

# 近红外定量模型快速测定大米的营养成分

黄林森<sup>1,2</sup>, 刘冬<sup>1</sup>, 覃统佳<sup>1,2</sup>, 林秀文<sup>1,2</sup>, 周志航<sup>1</sup>, 黄锦玲<sup>1</sup>, 从彦丽<sup>1</sup>

(1. 深圳职业技术学院应用化学与生物技术学院, 广东深圳 518000)

(2. 深圳大学化学与环境工程学院, 广东深圳 518000)

**摘要:** 大米中的蛋白质、脂肪、碳水化合物和水分是大米的重要营养成分, 这些组分的检测需要依赖国标法, 检测过程繁琐冗长, 污染大, 实现近红外快速检测具有重要的现实意义。本实验以 75 个大米样品为研究对象, 用国标法测定其蛋白质、脂肪、碳水化合物和水分含量, 用 Antares II 型近红外光谱仪采集光谱信息。将样品分为校正集和验证集, 其中校正集 63 个, 验证集 12 个, 通过 6 种光谱预处理方法和 3 种建模方法分别建立大米的蛋白质、脂肪、碳水化合物和水分定量模型, 用 Workflow 调用模型来实现四种组分的快速检测。结果表明, 蛋白质、碳水化合物通过一阶导数与 Savitzky-Golay 滤波对光谱进行处理; 脂肪、水分通过二阶导数与 Savitzky-Golay 滤波对光谱进行处理; 再通过 PLS 回归与国标法的测定值相关联建立的大米中蛋白质、脂肪、碳水化合物及水分的定量模型具有较高的预测精度。蛋白质、脂肪、碳水化合物和水分的预测模型的内部交叉验证的相关系数 R 分别为 0.9266、0.9333、0.9198、0.9175, RMSECV 分别为 0.280、0.133、0.362、0.270, 内部交叉验证的相关系数 R 均在 0.91 以上, RMSECV 均小于 0.37, 可实现对大米中蛋白质、脂肪、碳水化合物和水分的快速检测。

**关键词:** 近红外光谱分析技术; 定量模型; 快速检测

文章编号: 1673-9078(2019)08-317-324

DOI: 10.13982/j.mfst.1673-9078.2019.8.044

## Near Infrared Spectral Quantitative Model of the Nutrient Content in Rice

HUANG Lin-sen<sup>1,2</sup>, LIU Dong<sup>1</sup>, QIN Tong-jia<sup>1,2</sup>, LIN Xiu-wen<sup>1,2</sup>, ZHOU Zhi-hang<sup>1</sup>, HAUNG Jin-ling<sup>1</sup>,  
CONG Yan-li<sup>1</sup>

(1. School of Applied Chemistry and Biological Technology, Shenzhen Polytechnic, Shenzhen 518000, China)

(2. College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen 518000, China)

**Abstract:** Protein, fat, carbohydrate and moisture in rice are important nutrients in rice. The detection of these components needs to rely on the national standard method. The detection process is tedious and has large pollution. In this experiment, 75 rice samples were studied, and their contents of protein, fat, carbohydrate and moisture were measured by national standard method. Spectral information was collected by Antares II infrared spectrometer. Samples were divided into verification set and correction set, including 63 samples in correction set and 12 samples in verification set. A quantitative model of protein, fat, carbohydrate and moisture of rice was established through six spectral pretreatment methods and three modeling methods, and the model by running workflow was used to realize rapid detection of four components. The results showed that proteins and carbohydrates followed by the first derivative and Savitzky-Golay filtering, while fat and water followed by the second derivative and Savitzky-Golay filtering. The quantitative model of protein, fat, carbohydrate and water content in rice was established by Least partial square method in combination with the measured values by national standard method and has high prediction accuracy. The internal cross-validation coefficient of the prediction model of protein, fat, carbohydrate and moisture is 0.9266, 0.9333, 0.9198, 0.9175, and RMSECV is 0.280, 0.133, 0.362 and 0.270, respectively. The internal cross-validation R value is above 0.91 and RMSECV is less than 0.35, which can realize the rapid detection of protein, fat, carbohydrate and moisture in rice.

**Key words:** near-infrared spectroscopy; quantitative model; rapid detection

收稿日期: 2019-03-20

基因项目: 深圳市科学计划项目 (601821K27038); 深圳职业技术学院校级重点项目 (601722K27014)

作者简介: 黄林森 (1996-), 男, 本科, 研究方向: 食品生物技术

通讯作者: 刘冬 (1968-), 男, 博士, 教授, 研究方向: 食品生物技术

我国是亚洲地区最大的大米生产国和消费国<sup>[1]</sup>。大米的主要成分是淀粉, 同时含有丰富的蛋白质, 氨基酸, 还含有少量的粗纤维、脂肪、钙、磷、维生素等营养成分<sup>[2,3]</sup>。其中蛋白质、脂肪、碳水化合物和水分含量是评价大米品质的重要指标<sup>[4]</sup>。国标法测定这

些组分存在过程繁琐、条件不易控制、试剂消耗大、耗时长等不足。近红外(Near Infrared, NIR)光谱分析是近年来迅速发展的一种绿色分析技术,具有样品前处理简单、分析速度快、无需化学试剂、非破坏性和多成分同时分析等优点,为大米营养成分的快速检测提供了新的途径。

