

基于近红外光谱技术快速预测菊芋主要理化指标

李心怡¹, 徐小航¹, 韩少钦², 阚启鑫¹, 谭悦¹, 王祺皓¹, 刘果³, 李院平², 韩勇军², 肖苏尧¹,
宋明月¹, 段杉^{1*}, 曹庸^{1*}

(1. 华南农业大学食品学院, 广东省功能食品活性物重点实验室, 广东广州 510642)

(2. 湖南贵威生物科技有限公司, 湖南永州 425500) (3. 仲恺农业工程学院轻工食品学院, 广东广州 510225)

摘要: 菊糖、蛋白质、总多酚、灰分是菊芋中重要的理化成分, 会影响菊芋的营养特性和加工特性, 而这些成分的传统检测方法往往耗费大量化学试剂, 对设备要求较高。为了实现菊芋中主要理化指标的快速预测, 该研究以 56 份菊芋为研究对象, 利用近红外 (Near-Infrared, NIR) 光谱仪采集菊芋 900~2 500 nm 的光谱数据, 并结合其菊糖、蛋白质、总多酚、灰分的实测含量, 建立偏最小二乘回归 (Partial Least Squares Regression, PLSR) 预测模型。结果表明, 经过异常样本剔除、光谱预处理、特征波段筛选等模型建立过程, 最终得到菊糖、蛋白质、总多酚、灰分最优定量模型的预测相关系数 (R^2_p) 分别为 0.869 5、0.930 7、0.943 3、0.920 6, 预测均方根误差 ($RMSEP$) 分别为 2.237 2 g/100 g、0.363 4 g/100 g、0.107 7 mg/g、0.158 1 g/100 g, 模型的预测能力良好。外部验证结果显示, 通过近红外光谱和标准化学方法得到的检测结果之间无显著差异 ($P>0.05$)。该研究为 NIR 应用于菊芋理化成分的快速预测提供了理论依据。

关键词: 菊芋; 近红外漫反射法; 含量预测; 特征波段筛选; 偏最小二乘回归

文章编号: 1673-9078(2026)4-313-326

DOI: 10.13982/j.mfst.1673-9078.2026.4.0130

Rapid Prediction of the Main Physicochemical Indicators in Jerusalem Artichoke by Using Near-infrared Spectroscopy

LI Xinyi¹, XU Xiaohang¹, HAN Shaoqin², KAN Qixin¹, TAN Yue¹, WANG Qihao¹, LIU Guo³, LI Yuanping²,
HAN Yongjun², XIAO Suyao¹, SONG Mingyue¹, DUAN Shan^{1*}, CAO Yong^{1*}

(1. College of Food Science, Guangdong Provincial Key Laboratory of Nutraceuticals and Functional Foods, South China Agricultural University, Guangzhou 510642, China) (2. Hunan Guiwei Biotechnology Co. Ltd., Yongzhou 425500, China) (3. College of Light Industry and Food Science, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China)

Abstract: Inulin, protein, total polyphenols, and ash are important physicochemical components of the Jerusalem artichoke. They have an impact on Jerusalem artichoke's processing and nutritional characteristics. Traditional detection methods for these components often require a large number of chemical reagents and high-standard equipment. To achieve rapid prediction of the main physicochemical indicators in the Jerusalem artichoke, 56 Jerusalem artichoke samples were examined in this study. Spectral

引文格式:

李心怡, 徐小航, 韩少钦, 等. 基于近红外光谱技术快速预测菊芋主要理化指标[J]. 现代食品科技, 2026, 42(4): 313-326.

LI Xinyi, XU Xiaohang, HAN Shaoqin, et al. Rapid prediction of the main physicochemical indicators in jerusalem artichoke by using near-infrared spectroscopy [J]. Modern Food Science and Technology, 2026, 42(4): 313-326.

收稿日期: 2025-01-27; 修回日期: 2025-04-21; 接受日期: 2025-05-01

基金项目: 贵威生物营养与健康研究院项目 (5100-PH23004); 广东省功能食品活性物重点实验室 (2018B030322010); 广东省引进创新创业团队项目 (2019ZT08N291)

作者简介: 李心怡 (1999-), 女, 硕士研究生, 研究方向: 食品新原料与功能性食品, E-mail: lixinyi822@163.com

通讯作者: 段杉 (1966-), 男, 博士, 副教授, 研究方向: 食品生物技术, E-mail: duanshan@scau.edu.cn; 共同通讯作者: 曹庸 (1966-), 男, 博士, 教授, 研究方向: 食品化学, E-mail: caoyong2181@scau.edu.cn

data were collected in the range of 900~2 500 nm by using a near-infrared (NIR) spectrometer, and partial least squares regression prediction models were established in combination with the measured contents of inulin, protein, total polyphenols, and ash. The results showed that, after outlier elimination, spectral preprocessing, and characteristic band screening, prediction correlation coefficients (R^2_p) for the optimal quantitative models of inulin, protein, total polyphenols, and ash were 0.869 5, 0.930 7, 0.943 3, and 0.920 6, respectively. The corresponding root mean square errors of prediction were 2.237 2 g/100 g, 0.363 4 g/100 g, 0.107 7 mg·g⁻¹, and 0.158 1 g/100 g, respectively, indicating good predictive ability of the models. External validation showed no significant difference ($P>0.05$) between the detection results obtained using NIR spectroscopy and standard chemical methods. This study provides a theoretical basis for the application of NIR in the rapid prediction of physicochemical components in the Jerusalem artichoke.

Key words: Jerusalem artichoke; near-infrared diffuse reflectance spectrometry; content prediction; characteristic band screening; partial least squares regression

菊芋 (*Helianthus tuberosus* L.) 是一种多年生菊科向日葵属宿根性草本植物, 因生态适应性强, 种植成本低, 且块茎中营养价值较高, 被联合国粮农组织称为“21 世纪人畜共用作物”。菊芋块茎中含有丰富的蛋白质、维生素、矿物质和碳水化合物等活性成分, 其中碳水化合物主要是菊糖, 在新鲜菊芋中菊糖含量可达 15%~20%。菊糖是菊芋中重要的储备多糖, 其结构是由 D- 呋喃果糖通过 β -2,1 糖苷键连接而成的一种果聚糖混合物, 通常末端带有葡萄糖残基^[1]。这种多糖可以被结肠中的有益细菌选择性代谢, 充当益生元成分, 对肠道菌群的调节具有突出作用。此外, 菊糖在调节脂质代谢、降低血糖、抑制炎症因子表达、降低结肠癌风险、增强矿物质吸收、改善便秘和缓解抑郁等方面也表现出很大的潜力^[2]。因此, 菊芋在挖掘功能活性肽^[3]、开发功能性植物饮料^[4]、作为食品营养补充剂^[5,6]等方面具有广泛应用。这些理化成分的含量信息对于菊芋生产加工至关重要, 但传统的化学检测方法往往耗费大量时间和成本。例如通常使用高效液相色谱法进行菊糖含量的检测^[7,8], 尽管检测精度较高, 但样品的前处理过程复杂, 仪器设备要求高, 不能满足实际生产中对原料质量快速检测的需求。

近红外 (Near-Infrared, NIR) 光谱技术是一种使用方便、分析快速、不破坏样品的新型分析技术, NIR 区域的光谱信息主要来源于甲基 (-CH, Methyl)、氨基 (-NH, Amino)、羟基 (-OH, Hydroxyl)、巯基 (-SH, Thiol) 等含氢基团的倍频与合频吸收^[9], 绝大多数的化合物在 NIR 区域均有相应的吸收带, 通过这些吸收信息可以对样品进行定性或定量分析。因此 NIR 在食品的多种成分预测中具有很大的应用潜力, 还可以实现食品生产中的在线应用^[10,11]。目前, NIR 在薯芋类作物的理化指标检测中的应用逐渐成熟, He 等^[12]在 900~1 700 nm 范围构建还原糖的偏最小二乘模型, 具有较好的预测性能, 其中预测相关

系数 (Coefficient of Determination for Prediction, R^2_p) 为 0.950, 预测均方根误差 (Root Mean Square Error of Prediction, *RMSEP*) 为 0.253 g/100 g, 残差预测偏差 (Residual Predictive Deviation, *RPD*) 为 3.164, 实现甘薯中还原糖的快速测定; Wang 等^[13]使用 NIR 光谱和线性非线性算法对马铃薯淀粉和水分含量进行预测, 适用于快速预测马铃薯加工业中的原料质量; Li 等^[14]从中国 24 个产地收集 59 份菊芋, 通过建立偏最小二乘回归模型快速预测了菊芋的可溶性糖、纤维素、半纤维素、木质素、灰分等含量, 并结合生物质消化率, 实现快速筛选用于生物乙醇生产的理想原料, 但 NIR 在菊芋食品加工业中的应用尚未报道。

基于此, 本文采集全国不同产地、不同品种的具有代表性的 56 份菊芋样品, 通过将理化指标含量与 NIR 光谱值关联建立 PLSR 模型, 并采用不同光谱预处理方式和波段筛选方法对模型进行优化, 得到菊芋中菊糖、蛋白质、总多酚、灰分预测性能良好的定量模型, 以期能够实现食品加工中菊芋主要理化指标的快速预测。

1 材料与方法

1.1 材料与试剂

新鲜菊芋: 为提高预测模型的泛化能力, 综合考虑产地、品种、收获时间等影响菊芋成分的因素, 通过市场、网购平台两种渠道收集菊芋样品共 56 份。样品产地分布如图 1 所示, 具体信息见表 1。

