

# 奇亚籽水分、脂肪、蛋白质和灰分的近红外光谱快速检测模型优化

彭彬倩, 庞月红, 沈晓芳

(江南大学食品学院, 江苏无锡214122)

**摘要:**水分、脂肪、蛋白质、灰分是重要的奇亚籽品质指标, 目前主要依赖于化学法测定, 过程烦琐, 耗时费力, 且不能多指标同时监测。以103份不同产地的奇亚籽为建模样本, 通过19种光谱预处理方法和最佳谱区范围的筛选分别建立了奇亚籽中水分、脂肪、蛋白质、灰分的偏最小二乘模型。结果表明: 对于水分采用 Savitzky - Golay 滤波平滑(SG)对光谱进行预处理, 脂肪采用一阶导数(1st)和多元散射校正(MSC), 蛋白质采用1st、标准正态变化(SNV)和SG, 对于灰分采用1st、SNV和Norris微分平滑(ND)组合光谱进行预处理, 针对各参数最佳预处理光谱采用人工法进行谱区筛选建立模型, 得到水分、脂肪、蛋白质和灰分的验证集相关系数分别为0.993、0.972、0.925和0.923。结果显示, 利用近红外光谱可以实现对奇亚籽的水分、脂肪、蛋白质以及灰分的同时快速无损检测, 在大规模奇亚籽原料的分选中提高检测效率。

**关键词:**奇亚籽; 近红外光谱; 偏最小二乘法; 定量模型

中图分类号: TS222+.1; O657.3 文献标识码: A 文章编号: 1003-7969(2020)07-0137-08

## Model optimization for determination of moisture, fat, protein and ash in chia seed by near infrared spectroscopy

PENG Binqian, PANG Yuehong, SHEN Xiaofang

(School of Food Science and Technology, Jiangnan University, Wuxi 214122, Jiangsu, China)

**Abstract:** Determination of moisture, fat, protein and ash, four important quality indexes of chia seed, are based on chemical method, which is tedious, time-consuming, laborious, and can't monitor multiple indexes simultaneously. 103 chia seed samples from various producing areas were collected. A total of 19 kinds of spectral pretreatment and selection of optimal spectral range, the partial least-squares models of moisture, fat, protein and ash in chia seed were established. The results showed that the spectra were processed by Savitzky - Golay filter (SG) for moisture; the combination of first derivative (1st) and multiplicative signal correction (MSC) was found to be the best preprocessing method for fat model; protein was pretreated with 1st, standard normal variate (SNV) and SG; for the detection of ash, 1st, SNV and Norris derivative filter (ND) spectral pretreatment method was better. Furthermore, new models were developed with the selected optimal pretreatments by artificial process. For moisture, fat, protein and ash, the correlation coefficients of prediction set were 0.993, 0.972, 0.925 and 0.923, respectively. Accordingly, near infrared spectroscopy could achieve the simultaneous, rapid and nondestructive detection of moisture, fat, protein and ash in chia seed to improve the detection efficiency for massive samples.

**Key words:** chia seed; near infrared spectroscopy; partial least-squares; quantitative model

收稿日期: 2019-11-30; 修回日期: 2020-03-27

基金项目: 国家重点研发计划(2018YFC1602300)

作者简介: 彭彬倩(1994), 女, 在读硕士, 研究方向为食品检测(E-mail) pbq12@163.com。

通信作者: 沈晓芳, 教授(E-mail) xfshen@jiangnan.edu.cn。

奇亚籽(*Salvia hispanica* L.)原产于墨西哥南部等北美地区, 是美国食品和药物管理局(FDA)认证的安全食品。2014年我国国家卫生和计划生育委员会将其列为新食品原料, 正式进入我国市场。奇

亚籽营养价值高,不仅可以直接食用,还可作为面包<sup>[1]</sup>、蛋黄酱<sup>[2]</sup>、酸奶<sup>[3]</sup>等产品的辅料。奇亚籽被认为是蛋白质和植物油的良好来源,特别是奇亚籽油富含不饱和脂肪酸,其中 $\alpha$ -亚麻酸含量达60%以上,是 $n-3$ 脂肪酸的天然来源<sup>[4]</sup>,其应用前景和经济效益日益受到人们的重视。

近年来,国内对奇亚籽营养价值的深入研究<sup>[5-6]</sup>,大大促进了奇亚籽加工业的发展,并对奇亚籽原料快速检测分选提出了新的要求。国标法测定样品各组分过程烦琐、耗时耗力、费用昂贵。近红外光谱法作为一种无损检测技术,具有多指标同时检测、快速无损、成本低等优点<sup>[7]</sup>,在油脂<sup>[8]</sup>、谷物<sup>[9-11]</sup>、乳制品<sup>[12]</sup>等行业应用广泛,且已应用于多种谷物水分、蛋白质、脂肪和灰分等品质指标的检测<sup>[10, 13-14]</sup>,但目前还鲜见基于近红外的奇亚籽水分等多品质参数快速无损检测的相关研究报道。

本研究以不同产地的奇亚籽为研究对象,测定奇亚籽中水分、蛋白质、脂肪和灰分,同时采集同份样品的近红外漫反射光谱,为搭建品质模型提供基础数据。在近红外光谱全波段的基础上筛选最佳光谱预处理方式,通过人为逐一排除法剔除干扰波段建立偏最小二乘定量模型。本研究可为奇亚籽行业实现在线品控分选提供研究依据。

