

无人机高光谱条件下多预处理与特征筛选组合的玉米冠层氮含量反演模型优化研究

贾煜辉¹, 田丽^{1*}, 衣淑娟², 孔浩全¹, 刘勇智³

(1. 黑龙江八一农垦大学 信息与电气工程学院, 黑龙江 大庆 163000; 2. 黑龙江八一农垦大学 工程学院, 黑龙江 大庆 163000; 3. 大庆油田信息技术公司智能安防项目经理部, 黑龙江 大庆 163000)

摘要: 该研究以玉米为研究对象, 针对寒地种植条件下氮素含量监测精度不足的问题, 通过对多种预处理方式进行比较后, 采用小波变换(DWT)与标准正态变量变换(SNV)联合预处理方法, 系统比较了竞争性自适应重加权算法(CARS)、无信息变量消除算法(UVE)、连续投影算法(SPA)、遗传算法(GA)和遗传-连续投影算法(GA-SPA)等特征筛选算法, 并结合偏最小二乘回归(PLSR)、支持向量回归(SVR)、核极限学习机(KELM)、深度极限学习机(DELM)、极端梯度提升(XGBoost)、一维卷积神经网络(1D-CNN)等多种回归模型, 对玉米氮含量的高光谱反演性能进行了综合评估。结果表明, 全波段建模虽在训练集上表现出较高拟合度, 但测试集预测精度较低, 存在明显过拟合现象; 引入特征筛选后, 模型性能显著提升, 其中GA-SPA算法在特征提取与冗余抑制方面表现最佳。以GA-SPA筛选特征并结合DELM模型的预测效果最优, 测试集 R^2 达0.8111, RMSE为0.4930, 显著优于其他组合; 同时, GA-SPA在PLSR、KELM和SVR模型中亦保持了较高的预测稳定性。研究结果表明, 基于DWT+SNV预处理与GA-SPA特征优化的DELM建模框架能够有效提升玉米氮含量反演的精度与泛化能力, 为寒地玉米的养分监测和精准施肥提供了可靠的技术支撑。

关键词: 玉米; 氮含量; 高光谱技术; GA-SPA; DELM

中图分类号: O657.3; S513 **文献标识码:** A **文章编号:** 1004-4957(2026)04-0752-09

Optimizing Maize Canopy Nitrogen Inversion Models Using Multi-preprocessing and Feature-selection Combinations Under UAV Hyperspectral Data

JIA Yu-hui¹, TIAN Li^{1*}, YI Shu-juan², KONG Hao-quan¹, LIU Yong-zhi³

(1. College of Electrical and Information Engineering, Heilongjiang Bayi Agricultural University, Daqing 163000, China; 2. College of Engineering, Heilongjiang Bayi Agricultural University, Daqing 163000, China; 3. Intelligent Security Project Department, Daqing Oilfield Information Technology Company, Daqing 163000, China)

Abstract: This study focuses on maize and addresses the problem of insufficient accuracy in nitrogen content monitoring under cold-region cultivation conditions. After comparing multiple preprocessing methods, the combined use of discrete wavelet transform (DWT) and standard normal variate (SNV) was adopted as the preprocessing approach, and several feature selection algorithms including competitive adaptive reweighted sampling (CARS), uninformative variable elimination (UVE), successive projections algorithm (SPA), genetic algorithm (GA), and genetic algorithm-successive projections algorithm (GA-SPA) were systematically compared. Combined with multiple regression models such as partial least squares regression (PLSR), support vector regression (SVR), kernel extreme learning machine (KELM), deep extreme learning machine (DELM), extreme gradient boosting (XGBoost), and one-dimensional convolutional neural network (1D-CNN), the hyperspectral inversion performance for maize nitrogen content was comprehensively evaluated. The results showed that full-band modeling achieved high fitting accuracy in the training set but exhibited low predictive accuracy in the test set, indicating a clear overfitting issue. After introducing feature selection, model

收稿日期: 2025-10-10; 修回日期: 2026-01-14

基金项目: 国家重点研发项目(2023YFD1501005-06); 黑龙江省省属高等学校项目(ZRCPY2020)

* 通讯作者: 田丽, 博士, 副教授, 研究方向: 农业智能化与图像识别, E-mail: tianli19781015@163.com

网络首发日期: 2026-02-14

performance was significantly improved, with the GA-SPA algorithm performing best in feature extraction and redundancy suppression. The combination of GA-SPA and the DELM model achieved the highest prediction accuracy, with a test-set R^2 of 0.811 1 and RMSE of 0.493 0, outperforming other combinations. Moreover, GA-SPA also maintained high prediction stability in PLSR, KELM, and SVR models. The findings demonstrate that the DELM modeling framework based on DWT+SNV preprocessing and GA-SPA feature optimization can effectively enhance the accuracy and generalization ability of maize nitrogen content estimation, providing reliable technical support for nutrient monitoring and precision fertilization in cold-region maize cultivation.

Key words: maize; nitrogen content; hyperspectral technology; GA-SPA; DELM

玉米是我国重要的粮食作物,在保障我国粮食安全方面发挥着至关重要的作用,是重要的饲料来源^[1]。内蒙古、吉林、辽宁和黑龙江是主要的玉米生产省份,气候条件较为有利^[2]。然而,在中国北方的玉米种植中,由于缺乏科学的施肥知识和技术,普遍存在氮肥过量施用的问题,导致肥料利用效率低下和环境污染加剧^[3]。因此,在区域尺度上精准监测玉米氮素营养动态,优化氮肥管理策略,对实现寒地玉米高产、资源高效与生态可持续协同发展具有重要意义。

