Daily evapotranspiration estimation using limited meteorological data across diverse geographic regions of China

Wenxu Chen[1], Fei Wang[2], Xingcan Yuwen[2], Yi Shi[2], Xinbo Zhao[3*]

(1. *College of Cyber Security, Tarim University, Aral 843300, Xinjiang, China*;
2. *College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471000, China*;
3. *School of Agriculture, Nanjing Agricultural University, Nanjing 210095, China*)

Abstract: Evapotranspiration (ET) is a key component of the water cycle, and accurate estimation of reference crop evapotranspiration (ET_o) is essential for irrigation management. To build a precise, lightweight ET_o estimation model, this study takes key meteorological factors as inputs and applies machine learning models and hybrid models Crested Porcupine Optimizer kernel extreme learning machine (CPO-KELM), the Dung Beetle Optimizer Algorithm KELM (DBO-KELM), and Particle Swarm Optimization KELM (PSO-KELM) to estimate ET_o at 38 meteorological stations across China's seven major geographical regions. The results indicate that maximum temperature (T_{max}), average temperature (T_{ave}), and relative humidity (RH) are the primary factors affecting ET_o and were therefore used as model inputs. The standalone kernel extreme learning machine (KELM) model shows acceptable ET_o estimation performance, with R^2, RMSE, MAE, and NSE ranging from 0.802-0.885, 0.512-0.911, 0.464-0.970, and 0.802-0.885, respectively. Hybrid models outperform the standalone KELM, among which CPO-KELM is the most accurate: its R^2, RMSE, MAE, and NSE range from 0.881-0.942, 0.413-1.147, 0.284-0.763, and 0.881-0.942. At the regional scale, the CPO-KELM model exhibits its best performance in the Northeast and North China regions, with R^2, RMSE, MAE, and NSE ranging from 0.923-0.936, 0.413-0.511, 0.284-0.358, and 0.923-0.936, respectively. In contrast, its weakest performance is observed in parts of South China and Northwest China, with R^2, RMSE, MAE, and NSE ranging from 0.881-0.905, 0.675-1.147, 0.506-0.763, and 0.881-0.905. Compared to standalone KELM, CPO-KELM improves accuracy significantly: R^2 and NSE rise by 6.4%-9.9%, while MAE drops by 21.3%-38.8%. Thus, the hybrid CPO-KELM model effectively enhances ET_o estimation accuracy across China's regions. Therefore, the proposed CPO-KELM hybrid model provides a high-accuracy and lightweight alternative for ET_o estimation in data-scarce regions, and offers reliable technical support for intelligent water resources management and irrigation optimization across diverse climatic zones in China.

Keywords: ET_o, seven major geographical regions, machine learning algorithm, lightweight model
DOI: 10.25165/j.ijabe.20261901.10161

Citation: Chen W X, Wang F, Yu W X C, Shi Y, Zhao X B. Daily evapotranspiration estimation using limited meteorological data across diverse geographic regions of China. Int J Agric & Biol Eng, 2026; 19(1): 241–250.

1 Introduction

ET_o serves as a fundamental parameter in real-time irrigation forecasting and the efficient management of agricultural water resources[1]. It plays a critical role in assessing crop water requirements, monitoring soil moisture dynamics, evaluating irrigation efficiency, and forecasting crop growth. Furthermore, ET_o is inherently linked to atmospheric circulation, since its key driving variables—temperature, humidity, wind speed, and solar radiation—are shaped by large-scale climate dynamics[2]. Accurate ET_o prediction plays a crucial role in aiding agricultural producers to make informed decisions[3].

The Penman-Monteith (PM) model is the standard method recommended by the Food and Agriculture Organization (FAO) for estimating ET_o. This model can provide relatively accurate estimates of ET_o for common crops. However, this model relies on numerous input parameters, and the limited availability of weather stations in many regions hinders the acquisition of comprehensive meteorological data necessary for accurate ET_o estimation[4,5].

In recent years, machine learning (ML) models have been extensively applied for ET_o estimation. Yong et al.[6] utilized four types of meteorological data to evaluate the performance of three ML models in predicting ET_o, specifically light gradient boosting machine (LGBM), decision forest regression (DFR), and artificial neural network (ANN). The results indicate that both LGBM and ANN exhibit strong performance. Tejada et al.[7] employed various combinations of meteorological data to assess the performance of Support Vector Machines (SVMs) and Extreme Learning Machines (ELMs), and compared their performance with that of various empirical models. The results demonstrate that ML models using a limited set of meteorological parameters as inputs can yield more accurate daily ET_o estimates. Jimenez et al.[8] evaluated several ML models for ET_o prediction, and the results indicate that the Support Vector Machine (SVM), Random Forest (RF), and ELM outperformed baseline models at most locations.

Although ML algorithms demonstrate certain efficacy in ET_o estimation, the selection of model parameters often relies on trial and error, while the application of optimization algorithms to adjust these parameters can result in more accurate ET_o predictions[5,9]. Zhao et al.[10] combined the Golden Eagle Optimization Algorithm (GEO) and the Sparrow Search Algorithm (SSA) with the ELM model to optimize its parameters. The results indicate that the SSA-ELM model achieved the highest accuracy (R^2=0.998-0.999),

Received date: 2025-09-06 **Accepted date:** 2025-12-24
Biographies: Wenxu Chen, MS, research interest: agricultural engineering, Email: 120210091@taru.edu.cn; **Fei Wang**, MS, research interest: agricultural electrification, Email: wf01082021@163.com; **Xingcan Yuwen**, MS, research interest: agricultural electrification, Email: 15227166478@163.com; **Yi Shi**, PhD, research interest: agricultural electrification, Email: shiyigongteng@163.com.
***Corresponding author: Xinbo Zhao**, PhD, research interest: smart agriculture, School of Agriculture, Nanjing Agricultural University, Nanjing 210095, China. Tel: +86-19852866362, Email: 2023201129@stu.njau.edu.cn.

whereas the conventional ELM model had an accuracy range of R^2=0.804-0.983, with the hybrid model demonstrating higher accuracy and stability. Sebastian et al.[11] compared the adaptive controllers based on KELM, ELM, and Radial Basis Function Neural Network (RBFNN). The results indicate that KELM outperforms the other two algorithms. Tian et al.[12] employed the Genetic Algorithm (GA) to improve the Back Propagation Neural Network (BPNN), SVM, and KELM for cross-validation in predicting gas content. The results show that the average relative error for GASA-KELM is 10.6%, while for GASA-SVM, it is 13.04%, and for GASA-BPNN, it is 14.31%. GASA-KELM demonstrates higher accuracy, as well as stronger predictive and generalization capabilities. Zhao et al.[13] optimized the model parameters of KELM using three bio-inspired optimization algorithms, thereby improving the accuracy of ET_o predictions. KELM is widely used in research, and possesses notable advantages in regression and prediction tasks. Compared with conventional ELM, KELM incorporates kernel functions, which enhance its ability to capture nonlinear relationships between meteorological variables and ET_o, leading to higher prediction accuracy and better generalization performance[11–13]. Moreover, KELM has a relatively simple structure and fast training speed, making it suitable for large-scale and multi-site ET_o estimation across diverse climatic regions. Its performance, however, is sensitive to kernel parameters, which motivates the integration of metaheuristic optimization algorithms such as CPO, DBO, and PSO to further improve accuracy and robustness.

The performance of KELM is sensitive to kernel parameters. To address this limitation, metaheuristic optimization algorithms such as PSO, DBO, and CPO have been increasingly used to optimize model parameters and improve prediction accuracy. Fan et al.[14] optimized the model parameters of CNNL-STM-Attention (CLA) using the CPO algorithm, and the results indicate that CPO-CLA outperforms recent popular models in both prediction accuracy and stability. Zhang et al.[15] used the dung beetle optimization (DBO) algorithm to measure and analyze the WEFN resilience of the China Beidahuang Group. The results showed that the DBO-HKELM model has notable advantages over the FA-HKELM and NGO-HKELM models in terms of fitting accuracy, optimization rate, reliability, rationality, and robustness. Jia et al.[16] combined PSO with the Gaussian Exponential Model (GEM) and compared it with four other machine learning models, with the results showing that PSO-GEM achieved the highest accuracy across all climate conditions.

