

基于微型近红外光谱仪油菜籽粗脂肪 与粗蛋白校正模型的建立

王春峰¹, 白志杰¹, 孙武坚², 郭丽¹, 熊颖¹, 陆道礼¹, 陈斌^{1*}

(1. 江苏大学食品与生物工程学院, 镇江 212013; 2. 江苏大学机械工程学院, 镇江 212013)

摘要: 为了验证微型近红外光谱仪的现场分析实用性, 利用该光谱仪测定了油菜籽中粗脂肪与粗蛋白的含量。采集油菜籽样品的近红外反射光谱, 光谱经预处理和异常样本剔除后, 结合偏最小二乘法回归 (PLSR) 建立油菜籽的粗脂肪与粗蛋白定量分析模型。结果表明, 粗脂肪的模型校正相关系数 (R_c)、校正均方根误差 ($RMSEC$)、预测相关系数 (R_p) 和预测均方根误差 ($RMSEP$) 分别为 0.9187、1.1873、0.8162 和 1.3895; 粗蛋白的模型校正相关系数 (R_c)、校正均方根误差 ($RMSEC$)、预测相关系数 (R_p) 和预测均方根误差 ($RMSEP$) 分别为 0.8773、0.8153、0.8033 和 0.7532。验证了该光谱仪在油菜籽的粗脂肪含量和粗蛋白含量检测方面是可行的, 为进一步拓展微型近红外光谱仪的应用奠定了基础。

关键词: 油菜籽; 微型近红外光谱仪; 粗脂肪; 粗蛋白

中图分类号: TS203

文献标识码: A

文章编号: 1672-352X (2017)03-0541-05

Establishment of a calibration model for rapeseed crude fat and crude protein using a miniature near-infrared spectrometer

WANG Chunfeng¹, BAI Zhijie¹, SUN Wujian², GUO Li¹, XIONG Ying¹, LU Daoli¹, CHEN Bin¹

(1. School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013;

2. School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013)

Abstract: To verify the practicability of using a miniature near-infrared spectrometer for live analysis, the crude fat and crude protein contents in rapeseeds were determined. The near infrared reflectance spectra of the rapeseeds were collected followed by pretreatment and elimination of the abnormal samples. Partial least squares regression (PLSR) model was established for crude fat and crude protein of rapeseeds. The results showed that, for the crude fat model, the correlation coefficient (R_c) and root mean square error ($RMSEC$) of the calibration set and the correlation coefficient (R_p) and root mean square error of the prediction set were 0.9187, 1.1873, 0.8162 and 1.3895, respectively. For the crude protein model, the correlation coefficient (R_c) and root mean square error ($RMSEC$) of the calibration set and the correlation coefficient (R_p) and root mean square error of the prediction set were 0.8773, 0.8153, 0.8033 and 0.7532, respectively. These results demonstrated that detection of crude fat and crude protein in rapeseeds using the miniature near-infrared spectrometer is feasible, which built a solid foundation for the further application of the miniature near-infrared spectrometer.

Key words: rapeseed; miniature near-infrared spectrometer; crude fat; crude protein

油菜是我国主要农作物之一, 是世界上仅次于大豆的第二大植物油脂来源^[1], 油菜籽中的蛋白质也是饲料蛋白质和人类蛋白质的来源之一^[2], 所以油菜籽的粗脂肪含量和粗蛋白含量是评价油菜籽品

质的重要指标。传统的化学分析方法检测这 2 个指标, 耗时长、分析过程复杂和需化学试剂处理^[3], 而近红外光谱检测技术以其快速、无损及无污染等优点普遍应用于油菜籽的品质检测^[4]。国外方面,