没食子酸, 分析对照品 ($\geq 99\%$), 上海麦克林生化科技股份有限公司; 无水乙醇, 天津市富宇精细化工有限公司; 碳酸钠, 天津市大茂化学试剂厂; 福林酚 (生化试剂)、果糖、葡萄糖、蔗糖, 标准品 ($\text{HPLC} \geq 99\%$), 上海源叶有限公司; 乙腈, 色谱纯, 瑞典 Oceanpak 公司; 水为一级水, 未特别标注试剂为分析纯。

1.2 仪器与设备

LC-15C 高效液相色谱仪（配有示差检测器），上海岛津有限公司；S450 近红外光谱仪（配有 Prolab S450 近红外光谱采集软件），上海棱光技术有限公司；VersaMax 光栅型酶标仪，上海美谷分子仪器有限公司；SXL-1002 马弗炉，上海精宏实验设备有限公司；Rapid N exceed 杜马斯快速定氮分析仪，德国 Elementar 公司；AL104 万分之一电子分析天平，梅特勒－托利多仪器（上海）有限公司；DHG-9240A 电热鼓风干燥箱，上海一恒科学仪器有限公司。

1.3 实验方法

1.3.1 样品前处理

收集的新鲜菊芋需及时带回实验室处理，每个产地样品中挑选大小均匀、完整无机械损伤的菊芋约

500 g，洗净去皮，切片后置于 60 ℃干燥 24 h，粉碎后过 60 目筛^[15]，储存于干燥器内备用。

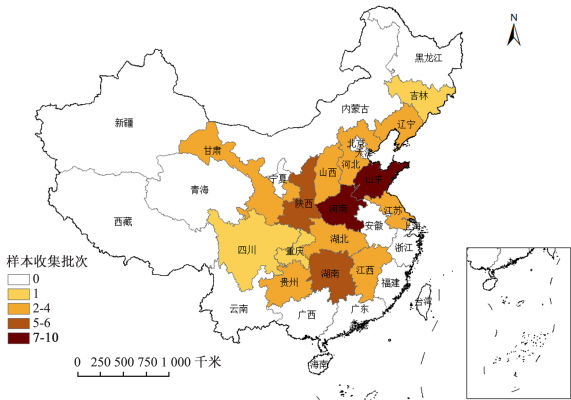


图 1 菊芋样品产地分布

Fig.1 Distribution of Jerusalem artichoke samples

注：地图通过 ArcGIS 10.8 桌面应用程序进行绘制。

表 1 菊芋样品具体信息

Table 1 Specific information of Jerusalem artichoke samples

编号	产地	品种	收集时间	编号	产地	品种	收集时间
J1	湖南省衡阳市	红皮	2023.11	J29	甘肃省定西市	白皮	2024.01
J2	陕西省安康市	白皮	2023.11	J30	山东省潍坊市	白皮	2024.01
J3	湖南省株洲市	红皮	2023.11	J31	湖北省宜昌市	白皮	2024.01
J4	湖南省永州市江华瑶族自治县	白皮	2023.11	J32	湖南省益阳市	白皮	2024.01
J5	湖南省常德市	红皮	2023.11	J33	重庆市梁平区	白皮	2024.01
J6	河南省商丘市	白皮	2023.12	J34	山东省青岛市	白皮	2024.01
J7	陕西省榆林市	红皮	2023.12	J35	四川省绵阳市	红皮	2024.01
J8	山东省济宁市	红皮	2023.12	J36	山东省烟台市	白皮	2024.01
J9	山东省菏泽市	白皮	2023.12	J37	河南省郑州市	白皮	2024.01
J10	江苏省徐州市	白皮	2023.12	J38	陕西省榆林市	白皮	2024.01
J11	江西省宜春市	白皮	2023.12	J39	河北省保定市	白皮	2024.01
J12	陕西省宝鸡市	白皮	2023.12	J40	山东省临沂市	白皮	2024.02
J13	河南省三门峡市	白皮	2023.12	J41	山东省滨州市	白皮	2024.02
J14	河南省开封市	白皮	2023.12	J42	辽宁省铁岭市	红皮	2024.02
J15	河南省开封市	红皮	2023.12	J43	江苏省徐州市	白皮	2024.02
J16	辽宁省锦州市	红皮	2023.12	J44	河南省商丘市	白皮	2024.02
J17	吉林省白山市	红皮	2023.12	J45	江西省赣州市	白皮	2024.02
J18	山西省运城市	红皮	2023.12	J46	河北省石家庄市	白皮	2024.02
J19	河南省三门峡市	红皮	2023.12	J47	贵州省六盘水市	白皮	2024.02
J20	河南省周口市	白皮	2023.12	J48	山西省运城市	红皮	2024.02
J21	河北省石家庄市	白皮	2023.12	J49	山东省潍坊市	红皮	2024.02
J22	陕西省汉中市	白皮	2024.01	J50	河南省开封市	白皮	2024.02
J23	贵州六盘水市	白皮	2024.01	J51	江西省宜春市	白皮	2024.02
J24	山东省济南市	白皮	2024.01	J52	山东省济南市	白皮	2024.02
J25	江苏省徐州市	红皮	2024.01	J53	河南省三门峡市	白皮	2024.02
J26	湖北省天门市	白皮	2024.01	J54	甘肃省定西市	白皮	2024.02
J27	湖南省岳阳市	白皮	2024.01	J55	辽宁省沈阳市	红皮	2024.02
J28	陕西省咸阳市	白皮	2024.01	J56	辽宁省大连市	红皮	2024.02

1.3.2 近红外光谱数据的采集

使用积分球漫反射模式采集菊芋的光谱信息,波长范围为 900~2 500 nm,扫描次数为 32 次,分辨率为 12 cm^{-1} ,采样间隔为 4 nm。将近红外光谱仪预热 30 min,待其稳定后进行光谱采集,避免因系统不稳定对实验结果产生影响。以仪器内置背景为参比,每次扫描前均自动进行参比校正。称取 15 g 的样品倒入样品杯内使底部铺满,每次测定时样品厚度基本保持一致,并在上方加盖反射部件防止样本透光。每个样品平行采集 3 次,取平均光谱作为该样品的原始光谱。

1.3.3 理化指标的测定

1.3.3.1 蛋白质的测定

参考 GB 5009.5-2016《食品安全国家标准 食品中蛋白质的测定》中的燃烧法^[16],准确称取 100 mg 样品,采用杜马斯快速定氮分析仪进行测定,以 6.25 为氮的换算系数。

1.3.3.2 总多酚的测定

参考 Kaszá 等^[17]的方法进行总多酚的提取,称取 1 g 样品,加入 1:20 ($\text{g}\cdot\text{mL}^{-1}$) 的 70% (φ) 乙醇,80 °C 超声辅助提取 30 min,抽滤,滤渣再次加入等量的 70% (φ) 乙醇重复提取,合并两次滤液,并用 70% 乙醇定容至 50 mL,备用。采用 Folin-酚法测定总多酚含量^[18],以不同浓度的没食子酸标准溶液绘制标准曲线,得到回归方程为 $y=3.6874x-0.1032$ ($R^2=0.9905$)。

1.3.3.3 菊糖的测定

采用热水浸提法进行菊糖的提取^[19],称取 1 g 样品,加入 1:20 ($\text{g}\cdot\text{mL}^{-1}$) 的水,90 °C 水浴提取 1 h,抽滤,滤渣再次加入等量的水重复提取,合并两次滤液,用水定容至 50 mL。参考 GB/T 41377-2022《菊粉质量要求》高效液相色谱法中的色谱条件^[20],采用 JADE-PAK 氨基色谱柱 (250 mm×4.6 mm, 5 μm),流动相为 75% (φ) 乙腈水溶液,进行等浓度洗脱,流速为 $1.0\text{ mL}\cdot\text{min}^{-1}$,进样体积为 20 μL ,柱温为 30 °C。分别称取葡萄糖、果糖、蔗糖标准品适量,加水溶解制成 $10\text{ mg}\cdot\text{mL}^{-1}$ 的混合标准储备溶液,取适量混合标准储备液加水稀释成 $5\text{ mg}\cdot\text{mL}^{-1}$ 的混合标准溶液。按照同样的检测条件进行测定。参考 GB/T 41377-2022《菊粉质量要求》中的计算方法,计算出样品中的菊糖含量,具体计算方法如下。

提取液中各糖(果糖、葡萄糖、蔗糖)的含量按式(1)计算:

$$\omega_i = \frac{A_i \times \rho_s}{A_s \times \rho_i} \times f \times 100\% \quad (1)$$

式中:

ω_i —提取液中某种糖的含量, %;

A_i —提取液中某种糖的峰面积;

ρ_s —标准液中某种糖的质量浓度, $\text{mg}\cdot\text{mL}^{-1}$;

A_s —标准液中某种糖的峰面积;

ρ_i —提取液的质量浓度, $\text{mg}\cdot\text{mL}^{-1}$;

f —提取液的稀释换算系数。

提取液中菊糖含量按式(2)计算:

$$\omega_0 = 100 - (\omega_1 + \omega_2 + \omega_3) \quad (2)$$

式中:

ω_0 —提取液中菊糖的质量分数, %;

ω_1 —提取液中果糖的质量分数, %;

ω_2 —提取液中葡萄糖的质量分数, %;