To build a lightweight and high-precision ET_o estimation model, this study used the Random Forest (RF) model to rank the feature importance of meteorological factors at the sites, identifying key factors influencing ET_o as inputs for the KELM model to estimate ET_o at 38 sites across China. Furthermore, CPO, DBO, and PSO optimization algorithms were combined with the KELM model to construct simplified ET_o models for 38 sites across China. The objectives of the study are: 1) to select key factors for estimating ET_o using the Random Forest (RF) algorithm; 2) to construct ET_o estimation models based on KELM biomimetic models and metaheuristic optimization algorithms, and to evaluate the model's simulation accuracy; 3) to assess the adaptability of the ET_o model across seven major geographical regions of China.

2 Materials and methods

2.1 Data sources

The basic meteorological data used in this study were sourced from the China Meteorological Data Network (http://data.cma.cn/). This study involves data from 38 meteorological stations in seven major geographical regions of China. Daily climatic variables, including atmospheric pressure (P), average temperature (T_{ave}), maximum and minimum air temperatures (T_{max}/T_{min}), average relative humidity (RH), wind speed at 2 m height (U_2), and sunshine hours (SH) from 1961 to 2016, were used to build the model, with 70% of the data used for training and 30% for testing. The daily mean meteorological data of 38 stations are listed in Table 1.

Table 1 The daily mean meteorological data for 38 stations

Station	Station number	P/kPa	T_{ave}/°C	T_{max}/°C	T_{min}/°C	RH/%	U_2/m·s^{-1}	SH/h
Aihui	50 468	97.28	–3.18	4.47	–11.66	0.69	1.87	6.61
Jiamusi	50 873	99.17	–0.73	5.9	–6.99	0.66	2.05	6.99
Harbin	50 953	94.25	–0.4	4.57	–8.74	0.67	2.06	7.25
Urumqi	51 463	99.31	0.72	6.65	–4.91	0.66	3.1	7.25
Hotan	51 828	98.47	0.71	7.25	–5.91	0.66	3.16	7.33
Dunhuang	52 418	94.9	1.68	7.76	–4.48	0.59	3.68	8.37
Minqin	52 681	93.88	0.36	6.75	–5.20	0.62	3.28	8.02
Xining	52 866	98.11	0.90	7.11	–4.85	0.69	2.65	7.06
Datong	53 487	89.62	–2.2	4.77	–9.40	0.69	2.51	7.15
Yinchuan	53 614	140.62	4.5	10.34	–1.46	0.56	3.06	7.36
Shanxi	53 772	99.42	2.87	9.03	–2.86	0.64	3.37	7.27
Houma	53 963	99.64	4.02	9.94	–1.48	0.6	3.23	7.75
Tongliao	54 135	98.5	2.12	7.74	–3.16	0.67	2.97	7.22
Changchun	54 161	98.41	3.37	8.43	–2.19	0.64	3.52	7.26
Yanji	54 292	98.48	1.26	8.4	–5.14	0.69	2.09	6.42
Shenyang	54 342	100.53	3.29	8.53	–1.89	0.68	3.68	6.67
Beijing	54 511	95.48	2.84	10.34	–3.98	0.57	2.82	7.75
Jinan	54 823	99.25	3.02	8.93	–2.5	0.67	3.53	7.38
Lhasa	55 591	98.84	1.91	8.34	–4.00	0.71	2.40	6.64
Yushu	56 029	100.36	3.64	9.49	–2.04	0.66	3.15	6.69
Changdu	56 137	100.2	3.71	9.52	–1.67	0.69	3.89	6.66
Emeishan	56 385	100.37	4.12	9.68	–1.32	0.64	3.56	6.82
Tengchong	56 739	99.57	5.35	11.88	–1.16	0.56	3.42	7.95
Kunming	56 778	99.79	4.93	10.97	–0.59	0.61	3.63	8.01
Yichang	57 461	99.66	4.49	10.22	–1.00	0.65	3.34	6.74
Wuhan	57 494	100.13	2.76	9.17	–3.21	0.73	3.75	6.69
Guiyang	57 816	99.21	3.13	9.83	–2.97	0.72	2.84	6.73
Guilin	57 957	98.48	6.32	9.85	–0.62	0.61	3.41	6.56
Ganxian	57 993	98.46	4.43	10.05	–0.96	0.64	3.33	7.08
Gushi	58 208	100.19	3.71	8.88	–1.36	0.7	2.96	6.6
Nanjing	58 238	95.76	5.00	10.96	–0.3	0.6	3.83	7.94
Hangzhou	58 457	96.17	4.40	11.18	–1.8	0.62	2.57	7.97
Nanchang	58 606	93.43	5.58	10.87	–1.28	0.58	2.27	8.21
Fuzhou	58 847	92.64	3.18	11.01	–3.48	0.59	1.77	7.82
Luogang	59 287	95.66	7.29	14.28	1.34	0.59	2.3	8.03
Shantou	59 316	87.24	3.89	10.27	–1.39	0.53	2.52	8.05
Nanning	59 431	98.56	8.99	14.44	4.31	0.53	5.58	7.12
Haikou	59 758	95.81	6.82	13.19	0.63	0.66	1.67	7.47

2.2 FAO-56 PM

The formula for calculating ET_o using the FAO-56 PM method is:

$$ET_o = \frac{0.408\Delta(R_n - G)}{\Delta + \gamma(1 + 0.34u_2)} + \frac{\frac{900}{T + 273} + \gamma u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

where, R_n represents net radiation, MJ/(m^2×d); G is soil heat flux, MJ/(m^2×d); T is the daily average temperature, °C; U_2 is the wind speed at 2 m height, m/s, e_s is the saturated vapor pressure, kPa; e_a

is the actual vapor pressure, kPa; Δ is the slope of the saturated vapor pressure-temperature curve, kPa/°C; and γ is the psychrometric constant, kPa/°C.

2.3 Machine learning algorithms

2.3.1 The kernel-based extreme learning machine

The KELM is an improved algorithm based on the extreme learning machine (ELM). KELM employs a kernel function to replace the feature mapping in ELM[17], which enhances the accuracy and generalization ability of KELM[18]. The KELM model used in this study employs the radial basis function (RBF) as its kernel function. The objective function $F(x)$ of ELM can be expressed as:

$$F(x) = h(x) \times \beta = H \times \beta = L \tag{2}$$

where, x represents the input variable, $h(x)$ and H are the outputs of the hidden layer nodes, β is the output weight, and L is the desired output.

The network training is transformed into a linear system-solving problem; H^* represents the generalized inverse of H. To enhance the stability of the neural network, a regularization coefficient C and the identity matrix I are introduced, and the least squares solution for the output weight is:

$$\beta = H^{\mathrm{T}}\left(HH^{\mathrm{T}} + \frac{I}{c}\right)^{-1} L \tag{3}$$

The kernel function is introduced into the ELM, and the kernel matrix is:

$$\Omega_{\mathrm{ELM}} = HH^{\mathrm{T}} = h(x_i)h(x_j) = K(x_i, x_j) \tag{4}$$

where, x_i and x_j represent the experimental input vectors, and Equation (4) can be expressed as:

$$\Omega_{\mathrm{ELM}} = HH^{\mathrm{T}} = h(x_i)h(x_j) = K(x_i, x_j) \tag{5}$$

where, $(x_1, x_2, ..., x_n)$ represents the given training samples, n is the number of samples, and $K()$ is the kernel function.

