

多源遥感与气象数据融合的油菜单产和总产预测

陈江^{1,2} 廖庆喜^{1,2} 许贝贝^{3,4} 廖宜涛^{1,2} 万星宇^{1,2}

(1. 华中农业大学工学院, 武汉 430070; 2. 农业农村部长江中下游农业装备重点实验室, 武汉 430070;
3. 养分资源高效利用全国重点实验室, 北京 100193; 4. 中国农业大学资源与环境学院, 北京 100193)

摘要: 油菜是全球重要的油料作物之一,其单产与总产的准确预测对粮油安全评估、农业生产管理及生物能源潜力分析具有重要意义。针对油菜单产与总产预测中样本规模受限、总产建模和多源信息协同效应缺乏系统评估的问题,本文收集县域油菜单产和总产数据,综合利用多源遥感与气象大数据,构建油菜单产与总产的协同预测框架。在此基础上,引入6种树结构机器学习模型,系统评估不同模型及不同数据源组合在油菜单产与总产预测中的性能差异。结果表明,遥感与气象数据联合输入在单产和总产预测中均优于单一遥感或气象数据源,且这一优势在不同树模型结构下表现出良好的一致性与稳定性。不同数据源组合在所有模型中的预测性能遵循一致排序:即联合输入表现最佳,其次为仅遥感输入,最低为仅气象输入;说明遥感与气象数据在生长状态与环境驱动信息上的互补性,是提高油菜单产和总产预测精度的重要基础。单产预测中,CatBoost表现最佳(R^2 为0.679、RMSE为0.269 t/hm²、MAE为0.194 t/hm²);总产预测中,Cubist表现最佳(R^2 为0.952、RMSE为9.42 kt、MAE为5.00 kt)。油菜单产与总产预测结果在空间分布上与观测值总体一致。本研究提出的多源遥感与气象信息融合结合树结构机器学习方法,能够有效提升区域尺度油菜单产与总产预测精度,为油菜产量监测、精准农业管理及农机收获调度决策提供了可靠的技术支撑。

关键词: 油菜; 多源大数据; 单产; 总产; 预测模型; 树结构机器学习

中图分类号: TP79 文献标识码: A 文章编号: 1000-1298(2026)11-0292-11 **OSID:** 

Prediction of Canola Yield and Production by Fusing Multisource Remote Sensing and Meteorological Data

CHEN Jiang^{1,2} LIAO Qingxi^{1,2} XU Beibei^{3,4} LIAO Yitao^{1,2} WAN Xingyu^{1,2}

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China
2. Key Laboratory of Agricultural Equipment in Mid-lower Yangtze River,
Ministry of Agriculture and Rural Affairs, Wuhan 430070, China
3. State Key Laboratory of Nutrient Use and Management, Beijing 100193, China
4. College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China)

Abstract: Canola is one of the most important oilseed crops worldwide, and accurate prediction of its yield and total production is of great significance for food and edible oil security assessment, agricultural production management, and bioenergy potential analysis. To address the limited sample size in canola yield and production prediction, as well as the lack of systematic evaluation of production modeling and multi-source information synergy, county-level canola yield and total production data were collected, and multi-source remote sensing and meteorological big data were integrated to develop a collaborative prediction framework for canola yield and total production. On this basis, six tree-based machine learning models were employed to systematically evaluate the performance differences among models and data-source combinations in predicting canola yield and production. The results demonstrated that the joint use of remote sensing and meteorological data consistently outperformed single-source inputs for both yield and

收稿日期: 2026-01-04 修回日期: 2026-01-24
基金项目: 国家重点研发计划项目(2021YFD1600503)、中央高校基本科研业务费专项资金资助项目(2662025GXPY004)和国家油菜产业技术体系专项(CARS-12-27)
作者简介: 陈江(1992—),男,副研究员,博士,主要从事多尺度作物信息感知与智能决策研究,E-mail: chenjiang@mail.hzau.edu.cn
通信作者: 廖宜涛(1982—),男,教授,博士,主要从事现代农业装备设计与测控研究,E-mail: liaotao@mail.hzau.edu.cn

production prediction, with this advantage showing strong consistency and robustness across different tree-based model structures. For yield prediction, CatBoost achieved the best performance ($R^2 = 0.679$, RMSE = 0.269 t/hm², MAE = 0.194 t/hm²), whereas for production prediction, Cubist performed best ($R^2 = 0.952$, RMSE = 9.42×10^3 t, MAE = 5.00×10^3 t). The spatial patterns of the predicted yield and production were generally consistent with the observations. Overall, the results indicated that the integration of multi-source remote sensing and meteorological information with tree-based machine learning methods can effectively improve the accuracy of regional-scale canola yield and production prediction. The research result can provide reliable technical support for canola yield monitoring, precision agricultural management, and harvest scheduling decisions for agricultural machinery.

Key words: canola; multisource big data; yield; production; prediction model; tree-based machine learning

0 引言

油菜是全球十分重要的油料作物之一,广泛种植于温带地区,包括加拿大、中国和澳大利亚等国家^[1]。作为主要的食用植物油来源,油菜在保障全球植物油供给和粮油安全方面发挥着重要作用。此外,油菜副产品(如菜粕)是优质的蛋白饲料资源,在畜牧业中具有重要经济价值。随着生物柴油和可再生能源需求的持续增长,油菜因其高含油率和成熟的加工体系,逐渐成为重要的生物柴油原料,其作为工业原料的潜在价值也日益凸显^[2]。在农业生产体系中,油菜通常被纳入轮作制度,有助于改善土壤结构,打破连作条件下病虫害的寄主连续性,从而降低病虫害发生风险,并提升农田生态系统的可持续性。在粮油安全、能源转型和农业可持续发展等多重需求驱动下,准确监测油菜单产与总产的时空变化,对于科学评估油菜生产能力与稳定性、验证与改进产量预测结果,以及支撑区域与国家尺度的粮油供给保障和生物能源潜力评估具有重要意义^[3]。

作物产量预测是智慧农业研究中的核心问题之一,相关研究在水稻、小麦和玉米等主粮作物上已取得较为丰富的成果^[4]。这些作物在全球粮食供给体系中占据主导地位,因而长期以来成为产量监测与预测研究的主要对象^[5]。现有研究多集中于作物单产预测,而针对总产量的系统性研究相对较少。相比之下,针对油菜的单产与总产预测研究明显不足,研究体系尚不完善^[6]。作为重要的油料作物,在油菜收获前对其产量进行准确预测,对于智慧农业背景下的生产决策具有重要支撑作用。可靠的油菜单产与总产预测信息不仅有助于指导田间精准管理措施的制定与优化,如施肥、灌溉和病虫害防控等,还能够为机械化作业与收获时序安排提供科学依据,降低因收获过早或过晚导致的产量损失和品质风险^[7]。此外,在区域尺度上,提前掌握油菜产量空间分布与规模变化,对于统筹农机资源配置、优