传统田间尺度的作物氮素诊断方法,如凯氏定氮法,需要大量的样本采集和复杂的处理步骤,耗时且成本高,限制了其在大规模农业中的应用^[4]。高光谱技术因其非破坏性、快速性和高灵敏度等优点,已广泛应用于农作物氮素含量等农学参数检测中,如王浩宇等^[5]以茄子为研究对象,采用多元散射校正(MSC)预处理方法,分别使用高光谱卷积自编码器(HSI-CAE)、竞争性自适应重加权算法(CARS)、RF特征提取算法,建立偏最小二乘回归(PLSR)模型预测叶片氮含量。结果表明:HSI-CAE特征提取算法测试集决定系数 R^2 达0.91,预测均方根误差(RMSEP)为 $3.12 \text{ mg} \cdot \text{g}^{-1}$,显著优于CARS、RF算法。匡润等^[6]以金钗石斛为对象,使用CARS筛选特征波段后,集成Bagging模型,结果表明,CARS-Bagging仅用9.3%的波段即可实现 $R^2=0.953$ 、 $\text{RMSEP}=0.083$,与全波段RF无显著差异($p>0.05$),建模耗时缩短66.7%,为田间精准施肥提供了轻量光谱方案。刘静等^[7]以辣椒叶片为研究对象,对原始光谱数据进行MSC、平滑滤波(SG)、一阶导数(FD)预处理,结合连续投影算法(SPA)、CARS特征筛选算法构建PLSR、RF模型。结果表明:SG平滑-连续投影-径向基函数神经网络(SG-SPA-RBFNN)组合模型反演精度最佳,验证结果 $R^2=0.98$,均方根误差(RMSE)=1.21,残差预测偏差(RPD)=3.08,为辣椒生长的精准管理和变量施肥工作提供了技术支撑。Hou等^[8]以苹果冠层叶片为研究对象,通过高光谱数据结合CARS-PLS特征提取和霜冰算法-最小二乘支持向量机(RIME-LSSVM)模型,实现了苹果叶氮含量的高精度反演($R^2=0.964$, $\text{RMSE}=0.052$),验证了高光谱技术在果树冠层氮含量预测中的可行性。

研究表明数据预处理与特征筛选在提升建模精度与泛化能力方面具有关键作用。常见的光谱预处理方法包括MSC^[9]、标准正态变量变换(SNV)^[10]、FD^[11]、SG^[12]以及小波变换(DWT)^[13]等,这些方法能够有效削弱光谱噪声、去除基线漂移并突出光谱的特征信息。在特征筛选方面,CARS^[14]、SPA^[15]、无信息变量消除(UVE)^[16]、遗传算法(GA)^[17]等算法广泛应用于压缩高光谱数据维度,减少冗余波段,提高模型计算效率与稳健性。

尽管已有研究证明预处理与特征筛选对模型精度具有显著影响,但目前多数工作仅聚焦于少量预处理方法或单一特征筛选策略,缺乏对预处理-特征筛选-建模的系统化组合评估;同时,不同特征筛选方法在不同模型中的适配性、传统模型与深度模型在无人机(UAV)高光谱条件下的性能差异仍未得到充分验证。此外,智能优化算法与传统波段选择方法的融合应用也相对有限,对于其是否能够在UAV高光谱农田场景中提升模型稳定性与泛化能力仍缺乏系统研究。因此,有必要构建覆盖多预处理、多特征选择、多建模算法的系统比较框架,以明确最优组合策略,并为UAV高光谱氮含量反演提供可推广的方法体系。

为弥补上述不足,本研究在确定DWT+SNV为最优预处理方法的基础上,系统比较CARS、UVE、SPA、GA、GA-SPA五类特征筛选方法,并结合PLSR、支持向量回归(SVR)、核极限学习机(KELM)、深度极限学习机(DELM)和一维卷积神经网络(1D-CNN)等线性与非线性模型,形成系统建模框架,旨

在探讨不同特征筛选与模型组合的适配性, 筛选出适用于 UAV 高光谱条件下玉米冠层氮含量预测的最优建模策略, 实现方法上的系统性创新和应用上的可推广性。

1 实验部分

1.1 试验地概况

试验地点位于黑龙江八一农垦大学安达农业科技园区(东经 $125^{\circ}21'4''$, 北纬 $46^{\circ}24'11''$), 园区规划面积 560 亩, 常年开展玉米规模化种植。试验区面积为 $2\ 800\ \text{m}^2$ 。图 1 为试验地点区位图, 试验地主要特征为: 冬季寒冷、干燥, 时间长达 200 天以上, 年平均气温为 $3.2\ ^{\circ}\text{C}$ 。年平均日照时数为 $2\ 659\ \text{h}$, 符合寒地要求。0~20 cm 土壤基础理化测定结果显示: 碱解氮含量为 $103.1\ \text{mg}\cdot\text{kg}^{-1}$, 有效磷含量为 $26.5\ \text{mg}\cdot\text{kg}^{-1}$, 有效钾含量为 $175\ \text{mg}\cdot\text{kg}^{-1}$; 同时, 土壤有机质含量为 $15.3\ \text{g}\cdot\text{kg}^{-1}$, pH 值为 7.86。试验于 2024 年 5 月实施, 数据采集时间为 2024 年 8 月。

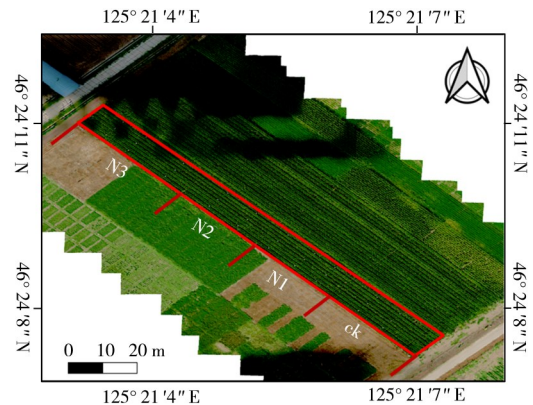


图 1 试验地点区位图

Fig. 1 Experimental site location map

1.2 试验材料与施氮处理设计

本研究选用玉米品种“金博士 825”作为供试作物, 所用肥料均为市售化肥, 包括尿素(含 N 46.4%)、磷酸二铵(含 P 48%)和氯化钾(含 K 60%)。试验共设置 4 个氮素施用水水平: 常规施氮量 CK ($225\ \text{kg}/\text{hm}^2$)、80%施氮量 N1 ($180\ \text{kg}/\text{hm}^2$)、70%施氮量 N2 ($158\ \text{kg}/\text{hm}^2$)和 60%施氮量 N3 ($135\ \text{kg}/\text{hm}^2$)。玉米播种前统一施入基肥, 包括 30%的尿素、全部的磷酸二铵和氯化钾, 并在拔节期和抽雄期分别追施 60%和 10%的氮肥。玉米定植前采用大田管道水肥一体化方式灌施基肥, 并完成整地做垄处理, 试验总面积为 $2\ 800\ \text{m}^2$ 。

