereas the fifth scenario had the worst results. As clearly seen from Figure 8, the fit line of RF-2 overlapped the ideal line (1:1 line), and it had a high correlation for all stations. All these results highly recommend the RF method in estimating daily ET_0.



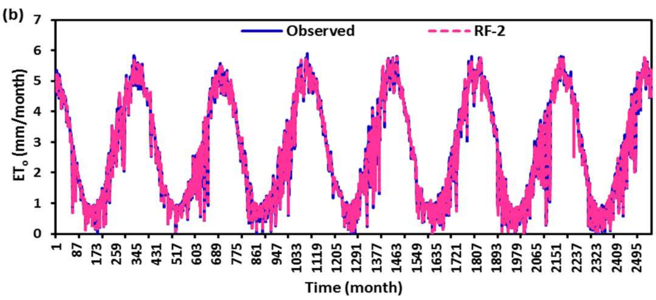


Figure 7. *Cont.*

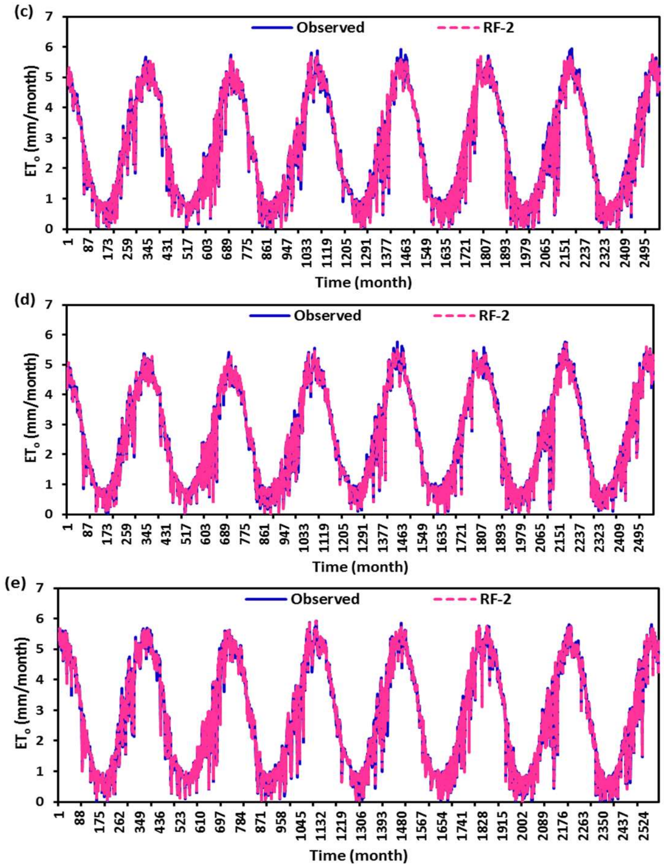


Figure 7. The temporal distribution of observed vs. estimated monthly ET_0 values yielded by the best RF-2 model corresponding to M2 at (**a**) Qarah-Tapah, (**b**) Mandali, (**c**) Kalar, (**d**) Iraq–Iran Border, and (**e**) Adhim stations during the testing period.