## 1 材料与方法

### 1.1 实验材料

奇亚籽样品,收集自墨西哥、阿根廷、澳大利亚、巴拉圭、玻利维亚、秘鲁、厄瓜多尔和尼加拉瓜等国,共计103份,样品状态为籽粒。

Antaris II 近红外分析仪,购自 Thermo Fisher 科技(中国)有限公司。光谱采集软件为 RESULT - Integration,建模软件为 TQ Analyst,均为 Thermo Fisher 科技(中国)有限公司开发。

### 1.2 实验方法

#### 1.2.1 理化指标的测定

参照 GB 5009.3—2016《食品安全国家标准 食品中水分的测定》,采用105℃直接干燥法测定奇亚籽中水分含量。参照 GB 5009.6—2016《食品安全国家标准 食品中脂肪的测定》,采用索氏抽提法测定奇亚籽中脂肪含量。参照 GB 5009.5—2016《食品安全国家标准 食品中蛋白质的测定》中半自动凯氏定氮仪法测定奇亚籽中蛋白质含量,折算系数取6.25。参照 GB 5009.4—2016《食品安全国家标准 食品中灰分的测定》,采用550℃马弗炉灼烧法测定奇亚籽中灰分含量。每个样品测3次取平均值作为实测值。

#### 1.2.2 光谱的采集

使用具有积分球的 Antaris II 近红外分析仪(配有 InGaAs 检测器)以漫反射模式采集每个样品的光谱。每份样品籽粒称取25g,置于圆形标准样品杯(直径5mm,高2mm)中,设定样品杯全程旋转扫描模式。整个实验在(25±1)℃的环境温度和38%左右的相对湿度下进行。所有光谱均采用背景空气光谱作为空白对照,每个光谱是32次扫描的平均值,光谱扫描范围为10 000~4 000  $\text{cm}^{-1}$ ,分辨率为8  $\text{cm}^{-1}$ ,增益为8×,原始数据集以3.857  $\text{cm}^{-1}$ 间隔测量,得到1 557个变量。将样品杯中样品倒回样品袋,重新称取25g样品重复扫描近红外光谱,3次采集得到平均光谱,以减少样品分布不均匀带来的误差。

#### 1.2.3 光谱预处理及建模

样品集划分:根据光谱数据计算马氏距离剔除异常点之后的样本随机划分校正集与验证集,各81份和22份,保证两集的样本化学值均匀分布。

光谱预处理与特征波段筛选:为了提高相关有效光谱信息,削减无效干扰光谱,达到放大和分离重叠光谱信息的效果,采用多元散射校正(MSC)、标准正态变化(SNV)、一阶导数(1st)、二阶导数(2nd)、Savitzky - Golay 滤波平滑(SG)和 Norris 微分平滑(ND)单一或组合方式,共计19种光谱预处理方法。其中,Savitzky - Golay 滤波平滑设定多项式的取值范围为7、阶数为3,Norris 微分平滑在段长为5、间距为5时进行。将全波段划分不同区间,采用手动法<sup>[15]</sup>筛选特征波段。

模型评判与验证:采用偏最小二乘(PLS)回归法建立奇亚籽水分、脂肪、蛋白质和灰分的校正模型。根据交叉验证均方根误差(RMSECV)确定主因子数和校正集相关系数( $r_c$ )作为直接衡量模型优劣的主要指标,并使用相对分析误差(RPD)对各组分模型进一步评估。若 $RPD \geq 2.0$ ,说明模型建立优良,可用于日常实际定量检测;若 $2.0 > RPD \geq 1.4$ ,说明模型较好,但用于实际应用需进一步改良;若 $RPD < 1.4$ ,说明模型不能用于定量检测<sup>[16-18]</sup>,进一步,相对标准偏差(RSD)应当小于10%,最后利用验证集样品(不参与建模)的近红外光谱图,通过比较模型预测值与化学值进行验证。

## 2 结果与分析

### 2.1 理化指标

Williams 等<sup>[19]</sup>认为样品化学值平均分布有利于建立稳健理想的校准模型,而不是一个高斯分布趋势,非高斯分布被认为是校准集的一个良好特征,

从而所建模型在未来分析偏向平均值。奇亚籽水分、脂肪、蛋白质和灰分在初始样本集的分布如图1所示。由图1可知:水分含量范围在2.45%~8.65%之间,其中含量在4.50%~7.50%之间的样品共有84份,占总样本量的81.6%;脂肪含量范围在26.4%~36.7%之间,其中含量在31.5%~36.5%之间的样品共有93份,占总样本量的90.3%;蛋白质含量范围在17.2%~27.6%之间,其中含量在18.0%~

26.0%之间的样品共有95份,占总样本量的92.2%;灰分含量范围在3.5%~5.6%之间,其中含量在4.1%~5.3%之间的样品共有91份,占总样本量的88.3%。本实验样本的各组份范围基本上覆盖了奇亚籽中可能出现的含量,数据离散度及波动性较大,虽然样品的分布不均匀,但样品在最大值和最小值之间连续存在。

