

考虑日光诱导叶绿素荧光的冬小麦蒸散量模拟

李尧^{1,2} 刘江舟^{1,2} 刘轩昂^{1,2} 赵政鑫^{1,2} 彭雄标^{1,2} 蔡焕杰^{1,2}

(1. 西北农林科技大学水利与建筑工程学院, 陕西杨凌 712100;

2. 西北农林科技大学中国旱区农业节水研究院, 陕西杨凌 712100)

摘要: 为探究在气象数据缺失情况下机器学习模型对冬小麦生育期实际蒸散量 (Actual evapotranspiration, ETa) 的模拟效果以及日光诱导叶绿素荧光 (Solar-induced chlorophyll fluorescence, SIF) 对于机器学习模型模拟 ETa 的优势, 将 SIF 与气象、作物生理指标、土壤水热条件等因素相结合, 构建梯度上升 (Gradient boosting, GB)、随机森林 (Random forest, RF) 和支持向量机 (Support vector machine, SVM) 3 种经典机器学习模型和线性回归 (Linear regression, LR) 模型模拟冬小麦生育期 ETa, 并与 Penman - Monteith (P - M) 模型计算得到的蒸散量 ET_{pm} 进行对比。结果表明: SIF 与 ETa 显著相关, 但仅通过 SIF 作为特征参数构建的机器学习模型拟合精度较低; 根据基于机器学习模型的特征参数重要度排序以及各情景下的模型模拟精度可知, SIF 对机器学习模型模拟 ETa 的精度有提升效果。机器学习模型在有足够的特征参数时拟合效果明显优于 P - M 模型, 且在平均温度、SIF、日照时数、叶面积指数 (Leaf area index, LAI) 和土壤含水率的基础上继续添加特征参数对模拟精度提升不大, 因此推荐使用上述 5 个特征参数组成的特征集构建机器学习模型进行 ETa 预测, 模型决定系数 R^2 分别为 0.92、0.91 和 0.91, 其中 GB 模型对冬小麦全生育期 ETa 的拟合效果最好。该研究可在气象数据缺失情况下为当地蒸散量的精准模拟和合理灌溉制度制定提供参考。

关键词: 冬小麦; 蒸散量; 涡度相关系统; 日光诱导叶绿素荧光; 机器学习模型

中图分类号: S275.3; S152.7 文献标识码: A 文章编号: 1000-1298(2025)05-0534-09

OSID:



Simulation of Evapotranspiration in Winter Wheat Considering Solar-induced Chlorophyll Fluorescence

LI Yao^{1,2} LIU Jiangzhou^{1,2} LIU Xuanang^{1,2} ZHAO Zhengxin^{1,2} PENG Xiongbiao^{1,2} CAI Huanjie^{1,2}

(1. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
2. Institute of Water-saving Agriculture in Arid Areas of China, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: In order to investigate the simulation effect of machine learning model on actual evapotranspiration (ETa) of winter wheat during the reproductive period and the effect of solar-induced chlorophyll fluorescence (SIF) on the simulation accuracy of machine learning model in the absence of meteorological data, SIF was combined with meteorological indicators, crop physiological indicators, soil thermal conditions and other factors, and three classical machine learning models, namely the gradient boosting (GB), random forest (RF), and support vector machine (SVM) were constructed, combined with linear regression (LR) model to simulate winter wheat ETa and compared with the evapotranspiration ET_{pm} calculated by Penman - Monteith (P - M) model. The results showed that although SIF was significantly correlated with ETa, the fitting accuracy of the machine learning model constructed only by using SIF as a feature parameter was low; according to the importance ranking of the feature parameters based on the machine learning model as well as the simulation accuracy of the model under each scenario, it was known that SIF had an enhancement effect on the accuracy of the machine learning model in simulating ETa. The machine learning model fit better than the P - M model when there were enough feature parameters, and adding feature parameters to the average temperature, SIF, sunshine hours, leaf

收稿日期: 2024-03-08 修回日期: 2024-05-08

基金项目: 国家自然科学基金项目 (52179046)

作者简介: 李尧 (1998—), 男, 博士生, 主要从事农业节水原理与技术研究, E-mail: 252933293@qq.com

通信作者: 蔡焕杰 (1962—), 男, 教授, 博士生导师, 主要从事农业节水原理与技术研究, E-mail: caihj@nwsuaf.edu.cn

area index (LAI) and soil moisture content did not improve the simulation accuracy, so it was recommended to use the feature set composed of the five feature parameters mentioned above to construct a machine learning model to predict ETa. The R^2 of the models were 0.92, 0.91 and 0.91, respectively, among which the GB model had the best fitting effect on the ETa of winter wheat during the whole reproductive period. The research result can provide a reference for the accurate simulation of local evapotranspiration and the development of rational irrigation system in the absence of meteorological data.

Key words: winter wheat; evapotranspiration; eddy-covariance system; solar-induced chlorophyll fluorescence; machine learning model

0 引言

农田实际蒸散量 (Actual evapotranspiration, ETa)是指地表和作物叶片表面的水分转变为水汽进入到大气的蒸发过程和植物体内的水分经过气孔向外扩散的蒸腾过程^[1]。作为农田生态系统水循环的重要组成部分,ETa 约占降水总量的 60%,且在干旱区可达到 90%,是研究作物生育期耗水量和生产力的重要指标^[2-3]。中国作为历史悠久的农业大国,农田面积占全国国土面积的 14%^[4]。因此,准确估计农田生态系统的 ETa 对于制定合理的灌溉制度,提高我国农业用水效率具有重要意义^[5]。

