doi:10.6041/j.issn.1000-1298.2026.22.025

基于 PROSAIL 与机器学习的无人机多光谱遥感 LAI 估算模型

张智韬[1,2]　董　宁[1,2]　刘　豪[1,2]　白旭乾[1,2]　陈俊英[1,2]　叶苏蒙[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. 西北农林科技大学旱区农业水土工程教育部重点实验室, 陕西杨凌 712100)

摘要: 叶面积指数(LAI)是表征作物冠层结构与生长状态的关键参数,对农田生态监测及精准农业管理意义重大。为提高复杂环境下 LAI 反演精度与稳定性,本文基于 PROSAIL 模型构建 114 210 组冠层反射率与 LAI 配对样本,利用光谱转换后的模拟数据计算 24 种植被指数,经 Pearson 相关性分析筛选出 8 个高相关指数作为建模特征。选取 KNN、RFR 与 XGBoost 3 种机器学习方法构建 LAI 反演模型,以无人机实测多光谱反射率及对应 LAI 为输入,评估各模型在 4 个土壤含水率梯度(T1 ~ T4)、3 个关键生育期(拔节期、抽穗期、灌浆期)及全含水率、全生育期混合条件下精度与适用性。结果表明:XGBoost 模型表现最优,非线性建模及泛化能力较强;RFR 模型在中高含水率与冠层稳定期反演效果较好;KNN 模型对土壤背景与特征扰动敏感,低含水率及早期生育期表现欠佳,混合条件下仍具一定稳定性。研究结果证实"物理模拟 + 特征筛选 + 机器学习建模"策略在多含水率、多时相农田环境中反演 LAI 的可行性,可为无人机多光谱遥感驱动的作物参数反演提供技术支撑。

关键词: 叶面积指数; PROSAIL 模型; 机器学习; 无人机; 多光谱遥感; 估算模型

中图分类号: S127; TP79　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)22-0295-11　　**OSID**:

Estimating LAI Model from UAV Multispectral Remote Sensing Based on PROSAIL Model and Machine Learning

Zhang Zhitao[1,2]　Dong Ning[1,2]　Liu Hao[1,2]　Bai Xuqian[1,2]　Chen Junying[1,2]　Ye Sumeng[1,2]

(1. *College of Water Resources and Architectural Engineering*, *Northwest A&F University*, *Yangling*, *Shaanxi* 712100, *China*
2. *Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas*, *Ministry of Education*, *Northwest A&F University*, *Yangling*, *Shaanxi* 712100, *China*)

Abstract: Leaf area index (LAI) is a key biophysical parameter that reflects crop canopy structure and growth status, playing an essential role in ecosystem monitoring and precision agriculture. To improve the accuracy and robustness of LAI inversion under complex conditions, it generated 114 210 paired samples of canopy reflectance and LAI using the PROSAIL radiative transfer model. From the simulated spectra, totally 24 vegetation indices were calculated, and eight with high correlation to LAI were selected through Pearson analysis as modeling features. Three machine learning algorithms, i. e., K-nearest neighbors (KNN), random forest regression (RFR), and extreme gradient boosting (XGBoost), were then applied to build LAI inversion models. These models were validated by using UAV-based multispectral reflectance and LAI observations across four soil moisture content (T1 ~ T4), three growth stages (jointing, heading, grain filling), and integrated conditions combining all moisture levels and stages. Results showed that XGBoost achieved the best performance with strong nonlinear modeling and generalization ability. RFR performed well under medium-to-high moisture and stable canopy stages. KNN was more sensitive to soil background and spectral disturbances, with reduced accuracy under low moisture and early growth stages, but retained moderate stability under integrated conditions. Research findings demonstrated the feasibility of integrating physical simulation, feature selection, and machine learning for LAI inversion, providing methodological support for UAV-based multispectral remote sensing of crop parameters.

Key words: leaf area index; PROSAIL model; machine learning; unmanned aerial vehicle (UAV); multispectral remote sensing; estimation model

收稿日期: 2025-07-29　修回日期: 2025-09-02

基金项目: 陕西省科技厅区域科技创新体系建设项目(2025ZY-QYCXYL-06)和国家自然科学基金项目(52479050)

作者简介: 张智韬(1976—),男,教授,博士,主要从事遥感技术在农业水土工程及水文水资源的应用研究,E-mail: zhitaozhang@126.com

0 引言

叶面积指数(Leaf area index, LAI)是反映作物冠层结构、光合作用能力和生物量累积水平的重要生物物理参数,广泛应用于作物生长监测、产量预测、胁迫诊断和精准灌溉调控等农业管理领域[1]。当前 LAI 获取方式主要依赖于人工采样或仪器间接测量,存在操作复杂、效率低、空间代表性差等问题,难以满足区域尺度、非接触式、高频次的农业信息获取需求[2]。

随着卫星遥感技术发展,LAI 估算方法日趋多样化,主要包括基于植被指数(如归一化植被指数(NDVI)、增强型植被指(EVI)、优化土壤调整植被指数(OSAVI)等)的经验模型,通过建立光谱指数与 LAI 之间的经验关系实现反演,计算简便但易受作物类型、土壤背景及观测条件影响[3];基于统计与机器学习的经验-半经验模型(如多元回归、随机森林、支持向量机等),能够刻画复杂的非线性特征关系,精度较高但对大量地面样本依赖性强;以及基于辐射传输模型(如 PROSAIL)的物理方法,可通过正向模拟与反演实现对 LAI 的可解释化估算[4]。然而,受制于卫星平台空间分辨率有限和重访周期较长,这些方法在田间尺度的高频动态监测中仍存在一定不足[5]。近年来,无人机遥感技术的快速发展,使基于多光谱影像的高分辨率作物参数反演成为研究热点[6]。无人机平台以灵活性高、成像精度高与操作周期短等显著优势,已成为田间尺度冠层动态监测的理想工具[7]。多光谱影像在特定波段对叶片色素、水分及结构变化敏感,为 LAI 遥感反演提供了重要基础[8]。然而,作物反射率受生育进程、土壤背景、水分胁迫等因素影响显著,致使 LAI 反演模型在多时空条件下存在精度不稳、泛化性差等问题[9]。

辐射传输模型作为连接植被生理特性与遥感观测的物理工具,为构建可解释的 LAI 反演模型提供了理论支撑[10]。其中,PROSAIL 模型通过耦合叶片尺度的 PROSPECT 模型与冠层尺度的 SAIL 模型,可模拟冠层在不同结构、生化组分及观测条件下反射率响应[11],在农业遥感中已被广泛用于正向模拟与反演建模[12]。然而,当前研究仍面临多方面的局限:①模型对冠层结构与生理特性随生育期演替的动态响应表达能力仍显不足,难以精确刻画不同生育阶段反射率响应的时变特征,从而制约了反演模型在多时相条件下鲁棒性[13]。②土壤水分变化对背景反射的扰动在低覆盖度场景中常被忽视,而条播作物冠层的聚集效应则易导致红光波段模拟偏差[14]。③在实际应用中还存在参数范围设定不确定、耦合关系简化、冠层结构非均匀性及土壤背景干扰等问题,均可能影响模拟反射率的精度与基于其生成数据的反演结果可靠性[15]。

近年来越来越多研究将物理模型模拟数据与机器学习方法相结合,提升模型适应能力与预测精度[16]。机器学习方法(如 K 近邻算法(KNN)、随机森林(RFR)与极端梯度提升(XGBoost)等)在处理非线性特征、复杂背景与高维输入方面具有显著优势,已被应用于多种作物参数反演任务[17]。然而,现有研究大多聚焦于单一环境条件或特定生育阶段建模[18],不同机器学习算法在多水分梯度胁迫、复杂生育期动态背景等综合环境情境下泛化能力、稳定性及适用边界仍缺乏系统性评估与深入解析[19]。