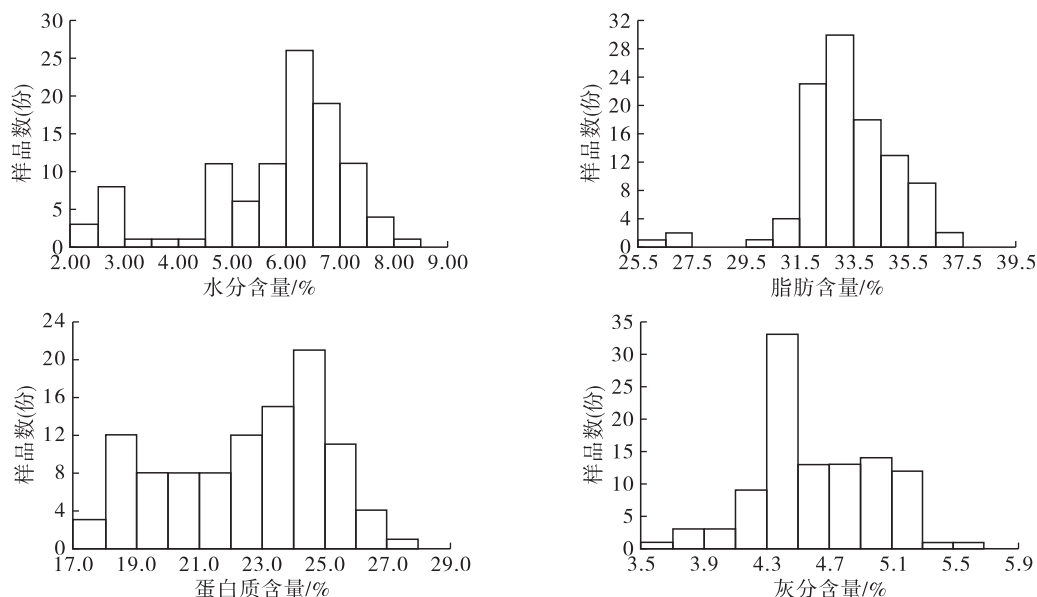


图1 奇亚籽水分、脂肪、蛋白质和灰分在初始样本集的分布

## 2.2 校正集与验证集的划分

对近红外光谱定量分析而言,其数据异常可分为光谱异常和实测值异常<sup>[8]</sup>,本实验借助 TQ Analyst 软件自带的马氏距离判定没有出现异常值。奇亚籽样本校正集和验证集参数见表1。由表1可知,103份奇亚籽的测量参数范围较广(样本生长地点、基因型和环境影响的变异性),水分的校正集和验证集化学值变幅范围分别为2.45%~8.65%、

2.56%~7.95%,脂肪的变幅分别为26.4%~36.7%、26.5%~36.5%,蛋白质的变幅分别为17.2%~27.6%、17.7%~26.5%,灰分的变幅分别为3.5%~5.6%、3.7%~5.3%。可见验证集化学值变幅范围包含在校正集内,校正集所建模型范围可以用于验证集样品,且校正集与验证集的平均值和标准偏差相近,可见校正集与验证集内的样本分布相似,选择合理。

表1 奇亚籽样本校正集和验证集参数

指标	校正集			验证集		
	化学值	平均值	标准偏差	化学值	平均值	标准偏差
水分	2.45~8.65	5.76	1.45	2.56~7.95	5.96	1.42
脂肪	26.4~36.7	33.2	1.76	26.5~36.5	33.1	2.05
蛋白质	17.2~27.6	22.6	2.55	17.7~26.5	22.2	2.68
灰分	3.5~5.6	4.6	0.39	3.7~5.3	4.6	0.39

## 2.3 光谱分析

图2为奇亚籽样本原始近红外光谱图,通过 Antaris II 近红外分析仪采集近红外光谱范围为10 000~4 000  $\text{cm}^{-1}$ ,12 500~10 000  $\text{cm}^{-1}$ 之间的区域光谱信息较少,因此此区域没有被用于开发校准模型。由图2可知,每条光谱趋势相似,难以用肉眼

评判特定波段峰值与指标含量的关系。在6 897  $\text{cm}^{-1}$ 和5 181  $\text{cm}^{-1}$ 处的谱带与—OH基团的第一倍频和组合频相对应,主要代表了水分含量;与脂类相关的谱带,即—CH<sub>2</sub>和—CH<sub>3</sub>官能团,在8 264  $\text{cm}^{-1}$ 是拉伸振动的第二倍频,5 780  $\text{cm}^{-1}$ 和5 681  $\text{cm}^{-1}$ 是拉伸振动的第一倍频,拉伸和弯曲振动的组合频出

现在  $4\ 329\ \text{cm}^{-1}$  和  $4\ 255\ \text{cm}^{-1}$  附近;  $4\ 859\ \text{cm}^{-1}$  和  $4\ 616\ \text{cm}^{-1}$  处吸收峰带与蛋白质含量有关<sup>[19-20]</sup>。剔除干扰光谱信息, 加强有效信息均有利于建立合适而精确的模型。通过 TQ Analyst 软件将光谱进行预处理及波段的剔除, 从而利用偏最小二乘法建立各组分的定量模型。

