

基于因子化和偏最小二乘算法的稻谷霉变快速分级预测

左 锋¹, 郝振宇², 章采东¹, 董志昊¹, 钱丽丽^{1*}, 张 涛³, 刘兴泉⁴, 周国鑫⁴

(1. 黑龙江八一农垦大学食品学院, 大庆 163319; 2. 江南大学食品学院, 无锡 214122; 3. 国家粮食和物资储备局, 北京 100834; 4. 浙江农林大学食品与健康学院, 杭州 311300)

摘 要: 粮食的霉变严重影响其品质与食品安全, 而常规检测手段存在速度慢以及需要大量专业实验室设备和操作人员等缺陷, 近红外光谱分析技术具有分析速度快、非破坏性、测试重现性好、易于实现在线分析和操作简单等诸多优点, 是一种很有潜力的快速检测方法。该研究基于近红外光谱分析技术建立了霉情检测模型, 对不同霉变程度稻谷进行近红外光谱的快速识别预测研究, 旨在开发一种可以快速鉴别霉变稻谷定性、定量模型。研究对 4 种 (2018 年牡丹江 27 号、2019 年牡丹江 27 号、龙粳长粒香和牡响 1 号) 不同霉变程度共 960 组霉变稻谷样品进行定性判别模型研究, 其中一阶导数+9 点平滑+因子化法建立的定性判别模型准确度较高, 样品之间的距离 S 均值大于 1, 分辨效果好, 通过留一交叉验证验证模型的平均准确率为 93.00%; 基于近红外光谱对 4 种不同霉变程度共 300 组霉变稻谷样品进行霉菌菌落总数的定量模型研究, 通过矢量归一化法+偏最小二乘法 (partial least squares, PLS) 定量分析建立定量判别模型, 其交叉验证的均方根误差 (root mean square error of cross-validation, RMSECV)、决定系数 (correlation coefficient of determination, R^2)、性能与偏差之比 (ratio of performance to deviation, RPD) 和预测均方根误差 (root mean squared error of prediction, RMSEP) 分别为 0.470、0.904 5、3.24 和 0.45, 模型精确度较好的。经过分析方法优选后而建立的霉变判别模型显示, 霉变是影响近红外光谱变化的主导因素, 而稻谷的品种与年际对其影响较小。研究结果可为基于近红外光谱分析技术对不同运输过程中的稻谷实现快速预测其霉变程度或其霉菌数量以及用于集装箱内粮食霉变情况监控在线实时监测装备的研究提供参考。

关键词: 近红外; 光谱技术; 分级预测; 稻谷; 霉变程度; 菌落总数; 矢量归一化; PLS 定量分析

doi: 10.11975/j.issn.1002-6819.202408027

中图分类号: O657.33

文献标志码: A

文章编号: 1002-6819(2025)-04-0299-10

左锋, 郝振宇, 章采东, 等. 基于因子化和偏最小二乘算法的稻谷霉变快速分级预测[J]. 农业工程学报, 2025, 41(4): 299-308. doi: 10.11975/j.issn.1002-6819.202408027 <http://www.tcsae.org>

ZUO Feng, HAO Zhenyu, ZHANG Caidong, et al. Rapid grading prediction of mould in rice grains based on factorisation and partial least squares algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(4): 299-308. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202408027 <http://www.tcsae.org>

0 引 言

近几年, 随着物流技术的发展, 稻谷的跨区域运输越来越频繁。黑龙江作为中国的商品粮基地, 其优质的粮食产品在南方市场逐年热销, “北粮南运”的需求在逐渐增加。然而, 在“北粮南运”过程中, 散粮在运输中因在不同季节跨越不同储粮生态区域常有霉变现象发生。同时, 稻谷在贮藏的过程中也有霉变现象发生, 霉变导致黄曲霉 B1、赭曲霉毒素和呕吐毒素含量超标, 给企业带来直接经济损失, 给消费者造成健康隐患^[1-6]。

光谱分析技术具有分析速度快、非破坏性、测试重现性好、易于实现在线分析和操作简单等诸多优点, 已经广泛应用于农业、食品领域^[7-9]。近红外光谱覆盖电磁 (electromagnetic, EM) 光谱的 780~2 500 nm 范围,

可捕获样品在反射、相互作用和透射模式下的光谱, 从而提供有关其元素分子键振动的丰富信息^[10-11]。在与电磁辐射相互作用时, 捕获的光谱包括泛音和组合带, 主要涉及氢键官能团 (-CH、-NH、-OH-SH)。这种宽带具有很高的信息量, 需要化学计量学工具来揭示光谱信息与样品结构特征之间的相关性^[12-13]。通过扫描样品, 可以获得粮食在霉变过程中有机化合物的含氢基团信息变化, 通过特性信息可以建立质量参数与光谱之间的对应关系, 结合化学计量学, 建立粮食霉变的判别模型^[14-15]。

近红外光谱仪是近红外分析的硬件, 主要由光源、分束系统、样品检测器、光学检测器以及数据处理分析系统组成^[16-17]。TAO 等^[18]使用近红外光谱来检测玉米粒表面被黄曲霉毒素 B1 污染, 对原始光谱进行导数处理, 与使用原始光谱相比, 导数处理后的光谱普遍提高了分类模型的性能。CHEN 等^[19]采用基于不同光谱预处理方法的近红外偏最小二乘法 (partial least squares, PLS) 模型来预测面团吸水率、面团醒发时间和面团稳定性, 每个参数的决定系数 (correlation coefficient of determination, R^2) 和预测均方根误差 (root mean squared error of prediction, RMSEP) 分别为 0.7 和 1.560、0.73 和 1.065

收稿日期: 2024-08-04 修订日期: 2024-11-02

基金项目: 国家重点研发计划项目 (2023YFD2301604-4)

作者简介: 左锋, 博士生, 研究方向为粮油加工与蛋白利用。

Email: zuofeng-518@126.com

*通信作者: 钱丽丽, 教授, 博士, 研究方向为农产品质量安全及产地溯源。Email: qianlili286@163.com

以及 0.79 和 1.090, 预测性能有待进一步提高。龚润华等^[20]对红小豆的近红外光谱进行标准正态变量变化法 (standard normal variable change method, SNV)、多元散射校正法 (multivariate scattering correction method, MSC) 及一阶导数预处理, 随后结合偏最小二乘分析法 (partial least squares, PLS), 建立红小豆样品不同品质指标的定量分析模型, 各 PLS 模型能较好地红小豆的品质指标预测, 虽然 PLS 和 PLS-DA 模型在其所选取的样本上取得了较好的预测和判别效果, 但模型的泛化能力仍需进一步验证。特别是在面对不同产地或不同年份的红小豆时, 模型的准确性和稳定性可能会受到影响。蒋雪松等^[21]通过阵列式的近红外光谱仪对小麦的霉变程度进行判别, 采用平滑、多元散射校正和导数变换等预处理方式消除光谱噪音, 再结合主成分分析 (principal component analysis, PCA)、线性判别分析 (linear discriminant analysis, LDA) 和偏最小二乘回归分析 (partial least squares regression, PLSR) 建立了小麦霉变程度的定性定量分析模型。虽然建立的 PLSR 模型在预测小麦菌落总数方面表现出较高的精度, 但模型的泛化能力尚需进一步验证, 对光谱预处理方法还需进行优化, 如尝试不同的平滑, 以提高光谱数据的准确性和可靠性。吕都等^[22]采用近红外光谱技术结合化学计量学方法建立快速鉴别稻谷霉菌污染的判别模型, 虽然采用了多种光谱预处理方法, 并确定了最佳预处理方式为 Quantile Normalization, 但其他预处理方法的组合或优化可能未被充分研究。近几年来, 采用近红外光谱预处理+化学计量学的分析模式在松子、柑橘、花生、番茄、菠萝等农产品的品质判别中得到了不同程度的研究, 但是对霉变稻谷进行快速分级预测仍需要进一步研究^[23-28]。