ω_3 —提取液中蔗糖的质量分数, %。

样品中菊糖含量按式(3)计算:

$$\omega = \frac{\omega_0 \times \rho_i \times V}{m \times 1000} \times 100\% \quad (3)$$

式中:

ω —样品中菊糖的含量, %;

ω_0 —提取液中菊糖的含量, %;

ρ_i —提取液的质量浓度, $\text{mg}\cdot\text{mL}^{-1}$;

V —提取液的总体积, mL;

m —样品的质量, g。

1.3.3.4 灰分的测定

参考 GB 5009.4-2016《食品安全国家标准 食品中灰分的测定》^[21],采用马弗炉灼烧法进行测定。

1.3.4 近红外定量模型的建立

1.3.4.1 异常数据的剔除

采用杠杆值与学生化残差法剔除存在异常数据的样本,相比于马氏距离法,它能够同时监测出化学值的异常情况。设定光谱杠杆值的阈值为各样本杠杆值均值的 2 倍,超过该阈值说明样本光谱值异常;学生残差的阈值设定为各样本学生残差绝对值均值的 3 倍,超过该阈值说明样本化学值异常^[22]。

1.3.4.2 样本集划分

采用基于 x-y 距离的样本集划分 (Sample Set Partitioning Based On Joint X-Y Distance, SPXY) 算法按照 3:1 比例将样本集划分为训练集和预测集^[23],SPXY 算法在肯纳德-斯通 (Kennard-Stone, KS) 算法的基础上发展而来,计算样品间距离时将光谱值和化学值变量同时考虑在内,使样本集划分更为科学。

1.3.4.3 光谱预处理

为了消除环境、仪器噪声等因素对光谱数据产生的干扰,需采用适当的方法对原始光谱进行预处

理。SG 卷积平滑 (Savitzky-Golay, SG) 对光谱进行平滑处理, 能够有效去除噪声的影响; 多元散射校正 (Multiplicative Scatter Correction, MSC)、标准正态变量变换 (Standard Normal Variate Transformation, SNV) 都有助于消除固体样品对光谱的散射影响; 一阶导数 (First Derivative, FD)、二阶导数 (Second Derivative, SD) 可以有效消除基线漂移和背景干扰带来的影响^[24]。本研究采用 SG 平滑、MSC、SNV、FD、SD 及它们的组合等多种方法对原始光谱进行预处理。

1.3.4.4 特征波段的筛选

特征波段一般是指光谱中对特定化学成分表现出明显吸收的特定波长, 这些特定波长对样品中存在物质的鉴定和定量分析具有重要价值^[25]。为了减少光谱数据中大量的冗余信息, 提高模型的预测性能与计算效率, 本研究采用竞争性自适应重加权 (Competitive Adaptive Reweighted Sampling, CARS)、连续投影算法 (Successive Projections Algorithm, SPA)、随机蛙跳 (Random Frog, RF) 三种常用的算法进行特征波段的筛选。CARS 使用蒙特卡洛采样与偏最小二乘回归模型的回归系数结合, 模拟达尔文“适者生存”原则, 筛选权重较低的波长, 识别出交叉验证均方根误差最低的波长; SPA 是一种前向迭代的选择方法, 即从一个波长开始, 在每次迭代中加入一个新变量, 直至所选变量数达到设定值, 实现选择冗余信息最少的波长以解决共线性问题; RF 模拟青蛙跳跃的随机性和多样性, 以 N 次迭代中变量出现概率为依据, 将具有最高概率的变量选为特征变量^[26-28]。

1.3.4.5 最优预测模型的建立

根据朗伯-比尔定律, 样品的一些物理化学性质与吸收峰之间存在线性关系, 因此可以用来预测理化成分的含量。偏最小二乘回归 (Partial Least Squares Regression, PLSR) 模型是一种常用的线性模型, 能够有效克服因变量和自变量之间存在的多重共线性问题, 在样本数量较少的情况下, 依然可以获得较为理想的预测效果。本研究利用光谱数据与 4 个理化指标的化学值建立 PLSR 模型, 通过训练集的相关系数 (Coefficient of Determination for Calibration, R^2_c)、训练集均方根误差 (Root Mean Square Error of Calibration, $RMSEC$)、预测集的相关系数 (Coefficient of Determination for Prediction, R^2_p)、预测均方根误差 (Root Mean Square Error of Prediction, $RMSEP$) 和残差预测偏差 (Residual Predictive Deviation, RPD) 等评价参数筛选出最优的预测模型。其中, 相关系数表示训练集或预测集中预测值和实测值之间线性关系的强

度; 均方根误差代表预测值和实测值之间的差异, 值越小表示预测准确性越好; RPD 是预测集中实测值的标准差与 $RMSEP$ 的比值, 可以评估预测模型的泛化能力, RPD 值大于 2.5 表明建立了一个优秀且可靠的模型^[29,30]。

1.3.4.6 外部验证

为了进一步验证定量模型的泛化能力, 本研究于 2024 年 3 月继续收集 5 份不参与建模的菊芋样品, 进行光谱采集与化学值测定, 通过预测值与实测值的差异来评价模型的可靠性和稳定性^[31]。

1.4 数据分析

每份样品的各理化指标均重复平行 3 次, 取平均值作为定量模型的化学值。光谱数据通过 OMNIC 软件转换成 csv 格式, 再使用 MATLAB R2021a 进行近红外模型的建立及相关图形的绘制, 使用 Excel 2010 进行表格的绘制。

2 结果与讨论

2.1 菊芋近红外图谱分析

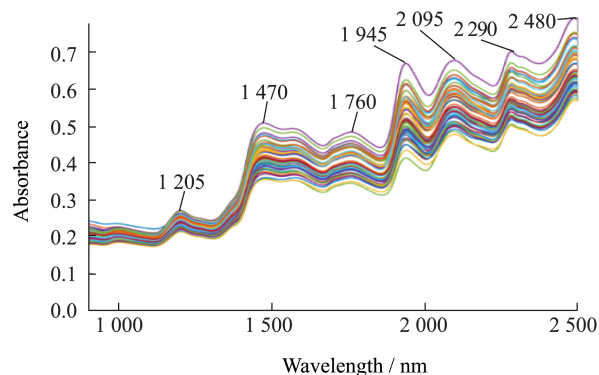


图 2 56 份菊芋的 NIR 原始图谱

Fig.2 56 sets of NIR raw spectra of Jerusalem artichoke

56 份菊芋样品的原始光谱如图 2 所示, 各样品图谱的走势基本一致, 但吸光值的大小有明显区别, 表明所含的化学成分种类相似, 但含量有所差异。所有样品在 1 200、1 500、1 950、2 100、2 300、2 500 nm 等附近呈现明显的吸收带, 包含了大量的基团吸收信息。例如, 1 195 nm 附近的吸收与 C-H 键拉伸振动的第一泛音有关, 1 385 nm 附近的吸收与 C-H 键的第二泛音以及 O-H 键有关, 这些与多糖的结构密切相关^[32]; 1 200 nm 附近的吸收对应于 C-H 键拉伸振动的第二泛音, 1 480 nm 附近的吸收与 O-H 键拉伸振动的基频有关, 1 420~1 520 nm 附近的吸收与 N-H 键有关, 这些与蛋白质结构中存在的 -NH、-NH₂ 和 -COOH 等

基团相关^[33]；酚类物质是含有一个或多个芳香环，且环上带有一个或多个羟基的化合物，其特征主要体现在 1 415~1 500 nm 和 1 955~2 035 nm 附近的区域^[34]。这些特征信息为近红外光谱与理化指标之间的关联提供了科学依据。