2.3.2 The crested porcupine optimizer algorithm

This study employed the CPO to optimize the hyperparameters of the KELM model. CPO is a novel metaheuristic optimization algorithm inspired by the defensive behavior of crested porcupines[19,20]. To enhance the search capability of the algorithm, the Levy flight mechanism and the Quasi-Opposition Based Learning (QOBL) strategy were incorporated into the original CPO. The mathematical model for circular population reduction is as follows:

$$N = N_{\min} + (N' - N_{\min}) \times \left(1 - \left(\frac{t\% \frac{T_{\max}}{T}}{\frac{T_{\max}}{T}}\right)\right) \tag{6}$$

where, T is the variable determining the number of cycles, t is the current function evaluation, $T_{\max}$ is the maximum number of function evaluations, N' is the population size, and $N_{\min}$ is the minimum number of individuals in the newly generated population.

In the first defensive behavior, when the CP detects a predator, it responds by agitating its quills. The predator has two choices: One is to approach the CP and reduce the distance between them, promoting exploration of the area between the predator and CP to accelerate convergence speed. Alternatively, it can move away from the CP to maximize the distance between them, encouraging the exploration of unexplored distant regions. This behavior is simulated using the following mathematical formula:

$$\overrightarrow{x_i^{t+1}} = \overrightarrow{x_i^t} + \tau_1 \times |2 \times \tau_2 \times \overrightarrow{x_{\mathrm{CP}}^t} - \overrightarrow{y_i^t}| \tag{7}$$

where, $\overrightarrow{x_{\mathrm{CP}}^t}$ is the optimal solution of the evaluation function at iteration t, $\overrightarrow{y_i^t}$ is the vector generated between the current CP and a randomly selected CP from the population, representing the predator's position at iteration t, τ_1 is a random number based on a normal distribution, and τ_2 is a random value in the interval [0,1].

In the second defensive behavior, the CP generates noise to threaten the predator, with the noise increasing as the predator approaches the CP. This behavior is simulated using the following mathematical formula:

$$\overrightarrow{x_i^{t+1}} = \left(1 - \overrightarrow{U_1}\right) \times \overrightarrow{x_i^t} + \overrightarrow{U_1} \times \left(\vec{y} + \tau_3 \times \left(\overrightarrow{x_{r1}^t} - \overrightarrow{x_{r2}^t}\right)\right) \tag{8}$$

where, $r1$ and $r2$ are two random integers between [1, N], and τ_3 is a random value generated between 0 and 1.

In the third defensive behavior, the CP secretes a foul odor that spreads around it to prevent the predator from approaching. This behavior is simulated using the following mathematical formula:

$$\overrightarrow{x_i^{t+1}} = \left(1 - \overrightarrow{U_1}\right) \times \overrightarrow{x_i^t} + \overrightarrow{U_1} \times \left(\overrightarrow{x_{r_1}^t} + S_i^t \times \left(\overrightarrow{x_{r_2}^t} - \overrightarrow{x_{r_3}^t}\right) - \tau_3 \times \vec{\delta} \times \gamma_t \times S_i^t\right) \tag{9}$$

where, r_3 is a random number between [1, N], $\vec{\delta}$ is a parameter used to control the search direction, $\overrightarrow{x_i^t}$ is the position of the i-th individual at iteration t, γ_t is the defense factor, τ_3 is a random value generated between 0 and 1, and S_i^t is the odor diffusion factor.

In the fourth defensive behavior, when the predator is very close to the CP, the CP employs a physical attack, representing a one-dimensional inelastic collision. This behavior is simulated using the following mathematical formula:

$$\overrightarrow{x_i^{t+1}} = \overrightarrow{x_{\mathrm{CP}}^t} + (\alpha(1 - \tau_4) + \tau_4) \times \left(\delta \times \overrightarrow{x_{\mathrm{CP}}^t} - \overrightarrow{x_i^t}\right) - \tau_5 \times \delta \times \gamma_t \times \overrightarrow{F_i^t} \tag{10}$$

where, $\overrightarrow{x_{\mathrm{CP}}^t}$ is the obtained optimal solution, α is the convergence speed factor, τ_4 is a random value between [0,1], and $\overrightarrow{F_i^t}$ is the average force affecting the CP.

2.3.3 The dung beetle optimizer algorithm

DBO is a novel metaheuristic optimization algorithm inspired by the natural behaviors of dung beetles[21]. DBO simulates five behaviors of dung beetles, including rolling, dancing, breeding, foraging, and stealing, and achieves global optimization through cooperative division of labor within the population. The position update method for ball-rolling is as follows:

$$x_i(t+1) = x_i(t) + \alpha \times k \times x_i(t-1) + b \times \Delta x \tag{11}$$

$$\Delta x = |x_i(t) - X^w| \tag{12}$$

where, α is a natural coefficient, X^w represents the global worst position, and Δx simulates changes in light intensity.

When encountering an obstacle, a dung beetle uses dancing to reorient itself and find a new path. The position update method for the dancing behavior is as follows:

$$x_i(t+1) = x_i(t) + \tan(\theta)| \tag{13}$$

Female dung beetles roll dung balls to a safe, egg-laying-suitable place and hide them there as a means of providing an appropriate habitat for their offspring as part of their reproductive behavior:

$$\mathrm{Lb}^* = \max(X^* \times (1 - R), \mathrm{Lb}) \tag{14}$$

$$\mathrm{Ub}^* = \min(X^* \times (1 + R), \mathrm{Ub}) \tag{15}$$

where, Lb^* and Ub^* represent the lower and upper bounds of the egg-laying area, X^* denotes the current local optimal solution, and

Lb and Ub are the lower and upper bounds of the optimization problem.

The area where egg-laying occurs is dynamically adjusted with the number of iterations:

$$B_i(t+1) = X^* + b_1 \times (B_i(t) - Lb^*) + b_2 \times (B_i(t) - Ub^*) \quad (16)$$

In the formula, $B_i(t)$ represents the position information of the i-th egg ball at the t-th iteration.

Foraging behavior: Some mature dung beetles emerge from the ground to search for food, and the optimal foraging area for young dung beetles is dynamically updated as follows:

$$\mathrm{Lb}^b = \max(X^b \times (1-R), \mathrm{Lb}) \quad (17)$$

$$\mathrm{Ub}^b = \min(X^b \times (1+R), \mathrm{Ub}) \quad (18)$$

where, X^b represents the global optimal position, and Lb and Ub denote the lower and upper bounds of the optimal foraging area, respectively.

The position update rule for young dung beetles is:

$$x_i(t+1) = x_i(t) + C_1 \times (x_i(t) - \mathrm{Lb}^b) + C_2 \times (x_i(t) - \mathrm{Ub}^b) \quad (19)$$

Stealing behavior: Some dung beetles will steal dung balls from other dung beetles, and the position update rule is as follows:

$$x_i(t+1) = X^b + S \times g \times \{|x_i(t) - X^*| + |x_i(t) - X^b|\} \quad (20)$$

where, g is a random vector of size 1×d following a normal distribution, and S is a constant.

2.3.4 The particle swarm optimization algorithm

PSO, proposed by Kennedy and Eberhart in 1995, is an optimization algorithm based on swarm intelligence[22]. PSO simulates the foraging behavior of birds, searching for the optimal solution through information sharing and collaboration among individuals. The mathematical model for this behavior is as follows:

$$v_i^{(t+1)} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (\mathrm{pbest}_i - x_i^t) + c_2 \cdot r_2 \cdot (\mathrm{gbest} - x_i^t) \quad (21)$$

In the equation, $v_i^{(t+1)}$ is the velocity of particle i at the next iteration $t+1$, w is the inertia weight, v_i^t is the velocity of particle i at the current iteration t, c_1 and c_2 are the acceleration coefficients, and r_1 and r_2 are random numbers between [0,1]. pbest_i represents the personal best position found by particle i up to now, gbest is the global best position found by the entire swarm up to now, and x_i^t is the position of particle i at the current iteration.

The mathematical model for updating the particle position is as follows:

$$x_i^{(t+1)} = x_i^t + v_i^{(t+1)} \quad (22)$$

In the equation, $x_i^{(t+1)}$ is the position of particle i at the next iteration $t+1$, and x_i^t is the position of particle i at the current iteration.