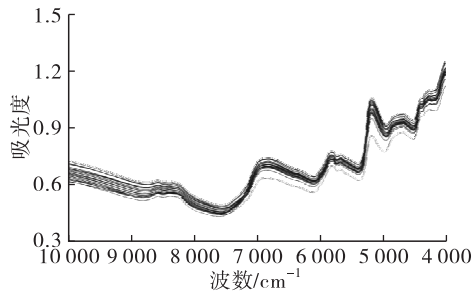


图2 奇亚籽样本近红外光谱图

#### 2.4 光谱预处理

从原始近红外光谱仪获取的 103 份奇亚籽光谱数据除了样品信息外, 还含有背景信息和噪声。为了获得优化的 PLS 校准模型, 建模前需要对光谱数据进行预处理。在没有特征光谱的情况下, 光谱预处理是建立 PLS 模型的变量之一。导数可重新求解重叠峰, 从而提高分辨率和灵敏度。随着光谱预处理的选择不同, 一阶导数主要解决基线偏移, 二阶导数则解决基线漂移<sup>[21]</sup>, 平滑可用于降低随机噪声, 从而提高光谱的信噪比, MSC 和 SNV 处理可以补偿由于颗粒大小等因素引起的光谱散射效应<sup>[22]</sup>。

水分、脂肪、蛋白质和灰分不同光谱预处理方法的 PLS 回归模型参数分别见表 2 ~ 表 5。由表 2 ~

表 5 可知, 优化后的模型具有较高的  $r$  值和较低的  $RMSECV$  和  $RMSEP$  (验证集预测均方根误差) 值, 可以较好地预测奇亚籽中 4 种分析物的含量。当不同光谱预处理下  $RMSECV$  接近的情况下, 如果一味追求校正集相关系数的最佳, 可能存在校正过度现象, 不利于验证, 因此在这种情况下选择低  $RMSEP$  为最优模型。优化结果如下:

水分: SG 处理下模型最佳。此时, 校正集相关系数为 0.976,  $RMSEP$  为 0.19%,  $RMSECV$  为 0.32%。校正集  $RPD$  为 4.6,  $RSD$  为 5.5%。

脂肪: 利用 1st 和 MSC 得到了最优 PLS 模型。此时,  $RMSEP$  为 0.54%, 验证集相关系数为 0.970,  $RMSECV$  为 0.95%, 校正集相关系数为 0.843,  $RPD$  为 1.9,  $RSD$  为 2.8%。

蛋白质: 1st、SNV 和 SG 组合处理结果最好。此时,  $RMSECV$  为 1.33%, 校正集相关系数为 0.850,  $RMSEP$  为 1.54%。

灰分: 1st、SNV 和 ND 组合处理结果最好。此时,  $RMSEP$  为 0.17%, 验证集相关系数为 0.899,  $RMSECV$  为 0.25%, 校正集相关系数为 0.798,  $RPD$  为 1.6,  $RSD$  为 5.5%。

主因子数变大可以防止欠配合, 然而, 主因子数过大往往会单方面使校正集搭建的模型过拟合, 不利于预测集的结果, 因此选择合适的主因子数很重要<sup>[23-24]</sup>。通过数据软件自动给出水分、脂肪、蛋白质和灰分含量模型主因子数分别为 10、6、5、9。

表2 基于全波段的近红外光谱预处理下水分指标建模结果

光谱预处理	校正集				验证集				主因子数
	$r$	$RMSECV/\%$	$RPD$	$RSD/\%$	$r$	$RMSEP/\%$	$RPD$	$RSD/\%$	
NONE	0.975	0.32	4.5	5.6	0.983	0.26	5.4	4.4	9
MSC	0.986	0.24	5.9	4.2	0.981	0.28	5.1	4.7	13
SNV	0.985	0.25	5.8	4.3	0.959	0.41	3.5	6.8	12
1st	0.980	0.29	5.0	5.0	0.984	0.26	5.5	4.3	5
2nd	0.867	0.74	2.0	12.7	0.893	0.72	2.0	12.0	5
SG	0.976	0.32	4.6	5.5	0.992	0.19	7.4	3.2	10
MSC + 1st	0.979	0.29	4.9	5.1	0.976	0.31	4.6	5.2	4
MSC + 2nd	0.877	0.70	2.1	12.1	0.897	0.65	2.2	10.8	3
MSC + SG	0.986	0.24	5.9	4.2	0.979	0.29	4.9	4.9	13
MSC + 1st + SG	0.980	0.29	5.1	4.9	0.980	0.28	5.0	4.7	5
MSC + 1st + ND	0.983	0.26	5.5	4.5	0.990	0.20	7.2	3.3	8
MSC + 2nd + SG	0.960	0.41	3.6	7.0	0.954	0.42	3.4	7.0	3
MSC + 2nd + ND	0.982	0.28	5.3	4.8	0.987	0.23	6.3	3.8	8
SNV + 1st	0.979	0.30	4.9	5.1	0.977	0.30	4.7	5.0	4
SNV + 2nd	0.877	0.71	2.1	12.2	0.885	0.68	2.1	11.3	3

续表 2

光谱预处理	校正集				验证集				主因子数
	<i>r</i>	RMSECV/%	RPD	RSD/%	<i>r</i>	RMSEP/%	RPD	RSD/%	
SNV + SG	0.985	0.25	5.7	4.4	0.959	0.41	3.5	6.8	12
SNV + 1st + SG	0.980	0.29	5.1	4.9	0.982	0.27	5.2	4.6	5
SNV + 1st + ND	0.983	0.26	5.5	4.6	0.983	0.26	5.5	4.3	9
SNV + 2nd + SG	0.961	0.40	3.6	6.9	0.950	0.44	3.2	7.3	3
SNV + 2nd + ND	0.982	0.28	5.3	4.7	0.988	0.22	6.5	3.7	8

