

## Research Article

# Coupling a Bat Algorithm with XGBoost to Estimate Reference Evapotranspiration in the Arid and Semiarid Regions of China

Yixiu Han,<sup>1</sup> Jianping Wu,<sup>2</sup> Bingnian Zhai,<sup>3</sup> Yanxin Pan,<sup>1</sup> Guomin Huang ,<sup>1</sup> Lifeng Wu,<sup>1</sup> and Wenzhi Zeng<sup>4</sup>

<sup>1</sup>School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China

<sup>2</sup>Laboratory of Ecology and Evolutionary Biology, Yunnan Key Laboratory of Plant Reproductive Adaption and Evolutionary Ecology, Yunnan University, Kunming 650500, China

<sup>3</sup>Key Laboratory of Plant Nutrition and Agri-Environment in Northwest China, Ministry of Agriculture, College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi 712100, China

<sup>4</sup>State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, Hubei 430072, China

Correspondence should be addressed to Guomin Huang; [g.huang@nit.edu.cn](mailto:g.huang@nit.edu.cn)

Received 25 May 2019; Accepted 17 September 2019; Published 17 October 2019

Academic Editor: Herminia García Mozo

Copyright © 2019 Yixiu Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate estimation of reference evapotranspiration (ET<sub>o</sub>) is key to agricultural irrigation scheduling and water resources management in arid and semiarid areas. This study evaluates the capability of coupling a Bat algorithm with the XGBoost method (i.e., the BAXGB model) for estimating monthly ET<sub>o</sub> in the arid and semiarid regions of China. Meteorological data from three stations (Datong, Yinchuan, and Taiyuan) during 1991–2015 were used to build the BAXGB model, the multivariate adaptive regression splines (MARS), and the gaussian process regression (GPR) model. Six input combinations with different sets of meteorological parameters were applied for model training and testing, which included mean air temperature ( $T_{\text{mean}}$ ), maximum air temperature ( $T_{\text{max}}$ ), minimum air temperature ( $T_{\text{min}}$ ), wind speed ( $U$ ), relative humidity (RH), and solar radiation ( $R_c$ ) or extraterrestrial radiation ( $R_a$ ,  $\text{MJ m}^{-2} \cdot \text{d}^{-1}$ ). The results indicated that BAXGB models (RMSE = 0.114–0.412  $\text{mm} \cdot \text{d}^{-1}$ , MAE = 0.087–0.302  $\text{mm} \cdot \text{d}^{-1}$ , and  $R^2 = 0.937$ –0.996) were more accurate than either MARS (RMSE = 0.146–0.512  $\text{mm} \cdot \text{d}^{-1}$ , MAE = 0.112–0.37  $\text{mm} \cdot \text{d}^{-1}$ , and  $R^2 = 0.935$ –0.994) or GPR (RMSE = 0.289–0.714  $\text{mm} \cdot \text{d}^{-1}$ , MAE = 0.197–0.564  $\text{mm} \cdot \text{d}^{-1}$ , and  $R^2 = 0.817$ –0.980) model for estimating ET<sub>o</sub>. Findings of this study would be helpful for agricultural irrigation scheduling in the arid and semiarid regions and may be used as reference in other regions where accurate models for improving local water management are needed.

## 1. Introduction

There are two ways for water entering the atmosphere from the land surface. One is the direct evaporation of water from soils and plant surfaces, while the other one is transpiration, by which water from soils passes through plants and then evaporates on the leaves and diffuses to the surrounding air through stomata. Technically, it is difficult to distinguish between these two different pathways, which are collectively defined as crop evapotranspiration (ET<sub>c</sub>). Although ET<sub>c</sub> can be directly measured by techniques such as lysimeter or eddy

covariance, these techniques cannot be applied widely due to their high costs in observation and maintenance. Therefore, ET<sub>c</sub> is generally estimated alternatively via multiplying the reference crop evapotranspiration (ET<sub>o</sub>) by the crop coefficient ( $K_c$ ). According to Allen et al. [1], ET<sub>o</sub> represents the evapotranspiration from an actively growing virtual vegetated surface that is 0.12 m tall, with a completely shaded ground and adequate water supply, and without pests or diseases. At present, ET<sub>o</sub> has become a fundamental parameter in the fields of regional water resource management and irrigation scheduling. The most widely accepted ET<sub>o</sub>

estimating method is the FAO-56 Penman–Monteith method (P-M), which has been proposed as a standard method for validating other methods because of its global applicability and simplicity (no need for parameter calibration) [1]. The method is function of basic meteorological parameters only, including temperature, solar radiation, wind speed, and relative humidity. However, even basic meteorological information in reality is often lacking, especially in developing countries, because of limited observational capacity and costs. For example, the lack of solar radiation records is a common problem in most parts of the world. Although there are over 2000 weather stations in China, less than 130 of them have observational records of radiation [2–4]. To solve the problem of lacking field observational data, various alternative methods, including empirical models, statistical models, artificial intelligence algorithms, and the satellite remote sensing technology, have been proposed and studied worldwide. Although these methods have different application scopes and characteristics, they all play a great role in improving the management of water resources.

Artificial intelligence (AI) algorithms are becoming popular recently because of their high flexibility and precision and have been applied widely in fields of hydrology, agriculture, and environment [5–8], such as drought forecasting [9], solar radiation modeling [2, 10], temperature estimation [11], precipitation modeling [12], soil moisture simulation [13], and reference evapotranspiration estimation [14–17]. In the early years, artificial neural networks (ANNs) were used to simulate ETo, generating more accurate results than the empirical model [18]. Kisi [19] compared least square support vector machine (LSSVM), multivariate adaptive regression splines (MARS), and M5 model tree (M5T) for estimating ETo in Antalya and Isparta of Iran and found that the MARS algorithm was most accurate when local input data were not sufficient. Sanikhani et al. [20] compared six AI algorithms, including multilayer perceptron (MLP), generalized regression neural network (GRNN), radial basis neural networks (RBNN), integrated ANFIS with grid partitioning and subtractive clustering (ANFIS-GP and ANFIS-SC), and GEP, with data from the Antalya and Isparta stations of Iran. The results tended to be station specific, of which GEP and GRNN produced better performance than other algorithms in the Antalya station, while RBNN and ANFIS-SC were superior to other algorithms in the other station. Saggi and Jain [21] used four AI algorithms (including deep learning-multilayer perceptron (DL), generalized linear model (GLM), random forest (RF), and gradient boosting decision tree (GBDT)) to determine daily ETo in Punjab, India. The DL algorithm presented high performance for modeling daily ETo, with  $NSE = 0.95–0.98$ ,  $RMSE = 0.192–0.269 \text{ mm}\cdot\text{d}^{-1}$ , and  $R^2 = 0.95–0.99$ . Landaras et al. [22] proposed an alternative strategy, using the combined data of local and its neighboring stations as input for ETo estimation with GEP and ANN, results of which showed that GEP had very good performance in all stations. Feng et al. [23] evaluated extreme learning machine (ELM), GRNN, and the Hargreaves model for the estimation of daily ETo using temperature data. They found that the Hargreaves

was worse than AI algorithms, while ELM and GRNN performed best in the local scenario and the cross station scenario, respectively. Fan et al. [24] evaluated four tree-based algorithms (including M5T, RF, GBDT, and XGBoost) for estimating ETo in eight stations of China and found that XGBoost could yield comfortable accuracy but with much less time cost when compared with other algorithms. Moreover, although the gaussian process regressive (GPR) algorithm with a clear mathematical basis has been applied widely, only a few studies to date used it for crop ETo estimation [25].

Parameters of AI models usually need to be calibrated, which is generally time-consuming, and requires a solid professional knowledge, especially when the inputs are complicated (e.g., input data from multiple sites or as multiple combinations) [26]. To overcome this problem, optimization with metaheuristic algorithms has been widely adopted by scholars recently, such as using the genetic algorithm (GA) to optimize ANN for ETo estimation in Iran [27] and hybridizing the firefly algorithm with ANFIS for estimating ETo in Burkina Faso [28]. The effects of metaheuristic optimization tend to be algorithm specific, with both positive and negative effects reported in various studies. For instance, Petković et al. [29] found that the radial basis function neural network (RBFNN) coupled with the particle swarm optimization (PSO) algorithm was more accurate than traditional artificial neural network algorithms.

In recent years, the bat algorithm (BA) has received a lot of attention because of its strong searching ability and robustness, and the application of this algorithm has been discussed in many fields [28, 30, 31]. In addition, as a tree-based assemble algorithm, XGBoost has been proved superior to other simple tree algorithms such as RF and M5T in terms of accuracy and efficiency (i.e., less computational cost) for ETo estimation [24]. However, XGBoost contains many parameters, which would have impacts on its stability and efficiency during processing. Therefore, using the bat algorithm to optimize XGBoost is expected to improve the accuracy and stability of the single XGBoost model. To our knowledge, studies using evolutionary algorithms to optimize model parameters of XGBoost in hydrology, agriculture, and environment fields are scarce to date, especially for ETo estimation. Thus, the aims of this study were (1) to assess the capability of XGBoost optimized by BA in estimating monthly ETo with limit meteorological data as inputs and (2) to evaluate the performance of the BAXGB model for ETo estimation in comparison with GPR and MARS models in the arid and semiarid regions of China.