2.4 The optimization processes of CPO, DBO, and PSO

The specific processes of optimizing KELM using CPO, DBO, and PSO are illustrated in Figure 1 below:

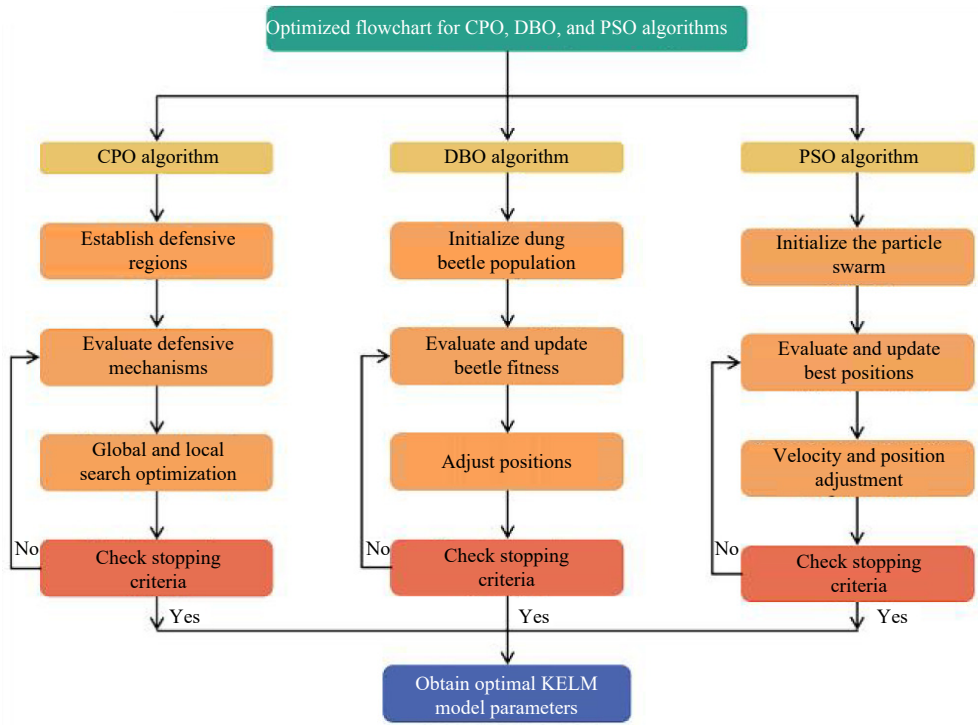


Figure 1 The flow chart of CPO, DBO, and PSO optimization methods process

In this study, three metaheuristic optimization algorithms—CPO, DBO, and PSO—were used to optimize the hyperparameters of the KELM. The core objective function for all optimization algorithms was to minimize the root mean square error (RMSE) of the KELM model on the training set, which is mathematically expressed as follows:

$$\text{Minimize: } f(C,\sigma) = \sqrt{\frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} (\hat{y}_i(C,\sigma) - y_i)^2} \quad (23)$$

In the equation, C is the regularization coefficient; σ is the RBF kernel parameter; $\hat{y}_i(C,\sigma)$ denotes the predicted $\mathrm{ET_o}$ value of the i-th training sample, calculated using the FAO-56 PM method; and N_{train} is the number of training samples, accounting for 70% of the total dataset.

2.5 Model performance evaluation metrics

The predictive performance of the model for $\mathrm{ET_o}$ was assessed using multiple statistical metrics, including the coefficient of determination (R^2), root mean square error (RMSE), Nash

coefficient (NSE), mean absolute error (MAE), and the global evaluation index (GPI). These metrics collectively provide a thorough evaluation of the model's accuracy and reliability.

The coefficient of determination, R^2, quantifies the proportion of variance in the dependent variable that can be explained by the independent variables. It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2} \tag{24}$$

where, $\hat{Y}_i$ is the predicted ET_o, Y_i is the observed value, $\bar{Y}$ is the mean of the observed values, and n is the total number of samples.

The RMSE provides a measure of the average magnitude of prediction errors and is computed as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\hat{Y}_i - Y_i\right)^2} \tag{25}$$

The NSE is a dimensionless statistic that evaluates how well the model's predictions match the observed data:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n} \left(\hat{Y}_i - Y_i\right)^2}{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2} \tag{26}$$

The MAE reflects the average absolute difference between predicted and observed values:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} \left|\hat{Y}_i - Y_i\right| \tag{27}$$

Finally, the GPI integrates the normalized scores of RMSE, R^2, MAE, and NSE, along with the median of observed data, into a single evaluation metric:

$$\mathrm{GPI} = \sum_{j=1}^{4} \alpha_j (g_j - y_j) \tag{28}$$

where, g_j denotes the normalized score of each metric ($g_j = -1$ for RMSE and MAE, and $g_j = 1$ for R^2 and NSE). A higher GPI value indicates superior model performance. In general, values of R^2 and NSE closer to 1 and RMSE and MAE approaching zero signify more accurate and reliable predictions.

3 Results and discussion

3.1 Importance ranking of meteorological factors influencing ET_o

In some underdeveloped regions, advanced meteorological observation equipment may not be available, resulting in incomplete meteorological data. These areas may only rely on basic conventional instruments to collect a limited range of meteorological data. Some scholars have pointed out that certain meteorological factors, such as air temperature and solar radiation, have a significant impact on ET_o, with their contribution to ET_o variation accounting for approximately 70%, while other factors have a smaller effect[23,24]. Excluding factors with minimal impact on the estimation target can improve the model's computational efficiency and reduce time costs. Therefore, this study employed the Random Forest algorithm to rank the importance of meteorological variables (P, T_{ave}, T_{max}, T_{min}, RH, U_2, SH), and learned about the importance of the characteristics of meteorological factors at each station (Table 2). The results indicate that T_{max} is the most influential factor on ET_o, with an importance range of 0.020-0.804. The second most important factor is T_{ave}, with an importance range of 0.024-0.809. The third important factor is RH, with an importance range of 0.021-0.118. P, T_{min}, U_2, and SH have minimal impact on importance. Emeka et al.[25] conducted a sensitivity analysis of meteorological data from multiple stations in Nigeria to assess the sensitivity coefficients for ET_o. Their results indicated that temperature and humidity are significant factors influencing ET_o, which is consistent with the findings of the present study. Zhao et al.[5] used Pearson correlation analysis to identify key factors influencing ET_o. Their results showed that temperature is the most significant meteorological factor affecting ET_o. In the analysis of this study, the top three factors in terms of feature importance, namely T_{max}, T_{ave}, and RH, accounted for 85.8% of the total importance in influencing ET_o. This indicates that T_{max}, T_{ave}, and RH play a dominant role among the many factors influencing ET_o. These three factors are subsequently used in the ET_o estimation model based on hybrid algorithms.

