

# 规模化农田无人机多光谱数据质量控制方法与小麦长势监测轻量化模型构建

冯庆贺 王紫晴 王春晓

青岛国测海遥信息技术有限公司 山东 青岛 266000

**摘要:** 针对规模化农田小麦长势监测中无人机多光谱数据易受光照、拼接、配准和特征冗余影响的问题, 本文以河南省武陟县冬小麦试验田为对象, 构建面向 DJI Mavic 3M 数据的质控与轻量化反演框架。研究建立飞行前、飞行中、处理后和建模前四级质量标签, 提取 NDVI、NDRE、GNDVI、OSAVI 等指数和 ROI 统计特征, 并通过相关性筛选、特征压缩和 LightGBM 参数约束建模。结果表明, LightGBM 对 LAI、AGB、LNC 和 SPAD 的验证集  $R^2$  分别为 0.84、0.83、0.79 和 0.81, 模型体积极约 142 KB, 单样本推理时间平均为 5 ms, 兼具精度与部署效率。

**关键词:** 无人机多光谱; 数据质量控制; 小麦长势; LightGBM; 轻量化模型

## 1 引言

无人机多光谱遥感兼具厘米级空间分辨率和灵活重复观测能力, 红边与近红外波段能够反映冠层叶绿素、氮素吸收和结构变化, 适合用于小麦长势监测<sup>[1]</sup>。但在实际应用中, 云影、曝光漂移、风致模糊、波段错位、正射拼接缝和样方边界偏移等因素, 都会造成植被指数在不同日期或地块间出现波动。

现有研究多关注模型精度, 对航测质控链条和部署成本描述不足, 导致田间复用稳定性下降<sup>[2]</sup>。本文将质量控制置于建模之前, 建立 Mavic 3M 多光谱数据质控流程, 并通过特征压缩、参数约束和效率评价构建面向 LAI、AGB、LNC 和 SPAD 的轻量机器学习模型<sup>[3]</sup>, 突出“数据可用性判读—特征压缩—轻量部署评价”的连续技术链。

## 2 研究区与数据来源

研究区位于河南省焦作市武陟县连片冬小麦试验田, 地势平坦、田块规整, 适配无人机航线规划和多时相观测。地面指标包括 LAI、AGB、LNC 和 SPAD; 试验设置郑麦 136、百农 4199、周麦 36 三个冬小麦品种, 播量为 180、225、270  $\text{kg} \cdot \text{hm}^{-2}$ , 纯氮为 0、120、180、240  $\text{kg} \cdot \text{hm}^{-2}$ , 覆盖低、中、高长势梯度<sup>[4]</sup>。

采用 DJI Mavic 3M 多光谱无人机作为航测平台, 搭载 RGB 相机及绿光 (约 560nm)、红光 (约 650nm)、红边 (约 730nm)、近红外 (约 860nm) 4 个光谱通道, 可用于构建 NDVI、NDRE、GNDVI 等植被指数。航测在晴朗或均匀薄云、太阳高度角稳定、低风速时段开展, 每次记录航高、航速、航向/旁向重叠度、气象条件、相机状态及辐射板影像, 作为质量判读依据。

地面采样与无人机航测原则上同日完成; 样方边界

采用 RTK 或固定地物辅助定位, 在正射影像中裁剪统一 ROI, 对畦埂、裸土、阴影等混合像元先做掩膜或标记低质量等级。数据设定原始样方 132 个, 质控剔除 12 个后保留 120 个, 其中拔节、孕穗、灌浆期各 40 个; 训练集和验证集按独立田块分组, 比例为 84:36。

## 3 多光谱数据质量控制方法

本文将质量控制划分为飞行前、飞行中、处理后和建模前四个阶段。飞行前控制包括检查镜头洁净度、电池状态、云台稳定性、辐射校正板、光照传感器和相机曝光模式; 同时固定航高、航向重叠度、旁向重叠度和飞行速度, 避免不同架次之间因空间分辨率和视角差异造成反射率不可比。针对本研究多时相生育期对比的核心需求, 航测时严格保持太阳高度角一致, 并将作业时段控制在当地时间 10:00–14:00。

