

# 基于机器学习的遮荫设施内参考作物蒸散量估算

陈绍民<sup>1</sup>, 李晓丽<sup>1</sup>, 杨启良<sup>1\*</sup>, 吴立峰<sup>2</sup>, 熊凯<sup>3</sup>, 刘小刚<sup>1</sup>

(1. 昆明理工大学现代农业工程学院, 昆明 650500; 2. 南昌工程学院水利与生态工程学院, 南昌 330099;  
3. 长沙理工大学水利与环境工程学院, 长沙 410114)

**摘要:** 为高效准确地估算遮荫设施内参考作物蒸散量 (Reference Evapotranspiration,  $ET_0$ ), 该研究通过分析三七栽培遮荫设施 (四周及顶部均由黑色遮阳网遮盖, 通风性较好) 内及设施外气象参数的关系, 采用 Sobol 敏感性分析方法筛选出设施外有效的气象参数, 并将其作为模型输入, 以 Penman-Monteith (FAO-56 PM) 模型计算的值为标准值, 采用贝叶斯优化 (Bayesian Optimization, BO) 算法优化机器学习方法 (支持向量回归机 (Support Vector Regression, SVR)、随机森林 (Random Forest, RF) 和极限学习机 (Extreme Learning Machine, ELM)) 中的参数, 建立 3 种遮荫设施内  $ET_0$  估算模型 (BO-SVR、BO-RF 和 BO-ELM)。结果表明: 遮荫设施内  $ET_0$  对设施外平均相对湿度、平均风速、最高气温和平均气温的敏感性较高, 一阶敏感系数分别为 0.450、0.304、0.064 和 0.026, 故基于 4 组气象参数建立模型。BO-ELM 模型的测试精度整体优于 BO-SVR 和 BO-RF, 其中 BO-ELM 模型基于平均相对湿度、平均风速、最高气温和平均气温的气象参数组合估算精度最高, 决定系数、均方根误差和平均绝对误差分别为 0.928、0.069 mm/d 和 0.046 mm/d, BO-ELM 模型也能很好地适应少量气象参数 (平均相对湿度和平均风速) 估算设施内  $ET_0$ , 决定系数、均方根误差和平均绝对误差分别为 0.910、0.078 mm/d 和 0.057 mm/d。综合考虑计算精度和计算代价, 可将 BO-ELM 模型作为气象参数缺失情况下遮荫设施内  $ET_0$  的估算方法。研究为遮荫设施内  $ET_0$  的估算提供有效方法。

**关键词:** 作物; 蒸散量; 模型; 机器学习; 敏感性分析; 贝叶斯优化算法; 三七

doi: 10.11975/j.issn.1002-6819.2022.11.012

中图分类号: S161.4

文献标志码: A

文章编号: 1002-6819(2022)-11-0108-09

陈绍民, 李晓丽, 杨启良, 等. 基于机器学习的遮荫设施内参考作物蒸散量估算[J]. 农业工程学报, 2022, 38(11): 108-116.

doi: 10.11975/j.issn.1002-6819.2022.11.012 <http://www.tcsae.org>

Chen Shaomin, Li Xiaoli, Yang Qiliang, et al. Estimation of reference evapotranspiration in shading facility using machine learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(11): 108-116. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.11.012 <http://www.tcsae.org>

## 0 引言

参考作物蒸散量 (Reference Evapotranspiration,  $ET_0$ ) 是计算作物需水量的重要参数, 是水循环的关键物理量, 是研究水量平衡的重要因素, 准确地测定或估算参考作物蒸散量对制订合理的农业用水计划具有重要意义<sup>[1]</sup>。1998 年联合国粮农组织 (Food and Agriculture Organization of the United Nations, FAO) 推荐 Penman-Monteith (FAO-56 PM) 模型为计算  $ET_0$  的标准方法<sup>[2]</sup>。该模型在不同区域都具有较强的适应性且估算精度较高, 但计算公式较为复杂, 需要的气象参数较多<sup>[3]</sup>, 因此多种简化模型在研究工作中应运而生, 如基于气温的经验模型 (Irmak 法<sup>[4]</sup>、Blaney-Criddle 法<sup>[5]</sup>、Hamon 法<sup>[6]</sup>、Hargreaves-Samani 法<sup>[7]</sup>等), 基于气温和辐射的经验模型 (Makkink 法<sup>[8]</sup>、Priestley-Taylor 法<sup>[9]</sup>、Hargreaves 法<sup>[10]</sup>等)。

蒙强等<sup>[11]</sup>对比分析了 5 种经验模型在西藏高原灌区的适用性, 发现 Makkink 模型的拟合结果更接近 FAO-56 PM 模型的计算值。但研究表明<sup>[12]</sup>不同区域适用的  $ET_0$  估算模型并不一致, 如四川盆地东部区域适合用 Hargreaves-Samani 法, 四川盆地其他区域更适合使用 Priestley-Taylor 法。

设施栽培环境中的太阳净辐射和风速相比于露天环境下的小, 因此经验模型在估算设施内  $ET_0$  时存在局限性。陈新明等<sup>[13]</sup>修正了 FAO-56 PM 公式, 在温室内取得了很好的拟合效果, 证实了在温室内将 FAO-56 PM 中与风速有关的空气动力学项设置为 0 是不准确的。Mahmoodi-Eshkaftaki 等<sup>[14]</sup>在温室内外环境中测定了不同的气象参数, 建立了预测  $ET_0$  的模型, 与 9 种经验模型作对比, 在温室内外均取得了很好的估算效果。Kitta 等<sup>[15]</sup>根据设施外气象参数和设施的 2 个特征 (设施对太阳辐射的透射率  $\tau$  和设施内外风速比率  $\omega$ ) 建立模型, 为计算设施内  $ET_0$  提供了简单方法。以上修正后的估算模型具有较高的精度, 但缺少其中任意气象参数都会对估算精度造成影响, 因此需要发展一种计算简便、精度较高的方法。

近年来, 机器学习被广泛应用在各个领域<sup>[16-18]</sup>, 在预测参考作物蒸散量上也取得了良好的进展<sup>[19-22]</sup>。Petković 等<sup>[23]</sup>采用 Makkink 法确定  $ET_0$ , 建立神经模糊逻辑模型, 确定

收稿日期: 2022-01-14 修订日期: 2022-05-10

基金项目: 国家自然科学基金项目 (51979134、51779113); 云南省高校特色作物高效用水与绿色生产重点实验室项目 (KKPS201923009); 云南省基础研究重点项目 (202201AS070034)

作者简介: 陈绍民, 博士, 讲师, 研究方向为农业水土资源高效利用。

Email: shaomin\_ly24@126.com

\*通信作者: 杨启良, 博士, 教授, 研究方向为智慧农业水利技术与装备。

Email: yangqilianglovena@163.com

了估算  $ET_0$  的最佳气象参数组合 (日平均气温、总辐射量)。崔宁博等<sup>[24]</sup>研究表明机器学习方法——基于思维进化算法 (Mind Evolutionary Algorithm, MEA) 优化的误差反向传播神经网络 (Back Propagation Neural Network, BPNN) 在缺乏气象资料时预测西北旱区  $ET_0$  较经验模型精度更高。毛亚萍等<sup>[25]</sup>用 3 类算法 (核函数算法、树的集成算法和神经网络算法) 对新疆地区的  $ET_0$  进行了建模, 在日尺度和月尺度上都有较好的拟合效果。董建华等<sup>[26]</sup>将本地站数据与邻站数据相结合, 利用支持向量机 (Support Vector Machine, SVM) 和极限梯度提升法 (Extreme Gradient Boosting, XGBoost) 2 种机器学习方法, 成功估算出目标站  $ET_0$ 。机器学习方法的应用往往需要的数据集样本数较大, 对于小样本数据, 机器学习的预测性能较差。针对这一现象, 模型参数调优逐渐受到研究者的关注。张千等<sup>[27]</sup>提出一种基于烟花优化的极限学习机方法, 解决了小样本预测精度低的问题, 对温室室内  $ET_0$  的估算具有很好的适应性。目前常用的优化算法有粒子群优化算法、遗传算法及差分进化算法等<sup>[28]</sup>, 为使模型有较好的计算精度, 优化算法通常需要经过数百次甚至上千次迭代才能得到一组较好的模型参数, 而基于贝叶斯理论的贝叶斯优化 (Bayesian Optimization, BO) 算法, 通过已有的先验信息得到后验分布, 可经过少数次迭代获得近似目标值, 为优化问题的处理提供了可靠思路<sup>[29]</sup>。从以上的研究来看, 前人多通过改进经验公式估算设施内  $ET_0$ , 精度高却没有减少气象参数的数量, 而且在设施内安装所有测量气象参数的仪器成本太高, 基于设施外气象参数和机器学习方法估算设施内  $ET_0$  的研究还比较少见。