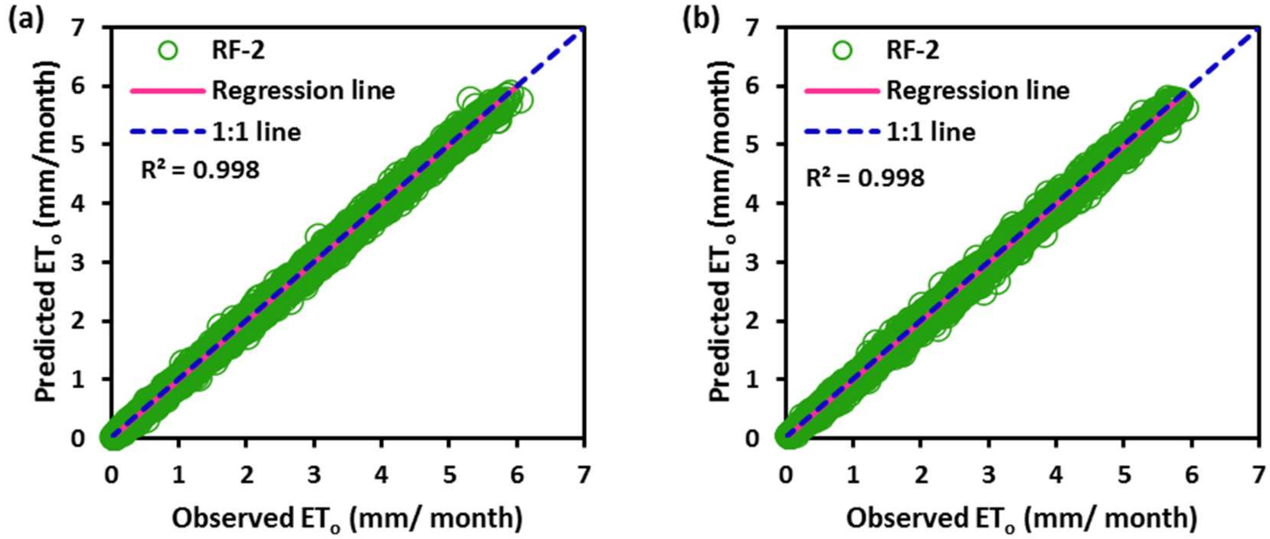


Figure 8. *Cont.*

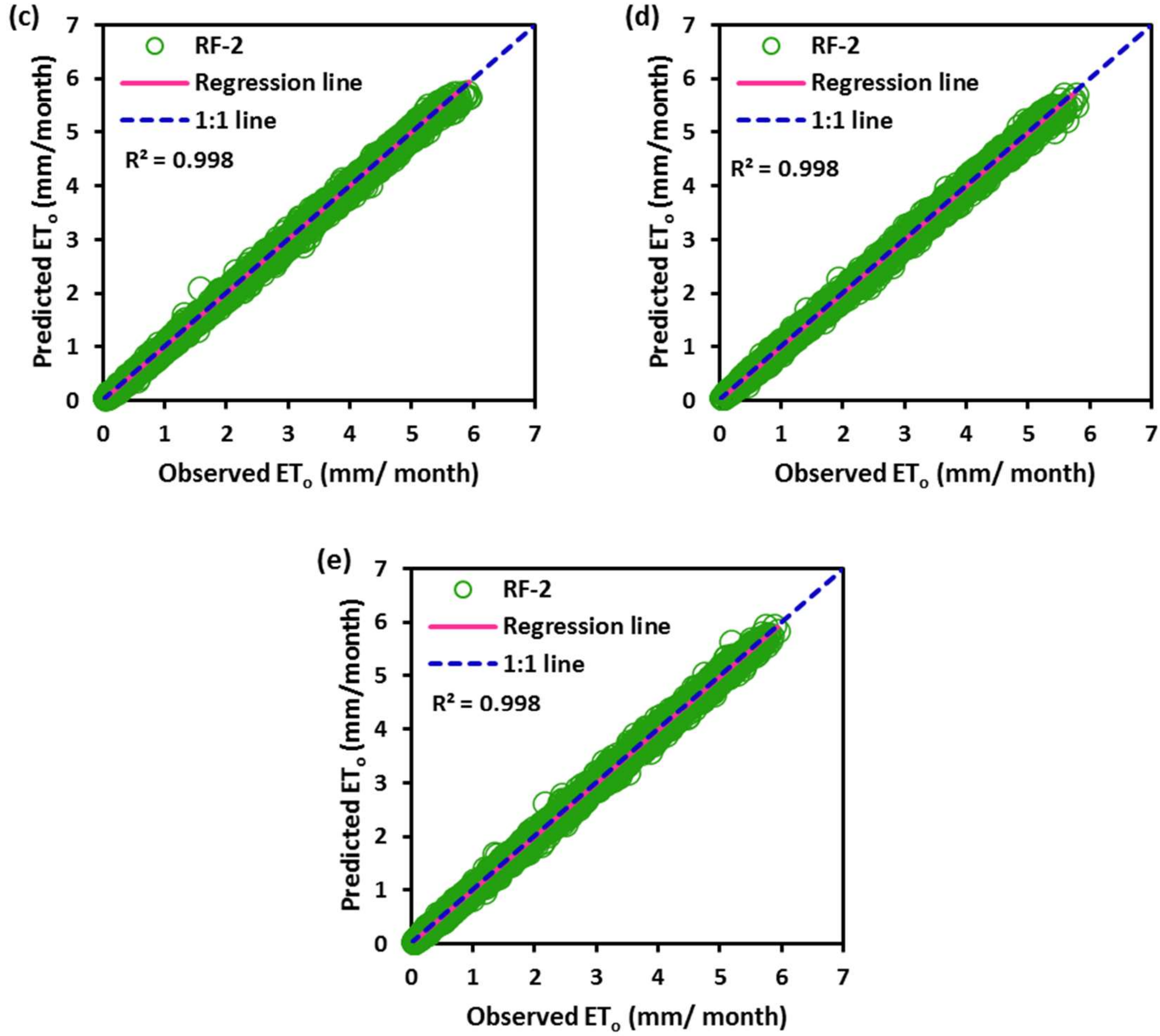


Figure 8. Scatter plots of the best RF-2 model corresponding to M2 for monthly ET_0 estimation at (**a**) Qarah-Tapah, (**b**) Mandali, (**c**) Kalar, (**d**) Iraq–Iran Border, and (**e**) Adhim stations during the testing period.