1.3 测定项目与方法

1.3.1 区域氮含量值测定 于 2024 年 8 月 16 日在寒地玉米的灌浆期使用 TYS-4N 氮含量测定仪进行玉米冠层叶片氮含量值测量。实验区域共计包含十垄玉米种植条带, 区域内含有 4 个氮含量梯度小区。将每一垄玉米均匀划分出 20 个采样区域, 十垄玉米种植区共划分出 200 个采样区域(其中: CK 小区划分出 40 个采样区域、N1 小区划分出 60 个采样区域、N2 小区划分出 40 个采样区域、N3 小区划分出 60 个采样区域)。每垄种有两排玉米, 在取样区域的两排玉米上方不同方位随机选取共计 4 片健康且未被遮挡的顶端叶片作为试验叶片, 分为上中下 3 个区域, 用测定仪夹取叶片进行测量, 将每个叶片 3 个测量结果的平均值作为该叶片的氮含量, 分别计算 4 片叶片的氮含量值后取平均值作为该区域的玉米叶片冠层氮含量值, 为后续玉米冠层高光谱数据建立感兴趣区域(Region of interest, ROI), 提取测量区域光谱反射率提供支持。玉米氮含量样本的统计分析结果表明: 共采集 200 个样本, 氮含量范围为 $14.6\sim 21.1\ \text{g}/\text{kg}$, 平均值为 $18.386\ \text{g}/\text{kg}$, 标准差为 $1.198\ \text{g}/\text{kg}$, 变异系数为 6.513%, 说明样本间氮含量差异较小, 整体分布较为集中。

1.3.2 冠层无人机高光谱数据的获取与处理 本研究采用 DJI Matrice 300 RTK 无人机搭载 GaiaSky-mini3-VN 高光谱成像系统, 数据获取时间选择 10:00~14:00 光照强度稳定时段(光照辐射值 $>1\ 000\ \text{W}/\text{m}^2$, 风速 $\leq 3\ \text{m}/\text{s}$), 通过 Specview 软件完成光谱仪参数配置: 采用 Auto Expostime 功能自动优化曝光时间, 执行 Gray Frame 和 Black Frame 程序完成灰板标定及暗电流采集, 方便后续反射率校准。

飞行任务通过 DJIPilot2 专业航测系统规划, 基于四角航点坐标计算生成 KML 航线, 设定飞行高度 50 m、航速 $5\ \text{m}/\text{s}$ 。每个架次执行前完成标准作业流程: 10 m 高度重力校准、50 m 基准高度标准板标定, 同步采集大气校正参考数据。

原始数据经 Specview 系统处理流程: 镜头畸变校正、基于标定板的反射率转换、大气校正消除气溶胶散射影响、高光谱影像拼接, 最终获取研究区完整的高光谱立方体数据($400\sim 1\ 000\ \text{nm}$ 波段范围, 光谱分辨率 $3\ \text{nm}$)。使用 ENVI 软件在每个小区样品采集处划分 ROI, 取所有光谱反射率平均值作为该

样本的高光谱数据，共计提取200条高光谱数据。实现氮含量与光谱特征的高精度空间同步获取。

1.4 高光谱特征波段筛选

高光谱数据通常具有高维度、高冗余和高共线性的特点，如果不加筛选直接建模，容易导致计算复杂度上升、模型稳定性下降。为此，本研究采用多种经典与改进的特征波段筛选方法，对不同算法的筛选机制和建模效果进行系统比较，为后续模型构建提供可靠的波段基础。

首先，CARS利用 Monte Carlo 随机采样和指数递减函数，根据 PLSR 模型的 RMSECV 指标逐步剔除权重较低的波段，从而保留与目标变量最相关的波段特征。其次，UVE通过量化每个波段在 PLSR 回归中的稳定性，并结合随机噪声变量作为基准，剔除贡献度低或不稳定的波段，以提升模型的可靠性。此外，SPA 采用逐步正交投影策略，优先选择彼此共线性最低的波段，最终构建信息互补性强、冗余度低的特征子集，有助于改善回归模型的稳健性。

在此基础上，研究进一步引入基于进化策略的 GA 进行全局波段组合搜索。GA 通过选择、交叉和变异机制，从高维候选空间中寻找预测性能最优的波段组合，以 PLSR 的交叉验证 R^2 、RMSE 作为适应度函数，具有较强的全局搜索能力。

为了兼顾局部精细筛选与全局优化能力，本研究采用 GA-SPA 联合特征筛选策略。该方法首先利用 SPA 从全谱中预筛选出中等规模且共线性较低的候选波段集合，随后以此为 GA 的搜索空间，由 GA 进一步迭代优化波段组合。最终得到的特征子集同时具备较低冗余性和优异的预测性能，可为模型构建提供更具稳定性与解释性的波段基础。

1.5 建模方法

本研究利用回归模型构建玉米区域冠层氮含量与无人机高光谱反射率之间的定量关系。所采用的算法包括 PLSR^[18]、SVR^[19]、极端梯度提升(XGBoost)^[20]、KELM^[21]、DELM^[22]以及1D-CNN^[23]。数据集通过随机抽样(RS)方法按7:3比例划分为训练集和测试集，并用于各模型的训练与性能验证。在本研究中，为实现作物冠层氮含量的高精度预测，结合高光谱数据高维、多重共线性和非线性关系特点，选择多种回归建模方法并针对数据特性进行参数设计。PLSR通过对自变量和因变量协同降维，提取主要成分建立线性回归模型，有效应对多重共线性问题，同时保留光谱主要信息，提高氮含量预测的稳定性与精度。SVR利用核函数将高维光谱映射到特征空间，通过优化回归平面拟合非线性关系，对小样本和噪声数据表现良好。XGBoost作为梯度提升集成学习方法，通过迭代训练弱学习器并加权组合，实现对光谱数据复杂非线性特征的高效拟合。KELM在ELM基础上引入核函数映射，捕捉光谱与氮含量间的非线性关系，同时保持快速训练和计算简便的优势。1D-CNN利用卷积与池化层自动提取连续波段的局部特征，能够捕捉氮素相关波段间的复杂非线性关系。