## 2. Materials and Methods

**2.1. Study Area and Data.** Monthly meteorological parameters during 1991–2015 from three weather stations in North China were selected in this study for model training and testing, including mean air temperature ( $T_{\text{mean}}$ , °C), maximum air temperature ( $T_{\text{max}}$ , °C), minimum air temperature ( $T_{\text{min}}$ , °C), wind speed ( $U$ ,  $\text{m s}^{-1}$ ), solar radiation ( $R_s$ ,  $\text{MJ m}^{-2}\cdot\text{d}^{-1}$ ) or extraterrestrial radiation ( $R_a$ ,  $\text{MJ m}^{-2}\cdot\text{d}^{-1}$ ), and relative humidity (RH, %). For the selected weather

stations (Figure 1), Datong and Yinchuan locate at the arid region of China, with a mean annual precipitation of 183 mm and 364 mm, respectively, while Taiyuan belongs to the semiarid region of China and has an average of 432 mm annual precipitation.

All meteorological data used in this study were obtained from the National Meteorological Information Center of China (NMICC) with quality control. In addition, a further quality control was applied to the data of solar radiation following guidelines in [3] before processing. The missing data of all datasets were no more than 1% in total. The geographic information and meteorological statistics during the training and testing stages are listed in Table 1. It is noted that, compared with the training stage, wind speed of Datong and Yinchuan decreased by 10% and 29.2% in the testing stage, respectively, while solar radiation of Taiyuan increased by 16% during the testing stage. Six input combinations that consisted of different meteorological parameters were used for model training and testing in this study (Table 2). The map road of this study is described in Figure 2.

**2.2. Multivariate Adaptive Regression Splines (MARS).** MARS is a nonlinear nonparametric regression technique introduced by Friedman [32], which uses the divide-and-conquer strategy for data processing. Specifically, data for model training are divided into subsets, each of which has its own function with at least two regression lines. The endpoints of a line segment are defined as nodes. During processing, specific assumptions regarding the basic functional relationship between input variables and outputs are not needed, but using the knotting method instead. For each input variable, knotting nodes can be made on any data point within the range of the variable, and the strategy of knotting is to search all possible knotting positions based on the adaptive regression algorithm and form a piecewise curve consequently. A knotting node marks the end of one data region and the beginning of another. Between the two adjoining knotting nodes, the regression of data forms a basis function (BF). Therefore, MARS generates BFs through the step-by-step searching, during which the node positions are automatically selected by an adaptive regression algorithm. If there are multiple input variables, it will search for all possible univariate node positions and interactions between all variables. However, these processes generally build an over fit model. To reduce the degree of overfitting, pruning is incorporated into MARS as the second phase of the model development. In summary, the MARS model is actually a weighed sum of BF, which makes the model more flexible. More details about MARS can be found in [32].

**2.3. Gaussian Process Regression (GPR).** Gauss process regression is a framework based on the Bayesian theory with wide application in many fields. A GPR model can be defined as a probability distribution over functions [33], which is denoted as

$$\begin{aligned} f(x) &= \text{GP}(m(x), k(x, x')), \\ m(x) &= E[f(x)], \\ k(x, x') &= E[(f(x) - m(x))(f(x') - m(x'))). \end{aligned} \quad (1)$$

Since the Gaussian process is a normal distribution in the domain of functions, it can be interpreted from the perspective of space. If there is a predictor group  $x$  consisting of  $d$  variables, it can be written as  $D = \{x_k, f(x_k), k = 1, 2, 3, \dots, n\}$ , where  $x_k \in R^n$  is the input vector and  $f(x_k) \in R^n$  is the corresponding output vector. In general, the Bayesian regression function can be expressed as

$$f(x) = \varphi(x)^T \omega, \quad (2)$$

where  $\omega$  is the weight vector, whose prior distribution is  $\omega \sim N(0, \Sigma_p)$ . For instance, if  $E(\omega) = 0$ , then  $E(\omega\omega^T) = \Sigma_p$ .

Since the expectation and variance can be expressed as

$$\begin{aligned} E(f(x)) &= \varphi(x)^T E(\omega) = 0, \\ E[f(x)f(x')] &= \varphi(x)^T E(\omega\omega^T)\varphi(x') = \varphi(x)^T \sum_p \varphi(x'), \end{aligned} \quad (3)$$

then we can see that when testing data and training data obey the same distribution, for a given testing vector  $x_*$ , the corresponding prediction vector  $f_*$  also belongs to the Gaussian distribution. Suppose they form a joint Gaussian distribution:

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K(x, x) & K(x, x_*) \\ K(x, x_*)^T & K(x_*, x_*) \end{bmatrix} \right\}, \quad (4)$$

where

$$\begin{aligned} K &= \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}, \\ k_* &= [k(x_*, x_1) \quad k(x_*, x_2) \quad \cdots \quad k(x_*, x_n)], \\ k_{**} &= k(x_*, x_*), \end{aligned} \quad (5)$$

in which  $k(x, x)$  is the correlation between  $x$  and itself, while  $K(x, x_*) = K(x_*, x)^T$  is the  $N * 1$  order covariance matrix between the input and the output, briefly recorded as  $K = K^T$ , and  $K(x_*, x_*)$  is the correlation of the output itself, which can be written in short as  $K_{**}$ .

The Bayesian posterior probability is as follows:

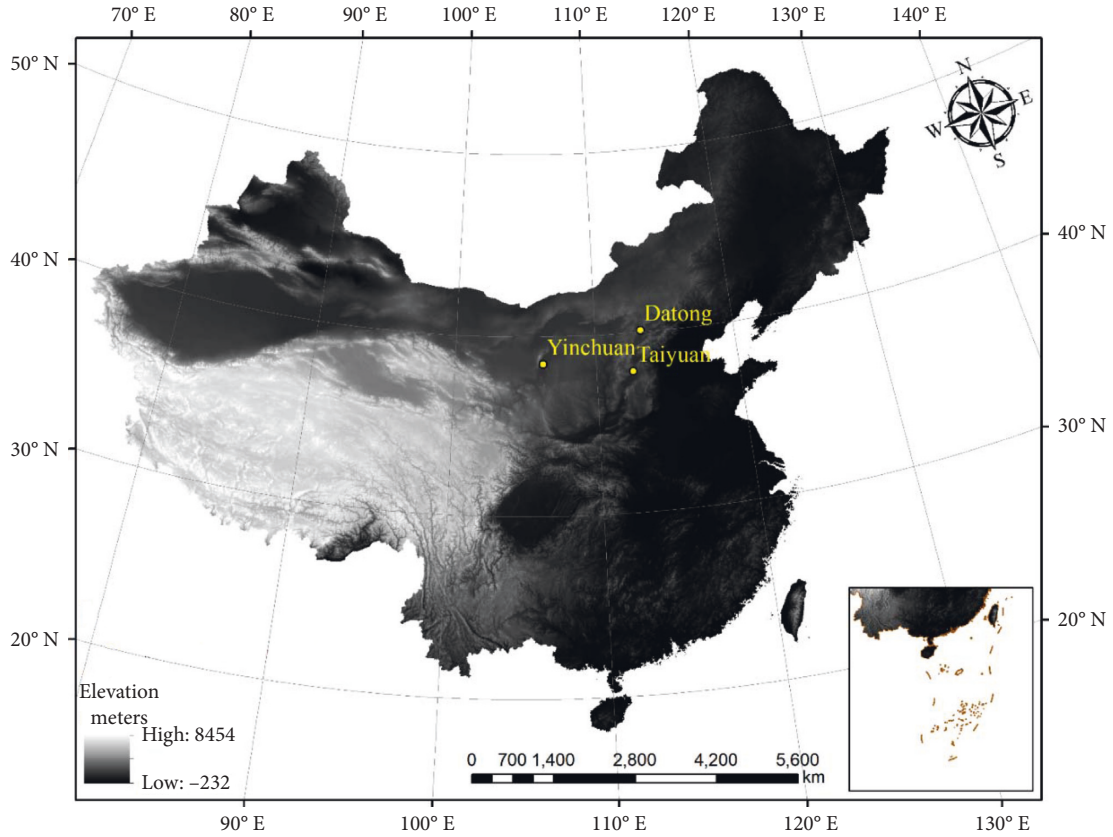
$$P(f_* | x_*, f) = N(\mu_*, \Sigma_*), \quad (6)$$

where

$$\begin{cases} \mu_* = \mu(x_*) + K_*^T K^{-1} (f - \mu(x)), \\ \Sigma_* = K_{**} - K_*^T K^{-1} K_*, \end{cases} \quad (7)$$

where  $\mu_*$  is the posterior probability means of  $f_*$  and  $\Sigma_*$  is the variance of posterior probability.

When there is noise in the regression model,



● Weather station series

FIGURE 1: The geographic information of the three weather stations in this study.

TABLE 1: Statistics of the meteorological data of the three stations during the training and testing stages in the study.