化跨区作业调度以及提升农业生产组织与管理效率具有重要意义^[8]。因此,有必要开展油菜单产与总产的协同预测研究。

遥感数据是作物产量预测中最常用的信息来源之一^[9]。遥感植被指数能够有效表征作物冠层结构、生物量累积及生长状态,因而被广泛应用于作物单产预测^[10]。常用的遥感植被指数包括归一化差异植被指数(Normalized difference vegetation index, NDVI)、增强型植被指数(Enhanced vegetation index, EVI)以及归一化差异水分指数(Normalized difference water index, NDWI)等^[11]。近年来,随着对植被光合作用过程认识的不断深化,新型遥感指标近红外反射率植被指数(Near-infrared reflectance of vegetation, NIRv)逐渐受到关注,并在作物产量预测中表现出较好的应用潜力。除光谱指数外,基于遥感反演的高级生物物理和生态产品,如叶面积指数(Leaf area index, LAI)、总初级生产力(Gross primary production, GPP)和蒸散发(Evapotranspiration, ET),也逐步被引入作物单产预测模型,以更直接刻画作物生长过程及其碳水循环特征。

作物生长和产量形成过程显著受到气象条件的影响,包括辐射、温度、降水和水汽压等关键因子^[12]。气象因子通过调控作物物候进程、光合作用效率和水分利用状况,对作物产量具有直接且持续的影响。单纯依赖遥感植被指数或遥感产品进行产量估算,往往难以充分刻画气象条件对作物生长和产量形成的独立调控作用,因而预测精度可能受到一定限制^[13]。尽管遥感植被指数在一定程度上综合反映了气象环境的影响,但其难以完全区分和量化不同气象因子对产量形成过程的具体贡献。因此,将遥感信息与气象数据相结合,构建多源数据驱动的作物产量预测框架,有助于更全面刻画作物生长环境与产量形成机制。已有研究表明,在遥感与气象数据协同作用下,作物产量预测的准确性和稳定性通常优于单一数据源情形^[14]。然而,多源遥感

与气象数据融合在油菜单产与总产预测中的综合效益及其相对优势仍缺乏系统评估,有必要进一步开展针对性的比较与分析。

作物产量预测模型主要分为 3 类:基于机理的作物生长模型、数据同化方法以及数据驱动的机器学习方法^[11]。作物生长模型通过显式描述作物生理过程与环境因子的相互作用来模拟产量形成机制,具有明确的物理和生物学含义,但通常对模型参数、管理信息和高质量输入数据依赖较强^[15]。数据同化方法通过融合观测数据与模型模拟结果,在一定程度上提高了产量估计的准确性,但其实现过程较为复杂,对模型结构和观测误差的设定具有较高要求^[16]。相比之下,机器学习方法无需显式刻画作物生长机理,能够从数据中自动学习作物生长环境与产量之间的非线性关系,因而在作物产量预测研究中得到了广泛应用^[17]。近年来,随着计算能力的提升和算法的发展,一些新兴的深度学习模型,如 Transformer 和 Kolmogorov-Arnold Network(KAN),在高维特征表达和复杂非线性建模方面展现出潜力^[18]。然而,深度学习模型通常对训练样本的数量和质量要求较高^[19],而在实际产量预测研究中,高质量、长时间序列且空间一致的产量样本往往难以获取。这在一定程度上限制了深度学习模型在区域尺度产量预测中的稳定应用。

在样本规模相对有限的情况下,基于树结构的机器学习模型因其对小样本的适应性、对特征尺度不敏感以及对非线性关系的良好刻画能力,也是作物产量预测研究中的可靠选择^[20]。这类模型能够有效处理多源异构数据,具有较强的鲁棒性和泛化能力。其中,随机森林(Random forest, RF)因结构简单、稳定性高且易于实现,在作物产量预测研究中应用最为广泛^[21]。在此基础上,近年来又发展出多种改进和扩展的树结构机器学习模型,包括极端梯度提升(eXtreme gradient boosting, XGBoost)、轻量级梯度提升机(Light gradient boosting machine, LightGBM)、类别提升算法(Categorical boosting, CatBoost)、规则模型与回归树结合方法(Cubist)以及极端随机树(Extremely randomized trees, ExtraTrees)等。这些模型也逐渐被引入作物产量预测研究,并在不同作物和区域尺度上表现出较好的预测性能。

本文以方法开发与评估为核心,构建一套融合多源遥感与气象大数据的油菜单产与总产协同预测框架,并系统评估其在区域尺度产量预测中的适用性与稳定性。研究引入 6 种树结构机器学习模型,评估不同模型在油菜单产与总产预测中的性能差

异,重点探讨多源遥感信息与气象信息的协同作用及其对预测精度的增益效果。以美国蒙大拿州和北达科他州油菜主要种植县域为案例区,对所提出的方法进行验证与性能评估。

1 研究区和数据

1.1 研究区域概况

为验证本文提出的方法,选取美国北部大平原地区的典型油菜种植县域作为案例区,主要包括蒙大拿州和北达科他州等油菜主产区(图 1)。该区域地处中高纬度温带大陆性气候区,地势相对平坦,农业生产以旱作为主,是美国油菜最为集中的种植区域之一。研究区油菜种植呈现显著的区域集聚特征,主要分布于北部和中部县域,形成连续且稳定的油菜生产带。该区域种植的油菜类型以春油菜为主,通常于每年 4—5 月播种,生长季集中在春末至夏季,并于 9 月左右完成收获。生长季内,春夏季温度变化、降水时空分布以及辐射条件共同调控油菜生长发育过程,而较强的年际气候波动使得油菜单产和总产在空间和时间尺度上均表现出显著差异。得益于种植规模大、空间分布相对集中以及物候规律清晰,该区域为融合多源遥感与气象大数据开展油菜单产和总产预测研究提供了理想的研究场景。基于上述特点,本文选取该区域为研究区,系统评估多源遥感和气象大数据在油菜单产与总产预测中的适用性与不确定性,为区域尺度油菜产量监测与农业生产管理提供科学依据。

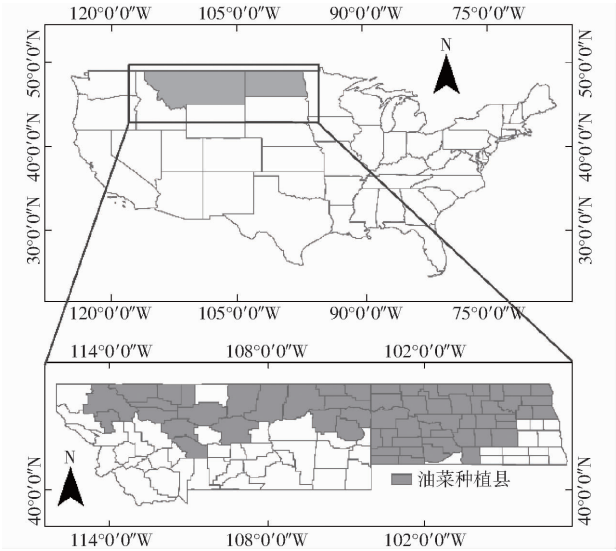


图 1 研究区和油菜种植区

Fig. 1 Study area and canola planting regions

1.2 数据获取与处理

1.2.1 数据源

本文所采用的数据主要包括多源遥感数据、气

象数据以及统计产量数据。遥感数据来自美国国家航空 航 天 局 (National Aeronautics and Space Administration, NASA) 发布的中分辨率成像光谱仪 (Moderate resolution imaging spectroradiometer, MODIS) 系列产品, 空间分辨率统一为 500 m。具体包括: 每日地表反射率产品, 即 MODIS 双向反射分布函数/反照率天底反射率产品 (BRDF/Albedo Nadir Reflectance, MCD43A4); 4 d 合成的叶面积与光能吸收参数, 即叶面积指数 LAI 和植被吸收光合有效辐射比例 (Fraction of photosynthetically active radiation, FPAR), 采用 MCD15A3H 产品; 植被碳通量相关变量采用 MODIS 陆地初级生产力产品 (MOD17A2HGF), 提供 8 d 合成的总初级生产力 GPP 和净初级生产力 (Net primary production, NPP); 地表水分与能量通量信息采用 MODIS 蒸散发产品 (MOD16A2GF), 提供 8 d 合成的实际蒸散发 ET 和潜在蒸散发 (Potential evapotranspiration, PET)。