目前国内外学者利用近红外分析技术做了一些关于粮食检测方面的研究,比如利用近红外光谱技术结合化学计量学方法对谷物中蛋白质<sup>[5-8]</sup>、淀粉<sup>[9-11]</sup>、脂肪<sup>[12,13]</sup>、水分<sup>[14]</sup>等营养成分含量建立数学预测模型。但主要集中在以下两个方面的研究:一是选取特征波长建立预测模型,具有模型简单、物理意义明确、计算量小的优点,但检测结果不稳定,精度不高;另一方面采用全波段范围进行建模,具有较好的精度,但是计算量大,模型复杂,不利于实际应用。

本试验在大米近红外光谱全波段的基础上,剔除部分干扰波长,分别比较各种化学计量学方法得到最佳的建模方法来建立的大米蛋白质、脂肪、碳水化合物和水分的快速检测定量模型,既能满足较好的预测精度,又能减小计算量。实现对大米中蛋白质、脂肪、碳水化合物和水分的快速同时测定,为大米生产企业实现在线检测和品质控制提供可能,也为食品质量监督部门的日常监管提供理论依据。

## 1 材料和方法

### 1.1 实验材料

#### 1.1.1 实验原料和试剂

大米样品 75 个,粉碎机磨成粉末,备用。

试剂:乙醚、石油醚(沸程 30 °C~60 °C)、乙醇、浓盐酸、浓硫酸、硼酸、氢氧化钠、硫酸铜、硫酸钾、溴甲酚绿、甲基红。以上试剂均采用国药集团化学试剂有限公司生产的分析纯试剂。

#### 1.1.2 仪器设备

近红外光谱仪(Antares II型), Thermo Fisher SCIENTIFIC 公司; FOSS 全自动凯氏定氮仪(Jewelled 8400 型), 福斯分析仪器公司; 电热恒温鼓风干燥箱(DHG-9123A 型), 上海精宏实验设备有限公司; 程控箱式电炉(SXL-1208 型), 上海精宏实验设备有限公司; 离心机(Centrifuge 5810 R), Peppercorn 公司。

### 1.2 实验方法

#### 1.2.1 国标法测定大米中蛋白质、脂肪、水分和碳水化合物的含量

大米中粗蛋白质含量测定参照《GB 5009.5-2016.

食品安全国家标准:食品中蛋白质的测定》全自动凯氏定氮法,粗脂肪含量依据《GB 5009.6-2010.食品国家安全标准:食品中脂肪的测定》酸水解法测定、水分含量应用《GB 5009.3-2016.食品国家安全标准:食品中水分的测定》直接干燥法测得、最后参照《GB 5009.4-2016.食品国家安全标准:食品中灰分的测定》得到粗灰分的含量。每个组分测四个平行数据,通过 Grubs 法在显著性  $\alpha=0.05$  下剔除可疑值,得平均值。再用大米的总量减去所测得蛋白质、脂肪、水分和灰分的含量求得碳水化合物的含量。

#### 1.2.2 样品近红外光谱的采集

近红外光谱仪采集光谱的工作参数设置如表 1。

表 1 近红外光谱仪参数设定

Table 1 Parameter Setting of NIR

参数	设定值
光谱范围	4000.00 cm <sup>-1</sup> ~10000.00 cm <sup>-1</sup>
增益	2x
光谱分辨率	8 cm <sup>-1</sup>
扫描次数	64 次
背景频率	每隔 4 h
样品杯是否旋转	是

大米样品用粉碎机磨成粉末后立即置于样品杯中,使固体粉末的厚度足以不让光透过(不低于 10 mm),于 Antars 傅里叶变换近红外光谱仪的积分球模块中,按照表 1 设定的参数采集样品的光谱信息,扣除背景后得到样品的近红外光谱。

#### 1.2.3 光谱预处理方法

样品的近红外光谱图通过一阶导数或二阶导数处理,放大和分离重叠的信息,再利用 Savitzky-Golay 滤波或 Norris Derivative 滤波的方法降低由于倒数处理后也被放大的噪声信息来提高信噪比。光谱的预处理采用以下六种:(1)一阶导数;(2)二阶导数;(3)一阶导数+Savitzky-Golay 滤波;(4)二阶导数+Savitzky-Golay 滤波;(5)一阶导数+Norris Derivative 滤波;(6)二阶导数+Norris Derivative 滤波。其中 Savitzky-Golay 平滑设定多项式的取值范围为 9、阶数为 3; Norris Derivative 滤波的段长: 5; 段间距: 5。

#### 1.2.4 建模方法

采集的大米近红外光谱通过导数处理和平滑处理,选择相关光谱中相关性高的波长范围,分别采用逐步多元线性回归(Stepwise Multiple Linear Regression, SMLR)、主成分回归(Principal Components Regression, PCR)、偏最小二乘法(Partial Least Squares, PLS)三种化学计量学方法,建立国标法测定的蛋白质、脂肪、水分和碳水化合物含量与近