为利用设施外气象参数估算遮荫设施内  $ET_0$ , 本文基于三七栽培遮荫设施内波文比观测系统和中国气象数据网提供的气象资料, 利用敏感性分析筛选有效的气象参数, 以 FAO-56 PM 模型的计算值作为标准值, 采用 BO 算法优化的机器学习方法建设施内  $ET_0$  的估算模型, 以期对遮荫设施内  $ET_0$  的估算提供有效方法。

## 1 材料与方法

### 1.1 试验地概况

本文试验地位于云南省红河哈尼族彝族自治州泸西县昆明理工大学三七控水减排提质增效关键技术研究示范基地 (24°50'N, 103°57'E, 1 796 m), 属于北亚热带低纬高原山地季风气候, 年均温度 15.2 °C, 年降雨量 878.5 mm。试验地土壤为微酸性红壤土, 试验地三七为设施栽培 (顶部铺设 3 层黑色遮阳网, 透光率为 8.3%, 四周铺设 2 层黑色遮阳网, 透光率为 24.5%)。

### 1.2 数据来源及预处理

#### 1.2.1 数据来源

三七田气象资料由安装在设施内的 1 000 型波文比观测系统提供, 测量的气象参数包括太阳净辐射 ( $R_n$ ,  $W/m^2$ )、气温 ( $T$ , °C)、相对湿度 (Relative Humidity, RH, %) 和风速 ( $u$ , m/s), 太阳净辐射传感器和风速传感器安装在离地面 1.5 m 的高度, 温湿度传感器距离地

面 1 m, 数据由 DL2e 数据采集器每 30 min 采集 1 次, 由于仪器损坏, 部分时段的数据未监测到, 主要监测到 2018 年 5 月 5 日—9 月 30 日和 2019 年 3 月 5 日—6 月 29 日的气象数据, 共 12 828 组。设施外气象数据来源于中国气象数据网 (<http://data.cma.cn/>), 气象站点为云南省泸西站 (24°32'N, 103°46'E, 1 704.3 m), 选取 2018—2019 年逐日气象资料, 包括平均气温 ( $T_{out}$ , °C)、最高气温 ( $T_{max, out}$ , °C)、最低气温 ( $T_{min, out}$ , °C)、平均相对湿度 ( $RH_{out}$ , %)、日照时数 ( $n_{out}$ , h) 和平均风速 ( $u_{out}$ , m/s)。

#### 1.2.2 数据预处理

将波文比采集的数据进行整理, 若一个指标 1 d 内的缺失量多于 3 次, 将该天所有数据舍弃, 缺失量少于 3 次或有异常值的用二次样条插值法补全, 日值数据由波文比 1 d 中采集的 48 组数据求均值得到。

根据 FAO-56 PM 计算指南<sup>[2]</sup>, 将太阳净辐射  $R_n$  乘以转换系数  $8.64 \times 10^{-2}$ , 使其单位由  $W/m^2$  转换为  $MJ/(m^2 \cdot d)$ 。

风速  $u$  转换为地表以上 2 m 处风速  $u_2$ , 转换式为

$$u_2 = \frac{4.87u}{\ln(67.8z - 5.42)} \quad (1)$$

式中  $u_2$  为地表以上 2 m 处风速, m/s;  $z$  为地表以上的高度, m;  $u$  为地表以上  $z$  m 处风速, m/s。

为避免各个参数因为量纲和单位的差别影响模型的训练效果, 在模型训练参数之前, 采用式 (2) 将参数做归一化处理。

$$X^* = \frac{X - \mu}{\sigma} \quad (2)$$

式中  $X$  为样本值;  $\mu$  为样本均值;  $\sigma$  为样本标准差;  $X^*$  为  $X$  归一化处理后的值。

为选取具有代表性的训练集和测试集, 提高模型的估算精度, 以  $ET_0$  为准将样本集从小到大排序, 采用等间隔抽样法, 以 2:1 的比例抽取训练集和测试集。

### 1.3 研究方法

#### 1.3.1 FAO-56 PM 模型

本文以 FAO-56 PM 模型计算得到的  $ET_0$  作为标准值, 计算式<sup>[2]</sup>为

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T_a + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (3)$$

式中  $ET_0$  为参考作物蒸散量, mm/d;  $G$  为土壤热通量,  $MJ/(m^2 \cdot d)$ ;  $T_a$  为平均温度, °C;  $e_s$  为饱和水汽压, kPa;  $e_a$  为实际水汽压, kPa;  $\Delta$  为饱和水汽压与温度关系曲线的斜率,  $kPa/^\circ C$ ;  $\gamma$  为干湿温度计常数,  $kPa/^\circ C$ 。  $ET_0$  可分为辐射项  $ET_{0rad}$  (mm/d) 和空气动力学项  $ET_{0aero}$  (mm/d), 计算式为

$$ET_{0rad} = \frac{0.408\Delta(R_n - G)}{\Delta + \gamma(1 + 0.34u_2)} \quad (4)$$

$$ET_{0aero} = \frac{\gamma \frac{900}{T_a + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (5)$$

### 1.3.2 敏感性分析

为避免气象参数过多造成信息冗余,本文基于设施内外  $ET_0$  关系式,采用 Sobol 敏感性分析方法筛选气象参数。Sobol 分析是由 Sobol<sup>[30]</sup>于 1993 年提出的一种基于方差分解的全局性敏感性分析方法,该方法的方差分解式为

$$D = \sum_i D_i + \sum_{i \neq j} D_{i,j} + \dots + \sum_{i \neq j \neq k} D_{i,j,k} \quad (6)$$

式中  $D$  为所有参数贡献的总方差; $D_i$  为第  $i$  个参数贡献的方差; $D_{i,j}$  为第  $i$  个和第  $j$  个参数相互作用贡献的方差; $k$  为参数个数。

不同阶数的敏感系数表示为

$$S_i = \frac{D_i}{D} \quad (7)$$

$$S_{ij} = \frac{D_{i,j}}{D} \quad (8)$$

$$S_{Ti} = \frac{D_i + D_{i,j} + \dots + D_{i,j,\dots,k}}{D} \quad (9)$$

式中  $S_i$  为第  $i$  个参数的一阶敏感系数; $S_{ij}$  为第  $i$  个与第  $j$  个参数的二阶敏感系数; $S_{Ti}$  为第  $i$  个参数的总阶敏感系数。

本文采用了 Tang 等<sup>[31]</sup>对敏感系数的分级标准,敏感系数在  $[0.1, 1)$  范围内为很敏感,在  $[0.01, 0.1)$  范围内为敏感,在  $[0, 0.01)$  范围内为不敏感。

### 1.3.3 建模方法

根据 Sobol 敏感性分析方法筛选有效的气象参数,剔除对  $ET_0$  影响较小的气象参数,选出一阶敏感系数最高的两个气象参数作为固定参数,其余参数自由组合,形成不同的特征组合,并分别输入到模型中训练。具体建模方法如下:

#### 1) 支持向量回归机

支持向量回归机 (Support Vector Regression, SVR) 是 SVM 的一个重要分支,其基本原理是将输入特征映射到高维空间中做线性回归<sup>[32]</sup>。为更好地处理非线性数据,本文采用径向基核函数 (Radial Basis Function, RBF) 进行映射。SVR 模型中有 2 个需要调优的参数,分别为惩罚因子 ( $c$ ) 和核函数参数 ( $g$ ),  $c$  过大或  $g$  过大,模型预测容易过拟合,  $c$  过小或  $g$  过小,模型预测容易欠拟合,因此这两个参数的过大或过小,均会使模型的泛化能力变差。

#### 2) 随机森林

随机森林 (Random Forest, RF) 是由 Breiman<sup>[33]</sup>于 2001 年提出的一种集成算法,基本流程为:通过自助抽样法 (bootstrap) 从训练集中抽取样本集,构建多个回归决策树模型,将所有回归决策树的预测值求均值,得到随机森林的最终预测值。本文选用 2 个参数对 RF 模型进行优化,分别为回归决策树的数量 ( $e$ ) 和树的深度 ( $d$ ), 树的数量太大,计算量也会很大,树的数量太小,模型预测容易欠拟合;树的深度越大,模型的拟合效果越好,能降低由于树的数量增加造成的过拟合,但会增加模型的复杂度。

#### 3) 极限学习机

极限学习机 (Extreme Learning Machine, ELM) 是由

Huang 等<sup>[34]</sup>于 2004 年提出的一种用于训练单隐含层前馈神经网络 (Single-hidden Layer Feedforward Neural Network, SLFN) 的算法,基本原理为:采用激活函数将输入数据的随机特征非线性映射到新的特征空间,得到隐含层的输出值  $H$ ,根据隐含层输出值  $H$  和目标值  $T$ ,通过  $H\beta=T$  求解连接隐含层和输出层的最小二乘解  $\hat{\beta}$ 。本文采用的激活函数为“sigmoid”,通过对隐含层神经元数 ( $h$ ) 调优来优化 ELM 的估算性能。

### 1.3.4 模型优化算法

各机器学习方法中的参数影响模型估算的性能,为使模型具有更好的泛化能力,需要对参数进行优化。本文采用贝叶斯优化 (Bayesian Optimization, BO) 算法对 SVR、RF 和 ELM 中的参数进行调优,该方法由概率代理模型 (probabilistic surrogate model) 通过先验概率分布得到含有更多信息的后验概率分布,采集函数 (acquisition function) 根据后验概率分布产生新的评估点<sup>[35]</sup>,本文以树形结构 Parzen 估计方法 (Tree-structured Parzen Estimator Approach, TPE)<sup>[36]</sup>作为概念代理模型。相比网格搜索和随机搜索,BO 迭代次数少、效率高,是一种全局优化算法<sup>[35]</sup>。表 1 为各模型需要调优的参数,其中 SVR 中的 2 个参数为浮点型,RF 中的 2 个参数和 ELM 中的隐含层神经元数 ( $h$ ) 均为整型。贝叶斯优化是一个迭代过程,本文以均方误差 (Mean Square Error, MSE) 最小为原则,迭代 100 次,寻找各模型的最优参数。

表 1 不同模型中待调优的参数

建模方法 Modeling methods	参数名称 Parameter names	数据类型 Data type	参数空间 Parameter space
支持向量回归机 Support Vector Regression (SVR)	惩罚系数 $c$	浮点型	[1.0, 10.0]
	核函数参数 $g$	浮点型	[0, 0.1]
随机森林 Random Forest (RF)	树的最大深度 $d$	整型	[10, 100]
	树的数量 $e$	整型	[10, 100]
极限学习机 Extreme Learning Machine (ELM)	隐含层神经元数 $h$	整型	[10, 100]

### 1.3.5 模型评价

为评估机器学习方法基于不同参数组合的估算精度,本文以决定系数 ( $R^2$ )、均方根误差 (Root Mean Square Error, RMSE) 和平均绝对误差 (Mean Absolute Error, MAE) 作为评价指标<sup>[20]</sup>。通常较好的模型具有接近于 1 的  $R^2$ , 较小的 RMSE 和 MAE。

### 1.3.6 数据统计

用 Excel 整理、预处理试验数据和作图,采用 Python 3.7 进行敏感性分析和模型构建。

## 2 结果与分析

### 2.1 设施内外气象参数的关系

设施外的太阳净辐射由气象站给出的日照时数计算得到,具体公式可参考 FAO-56 PM 计算指南<sup>[2]</sup>,其余气象因子均可从气象站下载,设施内外各气象因子的关系如图 1 所示。从图 1 可以看出,设施内外各气象因子

之间均有相关性，其中平均气温、最高气温、最低气温和平均相对湿度间的相关性显著，决定系数  $R^2$  分别为 0.914、0.721、0.925 和 0.923，设施内的平均气温高于设施外的平均气温，设施内的平均相对湿度整体低于设

施外的平均相对湿度，遮荫设施对太阳净辐射和平均风速的减弱程度较大，因设施等因素的影响，导致设施内外太阳净辐射和平均风速在所有气象因子中的相关程度较弱，决定系数  $R^2$  分别为 0.647 和 0.611。

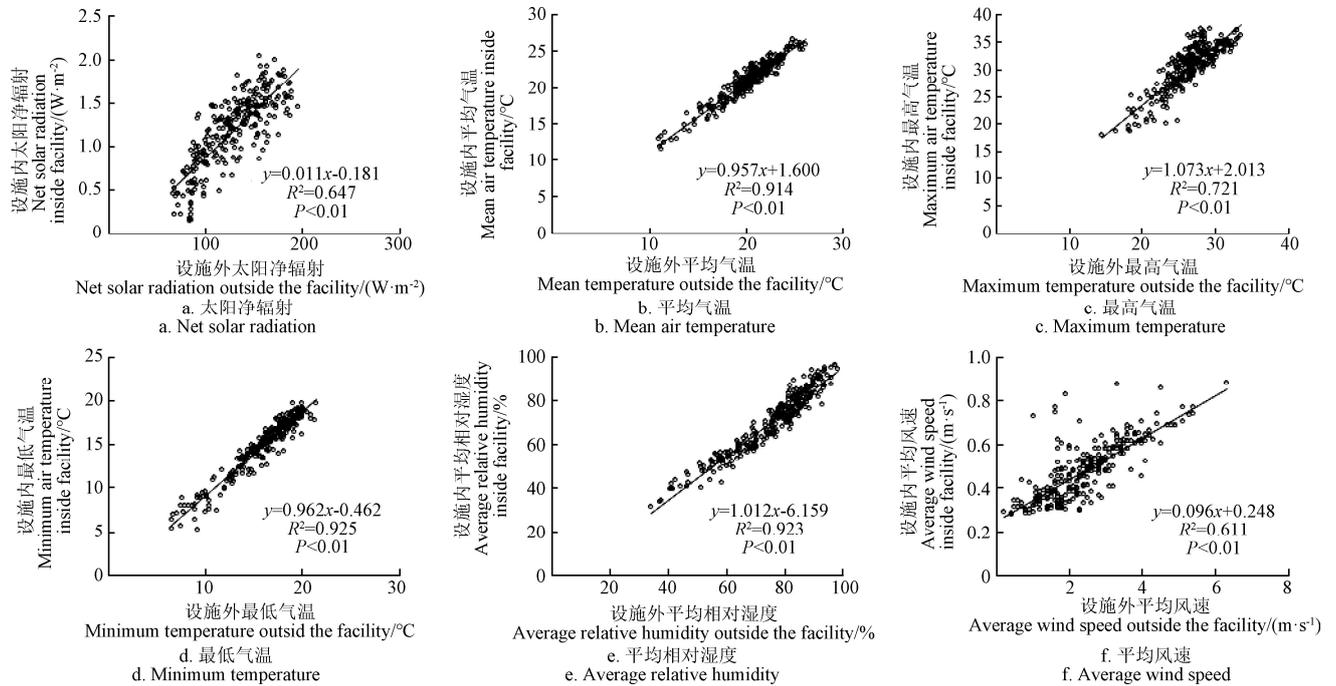


图 1 设施内外气象因子间的关系

Fig.1 Relationships between meteorological factors inside and outside the facility