本研究采用近红外光谱法结合不同光谱预处理算法构建定性定量模型, 能迅速准确地检测粮食内部质量和安全性, 减少人力物力的消耗, 为散粮的智能化管控运

输提供理论研究基础数据, 降低稻谷在运输中以及贮藏时的霉变损失。

1 材料与方法

1.1 试验材料

试验于 2018 年收获期采集牡丹江地区的稻谷, 品种为牡丹江 27 号, 2019 年收获期内采集牡丹江地区的 3 种稻谷, 品种分别为牡丹江 27 号、龙粳长粒香和牡响 1 号。将采集的样品进行晾晒, 挑除杂质, 控制稻谷含水率为 14.0%~14.5%, 常温储藏, 备用。

1.2 仪器设备

TENSOR II 傅里叶变换近红外光谱仪, 德国布鲁克-北京科技有限公司。

1.3 方法

1.3.1 不同程度霉变样品制备

按 1:5 (水:稻谷, L/g) 的比例将无菌水接入稻谷中, 拌匀, 分装并密封于 3 个无菌托盘后置于 35 °C 培养箱中。取样前将稻谷再次拌匀, 每个托盘取样 20 次, 每次取样 50 g, 一次取样产生 60 个样品。取样时间分别为 0 d (0 d 为未结露), 1 d, 7~8 d 和 12~13 d, 总计 240 组样品。

不同稻谷均采用此处理方法, 4 种稻谷共计 960 组样品。如表 1 所示, 从感官指标、霉变粒以及霉菌菌落总数界定稻谷的霉变程度 (下文中分别简称为未结露-霉变、结露-霉变前期、结露-霉变中期和结露-霉变后期) 划分依据为国家标准中对粮食霉变粒的规定, 参考 GB/T2715-2016。在制备不同霉变程度的稻谷样品时, 依据稻谷霉变粒比例和感官评价, 结合霉菌菌落总数的测定, 综合判断稻谷的霉变程度, 并进行相应的划分。由于样品是在同样湿度和温度下制备的, 且不同阶段取样时间间隔较长, 并未出现霉变程度与霉变时间不一致的样品。

表 1 结露稻谷霉变程度的界定

Table 1 Definition of mildew degree of dewy rice

霉变程度 Degree of mildew	取样时间 Sampling time/d	感官指标 Sensory index	霉变粒比例 Proportion of mildew granules/%	霉菌菌落总数 Total mold colonies/(CFU·g ⁻¹)
未结露且未霉变	0	色泽金黄、无异常霉变	0	10 ² ~10 ⁴
结露稻谷霉变前期	1	发软、颜色变暗	0~1	10 ⁴ ~10 ⁵
结露稻谷霉变中期	7~8	出现肉眼可见的霉菌生长, 影响了食用价值	1~5	10 ⁵ ~10 ⁷
结露稻谷霉变后期	12~13	出现大量霉菌生长, 彻底失去食用价值	>5	>10 ⁷

1.3.2 原始光谱的采集

每次取样完成后, 将样品置于室温 1 h 直至样品温度稳定。环境温度为 (25±1) °C, 相对湿度为 20%~30%。因霉变主要发生在稻谷表皮, 因此试验采用整粒稻谷进行近红外光谱扫描。在进行光谱扫描之前, 肉眼观察并使用显微镜对稻谷进行初步检查, 使用记号笔在不破坏表皮的情况下标记霉斑位置。将 TENSOR II 型傅里叶变换近红外光谱仪预热 1 h, 扫描波数范围设定为 11 000~4 000 cm⁻¹, 分辨率为 8 cm⁻¹, 扫描 64 次, 样品重复扫描 3 次, 取其平均光谱^[29]。

1.3.3 建立定性模型的试验设计

以梯度随机采样法抽取样品量的 2/3 作为建模样品集用于模型的建立, 1/3 作为预测样品集用于模型的验证。在每次取样时参照 GB/T 4789.15-2016^[30]测定稻谷的霉菌菌落总数 (以 3 个托盘中稻谷霉菌菌落总数平均值为准) 以确定霉变程度。定性建模集样品与预测集样品信息如表 2 所示。

1.3.4 建立定量模型的试验设计

为了平衡未发霉稻谷和发霉稻谷的比例, 按样品类型定量随机抽取 150 组未结露稻谷和 150 组霉变稻谷。300 组样品以 3:1 的比例划分为建模组 (200 个) 和预测

组（100 个），采用 K-折交叉验证法，将获得的 300 组样品数据集分成 4 份（ $K=4$ ），选择其中 1 份（ $K=1$, $n=100$ ）作为验证集检验该模型的判别正确率；3 份（ $K=1$, $n=200$ ）作为模型训练集（样本集）验证所建模型的准确性和有效性。参照 GB/T 4789.15-2016^[30] 检测 300 组稻谷样品的霉菌菌落总数。定量建模样品信息如表 3 所示。

表 2 基于近红外对稻谷的霉变程度定性判别模型的建模集与预测集样品信息表

Table 2 Sample information table of modeling set and prediction set of NIR qualitative discriminant model for rice with different degrees of rice

样品 Sample	建模样品数 Number of modeling samples				预测样品数 Predicted sample number				总计 Total
	2018 年牡丹江 27 号	2019 年牡丹江 27 号	龙粳长粒香 Lonjing long grain aromatic rice	牡响 1 号 Muxiang No. 1	2018 年牡丹江 27 号	2019 年牡丹江 27 号	龙粳长粒香 Lonjing long grain aromatic rice	牡响 1 号 Muxiang No. 1	
	2018 Mudanjiang 27	2019 Mudanjiang 27			2018 Mudanjiang 27	2019 Mudanjiang 27			
未结露-霉变	40	40	40	40	20	20	20	20	240
结露-霉变初期	40	40	40	40	20	20	20	20	240
结露-霉变中期	40	40	40	40	20	20	20	20	240
结露-霉变后期	40	40	40	40	20	20	20	20	240

表 3 基于近红外对稻谷的霉变程度定量判别模型的建模集与预测集样品信息表

Table 3 Sample information table of modeling set and prediction set of NIR quantitative discriminant model for rice with different degrees of rice

样品 Sample	建模样品数 Number of modeling samples				预测样品数 Predicted sample number				总计 Total
	2018 年牡丹江 27 号	2019 年牡丹江 27 号	龙粳长粒香 Lonjing long grain aromatic rice	牡响 1 号 Muxiang No. 1	2018 年牡丹江 27 号	2019 年牡丹江 27 号	龙粳长粒香 Lonjing long grain aromatic rice	牡响 1 号 Muxiang No. 1	
	2018 Mudanjiang 27	2019 Mudanjiang 27			2018 Mudanjiang 27	2019 Mudanjiang 27			
未结露-霉变	25	25	25	25	12	13	12	13	150
结露-霉变初期	9	9	9	9	4	5	4	5	
结露-霉变中期	8	8	8	8	4	4	4	4	150
结露-霉变后期	8	8	8	8	4	4	4	4	

1.3.5 近红外图谱预处理方法与分析判别方法

本研究采用 TENSOR II 傅里叶变换近红外光谱仪及其自带的分析软件 OPUS 7.5 进行光谱预处理与分析^[31]。鉴于近红外仪器检测时可能受自身及环境因素影响导致图谱带有噪声，且提取的弱信息易重合，增加了分析复杂性。因此，在建模过程中，采用了多种预处理方式，包括不进行预处理、矢量归一化、一阶/二阶导数结合平滑处理等多种组合方法，以优化光谱数据。