目前获得 ETa 的方法主要有田间测定、估算模型以及机器学习模型。田间实际测定法主要包括涡度相关法、蒸渗仪法以及波文比法等。如 WANG 等^[6]使用涡度相关系统测量了关中平原冬小麦夏玉米轮作各生育期的 ETa,分析了环境和生物因素对 ETa 的影响。袁宏伟等^[7]通过蒸渗仪法测量了淮北平原玉米的 ETa 并分析了不同受旱胁迫处理对 ETa 的影响。但田间实际测定法繁琐且有很强的地域局限性,难以实现大尺度的 ETa 测定^[8]。1980 年联合国粮食及农业组织将 Penman - Monteith (P - M)模型作为标准模型计算参考作物蒸散量,并在 FAO - 56 文件中提供了各类作物在不同生育阶段的作物系数推荐值进而计算作物生育期 ETa^[9]。宋孝玉等^[10]采用 P - M 模型和 FAO - 56 推荐的作物系数计算了鄂尔多斯高原鄂托克旗境内西部天然草地 ETa,分析了各草地类型生态需水和降雨资源之间的关系。但 P - M 模型对气象数据的完整性要求较高,在气象数据缺失地区,尤其是干旱或半干旱区,P - M 模型的应用受到限制,此外 FAO - 56 推荐的作物系数是关于时间的函数,对土壤环境、作物生长情况和水分胁迫等因素考虑不全面,进而使得在计算田间 ETa 时存在一定误差^[11]。随着机器学习算法的不断更新,一些学者开始使用机器学习模型解决气象、作物生理指标、土壤水热条件等数据与 ETa 之间的复杂非线性关系。如 MOHAMMADI 等^[12]使用鲸鱼优化算法优化后的支持向量回归模

型建立了部分气象数据与日蒸散量之间的机器学习模型,并与 P - M 模型计算的数据进行比较,证明了机器学习模型对日蒸散量有较好的拟合效果。SHI 等^[13]收集了已发表的机器学习模型模拟作物蒸散量的文献并进行 Meta 分析,证明了气象、作物生理指标和其他辅助预测因素的结合能有效模拟作物蒸散量,与其他模型相比,支持向量机和随机森林模型具有高使用频率和高模拟精度的优点。

日光诱导叶绿素荧光 (Solar-induced chlorophyll fluorescence, SIF)是指绿色植物通过光合作用吸收的有效辐射通过光合活动和非光化学淬灭过程被消耗后,有一小部分以红光和近红外光子(波长 640 ~ 850 nm)重新释放的过程^[14]。在以往的研究中,多位学者已证明 SIF 可以作为作物总初级生产力的替代参数^[15]。由于蒸散发和光合作用均为自然耦合的过程,因此研究者试图将 SIF 与 ETa 联系起来,揭示通过 SIF 估算 ETa 的潜力^[16]。如 WANG 等^[17]推导了冠层水汽传导与 SIF 之间的关系,提出了评估 ETa 的半机制模型。HAMED 等^[18]构建了水、能源、碳相结合的人工神经网络模型,进行地表湍流以及作物总初级生产力估算,间接表明了 SIF 与 ETa 有很强的非线性相关性。但将 SIF 作为特征参数建立机器学习模型直接估算 ETa 的研究相对较少。

因此,本文将涡度相关系统测量的 ETa 作为标准值,构建气象、作物生理指标、土壤水热条件和 SIF 等特征参数相结合的机器学习模型,对关中平原冬小麦全生育期的 ETa 进行逐日模拟,并与 P - M 模型计算得到的 ET_{pm} 结果进行精度比较,探究在气象数据缺失时机器学习模型对 ETa 的模拟效果以及引入 SIF 参数对模型精度的影响,为当地制定高效节水灌溉制度提供科学依据。

1 材料与方法

1.1 研究区概况

田间试验于 2020 年 10 月中下旬—2022 年 6 月初在西北农林科技大学农业水土工程重点实验室冬小麦农田(34°17'45"N, 108°04'07"E)进行 2 个生育期种植,如图 1 所示。研究区南北长约 200 m,东西

长约 250 m, 为半湿润干旱地区, 年平均日照时数 2 000 h 以上, 年均气温 12.9℃, 降雨量 560 mm, 蒸发量 1 500 mm, 土壤类型为粉质黏壤土, 田间持水量为 0.235 cm³/cm³, 容重为 1.35 g/m³^[19-20]。每年 10 月中下旬进行播种, 从播种后至次年 1 月中旬为冬小麦越冬期, 1 月下旬—3 月中旬为返青期, 3 月下旬—4 月中旬为拔节期, 4 月下旬—5 月上旬为抽穗—灌浆期, 5 月中旬—6 月上旬为乳熟—成熟期。播种前施用特殊氨基酸钙复混肥(氮肥 172.5 kg/hm², 磷肥 90 kg/hm², 钾肥 37.5 kg/hm²)。同时, 在每个生育期均施用 1 次除草剂和杀虫剂。

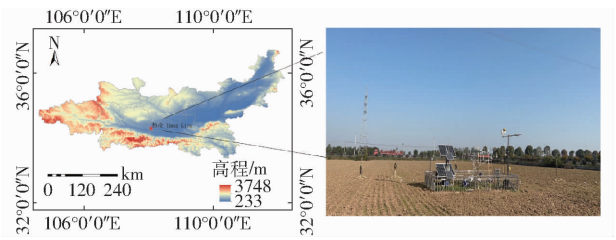


图 1 研究区概况图

Fig.1 Overview map of study area

1.2 数据测量和预处理

1.2.1 气象、土壤水热、作物生理指标和 ETa 数据

在田间放置 1 个开放路径涡度相关系统(Open path eddy-covariance system, OPEC)和 1 个气象环境观测系统获取气象和 ETa 数据, 经过计算, 冬小麦各生育期的能量平衡离散程度符合 FLUXNET 站点报告的能量平衡闭合(Energy balance closure, EBC)范围。使用介电常数计(5TM106 METER Environment, 美国)获取土壤水热数据。利用冠层分析仪(LAI-2200c 型, LI-COR Inc., 美国)测量冬小麦生长季节的叶面积指数(Leaf area index, LAI)。具体操作步骤见文献[21]。