收稿日期: 2016-12-02

基金项目: 国家自然科学基金项目 (31171697) 和国家重大科学仪器设备开发专项 (2014YQ491015) 共同资助。

作者简介: 王春峰, 硕士研究生。E-mail: wangkaka0320@163.com

* 通信作者: 陈斌, 教授, 博士生导师。E-mail: ncp@ujs.edu.cn

Hom 等^[5]研究小样品油菜籽含油率近红外光谱分析模型, Sidhu 等^[6]研究二极管阵列检测型近红外光谱仪的油菜籽含油率分析模型, Daszykowski 等^[7]研究了近红外光谱分析技术应用于油菜籽粗蛋白、粗脂肪模型建立; 国内, 高建芹等^[8]研究近红外光谱法测定油菜籽含油量及脂肪酸, 袁建等^[9]研究了油菜籽整籽粒和粉碎样品蛋白质、脂肪等含量。以上这些研究都是通过大型的台式仪器进行的, 而使用微型近红外光谱仪检测油菜籽粗脂肪和粗蛋白却鲜有报道。

近红外光谱分析技术 (near infrared spectroscopy, NIRS) 被誉为分析界的“巨人”, 它以高效、快速、绿色和无损等优点在农业、石化和制药等领域得到了广泛应用^[10-11]。但是目前绝大部分近红外光谱仪体积大、价格高且结构复杂, 难以满足现场分析的要求^[12]。

近红外光谱仪在不断提高信噪比、增加仪器稳定性和采样方便的方向发展的同时, 也在仪器的小型化、便携化和微型化方向发展。采用一些最新的光学原理、加工技术和 MEMS 技术, 使得仪器达到了小型化、微型化的阶段。如江苏大学食品与生物工程学院的近红外研究团队与济南海能仪器股份有限公司合作研制出一种微型近红外光谱仪。

本研究旨在验证微型近红外光谱仪的实用性。以研制的微型光谱仪 N100 样机 (简称 N100) 为基础, 检测油菜籽的粗脂肪、粗蛋白含量, 并与 FOSS 公司的 FOSS NIRSystem 6500 近红外光谱仪检测的结果比较。

1 材料与方法

1.1 试验材料

2016年5月20日到6月30日期间, 收集产自江苏、山东和重庆等地的油菜籽样品 104 个, 置于室内通风阴干、人工去壳过筛后, 依次编号并于 25℃ 恒温保存备用。

1.2 试验仪器

N100 型微型近红外光谱仪集光源、探测器、内置电路控制系统、电池和样品杯于一体, 无需任何移动部件^[13], 重量仅为 590 g 左右, 如图 1 所示。

光源: 光源采用 Welch Allyn 公司的 997418-21 型卤钨灯, 其参数: 电压 3.5 V, 电流 0.43 A, 色温 2 330 K, 寿命 10 000 h。其特点是光源稳定和一致性优良。

探测器: 仪器以日本滨松光子学株式会社生产的 C11708MA 近红外探测器为核心。探测器整合了

MEMS 技术和图像传感器技术, 其中 MEMS 技术具有微型化、智能化、多功能、高集成度和适于大批量生产等特点, 也是近红外光谱仪器发展的新趋势。具体参数如表 1 所列。

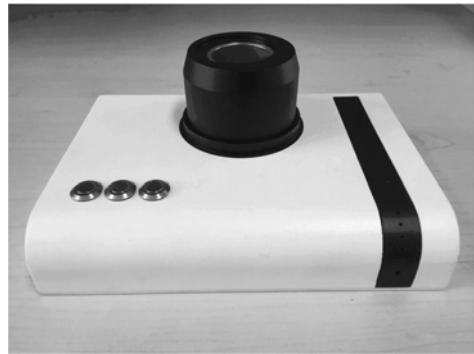


图 1 微型近红外光谱仪样机外形照片

Figure 1 The miniature near-infrared spectrometer prototype photos

表 1 型号 C11708MA 主要技术参数

Table 1 Main technology parameters of C11708MA

技术参数 Technical parameter	规格 Standard
波长范围 Spectral response range	640~1 050 nm
波长分辨率 Spectral resolution	≤20 nm
波长重现性 Wavelength reproducibility	-0.5~+0.5 nm
频谱杂散光 Wavelength temperature dependence	-0.05~+0.05 nm·°C ⁻¹
积分时间 Integral time	5~10 000 ms
信噪比 SNR	15 000:1
扫描次数 Scan times	1~32 767 次
电源电压 Supply voltage	4.75~5.25 V
功率 Power	30 mW
主机接口 Master interface	USB2.0
检测器 Detector	Si
检测器尺寸 Dimensions	27.6 mm×13 mm×16.8 mm
A/D	16 bit
运行环境温度 Operating temperature	+5~+40°C
仪器保管温度 Storage temperature	-20~70°C