Table 8. Ranks of ML models based on the five stations.

Model	ML Algorithm	Models' Ranks Based on Stations				
		Qarah-Tapah	Mandali	Kalar	Iraq–Iran Border	Adhim
M1	RF-1	2	2	2	2	2
	SVM-1	16	8	8	8	8
	BoT-1	6	15	16	15	15
	BaT-1	4	23	23	23	24
M2	RF-2	1	1	1	1	1
	SVM-2	17	9	9	9	9
	BoT-2	7	16	15	16	16
	BaT-2	3	22	22	22	22

Table 8. *Cont.*

Model	ML Algorithm	Models' Ranks Based on Stations				
		Qarah-Tapah	Mandali	Kalar	Iraq–Iran Border	Adhim
M3	RF-3	8	3	3	3	3
	SVM-3	18	10	10	10	10
	BoT-3	12	17	17	17	17
	BaT-3	14	26	26	26	26
M4	RF-4	11	5	5	5	5
	SVM-4	20	12	12	12	12
	BoT-4	15	19	19	19	19
	BaT-4	9	25	25	25	23
M5	RF-5	28	7	7	7	7
	SVM-5	27	14	14	14	14
	BoT-5	26	21	21	21	21
	BaT-5	25	28	28	28	28
M6	RF-6	24	6	6	6	6
	SVM-6	23	13	13	13	13
	BoT-6	22	20	20	20	20
	BaT-6	21	27	27	27	27
M7	RF-7	10	4	4	4	4
	SVM-7	19	11	11	11	11
	BoT-7	13	18	18	18	18
	BaT-7	5	24	24	24	25

4. Discussion

Evapotranspiration plays a significant role in the hydrological cycle and finds applications in water resource management, including irrigation, as well as in the assessment and surveillance of drought conditions [69–76]. Four different regression-based machine learning methods were compared for modeling daily reference evapotranspiration (ET0) for a semi-arid region (the Hemren catchment basin in Iraq). The comparison statistics indicated that the random forest method with Tmax, Tmin, Tmean, and SR inputs performed superior to the other methods in estimating the daily ET_0 at all stations, while the SVM had the worst accuracy. Random forest (RF) is a popular machine-learning algorithm known for its strong predictive performance, minimal overfitting, and ease of interpretability. RF constructs multiple decision trees and combines them using a voting mechanism, resulting in robust generalization capabilities. One key feature of RF is its use of embedded methods for feature selection, which combines the strengths of filter and wrapper methods. Embedded methods are known for their high accuracy, excellent generalization, and interpretability. RF builds multiple decision trees, each using a random subset of data observations and features. This ensures that the trees are uncorrelated and less prone to overfitting. Each tree consists of a series of questions based on features, with each question dividing the data into two groups based on their similarity, ultimately determining the importance of each feature. In summary, RF is a powerful machine learning algorithm valued for its predictive abilities, robustness, and interpretability, making it a popular choice in various applications.

In [55], ET_0 estimation was carried out using radial based artificial neural networks (RBNNs) and generalized regression artificial neural networks (GRNNs). The inputs for this estimation included daily mean relative humidity, sunshine duration, maximum and minimum air temperatures, mean air temperature, and wind speed. The study yielded the best R^2 values of 0.868 and 0.889 for the RBNN and GRNN, respectively. Granata and Nunno [77] adopted two deep learning methods, NARX and LSTM, to model ET_0. They experimented with various input combinations, including solar radiation, mean air temperature, sensible heat flux, relative humidity, and lagged ET_0 values. The results showed R^2 values of 0.687 and 0.837 as the best performance achieved by the LSTM and

NARX models, respectively. The table data in Tables 3–7 clearly demonstrate that the proposed methods achieved remarkable success in modeling ET_0.