红外光谱信息相关联的校正模型。根据校正集和验证集的相关系数 R 和均方根来确定最终的建模方法。

## 2 结果与分析

### 2.1 国标法测定结果分析

分别对 75 个大米样品的蛋白质、脂肪、水分和碳水化合物含量测定,共收集 300 个数据,数据处理结果如表 2。总样品分为验证集和校正集,其中校正集样品 63 个,验证集样品 12 个。

表 2 大米蛋白质、脂肪、水分、碳水化合物含量化学测定值的统计结果

Table 2 Statistical results of chemical determination of protein, fat, water and carbohydrate content in rice (g/100 g)

项目	组分	最小值	最大值	平均值	标准差
总样品集		4.81	8.66	7.02	0.84
校正集	蛋白质	4.81	8.66	6.95	0.77
验证集		5.39	8.12	7.08	0.77
总样品集		0.57	2.85	0.96	0.45
校正集	脂肪	0.57	2.85	0.90	0.37
验证集		0.74	2.59	1.20	0.62
总样品集		75.62	80.47	78.79	0.98
校正集	碳水化合物	75.62	80.47	78.83	0.95
验证集		76.71	80.38	78.75	0.95
总样品集		10.66	14.61	12.89	0.74
校正集	水分	10.66	14.61	12.89	0.68
验证集		10.79	14.47	12.93	1.02

由表 2 可知,总样品集大米的营养成分含量最高的成分是碳水化合物,其中最大值为 80.47,最小值为 75.62;其次是水分,最大值为 14.61,最小值为 10.66;蛋白质含量的最大值为 8.66,最小值为 4.81;脂肪的含量最低,在 0.57~2.85 之间。可见,总样品集中蛋白质、脂肪、碳水化合物以及水分的含量分布范围广,说明所取的样品具有一定的代表性。

由上表进一步可知校正集和验证集的蛋白质、脂肪、碳水化合物及水分的含量范围,其中蛋白质的为 4.81~8.66、5.39~8.12;脂肪的变幅为 0.57~2.85、0.74~2.59;碳水化合物的是 75.62~80.47、76.71~80.38;水分的变幅为 10.66~14.61,10.79~14.47。验证集中蛋白质、脂肪、水分和碳水化合物含量范围均在校正集的范围,所选校正集和验证集均合理。

### 2.2 近红外吸收光谱结果分析

75 个大米样品的近红外光谱信息如图 1。

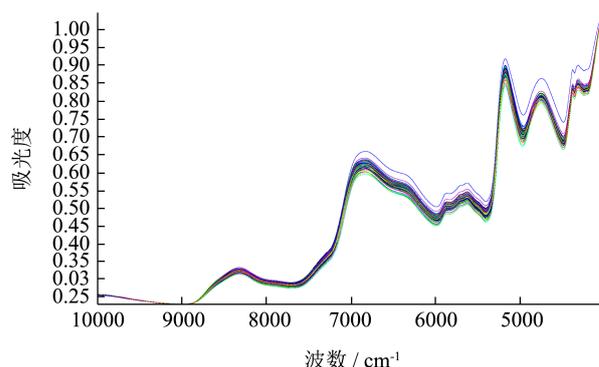


图 1 大米样品的近红外光谱图

Fig.1 Near-infrared spectrum of rice samples

图 1 中,大米样品集光谱的变化趋势基本相同,显示出样本群体的一致性,但未处理的光谱只是在吸光度上整体上下波动,很难区分样品光谱信息的差异而得到特征谱段,需要借助化学计量学的方法来放大和利用光谱信息。

### 2.3 光谱预处理方法、谱区的范围和建模方法对结果的影响

#### 2.3.1 谱区的选择

运用 TQ Analyst 9 定量分析软件,在近红外全光谱波段范围内(4000  $\text{cm}^{-1}$ ~10000  $\text{cm}^{-1}$ ),根据各组分的相关光谱,尽可能选择相关性高的光谱范围作为能代表大米品质的特征波长,运用特征波长进行建模可以减少计算量,同时满足模型有较高的预测精度和稳定性。蛋白质、脂肪、碳水化合物和水分的相关光谱如图 2;谱区范围,如表 3。