1.3 光谱采集

如图 2 所示, 样本杯中心位置是偏离光源中心的, 因此转动样品杯可采集样品不同部位的漫反射光谱。试验温度 25℃, 湿度 40%RH, 仪器的积分时间 16 ms, 扫描次数 32 次。以镀金白板作为背景, 将 104 个油菜籽样品依次均匀地装于样品杯中, 装样厚度接近样品杯高度的 2/3。每个样品漫反射光谱采集 3 次, 取 3 次平均值为最终的样本光谱, 共 104 条油菜籽样本光谱。

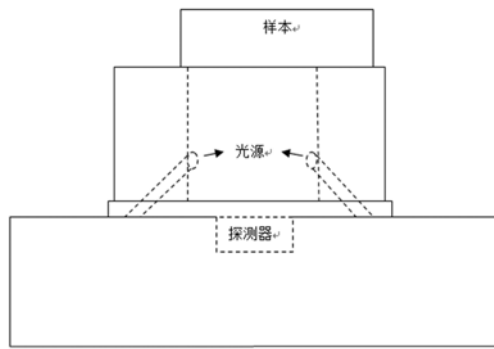


图 2 微型近红外光谱仪样机主要结构示意图

Figure 2 The main structure diagram of miniature near-infrared spectrometer prototype

1.4 粗脂肪与粗蛋白的测定

依据 GBT 5512-2008, 测得油菜籽样品的粗脂肪含量; 依据 GBT 14489.2-2008, 测得油菜籽样品的粗蛋白含量。使用的仪器分别是海能仪器股份有限公司生产的 SOX406 脂肪测定仪、K1100Q 全自动凯氏定氮仪。每个样本平行测 3 组, 取 3 组平均值做最终化学值。粗脂肪平行试验结果相对误差不大于 2%, 粗蛋白平行试验结果相对误差不大于

1.5%。

1.5 模型建立与评价

基于 Matlab 平台和江苏大学近红外工作室 NIRS Ver4.4.5 数据处理系统完成光谱数据的预处理和定量分析模型的建立。选用校正相关系数 (correlation coefficient of calibration, R_C)、预测相关系数 (correlation coefficient of prediction, R_P)、校正均方根误差 (root mean square error of calibration, $RMSEC$) 和预测均方根误差 (root mean square error of prediction, $RMSEP$) 作为模型的评价标准。好的模型应该有较低的均方根误差, 较高的相关系数: R_C 越高, $RMSEC$ 越小, 表明模型稳定性越好; R_P 越高, $RMSEP$ 越小, 表明模型预测能力越强^[14]。

2 结果与分析

2.1 校正集与预测集划分

校正集和预测集采用 Kennard-Stone 法进行划分, 粗脂肪、粗蛋白含量分布分别如表 2 和表 3 所示。校正集的粗脂肪、粗蛋白含量范围大于预测集的范围, 有利于使预测模型更加稳健。

表 2 油菜籽粗脂肪含量分布情况

Table 2 The content distribution of crude fat in rapeseeds

数据集 Dataset	样品数 Sample number	最小值 Minimum	最大值 Maximum	平均值 Mean	标准差 Standard deviation	变异系数 Variable coefficient	%
校正集 Calibration set	70	31.963	47.763	40.079	2.867	7.153	
预测集 Prediction set	31	36.308	44.332	39.624	1.733	4.373	

表 3 油菜籽粗蛋白含量分布情况

Table 3 The crude protein content distribution Content distribution in crude protein of rapeseeds

数据集 Dataset	样品数 Sample number	最小值 Minimum	最大值 Maximum	平均值 Mean	标准差 Standard deviation	变异系数 Variable coefficient	%
校正集 Calibration set	69	16.472	25.288	20.873	1.699	8.140	
预测集 Prediction set	31	19.040	23.415	21.425	1.260	5.881	

