

· 热带植物 ·

DOI: 10.15886/j.cnki.rdswwb.20230031



主持人:徐 冉

## 木薯块根灰分和水分近红外光谱预测模型的构建与优化

张逸杰<sup>1</sup>, 王思琦<sup>2</sup>, 陆小静<sup>3</sup>, 宋记明<sup>4</sup>, 王 睿<sup>2</sup>, 曹 敏<sup>2</sup>, 张 瑞<sup>2</sup>,  
王红刚<sup>2</sup>, 吴金山<sup>1</sup>

(1. 海南大学 林学院, 海口 570228; 2. 海南大学 热带作物学院, 海口 570228; 3. 中国热带农业科学院 热带作物品种资源研究所, 海口 571101; 4. 云南省农业科学院 热带和亚热带经济作物研究所, 云南 保山 678000)

**摘 要:** 为快速检测木薯的灰分和水分含量, 以同地块同一时期木薯种质资源为材料进行建模, 采用 GB 5009.3—2016 和 GB/T5009.4—2016 法对木薯灰分和水分含量进行测定, 同时使用近红外光谱分析仪对 137 份样品进行光谱采集, 利用 TQ Analyst 9.0 分析软件, 采用偏最小二乘法 (PLS) 构建木薯灰分和水分近红外定标模型。实验结果显示, 木薯灰分、水分模型相关系数 ( $R$ ) 分别为 0.94、0.93, 校正均方根误差 (RMSEC) 分别为 0.22、0.48, 预测均方根误差 (RMSEP) 分别为 0.21、1.46, 交叉验证均方差 (RMSECV) 分别为 0.40、1.54; 选用未参与建模的 20 份木薯种质资源对该模型进行外部验证, 预测值与真实值进行差异性分析 ( $P>0.05$ ),  $P$  值分别为 0.464、0.459 说明差异无显著性, 表明该模型可适用于木薯灰分和水分检测。

**关键词:** 木薯; 灰分; 水分; 近红外光谱; 预测模型; 快速检测

**中图分类号:** S533 **文献标志码:** A **文章编号:** 1674-7054 (2024) 03-0259-09

张逸杰, 王思琦, 陆小静, 等. 木薯块根灰分和水分近红外光谱预测模型的构建与优化 [J]. 热带生物学报, 2024, 15(3): 259-267. doi: 10.15886/j.cnki.rdswwb.20230031

木薯 (*Manihot esculenta* Crantz) 别名木蕃薯、树薯, 是世界三大薯类作物之一, 在食用、饲用和工业等开发利用上均有着较大的价值<sup>[1]</sup>。木薯的灰分和水分指标与木薯品质密切相关, 但在木薯品种选育过程中, 利用马弗炉法和烘干法检测其所含灰分和水分流程较为繁琐<sup>[2]</sup>, 效率较低, 因此现阶段迫切需要一种高效率、高精度和低成本的方法来检测批量木薯灰分和水分含量指标。随着近红外光谱技术 [near-infrared (NIR) spectroscopy] 与化学计量方法的快速发展, 近红外光谱技术被广泛应用于农产品品质分析领域, 并已成为不可或缺的快速准确的测定方法<sup>[3]</sup>, 是工业生产过程中质量控制的首选技术。近红外光谱技术能满足常规方法不能完成的实时监测需求, 同时具有

无损样品、操作简单、检测快速等特点, 在成品检测和在线检测中都有优势<sup>[3-5]</sup>。研究人员在玉米<sup>[6]</sup>、小麦<sup>[7]</sup>等饲料原料中利用近红外光谱技术检测其营养成分和利用近红外光谱技术对三七<sup>[8]</sup>、玉米<sup>[9]</sup>、棕榈<sup>[10]</sup>和橄榄<sup>[11-12]</sup>果实的含油量进行定量分析。Arslan 等<sup>[13]</sup>采用近红外光谱法对枸杞总黄酮和花青素的化学成分进行了定量分析。Jin 等<sup>[14]</sup>利用可见光和近红外光谱 (VIS-NIR) 可在 30 770~9 300  $\text{cm}^{-1}$  范围内对芒草叶片含水量进行预测。Heo 等<sup>[15]</sup>构建了 PLS 预测模型 ( $R_p^2=0.95$ ), 利用选定波长分别为 961.12、1 065.50、1 083.93、1 173.23、1 233.89 nm 的 HIS 系统预测了蒸干紫薯中的水分含量。Rabanera 等<sup>[16]</sup>对波长范围为 900~1 700 nm 的区段利用 HSI 系统, 测定了花生仁水分含量, 并

收稿日期: 2023-03-08

修回日期: 2023-05-06

基金项目: 国家现代木薯产业技术体系 (CARS-11-HNCYH)

作者简介: 张逸杰 (1998-) 男, 海南大学林学院 2020 级硕士研究生。E-mail: 530259881@qq.com

通信作者: 吴金山 (1984-) 男, 博士, 副教授, 研究方向: 主要研究方向为植物生理生化。E-mail: wujsh2007@163.com

构建了  $R^2_p$  为 0.94、 $RMSEP$  为 1.95% 的 PLS 预测模型。Rahman 等<sup>[17]</sup>在波段 1 000~1 550 nm 范围内利用 HSI 技术,开发出一套无损检测番茄中水分含量和可溶性固形物含量的方法。目前国内外基于近红外光谱技术对木薯的研究多集中在表征特征分析等方面,而对其灰分和水分含量进行快速检测模型的研究较少。本研究以海南省海南大学儋州校区农科基地木薯种质资源圃为依托,拟以 137 份不同木薯种质块根样品为定标和验证材料,进行样品光谱收集,获取木薯有效的近红外光谱信息,并运用各种光谱数据预处理技术和统计方法进行校正、验证,确定各参数的最佳优化设置,初步建立了木薯灰分和水分含量预测模型,以期为进一步建立其他重要品质性状指标模型提供基础,为木薯育种及资源品质改良研究提供高效准确快速的分析方法。

