DOI: 10.12398/j.issn.2096-7217.2025.02.009

基于随机森林及遥感植被指数的无人农场水稻产量预测研究

王俊[1]，吴振伟[1]，姜海[1]，柯娟[1]，FOYEZ Ahmed Prodhan[2]

(1. 江苏立卓信息技术有限公司，江苏常州，213000；2. 班加班杜谢赫·穆吉布尔·拉赫曼农业大学，孟加拉国加济布尔，1706，)

摘要：为确保国家粮食安全、制定有效的粮食策略以及促进农业的可持续发展，对区域内水稻产量进行及时且准确的预测至关重要。尽管利用遥感数据结合机器学习技术进行产量预测已有一定进展，但现有研究在深入解析机器学习模型机理及高时空分辨率数据的应用方面仍显不足。本研究基于2023年的Sentinel和MODIS *NDVI*数据及县级产量信息，应用随机森林模型从县级扩展至像素级，以江苏射阳无人农场为例对水稻产量进行预测，并探讨不同特征对模型学习机理的影响。研究结果表明：(1)2023年8—10月的*NDVI*数据与产量之间存在显著相关性，此阶段为作物生长周期中的关键时期，植被覆盖和生长状态在此期间对产量的预测具有重要作用。随机森林模型能够有效预测县级水稻产量，其中*RMSE*值为339.5 kg/hm^2。(2)此外，分析2023年射阳农场水稻产量空间分布图发现，水稻产量在该区域呈现显著的空间异质性，边缘地区产量较低，中心区域产量较高，产量主要集中在9 000～9 300 kg/hm^2。其研究成果不仅加深了对机器学习及植被数据在产量预测中的应用理解，也为提高模型准确性及制定科学的粮食生产策略提供了理论支持。

关键词：随机森林；机器学习；遥感；植被指数；产量预测；生育期

中图分类号：S24 **文献标识码：**A **文章编号：**2096-7217(2025)02-0097-08

王俊，吴振伟，姜海，柯娟，FOYEZ Ahmed Prodhan. 基于随机森林及遥感植被指数的无人农场水稻产量预测研究[J]. 智能化农业装备学报(中英文)，2025，6(2)：97－104

WANG Jun，WU Zhenwei，JIANG Hai，KE Juan，FOYEZ Ahmed Prodhan. Rice yield prediction of unmanned farm based on random forest and remote sensing vegetation index [J]. Journal of Intelligent Agricultural Mechanization，2025，6(2)：97－104

0 引言

随着全球人口的快速增长和气候变化的影响，粮食安全成为了国际社会面临的重大挑战之一。水稻作为世界上主要的粮食作物之一，其生产状况直接关系到数亿人的粮食供应与安全。因此，及时准确地监测和预测水稻产量对于国家粮食管理策略的制定、农业生产规划以及应对潜在的粮食危机具有至关重要的意义[1-3]。

传统的农作物产量预测方法通常依赖于地面调查和统计报告，这些方法不仅耗时耗力，而且容易受到人为因素的干扰。近年来，遥感技术以其高效、客观、覆盖范围广的特点，成为农业监测的重要工具[4-9]。特别是植被指数如归一化差值植被指数(normalized difference vegetation index，*NDVI*)，因其能反映地表植被生长状况而被广泛应用于农作物长势监测和产量估算中[10-11]。

机器学习作为一种强大的数据分析手段，在遥感数据处理与解译方面展现了巨大潜力[12-14]。随机森林(random forest，RF)模型作为一种集成学习方法，因其出色的分类和回归分析能力、对数据集中的异常值和噪声的鲁棒性，以及能够处理高维数据的特性，在众多领域得到了应用。然而，尽管利用遥感数据和机器学习进行产量预测的研究已经取得了进展[15-17]，但如何从机理上深入理解模型的预测性能，以及如何在高时空分辨率下有效利用这些数据，仍然是当前研究中的空白点。

针对上述问题，本研究旨在探索结合Sentinel和MODIS卫星平台的*NDVI*数据，使用随机森林模型对水稻产量进行县级到像素级的预测，以江苏射阳无人农场为例，并进一步解析不同特征对模型学习机制的影响。通过这种方法，我们期望为提高作物产量预测

收稿日期：2024-05-19 修回日期：2024-10-10

基金项目：国家自然科学基金(42071425)；江苏常州钟楼区揭榜挂帅项目(JBGS2023011)

第一作者：王俊，男，1987年生，江苏常州人，硕士，工程师；研究方向农业信息化、农业物联网。E-mail：wangjunrc@qq.com

的准确性提供科学依据,并为区域粮食安全和农业可持续发展政策制定提供支持。

1 材料与方法

1.1 试验地自然条件

试验地江苏射阳无人农场位于射阳现代农业产业园中,地处北纬33°45′,东经120°15′,海拔接近于海平面,属于季风气候区,历史最高温度为39 ℃,历史最低温度为-15 ℃。地处苏北平原东部,紧邻黄海,是北亚热带和暖温带的过渡地带。该地区受季风环流的影响,四季分明,夏季炎热,冬季较冷。年平均气温约为14 ℃,年降水量大约在1 000 mm。

