

尹 起, 周建平, 许 燕, 等. 基于粒子群优化(PSO)超限学习机预测新疆参考作物蒸散量[J]. 江苏农业学报, 2021, 37(3): 622-631.

doi: 10.3969/j.issn.1000-4440.2021.03.010

## 基于粒子群优化(PSO)超限学习机预测新疆参考作物蒸散量

尹 起<sup>1</sup>, 周建平<sup>1</sup>, 许 燕<sup>1</sup>, 李志磊<sup>2</sup>, 樊湘鹏<sup>1</sup>, 魏禹同<sup>1</sup>

(1. 新疆大学机械工程学院, 新疆 乌鲁木齐 830000; 2. 新疆大学工程训练中心, 新疆 乌鲁木齐 830000)

**摘要:** 参考作物蒸散量( $ET_0$ )的准确预测对于作物需水量预测、农田精准灌溉和提高水资源利用效率等具有重要意义。为了解决传统方法获取 $ET_0$ 的弊端, 本研究基于粒子群优化(Particle swarm optimization, PSO)-超限学习机(Extreme learning machine, ELM)预测 $ET_0$ 。通过选取新疆地区3个站点(乌鲁木齐、喀什、哈密)的最高气温( $T_{\max}$ )、最低气温( $T_{\min}$ )、平均相对湿度( $R_H$ )、风速( $u_2$ )、光照时间( $n$ )等气象数据, 建立PSO-ELM预测模型, 对模型精度和普适性进行研究, 并通过与ELM、Makink、I-A模型的对比, 探究不同气象因子组合模型的预测精度。结果表明, PSO-ELM模型在5种气象因子输入下具有最高预测精度(平均 $R^2=0.974\ 7$ , 平均 $MAE=0.252\ 0$  mm/d, 平均 $RMSE=0.364\ 3$  mm/d)。由PSO-ELM6模型与ELM、Makink、I-A模型的对比结果看出, 在相同的气象因子输入条件下, 3个站点用PSO-ELM6模型预测的效果最好。通过对PSO-ELM3模型在新疆地区普适性的研究发现, 该模型具有较高的预测精度(平均 $R^2=0.946\ 5$ , 平均 $MAE=0.307\ 0$  mm/d, 平均 $RMSE=0.356\ 9$  mm/d)。不同站点、不同气象因子输入的PSO-ELM模型能够较为精准地反映气象因子与 $ET_0$ 之间复杂的非线性关系, 且模型在新疆地区的普适性较好, 可以为新疆地区逐日 $ET_0$ 预测提供新的方法。

**关键词:** 新疆; 粒子群优化; 超限学习机; 参考作物蒸散量; 模型精度

**中图分类号:** S27; TP312 **文献标识码:** A **文章编号:** 1000-4440(2021)03-0622-10

## Prediction of reference crop evapotranspiration in Xinjiang based on particle swarm optimization(PSO) optimized extreme learning machine

YIN Qi<sup>1</sup>, ZHOU Jian-ping<sup>1</sup>, XU Yan<sup>1</sup>, LI Zhi-lei<sup>2</sup>, FAN Xiang-peng<sup>1</sup>, WEI Yu-tong<sup>1</sup>

(1. College of Mechanical Engineering, Xinjiang University, Urumqi 830000, China; 2. Engineering Training Center, Xinjiang University, Urumqi 830000, China)

**Abstract:** Accurate prediction of reference crop evapotranspiration ( $ET_0$ ) is of great significance in predicting crop water demand, precise irrigation of farmland and improving water resource utilization efficiency. To solve the disadvantages of traditional methods in obtaining  $ET_0$ ,  $ET_0$  was predicted based on particle swarm optimization (PSO)-extreme learning machine (ELM) in this study. By selecting meteorological data such as maximum temperature ( $T_{\max}$ ), minimum temperature ( $T_{\min}$ ), average relative humidity ( $R_H$ ), wind speed ( $u_2$ ) and illumination time ( $n$ ) of three stations in Xinjiang

(Urumqi, Kashgar and Hami), the PSO-ELM prediction model was established. The accuracy and universality of the model was studied, and the prediction accuracy of models combined with different meteorological factors was explored by comparing with ELM, Makink and I-A models. The results showed that, PSO-ELM model showed the highest prediction accuracy under the input condition of

收稿日期: 2020-08-27

基金项目: 国家级大学生创新创业训练计划项目(201810755079S);  
新疆维吾尔自治区研究生科研创新项目(XJ2019G033)

作者简介: 尹 起(1994-), 男, 河南新乡人, 硕士研究生, 主要从事智慧农业灌溉技术研究。(E-mail) 794760934@qq.com

通讯作者: 周建平, (E-mail) linkzhou@163.com

five meteorological factors (average  $R^2=0.974\ 7$ , average mean absolute error=0.252 0 mm/d, average root mean square error=0.364 3 mm/d). The prediction effect of PSO-ELM6 model was the best under the same meteorological factor input conditions of three stations by comparing the PSO-ELM6 model with ELM, Makkink, I-A models. The research on the universality of PSO-ELM3 model in Xinjiang showed that, the model had high prediction accuracy (average  $R^2=0.946\ 5$ , average mean absolute error=0.307 0 mm/d, average root mean square error=0.356 9 mm/d). The PSO-ELM model with different meteorological inputs at different stations can accurately reflect the complex non-linear relationship between meteorological factors and  $ET_0$ , and the model shows good generalizability in Xinjiang, which can provide new methods for daily  $ET_0$  prediction in Xinjiang.

**Key words:** Xinjiang; particle swarm optimization; extreme learning machine; reference crop evapotranspiration; model accuracy

参考作物蒸散量( $ET_0$ )是水循环研究中的重要组成部分,也是优化农业用水的重要变量,在水资源可持续管理及农业精准灌溉中起着重要作用<sup>[1]</sup>。在农业生态系统中,约 2/3 的降水量由作物蒸散过程损失<sup>[2-3]</sup>。由于频繁的干旱及农业、个人和工业用户之间对水资源的竞争,目前的农业生产用水量已经减少<sup>[4]</sup>。因此,有必要为灌溉管理者和水利研究人员提供一个准确的工具来估算参考作物蒸散量。 $ET_0$ 的获取方法较多,通常可以使用蒸渗仪或涡流协方差系统直接测量,但是使用、建设和维护蒸渗仪的成本较高。因此,使用基于气象因子经验和半经验建立的数学模型是一种更加符合实际且不需要额外代价的方法,如 Hargreaves-Samani 模型、Priestley-Taylor 模型、Makkink 模型等,但是这些数学模型受到多种因素影响,很难准确预测  $ET_0$ <sup>[5]</sup>。联合国粮食及农业组织(FAO)推荐 FAO-56 Penman-Monteith (PM)方法用于计算  $ET_0$  的标准值,该方法需要大量气象数据(相对湿度、最高气温、最低气温、太阳辐射和风速等),但是这些数据并不总能从气象站获取,在大多数发展中国家,这些气象数据往往是缺失或无法获取的<sup>[6]</sup>。