## 1 材料与方法

**1.1 实验材料** 本研究所用的 137 份木薯种质种植于海南大学儋州校区农科基地木薯种植资源圃(海拔:380 m, 109°49'E、19°51'N)。实验地块每年统一连续耕作,2021 年 4 月种植,2022 年 3 月收获,土质为红壤土。实验用的 137 份木薯块根取自同一地块,生长期同为 6 个月龄。取好样品后将木薯块根削皮、切段,并在 60 °C 烘箱干燥 24 h 后,用研磨机研磨成粉状,过 80 目(0.18 mm)筛。粉碎后每个样品采用四分法分为 2 个重复,装入自封袋干燥避光处保存。

**1.2 仪器设备** 布鲁克 MPA 近红外光谱仪(美国布鲁克公司);DHG-9070A 电热鼓风干燥箱(上海一恒科学仪器有限公司);EL204 电子分析天平[0.001, 梅特勒有限-托利多仪器(上海)有限公司];300Y 多功能粉碎机(中国伯欧五金制品有限公司);RE52AA 旋转蒸发仪(上海亚荣生化仪器厂)。

**1.3 光谱数据采集** 采用漫反射扫描(diffuse reflectance scanning)方式扫描样品。扫描样品前将光谱仪开机预热 30 min,以保证样品测试的稳定性<sup>[15]</sup>。在约 25 °C 环境下,取适量处理后的木薯样品置于 IN311/C 旋转台上的 IN311-S 样品杯(直径 50 mm)中,摊平,压实,以空气为空白背景扣除(每扫描 1 次样品粉末,均要以空气为空白进行背景扣

除),用近红外光谱仪采集 137 份木薯样品 NIRS 图和 CSV 格式的数据矩阵,并对 137 份样品进行全谱段光谱扫描,光谱范围为 12 500~3 600  $\text{cm}^{-1}$ ,测量通道为 Sphere Macrosample Rotating,分辨率为 16  $\text{cm}^{-1}$ ,样品扫描 64 次,光谱点数为 1 154。每个样品扫描 3 次,取 3 次的平均光谱。除所需样品信息外,获得的近红外光谱会因混入仪器和测试环境产生的无关信息,导致光谱基线漂移(spectral baseline drift)、散射和噪声等问题<sup>[16]</sup>。常规是通过多元散射校正(MSC)、一阶导数(1S)、二阶导数(2S)、Savitzky-Golay 平滑模式(SG)、标准正态变量(SNV)等方法予以校正<sup>[17-18]</sup>。导数处理常用来消除光谱中存在的背景干扰,Savitzky-Golay 平滑消除噪声,SNV 和 MSC 主要消除光散射<sup>[19]</sup>。在近红外光谱中,采用不同的前处理方法会得到不同的模型结果。

**1.4 水分及灰分的测定** 采用 GB5009.3—2016 和 GB/T5009.4—2016 法对木薯灰分和水分含量进行测定。将木薯样品在 105 °C 烘箱中烘干过夜,直至达到恒重,以获得水分含量。灰分含量是在 550 °C 下焚烧 4 h 后所得,每个样品重复测定 2 次。水分含量和灰分含量计算公式如下,

$$X_{\text{含水量}} = \frac{M_0 + M - M_1}{M} \times 100\%, \quad (1)$$

式(1)中, $X_{\text{含水量}}$ :样品含水量; $M_0$ :称量瓶恒重后的质量; $M$ :样品的质量; $M_1$ :称量瓶与样品恒重后的质量。

$$X_{\text{灰分含量}} = \frac{M_1 - M_0}{M \times (1 - W)} \times 100\%, \quad (2)$$

式(2)中, $X_{\text{灰分含量}}$ :样品灰分含量; $M_1$ :坩埚和燃烧后残渣的质量; $M_0$ :空坩埚的质量; $M$ :样品的质量; $W$ :样品的含水量。

## 1.5 定量分析模型与评价模型

**1.5.1 模型建立** 采用偏最小二乘回归(PLS)对定量数据进行校正。PLS 法是一种基于“精度矩阵(the concentration matrix)包含无用信息,分解精度矩阵,消除精度矩阵中无用信息”的回归方法。通过最小化误差平方和(sum of squares of errors)来确定目标数据的最佳匹配函数,然后用最简单的最小化误差平方和的方法来获得真值。

**1.5.2 模型评价** 模型 A 采用交叉验证均方根误差(RMSECV)评价模型的整体校正能力,由式(3)

计算。

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^N (Y-X)^2}{N-P}}, \quad (3)$$

式(3)中,  $N$ : 样本数量;  $P$ : 主成分数量;  $Y$ : 预测值;  $X$ : 真实值。

模型B: 预测均方根误差(RMSEP)是模型的整体预测能力评价指标, 由式(4)计算。

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^N (Y-X)^2}{N}}, \quad (4)$$

式(4)中,  $N$ : 样本数量;  $Y$ : 预测值;  $X$ : 真实值。

## 2 结果与分析

**2.1 水分和灰分的测定结果** 不同木薯品种的水分和灰分含量(表1)的对比分析结果表明, 在不同品种的淀粉灰分含量中, 14号(老板娘2号)含量最低, 为1.03%, 91号(50)含量最高, 为4.26%, 参试品种的灰分含量平均值为2.00%。通过对木薯不同品种的水分含量分析得知, 49号(南美9号)含量最低, 为1.57%, 94号(417)含量最高, 为8.45%, 参试品种的水分含量平均值为4.54%。