1.2 研究数据

1.2.1 Sentinel数据

本研究采用欧空局(ESA)的Sentinel-2光学卫星数据,该数据适用于监测地表和植被的变化动态。Sentinel-2A携带高分辨率多光谱成像仪,覆盖13个光谱波段,能在10~60 m的不同分辨率下获取图像,它搭载的多光谱仪能够捕捉不同波长的光,这使其能够计算归一化植被指数(*NDVI*)。*NDVI*是一种衡量地表植被健康度的重要指标[10-11],它通过式(1)计算得出。

$$NDVI=\frac{NIR-Red}{NIR+Red} \tag{1}$$

式中:*NIR*——近红外光波段的反射值;

Red——可见光红色波段的反射值。

*NDVI*的值范围从-1~+1。较高的正值(接近+1)表示植被茂密,较低的值(接近0甚至为负)则表示较少或无植被。

1.2.2 MODIS数据

中分辨率成像光谱仪(moderate resolution imaging spectroradiometer,MODIS)是NASA的一种先进的大型空间遥感仪器。自1999年以来,MODIS提供了连续的数据,用于监测全球的大气、海洋和陆地条件。这些数据涵盖36个光谱波段,用于分析全球气候变化、环境退化、自然灾害和更多环境和生态动态。MODIS的全球覆盖和频繁的观测周期使其成为理解地球系统的宝贵资源。

本研究采用的MOD13Q1是MODIS仪器的一种数据产品,它提供了16 d/次的全球植被指数信息,主要包括归一化植被指数(*NDVI*)和增强型植被指数(*EVI*)。产品的空间分辨率为250 m,适用于从区域到全球尺度的植被监测。MOD13Q1数据集通过分析MODIS卫星捕获的光谱反射率数据来计算植被指数,这些指数反映了植物对光合作用的能力和生物量的多少。由于其更新频率高和空间分辨率较细,MOD13Q1提供了及时且准确的植被状态信息,适用于长期的生态监测和农业生产预测。

1.2.3 县级统计数据

本研究使用的水稻产量数据主要来源于江苏省射阳县人民政府政府公开栏目。2023年全县水稻种植面积为71.2 khm^2,同比增加0.28 khm^2;全县平均产量为9 285 kg/hm^2,同比下降75 kg/hm^2;总产661 kt,同比减少3.2 kt。

1.3 研究方法

本研究采用的机器学习方法是随机森林模型,广泛用于分类和回归任务。该模型通过构建并平均多个决策树(decision tree, DT)的预测结果以实现更高的准确性和稳定性,进而提升模型的泛化能力。随机森林的核心步骤是利用输入数据集构建决策树,本研究中根据输入*NDVI*特征和单产值构建决策树。在决策树的构建过程中,模型从数据集中随机抽取样本和特征,这种方法被称为自助聚合(bootstrap aggregating,或称bagging),见式(2)。模型从原始数据集中有放回地抽取样本,以此构建M个独立训练集并训练相应决策树。在每棵决策树的训练过程中,算法随机选择N个特征中的N'个候选特征用于模型训练。通过这种方式,每棵决策树独立对数据进行预测,而随机森林模型则通过对所有决策树的预测结果进行平均,得出最终产量预测值。

$$Y=\frac{1}{B}\sum_{b=1}^{B}T_b(X,\theta_b) \tag{2}$$

式中:B——树的数量;

T_b——第b个决策树的预测函数;

θ_b——决策树参数;

X——输入特征向量;

Y——输出预测。

随机森林的核心优势在于其对数据中的噪声和异常值具有较高的容忍度,同时避免了决策树常见的过拟合问题。在构建每棵树时,随机森林不仅对样本进行随机抽取,还对特征进行随机选择,这种双重随机性使得模型在多种数据集上都能表现出较好的泛化能力。通过增加树的数量,随机森林可以显著提高预测的准确度,但同时也会增加模型的计算复杂度。此外,随机森林提供内部可视化功能,能够通过分析特征的重要性来解释模型决策过程,增强模型的可解释性。

在实际应用中,随机森林模型的参数调整和优化也非常关键。因此,本研究选择的主要的参数包括树的数量B、每棵树的最大深度、在每个节点处考虑的最大特征数,见图1。这些参数直接影响到模型的训练效