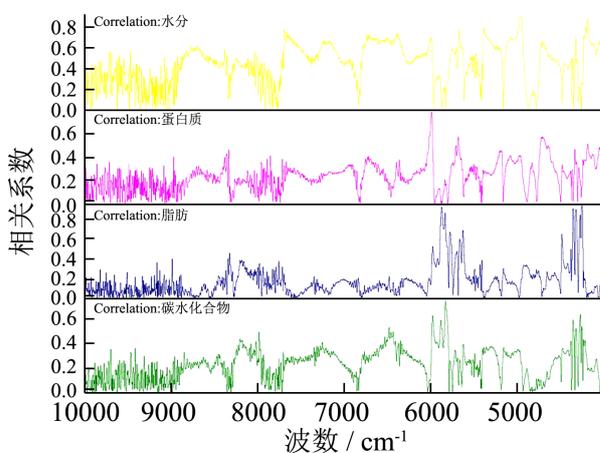


图 2 大米样品各组分的相关光谱

Fig.2 Correlation spectrum of each component in rice sample

表 3 蛋白质、脂肪、碳水化合物和水分的光谱范围

Table 3 The spectral range of protein, fat, carbohydrate and

water	
组分	光谱范围
蛋白质	6077.51 cm <sup>-1</sup> ~4114.88 cm <sup>-1</sup>
	7185.47 cm <sup>-1</sup> ~6275.88 cm <sup>-1</sup>
	8454.40 cm <sup>-1</sup> ~7779.38 cm <sup>-1</sup>
脂肪	7485.40 cm <sup>-1</sup> ~4352.42 cm <sup>-1</sup>
	4484.50 cm <sup>-1</sup> ~4183.99 cm <sup>-1</sup>
	6020.67 cm <sup>-1</sup> ~4940.73 cm <sup>-1</sup>
碳水化合物	7716.38 cm <sup>-1</sup> ~7055.52 cm <sup>-1</sup>
	9015.99 cm <sup>-1</sup> ~7906.71 cm <sup>-1</sup>
	4504.90 cm <sup>-1</sup> ~4024.84 cm <sup>-1</sup>
水分	6048.01 cm <sup>-1</sup> ~4701.53 cm <sup>-1</sup>
	7501.74 cm <sup>-1</sup> ~6700.05 cm <sup>-1</sup>

表 4 光谱预处理方法和建模方法组合结果

Table 4 Results of combination of spectral pretreatment and modeling method

组分	预处理方法	建模方法	校正集		验证集	
			R	RMSEC	R	RMSEP
蛋白质	一阶导数	SMLR	0.8658	0.370	0.8950	0.394
	一阶导数	PCR	0.9825	0.138	0.9890	0.155
	一阶导数	PLS	0.9978	0.0485	0.9356	0.281
	二阶导数	SMLR	0.6201	0.580	0.1067	0.849
	二阶导数	PCR	0.6333	0.572	0.2527	0.848
	二阶导数	PLS	0.4237	0.670	0.1117	0.775
	一阶导数+Savitzky-Golay滤波	SMLR	0.8877	0.341	0.9044	0.360
	一阶导数+Savitzky-Golay滤波	PCR	0.9830	0.136	0.9868	0.161
	一阶导数+Savitzky-Golay滤波	PLS	0.9969	0.0578	0.9706	0.196
	二阶导数+Savitzky-Golay平滑	SMLR	0.8111	0.433	0.9154	0.344
	二阶导数+Savitzky-Golay平滑	PCR	0.9538	0.222	0.9560	0.290
	二阶导数+Savitzky-Golay平滑	PLS	0.9993	0.0278	0.8084	0.475
	一阶导数+Norris Derivative滤波	SMLR	0.9251	0.281	0.9629	0.241
	一阶导数+Norris Derivative滤波	PCR	0.9836	0.133	0.9871	0.155
	一阶导数+Norris Derivative滤波	PLS	0.9830	0.136	0.9942	0.0965
二阶导数+Norris Derivative滤波	SMLR	0.9292	0.273	0.8768	0.373	
二阶导数+Norris Derivative滤波	PCR	0.9827	0.137	0.9897	0.148	
二阶导数+Norris Derivative滤波	PLS	0.9766	0.159	0.9844	0.150	
脂肪	一阶导数	SMLR	0.9487	0.117	0.9797	0.156
	一阶导数	PCR	0.9685	0.0918	0.9788	0.160
	一阶导数	PLS	0.9717	0.0871	0.9819	0.154
	二阶导数	SMLR	0.9152	0.149	0.9846	0.157
	二阶导数	PCR	0.7698	0.235	0.1786	0.666
	二阶导数	PLS	0.9683	0.0921	0.6872	0.533
	一阶导数+Savitzky-Golay滤波	SMLR	0.9484	0.117	0.9840	0.157

转下页

### 2.3.2 光谱预处理方法和建模方法的确定

如表 4 所示, 18 种预处理和建模方法的组合, 建立大米的蛋白质、脂肪、水分和碳水化合物含量的校正模型。根据校正集和验证集的相关系数 R 和均方根的参数来确定光谱的预处理和建模方法, 一般认为相关系数 R 越接近 1 和均方根越接近 0, 模型效果越好<sup>[15,16]</sup>。由表 4 可知, 不同的光谱预处理方法, 同样很大程度的影响模型的预测能力, 通过 SMLR 建立的近红外校正模型与 PCR 和 PLS 建立的校正模型相比大体都不理想。通过 PCR 和 PLS 建立的模型中部分校正集和验证集有的具有比较相关系数高, 但是均方根偏大; 或者相关系数不高, 均方根偏大。通过筛选同时满足校正集和验证的相关系数应尽可能的接近 1 和均方根应尽可能的接近于 0, 得出最佳光谱预处理方法和建模方法, 结果如表 5 所示。