近年来,机器学习模型被应用于各个领域的科学和工程研究中以处理各种问题,如建模、优化和预测等<sup>[7-12]</sup>。使用机器学习模型通过选择输入(添加或删除输入因子),并找到变量(输入和目标变量)之间隐藏的复杂关系。此外,机器学习在  $ET_0$  的预测方面也得到了大量应用<sup>[13-20]</sup>,且气温是被证明与  $ET_0$  相关性最好的因子<sup>[21-22]</sup>。本研究选择基于最高气温( $T_{\max}$ )、最低气温( $T_{\min}$ )的多种气象因子组合模型对试验结果进行分析。

张皓杰等<sup>[23-24]</sup>采用超限学习机(Extreme learning machine, ELM)对中国不同地区的  $ET_0$  进行建模

预测,研究在不同气象因子输入下模型的拟合程度,并且得到了较好的结果。虽然 ELM 算法已经被许多研究者在不同的工程领域中得到验证,但是该算法的参数(随机初始化的输入权重和隐藏阈值)会限制模型的准确程度。在此背景下,本研究提出 1 种粒子群优化-超限学习机(PSO-ELM)预测  $ET_0$  的方法,并将该算法与传统 ELM 算法及 2 种经验模型进行比较,分析在不同地区输入不同组合气象因子的最佳预测模型,并通过对模型的普适性分析,筛选出新疆不同地区逐日  $ET_0$  的最佳计算方法,以期对  $ET_0$  的高精度预测提供一定的参考。

## 1 材料与方法

### 1.1 研究区域和数据集划分

新疆维吾尔自治区位于中国西北地区,地处亚欧大陆腹地,属于温带大陆性气候。考虑到新疆不同地区地理区域的差异性,选取站点需要有代表性,本研究选取的 3 个主要研究站点为新疆地区 6 个气象站点中的乌鲁木齐牧试站(地理位置为  $87^{\circ}11'E$ 、 $43^{\circ}27'N$ ,海拔为 1 930 m)、喀什站(地理位置为  $75^{\circ}45'E$ 、 $39^{\circ}29'N$ ,海拔为 1 385 m)、哈密站(地理位置为  $93^{\circ}31'E$ 、 $42^{\circ}49'N$ ,海拔为 737 m)。选取的用于模型普适性分析的辅助站点分别为阿勒泰站(地理位置为  $88^{\circ}05'E$ 、 $47^{\circ}44'N$ ,海拔为 735 m)、昭苏站(地理位置为  $81^{\circ}30'E$ 、 $43^{\circ}14'N$ ,海拔为 1 851 m)、和田站(地理位置为  $79^{\circ}56'E$ 、 $37^{\circ}08'N$ ,海拔为 1 375 m)。本研究站点的地理位置见图 1。本研究获取的平均气温( $T_{\text{mean}}$ )、最高气温( $T_{\max}$ )、最低气温( $T_{\min}$ )、相对湿度( $R_H$ )、风速( $u$ )、日照时长( $n$ )等逐日气象数据来源于中国气象数据网(<http://data.cma.cn>)。本研究站点的气象数据涵盖 2009

年 1 月至 2019 年 12 月的 11 年 (132 个月), 站点气象资料完整且经过人工审核。本研究中的数据采用留出法进行划分, 以 2009 年 1 月至 2017 年 12 月的数据作为训练集, 占总数据量的 82%, 以 2018 年 1 月至 2019 年 12 月的数据作为测试集, 占总数据量的 18%。取 3 次试验结果的平均值作为最终数据评价指标。

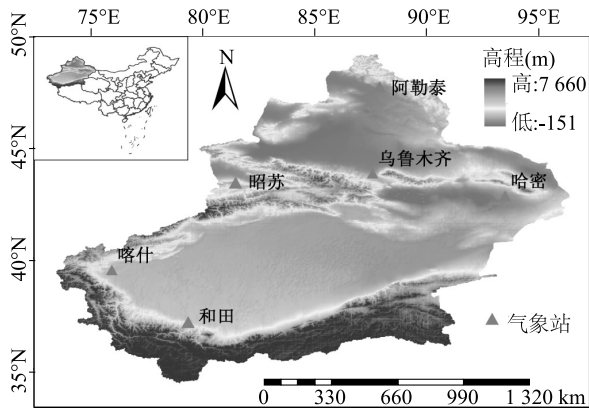


图 1 研究气象站地理位置示意

Fig.1 Geographical location map of meteorological stations selected in the study

## 1.2 参考作物蒸散量计算方法

以联合国粮食及农业组织推荐的 FAO-56 Penman-Monteith (PM) 方法作为  $ET_0$  计算方法获取标准  $ET_0$  值, 计算公式如下:

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

式中,  $ET_0$  为参考作物蒸散量, mm/d;  $T$  为平均气温,  $^{\circ}\text{C}$ ;  $\Delta$  为气温-饱和水汽压关系曲线上气温为  $T$  时的斜率,  $\text{kPa}/^{\circ}\text{C}$ ;  $R_n$  为净辐射量,  $\text{MJ}/(\text{m}^2 \cdot \text{d})$ ;  $G$  为土壤热通量,  $\text{MJ}/(\text{m}^2 \cdot \text{d})$ ;  $\gamma$  为湿度计常数,  $\text{kPa}/^{\circ}\text{C}$ ;  $u_2$  为 2 m 高度处的风速,  $\text{m/s}$ ;  $e_s$  为饱和水汽压,  $\text{kPa}$ ;  $e_a$  为实际水汽压,  $\text{kPa}$ 。上述每个变量的计算方法详见 FAO-56<sup>[25]</sup>。

另外选取 2 种经验模型:

Makkink 经验模型:

$$ET_0 = 0.61 \left( \frac{\Delta}{\Delta + \gamma} \right) \frac{R_s}{2.45} - 0.12 \quad (2)$$

式中,  $R_s$  为太阳总辐射,  $\text{MJ}/(\text{m}^2 \cdot \text{d})$ ; 其余变量的含义与公式 (1) 相同。

Irmak-Allen (I-A) 经验模型:

$$ET_0 = 0.489 + 0.289 R_n + 0.023 T_{\text{mean}} \quad (3)$$

式中,  $T_{\text{mean}}$  为当日平均气温,  $^{\circ}\text{C}$ ; 其余变量的含义与公式 (1) 相同。

## 1.3 超限学习机

ELM 是新加坡南洋理工大学黄广斌教授于 2006 年正式提出的<sup>[26]</sup>, 旨在克服传统机器学习算法存在的大量人为调整参数、计算速度缓慢、泛化性能较差等问题。该算法是一种单隐含层前馈传播的神经网络, 其输入层和隐含层之间的连接权重、隐含层神经元的阈值是随机生成的, 设定完成后就不再调整。隐含层和输出层的权重不需要迭代调整, 而是通过解广义逆求得, 因此该算法在计算时间上有较大优势。ELM 包括 3 层: 输入层、隐含层和输出层。