Station/ Meteorological data	Average		Maximum		Minimum		Skew		Kurtosis		CV	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
<i>Datong</i>												
$T_{max}$	14.52	14.54	32.38	29.05	-6.85	-5.13	-0.24	-0.31	-1.34	-1.41	0.79	0.82
$T_{min}$	1.44	0.37	19.6	18.01	-18.93	-18.06	-0.14	-0.14	-1.36	-1.26	7.97	30.97
$T_{mean}$	7.98	7.45	25.99	23.53	-12.89	-11.6	-0.19	-0.23	-1.35	-1.34	4.38	15.9
$R_s$	14.73	15.72	24.79	24.75	5.55	6.06	-0.05	-0.09	-1.28	-0.93	0.37	0.46
$U$	2.07	1.86	3.37	2.89	0.96	1.12	0.46	0.8	0.33	0.97	0.2	0.3
RH	51.34	54.32	76.94	78.47	26.47	29.53	-0.11	-0.25	-0.61	1.34	0.22	0.36
ETo	2.83	2.76	6.46	5.72	0.44	0.5	0.17	0.21	-1.33	-1.21	0.58	0.69
<i>Yinchuan</i>												
$T_{max}$	16.55	17.03	32.06	31.43	-5.45	-2.27	-0.29	-0.34	-1.29	-1.35	0.65	0.7
$T_{min}$	3.98	4.94	20.09	19.5	-17.27	-13.41	-0.16	-0.27	-1.33	-1.17	2.66	2.17
$T_{mean}$	10.26	10.98	26.08	25.47	-11.36	-7.84	-0.23	-0.3	-1.31	-1.26	1.66	1.43
$R_s$	15.93	15.4	25.85	24.05	6.4	7.02	0.04	-0.03	-1.28	-0.5	0.33	0.43
$U$	1.78	1.26	2.77	1.78	0.98	0.84	0.27	2.06	-0.23	5.59	0.21	0.54
RH	54.1	48.7	78.67	77.4	23.93	24.19	-0.46	0.03	-0.58	0.91	0.21	0.38
ETo	2.98	2.85	6.17	5.37	0.46	0.48	0.13	0.09	-1.39	-1.33	0.57	0.66
<i>Taiyuan</i>												
$T_{max}$	17.62	17.79	32.75	31.22	-1.89	-0.16	-0.25	-0.35	-1.34	-1.37	0.57	0.63
$T_{min}$	4.9	5.37	21.36	20.18	-13.84	-11.21	-0.08	-0.2	-1.37	-1.26	2.07	1.91
$T_{mean}$	11.26	11.58	27.06	25.7	-7.87	-5.69	-0.17	-0.27	-1.36	-1.32	1.32	1.27
$R_s$	13.29	15.46	22.3	27.81	4.41	6.08	0.03	0.12	-1.21	-0.66	0.35	0.47
$U$	1.44	1.39	2.6	2.49	0.57	0.82	0.6	0.89	0.23	0.48	0.27	0.35
RH	57.05	55.12	85.74	79.57	28.79	35.52	-0.09	0.03	-0.88	2.01	0.22	0.35
ETo	2.61	2.83	5.51	5.6	0.49	0.63	0.2	0.2	-1.29	-1.3	0.53	0.65

TABLE 2: Input combinations of different meteorological parameters used in this study.

Combinations	Models			Input combinations
	GPR	MARS	BAXGB	
1	GPR1	MARS1	BAXGB1	$T_{\text{mean}} \cdot R_a$
2	GPR2	MARS2	BAXGB2	$T_{\text{max}} \cdot T_{\text{min}} \cdot R_a$
3	GPR3	MARS3	BAXGB3	$T_{\text{max}} \cdot T_{\text{min}} \cdot R_s$
4	GPR4	MARS4	BAXGB4	$T_{\text{max}} \cdot T_{\text{min}} \cdot U \cdot R_a$
5	GPR5	MARS5	BAXGB5	$T_{\text{max}} \cdot T_{\text{min}} \cdot \text{RH} \cdot R_a$
6	GPR6	MARS6	BAXGB6	$T_{\text{max}} \cdot T_{\text{min}} \cdot \text{RH} \cdot U \cdot R_a$

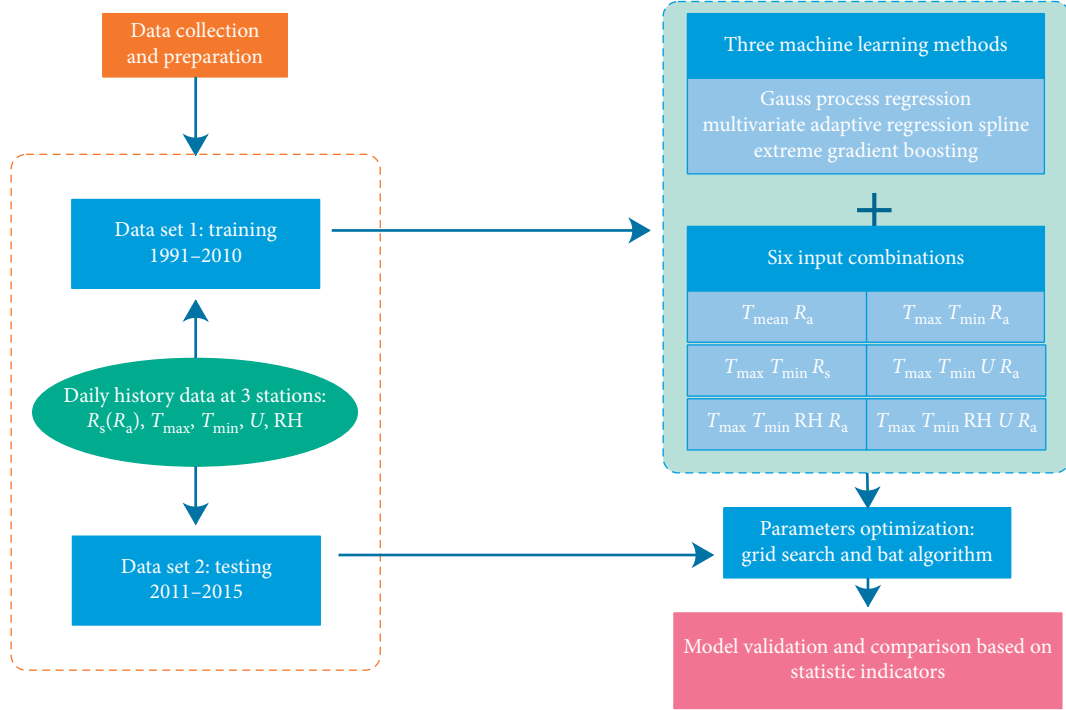


FIGURE 2: The technology roadmap of the three algorithms in this study.

$$y = f(x) + \varepsilon, \quad (8)$$

where  $x$  is the input vector,  $f$  is the function value vector, and  $y$  is the observations with noise. It is generally considered that noise obeys a normal distribution and has a mean value equal to 0, with the variation  $\sigma_n$  expressed as  $\varepsilon \sim N(0, \sigma_n^2)$ . The prior distribution of observed values can be obtained as

$$y \sim N(\mu(x), K(x, x) + \sigma_n^2 I_n), \quad (9)$$

where  $I_n$  is the  $n$ -dimensional identity matrix. Then, the joint distribution of the observation set  $(x, y)$  and the prediction set  $(x_*, y_*)$  can be written as follows:

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K(x, x) & \sigma_n^2 I_n K(x, x_*) \\ K(x, x_*)^T & K(x_*, x_*) \end{bmatrix} \right\}. \quad (10)$$

Therefore, the posterior probability can be calculated as follows:

$$P(y_* | x_*, y) = N(\mu_*, \Sigma_*), \quad (11)$$

where

$$\begin{cases} \mu_* = \mu(x_*) + K_*^T (K + \sigma_n^2 I_n)^{-1} (y - \mu(x)), \\ \Sigma_* = K_{**} - K_*^T (K + \sigma_n^2 I_n)^{-1} K_*. \end{cases} \quad (12)$$

In general, to make the symbol concise, data need to be preprocessed, and the mean value can be set as  $\mu(x) = \mu(x_*) = 0$ . Therefore, formula (11) can be written as

$$P(y_* | x_*, x, y) \sim N \left( K_*^T (K + \sigma_n^2 I_n)^{-1} (y - \mu(x)), K_{**} - K_*^T (K + \sigma_n^2 I_n)^{-1} K_* \right). \quad (13)$$

For regression problems, the mean value is used as predictive values ( $\hat{y}$ ), which can be written as

$$\hat{y} = K_*^T (K + \sigma_n^2 I_n)^{-1} y. \quad (14)$$

Further simplification is as follows:

$$\begin{aligned} (K + \sigma_n^2 I_n)^{-1} y &= (K + \sigma_n^2 I_n)^{-1} [y_1, y_2, \dots, y_n]^T = \alpha^T, \\ \hat{y} &= K_*^T \alpha^T = \sum_{i=1}^n K(x_*, x_i) \alpha_i. \end{aligned} \quad (15)$$

According to the Mercer theorem,  $K(x_*, x_i)$  can be changed to the Gaussian kernel method:

$$K(x_*, x_i) = \gamma \exp\left(-\frac{\|x_* - x_i\|^2}{2\sigma_k^2}\right), \quad (16)$$

where  $\gamma$  and  $\sigma_k^2$  can be determined automatically by the Bayesian method.

**2.4. XGBoost, Bat Algorithm, and BAXGB.** XGBoost, proposed by Chen and Guestrin [34], is a scalable artificial intelligence algorithm for tree boosting. The original model of XGBoost is gradient boosting decision trees, which combines many “weak” learning models into a “strong” learner through an iterative method. As shown in Figure 3, during each iteration of XGBoost, the previous predictors are corrected with residuals. The algorithm can independently determine the types of loss functions used for model evaluation. To reduce the risk of overfitting, an additional regularization term is added to the model. The mean score of each tree is used as the predictive value for regression. For the  $m$ -th decision tree, its calculation formula can be expressed as

$$\hat{y}_i = \sum_{m=1}^m f_m(x_i), \quad f_m \in W, \quad (17)$$

where  $m$  is the number of trees,  $f_m$  is a function in the functional space  $W$ , and  $W$  is the space of all decision trees.