气象数据来源于 Daymet 逐日气象数据集, 该数据由美国橡树岭国家实验室 (Oak Ridge National Laboratory, ORNL) 发布, 空间分辨率为 1 km。本文选取的气象变量包括日下行短波辐射 (Downward shortwave radiation, DSR)、最低气温 (Daily minimum air temperature, Tmin)、日最高气温 (Daily maximum air temperature, Tmax)、日降水量 (Daily precipitation, DP) 以及大气水汽压 (Vapor pressure, VP), 用于刻画油菜生长季内的热量、水分和大气湿度条件。

所有遥感与气象数据均通过 Google Earth Engine (GEE) 云计算平台获取, 时间范围覆盖 2008—2023 年。油菜单产和总产统计数据来源于美国国家农业统计局 (National Agricultural Statistics Service, NASS), 提供了蒙大拿州和北达科他州县尺度的油菜单产和总产数据。

1.2.2 油菜种植区识别

采用 NASS 发布的农作物数据层 (Cropland data layer, CDL) 对研究区油菜种植区进行识别。CDL 数据基于多源遥感影像与统计样本, 通过监督分类方法生成, 提供覆盖美国本土的年度作物类型分布图, 空间分辨率为 30 m。CDL 数据在农作物类型识别方面具有较高精度, 尤其适用于种植规模较大、空间分布集中的作物区。凭借其长期连续、高空间分辨率和明确的作物类型标注特征, CDL 已被广泛用于作物分布制图、作物识别精度验证、种植结构变化分析以及区域尺度作物产量预测等研究^[22]。本文使用 2008—2023 年逐年的 CDL 数据, 提取当年油菜

对应的作物类别信息, 以表征研究区内油菜种植区的空间分布及其年际变化特征。

1.2.3 遥感数据处理

基于 MODIS 多源遥感产品构建油菜生长与碳水过程相关的遥感特征变量。利用 MCD43A4 产品提供的日尺度反射率数据, 计算得到日尺度的植被指数。基于红光 (Red, R)、近红外 (Near-infrared, NIR)、绿光 (Green, G) 和短波红外 (Shortwave infrared, SWIR) 波段, 逐日计算增强型植被指数 (Enhanced vegetation index 2, EVI2)、绿叶绿素指数 (Green chlorophyll index, GCI)、归一化差异水分指数 NDWI 以及近红外植被指数 NIRv, 以综合表征油菜冠层结构、叶绿素含量、叶片水分状况及光合活性。

在获得日尺度植被指数后, 为实现多源遥感产品在时间分辨率上的一致性, 进一步对数据进行时间尺度重构。将日尺度的 EVI2、GCI、NDWI 和 NIRv 按 8 d 时间窗口进行合成, 同时将 4 d 分辨率的 LAI/FPAR 产品合成为 8 d 尺度, 从而与 MOD17A2HGF 和 MOD16A2GF 产品在时间上严格匹配。通过该处理, 不同遥感变量在同一时间尺度下具有一致的时间信息, 为后续特征融合和模型构建提供基础。

在空间层面, 为确保遥感特征能够真实反映油菜生长状况, 基于 2008—2023 年逐年的 CDL 油菜分类结果, 构建年度油菜种植掩膜, 并将其应用于所有 8 d 尺度的遥感产品中, 以提取油菜种植区内的遥感参数。最终, 在油菜种植县区内提取 5—8 月生长季期间 8 d 分辨率的油菜区时序 10 个遥感参数, 用于构建油菜单产预测模型; 同时, 对同一时期内的遥感参数进行时序累积与汇总, 形成累积特征变量, 用于油菜总产预测模型的输入。单产预测基于县域尺度油菜种植区内遥感与气象数据的空间均值时序特征, 而总产预测则采用油菜种植区内遥感与气象变量的空间累积量 (和值) 特征, 其中包含了油菜种植面积信息。

1.2.4 气象数据处理

利用 Daymet 提供的逐日气象数据, 对油菜生长季内的气象变量进行时间尺度与空间范围上的统一处理。选取 DSR、Tmin、Tmax、DP 和 VP 等气象变量, 按 8 d 时间窗口进行合成, 以匹配遥感产品的时间分辨率。在空间处理方面, 结合逐年 CDL 提供的油菜种植分布信息, 对合成后的 8 d 尺度气象数据进行掩膜, 仅保留油菜种植县区内的气象信息。在油菜 5—8 月生长季内, 于油菜种植县区提取 8 d 分辨率的时序 5 个气象参数时序数据, 用于油菜单产

预测模型的输入;同时,对生长季内气象变量进行时间累积,构建年度尺度气象特征,用于油菜总产建模。

2 研究方法

2.1 树结构模型

本文采用 6 种基于树结构的回归学习模型,包括 CatBoost、XGBoost、RF、LightGBM、Cubist 以及 ExtraTrees。其中,集成学习模型的回归过程均通过构建多棵回归决策树并对其输出进行集成实现;对于基于梯度提升框架的模型,则通过逐步优化目标函数并累积弱学习器完成回归建模。

2.2 模型参数

本文对 6 种树结构模型进行网格搜索以优化超参数组合,最小化建模误差以选择最优参数设置。CatBoost 回归模型采用的主要参数包括:学习率为 0.05,树深度为 6,叶节点正则化参数为 3.0,损失函数采用均方根误差(Root mean square error, RMSE),并统一设置单线程训练。XGBoost 回归模型的参数为:树的数量为 80,学习率为 0.05,最大树深度为 5,最小叶节点权重为 5,样本与特征子采样比例均为 0.8,并引入 L2 正则化项(reg_lambda 为 1.0)。RF 回归模型采用的参数为:决策树数量为 40,节点分裂时的特征选择方式为平方根策略,叶节点最小样本数为 2。LightGBM 回归模型采用的参数包括:树的数量为 150,学习率为 0.05,叶节点最小样本数为 20,样本与特征子采样比例均为 0.8,正则化参数设为 1.0。Cubist 回归模型的参数为:回归树数量设为 10,邻域样本数为 5。ExtraTrees 回归模型的参数包括:决策树数量为 120,特征分裂采用平方根策略,叶节点最小样本数为 2,且不进行自助采样。

2.3 模型性能评价

为评估不同树结构回归模型的预测性能,本文对所有模型统一采用 10 折交叉验证方法。将样本集随机划分为 10 个均匀的子集,每次选取其中 1 个子集作为验证集,其余 9 个子集作为训练集,重复该过程直至所有子集均参与一次验证,从而获得模型在独立验证样本上的预测结果。模型性能通过计算观测的油菜单产和总产与对应预测值之间的决定系数(Coefficient of determination, R^2)、RMSE 和平均绝对误差(Mean absolute error, MAE)进行评价。

为评估不同数据源组合对油菜单产与总产预测精度的影响,分别构建了 3 类输入方案:遥感数据与气象数据联合输入、仅使用遥感数据以及仅使用气象数据。利用 6 种不同树结构回归模型,同时评估了 3 种不同数据源组合对单产和总产预测的整体表