对于输入  $N$  个不同的样本  $(x_i, y_i)$ ,  $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ ,  $t_i = [t_{i1}, t_{i2}, \dots, t_{ik}]^T \in R^k$ , 设隐含层节点数为  $L$  个, 激活函数为  $G(x)$ , 则该算法的表达式为:

$$t_i = f(x_i) = \sum_{j=1}^L \beta_j G(w_j \cdot x_i + b_j), w_j \in R^n, \beta_j \in R^k \quad (4)$$

式中,  $i = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, L$ ;  $w_j = [w_{j1}, w_{j2}, \dots, w_{jn}]$  表示连接输入层到第  $j$  个隐含层节点的输入权重;  $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jk}]^T$  表示连接第  $j$  个隐含层节点到输出节点的输出权重; 激活函数  $G(x)$  可选择 Sigmoid 函数、Sin 函数等;  $w_j \cdot x_i$  为向量  $w_j$  和  $x_i$  的内积;  $b_j$  为隐含层第  $j$  个神经元的阈值。

将该表达式转为矩阵, 可得:

$$H\beta = T \quad (5)$$

由于  $H$  在训练前是已经确定的矩阵, 该网络的训练转化为求输出权重的最小二乘解的问题, 输出权重可表示为下式:

$$\beta = H^+ T \quad (6)$$

式中,  $H^+$  为隐含层输出矩阵  $H$  的 Moore-Penrose 广义逆<sup>[26]</sup>。

## 1.4 粒子群优化-超限学习机算法

由于 ELM 连接的输入层与隐含层的权重 ( $w$ ) 和隐藏神经元阈值 ( $b$ ) 是随机生成的, 从而可能使模型进入局部最优, 导致预测结果精度较差。粒子群优化算法是 1 种应用广泛的优化算法, 是由 Eberhart 和 Kennedy 从生物群体的协作行为出发提出的, 多用于调整机器学习模型参数, 其在农业方面的应用也被证明是有前景的<sup>[27-28]</sup>。因此本研究利用

PSO 算法优化 ELM 的输入权重和隐含神经元的阈值。PSO 中的每个粒子被认为是群中的一个个体,粒子的好坏用适应度函数的值判断,并通过个体和全局每步的最优位置不断更新,最终获得最优解。

在 1 个  $D$  维空间中,设含有  $n$  个粒子群  $X = (X_1, X_2, \dots, X_n)$ , 粒子  $i$  的位置  $(X_i) = (X_{i1}, X_{i2}, \dots, X_{iD})^T$ , 速度  $(V_i) = (V_{i1}, V_{i2}, \dots, V_{iD})^T$ , 粒子  $i$  和整个群的最佳位置分别为  $P_i = (P_{i1}, P_{i1}, \dots, P_{iD})^T$  和  $P_g = (P_{g1}, P_{g2}, \dots, P_{gD})^T$ , 在粒子群迭代优化过程中,粒子  $i$  的位置和速度的更新表示为:

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (7)$$

$$V_{id}^{k+1} = \omega V_{id}^k + c_1 r_1 (P_{id}^k - X_{id}^k) + c_2 r_2 (P_{gd}^k - X_{id}^k) \quad (8)$$

式中,  $d = 1, 2, \dots, D; i = 1, 2, \dots, n; \omega$  为惯性权重,其值代表偏重于全局搜索和局部搜索的能力; $k$  为当前迭代次数; $c_1, c_2$  为学习因子; $r_1, r_2$  为 0~1 中的随机数。

PSO-ELM 算法的具体步骤如下:

(1) 对数据集进行归一化处理,并划分训练集和测试集。

(2) 设置 ELM 输入层、隐含层、输出层维度,设置 PSO 种群大小、最小误差、迭代速度、迭代次数、惯性权重、学习因子。

(3) 计算粒子的适应度,选择均方误差 (Mean squared error, MSE) 作为适应度函数  $S_{MSE}$ , 表达式如下:

$$S_{MSE} = \frac{1}{N} \sum_{i=1}^N (T_i - Y_i)^2 \quad (9)$$

式中:  $N$  为总样本数;  $T_i$  为样本的实际值;  $Y_i$  为样本的预测值。

根据上式计算个体的适应度,迭代优化个体和全局最优粒子的位置和速度;当迭代到达设置的最小误差或一定的迭代次数时终止,得到输入权重和隐含层阈值的最优参数;将最优参数带入 ELM 进行计算,输出最优解。

PSO-ELM 算法的流程见图 2。

### 1.5 算法参数设置

设算法模型输入气象因子的数量为  $m$  个, PSO-ELM 算法和 ELM 算法的参数设置见表 1。

### 1.6 模型评价指标

为了评价预测结果,选择决定系数 ( $R^2$ )、平均绝对误差 (MAE)、均方根误差 (RMSE)、整体评价指标 (GPI) 作为预测结果的评价指标,相关公式如下:

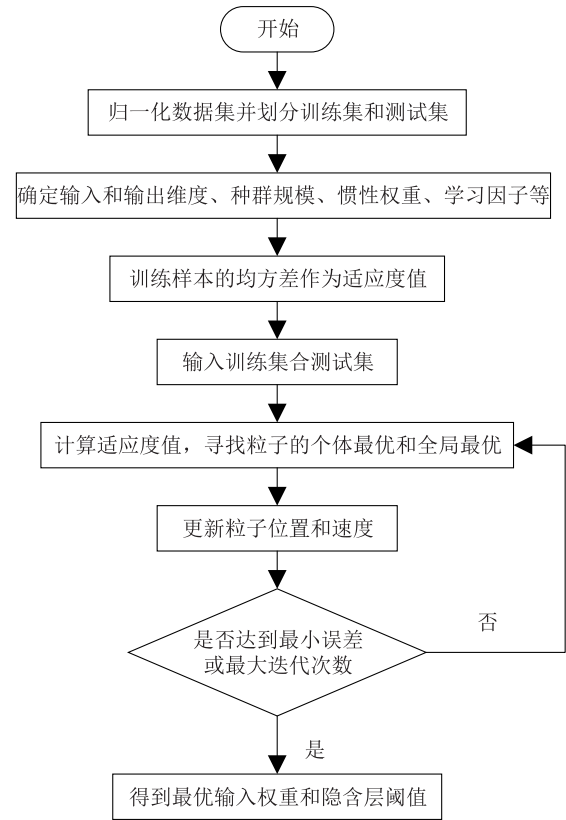


图 2 PSO-ELM 算法的流程

Fig.2 Algorithm flow chart of particle swarm optimization-extreme learning machine (PSO-ELM)

