

基于积分球-反向倍加法反演的棉种近红外光学特性分析与品种鉴别

张朕嘉¹, 罗真伟¹, 李林杰¹, 汪尹鹏², 张洪洲^{1*}, 李庆旭^{2*}

(1. 塔里木大学 机械电气化工程学院, 新疆 阿拉尔 843300; 2. 安徽财经大学 棉花工程研究所, 安徽 蚌埠 233030)

摘要: 以3个新疆典型棉种品种为对象, 采用近红外光谱-积分球测量系统采集120份棉籽仁样本及相应棉籽壳样本的总反射率和总透射率。利用反向倍加法反演获得吸收系数(μ_a)和约化散射系数(μ_s')光谱。分别以棉籽仁 μ_a 和 μ_s' 光谱为输入, 采用无信息变量消除法和竞争性自适应重加权采样进行变量筛选, 并建立偏最小二乘判别分析、支持向量机和随机森林品种鉴别模型。以嵌套交叉验证准确率、测试集准确率及分类评价指标评价模型性能。结果表明, 基于棉籽仁 μ_a 光谱的竞争性自适应重加权采样-偏最小二乘判别分析模型综合表现较好, 筛选出44个特征波长, 测试集准确率为97.22%, 嵌套交叉验证准确率为(95.83±5.89)%。该方法可实现棉种品种有效鉴别, 并为棉籽品种差异分析提供一定的光学依据。

关键词: 光学特性; 积分球; 近红外光谱; 反向倍加法; 棉种; 品种鉴别

中图分类号: O657.33; S339.3 **文献标识码:** A **文章编号:** 1004-4957(XXXX)XX-0001-09

Near-infrared Optical-Property Analysis and Variety Discrimination of Cottonseed Based on Integrating Sphere-Inverse Adding-Doubling

ZHANG Zhen-jia¹, LUO Zhen-wei¹, LI Lin-jie¹, WANG Yin-peng², ZHANG Hong-zhou^{1*}, LI Qing-xu²

(1. College of Mechanical and Electrical Engineering, Tarim University, Aral 843300, China; 2. Cotton Engineering Research Institute, Anhui University of Finance and Economics, Bengbu 233030, China)

Abstract: The identification of cottonseed cultivars is important for evaluating seed quality and variety traceability. Conventional near-infrared spectral analysis typically uses reflectance or transmittance spectra directly. However, these spectra exhibit apparent optical responses that contain information regarding coupled absorption and scattering. This study proposes a method based on optical parameters to identify cottonseed cultivars. Three typical cottonseed cultivars from Xinjiang were selected as research subjects. A total of 120 cottonseed kernel samples were used for discrimination modeling. The corresponding cottonseed shell samples were used for supplementary analysis of outer structural optical responses. The total reflectance and transmittance spectra were acquired using a near-infrared spectroscopy-integrating sphere measurement system. The spectral range was 900 - 1650 nm. The absorption coefficient (μ_a) and reduced scattering coefficient (μ_s') were obtained using the inverse adding-doubling method. The μ_a and μ_s' spectra were used to describe absorption- and scattering-related optical information, respectively. Thus, the original optical response was expressed using physically meaningful parameters. The μ_a and μ_s' spectra of cottonseed kernels were used as model inputs. Uninformative variable elimination and competitive adaptive reweighted sampling (CARS) were performed for variable selection. Partial least squares-discriminant analysis (PLS-DA), support vector machine, and random forest models were established for cultivar discrimination. A nested cross-validation strategy was used for model development. Outer 10-fold

收稿日期: 2026-05-08; 修回日期: 2026-06-18

基金项目: 国家自然科学基金资助项目(32501790)

* 通讯作者: 张洪洲, 博士, 教授, 研究方向: 农产品智能化检测与装备研究, E-mail: 120100012@taru.edu.cn;

李庆旭, 博士, 副教授, 研究方向: 农产品智能化检测与光谱分析研究, E-mail: 120220059@aufe.edu.cn

cross-validation was conducted to evaluate model stability. The inner procedure was used for variable selection and parameter optimization. After optimization, the final model was trained using the complete training set. An independent test set was used for external validation. The model performance was evaluated based on the nested cross-validation accuracy, test accuracy, precision, recall, and F1-score. The results showed that the optical parameters provided useful cultivar-related information. The μ_a -based models demonstrated good discrimination, while the CARS-PLS-DA model based on μ_a spectra exhibited the best comprehensive performance. By selecting 44 characteristic wavelengths from 231 full-spectrum variables, the test accuracy reached 97.22%, while the nested cross-validation accuracy was $(95.83 \pm 5.89)\%$. These results indicate that CARS retained its effective absorption features and that variable selection reduced spectral redundancy. Similarly, the μ_s' -based models demonstrated discrimination ability. This suggests that scattered information contributes to the differentiation of cultivars. Direct modeling using the original reflectance and transmittance spectra demonstrated discrimination potential. However, inverse adding-doubling (IAD)-derived parameters provided a clearer optical interpretation by separately describing the contributions of absorption and scattering. Therefore, the proposed method can be used to identify cottonseed cultivars effectively. Additionally, it provides optical evidence for analyzing cultivar differences in cottonseed samples.