在定性模型建立中，采用标准算法和因子化法两种模式识别算法，通过比较测试图谱与模型参考图谱之间的距离进行匹配，距离越短，匹配度越高^[32-33]。

在偏最小二乘法（partial least squares,PLS）定量分析模型的构建中，本研究融合了主成分分析（principal component analysis, PCA）、典型相关分析（canonical correlation analysis, CCA）及多元线性回归分析（multivariate linear regression, MLR），目的在于消除复共线关系，简化数据结构，并适用于非线性关系建模。PLS 是最流行的多变量校准技术，实际上是一种监督投影技术，类似于 PCA。该技术还将新的潜在变量（称为 PLS 因子）定义为显变量的线性组合，计算结果表示数据和响应之间的最高（PLS1）或剩余（较高 PLS 因子）协方差^[34-35]。为了实现数据可视化，将无监督多变量分析方法主成分分析（PCA）应用于光谱数据。这是一种降维方法，通过将更多的变量转换为更少的数字来减少大型数据集，而不会减少大型数据集中的信息^[36-40]。PLS 以其高准确

度和强稳定性，为粮食霉变的定量分析提供了可靠手段^[41-42]。这些处理方法和算法的应用，为粮食霉变的无损检测提供了科学有效的技术支撑^[43-44]。

评估 PLS 模型的关键指标包括决定系数（correlation coefficient of determination, R^2 ）、校正均方根误差（root mean square error of calibration, RMSEC）、预测均方根误差（root mean squared error of prediction, RMSEP）、交叉验证的均方根误差（root mean square error of cross-validation, RMSECV）和性能与偏差之比（ratio of performance to deviation, RPD）^[45]。其中， R^2 、RMSECV 和 RPD 共同反映模型精度与稳定性，而 RMSEP 则直接体现模型实用性。优质 PLSR 模型应具备高 RPD 和 R^2 ，低 RMSECV 和 RMSEP，且 $RMSECV>RMSEP$ 。更高的 R^2 和 RMSE 越低表示模型越好，RPD 越高，RPD 越高，模型可靠性越强^[46-47]。表 4 展示了 RPD 与模型精度的关系。

表 4 剩余预测偏差与模型精度的关系

Table 4 Relation between residual prediction deviation(RPD) and model accuracy

RPD	模型精度 Model accuracy
RPD<1.5	精度低，难以用于定性分析
2>RPD>1.5	模型可用于定性分析
2.5>RPD>2	精度一般，模型具有定量分析的潜力
3>RPD>2.5	精度良好，模型具有定量分析的可行性
RPD>3	精度高，模型可用于实际的定标分析

准确度计算方式为：正确的样本数除以整个样本数。样品之间的距离 S 值可以衡量定性模型的可行性，当

$S < 1$ 时, 表示两类样品之间存在相交关系, 模型的鉴别效果极差; 当 $S = 1$ 时, 表示两类样品之间是相切的关系, 模型的鉴别效果一般; 当 $S > 1$ 时, 表示两类样品是相离的关系, 数值越大, 模型的鉴别效果越好。判别准确率验证模型的精确度, 在满足 $S > 1$ 的条件下, 判别准确率越高模型的精确度越高。

2 结果与分析

稻谷在结露后容易导致霉变的发生, 因此, 需要对霉变稻谷进行快速识别。在实际的运输中, 稻谷的品种和年际不受限制, 而这两个因素对近红外光谱信息影响较大。在建立基于近红外光谱对结露稻谷的霉变判别模型时, 引入不同品种和年际的稻谷, 探究两因素对霉变的判别产生的影响, 将有助于提升霉变判别模型的普适性。

2.1 近红外光谱分析

图 1 展示了结露稻谷在不同霉变阶段的平均光谱图, 这些峰谱的变化代表着结露稻谷霉变过程中水分、脂质、淀粉和蛋白质的变化^[48]。这些峰谱的相似性反映了稻谷霉变的共性特征。在霉变过程中, 尽管霉菌分解了稻谷的营养物质, 但由于稻壳的保护, 谱峰形态保持相对稳定。为提升模型精度和效率, 建模时选择了 $4\,000 \sim 9\,500\text{ cm}^{-1}$ 波数范围, 避开了噪声较大的 $9\,500 \sim 11\,000\text{ cm}^{-1}$ 区域。

在定性模型建立中, 通过标准算法和因子化法评估了不同预处理方式的模型性能 (表 5~表 6)。

结果表明, 因子化法相较于标准算法表现出更好的模型稳定性和精确度。特别是在一阶导数+平滑预处理后

结合因子化法, 模型的鉴别效果最佳, 对预测集的判别准确率高达 93.00%, 表明此方法能有效区分不同霉变程度的稻谷。这归因于因子化法能够提取光谱的特征主成分, 而标准算法仅依赖平均值尺度识别, 提取的关键光谱信息有限。因此, 采用一阶导数+平滑+因子化法的建模策略, 为结露稻谷霉变程度的定性分析提供了高效准确的手段。

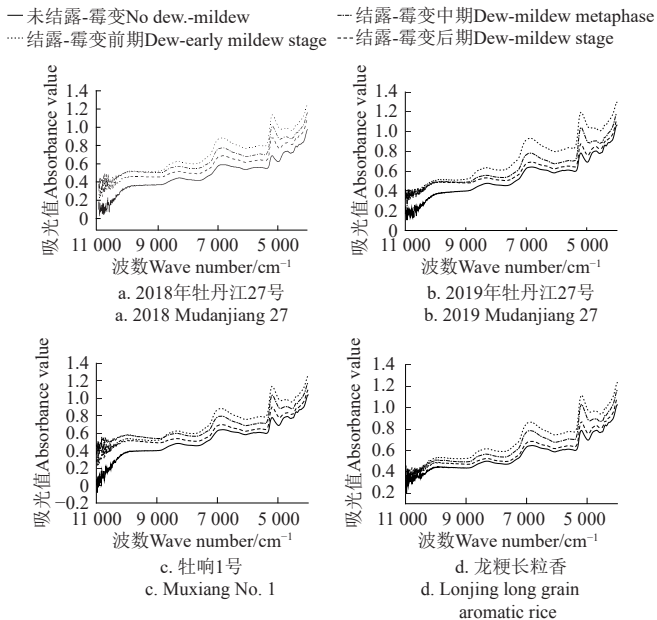


图 1 不同品种结露稻谷不同霉变时期的近红外平均光谱
Fig.1 Near-infrared average spectrum of different varieties of dewy rice in different mildew period

表 5 不同光谱预处理对标准算法建模准确度的影响
Table 5 Effect of different spectral data preprocessing methods on accuracy of standard algorithm modeling

