

近红外光谱分析法测定菜籽油中芥酸的含量

陈 蛋¹, 陈 斌^{1*}, 陆道礼¹, 钟旭美²

(1. 江苏大学生物与环境工程学院, 镇江 212013; 2. 陕西师范大学食品工程系, 西安 710062)

摘 要: 采用多通道 PDA 型近红外光谱仪, 应用偏最小二乘法建立了菜籽油中芥酸含量与近红外透射光谱的校正模型, 讨论多项式求导及平滑的窗口宽度和相关系数法筛选有效波长对校正模型的影响, 并对 10 个预测集样品利用预测相关系数 R_p 和预测均方根误差 $RMSEP$ 指标进行了预测精度分析, 结果发现: 在使用全谱数据进行偏最小二乘回归建模时, 一阶 7 点求导及平滑的预处理方法结果最佳, 此时建模效果为: $R_p = 0.739, RMSEP = 1.659$; 在此基础上通过相关系数法筛选波长后的建模效果为: $R_p = 0.958, RMSEP = 0.963$ 。后者 R_p 提高 29%, $RMSEP$ 减少 42%。由此可得出多项式求导及平滑法和相关系数法相结合对校正模型稳健性, 预测精度都有较大提高的结论。研究证明: 多通道近红外光谱仪快速测定菜籽油中芥酸含量的方法是可行的。

关键词: 芥酸; 多通道近红外光谱仪; 多项式求导; 相关系数法

中图分类号: TN219; TP274.52

文献标识码: A

文章编号: 1002-6819(2007)1-0234-04

陈 蛋, 陈 斌, 陆道礼, 等. 近红外光谱分析法测定菜籽油中芥酸的含量[J]. 农业工程学报, 2007, 23(1): 234-237.

Chen Dan, Chen Bin, Lu Daoli, et al. Determination of the erucic acid content in rapeseed oil by near-infrared spectroscopy [J]. Transactions of the CSAE, 2007, 23(1): 234-237. (in Chinese with English abstract)

0 引 言

随着化学工业的发展, 芥酸及其衍生物的工业用途的重要性也越来越突出, 被誉为 21 世纪的原料^[1]。菜籽油是芥酸的主要来源, 因此快速检测菜籽油中的芥酸含量成为了在菜籽油收购过程中的一种需要。目前检测菜籽油中的芥酸的主要方法是气相色谱法^[2], 但气相色谱法耗时、耗材、操作过程复杂, 只适合于实验室的检测, 难以满足生产过程的在线、实时检测。

近红外(NIR)分析方法是一种具有无损、快速、准确、简便等特点的新型分析技术, 被广泛应用于农产品、食品、药物和化工产品的定性定量分析^[3]。Sato, Tetsuo^[4]使用近红外光谱法检测芸苔中芥酸的含量, 得到了较好的结果; 顾伟株等人^[5]使用多元回归法分析菜籽油含油量的近红外光谱数据; 李廷莉等人^[6]对使用近红外光谱法进行油菜籽中芥酸含量检测进行了初步探讨, 而针对使用近红外光谱法检测菜籽油中芥酸含量的尚鲜见报道。本文从理论与实际生产相结合的角度出发, 使用带光纤探头的 CCD 型多通道近红外光谱仪, 运

用偏最小二乘回归(Partial Least Squares, PLS)^[7-9]方法建立校正模型, 对原始吸收光谱采用多项式求导及平滑进行了预处理, 以相关系数为入选波长法的准则选择波长, 并对其建立的校正模型精度进行了研究, 为菜籽油收购、生产等过程快速检测芥酸提供了一种方便、快速、准确的新方法。

1 试验仪器、材料和方法

1.1 检测仪器及附件

气相色谱分析仪采用美国惠普公司的 HP5890;

近红外光谱检测采用美国 Control Development 公司生产的 NIR-256L-1.7T1 多通道近红外光谱仪, 该仪器的探测器为单级半导体致冷 256 单元铟镓砷(InGaAs) PDA 型阵列探测器, 波长准确度优于 0.5 nm。在两个探测器单元间采用线性插值法插入 2 个光谱数据点, 每个样品的透射光谱共有 777 个光谱数据点。

1.2 物料

由成都西普油脂化工有限公司提供菜籽油样品 35 个, 芥酸浓度范围在 35%~54%, 其中 1~25 号样品作为校正集, 26~35 号样品作为预测集。

1.3 检测方法

1.3.1 化学值测定

菜籽油芥酸含量的化学方法根据 GB10219-88 进行, 色谱仪条件: 柱温 210℃, 气化室温度 250℃, 检测室温度 250℃, 进样量 1 μL。同一个样品进行 2 次平行测