注: NONE 表示无光谱预处理。下同。

表 3 基于全波段的不同光谱预处理下脂肪指标建模结果

光谱预处理	校正集				验证集				主因子数
	<i>r</i>	RMSECV/%	RPD	RSD/%	<i>r</i>	RMSEP/%	RPD	RSD/%	
NONE	0.878	0.84	2.2	2.5	0.955	0.64	3.1	1.9	6
MSC	0.883	0.82	2.2	2.5	0.935	0.73	2.7	2.2	7
SNV	0.884	0.82	2.2	2.5	0.945	0.70	2.9	2.1	7
1st	0.815	1.02	1.8	3.1	0.969	0.55	3.6	1.7	6
2nd	0.576	1.43	1.3	4.3	0.602	1.61	1.2	4.9	3
SG	0.872	0.85	2.1	2.6	0.955	0.64	3.1	1.9	6
MSC + 1st	0.843	0.95	1.9	2.8	0.970	0.54	3.7	1.6	6
MSC + 2nd	0.601	1.40	1.3	4.2	0.590	1.62	1.2	4.9	1
MSC + SG	0.876	0.84	2.1	2.5	0.939	0.72	2.8	2.2	6
MSC + 1st + SG	0.857	0.91	2.0	2.7	0.968	0.55	3.7	1.7	6
MSC + 1st + ND	0.902	0.75	2.4	2.3	0.948	0.65	3.1	2.0	6
MSC + 2nd + SG	0.684	1.27	1.4	3.8	0.624	1.57	1.3	4.8	1
MSC + 2nd + ND	0.895	0.78	2.3	2.4	0.912	0.83	2.4	2.5	7
SNV + 1st	0.820	1.01	1.8	3.0	0.967	0.57	3.5	1.7	6
SNV + 2nd	0.600	1.40	1.3	4.2	0.589	1.62	1.2	4.9	1
SNV + SG	0.873	0.86	2.1	2.6	0.945	0.70	2.9	2.1	7
SNV + 1st + SG	0.836	0.97	1.9	2.9	0.969	0.55	3.7	1.7	6
SNV + 1st + ND	0.904	0.75	2.4	2.3	0.953	0.62	3.2	1.9	7
SNV + 2nd + SG	0.682	1.28	1.4	3.9	0.626	1.57	1.3	4.8	1
SNV + 2nd + ND	0.899	0.77	2.3	2.3	0.935	0.72	2.8	2.2	8

表 4 基于全波段的不同光谱预处理下蛋白质指标建模结果

光谱预处理	校正集				验证集				主因子数
	<i>r</i>	RMSECV/%	RPD	RSD/%	<i>r</i>	RMSEP/%	RPD	RSD/%	
NONE	0.773	1.63	1.6	7.2	0.751	1.91	1.4	8.6	6
MSC	0.820	1.46	1.8	6.5	0.741	1.88	1.4	8.5	6
SNV	0.806	1.51	1.7	6.7	0.738	1.91	1.4	8.6	6
1st	0.850	1.33	2.0	5.9	0.781	1.72	1.6	7.7	6
2nd	0.690	1.83	1.4	8.1	0.624	2.16	1.3	9.7	2
SG	0.773	1.63	1.6	7.2	0.751	1.91	1.4	8.6	6
MSC + 1st	0.851	1.33	2.0	5.9	0.808	1.63	1.7	7.3	5
MSC + 2nd	0.697	1.81	1.4	8.0	0.552	2.32	1.2	10.5	2
MSC + SG	0.820	1.46	1.8	6.5	0.740	1.88	1.4	8.5	6
MSC + 1st + SG	0.855	1.31	2.0	5.8	0.820	1.58	1.7	7.1	5
MSC + 1st + ND	0.852	1.33	2.0	5.9	0.765	1.83	1.5	8.2	7
MSC + 2nd + SG	0.768	1.63	1.6	7.2	0.661	2.13	1.3	9.6	4
MSC + 2nd + ND	0.843	1.36	1.9	6.0	0.824	1.59	1.7	7.2	6
SNV + 1st	0.844	1.35	1.9	6.0	0.821	1.58	1.7	7.1	5

续表 4

光谱预处理	校正集				验证集				主因子数
	<i>r</i>	<i>RMSECV</i> /%	<i>RPD</i>	<i>RSD</i> /%	<i>r</i>	<i>RMSEP</i> /%	<i>RPD</i>	<i>RSD</i> /%	
SNV + 2nd	0.694	1.82	1.4	8.1	0.554	2.32	1.2	10.5	2
SNV + SG	0.806	1.51	1.7	6.7	0.738	1.91	1.4	8.6	6
SNV + 1st + SG	0.850	1.33	2.0	5.9	0.831	1.54	1.8	6.9	5
SNV + 1st + ND	0.843	1.36	1.9	6.0	0.763	1.81	1.5	8.2	6
SNV + 2nd + SG	0.760	1.65	1.6	7.3	0.672	2.10	1.3	9.5	4
SNV + 2nd + ND	0.840	1.38	1.9	6.1	0.825	1.59	1.7	7.2	6