Key words: optical properties; integrating sphere; near-infrared spectroscopy; inverse adding-doubling (IAD); cottonseed; variety discrimination

棉籽是棉花生产与加工过程中的重要副产物,富含油脂和蛋白质,在油脂加工、饲料利用和综合开发等方面具有较高利用价值^[1-2]。不同棉种品种所对应棉籽的内部组分及组织结构受遗传特性和生长环境等因素影响,往往存在一定差异^[3-4]。这些差异不仅关系到棉籽原料的分级利用,也会影响后续加工产品的品质稳定性。新疆是我国棉花主产区,棉种品种类型较多,棉籽在流通和加工过程中容易发生品种混杂,而不同品种棉籽外观相近,仅依靠人工经验或常规理化检测难以实现快速、客观鉴别。因此,建立适用于棉种品种快速识别的分析方法具有一定应用意义。

近红外光谱(NIR)技术具有分析速度快、样品前处理少和适于多组分信息获取等特点,已被用于棉籽品质检测及农产品快速识别研究^[5-6]。然而,常规近红外检测多直接利用反射或透射光谱强度进行建模,光谱信号中化学成分吸收和物理结构散射信息相互叠加。对于棉籽这类富含油脂、蛋白质且内部结构复杂的样本,散射背景可能掩盖与品种差异相关的有效吸收信息,从而影响模型的稳定性和解释性。相关研究表明,探明光在生物样品中的传输规律,有助于更准确地表征样品内部信息^[7]。因此,从光学特性角度分离并解析吸收与散射响应,可为棉种品种鉴别提供更具物理意义的特征依据。

吸收系数(μ_a)和约化散射系数(μ_s')是描述样品光学特性的参数。 μ_a 主要反映样品对光能量的吸收能力,通常与油脂、蛋白质、水分等化学组分的近红外吸收有关; μ_s' 主要反映样品对光传播方向和路径的散射影响,与硬度、细胞结构及表面状态等物理结构因素有关^[8]。积分球测量结合反向倍加法(IAD)可基于样品总反射率、总透射率和厚度信息反演获得光学特性参数,并已用于水果、蔬菜等农产品品质检测^[9-10]。相比之下,棉种品种鉴别的相关研究仍较少,利用IAD反演光学特性参数解析棉种品种差异的研究尚不充分。

基于此,本研究以3个新疆典型棉种品种为对象,采用近红外光谱-积分球测量系统采集棉籽仁和棉籽壳样本的总反射率和总透射率,并利用IAD反演获得 μ_a 和 μ_s' 光谱。在分析光学特性差异的基础上,结合特征波长筛选和分类算法建立棉种品种鉴别模型,以期能为棉种快速鉴别提供一种具有较好解释性的近红外光学特性分析方法。

1 实验部分

1.1 样品制备

本研究选取新疆地区3个典型棉种品种“军海1号”“新海陆45号”和“新陆中45号”为实验材料。每个品种筛选200粒籽粒饱满、表面无破损的棉籽,每5粒为一组,共获得120组棉籽样本。样品经手工剥壳后分别获得棉籽仁和棉籽壳,并置于电热恒温鼓风干燥箱中烘干至恒重,以降低水分波动

对光学参数反演结果的影响。

为满足积分球测量对样品平整度和厚度一致性的要求,对烘干后的棉籽仁进行切片处理。切片时横向去除棉籽仁两端不规则部分,截取中间区域,制备成厚度为 2.5 mm (± 0.1 mm) 的棉籽仁样本,如图 1 所示。棉籽壳去除内侧残留棉籽仁后,在相同条件下干燥处理,并选取相对完整、平整且能够覆盖积分球端口的壳片用于光学信号采集。棉籽壳样本仅用于外层结构光学特性补充分析,不参与后续品种鉴别模型的训练与测试。

1.2 光学特性参数采集与 IAD 反演

样品光学特性参数由近红外光谱仪-积分球测量系统采集,反射和透射模式下的实物装置及分时光纤连接方式分别如图 2 和图 3 所示。测量系统主要包括暗箱、HL2000-12 卤素光源、NIR17S 近红外光谱仪、IS-20-T 透射型积分球、IS-20-R 反射型积分球以及安装有 Morpho 光谱仪控制软件的计算机。其中,暗箱由铝型材和 2 mm 厚黑色聚碳酸酯实心板搭建,用于减少外部杂散光干扰。光谱采集范围为 900 ~ 1650 nm,积分时间为 1 ms。反射模式和透射模式采用同一台近红外光谱仪分时采集,并通过切换光纤连接分别获得总反射率和总透射率光谱。