## 2.2 设施内外 $ET_0$ 的关系

通过计算遮荫设施外  $ET_0$  的辐射项 ( $ET_{0rad, out}$ ) 和空气动力学项 ( $ET_{0aero, out}$ )，得到两者与设施内  $ET_0$  之间的关系如下：

$$ET_0 = 0.012ET_{0rad, out} + 0.291ET_{0aero, out} + 0.105 \quad (10)$$

拟合方程式(10)的  $R^2$  为 0.858, RMSE 为 0.099 mm/d, MAE 为 0.075 mm/d, 从式(10)可以看出，设施外  $ET_0$  的辐射项和空气动力学项拟合直线的斜率分别为 0.012 和 0.291，辐射项所占比例很小。进一步将遮荫设施内的  $ET_0$  分为辐射项和空气动力学项。以 2018 年 5 月为例 (图 2)，在选中的 27 d 内，晴天、阴天和雨天的比例为 8:11:8。

从图 2 中可以看出，设施内辐射项  $ET_{0rad, in}$  整体接近 0，空气动力学项  $ET_{0aero, in}$  与  $ET_0$  的变化趋势相同。5 月 27 日降雨量达到 40 mm，降雨使设施内相对湿度上升，进而导致与相对湿度相关的实际水汽压上升，水汽压差 (饱和水汽压  $e_s$  与实际水汽压  $e_a$  的差值) 变小，基于  $ET_0$  空气动力学项 (式 (5)) 可知， $ET_{0aero, in}$  在雨天有下降趋势。空气动力学项  $ET_{0aero, in}$  与  $ET_0$  之间的  $R^2$  为 0.999，RMSE 为 0.008 mm/d, MAE 为 0.006 mm/d, 因此，设施内的  $ET_0$  可以近似表示为

$$ET_0 \approx ET_{0aero, in} = \frac{\gamma \frac{900}{T_a + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (11)$$

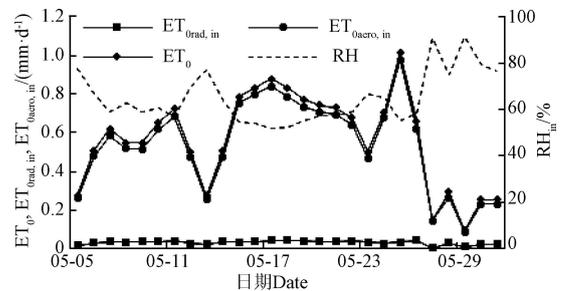


图 2 遮荫设施内  $ET_0$  各项值

Fig.2 Values for each part of  $ET_0$  in shading facility

为利用设施外气象参数计算遮荫设施内的  $ET_0$ ，将设施外空气动力学项  $ET_{0aero, out}$  与设施内  $ET_{0aero, in}$  做拟合 (图 3)， $ET_{0aero, out}$  与  $ET_{0aero, in}$  之间的关系为

$$ET_{0aero, in} = 0.291ET_{0aero, out} + 0.110 \quad (12)$$

遮荫设施外空气动力学项  $ET_{0aero, out}$  与设施内  $ET_{0aero, in}$  之间的  $R^2$  为 0.856，RMSE 为 0.097 mm/d, MAE 为 0.073 mm/d. 拟合结果表明，在设施内气象参数缺失严重的情况下，可通过式 (12) 利用设施外气象参数估算遮荫设施内  $ET_0$ 。

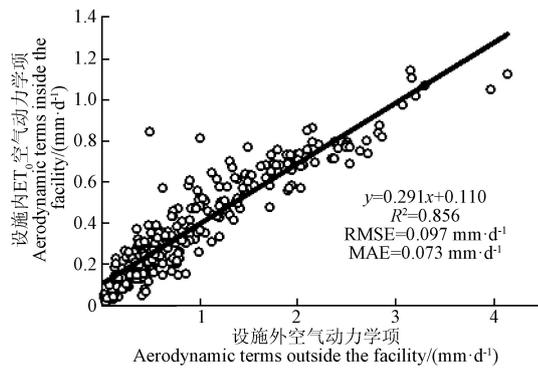


图3 设施内外 ET<sub>0</sub> 空气动力学项的拟合图

Fig.3 Fitting diagram of aerodynamic terms inside and outside the facility

### 2.3 敏感性分析

为筛选出有效的气象参数,采用 Sobol 敏感性分析方法,基于式(10)对设施外的气象参数作敏感性分析,结果如表2所示。由表2可知,设施外平均相对湿度(RH<sub>out</sub>)的一阶敏感系数最高,为0.450,总阶敏感系数达到0.581;设施外平均风速(u<sub>out</sub>),最高气温(T<sub>max, out</sub>),平均气温(T<sub>out</sub>),一阶敏感系数介于0.026~0.304,总阶敏感系数介于0.046~0.440。设施内 ET<sub>0</sub>对设施外最低气温 T<sub>min, out</sub>和日照时数 n<sub>out</sub>不敏感,一阶敏感系数均低于0.01,将不考虑这2个气象参数对遮荫设施内 ET<sub>0</sub>的影响。

根据遮荫设施内 ET<sub>0</sub>对设施外气象参数的敏感程度,本文选用 RH<sub>out</sub>和 u<sub>out</sub>作为固定参数,并与 T<sub>max, out</sub>或 T<sub>out</sub>自由组合,形成4种气象参数组合作为机器学习建模的输入项(表3)。

表2 遮荫设施外气象参数的敏感性分析结果

Table 2 The results of sensitivity analysis of meteorological parameters outside the shading facility

气象参数 Meteorological parameters	总阶敏感系数 Total order sensitivity coefficient	一阶敏感系数 First order sensitivity coefficient
设施外平均相对湿度 Average relative humidity outside the facility	0.581	0.450
设施外平均风速 Average wind speed outside the facility	0.440	0.304
设施外最高气温 Maximum air temperature outside the facility	0.096	0.064
设施外平均气温 Average air temperature outside the facility	0.046	0.026
设施外最低气温 Minimum air temperature outside the facility	0.020	0.002
设施外日照时数 Sunshine duration outside the facility	4.971×10 <sup>-4</sup>	8.228×10 <sup>-4</sup>

表3 遮荫设施外气象参数组合

Table 3 The combinations of meteorological parameters outside the shading facility

序号 Serial number	气象参数组合 Combination of meteorological parameters
1	平均相对湿度, 平均风速, 最高气温, 平均气温
2	平均相对湿度, 平均风速, 最高气温
3	平均相对湿度, 平均风速, 平均气温
4	平均相对湿度, 平均风速

### 2.4 模型建立及评价

#### 2.4.1 数据划分

将样本集以2:1的比例划分为训练集和测试集,其中训练集178组,用于敏感性分析和模型参数调优,并建立估算模型,测试集共有88组,用于测试模型的估算性能。样本集的划分结果见表4,训练集和测试集的数据统计特征相近,在一定程度上具有代表性。

表4 ET<sub>0</sub>在训练集和测试集上的统计特征

Table4 Statistical characteristics of ET<sub>0</sub> in training set and test set