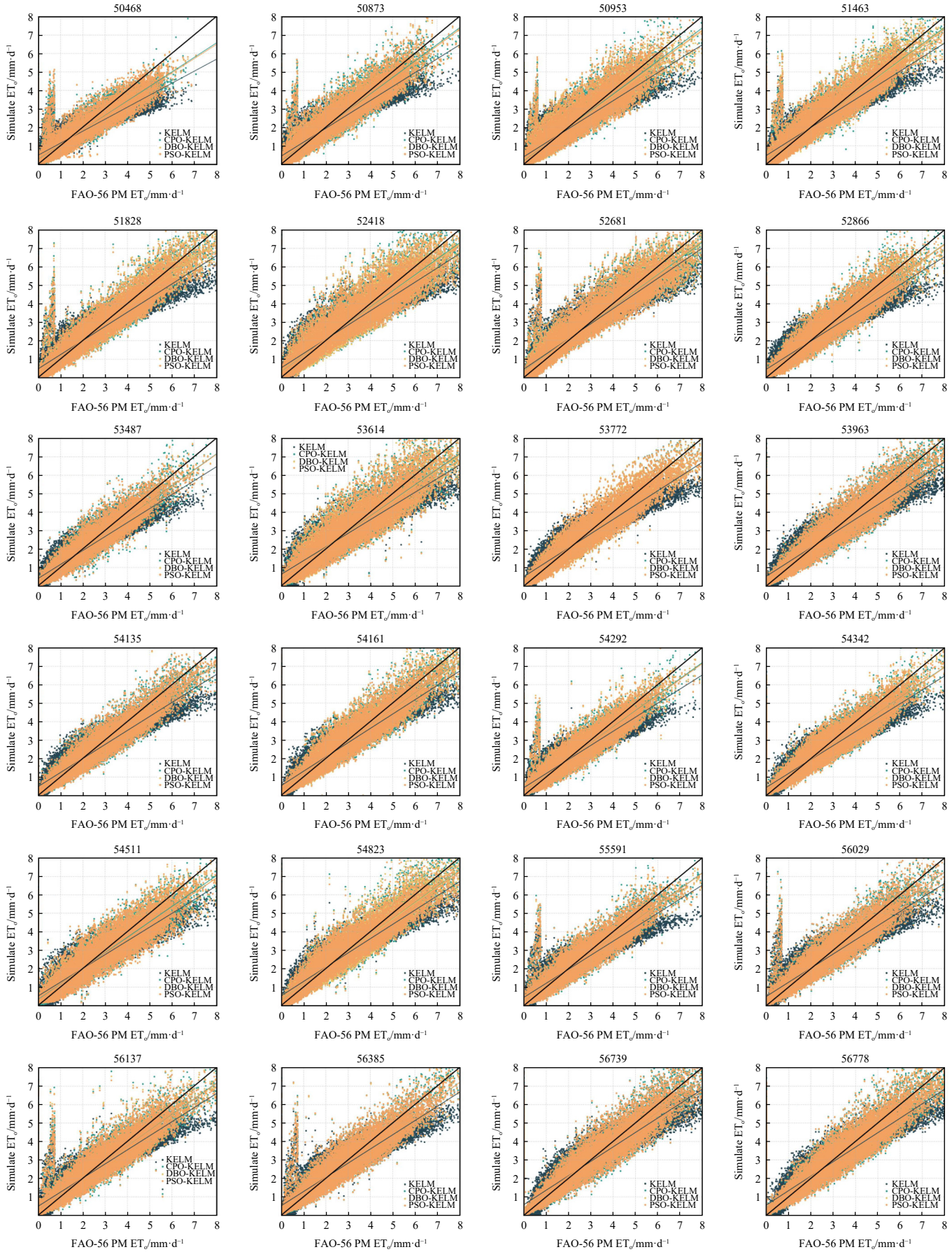
Table 2 The importance of the characteristics of meteorological factors at each station

Geographical Regions	Station	P	T_{ave}	T_{max}	T_{min}	RH	U_2	SH
Northeast	50 468	0.025	0.671	0.109	0.014	0.082	0.022	0.050
	50 873	0.014	0.587	0.213	0.010	0.077	0.028	0.055
	50 953	0.011	0.713	0.107	0.008	0.077	0.036	0.038
	54 161	0.018	0.024	0.741	0.009	0.098	0.038	0.054
	54 292	0.020	0.411	0.386	0.011	0.074	0.022	0.059
	54 342	0.024	0.027	0.720	0.013	0.076	0.024	0.097
North China	53 487	0.016	0.716	0.094	0.011	0.062	0.021	0.063
	53 772	0.017	0.048	0.728	0.010	0.103	0.034	0.041
	53 963	0.018	0.031	0.758	0.009	0.087	0.035	0.046
	54 135	0.021	0.033	0.749	0.011	0.086	0.027	0.054
	54 511	0.017	0.668	0.078	0.010	0.096	0.035	0.079
East China and Central China	54 823	0.019	0.025	0.737	0.011	0.106	0.033	0.050
	57 461	0.020	0.036	0.715	0.012	0.110	0.035	0.052
	57 494	0.018	0.064	0.702	0.013	0.089	0.022	0.073
	57 993	0.027	0.039	0.676	0.014	0.073	0.046	0.104
	58 208	0.027	0.048	0.676	0.014	0.044	0.026	0.148
	58 238	0.013	0.039	0.783	0.011	0.045	0.048	0.048
	58 457	0.010	0.761	0.066	0.010	0.059	0.05	0.033
	58 606	0.013	0.432	0.404	0.010	0.021	0.056	0.053
	58 847	0.012	0.809	0.020	0.009	0.022	0.065	0.052
South China	57 957	0.022	0.045	0.699	0.014	0.094	0.039	0.066
	59 287	0.012	0.435	0.393	0.012	0.024	0.049	0.063
	59 316	0.013	0.558	0.250	0.010	0.026	0.077	0.052
	59 431	0.008	0.078	0.727	0.024	0.041	0.091	0.020
	59 758	0.012	0.060	0.802	0.013	0.022	0.029	0.051
Northwest	51 463	0.018	0.073	0.692	0.010	0.078	0.027	0.086
	51 828	0.013	0.087	0.698	0.009	0.112	0.030	0.036
	52 418	0.013	0.657	0.158	0.009	0.071	0.047	0.034
	52 681	0.014	0.030	0.804	0.008	0.058	0.041	0.034
	52 866	0.020	0.056	0.732	0.011	0.067	0.029	0.067
	53 614	0.018	0.082	0.652	0.014	0.118	0.06	0.041
	56 029	0.020	0.046	0.714	0.011	0.104	0.035	0.052
Southwest	55 591	0.022	0.064	0.734	0.012	0.067	0.020	0.063
	56 137	0.021	0.048	0.699	0.012	0.106	0.039	0.057
	56 385	0.020	0.037	0.712	0.015	0.086	0.036	0.076
	56 739	0.017	0.049	0.699	0.012	0.113	0.048	0.049
	56 778	0.019	0.028	0.738	0.011	0.108	0.039	0.043
	57 816	0.020	0.210	0.558	0.014	0.080	0.021	0.078

3.2 Performance evaluation of ET_o estimation models

The statistical values of the ET_o model are presented in Tables 3-8. To comprehensively evaluate model performance across different metrics, the GPI, which synthesizes multiple statistical indicators into a single score, was calculated. A higher GPI value indicates better overall model performance.

The fitting performance of daily ET_o estimation models in China is shown in Figure 2.