现并展开分析。

3 结果与分析

3.1 观测的单产和总产

图 2 展示了研究区内油菜观测单产与总产的频率分布特征。两者在分布形态及离散程度上表现出明显差异,反映了不同油菜生产变量的空间异质性。油菜单产的分布相对集中,呈现单峰特征,主要分布于 1.0 ~ 2.5 t/hm² 区间。统计结果表明,单产的最小值为 0.426 t/hm²,最大值为 3.127 t/hm²,平均值为 1.737 t/hm²,标准差为 0.476 t/hm²,对应的变异系数为 27.4%。

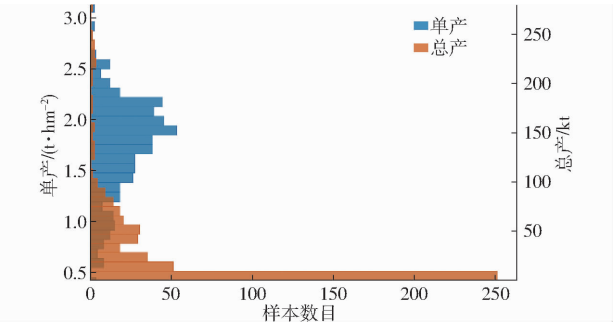


图 2 观测的油菜单产和总产频率分布直方图
Fig. 2 Frequency histograms of observed canola yield and production

与单产相比,油菜总产的分布呈现明显的右偏特征。总产最小值为 0.29 kt,最大值达到 279.53 kt,平均值为 29.21 kt,标准差为 43.04 kt,其变异系数高达 147.3%。该结果表明,油菜总产在空间尺度上的离散程度更大,其变化不仅受单产水平影响,同时强烈受种植面积差异的调制。

3.2 不同数据源组合下单产与总产预测性能对比

在单产预测中,遥感与气象数据联合输入在 6 种树结构回归模型中均表现出最优的预测性能(表 1)。与仅使用遥感数据相比,联合输入遥感与气象数据使单产预测的 RMSE 降低 0.31% ~ 6.57%,MAE 降低 2.38% ~ 8.20%。相较仅使用气象数据,遥感与气象数据联合输入下 RMSE 和 MAE 的降低幅度进一步扩大,分别为 13.36% ~ 20.65% 和 11.34% ~ 17.62%,表明单一气象数据在刻画单产空间与年际差异方面存在明显局限。

同时,遥感与气象数据联合输入在 R^2 指标上同样表现出系统性提升,6 种模型在联合输入条件下均获得了更高的决定系数(图 3)。这表明,多源信息融合有效增强了模型对影响单产形成关键要素的综合表征与解释能力,包括作物生长状况、冠层结构与生物量积累特征,以及温度、水分条件和辐射环境等气象因子的共同作用。相较之下,仅输入遥感数

表 1 不同数据源组合下预测油菜单产验证指标

Tab.1 Validation indicators of canola yield prediction under different input data combinations

模型	输入数据类型	R^2	RMSE/ ($t \cdot hm^{-2}$)	MAE/ ($t \cdot hm^{-2}$)
CatBoost	遥感 + 气象	0.679	0.269	0.194
	遥感	0.645	0.283	0.204
	气象	0.542	0.322	0.227
XGBoost	遥感 + 气象	0.651	0.281	0.202
	遥感	0.618	0.294	0.215
	气象	0.508	0.334	0.236
RF	遥感 + 气象	0.651	0.281	0.202
	遥感	0.624	0.291	0.216
	气象	0.543	0.321	0.228
LightGBM	遥感 + 气象	0.636	0.287	0.205
	遥感	0.603	0.299	0.222
	气象	0.476	0.344	0.241
Cubist	遥感 + 气象	0.649	0.282	0.201
	遥感	0.647	0.283	0.206
	气象	0.489	0.340	0.234
ExtraTrees	遥感 + 气象	0.664	0.276	0.198
	遥感	0.618	0.294	0.215
	气象	0.568	0.312	0.221

据在单产预测中的整体性能优于仅输入气象数据,表明遥感观测在表征作物生长状态、冠层结构及生产力差异方面具有更直接的信息优势。然而,仅依赖单一遥感或气象数据仍难以全面刻画影响单产形成的多重驱动因素,其预测误差和不确定性均明显高于多源遥感和气象数据融合方案。

在总产预测中,不同数据源组合之间的性能也存在明显差异(表 2)。在 6 种树结构回归模型中,遥感与气象数据联合输入均取得了最高的 R^2 以及最低的 RMSE 和 MAE,表现出稳定的预测优势。与仅输入遥感数据相比,联合输入遥感与气象数据使总产预测的 RMSE 降低 0.72% ~ 9.13%,MAE 降低 2.99% ~ 11.80%;相较仅输入气象数据,误差降低幅度进一步扩大,RMSE 和 MAE 的降低幅度在不同模型中普遍超过 143%。

上述结果在 6 种不同树结构回归模型中均表现出高度一致性。尽管模型结构和学习机制存在差异,所有模型在单产与总产预测任务中均呈现出联合输入遥感与气象数据优于仅输入遥感数据,仅输入遥感数据优于仅输入气象数据的稳定排序关系

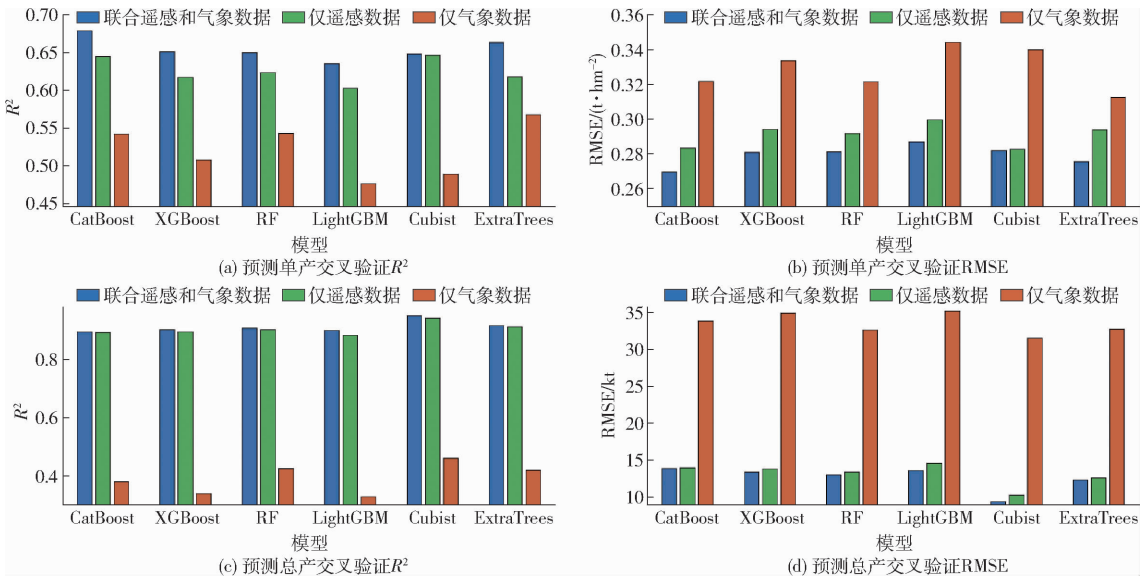


图 3 不同数据源组合下油菜单产与总产预测的验证结果

Fig.3 Validation results of canola yield and production prediction under different input data combinations