The objective function at the  $t$ -th iteration can be written as

$$\begin{aligned} \Theta(t) &= \Phi(t) + \Omega(t), \\ \Theta(t) &= \sum_{i=1}^n \Phi(y_i, \hat{y}_i) + \sum_{k=1}^t \Omega(f_k), \end{aligned} \quad (18)$$

where  $n$  is the  $n$ -th prediction and  $\hat{y}_i^{(t)}$  can be given as

$$\hat{y}_i^{(t)} = \sum_{m=1}^t f_m(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i). \quad (19)$$

The regularization term  $\Omega(f_k)$  for a decision tree is defined by Chen and Guestrin [34] as follows:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2, \quad (20)$$

where  $\gamma$  is the complexity of each leaf,  $T$  is the number of leaves in a tree,  $\lambda$  is a parameter to scale the penalty, and  $w$  is the vector of scores on the leaves. Then, the first-order along with the second-order Taylor expansions are taken to the loss function in XGBoost. For example, if the loss function is

the mean square error (MSE), then the objective function can finally be derived as

$$\Theta^{(t)} \approx \sum_{i=1}^n \left[ g_i w_{q(x_i)} + \frac{1}{2} \left( h_i w_{q(x_i)}^2 \right) \right] + \gamma T + \frac{1}{2} \sum_{j=1}^T w_j^2, \quad (21)$$

where  $q(\cdot)$  is a function that can assign data points to corresponding leaves and  $g_i$  and  $h_i$  are the first and second derivatives of MSE loss function, respectively. In formula (21), the loss function is determined by the sum of loss values for each data sample. Since each data sample corresponds to only one leaf node, the loss function can also be expressed as the sum of loss values for each leaf node, which is

$$\Theta^{(t)} \approx \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T, \quad (22)$$

where  $I_j$  represents all data samples in the leaf node  $j$ . Hence, optimization of the objective function is now transformed into finding the minimum value of a quadratic function. In other words, the change of model performance caused by a certain node split in the decision tree now can be evaluated based on the objective function. If the model performance of the decision tree is improved after the node split, then the change will be adopted; otherwise, the split will be stopped. The structure of the XGBoost algorithm can be seen in Figure 3. Besides, when optimizing the objective function, a predictive classifier can be trained against overfitting because of the regularization terms. In this study, three parameters were optimized, including the number of trees (nrounds), the ratio of subdatasets to all data in training the model (subset), and the minimum sum of instance weight (hessian) needed in a child (min\_child\_weight).

The bat algorithm (BA) is a metaheuristic algorithm developed by Yang [35], which mirrors the echolocation behavior of microbats. Microbats emit sound pulses to search for prey, and each of them has its own echolocation characteristics (i.e., pulse emission rates and loudness), the feature of which can be used for developing various BA. The main processes of constructing a BA can be summarized as follows:

- (1) A population of virtual bats is first generated for simulations. For each virtual bat, assume that it flies randomly with a velocity  $v_i$  and a frequency  $f_i$  at position  $x_i$ .
- (2) From the first iteration to the maximum iteration, positions and velocities of bats at time  $t$  are updated by

$$\begin{aligned} f_i &= f_{\min} + (f_{\max} - f_{\min}) \beta, \\ v_i^t &= v_i^{t-1} + (x_i^t - x_*) f_i, \\ x_i^t &= x_i^{t-1} + v_i^t, \end{aligned} \quad (23)$$

where  $\beta \in [0, 1)$  is a random vector from a normal distribution,  $x_i^t$  and  $v_i^t$  are the updated positions and velocities of

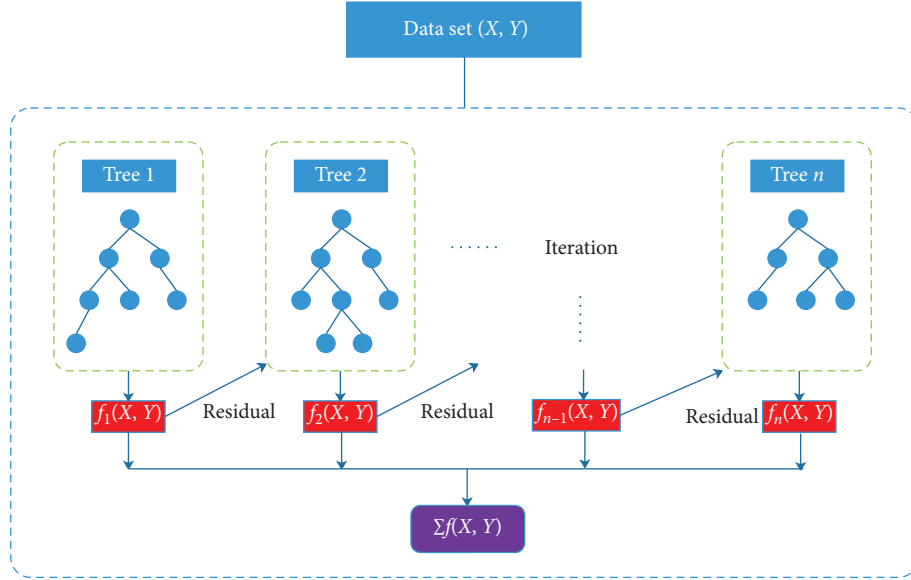


FIGURE 3: The structure of the XGBoost algorithm.

bats at time  $t$ , respectively, and  $x_*$  is the current best position (solution) a bat is located after comparing all the solutions among all the bats within the population.

- (3) Once the current best solution is selected, the bats will update their positions based on a random walk:

$$x_{\text{new}} = x_{\text{old}} + \text{rand} \cdot A^t, \quad (24)$$

where  $\text{rand} \in [-1, 1]$  is a random number and  $A^t$  is the mean loudness of all bats at time  $t$ .

- (4) As the iteration proceed, the rate of pulse emission  $r_i$  and the loudness  $A_i$  of each bat within the population will be updated according to the following rules:

$$\begin{aligned} A_i^{t+1} &= \alpha A_i^t, \\ r_i^{t+1} &= r_i^0 [1 - \exp(-\gamma t)], \end{aligned} \quad (25)$$

where  $\alpha$  and  $\gamma$  are both constants. The emission rates and loudness of all bats will be updated until the best solutions are found, and the iteration will be terminated by then.

In this study, a hybrid model integrating BA and XGBoost (i.e., BAXGB) was developed for ETo estimation. As stated previously, the performance of XGBoost models is largely dependent on the appropriate choice of parameters. In BAXGB, the parameter choosing for XGBoost was determined and optimized by BA. This paper optimizes three important parameters of the XGBoost algorithm, which are the number of trees (nrounds), the minimum sum of instance weight needed in a child (Min\_child\_weight), and the study rate (eta). The parameter values are shown in Table 3.

**2.5. Model Performance Evaluation.** Five statistical indicators were used for the performance evaluation of GPR, MARS, and BAXGB models in this study, which were as follows.

TABLE 3: Parameter values of the Bat-XGBoost algorithm.

Parameter	Value
Population	50
Max generation	200
A	0.98
R	0.98
nrounds	[20, 2000]
Min_child_weight	[1, 15]
Eta	[0.01, 0.3]

- (1) Coefficient of determination:

$$R^2 = \frac{[\sum_{i=1}^n (O_{i,m} - \bar{O}_{i,m})(O_{i,e} - \bar{O}_{i,e})]^2}{\sum_{i=1}^n (O_{i,m} - \bar{O}_{i,m})^2 \sum_{i=1}^n (O_{i,e} - \bar{O}_{i,e})^2}. \quad (26)$$

- (2) Root-mean-square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_{i,m} - O_{i,e})^2}. \quad (27)$$

- (3) Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |O_{i,m} - O_{i,e}|. \quad (28)$$

- (4) The maximum 10% RMSE ( $\text{RMSE}_{\text{max } 10\%}$ ):

$$\text{RMSE}_{\text{max } 10\%} = \sqrt{\frac{1}{z} \sum_{i=1}^z (E_{\text{max } 10\%})^2}. \quad (29)$$

- (5) The maximum 10% MAE ( $\text{MAE}_{\text{max } 10\%}$ ):

$$\text{MAE}_{\max 10\%} = \frac{1}{z} \sum_{i=1}^z |E_{\max 10\%}|, \quad (30)$$

where  $n$  is the total number of test data,  $O_e$  and  $O_m$  are the predicted ETo values by artificial intelligence methods and the FAO-56 P-M ETo values, respectively,  $\bar{O}_m$  is the mean P-M ETo value, and  $E_{\max 10\%}$  is the maximum 10% absolute error of the test data and the FAO-56 P-M ETo.

The value of  $R^2$  is between 0 and 1, and the value closer to 1 indicates better regression fitting between predicted ETo and P-M ETo and hence the better model performance. On the contrary, if the value of RMSE or MAE is closer to 0, the model performs better. RMSE is generally useful when model errors follow a normal distribution, whereas MAE suits better for models with a uniform error distribution [24]. In addition, considering people's different acceptability of the ETo error in different months, as well as the uncertainty caused by the water resource management process, it is necessary to evaluate not only the overall performance of the model but also whether the performance of the model meets the expectation in extreme cases. Therefore, two new statistical indicators using the maximum error point of 10% were adopted in this study to evaluate the model performance in extreme cases.