果和运行时间。一般情况下，增加树的数量和特征数量可以提高模型的表现，但过多则可能导致计算资源的大量消耗，因此需要在模型性能和计算效率之间找到合适的平衡点。

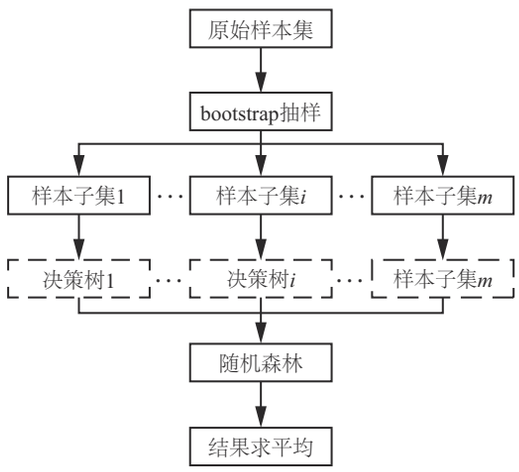


图 1　随机森林模型框架

Figure 1　Random forest modelling framework

1.4　模型评估

本研究选取 3 种广泛认可的评估指标以全面评估模型的准确性和适用性：决定系数（R^2）、均方根误差（*RMSE*）以及相对均方根误差（*RRMSE*），具体详见式（3）～式（5）。

决定系数（R^2）用于评估模型预测值与实际观测值之间的相关性，反映模型对数据变化的解释能力，是衡量模型解释效果的重要指标。*RMSE* 通过直接量化预测误差，成为衡量模型预测准确度的关键指标。*RRMSE* 则评估模型的相对误差，有助于消除数据规模对误差评估的潜在影响。这 3 个指标的综合应用，能够从多个维度全面评估模型的性能。

$$R^2 = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \bar{y}\right)^2} \tag{3}$$

$$RMSE = \sqrt{\left[\frac{1}{n}\sum\left(y_i - \hat{y}_i\right)^2\right]} \tag{4}$$

$$RRMSE = \frac{RMSE}{\bar{y}} \tag{5}$$

式中：y_i——第 i 个数据的真实值；

$\hat{y}_i$——第 i 个数据的预测值；

$\bar{y}_i$——所有数据的均值。

1.5　试验方案

在本研究中，我们构建一个综合数据集，该数据集融合了 Sentinel 2A 和 MODIS 精细处理的月均归一化植被指数（*NDVI*）与中国土地利用/土地覆盖变化（*CNLUCC*）数据，旨在为水稻产量预测提供更准确的县级输入特征。利用这些县级月均 *NDVI* 特征作为输入变量，结合从历史记录中获得的县级水稻产量数据作为响应变量，应用随机森林模型进行模拟。

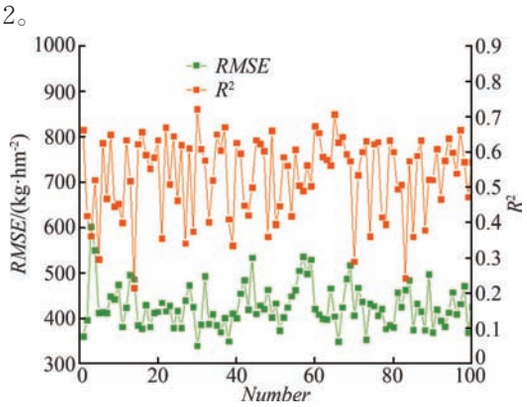
在模型训练阶段，通过选择决策树的数量和树的深度，并执行广泛的交叉验证过程，对模型参数进行了优化。这一步骤不仅确保了模型在未见数据上的泛化能力，也保证了预测结果的准确性。此外，模型的性能经过多个度量标准进行评估，以验证其在复杂农业环境中的实用性。

经过验证的模型被进一步迁移到更细粒度的像素级特征数据中，这一转换使得从较粗糙的县级模拟结果扩展到更为精细的像素级产量估计。此举不仅增强了模型的适用性，也为未来在其他区域应用提供了可行性的评估，展示了随机森林模型在地理空间数据分析中的强大潜力。

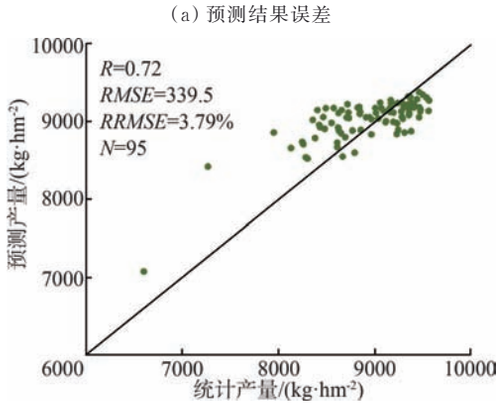
2　结果分析

2.1　模型结果分析

随机森林模型预测水稻产量与统计产量比较见图 2。



（a）预测结果误差



（b）预测结果和统计相关性

图 2　随机森林模型预测水稻产量与统计产量比较

Figure 2　Comparison of Random Forest forecasting results of rice yield

采用网格搜索方法对 RF 模型的超参数进行优化选择。为了确保研究结果的准确性和可靠性，模型共

重复运行100次。经过反复试验,随机森林模型在第30次运行时达到最优性能,*RMSE*最低且R^2最高,见图2(a),模型最优超参数配置详见表1。

表1 RF模型的参数值

Table 1 Parameter values of RF model

特征名称	特征值	最优值
N_estimators	[100, 200, 300, 400]	200
Max_depth	[10, 20, 30, 40]	20
Max_features	[*sqrt*, log2]	*sqrt*