接上页

	一阶导数+Savitzky-Golay滤波	PCR	0.9686	0.0917	0.9794	0.161	
	一阶导数+Savitzky-Golay滤波	PLS	0.9681	0.0924	0.9828	0.146	
	二阶导数+Savitzky-Golay平滑	SMLR	0.9508	0.114	0.9744	0.186	
	二阶导数+Savitzky-Golay平滑	PCR	0.9626	0.0999	0.9770	0.148	
	二阶导数+Savitzky-Golay平滑	PLS	0.9901	0.0517	0.9786	0.143	
	一阶导数+Norris Derivative滤波	SMLR	0.9485	0.117	0.9873	0.136	
	一阶导数+Norris Derivative滤波	PCR	0.9542	0.110	0.9850	0.141	
	一阶导数+Norris Derivative滤波	PLS	0.9527	0.112	0.9837	0.141	
	二阶导数+Norris Derivative滤波	SMLR	0.9635	0.0987	0.9888	0.116	
	二阶导数+Norris Derivative滤波	PCR	0.9582	0.106	0.9860	0.149	
	二阶导数+Norris Derivative滤波	PLS	0.9595	0.104	0.9843	0.144	
	一阶导数	SMLR	0.8494	0.486	0.8307	0.546	
	一阶导数	PCR	0.9557	0.271	0.9606	0.275	
	一阶导数	PLS	0.9734	0.211	0.7178	0.646	
	二阶导数	SMLR	0.7248	0.635	0.5404	0.907	
	二阶导数	PCR	0.7448	0.615	0.6650	0.685	
	二阶导数	PLS	0.9734	0.211	0.7178	0.646	
碳水化合物	一阶导数+Savitzky-Golay滤波	SMLR	0.8526	0.481	0.8464	0.509	
	一阶导数+Savitzky-Golay滤波	PCR	0.9565	0.269	0.9607	0.276	
	一阶导数+Savitzky-Golay滤波	PLS	0.9658	0.239	0.9639	0.261	
	二阶导数+Savitzky-Golay平滑	SMLR	0.7723	0.585	0.6866	0.684	
	二阶导数+Savitzky-Golay平滑	PCR	0.8961	0.409	0.7648	0.588	
	二阶导数+Savitzky-Golay平滑	PLS	0.9754	0.203	0.8869	0.487	
	一阶导数+Norris Derivative滤波	SMLR	0.8685	0.457	0.9366	0.350	
	一阶导数+Norris Derivative滤波	PCR	0.9471	0.296	0.9495	0.298	
	一阶导数+Norris Derivative滤波	PLS	0.9524	0.281	0.9628	0.256	
	二阶导数+Norris Derivative滤波	SMLR	0.8067	0.545	0.7427	0.664	
	二阶导数+Norris Derivative滤波	PCR	0.9399	0.315	0.9702	0.223	
	二阶导数+Norris Derivative滤波	PLS	0.9621	0.251	0.9655	0.323	
		一阶导数	SMLR	0.9267	0.254	0.9648	0.313
		一阶导数	PCR	0.9499	0.212	0.9588	0.299
	一阶导数	PLS	0.9461	0.219	0.9635	0.291	
	二阶导数	SMLR	0.7192	0.470	0.7415	0.730	
	二阶导数	PCR	0.7789	0.425	0.8358	0.656	
	二阶导数	PLS	0.9960	0.0606	0.9056	0.582	
水分	一阶导数+Savitzky-Golay滤波	SMLR	0.9292	0.250	0.9486	0.383	
	一阶导数+Savitzky-Golay滤波	PCR	0.9499	0.212	0.9581	0.301	
	一阶导数+Savitzky-Golay滤波	PLS	0.9465	0.218	0.9626	0.292	
	二阶导数+Savitzky-Golay平滑	SMLR	0.8633	0.342	0.8697	0.514	
	二阶导数+Savitzky-Golay平滑	PCR	0.9527	0.206	0.9522	0.365	
	二阶导数+Savitzky-Golay平滑	PLS	0.9800	0.135	0.9664	0.301	
	一阶导数+Norris Derivative滤波	SMLR	0.9272	0.254	0.9589	0.316	
	一阶导数+Norris Derivative滤波	PCR	0.9594	0.191	0.9561	0.305	

转下页

接上页

一阶导数+Norris Derivative滤波	PLS	0.9463	0.219	0.9578	0.302
二阶导数+Norris Derivative滤波	SMLR	0.9251	0.257	0.9473	0.355
二阶导数+Norris Derivative滤波	PCR	0.9590	0.192	0.9723	0.242
二阶导数+Norris Derivative滤波	PLS	0.9524	0.206	0.9605	0.294

表 5 最佳光谱预处理方法和建模方法组合结果

Table 5 Combination of Optimal Spectral Pretreatment and Modeling Methods

组分	最佳预处理方法	最佳建模方法	校正集		验证集		因子数
			R	RMSEC	R	RMSECV	
蛋白质	一阶导数+Savitzky-Golay滤波	PLS	0.9969	0.0578	0.9706	0.196	9
碳水化合物			0.9658	0.239	0.9639	0.261	8
水分	0.9800		0.135	0.9664	0.301	7	
脂肪	0.9901		0.0517	0.9786	0.143	7	

由表 5 可知, 通过导数和 Savitzky-golay 滤波对近红外光谱进行预处理, 再通过 PLS 进行建模, 结果显示的 R 均大于 0.96, RMSEC 均小于 0.25, 所得的校正模型效果优于其他的预处理和建模方法的组合。因此蛋白质、碳水化合物选择一阶导数与 Savitzky-Golay 滤波的光谱预处理方法; 脂肪、水分选择二阶导数和 Savitzky-Golay 滤波的光谱预处理方法; 再通过 PLS 的建模方法分别建立大米中蛋白质、脂肪、碳水化合物及水分含量测定的定量模型, 具有较高的预测性能。