表 1 PSO-ELM 算法和 ELM 算法的参数设置

Table 1 Algorithm parameter setting of particle swarm optimization-extreme learning machine (PSO-ELM) and ELM

参数	PSO-ELM	ELM
输入层维度	$m$	$m$
隐含层维度	16.0	16.0
种群规模	30.0	
最大迭代次数	100.0	
学习因子	1.5	
初始惯性权重	0.8	
终止惯性权重	0.4	

PSO-ELM: 粒子群优化-超限学习机算法; ELM: 超限学习机算法。

$$R^2 = 1 - \frac{\sum_{i=1}^M (\hat{y}_i - y_i)^2}{\sum_{i=1}^M (\bar{y}_i - y_i)^2} \quad (10)$$

$$MAE = \frac{1}{M} \sum_{i=1}^M |y_i - \hat{y}_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_i - \hat{y}_i)^2} \quad (12)$$



$$GPI_j = \sum_{k=1}^3 \alpha_k (Z_{jk} - Z_k) \quad (13)$$

式中,  $y_i$  为真实值;  $\hat{y}_i$  为预测值;  $\bar{y}_i$  为真实值的平均值;  $M$  为预测模型的天数;  $GPI_j$  为模型  $j$  的  $GPI$  值;  $Z_{jk}$  为模型  $j$  参数  $k$  的值;  $Z_k$  为模型参数  $k$  的中位数, 当  $k = R^2$  时,  $\alpha_k = 1$ , 当  $k = RMSE$  或  $MAE$  时,  $\alpha_k = -1$ ;  $GPI$  越大, 预测结果的准确度越高。

## 2 结果与分析

### 2.1 不同输入因子下 PSO-ELM 模型的预测精度分析

本研究是基于气温因子 ( $T_{\max}$ 、 $T_{\min}$ ) 进行分析

表 2 不同气象因子输入下的 PSO-ELM 模型精度评价

Table 2 Accuracy evaluation of particle swarm optimization-extreme learning machine (PSO-ELM) model under the condition of different meteorological factors input

模型	输入的气象因子	$R^2$			$MAE$ (mm/d)			$RMSE$ (mm/d)		
		乌鲁木齐	喀什	哈密	乌鲁木齐	喀什	哈密	乌鲁木齐	喀什	哈密
PSO-ELM1	$T_{\max}, T_{\min}, R_H, u_2, n$	0.989 2	0.985 2	0.949 7	0.094 6	0.196 7	0.464 7	0.131 4	0.390 4	0.571 1
PSO-ELM2	$T_{\max}, T_{\min}, u_2, n$	0.941 5	0.910 1	0.886 1	0.219 9	0.656 9	0.962 9	0.306 0	0.962 8	0.859 5
PSO-ELM3	$T_{\max}, T_{\min}, R_H, u_2$	0.987 0	0.983 2	0.946 9	0.099 0	0.209 2	0.462 2	0.144 1	0.293 0	0.587 0
PSO-ELM4	$T_{\max}, T_{\min}, R_H, n$	0.968 5	0.764 6	0.875 1	0.151 4	1.088 6	0.616 2	0.224 5	1.558 1	0.900 1
PSO-ELM5	$T_{\max}, T_{\min}, u_2$	0.930 0	0.867 8	0.871 5	0.237 6	0.765 4	0.652 0	0.334 8	1.167 5	0.913 0
PSO-ELM6	$T_{\max}, T_{\min}, n$	0.909 7	0.794 8	0.808 2	0.274 8	1.021 0	0.752 5	0.388 3	1.588 4	1.174 7
PSO-ELM7	$T_{\max}, T_{\min}, R_H$	0.963 4	0.777 0	0.870 9	0.160 0	1.040 6	0.611 6	0.242 1	1.516 3	0.915 3
PSO-ELM8	$T_{\max}, T_{\min}$	0.871 4	0.786 6	0.790 2	0.326 9	0.995 8	0.768 4	0.453 9	1.568 3	1.180 0
PSO-ELM9	$R_H, u_2, n$	0.515 1	0.667 9	0.568 8	0.662 6	1.149 4	1.345 8	0.881 3	1.746 7	1.512 5
PSO-ELM10	$u_2, n$	0.346 0	0.320 1	0.363 9	0.724 3	1.544 4	1.336 7	1.023 4	2.221 8	2.031 4
PSO-ELM11	$R_H, u_2$	0.368 2	0.590 7	0.484 6	0.799 4	1.457 6	1.407 7	1.005 9	2.052 9	1.828 1
PSO-ELM12	$R_H, n$	0.494 3	0.416 9	0.431 0	0.662 4	1.648 5	1.284 3	0.899 9	2.452 2	1.921 2

$T_{\max}$ : 最高气温;  $T_{\min}$ : 最低气温;  $R_H$ : 平均相对湿度;  $u_2$ : 风速;  $n$ : 光照时间;  $R^2$ : 决定系数;  $MAE$ : 平均绝对误差;  $RMSE$ : 均方根误差。模型 PSO-ELM9~PSO-ELM12 为不含气温因子对照。

对 3 个站点基于气温气象因子的 8 种模型 (PSO-ELM1~PSO-ELM8) 进行分析。图 3 为未经处理的原始折线, 可见  $ET_0$  在不同站点及不同模型下波动较大。由于图 3 具有伪噪声特性, 不利于直观展示各个模型与 PM 标准值的拟合度, 因此对预测值进行高斯平滑处理。图 4 为高斯平滑处理后基于 PM 标准值的 2019 年乌鲁木齐站点、喀什站点、哈密站点 PSO-ELM 模型日  $ET_0$  的预测曲线, 可以看出, 各站点  $ET_0$  表现为喀什站点>哈密站点>乌鲁木齐站点, 3 个站点的年均  $ET_0$  需求量分别为 4.058 6 mm/d、3.034 6 mm/d、1.573 8 mm/d。如表 2 所示,

的。由表 2 可知, 包含气温因子且表现较差的模型在乌鲁木齐、喀什、哈密地区分别为 PSO-ELM8、PSO-ELM4、PSO-ELM8, 但模型的各项评价指标均优于不包含气象因子的模型 (PSO-ELM9~PSO-ELM12), 在这 3 个地区含有气象因子且表现最差的模型 (喀什地区的 PSO-ELM4 模型) 的精度也比不含有气温因子的最好的模型 (喀什地区的 PSO-ELM9 模型) 高。可以推测, 气温因子与  $ET_0$  的相关性最高, 这与李志磊等<sup>[21-22]</sup>的研究结论一致。

