


## Article

# Estimation of Spring Maize Evapotranspiration in Semi-Arid Regions of Northeast China Using Machine Learning: An Improved SVR Model Based on PSO and RF Algorithms

Wenjie Hou <sup>1,2</sup> , Guanghua Yin <sup>1,\*</sup>, Jian Gu <sup>1,3,\*</sup> and Ningning Ma <sup>1</sup><sup>1</sup> Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China<sup>3</sup> Tillage and Cultivation Research Institute, Liaoning Academy of Agricultural Science, Shenyang 110116, China

\* Correspondence: ygh006@163.com (G.Y.); sygj981@163.com (J.G.)

**Abstract:** Accurate estimation of crop evapotranspiration ( $ET_c$ ) is crucial for effective irrigation and water management. To achieve this, support vector regression (SVR) was applied to estimate the daily  $ET_c$  of spring maize. Random forest (RF) as a data pre-processing technique was utilized to determine the optimal input variables for the SVR model. Particle swarm optimization (PSO) was employed to optimize the SVR model. This study used data obtained from field experiments conducted between 2017 and 2019, including crop coefficient and daily meteorological data. The performance of the innovative hybrid RF–SVR–PSO model was evaluated against a standalone SVR model, a back-propagation neural network (BPNN) model and a RF model, using different input meteorological variables. The  $ET_c$  values were calculated using the Penman–Monteith equation, which is recommended by the FAO, and used as a reference for the models' estimated values. The results showed that the hybrid RF–SVR–PSO model performed better than all three standalone models for  $ET_c$  estimation of spring maize. The Nash–Sutcliffe efficiency coefficient (NSE), root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination ( $R^2$ ) ranges were 0.956–0.958, 0.275–0.282 mm d<sup>-1</sup>, 0.221–0.231 mm d<sup>-1</sup> and 0.957–0.961, respectively. It is proved that the hybrid RF–SVR–PSO model is appropriate for estimation of daily spring maize  $ET_c$  in semi-arid regions.

**Keywords:** spring maize; crop evapotranspiration; support vector regression; particle swarm optimization; random forest



**Citation:** Hou, W.; Yin, G.; Gu, J.; Ma, N. Estimation of Spring Maize Evapotranspiration in Semi-Arid Regions of Northeast China Using Machine Learning: An Improved SVR Model Based on PSO and RF Algorithms. *Water* **2023**, *15*, 1503. <https://doi.org/10.3390/w15081503>

Academic Editor: Josef Tanny

Received: 22 February 2023

Revised: 7 April 2023

Accepted: 8 April 2023

Published: 12 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, the issue of food security has become increasingly important because of the growing population and the need to maintain sustainable development [1]. Global climate change has resulted in frequent droughts and water shortages, posing a serious threat to food security [2]. In semi-arid regions where rainfall is unevenly distributed and droughts are frequent, water scarcity is particularly severe. Rational irrigation is a viable solution for dealing with drought and reducing water shortages in agriculture [3,4]. Crop evapotranspiration ( $ET_c$ ), which mainly consists of soil surface evaporation and vegetation transpiration, is an integral part of the farmland water balance and hydrological cycle. As a critical indicator in the determination of irrigation regimes, determining crop water requirements is of utmost importance. Therefore, research on  $ET_c$  is crucial for improving agricultural water productivity, conserving irrigation water resources and ensuring food security [5,6].

As a major food crop over the world, maize (*Zea mays* L.) plays a vital role in ensuring food security [7]. China has the largest maize acreage in the world, with more than 41.29 Mha planted and over 260 Mt produced [8]. The western part of Northeast China is

a typical semi-arid agricultural area, and is also one of the key areas for planting spring maize in China [9]. Nevertheless, spring maize is highly sensitive to water stress and its yield can be severely impacted by drought and water deficit [10]. Thus, it is of paramount importance to precisely estimate  $ET_c$  and optimize irrigation water utilization to guarantee the quality and yield of spring maize.

To calculate  $ET_c$ , the crop coefficient ( $K_c$ ) is used in conjunction with reference crop evapotranspiration ( $ET_0$ ) [11]. However, determining  $ET_c$  is challenging due to its dependence on various meteorological variables, soil conditions and crop growth indicators [12,13]. Several empirical models have been developed over the years to estimate daily  $ET_0$ , which can be broadly categorized as temperature-based—such as the Hargreaves–Samani model [14], radiation-based—such as the Priestley–Taylor (P–T) model and Jensen–Haise (H–S) model [15,16], and the principles of energy balance and water vapor diffusion-based Penman–Monteith (PM) equation [11]. Among these empirical models, the PM equation has a wider range of applications and higher estimation accuracy and is recommended by the Food and Agricultural Organization (FAO) for daily  $ET_0$  estimation in various regions [10,17–19]. However, its use is restricted in areas where complete meteorological data are unavailable. Although the P–T and H–S models based on temperature or radiation data can be useful, their estimation accuracy is suboptimal. As a result, it is essential to develop an  $ET_c$  estimation model that requires minimal meteorological data input while still achieving high estimation accuracy.

Recently, with advancements in computer technology and artificial intelligence, machine learning models have been widely used in  $ET_c$  estimation owing to their ability to model complex nonlinear relationships. For example, Saggi and Jain [6] developed an ensemble model consisting of a regularization random forest and hybrid fuzzy–genetic model to estimate  $ET_c$  for maize and wheat, with the results demonstrating superior performance of the ensemble model. In another study by Yamaç [20], the performance of four machine learning models, namely support vector machine (SVM), k-nearest neighbor, random forest (RF) and adaptive boosting models, was compared for sugar beet evapotranspiration estimation under different weather data input conditions. The results demonstrated that the SVM model outperformed the other three models in various conditions. Han, et al. [21] applied the back-propagation neural network (BPNN) model for the  $ET_c$  of wheat, maize and soybean prediction. The BPNN model was verified using eddy correlation measurement of  $ET_c$ , and the results of the BPNN model were found to be satisfactory.

The selection of meteorological input data in these studies was usually determined empirically or by statistical analysis [20,22–25]. However, the performance of machine learning models is highly dependent on the input variables. As a novel machine learning algorithm, RF not only performs regression and classification, but also can be used for variable selection by examining the importance scores assigned to each input variable [26]. The RF algorithm builds a collection of decision trees and aggregates their predictions to improve accuracy and reduce overfitting. By ranking the features based on their importance, RF can help identify the most relevant features for a given task, which can be useful for reducing the dimensionality of the data and improving model performance. Mohammadi and Mehdizadeh [27] compared the accuracy of SVM models that were built with four different data pre-processing methods, including RF, relief, Pearson’s correlation and principal component analysis, in a simulation study of daily  $ET_0$ . The results showed that the prediction accuracy of the model established with meteorological input variables determined by the RF method is higher than that of the other three methods. In another study, Pinos, et al. [28] used the RF method to determine the most important variables of  $ET_0$  as input to the artificial neural network (ANN) model and found that selecting the input variables on the basis of quantification not only reduced the complexity of the model, but also improved its accuracy.

Moreover, the selection of hyperparameters in machine learning models has a significant impact on their accuracy and efficiency. Hyperparameters are settings that are manually adjusted in a machine learning model before it starts learning from data. For

example, one hyperparameter called the “learning rate” determines how quickly the model should adjust its settings during training. Another hyperparameter called the “regularization coefficient” determines how much importance should be given to preventing overfitting. By tuning these hyperparameters, researchers can make their models work better and more effectively. Traditional methods for hyperparameter selection, such as the gradient descent algorithm and grid search algorithm, can be complex and may easily result in local convergence. In recent years, bio-inspired algorithms have gained popularity in the optimization of machine learning models due to their exceptional computational efficiency and ability to find global optimal solutions. Petković, et al. [29] utilized the radial basis function neural network model, which is a type of ANN model, to predict daily  $ET_0$  and optimized the model using particle swarm optimization (PSO). Their results demonstrated that the optimized model outperformed the standalone SVM model in terms of prediction accuracy. Wu, et al. [30] optimized the extreme learning machine (ELM) model for estimating daily  $ET_0$  in various climatic regions of China, using three bio-inspired optimization algorithms, including genetic algorithm (GA), PSO and artificial bee colony (ABC) algorithms. The results demonstrated the effectiveness of bio-inspired heuristic optimization algorithms, particularly the PSO algorithm, in optimizing machine learning models for hydrological applications. In another study, Zhang, et al. [31] used the PSO algorithm to optimize a BPNN model for the prediction of total daily solar radiation and found that the accuracy of the model was significantly enhanced. We also found several other studies that used PSO to optimize machine learning models for other applications, such as flood forecasting, water quality modeling and materials science and engineering [32–34]. These studies further highlight the potential of PSO as a powerful optimization technique for improving the performance of machine learning models in various applications.

A classical machine learning model, support vector regression (SVR), has been implemented in meteorological hydrology owing to its outstanding computational speed and accurate regression prediction for complex high-dimensional data. However, the performance of the radial basis function (RBF) kernel-based SVR model is limited by the optimal choice of input variables and hyperparameters, including regularization coefficient ( $C$ ) and radial basis function (RBF) parameters, which are represented by  $\gamma$  and  $\epsilon$ . [35]. Thus, developing an improved SVR model is crucial to obtain optimal results [29]. PSO is a well-established algorithm for model optimization, while RF has been proven effective in selecting the most relevant input variables for machine learning models in various fields. As such, based on spring maize  $ET_c$  calculated by the FAO method using complete meteorological data and empirical values of  $K_c$ , we established a hybrid RF–SVR–PSO model and compared it with a standalone SVR model, RF model and BPNN model in order to verify the optimization effect of PSO and RF and the estimation accuracy of the hybrid model. The main purposes of this study were to: (1) apply a hybrid RF–SVR–PSO model to estimate the daily  $ET_c$  of spring maize, (2) compare the estimation performance of the hybrid model with the standalone SVR, BPNN and RF models under the input parameters determined by the RF method, (3) recommend the optimal  $ET_c$  estimation model and meteorological input variables for spring maize in the semi-arid region of Northeast China.

## 2. Materials and Methods

### 2.1. Experimental Site and Data Source

The field experiment data were collected in Fumeng County of Northeast China from the spring maize growing seasons between 2017 and 2019. This area, located at  $40^{\circ}41'–42^{\circ}56'$  N,  $121^{\circ}01'–122^{\circ}56'$  E, is known for being a vital producer of spring maize. Throughout the spring maize growth period, the average temperature was  $20.2^{\circ}\text{C}$ , with 169 days of accumulated temperature above  $10^{\circ}\text{C}$ , and the sunshine time in the growth period was 1295.8 h (accounting for 45.2% of the total annual sunshine time). This region is typically semi-arid, experiencing frequent droughts and intense evaporation averaging over 1900 mm per year. Table 1 illustrates the daily meteorological data for the experimental site during the spring maize growth stages in 2017–2019.