## 2.4 模型的建立与检验

### 2.4.1 模型的建立

蛋白质、碳水化合物通过一阶导数与 Savitzky-Golay 滤波的光谱预处理; 脂肪、水分通过二阶导数和 Savitzky-Golay 滤波的光谱预处理, 再通过 PLS 的建模方法分别建立大米中蛋白质、脂肪、碳水化合物及水分含量测定的定量模型。分别以蛋白质、脂肪、水分及碳水化合物含量的化学测定值为横坐标, 近红外预测值为纵坐标, 建立的校正模型如图 3~6。

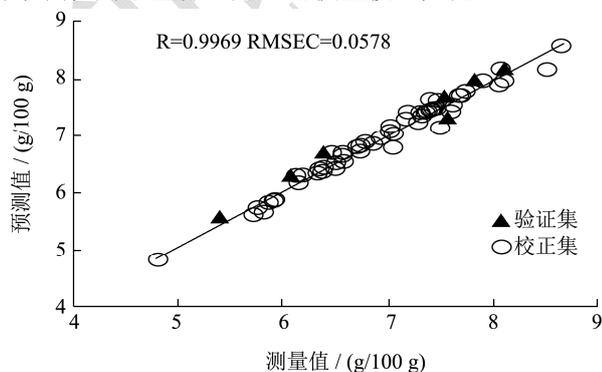


图 3 蛋白质的定量模型  
Fig.3 Quantitative model of protein

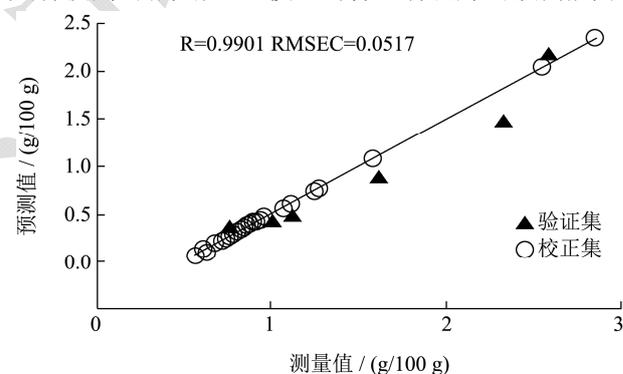


图 4 脂肪的定量模型

Fig.4 Quantitative model of fat

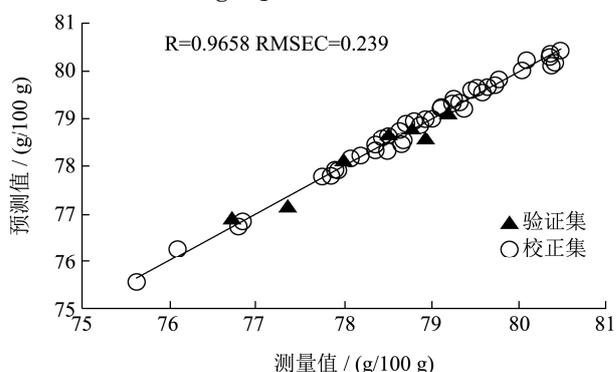


图 5 碳水化合物的定量模型

Fig.5 Quantitative model of carbohydrate

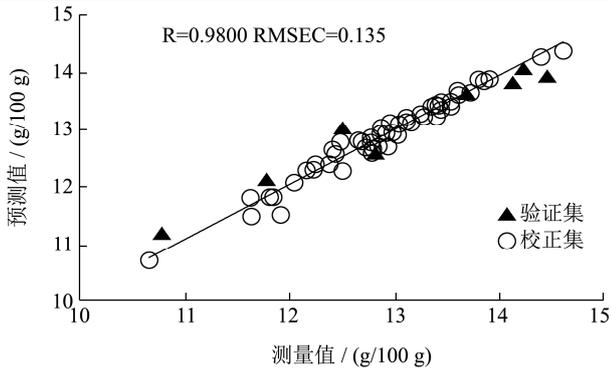


图6 水分的定量模型

Fig.6 Quantitative model of water

2.4.2 模型的优化与验证

根据光谱的分布差异计算马氏距离的方法鉴别异常光谱, 鉴别结果如图7所示。对图中马氏距离最大的异常光谱的样品进行重新扫描光谱再计算其马氏距离, 结果仍如图7所示, 故舍去该光谱。