表1 不同木薯品种的水分和灰分含量

Tab. 1 Water and ash contents of cassava Germplasm Resources

编号 No.	品种 Name	灰分/% Ash/%	水分/% Water/%	编号 No.	品种 Name	灰分/% Ash/%	水分/% Water/%
1	F2000	2.78 ± 0.04	4.24 ± 0.25	28	SC201	1.98 ± 0.06	2.95 ± 0.33
2	东莞红尾	1.94 ± 0.04	4.19 ± 0.37	29	F202	2.14 ± 0.05	3.60 ± 0.16
3	Q10	1.89 ± 0.00	4.86 ± 0.31	30	F03P	2.34 ± 0.02	4.50 ± 0.35
4	JG1301	2.53 ± 0.03	3.65 ± 0.39	31	COL712	2.06 ± 0.04	4.71 ± 1.39
5	SC9	2.14 ± 0.06	4.81 ± 0.25	32	SC2	1.71 ± 0.01	4.14 ± 0.86
6	MF-元引983	1.19 ± 0.02	3.34 ± 0.81	33	CH20	1.39 ± 0.01	4.87 ± 1.27
7	2B	1.75 ± 0.02	3.58 ± 2.13	34	SC6	2.04 ± 0.06	5.13 ± 1.44
8	老板娘3号	1.68 ± 0.01	5.21 ± 0.07	35	COL715	1.93 ± 0.05	4.95 ± 0.65
9	GR3	2.24 ± 0.01	5.05 ± 0.33	36	T12	2.00 ± 0.02	3.39 ± 0.57
10	ME191	1.18 ± 0.03	5.43 ± 0.28	37	8229	2.07 ± 0.02	4.29 ± 0.78
11	C3	1.29 ± 0.03	3.14 ± 2.52	38	ITBB01	1.97 ± 0.03	3.90 ± 0.30
12	泰国	1.15 ± 0.00	4.73 ± 0.08	39	KU58	2.14 ± 0.05	4.02 ± 0.24
13	云南思茅	1.72 ± 0.00	3.56 ± 0.04	40	ME190	2.26 ± 0.05	3.31 ± 0.05
14	老板娘2号	1.03 ± 0.03	3.22 ± 0.78	41	CM3992	1.97 ± 0.05	4.18 ± 0.66
15	GR911	1.95 ± 0.04	4.46 ± 0.30	42	SC205	1.72 ± 0.00	4.31 ± 0.10
16	KU50	1.83 ± 0.03	3.90 ± 0.47	43	C222	1.74 ± 0.00	5.92 ± 3.29
17	SC124多倍体	1.84 ± 0.06	3.27 ± 0.98	44	SC10	1.64 ± 0.02	3.47 ± 1.24
18	775	1.72 ± 0.00	2.36 ± 1.18	45	F297	2.06 ± 0.02	4.74 ± 0.17
19	Dec-46	1.16 ± 0.03	4.45 ± 0.71	46	水果木薯	1.72 ± 0.01	3.87 ± 0.61
20	SC16	2.03 ± 0.04	4.47 ± 0.52	47	SC124	1.64 ± 0.01	3.20 ± 0.11
21	哥伦比亚18R	2.10 ± 0.05	2.43 ± 0.05	48	SC12	2.16 ± 0.04	4.42 ± 2.23
22	SC4	1.41 ± 0.05	1.89 ± 1.21	49	南美9号	2.25 ± 0.04	1.57 ± 0.42
23	利民	1.73 ± 0.01	5.46 ± 2.86	50	花叶木薯	1.93 ± 0.04	4.21 ± 0.28
24	罗勇80	1.23 ± 0.00	3.30 ± 0.45	51	罗勇90	1.72 ± 0.01	7.07 ± 0.27
25	COL713	1.05 ± 0.02	3.54 ± 0.40	52	宝岛9-5	1.96 ± 0.01	5.78 ± 0.01
26	SC5	1.12 ± 0.03	4.47 ± 0.26	53	E1340	1.52 ± 0.01	6.08 ± 0.74
27	GR6	1.91 ± 0.04	4.16 ± 1.00	54	ME189	1.39 ± 0.01	4.78 ± 0.55