**Table 1.** Daily meteorological data for the experimental site during the spring maize growth periods from 2017 to 2019.

Years	Variables	Max	Min	Average	Sd
2017	T <sub>ave</sub> , °C	30.4	9.6	22.2	4.4
	T <sub>max</sub> , °C	40.0	14.9	28.6	4.4
	T <sub>min</sub> , °C	26.2	0.0	16.0	5.9
	n, h d <sup>-1</sup>	13.6	0.0	8.2	3.8
	RH, %	92.0	18.0	59.2	17.0
	U, m s <sup>-1</sup>	6.6	1.1	2.9	1.2
	hP, hPa	1001.8	979.0	989.0	4.4
	Precipitation, mm d <sup>-1</sup>	62.8	9.0	2.0	7.1
2018	T <sub>ave</sub> , °C	31.2	9.7	21.7	4.7
	T <sub>max</sub> , °C	38.1	15.6	27.7	4.4
	T <sub>min</sub> , °C	27.6	1.7	16.2	6.1
	n, h d <sup>-1</sup>	13.3	0.1	7.6	3.8
	RH, %	94.0	18.0	62.6	17.0
	U, m s <sup>-1</sup>	6.4	0.9	3.2	1.2
	hP, hPa	1004.4	977.3	989.6	5.5
	Precipitation, mm d <sup>-1</sup>	48.3	0.0	2.0	6.3
2019	T <sub>ave</sub> , °C	29.8	13.4	22.0	3.7
	T <sub>max</sub> , °C	38.4	19.3	28.0	3.7
	T <sub>min</sub> , °C	26.1	3.1	16.6	4.9
	n, h d <sup>-1</sup>	14.0	0.0	6.9	4.3
	RH, %	96.0	26.0	69.0	16.6
	U, m s <sup>-1</sup>	6.4	1.2	2.8	1.2
	hP, hPa	1004.2	975.8	987.6	5.9
	Precipitation, mm d <sup>-1</sup>	78.3	0.0	4.7	12.4

Note: Max, Min, Average and Sd mean the maximum, minimum, mean and standard deviation of each daily meteorological variable.

The daily meteorological data are sourced from the National Meteorological Information Center of the China Meteorological Administration (<https://data.cma.cn> (accessed on 22 August 2022)). The daily meteorological data included average, minimum and maximum temperature (T<sub>ave</sub>, T<sub>min</sub> and T<sub>max</sub>, °C), precipitation (mm), wind speed at 2 m height (U, m s<sup>-1</sup>), daily duration of sunshine (n, h), average relative humidity (RH, %) and average vapor pressure (hP, hPa). The Fumeng County meteorological station is a national ecological and agricultural meteorological observation station. The ground observation is responsible for 24 h monitoring and automatic uploading of station meteorological data at 10 min intervals to participate in global sharing.

The maize varieties were “Yufeng 303” and “Zhengdan 958”, which were used alternately between years. The maize was planted using both wide and narrow rows, with row widths of 60 cm and 40 cm, respectively. The planting density was 60 thousand plants ha<sup>-1</sup>, and shallowly buried drip irrigation tape was used for sowing via an integrated machine. The irrigation method employed was shallow buried drip irrigation, with the embedded drip irrigation tape only laid in the center of the narrow row, approximately 3–5 cm beneath the surface. For further information on the field experiment, please refer to Wang, et al. [36].

## 2.2. Maize Crop Evapotranspiration Calculation

The daily ET<sub>0</sub> was calculated by the FAO56 Penman–Monteith equation [11]

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

where ET<sub>0</sub> is the reference evapotranspiration (mm/d), R<sub>n</sub> is the net radiation (MJ m<sup>-2</sup> d<sup>-1</sup>), G is the soil heat flux density (MJ m<sup>-2</sup> d<sup>-1</sup>), γ is the psychrometric constant, T is the mean

air temperature ( $^{\circ}\text{C}$ ),  $u_2$  is the daily wind speed at 2 m height ( $\text{m s}^{-1}$ ),  $e_s$  is the saturation vapor pressure (kPa),  $e_a$  is the actual vapor pressure (kPa) and  $\Delta$  is the slope of the saturation vapor pressure–temperature curve ( $\text{kPa } ^{\circ}\text{C}^{-1}$ ).

In this study, the computation period was 24 h, the surface was covered by vegetation and the amount of heat transferred through the soil surface was generally small compared to  $R_n$ . Therefore, the soil heat flux was considered negligible according to the recommendation of FAO56 [11]. However, it is important to note that under certain conditions, such as those of bare surfaces or shorter computation periods, the soil heat flux can become more significant and should be measured or estimated using appropriate methods. According to the Ångström–Prescott (A–P) formula (Equation (2)), the daily solar radiation ( $R_s$ ) can be estimated using extraterrestrial radiation ( $R_a$ ), the actual daily duration of sunshine ( $n$ ) and the maximum possible duration of sunshine ( $N$ ) [11,37].

$$\frac{R_s}{R_a} = a + b \frac{n}{N} \quad (2)$$

$$N = 24 \times \omega_s / \pi \quad (3)$$

$$R_a = (24 \times 60 / \pi) G_{sc} d_r (\omega_s \sin \varphi \sin \delta + \cos \varphi \cos \delta \sin \omega_s) \quad (4)$$

$$d_r = 1 + 0.033 \cos(2\pi \times J/365) \quad (5)$$

$$\delta = 0.409 \sin(2\pi \times J/365 - 1.39) \quad (6)$$

$$\omega_s = \arccos(-\tan \varphi \tan \delta) \quad (7)$$

where  $a$  and  $b$  are regression constants with FAO56-recommended values of 0.25 and 0.5;  $\omega_s$  is the sunset hour angle (rad);  $G_{sc}$  is the solar constant with a value of  $0.0820 \text{ MJ m}^{-2}$ ;  $J$  is the day of the year;  $d_r$  is the inverse relative earth–sun distance;  $\delta$  is the solar declination (rad);  $\varphi$  is the latitude (rad) [30].

The spring maize  $ET_c$  was calculated by multiplying  $ET_0$  and  $K_c$ :

$$ET_c = K_c ET_0 \quad (8)$$

where  $ET_0$  is the reference evapotranspiration ( $\text{mm d}^{-1}$ ) calculated by Equation (1) and  $K_c$  is the crop coefficient.

During the growing seasons of spring maize from 2017 to 2019, the growing stages were separated into sowing to jointing, jointing to tasseling, tasseling to filling, and filling to maturity stages, according to the phenological characteristics of spring maize, corresponding to initial, crop development, mid-season, and late-season stages respectively [11]. The meteorological data and crop coefficient values collected during spring maize growing stages from 2017 to 2018 were used to train the machine learning model, while meteorological data and crop coefficient values collected during spring maize growing stages from 2019 were used to test the machine learning model (Table 2). The spring maize  $K_c$  values of initial, mid-season and late-season stages suggested by FAO56 are 0.3, 1.2 and 0.6. The  $K_c$  values of mid-season and late-season stages were adjusted due to the actual conditions of the experimental site according to the following equations:

$$K_{c-mid} = K_{c-mid-FAO56} + [0.04(u_2 - 2) - 0.004(RH_{min} - 45)] \left(\frac{h}{3}\right)^{0.3} \quad (9)$$

$$K_{c-end} = K_{c-end-FAO56} + [0.04(u_2 - 2) - 0.004(RH_{min} - 45)] \left(\frac{h}{3}\right)^{0.3} \quad (10)$$

where  $K_{c-mid-FAO56} = 1.2$  and  $K_{c-end-FAO56} = 0.6$  are the mid- and late-season crop coefficients for spring maize,  $h$  is the mean height of the spring maize during the growth stage,  $RH_{min}$  is the mean value for daily minimum relative humidity (%) during the growth stage, and  $u_2$  is the daily average wind speed ( $m\ s^{-1}$ ) at 2 m above ground during the growth stage from the experiment site [11].

**Table 2.** The stages of spring maize growing seasons between 2017 and 2019.

Crop Growth Stages	Training Period		Testing Period
	2017	2018	2019
Initial	1 May–31 May (31 d)	28 April–26 May (29 d)	14 May–10 June (28 d)
Crop development	1 June–20 July (50 d)	27 May–15 July (50 d)	11 June–28 July (48 d)
Mid-season	21 July–1 September (43 d)	16 July–30 August (46 d)	29 July–7 September (41 d)
Late-season	2 September–27 September (26 d)	31 August–4 October (35 d)	8 September–28 September (21 d)
total days (d)	150	160	138

### 2.3. Support Vector Regression

The support vector machine algorithm proposed by Vapnik [38] is a powerful supervised machine learning model based on mathematical–statistical theory. Building on Vapnik’s concept, Drucker, et al. [39] further developed the support vector regression (SVR) technique for solving the regression problem. The SVR model has an advantage in that it is based on a series of kernel functions that are independent of the dimension of the input space, allowing for effective modeling of nonlinear relationships in higher-dimensional feature space [23]. As a result, the SVR model has been widely used in various fields, including hydrology, agriculture and meteorology, for prediction and estimation. As mentioned before, the regression ability of the SVR model depends on the kernel function; therefore, the choice of the kernel is crucial to the construction of the SVR model, and the performance of different kernels varies. Some previous studies [40–43] have shown that the RBF (radial basis function) kernel outperforms other kernels, and it was, thus, used in this study. The SVR model’s hyperparameters, including  $C$ ,  $\gamma$  and  $\varepsilon$ , play a critical role in determining the trade-off between the model’s accuracy and complexity, ultimately affecting its performance. In this study, the R package “e1071” [44], an open-source software package, was used to construct the SVR model for the estimation of the spring maize daily  $ET_c$ . The details about the mathematical–statistical theory of the SVR model can be found in the Supplementary Material.

### 2.4. Particle Swarm Optimization Algorithm

The particle swarm optimization (PSO) algorithm is a stochastic heuristic optimization algorithm that simulates the foraging behavior of bird populations, developed by Kennedy, et al. [45], and is applied to solving function optimization problems. The PSO algorithm has excellent global search and optimization abilities, and has been widely used in the parameter optimization of various machine learning models [46]. The PSO algorithm forms a swarm of particles, where each particle represents a potential solution in the solution space of the optimization problem [30]. A fitness function is defined to evaluate the fitness value of each particle, and the PSO algorithm constantly changes the flight direction and velocity of each particle; then, the algorithm constantly iterates to obtain the optimal value of the fitness function to search for the optimal solution to the actual problem. In this study, the PSO algorithm is implemented by using the R package “pso” [47] to optimize the SVR model. More detailed information about the computation procedure of the PSO algorithm can be found in the Supplementary Material.