收稿日期: 2005-07-01 修订日期: 2006-01-22

基金项目: 江苏省 2004 年度博士后科研资助项目

作者简介: 陈 蛋(1979-), 男, 研究方向是近红外技术在食品的无损检测应用研究。镇江 江苏大学生物与环境工程学院, 212013

*通讯作者: 陈 斌(1960-), 男, 博士, 教授, 博士生导师, 主要从事近红外光谱技术在食品和农产品应用的研究。镇江 江苏大学生物与环境工程学院, 212013。Email: ncp@ujs.edu.cn

定,其相对相差不超过 1.5%。

1.3.2 光谱采集

采用 NIR-256L-1.7T1 多通道近红外光谱仪扫描样品,为降低随机噪声的影响,扫描 40 次后取平均,积分时间设为 0.025 s,扫描区间 900~1700 nm。光纤探头直接插入油样中采集光谱。图 1 是 3 个不同芥酸含量的菜籽油样品近红外光谱图。

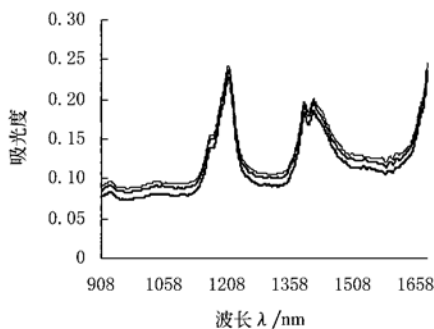


图 1 菜籽油的近红外光谱

Fig. 1 Near infrared spectra of rapeseed oil

2 数据处理与分析

试验数据的处理采用江苏大学近红外光谱分析课题组编写的《常用化学计量学方法软件》系统完成(NIRSA 数据处理系统)。软件中的 PLS 使用非线性迭代算法(nonlinear iterative partial least squares, NIPALS)^[10,11]。

由于光纤探头的灵敏度很高,因此多通道近红外光谱仪测采集光谱时易受到试验条件、环境条件等因素的影响,同时在检测透射光谱的同时也存在一定的漫反射现象,因此透射光谱中存在基线漂移现象,在建模校正模型之前需要对透射光谱进行一定的预处理,以消除各种干扰引起的基线漂移。

常用的预处理方法是多项式平滑及求导处理等^[12],在平滑、求导处理时窗口的大小(或点数)直接影响预处理的效果。相关系数法是将样本的光谱数据和化学组分含量相关联,得到相关系数,以确定信息最强的光谱范围。适当舍弃那些相关性较差的波长,选择相关系数大的波段进行建模,可以提高校正模型的预测精度^[13]。模型的评价指标使用校正相关系数 R_c ,校正均方根误差(RMSEC),预测相关系数 R_p ,预测均方根误差(RMSEP)。

2.1 全光谱预处理建模

35 个样品的光谱经过求导及平滑预处理后,随机选择其中的 25 个样品做校正集,余下的 10 个样品做预

测集,采用不同窗口宽度的导数和平滑处理后的光谱,通过 PLS 进行建立校正模型,主因子数由通过交互验证法^[7,14,15]来确定,建立校正模型的预测效果见表 1。从表中可以看到,当采用窗口宽度 7 时,校正模型的预测效果最好。

表 1 全谱的不同预处理方法的 PLS 建模结果

Table 1 PLS results of different pretreatments in the whole spectra

预处理方法	主因子数	校正集		预测集	
		R_c	RMSEC	R_p	RMSEC
无	4	0.832	1.625	0.726	1.944
一阶导数+平滑 (窗口宽度 5)	4	0.853	1.241	0.724	1.691
一阶导数+平滑 (窗口宽度 7)	4	0.865	1.207	0.739	1.659
一阶导数+平滑 (窗口宽度 9)	4	0.840	1.603	0.703	1.735

2.2 相关系数法筛选变量

对采用窗口宽度 7 的预处理方法得到的光谱中的每个波长点处的光谱变量(吸光度)与芥酸的实测值进行相关系数分析,通过反复测试,设置相关系数阈值为 0.165,把大于阈值的光谱变量保留,小于阈值的光谱变量去除,这样保留下来的光谱变量由原来的 777 个变成 334 个,对保留下来的光谱变量再采用法 PLS 建立校正模型并对预测集样品进行预测,波长筛选前后的预测效果在表 2 中列出。相关系数法运算的流程图见图 2。芥酸含量的预测值与实测值之间的关系如图 3 和图 4 所示。