对图2(a)进行分析,结果显示RF模型在整个试验过程中表现较为稳定,其均方根误差(*RMSE*)值大多数在300~600 kg/hm^{2}之间波动,表明模型在产量预测方面具有较高的一致性。使用表现最佳的RF模型进行产量预测,发现模型预测结果与真实值之间的偏差较小,展现较高的拟合度,如图2(b)所示。试验使用95个县级数据为测试集,测试结果显示模型的R达到0.72,*RMSE*低于350 kg/hm^2,*RRMSE*低于4%,充分说明模型预测县级水稻产量与实际产量之间存在高度一致性。

2.2 最优模型可视化

射阳无人农场作为现代农业的前沿实践基地,正积极借助先进的模型技术迈向更高的智能化水平。将模型的内部结构进行可视化。每个节点表示一个决策点,其中包含用于分割数据的条件或判断标准。在这些决策树中,根节点位于顶部,是数据分割的起点。从根节点开始,数据根据设定的条件分流到不同的子节点。这些分支过程一直持续到达叶节点,叶节点通常代表了最终的预测值。

例如,如果*NDVI* 10月份的值大于0.2,则数据流向右子节点,否则流向左子节点。决策树的每一层代表了决策过程中的一个更细致的划分级别,能够观察到模型如何基于输入特征的不同值进行决策路径的选择。此可视化有助于理解随机森林中单个决策树的构造和决策逻辑,同时展示了模型的复杂性和决策深度。通过分析这些树的结构,可以更好地理解模型的预测行为和可能的改进方向。

通过观察决策树的各层结构,从根节点起始的粗粒度判断,到逐步细化的各级子节点分支,农场的智能化管理系统可以准确地根据实时收集的数据进行决策。这种可视化不仅展示了模型本身的复杂性和决策深度,更是让农场技术人员能够深入理解模型的预测行为。比如,通过分析决策树,技术人员可以知晓在何种天气、土壤养分等条件组合下,农作物产量可能会达到最佳状态,或者出现异常波动,进而针对性地对农场的灌溉系统、施肥方案进行改进,让射阳无人农场在精准农业的道路上越走越稳,实现更高的生产效益和资源利用效率。

2.3 水稻全生育期长势监测动态图

在利用卫星采集水稻数据前,需要根据水稻种植区域的地理位置等因素,规划卫星的轨道和确定合适的观测窗口。天气状况良好(避免云层遮挡等情况影响光谱数据获取的准确性),同时还要考虑太阳高度角等因素,因为不同的太阳光照条件下,水稻对光的反射情况会有所不同,一般会选择在太阳高度角适中的时候进行观测,这样获取的反射光谱信息更具代表性和稳定性。

当卫星飞临水稻种植区域上空时,其搭载的高光谱传感器开始工作,从400 nm到2 500 nm左右的各个光谱通道依次收集数据。每个光谱通道就像是一个“过滤器”,只允许特定波长范围的光通过并记录其反射强度。例如,在550 nm左右的绿光波段通道,会记录下水稻田在这个波段下的反射光强度,众多这样的通道数据组合起来,就形成了一个完整的、反映水稻当时光谱特征的数据集,包含了水稻叶片颜色、植株结构、含水量等诸多信息所对应的光谱信息,为后续准确分析水稻生长状况提供了丰富的原始素材。

图3显示2023年5月中旬至5月下旬时育秧移栽阶段大田作物长势状况,小麦处于拔节期下限或孕穗初期偏低水平,反映出植株营养不足或环境压力,此时*NDVI*值:0.2~0.35。



图3 育秧移栽阶段大田作物状况

Figure 3 Field crop condition in seedling raising and transplanting stage

图4显示2023年6月至7月初时分蘖生长期水稻长势状况,此时植被覆盖度不足。梅雨后分蘖盛期因光照不足,分蘖发生缓慢且田间不平衡性大,此时*NDVI*值:0.1~0.2。