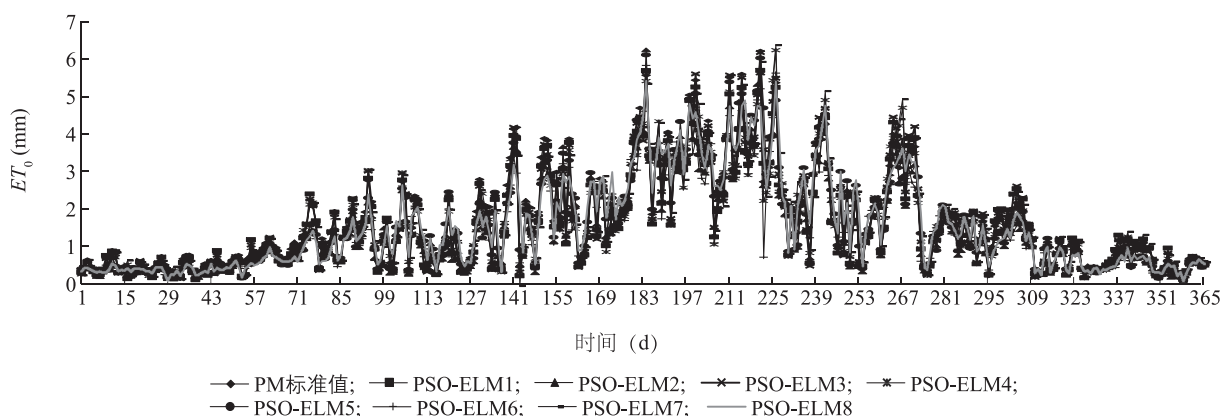
3 个站点的  $R^2$  为 0.777 0~0.989 2,  $MAE$  为 0.094 6~1.088 6 mm/d,  $RMSE$  为 0.131 4~1.588 4 mm/d。

当输入的气象因子数量为 5 个时, 乌鲁木齐站点、喀什站点、哈密站点的 PSO-ELM1 模型的  $R^2$  分别为 0.989 2、0.985 2、0.949 7,  $MAE$  分别为 0.094 6 mm/d、0.196 7 mm/d、0.464 7 mm/d,  $RMSE$  分别为 0.131 4 mm/d、0.390 4 mm/d、0.571 1 mm/d, 平均  $R^2=0.974 7$ , 平均  $MAE=0.252 0$  mm/d, 平均  $RMSE=0.364 3$  mm/d。虽然哈密站点模型的精度略低于其他 2 个站点, 但相比于其他模型的精度已经达到最高水平。以上结果表明, PSO-ELM1 模型的综合判

断精度在所有模型中最高,能够精准反映气象因子与  $ET_0$  之间的关系。

当输入的气象因子数量为 4 个时,即缺失  $u_2$ 、 $n$ 、 $R_H$  其中 1 个,此时 PSO-ELM 模型的精度较输入 5 个气象因子的模型略有下降,在喀什站点 PSO-ELM4 模型的  $R^2$  下降了 0.220 6,其预测效果相对较差。而在乌鲁木齐站点,PSO-ELM4 模型的  $R^2$  为 0.968 5,表明该模型在乌鲁木齐站点的适用性大于喀什站点及哈密站点。PSO-ELM3 模型的精度明显优于 PSO-ELM2、PSO-ELM4 模型,与 PSO-ELM1 模型相比, $R^2$ 、 $MAE$ 、 $RMSE$  的差值分别小于 0.002 9 mm/d、0.012 6

mm/d、0.097 5 mm/d,可以作为 PSO-ELM1 的近似模型。乌鲁木齐站点模型按精度排序为 PSO-ELM3>PSO-ELM4>PSO-ELM2,其中气象因子  $u_2$  和  $R_H$  对模型的正向影响较大,喀什站点、哈密站点模型按精度排序为 PSO-ELM3>PSO-ELM2>PSO-ELM4,其中气象因子  $u_2$  和  $n$  对模型的正向影响较大。在本输入因子条件下,3 个站点中表现最好的模型均含有  $u_2$  因子。综合考虑得出,在 4 个气象因子条件下,气象因子  $u_2$  是除气温因素外对  $ET_0$  影响最大的,这与张皓杰等<sup>[23]</sup>研究得出的除气温因素外  $u_2$  因素为西北旱区  $ET_0$  的主要驱动因子的结论一致。



