

用 Shenk's 方法提高芥酸 NIR 模型预测精度^{*}

陈 斌 吴继明

【摘要】 在光谱测量的过程中,由于光谱仪器本身的精度和测量环境等因素,经常会出现光谱的漂移、线性或非线性改变。以基于近红外光谱分析建立菜籽油中芥酸预测模型为例,将多元校正 Shenk's 算法用于修正同台仪器测量的光谱差异,大大提高了校正模型的预测精度。结果表明,对于光谱修正后再建模预测,预测均方差从 2.263 下降到 1.569,平均相对误差从 4.6% 下降到 3.218%,相关系数也由 0.780 提高到了 0.913,模型的预测精度显著提高。

关键词: Shenk's 算法 光谱修正 校正模型

中图分类号: O657.33 **文献标识码:** A

Study on the Prediction Precision Improvement of the Erucic Acid Near-infrared Spectrum Model by Using Shenk's Algorithm

Chen Bin Wu Jiming
(Jiangsu University)

Abstract

As some existed influencing factors resulting from the precise and the measure environments of spectrum instrument, the spectrum drifts, the linear or nonlinear conversion can be found in the process of the measure. Prediction errors will be made if we directly apply the results of the sample's measurement in the construction and prediction of the model. Therefore, multivariate calibration algorithm—Shenk's algorithm was adopted to correct the discrepancy of the erucic acid near-infrared spectrums and it was indicated that predication precise would be improved greatly, root mean square errors of prediction (RMSEP) has been reduced to 1.569 from 2.263, and average relative error was dropped to 3.218% from 4.6%, with the relative coefficient improved to 0.913 from 0.780.

Key words Shenk's algorithm, Spectrum correction, Calibration model

引言

随着计算机技术的发展,光谱结合化学计量学建立的快速分析方法已成功地应用于多个领域,这些分析方法多为多元校正方法^[1]。

在近红外光谱分析技术的应用过程中,当采用多元校正方法建立校正模型时,无论在建模过程还是在预测过程中,均需要用仪器测量光谱,而测量时并不能完全保证仪器检测条件固定不变,仪器输出

的信号是仪器量测信号和样品光谱信号的组合,如非线性的环境温度变化能使测量结果有很大差异^[2]。将外部条件变化的因素包含在所建模型中可以解决此问题,然而这种方法给建立模型带来较大的困难^[3]。

通过多元校正 Shenk's 算法边测量、边修正光谱的方法也可以解决仪器外部因素的变化,对修正后的光谱再建模预测,可以提高预测结果的精度。其实质是克服样品量测信号(或光谱)间的不一致

收稿日期: 2006-03-31

^{*} 2004 年度江苏省博士后科研资助计划项目

陈 斌 江苏大学食品与生物工程学院 教授 博士生导师, 212013 镇江市

吴继明 江苏大学食品与生物工程学院 硕士生

性^[4~5]。本研究是将 Shenk's 校正算法^[6~8]用于修正同台仪器测量光谱间的差异,使模型的预测精度显著提高。

1 校正原理与试验

1.1 Shenk's 算法原理

Shenk's 算法也称为专利方法。它包括波长校正和吸光度校正。

1.1.1 波长校正

(1)对应于 $R_{ms,j}$ (源机标准样品光谱阵第 i 个波长点),在 R_{ss} (目标机标准样品光谱阵)上选取窗口为 $k+j+1$ 的光谱段 $R_{ss,j+k+1}$,分别计算 $R_{ms,i}$ 与 $R_{ss,i-j}, R_{ss,i-j+1}, \dots, R_{ss,i+k-1}, R_{ss,i+k}$ 的相关系数,得到 $R_{ss,l}(i-j < l < i+k)$ 与 $R_{ms,i}$ 相关系数 r_i 最大,说明目标机上所测光谱的波长 l 与源机所测光谱的波长 i 相同。为获得更加精确的结果,选取波长 $l-1, l, l+1$ 与对应的相关系数 r_{l-1}, r_l, r_{l+1} 建立一元二次抛物线模型 $r = a_i + b_i i + c_i i^2$,由抛物线模型在目标机上得到与源机波长 i 相同的波长 i' 。

(2)循环 i ,求出所有对应的 i' 。

(3)用求得的 i 和 i' 建立一元二次抛物线波长校正模型 $i' = A + Bi + Ci^2$ 。

1.1.2 吸光度校正

波长校正后,首先用插值法计算目标波长 i' 的吸光度矩阵 $R_{ss,i}^{\#}$,并采用线性公式 $R_{ms,i} = s_{a_i} + s_{b_i} R_{ss,i}^{\#}$,求出回归系数 s_{a_i} 和 s_{b_i} 。对未知样品光谱 $R_{s,un}^{\#}$ 用波长校正模型 $i' = A + Bi + Ci^2$ 对波长进行校正,然后用插值法计算 $R_{s,un}^{\#}$,最后由 $R_{s,un}^p = s_{a_i} + s_{b_i} R_{s,un}^{\#}$ 得到转换结果。