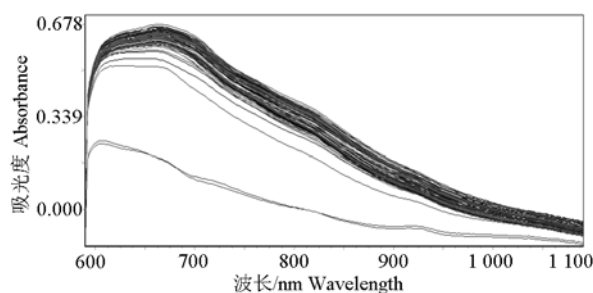


图 3 油菜籽原始光谱

Figure 3 Original spectra of rapeseeds

2.2 光谱预处理与异常样本剔除

采集得到的原始光谱 (图 3) 中不仅携带样本的化学成分信息, 还包含了其他无关的信息和噪声^[15];

在建立油菜籽的近红外光谱定量分析模型时, 异常样本的存在则会降低校正模型的相关性。因此建立模型前, 对原始光谱进行预处理和异常样本剔除就变得十分关键。预处理可以减弱甚至消除非样本信息; 剔除异常样本可以提高光谱与化学值相关性。对采集到的光谱数据进行移动平滑滤波 (moving average filter, MAF)、卷积平滑滤波 (savitzky-golay filter, SGF)、归一化 (normalization)、中心化 (centralization)、多元散射校正 (multiplicative scatter correction, MSC) 和标准正态变换 (standard normal variate transformation, SNV) 等预处理后, 再利用 PLS 结合马氏距离, 进行识别异常样本并剔除处理, 图 4 和图 5 是对原始光谱采用归一化+SGF+剔除 3

个异常样本和 MSC+剔除 4 个异常样本方法处理后的光谱。

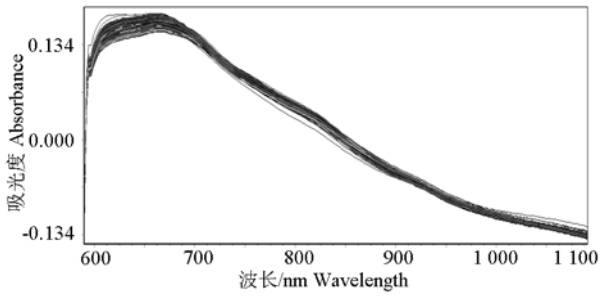


图 4 归一化+SGF 预处理+异常样本剔除后油菜籽光谱
Figure 4 The spectrum of rapeseeds after Normalization + SGF + Abnormal samples disposed

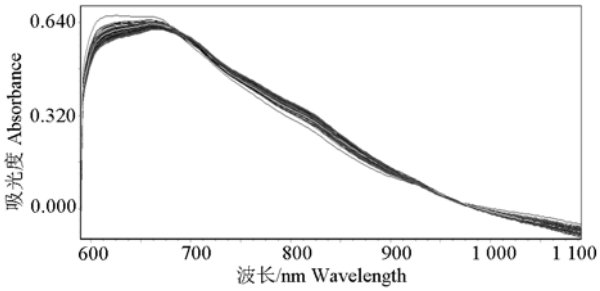


图 5 MSC 预处理+异常样本剔除后油菜籽光谱
Figure 5 The spectrum of rapeseeds after MSC + Abnormal samples disposed

2.3 校正模型建立

偏最小二乘回归 (partial least squares regression, PLSR) 可以在自变量存在多重相关和样本数少于变量个数的条件下进行回归建模。它不是直接考虑因变量与自变量的建模关系, 而是将变量中的信息进行综合筛选后, 选取最具有解释能力的新综合变量进行建模^[7]。表 4 和表 5 分别是经预处理和异常样本剔除后, 粗脂肪、粗蛋白的 PLSR 校正模型和预测结果。对于粗脂肪模型, 虽然经过 SGF、归一化、中心化和 MSC 预处理后, 模型的校正相关系数 (R_c) 要好于归一化+SGF, 但是其预测相关系数却劣于归一化+SGF, 但出现了过拟合现象。分析原因可能是粗脂肪模型的主成分数过多, 将噪声信息当作样本的化学成分信息参与了模型的建立, 而经过归一化+SGF 预处理后, 模型的主成分数由最高的 14 个减少到 10 个, 有效地简化了模型的复杂性和冗余性, 提高了模型的预测能力和稳定性; 同样对于粗蛋白模型, 也出现相似结果, 经过 MSC 预处理后, 模型的主成分数由最高的 13 个减少到 9 个, 模型效果较优。因此可以确定经归一化+SGF 预处理+异常样本剔除后的粗脂肪模型和经 MSC 预处理+异常样本剔除后粗蛋白的模型效果较优。