样本集 Sample set	样本数量 Sample size	最大值 Maximum values/ (mm·d <sup>-1</sup> )	最小值 Minimum values/ (mm·d <sup>-1</sup> )	平均值 Mean/ (mm·d <sup>-1</sup> )	标准差 Standard deviation/ (mm·d <sup>-1</sup> )
训练集 Training set	178	1.189	0.023	0.422	0.266
测试集 Test set	88	1.155	0.043	0.413	0.254

#### 2.4.2 模型参数的选择

采用贝叶斯优化算法根据训练集对SVR、RF和ELM中的参数进行优化,迭代次数均为100次。各模型优化后的参数见表5,其余参数均采用默认值。

表5 基于不同气象参数组合的不同模型优化后的参数

Table 5 The optimized parameters of the different models based on different meteorological parameter combinations

序号 Serial number	SVR		RF		ELM
	惩罚系数 Penalty parameter	核函数参数 Kernel function parameter	树的最大深度 Maximum depth of tree	树的数量 Number of tree	隐含层神经元数 Number of nodes in hidden layer
1	9.86	0.006	31	15	18
2	8.72	0.009	95	38	13
3	9.76	0.026	88	16	15
4	9.27	0.002	48	23	11

#### 2.4.3 模型估算精度

基于4种气象参数组合(表3),采用BO优化参数建立BO-SVR模型、BO-RF模型和BO-ELM模型估算遮荫设施内 ET<sub>0</sub>(表6),机器学习模型分别用BO-SVR1~BO-SVR4、BO-RF1~BO-RF4和BO-ELM1~BO-ELM4表示。BO-SVR和BO-ELM对输入的气象参数数量较为敏感,估算精度随输入参数数量的增多而提升,对比训练集和测试集的估算结果,没有明显的过拟合或欠拟合现象;BO-RF1~BO-RF4模型测试集RMSE均比训练集RMSE大。总体上,BO-RF相对于BO-SVR和BO-ELM训练效果较好,而测试效果较差。

在输入RH<sub>out</sub>、u<sub>out</sub>、T<sub>max, out</sub>和T<sub>out</sub>时,各估算模型测试集的RMSE介于0.069~0.082 mm/d,训练集和测试集的估算值散点图如图4所示。3种机器学习估算模型测试集的估算值散点分布于1:1线两侧,拟合直线斜率分别为0.932、0.944和0.951,可见各模型均有很好的拟合效果。其中BO-ELM1的估算效果最好,训练集和测试集R<sup>2</sup>分别为0.938和0.928, RMSE分别为0.066和0.069 mm/d,

MAE 分别为 0.050 和 0.046 mm/d，表明 BO-ELM1 模型估算设施内  $ET_0$  具有较好的精度和稳定性。因此可将  $RH_{out}$ 、 $u_{out}$ 、 $T_{max, out}$  和  $T_{out}$  作为估算遮荫设施内  $ET_0$  的首选参数组合。

在采用最小数量气象参数组合 ( $RH_{out}$  和  $u_{out}$ ) 参与模型建立时，3 种机器学习模型测试集的 RMSE 介于 0.078~0.091 mm/d (表 6)。BO-ELM4 模型的估算精度最高，测试集  $R^2$  为 0.910；BO-RF4 模型的估算精度最低，测试集  $R^2$  为 0.875，远小于其训练集  $R^2$  (0.981)，表明 BO-RF4 模型存在过拟合现象；BO-ELM4 和 BO-SVR4 模型同时具有较稳定的性能和相近的估算精度 (表 6)，能很好地适应少量气象参数输入估算遮荫设施内  $ET_0$ 。

表 6 不同模型对  $ET_0$  的估算结果

Table 6 The estimation results of  $ET_0$  by different models

估算模型 Estimation models	训练集 Training set			测试集 Test set		
	$R^2$	RMSE/ (mm·d <sup>-1</sup> )	MAE/ (mm·d <sup>-1</sup> )	$R^2$	RMSE/ (mm·d <sup>-1</sup> )	MAE/ (mm·d <sup>-1</sup> )
BO-SVR1	0.928	0.072	0.054	0.927	0.069	0.045
BO-RF1	0.985	0.034	0.025	0.898	0.082	0.064
BO-ELM1	0.938	0.066	0.050	0.928	0.069	0.046
BO-SVR2	0.916	0.077	0.060	0.918	0.073	0.051
BO-RF2	0.987	0.031	0.023	0.897	0.082	0.065
BO-ELM2	0.912	0.079	0.061	0.918	0.074	0.053
BO-SVR3	0.916	0.078	0.062	0.895	0.083	0.060
BO-RF3	0.984	0.034	0.026	0.890	0.085	0.064
BO-ELM3	0.911	0.079	0.063	0.905	0.080	0.057
BO-SVR4	0.897	0.086	0.068	0.907	0.080	0.055
BO-RF4	0.981	0.038	0.028	0.875	0.091	0.069
BO-ELM4	0.910	0.080	0.064	0.910	0.078	0.057

注：估算模型后的数字 1~4 为气象参数组合序号。BO 为贝叶斯优化。下同。  
Note: The numbers 1-4 after the estimation models are the serial numbers of meteorological parameter combination. BO is bayesian optimization. Same as below.

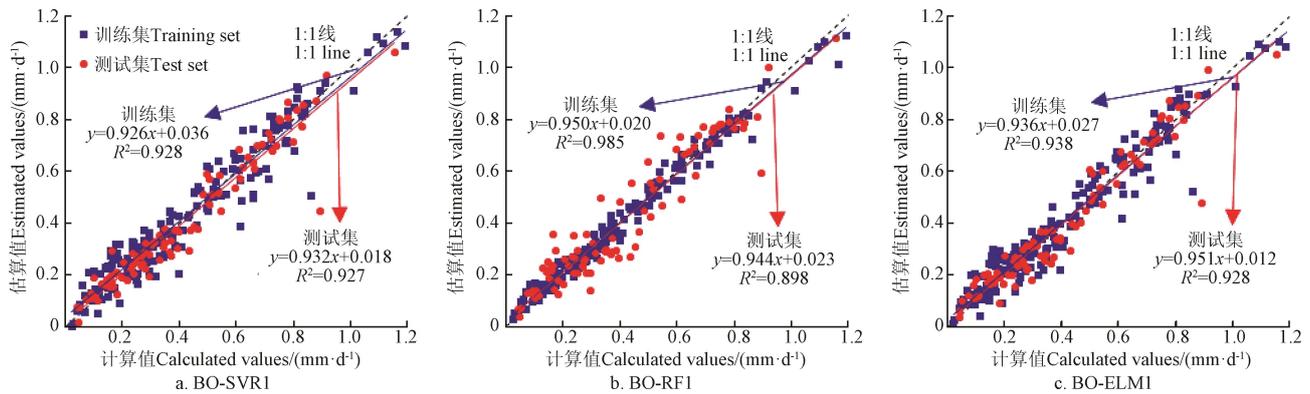


图 4  $ET_0$  估算值与计算值散点图

Fig.4 Scatter diagrams of the estimated and calculated values of  $ET_0$

2.4.4 计算代价

本文分别从模型的参数调优时间和建模时间统计了各估算模型的计算代价 (表 7)。

表 7 不同模型的计算代价

Table 7 The calculation cost of different models

估算模型 Estimation models	参数调优时间 Parameter tuning time/s	建模时间 Modeling time/s	总时长 Total time/s	各模型时长均值 Mean running time/s
BO-SVR1	0.96	0.00	0.96	0.97
BO-SVR2	0.96	0.00	0.96	
BO-SVR3	1.02	0.00	1.02	
BO-SVR4	0.94	0.00	0.94	
BO-RF1	28.35	0.02	28.36	27.46
BO-RF2	28.73	0.05	28.78	
BO-RF3	26.33	0.02	26.35	
BO-RF4	26.33	0.02	26.35	
BO-ELM1	2.22	0.02	2.23	2.22
BO-ELM2	2.16	0.00	2.16	
BO-ELM3	2.19	0.00	2.19	
BO-ELM4	2.30	0.00	2.30	