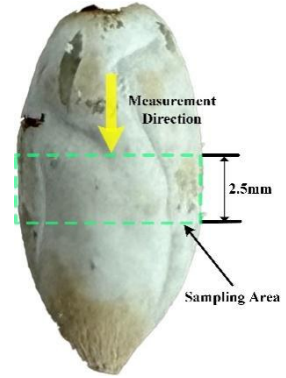


图 1 棉籽仁样品切片示意图

Fig. 1 Schematic diagram of a sliced cottonseed kernel sample

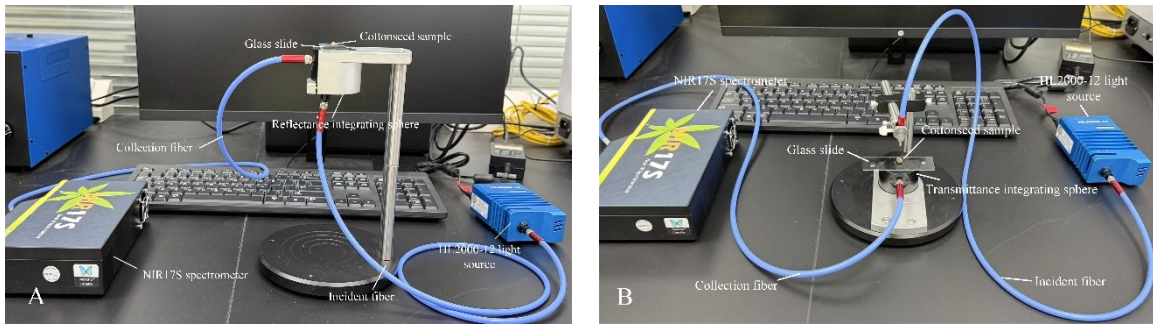


图 2 棉种光谱采集装置实物图

Fig. 2 Photographs of cottonseed spectral acquisition setups:

A. reflectance mode; B. transmittance mode

测量总反射率时,先以标定反射率为 98% 的漫反射白板和空载状态分别获取参考信号与杂散光背景,再将样品置于反射端口采集反射信号;测量总透射率时,先完全遮挡入射光束获取系统暗背景信号,再将样品置于透射端口采集透射信号,公式如下:

总反射率 (R_t):

$$R_t = r_{std} \times \frac{R_s - R_0}{R_{std} - R_0} \quad (1)$$

总透射率 (T_t):

$$T_t = \frac{T_s - T_{dark}}{T_0 - T_{dark}} \quad (2)$$

其中, r_{std} 为标准白板的标定反射率; R_s 和 T_s 为测得的反射和透射信号; R_{std} 为标准白板的反射信号; R_0 为反射端口的杂散光信号; T_0 为透射端口的透射信号; T_{dark} 为系统暗噪声信号。

将测得的总反射率、总透射率及样品实测厚度作为输入参数,利用 IAD 反演获得样品的 μ_s 和 μ_s' 光谱。IAD 算法通过迭代调整光学参数,使理论总反射率和总透射率与实测值达到设定误差范围内匹配,并对近红外范围内各波长点分别反演。所有样品均采用一致的系统参数、校正流程和反演条件。

1.3 理化指标测定

为分析不同品种棉籽仁样本的物质差异基础,对完成光学特性采集后的棉籽仁样本进行油脂含量、蛋白质含量和硬度测定。蛋白质含量参照 GB 5009.5-2016 《食品中蛋白质的测定》^[11],采用凯氏定氮法测定。将研磨后的棉籽仁粉末经浓硫酸和凯氏定氮催化剂消化后,利用 N610 凯氏定氮仪测定蛋白质

含量。

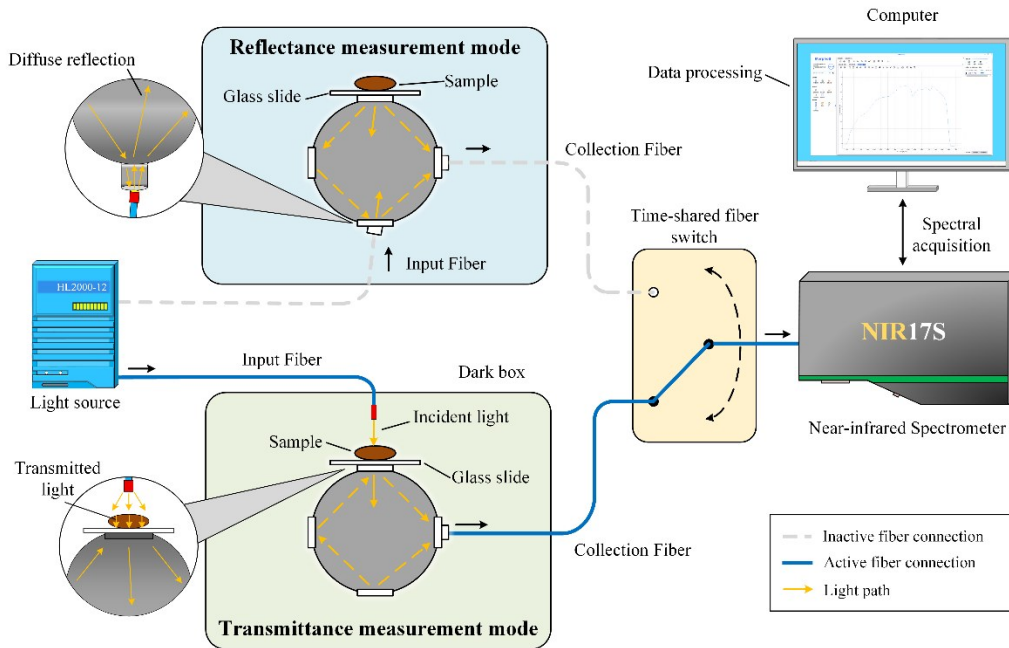


图3 棉种反射/透射测量模式及分时光纤连接示意图

Fig. 3 Schematic diagram of reflectance/transmittance measurement modes and time-shared fiber connections for cottonseed optical measurement