### 3. Results and Discussion

**3.1. Model Performance in the Datong Station.** Performance of the three artificial intelligence models used in this study (i.e., GPR, MARS, and BAXGB) during the training and testing stages for the Datong station is listed in Table 4. Both air temperature and extraterrestrial radiation were used for the first two input combinations. The difference was that the input combination 1 used the monthly average temperature, while the input combination 2 used the monthly maximum and minimum temperatures. Under these two input combinations, the accuracies of the three types of models varied. For GPR and BAXGB models, input combination 2 had higher accuracy than input combination 1. However, MARS models yielded an opposite result. Input combinations 3, 4, and 5 represented the effects of solar radiation, wind speed, and relative humidity or temperature data, respectively. Considering that the extraterrestrial radiation is only related to the latitude of the Earth and does not belong to the meteorological parameters to be observed, this information was added to the combinations 4 and 5. The importance of meteorological factors given by different models was different. MARS and BAXGB models showed that relative humidity was the most important, followed by wind speed and solar radiation. However, GPR model results showed that wind speed was the most critical, followed by relative humidity and solar radiation. Although there was difference to some degree in the importance between different meteorological parameters, it was still clear that relative humidity and wind speed were more important than solar radiation for the Datong station. Input combination 6 was a combination of temperature, wind speed, and relative humidity. For the three artificial intelligence models, results of input combination 6 were better than other combinations that with relative humidity or wind speed alone. Comparisons between the three artificial

intelligence models under the same input combination indicated that BAXGB (RMSE = 0.116–0.438 mm·d<sup>-1</sup> and MAE = 0.087–0.311 mm·d<sup>-1</sup>) always performed better than MARS (RMSE = 0.153–0.512 mm·d<sup>-1</sup> and MAE = 0.221–0.559 mm·d<sup>-1</sup>) and GPR (RMSE = 0.297–0.695 mm·d<sup>-1</sup> and MAE = 0.120–0.379 mm·d<sup>-1</sup>) under each input combination. In addition, the MARS model generally performed better than the GPR model, except in combinations 2 and 3. Across algorithms and input combinations, values of RMSE<sub>max10%</sub> and MAE<sub>max10%</sub> during the testing stage were 2–2.5 times of their corresponding values of RMSE and MAE, suggesting that model performance in extreme cases would be worse than normal conditions. Although the overall performance of the BAXGB model was good, the maximum error point of 10% with input combination 3 (i.e., BAXGB3) exceeded 1.0 mm·d<sup>-1</sup>. This may be due to the difference in distribution of solar radiation data between the training and the testing stage.

Figure 4 presents scatter diagrams between the estimated ETo by the three artificial intelligence algorithms and the P-M ETo (mm·d<sup>-1</sup>) during the testing stage for the Datong station. In Figure 4, the  $R^2$  value ranged at 0.832–0.978, 0.953–0.992, and 0.955–0.996 for GPR, MARS, and BAXGB for all input combinations, respectively. The slopes of the GPR models were all less than 0.95 except with the input combination 3, which led to a problem of underestimation. These results clearly indicated that, for the Datong station, BAXGB was the most superior, while GPR was the worst among the AI algorithms, in terms of the performance in ETo estimation.

Figure 5 presents comparisons between the estimated ETo by AI algorithms and the P-M ETo (mm·d<sup>-1</sup>) during the testing stage for the Datong station. P-M ETo peaked in May with a value over 5 mm·d<sup>-1</sup> while had the lowest value in December (0.6 mm·d<sup>-1</sup>). For the GPR model, the simulation effect of input combination 1 did not catch the trend well, resulting in a deviation more than 1 mm·d<sup>-1</sup>, presented as underestimation in May and overestimation in September, respectively. For the MARS model, input combinations 2 and 3 had a problem of overall overestimation, especially for months from August to October where the overestimation was more than 0.4 mm·d<sup>-1</sup>. For the BAXGB model, all input combination simulated an acceptable trend, except for the input combination 3 where an overestimation (more than 0.4 mm·d<sup>-1</sup>) was detected from July to October. Overall, among the AI methods in this study, the BAXGB model had the smallest deviation in all the 12 months, while GPR models performed worst.

To further evaluate the performance of different AI models, Taylor diagrams were used to analyze the performance of models under different combinations, and the results of the Datong station can be seen in Figure 6. It is not hard to see that combination 4 performs best in the GPR models, while combination 5 performs best in the MARS models. However, the combination 6 is superior to others in the BAXGB models. In addition, the point of the BAXGB model shows much closer than the other two models. This is consistent with the previous statistical indicators and scatter plot results.

**3.2. Model Performance in the Yinchuan Station.** The performance of the three artificial intelligence models in the



TABLE 4: Results of the statistical indicators for all models in the Datong station during the training and testing stages.

Model	Training					Testing				
	$R^2$	RMSE	MAE	RMSE <sub>max10%</sub>	MAE <sub>max10%</sub>	$R^2$	RMSE	MAE	RMSE <sub>max10%</sub>	MAE <sub>max10%</sub>
GPR1	0.841	0.66	0.519	1.392	1.373	0.832	0.695	0.559	1.333	1.328
GPR2	0.978	0.258	0.196	0.586	0.563	0.972	0.372	0.302	0.744	0.738
GPR3	0.98	0.242	0.181	0.565	0.547	0.956	0.444	0.35	0.898	0.891
GPR4	0.974	0.279	0.202	0.673	0.645	<b>0.978</b>	<b>0.297</b>	<b>0.221</b>	<b>0.668</b>	<b>0.664</b>
GPR5	<b>0.988</b>	<b>0.197</b>	<b>0.138</b>	<b>0.483</b>	<b>0.46</b>	0.951	0.398	0.251	1.089	1.019
GPR6	0.986	0.209	0.139	0.548	0.517	0.958	0.37	0.25	0.919	0.877
MARS1	0.964	0.315	0.236	0.723	0.71	0.97	0.292	0.227	0.649	0.631
MARS2	0.986	0.199	0.15	0.465	0.453	0.962	0.421	0.337	0.821	0.812
MARS3	0.984	0.209	0.157	0.481	0.468	0.953	0.512	0.379	1.158	1.148
MARS4	0.99	0.162	0.121	0.371	0.363	0.98	0.271	0.216	0.559	0.552
MARS5	0.992	0.154	0.11	0.377	0.362	<b>0.992</b>	<b>0.153</b>	<b>0.12</b>	<b>0.321</b>	<b>0.32</b>
MARS6	<b>0.994</b>	<b>0.122</b>	<b>0.087</b>	<b>0.3</b>	<b>0.292</b>	0.992	0.189	0.121	0.509	0.483
BAXGB1	0.978	0.238	0.179	0.541	0.531	0.972	0.291	0.23	0.592	0.587
BAXGB2	0.988	0.209	0.148	0.518	0.505	0.974	0.284	0.227	0.576	0.568
BAXGB3	0.998	0.076	0.057	0.175	0.168	0.955	0.438	0.311	1.067	1.057
BAXGB4	0.996	0.092	0.069	0.214	0.207	0.988	0.191	0.153	0.403	0.398
BAXGB5	0.996	0.108	0.076	0.266	0.255	0.992	0.146	0.112	0.337	0.328
BAXGB6	<b>1</b>	<b>0.03</b>	<b>0.022</b>	<b>0.068</b>	<b>0.066</b>	<b>0.996</b>	<b>0.116</b>	<b>0.087</b>	<b>0.272</b>	<b>0.258</b>

Note. Bold font indicates the best model.

training and testing stages for the Yinchuan station is compared in Table 5. In the first two input combinations, combination 2 had higher accuracy than combination 1 for all the three algorithms used in Yinchuan, which was different from Datong. However, the rank of importance of meteorological factors in different models showed similar trends with the Datong station. In addition, the combination including both wind speed and relative humidity (i.e., input combination 6) was still better than either factor alone for MARS and BAXGB models. It can be seen from the comparison of the three artificial intelligence models under the same input combination that BAXGB (RMSE = 0.175–0.261 mm·d<sup>-1</sup> and MAE = 0.123–0.204 mm·d<sup>-1</sup>) performed better than MARS (RMSE = 0.172–0.340 mm·d<sup>-1</sup> and MAE = 0.108–0.266 mm·d<sup>-1</sup>) and GPR (RMSE = 0.289–0.663 mm·d<sup>-1</sup> and MAE = 0.195–0.511 mm·d<sup>-1</sup>) under each combination. In general, the BAXGB model was better than the MARS model, and the GPR model was the worst, except for some individual combinations. Similar to the Datong station, RMSE<sub>max10%</sub> and MAE<sub>max10%</sub> during the testing stage were more than twice as high as the corresponding RMSE and MAE in the Yinchuan station. However, compared with Datong, values of RMSE<sub>max10%</sub> and MAE<sub>max10%</sub> were generally smaller in Yinchuan.

Figure S1 presents scatter diagrams between the estimated ETo and the P-M ETo (mm d<sup>-1</sup>) during the testing stage for the Yinchuan station. In Figure S1, the  $R^2$  value ranged at 0.871–0.974, 0.974–0.994, and 0.978–0.988 for GPR, MARS, and BAXGB among the input combinations, respectively. The slopes of the GPR models were less than 0.95 for input combinations 1, 3, 4, and 6, which led to a problem of underestimation. The slopes of MARS models were greater than 1 in all input combinations except the combination 3, suggesting a problem of overestimation. Overall, BAXGB and MARS models were superior to GPR models in the Yinchuan station.

Comparisons between the estimated ETo and the P-M ETo during the testing stage for the three AI algorithms in Yinchuan are presented in Figure S2. P-M ETo had a peak in June (nearly 5 mm·d<sup>-1</sup>) but was generally small in winter (averaged at about 1 mm·d<sup>-1</sup>). For GPR models, the deviations of input combinations 1 and 5 were the most obvious, in which the combination 1 was seriously overestimated by 0.6 mm·d<sup>-1</sup> or more from August to November, while the combination 5 was overvalued every month at a range of 0.2–0.6 mm·d<sup>-1</sup>. For MARS models, input combinations were performed similarly to the GPR model, but the extent of overestimation was less. In addition, MARS models in combinations 1 and 2 were significantly overestimated from May to July. For the BAXGB model, it had the smallest deviations (less than 0.4 mm·d<sup>-1</sup>) in each month among the three AI algorithms.