The PSO algorithm can be summarized in the following steps:

Step 1: Define the fitness function and population size  $n$ , and randomly initialize the velocity and position of each particle;

Step 2: Evaluate the fitness value of each particle using the fitness function;

Step 3: Update the personal best position for each particle based on its own best fitness value;

Step 4: Update the global best position for all particles;

Step 5: Update the velocity and position for each particle using the personal best and global best velocity and position;

Step 6: Judge whether the maximum number of iterations has been reached, or if the fitness value meets the requirements. If so, terminate the iteration and output the result. Otherwise, return to Step 2.

### 2.5. Random Forest

The RF model was utilized both to determine the input variables for all machine learning models and as a standalone regression model to estimate the daily  $ET_c$  of spring maize. The random forest (RF) method is a widely used tree-based machine learning algorithm for constructing classification and regression models [26]. It is an extended variant of Bagging [48], which employs decision trees as basic learners and introduces random attribute selection into the training process. The RF model has been applied to the estimation of  $ET_0$  in many studies [23,27,49]. In addition, the RF method possesses exceptional capability in determining the importance of variables [28]. To ascertain the importance of input variables, all meteorological variables and crop coefficient values from the training period were employed as the input training data to construct each tree. In the tree generation, a random bootstrap sampling of each point of input training data was conducted, resulting in approximately 37% of the input training data being excluded from tree generation and classified as out-of-bag (OOB) observations [50]. The RF model determined the importance of each input variable by measuring the mean decrease in prediction accuracy when samples of a variable in the OOB dataset were randomly permuted [51]. In this study, the RF was constructed using the R package “rfPermute” [52].

### 2.6. Back-Propagation Neural Network

The back-propagation neural network (BPNN) is a multilayer feedforward neural network that uses the error back-propagation training method to train itself. The algorithm used for training is known as the BP algorithm, and employs gradient descent techniques to adjust the connection weights and thresholds of each layer in order to minimize the difference between the actual and desired output values of the network [53]. Hornik, et al. [54] proved that the BP neural network is capable of simulating continuous functions of arbitrary complexity with just one hidden layer containing enough neurons. In light of this, R package “neuralnet” [55] was utilized in this study to develop a standalone BP neural network model for estimating the daily  $ET_c$  of spring maize, and its estimation accuracy was compared with the hybrid model.

### 2.7. Hybrid Model Building

The hybrid model combines random forest, support vector regression and particle swarm optimization algorithms. The hybrid model consists of three parts, the first part involves data selection and pre-processing, in which variable importance is ranked using the random forest method. Subsequently, input data for the SVR model are selected based on variable importance. To enhance computational efficiency, convergence accuracy and estimation precision, it was necessary to normalize the input variables of the model due to their different dimensions. This was achieved by normalizing the variables according to Equation (11).

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (11)$$

where  $X_{\min}$  and  $X_{\max}$  are the minimum and maximum of all input variable values;  $X$  is the measured values of all the input variables;  $X'$  is the normalized values of measured values.

The second part involves the model simulation, which begins with initialization of the hyperparameters of the SVR model, including regulation coefficient (C) and RBF

parameters  $\gamma$  and  $\varepsilon$ . Additionally, the optimization accuracy requirement is defined, after which the model estimates  $ET_c$  and computes the corresponding error.

If the desired target is not achieved, the model proceeds to the third part, which involves optimizing the hyperparameters of the SVR model using the PSO algorithm. The SVR hyperparameters with the highest fitness are returned, thereby achieving the objective of optimizing the SVR model and improving the accuracy of  $ET_c$  estimation. R version 4.1.1 [56] was used to build and implement crop evapotranspiration estimation models, and the structure of the hybrid RF–SVR–PSO model is illustrated in Figure 1.

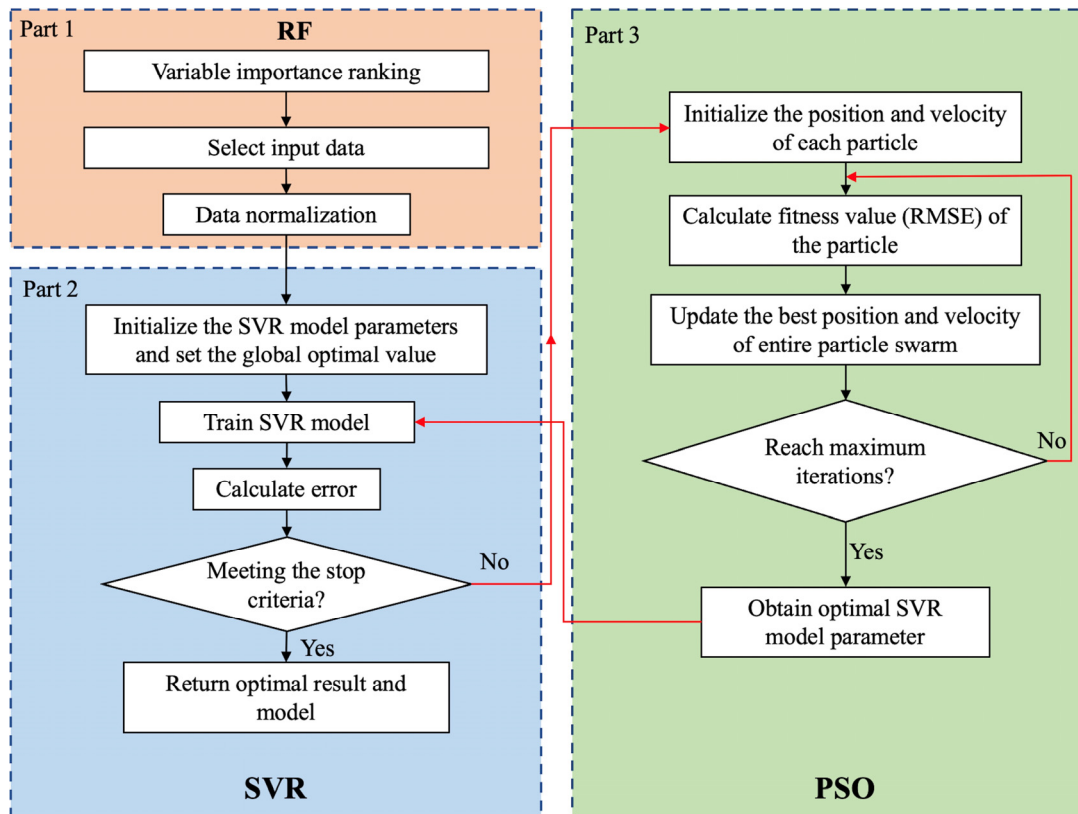


Figure 1. The flowchart of hybrid RF–SVR–PSO model.

### 2.8. Evaluation Criteria of Model Performance

Model performance was assessed by using the root mean square error (RMSE), mean absolute error (MAE), coefficient of determination ( $R^2$ ) and Nash–Sutcliffe efficiency coefficient (NSE) [57]. These statistical criteria were calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \tag{12}$$

$$MAE = \left| \frac{\sum_{i=1}^n (O_i - P_i)}{n} \right| \tag{13}$$

$$R^2 = \frac{[\sum_{i=1}^n (O_i - \bar{O}_i)(P_i - \bar{P}_i)]^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2 \sum_{i=1}^n (P_i - \bar{P}_i)^2} \tag{14}$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2} \tag{15}$$

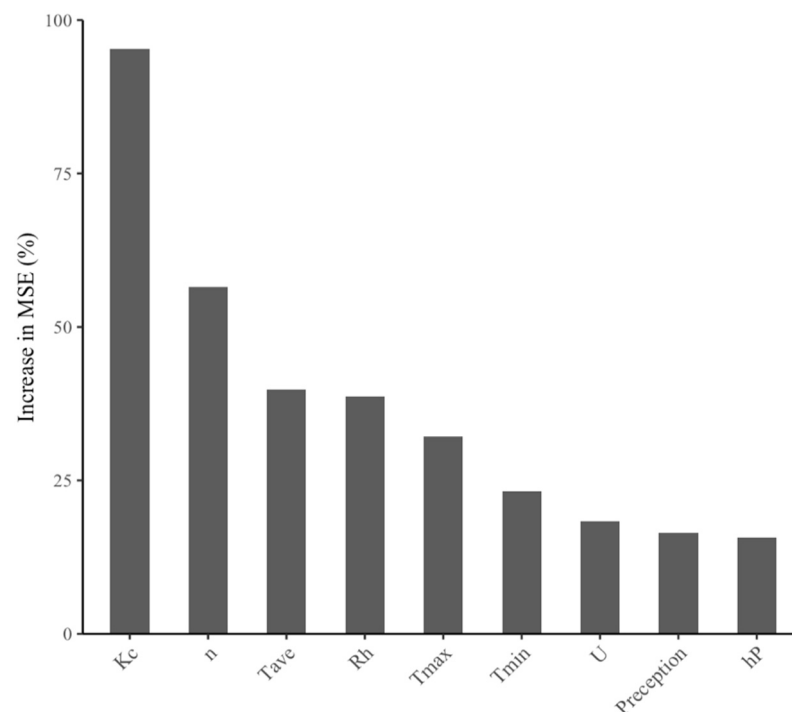


where  $n$  is the sample number,  $O_i$  denotes the observed values and  $P_i$  denotes the estimated values.  $\bar{O}_i$  and  $\bar{P}_i$  are the mean values of the observed and estimated values.

### 3. Results

#### 3.1. The Variables for Determining Crop Evapotranspiration

To identify the principal variables affecting  $ET_c$ , we fitted the model using all nine variables, including eight meteorological variables and  $K_c$ , based on RF analysis. The ranking of importance of the variables showed that  $K_c$  had the greatest impact on  $ET_c$ , reaching 95.23%, followed by  $n$ ,  $T_{ave}$ ,  $RH$ ,  $T_{max}$ ,  $T_{min}$ ,  $U$ ,  $hP$  and Precipitation (Figure 2). This result shows that the most important factors affecting  $ET_c$ , in addition to crop coefficients, are sunshine hours, temperature, and relative humidity, which is similar to what was found in the study conducted by Pinos, Chacón and Feyen [28]. Therefore, we added  $K_c$  to all machine learning models. We then added the three, four and five variables with the highest importance outside of  $K_c$  as input variables to the machine learning model.



**Figure 2.** Plot of variable importance ranking, where variable importance is expressed as the percentage increase in mean squared error (%IncMSE). Each value represents the increase in prediction error of the same model after a variable is omitted.

