

傅里叶近红外光谱的雪青梨酸度偏最小二乘法定量分析

刘燕德^{1, 2}, 应义斌¹

1 浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029

2 江西农业大学工学院, 江西 南昌 310045

摘要 利用近红外漫反射光谱测定法获取了完整雪青梨的近红外光谱($12\,500\sim 4\,000\text{ cm}^{-1}$), 采用多元校正算法偏最小二乘法(PLS)方法, 选取不同的波段范围对漫反射光谱进行有效信息提取和分析, 得出了不同因子数时 PLS 方法进行酸度分析的结果及其因子数与交互有效检验标准偏差(RMSECV)关系, 确定了最佳回归的因子数和用于定量分析的最佳波段范围。实验结果表明: 校正模型的预测精度在 $5\,452\sim 12\,285\text{ cm}^{-1}$ 波段范围内, 最佳主因子数为7时, 雪青梨总酸的预测精度最好, 其预测集的相关系数达到了0.79, 预测标准偏差为0.0186。

主题词 傅里叶近红外光谱; 完整梨; 无损检测; 偏最小二乘法; 酸度

中图分类号: O433.5 **文献标识码:** A **文章编号:** 1000-0593(2006)08-1454-03

引言

近红外光谱分析具有样品处理简单, 分析速度快, 可以同时测定多种组分, 实现非破坏性和无污染性等优点, 在食品和农产品品质检测中得到广泛的应用^[1, 2]。

近红外光谱所含信息极其丰富, 包括键强度、化学组分、电负性和氢键信息, 对固体样品还包括散射、漫反射、反射光的偏振、样品的颗粒和尺寸等信息。通常的光谱分析技术是无法将近红外较复杂的弱信息提取出来, 因此, 必须将化学计量学方法融入到近红外光谱分析之中, 才能有效地提取信息^[3]。

光谱仪采集的水果吸光度光谱信号, 其样品的光谱严重重叠, 吸收较弱, 易受噪声和背景的干扰, 而采用化学计量学方法可有效提取成分复杂而谱峰重叠的光谱信息, 其中偏最小二乘法是最普遍使用的定量分析方法, 该方法可有效地除去噪声, 解决光谱共线问题, 并有效提取光谱信息^[4, 5]。

为了更好地寻找并建立用于水果近红外光无损检测内部品质的新方法, 寻找水果的有效光谱范围和确定最佳主因子数, 本文利用偏最小二乘法, 对完整梨的近红外光谱数据采用不同的波段范围对梨总酸度进行了定量分析并对未知样本进行预测, 给出了校正模型的预测精度, 其结果较为理想。

1 实验部分

1.1 近红外光谱数据的获取

利用 Nexus FTIR 型傅里叶快速变换近红外光谱分析仪(美国 Nicolet 公司)来采集水果光谱, 室温条件下, 测量波段范围设为 $12\,500\sim 3\,800\text{ cm}^{-1}$, 采样间隔设为 $2\,0\text{ cm}^{-1}$, 扫描次数设为64, 仪器分辨率设为 $16\,0\text{ cm}^{-1}$ 。检测器用铟镓砷(InGaAs)。每个梨需进行三次光谱测量, 分别位于梨的赤道等间隔的三个位置。所有的梨检测均为赤道不同部位测量三个点, 把三次测量中反射光谱进行平均, 就得到每个梨样品平均光谱(见图1)。

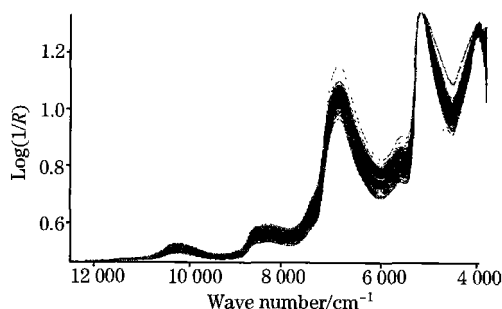


Fig 1 Original spectra of 248 samples with 16 cm^{-1} resolution

收稿日期: 2004-11-30, 修订日期: 2005-04-15

基金项目: 国家自然科学基金(60468002, 30560064)资助项目

作者简介: 刘燕德, 女, 1967年生, 浙江大学生物系统工程与食品科学学院教授, 博士研究生

1.2 样本及酸度测定

试验用样本均为从浙江大学果园购买的浙江地方主栽品种——雪青梨，样品分为校正组、检验组两组，在 26℃ 环境温度和相对湿度为 68% 的条件下采集其近红外漫反射光谱。每个梨酸度测定具体参照食品卫生检验方法理化部分总则 (GB/T 5009 1—2003) 和水果中酸度分析标准 (GB/T 12456—90)，用自动电位滴定仪 (型号: ZD-2 Titrino, 上海, 中国) 测定。具体酸度的测定结果见表 1。