预处理方式 Pretreatment mode	S 值 S value				判别准确率 Discrimination accuracy rate /%				平均判别准确率 Average discriminant accuracy/%
	未结露-霉变 No dew - mildew	结露-霉变初期 Dew - early mildew stage	结露-霉变中期 Dew - mildew metaphase	结露-霉变后期 Dew - mildew stage	未结露-霉变 No dew. - mildew	结露-霉变初期 Dew - early mildew stage	结露-霉变中期 Dew - mildew metaphase	结露-霉变后期 No dew. - mildew	
无预处理	0.582	0.908	0.905	0.587	67	92	79	21	64.75
矢量归一化	1.161	1.049	1.044	1.166	100	92	93	86	92.75
一阶导数+5 点平滑	0.545	0.272	0.273	0.543	87	38	43	93	65.25
一阶导数+9 点平滑	0.776	0.414	0.412	0.773	87	54	71	93	76.25
一阶导数+13 点平滑	0.947	0.532	0.532	0.942	93	77	79	86	83.75
一阶导数+17 点平滑	1.035	0.604	0.604	1.035	93	77	86	86	85.50
一阶导数+矢量归一化+5 点平滑	1.109	0.974	0.974	1.109	93	92	86	89	90.0
一阶导数+矢量归一化+9 点平滑	0.635	0.359	0.359	0.635	93	23	86	100	75.50
一阶导数+矢量归一化+13 点平滑	0.806	0.493	0.493	0.806	74	77	73	72	74.00
一阶导数+矢量归一化+17 点平滑	0.909	0.609	0.609	0.909	69	62	78	78	71.75
二阶导数+5 点平滑	0.089	0.072	0.072	0.089	0	0	7	0	1.75
二阶导数+9 点平滑	0.098	0.076	0.076	0.098	0	0	0	0	0.00
二阶导数+13 点平滑	0.131	0.091	0.091	0.131	0	0	0	0	0.00
二阶导数+17 点平滑	0.194	0.121	0.121	0.194	0	8	0	0	2.00
二阶导数+矢量归一化+5 点平滑	0.214	0.179	0.179	0.214	0	0	0	14	3.50
二阶导数+矢量归一化+9 点平滑	0.153	0.124	0.124	0.153	0	0	7	0	1.75
二阶导数+矢量归一化+13 点平滑	0.144	0.111	0.111	0.144	0	0	7	0	1.75
二阶导数+矢量归一化+17 点平滑	0.180	0.130	0.130	0.180	0	8	7	7	5.50

表 6 不同光谱预处理对因子化法建模准确度的影响
Table 6 Effects of different spectral data preprocessing methods on the accuracy of factorization modeling

预处理方式 Pretreatment mode	S 值 S value				判别准确率 Discrimination accuracy rate /%				平均判别 准确率 Average discriminant accuracy/%
	未结露-霉变 No dew - mildew	结露-霉变 初期 Dew - early mildew stage	结露-霉变 中期 Dew - mildew metaphase	结露-霉变 后期 Dew - mildew stage	未结露-霉变 No dew - mildew	结露-霉变 初期 Dew - early mildew stage	结露-霉变 中期 Dew - mildew metaphase	结露-霉变 后期 No dew. - mildew	
无预处理	0.942	0.530	0.533	0.945	93	77	79	86	83.75
矢量归一化	1.091	0.812	0.812	1.096	93	62	79	93	81.75
一阶导数+5 点平滑	1.042	1.042	1.127	1.204	93	100	86	86	91.25
一阶导数+9 点平滑	1.013	1.013	1.128	1.211	93	100	93	86	93.00
一阶导数+13 点平滑	1.001	1.005	1.129	1.213	93	100	93	79	91.25
一阶导数+17 点平滑	0.998	0.998	1.122	1.202	93	100	86	79	89.50
一阶导数+矢量归一化+ 5 点平滑	1.138	0.942	0.942	1.138	95	92	86	96	92.25
一阶导数+矢量归一化+ 9 点平滑	1.109	0.974	0.974	1.109	93	92	86	94	91.25
一阶导数+矢量归一化+ 13 点平滑	1.100	0.978	0.978	1.100	93	92	86	98	92.25
一阶导数+矢量归一化+ 17 点平滑	1.097	0.979	0.979	1.097	93	90	82	100	91.25
二阶导数+5 点平滑	0.403	0.057	0.057	0.216	7	0	21	0	7.00
二阶导数+9 点平滑	0.655	0.100	0.100	0.445	73	0	21	0	23.50
二阶导数+13 点平滑	0.651	0.434	0.434	0.651	87	15	43	71	54.00
二阶导数+17 点平滑	0.888	0.666	0.666	0.888	80	85	57	100	80.50
二阶导数+矢量归一化+ 5 点平滑	0.002	0.001	0.001	0.008	0	0	7	0	1.75
二阶导数+矢量归一化+ 9 点平滑	0.005	0.005	0.002	0.002	0	0	7	0	1.75
二阶导数+矢量归一化+ 13 点平滑	0.736	0.424	0.424	0.736	67	0	50	86	50.75
二阶导数+矢量归一化+ 17 点平滑	1.045	0.707	0.707	1.045	93	38	79	100	77.50

2.2 结露稻谷不同霉变程度定性模型的建立

在以标准算法为分析方法的定性模型中（表 5），只有矢量归一化法预处理下 S 值均大于 1，能够将建模集中结露稻谷的不同霉变程度进行区分，且对预测集判别准确率较高，为 92.75%。以因子化法分析的定性模型中（表 6），一阶导数+平滑的预处理方式具备较好的鉴别效果（ $S>1$ ）。一阶导数+5 点平滑/9 点平滑/13 点平滑均可以将结露稻谷的不同霉变程度的差异最大化，保证模型的精确度，对预测集的判别准确率最高可达 93.00%（一阶导数+9 点平滑+因子化法），这说明一阶导数+平滑+因子化法的建模方式对不同霉变程度的判别稳定性较好。

综合来看，因子化法比标准算法具备更好的模型稳定性和精确度，这主要归功于因子化法可以提取光谱的特征主成分参与到模式识别中，而标准算法只是平均值尺度的识别，提取的关键光谱信息不足

采用一阶导数+9 点平滑+因子化法建立的判别模型，结果如图 2 所示，能明显区分结露稻谷的不同霉变程度，展现出高稳定性和准确性。尽管“结露-霉变前期”与“结露-霉变中期”略有重叠，这可能与霉菌生长速度和人工分辨误差有关。稻谷品种和年限对聚类区块分离趋势产生影响，尤其在“未结露-霉变”和“结露-霉变后期”最为显著。

2.3 结露稻谷霉菌总数的定量模型的建立

采用不同的光谱预处理方式建立稻谷霉菌总数的定量分析模型如表 7 所示，结果显示，不进行预处理、消

除常数偏移量、减去一条直线、最小-最大归一化法和多元散射校正等方法对建模集均有良好表现。但求导预处理导致模型精度和稳定性下降^[49]。相比而言，矢量归一化法、最小-最大归一化法和多元散射校正提高了模型精度。

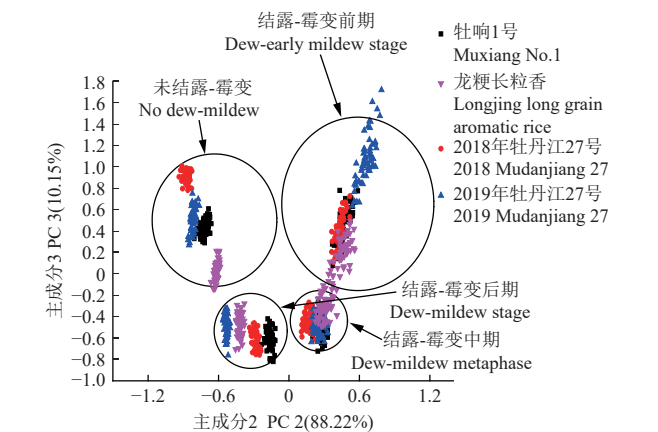


图 2 结露稻谷不同霉变程度的因子分析判别函数得分图
Fig.2 Factor analysis discriminant function score of rice with different degrees of mildew after dew-formation

以不同光谱预处理方式建立稻谷霉菌总数的定量分析模型，矢量归一化法、最小-最大归一化法和多元散射校正则都提高了模型的精度。但对光谱进行一阶导数加平滑处理后，建模维数降低，说明求导处理减少了原始图谱的信息，有效信息也同时减少，降低了模型的精度和稳定性。因此结合建模结果采用适量归一化法建立定量模型。在定性模型建立过程中，因子化法可以提取光谱的特征主成分参与到模式识别，一阶导数平滑均可以