The stages of the spring maize growing seasons from 2017 to 2019 are shown in Table 2. The growth stages of spring maize were 150 d in 2017, 160 d in 2018 and 138 d in 2019. The lengths of the four major growth stages (initial stage, crop development stage, mid-season stage and late-season stage) were 31, 50, 43 and 26 d in 2017; 29, 50, 46 and 35 d in 2018; and 28, 48, 41 and 21 d in 2019. The values of  $K_{c-mid}$  and  $K_{c-end}$  needed to be calibrated by the actual crop height  $RH_{min}$  and  $u_2$  in mid- and late-season crop growth conditions and climatic conditions. The mid- and late-season crop growth stages' heights were the 3-year average heights of 2.3 m and 2.8 m, respectively. The calibrated  $K_{c-mid}$  and  $K_{c-end}$  values according to Equations 9 and 10 were 1.19 and 0.6 in 2017; 1.22 and 0.65 in 2018; and 1.14 and 0.63 in 2019, close to the maize crop coefficients recommended by Ji, et al. [58].

#### 3.2. Performance Assessment

Table 3 presents the performance metrics of the four machine learning models, including the hybrid RF-SVR-PSO model and the three standalone models, evaluated for

different input variable combinations determined by the RF method during both training and testing periods. During the training and testing periods, all machine learning models were able to produce accurate estimates of spring maize daily  $ET_c$  with NSE, RMSE, MAE and  $R^2$  ranging from 0.858–0.986, 0.206–0.508 mm d<sup>-1</sup>, 0.152–0.426 mm d<sup>-1</sup> and 0.915–0.989, respectively.

**Table 3.** Statistical performance of hybrid RF–SVR–PSO model, standalone SVR model, BPNN model and RF model with three different variables input for training and testing periods.

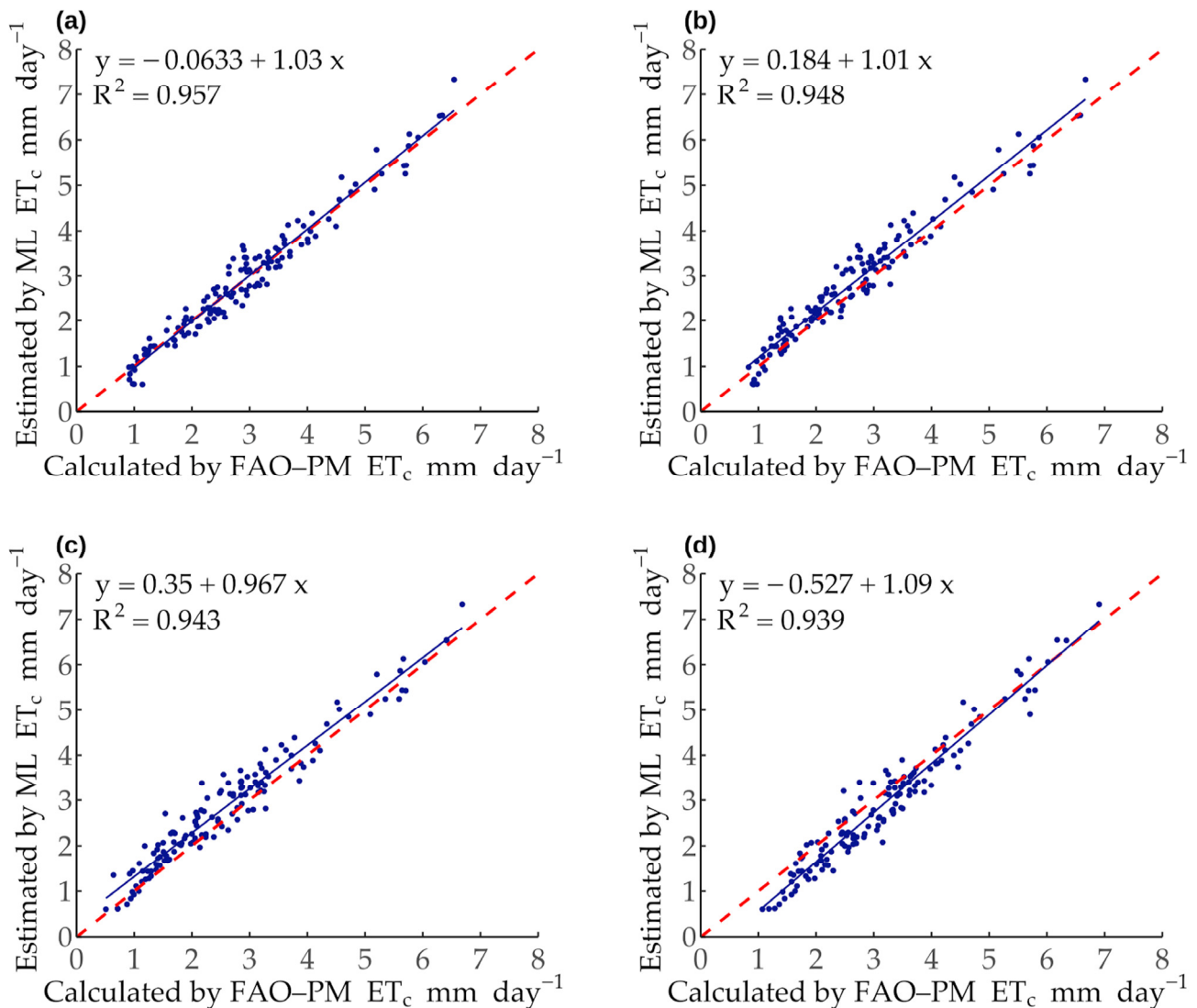
Input/Model	Training Periods				Testing Periods			
	$R^2$	RMSE (mm d <sup>-1</sup> )	MAE (mm d <sup>-1</sup> )	NSE	$R^2$	RMSE (mm d <sup>-1</sup> )	MAE (mm d <sup>-1</sup> )	NSE
$K_c, T_{ave}, n, RH$								
RF–PSO–SVR	0.970	0.396	0.329	0.949	0.957	0.282	0.231	0.956
SVR	0.979	0.252	0.194	0.979	0.948	0.365	0.289	0.927
BPNN	0.976	0.271	0.211	0.976	0.943	0.418	0.333	0.904
RF	0.989	0.234	0.169	0.983	0.939	0.434	0.369	0.897
$K_c, T_{ave}, T_{max}, n, RH$								
RF–PSO–SVR	0.970	0.366	0.297	0.956	0.959	0.278	0.225	0.958
SVR	0.985	0.216	0.165	0.985	0.957	0.320	0.263	0.944
BPNN	0.984	0.221	0.169	0.984	0.960	0.341	0.277	0.936
RF	0.988	0.238	0.177	0.982	0.915	0.508	0.426	0.858
$K_c, T_{ave}, T_{max}, T_{min}, n, RH$								
RF–PSO–SVR	0.965	0.388	0.319	0.951	0.961	0.275	0.221	0.958
SVR	0.986	0.210	0.162	0.986	0.948	0.340	0.281	0.936
BPNN	0.986	0.209	0.158	0.986	0.955	0.341	0.279	0.936
RF	0.988	0.206	0.152	0.982	0.918	0.486	0.409	0.870

Comparing the performance indexes of the three standalone machine learning models in the test period, the SVR model outperforms the BPNN and RF models in terms of RMSE, MAE and NSE, with the  $R^2$  being slightly smaller than that of the BPNN model for four and five input variables, but significantly higher than that of the RF model. The RF model performed the poorest, with  $R^2$ , RMSE, MAE and NSE ranging from 0.915–0.939, 0.434–0.508 mm d<sup>-1</sup>, 0.369–0.409 mm d<sup>-1</sup> and 0.858–0.897, respectively. The SVR model performed best when the input variables were  $K_c, T_{ave}, T_{min}, n$  and RH, with  $R^2$ , RMSE, MAE and NSE values of 0.957, 0.320 mm d<sup>-1</sup>, 0.263 mm d<sup>-1</sup> and 0.944.

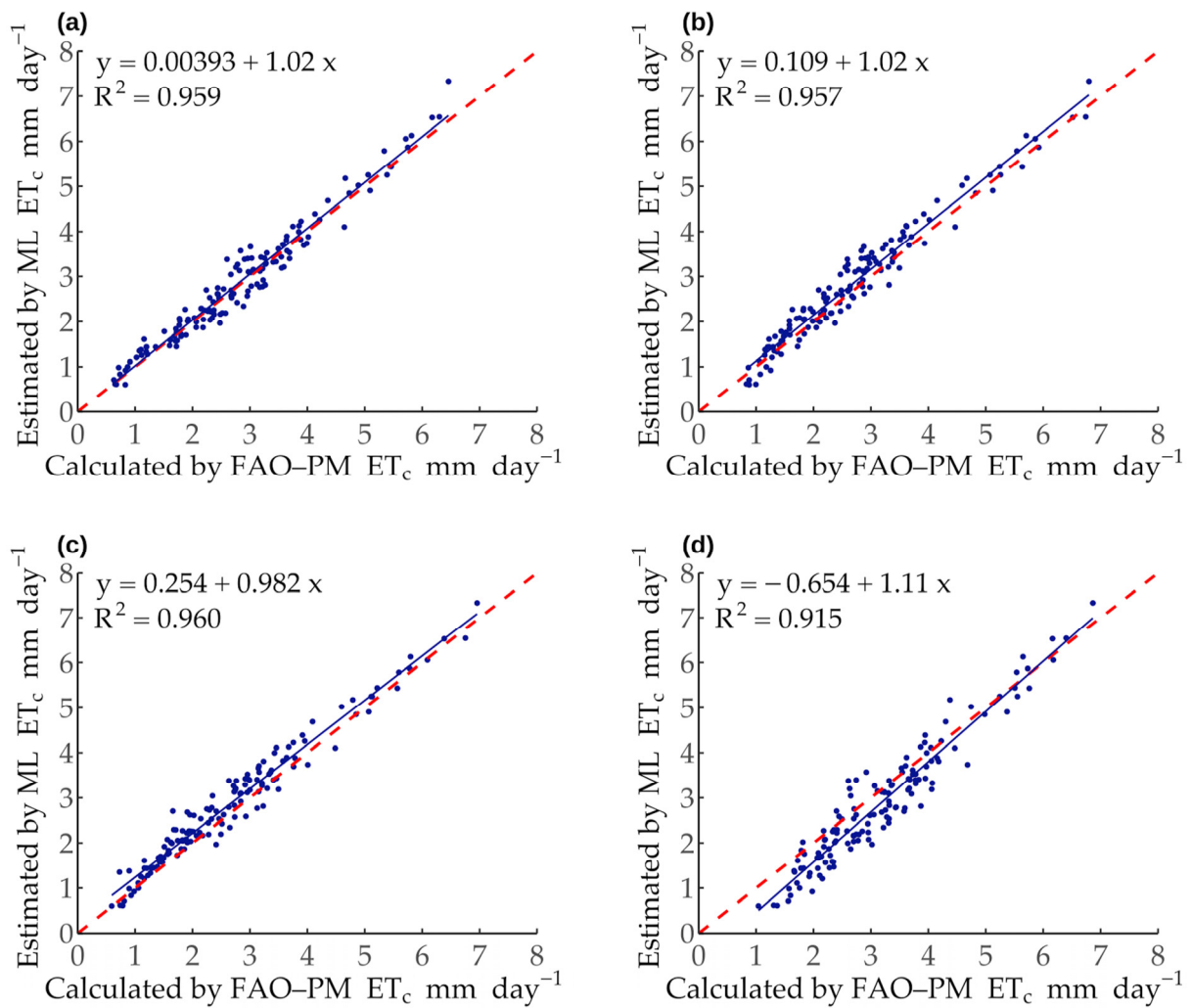
Considering the need for optimization, we used the PSO algorithm to optimize the SVR model and enhance the accuracy of the models and subsequently established the RF–SVR–PSO hybrid model accordingly. For the hybrid model, we set the parameters of the PSO algorithm as follows: maximum number of iterations: 50, swarm size: 20, tolerance error: 0.01, inertia weights: 0.8, learning factor: 1.2. The initial particle velocity and position were randomly assigned values between intervals [0.01,0.01,0.1] and [0.1,0.5,8], representing the values of hyperparameters  $\gamma, \epsilon$  and C of SVR model, respectively. Table 3 presents the performance metric values of the hybrid model for estimating spring maize evapotranspiration during the test period. The hybrid RF–SVR–PSO model provided better estimates of spring maize daily  $ET_c$  for all three combinations of input variables compared to the standalone SVR model, with  $R^2$ , RMSE, MAE and NSE values of 0.957–0.961, 0.275–0.282 mm d<sup>-1</sup>, 0.221–0.231 mm d<sup>-1</sup> and 0.956–0.958, respectively. As the number of input variables increased, the model performance and the accuracy of estimated  $ET_c$  values was only slightly improved.