$ET_0$ : 参考作物蒸散量。横坐标的数字 1 代表 1 月 1 日,365 代表 12 月 31 日,其他数字以此类推。

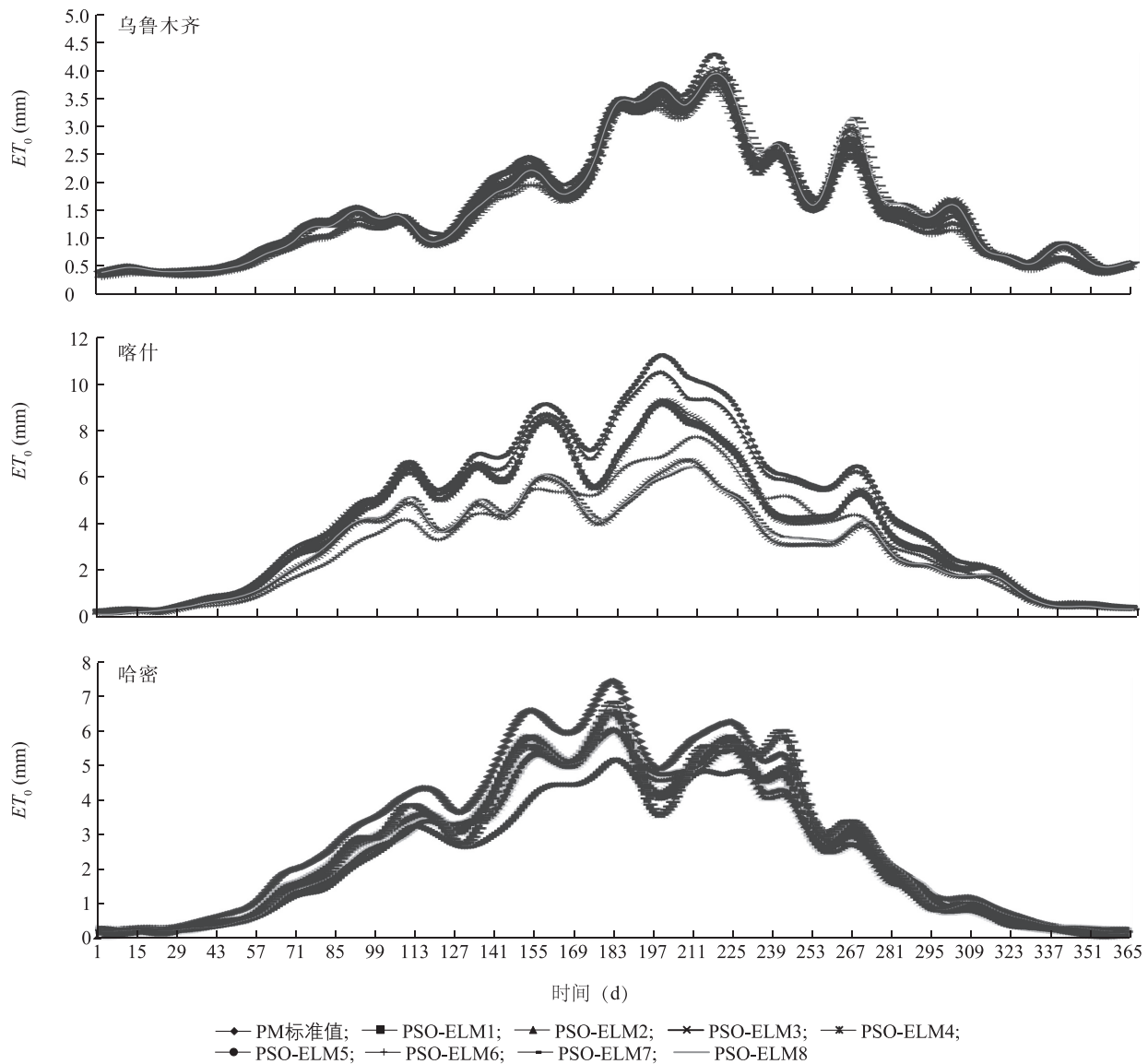
图 3 未经处理的原始折线

Fig.3 Raw line graph without processing

当输入的气象因子数量为 3 个时,即缺失  $u_2$ 、 $n$ 、 $R_H$  其中的 2 个,PSO-ELM6 模型的精度下降最为明显,其中喀什站点、哈密站点的  $R^2$  分别降至 0.794 8、0.808 2,  $MAE$  分别提高至 1.021 0 mm/d、0.752 5 mm/d,  $RMSE$  分别提高至 1.588 4 mm/d、1.174 7 mm/d。值得关注是,在喀什地区,与 PSO-ELM4 模型相比可知,PSO-ELM6 模型在缺少输入因子  $R_H$  的情况下综合指标反而优于 PSO-ELM4 模型,说明在某些地区增加输入因子不一定能提升预测精度。模型 PSO-ELM5 在喀什站点、哈密站点表现出较为理想的拟合效果。与 PSO-ELM2 相比,PSO-ELM5 在少了因子  $n$  后,3 个站点的  $R^2$ 、 $MAE$ 、 $RMSE$  变化较小,表明输入不同气象因子对  $ET_0$  的影响水平与贡献率大小不同,得出  $n$  因子的贡献率较其余 2 个因子小。PSO-ELM7 模型在乌鲁木齐站点的表现优于 PSO-ELM5 且异于喀什站点、哈密站点的较优模型,  $MAE$ 、 $RMSE$  分别为 0.160 0 mm/d、0.242 1

mm/d,是 3 因子模型中的最优值,表明在乌鲁木齐地区,气象因子  $R_H$  对  $ET_0$  的贡献率起主导作用,PSO-ELM7 可作为该地区  $ET_0$  的推荐模型。

当输入的气象因子数量为 2 个时,即只有  $T_{\max}$ 、 $T_{\min}$  2 个因子,仅在输入气温因子时模型 PSO-ELM8 也表现出一定的预测能力,其中乌鲁木齐站点的  $R^2$  高达 0.871 4,  $MAE$  为 0.326 9 mm/d,  $RMSE$  为 0.453 9 mm/d,与喀什站点、哈密站点的相关指标相比最好,说明该模型在新疆平均气温较低的地区(乌鲁木齐、喀什、哈密,年均气温分别为 8.35 °C、12.95 °C、11.00 °C)适用性好;相对于 PSO-ELM6 而言少了气象因子  $n$ ,但相对应的预测精度仅略微下降,这与上文中气温因子和  $ET_0$  相关性最高的结论一致;在喀什站点,与 PSO-ELM7 模型相比,PSO-ELM8 模型在缺少输入因子  $R_H$  情况下综合指标反而也优于 PSO-ELM7 模型,同样印证了上文所提到的结论——在某些地区增加输入因子不一定能提升预测精度。



$ET_0$ : 参考作物蒸散量。横坐标的数字 1 代表 1 月 1 日, 365 代表 12 月 31 日, 其他数字以此类推。

图 4 各站点 2019 年粒子群优化-超限学习机 (PSO-ELM) 模型的参考作物蒸散量预测值

Fig.4 Predictive value of reference crop evapotranspiration ( $ET_0$ ) of particle swarm optimization-extreme learning machine (PSO-ELM) model for each station in 2019

## 2.2 PSO-ELM 模型与其他模型预测精度的比较分析

以 PM 公式计算得出的  $ET_0$  作为标准值, 在上述模型精度的评价中选择精度较差的 PSO-ELM6 模型与 ELM、Makkink、I-A 3 种模型进行综合对比, Makkink、I-A 模型的计算方法分别见公式(2)和公式(3)。

选择气象因子  $T_{\max}$ 、 $T_{\min}$ 、 $n$  作为输入参数。表 3 显示, 3 个站点 PSO-ELM6 模型的  $R^2$ 、 $MAE$ 、 $RMSE$  范围分别为 0.794 8~0.907 9、0.274 8~1.021 0 mm/d、

0.388 3~1.588 4 mm/d。ELM 模型的精度较 PSO-ELM6 低, 其中  $R^2$ 、 $MAE$ 、 $RMSE$  的范围分别为 0.704 1~0.897 5、0.287 3~1.106 7 mm/d、0.405 1~1.719 7 mm/d。PSO-ELM6、ELM 的  $GPI$  分别为 1.730 3、0.916 5, 结果表明, 粒子群优化后的 ELM 在精度方面有了明显提升。在经验模型方面, 虽然乌鲁木齐站点经验模型 I-A 的  $R^2$  为 0.791 4, 小于 Makkink 的 0.857 4, 但两者的  $GPI$  分别为 -1.135 1、-8.880 7, 表明 I-A 模型的综合精度比 Makkink 模型高。4 种模

型的  $GPI$  排序为  $PSO-ELM6 > ELM > I-A > Makkink$ 。图 5 为在 3 个站点由 4 种不同模型的预测结果与 PM 公式计算的  $ET_0$  标准值分布的箱线图,可以看出,PSO-ELM6 模型的箱体位置分布、中位线、上下边缘线与 PM 标准模型非常相近且相似度大于 ELM,与 I-A、