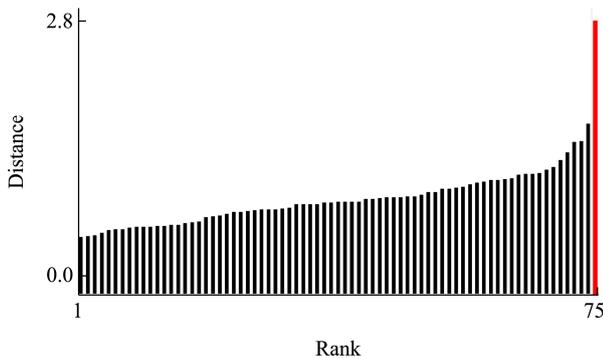


图7 近红外光谱图的马氏距离

Fig.7 The mahalanobis distance of the near infrared spectrogram

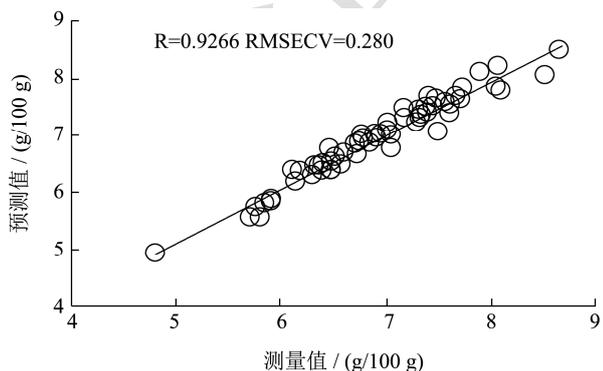


图8 蛋白质定量模型内部交叉验证结果

Fig.8 Results of cross-validation of protein quantitative model

对建立的模型进行内部交叉验证, 由此判断模型的预测能力。内部交叉验证结果分别以蛋白质、脂肪、水分及碳水化合物含量的化学测定值为横坐标, 近红外预测值为纵坐标作图, 并计算相关系数(R)与预测均方根误差(RMSECV)如图8~11。

内部交叉验证的结果表明, 预测大米中蛋白质、

脂肪、碳水化合物、水分的含量的近红外光谱模型RMSEC 分别为 0.9266、0.9333、0.9198、0.9175, RMSECV 分别为 0.280、0.133、0.362、0.270, 相关系数均达到 0.91 以上, 且均方根均低于 0.37, 结果表明模型具有很好的预测性能。

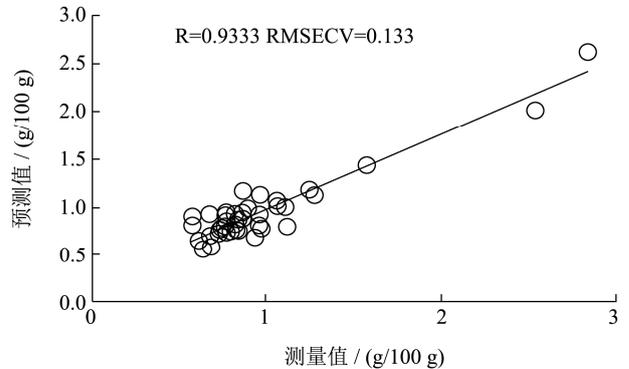


图9 脂肪的定量模型内部交叉验证结果

Fig.9 Results of cross-validation of fat quantitative model

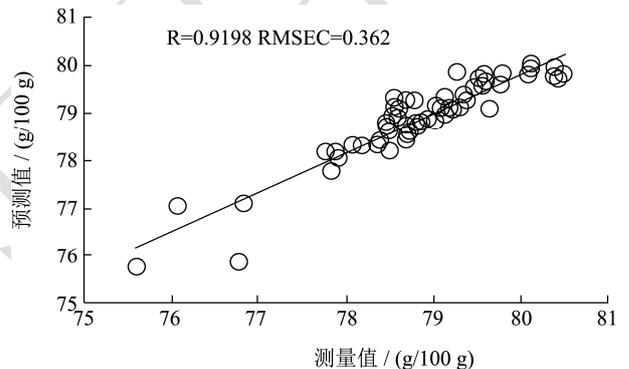


图10 碳水化合物的定量模型内部交叉验证结果

Fig.10 Results of cross-validation of carbohydrate quantitative model

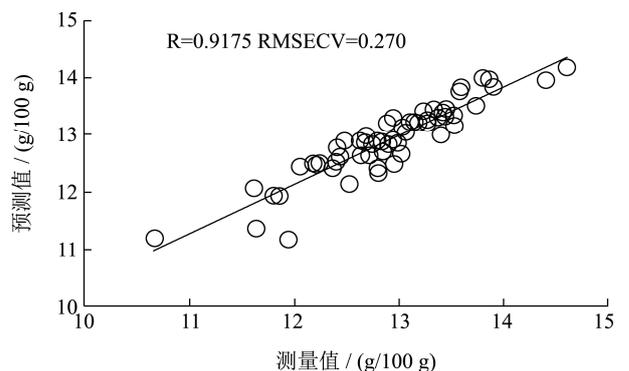


图11 水分的定量模型内部交叉验证结果

Fig.11 Results of cross-validation of water quantitative model

3 结论

3.1 本实验应用近红外光谱(NIRS)分析技术, 通过比较不同的光谱预处理和建模方法确定最佳建模途径为: 大米中的蛋白质、碳水化合物通过一阶导数与Savitzky-Golay滤波对光谱进行处理; 脂肪、水分通过

二阶导数和 Savitzky-Golay 滤波对光谱进行处理;再通过 PLS 回归与国标法的测定值相关联建立定量模型。内部交叉验证表明预测值和真实值之间的相关系数(R)均达到 0.91 以上,均方根(RMSECV)均低于 0.37。