将结露稻谷的不同霉变程度的差异最大化，保证模型的精确度，对预测集的判别准确率最高。

表 7 不同光谱预处理方式对定量分析模型的影响
Table 7 Effects of different spectral data preprocessing methods on quantitative analysis models

光谱预处理方式 Spectral pretreatment method	交叉验证的 均方根误差 RMSECV	R^2	维数 Dimension	RPD
无预处理 No preprocessing	0.576	0.866 5	9	2.74
消除常数偏移量 Eliminate constant offsets	0.577	0.865 7	9	2.73
减去一条直线 Subtract a line	0.599	0.855 5	8	2.63
矢量归一化法 Vector normalization method	0.546	0.879 7	9	2.88
最小-最大归一化法 Minimum-maximum normalization method	0.542	0.881 6	8	2.91
多元散射校正 Multivariate scattering correction	0.558	0.874 7	9	2.83
一阶导数+5 点平滑 First derivative +5 points smooth	0.801	0.741 5	3	1.97
一阶导数+9 点平滑 First derivative +9 point smoothing	0.774	0.758 4	4	2.04
一阶导数+13 点平滑 First derivative +13 points smooth	0.759	0.768 1	7	2.08
一阶导数+17 点平滑 First derivative +17 points smooth	0.694	0.805 9	7	2.27
二阶导数+5 点平滑 Second derivative +5 points smooth	0.352	0.712 9	5	1.21
二阶导数+9 点平滑 Second derivative +9 points smooth	0.389	0.762 3	4	1.28
二阶导数+13 点平滑 Second derivative +13 points smooth	0.432	0.811 7	3	1.33
二阶导数+17 点平滑 Second derivative +17 points smooth	0.497	0.822 3	2	1.41
一阶导数+减去一条直线+5 点平滑 First derivative + minus a straight line +5 points of smoothing	0.718	0.837 0	3	1.88
一阶导数+减去一条直线+9 点平滑 First derivative + minus a straight line +9 points of smoothing	0.702	0.860 0	3	1.83
一阶导数+减去一条直线+13 点平滑 First derivative + minus a straight line +13 points of smoothing	0.729	0.820 0	6	1.92
一阶导数+矢量归一化+5 点平滑 First derivative + vector normalization + 5-point smoothing	0.769	0.758 0	3	2.08
一阶导数+矢量归一化+9 点平滑 First derivative + vector normalization +9 point smoothing	0.752	0.784 0	3	2.01
一阶导数+矢量归一化+13 点平滑 First derivative + vector normalization + 13-point smoothing	0.756	0.778 0	4	2.02
一阶导数+多元散射校正+5 点平滑 First derivative + multiple scattering correction +5 point smoothing	0.764	0.766 0	3	2.06
一阶导数+多元散射校正+9 点平滑 First derivative + multiple scattering correction +9 point smoothing	0.745	0.796 0	3	1.98
一阶导数+多元散射校正+13 点平滑 First derivative + multivariate scattering correction + 13-point smoothing	0.747	0.792 0	4	1.99

运用近红外光谱建立预测模型时，化学值异常和光谱异常的存在会严重影响模型的精度，需剔除化学值和光谱异常数据以确保模型精度^[50]。如表 8 所示，经过优化，矢量归一化法成为最优预处理方式。使用该方法建立的 PLS 分析模型在预测集上表现出色，RPD 为 3.24，RMSEP 为 0.450，证明了其高预测准确性。

关于 PLS 选择的变量数，首先通过不同的光谱预处

理方法（矢量归一化法、最小-最大归一化法和多元散射校正）建立了稻谷霉菌总数的定量分析模型。其中化学值异常和光谱异常的存在会对其模型产生影响，因此需剔除这些异常数据以确保模型的准确性。通过研究发现，矢量归一化法是最优的预处理方式，其关键指标为：RMSECV 为 0.546， R^2 为 0.8797，维数为 9，RPD 为 2.88。经过多次调整维数，得到不同的均方根误差（RMSE）和 R^2 值，选择 RMSE 最低、 R^2 最高的维数作为最佳维数。而维数的选择直接影响模型的复杂性和预测能力，为避免过拟合或欠拟合，进一步地进行逐步实验，最终确定维数为 9 时模型性能最佳。因此，PLS 的变量数为 9。

通过矢量归一法预处理和 PLS 定量分析拟合的稻谷菌落总数与近红外光谱预测值的相关性结果如图 3，建模集和验证集的样品点均匀分布在拟合线两侧，显示出高相关性。不同年际和品种的稻谷样品在各菌落总数区段分布均匀，表明近红外光谱结合 PLS 模型可定量检测结露稻谷的菌落总数，且受“年际”和“品种”因素影响较小。

表 8 校正后的定量分析模型的精度
Table 8 Precision of the quantitative analysis model after correction

光谱预处理方式 Spectral pretreatment method	RMSECV	R^2	RPD	预测均方 根误差 RMSEP
矢量归一化法 Vector normalization method	0.470	0.904 5	3.24	0.450
最小-最大归一化法 Minimum-maximum normalization method	0.517	0.886 9	3.01	0.490
多元散射校正 Multivariate scattering correction	0.501	0.898 5	3.14	0.466

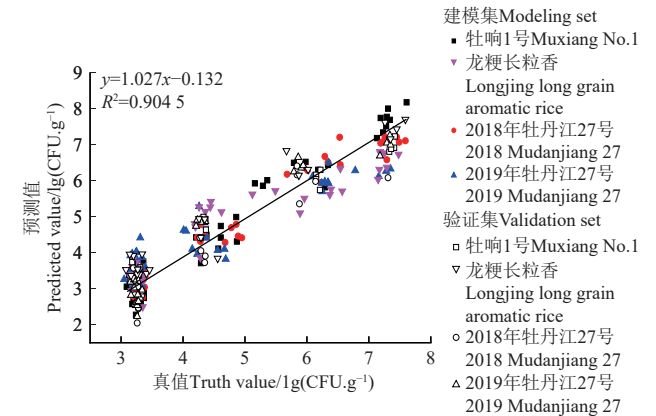


图 3 稻谷菌落总数与近红外光谱预测值的相关关系
Fig.3 Correlation between total number of rice colonies and predicted value of near-infrared spectrum

3 结论

以不同温湿度条件下稻谷的霉变程度作为建模研究样品，建立了不同霉变程度的定性和定量分析，模型可用于不同温湿度运输条件下稻谷的霉变情况判别，霉变依据国家标准霉变粒率，以及稻谷霉菌数用于样品建模分级，本研究制备不同霉变程度稻谷作为研究对象，通过从感官指标、霉变粒以及霉菌菌落总数界定稻谷的霉变程度。基于因子化和偏最小二乘算法对稻谷霉变快速

分级预测, 主要结论如下:

1) 基于近红外光谱对4种稻谷样品建立了其不同霉变程度的定性判别模型, 其中一阶导数+9点平滑+因子化法建立的定性判别模型准确度较高, 样品之间的距离 S 均值大于1, 分辨效果好, 借助留一交互验证验证模型的平均准确率为93.00%。

2) 基于近红外光谱对4种共300组稻谷样品建立了霉菌菌落总数的定量模型, 通过矢量归一化法+偏最小二乘法(partial least squares, PLS)定量分析建立的定量判别模型, 其交叉验证的均方根误差(root mean square error of cross-validation, RMSECV)、决定系数(correlation coefficient of determination, R^2)、性能与偏差之比(ratio of performance to deviation, RPD)和预测均方根误差(root mean squared error of prediction, RMSEP)分别为0.47、0.904 5、3.24和0.45, 经过分析方法优选后而建立的霉变判别模型显示, 霉变是影响近红外光谱变化的主导因素, 而稻谷的品种与年际对其影响较小。