基于此,本研究提出一种融合 PROSAIL 模拟数据、无人机多光谱遥感观测与机器学习方法的 LAI 反演策略。通过构建大规模模拟数据集,并结合特征筛选与多模型性能对比,系统评估不同方法在多环境情境下适用性与精度表现。

1 材料与方法

1.1 试验区概况

研究区位于中国旱区节水农业研究院(图 1a),坐落于陕西省杨凌示范区(34°17′42.17″N,108°4′20″E,海拔 525 m)。该地区是中国西北部典型的旱地农业区,多年平均降水量约 635 mm,主要集中在 7—9 月,属于温带季风性半湿润气候区。试验地土壤类型为中壤土,平均田间持水率约 26%,凋萎系数约 8.6%,土壤干容重约 1.6 g/cm^3,土壤 pH 值约 8.1,有机质含量约 13.3 g/kg,总氮含量约 0.8 g/kg。地下水埋藏较深,向上补给量可忽略不计。

1.2 试验设计

试验设计如图 1b 所示,选用黄土高原地区广泛种植的冬小麦品种“小偃 22”为研究对象,设置 4 种土壤含水率梯度处理:T1((40% ~50%)田间持水率)、T2((50% ~65%)田间持水率)、T3((65% ~80%)田间持水率)和 T4((80% ~95%)田间持水率),分别表示干旱、半干旱、适宜湿度和土壤过湿的土壤湿度水平,以模拟不同水分状况对冬小麦 LAI 的影响,每处理设 3 个重复小区(用 -1、-2、-3 表示),共 12 个试验小区,单个小区面积为 4 m×4 m。为消除边际效应,在各小区间设置宽 1.5 m 保护行,以减少水分渗透干扰。

冬小麦分别于 2022、2023、2024 年 10 月中旬播种,播种深度 4 cm,行距 25 cm,播种量 200 kg/hm^2。

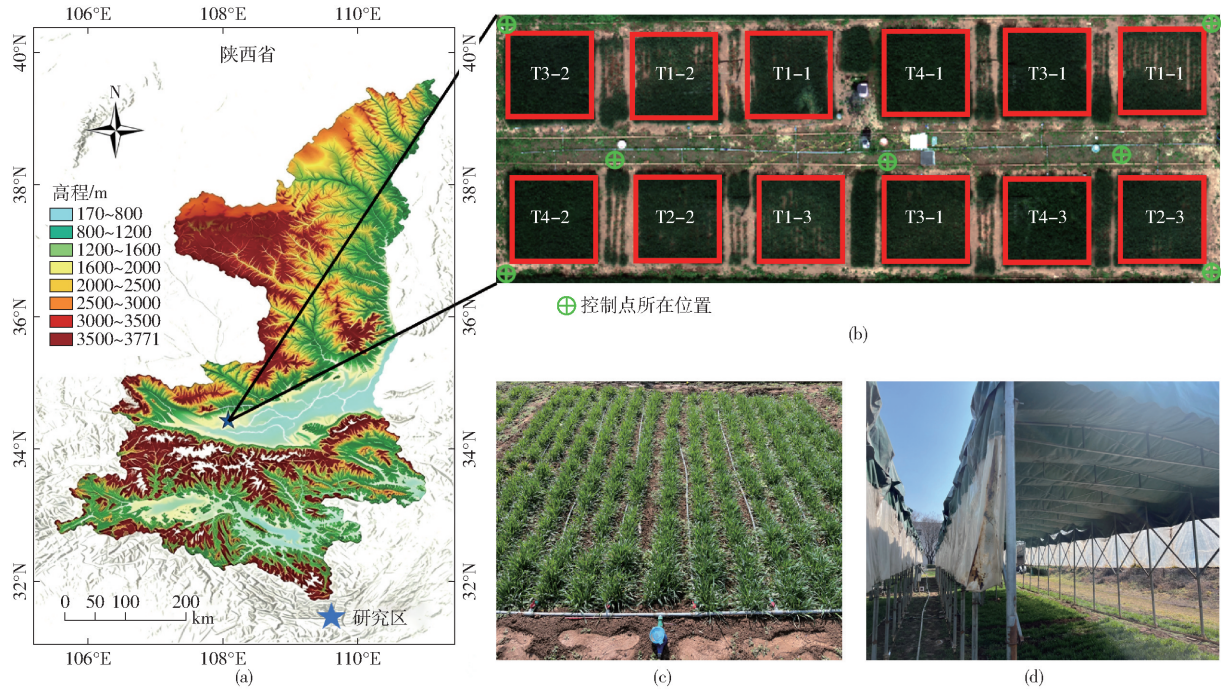


图 1　试验区概况

Fig. 1　Overview of pilot area

为实现精准灌溉管理，采用 0 ~ 60 cm 土层钻探采样[20]结合烘干法监测土壤含水率，每 5 d 采集 1 次。当土壤含水率降至各处理设定下限时启动灌溉，灌溉方式为滴灌（图 1c），灌水量通过小区水表监测控制，平均灌水量如表 1 所示。为避免降雨干扰，试验区设置移动式避雨棚（图 1d）。

表 1　平均灌水量

Tab. 1　Average irrigation water

处理	T1 - 1	T1 - 2	T1 - 3	T2 - 1	T2 - 2	T2 - 3
灌水量/m^3	0	0	0	0. 249	0. 253	0. 247
处理	T3 - 1	T3 - 2	T3 - 3	T4 - 1	T4 - 2	T4 - 3
灌水量/m^3	0. 564	0. 558	0. 566	0. 869	0. 862	0. 854

1.3　多光谱影像采集与处理

多光谱影像采用大疆 Matrice600Pro 型无人机（图 2a）搭载 MS600V2 型多光谱相机（图 2b）获取。该相机包含 6 个波段，中心波长分别为 490、555、680、720、800、900 nm。飞行高度设为 20 m，地面分辨率为 0. 014 m/像素，影像前向与侧向重叠度均为 80%。拍摄时间固定在 12:45—13:15 之间，最大限度减少太阳高度角变化带来的辐射误差。

飞行前布设 7 个地面控制点（图 1b），使用 Hi - TargetiRTK3 型差分 GNSS 设备（图 2c）采集控制点坐标，影像拼接与配准通过 Pix4Dmapper 软件完成。辐射校准采用标准反射率面板（图 2d），在 ENVI 5. 3 软件完成 6 波段影像合成与定标，提取冠层反射率数据。

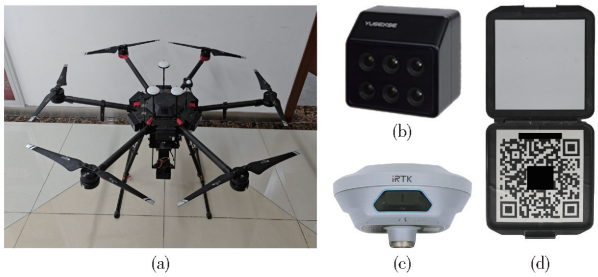


图 2　数据采集设备

Fig. 2　Data acquisition equipment

1.4　叶面积指数数据获取

多光谱影像采集后，在试验小区内随机选取 3 株具有代表性的小麦植株，进行茎叶分离处理后拍照。图像处理通过 Matlab R2022b 实现，利用颜色阈值法（图 3）提取叶片区域，计算单株叶面积。结合田间样本调查获得的单位面积单茎数，计算单个小区 LAI。最终共采集实测样本 348 组，涵盖不同水分梯度与生育阶段。单个小区叶面积指数计算式[21]为

$$LAI = \frac{\rho L_A}{S} \tag{1}$$

式中　LAI——叶面积指数

L_A——单植株叶面积，m^2/株

S——小区面积，m^2

ρ——单个小区植株数，株

1.5　模拟数据构建

采用 PROSAIL 模型生成冠层高光谱反射率数据，模拟波长为 400 ~ 2 500 nm，光谱间隔为 1 nm。



(a) 原值图像 (b) 叶片面积选择

图 3 叶面积提取示意图

Fig. 3 Schematics of leaf area extraction

该模型耦合了叶片光学模型 PROSPECT 与冠层辐射传输模型 SAIL,输入参数涵盖叶面积指数、叶绿素含量、平均叶倾角、干物质含量、等效水厚度、土壤反射率及观测几何等 13 项参数(表 2)。依据文献[9,12,22]与实测范围设置参数取值区间,组合生成 114 210 组模拟样本。