从表 7 中可知，BO-SVR 和 BO-ELM 的运行时间较短，BO-RF 的计算代价较高。BO-SVR 在输入 4 组气象参数时的运行时长差距较小，平均总时长为 0.97 s。RF 中有 2 个参数需要调优，参数空间均为 [10, 100]，树的数量和树的深度太大均会增加计算量，使 RF 的参数调优时

间过长，介于 26.35~28.78 s，从表 5 中可知，BO-RF2 中树的最大深度和树的数量分别为 95 和 38，相较于 BO-RF1、BO-RF3 和 BO-RF4 均较大，所以 BO-RF2 的建模时间相对较长。BO-ELM1 的隐含层神经元数比其他 3 组多，致使建模时间长，说明隐含层神经元数过大也会对 ELM 的计算代价产生影响。因此，在样本数据量较少时，可适当缩小参数空间，减少运行时间。

3 讨论

1) 由于设施环境等因素的影响，风速的大小和方向变化较大，遮荫设施对太阳净辐射有减弱作用，导致设施内外太阳净辐射和平均风速的相关性较弱。与刘海军等<sup>[37]</sup>采用的网室设施相似，三七栽培遮荫设施内的气温相较于设施外的气温偏高，香蕉种植网室内的相对湿度增加，而三七栽培遮荫设施内的相对湿度偏低，因为香蕉和三七的灌溉制度不同，另一方面，覆盖网的颜色可能也起到一定作用，还需作进一步研究。

2) 太阳辐射是  $ET_0$  辐射项的重要参数，通常对  $ET_0$  有较大的影响<sup>[23,38-39]</sup>，刘景杉等<sup>[40]</sup>分析了云贵高原参考作物蒸散量的成因，结果表明，从年尺度来看，太阳辐射对  $ET_0$  的作用最大，而在本文中，三七栽培遮荫设施内  $ET_0$  对设施外气象参数日照时数  $n_{out}$  (与太阳辐射直接相关) 不敏感，Möller 等<sup>[41]</sup>研究发现，设施内的太阳辐射主要取决于太阳高度角，太阳辐射的透过率随太阳高度

角的减小而降低, 试验地四周及顶部铺设黑色遮阳网, 使太阳辐射的透射率和辐射强度均有所减弱, 设施顶部透光率为 8.3%, 四周透光率为 24.5%, 因此本文中太阳辐射相关的日照时数对遮荫设施内的  $ET_0$  影响较小。

3) RF 中树的数量和树的深度以及 ELM 中隐含层神经元数均对模型的运行时间有影响, 因此在通过优化参数的方式解决样本数量少的问题时, 可适当缩小参数空间, 以减少模型运行的时间代价。在 3 种估算模型中, SVR 的表现稳定, 这是由于 SVR 有一个间隔带, 提升了模型的容错能力, 惩罚系数  $c$  反映了估算值超出间隔带的惩罚程度, 影响着模型的稳定性<sup>[32]</sup>, 在 BO 调优后, 提升了 SVR 的估算精度; ELM 可通过非线性映射得到输入与输出的潜在关系, 前人通过优化 ELM 中参数的方式提升模型的估算精度<sup>[27,42]</sup>, 本文利用优化的 ELM 对设施内  $ET_0$  的估算也取得了较好的结果。BO 优化后的 SVR 和优化后的 ELM 在训练集上的训练精度与在测试集上的估算精度基本一致, 没有明显的过拟合或欠拟合现象, 而优化后的 RF 在训练集上的估算精度远高于测试集, 稳定性较差, 与毛亚萍等<sup>[25]</sup>研究结果一致, RF 模型多用于处理分类问题, 可高效提取有效特征, 模型的估算结果由多棵决策树取均值得到, 对于小样本数据, 可能出现多棵相似的回归决策树, 致使 RF 模型在训练集和测试集上的估算效果差异较大。

## 4 结论

本文以 FAO-56 PM 的计算值为标准值, 通过敏感性分析筛选设施外有效的气象参数, 利用 3 种基于贝叶斯优化 (Bayesian Optimization, BO) 算法优化的机器学习方法, 估算了遮荫设施内参考作物蒸散量 (Reference Evapotranspiration,  $ET_0$ ), 结果表明:

1) 设施内外各气象因子之间均有相关性, 其中平均气温、最高气温、最低气温和平均相对湿度间的相关性较高,  $R^2 \geq 0.721$ , 设施内的平均气温高于设施外的平均气温, 设施内的平均相对湿度整体低于设施外的平均相对湿度。遮荫设施内  $ET_0$  的辐射项 ( $ET_{0rad,in}$ ) 接近 0,  $ET_0$  近似等于空气动力学项 ( $ET_{0aero,in}$ ),  $ET_{0aero,in}$  与设施外空气动力学项 ( $ET_{0aero,out}$ ) 之间的  $R^2$  为 0.856, RMSE 为 0.097 mm/d, MAE 为 0.073 mm/d, 可通过计算设施外空气动力学项  $ET_{0aero,out}$  估算遮荫设施内  $ET_0$ 。

2) 设施外的 6 种气象参数中, 平均相对湿度 ( $RH_{out}$ )、平均风速 ( $u_{out}$ )、最高气温 ( $T_{max,out}$ ) 和平均气温 ( $T_{out}$ ) Sobol 敏感性分析的一阶敏感系数分别为 0.450、0.304、0.064 和 0.026, 对设施内  $ET_0$  影响较大, 可作估算遮荫设施内  $ET_0$  的机器学习模型建模输入项。

3) 以  $RH_{out}$ 、 $u_{out}$ 、 $T_{max,out}$  和  $T_{out}$  作为估算遮荫设施内  $ET_0$  的首选气象参数组合, BO 优化的 ELM1 模型具有较小的计算代价, 较高的估算精度和稳定性, 测试集  $R^2$ 、RMSE 和 MAE 分别为 0.928、0.069 mm/d 和 0.046 mm/d。BO-ELM4 模型也能很好地适应输入少量气象参数 ( $RH_{out}$  和  $u_{out}$ ) 估算设施内  $ET_0$ ,  $R^2$ 、RMSE 和 MAE 分别为 0.910、