油脂含量参照 GB/T 14488.1-2008《植物油料 含油量测定》^[12], 采用索氏提取法测定。将棉籽仁样本烘干、研磨并装入滤纸包, 以 30 ~ 60 °C 石油醚为抽提溶剂, 利用 FT-160 粗脂肪测定仪进行抽提。

硬度采用 TA.XTC-18 质构仪测定。将棉籽仁样本水平放置于测试平台, 采用 TA/2 柱形探头进行下压测试, 触发力为 2 gf, 测试距离为 10 mm, 探头速度为 1 mm/s, 记录力~时间曲线中的首次峰值作为硬度值。

1.4 数据处理与建模方法

建模前, 采用主成分分析 (PCA) 结合 Hotelling's T^2 统计量和 Q 残差对棉籽仁光学参数光谱进行异常样本筛查, 未发现明显异常样本。随后采用 Kennard-Stone (KS) 算法将 120 份棉籽仁样本按 7 : 3 的比例划分为训练集和测试集, 为保证不同品种样本在训练集和测试集中的均衡分布, 本研究按品种分别进行 KS 划分后再合并样本集, 每个品种训练集和测试集样本数分别为 28 和 12。训练集用于特征波长筛选、模型参数优化和判别模型建立, 测试集仅用于最终外部验证。

棉籽仁 μ_a 和 μ_s' 光谱均包含 231 个变量。为比较不同变量模式对模型性能的影响, 分别采用全光谱输入、无信息变量消除法 (UVE) 和竞争性自适应重加权采样 (CARS) 进行建模分析^[13]。在不同变量模式下, 分别建立偏最小二乘判别分析 (PLS-DA)、支持向量机 (SVM) 和随机森林 (RF) 品种鉴别模型^[14]。

为避免数据泄漏, 特征波长筛选和模型参数优化均在训练集内部完成。模型训练采用嵌套交叉验证 (NCV) 策略, 外层采用 10 折交叉验证评价模型稳定性, 内层用于特征筛选和参数寻优。交叉验证结果以平均准确率评价模型性能, 并以标准差反映折间波动。完成参数优化后, 基于完整训练集建立最终模型, 并利用独立测试集评价外部预测性能。模型性能利用嵌套交叉验证准确率、测试集准确率 (Acc) 进行综合评价, 并基于混淆矩阵计算精确率、召回率和 F1 分数。公式如下:

精确率 (Precision, P_i):

$$P_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ki}} \times 100\% \quad (3)$$

召回率 (Recall, R_i):

$$R_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ik}} \times 100\% \quad (4)$$

F1 分数 (F1-score, $F1_i$):

$$F1_i = \frac{2 \times P_i \times R_i}{P_i + R_i} \quad (5)$$

准确率 (Accuracy, Acc):

$$\text{Acc} = \frac{\sum_{i=1}^n N_{ii}}{\sum_{i=1}^n \sum_{j=1}^n N_{ij}} \times 100\% \quad (6)$$

2 结果与讨论

2.1 不同品种棉种的物质差异基础

为分析不同品种棉籽仁样本的物质差异基础, 对其油脂含量、蛋白质含量和硬度进行测定, 结果见表 1。3 个品种棉籽仁样本在上述指标上均存在一定差异, 说明不同品种棉籽仁在化学组成和物理特性方面具有差异基础, 可为后续光学特性分析和品种鉴别模型建立提供参考。

表 1 不同品种棉种样本的理化指标

Table 1 Physicochemical indicators of cottonseed samples from different cultivars

Cultivar	Oil content/%	Protein content/%	Hardness/N
JH-1	28.73 ± 0.65 ^b	44.26 ± 0.28 ^a	48.99 ± 4.85 ^{ab}
XHL-45	28.45 ± 0.51 ^b	43.88 ± 0.37 ^b	49.61 ± 6.21 ^a
XLZ-45	29.97 ± 1.13 ^a	41.74 ± 0.26 ^c	46.26 ± 4.21 ^b

Values are mean ± SD. Different lowercase letters within a column indicate significant differences ($p < 0.05$).