图5显示2023年7月中旬至8月初:拔节生长期水稻长势状况,此时植被覆盖度不足,可能处于生长受阻状态。2023年7月盐城降水量达210 mm(常年同期150 mm),且伴有连续5 d以上阴雨(日照时数仅80 h)。台风“杜苏芮”残余环流(7月29日)加剧局部内涝,影响水稻根系发育,此时*NDVI*值:0.15~0.2。

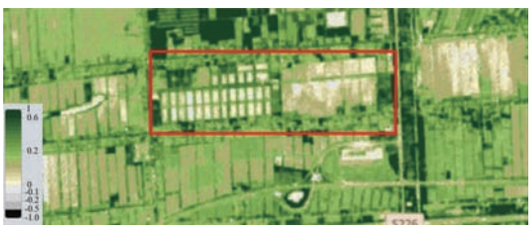


图 4　分蘖生长期水稻长势状况

Figure 4　Rice growth status in tillering period



图 5　拔节生长期水稻长势状况

Figure 5　Growth status of rice in jointing period

图 6 显示 2023 年 8 月中旬至 8 月下旬：孕穗生长期水稻长势总体良好，经历了连续阴雨寡照天气，导致光照不足、干物质积累稍有减少，此时 *NDVI* 值：0.35～0.5。



图 6　孕穗生长期水稻长势状况

Figure 6　Growth status of rice during booting period

图 7 显示 2023 年 9 月初至 9 月中旬：抽穗扬花期水稻长势状况：植被覆盖度和生物量积累存在一定压力，但未达到严重受损水平，此时 *NDVI* 值：0.3～0.45。

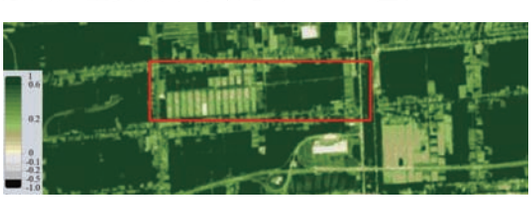


图 7　抽穗扬花期水稻长势状况

Figure 7　Rice growth status at heading and flowering stage



图 8　灌浆期至成熟期水稻长势状况

Figure 8　Rice growth from filling stage to mature stage

图 8 显示 2023 年 9 月下旬至 10 月中旬：灌浆期至成熟期水稻长势状况为植株处于严重受损状态，主要受极端气候、病虫害扩散及管理滞后共同影响，此时 *NDVI* 值：0.15～0.3。

图 9 显示 2023 年 10 月下旬至 11 月上旬，收获期水稻大田状况为已完成收割，裸露地表或秸秆覆盖会导致 *NDVI* 值偏低，此时 *NDVI* 值：0.1～0.2。



图 9　收获期水稻大田状况

Figure 9　Rice planting conditions at harvest stage

2.4　特征重要性分析

图 10 显示 *NDVI*_08、*NDVI*_09 和 *NDVI*_10 具有较高的特征重要性，这表明这些月份的 *NDVI* 数据对水稻产量的预测具有显著的影响。这一发现与作物生长的关键阶段相吻合，因为 8～10 月通常是水稻的拔节孕穗期至成熟期，此阶段作物的生物量增长最为迅速，植被覆盖面积和叶绿素含量达到峰值，因此 *NDVI* 作为反映植被状况的指标，在模型中显示出高重要性。

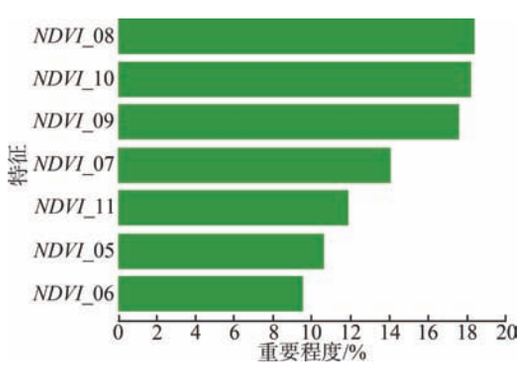


图 10　随机森林模型特征重要性分析

Figure 10　Analysis of the feature importance of Random Forest model

其次，2023 年 5—7 月和 11 月的 *NDVI* 特征重要性较低。5—7 月是作物的生长初期，此时作物处于萌发和幼苗生长阶段，植被覆盖较少，*NDVI* 值相对较低，且此阶段的生物量和作物最终产量的相关性可能较低。11 月接近作物收获期，虽然植被覆盖可能仍然很高，但其对最终产量的影响相对减弱，因为此时作物已经完成了主要的生物量积累。

此外，从模型的角度来看，这种时间相关的特征重要性变化可能还受到气候和管理措施等其他季节性因素的影响。例如，8—10 月可能是某些地区灌溉和施肥等农业管理措施集中的时期，这些措施会直接影响 *NDVI* 的读数和作物的生长状况。