Table 1 Statistic data of acidity in pear samples			
Samples		Calibration set	Prediction set
Indexes		Acidity/ %	
No. of samples		166	82
Mean		0.140	0.139
Range		0.253~0.095	0.234~0.098

1.3 数据处理与分析

采用定量分析软件 TQ Analyst v6.01 中的 PLS 方法并结合 Matlab6.3 自行编软件完成。通过留一法交叉验证来确定最佳因子数，波段选优以预测样本的标准误差 (RMSEP) 来评价与对比分析。

2 结果与讨论

2.1 近红外漫反射光谱区间的选取

本次研究测量了 248 个完整雪青梨样品，首先将上述得到的光谱数据进行中心化处理，再从 248 个样品中选取酸度分布均匀的特征样品 166 个作为校正样品，其余 82 个作为预测样品。利用 TQ Analyst 定量软件的 PLS 定量分析来进行波段的选择和优化。

根据 TQ 软件的初步分析结果，找到了用于定量分析的三个波段范围：5 452~12 285 cm⁻¹，5 452~8 658 cm⁻¹ 和 8 658~12 285 cm⁻¹。为了优化光谱波段范围，确定最佳回归和预测用光谱区间，首先进行全光谱回归，分析预测效果，然后再逐步缩短光谱范围，如果预测的标准偏差过大，则说明该光谱区间不适合定量建模分析或根据水果光谱的吸收特征来划分不同光谱区间范围。表 2 为分别在三个不同光谱区间内雪青梨酸度 PLS 定量分析校正和预测结果。

Table 2 The results for calibration models of PLS regression methods for the log (1/R), mean values of RMSEC, RMSEP and RMSECV, the correlation coefficients of determination, *r* with 5 452~12 285 cm⁻¹, 5 452~8 658 cm⁻¹ and 8 658~12 285 cm⁻¹ wavelength ranges

Parameter	Wave number range/cm ⁻¹	No. of factors	Correlation	RMSECV	RMSEC	RMSEP
TA	5 442~12 285	7	0.815	0.022	0.020	0.019
	5 452~8 658	7	0.808	0.022	0.020	0.020
	8 658~12 285	6	0.807	0.023	0.018	0.021

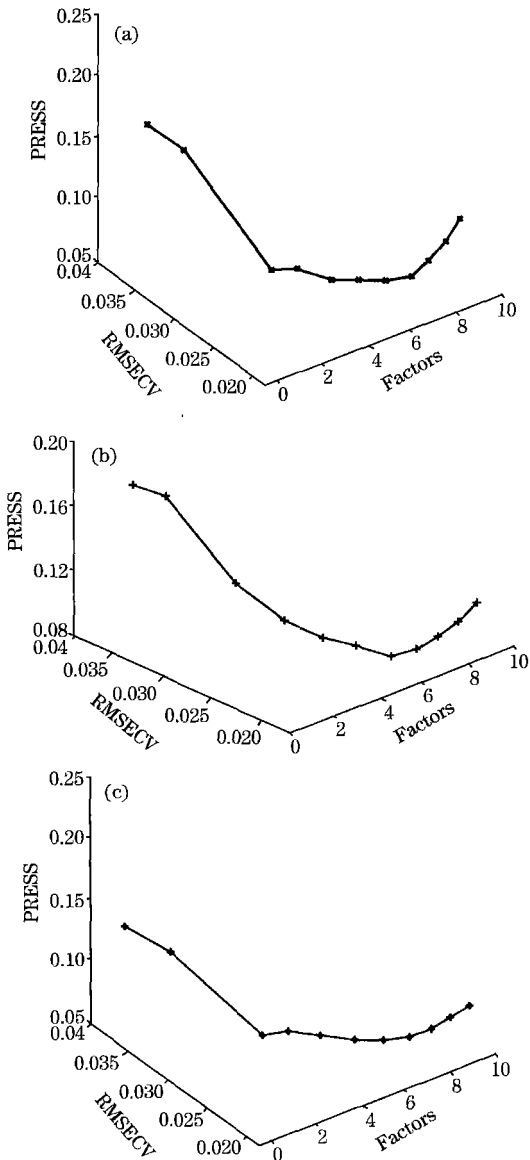


Fig 2 Relation among factors, RMSECV and PRESS in three wave number ranges PLS regression
(a): 5 452~12 285 cm⁻¹; (b): 8 658~12 285 cm⁻¹;
(c): 5 254~8 658 cm⁻¹