表 2 PROSAIL 模型参数

Tab. 2 PROSAIL model parameters

参数	数值
叶面积指数(LAI)/($m^2 \cdot m^{-2}$)	1 ~ 8
叶绿素含量(Cab)/($\mu g \cdot cm^{-2}$)	30 ~ 70
叶片结构参数(N)	1 ~ 3
平均叶倾角(ALA)/(°)	20 ~ 60
干物质含量(Cm)/($g \cdot cm^{-2}$)	0.005 ~ 0.01
土壤反射率(psoil)	0 ~ 1
类胡萝卜素含量(Car)/($\mu g \cdot cm^{-2}$)	Cab/6
等效水厚度(Cw)/cm	0.2
热点系数(hspot)	0.3
相对方位角(psi)/(°)	0
棕色物质含量(Cbrown)	0
太阳天顶角(tts)/(°)	0
观测天顶角(tto)/(°)	0

1.6 光谱转换

由于无人机获取的是窄波段多光谱影像,而 PROSAIL 模型模拟的冠层光谱反射率是在 400 ~ 2 500 nm 内以 1 nm 为间隔的连续高光谱数据,二者存在明显的尺度效应,需将模拟高光谱反射率依据 MS600V2 型多光谱相机响应函数(图 4)进行波段匹配。转换方法采用加权积分形式,转换公式为

$$R = \frac{\int_{\lambda_{min}}^{\lambda_{max}} f_i(\lambda) R_{PROSAIL} d\lambda}{\int_{\lambda_{min}}^{\lambda_{max}} f_i(\lambda) d\lambda} \quad (2)$$

式中 R——模拟 MS600V2 型多光谱相机的波段反射率

λ_{max}——不同波段最大波长

λ_{min}——不同波段最小波长

$R_{PROSAIL}$——PROSAIL 模型模拟的连续高光谱反射率

$f_i(\lambda)$——MS600V2 型多光谱相机光谱响应函数

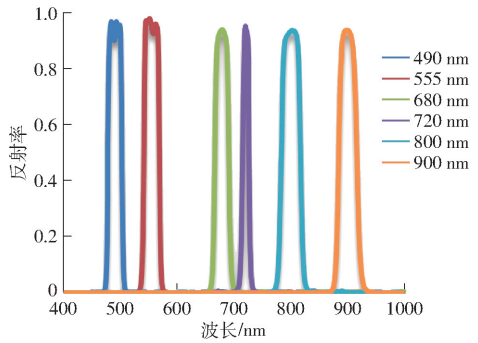


图 4 MS600V2 型多光谱相机光谱响应函数曲线

Fig. 4 Spectral response function curves of MS600V2 multispectral camera

1.7 植被指数构建与筛选

基于处理后的 6 个波段反射率,计算 24 种常见植被指数(表 3),涵盖归一化类(NDVI、GNDVI)、差值类(DVI、GWDVI)、土壤调节类(SAVI、MSAVI)等不同类型,这些指数已被广泛验证适用于植被生长监测和 LAI 估算研究[23]。考虑到 MS600V2 型多光谱相机包含 2 个近红外波段(800、900 nm),所有指数均构建 NIR1 与 NIR2 2 种形式,以增强建模特征的丰富性和适应性。

通过皮尔逊相关性分析方法(Pearson correlation analysis),筛选出与 LAI 相关性系数高于 0.7 的植被指数,用于后续模型训练输入变量。

1.8 模型构建与精度评价

1.8.1 模型构建

为实现冬小麦叶面积指数(LAI)高精度反演,选取 K 近邻回归(K-nearest neighbors, KNN)、随机森林回归(Random forest regression, RFR)和极端梯度提升回归(Extreme gradient boosting, XGBoost)3 种常用的机器学习算法进行建模。以 PROSAIL 模拟样本的 70% 作为训练集、30% 作为验证集,建模前对所有特征进行标准化处理,训练过程采用 Optuna 框架进行超参数搜索,并重复运行 3 次取平均值,以提高精度与稳定性。

(1)K 近邻回归(KNN)

KNN 属于基于实例的非参数回归方法,通过在特征空间内计算待预测样本与训练样本距离,选取最近的 k 个邻居样本,根据输出值加权平均作为预测结果。采用闵可夫斯基距离($p=2$,即欧氏距离)作为度量方式。

核心参数包括邻居数 k、距离权重方式(均匀权重或距离倒数权重)及搜索算法(如 ball-tree 或 kd-tree)。通过 Optuna 在训练集上利用交叉验证确定最优参数组合,并在建模前对特征进行 Z-score 标准化以减弱尺度差异的影响。

表 3　植被指数计算公式

Tab. 3　Vegetation index calculation formulas

植被指数	计算公式	文献序号
归一化植被指数(NDVI)	$NDVI=(NIR-Red)/(NIR+Red)$	[24]
增强植被指数(EVI2)	$EVI_2=2.5(NIR-Red)/(NIR+2.4Red+1)$	[25]
增强植被指数(EVI)	$EVI=2.5(NIR-Red)/(NIR+6Red-7.5Blue+1)$	[26]
叶绿素植被指数(GCVI)	$GCVI=NIR/Green-1$	[27]
土壤调节植被指数(SAVI)	$SAVI=(NIR-Red)(1+L)/(NIR+Red+L)$	[28]
改进型土壤调节植被指数(MSAVI)	$MSAVI=[2NIR+1-\sqrt{(2NIR+1)^2-8(NIR-Red)}]/2$	[29]
差值植被指数(DVI)	$DVI=NIR-Red$	[30]
冠层植被指数(CVI)	$CVI=NIR/Red$	[31]
优化型土壤调节植被指数(OSAVI)	$OSAVI=(NIR-Red)/(NIR+Red+0.16)$	[32]
绿度归一化植被指数(GNDVI)	$GNDVI=(NIR-Green)/(NIR+Green)$	[33]
广动态植被指数(WDRVI)	$WDRVI=(\alpha NIR-Red)/(\alpha NIR+Red)$	[34]
绿色广动态植被指数(GWDRVI)	$GWDRVI=(\alpha NIR-Green)/(\alpha NIR+Green)$	[34]

注：*Blue*、*Green*、*Red*、*NIR* 分别表示蓝、绿、红和近红外波段光谱反射率。L 表示土壤亮度调节系数，取 0.5；α 表示 NIR 波段加权系数，取 0.1。

(2)随机森林回归(RFR)

RFR 是一种基于 Bagging (Bootstrap aggregating)的集成学习方法，通过构建多棵决策树并对预测结果取平均值降低方差、提升泛化能力。每棵树在训练时从原始数据中进行自助采样(Bootstrap)，并在节点分裂时随机选择部分特征参与划分。