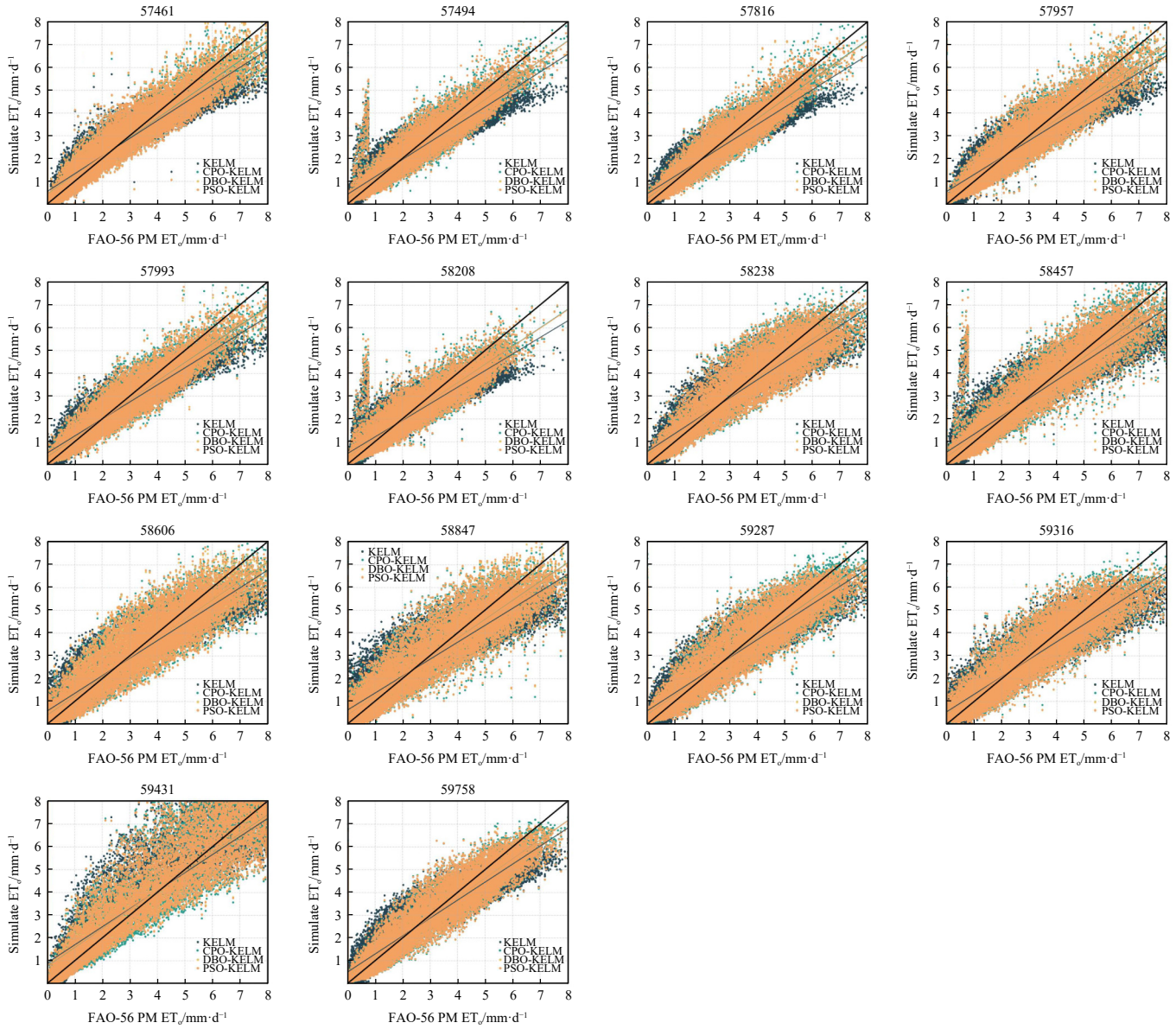


Figure 2 The fitting performance of daily ET_o estimation models in China

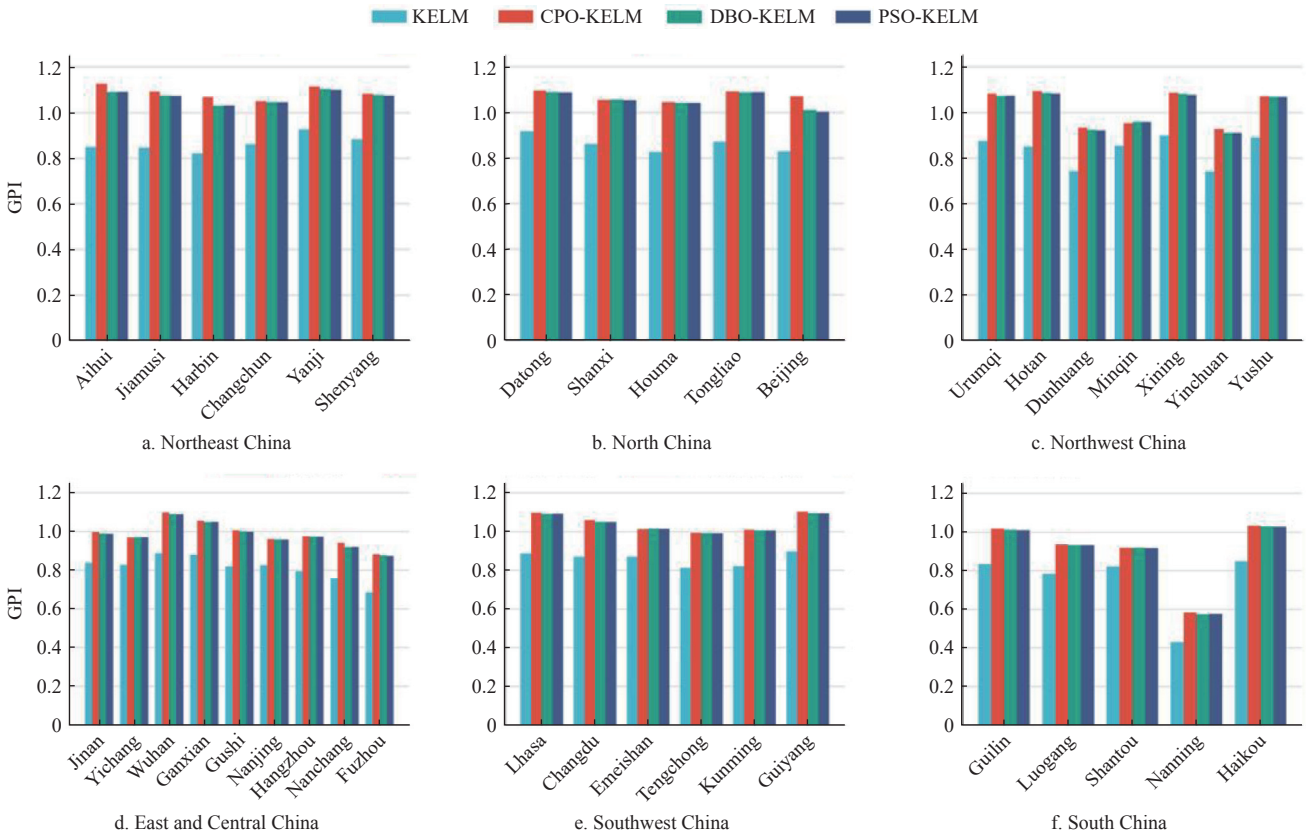
The KELM model demonstrates good accuracy at most stations, with RMSE=0.512-0.911, R^2=0.802-0.885, MAE=0.464-0.970, and NSE=0.802-0.885. Compared to traditional empirical models, the KELM model demonstrates superior accuracy and enhanced stability[26]. This is because KELM enhances the model's ability to process nonlinear data through kernel functions, while empirical models typically predict based on linear or simple nonlinear problems. The optimization algorithm can be used to refine the parameters of KELM, leading to the identification of a more optimal parameter combination. Compared to the standalone KELM algorithm, the optimized version further enhances the model' s stability and generalization capability. Among the three hybrid models, the accuracy of the CPO-KELM model is higher, with RMSE=0.413-1.147, R^2=0.881-0.942, MAE=0.284-0.763, and NSE=0.881-0.942. The model accuracy of DBO-KELM is as follows: RMSE=0.434-1.149, R^2=0.877-0.941, MAE=0.308-0.783, and NSE=0.877-0.940. For PSO-KELM, the model accuracy is RMSE=0.438-1.149, R^2=0.877-0.940, MAE=0.307-0.777, and NSE=0.877-0.940. The results indicate that the three optimization algorithms enhance the accuracy and stability of the KELM model in estimating ET_o. In recent studies, numerous researchers have also utilized optimization algorithms to refine the parameters of machine learning models, yielding favorable results. For instance, Wu et al.[27] developed the KELM model and its optimized version, achieving higher accuracy and training speed with fewer input feature wavelengths. Suo et al.[28] developed a feature-layer information fusion recognition model with four different types to identify sunflowers. The results indicate that the KELM algorithm optimized by Competitive Adaptive Reweighted Sampling (CARS) significantly improved the model's recognition accuracy. Peng et al.[29] optimized the KELM algorithm using the Sine-Cosine Algorithm (SCA), significantly enhancing its performance. The above results demonstrate that the optimized models exhibit improved accuracy and applicability, indicating that CPO, DBO, and PSO have a beneficial effect on optimizing the KELM model.