2.2 不同品种棉种的光学特性分析

为比较不同品种棉种在近红外范围内的光学响应差异, 利用 IAD 算法分别获得棉籽仁和棉籽壳样本的 μ_a 和 μ_s' 光谱, 结果见图 4。由图 4A 可知, 不同品种棉籽仁 μ_a 光谱整体变化趋势较为一致, 但在 1 200 nm 和 1 450 ~ 1 530 nm 附近存在较明显的吸收响应。其中, 1 200 nm 附近吸收峰主要与脂质分子中 C—H 键的二级倍频伸缩振动有关^[15]; 1 450 ~ 1 530 nm 区域的宽吸收带主要受水分 O—H 键第一倍频振动及蛋白质 N—H 键吸收影响^[16]。不同品种在上述波段的吸收强度存在差异, 说明 μ_a 光谱能够在一定程度上反映棉籽仁内部油脂、蛋白质及含水相关组分差异。

由图 4B 可知, 棉籽仁 μ_s' 光谱整体随波长增加呈下降趋势, 符合生物组织内部散射随波长衰减的规律^[17]。同时, μ_s' 光谱在 1 200 nm 和 1 500 nm 附近出现波谷, 其位置与 μ_a 光谱中的主要吸收区域基本对应, 说明强吸收波段附近的散射响应也会发生变化, 该现象与相关农产品光学特性研究结果相似^[18]。从品种间差异来看, 3 个品种棉籽仁 μ_s' 光谱在整体强度上存在一定分层, 但样本曲线仍有交叠, 表明仅依靠单一光学参数或局部波段难以实现稳定区分, 仍需结合多变量模型进一步提取品种差异信息。

由图 4C 和图 4D 可知, 棉籽壳与棉籽仁的光学响应存在差异。棉籽壳 μ_a 光谱整体较平缓, 1 200 nm 附近吸收峰减弱, 主要吸收变化集中在 1 400 ~ 1 500 nm 附近; 其 μ_s' 光谱整体高于棉籽仁, 表现出更强散射特征。不同品种棉籽壳 μ_s' 曲线存在分层, 并在约 1 400 nm 附近出现交叉, 说明棉籽壳作为完整棉籽的重要外层结构, 也包含一定品种相关的散射响应信息。但与棉籽仁相比, 棉籽壳光学响应更易受壳层厚度、颜色及表面磨损等因素影响, 差异来源较复杂。为减少非品种因素干扰, 本文将棉籽壳作为外层结构光学响应进行比较分析, 后续模型仍以棉籽仁 μ_a 和 μ_s' 光谱作为主要输入。