核心参数包括树数量(n_estimators)、最大树深度(max_depth)、节点最小分裂样本数(min_samples_split)、叶节点最小样本数(min_samples_leaf)及最大特征数(max_features)。利用 Optuna 在训练集上自动搜索最优参数组合，以在保证精度的同时控制模型复杂度。

(3)极端梯度提升回归(XGBoost)

XGBoost 基于梯度提升(Gradient boosting)思想，采用加法模型迭代拟合残差，通过二阶梯度信息实现快速收敛，并引入正则化项以抑制过拟合。

核心参数包括学习率(eta)、最大树深度(max_depth)、最小叶节点权重(min_child_weight)、子采样比例(subsample)、特征采样比例(colsample_bytree)及正则化系数(lambda、alpha)。通过 Optuna 在训练集上搜索最优参数，并采用早停策略避免过拟合。

1.8.2　精度评价

模型性能采用决定系数(R^2)、均方根误差(RMSE)、平均绝对误差(MAE)和归一化均方根误差(NRMSE)4 项指标进行评价。其中，R^2 反映模型拟合程度，RMSE 衡量整体误差水平，MAE 体现平均预测偏差，NRMSE 能有效消除量纲影响，便于不同情景下误差的横向比较，四者综合反映模型反演能力。

此外，为进一步评估预测结果的统计显著性，采用配对 t 检验计算预测值与实测值之间的 P 值，并在拟合图中进行显著性标注($P<0.05$ 标记为 *，$P<0.01$ 标记为 **，$P\geqslant 0.05$ 记为 ns)。

2　结果与分析

2.1　基于 Pearson 相关性分析方法的植被指数筛选

Pearson 相关性分析结果(表 4)表明，多数归一化类和土壤调节类植被指数与 LAI 具有较强正相关性，尤其是 GWDVI、GNDVI、GVI 和 GCVI，表现出显著正相关性。为确保输入机器学习模型的特征对 LAI 有较强的表征能力，选取 Pearson 相关系数大于 0.7 植被指数作为模型输入特征，用于后续的 LAI 反演建模，分别为 GWDRVI_NIR2、GCVI_NIR2、GWDRVI_NIR1、GCVI_NIR1、GNDVI_NIR2、GNDVI_NIR1、MSAVI_NIR1、EVI2_NIR1。

表 4　Pearson 相关性分析结果

Tab. 4　Pearson correlation analysis results

植被指数	Pearson 相关系数	植被指数	Pearson 相关系数
GWDRVI_NIR2	0.781	OSAVI_NIR1	0.692
GCVI_NIR2	0.775	CVI_NIR1	0.691
GWDRVI_NIR1	0.770	DVI_NIR1	0.691
GCVI_NIR1	0.766	OSAVI_NIR2	0.690
GNDVI_NIR2	0.765	MSAVI_NIR2	0.690
GNDVI_NIR1	0.751	SAVI_NIR2	0.682
MSAVI_NIR1	0.705	EVI2_NIR2	0.680
EVI2_NIR1	0.701	EVI_NIR1	0.675
SAVI_NIR1	0.699	NDVI_NIR2	0.667
WDRVI_NIR2	0.698	NDVI_NIR1	0.659
CVI_NIR2	0.696	EVI_NIR2	0.651
WDRVI_NIR1	0.694	DVI_NIR2	0.646

2.2　不同土壤含水率下 LAI 反演效果

在不同土壤含水率(T1 ~ T4)条件下，采用

KNN、RFR 和 XGBoost 模型反演 LAI，绘制实测值与反演值散点拟合图（图 5），并结合拔节期、抽穗期、灌浆期分析同一水分条件下模型对冬小麦冠层变化的响应。

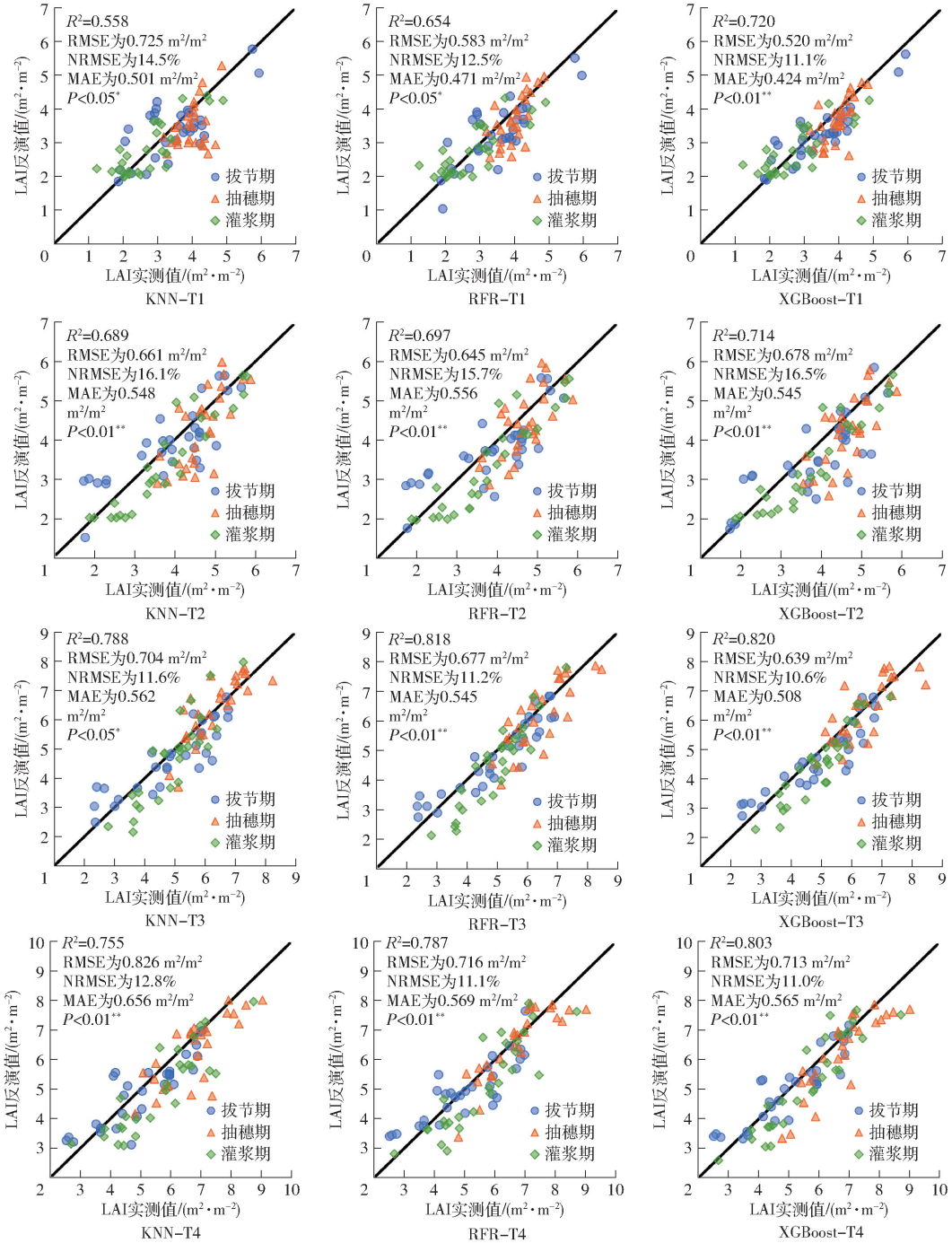


图 5 不同土壤含水率下 KNN、RFR、XGBoost 模型的 LAI 实测值与反演值

Fig. 5 Measured and inversion LAI values using KNN, RFR, and XGBoost models under different soil moisture gradients