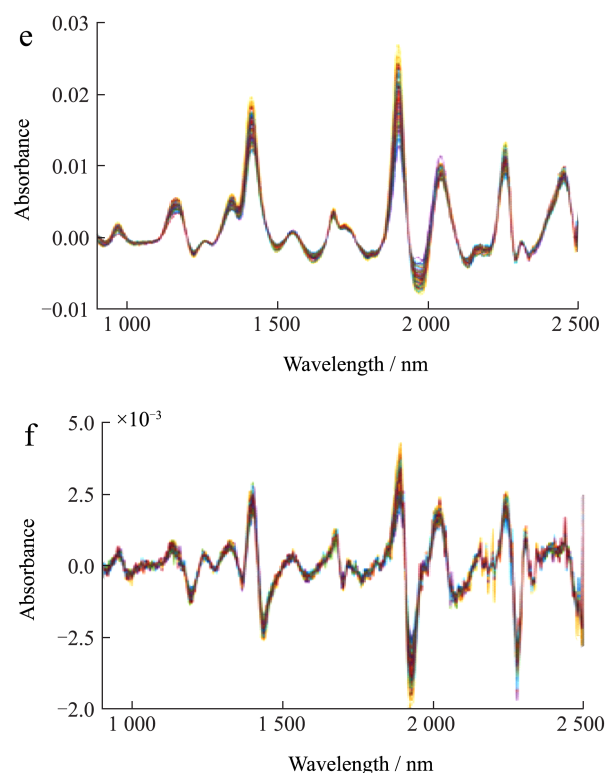
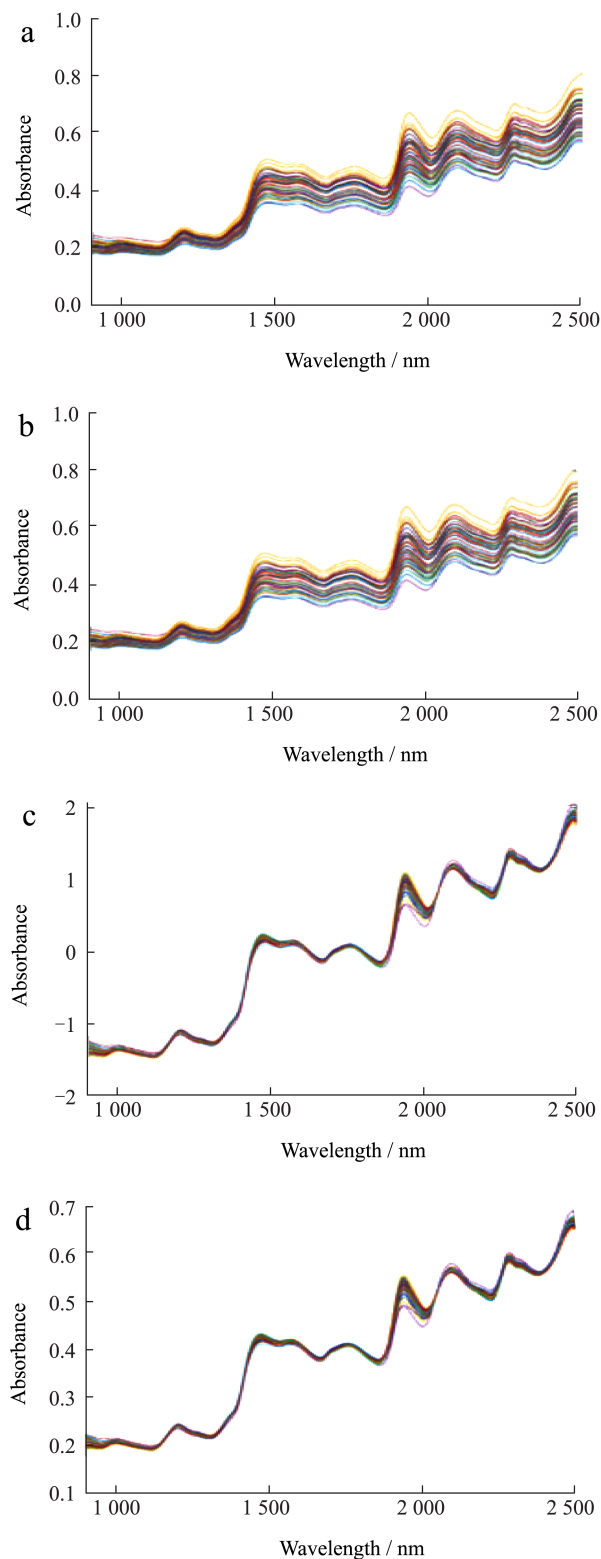


图 3 经不同预处理方法的近红外图谱

Fig.3 NIR Spectra after different spectral preprocessing methods

注：(a) 原始图谱；(b) SG 平滑；(c) SNV；(d) MSC；(e) FD；(f) SD。

此外，原始光谱中也表现出明显的散射效应和基线漂移。为减少这些干扰因素的影响，采用 SG 平滑、MSC、SNV、FD、SD 及它们的组合等多种方法对原始光谱进行预处理。SG 处理消除了光谱中的噪声，使光谱更加平滑，见图 3b。但同时也会去除光谱中的一些细节，削弱光谱的整体特征^[35]；SNV、MSC 处理消除了固体样品由于颗粒大小和分布不均匀而导致的散射影响，使图谱中 1 200、1 700 nm 附近的吸收峰更加明显，分别见图 3c、3d。由于两种处理方法起到相似作用，因此 SNV 和 MSC 处理后的图谱在趋势上较为一致；导数处理消除基线漂移并放大了原本不太突出的波峰和波谷信息，使处理后的图谱中出现一些新的原本隐藏在原始光谱中的吸收峰，FD、SD 处理后的光谱分别见图 3e、3f。同时导数处理后改变了光谱的形状，可能会使光谱分析变得更为复杂^[36]。

2.2 异常样本的剔除

在光谱采集和化学检测过程中，往往会因为仪器、环境、操作人员等因素引入异常数据。为了避免异常样本对模型的准确性造成影响，采用杠杆值与学生化

残差法剔除异常样本,结果如图4所示。在菊糖样本集中,18号样本的残差值超过了阈值上限,13、24号样本的残差值低于阈值下限,说明其化学值异常,予以剔除;在蛋白质样本集中,36号样本的残差值超过了阈值上限,说明其化学值异常,予以剔除;在总多酚样本集中,15、24、39号样本的残差值超过了阈值上限,说明其化学值异常,予以剔除;在灰分样本集中,43号样本的残差值低于阈值下限,说明其化学值异常,予以剔除。化学值出现异常的原因有很多,可能是由于实验操作不当、仪器误差、样本自身特殊性等多种因素导致,对异常样本的剔除有利于提高模型的稳定性。

2.3 样本集划分

在近红外定量模型的建立中,需将数据集划分为训练集和预测集,其中训练集用于模型的开发,预测集则用于模型的验证,对模型实际预测能力进行评估。本研究采用SPXY算法按照3:1的比例,将剔除异常样本后的各指标样本集划分为训练集与预测集,划分结果如表2所示。结果显示,训练集中各理化指标含量的范围较广,能够代表菊芋主要理化指标的一般含量范围,这有助于构建稳定的校准模型。预测集中4个理化指标的含量均在训练集的范围之内,说明样本划分是合理的^[37]。

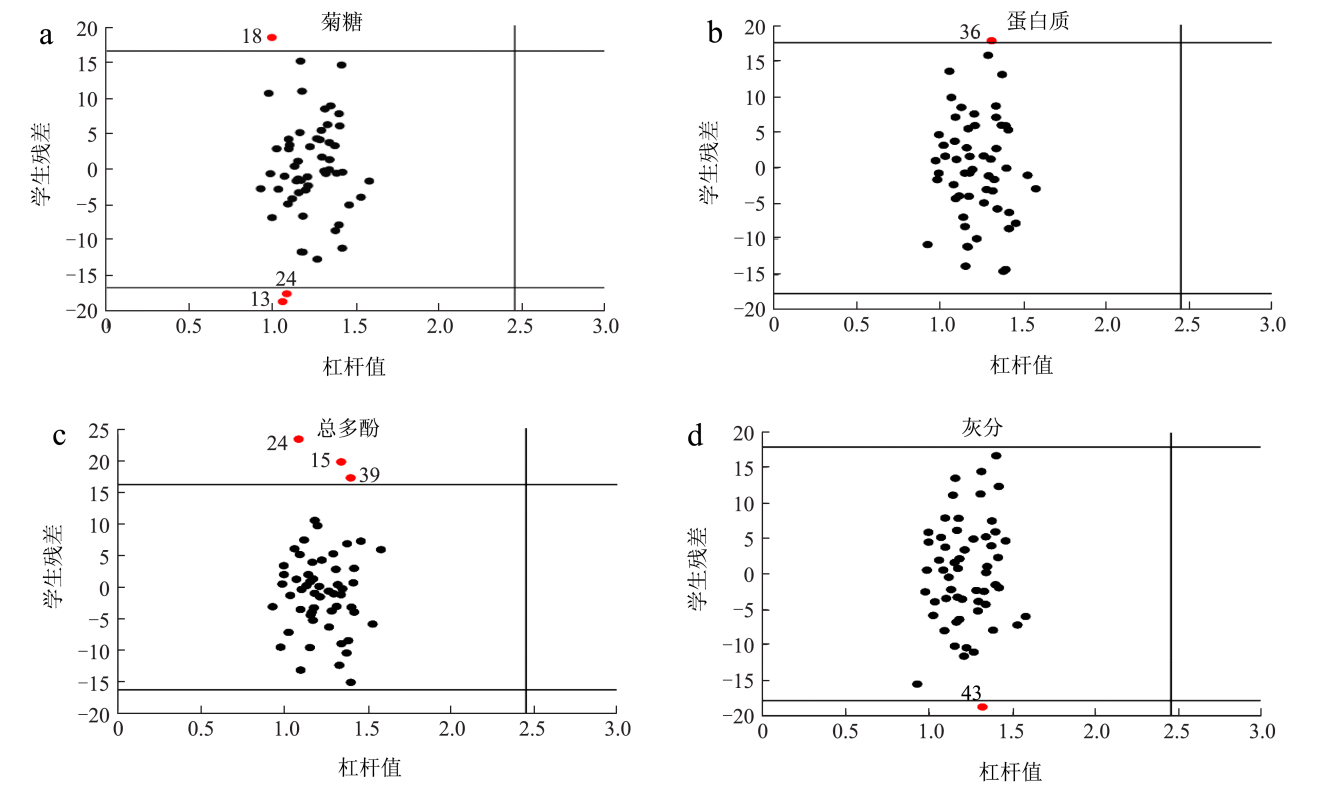


图4 剔除异常样本结果

Fig.4 Results of eliminating outlier samples

表2 样本集划分

Table 2 Division of sample sets							
理化指标	组别	样本数	最小值	最大值	平均值	标准差	变异系数/%
菊糖/(g/100 g)	训练集	40	49.05	82.60	65.45	9.04	13.81
	预测集	13	52.96	73.46	64.02	5.62	8.78
蛋白质/(g/100 g)	训练集	41	5.66	12.99	9.46	1.84	19.45
	预测集	14	6.41	11.42	8.92	1.55	17.38
总多酚/(mg/g)	训练集	40	2.02	4.12	2.67	0.49	18.23
	预测集	13	2.00	3.68	2.60	0.48	18.46
灰分/(g/100 g)	训练集	41	3.19	8.50	5.97	1.06	17.77
	预测集	14	4.71	7.00	6.09	0.62	10.26