**3.3. Model Performance in the Taiyuan Station.** The performance of the three artificial intelligence algorithms in the training and testing stages for the Taiyuan station is compared in Table 6. The statistical results in Taiyuan were similar to those in Datong because Datong is geographically closer to Taiyuan when compared with Yinchuan. The most significant difference between Taiyuan and Datong was that MARS models showed a difference in the importance rank of different meteorological parameters. However, the BAXGB model showed that solar radiation was the most important, followed by wind speed and relative humidity. Considering that each statistical index of the BAXGB model was better than the other two models, it is fair to suggest that solar radiation in this area was more important than other meteorological parameters.

Figure S3 presents scatter diagrams between the estimated ETo and the P-M ETo (mm d<sup>-1</sup>) during the testing stage for the Yinchuan station. In Figure S3, the  $R^2$  value ranged at 0.817–0.981, 0.935–0.986, and 0.940–0.987 for

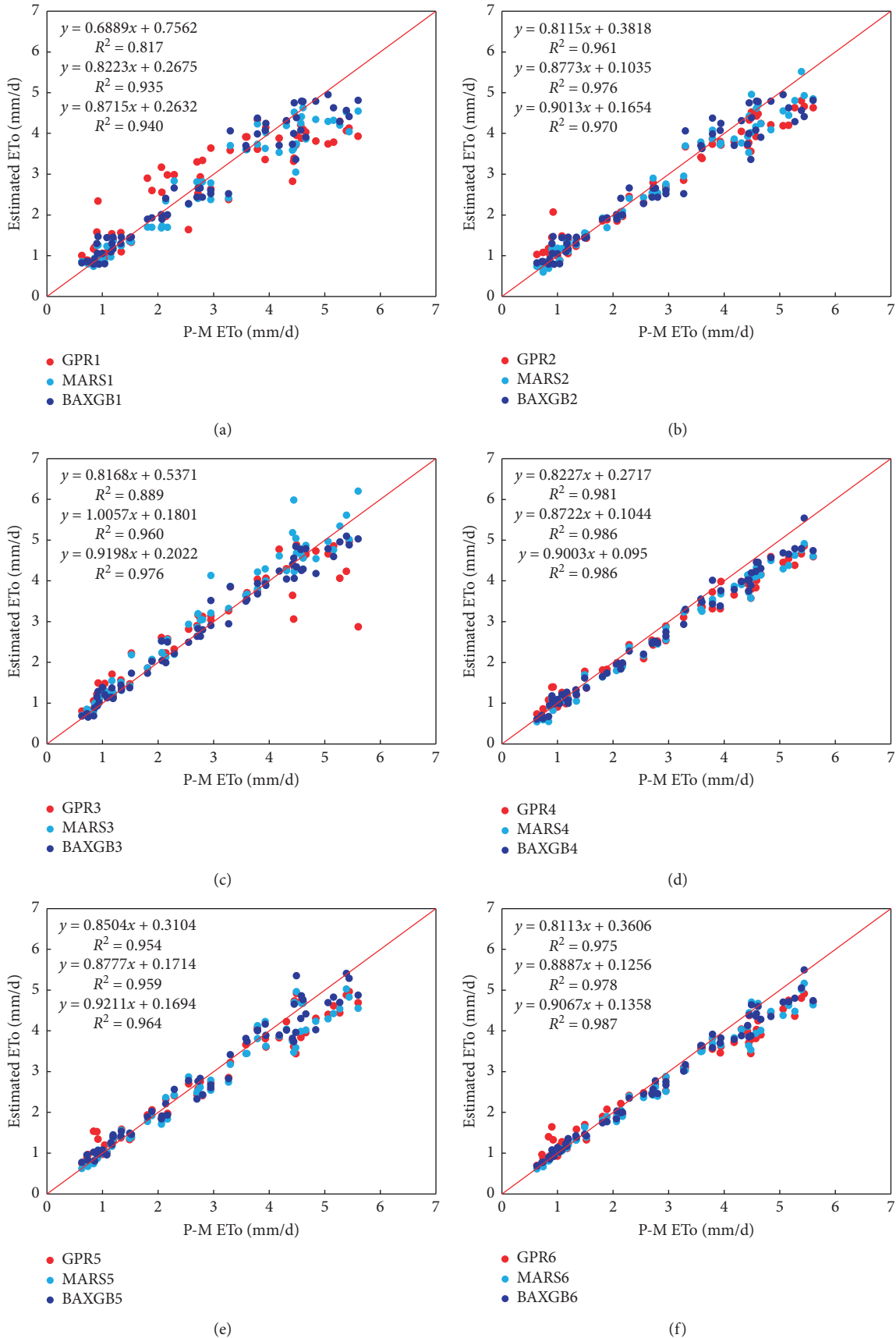


FIGURE 4: Scatter plots between the estimated ETo by the three algorithms and the FAO-56 P-M ETo in the testing stage for the Datong station.

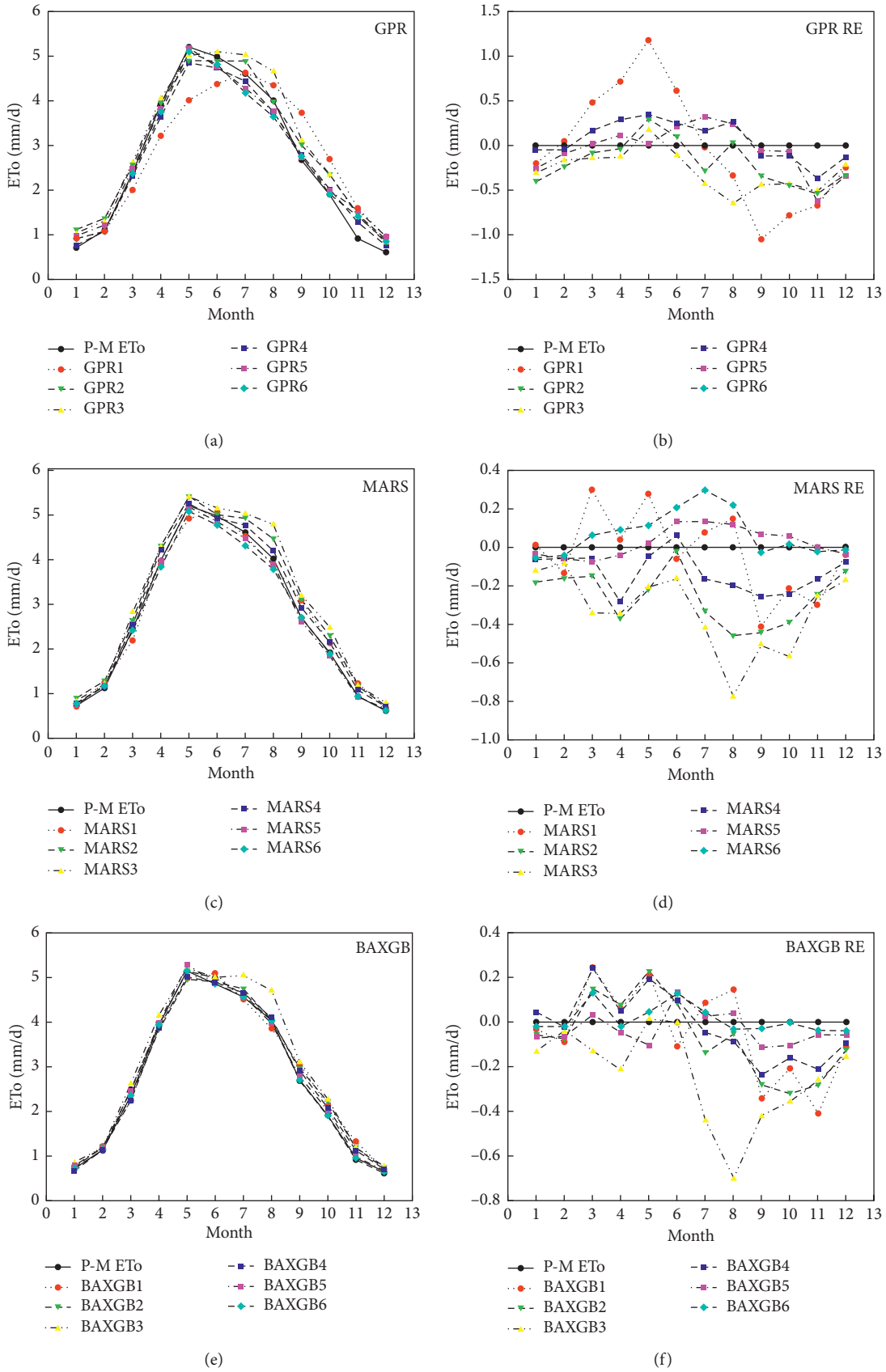


FIGURE 5: Relative errors by the different algorithms at each month for the Datong station.



TABLE 5: Results of the statistical indicators for all models in the Yinchuan station during the training and testing stages.