In the construction of the ET_o model, the accuracy of the CPO-KELM model developed in this study demonstrates a clear advantage over recent research. Long et al.[12] optimized the KELM model for ET_o prediction using algorithms such as the Sparrow Search Optimization Algorithm (SSA); however, the model's accuracy and applicability were lower than those of the models presented in this study (CPO-KELM, DBO-KELM, and PSO-

KELM). Wu et al.[30] combined ELM with three bio-inspired algorithms (PSO, GA, and ABC) to predict ET_o. In comparison, the three hybrid models proposed in this study exhibit superior accuracy. Hu et al.[31] integrated three bio-inspired algorithms—Aquila Optimizer (AO), Tuna Swarm Optimization (TSO), and Sparrow Search Algorithm (SSA)—with the ELM model to assess the accuracy of ET_o estimation. The results indicate that the AO-ELM model achieved the highest accuracy, with an R^2 of 0.9139. The accuracy of the CPO-KELM model developed in this study is higher, with an R^2 ranging from 0.881 to 0.942.

3.3 Applicability of the ET_o model across various regions

ET_o models were constructed using meteorological factors screened via Random Forest, and the statistical performance of KELM and three hybrid optimization models (CPO-KELM, DBO-KELM, and PSO-KELM) across seven major geographical regions was evaluated based on the comprehensive evaluation index (GPI) (Figure 3). The results indicate that CPO-KELM exhibits the highest prediction accuracy nationwide (RMSE: 0.413-1.147, R^2: 0.881-0.942). It achieves the lowest RMSE and MAE values and the highest R^2 and NSE indices across all regions. Next is DBO-KELM, which performs well in Northeast China (RMSE: 0.434-0.536, R^2: 0.910-0.925), North China (RMSE: 0.457-0.553, R^2: 0.910-0.935), South China (RMSE: 0.540-1.149, R^2: 0.881-0.934), Northwest China (RMSE: 0.471-0.683, R^2: 0.877-0.941), and Southwest China (RMSE: 0.447-0.582, R^2: 0.903-0.927). PSO-KELM ranks second only to CPO-KELM, particularly in East China and Central China (RMSE: 0.459-0.730, R^2: 0.886-0.926).



a. Northeast China b. North China c. Northwest China

d. East and Central China e. Southwest China f. South China

Figure 3 GPI indicator estimates of the ET model in different regions

Nationally, CPO-KELM exhibited the most superior comprehensive performance, with RMSE (0.413-1.147), R^2 (0.881-0.942), MAE (0.284-0.763), and NSE (0.881-0.942) all falling within the optimal range. This was followed by DBO-KELM (RMSE: 0.434-1.149, R^2: 0.877-0.941) and PSO-KELM (RMSE: 0.438-1.149, R^2: 0.877-0.940), whereas the original KELM model showed significantly lower accuracy than the optimized variants (RMSE: 0.512-0.911, R^2: 0.802-0.885).

Analysis of typical sites revealed that models demonstrated favorable applicability at 31 sites, including Aihui, Jiamusi, and Harbin, with GPI values for CPO-KELM, DBO-KELM, and PSO-KELM ranging from 0.952-1.125, 0.957-1.103, and 0.957-1.099, respectively. In contrast, model performance was suboptimal at seven sites, including Dunhuang (Northwest China) and Nanning (South China) (Tables 3-8). A cross-analysis of raw meteorological data (Table 1) reveals that the underperformance at these sites is not primarily attributed to "extreme climatic conditions", but essentially stems from the limitation of data dimensionality in the current model. For Nanning, the daily average wind speed (U_2=5.58 m/s) is the highest among all 38 stations—far exceeding the global average U_2 (2.84 m/s) of the studied sites—indicating that wind speed, an unincorporated factor in our model, serves as a dominant driver of local ET_o dynamics. Similarly, Dunhuang exhibits the highest daily average sunshine hours (SH=8.37 h) across all sites, with solar radiation (SH) being a pivotal regulator of ET_o in this arid region (Table 1).

The model in this study was constructed exclusively based on three key factors (T_{max}, T_{ave}, RH) identified by Random Forest analysis, which collectively account for 85.8% of ET_o variability across most studied regions (Section 3.1). However, in areas where ET_o is predominantly governed by factors other than temperature and humidity (e.g., wind speed in Nanning, solar radiation in Dunhuang), the exclusion of these critical variables inevitably leads to incomplete capture of ET_o response mechanisms, thereby causing performance fluctuations.

Notably, this underperformance does not constitute a flaw of

the proposed models, but rather an important scientific finding that clarifies the applicability boundary of the T/RH-based ET_o estimation framework. Specifically, the CPO-KELM, DBO-KELM, and PSO-KELM models demonstrate superior applicability in regions where temperature and humidity are the primary ET_o drivers (e.g., North China, most of Northeast China), as evidenced by their high R^2 (0.927-0.936) and low RMSE (0.449-0.522) (Table 4). For regions dominated by wind speed, solar radiation, or other non-T/RH factors, the model's accuracy may decrease, which highlights the necessity of tailoring variable selection to regional climatic characteristics.

This insight provides valuable guidance for future optimization: for microclimates governed by site-specific dominant factors (e.g., high-wind regions like Nanning, high-radiation areas like Dunhuang), supplementing corresponding key variables (U_2, SH, etc.) into the hybrid model framework will further enhance its robustness and generalization capability. This is particularly meaningful for precision agriculture and water resource management in diverse geographical contexts, as it enables the development of region-adapted ET_o estimation tools that balance simplicity (fewer variables for data-scarce regions) and accuracy (supplementary variables for complex microclimates).

4 Conclusions

This study developed models for estimating ET_o based on meteorological data from 38 sites across seven major geographical regions of China, using four machine learning algorithms: KELM, CPO-KELM, DBO-KELM, and PSO-KELM. The results indicate that:

1) When calculating the influence of meteorological factors on ET_o at each site using the Random Forest algorithm, T_{max}, T_{ave}, and RH were identified as the primary factors affecting ET_o results, with importance ranges of 0.020 to 0.804, 0.024 to 0.809, and 0.021 to 0.118, respectively.

2) The optimization algorithms significantly improved the performance of the KELM model for estimating ET_o, with the CPO-KELM model achieving the highest estimation accuracy. Compared to the independent KELM model, the CPO-KELM model has improved accuracy, with R^2 and NSE improvements ranging from 6.4% to 9.9% and 6.4% to 9.9%, respectively. MAE decreased by 21.3%-38.8%.

3) The CPO-KELM model performs optimally in ET_o estimation across all geographical regions in China, with its comprehensive statistical indicator GPI outperforming those of the DBO-KELM, PSO-KELM, and the original KELM models. Furthermore, the CPO-KELM model achieves the best performance in North China (R^2: 0.927-0.936, RMSE: 0.449-0.522, MAE: 0.317-0.370, NSE: 0.927-0.936, GPI: 1.044-1.095). Although the ET_o estimation models demonstrate good applicability at most stations, the models perform poorly at specific stations such as Dunhuang, Yinchuan, and Luogang due to extreme climatic conditions and complex relationships among meteorological factors. This limitation is mainly attributed to the lightweight input structure of the model, which may not fully capture ET_o dynamics in regions where wind speed or solar radiation dominates. These results clarify the applicability boundary of the proposed model under extreme or highly specialized climatic conditions.

4) The lightweight and high-precision CPO-KELM model, using only a few commonly available meteorological inputs, shows strong potential for irrigation management and efficient water use. Its adaptability across different climatic regions indicates broad applicability and supports data-driven and sustainable agricultural water management under changing climate conditions.

Acknowledgements

This work was supported by Key R&D and Promotion Projects in Henan Province (Science and Technology Development) (Grant No. 222102110452), Henan University of Science and Technology Young Backbone Teacher Project (Grant No. 4004-13450010), and Key Laboratory of Tarim Oasis Agriculture (Tarim University), Ministry of Education (Grant No: TOlab2025001).

[References]

[1] Wu Z J, Cui N B, Zhu B, Zhao L, Wang X K, Hu X T, et al. Improved hargreaves model based on multiple intelligent optimization algorithms to estimate reference crop evapotranspiration in humid areas of southwest China. Atmosphere, 2021; 12(1): 15.