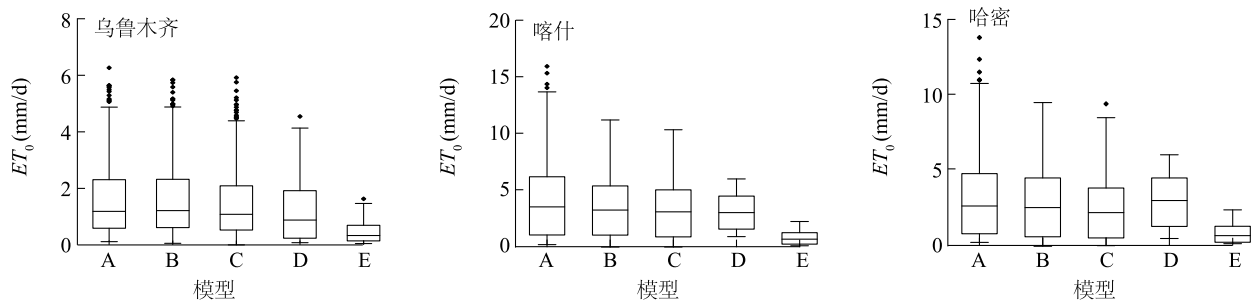
Makkink 传统模型相比,机器学习模型提供了更接近 PM 标准值的  $ET_0$  分布。根据上述分析可知,基于机器学习的模型比传统经验模型具有更好的性能,与 ELM、I-A、Makkink 模型相比,PSO-ELM6 模型在所有站点提供了最好的  $ET_0$  预测。

表 3 相同气象因子输入下的 PSO-ELM 模型与其他模型精度对比评价

Table 3 Comparison of precision between particle swarm optimization-extreme learning machine (PSO-ELM) model and other models with the same meteorological factor input

模型	输入的气象因子	$R^2$			MAE (mm/d)			RMSE (mm/d)		
		乌鲁木齐	喀什	哈密	乌鲁木齐	喀什	哈密	乌鲁木齐	喀什	哈密
PSO-ELM6	$T_{\max}, T_{\min}, n$	0.907 9	0.794 8	0.808 2	0.274 8	1.021 0	0.752 5	0.388 3	1.588 4	1.174 7
ELM	$T_{\max}, T_{\min}, n$	0.897 5	0.713 3	0.704 1	0.287 3	1.106 7	0.914 0	0.405 1	1.719 7	1.385 6
I-A	$T_{\max}, T_{\min}, n$	0.791 4	0.647 8	0.691 3	0.519 4	1.627 7	0.955 0	0.695 1	2.407 3	1.481 0
Makkink	$T_{\max}, T_{\min}, n$	0.857 4	0.736 3	0.755 7	1.137 6	3.335 7	2.304 0	1.476 0	4.314 5	3.082 2

$T_{\max}$ :最高气温; $T_{\min}$ :最低气温; $n$ :光照时间; $R^2$ :决定系数;MAE:平均绝对误差;RMSE:均方根误差。



A: PM; B: PSO-ELM6; C: ELM; D: I-A; E: Makkink; ◆: 异常值。

图 5 各台站不同模型  $ET_0$  分布

Fig.5  $ET_0$  distribution of different models at different stations

### 2.3 PSO-ELM 模型的普适性分析

由上述分析结果可知,PSO-ELM3 模型(输入气象因子为  $T_{\max}, T_{\min}, R_H, u_2$ )在输入因子较少的情况下表现出较高精度。为了对模型在新疆地区的普适性进行分析,以 PSO-ELM3 作为乌鲁木齐、喀什、哈密训练站点的模型,在 2018–2019 年选取乌鲁木齐、喀什、哈密、阿勒泰、昭苏、和田中的 5 个站点作为预测站点建立预测模型,共构建 15 个预测模型。

由表 4 可知,模型 PSO-ELM3 在新疆地区各个站点的  $R^2$  均大于 0.905 0, MAE 均小于 0.456 7 mm/d, RMSE 均小于 0.596 8 mm/d,其中阿勒泰站点的  $ET_0$  误差相对较大,平均  $R^2$ 、MAE、RMSE 分别为 0.923 4、0.345 9 mm/d、0.403 3 mm/d,模型精度稍低,这是由于阿勒泰地区处于较高的纬度且年平均气温(4.25℃)与其他站点的差距较大。整体上看,与原始训练

站点模型精度对比可知,预测站点的模型预测精度仅略微下降,各个站点间模型的可移植性较好,都达到了较高水平。结果表明,PSO-ELM3 模型在新疆各地区的普适性较好,可以将该模型作为预测新疆地区其他站点  $ET_0$  值的推荐模型。

## 3 结 论

本研究基于粒子群优化算法<sup>[29]</sup>优化超限学习机的权重和阈值,利用新疆各地区 2009–2019 年的气象数据建立了 1 个用于逐日  $ET_0$  预测的 PSO-ELM 模型,将该模型用于不同地区的  $ET_0$  预测。结果表明,PSO-ELM 在 3 个主研究地区的 8 种气象因子组合条件下都表现出了较好的预测精度,通过与其他模型的对比及普适性分析,验证了该模型在不同地区、不同条件下的适用性和普适性。



表 4 不同站点间 PSO-ELM3 模型的普适性分析结果

Table 4 Universality results of PSO-ELM3 model among different stations

训练站点	预测站点	$R^2$	MAE (mm/d)	RMSE (mm/d)
乌鲁木齐	喀什	0.965 3	0.136 5	0.154 3
	哈密	0.982 3	0.103 4	0.153 0
	阿勒泰	0.925 7	0.234 6	0.223 9
	昭苏	0.933 0	0.268 9	0.234 2
	和田	0.930 5	0.257 0	0.243 5
喀什	乌鲁木齐	0.967 8	0.267 8	0.344 7
	哈密	0.952 3	0.304 5	0.350 4
	阿勒泰	0.920 7	0.367 8	0.389 4
	昭苏	0.980 1	0.230 8	0.299 8
	和田	0.976 9	0.257 4	0.325 3
哈密	乌鲁木齐	0.942 1	0.432 4	0.483 0
	喀什	0.925 1	0.453 6	0.537 7
	阿勒泰	0.923 7	0.435 4	0.596 7
	昭苏	0.935 7	0.397 8	0.523 4
	和田	0.936 5	0.456 6	0.494 3

$R^2$ : 决定系数; MAE: 平均绝对误差; RMSE: 均方根误差。

对 5 种不同的气象因子进行组合, 得到 3 个站点的 8 种不同 PSO-ELM 模型, 不同模型预测乌鲁木齐、喀什、哈密地区  $ET_0$  的精度有差异, 其中 PSO-ELM1 的预测精度最高, 在对应站点的平均  $R^2$  为 0.974 7, MAE、RMSE 表现出较小误差 (平均 MAE = 0.252 0 mm/d, 平均 RMSE = 0.364 3 mm/d)。当输入因子数为 4 个时, 模型 PSO-ELM2、PSO-ELM3、PSO-ELM4 的分析结果表明, 在新疆各站点中除气温因子外气象因子  $u_2$  对  $ET_0$  预测影响最大, 在只有气温因子输入的情况下, 模型也能表现出足够的精度。结果表明, PSO-ELM 模型在不同气象因子输入组合下是一种能够作为替代 PM 公式计算  $ET_0$  的模型。