1.2.2 SIF 数据

将自动六通道塔基 SIF 仪器(AUTOSIF-2-8 型)部署在植物冠层上方约 2 m 处的 6 个三脚架上, 对涡度相关系统周围的出射冠层辐射进行采样。与只有 1 个单一植被目标的 SIF 测量系统相比, 该六通道系统减少了 SIF 和通量测量之间的足迹失配。该系统包含定制的分光辐射计(QE Pro, Ocean Optics, 美国), 信噪比为 1 000, 光谱分辨率为 0.34 nm, 采样间隔为 0.17 nm, 光谱范围为 645 ~ 805 nm。使用光学多路复用器(MPM-2000 型, Ocean Optics, 美国)和余弦校正器(CC-3 型, Ocean Optics, 美国)校正暗电流的盲通道和具有 25 nm 视场的 6 个最低点裸光纤之间切换以获得上升流辐射。因此, SIF 仪器具有 6 个测量足迹圆, 每个测量足迹圆直径约为 0.9 m。分光辐射计和多路

复用器都被封装在(25 ± 1)℃ 的干燥温控容器中。在校正暗电流和优化积分时间后, 应用“sandwich”方法来减少不稳定天气条件的负面影响^[22]。在安装前, 对系统进行辐射和光谱校准。

使用奇异向量分解(SVD)法从连续光谱测量中检索 SIF, 该方法假设无 SIF 频谱可以表示为从无 SIF 训练数据集中获得的奇异向量的线性组合^[23]。采用 740 ~ 780 nm 光谱拟合窗口反演近红外波段冠层 SIF 观测值, 然后将检索到的 SIF 值汇总为日平均数据^[15]。

1.2.3 归一化处理

收集 4 个气象指标(平均温度、相对湿度、风速、日照时数)、2 个土壤水热指标(土壤含水率、土壤温度)、1 个作物生理指标(LAI)和 SIF 作为建立机器学习模型的特征参数, 为降低不同特征数量级对模型精度的影响, 在进行特征参数重要度评估和建立机器学习模型前对所选的 8 个特征参数进行标准化处理^[8]。

1.3 研究方法

1.3.1 相关性分析

为评估 SIF 与 ETa 之间的相关性, 计算了两者之间的 Pearson 和 Spearman 相关系数, 其中 Pearson 相关系数衡量的是线性相关关系, Spearman 相关系数衡量分级定序变量之间的相关程度^[24]。

1.3.2 特征参数重要度评估

为筛选出有效的特征参数进而减少收集数据的工作量, 并检验引入 SIF 指标是否有利于提升 ETa 的预测精度, 使用基于机器学习模型的特征参数重要度评估法(Model based ranking)对收集到的 8 个特征参数在包括 SIF 和剔除 SIF 情况下进行 2 次重要度排序。重要度数值能够体现所选特征与 ETa 之间的线性或非线性关系强度, 所有特征的重要度数值之和为 1, 其中单个特征的数值越大证明对 ETa 的重要程度越高^[25]。

1.3.3 机器学习模型

使用梯度上升(Gradient boosting, GB)、随机森林(Random forest, RF)和支持向量机(Support vector machine, SVM)3 种经典机器学习模型和线性回归(Linear regression, LR)模型进行 ETa 的模拟。其中 GB 模型可以通过每步都沿着目标函数的梯度方向移动一定长度, 使目标函数值增大, 进而求解模型参数的最优值^[26]。RF 模型以决策树为基本学习器, 从一系列有特征和标签的数据中总结出决策规则, 用树状图结构来呈现这些规则并在 Bagging 集成学习算法的基础上, 进一步采用自助采样法对特征进行随机选择以解决分类和回归问题^[27]。SVM 模型

可通过核方法将给定的数据集进行非线性分类,使不同类别之间的间隔最大化,进而解决高维特征的分类回归问题^[28]。LR 模型是利用最小平方函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析,是一个或多个回归系数的模型参数的线性组合^[26]。

为了检验机器学习模型的实际性能,提高模型泛化能力,首先将本研究收集到的 2017 年数据(175 组)在冬小麦生育期各阶段随机抽取 70% 作为训练集,30% 作为验证集,通过十重交叉验证和网格搜索找到最佳超参数进行模型的参数优化,随后将 2017 年所有数据作为新的训练集,2018 年数据(201 组)作为测试集预测 ETa,各模型的超参数范围见表 1。

表 1 机器学习模型参数

Tab.1 Machine learning models parameters

机器学习模型	参数
GB	n_estimators:[10, 50, 100]
	learning_rate:[0.01, 0.1, 0.5]
	max_depth:[3, 5, 10]
RF	n_estimators:[10, 50, 100]
	max_depth:[None, 10, 20]
	min_samples_split:[2, 5, 10]
	min_samples_leaf:[1, 2, 4]
SVM	C:[0.1, 1, 10]

1.3.4 Penman – Monteith 模型

根据气象环境观测系统监测的气象数据应用 Penman – Monteith 公式计算出参考作物蒸发蒸腾量 ET₀,然后根据单作物系数法计算得出实际蒸散量 ET_{pm}^[29]。

1.4 模型验证

使用归一化均方根误差 (Normal root mean square error, NRMSE) 和决定系数 (R^2)^[30-32] 评价模型精度,其中 NRMSE 越小证明模拟效果越好, R^2 越大证明模拟效果越好。

使用 Excel 2020 进行数据采集和分类;Python 3.8 中的 Scikit – Learn 库进行特征参数重要度排序、机器学习模型和线性回归模型建立;Cropwat 进行 ET₀ 计算;SPSS 22.0 进行相关性分析;Origin 2018 进行图表绘制。