续表1 Tab.1 Continued

编号 No.	品种 Name	灰分/% Ash/%	水分/% Water/%	编号 No.	品种 Name	灰分/% Ash/%	水分/% Water/%
55	NZ188	1.84 ± 0.06	5.76 ± 0.43	92	196	3.67 ± 0.33	4.04 ± 0.11
56	琼中1号	1.63 ± 0.04	4.71 ± 0.21	93	571	4.24 ± 0.19	3.86 ± 0.16
57	泰国种	1.57 ± 0.02	5.82 ± 0.00	94	417	3.05 ± 0.07	8.45 ± 0.21
58	新选048	1.65 ± 0.01	4.22 ± 1.94	95	ZM8229	4.12 ± 0.16	4.38 ± 0.21
59	H660	1.54 ± 0.05	5.3 ± 0.39	96	521	3.88 ± 0.00	4.17 ± 0.08
60	SC8辐射	2.41 ± 0.02	3.55 ± 2.64	97	SC8013	3.06 ± 0.08	3.33 ± 0.15
61	NZ199	1.23 ± 0.03	4.01 ± 1.69	98	274	3.27 ± 0.07	4.52 ± 0.07
62	ME187	1.89 ± 0.02	6.27 ± 0.29	99	428	4.20 ± 0.04	3.51 ± 0.05
63	276	1.78 ± 0.03	3.86 ± 1.00	100	6068	3.00 ± 0.05	6.42 ± 0.22
64	18R	2.09 ± 0.03	4.85 ± 0.11	101	MB	3.69 ± 0.14	3.61 ± 0.14
65	COC173	2.11 ± 0.06	3.49 ± 2.23	102	T3B1	2.35 ± 0.06	6.1 ± 0.04
66	C9	1.39 ± 0.01	5.22 ± 0.51	103	T2B	1.46 ± 0.03	6.24 ± 0.00
67	M-SC124	1.77 ± 0.06	5.64 ± 0.01	104	T1A1	2.63 ± 0.04	5.20 ± 0.21
68	MF556	1.25 ± 0.01	3.1 ± 2.34	105	T4C1	2.42 ± 0.03	5.47 ± 0.14
69	老板娘1号	1.16 ± 0.03	6.39 ± 1.2	106	T2A1	1.43 ± 0.03	6.01 ± 0.05
70	MF81	1.57 ± 0.04	4.91 ± 0.04	107	T3A1	1.98 ± 0.05	5.08 ± 0.07
71	F539	1.65 ± 0.01	4.95 ± 0.46	108	T4B1	1.32 ± 0.02	5.59 ± 0.07
72	M-GR911	1.65 ± 0.04	3.35 ± 1.16	109	T1C1	1.92 ± 0.05	5.01 ± 0.17
73	GR891	1.97 ± 0.03	4.24 ± 0.13	110	T4A1	1.26 ± 0.03	5.32 ± 0.35
74	MF648	1.41 ± 0.02	4.19 ± 0.75	111	T3C1	3.36 ± 0.04	5.35 ± 0.14
75	MF671	1.29 ± 0.02	5.10 ± 0.30	112	T2C1	2.66 ± 0.05	4.76 ± 0.28
76	SC7	2.92 ± 0.04	4.98 ± 0.50	113	T1B	1.96 ± 0.04	5.18 ± 0.04
77	M 糯米	1.49 ± 0.04	4.55 ± 0.52	114	1 崖州扁豆左6	2.36 ± 0.03	5.46 ± 0.21
78	C576	1.84 ± 0.03	4.34 ± 1.47	115	1 撒播黑豆右3	3.08 ± 0.02	4.33 ± 0.07
79	C1030	1.56 ± 0.05	7.70 ± 0.90	116	1 柱花草左7	2.67 ± 0.01	3.72 ± 0.14
80	MF532	1.99 ± 0.03	4.29 ± 0.93	117	1 崖州扁豆左5	2.73 ± 0.09	4.22 ± 0.28
81	SC8	1.64 ± 0.03	4.50 ± 0.63	118	1 花生左6	2.29 ± 0.06	4.53 ± 0.32
82	MF359	1.19 ± 0.01	6.13 ± 1.75	119	1 花生右1	3.26 ± 0.01	4.36 ± 0.24
83	M-FC10	1.67 ± 0.01	3.50 ± 0.69	120	1 对照2右3	1.99 ± 0.05	2.41 ± 0.08
84	MSC5	1.07 ± 0.02	3.94 ± 0.93	121	1 崖州扁豆右4	1.99 ± 0.02	3.61 ± 0.08
85	E407	1.77 ± 0.00	3.61 ± 0.96	122	1 柱花草右3	1.85 ± 0.03	3.25 ± 0.13
86	MF181	1.12 ± 0.02	4.02 ± 1.17	123	1 崖州扁豆右6	1.20 ± 0.02	3.16 ± 0.11
87	M-SC4	1.41 ± 0.01	2.47 ± 0.01	124	1 油菜花3左1	2.72 ± 0.05	3.37 ± 0.23
88	OOO	1.25 ± 0.01	3.96 ± 0.42	125	1 左撒崖州扁豆	1.78 ± 0.03	4.24 ± 0.09
89	SC11	1.79 ± 0.00	5.68 ± 0.02	126	1 黑豆撒左4	2.59 ± 0.08	3.59 ± 0.35
90	F274	2.26 ± 0.02	4.10 ± 1.68	127	1 花生左4	2.48 ± 0.01	3.17 ± 0.20
91	50	4.26 ± 0.07	4.07 ± 0.07	128	1 对照左1	2.04 ± 0.06	3.61 ± 0.23

续表1 Tab.1 Continued

编号 No.	品种 Name	灰分/% Ash/%	水分/% Water/%	编号 No.	品种 Name	灰分/% Ash/%	水分/% Water/%
129	1 崖州扁豆撒右2	2.61 ± 0.07	2.56 ± 0.40	134	1 油菜间作撒右1	2.26 ± 0.06	6.35 ± 0.13
130	1 柱花草2右2	2.40 ± 0.02	6.44 ± 0.28	135	1 柱花草1左2	2.20 ± 0.02	7.63 ± 0.10
131	1 柱花草右5	1.90 ± 0.03	6.39 ± 0.08	136	1 右8对照	1.06 ± 0.02	7.14 ± 0.09
132	1 撒播柱花草右4	2.29 ± 0.01	6.49 ± 0.07	137	1 对照左8	1.07 ± 0.01	8.15 ± 0.11
133	1 花生右7	1.17 ± 0.04	7.56 ± 0.57				

注:崖州扁豆、柱花草、花生、油菜花分别代表与木薯SC9品种间作的不同植物种类。

Note: *Macrotyloma uniflorum* 'Yazhou', stylosanthes, peanut and rape flower represent different plant species intercropping with cassava SC9 variety respectively