由图 5 可知，T1～T4 区间，随含水率提升，植被覆盖增加，土壤背景反射干扰减弱，3 种模型 R^2 和精度指标整体提升，NRMSE 整体呈下降趋势，P 值在多数情境下均小于 0.01，表明反演值与实测值的差异具有显著性。低含水率（T1、T2）时，各模型反演点分布离散，高估或低估现象较多，R^2 低，反演精度不佳。T3 为接近最优水分状态，模型精度达峰值，KNN、RFR、XGBoost R^2 分别为 0.778、0.818 和 0.820，NRMSE 分别为 11.6%、11.2% 和 10.6%，反演点高度集中于 1∶1 线附近，P 值显著。高水分（T4）条件下，虽植被生长良好、LAI 高且土壤干扰弱，但冠层结构复杂性上升（叶片重叠、反射饱和）使反射率与 LAI 非线性增强，反演精度略降，KNN、RFR、XGBoost R^2 分别下降 0.033、0.031 和 0.017，XGBoost 下降幅度最小，鲁棒性更强。同一含水率梯度下，随着生育期从拔节期推进至抽穗期、灌浆

期，植被覆盖度提高，土壤干扰降低。点云分布显示，拔节期点分布离散，抽穗期和灌浆期点较为聚集，反演精度显著提高。XGBoost 在同一含水率梯度不同生育期的表现依旧优于 KNN 和 RFR，有更多点聚集在 1∶1线附近。

综上，XGBoost 在不同含水率条件下均表现最优，反演能力稳定。同一含水率梯度下，XGBoost 对不同生育期的反演能力也优于另 2 种模型，表明对复杂冠层的处理能力优于 KNN 和 RFR。

2.3　不同生育期 LAI 反演效果

为评估不同生育期 LAI 反演模型适用性，选取拔节期、抽穗期、灌浆期 3 个关键阶段，采用 KNN、RFR 和 XGBoost 模型反演 LAI，绘制实测值与反演值散点拟合图（图 6），结合土壤含水率梯度（T1 ~ T4）分析模型在同一生育期对含水率变化的响应。

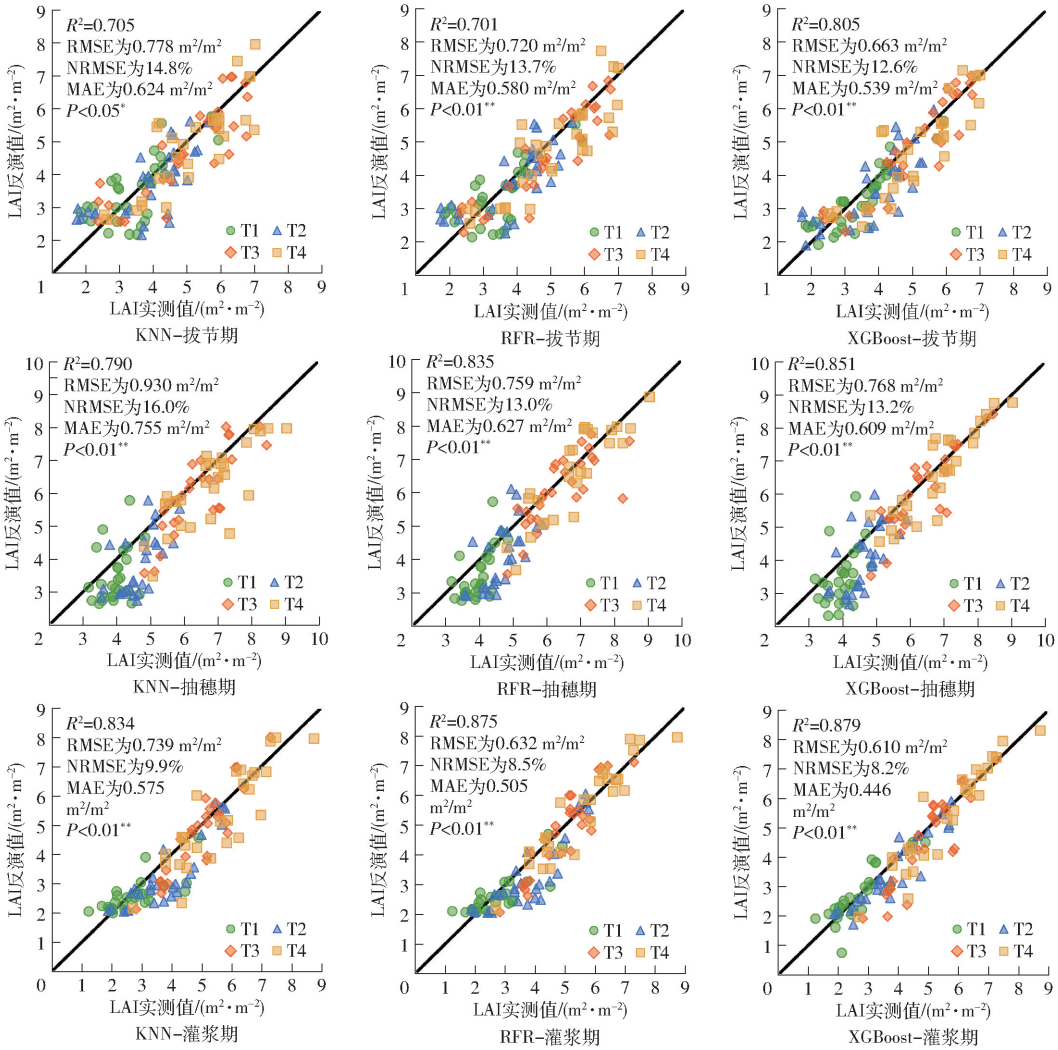


图 6　不同生育期 KNN、RFR、XGBoost 模型的 LAI 实测值与反演值

Fig. 6　Measured and inversion LAI values for KNN, RFR, and XGBoost models at different growth stages

由图 6 可知，随生育期推进，3 种模型 R^2 提升、NRMSE 整体呈下降趋势，P 值在绝大多数情境下小于 0.01。拔节期植被冠层未闭合，土壤背景干扰强，模型反演误差大：KNN 和 RFR 反演点较为离散；XGBoost 反演点相对集中，精度显著优于其他模型。抽穗期植被覆盖增强、冠层稳定，模型性能明显改善：XGBoost 拟合最佳，点云密集于 1∶1线周围；RFR 整体稳定，误差小；KNN 虽有提升，但低 LAI 时偏差仍较大。灌浆期冠层略有衰退但 LAI 仍较高，模型反演精度总体提升：XGBoost 继续最优，R^2 达到 0.879（NRMSE 为 8.2%），P 极显著，点云聚集在 1∶1 线附近；RFR 次之；KNN 反演点略显离散。同一生育期不同含水率梯度影响显示：低含水率（T1）时反演误差普遍偏大，对 KNN 影响最显著；适宜含水率（T2、T3）时，反演点分布集中，模型性能显著提升。XGBoost 对土壤含水率变化不敏感，在不同生育期和含水率条件下保持稳定高精度，体现强适应性与鲁棒性。

综上，XGBoost 在各生育期表现最优，尤其拔节期和干旱条件下鲁棒性强。同一生育期内，XGBoost 对不同含水率梯度的反演能力亦优于另 2 种模型，表明受土壤背景的影响小于 KNN 和 RFR。