(图 3)。这一现象表明,多源数据融合所带来的性能提升并非依赖于特定模型结构,而是源于遥感与气象信息在作物产量形成机理上的互补性。

3.3 不同树结构模型预测单产和总产性能对比

图 4 展示了 6 种树结构回归模型在单产预测中的交叉验证散点分布结果。整体来看,各模型的预测值与观测值均沿 1:1 参考线分布,表明所有模型均能够较好地捕捉单产的总体变化特征。然而,不同模型在拟合精度和误差水平上仍存在一定差异。

各模型在单产预测中 R^2 为 0.636 ~ 0.679, RMSE 和 MAE 分别为 0.269 ~ 0.287 t/hm^2 和 0.194 ~ 0.205 t/hm^2 。其中, CatBoost 模型在单产预测中表现最佳,其 R^2 最高(0.679),同时具有较低的 RMSE (0.269 t/hm^2) 和 MAE (0.194 t/hm^2),显示出更强的建模能力。ExtraTrees 模型的整体表现次之,其 R^2 和误差指标同样处于较优水平,而 LightGBM 在单产预测中的解释能力相对较低。从散点分布特征可以看出,在中等单产区间(1 ~ 2.5 t/hm^2),各模型预测结果与观测值吻合较好;而在低单产和高单产

表 2 不同数据源组合下油菜总产预测验证指标

Tab.2 Validation indicators of canola production prediction under different input data combinations

模型	输入数据类型	R^2	RMSE/ kt	MAE/ kt
CatBoost	遥感 + 气象	0.896	13.89	6.95
	遥感	0.894	13.99	7.40
	气象	0.381	33.82	17.28
XGBoost	遥感 + 气象	0.903	13.39	6.75
	遥感	0.897	13.82	7.43
	气象	0.338	34.97	18.70
RF	遥感 + 气象	0.908	13.01	6.53
	遥感	0.903	13.40	6.99
	气象	0.424	32.63	16.79
LightGBM	遥感 + 气象	0.900	13.58	7.46
	遥感	0.884	14.62	7.24
	气象	0.329	35.21	19.04
Cubist	遥感 + 气象	0.952	9.42	5.00
	遥感	0.943	10.28	5.59
	气象	0.462	31.53	16.17
ExtraTrees	遥感 + 气象	0.917	12.36	6.41
	遥感	0.913	12.66	6.82
	气象	0.419	32.77	16.57

极值区间,不同模型的预测离散程度存在差异,反映出模型结构在刻画极端单产条件方面的能力仍有所不同。RF 与 XGBoost 模型的评价指标在 3 位小数精度下数值一致,但在更高精度下仍存在细微差异,且其散点分布形态在不同单产区间表现出一定差别。

图 5 给出了 6 种树结构回归模型在总产预测中的散点验证结果。相较单产预测,各模型在总产预测中的整体建模效果更为显著。各模型 R^2 为 0.896 ~ 0.952, RMSE 和 MAE 分别为 9.42 ~ 13.89 kt 和 5.00 ~ 7.46 kt。其中, Cubist 模型在总产预测中表现最优,其 R^2 达到 0.952,同时 RMSE 和 MAE 分别降至 9.42 kt 和 5.00 kt,显著优于其他模型,表明该模型在刻画县域尺度总产变异方面具有更强的表达能力。ExtraTrees 与 RF 模型的预测性能亦较为突出,而 CatBoost、XGBoost 和 LightGBM 在高总产区间的预测离散度相对较大。

从散点分布可以发现,在低至中等总产区间

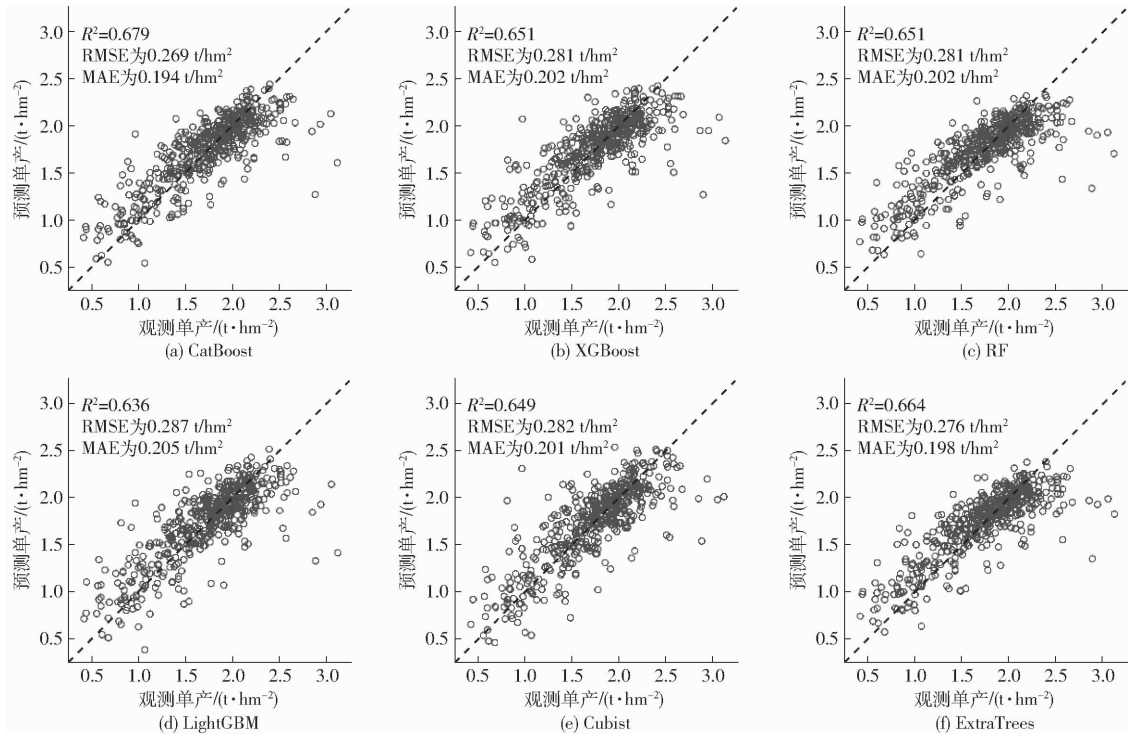


图 4 不同树结构模型油菜单产预测验证散点图

Fig.4 Validation scatter plots of canola yield predictions from different tree-based models

(0 ~ 100 kt),各模型预测结果与观测值高度一致;而在高总产区间(>150 kt),模型间差异逐渐显现,部分模型对高产县域存在一定低估或高估现象。先预测单产并结合种植面积估算总产的精度 R^2 为 0.896、RMSE 为 12.59 kt、MAE 为 6.06 kt,低于本文提出的直接利用遥感和气象数据预测总产的方案。这种差异可能与误差累积有关,即单产和种植面积估算中引入的不确定性在乘积过程中被放大,从而

增加了总产预测误差(RMSE 增加 25.18%,MAE 增加 17.49%)。

从评价指标上看,总产预测的 R^2 高于单产预测,但这一差异需要结合数据本身的统计特征进行理解。 R^2 反映的是模型对观测值总方差的解释比例,其数值受观测数据变异程度影响较大。本研究中,单产观测数据的变异系数为 27.4%,而总产观测数据的变异系数高达 147.3%,表明总产在年际