**2.2 样本光谱采集结果分析** 在4 000~12 000 cm<sup>-1</sup>范围内计算了137个样品的平均光谱。由原始光谱结果(图1)可知,137个样品的吸收光谱变化趋势相似。但由于木薯成分复杂,不同碱基对应不同组分(如甲基和亚甲基)或相同碱基在不同化学环境下的近红外吸收波长和强度有显著差异,因此木薯各成分的光谱吸收强度不同。图1分布在9 000~12 000 cm<sup>-1</sup>波段范围内的谱线明显可分为两类,由此表明根据化学成分和分子结构可将137个样品分为两类。

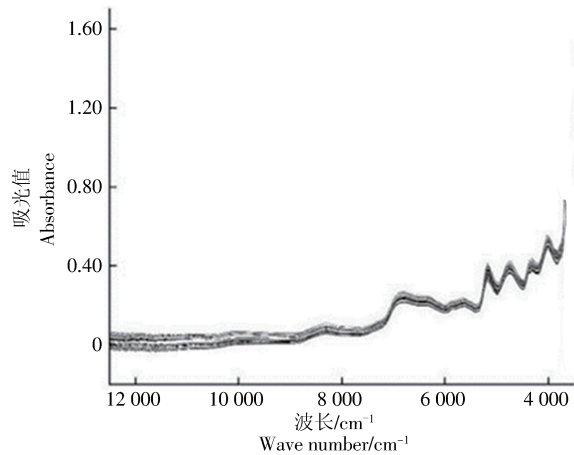


图1 参试木薯种质资源近红外光谱图

Fig. 1 Near infrared spectrum of cassava Germplasm Resources

**2.3 近红外定量模型的建立**

**2.3.1 校正集和验证集** 在TQ Analyst 9.0分析软件中,采用偏最小二乘法(PLS)对137份木薯样品进行光谱离群值剔除(Spectrum Qutlier)分析,未发现存在影响定量模型建立的异常样本<sup>[20]</sup>,同时根据软件优化选择测定灰分和水分2个指标的校正集与验证集,保证将最小和最大含量值的样本划入校正集中,具体划分见表2。

表2 校正集与验证集样本信息

Tab. 2 Sample information of correction set and verification set

成分 Component	校正集 Correction set		验证集 Validation set	
	样本量 sample size	范围 Range	样本量 sample size	范围 Range
灰分/% Ash/%	117	1.030~4.260	20	1.150~4.200
水分/% Water/%	117	1.570~8.450	20	2.470~7.140

**2.3.2 优化预处理方法和建模间隔** 对木薯样品光谱数据在测定波长范围内进行总体方差计算,以每一波长点的方差值[D(X)]对波长作图(图2),观察光谱数据变异性大小<sup>[21]</sup>。由图2可见,在6 184.92~4 309.97 cm<sup>-1</sup>和3 750.18~3 594.66 cm<sup>-1</sup>波段内曲线波动明显,数据变异性强,可见在此区间内的信息量丰富。以此波段为特征区间,考察全光谱波段(用a表示)与此特征区间(用b表示)的建模效果。

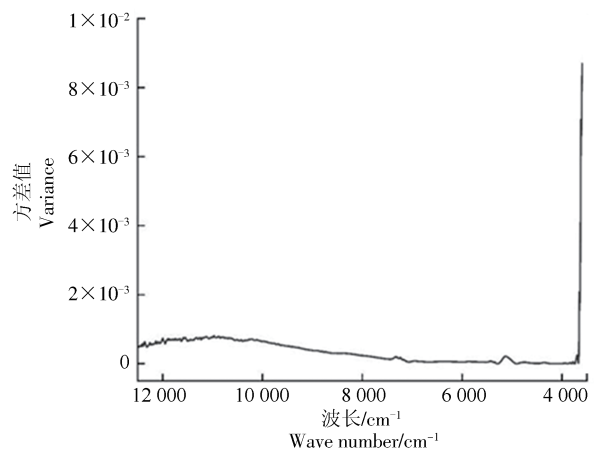


图2 木薯样品光谱数据方差分析图

Fig. 2 Analysis of variance chart



为了将全光谱波段与模型优化预测波段进行比较,确定最佳建模波段,分别考察了全光谱波段和光谱波段特征区间(5 800~4 300 $\text{cm}^{-1}$ 和4 000~3 600  $\text{cm}^{-1}$ )的建模效果(表3)。检测 MSC、1S、2S、SG 和 SNV 5种预处理方法对定量建模影响,以校正模型标准差(RMSEC)、预测均方根误差(RMSEP)、

交叉验证均方差(RMSECV)和相关系数(R)作为模型的评价指标,结果表明,对于水分和灰分,经 MSC+2S+SG 预处理后,在选择b波长区间分别建立模型时,R值最大,RMSEC、RMSEP以及RMSECV接近于0。这说明经 MSC+2S+SG 预处理后,选择b为建模区间得到的模型预测结果更准确<sup>[22]</sup>。