2 结果与分析

2.1 SIF – ETa 相关性及模拟

2020—2022 年冬小麦生育期的 SIF 和 ETa 变化规律如图 2 所示。由图 2 可知,SIF 与 ETa 的整体变化趋势基本相同,但 SIF 数据的峰值一般出现在 4—5 月并在 5 月之后有逐渐下降趋势,而 ETa 数

据的峰值一般出现在 5 月中下旬,因此在每年的 5 月 1 日至冬小麦成熟,ETa 归一化数据的日平均值超出 SIF 数据 0.16。通过分析两组数据的相关性可知,SIF 与 ETa 的 Pearson 和 Spearman 相关系数分别为 0.80 和 0.81,可见两组数据相关性极强,因此考虑将 SIF 数据单独作为特征参数建立机器学习模型预测 ETa,得到的各模型模拟精度见表 2。

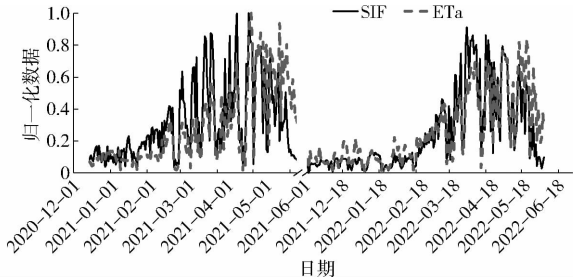


图 2 冬小麦生育期归一化 SIF 和 ETa 变化曲线

Fig.2 Variation curves of normalized SIF and ETa during growth period of winter wheat

表 2 基于 SIF 的各模型模拟精度

Tab.2 Simulation accuracy of SIF-based models

模型	NRMSE	R^2
GB	0.21	0.67
RF	0.24	0.62
SVM	0.23	0.63
LR	0.19	0.70

通过比较 4 个模型评价指标(表 2)可知,各模型对 ETa 的模拟效果均不理想, R^2 均值仅为 0.66。由于给出的特征参数过少,导致机器学习模型的模拟精度不高,LR 模型的模拟精度优于机器学习模型, R^2 为 0.7,机器学习模型中 GB 模型模拟精度最高, R^2 为 0.67,RF 模型精度最低, R^2 为 0.62。综上,仅用 SIF 作为输入特征建立的机器学习模型难以实现对 ETa 的精确模拟,需要考虑将 SIF 与气象、作物生理指标和土壤水热条件等因素相结合进而提高各模型拟合精度。

2.2 模拟情景设置

所选特征参数在机器学习模型中对预测 ETa 的重要度排序如表 3 所示,在包含 SIF 时,重要度超过 0.1 的特征参数分别为平均温度(0.510)、SIF(0.240)和日照时数(0.141),而在不包含 SIF 时,重要度超过 0.1 的特征参数分别为平均温度(0.621)、日照时数(0.195)和 LAI(0.126),可见 SIF 在预测 ETa 的机器学习模型中重要度仅次于平均温度。

在两组特征中,单个特征的重要度最大值仅为 0.621,不存在单个特征重要度过大而不必考虑其他特征的情况,各组中排名前 3 的特征重要度均超过

表 3 机器学习模型中特征参数重要度
Tab.3 Importance of feature parameters in machine learning models

特征参数有无 SIF	特征参数	特征重要度
有 SIF	平均温度	0.510
	SIF	0.240
	日照时数	0.141
	LAI	0.057
	土壤含水率	0.016
	土壤温度	0.014
	空气相对湿度	0.013
	风速	0.009
无 SIF	平均温度	0.621
	日照时数	0.195
	LAI	0.126
	空气相对湿度	0.019
	土壤温度	0.017
	土壤含水率	0.011
	风速	0.010

0.1 且之和分别为 0.891 和 0.942,因此为了得到模拟精度较高且特征参数数量较少的机器学习模型,首先选择两组中重要度排名前 3 的特征参数组成 S1 和 NS1 情景,并在之后的模拟情景中按重要度从大到小依次增加一个特征参数,共设置 11 个模拟情景进行冬小麦生育期的 ETa 模拟,情景设置见表 4。

表 4 模拟 ETa 的机器学习模型情景设置
Tab.4 Machine learning model scenarios settings for simulating ETa

特征参数有无 SIF	模拟情景	特征参数
有 SIF	S1	平均温度、SIF、日照时数
	S2	S1、LAI
	S3	S2、土壤含水率
	S4	S3、土壤温度
	S5	S4、空气相对湿度
	S6	S5、风速
无 SIF	NS1	平均温度、日照时数、LAI
	NS2	NS1、空气相对湿度
	NS3	NS2、土壤温度
	NS4	NS3、土壤含水率
	NS5	NS4、风速

2.3 机器学习模型与模拟情景选择

各模拟情景的机器学习和 LR 模型拟合精度如表 5 所示,GB、RF 和 SVM 模型在各模拟情景中验证集、测试集和全集中 R^2 均值分别为 0.89、0.87 和 0.88,NRMSE 均值分别为 0.08、0.09 和 0.08,而 LR 模型在全集中各情景的 R^2 均值仅为 0.72,NRMSE 为 0.19。可见在各情景中,机器学习模型的拟合效果均优于 LR 模型,且机器学习模型中的 GB 模型拟合精度最高,RF 模型拟合精度最低。

表 5 各情景下机器学习模型对冬小麦生育期 ETa 的模拟精度
Tab.5 Simulation accuracy of machine learning models for winter wheat fertility ETa under various scenarios