在本研究中,从样品总体中挑选出一个样品作为标准样品,以首次测量的标准样品光谱记为 R_{ms} ,随后对本标准样品测量的光谱记为 R_{ss} 。采用 Shenk's 算法对不同时刻在同一台仪器上扫描的样品光谱进行修正,然后对修正后的光谱和修正前的光谱分别建立校正模型并对预测样本进行预测。通过预测均方差、平均相对误差以及相关系数指标,评价修正前、后所建模型的预测能力。

1.2 光谱测量

1.2.1 仪器及试验参数

仪器:北京瑞利分析仪器公司生产的 WQF-400N 型傅里叶快速变换近红外分析仪,仪器的光谱范围 $11\ 300 \sim 3\ 500\ \text{cm}^{-1}$,选用硫化铅(PbS)探测器。

试验条件:环境温度在 $26\ \text{C}$ 左右,相对湿度在 50% 左右,光源功率 $6\ \text{W}$ 、电压 $6\ \text{V}$ 卤钨灯。

试验对象:57 个不同产地的菜籽油样品,由四川西普油脂化工有限公司提供,以菜籽油中的芥酸为检测指标。

仪器参数:采样区间 $6\ 394 \sim 3\ 502\ \text{cm}^{-1}$,采样分辨率为 $4\ \text{cm}^{-1}$,采样次数 64 次,以空气为本底。每个样品均扫描 2 次,最终的光谱数据由 2 次光谱数据的平均值计算得到。

化学值测定:菜籽油中芥酸含量按 GB 10219-88 标准测定。色谱仪条件:柱温 $210\ \text{C}$,气化室温度 $250\ \text{C}$,检测室温度 $250\ \text{C}$,进样量 $1\ \mu\text{L}$ 。同一个样品进行 2 次平行测定取平均值。

1.2.2 样品光谱测量

对 57 个菜籽油样品编号,选择其中一个样品标记为标准样。在测第 1 个样品前,首先扫描标准样,以后每采样 5 个菜籽油样品光谱后重新采样 1 次标准样的光谱,以备修正光谱时使用。即:分别第 1、6、11、16、21、26、31、36、49、52 号样品前扫描标准样。

1.2.3 光谱修正

采用由江苏大学开发的近红外光谱仪器模型转移处理软件进行光谱修正。该软件系统的主界面如图 1 所示,软件分为数据导入、模型转移、模型预测、图谱比较、效果验证、图谱打印等几个功能模块。

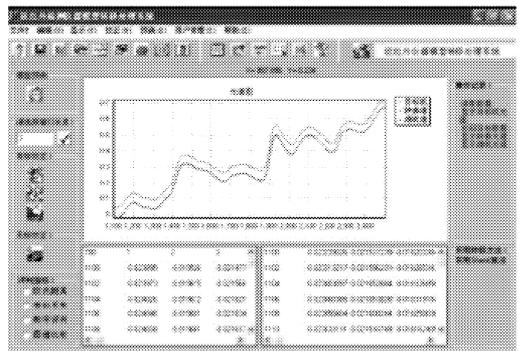


图 1 模型转移软件主界面

Fig.1 Main form of the model transfer software

采用 Shenk's 校正算法,窗口长度设为 21($k=j=10$),分别将 1、6、11、16、21、26、31、36、49、52 号样品前测得的标准样光谱视为源机光谱 R_{ms} ,将 1~5、6~10、11~15、16~20、21~25、26~30、31~35、36~48、49~51、52~57 号样品光谱分为 10 个集合,将其视为目标机光谱 R_{ss} ,将 1 号样品光谱和 6、11、16、21、26、31、36、49、52 号样品前测得的标准样光谱分别建立关联并对目标机光谱进行修正,得到修正后的光谱数据。如图 2 所示,将在 1 号和 6 号样品前扫描得到的标准样光谱通过 Shenk's 算法建立关联,再对第 1 个光谱集合通过此关联进行修正,待 10 个集合完全修正完毕后将修正后的光谱数据