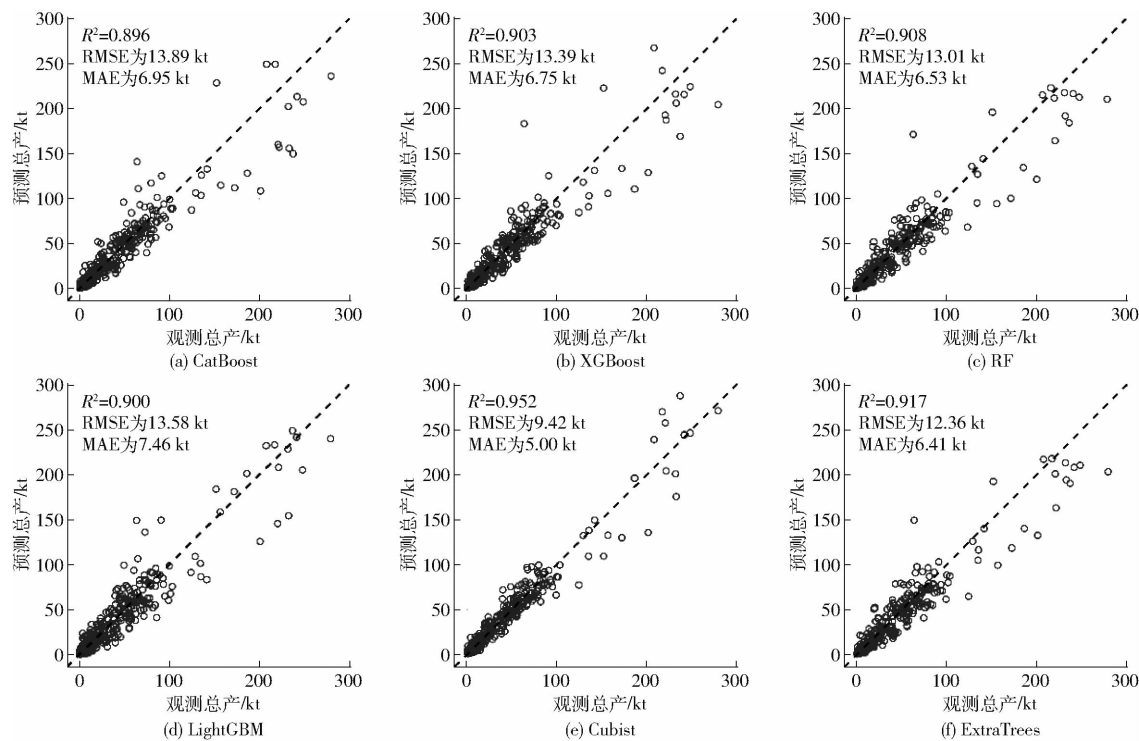


图 5 不同树结构模型油菜总产预测验证散点图

Fig. 5 Validation scatter plots of canola production predictions from different tree-based models

和空间尺度上的波动显著强于单产。总产数据中叠加的种植面积信息在统计上进一步放大了观测值的变化幅度,使模型在刻画年际变化时更容易获得较高的 R^2 。另一方面,由于单产与总产在物理量纲和数量级上存在显著差异, RMSE 和 MAE 等绝对误差指标不具备直接可比性。为此,通过引入相对均方根误差 (Relative root mean square error, RRMSE) 对两类预测结果进行归一化比较,结果显示单产预测的 RRMSE 为 15.50%, 低于总产预测的 32.24%, 说明在相对误差意义上,单产预测具有较好的稳定性,其预测精度并不低于总产预测。

3.4 单产和总产预测模型误差分析与空间制图

图 6 展示了单产与总产预测误差随播种面积变化的统计关系。结果表明,单产预测误差与播种面积之间不存在显著相关性,其皮尔逊相关系数 r 为 -0.021 , 且未通过显著性检验 (p 为 0.637)。这说明在不同种植规模条件下,单产预测误差总体上呈随机分布,模型在小面积与大面积县域中的单产预测稳定性较为一致,未表现出明显的尺度依赖性偏差。相比之下,总产预测误差与播种面积之间呈现出显著但相对较弱的负相关关系 (r 为 -0.111 、 p 为 0.013)。从误差散点分布来看,单产预测误差主要集中在零值附近,极端正负误差数量较少;而总产预测误差在高播种面积区间呈现出更大的数值跨度,反映出总产预测在极端高产或大面积县域中仍存在较高不确定性。

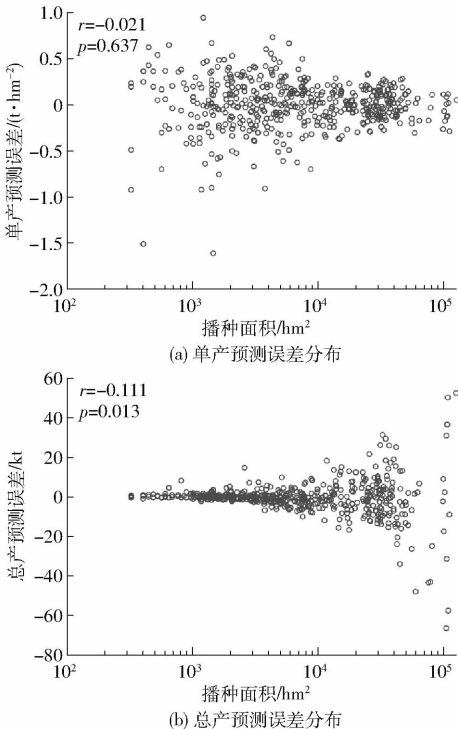


图 6 油菜单产和总产预测误差与种植面积关系
Fig. 6 Relationships between prediction errors of canola yield and production and planting area

图 7 分别显示了 2008—2023 年县域尺度单产和总产观测值、预测值和预测误差的空间分布图。总体上单产的预测值与观测值接近,单产预测误差在研究区内呈现出明显的空间异质性,但大多数县域的预测误差处于较低水平。定量上,多数县域的单产预测误差集中在 $0 \sim 0.3 \text{ t/hm}^2$ 范围,仅少数区域出现超过

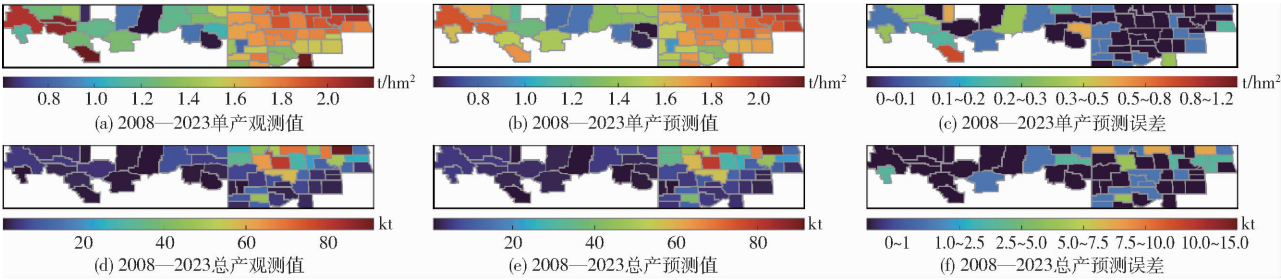


图 7 2008—2023 年油菜单产与总产预测值及其误差空间格局

Fig. 7 Spatial patterns of predicted canola yield and production and their errors from 2008 to 2023