数据集	评价指标	模型	模拟情景										
			S1	S2	S3	S4	S5	S6	NS1	NS2	NS3	NS4	NS5
验证集	NRMSE	GB	0.08	0.08	0.05	0.05	0.04	0.04	0.10	0.08	0.07	0.06	0.05
		RF	0.10	0.08	0.06	0.06	0.06	0.05	0.12	0.10	0.07	0.07	0.07
		SVM	0.09	0.08	0.05	0.05	0.05	0.04	0.11	0.09	0.07	0.07	0.05
	R ²	GB	0.88	0.89	0.93	0.93	0.95	0.96	0.85	0.88	0.91	0.92	0.94
		RF	0.86	0.88	0.92	0.92	0.92	0.93	0.82	0.86	0.9	0.9	0.91
		SVM	0.87	0.89	0.93	0.93	0.94	0.95	0.84	0.87	0.91	0.91	0.93
测试集	NRMSE	GB	0.12	0.10	0.08	0.08	0.07	0.07	0.13	0.11	0.10	0.08	0.08
		RF	0.13	0.11	0.08	0.08	0.07	0.07	0.13	0.12	0.11	0.09	0.09
		SVM	0.12	0.10	0.10	0.10	0.08	0.08	0.13	0.11	0.10	0.09	0.09
	R ²	GB	0.82	0.86	0.89	0.89	0.9	0.91	0.81	0.84	0.86	0.88	0.88
		RF	0.8	0.83	0.88	0.88	0.9	0.9	0.8	0.82	0.84	0.87	0.87
		SVM	0.82	0.85	0.86	0.86	0.88	0.88	0.81	0.83	0.85	0.87	0.87
全集	NRMSE	GB	0.10	0.08	0.06	0.06	0.05	0.05	0.11	0.10	0.08	0.07	0.07
		RF	0.11	0.10	0.07	0.07	0.07	0.07	0.13	0.11	0.10	0.08	0.08
		SVM	0.10	0.09	0.07	0.06	0.06	0.05	0.12	0.10	0.08	0.08	0.08
		LR	0.21	0.19	0.18	0.16	0.15	0.14	0.28	0.23	0.20	0.18	0.15
	R ²	GB	0.86	0.88	0.92	0.92	0.93	0.94	0.83	0.86	0.89	0.9	0.91
		RF	0.84	0.85	0.90	0.9	0.91	0.91	0.81	0.84	0.86	0.88	0.88
		SVM	0.85	0.87	0.91	0.92	0.92	0.93	0.82	0.85	0.88	0.89	0.89
		LR	0.7	0.7	0.72	0.75	0.76	0.78	0.65	0.69	0.72	0.74	0.76

与 NS1 ~ NS5 情景相比,S1 ~ S5 情景在特征数量相同时的验证集、测试集和全集均表现出更高的拟合精度, R^2 平均增加 0.02,NRMSE 平均降低 0.01。说明通过将 SIF 与气象、作物生理指标和土

壤水热等因素相结合对机器学习模型模拟 ET_a 的精度有提升效果。

通过比较考虑 SIF 的 6 个情景可知, $S1 \sim S3$ 情景模型的拟合精度提升较大, 各模型 R^2 平均增加 0.06, NRMSE 平均减小 0.04, 而 $S3 \sim S6$ 情景的拟合精度变化不大, 各模型 R^2 平均增加 0.02, NRMSE 平均减小 0.01。此外, 3 个机器学习模型在全集中的 R^2 在 $S3$ 情景均达 0.9 以上。因此, 推荐使用平均温度、SIF、日照时数、LAI 和土壤含水率 5 个特征组成的 $S3$ 情景构建机器学习模型进行 ET_a 预测。

2.4 机器学习模型与 P-M 模型模拟精度对比

机器学习模型在 $S3$ 情景下的模拟值、P-M 模

型模拟值和当地涡度相关系统的实测值如图 3 所示。与实测值相比, 机器学习模型的模拟值与实测值和变化幅度基本相同, 模拟值与实测值基本重合, 而 P-M 模型的模拟值普遍小于实测值。通过计算可知 GB、RF 和 SVM 模型的日均模拟值分别比实测值小 0.03、0.01、0.12 mm。图 3 显示, 虽然 SVM 模型的拟合精度较高, 但对冬小麦第 2 个生育期中乳熟-成熟期的 ET_a 模拟值普遍偏小。P-M 模型的日均模拟值比实测值小 0.7 mm, 在冬小麦的越冬期和乳熟-成熟期, ET_{pm} 对 ET_a 的低估更明显。综上, GB 模型对 ET_a 的拟合精度最高且模拟值与实测值的生育期变化趋势基本相同, 不存在明显的高估或低估现象。

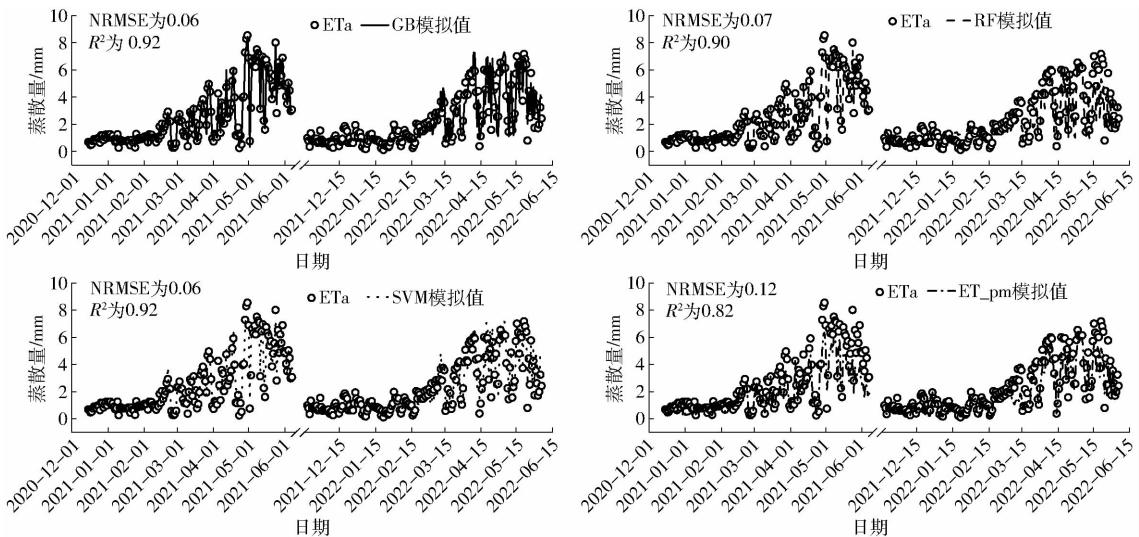


图 3 实测 ET_a 、P-M 模型计算的 ET_{pm} 与 GB、RF 和 SVM 模型在 $S3$ 情景的模拟值

Fig.3 Measured ET_a , ET_{pm} calculated according to P-M model and simulated values of GB, RF and SVM models in $S3$ scenario