表 4 多种预处理方法所建粗脂肪 PLSR 校正模型和预测结果

Table 4 Results of calibration and prediction for the crude fat PLSR model established using different pretreatment methods

预处理方法 Preprocessing	主成分数 PLS component number	校正集 Calibration set		预测集 Prediction set	
		R_c	RESEC	R_p	RMSEP
Raw	13	0.9708	0.6327	0.5994	1.6541
MAF	11	0.8956	1.1733	0.4790	1.6348
SGF	14	0.9626	0.7144	0.6414	1.5399
归一化 Normalization	11	0.9519	0.8345	0.8100	1.3660
中心化 Centralization	13	0.9708	0.6327	0.5994	1.6541
MSC	10	0.9700	0.6525	0.5604	1.6709
SNV	10	0.9445	0.8942	0.7452	1.5330
归一化+SGF	10	0.9187	1.0772	0.8162	1.3895

表 5 多种预处理方法所建粗蛋白 PLSR 校正模型和预测结果

Table 5 Results of calibration and prediction for the crude protein PLSR model established using different pretreatment methods

预处理方法 Preprocessing	主成分数 PLS Component number	校正集 Calibration set		预测集 Prediction set	
		R_c	RESEC	R_p	RMSEP
Raw	11	0.9173	0.6469	0.5541	0.5541
MAF	13	0.8785	0.7668	0.5692	1.2887
SGF	11	0.8881	0.7467	0.6165	1.1824
归一化 Normalization	11	0.9365	0.6011	0.6820	0.9575
中心化 Centralization	11	0.9173	0.6469	0.5541	0.5541
MSC	9	0.8773	0.8153	0.8033	0.7532
SNV	11	0.9484	0.5034	0.6213	0.9663

表 6 2 种仪器的粗脂肪模型精度比较

Table 6 The comparison of crude fat model precision between N100 and FOSS 6500

仪器 Instrument	主成分数 PLS Components	校正集 Calibration set		预测集 Prediction set	
		R_c	$RESEC$	R_p	$RMSEP$
N100	10	0.9187	1.0772	0.8162	1.3895
FOSS 6500	11	0.8739	1.3281	0.7312	1.4892

表 7 2 种仪器的粗蛋白模型精度比较

Table 7 The comparison of crude protein model precision between N100 and FOSS 6500

仪器 Instrument	主成分数 PLS component number	校正集 Calibration set		预测集 Prediction set	
		R_c	$RESEC$	R_p	$RMSEP$
N100	9	0.8773	0.8153	0.8033	0.7532
FOSS 6500	14	0.9265	0.6356	0.8404	0.7412

2.4 仪器模型精度比较

为检验微型光谱仪建模精度,同时采用 FOSS NIRSystem 6500 近红外光谱仪(简称 FOSS 6500)建立油菜籽的粗脂肪、粗蛋白定量分析模型,将两种仪器建立的模型进行比较,结果如表 6 和表 7 所示。FOSS 6500 的波长范围为 400~2 500 nm,为保证对比一致性,选取与微型近红外光谱仪样机相同波长范围(640~1 050 nm)。对于粗脂肪模型,光谱仪样机模型精度优于 FOSS 6500 的建模效果;对于粗蛋白模型,FOSS 6500 的模型精度优于 N100 的模型精度,但二台仪器所建立的模型优劣相差不大。