DELM通过多层隐含层增强特征提取能力，能够更精准地拟合高光谱数据的非线性特征。DELM在处理高维、多重共线性以及噪声干扰时表现出更高的预测精度和稳健性。其多层结构可逐步提取低级到高级特征，使模型能够充分利用氮素相关波段间的复杂关系，从而提升冠层氮含量预测性能。通过这些方法的组合比较与优化，全面评估不同建模策略在氮素预测中的适用性，为研究数据特征选择最优建模方法提供依据。

1.6 模型评价指标

(1) 决定系数(R^2)

R^2 用于衡量模型对观测数据的拟合程度，反映因变量与自变量之间的整体相关性，其取值范围为0~1。当 R^2 越接近1，说明模型预测值在实测值附近分布均匀且密集，模型估测精度越高。其计算公式为：

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

式中 y_i 为样本的实测值， \hat{y}_i 为预测值， \bar{y} 为实测值的均值， n 为样本数量。

(2) 均方根误差(RMSE)

RMSE用于衡量预测值与实际观测值之间的偏差，是机器学习模型常用的性能指标。RMSE值越小，说明预测值与实测值之间的误差越小，模型估测精度越高；反之，RMSE越大，表示模型预测精度越低。其计算公式为：

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

式中 y_i 为样本的实测值, \hat{y}_i 为预测值, n 为样本数量。

(3) 残差预测偏差 (RPD)

RPD 指的是样本真实值的标准差与预测误差之间的比值。它反映了模型将样本差异转换为预测区分能力的强弱, 是光谱分析模型可靠性的核心指标之一。其计算公式为:

$$\text{RPD} = \frac{\text{SD}}{\text{RMSE}}$$

式中 SD 为测试集真实值的标准差, 用于评价数据集本身的离散程度, RMSE 为均方根误差。

(4) 交叉验证均方根误差 (RMSECV)

RMSECV 指的是训练集样本在交叉验证过程中, 实测值与预测值之间偏差的平均水平。它反映了模型在训练阶段的内部稳健性, 以及模型在面对样本子集变异时保持预测一致性的能力。其计算公式为:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_{i,cv})^2}{n}}$$

式中 n 为训练集样本量; y_i 为样本 i 的化学实测值; $\hat{y}_{i,cv}$ 为样本 i 作为验证组时, 由模型计算出的预测值。

2 结果与讨论

2.1 光谱预处理

在光谱采集过程中, 环境光照不均与仪器噪声等因素会引入干扰, 从而影响建模精度。为提升光谱质量与信噪比, 本文对原始数据进行了多种预处理。基本方法包括 SG、SNV、MSC、FD、二阶导数 (SD)、去趋势 (DT) 及 DWT; 在此基础上, 采用 SG+FD、SG+SD、DWT+SNV、DWT+MSC、DWT+FD、DWT+SD 等组合方案, 以筛选出适合本研究数据特征的最优预处理方式, 图 2 为基本方法预处理后的样本光谱。

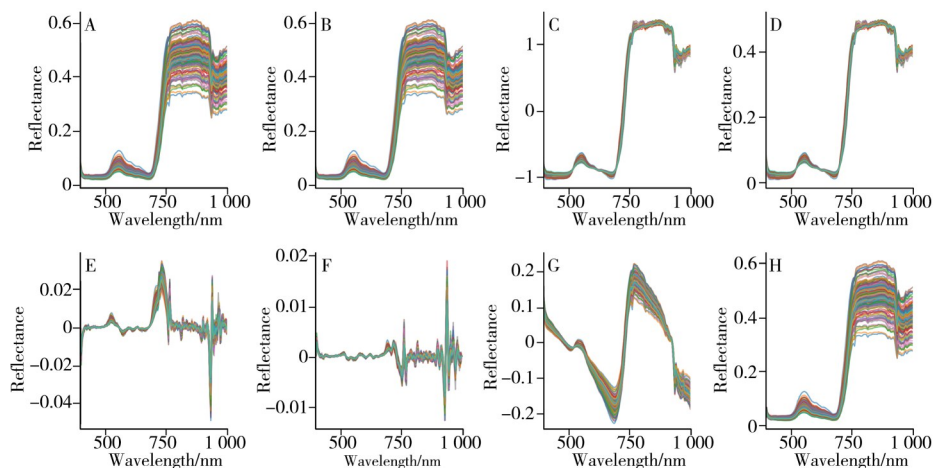


图 2 基本方法预处理后的样本光谱

Fig. 2 Sample spectra preprocessed by basic methods

A: 原始光谱, B: SG, C: SNV, D: MSC, E: FD, F: SD, G: DT, H: DWT

图 2 展示了原始光谱及不同预处理方式下的光谱变化。原始光谱图 2A 存在一定噪声与基线差异。经 SG 平滑 (图 2B) 后噪声有所降低, 但基线漂移仍未完全消除; SNV (图 2C) 和 MSC (图 2D) 有效减少了样本间的散射效应, 使光谱整体形态更一致。一阶与二阶导数 (图 2E、F) 能增强特征变化, 但同时放大了噪声, 曲线波动较大。DT (图 2G) 强化了光谱差异, 但噪声累积明显。相比之下, DWT (图 2H) 可在保留主要光谱特征的同时显著抑制高频噪声, 使曲线更加平滑、稳定。综合来看, SNV 有助于校正尺度差异, 而 DWT 可在降噪与特征保持之间取得更好平衡。因此, DWT+SNV 联合预处理能够同时提

升光谱稳定性与特征表达能力，是后续建模的较优预处理方案。

2.2 不同光谱预处理方法的性能比较

为评估不同预处理方法的效果，将原始光谱及各预处理后的光谱数据作为PLSR模型的输入，以玉米冠层氮含量为输出，确定最佳主成分数，并利用测试集对模型预测性能进行评估，从而筛选出最优光谱预处理方法。表1为不同预处理下光谱对氮含量的预测性能，其中，LV为PLSR模型提取的潜变量数量。