通过对比各模型的 NRMSE 和 R^2 可知, P-M 模型的 R^2 最小, NRMSE 最大, 可见在 $S3$ 情景下, GB、RF 和 SVM 模型的拟合效果明显优于 P-M 模型。将 P-M 模型与表 5 中的其他模型对比可知, P-M 模型拟合效果优于所有情景的 LR 模型和 $NS1$ 情景的 RF 模型。

3 讨论

KHALID 等^[33]在机器学习的特征选择和特征提取研究中指出, 在数据维度较多时, 应在尽量保持模型较高精度的情况下降低维度。本文使用基于机器学习模型的特征参数重要度评估法进行特征的排序和筛选, 进而降低特征集维度。从特征参数重要度的排序中发现, 平均温度对预测 ET_a 的重要度最高, 证明平均温度与 ET_a 有很强的相关性, 这与陈志君等^[29]的研究结论类似, 这是因为在冬小麦生育期内温度的增加一方面会加速土壤和冠层表面水分

的蒸发进而直接影响 ET_a , 另一方面也可以通过调节作物气孔导度、土壤表面温度间接影响 ET_a 。WANG 等^[34]进行通径分析研究了气象、作物生理指标、土壤水热等因素对冬小麦生育期 ET_a 的影响, 证明了空气温度为对 ET_a 影响最明显的气象因素, 同时也证明了 ET_a 与各因素之间的复杂关系, 不能通过单个因素对 ET_a 进行预测。此外, MAES 等^[35]研究表明, 可以通过 GPP(作物总初级生产力)为中介进而构建 SIF 与 ET_a 的耦合关系, SIF 具有进一步提升 ET_a 预测精度的潜力, 本研究通过相关性检验证明了 SIF 与 ET_a 显著相关, 仅通过 SIF 数据建立的机器学习模型 R^2 均未超过 0.7, 但将 SIF 与气象、土壤水热、作物生理指标相结合建立的机器学习模型对 ET_a 的模拟精度高于无 SIF 参与的机器学习模型, 进而验证了 MAES 等^[35]的观点。ALLEN 等^[9]认为土壤的水热性质是影响 ET_a 的关键因素, XIAO 等^[36]的研究也指出当遇到干旱时, 作物会优化调节

气孔使光合作用最大化,水分散失最小化,此时空气相对湿度为影响 ETa 的重要指标。但在本研究的特征参数重要度排序中,重要度排名前 4 的指标分别为平均温度、SIF、日照时数和 LAI,而土壤水热性质相关指标和空气相对湿度的重要度均低于 0.1,对预测 ETa 的贡献度不高,这是由于本研究中的 2 个冬小麦生育期降雨量充足,分别为 305.3 mm 和 294.3 mm,且各次降雨在时间分配上也十分合理,使得土壤水热性质和空气相对湿度适宜冬小麦生长,并未产生干旱胁迫进而导致气孔关闭,影响作物蒸散。通过对比有 SIF 与无 SIF 的两组特征集的重要度可知,在添加 SIF 作为特征后,平均温度、日照时数等因素的特征重要度下降,土壤含水率的特征重要度上升。这是因为冬小麦进行光合作用需要足够的土壤水分,而 SIF 可以捕捉到冬小麦因水分胁迫而引起的光合作用变化。因此,当土壤含水率不足时,冬小麦会限制气孔开放以减少水分散失,导致光合作用降低,进而导致 SIF 下降。可见 SIF 提供了一个指示冬小麦水分利用和光合活动状态的额外维度,与土壤含水率相结合,可以更全面地反映冬小麦的蒸散发过程。

在机器学习模型选择中,本文发现 GB、RF 和 SVM 模型对于 ETa 的拟合效果均较好,各模型在 S3 情景下模拟值与实测值的 R^2 均在 0.9 以上,这与 SHI 等^[13]对预测蒸散量的机器学习模型进行荟萃分析(META)后推荐的蒸散量预测模型相同,META 分析结果表明在收集到的 139 个机器模型中,RF 和 SVM 的样本数量足够多且 R^2 集中分布在 0.75 ~ 0.95 之间且均值约为 0.82。本研究 GB、RF 和 SVM 模型在 S3 情景的 R^2 分别为 0.92、0.90 和 0.91,这在一定程度上侧面证明了本研究所选择的地区与气候类型和作物类型具有一定的普遍性,与大多数学者得出的模拟精度和推荐的模型均相同,并且模型模拟精度略高于前人模拟的均值。WANG 等^[34]在研究中证明了 ETa 与环境变量之间存在复杂的非线性关系,因此 LR 模型在本研究中对 ETa 的拟合精度最低。本研究中虽然 3 种机器学习模型的拟合精度均较高,但仅进行了单一地区单一作物的 ETa 模拟,并未改变地区和作物类型进行空间的交叉验证,因此在未来研究中应考虑对多种植被类型和地区的 ETa 进行模拟,进而评估机器学习模型对其他地区和作物