表 5 基于全波段的不同光谱预处理下灰分指标建模结果

光谱预处理	校正集				验证集				主因子数
	<i>r</i>	<i>RMSECV</i> /%	<i>RPD</i>	<i>RSD</i> /%	<i>r</i>	<i>RMSEP</i> /%	<i>RPD</i>	<i>RSD</i> /%	
NONE	0.764	0.28	1.5	6.0	0.721	0.27	1.5	5.8	7
MSC	0.795	0.26	1.5	5.7	0.762	0.25	1.6	5.4	9
SNV	0.778	0.25	1.6	5.5	0.684	0.30	1.4	6.4	5
1st	0.651	0.30	1.3	6.5	0.659	0.29	1.4	6.3	3
2nd	0.060	0.43	0.9	9.4	0.216	0.39	1.0	8.5	1
SG	0.764	0.28	1.5	6.0	0.721	0.27	1.5	5.8	7
MSC + 1st	0.709	0.28	1.5	6.0	0.697	0.27	1.5	6.0	5
MSC + 2nd	0.102	0.42	1.0	9.1	0.265	0.40	1.0	8.6	1
MSC + SG	0.795	0.26	1.5	5.7	0.761	0.25	1.6	5.4	9
MSC + 1st + SG	0.732	0.27	1.5	5.8	0.698	0.27	1.5	6.0	5
MSC + 1st + ND	0.800	0.24	1.7	5.2	0.880	0.18	2.2	4.0	9
MSC + 2nd + SG	0.581	0.32	1.3	6.9	0.403	0.35	1.1	7.7	3
MSC + 2nd + ND	0.788	0.24	1.7	5.3	0.841	0.21	1.9	4.6	8
SNV + 1st	0.691	0.28	1.4	6.2	0.687	0.28	1.4	6.0	5
SNV + 2nd	0.125	0.42	1.0	9.1	0.259	0.40	1.0	8.7	1
SNV + SG	0.778	0.25	1.6	5.5	0.684	0.30	1.4	6.4	5
SNV + 1st + SG	0.713	0.28	1.4	6.0	0.690	0.28	1.4	6.0	5
SNV + 1st + ND	0.798	0.25	1.6	5.5	0.899	0.17	2.4	3.7	9
SNV + 2nd + SG	0.586	0.32	1.3	6.9	0.383	0.36	1.1	7.7	3
SNV + 2nd + ND	0.781	0.25	1.6	5.3	0.838	0.22	1.9	4.7	8

## 2.5 特征波段的筛选

在预测不同固体样品中的花青素、可溶性固形物以及总糖含量<sup>[11]</sup>、藏红花素<sup>[25]</sup>、水分、脂肪、蛋白质、总酚类化合物、咖啡因和可可碱<sup>[26]</sup>等,均可以证明光谱范围的选择可以有效地提取对应分量的光谱信息,避免过拟合,大大提高了模型的精度。本研究采用人工筛选方法<sup>[15]</sup>将全波段分为 11 个大的区间:4 000~4 500  $\text{cm}^{-1}$ ,4 500~5 000  $\text{cm}^{-1}$ ,5 000~5 500  $\text{cm}^{-1}$ ,5 500~6 000  $\text{cm}^{-1}$ ,6 000~6 500  $\text{cm}^{-1}$ ,6 500~7 500  $\text{cm}^{-1}$ ,7 500~8 000  $\text{cm}^{-1}$ ,8 000~8 500  $\text{cm}^{-1}$ ,8 500~9 000  $\text{cm}^{-1}$ ,9 000~9 500  $\text{cm}^{-1}$ ,9 500~10 000  $\text{cm}^{-1}$ 。结合 TQ Analyst 软件对波段进行了严格筛选,排除了与组分无关的波段。最后,将选定的谱区用于单谱区或组合谱区进行比较。

为了进一步评估光谱范围变量选择的效果,表 6 分别对 4 个优化 PLS 模型的选择范围、*RMSECV*、*RMSEP*、*RPD* 和 *RSD* 进行了汇总。通过对比表 2~表 6 中的 *RMSECV* 和 *r* 等参数,可以清楚地看出光谱范围选择对模型的影响。例如校正集中,水分模型的 *RMSECV*、*r*、*RPD* 和 *RSD* 在表 2 中分别为 0.32%、0.976、4.6 和 5.5%。表 6 校正集中,在 9 826~9 729  $\text{cm}^{-1}$ ,8 665~8 217  $\text{cm}^{-1}$  和 7 388~6 095  $\text{cm}^{-1}$  谱区下,*RMSECV* 和 *RSD* 进一步降低到 0.23% 和 4.0%,而 *r* 和 *RPD* 增加到 0.987 和 6.3。对于脂肪、蛋白质和灰分,也得到了类似的结果。

由表 6 可以看出,*RMSECV* 和 *RMSEP* 的值,校正集和验证集中的 *r*、*RPD* 和 *RSD* 值比较接近。对于水分和脂肪模型,*RMSEP* 略小于 *RMSECV*,对于蛋白质和灰分模型,验证集 *r* 相较校正集的略小,均

方根误差略大。通过比较表2~表5和表6中均方根误差、 $r$ 和 $RPD$ ,光谱预处理和谱区筛选使模型有显著提升的效果。以水分模型为例,未经预处理光谱的 $RMSECV$ 、校正集 $RPD$ 、校正集和验证集 $r$ 值分别为0.32%、4.5、0.975和0.983;采用SG预处理方