5. Conclusions

In this study, the applicability of four different regression-based machine learning methods in estimating ET_0 was investigated. Climatic data from five stations located in a semi-arid region of Iraq was used as inputs to the models. According to comparison statistics (R^2, MSE, RMSE, and MAE) and graphical inspection, the random forest model offered the best ET_0 estimates in all stations, while the SVM provided the worst accuracy. Employing various combinations of climatic inputs revealed that the models with Tmax, Tmin, Tmean, and SR inputs produced the best estimations. The best random forest model with Tmax, Tmin, Tmean, and SR improved the estimation accuracy of the SVM, BoT, and BaT models by 94%, 83%, and 38% for Qarah-Tapah, by 68%, 50%, and 8% for Mandali, by 73%, 49%, and 8% for Kalar, by 72%, 51%, and 9% for Iran–Iraq Border, and by 93%, 81%, and 73% for Adhim with respect to RMSE in the test period, respectively. The outcomes of the study recommend random forest for estimating ET_0 in a semi-arid region. The study used data from one region, and more data can be used to assess the regression-based machine learning methods in future studies. The regression-based methods considered in this study may be compared with more complex machine learning methods.

Author Contributions: Conceptualization: S.S.S., A.M.S.A.-J., O.K., and A.E.; data creation: O.K. and S.S.S.; formal analysis: O.K. and S.S.S.; investigation: O.K. and A.M.S.A.-J.; methodology: S.S.S. and A.M.S.A.-J.; software: A.E. and O.K.; validation: S.S.S. and O.K.; writing—original draft: O.K., A.M.S.A.-J. and S.S.S.; writing—review and editing, S.S.S., A.M.S.A.-J., O.K., A.E and M.Z.-K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: It will be available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. French, A.; Hunsaker, D.; Bounoua, L.; Karnieli, A.; Luckett, W.; Strand, R. Remote Sensing of Evapotranspiration over the Central Arizona Irrigation and Drainage District, USA. *Agronomy* **2018**, *8*, 278. [CrossRef]
2. Calera, A.; Campos, I.; Osann, A.; D'Urso, G.; Menenti, M. Remote Sensing for Crop Water Management: From ET Modelling to Services for the End Users. *Sensors* **2017**, *17*, 1104. [CrossRef] [PubMed]
3. Anderson, M.C.; Zolin, C.A.; Sentelhas, P.C.; Hain, C.R.; Semmens, K.; Tugrul Yilmaz, M.; Gao, F.; Otkin, J.A.; Tetrault, R. The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. *Remote Sens. Environ.* **2016**, *174*, 82–99. [CrossRef]
4. Senay, G.B.; Velpuri, N.M.; Bohms, S.; Budde, M.; Young, C.; Rowland, J.; Verdin, J.P. Drought Monitoring and Assessment. In *Hydro-Meteorological Hazards, Risks and Disasters*; Shroder, J.F., Paron, P., Di Baldassarre, G., Eds.; Elsevier: Amsterdam, The Netherlands, 2015; pp. 233–262.
5. Yusuf, B.; Al-Janabi, A.M.S.; Ghazali, A.H.; Al-Ani, I. Variations of infiltration capacity with flow hydraulic parameters in permeable stormwater channels. *ISH J. Hydraul. Eng.* **2022**, *28*, 234–242. [CrossRef]
6. Khan, A.; Stöckle, C.O.; Nelson, R.L.; Peters, T.; Adam, J.C.; Lamb, B.; Chi, J.; Waldo, S. Estimating Biomass and Yield Using METRIC Evapotranspiration and Simple Growth Algorithms. *Agron. J.* **2019**, *111*, 536–544. [CrossRef]
7. Doorenbos, J.; Kassam, A.H.; Bentvelsen, C.; Uittenbogaard, G. Yield Response to Water. In *Irrigation and Agricultural Development*; Johl, S.S., Ed.; Pergamon Press: Oxford, UK, 1980; pp. 257–280.
8. Trenberth, K.E.; Smith, L.; Qian, T.; Dai, A.; Fasullo, J. Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data. *J. Hydrometeorol.* **2007**, *8*, 758–769. [CrossRef]
9. Wang, K.; Dickinson, R.E. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.* **2012**, *50*. [CrossRef]
10. Granata, F. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agric. Water Manag.* **2019**, *217*, 303–315. [CrossRef]
11. Jung, M.; Reichstein, M.; Ciais, P.; Seneviratne, S.I.; Sheffield, J.; Goulden, M.L.; Bonan, G.; Cescatti, A.; Chen, J.; de Jeu, R.; et al. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **2010**, *467*, 951–954. [CrossRef]