将使用  $T_{\max}$ 、 $T_{\min}$ 、 $n$  作为气象输入因子的 PSO-ELM6、ELM、Makkink、I-A 模型与由 PM 公式计算出的标准值进行对比发现, PSO-ELM6 的预测效果好于其他 3 种模型, 乌鲁木齐站点、喀什站点、哈密站点的  $R^2$  分别为 0.907 9、0.794 8、0.808 2, 且平均 MAE 为 0.682 8 mm/d, 平均 RMSE 为 1.050 5 mm/d, 均为最小值。因此 PSO-ELM6 更适合气象因子

( $T_{\max}$ 、 $T_{\min}$ 、 $n$ ) 输入条件下乌鲁木齐、喀什、哈密地区的逐日  $ET_0$  预测。

通过对模型的普适性分析, 将乌鲁木齐站点、喀什站点、哈密站点的 PSO-ELM3 模型应用到疆地区其他站点进行  $ET_0$  预测, 预测结果表现较好,  $R^2$  为 0.920 7~0.982 3, MAE 为 0.103 4~0.456 6 mm/d, RMSE 为 0.153 0~0.596 7 mm/d, 整体表现出较高的预测精度, 表明该模型在新疆不同地区的普适性较好, 可以作为新疆缺乏历史数据地区的  $ET_0$  预测模型使用。

#### 参考文献:

- [1] 康绍忠. 农业水土工程概论[M]. 北京: 中国农业出版社, 2007: 115-118.
- [2] SHIRI J, NAZEMI A H, SADRAADDINI A A, et al. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran[J]. Computers and Electronics in Agriculture, 2014, 108: 230-241.
- [3] KISI O, SANIKHANI H, ZOUNEMAT-KERMANI M, et al. Long-term monthly evapotranspiration modeling by several data-driven methods without climatic data[J]. Computers and Electronics in Agriculture, 2015, 115: 66-77.
- [4] MARTI P, GONZÁLEZ-ALTOZANO P, LÓPEZ-URREA R, et al. Modeling reference evapotranspiration with calculated targets. Assessment and implications[J]. Agricultural Water Management, 2015, 149: 81-90.
- [5] FERREIRA L B, DA CUNHA F F, DE OLIVEIRA R A, et al. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM-A new approach[J]. Journal of Hydrology, 2019, 572: 556-570.
- [6] BANDA P, CEMEK B, KUCUKTOPCU E, et al. Estimation of daily reference evapotranspiration by neuro computing techniques using limited data in a semi-arid environment[J]. Archives of Agronomy and Soil Science, 2018, 64(7): 916-929.
- [7] 冀秀梅, 王 龙, 高克伟, 等. 极限学习机在中厚板轧制力预报中的应用[J]. 钢铁研究学报, 2020, 32(5): 393-399.
- [8] 卢宏亮, 赵明松. 基于神经网络模型的安徽省土壤 pH 预测[J]. 江苏农业学报, 2019, 35(5): 1119-1123.
- [9] 李 晨, 崔宁博, 魏新平, 等. 改进 Hargreaves 模型估算川中丘陵区参考作物蒸散量[J]. 农业工程学报, 2015, 31(11): 129-135.
- [10] LI H J, XU Q, HE Y S, et al. Prediction of landslide displacement with an ensemble-based extreme learning machine and copula models[J]. Landslides, 2018, 15(10): 2047-2059.
- [11] 张 千, 魏正英, 张育斌, 等. 基于烟花算法优化极限学习机的温室参考作物蒸散量预测研究[J]. 中国农村水利水电, 2020(3): 29-32, 38.
- [12] 许伟栋, 赵忠盖. 基于卷积神经网络和支持向量机算法的马铃

- 薯表面缺陷检测[J].江苏农业学报,2018,34(6):1378-1385.
- [13] 魏 俊,崔宁博,陈雨霖,等. 基于极限学习机模型的中国西北地区参考作物蒸散量预报[J].中国农村水利水电,2018(8):35-39.
- [14] 刘小华,魏炳乾,吴立峰,等. 4 种人工智能模型在江西省参考作物蒸散量计算中的适用性[J].排灌机械工程学报,2020,38(1):102-108.
- [15] 樊湘鹏,许 燕,周建平,等. 遗传算法与小波神经网络在 ETO 预测中的应用[J].燕山大学学报,2019,43(2):182-188.
- [16] RAWAT K S, SINGH S K, BALA A, et al. Estimation of crop evapotranspiration through spatial distributed crop coefficient in a semi-arid environment [J]. Agricultural Water Management, 2019,213: 922-933.
- [17] 邢立文,崔宁博,董 娟. 基于 LSTM 深度学习模型的华北地区参考作物蒸散量预测研究[J].水利水电技术,2019,50(4):64-72.
- [18] 李可利,张 鑫. 基于 ANFIS 的陕西省参考作物蒸散量计算[J].自然资源报,2020,35(6):1472-1483.
- [19] GRANATA F. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study [J]. Agricultural Water Management, 2019,217: 303-315.
- [20] ANTONOPOULOS V Z, ANTONOPOULOS A V. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables[J]. Computers and Electronics in Agriculture, 2017,132: 86-96.
- [21] 李志磊,周建平,魏正英,等.  $ET_0$  预测的卡尔曼滤波修正 ANFIS 模型研究[J].干旱地区农业研究,2017,35(3):114-119.
- [22] 冯 禹,崔宁博,龚道枝,等. 基于极限学习机的参考作物蒸散量预测模型[J].农业工程学报,2015,31(增刊1):153-160.
- [23] 张皓杰,崔宁博,徐 颖,等. 基于 ELM 的西北旱区参考作物蒸散量预报模型[J].排灌机械工程学报,2018,36(8):779-784.
- [24] 吴立峰,鲁向晖,刘小强,等. 蝙蝠算法优化极限学习机模拟参考作物蒸散量[J].排灌机械工程学报,2018,36(9):802-805, 829.
- [25] ALLEN R G, PEREIRA L S, RAES D, et al. Crop evapotranspiration: guide-lines for computing crop water requirements [M]// FAO. Irrigation and Drainage. Rome:FAO, 1998.
- [26] HUANG G, ZHU Q, SIEW C K, et al. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1): 489-501.
- [27] 王 俊,刘 刚. 基于粒子群优化聚类的温室无线传感器网络节能方法[J].农业工程学报,2012,28(7):172-177.
- [28] 刘环宇,陈海涛,闵诗尧,等. 基于 PSO-SVR 的植物纤维地膜抗张强度预测研究[J].农业机械学报,2017,48(4):118-124.
- [29] 郭亚菲,樊 超,闫洪涛. 基于主成分分析和粒子群优化神经网络的粮食产量预测[J].江苏农业科学,2019,47(19):241-245.

(责任编辑:徐 艳)