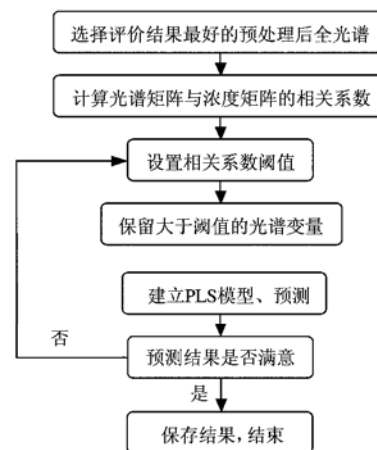


图 2 相关系数分析法筛选变量流程图

Fig. 2 Flowchart of the algorithm of correlation coefficient analysis for selecting variables

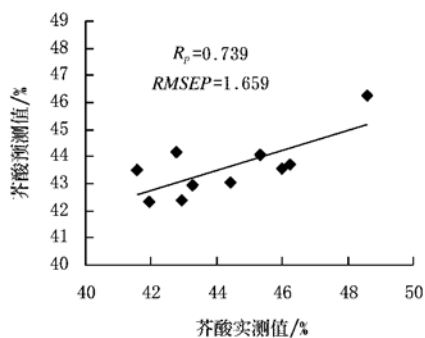


图3 采用相关系数法前的芥酸预测结果

Fig. 3 Predicted results before using correlation coefficients

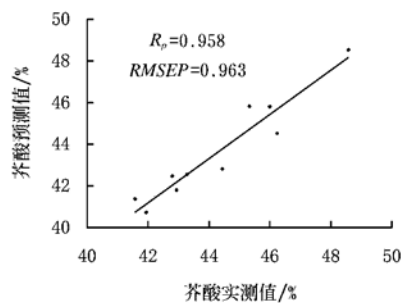


图4 使用相关系数法的芥酸预测效果

Fig. 4 Predicted results after using correlation coefficients

表2 相关系数法筛选变量前后的PLS建模结果

Table 2 PLS modeling results before and after using correlation coefficients

筛选波长法	预处理方法	变量数	校正集		预测集		
			R_c	$RMSEC$	R_p	$RMSEP$	E
无	一阶导数+平滑(窗口宽度7)	777	0.865	1.207	0.739	1.659	3.239%
相关系数法	一阶导数+平滑(窗口宽度7)	334	0.965	0.618	0.958	0.963	1.730%

3 结果与分析

通过比较表1中不同光谱预处理方法对PLS校正模型预测的影响,可见发现采用一阶导数及平滑的预处理方法对预测效果有较大的提高。这是因为在光谱的采集过程中,仪器本身的温度变化、环境的温度的波动,以及菜籽油的黏度受环境温度的影响、色泽等因素都会对样品的透射光谱产生影响使得光谱发生基线的漂移。光谱经过求一阶导数后可以消除基线的平移误差。虽然导数光谱光谱可以消除基线的平移,但同时也将随机噪声放大,必须经过一定光谱平滑处理才能有效地提高导数光谱的信噪比。在进行平滑处理时,合理选择窗口大小是影响平滑效果的关键。当窗口较小时,平滑的效果与导数光谱相比没有明显的改善,较大的窗口宽度可以提高光谱的信噪比,但窗口宽度太大会导致光谱失真,只有通过校正模型的预测效果的比较才能确定出最佳的窗口宽度。经反复比较得出窗口宽度为7时效果最好的结论(见表1)。

表2是采用相关系数法筛选变量前后校正模型的预测效果比较。从表中可见,筛选波长前预测集的 $R_p = 0.738$, $RMSEP = 1.659$, $E = 3.239\%$; 筛选波长后 $R_p = 0.958$, $RMSEP = 0.963$, R_p 改善了29%, $RMSEP$ 减少了42%, $E = 1.730\%$ 。表明光谱经过预处理后运用相关系数法,预测结果有了很大提高。经过波长的筛选,保留了能反映菜籽油中芥酸信息的主要光谱变量,

删除了一些背景干扰和次要光谱变量,优化了光谱数据,将原来的777个光谱变量减少到334个,不但简化了计算,而且也使得校正模型的抗干扰能力得到提高,校正模型对预测集样品的预测精度达到平均相对误差小于2%。图3和图4分别是筛选波长前和后的芥酸浓度的实测值与预测值的相关图。

4 结论

1) 在使用PLS法建立校正模型前,采用多项式求导及平滑对光谱进行预处理,比直接使用原始透射光谱建立校正模型的预测效果好。当使用窗口宽度7时,对原始透射光谱进行一阶导数及平滑处理后,校正集和预测集的回判精度和预测精度都得到了提高,窗口宽度过小或窗口宽度过大时,预测精度会变差。因而选择最佳的窗口宽度应该根据建立的校正模型的预测精度来定。