0.5 t/hm² 的较大误差。多数县域的总产预测误差低于 5 kt, 而误差超过 10 kt 的县域数量较少。

从空间格局上看, 单产预测误差在边缘或零散种植区相对较高, 而在油菜种植较为集中、面积较大的核心产区误差相对较小。这表明模型在主产区内部能够较稳定地捕捉单产空间分布特征, 而在非核心或种植规模较小的区域, 单产预测仍受到局地环境异质性和观测不确定性的影响。总产预测误差在研究区中呈现出一定的区域聚集特征, 高误差区往往对应于总产水平较高的县域。这一现象与前述误差和播种面积关系分析结果相一致, 表明在高总产或大规模种植条件下, 总产预测误差增大。

4 讨论

本研究基于 2008—2023 年县域尺度油菜单产与总产观测数据, 系统评估了多源遥感与气象数据融合条件下, 不同树结构机器学习模型对油菜单产和总产的预测能力。研究表明, 多源遥感与气象信息的融合能够更全面地表征作物生长状态与气候驱动条件, 是提高区域尺度油菜单产和总产预测精度的关键途径。另一方面, 单一遥感数据源输入对油菜单产和总产的预测精度高于单一气象数据源。然而, 本研究所采用的遥感数据仍主要以光学遥感为主, 对作物结构和水分状况的刻画存在一定局限。未来可进一步引入更多类型的多源遥感观测数据^[23], 例如被动微波遥感对植被含水量和土壤水分的敏感信息、主动微波合成孔径雷达对冠层结构的全天候观测能力, 以及激光雷达提供的植被垂直结构信息, 从而在物理和结构层面进一步约束油菜生长过程, 提升单产和总产预测的精度与稳健性。在总产预测中, 本文采用油菜种植区内遥感与气象变量的空间累积量(区域和值)作为模型输入特征。该类特征在统计意义上同时反映了油菜种植面积及其生长强度的综合效应, 参与累积计算的像元数量随油菜种植面积变化而变化, 而单个像元的变量值则表征油菜生长状况。因此, 即使未显式引入面积变量, 模型输入已隐式包含面积信息, 能够直接表征油菜总产。

在算法开发过程中, 也尝试了多种深度学习模型, 包括 Transformer 和 KAN, 但在本研究的数据条件下, 其预测效果并未优于最终采用的树结构机器学习模型。这可能与油菜单产和总产样本量相对有限有关, 使得深度学习模型难以充分发挥优势。这说明在农业产量建模中, 不同方法在不同数据规模和应用场景下各有其适用性。与以往研究多聚焦单一单产指标不同^[24], 本研究同时对油菜单产和总产进行了系统预测与评估。这种联合预测框架具有重要的理论意义和应用价值。单产主要反映作物生长状况和单位面积生产能力, 而总产则综合体现了单产水平与种植面积的共同作用, 是农业生产决策和粮油安全评估中的关键指标。同时预测单产和总产, 有助于更全面地理解区域尺度油菜生产格局及其驱动机制。此外, 总产预测的不确定性不仅来源于单产预测误差, 还受到播种面积估计误差及其空间异质性的影响。这一发现强调了样本数量和样本空间代表性的重要性。未来研究可通过整合更多区域的油菜单产和总产观测数据, 扩大样本规模和空间覆盖范围, 从而提升模型应用范围, 为跨区域、跨气候带的油菜产量预测提供支撑。

本文在 MODIS 分辨率 500 m 下基于 CDL 构建油菜掩膜进行像元筛选, 只要 MODIS 像元内包含油菜像元, 即保留其 MODIS 观测值参与县域统计分析。鉴于研究区县域油菜种植面积较大(平均为 15 487 hm²)且分布连续, 该处理方式在县域尺度下对平均与累积特征的影响相对有限。本研究在县域尺度上开展油菜单产和总产预测, 能够有效刻画区域尺度的总体变化特征, 但仍难以反映田块尺度的精细差异。随着中高分辨率卫星数据的持续发展, 未来有望通过多尺度数据融合实现更精细的油菜产量预测。例如, PlanetScope 提供的高时空分辨率影像可用于捕捉作物快速生长变化, Landsat-8/9 和 Sentinel-2 可提供稳定的中高分辨率光学观测^[25], 而 Sentinel-1 合成孔径雷达数据则可在多云条件下补充作物结构与水分信息^[26]。通过联合多源中高分辨率卫星数据, 并结合田块尺度的地块边界信

息,未来研究有望实现从县域到田块尺度的油菜单产和总产精细化预测。这不仅有助于提高模型精度,也为精准农业管理和农田生产决策提供更具操作性的空间信息支持。

本研究构建的多源数据融合与多模型评估框架,对我国油菜单产和总产预测具有重要借鉴意义。我国油菜种植区域主要集中在长江中下游平原等地区,种植制度复杂、空间异质性强,对产量预测提出了更高要求。未来可收集我国油菜地面观测资料与多尺度遥感数据,构建适用于我国油菜主产区的单产与总产联合预测模型。通过融合卫星遥感、无人机观测以及地面传感器等多尺度观测数据,有望在我国油菜主产区实现高时空分辨率的单产和总产动态监测。这将为油菜种植精准管理、收获决策优化以及粮油安全评估提供重要的科学依据。

5 结论

(1)本研究构建了多源遥感与气象数据驱动的

树结构机器学习预测框架。该方法能够准确地刻画区域尺度油菜单产与总产的时空变化特征,预测结果与观测值在统计特征和空间分布上保持良好一致性。

(2)遥感与气象数据联合输入在单产和总产预测中均优于单一数据源输入。多源信息融合在 6 种树结构回归模型中均降低了预测误差并提高了决定系数,说明遥感数据对作物生长状态的表征能力与气象数据对环境驱动因素的刻画具有明显互补性,是提升油菜单产和总产预测精度的关键。

(3)CatBoost 在油菜单产预测中表现最优(R^2 为 0.679、RMSE 为 0.269 t/hm²、MAE 为 0.194 t/hm²),而 Cubist 在油菜总产预测中表现最优(R^2 为 0.952、RMSE 为 9.42 kt、MAE 为 5.00 kt)。油菜单产和总产预测值与观测值空间制图具有较高一致性,多数县域的单产预测误差低于 0.3 t/hm²,总产预测误差低于 5 kt,表明所提出的方法能够稳定、可靠地支持区域尺度油菜产量监测与评估。

参 考 文 献

[1] ZENG F, WANG R, JIANG Y, et al. Growth monitoring of rapeseed seedlings in multiple growth stages based on low-altitude remote sensing and semantic segmentation[J]. Computers and Electronics in Agriculture, 2025, 232: 110135.

[2] CONFORTIN T, TODERO I, LUFT L, et al. Oil yields, protein contents, and cost of manufacturing of oil obtained from different hybrids and sowing dates of canola[J]. Journal of Environmental Chemical Engineering, 2019, 7: 102972.

[3] XU S, ZHU X, CAO R, et al. Automatic SAR-based rapeseed mapping in all terrain and weather conditions using dual-aspect Sentinel-1 time series[J]. Remote Sensing of Environment, 2025, 318: 114567.

[4] MA Y, LIANG S, MYERS D, et al. Subfield-level crop yield mapping without ground truth data: a scale transfer framework [J]. Remote Sensing of Environment, 2024, 314: 114427.