表3 不同预处理以及建模区间的考察结果

Tab. 3 Investigation results of different pretreatment and modeling intervals

指标 Index	预处理 Pretreatment	区间 Section	相关系数 R	校正标准差 RMSEC	预测标准差 RMSEP	交叉标准差 RMSECV	
灰分 Ash	Constant(原始)	a	0.742 8	0.457 0	0.636 0	0.514 0	
		b	0.759 8	0.444 0	0.591 0	0.498 0	
	MSC	a	0.769 4	0.436 0	0.403 0	0.509 0	
		b	0.774 4	0.432 0	0.675 0	0.575 0	
	SNV	a	0.770 8	0.435 0	0.404 0	0.518 0	
		b	0.729 4	0.467 0	0.633 0	0.579 0	
	MSC+1S+SG	a	a	0.826 39	0.384 0	0.636 0	0.515 0
			b	0.904 0	0.292 0	0.521 0	0.535 0
		SNV+1S+SG	a	0.828 8	0.382 0	0.628 0	0.513 0
			b	0.906 2	0.289 0	0.554 0	0.495 0
		MSC+2S+SG	a	0.873 5	0.332 0	0.820 0	0.590 0
			b	0.944 9	0.223 0	0.210 0	0.398 0
SNV+2S+SG	a	0.876 0	0.329 0	0.810 0	0.623 0		
	b	0.925 7	0.270 00	0.247 0	0.495 0		
水分 Water	Constant(原始)	a	0.788 8	0.784 0	1.280 0	1.440 0	
		b	0.693 9	0.919 0	1.300 0	1.440 0	
	MSC	a	0.778 0	0.801 0	1.340 0	1.280 0	
		b	0.685 5	0.929 0	1.280 0	1.420 0	
	SNV	a	0.784 1	0.792 0	1.380 0	1.440 0	
		b	0.696 2	0.916 0	1.400 0	1.470 0	
	MSC+1S+SG	a	a	0.841 4	0.689 0	1.450 0	1.500 0
			b	0.820 7	0.729 0	1.220 0	1.530 0
		SNV+1S+SG	a	0.844 8	0.683 0	1.450 0	1.480 0
			b	0.884 7	0.595 0	1.370 0	1.640 0
		MSC+2S+SG	a	0.863 5	0.643 0	1.710 0	1.620 0
			b	0.926 4	0.480 0	1.460 0	1.540 0
SNV+2S+SG	a	0.862 8	0.645 0	1.710 0	1.620 0		
	b	0.892 6	0.575 0	1.340 0	1.590 0		

注:区间a:全光谱波段;区间b:光谱波段特征区间(5 800~4 300  $\text{cm}^{-1}$ 和4 000~3 600  $\text{cm}^{-1}$ )。

Note: Interval a: full spectrum band; Interval b: spectral band characteristic interval (5 800~4 300  $\text{cm}^{-1}$  and 4 000~3 600  $\text{cm}^{-1}$ )

**2.4 定量模型** 根据上述最佳光谱预处理法和最优建模区间,采用PLS建立灰分和水分的近红外定量分析模型。最优定量模型均为通过MSC+2S+SG处理,以 $5\ 800\sim 4\ 300\ \text{cm}^{-1}$ 和 $4\ 000\sim 3\ 600\ \text{cm}^{-1}$ 为建模区间建立而成。通过RMSECV对主因子数作

图获得最佳主因子数分别为14、16;模型相关系数(R)分别为0.944 9、0.926 4;RMSEC依次为0.223 0、0.480 0;RMSEP依次为0.210 0、1.460;RMSECV依次为0.398 0和1.540。R值均接近于1,表明预测值与参考值间存在良好的相关性(图3)。

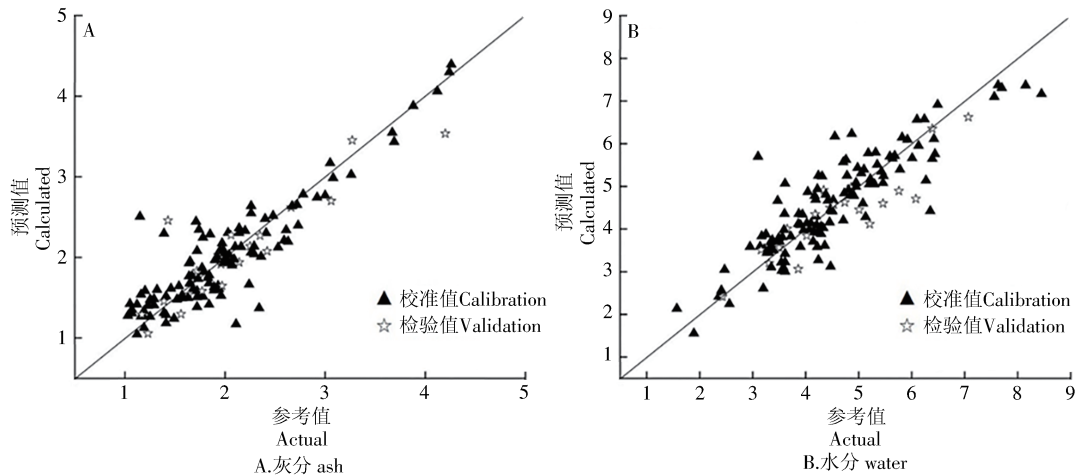


图3 木薯样品预测值和参考值的相关图

Fig. 3 Correlation diagram between predicted value and reference value of cassava sample

**2.5 可行性判断** 为进一步研究灰分与水分2种优化模型的可预测性,分析偏最小二乘法拟合得到验证集指标值与真实值的拟合程度,在95%置信区间内进行配对t检验,分别检验最优模型的拟合值,验证定量模型可靠性,可在实际检测中实时反馈木薯品质。选用未参与建模的20份木薯种质资源对该模型进行外部验证,预测值与真实值进行差异性分析, ( $P>0.05$ ),  $P$ 值分别为0.464、0.459,说明差异无显著性。由此表明,实验建立的预测模型可实现对不同产地木薯的灰分和水分的快速无损检测。

### 3 讨论

由于化学成分在不同波长处吸收强度差异,不同波长段所蕴藏的信息差异明显,选择合理波长范围和预处理方法可有效地提高模型性能。近红外光谱(NIRS)与多变量分析相结合已被用于测定木薯块根中的类胡萝卜素、纤维素、氰化物、干物质、淀粉、总糖和总氮含量等多种营养成分<sup>[23-25]</sup>,但对灰分和水分的测定模型构建仍然有待研究,本研究结果可以对其进行补充完善。李荣<sup>[26]</sup>在对波长 $5\ 296\sim 12\ 000\ \text{cm}^{-1}$ 内益智水分和灰分含量进行NIRS定量分析模型构建后发现,MSC+2S+