类型 ETa 拟合效果。

通过比较机器学习模型和 P-M 模型的拟合效果可知,GB、RF 和 SVM 模型在各情景下拟合效果均优于 P-M 模型,且 P-M 模型在整个冬小麦生育期的模拟值普遍小于实测值。PENG 等^[37]在研究关中平原冬小麦农田蒸散量分配时,使用 P-M 模型计算 ETa 同样证明了计算值小于当地涡度相关系统的实测值。这是由于 FAO-56 推荐的作物系数也是时间的函数,但在各生育阶段的作物系数确定后并不会再考虑实际生育期气候、种植品种、土壤理化性质以及田间管理措施等对模型准确度的影响,因此给 ETa 预测带来了较大误差^[29]。刘春伟等^[38]提出由于影响冬小麦全生育期的因素较多,应该分时段进行冬小麦不同生育阶段 ETa 模拟,进而得到预测全生育期 ETa 的最优方法。本研究中,SVM 模型在 S3 情景的拟合精度较高,但对冬小麦第 2 个生育期中乳熟-成熟期的 ETa 模拟值普遍偏小,这可能是由于冬小麦的越冬期较长且越冬期的 ETa 波动幅度并不大,而乳熟-成熟期的 ETa 虽然波动较大但数据点较少,导致用于训练模型的样本有限。因此有必要在本研究区继续收集数据,增加模型的样本集数量,以期实现利用更少的特征参数建立机器学习模型且对冬小麦各生育期 ETa 模拟均能取得更好的效果。

4 结 论

(1)冬小麦全生育期 SIF 与 ETa 有极强的相关性,但 SIF 达到峰值时间略早于 ETa 且仅用 SIF 构建机器学习模型对 ETa 的拟合精度较低,各机器学习模型的 R^2 均不高于 0.7。

(2)选取的 8 个特征参数重要度排序中,包含 SIF 的特征集重要度分布更均匀,排名前 3 的特征重要度之和高于不含 SIF 的特征集。在特征数量相同的情景中,包含 SIF 的情景对应的机器学习模型拟合精度更高, R^2 平均增加 0.02,可见将 SIF 加入机器学习模型的特征集中能增加模型拟合精度。

(3)3 种机器学习模型在 S1 ~ S6 情景的拟合效果均优于 P-M 模型,因此可以作为当地气象资料缺失时预测 ETa 的另一种方法。S4 情景构建的机器学习模型同时具有特征参数少和模拟精度高的优点,其中 GB 模型对冬小麦全生育期 ETa 的拟合效果最好。

参 考 文 献

[1] FISHER J B, WHITTAKER R J, MALHI Y. ET come home: potential evapotranspiration in geographical ecology[J]. Global Ecology and Biogeography, 2015, 20(1):1-18.

- [2] WANG Sheng, LIAN Jinjiao, PENG Yuzhong, et al. Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China[J]. *Agricultural Water Management*, 2019, 221: 220 – 230.
- [3] HU Zhongming, YU Guirui, ZHOU Yanliang, et al. Partitioning of evapotranspiration and its controls in four grassland ecosystems: application of a two-source model[J]. *Agricultural & Forest Meteorology*, 2009, 149(9): 1410 – 1420.
- [4] 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)
- [5] DOU Xianming, YANG Yongguo. Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems[J]. *Computers and Electronics in Agriculture*, 2018, 148: 95 – 106.
- [6] WANG Yunfei, ZOU Yufeng, CAI Huanjie. Seasonal variation and controlling factors of evapotranspiration over dry semi-humid cropland in Guanzhong Plain, China[J]. *Agricultural Water Management*, 2022, 259(1): 107242.
- [7] 袁宏伟, 崔毅, 蒋尚明, 等. 基于大型蒸渗仪和遗传算法的受旱玉米蒸发蒸腾量估算[J]. *农业机械学报*, 2018, 49(10): 326 – 335.
- YUAN Hongwei, CUI Yi, JIANG Shangming, et al. Estimation of maize evapotranspiration under drought stress based on large-scale lysimeter and genetic algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(10): 326 – 335. (in Chinese)
- [8] 李晨, 崔宁博, 魏新平, 等. 改进 Hargreaves 模型估算川中丘陵地区参考作物蒸散量[J]. *农业工程学报*, 2015, 31(11): 129 – 135.
- LI Chen, CUI Ningbo, WEI Xinping, et al. Improvement of Hargreaves method for reference evapotranspiration in hilly area of central Sichuan Basin[J]. *Transactions of the CSAE*, 2015, 31(11): 129 – 135. (in Chinese)
- [9] ALLEN R G, PEREIRA L S, RAES D, et al. *Crop evapotranspiration* [M]. Rome: United Nations FAO, 1998.
- [10] 宋孝玉, 刘雨, 覃琳, 等. 内蒙古鄂托克旗天然草地植被生态需水量研究[J]. *农业工程学报*, 2021, 37(3): 107 – 115.
- SONG Xiaoyu, LIU Yu, QIN Lin, et al. Ecological water requirement of natural grassland vegetation in the Otog Banner of Inner Mongolia[J]. *Transactions of the CSAE*, 2021, 37(3): 107 – 115. (in Chinese)
- [11] SHRESTHA N K, SHUKLA S. Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment[J]. *Bioresource Technology*, 2013, 128(2): 351 – 358.
- [12] 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.
- [13] SHI Haiyang, LUO Geping, OLAF H, et al. Evaluation of water flux predictive models developed using eddy-covariance observations and machine learning: a meta-analysis[J]. *Hydrology and Earth System Sciences*, 2022, 26: 4603 – 4618.
- [14] YANG J C, MAGNEY T S, ALBERT L P, et al. Gross primary production (GPP) and red solar induced fluorescence (SIF) respond differently to light and seasonal environmental conditions in a subalpine conifer forest[J]. *Agricultural and Forest Meteorology*, 2022, 317: 108904.
- [15] LIU Zhunqiao, ZHAO Feng, LIU Xinjie, et al. Direct estimation of photosynthetic CO₂ assimilation from solar-induced chlorophyll fluorescence (SIF)[J]. *Remote Sensing of Environment*, 2022, 271: 112893.
- [16] ZHOU Kai, ZHANG Quan, XIONG Lihua, et al. Estimating evapotranspiration using remotely sensed solar-induced fluorescence measurements[J]. *Agricultural and Forest Meteorology*, 2018, 49(9): 205 – 213.
- [17] WANG Renjun, ZHENG Jianghua, MAO Xurui, et al. Scaling solar-induced chlorophyll fluorescence by using VPD^{0.5} improves the simulation of reference crop evapotranspiration in the arid and semiarid regions of northern China[J]. *Journal of Hydrology*, 2023, 626: 130254.
- [18] HAMED A S, BIN F, KONINGS A G, et al. Water, energy, and carbon with artificial neural networks (WECANN): a statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence[J]. *Biogeosciences*, 2017, 14(18): 4101 – 4124.
- [19] ZHENG Jing, FAN Junliang, ZHANG Fucang, et al. Evapotranspiration partitioning and water productivity of rainfed maize under contrasting mulching conditions in Northwest China[J]. *Agricultural Water Management*, 2021, 243: 106473.
- [20] YU Lianyu, ZENG Yijian, SU Zhongbo, et al. The effect of different evapotranspiration methods on portraying soil water dynamics and ET partitioning in a semi-arid environment in Northwest China[J]. *Hydrology and Earth System Sciences*, 2016, 20(3): 975 – 990.
- [21] PENG Xiongbiao, WANG Yunfei, MA Jing, et al. Seasonal variation and controlling factors of carbon balance over dry semi-humid cropland in Guanzhong Plain[J]. *European Journal of Agronomy*, 2023, 149: 126912.
- [22] MERONI M, PICCHI V, ROSSINI M, et al. Leaf level early assessment of ozone injuries by passive fluorescence and photochemical reflectance index[J]. *Remote Sensing*, 2008, 29: 5409 – 5422.
- [23] CHANG C Y, GUANTER L, FRANGENBERG C, et al. Systematic assessment of retrieval methods for canopy far-red solar-induced chlorophyll fluorescence using high-frequency automated field spectroscopy[J]. *Journal of Geophysical Research: Biogeosciences*, 2020, 125(7): e2019JG005533.