[2] Xu Y, Xu Y P, Wang Y F, Wu L, Li G, Song S. Spatial and temporal trends of reference crop evapotranspiration and its influential variables in Yangtze River Delta, eastern China. Theoretical and Applied Climatology, 2017; 130: 945–958.

[3] Cunha A C, Gabriel Filho L R A, Tanaka A A, Goes B C, Putti F F. Influence of the estimated global solar radiation on the reference evapotranspiration obtained through the Penman-Monteith FAO 56 method. Agricultural Water Management, 2021; 243: 106491.

[4] He H, Wu Z, Li D D, Zhang T C, Pan F F, Yuan H W, et al. Characteristics of winter wheat evapotranspiration in eastern china and comparative evaluation of applicability of different reference evapotranspiration models. Journal of Soil Science and Plant Nutrition, 2022; 22: 2078–2091.

[5] Zhao L, Xing L W, Wang Y H, Cui N B, Zhou H M, Shi Y, et al. Prediction model for reference crop evapotranspiration based on the back-propagation algorithm with limited factors. Water Resources Management, 2023; 37: 1207–1222.

[6] Yong S L S, Ng J L, Huang Y F, Ang C K. Estimation of reference crop evapotranspiration with three different machine learning models and limited meteorological variables. Agronomy-Basel, 2023; 13(4): 1048.

[7] Tejada A T, Ella V B, Lampayan R M, Reano C E. Modeling reference crop evapotranspiration using support vector machine (SVM) and extreme learning machine (ELM) in region IV-A, Philippines. Water, 2022; 14(5): 754.

[8] Bellido-Jimenez J A, Estevez J, Vanschoren J, Garcia-Marin A P. AgroML: an open-source repository to forecast reference evapotranspiration in different geo-climatic conditions using machine learning and transformer-based models. Agronomy-Basel, 2022; 12: 656.

[9] Gong D Z, Hao W P, Gao L L, Feng Y, Cui N B. Extreme learning machine for reference crop evapotranspiration estimation: Model optimization and spatiotemporal assessment across different climates in China. Computers and Electronics in Agriculture, 2021; 187: 106294.

[10] Zhao X B, Li Y Z, Zhao Z H, Xing X G, Feng G H, Bai J Y, et al. Prediction model for daily reference crop evapotranspiration based on hybrid algorithm in semi-arid regions of China. Atmosphere, 2022; 13(6): 922.

[11] Sebastian J, Raj S K, Shihabudheen K V. Adaptive control of a nonaffine nonlinear system using self-organising kernel extreme learning machine. Isa Transactions, 2024; 146: 567–581.

[12] Tian S C, Ma L, Li H X, Tian F Y, Mao J R. Research on a coal seam gas content prediction method based on an improved extreme learning machine. Applied Sciences-Basel, 2023; 13(15): 8753.

[13] Zhao L, Zhao X B, Li Y Z, Shi Y, Zhou H M, Li X Z, et al. Applicability of hybrid bionic optimization models with kernel-based extreme learning machine algorithm for predicting daily reference evapotranspiration: a case study in arid and semiarid regions, China. Environmental Science and Pollution Research, 2023; 30: 22396–22412.

[14] Fan Y L, Ma Z, Tang W W, Liang J, Xu P F. Using crested porcupine optimizer algorithm and CNN-LSTM-attention model combined with deep learning methods to enhance short-term power forecasting in PV generation. Energies, 2024; 17(14): 3435.

[15] Zhang Z Q, Zhang L L, Liu D, Sun N, Li M, Faiz M A, et al. Measurement and analysis of regional water-energy-food nexus resilience with an improved hybrid kernel extreme learning machine model based on a dung beetle optimization algorithm. Agricultural Systems, 2024; 218: 103966.

[16] Jia Y, Wang H, Li P C, Su Y J, Wang F C, Huo S Y. Particle swarm optimization algorithm with Gaussian exponential model to predict daily and monthly global solar radiation in Northeast China. Environmental Science and Pollution Research, 2023; 30: 12769–12784.

[17] Fei S W, Liu Y Z. Fault diagnosis method of bearing utilizing GLCM and MBASA-based KELM. Scientific Reports, 2022; 12: 17368.

[18] Ma J, Liang S T, Du Z Y, Chen M. Compound fault diagnosis of rolling bearing based on ALIF-KELM. Mathematical Problems in Engineering, 2021; 2021: 2636302.

[19] Abdel-Basset M, Mohamed R, Abouhawwash M. Crested porcupine optimizer: A new nature-inspired metaheuristic. Knowledge-Based Systems, 2024; 284: 111257.

[20] Liu H J, Zhou R, Zhong X Y, Yao Y, Shan W F, Yuan J, et al. Multi-strategy enhanced crested porcupine optimizer: CAPCPO. Mathematics, 2024; 12(19): 3080.

[21] Xue J K, Shen B. Dung beetle optimizer: A new meta-heuristic algorithm for global optimization. The Journal of Supercomputing, 2022; 79: 7305–7336.

[22] Kennedy J, Eberhart R. Particle swarm optimization. In Proceedings of the Proceedings of ICNN'95-International conference on neural networks, Perth, WA, Australia: IEEE, 1995; pp.1942–1948. doi: 10.1109/ICNN.1995.488968

[23] Singh A, Kukal M S, Shapiro C A, Snow D D, Irmak S, Iqbal J. Growth phase-specific evaporative demand and nighttime temperatures determine Maize (*Zea Mays* L.) yield deviations as revealed from a long-term field experiment. Agricultural and Forest Meteorology, 2021; 308–309: 108543. doi: 10.1016/j.agrformet.2021.108543.

[24] Despotovic M, Nedic V, Despotovic D, Cvetanovic S. Review and statistical analysis of different global solar radiation sunshine models. Renewable & Sustainable Energy Reviews, 2015; 52: 1869–1880.

[25] Emeka N, Ikenna O, Okechukwu M, Chinenye A, Emmanuel E. Sensitivity of FAO Penman-Monteith reference evapotranspiration (ET_o) to climatic variables under different climate types in Nigeria. Journal of Water and Climate Change, 2021; 12(3): 858–878.

[26] Heramb P, Ramana Rao K V, Subeesh A, Srivastava A. Predictive modelling of reference evapotranspiration using machine learning models coupled with grey wolf optimizer. Water, 2023; 15(5): 856.

[27] Wu X J, Wu X M, Huang H C, Zhang F G, Wen Y X. Characterization of pepper ripeness in the field using hyperspectral imaging (HSI) with back propagation (BP) neural network and kernel based extreme learning machine (KELM) models. Analytical Letters, 2024; 57(3): 409–424.

[28] Suo L M, Liu H L, Ni J, Wang Z W, Zhao R. Sunflower origin identification based on multi-source information fusion technique of kernel extreme learning machine. Agronomy-Basel, 2024; 14(6): 1320.

[29] Peng S Q, Sun Q Z, Fan L, Zhou J, Zhuo X D. Optimized kernel extreme learning machine using Sine Cosine Algorithm for prediction of unconfined compression strength of MICP cemented soil. Environmental Science and Pollution Research, 2024; 31: 24868–24880.

[30] Wu Z J, Cui N B, Hu X T, Gong D Z, Wang Y S, Feng Y, et al. Optimization of extreme learning machine model with biological heuristic algorithms to estimate daily reference crop evapotranspiration in different climatic regions of China. Journal of Hydrology, 2021; 603: 127028.

[31] Hu J, Ma R, Jiang S Z, Liu Y L, Mao H Y. Prediction of reference crop evapotranspiration in China's climatic regions using optimized machine learning models. Water, 2024; 16: 3349.

[32] Yang L S, Feng Q, Adamowski J F, Yin Z L, Wen X H, Wu M, et al. Erratum to 'Spatio-temporal variation of reference evapotranspiration in northwest China based on CORDEX-EA' [Atmos. Res. 238 (2020) 104868]. Atmospheric Research, 2021; 252: 105425.