2.4 全含水率、全生育期混合条件下 LAI 反演效果

为评估 3 种模型在多含水率梯度与多生育期混合条件下整体反演能力，整合所有实测数据（涵盖 T1～T4 4 个土壤含水率水平及拔节期、抽穗期、灌浆期 3 个典型生育期），采用 KNN、RFR 和 XGBoost 模型反演 LAI，绘制实测值与反演值散点拟合图。

由图 7 可知，在混合条件下，XGBoost 反演点分布最集中于 1∶1 线附近，R^2 为 0.793，NRMSE 为 10.8%，$P<0.01$，表明反演值与实测值差异显著；RMSE 与 MAE 分别为 0.756、0.629 m^2/m^2，显示出较强的泛化能力。KNN 在混合条件下表现良好，R^2 达 0.781，NRMSE 为 11.0%，RMSE 与 MAE 分别为 0.771、0.628 m^2/m^2，3 项指标均优于 RFR，且 P 值极显著。RFR 精度相对较低（$R^2=0.746$，NRMSE 为 12.1%，RMSE 为 0.847 m^2/m^2，MAE 为 0.708 m^2/m^2），虽整体稳定，但在多源干扰下易出现误差波动。

综上，在全含水率、全生育期混合条件下，XGBoost 模型仍表现最优，预测精度与稳定性均领先；KNN 次之，整体性能优于 RFR，适用于中等复杂环境；RFR 精度相对较低，但仍具一定应用价值。

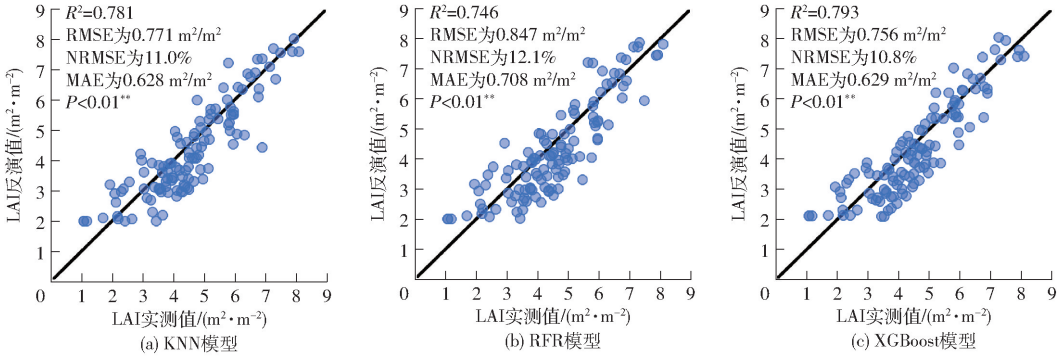


图 7 全含水率、全生育期混合条件下 KNN、RFR、XGBoost 模型 LAI 反演值与实测值

Fig. 7 LAI inversion values and measured values for KNN, RFR, and XGBoost models under mixed conditions of full moisture and full growth period

2.5 反演图

为直观呈现不同生育期小麦冠层 LAI 变化及同一生育期不同土壤含水率条件下 LAI 差异，选取拔节期、抽穗期和灌浆期的多光谱影像，基于最优模型 XGBoost 反演结果绘制 LAI 反演图，结果如图 8 所示。

图 8a 为拔节期 LAI 反演图，此时小麦处于早期生长阶段，冠层尚未完全闭合，土壤背景影响较明显。各小区 LAI 整体较低，空间异质性较弱，不同土壤含水率条件下小麦生长的初始差异较小。

图 8b 为抽穗期 LAI 反演图，随生育期推进，高水处理小区小麦冠层覆盖度显著提高，LAI 较拔节期大幅增加；低水处理小区冠层覆盖度提升较小，LAI 增幅有限。不同小区间 LAI 差异更清晰，体现土壤含水率条件对小麦生长的影响在抽穗期进一步显现。

图 8c 为灌浆期 LAI 反演图，此时小麦生长趋于成熟，LAI 达较高水平。各小区 LAI 空间分布与抽穗期存在差异：高水处理小区覆盖度趋于完全闭合，LAI 较抽穗期进一步增大；低水处理小区 LAI 较抽穗期无明显变化。

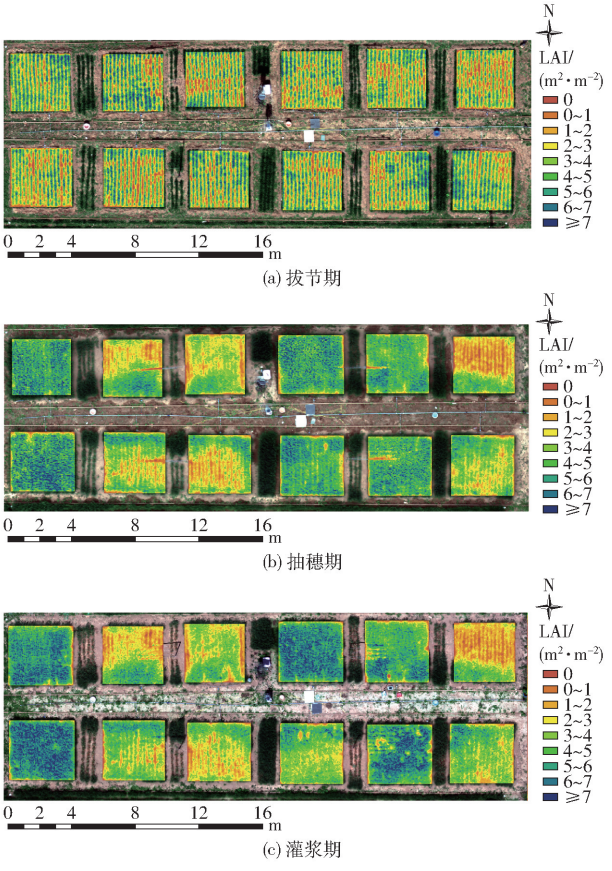
对比 3 幅反演图，结果均符合冬小麦各生育期生长特性，可清晰观察到小麦冠层 LAI 随生育期的动态变化及不同土壤含水率条件下空间差异，为理解小麦生长与土壤水分的关系提供了直观可视化依据。

3 讨论

3.1 PROSAIL 模型模拟数据适用性与局限性

PROSAIL 模型通过耦合叶片尺度的 PROSPECT 与冠层尺度的 SAIL 模型，为构建冠层反射率模拟样本提供了物理机制支撑[35]。本研究基于参数组合生成 114 210 组模拟数据，并通过光谱响应函数转换匹配无人机搭载的多光谱波段，作为机器学习建模的输入数据。在生育期明确、水分条件适中情境下，该模拟数据在各类模型中表现出较好的训练稳定性和泛化能力，说明 PROSAIL 模型可为 LAI 反演提供有效数据支持。然而，模型在实际应用中仍面临部分限制。

首先，参数设置对模拟数据的代表性与反演结果存在明显影响[35]。本研究采用文献[9,12,22]的参数取值范围，并结合实际建模需求固定部分参数（如 Car 为 Cab/6、Cbrown 为 0、psi 为 0°等），提升了建模效率。但模型参数间存在一定耦合性，如 Cab 与 Cm 的协同影响在实际冠层中具有复杂生理基础，固定简化设置虽便于模拟但可能掩盖生育期或



(a) 拔节期

(b) 抽穗期

(c) 灌浆期

图 8　基于最优反演模型的研究区冬小麦各生育期 LAI 反演图

Fig. 8　LAI inversion map of winter wheat in different growth stages in study area based on optimal inversion model

极端情境下真实光谱变化，降低模型对不典型样本的适应性。

其次，固定参数可能限制光谱表达的灵活性[36]。如 Cbrown、psi、tts 和 tto 等观测几何参数虽在本研究中取常量以简化建模，但在实际应用中可能随观测时间、角度及植被衰老程度变化。静态设定虽有助于减少变量间耦合与计算开销，但在多时相、强光照变化条件下可能引入额外误差。