2.4 基于不同光谱预处理方式的PLSR建模结果

由于大量基团的吸收信息存在重叠，因此还需要结合化学计量学方法如主成分分析（Principal Component Analysis, PCA）提取主成分来建立定量模型。交叉验证是一种常用的验证方法，其中一些样本被排除在校准之外并用于预测，重复此操作直到所有

样品都被排除在外一次，能够避免校准模型出现过拟合现象。为了筛选出最优的光谱预处理方式，分别使用不同的光谱预处理方法并结合 PCA 特征提取和交叉验证来建立 PLSR 模型。本研究设置最大主成分数为 20，通过 10 折交叉验证确定最佳的主成分数（PCs）建立 PLSR 模型。不同光谱预处理方式建立各理化指标 PLSR 模型的参数结果见表 3。

表 3 基于不同光谱预处理方式的PLSR建模结果
Table 3 The PLSR modeling results based on different spectral preprocessing methods

理化指标	预处理方式	PCs	R^2_c	RMSEC	R^2_p	RMSEP	RPD
菊糖	None	19	0.999 9	0.104 5	0.634 3	3.980 6	1.41
	SG	20	0.999 9	0.098 9	0.665 4	3.824 4	1.47
	MSC	10	0.895 5	2.745 9	0.395 5	5.933 0	0.95
	SNV	12	0.909 0	2.578 5	0.402 5	5.363 5	1.05
	FD	13	0.994 6	0.651 4	0.733 8	2.703 1	2.08
	SD	14	0.999 7	0.164 0	0.850 6	2.406 5	2.34
	SG+MSC	11	0.980 8	1.235 2	0.546 7	4.571 4	1.23
	SG+SNV	13	0.993 8	0.729 8	0.552 9	4.176 4	1.35
	SG+FD	12	0.990 7	0.855 9	0.766 7	2.903 2	1.94
	SG+SD	12	0.997 2	0.468 6	0.825 5	2.557 2	2.20
	MSC+FD	18	0.966 0	1.854 5	0.524 1	4.959 2	1.13
	MSC+SD	12	0.985 3	1.356 5	0.645 6	3.622 3	1.55
	SNV+FD	11	0.987 8	1.243 4	0.700 9	3.047 5	1.84
	SNV+SD	13	0.999 7	0.161 9	0.777 2	2.782 5	2.02
蛋白质	None	10	0.981 2	0.247 4	0.865 2	0.495 2	3.13
	SG	10	0.981 0	0.248 6	0.865 0	0.495 7	3.13
	MSC	12	0.984 5	0.225 1	0.929 2	0.368 1	4.21
	SNV	12	0.985 4	0.218 0	0.925 8	0.376 4	4.12
	FD	9	0.993 9	0.141 1	0.894 8	0.452 4	3.43
	SD	5	0.982 4	0.239 6	0.463 2	0.798 3	1.94
	SG+MSC	12	0.982 6	0.238 2	0.930 7	0.363 4	4.26
	SG+SNV	12	0.983 5	0.231 8	0.927 3	0.371 8	4.17
	SG+FD	9	0.989 6	0.185 0	0.880 4	0.469 7	3.30
	SG+SD	9	0.998 2	0.077 9	0.885 7	0.495 6	3.13
	MSC+FD	10	0.997 3	0.093 9	0.914 6	0.416 6	3.72
	MSC+SD	14	1.000 0	0.000 1	0.664 0	0.779 3	1.99
	SNV+FD	10	0.997 5	0.091 2	0.927 5	0.386 3	4.01
	SNV+SD	14	1.000 0	0.001 0	0.702 2	0.739 9	2.09

续表3

理化指标	预处理方式	PCs	R^2_c	RMSEC	R^2_p	RMSEP	RPD
总多酚	None	12	0.921 4	0.130 0	0.670 0	0.285 6	1.68
	SG	12	0.917 9	0.132 7	0.672 2	0.283 5	1.69
	MSC	6	0.780 1	0.204 5	0.761 3	0.232 3	2.07
	SNV	6	0.780 0	0.204 5	0.769 2	0.240 6	1.99
	FD	12	0.998 2	0.020 4	0.757 0	0.261 4	1.84
	SD	5	0.964 0	0.089 8	0.173 9	0.300 1	1.60
	SG+MSC	6	0.779 4	0.204 7	0.761 1	0.232 3	2.07
	SG+SNV	6	0.779 3	0.204 8	0.769 0	0.240 6	1.99
	SG+FD	5	0.776 9	0.205 7	0.739 8	0.240 0	2.00
	SG+SD	12	0.998 3	0.019 6	0.565 5	0.277 5	1.73
	MSC+FD	4	0.776 8	0.205 7	0.776 6	0.225 4	2.13
	MSC+SD	4	0.929 7	0.123 4	0.179 2	0.322 0	1.49
	SNV+FD	4	0.776 8	0.205 7	0.781 5	0.235 0	2.04
	SNV+SD	4	0.928 3	0.124 5	0.376 8	0.295 0	1.63
灰分	None	13	0.769 7	0.454 1	0.522 9	0.776 0	0.81
	SG	10	0.885 4	0.336 6	0.468 6	0.898 0	0.70
	MSC	12	0.846 0	0.383 4	0.347 4	0.956 9	0.65
	SNV	8	0.954 6	0.218 8	0.388 7	0.921 7	0.68
	FD	12	0.970 0	0.163 4	0.725 7	0.328 7	1.90
	SD	10	0.998 7	0.038 0	0.876 7	0.197 7	3.16
	SG+MSC	15	0.828 4	0.401 7	0.454 9	0.876 7	0.71
	SG+SNV	11	0.846 8	0.382 6	0.533 2	0.741 0	0.84
	SG+FD	10	0.994 3	0.078 9	0.684 2	0.376 1	1.66
	SG+SD	10	0.9771	0.157 0	0.779 2	0.273 4	2.29
	MSC+FD	15	0.999 3	0.027 3	0.706 5	0.356 4	1.75
	MSC+SD	10	0.998 8	0.036 7	0.773 4	0.284 7	2.19
	SNV+FD	9	0.957 3	0.152 1	0.589 3	0.687 9	0.91
	SNV+SD	9	0.986 4	0.121 5	0.680 1	0.380 8	1.64

可以发现，不同预处理方式对模型性能的影响不同，适当的预处理可以提高模型的准确性，而不恰当的预处理可能会导致预测效果更差。在菊糖指标中，SD 预处理方式效果最佳，经 SD 预处理后 R^2_p 从 0.634 3 上升至 0.850 6， $RMSEP$ 从 3.980 6 g/100 g 下降至 2.406 5 g/100 g， RPD 从 1.41 上升至 2.34，模型预测性能整体上得到显著提升，而 MSC、SNV 预处理效果最差；在蛋白质指标中，SG+MSC 的预处理方式效果最佳，经 SG+MSC 预处理后 R^2_p 从 0.865 2 上升至 0.930 7， $RMSEP$ 从 0.495 2 g/100 g 下降至 0.363 4 g/100 g， RPD 从 3.13 上升至 4.26，而 SD 处理效果最差；在总多酚指标中，SNV+FD 预处理方式效果

最佳，经 SNV+FD 预处理后 R^2_p 从 0.670 0 上升至 0.781 5， $RMSEP$ 从 0.285 6 mg·g⁻¹ 下降至 0.235 0 mg·g⁻¹， RPD 从 1.68 上升至 2.04，而 SD 处理效果最差；在灰分指标中，SD 预处理方式效果最佳，经 SD 预处理后 R^2_p 从 0.522 9 上升至 0.876 7， $RMSEP$ 从 0.776 0 g/100 g 下降至 0.197 7 g/100g， RPD 从 0.81 上升至 3.16，而 MSC、SNV 预处理效果最差。

因此，分别选用 SD、SG+MSC、SNV+FD、SD 作为菊糖、蛋白质、总多酚、灰分定量模型的最佳预处理方式。尽管原始光谱经过预处理后模型的性能有显著提升，但仍有一些模型的参数不在理想范围。

2.5 基于不同特征波段筛选方式的PLSR建模结果

为了进一步对模型进行优化,在各指标的最佳预处理方式下,分别使用CARS、RF、SPA三种波段筛选方法提取特征波段并建立PLSR模型,以筛选最优的波段筛选方式。不同特征波段筛选方式建立各理化指标PLSR模型的参数结果见表4。

在菊糖指标中,CARS波段筛选方式最佳,经CARS提取特征波段后 R^2_p 从0.850 6上升至0.869 5, $RMSEP$ 从2.406 5 g/100 g下降至2.237 2 g/100 g, RPD 从2.34上升至2.51,参与建模变量数减少了87.2%,提高了模型的预测性能和分析效率,而SPA波段筛选效果最差;在蛋白质指标中,经CARS、RF、SPA提取特征波段后,模型的训练集及预测集的相关系数(R^2)均下降,而均方根误差($RMSE$)均上升,表明模型性能比全波段所建模型有所下降,这可能与波段筛选过程中一些有用的信息也被筛除有关^[38];在总多酚指标中,CARS波段筛选方式最佳,经CARS提取特征波段后 R^2_p 从0.781 5上升至0.943 3, $RMSEP$ 从0.235 0 mg·g⁻¹下降至0.107 7 mg·g⁻¹, RPD 从2.04上