图 11 中，将左侧的现代农业产业园航拍影像与右

侧的产量分布图结合分析,我们可以进一步探索预测现代农业产业园水稻产量的空间分布规律及其背后的潜在因素。

将训练好的最优随机森林模型结合 Sentinel *NDVI* 数据进行的产量预测,结果显示研究区的水稻产量在中间区域较高,而四周区域则相对较低。中间的高产区域可能具有更优质的土壤条件、更均匀的水分分布或更有效的田间管理实践。例如,中间区域可能由于地形的原因获得了更好的水源供应,或者可能接受到更多的太阳光照,这些都是影响作物生长的重要因素。四周产量偏低的区域可能受到多种因素的影响,如边缘效应,其中土壤退化、水分胁迫或管理不足可能更为常见。此外,四周区域可能更受到风害或其他自然因素的影响,导致作物生长不良。

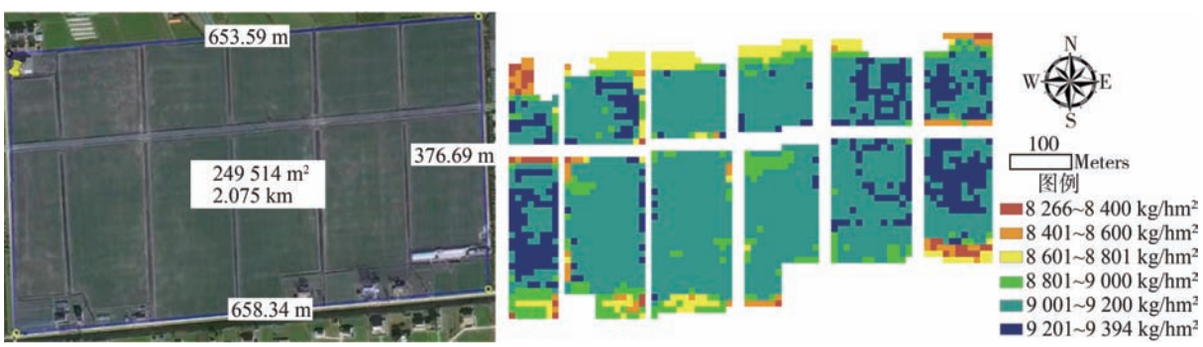


图 11 2023 年江苏射阳无人农场水稻产量分布图

Figure 11 The distribution of rice production from Sheyang unmanned farm of Jiangshu Province in 2023

2.5 产量分析结果

通过统计分析方法对模拟的射阳县现代农业产业园无人农场水稻产量进行深入分析,如图 12 所示。

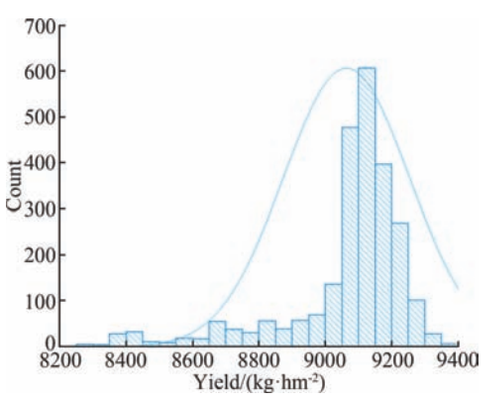


图 12 江苏射阳无人农场 2023 年水稻产量分布直方图

Figure 12 Histogram of rice yield distribution from Sheyang unmanned Farm of Jiangshu Province in 2023

水稻产量数据的分布通过直方图可视化呈现,其中大部分数据集中在 9 000~9 300 kg/hm^2,显示了该现代农业产业园无人农场内水稻产量的主要分布特征。

直方图上方的蓝色曲线表示拟合的正态分布,这进一步突出了数据的集中趋势和分散程度。尽管部分地区的产量低于 9 000 kg/hm^2,但整体产量分布表明了一种明显的中心趋势。通过对这些数据进行均值模拟,我们计算出现代农业产业园无人农场区域水稻的平均预测产量为 9 063 kg/hm^2。此外,为了验证模型的预测准确性,我们将模拟结果与实际收割测量出数据进行了对比,发现两者非常接近,实测数据平均产量为 9 120 kg/hm^2,从而证明了随机森林模型在本研究中的有效应用,并确认了其在农业产量预测方面的实用性。

3 讨论

本研究选取水稻生长的 6 个关键时期进行研究分析。水稻生长前期,由于作物的覆盖率小,土壤、杂草等信息干扰较大,该时期不适合水稻估产。在水稻生长中后期,从抽穗期到成熟期,作物由营养生长向生殖生长过渡,作物群体状况已趋于稳定,该阶段估产效果较好。水稻抽穗期是产量形成的关键期,利用该时期的光谱信息进行估产效果最好。基于植被指数对作物产量进行反演,是一种常见的监测手段。其中归一化植被指数 *NDVI* 使用最为广泛,很多研究认为,*NDVI* 与产量相关,进行产量估测最普遍的方式是在 *NDVI* 与作物产量之间建立经验模型,该方法操作简单,便于推广。在植被覆盖度较低时,*NDVI* 可以较为准确地估测作物长势和生物量,当覆盖度较高时,准确度会下降。