从表 2 可知不同的光谱波段范围，可以达到不同的回归效果，同时对样品的预测结果也不同。如果波段范围取得过宽，可能会引入不必要的变量，干扰所建立的模型，影响模型的预测能力；如果波段范围过窄，可能会人为丢失重要变量，造成预测偏差增大。因此波段范围选取适中，才能得到较好的预测结果。本实验研究通过选取有效的光谱区间，找到对预测结果较好的波段范围为 5 452~12 285 cm⁻¹。

2.2 最佳因子数的确定

因子数的合理选取与确定对一个定量模型来说是非常关键的，因子数选取较大，造成模型的过拟合现象，因子数过小模型欠拟合，因此合理选择用于定量分析用的最佳因子数对保证定量模型预测稳定性非常重要。本研究采用交叉验证

方法, 利用样品的交互验证预测标准偏差 RMSECV (root mean square errors of cross validation) 和预测残差之和 PRESS (prediction residual sum of squares) 来评价并确定其最佳因子数, RMSECV 和 PRESS 越小, 说明模型的预测能力越强。通过选取不同的因子数进行回归, 以最小的 RMSECV 值对应的因子数为最佳的回归因子数。如图 2 (a), (b), (c) 所示分别为 $5\,452 \sim 12\,285\text{ cm}^{-1}$, $5\,452 \sim 8\,658$

cm^{-1} 和 $8\,658 \sim 12\,285\text{ cm}^{-1}$ 三个波段的 RMSECV 与 PRESS 因子数的关系曲线图。从图中可见到, 该三波段的最佳因子数分别为 7, 6 和 7。

2.3 最佳定量模型的确定

通过 PLS 定量方法并比较不同波段范围的定量模型预测结果, 找出了用于最佳定量模型的有效光谱范围和最佳主因子数, 认为在波段范围为 $5\,452 \sim 12\,285\text{ cm}^{-1}$, 因子数为 7 时预测值与实际值的相关系数相对较高 ($r = 0.79$), 预测标准偏差较低 ($RMSEP = 0.019$)。图 3 为 82 个预测集雪青梨预测值与实际测量值的相关系数。

3 结 论

利用多元校正算法偏最小二乘法 (PLS) 方法, 选取不同的波段范围对梨漫反射光谱进行有效信息提取, 得出了不同因子数时 PLS 方法进行酸度定量分析的结果及其因子数与交互有效检验标准偏差 (RMSECV) 的相关关系, 确定了最佳回归的因子数为 7 和用于定量分析的最佳波段范围为 $5\,452 \sim 12\,285\text{ cm}^{-1}$ 。本文通过采用选取有效波段及最佳因子数的方法, 提高了定量分析精度, 具有实际应用价值。

参 考 文 献

- [1] Weyer L G. Applied Spectroscopy Reviews, 1985, 21(1, 2): 1.
- [2] Stark E, Luchter K, Margoshes M. Applied Spectroscopy Reviews, 1986, 22(4): 335.
- [3] ZHENG Yong mei, ZHANG Jun, CHEN Xing-dan, et al(郑咏梅, 张 军, 陈星旦, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(9): 1047.
- [4] ZHU Er yi, YANG Peng yuan(朱尔一, 扬芃原). Application and Technique of Chemertric(化学计量学技术及应用). Beijing: Science Press(北京: 科学出版社), 2001. 11.
- [5] Phil William Karl Norris. Near-Infrared Technology in the Agricultural and Food Industries. Minnesota, St. Paul: the American Association of Cereal Chemists, Inc. USA, 1987.

The Pear Acidity Quantified Analysis Using PLS Methods and Fourier Transform Near-Infrared Spectroscopy

LIU Yarr de^{1, 2}, YING Yi bin¹

1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China

2. College of Engineering, Jiangxi Agricultural University, Nanchang 330045, China

Abstract Fourier transform near infrared (FT-NIR) spectrum of intact Xueqing pear was obtained by diffusion reflectance. Partial least squares (PLS) regression was carried out, describing the relationships between the data sets of laboratory data and the FT-NIR spectra. Different wave number ranges were chosen for regression and spectral information abstraction. The 3D curves were shown for different factors, root mean square errors of cross validation (RMSECV), and prediction residual sum of squares (PRESS). Analysis results show that the best calibration model gave the relative high correlation coefficient of 0.79 and the low standard errors of prediction of 0.019 when the best wave number range was $5\,452 \sim 12\,285\text{ cm}^{-1}$ and the best factor was 7. The method of selecting advantageous wavelength ranges is feasible to obtain high prediction precision.

Keywords Fourier transform near infrared; Intact pear; Nondestructive measurement; PLS methods; Acidity

(Received Nov. 30, 2004; accepted Apr. 15, 2005)