0.078 和 0.057 mm/d。BO-ELM 模型可用于估算遮荫设施内  $ET_0$ 。

## [参 考 文 献]

- [1] 刘子豪, 陆建忠, 黄建武, 等. 基于 CMIP5 模式鄱阳湖流域未来参考作物蒸散量预估[J]. 湖泊科学, 2019, 31(6): 1685-1697.  
Liu Zihao, Lu Jianzhong, Huang Jianwu, et al. Prediction and trend of future reference crop evapotranspiration in the Poyang Lake Basin based on CMIP5 Models[J]. Journal of Lake Sciences, 2019, 31(6): 1685-1697. (in Chinese with English abstract)
- [2] Allan R G, Pereira L S, Raes D, et al. Crop evapotranspiration-guidelines for computing crop water requirements[R]. Rome: FAO Irrigation and Drainage Paper 56, 1998.
- [3] Bellido-Jiménez J A, Estévez J, García-Marín A P. New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain[J]. Agricultural Water Management, 2021, 245: 106558.
- [4] Irmak S, Irmak A, Allen R G, et al. Solar and net radiation-based equations to estimate reference evapotranspiration in humid climates[J]. Journal of Irrigation and Drainage Engineering, 2003, 129(5): 336-347.
- [5] Blaney H F, Criddle W D. Determining water requirements in irrigated areas from climatological and irrigation data[R]. Soil Conservation Service Technical Paper 96. Washington DC: US Department of Agriculture, 1950.
- [6] Hamon W R. Estimating potential evapotranspiration[J]. Journal of Hydraulics Division, Proceedings of the American Society of Civil Engineers, 1961, 87: 107-120.
- [7] Hargreaves G H, Samani Z A. Reference crop evapotranspiration from temperature[J]. Applied Engineering in Agriculture, 1985, 1(2): 96-99.
- [8] Makkink G F. Testing the Penman formula by means of lysimeters[J]. Journal of the Institution of Water Engineers, 1957, 11(3): 277-288.
- [9] Priestley C H B, Taylor R J. On the assessment of the surface heat flux and evaporation using large-scale parameters[J]. Monthly Weather Review, 1972, 100(2): 81-92.
- [10] Hargreaves H. Moisture availability and crop production[J]. Transactions of the ASAE, 1975, 18(5): 980-984.
- [11] 蒙强, 刘静霞, 李玉庆, 等. 西藏高原灌区参考作物蒸散量模型的适用性研究[J]. 节水灌溉, 2020(6): 61-67, 72.  
Meng Qiang, Liu Jingxia, Li Yuqing, et al. Applicability study of different estimating methods for reference crop evapotranspiration in irrigation district in Tibet Plateau[J]. Water Saving Irrigation, 2020(6): 61-67, 72. (in Chinese with English abstract)
- [12] 李晨, 崔宁博, 冯禹, 等. 四川省不同区域参考作物蒸散量计算方法的适用性评价[J]. 农业工程学报, 2016, 32(4): 127-134.  
Li Chen, Cui Ningbo, Feng Yu, et al. Adaptation evaluation for reference evapotranspiration methods in different regions of Sichuan[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(4): 127-134. (in Chinese with English abstract)
- [13] 陈新明, 蔡焕杰, 李红星, 等. 温室大棚内作物蒸发蒸腾量计算[J]. 应用生态学报, 2007, 18(2): 317-321.  
Chen Xinming, Cai Huanjie, Li Hongxing, et al. Calculation of crop evapotranspiration in greenhouse[J]. Chinese Journal of Applied Ecology, 2007, 18(2): 317-321. (in Chinese with English abstract)
- [14] Mahmoodi-Eshkaftaki M, Rafiee M R. A simple-to-use nomogram for reference evapotranspiration assessment using multivariate models: Study on outdoor and greenhouse

- environments[J]. *Environmental Technology & Innovation*, 2021, 22: 101421.
- [15] Kitta E, Baille A, Katsoulas N, et al. Predicting reference evapotranspiration for greenhouse-grown crops[J]. *Agricultural Water Management*, 2014, 143: 122-130.
- [16] 贺登辉, 李芮林, 孙帅辉, 等. 基于机器学习的离心泵气液两相压升预测[J]. *农业工程学报*, 2022, 38(4): 33-41. He Denghui, Li Ruilin, Sun Shuaihui, et al. Prediction of gas-liquid two-phase pressure increment of a centrifugal pump based on machine learning[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(4): 33-41. (in Chinese with English abstract)
- [17] 王晓东, 安瑞东. 基于机器学习的热轧带钢力学性能预测模型及应用[J]. *塑性工程学报*, 2021, 28(10): 155-165. Wang Xiaodong, An Ruidong. Mechanical properties prediction models based on machine learning for hot rolled strip and application[J]. *Journal of Plasticity Engineering*, 2021, 28(10): 155-165. (in Chinese with English abstract)
- [18] 齐国庆, 吴东, 胡博, 等. 机器学习在脑功能障碍磁共振成像诊断中的应用[J]. *磁共振成像*, 2021, 12(10): 85-88. Qi Guoqing, Wu Dong, Hu Bo, et al. Application of machine learning in the diagnosis of brain dysfunction by magnetic resonance imaging[J]. *Chinese Journal of Magnetic Resonance Imaging*, 2021, 12(10): 85-88. (in Chinese with English abstract)
- [19] 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. Feng Yu, Cui Ningbo, Gong Daozhi, et al. Prediction model of reference crop evapotranspiration based on extreme learning machine[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2015, 31(Supp.1): 153-160. (in Chinese with English abstract)
- [20] 姚付启, 董建华, 范军亮, 等. 基于数值天气预报后处理的参考作物蒸散量预报改进[J]. *农业机械学报*, 2021, 52(7): 293-303. Yao Fuqi, Dong Jianhua, Fan Junliang, et al. Improvement of reference crop evapotranspiration forecasting based on numerical weather prediction post processing[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(7): 293-303. (in Chinese with English abstract)
- [21] Ahmadi F, Mehdizadeh S, Mohammadi B, et al. Application of an artificial intelligence technique enhanced with intelligent water drops for monthly reference evapotranspiration estimation[J]. *Agricultural Water Management*, 2021, 244: 106622.
- [22] Mohammadi B, Mehdizadeh S. Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm[J]. *Agricultural Water Management*, 2020, 237: 106145.
- [23] Petković B, Petković D, Kuzman B, et al. Neuro-fuzzy estimation of reference crop evapotranspiration by neuro fuzzy logic based on weather conditions[J]. *Computers and Electronics in Agriculture*, 2020, 173: 105358.
- [24] 崔宁博, 魏俊, 赵璐, 等. 基于 MEA-BPNN 的西北旱区参考作物蒸散量预报模型[J]. *农业机械学报*, 2018, 49(8): 228-236, 307. Cui Ningbo, Wei Jun, Zhao Lu, et al. Reference crop evapotranspiration prediction model of arid areas of Northwest China based on MEA-BPNN[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(8): 228-236, 307. (in Chinese with English abstract)
- [25] 毛亚萍, 房世峰. 基于机器学习的参考作物蒸散量估算研究[J]. *地球信息科学学报*, 2020, 22(8): 1692-1701. Mao Yaping, Fang Shifeng. Research of reference evapotranspiration's simulation based on machine learning[J]. *Journal of Geo-Information Science*, 2020, 22(8): 1692-1701. (in Chinese with English abstract)
- [26] 董建华, 刘小刚, 吴立峰, 等. 基于机器学习的 ET<sub>0</sub> 跨站适应性研究[J]. *西北农林科技大学学报(自然科学版)*, 2021, 49(9): 144-154. Dong Jianhua, Liu Xiaogang, Wu Lifeng, et al. Cross-station adaptability of ET<sub>0</sub> based on machine learning[J]. *Journal of Northwest A&F University (Natural Science Edition)*, 2021, 49(9): 144-154. (in Chinese with English abstract)
- [27] 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. Zhang Qian, Wei Zhengying, Zhang Yubin, et al. A prediction of greenhouse reference evapotranspiration forecasting based on Fire Algorithm optimized Extreme Learning Machine[J]. *China Rural Water and Hydropower*, 2020(3): 29-32, 38. (in Chinese with English abstract)
- [28] 仇文岗, 唐理斌, 陈福勇, 等. 基于 4 种超参数优化算法及随机森林模型预测 TBM 掘进速度[J]. *应用基础与工程科学学报*, 2021, 29(5): 1186-1200. Zhang Wengang, Tang Libin, Chen Fuyong, et al. Prediction for TBM penetration rate using four hyperparameter optimization methods and random forest model[J]. *Journal of Basic Science and Engineering*, 2021, 29(5): 1186-1200. (in Chinese with English abstract)
- [29] 周闰明, 石朋, 瞿思敏, 等. 基于贝叶斯优化算法的新安江模型参数优化及应用[J]. *水电能源科学*, 2019, 37(5): 6-9, 127. Zhou Yanming, Shi Peng, Qu Simin, et al. Application of Bayesian optimization to calibrate parameters of Xin'anjiang hydrological model[J]. *Water Resources and Power*, 2019, 37(5): 6-9, 127. (in Chinese with English abstract)
- [30] Sobol I M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates[J]. *Mathematics and Computers in Simulation*, 2001, 55(1/2/3): 271-280.
- [31] Tang Y, Reed P, Wagener T, et al. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation[J]. *Hydrology and Earth System Sciences*, 2007, 11(2): 793-817.
- [32] Smola A J, Schölkopf B. A tutorial on support vector regression[J]. *Statistics and computing*, 2004, 14(3): 199-222.
- [33] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [34] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feedforward neural networks[C]// *IEEE International Joint Conference on Neural Networks*. Budapest, Hungary: IEEE, 2004.
- [35] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述[J]. *软件学报*, 2018, 29(10): 3068-3090. Cui Jiayu, Yang Bo. Survey on Bayesian optimization methodology and applications[J]. *Journal of Software*, 2018, 29(10): 3068-3090. (in Chinese with English abstract)
- [36] Bergstra J, Bardenet R, Bengio Y, et al. Algorithms for hyper-parameter optimization[C]// *Advances in Neural Information Processing Systems*. Granada, Spain: Curran Associates Inc, 2011: 2546-2554.
- [37] 刘海军, 黄冠华, Shabtai Cohen, 等. 网室内作物蒸散量的变化和影响因素研究[J]. *中国生态农业学报*, 2009, 17(3): 484-488. Liu Haijun, Huang Guanhua, Shabtai Cohen, et al. Change in crop evapotranspiration and associated influencing factors under greenhouse conditions[J]. *Chinese Journal of Eco-Agriculture*, 2009, 17(3): 484-488. (in Chinese with English abstract)
- [38] 夏兴生, 朱秀芳, 潘耀忠, 等. 地表太阳辐射经验值对参