[24] 张金屯. 数量生态学[M]. 3 版. 北京:科学出版社,2018:151 – 152.

[25] ZHOU Z, GILES H. Unbiased measurement of feature importance in tree-based methods[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2021, 15(2): 1 – 21.

[26] 周志华. 机器学习[M]. 北京:清华大学出版社,2016.

[27] WANG Yanwen, SONG Qingqing, DU Yanjun, et al. A random forest model to predict heatstroke occurrence for heatwave in China[J]. Science of the Total Environment, 2019, 650: 3048 – 3053.

[28] FAN Junliang, YUE Wenjun, WU Lifeng, et al. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China[J]. Agricultural and Forest Meteorology, 2018, 263: 225 – 241.

[29] 陈志君,朱振闯,孙仕军,等. Stacking 集成模型模拟膜下滴灌玉米逐日蒸散量和作物系数[J]. 农业工程学报,2021, 37(5): 95 – 104.

CHEN Zhijun, ZHU Zhenchuang, SUN Shijun. Estimation of daily evapotranspiration and crop coefficient of maize under mulched drip irrigation by Stacking ensemble learning model[J]. Transactions of the CSAE, 2021, 37(5): 95 – 104. (in Chinese)

[30] 周修理,秦娜,王开宇,等. 基于机器学习的机械压实对大豆产量的影响预测研究[J]. 农业机械学报,2023, 54(11): 139 – 147.

ZHOU Xiuli, QIN Na, WANG Kaiyu. Effect of mechanical compaction on soybean yeild based on machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(11):139 – 147. (in Chinese)

[31] MONTEITH J L. Principles of environmental physics[J]. Plant Growth Regulation, 1991, 10(2):177 – 178.

[32] 李尧,付玉娟,张旭东. 基于 Hydrus – 2D 分析农田覆膜对降雨入渗的影响[J]. 农业工程学报,2022,38(11):134 – 143.

LI Yao, FU Yujuan, ZHANG Xudong. Effects of farmland plastic mulching on rainfall infiltration using Hydrus – 2D model [J]. Transactions of the CSAE, 2022, 38(11): 134 – 143. (in Chinese)

[33] KHALID S, KHALIL T, NASREEN S. A survey of feature selection and feature extraction techniques in machine learning[C]// 2014 Science and Information Conference. IEEE, 2014.

[34] WANG Yu, ZHOU Li, JIA Qingyu, et al. Direct and indirect effects of environmental factors on daily CO₂ exchange in a rainfed maize cropland—a SEM analysis with 10 year observations[J]. Field Crops Research, 2019, 242: 107591.

[35] MAES W H, PAGAN B R, MARTENS B, et al. Sun-induced fluorescence closely linked to ecosystem transpiration as evidenced by satellite data and radiative transfer models[J]. Remote Sensing of Environment, 2020, 249:112030.

[36] XIAO Jingfeng, FISHER J B, HASHIMOTO H, et al. Emerging satellite observations for diurnal cycling of ecosystem processes[J]. Nature Plants, 2021(7):877 – 887.

[37] PENG Xiongbiao, LIU Xuanang, WANG Yunfei, et al. Evapotranspiration partitioning and estimation based on crop coefficients of winter wheat cropland in the Guanzhong Plain, China[J]. Agronomy, 2023, 13(12): 2982.

[38] 刘春伟,曾懿婷,邱让建. 用分时段修正双源模型估算南京地区冬小麦生育期蒸散量[J]. 农业工程学报,2016, 32(1): 80 – 87.

LIU Chunwei, ZENG Xieting, QIU Rangjian. Simulated total evapotranspiration of winter wheat with modified Shuttleworth – Wallace model in different stages in Nanjing[J]. Transactions of the CSAE, 2016, 32(1):80 – 87. (in Chinese)