Model	Training					Testing				
	$R^2$	RMSE	MAE	RMSE <sub>max10%</sub>	MAE <sub>max10%</sub>	$R^2$	RMSE	MAE	RMSE <sub>max10%</sub>	MAE <sub>max10%</sub>
GPR1	0.856	0.653	0.514	1.371	1.351	0.87	0.663	0.511	1.424	1.401
GPR2	0.982	0.248	0.191	0.568	0.548	0.974	0.334	0.259	0.73	0.719
GPR3	0.984	0.23	0.176	0.53	0.518	0.974	0.289	0.197	0.751	0.696
GPR4	0.98	0.257	0.196	0.588	0.574	0.966	<b>0.343</b>	<b>0.262</b>	<b>0.752</b>	<b>0.736</b>
GPR5	0.984	0.23	0.169	0.536	0.519	<b>0.972</b>	0.461	0.378	0.941	0.907
GPR6	<b>0.986</b>	<b>0.208</b>	<b>0.147</b>	<b>0.507</b>	<b>0.491</b>	0.97	0.384	0.285	0.888	0.878
MARS1	0.968	0.3	0.229	0.677	0.666	0.974	0.34	0.262	0.728	0.723
MARS2	0.988	0.187	0.142	0.422	0.409	0.984	0.262	0.194	0.609	0.602
MARS3	0.986	0.203	0.153	0.472	0.466	0.984	0.227	0.184	0.461	0.453
MARS4	0.992	0.159	0.12	0.368	0.36	0.986	0.219	0.163	0.523	0.513
MARS5	0.992	0.155	0.117	0.348	0.344	0.99	0.323	0.266	0.687	0.676
MARS6	<b>0.996</b>	<b>0.105</b>	<b>0.07</b>	<b>0.275</b>	<b>0.265</b>	<b>0.994</b>	<b>0.172</b>	<b>0.108</b>	<b>0.464</b>	<b>0.458</b>
BAXGB1	0.978	0.261	0.189	0.614	0.594	0.978	0.261	0.204	0.58	0.576
BAXGB2	0.988	0.203	0.146	0.49	0.468	0.982	0.229	0.173	0.524	0.51
BAXGB3	0.99	0.176	0.13	0.412	0.402	0.986	0.19	0.142	0.416	0.409
BAXGB4	<b>1</b>	<b>0.051</b>	<b>0.038</b>	<b>0.116</b>	<b>0.113</b>	<b>0.988</b>	0.204	0.156	0.467	0.452
BAXGB5	0.992	0.205	0.145	0.502	0.487	<b>0.988</b>	0.222	0.158	0.534	0.532
BAXGB6	0.996	0.162	0.113	0.393	0.378	<b>0.988</b>	<b>0.175</b>	<b>0.123</b>	<b>0.412</b>	<b>0.401</b>

Note. Bold font indicates the best model.

TABLE 6: Results of the statistical indicators for all models in the Taiyuan station during the training and testing stages.

Model	Training					Testing				
	$R^2$	RMSE	MAE	RMSE <sub>max10%</sub>	MAE <sub>max10%</sub>	$R^2$	RMSE	MAE	RMSE <sub>max10%</sub>	MAE <sub>max10%</sub>
GPR1	0.812	0.605	0.466	1.34	1.311	0.817	0.714	0.564	1.457	1.451
GPR2	0.972	0.248	0.19	0.561	0.55	0.96	0.418	0.311	0.949	0.942
GPR3	0.97	0.244	0.18	0.594	0.564	0.889	0.534	0.317	1.487	1.328
GPR4	0.972	0.243	0.176	0.582	0.564	<b>0.98</b>	0.403	0.31	<b>0.835</b>	<b>0.827</b>
GPR5	0.976	0.223	0.16	<b>0.545</b>	<b>0.534</b>	0.955	<b>0.391</b>	<b>0.297</b>	0.85	0.842
GPR6	<b>0.978</b>	<b>0.22</b>	<b>0.152</b>	0.559	0.546	0.976	0.397	0.301	0.858	0.851
MARS1	0.949	0.317	0.24	0.712	0.703	0.935	0.497	0.351	1.138	1.115
MARS2	0.97	0.241	0.187	0.542	0.534	0.976	0.377	0.275	0.865	0.853
MARS3	0.974	0.224	0.163	0.551	0.52	0.96	0.377	0.259	0.954	0.884
MARS4	0.984	0.178	0.132	0.422	0.406	<b>0.986</b>	0.364	0.284	<b>0.811</b>	<b>0.803</b>
MARS5	0.982	0.189	0.137	0.453	0.438	0.958	0.385	0.282	0.868	0.859
MARS6	<b>0.988</b>	<b>0.151</b>	<b>0.101</b>	<b>0.39</b>	<b>0.361</b>	0.978	<b>0.33</b>	<b>0.22</b>	0.817	0.809
BAXGB1	0.972	0.241	0.181	0.555	0.542	0.939	0.412	0.302	0.928	0.918
BAXGB2	0.994	0.115	0.087	0.266	0.256	0.97	0.314	0.24	0.712	0.699
BAXGB3	0.984	0.227	0.16	0.556	0.532	0.976	0.259	0.193	0.581	0.58
BAXGB4	0.996	0.097	0.071	0.232	0.221	0.986	0.297	0.235	0.643	0.632
BAXGB5	0.996	0.082	0.063	0.186	0.182	0.964	0.308	0.237	0.722	0.715
BAXGB6	<b>1</b>	<b>0.036</b>	<b>0.026</b>	<b>0.083</b>	<b>0.08</b>	<b>0.988</b>	<b>0.253</b>	<b>0.182</b>	<b>0.594</b>	<b>0.578</b>

Note. Bold font indicates the best model.

Figure S4 presents comparison between the estimated ETo and the P-M ETo ( $\text{mm d}^{-1}$ ) during the testing stage for the three AI algorithms in the Taiyuan station. The monthly fluctuation of P-M ETo in Taiyuan was similar to that of Datong, with the peak appeared in May (nearly  $5 \text{ mm d}^{-1}$ ) as well. For the GPR model, the combination 1 was underestimated in spring but overestimated in autumn, the maximum deviation of which was over  $1 \text{ mm d}^{-1}$ . For MARS models, the input combination 3 had a severe underestimation in April, while other combinations had overestimations at spring and summer. Although there was an overall trend of underestimation for the BAXGB models, combinations 1, 3, and 4 were underestimated by more than  $0.4 \text{ mm d}^{-1}$  in some months (in spring usually).

**3.4. Comparison between Different Algorithms.** Overall, the three AI algorithms presented different optimal combinations at different stations. For example, GPR models in combination 4 yielded better performance than in combination 6 across the three stations. On the contrary, BAXGB models identified that the combination 6 was the optimal input combination. This indicated that the BAXGB algorithm might have more advantages in identifying the contributions of multiple parameters compared with the other two algorithms, and therefore, its accuracy was the best.

Given the fact that BAXGB overall performed better than GPR and MARS across the three stations and the six input combinations, statistical results of BAXGB can be used for analyzing the importance of each major meteorological

parameter for estimating ETo. Based on the results of BAXGB models in this study, solar radiation is more important than the others in Yinchuan and Taiyuan stations. This is similar to other studies [24, 36], where they show that solar radiation is more important than wind speed and relative humidity in a semiarid environment of Qilian Mountain of Northwest China and Southeast China. However, relative humidity and wind speed seem more important than solar radiation for the Datong station with BAXGB models in this study, which may be related to the difference in the distribution of solar radiation data between the training and testing stages. Compared with the input combinations 4 and 5, statistical indicators of the input combination 3 are smaller during the training stage. These differences are much larger in the testing stage than the training stage, especially for values of  $RMSE_{max10\%}$  and  $MAE_{max10\%}$ .

Tao et al. [28] developed a coupling model based on the firefly algorithm and ANFIS in Burkina Faso, a semiarid region of Africa, in which the RMSE in the testing period ranged at  $0.24\text{--}1.10\text{ mm}\times\text{d}^{-1}$  with different input combinations. Shamsirband [37] found that the cuckoo search algorithm optimized ANN and ANFIS had an RMSE value of 0.330 and  $0.265\text{ mm d}^{-1}$  in the testing stage at Serbia, respectively. Another case study in Serbia showed that the genetic programming models coupling with GA generated RMSE at  $0.649\text{--}1.188$  and  $0.657\text{--}1.193\text{ mm d}^{-1}$  in the testing stage and the validation, respectively [38]. In addition, a hybrid model (GA-SVM) conducted in the Pailugou Watershed, Northwest China, showed that the RMSE values ranged at  $0.138\text{--}0.424\text{ mm d}^{-1}$  [21, 36]. Overall, RMSE values obtained in this study are similar to those achieved by other researchers using coupled artificial intelligence models in semiarid regions. Furthermore, in the case of limited observational data in this region, BAXGB models with only temperature and extraterrestrial radiation as inputs would offer an acceptable result and can be used as an alternative method for ETo estimation. On this basis, adding one or two meteorological factors would also be beneficial.

Over the past decade, many highly efficient evolutionary algorithms have been proposed. The XGBoost is also a relatively new technology, and the study of combining evolutionary algorithms with XGBoost is very limit to date. This study confirms that the bat algorithm is an effective method for optimizing XGBoost. However, whether other evolutionary algorithms would be more efficient than BA for XGBoost optimization needs to be further studied.

#### 4. Conclusions

Accurate and robust reference evapotranspiration (ETo) estimating models are of importance for agriculture and irrigation scheduling. In this study, a new model (BAXGB) that couples a bat algorithm with XGBoost was proposed for the estimation of monthly ETo. The developed BAXGB algorithm was studied by comparing with the FAO 56 Penman–Monteith model results in arid and semiarid environments of China. Data from three stations (Datong, Yinchuan, and Taiyuan) were used to build the BAXGB

model as well as the multivariate adaptive regression splines (MARS) and the gaussian process regression (GPR) models. Six models for each algorithm were developed using different input combinations including mean air temperature, maximum air temperature, minimum air temperature, wind speed, relative humidity, and solar radiation or extraterrestrial radiation. The results showed that the BAXGB models exhibited a suitable accuracy for estimating ETo. Except for the input combination 6 ( $T_{max}$ ,  $T_{min}$ ,  $U$ , and RH), BAXGB models were superior to MARS models. The GPR model was generally worse than the MARS model. For the Datong station, the influence of solar radiation on model accuracy was less than that of wind speed and relative humidity. However, for the other two stations, there was an opposite trend. Findings of this research will be helpful for the agricultural and irrigation scheduling and may offer reference to other regions where accurate models for improving the local water management are needed.