表1 不同预处理下光谱对氮含量的预测性能

Table 1 The predictive performance of spectra on nitrogen content under different preprocessing conditions

Preprocessing	R^2_t	RMSE _t	R^2_p	RMSE _p	LV	RPD
RAW	0.887 7	0.408 6	0.703 3	0.617 7	14	1.851
SG	0.849 7	0.472 8	0.712 9	0.607 7	13	1.882
SNV	0.848 9	0.474 0	0.713 3	0.607 2	10	1.884
MSC	0.760 9	0.596 3	0.728 2	0.591 3	6	1.934
FD	0.815 4	0.524 0	0.688 2	0.633 3	4	1.805
SD	0.775 9	0.577 4	0.668 0	0.653 5	3	1.749
DT	0.745 6	0.615 2	0.725 7	0.594 0	5	1.925
SG+FD	0.757 3	0.600 8	0.704 1	0.617 0	3	1.852
SG+SD	0.758 8	0.598 9	0.674 1	0.647 4	2	1.767
DWT	0.839 3	0.488 8	0.753 3	0.563 4	14	2.029
DWT+SNV	0.833 0	0.498 4	0.755 7	0.560 5	12	2.040
DWT+MSC	0.834 7	0.495 7	0.740 0	0.578 3	12	1.977
DWT+FD	0.748 2	0.612 0	0.685 6	0.636 0	3	1.798
DWT+SD	0.728 5	0.635 4	0.658 5	0.662 8	2	1.726

根据表1中数据可知，不同光谱预处理方法对模型的拟合效果存在明显差异。单一预处理方法在测试集上表现不同，而DWT的组合方案整体性能更优。其中，DWT与SNV结合的DWT+SNV方法，在训练集和测试集上均取得较高的 R^2 值(训练集 $R^2=0.8330$ ，测试集 $R^2=0.7557$ ，RPD=2.040)，且RMSE较低，预测能力强，RPD=2.040，表明模型的拟合精度和预测能力均优于其他方法。DWT+SNV方法表现优异的原因在于：DWT可在多尺度上有效分离光谱信号与噪声，保留关键特征；SNV能够消除散射效应与基线漂移，实现光谱标准化，从而提升冠层氮含量预测精度。因此，DWT+SNV被确定为本研究最优的光谱预处理方案。

2.3 特征波段筛选结果分析

对DWT+SNV预处理后的光谱数据，分别采用CARS、UVE、SPA、GA以及GA-SPA方法进行特征波段筛选，以提取与冠层氮含量高度相关的关键波段。

2.3.1 基于CARS、UVE、SPA和GA的特征波段选择 在波长筛选过程中，CARS算法设置Monte Carlo采样次数为100次，最大主成分数为20，交叉验证次数为10。在迭代过程中，随着采样次数的增加，特征波段数量呈递减趋势，当去除过多有效波段信息时，模型的RMSECV逐渐增大，迭代次数为34时选出最优特征子集，确定46个光谱特征波段。图3为CARS特征波段筛选结果。

UVE算法在波长筛选中采用5折交叉验证选择PLSR最优成分数，结合200次Bootstrapping采样构建回归模型，最终依据回归系数稳定性筛选出最优光谱特征波段，共筛选出83个光谱特征波段，从筛选结果来看，其中可能仍然包含一定的冗余信息。SPA算法以范数最大波段作为起始，最大端元素数20、最小端元素数5。前向投影迭代过程中，每次均选择与已选波段最不共线的新波段，最终筛选出45个光谱特征波段。GA算法以二进制编码表示特征子集，种群规模为100，进化次数为60，交叉概率为0.8，突变概率为0.2。适应度函数通过PLSR回归模型在5折交叉验证下评估子集的 R^2 ，最终选择最优波段组合以最大化预测性能，共筛选出80个光谱特征

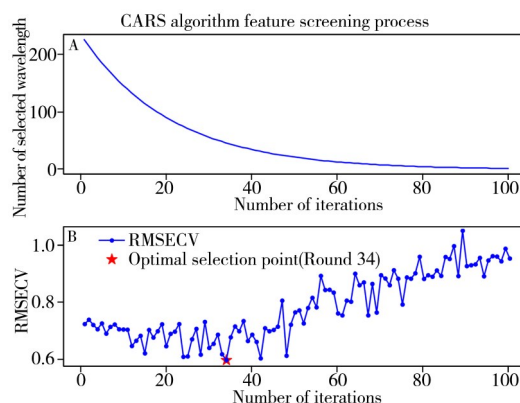


图3 CARS特征波段筛选结果

Fig. 3 CARS feature band screening results

波段。图 4 为各算法提取的特征波段分布。

由图 4 可知, CARS 主要保留 NIR 区(750~1 000 nm)波段, 能反映冠层结构和氮素积累, 但蓝光和红光区波段较少; UVE 覆盖蓝光(408~495 nm)、红边(666~715 nm)及部分近红外区(775~894 nm), 信息全面, 有助于构建稳定模型, 但波段较多可能存在一定冗余; SPA 集中在蓝光、红边和近红外区, 信息集中且冗余低, 但绿光和红光区波段较少, 可能影响模型对氮素的敏感性; GA 波段分布最均衡, 涵盖蓝光、红光、红边及近红外区, 能捕捉多区域氮素敏感信息, 但波段最多, 可能导致模型复杂度增加。