法后,4个值分别为0.32%、4.6、0.976和0.992;利用人工筛选谱区方法预处理之后, $RMSECV$ 进一步降低到0.23%,校正集和验证集 $r$ 值进一步提高到0.987和0.993,校正集 $RPD$ 进一步提高到6.3。脂肪、蛋白质和灰分模型也有类似的情况。

表6 基于最优光谱预处理的特征波段下建模结果

指标	光谱预处理	光谱范围/ $\text{cm}^{-1}$	校正集				验证集			
			$r$	$RMSECV/\%$	$RPD$	$RSD/\%$	$r$	$RMSEP/\%$	$RPD$	$RSD/\%$
水分	SG	9 826 ~ 9 729, 8 665 ~ 8 217, 7 388 ~ 6 095	0.987	0.23	6.3	4.0	0.993	0.21	6.8	3.5
脂肪	1st + MSC	9 136 ~ 8 838, 7 311 ~ 7 052, 5 601 ~ 4 136	0.894	0.78	2.3	2.4	0.972	0.52	3.8	1.6
蛋白质	1st + SNV + SG	6 003 ~ 5 150, 4 446 ~ 4 020	0.926	0.95	2.7	4.2	0.925	1.10	2.5	5.0
灰分	1st + SNV + ND	9 922 ~ 9 037, 8 250 ~ 8 155, 5 944 ~ 4 121	0.964	0.10	3.8	2.3	0.923	0.15	2.6	3.3

## 2.6 可行性研究

为了进一步研究4种优化模型的可预测性,采用偏最小二乘法拟合得到PLS模型和化学法分析得到校正集和验证集4个品质指标的拟合线,见图3。由图3可知,散点与拟合线之间距离不大,无明显偏离点,可见大部分预测结果与实测值无显著差异,验证

了定量模型可靠性,可在实际检测中实时反馈奇亚籽品质。虚线为验证集拟合线,实线为校正集拟合线,两条线高度重叠,其中水分模型最佳, $r$ 值达到0.98以上,其余指标的 $r$ 值在0.9左右或0.9以上。结果表明,实验所建立的预测模型可以实现对不同产地奇亚籽的水分、脂肪、蛋白质以及灰分的快速无损检测。

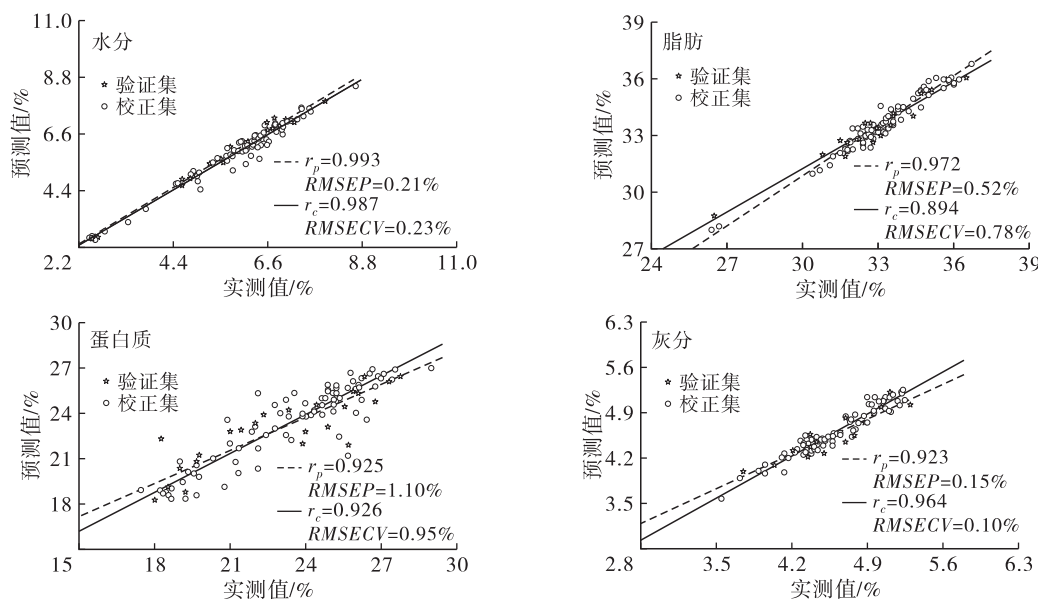


图3 奇亚籽水分、脂肪、蛋白质和灰分的模型拟合线

## 3 结论

本实验以产自墨西哥、阿根廷、澳大利亚、巴拉圭等地的奇亚籽为研究对象,探讨了基于近红外光谱的奇亚籽水分、脂肪、蛋白质和灰分的同时快速无损检测方法。选择合适的谱区范围和光谱预处理方法可以有效地提高模型的性能,4个模型验证结果为 $0.923 \leq r_p \leq 0.993$ ,  $0.15\% \leq RMSEP \leq 1.10\%$ ,

$2.5 \leq RPD \leq 6.8$ 。以上结果表明模型的预测值与实测值接近,预测效果良好。综上,该模型可用于食品工业的快速在线检测和评估,有利于奇亚籽原料的合理利用。

## 参考文献:

- [1] FERNANDES S S, SALAS - MELLADO M D L M. Addition of chia seed mucilage for reduction of fat content in