For the test period, taking the  $ET_c$  values using the PM–FAO equation as the target values, estimated  $ET_c$  values of the RF–PSO–SVR model, SVR model, BPNN model and RF model under different combinations of input variables were compared in the form of scatter and hydrograph plots, as shown in Figures 3–8. The scatter plots clearly depict that the scatter distribution of the hybrid RF–SVR–PSO model are more evenly concentrated on the ideal line (i.e., 1:1 line) than the other three standalone models, and the fit of the hybrid

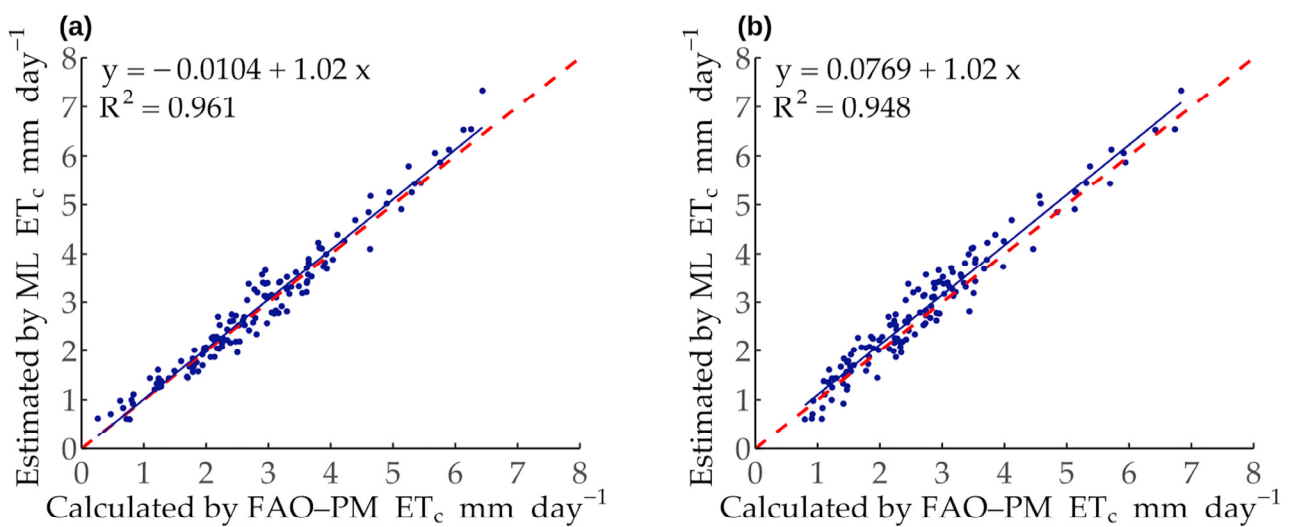
RF-SVR-PSO model improves slightly as the number of input variables increase. It was also evident from the scatter plot that the trend lines for the three independent models were generally above or below the ideal line (i.e., 1:1 line), with varying degrees of overestimation or underestimation of the  $ET_c$  estimates compared to the FAO-PM calculations. Compared with the calculated value of FAO-PM,  $ET_c$  estimates of the three standalone models are overestimated or underestimated to vary degrees. Further, the accuracy of the standalone models decreased when the input variables were  $K_c$ ,  $T_{ave}$ ,  $T_{max}$ ,  $T_{min}$ ,  $n$  and  $RH$ . As shown in the hydrograph plots, the hybrid RF-PSO-SVR model outperformed the standalone models, both in capturing the peaks and in terms of the overall individual values estimated.



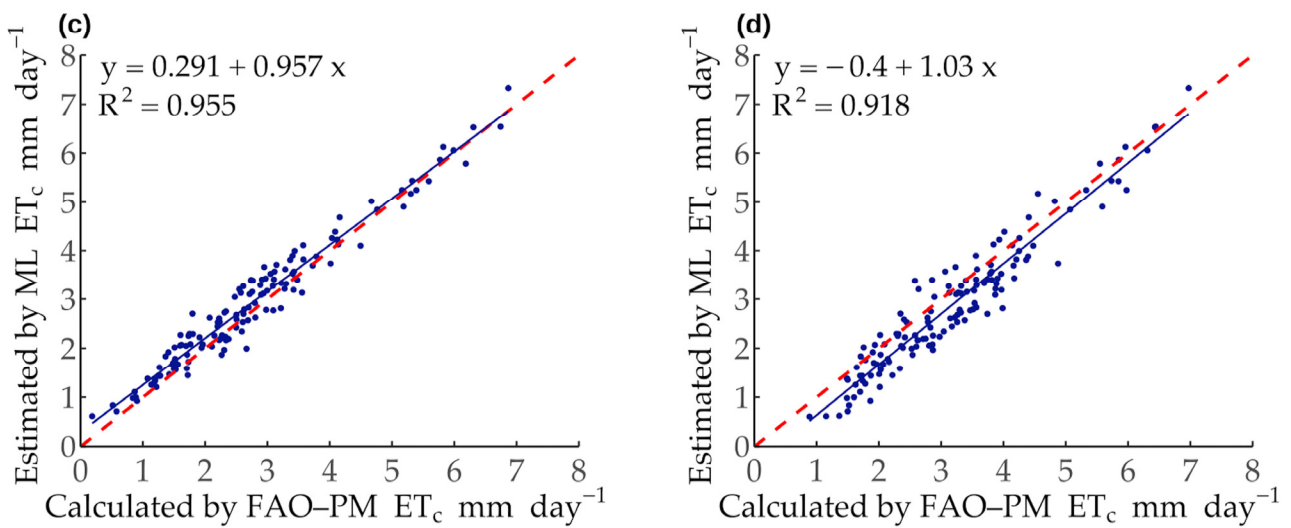
**Figure 3.** Scatter plots of calculated daily spring maize  $ET_c$  values by FAO-PM method compared with ML model-estimated values with  $K_c$ ,  $T_{ave}$ ,  $n$  and  $RH$  as the input. RF-SVR-PSO model (a), SVR model (b), BPNN model (c) and RF model (d).



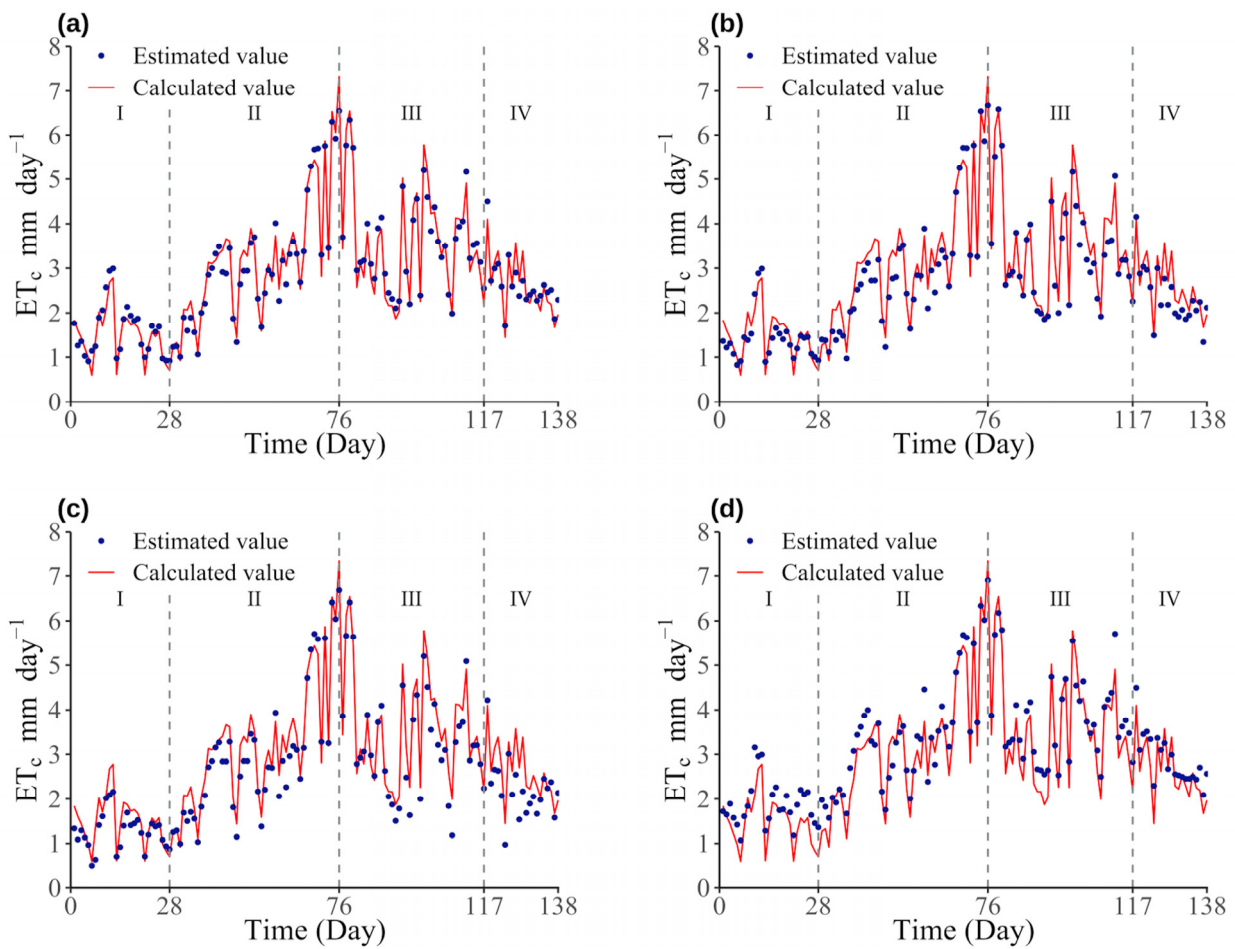
**Figure 4.** Scatter plots of calculated daily spring maize  $ET_c$  values by FAO-PM method compared with ML model-estimated values with  $K_c$ ,  $T_{ave}$ ,  $T_{max}$ ,  $n$  and  $RH$  as the input. RF-SVR-PSO model (a), SVR model (b), BPNN model (c) and RF model (d).