12. Mueller, B.; Hirschi, M.; Jimenez, C.; Ciais, P.; Dirmeyer, P.A.; Dolman, A.J.; Fisher, J.B.; Jung, M.; Ludwig, F.; Maignan, F.; et al. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 3707–3720. [CrossRef]
13. Zeng, Z.; Piao, S.; Lin, X.; Yin, G.; Peng, S.; Ciais, P.; Myneni, R.B. Global evapotranspiration over the past three decades: Estimation based on the water balance equation combined with empirical models. *Environ. Res. Lett.* **2012**, *7*, 14026. [CrossRef]
14. Zeng, Z.; Wang, T.; Zhou, F.; Ciais, P.; Mao, J.; Shi, X.; Piao, S. A worldwide analysis of spatiotemporal changes in water balance-based evapotranspiration from 1982 to 2009. *J. Geophys. Res. Atmos.* **2014**, *119*, 1186–1202. [CrossRef]
15. Al-Janabi, A.M.S.; Halim Ghazali, A.; Yusuf, B. Modified models for better prediction of infiltration rates in trapezoidal permeable stormwater channels. *Hydrol. Sci. J.* **2019**, *64*, 1918–1931. [CrossRef]
16. Fisher, J.B.; Melton, F.; Middleton, E.; Hain, C.; Anderson, M.; Allen, R.; McCabe, M.F.; Hook, S.; Baldocchi, D.; Townsend, P.A.; et al. The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resour. Res.* **2017**, *53*, 2618–2626. [CrossRef]
17. Petropoulos, G.P.; Ireland, G.; Lamine, S.; Griffiths, H.M.; Ghilain, N.; Anagnostopoulos, V.; North, M.R.; Srivastava, P.K.; Georgopoulou, H. Operational evapotranspiration estimates from SEVIRI in support of sustainable water management. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 175–187. [CrossRef]
18. Jung, M.; Reichstein, M.; Margolis, H.A.; Cescatti, A.; Richardson, A.D.; Arain, M.A.; Arneth, A.; Bernhofer, C.; Bonal, D.; Chen, J.; et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res.* **2011**, *116*. [CrossRef]
19. Xu, T.; Guo, Z.; Liu, S.; He, X.; Meng, Y.; Xu, Z.; Xia, Y.; Xiao, J.; Zhang, Y.; Ma, Y.; et al. Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale. *J. Geophys. Res. Atmos.* **2018**, *123*, 8674–8690. [CrossRef]
20. Cui, G.; Zhu, J. Infiltration model in sloping layered soils and guidelines for model parameter estimation. *Hydrol. Sci. J.* **2017**, *62*, 2222–2237. [CrossRef]
21. Zhao, B.; Mao, K.; Cai, Y.; Shi, J.; Li, Z.; Qin, Z.; Meng, X.; Shen, X.; Guo, Z. A combined Terra and Aqua MODIS land surface temperature and meteorological station data product for China from 2003 to 2017. *Earth Syst. Sci. Data* **2020**, *12*, 2555–2577. [CrossRef]
22. Ma, Y.; Liu, S.; Song, L.; Xu, Z.; Liu, Y.; Xu, T.; Zhu, Z. Estimation of daily evapotranspiration and irrigation water efficiency at a Landsat-like scale for an arid irrigation area using multi-source remote sensing data. *Remote Sens. Environ.* **2018**, *216*, 715–734. [CrossRef]
23. Miralles, D.G.; Holmes, T.R.H.; De Jeu, R.A.M.; Gash, J.H.; Meesters, A.G.C.A.; Dolman, A.J. Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 453–469. [CrossRef]