2) 在确定最佳的预处理方法后,用相关系数法筛选出对校正模型贡献较大的光谱变量,可以降低无关波长对建立校正模型的影响,不但可以减少建立校正模型的光谱变量数,而且可以大大提高模型的预测能力和稳健性。

3) 本研究建立的校正模型基本满足了实际的生产过程中检测菜籽油中芥酸的精度要求。

近红外光谱分析法具有方便简捷,无须任何化学试剂,不需要对样品做任何预处理的特点,是一种值得推广的环保型、无成本的快速监测方法。

[参 考 文 献]

- [1] 常致成. 芥酸及其衍生产品的开发应用[J]. 日用化学工业, 2000, 30(6): 26- 30.
- [2] 廖百森, 肖义夫, 赵 舰, 等. 气相色谱法快速测定菜籽油中芥酸[J]. 中国公共卫生, 2003, 19(2): 214- 214.
- [3] 陈 斌. 近红外光谱定量技术在方便面油份快速测定中的应用[J]. 中国粮油学报, 2002, 17(4): 44- 47.
- [4] Sato Tetsuo, Uezono Ichiro. Nondestructive estimation of fatty acid composition in seeds of *Brassica napus L.* by near-infrared spectroscopy[J]. Journal of the American Oil Chemists' Society, 1998, 75(12).
- [5] 顾伟株, 汪廷祥. 多元线性回归法分析菜籽油含油量的近红外光谱数据[J]. 中国粮油学报, 1995, 10(2): 57- 64.
- [6] 李廷莉, 孙超才, 钱小芳, 等. 油菜籽品质测定方法(近红外反射光谱法与传统化学方法)的比较[J]. 上海农业学报, 2003, 19(1): 11- 14.
- [7] 王惠文. 偏最小二乘回归方法及其应用[M]. 北京: 国防工业出版社, 2001.
- [8] Baibing Li. Model selection for partial least squares regression[J]. Chemometrics and Intelligent Laboratory Systems, 2002, 64: 79- 89.
- [9] Huang J. A comparison of calibration methods based on calibration dataset size and robustness[J]. Chemometrics and Intelligent Laboratory Systems, 2002, 62: 25- 35.
- [10] Dick Kleinknecht. A statistician's view of the single-Y PLS problem[Z]. Homepage of Chemometrics, editorial June 2002.
- [11] António S. Barros, Douglas N Rutledge. Principal components transform-partial least squares: a novel method to accelerate cross-validation in PLS regression[J]. Chemometrics and Intelligent Laboratory Systems, 2004, 73: 245 - 255.
- [12] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. 化学进展, 2004, 16(4): 528- 542.
- [13] 林 松. 近红外光谱分析技术在啤酒酒精度检测中的应用研究[D]. 镇江: 江苏大学, 2004. 31.
- [14] Harald A Martens, Pierre Dardenne. Validation and verification of regression in small data sets[J]. Chemometrics and Intelligent Laboratory Systems, 1998, 44: 99- 121.
- [15] 许 禄. 化学计量学—一些重要方法的原理及应用[M]. 北京: 科学出版社, 2004: 20- 23.

Determination of the erucic acid content in rapeseed oil by near-infrared spectroscopy

Chen Dan¹, Chen Bin^{1*}, Lu Daoli¹, Zhong Xumei²

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

2. Department of Food Engineering, Shaanxi Normal University, Xi'an 710062, China)

Abstract: In this paper, the erucic acid content of rapeseed oil and the correction model of near-infrared transmission spectrum were established based on partial least square (PLS) method by using multichannel PDA near-infrared spectroscopy instrument. The influences of window size of polynomial derivation and smoothing and wavelength selected by correlation coefficient on the correction model were discussed. The predicted precisions of ten samples were analyzed by the index of predicted relative coefficient (R_P) and root-mean-square error ($RMSEP$). The result showed that the first derivation with seven points and smooth pretreatment is the best while adopting PLS modeling on the whole spectra, and the results of R_P and $RMSEP$ are 0.379 and 1.659, respectively. When selecting wavelength by relative coefficient, R_P and $RMSEP$ are 0.958 and 0.963, respectively. Compared with the results of the two methods, the latter R_P improves by 29% and $RMSEP$ reduces by 42%. Therefore, the conclusions can be drawn that combining polynomial derivation and smooth pretreatment with correlation coefficients will improve the steadiness and predicted precisions of correction model. It shows that it is feasible to quickly determine the erucic acid content in rapeseed oil by multichannel near-infrared spectroscopy.

Key words: erucic acid; multichannel near-infrared spectroscopy; polynomial derivation; correlation coefficient