升至4.46,参与建模变量数减少了92.8%,而SPA波段筛选效果最差;在灰分指标中,CARS波段筛选方式最佳,经CARS提取特征波段后 R^2_p 从0.876 7上升至0.920 6, $RMSEP$ 从0.197 7 g/100 g下降至0.158 1 g/100 g, RPD 从3.16上升至3.95,参与建模变量数减少了84.4%,而SPA波段筛选效果最差。因此选取CARS作为菊糖、总多酚、灰分指标的最佳波段筛选方式,选取经SG+MSC处理的全波段用于蛋白质定量模型的建立,各指标参与建模的特征波段分布如图5所示。

特征变量选择有助于缩小检测的波段范围,就容易开发成本较低的专用近红外器件实现快速检测目的。但由于CARS、RF、SPA算法得到的特征波段以若干点的形式呈现,可能导致某些重要片段峰的信息丢失,降低模型精度。因此,后续应尝试基于区间的波段筛选算法,如区间偏最小二乘回归(Interval Partial Least Squares Regression, iPLSR)提取出特征带,探究能否进一步提升模型预测性能。目前还出现许多波段筛选的优化算法,如衰减消去蜻蜓算法^[39]、遗传算法^[40]等,还应尝试更多的算法,在保证模型性能的前提下,尽可能减少参与建模的波段数。

表 4 基于不同波段筛选方式的PLSR建模结果
Table 4 The PLSR modeling results based on different wavelength selection methods

理化指标	特征波段筛选方式	变量数	PCs	R^2_c	RMSEC	R^2_p	RMSEP	RPD
菊糖	Full spectrum	321	14	0.999 7	0.164 0	0.850 6	2.406 5	2.34
	CARS	41	13	0.999 2	0.254 7	0.869 5	2.237 2	2.51
	RF	20	20	0.831 6	3.389 0	0.431 3	5.208 4	1.08
	SPA	60	13	0.505 4	4.228 9	0.384 5	5.996 7	0.94
蛋白质	Full spectrum	321	12	0.982 6	0.238 2	0.930 7	0.363 4	4.26
	CARS	25	9	0.978 0	0.267 2	0.127 6	1.256 5	1.23
	RF	40	8	0.971 9	0.300 8	0.837 3	0.565 1	2.74
	SPA	27	8	0.926 7	0.475 4	0.905 2	0.439 5	3.53
总多酚	Full spectrum	321	4	0.776 8	0.205 7	0.781 5	0.235 0	2.04
	CARS	23	9	0.985 1	0.058 4	0.943 3	0.107 7	4.46
	RF	20	10	0.961 8	0.092 4	0.814 6	0.215 6	2.23
	SPA	48	6	0.790 7	0.200 3	0.699 8	0.288 0	1.67
灰分	Full spectrum	321	10	0.998 7	0.038 0	0.876 7	0.197 7	3.16
	CARS	50	20	1.000 0	0.003 8	0.920 6	0.158 1	3.95
	RF	40	20	0.989 9	0.105 1	0.643 2	0.587 9	1.06
	SPA	50	4	0.000 0	0.889 3	0.331 1	0.452 2	1.38

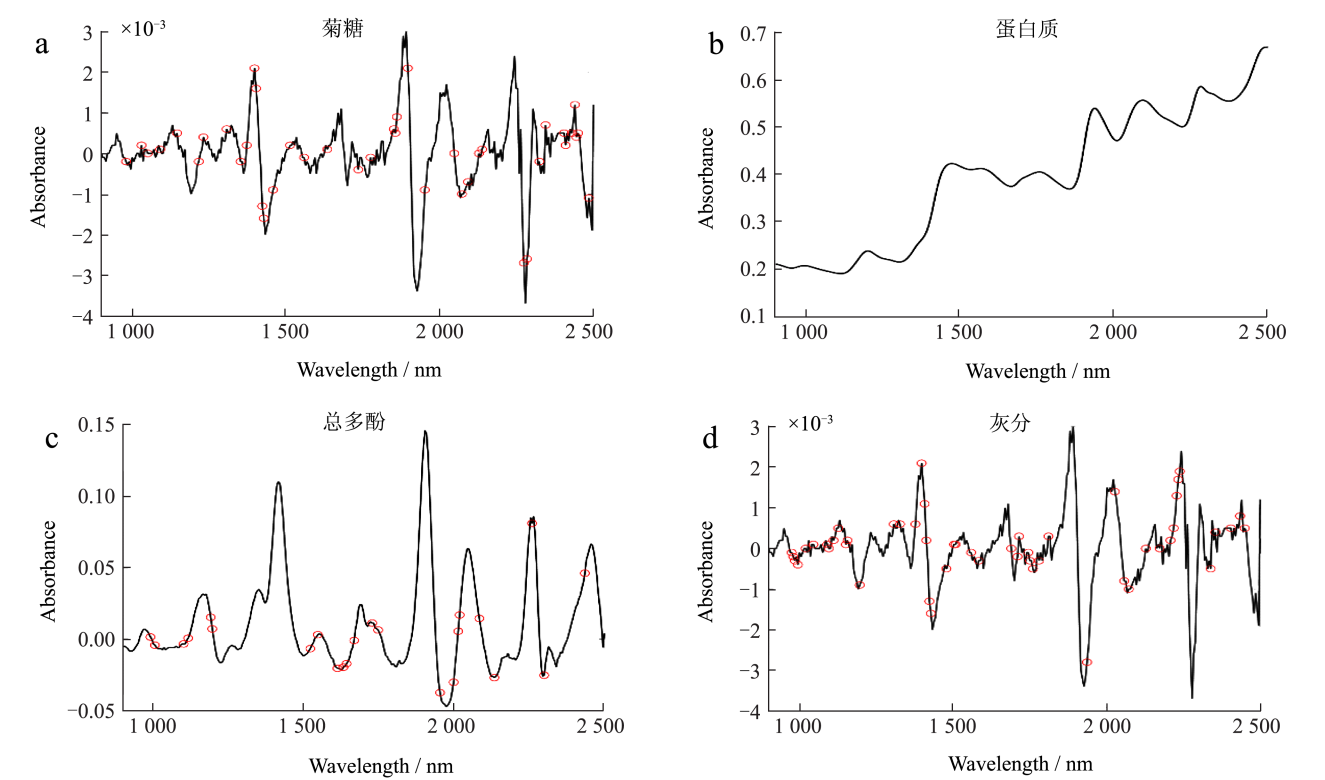


图 5 各指标参与建模波段分布图

Fig.5 Distribution of spectral band involved in modeling for each indicator

注：蛋白质指标为 900~2 500 nm 全波段建模。

表 5 各理化指标最优定量模型的建模结果

Table 5 The modeling results of the optimal quantitative models for various physicochemical indicators									
理化指标	预处理方式	波段筛选方式	变量数	PCs	R^2_c	RMSEC	R^2_p	RMSEP	RPD
菊糖	SD	CARS	41	13	0.999 2	0.254 7	0.869 5	2.237 2	2.51
蛋白质	SG+MSC	—	321	12	0.982 6	0.238 2	0.930 7	0.363 4	4.26
总多酚	SNV+FD	CARS	23	9	0.985 1	0.058 4	0.943 3	0.107 7	4.46
灰分	SD	CARS	50	20	1.000 0	0.003 8	0.920 6	0.158 1	3.95

2.6 最优定量模型的建立

经过光谱预处理、特征波段筛选分别对模型进行优化后，模型的预测性能逐渐得到提升，最终筛选得到各理化指标最优的建模方式及评价参数如表 5 所示。各理化指标定量模型的性能参数分别为：菊糖 $R^2_p=0.869\ 5$ 、 $RMSEP=2.237\ 2\ \text{g}/100\ \text{g}$ 、 $RPD=2.51$ ；蛋白质 $R^2_p=0.930\ 7$ 、 $RMSEP=0.363\ 4\ \text{g}/100\ \text{g}$ 、 $RPD=4.26$ ；总多酚 $R^2_p=0.943\ 3$ 、 $RMSEP=0.107\ 7\ \text{mg}\cdot\text{g}^{-1}$ 、 $RPD=4.46$ ；灰分 $R^2_p=0.920\ 6$ 、 $RMSEP=0.158\ 1\ \text{g}/100\ \text{g}$ 、 $RPD=3.95$ 。在模型的性能评价参数中，当 R^2 的值在 0.82~0.90 范围内时，表示模型良好且可接受； R^2 的值大于 0.91 意味着该模型非常出色^[41]。结果表明，4 个理化成分的

定量模型中 R^2 均大于 0.86， $RMSEP$ 均小于预测集实测化学值的标准差，且 RPD 均大于 2.5，因此所建的定量模型性能均达到良好或以上水平。

在已有的研究报道中，Meenu 等^[42]在 700~1 300 nm 建立了去壳燕麦中 β - 葡聚糖最佳 PLSR 预测模型 ($R^2_c=0.789$ ， $R^2_p=0.735$)；Fan 等^[43]在 940~1 660 nm 建立了玉米粉中蛋白质最佳 PLSR 预测模型 ($R^2_p=0.93$ ， $RPD=3$)；Meenu 等^[44]在 700~2 500 nm 分别建立了去壳燕麦和裸燕麦中总酚含量的 PLSR 预测模型， R^2 分别为 0.65 和 0.67，可见本研究建立的模型预测性能较为理想。各理化指标定量模型中训练集和预测集的散点分布图如图 6 所示，各样本点紧密分布在分界线两侧，表明模型的线性拟合关系较好。