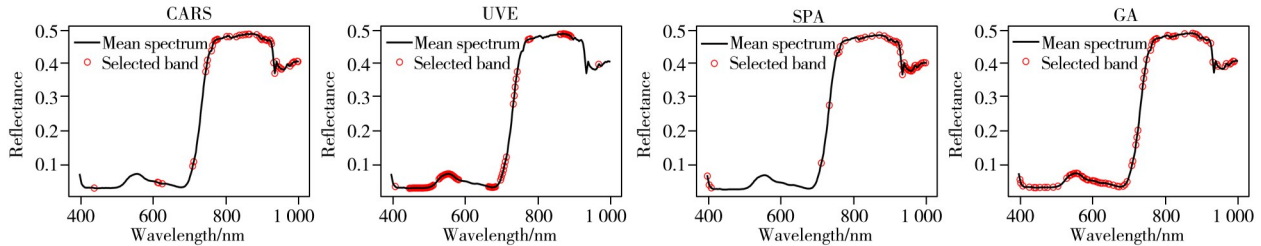


图 4 各算法提取的特征波段分布

Fig. 4 Distribution of feature bands extracted by each algorithm

2.3.2 基于 GA-SPA 的特征波段筛选 采用 GA-SPA 算法进行特征波段筛选。该方法先利用 SPA 预筛选 100 个候选波段, 有效降低多重共线性影响, 再在候选集上引入遗传算法(种群规模 100、迭代 100、交叉概率 0.8、变异概率 0.2、锦标赛选择)进行全局优化, 并根据子集规模自适应调整 PLSR 成分数以保证模型稳健性。图 5 为 GA-SPA 算法流程图。

通过 GA-SPA 算法筛选出 35 个波段, 分布在蓝光、绿光、红光、红边和近红外区域, 实现了从可见光到近红外的连续覆盖。各波段对应氮素相关的光谱特征如下: 蓝光区域(402~408 nm): 叶绿素 a/b 吸收峰区。氮素是叶绿素的重要组成元素, 因此该区域光谱能反映叶片氮素含量变化。绿光区域(465~558 nm): 叶绿素 b 和总叶绿素区, 对氮素变化敏感, 可补充蓝光区信息, 增强氮素预测能力。红光区域(636~647 nm): 叶绿素 a 吸收峰区, 与叶片氮素高度相关, 是氮素预测的关键波段。红边区域(704~775 nm): 氮素敏感区, 红边光谱受叶片氮素含量和叶片内部结构影响, 可反映氮素分布状态。近红外区域(778~992 nm): 反射受叶片结构和生物量主导, 间接反映冠层氮素分布, 同时可补充结构信息, 提升模型稳健性。该波段分布兼顾了氮素敏感区和冠层结构区, 避免了单一区域过度集中。相比单一特征选择方法, GA-SPA 先通过 SPA 降低多重共线性, 再利用遗传算法全局优化, 实现多区域连续波段的均衡选择, 从而提高了冠层氮含量预测模型的稳健性与精度。图 6 为 GA-SPA 的提取特征波段分布。

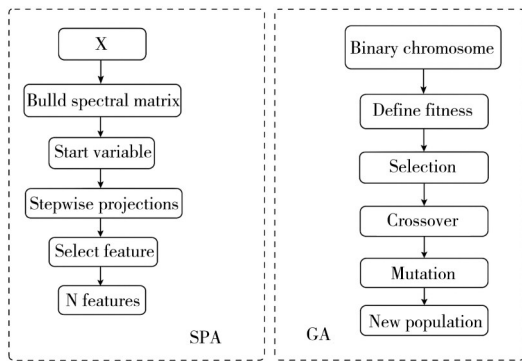


图 5 GA-SPA 算法流程图

Fig. 5 GA-SPA algorithm flowchart

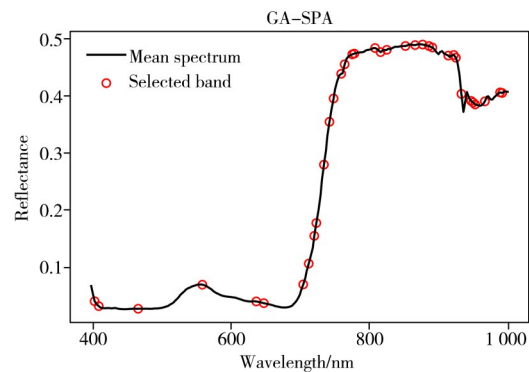


图 6 GA-SPA 提取特征波段分布

Fig. 6 GA-SPA extracts feature band distribution

2.4 模型分析

本研究采用 CARS、UVE、SPA、GA 和 GA-SPA 算法提取特征波段, 对比分析不同建模算法。

表 2 为基于 DWT+SNV 预处理方法和不同特征筛选方法的氮元素回归模型拟合效果, 由表可知, 使用原始光谱时, PLSR、SVR、KELM、DELM 和 1D-CNN 虽在训练集表现良好, 但测试集 R^2 存在较大差异, XGBoost 和 1D-CNN 在 RAW 条件下测试集 R^2 分别仅为 0.594 4 和 0.579 6, 存在过拟合风险。采用特征筛选算法后, 部分模型测试集 R^2 明显提升。其中使用 GA-SPA 特征波段的 DELM 回归模型效果最

好($R^2=0.8111$, $RMSE=0.4930$, $RPD=2.321$)。此外, GA-SPA 在 PLSR、KELM 和 SVR 等模型中同样表现出明显优势, 测试集 R^2 分别达到 0.7886、0.7965 和 0.7637, 显示该方法能够有效优化波段组合, 降低冗余信息对模型的干扰, 提高预测精度。相比之下, 虽然 XGBoost 和 1D-CNN 在经过 GA-SPA 优化后训练集 R^2 有所提高, 但测试集 R^2 提升有限, 仍存在一定过拟合风险, 说明复杂非线性模型在高维光谱数据下仍需结合正则化或数据增强等手段以增强泛化能力。综上所述, GA-SPA 特征筛选方法能够显著提升各类回归模型的预测性能, 其中与 DELM 结合的模型实现了训练拟合与测试泛化的最佳平衡, 是基于 DWT+SNV 预处理进行玉米氮元素预测的最优策略, 其中 Var 为特征筛选的变量数。