3.2 因此本实验建立的大米近红外预测模型的预测精度与现有同类模型相比有所提高;由于干扰波长的剔除大大减少了变量数,在保证预测精度的同时降低了计算量;利用 Workflow 软件调用建立的模型可实现四种组分的快速检测。因此本模型适用于大米蛋白质、脂肪、碳水化合物和水分的近红外快速检测,以节省大量的人力、物力,减少化学试剂的消耗和环境污染。

### 参考文献

- [1] 夏凡,董月,朱蕾等.大米理化性质与其食用品质相关性研究[J].粮食深加工及食品,2018,43(5):100-106  
XIA Fan, DONG Yue, ZHU Lei, et al. Correlation between the physical and chemical properties of rice and its edible quality [J]. Grain Deep Processing and Food, 2018, 43(5): 100-106
- [2] 杨福臣,孙兆远,孙芝杨.乳粉中蛋白质和脂肪近红外测定模型研究[J].食品研究与开发,2017,38(2):169-173  
YANG Fu-cheng, SUN Zhao-yuan, SUN Zhi-yang. Study on the determination model of protein and fat in milk powder by NIR [J]. Food Research and Development, 2017, 38(2): 169-173
- [3] Zhang Z H, Sun D W. Effects of cooling methods on the coling efficiency and quality of cooked rice [J]. Journal of Food Engineering, 2006, 77: 269-274
- [4] 黄丽,柏芸,赵思明,等.稻米质量对食品安全的影响[J].中国粮油学报,2013,28(4):113-117  
HUANG Li, BAI Yun, ZHAO Si-ming, et al. Effect of rice quality on the food safety [J]. Journal of the Chinese Cereals and Oils Association, 2013, 28(4): 113-117
- [5] 周子立,冯雷.光谱技术结合数据处理技术测定大米中淀粉和蛋白含量[J].软件工程师,2014,17(7):58-60  
ZHOU Zi-li, FENG Lei. Determination of starch contents and protein contents in rice using spectroscopy combined with data processing technique [J]. Software Engineer, 2014, 17(7): 58-60
- [6] 刘明博,唐延林,李晓利,等.大米蛋白质含量近红外光谱检测模型研究[J].中国农学通报,2013,29(12):212-216  
LIU Ming-bo, TANG Yan-lin, LI Xiao-li, et al. Study on NIR spectral detection model of rice protein content [J]. Chinese Agricultural Science Bulletin, 2013, 29(12): 212-216
- [7] 郭咏梅,李华慧,李少明,等.糙米蛋白质含量与矿物质元素含量的相关分析及 NIRS 模型的建立[J].植物遗传资源学报,2013,1(14):174-180  
GUO Yong-mei, LI Hua-hui, LI Shao-ming, et al. Correlation analysis of protein content and mineral content in brown rice and establishment of the math model for the NIRS analysis [J]. Journal of Plant Genetic Resources, 2013, 1(14): 174-180
- [8] 俞法明,陆艳婷,严文潮,等.用近红外光谱技术快速测定籼稻品种的蛋白质含量[J].中国粮油学报,2009,24(5):134-138  
YU Fa-ming, LU Yan-ting, YAN Wen-chao, et al. Rapid determination of protein content in long grain rice by near infrared spectroscopy [J]. Journal of the Chinese Cereals and Oils Association, 2009, 24(5): 134-138
- [9] 林葵,陈建红,黄岛平.近红外快速检测广西大米中淀粉及直链淀粉成分的研究[J].大众科技,2011,138(2):95-96  
LIN Kui, CHEN Jian-hong, HUANG Dao-ping. Study on NIR spectral rapid detection of starch and amylose in guangxi rice [J]. Popular Science & Technology, 2011, 138(2): 95-96
- [10] 徐泽林,金华丽.近红外光谱法测定大米中的淀粉含量[J].农产品加工学刊,2011,238(3):74-76  
XU Ze-lin, JIN Hua-li. Determination of starch content in rice by near-infrared spectroscopy [J]. Academic Periodical of Farm Products Processing, 2011, 238(3): 74-76
- [11] 陆艳婷,金庆生,叶胜海,等.应用近红外光谱分析技术快速测定粳稻品种的直链淀粉含量[J].中国粮油学报,2007,22(3):149-153  
LU Yan-ting, JIN Qing-sheng, YE Sheng-hai, et al. Application of near infrared spectroscopy for rapid analysis of amylose content in short-grain rice [J]. Journal of the Chinese Cereals and Oils Association, 2007, 22(3): 149-153
- [12] 王传梁,陈坤杰.基于近红外漫反射技术的大米脂肪含量的研究[J].粮油加工,2007,2:62-64  
WANG Chuan-liang, CHEN Kun-jie. Study on fat content in rice by near-infrared reflectance spectroscopy [J]. Grain Processing, 2007, 2: 62-64
- [13] 王海莲,万向元,胡培松,等.稻米脂肪含量近红外光谱分析技术研究[J].中国农业科学,2005,38(8):1540-1546  
WANG Hai-lian, WAN Xiang-yuan, HU Pei-song, et al. Quantitative analysis of fat content in rice by near-infrared spectroscopy technique [J]. Scientia Agriculture Science, 2005, 38(8): 1540-1546