以上定性定量模型可以对不同运输过程中的稻谷实现快速预测其霉变程度或其霉菌数量, 未来可研发模型搭载平台由实验室大型精密设备转化为便携式小型设备, 研制在线实时监测装备用于集装箱内粮食霉变情况监控。

[参 考 文 献]

- [1] 郑凯闻. 宏观调控: 粮食流通顺畅有序市场运行总体平稳[J]. *中国粮食经济*, 2020(2): 38-40.
ZHENG Kaiwen. Macro-control: Smooth and orderly grain circulation and stable market operation on the whole[J]. *China Grain Economy*, 2020(2): 38-40.
- [2] 温富荣, 王斌兴, 张镇. 港口散粮物流中长运距、大运量输送系统的应用[J]. *现代食品*, 2019(20): 16-18.
WEN Furong, WANG Binxing, ZHANG Zhen. Application of long-haul and large-volume conveying system in port bulk grain logistics[J]. *Modern Food*, 2019(20): 16-18.
- [3] 向长琼, 周浩, 张华昌, 等. “北粮南运”散粮集装箱储运建设的思考[J]. *粮油仓储科技通讯*, 2018, 34(6): 1-2, 11.
XIANG Changqiong, ZHOU Hao, ZHANG Huachang, et al. Thinking on the construction of bulk grain container storage and transportation of "grain from North to South"[J]. *Grain and Oil Storage Technology Bulletin*, 2018, 34(6): 1-2, 11.
- [4] 孙丽颖. 着力化解“北粮南运”融资难题[N]. *金融时报*, 2018-06-25(010).
- [5] 张婷, 徐圆程, 王光宇, 等. 中国南方稻谷主产区黑曲霉菌群的流行特征及产毒能力研究[J]. *中国粮油学报*, 2022, 37(12): 1-8.
ZHANG Ting, XU Yuancheng, WANG Guangyu, et al. Study on epidemic characteristics and toxic production capacity of *Aspergillus Niger* in major rice producing areas of Southern China [J]. *Journal of China grain and oil*, 2022 ((12) : 1-8. (in Chinese with English abstract)
- [6] 文韬, 洪添胜, 李立君, 等. 基于高光谱技术的霉变稻谷脂肪酸含量无损检测[J]. *农业工程学报*, 2015, 31(18): 233-239.
WEN Tao, HONG Dingsheng, LI Lijun, et al. Non-destructive detection of fatty acid content in mildew rice based on hyperspectral technology[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2015, 31(18): 233-239. (in Chinese with English abstract)
- [7] 马世榜, 汤修映, 徐杨, 等. 可见/近红外光谱结合遗传算法无损检测牛肉 pH 值[J]. *农业工程学报*, 2012, 28(18): 263-268.
MA Shibang, TANG Xiuying, XU Yang, et al. Non-destructive detection of beef pH value by visible/near-infrared spectroscopy combined with genetic algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2012, 28(18): 263-268. (in Chinese with English abstract)
- [8] 彭望, 王前, 廖庆喜, 等. 基于近红外光谱的便携式油菜籽品质检测装置研制[J]. *农业工程学报*, 2024, 40(18): 13-22.
PENG Wang, WANG Qian, LIAO Qingxi, et al. Development of portable rapeseed quality detection device based on near infrared spectroscopy [J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2019, 40(18): 13-22. (in Chinese with English abstract)
- [9] 朱文杰, 黄文倩, 祝清震, 等. 手持式黄油桃可溶性固形物可见近红外光谱检测设备研制[J]. *农业工程学报*, 2024, 40(01): 286-292.
ZHU Wenjie, HUANG Wenqian, ZHU Qingzhen, et al. Development of handheld visible near-infrared spectroscopy detection equipment for soluble solid of nectarine [J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2019, 40(01): 286-292. (in Chinese with English abstract)
- [10] QU J H, LIU D, CHENG J H, et al. Applications of near-infrared spectroscopy in food safety evaluation and control: A review of recent research advances[J]. *Critical Reviews in Food Science and Nutrition*. 2015;55(13): 1939-1954.
- [11] 刘爽, 柴春祥. 近红外光谱技术在水产品检测中的应用进展[J]. *食品安全质量检测学报*, 2021, 12(21): 8590-8596.
LIU Shuang, CHAI Chunxiang. The progress in the application of near infrared spectroscopy in the detection of aquatic products [J]. *Journal of food safety and quality testing*, 2021, 12 (21) : 8590-8596.(in Chinese with English abstract)
- [12] OZAKI Y, MCCLURE W F, CHRISTY A A. Near-infrared Spectroscopy in Food Science and Technology[M]. Hoboken: John Wiley & Sons, 2006: 407.
- [13] 王一博, 孙瑞鸿, 王长生, 等. 采用量子力学方法研究带正电氨基酸分子复合物间非键作用的强度[J]. *化学研究*, 2024, 35(01): 61-67.
WANG Yibo, SUN Ruihong, WANG Changsheng, et al. Study on the strength of non-bond interaction between positively charged amino acid molecular complexes by quantum mechanics [J]. *Chemical Research*, 2024, 35 (1) : 61-67. (in Chinese with English abstract)
- [14] 杨东, 王舒卉, 吴建华, 等. 玉米籽粒霉变等级高光谱图像检测方法研究[J]. *中国粮油学报*, 2022, 37(11): 46-53.
YANG Dong, WANG Shuhui, WU Jianhua, et al. Level of corn grain mildew hyperspectral image detection method research [J]. *Chinese journal of grain and oil*, 2022 ,37(11) : 46-53. (in Chinese with English abstract)
- [15] 杨思成, 曹阳, 赵会义, 等. 粮食霉变及其快速检测技术研究进展[J]. *粮食与油脂*, 2018, 31(7): 21-23.