SG和SNV+2S分别为益智水分和灰分建模的最佳预处理方法,这与本实验结果相类似<sup>[27-29]</sup>。定标模型建好后,检测木薯样品中的水分、灰分的含量仅需光谱扫描和软件分析2步即可完成。此外,利用近红外方法进行分析测定,所得结果较为准确,且稳定性较高,可以有效减少样品的损耗,具有较好的应用前景<sup>[30-33]</sup>。

本研究采集的木薯样品均来自同一产地,所测灰分和水分含量分布范围广,具有一定的代表性,可用于分析模型的建立。考虑到木薯产地分布广泛,种属较多,定量分析模型在实际应用过程中需补充新的样品对模型进行完善,以扩大模型的适应性。

### 4 结论

本研究建立了新型的近红外光谱分析新方法,能够成功用于快速检测木薯水分、灰分的含量。结果显示,水分和灰分的最优定量模型均为通过MSC+2S+SG处理,以 $5\ 800\sim 4\ 300\ \text{cm}^{-1}$ 和 $4\ 000\sim 3\ 600\ \text{cm}^{-1}$ 为建模区间建立而成。预测值与真实值之间无显著差异,由此表明2种方法之间不存在明显系统误差。

## 参考文献:

- [1] SADRAS V O, CALDERINI D F. Preface [M]//Crop Physiology Case Histories for Major Crops. Amsterdam: Elsevier, 2021: xvii–xix.
- [2] SUMARDIONO S, JOS B, PUDJIHASTUTI I, et al. Effect of chemical modification, drying method, and drying temperature on baking expansion and the physicochemical properties of cassava starch [J]. *Journal of Food Processing and Preservation*, 2022, 46(1): e16111.
- [3] PITAK L, SIRISOMBOON P, SAENGPRACHATANARUG K, et al. Rapid elemental composition measurement of commercial pellets using line-scan hyperspectral imaging analysis [J]. *Energy*, 2021, 220: 119698.
- [4] CHEN H, TAN C, LIN Z, et al. Quantifying several adulterants of notoginseng powder by near-infrared spectroscopy and multivariate calibration [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2019, 211: 280–286.
- [5] JIANG D, QI G, HU G, et al. A residual neural network based method for the classification of tobacco cultivation regions using near-infrared spectroscopy sensors [J]. *Infrared Physics & Technology*, 2020, 111: 103494.
- [6] QIU G, LÜ E, WANG N, et al. Cultivar classification of single sweet corn seed using Fourier transform near-infrared spectroscopy combined with discriminant analysis [J]. *Applied Sciences*, 2019, 9(8): 1530.
- [7] BASATI Z, JAMSHIDI B, RASEKH M, et al. Detection of sunn pest-damaged wheat samples using visible/near-infrared spectroscopy based on pattern recognition [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2018, 203: 308–314.
- [8] CHEN H, TAN C, LIN Z, et al. Quantifying several adulterants of notoginseng powder by near-infrared spectroscopy and multivariate calibration [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2019, 211: 280–286.
- [9] TALLADA J G, PALACIOS-ROJAS N, ARMSTRONG P R. Prediction of maize seed attributes using a rapid single kernel near infrared instrument [J]. *Journal of Cereal Science*, 2009, 50(3): 381–387.
- [10] SUDARNO, SILALAH D D, RISMAN T, et al. Rapid determination of oil content in dried-ground oil palm mesocarp and kernel using near infrared spectroscopy [J]. *Journal of Near Infrared Spectroscopy*, 2017, 25(5): 338–347.
- [11] SAHA U, JACKSON D. Analysis of moisture, oil, and fatty acid composition of olives by near-infrared spectroscopy: development and validation calibration models [J]. *Journal of the Science of Food and Agriculture*, 2018, 98(5): 1821–1831.
- [12] ALTIERI G, MATERA A, GENOVESE F, et al. Models for the rapid assessment of water and oil content in olive pomace by near-infrared spectrometry [J]. *Journal of the Science of Food and Agriculture*, 2020, 100(7): 3236–3245.
- [13] ARSLAN M, ZOU X, HU X, et al. Near infrared spectroscopy coupled with chemometric algorithms for predicting chemical components in black goji berries (*Lycium ruthenicum* Murr.) [J]. *Journal of Near Infrared Spectroscopy*, 2018, 26(5): 275–286.
- [14] JIN X, SHI C, YU C Y, et al. Determination of leaf water content by visible and near-infrared spectrometry and multivariate calibration in *Miscanthus* [J]. *Frontiers in Plant Science*, 2017, 8: 721.
- [15] HEO S, CHOI J Y, KIM J, et al. Prediction of moisture content in steamed and dried purple sweet potato using hyperspectral imaging analysis [J]. *Food Science and Biotechnology*, 2021, 30(6): 783–791.
- [16] RABANERA J D, GUZMAN J D, YAPTENCO K F. Rapid and Non-destructive measurement of moisture content of peanut (*Arachis hypogaea* L.) kernel using a near-infrared hyperspectral imaging technique [J]. *Journal of Food Measurement and Characterization*, 2021, 15(4): 3069–3078.
- [17] RAHMAN A, KANDPAL L, LOHUMI S, et al. Nondestructive estimation of moisture content, pH and soluble solid contents in intact tomatoes using hyperspectral imaging [J]. *Applied Sciences*, 2017, 7(1): 109.
- [18] ZOU X, ZHAO J, POVEY M J W, et al. Variables selection methods in near-infrared spectroscopy [J]. *Analytica Chimica Acta*, 2010, 667(1/2): 14–32.
- [19] OZAKI Y, CHRISTY A A, MCCLURE W F. Near-infrared spectroscopy in food science and technology[M]. New York: John Wiley & Sons, 2006: 23–25.
- [20] FAZELI BURESTAN N, AFKARI SAYYAH A H, SAFI M. Prediction of amylose content, protein content, breakdown, and setback viscosity of *Kadus* rice and its flour by near-infrared spectroscopy (NIRS) analysis [J]. *Journal of Food Processing and Preservation*, 2021, 45(1): e15069.
- [21] OZAKI Y, HUCK C, TSUCHIKAWA S, et al. Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications [M]. Singapore: Springer Singapore, 2021: 57–63.
- [22] JIAO Y, LI Z, CHEN X, et al. Preprocessing methods for near-infrared spectrum calibration [J]. *Journal of Chemometrics*, 2020, 34(11): e3306.
- [23] LEBOT V, CHAMPAGNE A, MALAPA R, et al. NIR determination of major constituents in tropical root and *Tuber* crop flours [J]. *Journal of Agricultural and Food Chemistry*, 2009, 57(22): 10539–10547.
- [24] LEBOT V, MALAPA R, JUNG M. Use of NIRS for the rapid prediction of total N, minerals, sugars and starch in tropical root and tuber crops [J]. *New Zealand Journal of Crop and Horticultural Science*, 2013, 41(3): 144–153.
- [25] SÁNCHEZ T, CEBALLOS H, DUFOUR D, et al. Predic-