- bread and cakes[J]. *Food Chem*, 2017, 227: 237 – 244.
- [2] ROJAS V M, MARCONI L F D C B, GUIMARES – INCIO A, et al. Formulation of mayonnaises containing PUFAs by the addition of microencapsulated chia seeds, pumpkin seeds and baru oils [J]. *Food Chem*, 2019, 274: 220 – 227.
- [3] KWON H C, BAE H, SEO H G, et al. Short communication: chia seed extract enhances physicochemical and antioxidant properties of yogurt [J]. *J Dairy Sci*, 2019, 102 (6): 4870 – 4876.
- [4] 姚宏燕, 罗文涛, 杨成, 等. 奇亚籽油的品质特性及提取工艺研究进展 [J]. *中国油脂*, 2019, 44 (4): 46 – 49.
- [5] 荣旭, 陶宁萍, 李玉琪, 等. 奇亚籽营养成分分析与评价 [J]. *中国油脂*, 2015, 40(9): 89 – 93.
- [6] 文宁, 宁伟伟, 陶宁萍. 奇亚籽蛋白的提取和营养评价 [J]. *中国油脂*, 2018, 43(9): 44 – 48.
- [7] 陆婉珍. 现代近红外光谱分析技术 [M]. 北京: 中国石化出版社, 2007.
- [8] 洛曲, 于修焯, 张建新, 等. 基于近红外光谱的藏区酥油脂肪和蛋白质含量快速检测分析 [J]. *中国油脂*, 2018, 43(3): 136 – 140.
- [9] 王乐, 史永革, 李勇, 等. 在线近红外过程分析技术在豆粕工业生产上的应用 [J]. *中国油脂*, 2015, 40(1): 91 – 94.
- [10] 李路, 黄汉英, 赵思明, 等. 大米蛋白质、脂肪、总糖、水分近红外检测模型研究 [J]. *中国粮油学报*, 2017, 32(7): 121 – 126.
- [11] 卜晓朴, 彭彦昆, 王文秀, 等. 生鲜紫薯花青素等多品质参数的可见 – 近红外快速无损检测 [J]. *食品科学*, 2018, 39(16): 227 – 232.
- [12] CAREDDA M, ADDIS M, IBBA I, et al. Building of prediction models by using Mid – Infrared spectroscopy and fatty acid profile to discriminate the geographical origin of sheep milk [J]. *LWT – Food Sci Technol*, 2017, 75: 131 – 136.
- [13] 乔瑶瑶, 赵武奇, 胡新中, 等. 近红外光谱技术检测燕麦中蛋白质含量 [J]. *中国粮油学报*, 2016, 31(8): 138 – 142.
- [14] 钱丽丽, 宋雪健, 张东杰, 等. 基于傅里叶变换近红外光谱法鉴别五常大米 [J]. *食品科学*, 2018, 39(8): 231 – 236.
- [15] SINELLI N, CERRETANI L, EGIDIO V D, et al. Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity [J]. *Food Res Int*, 2010, 43(1): 369 – 375.
- [16] 李娟, 李忠海, 付湘晋. 稻谷新陈度近红外快速无损检测的研究 [J]. *光谱学与光谱分析*, 2012, 32(8): 2126 – 2130.
- [17] WANG L, WANG Q, LIU H, et al. Determining the contents of protein and amino acids in peanuts using near – infrared reflectance spectroscopy [J]. *J Sci Food Agric*, 2013, 93(1): 118 – 124.
- [18] 母建华. 基于光谱分析的茶鲜叶全氮含量快速检测技术 [D]. 江苏 镇江: 江苏大学, 2008.
- [19] WILLIAMS P, NORRIS K. Near – infrared technology in the agricultural and food industries [M]. USA: American Association of Cereal Chemists, 1987: 145 – 169.
- [20] SNCHEZ – CARNERERO C C, NEZ – SNCHEZ N, CASANO S, et al. The potential of near infrared spectroscopy to estimate the content of cannabinoids in *Cannabis sativa* L.: a comparative study [J]. *Talanta*, 2018, 190: 147 – 157.
- [21] 牛智有. 鱼粉、精料补充料及其中肉骨粉含量的近红外漫反射光谱分析 [D]. 北京: 中国农业大学, 2005.
- [22] ZHAN H, FANG J, TANG L, et al. Application of near – infrared spectroscopy for the rapid quality assessment of *Radix Paeoniae Rubra* [J]. *Spectrochim Acta A*, 2017, 183: 75 – 83.
- [23] CASCANT M M, BREIL C, FABIANO – TIXIER A S, et al. Determination of fatty acids and lipid classes in salmon oil by near infrared spectroscopy [J]. *Food Chem*, 2018, 239: 865 – 871.
- [24] KAUR B, SANGHA M K, KAUR G. Calibration of NIRS for the estimation of fatty acids in *Brassica juncea* [J]. *J Am Oil Chem Soc*, 2016, 93(5): 673 – 680.
- [25] LI S L, SHAO Q S, LU Z H, et al. Rapid determination of crocins in saffron by near – infrared spectroscopy combined with chemometric techniques [J]. *Spectrochim Acta A*, 2018, 190: 283 – 289.
- [26] HASHIMOTO J C, LIMA J C, CELEGHINI R M S, et al. Quality control of commercial cocoa beans (*Theobroma cacao* L.) by near – infrared spectroscopy [J]. *Food Anal Method*, 2018, 11(5): 1510 – 1517.