- YANG Sicheng, CAO Yang, ZHAO Huiyi, et al. Research progress of grain mildew and its rapid detection technology[J]. *Food and Oil*, 2018, 31(7): 21-23. (in Chinese with English abstract)
- [16] CEN, H Y, YONG H. Theory and application of near infrared reflectance Spectroscopy in determination of food quality[J]. *Trends in food science & Technology*. 2007, 18(2): 72-83.
- [17] 张仲源, 刘静, 管晓, 等. 近红外光谱分析技术在食品检测中的应用研究进展[J]. *食品与发酵工业*, 2011, 37(11): 159-165.
- ZHANG Zhongyuan, LIU Jing, GUAN Xiao, et al. Application of near infrared spectral analysis technology in food detection research progress [J]. *Journal of Food and Fermentation Industry*, 2011 (11): 159-165. (in Chinese with English abstract)
- [18] TAO F, YAO H, HRUSKA Z, et al. Detection of aflatoxin B1 on corn kernel surfaces using visible-near infrared spectroscopy[J]. *Journal of Near Infrared Spectroscopy*, 2020, 28(2): 59-69.
- [19] CHEN J, ZHU S P, ZHAO G H. Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR[J]. *Food chemistry*. 2017, 221: 1939-1946.
- [20] 龚润华, 杨信廷, 郭晓晖, 等. 基于近红外光谱技术的红小豆品质多指标无损检测及品种鉴别方法研究[J]. *中国粮油学报*, 2024, 39(10): 190-197.
- GONG Runhua, YANG Xinting, GUO Xiaohui, et al. Study on multi-index nondestructive testing and variety identification method of red bean based on near infrared spectroscopy [J]. *Journal of China Grain and Oil*, 2024, 33 (10) 6: 190-197. (in Chinese with English abstract)
- [21] 蒋雪松, 赵天霞, 刘潇, 等. 阵列式光纤光谱仪的小麦霉变在线检测[J]. *光谱学与光谱分析*, 2018, 38(12): 3729-3735.
- JIANG Xuesong, ZHAO Tianxia, LIU Xiao, et al. On-line detection of wheat mildew by array optical fiber spectrometer[J]. *Spectroscopy and Spectral Analysis*, 2018, 38(12): 3729-3735. (in Chinese with English abstract)
- [22] 吕都, 唐健波, 徐廷霞, 等. 近红外光谱技术快速鉴别稻谷霉菌污染的研究[J]. *食品科技*, 2021, 46(6): 301-306.
- LV Du, TANG Jianbo, XU Tingxia, et al. Mould contamination of near infrared spectrum technology quickly identifying rice research [J]. *Journal of Food Science and Technology*, 2021, 46(6) : 301-306. (in Chinese with English abstract)
- [23] 蒋大鹏, 张冬妍, 李丹丹, 等. 运用近红外光谱技术对松子霉变的快速检测[J]. *东北林业大学学报*, 2019, 47(5): 83-87, 99.
- JIANG Dapeng, ZHANG Dongyan, LI Dandan, et al. Rapid detection of pine nut mildew by near infrared spectroscopy[J]. *Journal of Northeast Forestry University*, 2019, 47(5): 83-87, 99. (in Chinese with English abstract)
- [24] 蒋大鹏. 基于改进近红外漫反射技术与流形学习的东北松子霉变分类研究[D]. 哈尔滨: 东北林业大学, 2019.
- JIANG Dapeng. Research on Mildew Classification of Pine Nuts in Northeast China Based on Improved Near-infrared Diffuse Reflection Technology and Manifold Learning [D]. Harbin: Northeast Forestry University, 2019. (in Chinese with English abstract)
- [25] 李晋, 张琛, 刘红, 等. 近红外光谱联合化学计量学在柑橘类水果质量无损检测方面的最新研究及应用进展[J]. *食品与发酵工业*, 2024, 50(5): 367-379.
- LI Jin, ZHANG Chen, LIU Hong, et al. Recent research and application progress of near infrared spectroscopy combined with chemometrics in non-destructive detection of citrus fruit quality [J]. *Food and Fermentation Industry*, 2024, 50 (5) : 367-379. (in Chinese with English abstract)
- [26] BILAL M. 基于近红外光谱、中红外光谱和比色传感器阵列的花生 (*Arachis hypogaea* L.) 品质无损检测研究[D]. 镇江: 江苏大学, 2023.
- BILAL M. Study on Non-destructive Detection of Peanut (*Arachis hypogaea* L.) Quality Based on Near Infrared Spectroscopy, Mid-infrared Spectroscopy and Colorimetric Sensor Array [D]. Zhenjiang: Jiangsu university, 2023. (in Chinese with English abstract)
- [27] 陈昂. 番茄内外品质在线无损检测与分级系统研发[D]. 杨凌: 西北农林科技大学, 2023.
- CHEN Ang. Development of Online Non-destructive Testing and Grading System for Internal and External Quality of Tomato [D]. Yangling: Northwest A&F and Forestry University of Science and Technology, 2023. (in Chinese with English abstract)
- [28] 徐赛, 陆华忠, 王旭, 等. 基于可见/近红外光谱的菠萝水心病无损检测[J]. *农业工程学报*, 2021, 37(21): 287-294.
- XU Sai, LU Huazhong, WANG Xu, et al. Non-destructive detection of heart disease in pineapple water based on visible/near-infrared spectroscopy [J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 201, 37(21): 287-294. (in Chinese with English abstract)
- [29] 郭俊先, 张振振, 韩景, 等. 基于傅里叶变换近红外光谱的籽棉含水率无损检测[J]. *农业工程学报*, 2023, 39(21): 152-160.
- GUO Junxian, ZHANG Zhenzhen, HAN Jing, et al. Nondestructive detection of moisture content of seed cotton based on Fourier transform near infrared spectroscopy[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2023, 39(21): 152-160. (in Chinese with English abstract)
- [30] 马群飞. GB 4789.15-2016《食品安全国家标准食品微生物学检验霉菌和酵母计数》标准解读[J]. *中国卫生标准管理*, 2018, 9(05): 1-3.
- [31] 杨琼. 近红外光谱法定量分析及其应用研究[D]. 重庆: 西南大学, 2009.
- YANG Qiong. Research on Quantitative Analysis and Application of Near Infrared Spectroscopy [D]. Chongqing: Southwest University, 2009. (in Chinese with English abstract)
- [32] 崔程, 刘翠玲, 孙晓荣, 等. 基于近红外高光谱成像技术的花生冻伤检测方法研究[J]. *食品工业科技*, 2024, 45(6): 226-233.
- CUI Cheng, LIU Cuiling, SUN Xiaorong, et al. Set based on near infrared specular like technology of peanut frostbite detection method research [J]. *Journal of food science and technology*, 2024, 45(6) : 226-233. (in Chinese with English abstract)
- [33] 吴静珠, 李慧, 王克栋, 等. 光谱预处理在农产品近红外模型优化中的应用研究[J]. *农机化研究*, 2011, 33(3): 178-181.