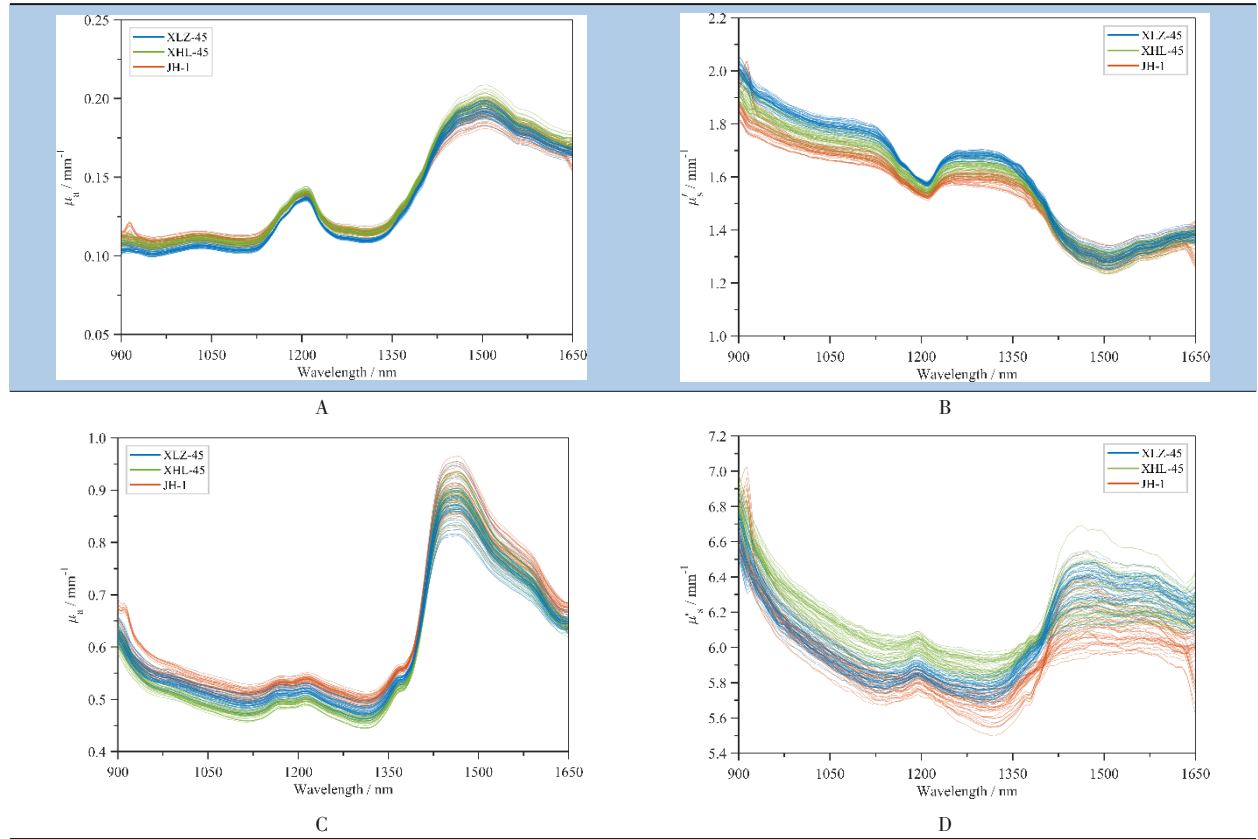
2.3 特征波长筛选与模型结果

结合建模结果, 以综合表现较好的 μ_a -CARS-PLS-DA 模型展示 CARS 特征变量筛选过程, 结果见图 5。随着蒙特卡罗采样次数增加, 保留变量数逐渐减少, 交叉验证均方根误差 (RMSECV) 先降低后升高, 并在第 18 次采样时达到最小。回归系数路径图显示, 蓝色虚线所示最优采样点之后, 部分变量回归系数波动明显增大, 说明继续剔除变量可能导致模型稳定性下降。最终, CARS 从 231 个变量中筛选出 44 个特征波长, 主要分布于 950 ~ 1060 nm、1175 ~ 1220 nm 和 1430 ~ 1520 nm 等区域, 与棉籽仁

的 μ_a 光谱主要吸收响应区基本一致,表明筛选结果能够保留与品种差异相关的有效吸收信息。

图4 不同品种棉籽仁与棉籽壳的光学特性光谱图

Fig. 4 Optical property spectra of cottonseed kernels and hulls from different cultivars



A-B. μ_a spectra and μ_s' spectra of kernel; C-D. μ_a spectra and μ_s' spectra of hull

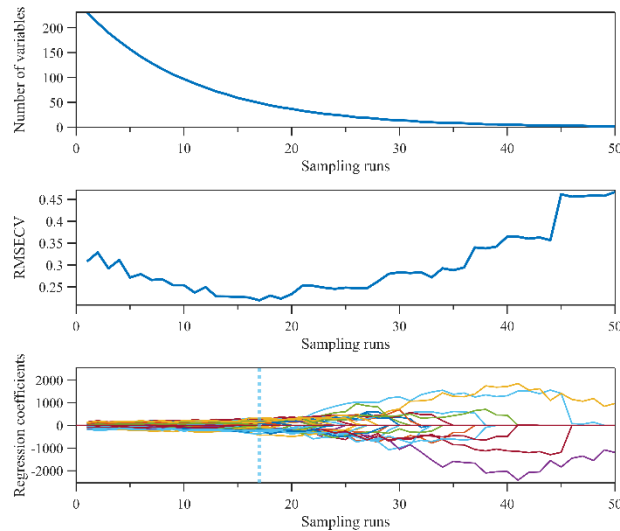


图5 基于CARS算法的棉籽仁 μ_a 光谱特征波长筛选过程

Fig. 5 CARS-based wavelength selection for cottonseed kernel μ_a spectra

不同光学参数下的品种鉴别结果见表2和表3。由表2可知,基于 μ_a 光谱的CARS-PLS-DA模型从231个全谱变量中筛选出44个特征变量,外层10折交叉验证准确率为 $(95.83 \pm 5.89)\%$,测试集准确率为97.22%,在 μ_a 光谱模型中表现最优。为进一步观察该模型的具体分类情况,绘制了其测试集混淆矩阵,结果见图6。该模型在36个测试样本中正确识别35个,仅有1个军海1号样本被误判为新海陆45号,说明 μ_a 光谱中的特征变量能够有效表征棉籽仁品种间的吸收特性差异,并具有较好的品种识别

能力。

由表3可知, μ_s' 所反映的结构散射信息同样能够体现棉籽仁的品种间差异。其中, 全光谱SVM和RF模型测试集准确率均为97.22%, 外层10折交叉验证准确率分别为 $(92.50 \pm 6.15)\%$ 和 $(90.83 \pm 8.29)\%$ 。二者测试集准确率相同, 可能与两种模型对全波段 μ_s' 信息的判别结果较为接近有关。相比之下, μ_a -CARS-PLS-DA模型在相同测试集准确率下具有更高的平均准确率, 并将变量数减少至44个, 说明其在保持判别效果的同时有利于简化模型。经UVE或CARS筛选后, 部分模型准确率下降, 说明 μ_s' 相关信息可能分布于较宽波段范围, 过度压缩变量易造成有效散射信息损失。由表4可知, 代表性模型在各类别上均取得较高精确率、召回率和F1值, 进一步表明IAD反演得到的 μ_a 和 μ_s' 均可作为棉籽品种鉴别提供有效光学信息。