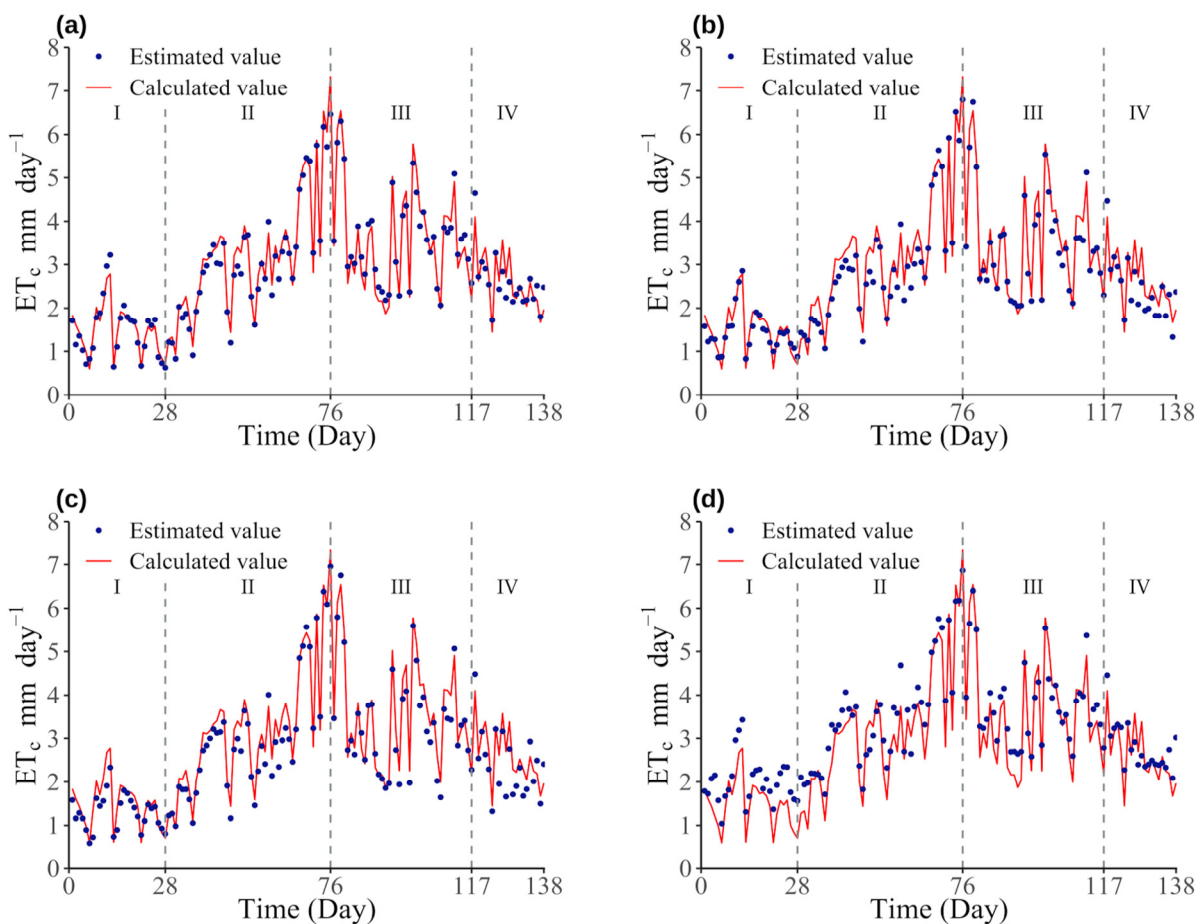
**Figure 5.** Cont.



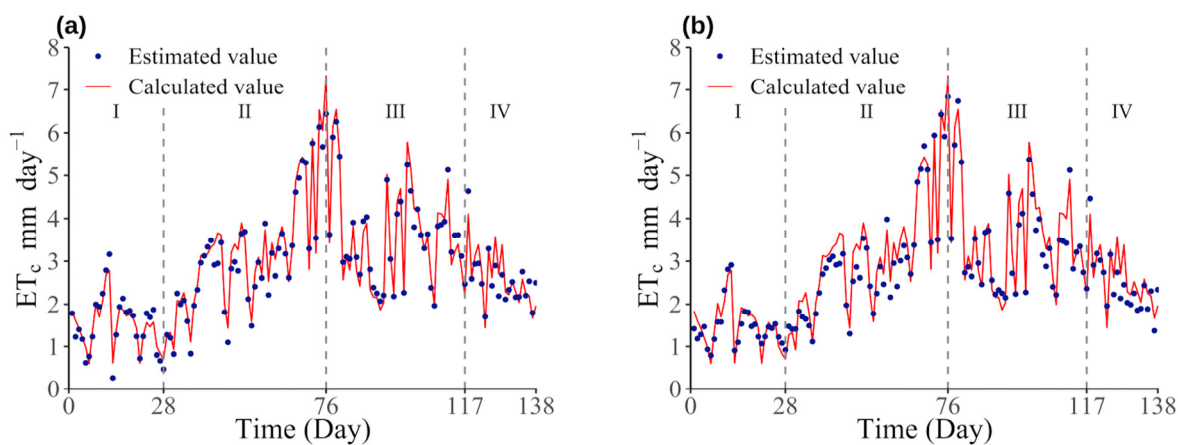
**Figure 5.** Scatter plots of calculated daily spring maize  $ET_c$  values by FAO-PM method compared with ML model-estimated values with  $K_c$ ,  $T_{ave}$ ,  $T_{max}$ ,  $T_{min}$ ,  $n$  and RH as the input. RF-SVR-PSO model (a), SVR model (b), BPNN model (c) and RF model (d).



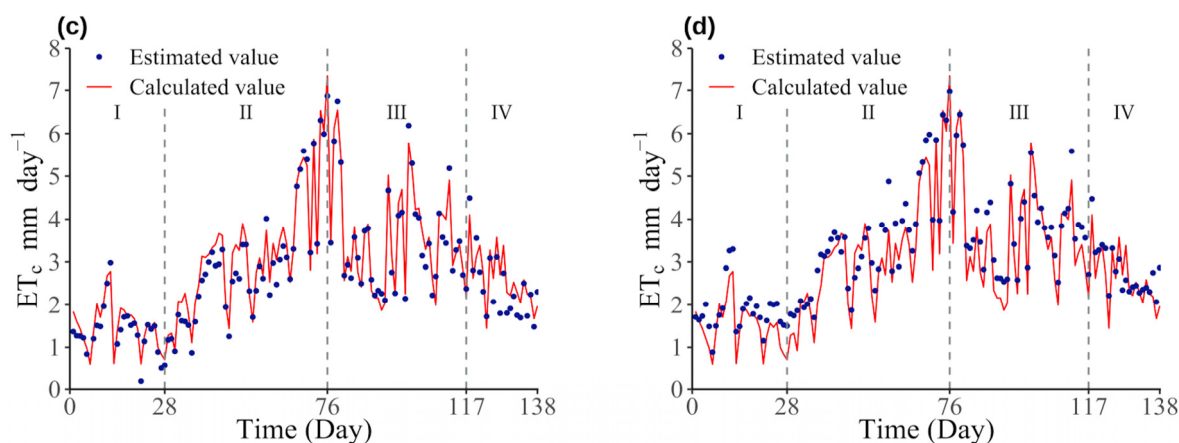
**Figure 6.** Comparison of simulated and calculated values of spring maize  $ET_c$  in 2019 of different models with  $K_c$ ,  $T_{ave}$ ,  $n$  and RH as the input: RF-SVR-PSO model (a), SVR model (b), BPNN model (c) and RF model (d). The areas labeled I, II, III and IV indicate the initial stage, crop development stage, mid-season stage and late-season stage, respectively.



**Figure 7.** Comparison of simulated and calculated values of spring maize  $ET_c$  in 2019 of different models with  $K_c$ ,  $T_{ave}$ ,  $T_{max}$ ,  $n$  and  $RH$  as the input: RF-SVR-PSO model (a), SVR model (b), BPNN model (c) and RF model (d). The areas labeled I, II, III and IV indicate the initial stage, crop development stage, mid-season stage and late-season stage, respectively.



**Figure 8.** Cont.



**Figure 8.** Comparison of simulated and calculated values of spring maize  $ET_c$  in 2019 of different models with  $K_c$ ,  $T_{ave}$ ,  $T_{max}$ ,  $T_{min}$ ,  $n$  and RH as the input: RF–SVR–PSO model (a), SVR model (b), BPNN model (c) and RF model (d). The areas labeled I, II, III and IV indicate the initial stage, crop development stage, mid-season stage and late-season stage, respectively.

#### 4. Discussion

In this study, it was found that, taking RMSE, MAE and NSE as the performance evaluation indexes of the model, the daily  $ET_c$  estimation accuracy of the standalone SVR model was higher than that of the standalone BPNN and RF models. The study demonstrated the superiority of the SVR model in handling the complex nonlinear relationship between  $ET_c$  and meteorological variables, and its high accuracy and computational efficiency in estimating  $ET_c$  [22,27,37].