- 考作物需水量计算的影响[J]. 农业机械学报, 2020, 51(2): 254-266.
- Xia Xingsheng, Zhu Xiufang, Pan Yaozhong, et al. Influence of solar radiation empirical values on reference crop evapotranspiration calculation in different regions of China[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(2): 254-266. (in Chinese with English abstract)
- [39] 张薇, 韦群, 吴天傲, 等. 基于 GBDT 算法的参考作物蒸散量模型在江苏省的预测[J]. 江苏农业学报, 2020, 36(5): 1169-1180.
- Zhang Wei, Wei Qun, Wu Tian'ao, et al. Prediction models of reference crop evapotranspiration based on gradient boosting decision tree(GBDT) algorithm in Jiangsu province[J]. Jiangsu Journal of Agricultural Sciences, 2020, 36(5): 1169-1180. (in Chinese with English abstract)
- [40] 刘泉杉, 吴宗俊, 崔宁博, 等. 云贵高原参考作物蒸散量时空特征及成因分析[J]. 排灌机械工程学报, 2022, 40(3): 302-310.
- Liu Quanshan, Wu Zongjun, Cui Ningbo, et al. Spatial-temporal distribution characteristics and attribution analysis of reference crop evapotranspiration in Yunnan-Kweichow Plateau[J]. Journal of Drainage and Irrigation Machinery Engineering, 2022, 40(3): 302-310. (in Chinese with English abstract)
- [41] Möller M, Cohen S, Pirkner M, et al. Transmission of short-wave radiation by agricultural screens[J]. Biosystems Engineering, 2010, 107(4): 317-327.
- [42] 尹起, 周建平, 许燕, 等. 基于粒子群优化 (PSO) 超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.
- Yin Qi, Zhou Jianping, Xu Yan, et al. Prediction of reference crop evapotranspiration in Xinjing based on particle swarm optimization (PSO) optimized extreme learning machine[J]. Jiangsu Journal of Agricultural Sciences, 2021, 37(3): 622-631. (in Chinese with English abstract)

## Estimation of reference evapotranspiration in shading facility using machine learning

Chen Shaomin<sup>1</sup>, Li Xiaoli<sup>1</sup>, Yang Qiliang<sup>1\*</sup>, Wu Lifeng<sup>2</sup>, Xiong Kai<sup>3</sup>, Liu Xiaogang<sup>1</sup>

(1. Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China;

2. School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China;

3. School of Hydraulic and Environmental Engineering, Changsha University of Science & Technology, Changsha 410114, China)

**Abstract:** Reference evapotranspiration ( $ET_0$ ) is one of the most important parameters to calculate the crop water demand. The key physical quantity of the water cycle can also pose a great challenge to the water balance for the decision-making on the agricultural water use plan at present. The  $ET_0$  in the facilities can be generally estimated by the improved empirical formula with high accuracy. But, much more meteorological parameters are required during estimation. It is prohibitively expensive for the experimental cost of  $ET_0$  estimation using measurement instruments for the meteorological parameters in the shading facilities. Alternatively, machine learning can be expected to easily obtain the meteorological parameters outside the facilities. However, only a few studies were focused on the estimation of  $ET_0$  in this case. In this study, an efficient and accurate estimation of the  $ET_0$  was proposed to clarify the relationship between the meteorological parameters inside and outside of *Panax notoginseng* shading facility. A Sobol sensitivity analysis was implemented to determine the effective meteorological parameters outside the facility as the model input. A Penman-Monteith model was used to calculate the standard values. Three  $ET_0$  estimation models (BO-SVR, BO-RF, and BO-ELM) were established, where the Bayesian Optimization (BO) was used to optimize the parameters in the Support Vector Regression (SVR), Random Forest (RF), and Extreme Learning Machine (ELM). The results showed that there was a strong correlation between six meteorological parameters inside and outside the shading facility, among which the average temperature, the maximum temperature, the minimum temperature, and average relative humidity were significantly correlated, and the coefficient of determination ( $R^2$ ) values were 0.914, 0.721, 0.925 and 0.923, respectively. The radiation term was close to 0 in the shading facility. The  $ET_0$  was approximately equal to the aerodynamic term, where the  $R^2$ , the Root-Mean-Square Error (RMSE), and the Mean Absolute Error (MAE) were 0.999, 0.008 mm/d, and 0.006 mm/d, respectively. There was a strong correlation between the aerodynamic terms inside and outside the shading facility, where the  $R^2$ , RMSE, and MAE were 0.856, 0.097 mm/d, and 0.073 mm/d, respectively. Therefore, it was feasible to estimate the  $ET_0$  in the shading facility of *Panax notoginseng* using the meteorological factors outside the facility. In Sobol sensitivity analysis, the  $ET_0$  in the shading facility was highly sensitive to the average relative humidity, average wind speed, maximum temperature, and average temperature, with the first-order sensitivity coefficients of 0.450, 0.304, 0.064, and 0.026, respectively. There was a small influence of the minimum temperature and sunshine duration on the  $ET_0$ , where the first-order sensitivity coefficients were less than 0.01. Therefore, an optimal combination of four meteorological parameters was constructed for the improved model. The overall performance of the BO-ELM model in the test accuracy was better than those of the BO-SVR and BO-RF models. The highest accuracy was achieved for the BO-ELM model using the optimal combination of average relative humidity, average wind speed, the maximum temperature, and average temperature, particularly with the  $R^2$ , RMSE, and MAE of 0.928, 0.069 mm/d, and 0.046 mm/d, respectively. The BO-ELM model was also well adapted to estimate the  $ET_0$  in the facility with a small number of meteorological parameters (average relative humidity, and average wind speed), with the  $R^2$ , RMSE, and MAE of 0.910, 0.078 mm/d, and 0.057 mm/d, respectively. The computational cost of each estimation model was calculated from the parameter tuning time and modeling time of the model. Overall, the BO-SVR and BO-ELM models presented relatively short running time of 0.97 and 2.22 s, respectively. By contrast, the longest running time of 27.46 s was obtained in the BO-RF model. Therefore, the BO-ELM model can be expected to serve as the  $ET_0$  estimation in the shading facility in the absence of some meteorological parameters, fully considering the calculation accuracy and cost of the simulation. The findings can also provide an effective way for the estimation of  $ET_0$  in the shading facilities.

**Keywords:** crop; evapotranspiration; models; machine learning; sensitivity analysis; Bayesian optimization algorithm; *Panax notoginseng*