再进行建模预测。

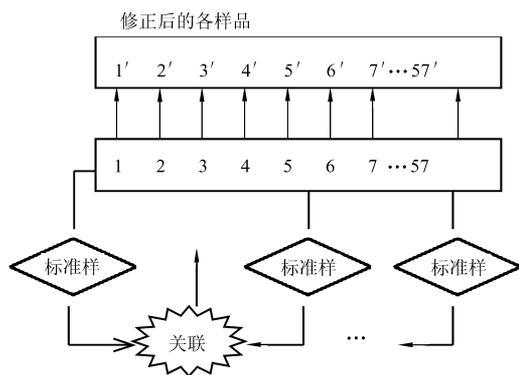


图 2 光谱修正处理方法流程图

Fig. 2 Method flow of spectrum correction

1.2.4 数据预处理和模型建立

在光谱修正时首先对目标机光谱全谱进行修正,由于仪器原因在光谱后半段出现仪器噪声,如图 3 所示。由于芥酸主要在 $5\ 400\sim 5\ 850\ \text{cm}^{-1}$ 处出现峰值,所以截取此段波段进行建模预测。

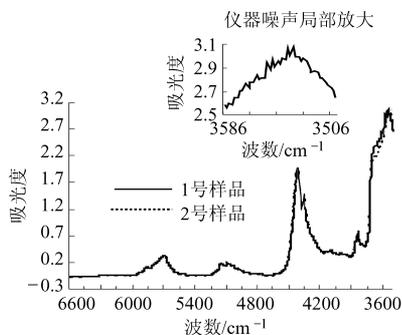


图 3 1 号和 2 号样品的全光谱图

Fig. 3 All wavelength spectrum of the No. 1 and No. 2 samples

在建模时,对校正集的划分至关重要。通常有杠杆点算法(lev 算法)、MSID 算法(maximisation of smallest interpoint distance)、Kennard-Stone 算法等^[9]。本文采用 Kennard-Stone 算法把 57 个菜籽油样品分为 2 个子集,其中选取 48 个样品作为校正集,其余 9 个样品作为预测集,预测集编号为:4、13、15、18、35、36、42、56、57。

建立校正模型时所用的数据预处理方法为 5 点一阶微分处理和 5 点 MAF 平滑处理(先微分后平滑)。分别采用偏最小二乘法(提取 15 个主成分)对修正前和修正后的光谱数据进行建模预测。采用由江苏大学开发的化学计量学软件进行光谱预处理和建立预测模型。

2 结果与分析

2.1 标准样谱图差异比较

图 4 为标准样品在 1 号和 16 号样品前扫描到

的光谱图。从图 4 可以看出,由于光谱仪器本身的光谱重现性和测量环境等因素,在光谱测量的过程中,同一样品不同时刻测量的光谱出现了较明显的差异,最大相对差值已经达到 64.67%(出现在 $5\ 774\ \text{cm}^{-1}$ 处)。

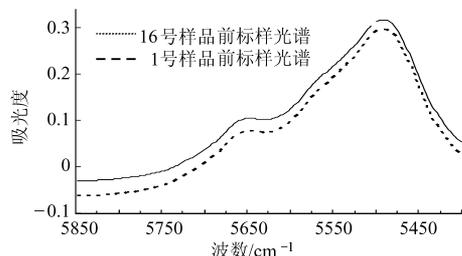


图 4 标准样谱图差异比较

Fig. 4 Differences of the standard samples spectrums

2.2 修正后的光谱图差异比较

图 5 为 16 号样品光谱经 Shenk's 修正后在峰值处差异图。从图 5 可以看出,标准样光谱经修正后光谱图与源机光谱图以及待修正的目标机光谱图间出现了较为明显的差异,修正后的光谱图已经接近于修正前的光谱。图 6 为 16 号样品修正后的光谱和源机光谱的差谱以及目标机和源机光谱的差谱比较图。从图中可以更明显地看出,修正后的光谱已经十分接近源机光谱。经过进一步计算得知,修正后的光谱在 $5\ 682\ \text{cm}^{-1}$ 处最大差值已经从 0.175 2(相对差值 9.622 0%)降到了 0.048 5(相对差值 2.663 6%)。

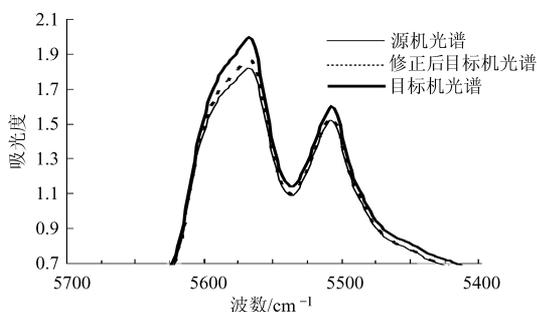


图 5 16 号样品光谱经 Shenk's 修正后峰值差异

Fig. 5 Differences of No. 16 spectrum peak value modified by Shenk's

2.3 预测结果比较

本研究中除了可以从修正后的图谱和未经修正前的图谱比较中可以直观地看出其效果,还可以从预测结果中看出其差异。将修正后的光谱数据和未经修正的光谱数据分别采用偏最小二乘法(提取 15 个主成分)建立校正模型并对同一个预测集样品集进行预测,其预测结果如表 1~2 所示。从表中可以明显看出,修正后的光谱预测精度明显优于未经修正的预测精度。对于光谱修正后再建模预测,预测均