飞行中控制强调现场记录和异常架次识别。云影移动、阵风、相机曝光突变或航线中断, 会在正射影像中形成条带、模糊或拼接断裂。对这类情况应在飞行日志中设置质量标记。若关键样方受到明显云影覆盖或航线缺失影响, 应优先重飞; 若无法重飞, 标记为低质量样本, 酌情剔除或单独划为异常集。

处理后控制包括辐射校正、波段配准、正射拼接、坐标统一和样方 ROI 检查。Mavic 3M 数据可通过辐射板和光照信息进行相对反射率校正, 校正后检查同质地物的反射率稳定性。波段配准重点关注红边和近红外通道, 因为冠层边界、畦沟和标志物处的错位会显著影响指数值。正射拼接完成后, 利用道路、田埂或标志点检查几何偏移。

建模前对 ROI 内裸土、阴影、田埂和边界混合像元进行掩膜, 对病虫害斑块等真实长势异常单独标记, 采

用箱线图、马氏距离或初始模型残差识别异常样本；保留质量标签而不是简单删除可疑样本。轻微阴影样本可标记为低质量样本；边界明显错位或地面指标日期不匹

配的样本应剔除。多时相航测还应比较稳定地物和同一小区反射率序列，识别整体漂移，并记录具体原因，便于后续追溯。如图1所示：

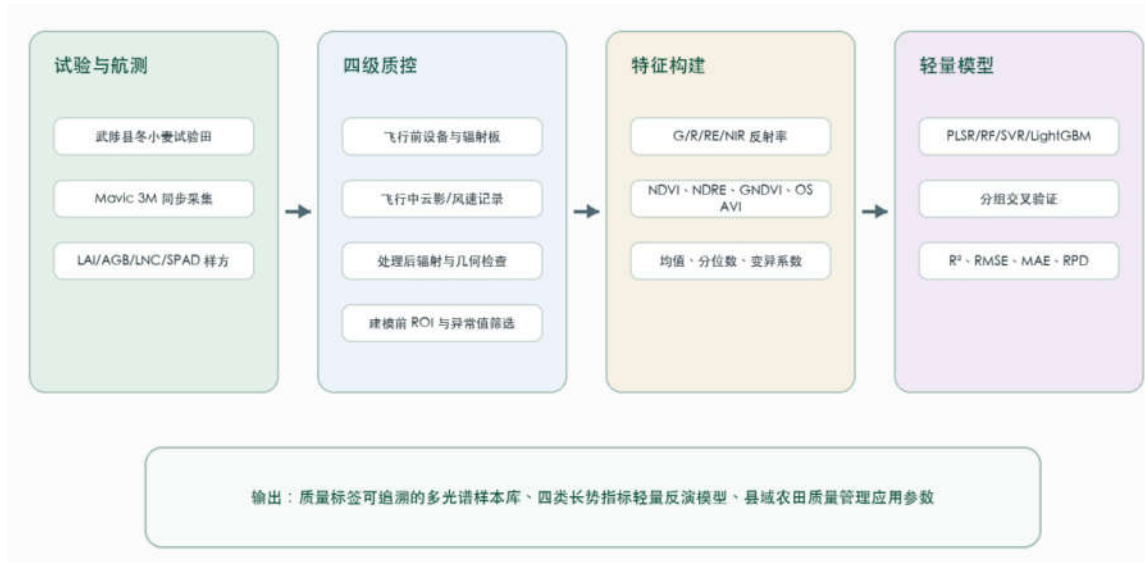


图1 Mavic 3M 多光谱质量控制与长势建模流程

#### 4 特征构建与轻量化模型

特征构建以绿、红、红边和近红外反射率为基础，计算 NDVI、NDRE、GNDVI、OSAVI、RVI、红边比值指数，并提取 ROI 均值、中位数、标准差、分位数和变异系数。LAI 与 AGB 主要受冠层覆盖度、叶片层叠和近红外散射影响，因此近红外均值、NDVI、OSAVI 和高分位反射率响应较强；LNC 与 SPAD 更接近叶片生理和色素状态，红边与近红外组合指数更敏感<sup>[5]</sup>。