此外，模型对土壤背景反射的描述能力有限。虽 psoil 设为动态参数，用于模拟不同含水率条件下背景反射变化，但未引入实测土壤光谱信息，导致在 T1（干旱）与 T4（高湿）含水率梯度条件下，模拟土壤反射率的动态响应不足，降低了数据在低植被覆盖条件下真实性。土壤背景反射误差是影响拔节期模型性能的重要因素，需在未来引入实测土壤光谱曲线或构建土壤-水分动态函数加以改进[37]。

综上，PROSAIL 模型在构建统一、可控的模拟样本方面表现良好，尤其适用于中等含水率与结构稳定情境。但在土壤含水率变异显著、生育期跨度较大或冠层结构动态性强的场景下，仍需进一步提升对实际变化的表达能力。未来可结合实测土壤光谱数据、引入时序参数或构建动态结构函数，以提高模拟数据在复杂环境下适应性和反演建模价值。

3.2　机器学习模型在 LAI 反演中的响应机制与差异性

本研究基于模拟样本与筛选的 8 个植被指数特征，构建 XGBoost、RFR 和 KNN 3 种 LAI 反演模型，系统评估在不同土壤含水率与生育期条件下性能差异。

XGBoost 在多数含水率和生育期条件下表现最优，拟合能力与稳定性突出。拔节期、抽穗期、灌浆期 R^2 均超 0.8，灌浆期达 0.879，混合样本下 R^2 仍保持 0.793。其优势源于梯度提升串行建模策略，可逐步拟合误差残差，有效捕捉 LAI 与植被指数间的非线性关系及特征交互效应，通过特征重要性排序与增益计算自适应选择最优特征组合[38]。同时，该模型在样本不均衡或含噪声时鲁棒性较强，故在低含水率（T1）条件下仍能保持较高精度，特别适用于特征耦合强、光谱背景复杂、时序差异大的 LAI 反演场景。

RFR 模型在多数情境下适用性良好，尤其在适中含水率（T3）与灌浆期等冠层稳定阶段，反演精度接近 XGBoost，R^2 超 0.81；但全含水率-全生育期混合样本下表现略逊，R^2 降至 0.746。这与随机特征选择及样本抽样机制相关：虽能建模变量非线性关系，但树结构并行训练且单树仅处理部分特征与样本[39]，对主导特征变化的敏感性低于 XGBoost，在生育期导致 LAI 分布差异显著时，难以充分提取子空间最优判别特征，泛化能力下降；极端含水率（T1、T4）条件下，对土壤背景干扰的适应性较弱，易出现预测波动。

KNN 模型整体精度最低，低含水率（T1）条件下 R^2 仅 0.558。其依赖欧氏距离度量样本相似性，通过局部邻域回归预测，对样本分布均匀性与特征尺度高度敏感[40]。虽经特征标准化与 k 值优化，但样本密度不足或特征差异大时，反演误差易放大。拔节期冠层未闭合、土壤干扰强，样本局部分布特征不显著，难以找到“真实相似”邻域点，导致局部拟合失效；且因不具备外推能力，灌浆期 LAI 较大或结构剧变时存在低估或偏移，适用性不足。

综上，XGBoost 凭借结构优势、强非线性建模能力及特征交互学习机制，在多变量耦合、背景复杂、时序差异大的遥感环境中表现最稳定；RFR 结构简洁、计算高效，适用于变量关系明确、样本分布均衡场景，但混合样本与边界样本学习能力不足；KNN 实现简单，适合样本密集、特征差异小的局部建模，

泛化能力弱且对局部样本敏感,难以胜任复杂农田 LAI 估算。未来可考虑集成优化或多模型融合,以提升复杂情境下反演精度与稳定性。

4 结论

(1)在不同含水率梯度条件下,XGBoost 模型表现最优,T3 含水率下 R^2 最高达 0.820,且在 T1 低含水率条件下仍能保持较高精度;RFR 在中高含水率(T3、T4)条件下表现良好,但在低含水率下误差显著增大;KNN 在 T1 条件下表现最差,R^2 仅为 0.558,说明对背景扰动敏感,稳定性不足。

(2)在不同生育期条件下,3 种模型精度均随生育期推进而提升。XGBoost 在拔节、抽穗和灌浆期均具较高精度,灌浆期 R^2 达 0.879;RFR 在灌浆期精度提升明显;KNN 在拔节期误差较大,对冠层结构变化适应性差。整体来看,冠层结构稳定有助于模型反演精度提升。

(3)在混合建模情境下,3 种模型精度均略有下降。XGBoost 依然保持最优表现,R^2 达 0.793,泛化能力较强;KNN 在混合样本中反演稳定性优于 RFR,而 RFR 精度下降更明显,表现出一定的建模局限性。

(4)XGBoost 适用于多含水率、多生育期及复杂环境下 LAI 反演任务,RFR 适合特征明确、样本均衡的条件,KNN 在混合情境下具备一定稳定性但对特征敏感。验证了“物理建模 + 特征筛选 + 机器学习建模”的有效融合,为农田遥感监测与精准农业提供了技术支撑。

参 考 文 献

[1] 于海琳,兰玉彬,李京谦,等. 基于无人机遥感数据和机器学习的向日葵 LAI 反演[J]. 农业机械学报,2025,56(1):356 - 365.
Yu Hailin, Lan Yubin, Li Jingqian, et al. Sunflower LAI inversion based on unmanned aerial vehicle remote sensing data and machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(1): 356 - 365. (in Chinese)

[2] Fang H, Baret F, Plummer S, et al. An overview of global leaf area index (LAI): methods, products, validation, and applications[J]. Reviews of Geophysics, 2019, 57(3): 739 - 799.

[3] 俞徐苗. 基于 PROSAIL 辐射传输模型的城市灌木滞尘分布研究[D]. 上海:上海师范大学,2022.
Yu Xumiao. Estimation of the distribution of dust deposition in urban shrubs based on the PROSAIL radiative transfer model [D]. Shanghai: Shanghai Normal University, 2022. (in Chinese)

[4] 李晓宇. 基于遥感地面试验的 FPAR 反演与验证[D]. 呼和浩特:内蒙古师范大学,2015.
Li Xiaoyu. Inversion and validation of FPAR based on remote sensing data and field sampling[D]. Hohhot: Inner Mongolia Normal University, 2015. (in Chinese)

[5] 史舟,梁宗正,杨媛媛,等. 农业遥感研究现状与展望[J]. 农业机械学报,2015,46(2):247 - 260.
Shi Zhou, Liang Zongzheng, Yang Yuanyuan, et al. Status and prospect of agricultural remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(2): 247 - 260. (in Chinese)

[6] 谭丞轩,张智韬,许崇豪,等. 无人机多光谱遥感反演各生育期玉米根域土壤含水率[J]. 农业工程学报,2020,36(10):63 - 74.
Tan Chengxuan, Zhang Zhitao, Xu Chonghao, et al. Soil water content inversion model in field maize root zone based on UAV multispectral remote sensing[J]. Transactions of the CSAE, 2020, 36(10): 63 - 74. (in Chinese)

[7] Peng X, Han W, Ao J, et al. Assimilation of LAI derived from UAV multispectral data into the SAFY model to estimate maize yield[J]. Remote Sensing, 2021, 13(6): 1094.

[8] Guo A, Ye H, Huang W, et al. Inversion of maize leaf area index from UAV hyperspectral and multispectral imagery[J]. Computers and Electronics in Agriculture, 2023, 212: 108020.