The performance of machine learning models can be improved using bio-inspired optimization algorithms. Many studies have shown that optimization algorithms can effectively enhance the performance of machine learning models. For instance, Zhu, et al. [59] proposed a hybrid PSO–ELM model to estimate daily  $ET_0$ , which showed a 13% lower RMSE than the standalone ELM model. Jia, et al. [60] used the sparrow search algorithm (SSA) to optimize the ELM model, resulting in a significant improvement in the model's performance. Wen and Yuan [61] optimized the BPNN model for  $CO_2$  emissions forecasting using the PSO algorithm, and the results indicated a positive effect on optimization. Given the simplicity of the PSO algorithm and its good optimization results, this study utilized the PSO algorithm to determine and optimize the hyperparameters ( $C$ ,  $\gamma$  and  $\epsilon$ ) of the SVR model, resulting in the development of a hybrid RF–SVR–PSO model for spring maize daily  $ET_c$  estimation. The RMSE and MAE of the hybrid RF–SVR–PSO model computed with three different meteorological input variables decreased by 13.2% to 22.8% and 14.6% to 21.2%, and NSE improved by 1.5% to 3.1% compared to the standalone SVR model (testing period). While it may seem that all machine learning models perform well, the PSO algorithm led to substantial improvements in the accuracy of  $ET_c$  estimates compared to the standalone machine learning models. Specifically, the PSO algorithm significantly reduced the overestimation and underestimation of  $ET_c$  estimates by standalone machine learning models, which is critical for guiding actual maize production practices. Spring maize is particularly sensitive to water stress, and when the model underestimates  $ET_c$ , the recommended irrigation amount may be lower than the amount of water required for maize production, resulting in reduced maize yields and affecting food security. Conversely, if the model overestimates  $ET_c$ , the recommended irrigation amount may be higher than the amount of water required for maize production, resulting in wasted water and reduced water productivity.

The choice of appropriate meteorological input variables for ML models plays a crucial role in the accuracy of estimating spring maize daily  $ET_c$  [25]. The selection of suitable inputs for the ML models can effectively improve the accuracy of the results. This study used the RF method as a data pre-processing method to determine the importance of the

meteorological input variables and to identify suitable inputs for the ML model. The RF method ranked the importance of estimating spring maize daily  $ET_c$  variables in the study area. By taking the top four, five and six variables of highest importance as the input of the ML models, we determined the optimal input of the model. With an increase in the number of input meteorological variables, the accuracy of the hybrid RF–SVR–PSO model in estimating the spring maize daily  $ET_c$  was only slightly improved. Although the accuracy of the hybrid RF–SVR–PSO model improved only slightly with an increase in the number of input variables, considering the computational efficiency and estimation accuracy, the input variables  $K_c$ ,  $n$ ,  $T_{ave}$  and RH were determined as the best combination for the hybrid RF–SVR–PSO model, with  $R^2$ , RMSE, MAE and NSE values of 0.957, 0.282, 0.231 and 0.956, respectively.

The results of the proposed hybrid RF–SVR–PSO model’s performance indices for estimating  $ET_c$  were compared with those of other approaches and are presented in Table 4 for the testing period. Jia et al. [60] proposed a hybrid SSA–ELM model using  $T_{max}$ ,  $T_{min}$ ,  $n$ , maize leaf area index (GLAI) and plant height (h) as the input variables to estimate the spring maize  $ET_c$ , and obtained RMSE, MAE and  $R^2$  values of 0.433 mm d<sup>−1</sup>, 0.342 mm day<sup>−1</sup> and 0.895, respectively. In a study conducted by Yamaç [20], RMSE, MAE and  $R^2$  values were reported for the adaptive boosting (AB) and SVM models using  $K_c$ ,  $T_{max}$ ,  $T_{min}$ , RH and U as the input variables as 0.954 mm day<sup>−1</sup>, 0.688 mm day<sup>−1</sup> and 0.856, and 0.699 mm day<sup>−1</sup>, 0.557 mm day<sup>−1</sup> and 0.923, respectively. Thus, it can be seen that the hybrid RF–SVR–PSO has a high accuracy for spring maize  $ET_c$  estimation. Furthermore, the hybrid model can be used in semi-arid regions as an alternative to the widely used FAO56-recommended approach, the PM equation, to obtain satisfactory  $ET_c$  estimation. However, as a machine learning model, the hybrid RF–SVR–PSO operates as a black box, and its parameters must be re-determined for use in different locations with varying meteorological conditions. Additionally, in areas where eddy covariance or lysimeter data are obtainable, utilizing such measurement data as the reference could avoid the “double bias” caused by using FAO–PM approach as a reference.

**Table 4.** Performance indices for  $ET_c$  estimation using the hybrid model and different approaches.

Model	Input	Performance Indicator			Reference
		RMSE (mm day <sup>−1</sup> )	MAE (mm day <sup>−1</sup> )	$R^2$	
RF–SVR–PSO	$K_c$ , $T_{ave}$ , $n$ , RH	0.282	0.231	0.957	
SSA–ELM	$T_{max}$ , $T_{min}$ , $n$ , GLAI, h	0.433	0.342	0.895	Jia et al. [60]
AB	$K_c$ , $T_{max}$ , $T_{min}$ , RH, U	0.954	0.688	0.856	Yamaç [20]
SVM	$K_c$ , $T_{max}$ , $T_{min}$ , RH, U	0.699	0.557	0.923	Yamaç [20]

The proposed novel hybrid RF–SVR–PSO model can provide systematic support for spring maize  $ET_c$  estimation and irrigation management in semi-arid regions. In this study, the training and testing datasets were divided only by a simple hold-out method, which may result in a reduced generalization ability. Therefore, in the forthcoming study, the hybrid RF–SVR–PSO model can be combined with the K-fold cross-validation method to improve the estimation accuracy, generalization and robustness of the model.

## 5. Conclusions

In the hybrid model proposed in this paper, the RF model was used to rank the importance of variables, and determine the input variables of the hybrid model, and the PSO algorithm was used to enhance the estimation performance of crop daily evapotranspiration of the SVR model. The performance of the hybrid RF–SVR–PSO model was compared with three standalone models (including the SVR model, BPNN model and RF model) using four evaluation indicators, namely  $R^2$ , RMSE, MAE and NSE. The results demonstrated that, using the same input variables, the estimation accuracy of the spring maize daily  $ET_c$



of RF–SVR–PSO model was better than that of the standalone models. The RF–SVR–PSO model with  $K_c$ ,  $T_{ave}$ ,  $n$  and RH as the input variables can be utilized to estimate spring maize daily  $ET_c$  and provide a precise and accurate basis for agricultural water resource management and decision making. This conclusion can promote the development of water-saving agriculture and efficient utilization of agricultural water in arid and semi-arid areas of Northeast China, and it is also valuable for regions with different climatic conditions. In future studies, the model will be used in regions with different climatic conditions to improve its effectiveness at different stations.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/w15081503/s1>. Detailed mathematical theoretical derivation of the SVR model and the PSO algorithm can be found in the supplementary material.

**Author Contributions:** Conceptualization, W.H.; methodology, W.H.; software, W.H.; validation, J.G. and N.M.; formal analysis, W.H.; investigation, W.H.; resources, J.G. and G.Y.; data curation, W.H. and N.M.; writing—original draft preparation, W.H.; writing—review and editing, W.H. and J.G.; visualization, W.H.; supervision, J.G. and G.Y.; project administration, G.Y. and J.G.; funding acquisition, G.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National “Fourteenth Five-Year Plan” Key R&D Program (grant numbers 2022YFD1500604), the Black Land Protection and Utilization Science and Technology Innovation Project of the Chinese Academy of Sciences (XDA28090200, XDA28120100), the Liaoning Province Applied Basic Research Program (2022JH2/101300195), the Liaoning Province Young Top Talent Program (XLYC1907106) and the Liaoning Outstanding Innovation Team (XLYC2008015).

**Data Availability Statement:** The dataset used during this study can be obtained from the National Meteorological Information Center of China Meteorological Administration or from the corresponding author upon reasonable request.

**Acknowledgments:** We would like to thank the National Climatic Centre of the China Meteorological Administration for providing the climate database used in this study.

**Conflicts of Interest:** The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Kang, S. Towards water and food security in China. *Chin. J. Eco-Agric.* **2014**, *22*, 880–885.
2. Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Pretty, J.; Robinson, S.; Thomas, S.M.; Toulmin, C. Food Security: The Challenge of Feeding 9 Billion People. *Science* **2010**, *327*, 812–818. [[CrossRef](#)] [[PubMed](#)]
3. Wang, D.; Li, G.; Mo, Y.; Zhang, D.; Xu, X.; Wilkerson, C.J.; Hoogenboom, G. Evaluation of subsurface, mulched and non-mulched surface drip irrigation for maize production and economic benefits in northeast China. *Irrig. Sci.* **2021**, *39*, 159–171. [[CrossRef](#)]
4. Zou, H.; Fan, J.; Zhang, F.; Xiang, Y.; Wu, L.; Yan, S. Optimization of drip irrigation and fertilization regimes for high grain yield, crop water productivity and economic benefits of spring maize in Northwest China. *Agric. Water Manag.* **2020**, *230*, 105986. [[CrossRef](#)]
5. Tuo, Y.; Wang, Q.; Zhang, L.; Shen, F.; Wang, F.; Zheng, Y.; Wang, Z. Establishment of a crop evapotranspiration calculation model and its validation. *J. Agron. Crop. Sci.* **2022**, *209*, 251–260. [[CrossRef](#)]
6. Saggi, M.K.; Jain, S. Application of fuzzy-genetic and regularization random forest (FG-RRF): Estimation of crop evapotranspiration (ET) for maize and wheat crops. *Agric. Water Manag.* **2020**, *229*, 105907. [[CrossRef](#)]
7. Liu, X.; Fu, B. Drought impacts on crop yield: Progress, challenges and prospect. *Acta Geogr. Sin.* **2021**, *76*, 2632–2646.
8. FAOSTAT. Food and Agricultural Organization of the United Nations: Major Food and Agricultural Commodities and Producers. 2020. Available online: <http://www.fao.org/faostat/en/#data/QC/visualize> (accessed on 18 August 2022).
9. Hou, Y.; Kong, L.; Cai, H.; Liu, H.; Gao, Y.; Wang, Y.; Wang, L. The Accumulation and Distribution Characteristics on Dry Matter and Nutrients of High-Yielding Maize Under Drip Irrigation and Fertilization Conditions in Semi-Arid Region of Northeastern China. *Sci. Agric. Sin.* **2019**, *52*, 3559–3572.
10. Yang, X.; Ming, B.; Tao, H.; Wang, P. Spatial distribution characteristics and impact on spring maize yield of drought in Northeast China. *Chin. J. Eco-Agric.* **2015**, *23*, 758–767.
11. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; FAO—Food and Agriculture Organization of the United Nations: Rome, Italy, 1998; Available online: <https://www.fao.org/3/X0490E/x0490e00.htm> (accessed on 18 August 2022).