3 结论

对于油菜籽粗脂肪,经归一化+卷积平滑(SGF)光谱预处理+异常样本剔除后的校正模型和预测结果较优:校正相关系数(R_c)、校正均方根误差($RMSEC$)、预测相关系数(R_p)和预测均方根误差($RMSEP$)分别为 0.9187、1.0772、0.8162 和 1.3895;对于油菜籽粗蛋白,经多元散射校正(MSC)光谱预处理+异常样本剔除后的校正模型和预测结果较优:校正相关系数(R_c)、校正均方根误差($RMSEC$)、预测相关系数(R_p)和预测均方根误差($RMSEP$)分别为 0.8773、0.8153、0.8033 和 0.7532。

N100 建立的油菜籽粗脂肪和粗蛋白的模型效果与 FOSS 6500 建立的模型效果相差不大,甚至粗脂肪模型还要优于 FOSS 6500 建立的模型,完全满足实际生活中农产品的检测要求。

N100 的波长范围属于短波段,与全波段仪器(如 FOSS 6500)相比,信息量相对小些,测量低含量组分精度可能会下降。但综合考虑仪器的价格及实用性,N100 体积小、成本低和方便便携的特点更适合用于实际应用中,其分析精度基本能适应生产过程的质量检测。

参考文献:

- [1] FU D, JIANG L, MASON A S, et al. Research progress and strategies for multifunctional rapeseed: A case study of China[J]. J Integr Agr, 2016, 15(8): 1673-1684.
- [2] KOZŁOWSKI K, JEROCH H. Enhancing the nutritional value of poultry feedstuffs using the Example of rapeseed products—a review[J]. Ann Anim Sci, 2014, 14(2): 245-256.
- [3] 陆字振, 杜昌文, 余常兵, 等. 红外光谱在油菜籽快速无损检测中的应用[J]. 植物营养与肥料学报, 2013, 19(5): 1257-1263.
- [4] 陈斌, 卢丙, 陆道礼. 基于微型近红外光谱仪的油菜籽含油率模型参数优化研究[J]. 现代食品科技, 2015 (8): 286-292.
- [5] HOM N H, BECKER H C, MÖLLERS C. Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds[J]. Euphytica, 2007, 153(1): 27-34.
- [6] SIDHU H K, HAAGENSEN D M, RAHMAN M, et al. Diode array near infrared spectrometer calibrations for composition analysis of single plant canola (*Brassica napus*) seed[J]. Appl Eng Agric, 2014, 30(1): 69-76.
- [7] DASZYKOWSKI M, WROBEL M S, CZARNIK-MATUSEWICZ H, et al. Near-infrared reflectance spectroscopy and multivariate calibration techniques applied to modelling the crude protein, fibre and fat content in rapeseed meal[J]. Analyst, 2008, 133(11): 1523-1531.
- [8] 高建芹, 张洁夫, 浦惠明, 等. 近红外光谱法在测定油菜籽含油量及脂肪酸组成中的应用[J]. 江苏农业学报, 2007, 23(3): 189-195.
- [9] 袁建, 朱贞映, 鞠兴荣, 等. FT-NIR 在油菜籽品质指标快速检测中的应用研究[J]. 中国粮油学报, 2016, 31(6): 158-162.
- [10] 褚小立, 王艳斌, 陆婉珍. 近红外光谱仪国内外现状与展望[J]. 分析仪器, 2007, 2007(4): 1-4.
- [11] 于新洋, 卢启鹏, 高洪智, 等. 便携式近红外光谱仪器现状及展望[J]. 光谱学与光谱分析, 2013, 33(11): 2983-2988.
- [12] 尹慧敏, 陈思羽. 便携式谷物成分近红外检测仪设计[J]. 粮油加工(电子版), 2015 (8): 44-45.
- [13] LUTZ O M D, BONN G K, RODE B M, et al. Reproducible quantification of ethanol in gasoline via a customized mobile near-infrared spectrometer[J]. Anal Chim Acta, 2014, 826: 61-68.
- [14] 樊书祥, 黄文倩, 李江波, 等. 特征变量优选在苹果可溶性固形物近红外便携式检测中的应用[J]. 光谱学与光谱分析, 2014, 34(10): 2707-2712.
- [15] 吴静珠, 李慧, 王克栋, 等. 光谱预处理在农产品近红外模型优化中的应用研究[J]. 农机化研究, 2011, 33(3): 178-181.