[9] Atzberger C. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs[J]. Remote Sensing, 2013, 5(2): 949 - 981.

[10] Li H, Liu G, Liu Q, et al. Retrieval of winter wheat leaf area index from Chinese GF - 1 satellite data using the PROSAIL model[J]. Sensors, 2018, 18(4): 1120.

[11] Jacquemoud S, Verhoef W, Baret F, et al. PROSPECT + SAIL models: a review of use for vegetation characterization[J]. Remote Sensing of Environment, 2009, 113: 56 - 66.

[12] 汪彦龙,王钧,李广,等. 采用机器学习优化 PROSAIL 模型的青贮玉米叶面积指数反演[J]. 农业工程学报,2025,41(9):134 - 142.
Wang Yanlong, Wang Jun, Li Guang, et al. Inversion of silage maize leaf area index based on machine learning optimized PROSAIL model[J]. Transactions of the CSAE, 2025, 41(9): 134 - 142. (in Chinese)

[13] Darvishzadeh R, Skidmore A, Schlerf M, et al. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland[J]. Remote Sensing of Environment, 2008, 112(5): 2592 - 2604.

[14] Chaabouni S, Kallel A, Houborg R. Improving retrieval of crop biophysical properties in dryland areas using a multi-scale variational RTM inversion approach[J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 94: 102220.

[15] 王枭轩,卢小平,杨泽楠,等. 基于 PROSAIL 结合 VMG 模型的冬小麦叶面积指数反演方法[J]. 农业机械学报,2022,53(6):209 - 216.
Wang Xiaoxuan, Lu Xiaoping, Yang Zenan, et al. Retrieving method for leaf area index of winter wheat by combining PROSAIL model with VMG model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(6): 209 - 216. (in Chinese)

[16] Duan X, Maqsoom A, Khalil U, et al. Enhancing soil moisture retrieval in semi-arid regions using machine learning algorithms and remote sensing data[J]. Applied Soil Ecology, 2024, 204: 105687.
[17] 张智韬,魏广飞,姚志华,等. 基于无人机多光谱遥感的土壤含盐量反演模型研究[J]. 农业机械学报,2019,50(12):151 - 160.
Zhang Zhitiao, Wei Guangfei, Yao Zhihua, et al. Soil salt inversion model based on UVA multispectral remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 151 - 160. (in Chinese)
[18] Jin X, Kumar L, Li Z, et al. A review of data assimilation of remote sensing and crop models[J]. European Journal of Agronomy, 2018, 92: 141 - 152.
[19] Berger K, Verrelst J, Féret J B, et al. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method [J]. International Journal of Applied Earth Observation and Geoinformation, 2020, 92: 102174.
[20] Yang N, Zhang Z, Yang X, et al. Evaluation of crop water status using UAV-based images data with a model updating strategy [J]. Agricultural Water Management, 2025, 312: 109445.
[21] 杨宁. 表型参数对无人机遥感诊断冬小麦水分状况的影响及模型研究[D]. 杨凌:西北农林科技大学,2025.
Yang Ning. The impact of phenotypic parameters on UAV remote sensing for diagnosing winter wheat water status and model research[D]. Yangling: Northwest A&F University,2025. (in Chinese)
[22] Campos-Taberner M, García-Haro F J, Camps-Valls G, et al. Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring[J]. Remote Sensing of Environment, 2016, 187: 102 - 118.
[23] 付新阳,崔利华,董雨昕,等. 去除土壤背景影响的多光谱遥感影像玉米叶面积指数估算[J]. 农业机械学报,2025,56(5):384 - 394.
Fu Xinyang, Cui Lihua, Dong Yuxin, et al. Estimation of maize leaf area index from multi-spectral remote sensing with soil background effects removed[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(5): 384 - 394. (in Chinese)
[24] Blackburn G A. Quantifying chlorophylls and caroteniods at leaf and canopy scales: an evaluation of some hyperspectral approaches[J]. Remote Sensing of Environment, 1998, 66(3): 273 - 285.
[25] Jiang Z, Huete A R, Didan K, et al. Development of a two-band enhanced vegetation index without a blue band[J]. Remote Sensing of Environment, 2008, 112(10): 3833 - 3845.
[26] Huete A, Didan K, Miura T, et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices [J]. Remote Sensing of Environment, 2002, 83(1 - 2): 195 - 213.
[27] Zhang L, Zhang Z, Luo Y, et al. Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning[J]. Agricultural and Forest Meteorology, 2021, 311: 108666.
[28] Ren H, Zhou G, Zhang F. Using negative soil adjustment factor in soil-adjusted vegetation index (SAVI) for aboveground living biomass estimation in arid grasslands[J]. Remote Sensing of Environment, 2018, 209: 439 - 445.
[29] 周鹏. 基于 Albedo - MSAVI 特征空间的奈曼旗沙漠化遥感监测[J]. 科学技术创新,2021(32):78 - 81.
Zhou Peng. Remote sensing monitoring of desertification in the Naimanqi Desert based on the Albedo - MSAVI feature space [J]. Scientific and Technological Innovation, 2021(32): 78 - 81. (in Chinese)
[30] Xu L Y, Xie X D, Li S. Correlation analysis of the urban heat island effect and the spatial and temporal distribution of atmospheric particulates using TM images in Beijing[J]. Environmental Pollution, 2013, 178: 102 - 114.
[31] Clemente A A, Maciel G M, Siquieroli A C S, et al. High-throughput phenotyping to detect anthocyanins, chlorophylls, and carotenoids in red lettuce germplasm[J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 103: 102533.
[32] Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices[J]. Remote Sensing of Environment, 1996, 55(2): 95 - 107.
[33] 任枫荻. 作物涝害光谱监测方法研究[D]. 长春:吉林大学,2021.
Ren Fengdi. Research on spectral monitoring method of crop waterlogging[D]. Changchun: Jilin University, 2021. (in Chinese)
[34] Gitelson A A. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation[J]. Journal of Plant Physiology, 2004, 161(2): 165 - 173.
[35] 李卓琳,袁金国,杨紫妍,等. 基于 PROSAIL 模型和 Sentinel - 2A 影像的夏玉米叶绿素含量反演比较研究[J]. 遥感技术与应用,2025,40(3):621 - 635.
Li Zhuolin, Yuan Jinguo, Yang Ziyan, et al. Comparative study on chlorophyll content inversion of summer maize based on PROSAIL model and Sentinel - 2A image[J]. Remote Sensing Technology and Application, 2025, 40(3): 621 - 635. (in Chinese)
[36] Jiao Q, Sun Q, Zhang B, et al. A random forest algorithm for retrieving canopy chlorophyll content of wheat and soybean trained with PROSAIL simulations using adjusted average leaf angle[J]. Remote Sensing, 2021, 14(1): 98.
[37] Quan X, He B, Yebra M, et al. Retrieval of forest fuel moisture content using a coupled radiative transfer model[J]. Environmental Modelling & Software, 2017, 95: 290 - 302.
[38] Sudu B, Rong G, Guga S, et al. Retrieving SPAD values of summer maize using UAV hyperspectral data based on multiple machine learning algorithm[J]. Remote Sensing, 2022, 14(21): 5407.
[39] Nile W, Rina S, Mula N, et al. Inversion of leaf chlorophyll content in different growth periods of maize based on multi-source data from "Sky-Space-Ground"[J]. Remote Sensing, 2025, 17(4): 572.
[40] Zhang J, Zhang W, Xiong S, et al. Comparison of new hyperspectral index and machine learning models for prediction of winter wheat leaf water content[J]. Plant Methods, 2021, 17(1): 34.