24. Yao, Y.; Liang, S.; Li, X.; Chen, J.; Wang, K.; Jia, K.; Cheng, J.; Jiang, B.; Fisher, J.B.; Mu, Q.; et al. A satellite-based hybrid algorithm to determine the Priestley–Taylor parameter for global terrestrial latent heat flux estimation across multiple biomes. *Remote Sens. Environ.* **2015**, *165*, 216–233. [CrossRef]
25. Zhang, K.; Kimball, J.S.; Running, S.W. A review of remote sensing based actual evapotranspiration estimation. *WIREs Water* **2016**, *3*, 834–853. [CrossRef]
26. Bateni, S.M.; Entekhabi, D.; Jeng, D.-S. Variational assimilation of land surface temperature and the estimation of surface energy balance components. *J. Hydrol.* **2013**, *481*, 143–156. [CrossRef]
27. He, X.; Xu, T.; Xia, Y.; Bateni, S.M.; Guo, Z.; Liu, S.; Mao, K.; Zhang, Y.; Feng, H.; Zhao, J. A Bayesian Three-Cornered Hat (BTCH) Method: Improving the Terrestrial Evapotranspiration Estimation. *Remote Sens.* **2020**, *12*, 878. [CrossRef]
28. Lu, Y.; Steele-Dunne, S.C.; Farhadi, L.; van de Giesen, N. Mapping Surface Heat Fluxes by Assimilating SMAP Soil Moisture and GOES Land Surface Temperature Data. *Water Resour. Res.* **2017**, *53*, 10858–10877. [CrossRef]
29. Xu, T.; Bateni, S.M.; Liang, S.; Entekhabi, D.; Mao, K. Estimation of surface turbulent heat fluxes via variational assimilation of sequences of land surface temperatures from Geostationary Operational Environmental Satellites. *J. Geophys. Res. Atmos.* **2014**, *119*, 10780–10798. [CrossRef]
30. Xia, Y.; Hao, Z.; Shi, C.; Li, Y.; Meng, J.; Xu, T.; Wu, X.; Zhang, B. Regional and Global Land Data Assimilation Systems: Innovations, Challenges, and Prospects. *J. Meteorol. Res.* **2019**, *33*, 159–189. [CrossRef]
31. Zhang, B.; Xia, Y.; Long, B.; Hobbins, M.; Zhao, X.; Hain, C.; Li, Y.; Anderson, M.C. Evaluation and comparison of multiple evapotranspiration data models over the contiguous United States: Implications for the next phase of NLDAS (NLDAS-Testbed) development. *Agric. For. Meteorol.* **2020**, *280*, 107810. [CrossRef]
32. Long, D.; Longuevergne, L.; Scanlon, B.R. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res.* **2014**, *50*, 1131–1151. [CrossRef]
33. Kalma, J.D.; McVicar, T.R.; McCabe, M.F. Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data. *Surv. Geophys.* **2008**, *29*, 421–469. [CrossRef]
34. Velpuri, N.M.; Senay, G.B.; Singh, R.K.; Bohms, S.; Verdin, J.P. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. *Remote Sens. Environ.* **2013**, *139*, 35–49. [CrossRef]
35. Anapalli, S.S.; Ahuja, L.R.; Gowda, P.H.; Ma, L.; Marek, G.; Evett, S.R.; Howell, T.A. Simulation of crop evapotranspiration and crop coefficients with data in weighing lysimeters. *Agric. Water Manag.* **2016**, *177*, 274–283. [CrossRef]