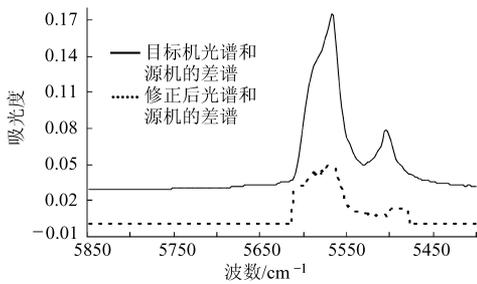


图6 源机、目标机以及修正后的光谱差谱

Fig. 6 Subtractive spectrum of transfer slaver and master spectrums

方差从 2.263 下降到 1.569, 平均相对误差从 4.600% 下降到 3.218%, 相关系数也由 0.780 提高到了 0.913, 模型的预测精度显著提高。

表1 光谱修正后和修正前模型预测结果比较

Tab. 1 Comparisons of the results between the two models

样品号	实测值	预测值		相对误差	
		光谱修正后	原光谱	光谱修正后	原光谱
4	44.726	42.565	43.290	4.832	3.212
13	42.621	40.609	41.224	4.720	3.277
15	41.744	42.840	42.120	2.625	0.901
18	44.273	41.724	41.863	5.757	5.444
35	43.271	43.023	46.890	0.572	8.364
36	46.230	45.861	46.785	0.796	1.203
42	34.986	35.988	37.000	2.864	5.756
56	47.528	45.827	43.419	3.577	8.645
57	46.633	47.918	47.960	2.757	2.846

表2 光谱修正后和修正前模型预测结果误差分析

Tab. 2 Model prediction error analysis between the two models

光谱集	平均误差		最大误差		预测均方差	相关系数
	绝对	相对/%	绝对	相对/%		
光谱修正后预测集	1.392	3.218	2.550	5.757	1.569	0.913
原光谱预测集	1.990	4.600	4.109	8.645	2.263	0.780

3 结束语

在北京瑞利分析仪器公司生产的 WQF-400N 型傅里叶快速变换近红外分析仪上检测了 57 个菜籽油样品, 借助 Kennard-Stone 算法将 48 个样品划为校正集, 9 个划为预测集; 以 1 号样品作为标准样品, 借助多元校正 Shenk's 算法不断修正其他样品的近红外光谱, 通过对比提高了修正前后模型的预测精度, 光谱修正后可明显提高预测精度。

试验证明, 通过用标准样品的光谱作为标准光谱, 定时检测系统的光谱变化和对样品光谱的不断修正可以降低仪器外部因素变化及测量过程中光谱的漂移、线性或非线性改变的影响, 及时准确修正仪器漂移等原因引起的测量光谱误差, 采用软件修正的方法提高了仪器的稳定性, 实现了校正模型标准化的目的。

参 考 文 献

- 陆婉珍, 袁洪福, 徐广通. 现代近红外光谱分析技术[M]. 北京: 中国石化出版社, 2000.
- Bouveresse E, Massart D L, Dardenne P. Calibration transfer across near-infrared spectrometric instruments using Shenk's algorithm; effects of different standardization samples[J]. Analytica. Chimica Acta., 1994, 297(3): 405~416.
- Wulfert F, Wim T K, Noord O E, et al. Coercion of temperature-induced spectral variation by continuous piecewise direct standardization[J]. Anal. Chem., 2000, 72(7): 1639~1644.
- 褚小立, 袁洪福, 陆婉珍. 光谱多元校正中的模型传递[J]. 光谱学与光谱分析, 2001, 21(6): 881~885.
- Ritchie G E, Ciurczak E W, Mark H. Books of abstracts presented at PITCON 2000 science for the 21st century [C]. New Orleans; [s. n.], 2000.
- Shenk J S, Westerhaus M O. Optical instrument calibration system: US patent, 4866644[P]. 1989.
- Park R S, Agnew R E, Barnes R J. The development of near infrared reflectance spectroscopy calibrations for undried grass silage and their transfer to another instrument using multiple and single sample standardization[J]. Near Infrared Spectrosc, 1999, 7: 117~131.
- Hoffmann U, Szydlowski N Z. Portability of near infrared spectroscopic calibrations for petrochemical parameters[J]. Near Infrared Spectrosc, 1999, 7: 733.
- Heronides, Roberto. A strategy for selecting calibration samples for multivariate modeling[J]. Chemom. Intel. Lab. Syst., 2004, 72(1): 83~91.
- 陈斌, 黄传旭, 陆道礼. 多尺度分解和主成分法在近红外光谱分析中的应用[J]. 江苏大学学报: 自然科学版, 2004, 25(2): 105~107.