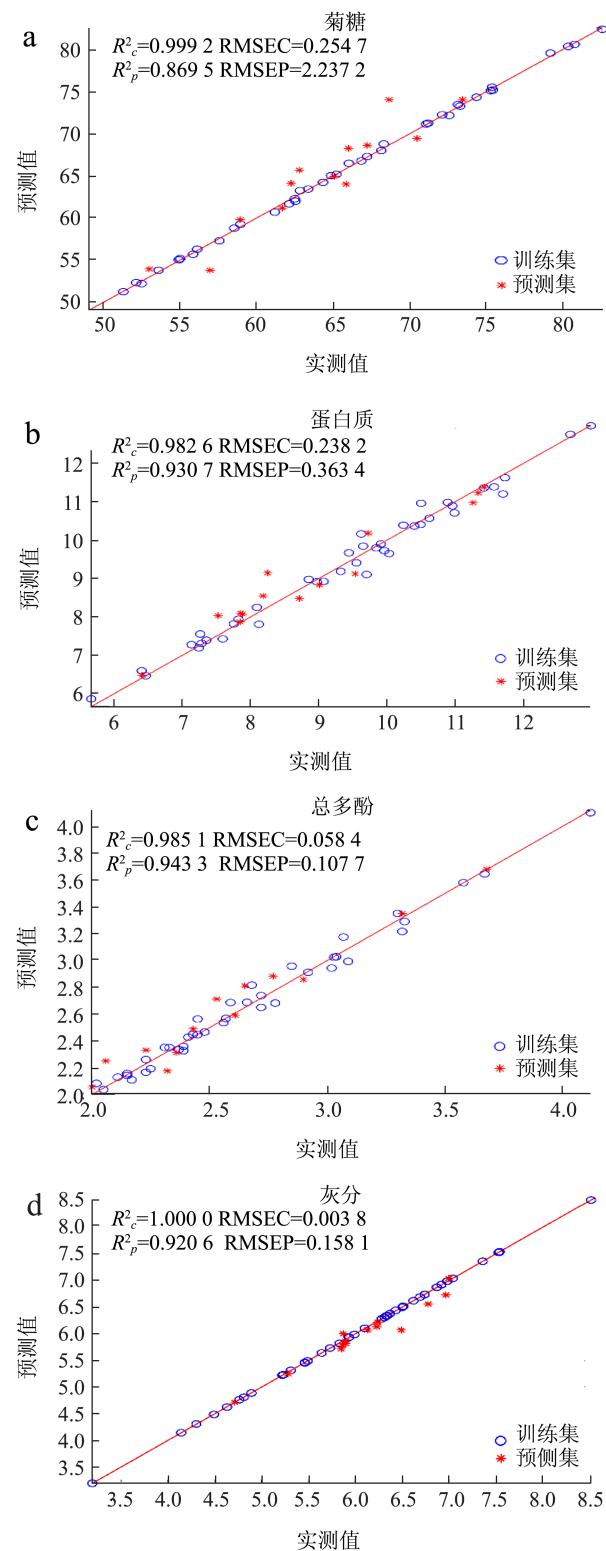


图 6 定量模型中预测值与实测值的散点分布图

Fig.6 The scatter plot of predicted versus actual values in the quantitative model

2.7 外部验证

为了探究模型对未知样品的适用性，利用所建立的定量模型对 5 个未知样品的理化指标进行含量预测，同时使用标准的化学方法检测其含量，得到预测

值与实测值的结果如表 6 所示。一般来说，两个测试结果的相对误差在 10% 以内表示该测试方法具有良好的准确性。结果显示，各理化指标中预测值与实测值的相对误差范围为 -14.61%~9.64%，其中菊糖、灰分指标中各有 1 个样本相对误差的绝对值超过 10%，分别为 -10.13% 和 -14.61%，表明菊糖、灰分定量模型的预测精度还有待提升，而蛋白质、总多酚定量模型对未知样品的预测效果较好。分析其原因主要有以下两点：一是菊糖、灰分预测模型中 $RMSEP$ 显著高于 $RMSEC$ ，其中灰分模型的 R^2_c 达到 1.000，这表示可能出现一定过拟合风险，需要进一步排除原因并采取相应的措施，如考虑进行多年采样补充模型样本量、使用正则化技术或集成技术来降低模型的复杂度。二是可能与菊粉果聚糖、灰分自身结构的复杂性有关，同时灰分中的大多数无机盐成分在近红外区域没有特征吸收，需与其他成分关联间接预测，增加了模型的复杂性，因此这两个指标的预测具有一定挑战性。

表 6 未知样品的外部验证结果

Table 6 External validation results of the unknown sample					
理化指标	样品编号	实测值	预测值	绝对误差	相对误差/%
菊糖 (g/100 g)	1	71.77	64.50	7.27	-10.13
	2	60.10	65.11	5.01	8.34
	3	67.28	62.24	5.04	-7.49
	4	74.04	67.76	6.28	-8.48
	5	67.80	67.66	0.15	-0.22
蛋白质 (g/100 g)	1	9.85	9.88	0.03	0.33
	2	9.14	8.53	0.61	-6.71
	3	12.04	12.44	0.39	3.26
	4	5.96	6.08	0.12	1.97
	5	13.21	12.64	0.57	-4.31
总多酚 (mg·g ⁻¹)	1	2.46	2.26	0.21	-8.39
	2	2.49	2.73	0.24	9.64
	3	2.60	2.37	0.24	-9.07
	4	2.50	2.37	0.13	-5.37
	5	2.54	2.32	0.22	-8.71
灰分 (g/100 g)	1	5.39	5.42	0.03	0.53
	2	5.67	5.81	0.14	2.52
	3	5.87	5.49	0.38	-6.49
	4	4.71	4.93	0.22	4.60
	5	6.93	5.91	1.01	-14.61

应用配对样本 t 检验将预测值与实测值进行比较，表 7 的结果表明，近红外光谱和标准化学方法得到的检测结果之间没有显著差异 ($P=0.194>0.05$)^[45,46]。因此，尽管菊糖、灰分定量模型的泛化能力稍差，但 t 检验结果进一步证实了利用近红外光谱预测菊芋中菊糖、蛋白质、总多酚、灰分含量的可行性。

表 7 实测值与预测值配对样本t检验结果

Table 7 Paired sample t-test results of measured versus predicted values

配对样本	平均值	标准差	标准误差平均值	差值 95% 置信区间		t	自由度	显著性 (双尾)
				上限	下限			
实测值 - 预测值	0.795 00	2.640 91	0.590 53	-0.440 98	2.030 98	1.346	19	0.194

由于 NIR 技术是一种大数据模拟技术，在未来应用中仍需关注和考虑定量模型存在的局限性，参与建模的样本数量越多，越有利于提高模型的稳定性和准确性。因此要不断引入新样本来更新模型的校准范围，提高模型的泛化能力。此外，还要考虑将复杂的建模代码转化为易于使用的系统软件，才能更好地在实际生产中进行应用^[47]。

3 结论

理化成分的检测对菊芋的加工应用和品质评价具有重要作用，目前传统的检测分析方法有高效液相色谱法、凯氏定氮法、分光光度法等，尽管检测的精确度较高，但前处理复杂，耗时长且对样品具有破坏性，不适合在实际生产中应用。本研究利用近红外光谱技术在 900~2 500 nm 范围内建立菊芋中菊糖、蛋白质、总多酚、灰分含量的 PLSR 预测模型，分别经过剔除异常样本、划分样本集、光谱预处理、特征波段筛选等模型建立过程，最终得到了预测性能良好的定量模型。其中菊糖 R^2_p 为 0.869 5，蛋白质 R^2_p 为 0.930 7，总多酚 R^2_p 为 0.943 3，灰分 R^2_p 为 0.920 6。利用未知样品进行外部验证，进一步证实了近红外光谱技术可作为传统化学分析方法的替代方案，实现了菊芋 4 种理化指标的同步快速测量，提高了分析效率，降低了成本，为菊芋理化指标的检测提供了新的参考。但为了提高模型的泛化能力，仍需要不断丰富建模的样本量以及尝试更多的建模方法，以便更好地在实际生产中进行应用。

参考文献

[1] APOLINÁRIO A C, de LIMA DAMASCENO B P G, de MACÊDO BELTRÃO N E, et al. Inulin-type fructans: a review on different aspects of biochemical and pharmaceutical technology [J]. Carbohydrate Polymers, 2014, 101: 368-378.

[2] QIN Y, WANG L, YANG X, et al. Inulin: properties and health benefits [J]. Food & Function, 2023, 14(7): 2948-2968.

[3] MAUMELA P, van RENSBURG E, CHIMPHANGO A F A, et al. Sequential extraction of protein and inulin from the tubers of Jerusalem artichoke (*Helianthus tuberosus* L.) [J]. Journal of Food Science and Technology, 2020, 57(2): 775-786.

[4] DIMITROVSKI D, VELICKOVA E, DIMITROVSKA M, et al. Synbiotic functional drink from Jerusalem artichoke juice fermented

by probiotic *Lactobacillus plantarum* PCS26 [J]. Journal of Food Science and Technology, 2016, 53(1): 766-774.