36. Liu, X.; Xu, C.; Zhong, X.; Li, Y.; Yuan, X.; Cao, J. Comparison of 16 models for reference crop evapotranspiration against weighing lysimeter measurement. *Agric. Water Manag.* **2017**, *184*, 145–155. [CrossRef]
37. Stanhill, G. Evapotranspiration. In *Encyclopedia of Soils in the Environment*; Hillel, D., Rosenzweig, C., Powlson, D., Scow, K., Singer, M., Sparks, D., Eds.; Academic Press: Cambridge, MA, USA, 2005; pp. 502–506.
38. Chen, Z.; Shi, R.; Zhang, S. An artificial neural network approach to estimate evapotranspiration from remote sensing and AmeriFlux data. *Front. Earth Sci.* **2012**, *7*, 103–111. [CrossRef]
39. Sihag, P.; Kumar, M.; Sammen, S.S. Predicting the infiltration characteristics for semi-arid regions using regression trees. *Water Supply* **2021**, *21*, 2583–2595. [CrossRef]
40. Malik, A.; Tikhamarine, Y.; Sammen, S.S.; Abba, S.I.; Shahid, S. Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. *Environ. Sci. Pollut. Res.* **2021**, *28*, 39139–39158. [CrossRef]
41. Sammen, S.S.; Ehteram, M.; Abba, S.I.; Abdulkadir, R.A.; Ahmed, A.N.; El-Shafie, A. A new soft computing model for daily streamflow forecasting. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 2479–2491. [CrossRef]
42. Abba, S.I.; Abdulkadir, R.A.; Sammen, S.S.; Pham, Q.B.; Lawan, A.A.; Esmaili, P.; Malik, A.; Al-Ansari, N. Integrating feature extraction approaches with hybrid emotional neural networks for water quality index modeling. *Appl. Soft Comput.* **2021**, *114*, 108036. [CrossRef]
43. Ebtehaj, I.; Sammen, S.S.; Sidek, L.M.; Malik, A.; Sihag, P.; Al-Janabi, A.M.S.; Chau, K.-W.; Bonakdari, H. Prediction of daily water level using new hybridized GS-GMDH and ANFIS-FCM models. *Eng. Appl. Comput. Fluid Mech.* **2021**, *15*, 1343–1361. [CrossRef]
44. Hashim, B.M.; Al Maliki, A.; Alraheem, E.A.; Al-Janabi, A.M.S.; Halder, B.; Yaseen, Z.M. Temperature and precipitation trend analysis of the Iraq Region under SRES scenarios during the twenty-first century. *Theor. Appl. Climatol.* **2022**, *148*, 881–898. [CrossRef]
45. Mdegela, L.; Municio, E.; De Bock, Y.; Luhanga, E.; Leo, J.; Mannens, E. Extreme Rainfall Event Classification Using Machine Learning for Kikuletwa River Floods. *Water* **2023**, *15*, 1021. [CrossRef]
46. Sihag, P.; Dursun, O.F.; Sammen, S.S.; Malik, A.; Chauhan, A. Prediction of aeration efficiency of Parshall and Modified Venturi flumes: Application of soft computing versus regression models. *Water Supply* **2021**, *21*, 4068–4085. [CrossRef]
47. Alomari, N.K.; Sihag, P.; Sami Al-Janabi, A.M.; Yusuf, B. Modeling of scour depth and length of a diversion channel flow system with soft computing techniques. *Water Supply* **2023**, *23*, 1267–1283. [CrossRef]
48. Sammen, S.S.; Ghorbani, M.A.; Malik, A.; Tikhamarine, Y.; AmirRahmani, M.; Al-Ansari, N.; Chau, K.-W. Enhanced Artificial Neural Network with Harris Hawks Optimization for Predicting Scour Depth Downstream of Ski-Jump Spillway. *Appl. Sci.* **2020**, *10*, 5160. [CrossRef]
49. Granata, F.; Di Nunno, F.; Modoni, G. Hybrid Machine Learning Models for Soil Saturated Conductivity Prediction. *Water* **2022**, *14*, 1729. [CrossRef]
50. Ehteram, M.; Sammen, S.S.; Panahi, F.; Sidek, L.M. A hybrid novel SVM model for predicting CO2 emissions using Multiobjective Seagull Optimization. *Environ. Sci. Pollut. Res.* **2021**, *28*, 66171–66192. [CrossRef]
51. Abba, S.I.; Abdulkadir, R.A.; Sammen, S.S.; Usman, A.G.; Meshram, S.G.; Malik, A.; Shahid, S. Comparative implementation between neuro-emotional genetic algorithm and novel ensemble computing techniques for modelling dissolved oxygen concentration. *Hydrol. Sci. J.* **2021**, *66*, 1584–1596. [CrossRef]
52. Alali, Y.; Harrou, F.; Sun, Y. Unlocking the Potential of Wastewater Treatment: Machine Learning Based Energy Consumption Prediction. *Water* **2023**, *15*, 2349. [CrossRef]
53. Ni, J.; Liu, R.; Li, Y.; Tang, G.; Shi, P. An Improved Transfer Learning Model for Cyanobacterial Bloom Concentration Prediction. *Water* **2022**, *14*, 1300. [CrossRef]
54. Yang, F.; White, M.A.; Michaelis, A.R.; Ichii, K.; Hashimoto, H.; Votava, P.; Zhu, A.-X.; Nemani, R.R. Prediction of Continental-Scale Evapotranspiration by Combining MODIS and AmeriFlux Data Through Support Vector Machine. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3452–3461. [CrossRef]
55. Ladlani, I.; Houichi, L.; Djemili, L.; Heddam, S.; Belouz, K. Modeling daily reference evapotranspiration (ET0) in the north of Algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): A comparative study. *Meteorol. Atmos. Phys.* **2012**, *118*, 163–178. [CrossRef]
56. Wen, X.; Si, J.; He, Z.; Wu, J.; Shao, H.; Yu, H. Support-Vector-Machine-Based Models for Modeling Daily Reference Evapotranspiration with Limited Climatic Data in Extreme Arid Regions. *Water Resour. Manag.* **2015**, *29*, 3195–3209. [CrossRef]
57. Antonopoulos, V.Z.; Antonopoulos, A.V. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Comput. Electron. Agric.* **2017**, *132*, 86–96. [CrossRef]
58. Adnan, M.; Latif, M.A.; Nazir, M. Estimating Evapotranspiration using Machine Learning Techniques. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 108–113. [CrossRef]
59. Carter, C.; Liang, S. Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 86–92. [CrossRef]
60. Özgür, A.; Yamaç, S.S. Modelling of daily reference evapotranspiration using deep neural network in different climates. *arXiv* **2020**, arXiv:2006.01760.
61. Wu, L.; Peng, Y.; Fan, J.; Wang, Y. Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. *Hydrol. Res.* **2019**, *50*, 1730–1750. [CrossRef]