[5] 刘维, 郑昌玲, 孙少杰, 等. 近十年国家气象中心作物产量预报业务技术进展[J]. 中国农业气象, 2025, 46(5): 694-703. LIU Wei, ZHENG Changling, SUN Shaojie, et al. Advance in operational technology of yield forecasting in national meteorological centre in recent 10 years[J]. Chinese Journal of Agrometeorology, 2025, 46(5): 694-703. (in Chinese)

[6] FERNANDO H, HA T, NKETIA K, et al. Machine learning approach for satellite-based subfield canola yield prediction using floral phenology metrics and soil parameters[J]. Precision Agriculture, 2024, 25: 1386-1403.

[7] 马瑜蔓,段博,徐宾灿,等. 基于分数阶微分和无人机高光谱指数优选的油菜产量预测[J]. 农业工程学报, 2025, 41(10): 166-175. MA Yuman, DUAN Bo, XU Bincan, et al. Rapeseed yield prediction based on fractional-order differentiation and UAV hyperspectral index optimization[J]. Transactions of the CSAE, 2025, 41(10): 166-175. (in Chinese)

[8] 万星宇,舒彩霞,廖庆喜,等. 高地隙履带自走式中间条铺油菜割晒机设计与试验[J]. 农业机械学报, 2022, 53(9): 110-120. WAN Xingyu, SHU Caixia, LIAO Qingxi, et al. Design and experiment of self-propelled middle-placement rape windrower with high ground clearance crawler[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(9): 110-120. (in Chinese)

[9] 吴炳方,田富有,曾红伟,等. 全球农情遥感监测云服务平台的新进展[J]. 遥感学报, 2025, 29(6):1918-1937. WU Bingfang, TIAN Fuyou, ZENG Hongwei, et al. Recent advancements of cloud-based global crop monitoring system (CropWatch)[J]. National Remote Sensing Bulletin, 2025, 29(6): 1918-1937. (in Chinese)

[10] WANG C, ZHANG J, WU H, et al. A band selection method for consumer-grade camera modification for UAV-based rapeseed growth monitoring[J]. Smart Agricultural Technology, 2025, 10: 100830.

[11] XIAO G, HUANG J, ZHUO W, et al. Progress and perspectives of crop yield forecasting with remote sensing: a review[J]. IEEE Geoscience and Remote Sensing Magazine, 2025, 3(13): 338-368.

[12] ZHOU W, LIU Y, ATA-UL-KARIM S, et al. Integrating climate and satellite remote sensing data for predicting county-level wheat yield in China using machine learning methods [J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 111: 102861.

- [13] 王鹏新, 王静怡, 郭丰玮, 等. 基于遥感多参数和 Stacking 集成学习的冬小麦单产估测[J]. 农业机械学报, 2025, 56(11): 369–377.
WANG Pengxin, WANG Jingyi, GUO Fengwei, et al. Yield estimation of winter wheat based on multiple remotely sensed parameters and Stacking ensemble learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(11): 369–377. (in Chinese)
- [14] CAI Y, GUAN K, LOBELL D, et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches[J]. Agricultural and Forest Meteorology, 2019, 274: 144–159.
- [15] 杨广, 雷杰, 孔春贤, 等. 膜下滴灌水源矿化度对棉花生长的影响及 AquaCrop 模拟[J]. 农业工程学报, 2022, 38(21): 83–92.
YANG Guang, LEI Jie, KONG Chunxian, et al. Effects of water salinity on cotton growth in mulched drip irrigation and its simulation by AquaCrop model[J]. Transactions of the CSAE, 2022, 38(21): 83–92. (in Chinese)
- [16] 彭慧文, 赵俊芳, 谢鸿飞, 等. 作物模型应用与遥感信息集成技术研究进展[J]. 中国农业气象, 2022, 43(8): 644–656.
PENG Huiwen, ZHAO Junfang, XIE Hongfei, et al. Progresses of crop model application and its integration with remote sensing technology[J]. Chinese Journal of Agrometeorology, 2022, 43(8): 644–656. (in Chinese)
- [17] 竞霞, 张杰, 王娇娇, 等. 水稻产量遥感监测机器学习算法对比[J]. 光谱学与光谱分析, 2022, 42(5): 1620–1627.
JING Xia, ZHANG Jie, WANG Jiaojiao, et al. Comparison of machine learning algorithms for remote sensing monitoring of rice yields[J]. Spectroscopy and Spectral Analysis, 2022, 42(5): 1620–1627. (in Chinese)
- [18] 王鹏新, 杜江莉, 张悦, 等. 基于遥感多参数和 CNN-Transformer 的冬小麦单产估测[J]. 农业机械学报, 2023, 55(3): 173–182.
WANG Pengxin, DU Jiangli, ZHANG Yue, et al. Yield estimation of winter wheat based on multiple remotely sensed parameters and CNN-Transformer[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 55(3): 173–182. (in Chinese)
- [19] 王鹏新, 田惠仁, 张悦, 等. 基于深度学习的作物长势监测和产量估测研究进展[J]. 农业机械学报, 2022, 53(2): 1–14.
WANG Pengxin, TIAN Huiren, ZHANG Yue, et al. Crop growth monitoring and yield estimation based on deep learning: state of the art and beyond[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2): 1–14. (in Chinese)
- [20] 王鹏新, 王颖, 田惠仁, 等. 基于 LightGBM 的冬小麦产量估测与可解释性研究[J]. 农业机械学报, 2023, 54(2): 197–206.
WANG Pengxin, WANG Ying, TIAN Huiren, et al. Interpretability on yield estimation of winter wheat based on LightGBM[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 197–206. (in Chinese)
- [21] CHEN J, YU T, CHERNEY J, et al. Optimal integration of optical and SAR data for improving alfalfa yield and quality traits prediction: new insights into satellite-based forage crop monitoring[J]. Remote Sensing, 2024, 16: 734.
- [22] LARK T, MUELLER R, JOHNSON D, et al. Measuring land-use and land-cover change using the U. S. department of agriculture's cropland data layer: cautions and recommendations[J]. International Journal of Applied Earth Observation and Geoinformation, 2017, 62: 224–235.
- [23] ROßBERG T, SCHMITT M. Comparing the relationship between NDVI and SAR backscatter across different frequency bands in agricultural areas[J]. Remote Sensing of Environment, 2025, 319: 114612.
- [24] 王鹏新, 李明启, 张悦, 等. 基于遥感多参数和 IPSO-WNN 的冬小麦单产估测[J]. 农业机械学报, 2023, 55(1): 154–163.
WANG Pengxin, LI Mingqi, ZHANG Yue, et al. Yield estimation of winter wheat based on remotely sensed multi-parameters and IPSO-WNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 55(1): 154–163. (in Chinese)
- [25] 陈悦, 赵庚星, 常春艳, 等. 基于 Sentinel-2 多光谱影像的小麦-玉米轮作耕地粮食产量估测——以曹县为例[J]. 应用生态学报, 2023, 34(12): 3347–3356.
CHEN Yue, ZHAO Gengxing, CHANG Chunyan, et al. Grain yield estimation of wheat-maize rotation cultivated land based on Sentinel-2 multi-spectral image: a case study in Caoxian County, Shandong, China[J]. Chinese Journal of Applied Ecology, 2023, 34(12): 3347–3356. (in Chinese)
- [26] CHEN J, ZHANG Z. An improved fusion of Landsat-7/8, Sentinel-2, and Sentinel-1 data for monitoring alfalfa: implications for crop remote sensing[J]. International Journal of Applied Earth Observation and Geoinformation, 2023, 124: 103533.