研究基于关键生育期 NDVI 植被指数估测水稻产量,结果表明,单一植被指数建立的线性模型在面对作物覆盖不均匀的区域时,其精度和灵敏度降低,不能全面准确地反应产量信息,而采用随机森林模型实现了从县级到像素级的水稻产量预测相结合,在一定程度上提高了产量估算的精度。

4 结论

本研究利用 2023 年的 Sentinel 和 MODIS *NDVI*

数据以及县级产量信息，采用随机森林模型实现了从县级到像素级的水稻产量预测，并分析了不同特征对模型学习机制的影响。研究的主要结论有：

1）2023年8—10月的*NDVI*数据与水稻产量之间存在显著相关性。这一时期是作物生长周期的关键阶段，植被覆盖度和生长状况对产量预测至关重要。随机森林模型在预测县级水稻产量方面表现出色，具有较低的均方根误差，为339.5 kg/hm^2。

2）通过对射阳农场的水稻产量空间分布进行分析，发现该区域水稻产量呈现明显的空间异质性。边缘地区的产量较低，而中心区域的产量较高，主要产量范围集中在9 000～9 300 kg/hm^2。

这些成果不仅提升了我们对机器学习和植被数据在作物产量预测应用中的理解，还为提高预测模型的准确性、制定科学的粮食生产策略以及促进农业可持续发展提供了重要的理论支持。在未来展望中，研究可进一步深化的方面可能包括：

1）模型优化：通过调整随机森林模型的参数，或尝试其他先进的机器学习算法，比如支持向量机、深度学习网络等，来进一步提升预测的准确性。

2）数据融合：整合更多类型的遥感数据（如土壤湿度、气象数据、高时空分辨率的植被指数等）和地面观测数据，以实现更全面的产量预测。

3）时空分析：进行时间序列分析以及更细致的空间分析，以揭示作物生长的时空动态变化规律，并识别影响产量的关键因素。

4）应用拓展：将研究成果应用于其他作物的产量预测，或扩大到更大的地理范围和不同的气候类型区域。

5）策略制定：基于高精度的产量预测结果，协助政府和农业部门制定更为精准和高效的粮食管理策略，确保国家粮食安全。

6）灾害应对：利用预测模型辅助在自然灾害（如干旱、洪水）发生前的粮食产量评估，以提前做好灾害应对和缓解策略。

7）实时监控系统开发：建立实时的农作物生长监控和产量预测系统，为农业生产者提供即时的数据服务和决策支持。

通过这些未来工作，研究不仅可以在科学上取得进展，还能更好地服务于实际的农业生产和粮食安全管理。

参 考 文 献

[1] 张佳华. 遥感技术在作物生长模式及农业气象预报中的应用研究[M]. 北京：中国农业科学技术出版社，2021.

[2] WANG J, CHEN J, ZHANG J, et al. Consistency and uncertainty of remote sensing-based approaches for regional yield gap estimation: A comprehensive assessment of process-based and data-driven models [J]. Field Crops Research, 2023, 302: 109088.

[3] 张佳华，张健南，姚凤梅，等. 开放式增温对东北稻田生态系统作物生长与产量的影响[J]. 生态学杂志，2013，32(1)：15－21.
ZHANG Jiahua, ZHANG Jiannan, YAO Fengmei, et al. Effects of free air temperature increasing on the rice growth and grain yield in northeast China [J]. Chinese Journal of Ecology, 2013, 32(1): 15－21.

[4] 赵春江. 农业遥感研究与应用进展 [J]. 农业机械学报，2014，45(12)：277－293.
ZHAO Chunjiang. Advances of research and application in remote sensing for agriculture [J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(12): 277－293.

[5] 聂鹏程，钱程，覃锐苗，等. 天空地一体化信息感知与融合技术发展现状与趋势[J]. 智能化农业装备学报(中英文)，2023，4(2)：1－11.
NIE Pengcheng, QIAN Cheng, QIN Ruimiao, et al. Development status and trends of space-air-ground integrated information sensing and fusion technology [J]. Journal of Intelligent Agricultural Mechanization, 2023, 4(2): 1－11.

[6] BAI Y, ZHANG J, ZHANG S, et al. A remote sensing-based two-leaf canopy conductance model: Global optimization and applications in modeling gross primary productivity and evapotranspiration of crops [J]. Remote Sensing of Environment, 2018, 215: 411－437.