12. 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](#)]
13. Kumar, R.; Jat, M.K.; Shankar, V. Methods to estimate irrigated reference crop evapotranspiration—A review. *Water Sci. Technol.* **2012**, *66*, 525–535. [[CrossRef](#)]
14. Najafi, P.; Tabatabaei, S.H. Comparison of different Hargreaves-Samani methods for estimating potential evapotranspiration in arid and semi-arid regions of Iran. *Res. Crops* **2009**, *10*, 441–447.
15. Ai, Z.; Yang, Y. Modification and Validation of Priestley-Taylor Model for Estimating Cotton Evapotranspiration under Plastic Mulch Condition. *J. Hydrometeorol.* **2016**, *17*, 1281–1293. [[CrossRef](#)]
16. Al-Ghobari, H.M. Estimation of reference evapotranspiration for southern region of Saudi Arabia. *Irrig. Sci.* **2000**, *19*, 81–86. [[CrossRef](#)]
17. Xu, Z.; Yi, L.I.; Liu, J. Application of stochastic model to simulation of reference crop evapotranspiration in grassland of arid region. *J. Hydraul. Eng.* **2008**, *39*, 1267–1272, 1278.
18. Wang, W.; Peng, S.Z.; Luo, Y.F. Chaotic behavior analysis and prediction of reference crop evapotranspiration. *J. Hydraul. Eng.* **2008**, *39*, 1030–1036.
19. Pinos, J. Estimation methods to define reference evapotranspiration: A comparative perspective. *Water Pract. Technol.* **2022**, *17*, 940–948. [[CrossRef](#)]
20. Yamaç, S.S. Artificial intelligence methods reliably predict crop evapotranspiration with different combinations of meteorological data for sugar beet in a semiarid area. *Agric. Water Manag.* **2021**, *254*, 106968. [[CrossRef](#)]
21. Han, X.; Wei, Z.; Zhang, B.; Li, Y.; Du, T.; Chen, H. Crop evapotranspiration prediction by considering dynamic change of crop coefficient and the precipitation effect in back-propagation neural network model. *J. Hydrol.* **2021**, *596*, 126104. [[CrossRef](#)]
22. Yin, Z.; Wen, X.; Feng, Q.; He, Z.; Zou, S.; Yang, L. Integrating genetic algorithm and support vector machine for modeling daily reference evapotranspiration in a semi-arid mountain area. *Hydrol. Res.* **2017**, *48*, 1177–1191. [[CrossRef](#)]
23. Fan, J.; Yue, W.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol.* **2018**, *263*, 225–241. [[CrossRef](#)]
24. Xing, X.; Ma, X.; Yu, M.; Liu, Y. Estimating models for reference evapotranspiration with core meteorological parameters via path analysis. *Hydrol. Res.* **2017**, *48*, 340–354. [[CrossRef](#)]
25. Zhao, L.; Zhao, X.; Zhou, H.; Wang, X.; Xing, X. Prediction model for daily reference crop evapotranspiration based on hybrid algorithm and principal components analysis in Southwest China. *Comput. Electron. Agric.* **2021**, *190*, 106424. [[CrossRef](#)]
26. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
27. Mohammadi, B.; Mehdizadeh, S. Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. *Agric. Water Manag.* **2020**, *237*, 5–32. [[CrossRef](#)]
28. Pinos, J.; Chacón, G.; Feyen, J. Comparative analysis of reference evapotranspiration models with application to the wet Andean páramo ecosystem in southern Ecuador. *Meteorologica* **2020**, *45*, 25–45.
29. Petković, D.; Gocic, M.; Shamshirband, S.; Qasem, S.N.; Trajkovic, S. Particle swarm optimization-based radial basis function network for estimation of reference evapotranspiration. *Theor. Appl. Climatol.* **2016**, *125*, 555–563. [[CrossRef](#)]
30. Wu, Z.; Cui, N.; Hu, X.; Gong, D.; Wang, Y.; Feng, Y.; Jiang, S.; Lv, M.; Han, L.; Xing, L.; et al. Optimization of extreme learning machine model with biological heuristic algorithms to estimate daily reference crop evapotranspiration in different climatic regions of China. *J. Hydrol.* **2021**, *603*, 127028. [[CrossRef](#)]
31. Zhang, Y.; Cui, N.; Feng, Y.; Gong, D.; Hu, X. Comparison of BP, PSO-BP and statistical models for predicting daily global solar radiation in arid Northwest China. *Comput. Electron. Agric.* **2019**, *164*, 104905. [[CrossRef](#)]
32. Singh, A.; Sharma, A.; Rajput, S.; Bose, A.; Hu, X. An Investigation on Hybrid Particle Swarm Optimization Algorithms for Parameter Optimization of PV Cells. *Electronics* **2022**, *11*, 909. [[CrossRef](#)]
33. Xu, Y.; Hu, C.; Wu, Q.; Jian, S.; Li, Z.; Chen, Y.; Zhang, G.; Zhang, Z.; Wang, S. Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. *J. Hydrol.* **2022**, *608*, 127553. [[CrossRef](#)]
34. Li, W.; Zhang, L.; Chen, X.; Wu, C.; Cui, Z.; Niu, C. Predicting the evolution of sheet metal surface scratching by the technique of artificial intelligence. *Int. J. Adv. Manuf. Technol.* **2020**, *112*, 853–865. [[CrossRef](#)]
35. Baydaroglu, Ö.; Koçak, K. SVR-based prediction of evaporation combined with chaotic approach. *J. Hydrol.* **2014**, *508*, 356–363. [[CrossRef](#)]
36. Wang, Z.; Yin, G.; Gu, J.; Wang, S.; Ma, N.; Zhou, X.; Liu, Y.; Zhao, W. Effects of Water, Nitrogen and Potassium Interaction on Water Use Efficiency of Spring Maize Under Shallow-buried Drip Irrigation. *J. Soil Water Conserv.* **2022**, *36*, 316–324. [[CrossRef](#)]
37. Chen, S.; He, C.; Huang, Z.; Xu, X.; Jiang, T.; He, Z.; Liu, J.; Su, B.; Feng, H.; Yu, Q.; et al. Using support vector machine to deal with the missing of solar radiation data in daily reference evapotranspiration estimation in China. *Agric. For. Meteorol.* **2022**, *316*, 108864. [[CrossRef](#)]
38. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Vapnik, V.N., Ed.; Springer: Berlin/Heidelberg, Germany, 1995. [[CrossRef](#)]
39. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In Proceedings of the 10th Annual Conference on Neural Information Processing Systems (NIPS), Denver, CO, USA, 1996; pp. 155–161.
40. Shrestha, N.K.; Shukla, S. Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment. *Agric. For. Meteorol.* **2015**, *200*, 172–184. [[CrossRef](#)]

41. Zuo, R.; Carranza, E.J.M. Support vector machine: A tool for mapping mineral prospectivity. *Comput. Geosci.* **2011**, *37*, 1967–1975. [[CrossRef](#)]
42. Abdollahi, S.; Pourghasemi, H.R.; Ghanbarian, G.A.; Safaeian, R. Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4017–4034. [[CrossRef](#)]
43. Pal, M.; Maxwell, A.E.; Warner, T.A. Kernel-based extreme learning machine for remote-sensing image classification. *Remote Sens. Lett.* **2013**, *4*, 853–862. [[CrossRef](#)]
44. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. 2022. Available online: <https://CRAN.R-project.org/package=e1071> (accessed on 24 August 2022).
45. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the 1995 IEEE International Conference on Neural Networks (ICNN 95), Perth, WA, Australia, 27 November–1 December 1995; pp. 1942–1948.
46. Gu, W.; Chai, B.; Teng, Y. Research on Support Vector Machine Based on Particle Swarm Optimization. *Trans. Beijing Inst. Technol.* **2014**, *34*, 705–709.
47. Bendtsen, C. pso: Particle Swarm Optimization. 2022. Available online: <https://CRAN.R-project.org/package=pso> (accessed on 24 August 2022).
48. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
49. Karimi, S.; Shiri, J.; Marti, P. Supplanting missing climatic inputs in classical and random forest models for estimating reference evapotranspiration in humid coastal areas of Iran. *Comput. Electron. Agric.* **2020**, *176*, 105633. [[CrossRef](#)]
50. Heung, B.; Bulmer, C.E.; Schmidt, M.G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. *Geoderma* **2014**, *214–215*, 141–154. [[CrossRef](#)]
51. Wang, Z.; Zhang, X.; Chhin, S.; Zhang, J.; Duan, A. Disentangling the effects of stand and climatic variables on forest productivity of Chinese fir plantations in subtropical China using a random forest algorithm. *Agric. For. Meteorol.* **2021**, *304–305*, 108412. [[CrossRef](#)]
52. Archer, E. rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics. 2022. Available online: <https://CRAN.R-project.org/package=rfPermute> (accessed on 24 August 2022).
53. Zhang, D.; Lin, J.; Peng, Q.; Wang, D.; Yang, T.; Sorooshian, S.; Liu, X.; Zhuang, J. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J. Hydrol.* **2018**, *565*, 720–736. [[CrossRef](#)]
54. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [[CrossRef](#)]
55. Fritsch, S.; Guenther, F.; Wright, M.N. neuralnet: Training of Neural Networks. 2019. Available online: <https://github.com/bips-hb/neuralnet> (accessed on 24 August 2022).
56. R Core Team. R: A Language and Environment for Statistical Computing. 2021. Available online: <https://www.R-project.org/> (accessed on 20 August 2022).
57. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
58. Ji, R.; Ban, X.; Zhang, S. Ascertainment of Crop Coefficients of Maize in Liaoning Area. *Chin. Agric. Sci. Bull.* **2004**, *20*, 246–248+268.
59. Zhu, B.; Feng, Y.; Gong, D.; Jiang, S.; Zhao, L.; Cui, N. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Comput. Electron. Agric.* **2020**, *173*, 105430. [[CrossRef](#)]
60. Jia, Y.; Su, Y.; Zhang, R.; Zhang, Z.; Lu, Y.; Shi, D.; Xu, C.; Huang, D. Optimization of an extreme learning machine model with the sparrow search algorithm to estimate spring maize evapotranspiration with film mulching in the semiarid regions of China. *Comput. Electron. Agric.* **2022**, *201*, 107298. [[CrossRef](#)]
61. Wen, L.; Yuan, X. Forecasting CO2 emissions in Chinas commercial department, through BP neural network based on random forest and PSO. *Sci. Total Environ.* **2020**, *718*, 137194. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.