为避免特征冗余，先以  $|r| > 0.85$  剔除高度相关且物理意义重复的指数，再在训练集内交叉验证筛选稳定变量。其中初始特征 48 个，相关性筛选后 18 个，最终保留 NDVI、NDRE、GNDVI、OSAVI、近红外均值、红边均值、红边比值、近红外 P75、NDRE P75 和变异系

数 10 个特征。LightGBM 设置为  $n\_estimators = 120$ 、 $max\_depth = 3$ 、 $num\_leaves = 8$ 、 $learning\_rate = 0.05$ 、 $min\_child\_samples = 8$ ；PLSR 主成分数为 5，RF 为  $n\_estimators = 300$ 、 $max\_depth = 6$ ，SVR 采用 RBF 核， $C = 10$ 、 $gamma = 0.1$ 、 $epsilon = 0.1$ 。

数据划分采用分组抽样方式，按独立田块或处理小区划分训练集与验证集，同一田块样方仅进入其中一组，并保持不同生育期和长势水平分布均衡，避免空间自相关造成精度高估。模型评价以验证集  $R^2$  为主要指标；RMSE、MAE、RPD、模型体积和推理时间作为补充指标，用于刻画误差、估算能力、运行开销和轻量化部署可行性。如图2所示：

| 环节      | 输入/对象                                  | 主要质量控制内容                        | 模型与输出   |
|---------|--|---------------------------------|---|
| 数据获取    | Mavic 3M 多光谱影像；<br>LAI、AGB、LNC、SPAD 样方 | 统一航高、重叠度和太阳高度角；记录云量、风速、曝光和辐射板影像 | 形成带日志的航测样本  |
| 影像处理    | 绿、红、红边、近红外反射率与正射影像                     | 辐射校正、波段配准、拼接检查、坐标统一、ROI 裁剪和掩膜   | 输出标准化反射率与样方统计量  |
| 特征压缩建模  | NDVI、NDRE、GNDVI、OSAVI、RVI、分位数和变异系数     | 48 个特征压缩至 10 个；分组抽样防止空间泄漏       | 比较 PLSR、RF、SVR、LightGBM；约束 LightGBM 参数                |
| 模型验证与评价 | 四类长势指标验证结果；模型体积与推理时间                   | 验证精度、稳定性与轻量化部署可行性               | LightGBM $R^2$ : LAI 0.84、AGB 0.83、LNC 0.79、SPAD 0.81 |

图2 数据质量控制、特征建模与模型精度综合说明

## 5 结果与讨论

结果表明,云影样方、边界混合像元和拼接缝附近样方会增大误差,尤其影响 LAI 和 AGB。原始 132 个样方中,质控剔除云影覆盖样方 4 个、边界错位样方 5 个、地面指标日期不匹配样方 3 个,另有 14 个轻微阴影或拼接边缘样方标记为低质量样本并保留质量标签。经质控后,LightGBM 对 LAI、AGB、LNC 和 SPAD 的  $R^2$  由 0.76、0.75、0.71、0.73 提高至 0.84、0.83、0.79、0.81, RMSE 分别降低 26.2%、24.7%、21.6% 和 21.4%。

模型比较显示,PLSR 表达清晰、参数少,但对冠

层饱和后的非线性关系刻画不足;随机森林在 AGB 和 SPAD 估算中较稳定,但解释受树结构影响;SVR 对小样本非线性关系具有拟合潜力,但核函数和惩罚参数会影响泛化;LightGBM 在 LAI、AGB、LNC 和 SPAD 上分别取得 0.84、0.83、0.79 和 0.81 的验证集  $R^2$ ,平均  $R^2$  为 0.818,模型体积约 142 KB,单样本为单个样方的 10 维输入特征向量,测试环境为 Intel Core i5-10400F CPU、8GB 内存的普通办公计算机,单线程运行下平均推理时间为 5 ms,精度和部署适应性较好。如图3所示:

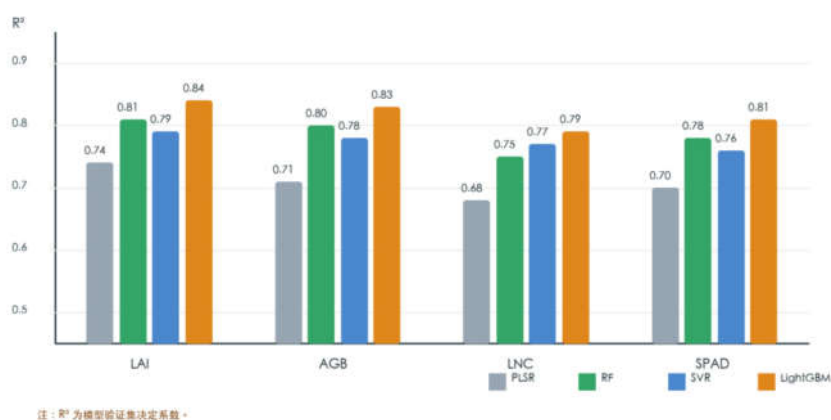


图3 四类长势指标建模精度与特征响应

特征响应方面,NDRE、红边比值指数和近红外分位数对 LNC 与 SPAD 的贡献较大,NDVI、OSAVI 与近红外均值对 LAI 和 AGB 更有效。其原因在于拔节至灌浆期冠层逐渐闭合,红光波段易饱和,而红边区域仍能反映叶绿素和氮素变化;近红外则更多体现冠层结构和生物量积累。工程应用不宜仅依赖单一 NDVI,而应结合红边和近红外信息,以提高不同生育期和管理处理下的稳定性。

在县域推广场景下,模型输出可转化为分区管理信息。将 LAI 与 AGB 偏低区域作为苗情补救和灌溉检查重点,将 LNC 与 SPAD 偏低区域作为氮素诊断依据。效率评价显示,特征数量由 48 个压缩至 10 个,降幅为 79.2%;LightGBM 模型体积约 142 KB,单样本推理时间平均为 5 ms,可满足普通工作站或县域农技服务平台批量处理需求。

## 6 结论

本文构建了适用于 DJI Mavic 3M 多光谱数据的四级质量控制方法,并提出植被指数、样方统计特征与轻量机器学习模型相结合的技术框架。结果表明,红边和近红外特征能够支撑 LAI、AGB、LNC 和 SPAD 估算,LightGBM 在特征压缩和参数约束后兼顾精度与效率。该

方法可为武陟县小麦试验田及类似平原农区内的长势诊断、田间质量管理和变量施肥决策提供参考,并有助于形成可复核的无人机农情监测作业规范。

## 参考文献

- [1]齐浩,孙海芳,吕亮杰,等.基于无人机多光谱信息与纹理特征融合的小麦叶面积指数估测[J].农业机械学报,2025,56(3):334-344.
- [2]王文辉,田超,鲁圣权,等.基于无人机多光谱影像和冠层形态的冬小麦叶绿素含量监测研究[J].麦类作物学报,2026,46(3):404-412.
- [3]Zu J, Yang H, Wang J, Cai W, Yang Y. Inversion of winter wheat leaf area index from UAV multispectral images: classical vs. deep learning approaches[J]. Frontiers in Plant Science, 2024, 15: 1367828.
- [4]朱永基,陶新宇,陈小芳,等.基于无人机多光谱影像植被指数与纹理特征的冬小麦地上部生物量估算[J].浙江农业学报,2023,35(12):2966-2976.
- [5]Su X, Nian Y, Shaghaleh H, et al. Winter wheat SPAD estimation based on UAV multispectral imagery[J]. Frontiers in Plant Science, 2024, 15: 1404238.