表2 回归模型拟合情况

Table 2 Fitting performance of regression models

Preprocessing	Method	Model	R^2_t	RMSE _t	R^2_p	RMSE _p	RPD	Var	
DWT+SNV	RAW	PLSR	0.833 0	0.498 4	0.755 7	0.560 5	2.040	223	
			CARS	0.790 0	0.558 9	0.780 3	0.531 6	2.151	46
			UVE	0.764 8	0.591 4	0.760 8	0.554 7	2.062	83
	SPA	PLSR	0.800 2	0.545 1	0.720 1	0.600 1	1.906	45	
			GA	0.797 0	0.549 4	0.781 8	0.529 7	2.159	80
			GA-SPA	0.800 6	0.544 5	0.788 6	0.521 4	2.193	35
	RAW	SVR	0.741 9	0.619 6	0.669 7	0.651 8	1.754	223	
			CARS	0.754 0	0.604 8	0.742 0	0.576 1	1.985	46
			UVE	0.679 0	0.690 9	0.651 1	0.670 0	1.707	83
	SPA	SVR	0.736 6	0.625 9	0.705 2	0.615 8	1.856	45	
			GA	0.747 4	0.612 9	0.719 0	0.601 2	1.902	80
			GA-SPA	0.803 5	0.540 5	0.763 7	0.551 3	2.074	35
	RAW	XGBoost	0.876 1	0.429 2	0.594 4	0.722 3	1.584	223	
			CARS	0.852 7	0.467 9	0.538 3	0.770 5	1.484	46
			UVE	0.812 0	0.528 7	0.580 3	0.734 7	1.556	83
	SPA	XGBoost	0.745 0	0.615 8	0.540 5	0.768 8	1.487	45	
			GA	0.856 6	0.461 8	0.610 0	0.708 2	1.615	80
			GA-SPA	0.847 9	0.475 7	0.612 1	0.706 3	1.619	35
	RAW	KELM	0.797 7	0.548 5	0.763 3	0.551 8	2.073	223	
			CARS	0.791 8	0.556 4	0.780 2	0.531 7	2.150	46
			UVE	0.811 0	0.530 2	0.774 1	0.539 0	2.123	83
	SPA	KELM	0.783 7	0.567 1	0.718 4	0.601 8	1.901	45	
			GA	0.819 7	0.517 9	0.787 4	0.523 0	2.187	80
			GA-SPA	0.833 5	0.497 6	0.796 5	0.511 6	2.235	35
	RAW	DELM	0.831 7	0.500 3	0.757 6	0.558 4	2.048	223	
			CARS	0.833 4	0.497 7	0.767 8	0.546 5	2.093	46
			UVE	0.776 4	0.576 7	0.758 6	0.557 2	2.053	83
	SPA	DELM	0.796 8	0.549 8	0.776 1	0.536 6	2.132	45	
			GA	0.829 4	0.503 7	0.796 1	0.512 2	2.232	80
			GA-SPA	0.830 0	0.502 8	0.811 1	0.493 0	2.321	35
	RAW	1D-CNN	0.838 8	0.489 6	0.579 6	0.735 3	1.556	223	
			CARS	0.804 2	0.539 6	0.664 8	0.656 7	1.742	46
			UVE	0.794 9	0.552 3	0.650 7	0.670 3	1.706	83
	SPA	1D-CNN	0.824 9	0.510 3	0.626 1	0.693 5	1.648	45	
			GA	0.822 7	0.513 4	0.670 8	0.650 8	1.757	80
			GA-SPA	0.821 9	0.514 7	0.677 7	0.643 9	1.777	35

图7展示了基于DWT+SNV预处理和GA-SPA特征筛选的DELM模型对玉米冠层氮含量的预测值与参考值之间的关系。可以看出, 预测值整体分布于1:1线两侧, 未出现明显系统性高估或低估现象, 且在不同氮含量区间内散点分布相对均匀, 表明模型对玉米冠层氮含量变化具有较好的响应能力。

2.5 基于无人机高光谱和最优反演模型的寒地玉米氮含量空间反演

将基于DWT+SNV预处理并经GA-SPA特征筛选的DELM模型确立为氮含量预测反演最佳模型。为实现区域玉米氮含量的空间反演, 首先使用阈值算法去除地面影像, 然后将玉米冠层高光谱影像中不同波段下的光谱反射率代入最佳模型中, 计算出每个像素点对应的氮含量值, 最终生成玉米氮含量空间分布图, 如图8所示。图像显示玉米冠层氮含量呈现明显的条带状空间异质性, 田块纵向(左上至右下)

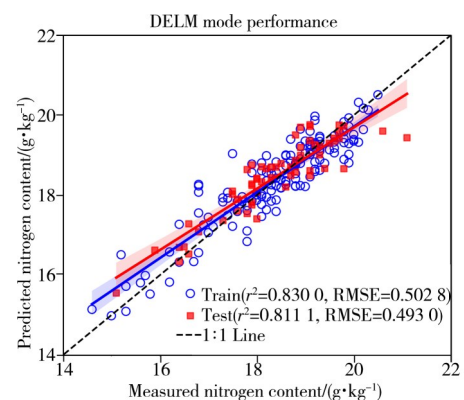


图7 DELM模型实测值与预测值散点图

Fig. 7 Scatter plot of measured and predicted values of DELM model

图7 DELM模型实测值与预测值散点图

总体表现为由低氮向高氮过渡的梯度分布,且种植行间存在显著的颜色差异,反映了田块尺度氮素营养的空间变异特征。背景蓝色区域为阈值法剔除的土壤及非植被像素。

3 结论

本研究通过比较多种光谱预处理方法,确定DWT+SNV为预处理最优方案;在此基础上,系统评估了多种特征筛选与建模策略在作物氮含量预测中的表现。结果表明,全波段建模易出现过拟合,而引入特征筛选可显著提升模型精度与稳定性。其中,GA-SPA方法在特征优化方面最为突出,结合DELM模型时测试集 $R^2=0.8111$ 、 $RMSE=0.4930$ 、 $RPD=2.321$,表现出最佳预测性能和泛化能力。GA-SPA在PLSR、KELM和SVR中亦展现较强适用性,而XGBoost与1D-CNN在有限样本下提升有限。总体而言,基于DWT+SNV预处理与GA-SPA特征选择的建模框架能够有效提升氮含量估测精度,可用于玉米冠层氮含量的快速检测,为高光谱驱动的精准农业提供了可靠技术支撑。

需要指出的是,本研究尚未引入实验室标准化学方法对叶片氮含量进行对比验证,这在一定程度上限制了对模型绝对精度的评价。后续研究将结合凯氏定氮或元素分析仪测定结果,对高光谱反演模型进行进一步的定量验证与校准。

参考文献:

- [1] He H Y, Hu Q, Li R, Pan X B, Huang B X, He Q J. *Field Crops Res.*, **2020**, 254: 107830.
- [2] Ye F, Qin S Z, Li H J, Li Z L, Tong T. *Front. Sustain. Food Syst.*, **2024**, 8: 1349765.
- [3] Yu X M, Schweikert K, Li Y J, Ma J, Doluschitz R. *J. Environ. Manag.*, **2023**, 325: 116347.
- [4] Li D L, Zhang P, Chen T, Qin W. *Remote. Sens.*, **2020**, 12(16): 2578.
- [5] Wang H Y, Wei Z Y, Yang Y X, Hou J Y, Sun Z T, Hu J. *Spectrosc. Spectral Anal.* (王浩宇, 魏子渊, 杨永霞, 侯军英, 孙章彤, 胡瑾. 光谱学与光谱分析), **2024**, 44(8): 2208-2215.
- [6] Kuang R, Long T, Liu H L, Wu J H, Lü J S, Xie Z R, Liu W T, Lan Y B, Long Y B, Wang Z H, Zhao J. *Spectrosc. Spectral Anal.* (匡润, 龙腾, 刘海林, 吴继辉, 吕金胜, 谢自然, 刘文涛, 兰玉彬, 龙拥兵, 王再花, 赵静. 光谱学与光谱分析), **2024**, 44(7): 1918-1927.
- [7] Liu J, Wang H, Zhang L, Xiao J J, Wu J G, Gong M C. *Chin. Agric. Sci.* (刘静, 汪泓, 张磊, 肖玖军, 吴建高, 龚明冲. 中国农业科学), **2025**, 58(2): 252-265.
- [8] Hou K Y, Bai T C, Li X, Shi Z Y, Li S W. *Forests*, **2024**, 15(2): 268.
- [9] Liu A R, Meng X, Liu Z G, Song Y F, Zhao X M, Zhi D N. *Spectrosc. Spectral Anal.* (刘傲然, 孟惜, 刘智国, 宋宇斐, 赵雪曼, 智丹宁. 光谱学与光谱分析), **2025**, 45(8): 2326-2334.
- [10] Yang W B, Li Y X, Wu F, Zou X C. *Trans. Chin. Soc. Agric. Eng.* (杨伟博, 李映雪, 吴芳, 邹晓晨. 农业工程学报), **2025**, 41(15): 165-173.
- [11] Li Y M, Wang H, Zhao H L, Zhang L G, Xia W J. *Agronomy*, **2025**, 15(9): 2072.
- [12] Liu X G, Zhang S, Chen S M, Tuo Y F, Peng K L, Tan S, Lou F C, Jiang T T, Ma X D. *Int. J. Remote. Sens.*, **2024**, 45(22): 8471-8498.
- [13] Guo Y P, Wang X M, Zhao F, Li P P. *Trans. Chin. Soc. Agric. Eng.* (郭艳萍, 王雪梅, 赵枫, 李平平. 农业工程学报), **2025**, 41(3): 83-93.
- [14] Zhu S F, Bai J C, Jin Z Y, Xiang S, Yu F H. *Trans. Chin. Soc. Agric. Mech.* (朱盛凡, 白驹驰, 金忠煜, 相爽, 于丰华. 农业机械学报), **2025**, 56(9): 576-584, 637.
- [15] Wen Z, Guo S, Shu T, Wu Y H, Huang R. *Spectrosc. Spectral Anal.* (文竹, 郭松, 舒田, 伍玉菡, 黄瑞. 光谱学与光谱分析), **2025**, 45(9): 2590-2596.
- [16] Wang Z L, Chen J X, Zhang J W, Tan X M, Ali Raza M, Ma J, Zhu Y, Yang F, Yang W Y. *Crop J.*, **2022**, 10(5): 1224-1238.
- [17] Wang X, Zhang G Y, Ma T M, Wang Z M, Zhao X D. *Trans. Chin. Soc. Agric. Eng.* (王雪, 张广月, 马铁民, 王志明, 赵孝东. 农业工程学报), **2025**, 41(8): 291-300.
- [18] Ji F J, Li F, Hao D L, Shiklomanov A N, Yang X, Townsend P A, Dashti H, Nakaji T, Kovach K R, Liu H R, Luo M, Chen M. *New Phytol.*, **2024**, 243(1): 111-131.
- [19] Yang F Q, Chen R Q, Liu Y, Yang Y. *Spectrosc. Spectral Anal.* (杨福芹, 陈日强, 刘杨, 杨昱. 光谱学与光谱分析), **2025**, 45(6): 1729-1738.
- [20] Li S, Cui L, Fu P. *J. Instrum. Anal.* (李硕, 崔岚, 付沛. 分析测试学报), **2024**, 43(4): 523-531.
- [21] Li X K, Yu H Y, Yu Y, Wang H J, Zhang L, Zhang X, Sui Y Y. *Spectrosc. Spectral Anal.* (李晓凯, 于海业, 于跃, 王洪健, 张蕾, 张昕, 隋媛媛. 光谱学与光谱分析), **2023**, 43(1): 93-99.
- [22] Gao S Y, Qiao J L, Zhou L N, Liu S, Chen L M, Yu Y, Kong L J. *Agronomy*, **2025**, 15(4): 828.
- [23] Li C S, Gao S H. *J. Instrum. Anal.* (李昌盛, 高树辉. 分析测试学报), **2025**, 44(5): 781-793.

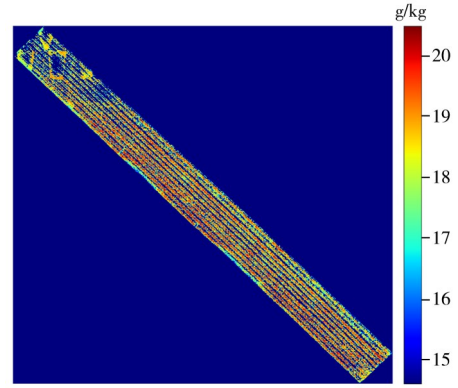


图 8 玉米冠层氮含量空间分布反演图
Fig. 8 Inverse map of spatial distribution of nitrogen content in maize canopy