[5] SAED B, EL-WASEIF M, FAHMY H, et al. Physicochemical and sensory characteristics of instant mushroom soup enriched with Jerusalem artichoke and cauliflower [J]. Foods, 2022, 11(20): 3260.

[6] OZGOREN E, ISIK F, YAPAR A. Effect of Jerusalem artichoke (*Helianthus tuberosus* L.) supplementation on chemical and nutritional properties of crackers [J]. Journal of Food Measurement and Characterization, 2019, 13(4): 2812-2821.

[7] PEROVIĆ J, KOJIĆ J, KRULJ J, et al. Inulin determination by an improved HPLC-ELSD Method [J]. Food Analytical Methods, 2022, 15(4): 1001-1010.

[8] LI S, WU Q, YIN F, et al. Development of a combined trifluoroacetic acid hydrolysis and HPLC-ELSD method to identify and quantify inulin recovered from Jerusalem artichoke assisted by ultrasound extraction [J]. Applied Sciences, 2018, 8(5): 710.

[9] KUTSANEDZIE F Y H, GUO Z, CHEN Q. Advances in nondestructive methods for meat quality and safety monitoring [J]. Food Reviews International, 2019, 35(6): 536-562.

[10] POREP J U, KAMMERER D R, CARLE R. On-line application of near infrared (NIR) spectroscopy in food production [J]. Trends in Food Science & Technology, 2015, 46(2): 211-230.

[11] BEĆ K B, GRABSKA J, HUCK C W. Miniaturized NIR spectroscopy in food analysis and quality control: promises, challenges, and perspectives [J]. Foods, 2022, 11(10): 1465.

[12] HE H, WANG Y, ZHANG M, et al. Rapid determination of reducing sugar content in sweet potatoes using NIR spectra [J]. Journal of Food Composition and Analysis, 2022, 111: 104641.

[13] WANG Y, XU Y, WANG X, et al. Optimizing the effects of potato size and shape on near-infrared prediction models of potato quality using a linear-nonlinear algorithm [J]. Journal of Food Composition and Analysis, 2024, 135: 106679.

[14] LI M, HE S, WANG J, et al. An NIRS-based assay of chemical composition and biomass digestibility for rapid selection of Jerusalem artichoke clones [J]. Biotechnology for Biofuels, 2018, 11(1): 334.

[15] 李鑫,沈晓君,王媛媛,等.近红外光谱法快速测定晒青毛茶三种儿茶素组分含量[J].现代食品科技,2024,40(3):326-332.

[16] GB5009.5-2016,食品安全国家标准 食品中蛋白质的测定[S].

[17] KASZÁS L, KOVÁCS Z, KOROKNAI J, et al. Identification of key parameters and phytochemical composition of fiber and brown juice by-products from green biomass of Jerusalem artichoke (*Helianthus tuberosus* L.) [J]. Biomass and Bioenergy, 2024, 188: 107332.

[18] SINGLETON V L, ORTHOFER R, LAMUELA-RAVENTÓS R M. Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin-Ciocalteu reagent [J]. Methods Enzymol, 1999, 299: 152-178.

[19] DU M, CHENG X, QIAN L, et al. Extraction, physicochemical

- properties, functional activities and applications of inulin polysaccharide: a review [J]. *Plant Foods for Human Nutrition*, 2023, 78(2): 243-252.
- [20] GB/T 41377-2022, 菊粉质量要求[S].
- [21] GB 5009.4-2016, 食品安全国家标准 食品中灰分的测定[S].
- [22] 肖紫鸣, 李建勋, 张娅, 等. 基于近红外光谱技术快速测定甘薯多糖含量[J]. *食品安全质量检测学报*, 2022, 13(10): 3228-3236.
- [23] CHEN J, REN X, ZHANG Q, et al. Determination of protein, total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy [J]. *Journal of Cereal Science*, 2013, 58(2): 241-247.
- [24] XIE C, QIAO M, YANG L, et al. Establishment of a general prediction model for protein content in various varieties and colors of peas using visible-near-infrared spectroscopy [J]. *Journal of Food Composition and Analysis*, 2024, 127: 105965.
- [25] LI L, LI L, GOU G, et al. A nondestructive detection method for the multi-quality attributes of oats using near-infrared spectroscopy [J]. *Foods*, 2024, 13(22): 3560.
- [26] 田文强, 王巧华, 徐步云, 等. 基于近红外光谱的腌制期咸鸭蛋理化指标无损检测[J]. *食品科学*, 2023, 44(2): 319-326.
- [27] 陈蓓, 郑恩让, 郭拓. 多种光谱变量筛选算法在红参提取近红外建模中的应用[J]. *光谱学与光谱分析*, 2021, 41(8): 2443-2449.
- [28] LI Q, ZHOU W, WANG Q, et al. Research on online nondestructive detection technology of duck egg origin based on visible/near-infrared spectroscopy [J]. *Foods*, 2023, 12(9): 1900.
- [29] WU Z, LI C, LIU H, et al. Quantification of caffeine and catechins and evaluation of bitterness and astringency of Pu-erh ripen tea based on portable near-infrared spectroscopy [J]. *Journal of Food Composition and Analysis*, 2024, 125: 105793.
- [30] 张振宇, 常相伟, 严辉, 等. 基于近红外光谱分析技术的干姜质量快速评价研究[J]. *中草药*, 2022, 53(23): 7516-7523.
- [31] ZHU M, LONG Y, CHEN Y, et al. Fast determination of lipid and protein content in green coffee beans from different origins using NIR spectroscopy and chemometrics [J]. *Journal of Food Composition and Analysis*, 2021, 102: 104055.
- [32] ZHU G, TIAN C. Determining sugar content and firmness of 'Fuji' apples by using portable near-infrared spectrometer and diffuse transmittance spectroscopy [J]. *Journal of Food Process Engineering*, 2018, 41(6): e12810.
- [33] 仇逊超, 曹军. 近红外光谱波段优化在东北松子蛋白质定量检测中的应用[J]. *现代食品科技*, 2016, 32(11): 303-309.
- [34] ZHANG C, SHEN Y, CHEN J, et al. Nondestructive prediction of total phenolics, flavonoid contents, and antioxidant capacity of rice grain using near-infrared spectroscopy [J]. *Journal of Agricultural and Food Chemistry*, 2008, 56(18): 8268-8272.
- [35] 杨焰婷, 江谧, 詹瑞玲, 等. 青稞特征营养成分含量快速检测模型的建立及优化[J]. *食品工业科技*, 2024, 45(18): 228-238.
- [36] WIDYANINGRUM, PURWANTO Y A, WIDODO S, et al. Rapid assessment of vanilla (*Vanilla planifolia*) quality parameters using portable near-infrared spectroscopy combined with random forest [J]. *Journal of Food Composition and Analysis*, 2024, 133: 106346.
- [37] CHEN J, REN X, ZHANG Q, et al. Determination of protein, total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy [J]. *Journal of Cereal Science*, 2013, 58(2): 241-247.
- [38] YU Y, CHAI Y, LI Z, et al. Quantitative predictions of protein and total flavonoids content in Tartary and common buckwheat using near-infrared spectroscopy and chemometrics [J]. *Food Chemistry*, 2025, 462: 141033.
- [39] 陈勇, 吴彩娥, 熊智新. 基于衰减消去蜻蜓算法的小麦粉蛋白质近红外特征波长优选[J]. *食品科学*, 2022, 43(14): 219-225.
- [40] XIAOWEI H, XIAOBO Z, JIEWEN Z, et al. Measurement of total anthocyanins content in flowering tea using near infrared spectroscopy combined with ant colony optimization models [J]. *Food Chemistry*, 2014, 164: 536-543.
- [41] CHENG J, SUN D. Partial least squares regression (PLSR) applied to NIR and HSI spectral data modeling to predict chemical properties of fish muscle [J]. *Food Engineering Reviews*, 2017, 9(1): 36-49.
- [42] MEENU M, ZHANG Y, KAMBOJ U, et al. Rapid determination of β -glucan content of hulled and naked oats using near infrared spectroscopy combined with chemometrics [J]. *Foods*, 2022, 11(1): 43.
- [43] FAN C, LIU Y, CUI T, et al. Quantitative prediction of protein content in corn kernel based on near-infrared spectroscopy [J]. *Foods*, 2024, 13(24): 4173.
- [44] MEENU M, COZZOLINO D, XU B. Non-destructive prediction of total phenolics and antioxidants in hulled and naked oat genotypes with near-infrared spectroscopy [J]. *Journal of Food Measurement and Characterization*, 2023, 17(5): 4893-4904.
- [45] YI J, SUN Y, ZHU Z, et al. Near-infrared reflectance spectroscopy for the prediction of chemical composition in walnut kernel [J]. *International Journal of Food Properties*, 2017, 20(7): 1633-1642.
- [46] D.D. D, SHARMA V, MANGAL M, et al. NIR spectroscopy prediction model for capsaicin content estimation in chilli: a rapid mining tool for trait-specific germplasm screening [J]. *Journal of Food Composition and Analysis*, 2025, 137: 106915.
- [47] WEI C, LI X. A near-infrared spectroscopy method for the detection of texture profile analysis of *Litopeno vannamei* based on partial least squares regression [J]. *Journal of Food Process Engineering*, 2022, 45(10): e14140.