62. Wu, T.; Zhang, W.; Jiao, X.; Guo, W.; Hamoud, Y.A. Comparison of five Boosting-based models for estimating daily reference evapotranspiration with limited meteorological variables. *PLoS ONE* **2020**, *15*, e0235324. [CrossRef]
63. Shiri, J.; Zounemat-Kermani, M.; Kisi, O.; Mohsenzadeh Karimi, S. Comprehensive assessment of 12 soft computing approaches for modelling reference evapotranspiration in humid locations. *Meteorol. Appl.* **2020**, *27*, e1841. [CrossRef]
64. Zounemat-Kermani, M.; Batelaan, O.; Fadaee, M.; Hinkelmann, R. Ensemble machine learning paradigms in hydrology: A review. *J. Hydrol.* **2021**, *598*, 126266. [CrossRef]
65. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]
66. Mostafa, R.R.; Kisi, O.; Adnan, R.M.; Sadeghifar, T.; Kuriqi, A. Modeling Potential Evapotranspiration by Improved Machine Learning Methods Using Limited Climatic Data. *Water* **2023**, *15*, 486. [CrossRef]
67. Dimitriadou, S.; Nikolakopoulos, K.G. Multiple Linear Regression Models with Limited Data for the Prediction of Reference Evapotranspiration of the Peloponnese, Greece. *Hydrology* **2022**, *9*, 124. [CrossRef]
68. Dimitriadou, S.; Nikolakopoulos, K.G. Artificial Neural Networks for the Prediction of the Reference Evapotranspiration of the Peloponnese Peninsula, Greece. *Water* **2022**, *14*, 2027. [CrossRef]
69. Vangelis, H.; Tigkas, D.; Tsakiris, G. The effect of PET method on Reconnaissance Drought Index (RDI) calculation. *J. Arid. Environ.* **2013**, *88*, 130–140. [CrossRef]
70. Tegos, A.; Stefanidis, S.; Cody, J.; Koutsoyiannis, D. On the Sensitivity of Standardized-Precipitation-Evapotranspiration and Aridity Indexes Using Alternative Potential Evapotranspiration Models. *Hydrology* **2023**, *10*, 64. [CrossRef]
71. Sang, L.; Zhu, G.; Xu, Y.; Sun, Z.; Zhang, Z.; Tong, H. Effects of Agricultural Large-And Medium-Sized Reservoirs on Hydrologic Processes in the Arid Shiyang River Basin, Northwest China. *Water Resour. Res.* **2023**, *59*, 2. [CrossRef]
72. Li, J.; Wang, Z.; Wu, X.; Xu, C.; Guo, S.; Chen, X. Toward Monitoring Short-Term Droughts Using a Novel Daily Scale, Standardized Antecedent Precipitation Evapotranspiration Index. *J. Hydrometeorol.* **2020**, *21*, 891–908. [CrossRef]
73. Yin, L.; Wang, L.; Keim, B.D.; Konsoer, K.; Yin, Z.; Liu, M.; Zheng, W. Spatial and wavelet analysis of precipitation and river discharge during operation of the Three Gorges Dam, China. *Ecol. Indic.* **2023**, *154*, 110837. [CrossRef]
74. Liu, Z.; Xu, J.; Liu, M.; Yin, Z.; Liu, X.; Yin, L.; Zheng, W. Remote sensing and geostatistics in urban water-resource monitoring: A review. *Mar. Freshw. Res.* **2023**. [CrossRef]
75. Tian, H.; Huang, N.; Niu, Z.; Qin, Y.; Pei, J.; Wang, J. Mapping Winter Crops in China with Multi-Source Satellite Imagery and Phenology-Based Algorithm. *Remote Sens.* **2019**, *11*, 820. [CrossRef]
76. Wu, B.; Quan, Q.; Yang, S.; Dong, Y. A social-ecological coupling model for evaluating the human-water relationship in basins within the Budyko framework. *J. Hydrol.* **2023**, *619*, 129361. [CrossRef]
77. Granata, F.; Nunno, F.D. Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. *Agric. Water Manag.* **2021**, *255*, 107040. [CrossRef]