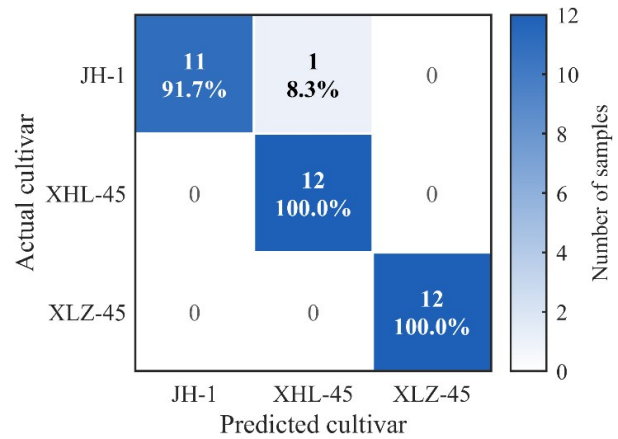


图6 μ_a -CARS-PLS-DA模型测试集混淆矩阵

Fig. 6 Confusion matrix of μ_a -CARS-PLS-DA model on test set

表2 基于棉籽仁 μ_a 光谱的棉种品种鉴别模型总体性能

Table 2 Overall performance of cottonseed variety-discrimination models based on cottonseed kernel μ_a spectra

Model	Feature variable	VN	NCV Acc/% (mean $\pm$ SD)	Test Acc/%
PLS-DA	Full	231	82.50 $\pm$ 13.86	94.44
	CARS	44	95.83 $\pm$ 5.89	97.22
	UVE	154	82.50 $\pm$ 14.93	91.67
SVM	Full	231	83.33 $\pm$ 12.42	88.89
	CARS	40	84.17 $\pm$ 11.42	88.89
	UVE	81	78.33 $\pm$ 11.92	83.33
RF	Full	231	84.17 $\pm$ 9.17	94.44
	CARS	40	81.67 $\pm$ 12.91	88.89
	UVE	110	81.67 $\pm$ 12.91	83.33

表3 基于棉籽仁约化散射系数 (μ_s') 光谱的棉种品种鉴别模型总体性能

Table 3 Overall performance of cottonseed variety-discrimination models based on cottonseed kernel μ_s' spectra

Model	Feature variable	VN	NCV Acc/% (mean $\pm$ SD)	Test Acc/%
PLS-DA	Full	231	79.17 $\pm$ 4.39	77.78
	CARS	30	75.00 $\pm$ 11.11	83.33
	UVE	177	81.67 $\pm$ 10.24	83.33
SVM	Full	231	92.50 $\pm$ 6.15	97.22
	CARS	44	85.00 $\pm$ 6.57	88.89
	UVE	144	91.67 $\pm$ 5.56	94.44
RF	Full	231	90.83 $\pm$ 8.29	97.22
	CARS	45	90.00 $\pm$ 6.57	91.67
	UVE	96	88.33 $\pm$ 7.03	94.44

表4 代表性模型在测试集上的分类评价指标

Table 4 Classification evaluation metrics of representative models on test set

Optical parameter	Model	Feature variable	Category	Test set		
				P/%	R/%	F1/%
μ_a	PLS-DA	CARS	JH-1	100.00	91.67	95.65
			XHL-45	92.31	100.00	96.00
			XLZ-45	100.00	100.00	100.00
μ_s'	SVM	Full	JH-1	92.31	100.00	96.00
			XHL-45	100.00	91.67	95.65
			XLZ-45	100.00	100.00	100.00
μ_s'	RF	Full	JH-1	100.00	100.00	100.00
			XHL-45	100.00	91.67	95.65
			XLZ-45	92.31	100.00	96.00

为比较 IAD 反演参数与原始光谱的建模效果差异,采用未经反演的总反射率和总透射率光谱进行直接建模,结果见表 5。由建模结果可知,原始总反射率和总透射率光谱均表现出一定的品种判别能力,其中,总反射率光谱的 CARS-SVM 模型测试集准确率最高,为 94.44%;总透射率光谱的 Full-PLS-DA 模型测试集准确率最高,为 91.67%,表明两类原始光谱中均包含可用于品种区分的信息。相较于原始光谱的表观响应,IAD 反演得到的 μ_a 和 μ_s' 能够分别表征样品的吸收和散射特性,使模型结果具有更明确的光学意义,并为棉籽品种差异分析提供更具物理意义的解释依据。

表 5 原始总反射率和总透射率光谱直接建模结果

Table 5 Direct modeling results based on original total reflectance and total transmittance spectra

Model	Spectral input	Feature variable	VN	NCV Acc/% (mean $\pm$ SD)	Test Acc/%
PLS-DA	R	Full	231	77.50 $\pm$ 7.91	86.11
		CARS	59	81.67 $\pm$ 12.30	86.11
	T	Full	231	80.00 $\pm$ 8.96	91.67
		CARS	14	78.33 $\pm$ 7.03	88.89
SVM	R	Full	231	88.33 $\pm$ 8.05	91.67
		CARS	33	85.83 $\pm$ 9.66	94.44
	T	Full	231	87.50 $\pm$ 5.89	88.89
		CARS	20	72.50 $\pm$ 13.06	75.00
RF	R	Full	231	87.50 $\pm$ 9.00	91.67
		CARS	44	85.83 $\pm$ 9.66	91.67
	T	Full	231	85.00 $\pm$ 7.66	86.11
		CARS	11	89.17 $\pm$ 6.86	83.33