[7] ZHANG S, ZHANG J, BAI J, et al. Evaluation and improvement of the daily boreal ecosystem productivity simulator in simulating gross primary productivity at 41 flux sites across Europe [J]. Ecological Modelling, 2018, 368: 205－232.

[8] CAO D, FENG J, BAI L, et al. Delineating the rice crop activities in northeastern China through regional parametric synthesis using satellite remote sensing time-series data from 2000 to 2015 [J]. Journal of Integrative Agriculture. 2021, 20(2): 424－437.

[9] WANG P J, SUN R, ZHANG J H, et al. Yield estimation of winter wheat in the north China Plain using the remote-sensing-photosynthesis-yield estimation for crops (RS-P-YEC) model [J]. International Journal of Remote Sensing, 2011, 32: 6335－6348.

[10] 刘珊珊，牛超杰，边琳，等. 基于NDVI的水稻产量遥感估测[J]. 江苏农业科学，2019，47(3)：193－198.

[11] 许童羽，洪雪，陈春玲，等. 基于冠层NDVI数据的北方粳稻产量模型研究[J]. 浙江农业学报，2016，28(10)：

1790－1795.
XU Tongyu, HONG Xue, CHEN Chunling, et al. Study on northern *japonica* rice yield model based on canopy date of NDVI [J]. Acta Agriculturae Zhejiangensis, 2016, 28(10): 1790－1795.
[12] YAO F, TANG Y, WANG P, et al. Estimation of maize yield by using a process-based model and remote sensing data in the northeast China Plain [J]. Physics and Chemistry of the Earth, Parts A/B/C, 2015, 87: 142－152.
[13] ZHANG S, BAI Y, ZHANG J H. Remote sensing-based quantification of the summer maize yield gap induced by suboptimum sowing dates over north China Plain [J]. Remote Sensing, 2021, 13: 3582.
[14] PRODHAN F A, ZHANG J, SHARMA T P P, et al. Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach [J]. Science of the Total Environment, 2022, 807: 151029
[15] 李世隆, 许辰一, 王楠, 等. 基于BWO－ELM的水稻氮素无人机高光谱反演研究[J]. 智能化农业装备学报(中英文), 2024, 5(3): 14－21.
LI Shilong, XU Chenyi, WANG Nan, et al. Research on rice nitrogen unmanned aerial vehicle hyperspectral inversion based on BWO-ELM [J]. Journal of Intelligent Agricultural Mechanization, 2024, 5(3): 14－21
[16] AGHIGHI H, AZADBAKHT M, ASHOURLOO D, et al. Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11, 4563－4577.
[17] YANG X, ZHANG J, YANG S, et al. Modelling the crop yield gap with a remote sensing-based process model: A case study of winter wheat in the north China Plain [J]. Journal of Integrative Agriculture, 2023, 22(10): 2993－3005.

Rice yield prediction of unmanned farm based on random forest and remote sensing vegetation index

WANG Jun[1], WU Zhenwei[1], JIANG Hai[1], KE Juan[1], FOYEZ Ahmed Prodhan[2]

(1. *Jiangsu Lizhuo Information Technology Co., Ltd., Changzhou* 213000, *China*; 2. *Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur* 1706, *Bangladesh*)

Abstract: In order to ensure national food security and promote sustainable agricultural development, it is crucial to accurately predict regional rice yield by formulating effective food strategies. While advancemnets have been achieved by integrating remote sensing data with machine learning technology for yield prediction, a more in-depth analysis of the machine learning model mechanism and the application of high spatio-temporal resolution data remains necessary. This study utilized Sentinel and MODIS Normalized Difference Vegetation Index (*NDVI*) data along with county-level yield statistics from 2023 to implement a Random Forest (RF) model for predicting yield from county level to pixel level with case study of Sheyang unmanned farm in Jiangsu Province, and investigated the impact of different features on the model's learning mechanism. The results demonstrated that there was a significant correlation between *NDVI* data and rice yield during the period of August to October 2023, which is a critical period in the crop phenological cycle. The RF model effectively predicted rice yield at the county level, with a Root Mean Square Error (*RMSE*) value of 339.5 kg /hm^2. Furthermore, spatial distribution mapping of the predicted yields within Sheyang County for 2023 revealed significant heterogeneity, indicating lower yields in marginal regions and higher yields in central areas ranging from 9 000 to 9 300 kg/hm^2. This study not only enhances understanding of machine learning application and vegetation data in yield prediction but also provides theoretical support for improving model accuracy and formulating scientific food production strategies.

Keywords: Random Forest; machine learning; remote sensing; vegetation index; yield forecast; growth stage

Foundation items: National Natural Science Foundation of China (42071425); Jiangsu Changzhou Zhonglou District Challenge-Response System Project (JBGS2023011)

First Author: WANG Jun, E-mail: wangjunrc@qq.com