#### Data Availability

The statistical results are within the manuscript and its Supporting Information file. However, raw data underlying the results presented in the study are available from the National Meteorological Information Center (NMIC) of China Meteorological Administration (CMA) (<http://data.cma.cn/>).

#### Conflicts of Interest

The authors declare that there are no conflicts of interest.

#### Authors' Contributions

Y. Han did conceptualization, writing, and original draft preparation. J. Wu was involved in funding acquisition. B. Zhai edited the manuscript. G. Huang was responsible for methodology and validation. Y. Pan cured the data. L. Wu performed the analysis using software. W. Zeng revised the manuscript.

#### Acknowledgments

The authors would like to thank the China Meteorological Administration (CMA) for providing the meteorological and radiation data. This research was funded by National Natural Science Foundation of China (nos. 51790533, 51709144, and 31570444) and the Natural Science Foundation of Jiangxi Province, China (no. 20171BAB216051).

#### Supplementary Materials

Figure S1: scatter plots between the estimated ETo by the three algorithms and the FAO-56 P-M ETo in the testing stage for the Yinchuan station. Figure S2: relative errors by the different algorithms at each month for the Yinchuan station. Figure S3: scatter plots between the estimated ETo by the three algorithms and the FAO-56 P-M ETo in the testing stage for the Taiyuan station. Figure S4: relative errors by the

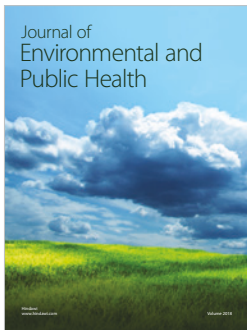
different algorithms at each month for the Taiyuan station. (*Supplementary Materials*)

## References

- [1] R. G. Allen, L. S. Pereira, D. Raes et al., *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements-FAO Irrigation and Drainage Paper*, Vol. 56, FAO, Rome, Italy, 1998.
- [2] J. Fan, L. Wu, F. Zhang et al., "Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: a review and case study in China," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 186–212, 2019.
- [3] J. Fan, L. Wu, F. Zhang, H. Cai, X. Ma, and H. Bai, "Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 168–186, 2019.
- [4] J. Fan, X. Wang, L. Wu et al., "Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China," *Energy Conversion and Management*, vol. 164, pp. 102–111, 2018.
- [5] X. Ma, X. Mei, W. Wu, X. Wu, and B. Zeng, "A novel fractional time delayed grey model with grey wolf optimizer and its applications in forecasting the natural gas and coal consumption in Chongqing China," *Energy*, vol. 178, pp. 487–507, 2019.
- [6] X. Ma, W. Wu, B. Zeng, Y. Wang, and X. Wu, "The conformable fractional grey system model," *ISA Transactions*, 2019, In press.
- [7] Z. M. Yaseen, S. M. Awadh, A. Sharafati, and S. Shahid, "Complementary data-intelligence model for river flow simulation," *Journal of Hydrology*, vol. 567, pp. 180–190, 2018.
- [8] Q. Xiao, C. Li, Y. Tang, L. Li, and L. Li, "A knowledge-driven method of adaptively optimizing process parameters for energy efficient turning," *Energy*, vol. 166, pp. 142–156, 2019.
- [9] M. Ali, R. C. Deo, N. J. Downs, and T. Maraseni, "Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting," *Computers and Electronics in Agriculture*, vol. 152, pp. 149–165, 2018.
- [10] L. Wu, G. Huang, J. Fan, F. Zhang, X. Wang, and W. Zeng, "Potential of kernel-based nonlinear extension of Arps decline model and gradient boosting with categorical features support for predicting daily global solar radiation in humid regions," *Energy Conversion and Management*, vol. 183, pp. 280–295, 2019.
- [11] S. Mehdizadeh, "Assessing the potential of data-driven models for estimation of long-term monthly temperatures," *Computers and Electronics in Agriculture*, vol. 144, pp. 114–125, 2018.
- [12] A. Azad, M. Manoochehri, H. Kashi et al., "Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling," *Journal of Hydrology*, vol. 571, pp. 214–224, 2019.
- [13] R. Prasad, R. C. Deo, Y. Li, and T. Maraseni, "Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors," *Soil and Tillage Research*, vol. 181, pp. 63–81, 2018.
- [14] L. Wu, H. Zhou, X. Ma, J. Fan, and F. Zhang, "Daily reference evapotranspiration prediction based on hybridized extreme learning machine model with bio-inspired optimization algorithms: application in contrasting climates of China," *Journal of Hydrology*, vol. 577, article 123960, 2019.
- [15] L. Wu and J. Fan, "Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration," *PLoS One*, vol. 14, no. 5, Article ID e0217520, 2019.
- [16] G. Huang, L. Wu, X. Ma et al., "Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions," *Journal of Hydrology*, vol. 574, pp. 1029–1041, 2019.
- [17] L. Wu, X. Lu, X. Liu et al., "Simulation of reference crop evapotranspiration by using bat algorithm optimization based extreme learning machine," *Journal of Drainage and Irrigation Machinery Engineering (JDIME)*, vol. 36, no. 9, pp. 802–805, 2018, in Chinese.
- [18] S. Trajkovic, B. Todorovic, and M. Stankovic, "Forecasting of reference evapotranspiration by artificial neural networks," *Journal of Irrigation and Drainage Engineering*, vol. 129, no. 6, pp. 454–457, 2003.
- [19] O. Kisi, "Modeling reference evapotranspiration using three different heuristic regression approaches," *Agricultural Water Management*, vol. 169, pp. 162–172, 2016.
- [20] H. Sanikhani, O. Kisi, E. Maroufpoor, and Z. M. Yaseen, "Temperature-based modeling of reference evapotranspiration using several artificial intelligence models: application of different modeling scenarios," *Theoretical and Applied Climatology*, vol. 135, no. 1–2, pp. 449–462, 2019.
- [21] M. K. Saggi and S. Jain, "Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning," *Computers and Electronics in Agriculture*, vol. 156, pp. 387–398, 2019.
- [22] G. Landaras, E. Bekoe, J. Ampofo et al., "New alternatives for reference evapotranspiration estimation in West Africa using limited weather data and ancillary data supply strategies," *Theoretical and Applied Climatology*, vol. 132, no. 3–4, pp. 701–716, 2018.
- [23] Y. Feng, Y. Jia, N. Cui, L. Zhao, C. Li, and D. Gong, "Calibration of Hargreaves model for reference evapotranspiration estimation in Sichuan basin of southwest China," *Agricultural Water Management*, vol. 181, pp. 1–9, 2017.
- [24] J. Fan, W. Yue, L. Wu et al., "Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China," *Agricultural and Forest Meteorology*, vol. 263, pp. 225–241, 2018.
- [25] M. Karbasi, "Forecasting of multi-step ahead reference evapotranspiration using wavelet- Gaussian process regression model," *Water Resources Management*, vol. 32, no. 3, pp. 1035–1052, 2018.
- [26] Y. Feng, Y. Jia, Q. Zhang, D. Gong, and N. Cui, "National-scale assessment of pan evaporation models across different climatic zones of China," *Journal of Hydrology*, vol. 564, pp. 314–328, 2018.
- [27] S. S. Eslamian, S. A. Gohari, M. J. Zareian, and A. Firoozfar, "Estimating Penman-Monteith reference evapotranspiration using artificial neural networks and genetic algorithm: a case study," *Arabian Journal for Science and Engineering*, vol. 37, no. 4, pp. 935–944, 2012.
- [28] H. Tao, L. Diop, A. Bodian, K. Djaman, P. M. Ndiaye, and Z. M. Yaseen, "Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: regional case study in Burkina Faso," *Agricultural Water Management*, vol. 208, pp. 140–151, 2018.

- [29] D. Petković, M. Gocic, S. Shamshirband, S. N. Qasem, and S. Trajkovic, "Particle swarm optimization-based radial basis function network for estimation of reference evapotranspiration," *Theoretical and Applied Climatology*, vol. 125, no. 3-4, pp. 555-563, 2016.
- [30] J. Fan, L. Wu, X. Ma, H. Zhou, and F. Zhang, "Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions," *Renewable Energy*, vol. 145, pp. 2034-2045, 2020.
- [31] L. Xiao, F. Qian, and W. Shao, "Multi-step wind speed forecasting based on a hybrid forecasting architecture and an improved bat algorithm," *Energy Conversion and Management*, vol. 143, pp. 410-430, 2017.
- [32] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1-67, 1991.
- [33] C. E. Rasmussen, *Gaussian Processes in Machine Learning. Summer School on Machine Learning*, Springer, Berlin, Germany, 2003.
- [34] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, ACM, San Francisco, CA, USA, August 2016.
- [35] X. S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*, pp. 65-74, Springer, Berlin, Germany, 2010.
- [36] Z. Yin, X. Wen, Q. Feng, Z. He, S. Zou, and L. Yang, "Integrating genetic algorithm and support vector machine for modeling daily reference evapotranspiration in a semi-arid mountain area," *Hydrology Research*, vol. 48, no. 5, pp. 1177-1191, 2017.
- [37] S. Shamshirband, M. Amirmojahedi, M. Gocić et al., "Estimation of reference evapotranspiration using neural networks and cuckoo search algorithm," *Journal of Irrigation and Drainage Engineering*, vol. 142, no. 2, article 04015044, 2015.
- [38] S. Jovic, B. Nedeljkovic, Z. Golubovic, and N. Kostic, "Evolutionary algorithm for reference evapotranspiration analysis," *Computers and Electronics in Agriculture*, vol. 150, pp. 1-4, 2018.





Hindawi

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