3 结 论

本研究采用近红外光谱-积分球测量系统结合 IAD 反演方法,获得 3 个新疆典型棉种品种棉籽仁和棉籽壳样本的 μ_a 和 μ_s' 光谱,并基于棉籽仁光学参数建立品种鉴别模型。结果表明,不同品种棉籽仁在 μ_a 和 μ_s' 光谱中均存在差异,说明吸收特性和散射特性均可反映棉籽品种间差异;棉籽壳 μ_s' 光谱表现出较强散射特征及波段依赖性,可作为外层结构光学响应的补充信息。基于棉籽仁 μ_a 光谱的 CARS-PLS-DA 模型从 231 个全谱变量中筛选出 44 个特征波长,测试集准确率为 97.22%,嵌套交叉验证准确率为 (95.83 $\pm$ 5.89)%,可有效实现棉种品种鉴别。研究结果说明,IAD 反演获得的光学特性参数可为棉种品种鉴别提供有效特征,并为棉籽品种差异分析提供一定的光学依据。后续仍需增加品种、产地和完整棉籽样本,进一步验证该方法的实际适用性。

参考文献:

- [1] Wu M, Pei W, Wedegaertner T, Zhang J, Yu J. *Front. Plant Sci.*, **2022**, 13: 864850.
- [2] Gurusamy M, Subramanian B. *Fuel*, **2025**, 383: 133789.
- [3] Le Y, Zhao W, Liu X, Chen M, Xiong X, Zhang X, Lin Z. *Plant Physiol. Biochem.*, **2024**, 216: 109159.
- [4] Gong J, Kong D, Liu C, Li P, Liu P, Xiao X, Liu R, Lu Q, Shang H, Shi Y, Li J, Ge Q, Liu A, Deng X, Fan S, Pan J, Chen Q, Yuan Y, Gong W. *J. Agric. Food Chem.*, **2022**, 70 (8): 2529-2544.
- [5] Dong X F, Li Q X, Luo Z W, Zhang S, Zhang H Z, Jin G Q. *Processes*, **2025**, 13 (10): 3221.
- [6] Li Q X, Zhang H Z, Zhou W H, Liu R H, Dong X F, Jin G Q. *Microchem. J.*, **2024**, 207: 112005.
- [7] Xia Y, Liu W, Meng J, Hu J, Liu W, Kang J, Luo B, Zhang H, Tang W. *Front. Plant Sci.*, **2024**, 14: 1324881.
- [8] Xie D, Liu D, Guo W. *Postharvest Biol. Technol.*, **2021**, 179: 111569.
- [9] Ouyang A G, Wang Y, Liu Y D, Hou Y F, Wang G T. *Smart Agric.* (欧阳爱国, 王阳, 刘燕德, 侯佑飞, 王观田. 智慧农业 (中英文)), **2025**, 7 (4): 47-57.
- [10] Xie D D, Guo W C, Gao M J, Liu D Y. *Food Sci.* (谢丹丹, 郭文川, 高梦杰, 刘大洋. 食品科学), **2021**, 42 (1): 41-46.
- [11] GB 5009.5-2016. National Food Safety Standard-Determination of Protein in Foods. National Standards of the People's Republic of China (食品安全国家标准-食品中蛋白质的测定. 中华人民共和国国家标准).
- [12] GB/T 14488.1-2008. Oilseeds-Determination of Oil Content. National Standards of the People's Republic of China (植物油料-含油量测定. 中华人民共和国国家标准).
- [13] Liang X J, Wang Y N, Ma J F, Sun P, Guo T, Yan S K, Xiao X. *J. Instrum. Anal.* (梁小娟, 王娅妮, 马晋芳, 孙鹏, 郭拓, 严诗楷, 肖雪. 分析测试学报), **2023**, 42 (3): 307-314.
- [14] Liu W, Qiu X W, Jiang L W, Fan W, Li P, Zheng Y. *J. Instrum. Anal.* (柳薇, 邱熙文, 蒋立文, 范伟, 李跑,

- 郑郁. 分析测试学报), **2025**, 44 (2): 246-252.
- [15] Cazzaniga E, Cavallini N, Giraudo A, Gavoci G, Geobaldo F, Pariani M, Ghirardello D, Zeppa G, Savorani F. *Foods*, **2023**, 12 (1): 34.
- [16] Medeiros D T D, Ramalho F M G, Batista F G, Mascarenhas A R P, Chaix G, Hein P R G. *Ind. Crops Prod.*, **2023**, 192: 115989.
- [17] Lu R, Van Beers R, Saeys W, Li C, Cen H. *Postharvest Biol. Technol.*, **2020**, 159: 111003.
- [18] Gao M, Guo W, Huang X, Du R, Zhu X. *Postharvest Biol. Technol.*, **2021**, 172: 111385.

(责任编辑: 盛文彦)