- tion of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques [J]. *Food Chemistry*, 2014, 151: 444–451.
- [26] 李荣. 近红外光谱技术对益智质量评价的研究 [D]. 广州: 广东药科大学, 2020.
- [27] MISHRA P, ROGER J M, MARINI F, et al. Parallel pre-processing through orthogonalization (PORTO) and its application to near-infrared spectroscopy [J]. *Chemometrics and Intelligent Laboratory Systems*, 2021, 212: 104190.
- [28] OLIVEIRA U F, COSTA A M, ROQUE J V, et al. Predicting oil content in ripe Macaw fruits (*Acrocomia aculeata*) from unripe ones by near infrared spectroscopy and PLS regression [J]. *Food Chemistry*, 2021, 351: 129314.
- [29] LIU Y, SUN L, DU C, et al. Near-infrared prediction of edible oil frying times based on Bayesian Ridge Regression [J]. *Optics*, 2020, 218(12): 164950.
- [30] 高美凤, 陶焕明. 改进团队进步算法的近红外光谱波长筛选 [J]. *光谱学与光谱分析*, 2021, 41(10): 3032–3038.
- [31] 陶焕明. 光谱波长筛选方法及其在近红外检测中的应用 [D]. 无锡: 江南大学, 2021.
- [32] 洪士军, 黄雯, 张立国, 等. 基于尺度不变特征变换筛选稳定特征波长的近红外光谱模型传递方法 [J]. *分析测试学报*, 2020, 39(10): 1260–1266.
- [33] 吕晓菡, 蒋锦琳, 杨静, 等. 基于特征波长建模的近红外光谱技术检测辣椒素含量 [J]. *浙江大学学报(农业与生命科学版)*, 2019, 45(6): 760–766.

## Construction and optimization of prediction models for cassava ash and water content by using near-infrared spectroscopy

ZHANG Yijie<sup>1</sup>, WANG Siqi<sup>2</sup>, LU Xiaojing<sup>3</sup>, SONG Jiming<sup>4</sup>, WANG Rui<sup>2</sup>, CAO Min<sup>2</sup>,  
ZHANG Rui<sup>2</sup>, WANG Honggang<sup>2</sup>, WU Jinshan<sup>1</sup>

(1. School of Forestry, Hainan University, Haikou, Hainan 570228, China; 2. School of Tropical Crops, Hainan University, Haikou, Hainan 570228, China; 3. Research Institute of Tropical Crops Genetic Resources, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101, China; 4. Institute of Tropical and Subtropical Cash Crops, Yunnan Academy of Agricultural Sciences, Baoshan, Yunnan 678000, China)

**Abstract:** Ash content and moisture content are important indexes for cassava quality evaluation. It is of great significance to establish a method for rapid detection of ash and moisture contents in cassava by using near infrared spectroscopy combined with multivariate mathematical analysis. Tubers of 137 accessions of cassava were harvested at the cassava growing site and their ash and water contents were determined by using the methods described in the standards GB5009.3—2016 and GB/T5009.4—2016 and at the same time the near-infrared spectroscopy. The NIR spectral data collected were analyzed by using the software TQ Analyst 9.0. The NIR calibration models for ash and water contents of cassava were constructed by using the partial least squares (PLS) method. The results showed that correlation coefficients (R) of the ash and water content models were 0.94 and 0.93, respectively. Root mean square error of correction (RMSEC) was 0.22 and 0.48, respectively. Root mean square error of prediction (RMSEP) was 0.21 and 1.46, respectively. Root mean square error of cross-validation (RMSECV) was 0.40 and 1.54, respectively. Twenty accessions of cassava germplasm that were not involved in the modelling were selected for external validation of the model. The results showed that there is no significant difference between the predicted value and the actual value, with the p-values ( $P>0.05$ ) being 0.464 and 0.459, respectively, indicating that the model could be applied to the detection of cassava ash and water contents. The near-infrared quantitative detection models of ash and moisture established can be used for rapid detection.

**Keywords:** cassava; ash; water; near-infrared spectroscopy; prediction model; rapid detection method