- WU Jingzhu, LI Hui, WANG Kedong, et al. Spectral preprocessing in agricultural products for the application of near infrared model optimization [J]. *Journal of Agricultural Mechanization Research*, 2011, 33(3): 178-181. (in Chinese with English abstract)
- [34] DECONINCK E. Full characterisation of heroin samples using infrared spectroscopy and multivariate calibration[J]. *Molecules* (Basel, Switzerland). 2024, 29, (5): 1116.
- [35] 寇泽坤, 陈国通, 李思雨, 等. 拉曼光谱结合化学计量学方法鉴别糖浆掺假蜂蜜[J]. *食品科学*, 2024, 45(1): 254-260.
- KOU Zekun, CHEN Guotong, LI Siyu, et al. Identification of adulterated honey in syrup by Raman spectroscopy and chemometrics [J]. *Food Science*, 2024, 45(1): 254-260. (in Chinese with English abstract)
- [36] JOSHI R, SATHASIVAM R, PARK S U, et al. Application of fourier transform infrared spectroscopy and multivariate analysis methods for the non-Destructive evaluation of phenolics compounds in moringa powder[J]. *Agriculture* (Basel). 2022, 12, (1): 10.
- [37] JOSHI R, SATHASIVAM R, JAYAPAL P K, et al. Comparative determination of phenolic compounds in arabidopsis thaliana leaf powder under distinct stress conditions using fourier-transform infrared (FT-IR) and near-infrared (FT-NIR) Spectroscopy[J]. *Plants* (Basel). 2022, 11,(7): 836.
- [38] 郭茂, 黄忠宇, 汪杰, 等. 基于激光诱导击穿光谱技术结合机器学习算法的3种干腌火腿产地识别[J]. *中国食品学报*, 2022, 22(10): 279-285.
- GUO MAO, HUANG Zhongyu, WANG Jie, et al. Origin identification of three kinds of dry-cured ham based on laser induced breakdown spectroscopy combined with machine learning algorithm[J]. *Acta Food Sinica of China*, 2022, 22(10): 279-285. (in Chinese with English abstract)
- [39] 刘燕德, 刘良峰, 李雄, 等. 基于高光谱成像结合波段比和改进分水岭分割算法的草莓早期损伤检测[J]. *中国农业大学学报*, 2024, 29(11): 50-62.
- LIU Yande, LIU Liangfeng, LI Xiong, et al. Early damage detection of strawberry based on hyperspectral imaging combined with band ratio and improved watershed segmentation algorithm [J]. *Journal of China Agricultural University*, 2019, 29(11): 50-62. (in Chinese with English abstract)
- [40] 张乐, 吴静珠, 李江波, 等. 基于随机森林的单粒玉米种子水分近红外快速定量检测[J]. *中国粮油学报*, 2021, 36(12): 114-119.
- ZHANG Le, WU Jingzhu, LI Jiangbo, et al. Rapid near-infrared quantitative detection of water content of single maize seed based on random forest [J]. *Journal of Cereals and Oils*, 2019, 36(12): 114-119. (in Chinese with English abstract)
- [41] 柳艳云, 胡昌勤. 近红外分析中光谱波长选择方法进展与应用[J]. *药物分析杂志*, 2010, 30(5): 968-975.
- LIU Yanyun, HU Changqin. Progress and application of spectral wavelength selection methods in near-infrared analysis[J]. *Journal of Pharmaceutical Analysis*, 2010, 30(5): 968-975. (in Chinese with English abstract)
- [42] 梁浩, 黄生志, 孟二浩, 等. 基于多种混合模型的径流预测研究[J]. *水利学报*, 2020, 51(1): 112-125.
- LIANG Hao, HUANG Shengzhi, MENG Erhao, et al. Based on a variety of hybrid model of runoff prediction research [J]. *Journal of Water Conservancy*, 2020 51(1): 112-125. (in Chinese with English abstract)
- [43] 肖宏辉, 李春霖, 张永志, 等. 无损检测技术在水果品质安全和真实性的应用研究进展[J]. *核农学报*, 2024, 38(4): 736-745.
- XIAO Honghui, LI Chunlin, ZHANG Yongzhi, et al. Application of non-destructive testing technology in fruit quality safety and authenticity [J]. *Chinese Journal of Nuclear Agriculture*, 2019, 38(4): 736-745. (in Chinese with English abstract)
- [44] 李彦龙, 钱丽丽, 陈明明, 等. 元素指纹在茶叶产地溯源中的应用及其影响因素研究进展[J/OL]. *食品科学*, 1-14[2024-10-25]. <http://kns.cnki.net/kcms/detail/11.2206.TS.20240821.1951.085.html>.
- LI Yanlong, QIAN Lili, CHEN Mingming, et al. Application of element fingerprint in tea origin tracing and its influencing factors [J/OL]. *Food Science*, 1-14 [2024-10-25]. <http://kns.cnki.net/kcms/detail/11.2206.TS.20240821.1951.085.html>. (in Chinese with English abstract)
- [45] 张永立, 杨光辉, 王美蟠, 等. 鲜食玉米果穗含水率近红外光谱无损检测影响因素[J]. *农业工程学报*, 2024, 40(15): 262-270.
- ZHANG Yongli, YANG Guanghui, WANG Meipan, et al. Factors affecting water content of fresh corn ear by NIR spectroscopy [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2024, 40(15): 262-270. (in Chinese with English abstract)
- [46] FAZELI B N, AFKARI S A H, EBRAHIM T. Prediction of some quality properties of rice and its flour by near - infrared spectroscopy (NIRS) analysis[J]. *Food Science & Nutrition*. 2021, 9(2): 1099-1105.
- [47] SRISAWAS W, JINDAL V K, THANAPASE W. Relationship between sensory textural attributes and near infrared spectra of cooked rice[J]. *Journal of Near Infrared Spectroscopy*, 2007, 15(5): 333-340.
- [48] 万立昊, 渠琛玲, 薛飞, 等. 稻谷霉变程度与敏感品质变化关系研究[J]. *粮食与油脂*, 2019, 32(9): 71-75.
- WAN Lihao, QU Chenling, XUE Fei, et al. Study on the relationship between mildew degree and sensitive quality of rice[J]. *Grains and Fats*, 2019, 32(9): 71-75. (in Chinese with English abstract)
- [49] 姜微. 高光谱技术在马铃薯品种鉴别及品质无损检测中的应用研究[D]. 哈尔滨: 东北农业大学, 2017.
- JIANG Wei. Application of Hyperspectral Technology in Potato Variety Identification and Quality Nondestructive Testing [D]. Harbin: Northeast Agricultural University, 2017. (in Chinese with English abstract)
- [50] 黄瑞娟. 红外光谱技术在食品检测中的应用[J]. *现代测量与实验室管理*, 2015, 23(1): 9-14.
- HUANG Ruijuan. Application of infrared spectroscopy in food detection[J]. *Modern Measurement and Laboratory Management*, 2015, 23(1): 9-14. (in Chinese with English abstract)

Rapid grading prediction of mould in rice grains based on factorisation and partial least squares algorithm

ZUO Feng¹, HAO Zhenyu², ZHANG Caidong¹, DONG Zhihao¹, QIAN Lili^{1*}, ZHANG Tao³,
LIU Xingquan⁴, ZHOU Guoxin⁴

(1. College of Food Science, Heilongjiang Bayi Agricultural University, Daqing 163319, China; 2. College of Food Science, Jiangnan University, Wuxi 214122, China; 3. National Food and Strategic Reserves Administration, Beijing 100834, China; 4. College of Food and Health, Zhejiang A & F University, Hangzhou 311300, China)

Abstract: Food spoilage caused by mold has posed a major threat to grain quality and national food safety. Mold can also produce mycotoxins damage to the nutritional value of cereals, and even be harmful to human health. It is highly required for efficient and reliable detection. Traditional detection of the mold in grains can often involve time-consuming laboratory tests that rely heavily on specialized equipment. It is very necessary to realize the rapid on-site assessments. Alternatively, near-infrared spectroscopy (NIRS) can be expected to rapidly detect the mold of agricultural products, due mainly to the detection speed, non-destructive testing, repeatability, and easy online analysis. The purpose of this study was to establish a mold detection model in rice using NIRS. A systematic investigation was also implemented to rapidly distinguish rice from the different levels of mold. A dataset of 960 samples was taken from four varieties with different mold degrees (2018 Mudanjiang 27, 2019 Mudanjiang 27, Longjing long grain aromatic rice, and Muxiang 1). Qualitative discrimination models were then realized for the different degrees of mold contamination. The first derivative was combined with 9-point smoothing and factorization during preprocessing, in order to obtain the qualitative discriminant model with high accuracy. The mean S value greater than 1 represented the excellent performance to distinguish between mildewed and non-mildewed rice. The accuracy of the model was further verified by a leave-one cross-validation. The accuracy reached 93%. In addition, 300 independent data sets of rice samples were also utilized with the different degrees of mildew. The total number of mold colonies was quantitatively characterized using NIRs. A discrimination model was then established by vector normalization and partial least squares (PLS) method. Some indexes were calculated to evaluate the accuracy of cross-validation root-mean-square error (RMSECV), determination coefficient (R^2), and performance deviation ratio (RPD). The prediction root-mean-square error (RMSEP) was also used to evaluate the accuracy of the model. The results showed that the values of RMSECV, R^2 , RPD, and RMSEP were 0.470, 0.904 5, 3.24, and 0.45, respectively. The first three parameters indicated the high precision of the model, while the latter indicated the high accuracy of the model. Therefore, the improved model with high accuracy was achieved to rapidly predict the mold in grains. The mold was the main influencing factor on the variation of the NIR spectrum after optimization. On the contrary, there were relatively small effects of rice variety and harvest year on the spectral characteristics. Therefore, the NIRs can be highlighted to detect mold contamination in rice, regardless of variety or time differences. This finding can provide a strong reference for the rapid prediction of the mildew degree or amount of rice in different transportation using NIR spectroscopy. Online real-time monitoring equipment can also be offered to monitor the grain mildew in containers.

Keywords: near-infrared; spectroscopy; classification prediction; paddy; degree of mould; total number of bacterial colonies; vector normalization; PLS quantitative analysis