

近红外光谱定量分析技术在枇杷可溶性固形物无损检测中的应用

李建平, 傅霞萍, 周莹, 应义斌*, 谢丽娟, 牛晓颖, 闫战科, 于海燕

浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029

摘要 应用近红外漫反射光谱定量分析技术对两个产地三个品种枇杷的可溶性固形物进行无损检测试验研究。通过分析,发现在波长1400~1500 nm和1900~2000 nm两段范围,样品的可溶性固形物与光谱吸光度之间的相关系数较高;用偏最小二乘回归PLSR、逐步多元线性回归SMLR和主成分回归PCR三种方法分别建立这两个波段和全波段范围的模型,全波段的PLSR模型的效果较优。研究发现一阶和二阶微分光谱建立的模型均不如原始光谱建立的模型效果好。最终建立三个品种枇杷样品的原始光谱在全波段范围经17点平滑后的PLSR模型,模型的校正集和预测集的相关系数分别为0.96和0.95。研究表明近红外光谱检测技术可用于枇杷可溶性固形物含量的定量分析。

关键词 近红外; 光谱学; 定量分析; 可溶性固形物; 枇杷

中图分类号: S37; TP216 **文献标识码**: A **文章编号**: 1000-0593(2006)09-1605-05

引言

不同物质在近红外区域有着丰富的吸收光谱,应用现代近红外光谱(near infrared spectroscopy, NIR)分析技术可以充分利用全谱或多波长下的光谱数据,通过化学计量学建模方法快速实现特定成分的定性或定量分析^[1]。由于近红外光谱技术的方便、快速、无损伤、无需样品预处理等特点,近年来在农业和食品领域获得了广泛的应用^[2-4]。近红外光谱技术在水果内部品质检测方面的研究主要有以下指标:糖度(常用可溶性固形物含量(soluble solid content, SSC)表示)、酸度、坚实度、干物质含量等等^[5-6];研究的水果品种也有很多,如苹果、梨、芒果、桃子、猕猴桃、樱桃等等^[6-11]。

但到目前为止,关于应用NIR技术检测枇杷内部品质的报导几乎未曾见过。枇杷主要种植在我国江浙、福建一带,果实一般在春末夏初时节成熟,具有很好的药用价值。本文应用近红外光谱技术开展枇杷可溶性固形物含量的无损检测试验研究,并比较了不同的建模方法及光谱平滑和微分处理对检测结果的影响。

1 材料与方法

1.1 材料

实验所用的枇杷样品包括了采自杭州塘栖的软条白沙品

种和白沙品种各100个,及采自杭州淳安白沙品种70个,采摘时间分别是2006年5月27日、5月30日和6月5日。采下的枇杷拿回实验室后先放入冰箱冷藏(4℃),试验前一天从冰箱取出后在室温条件下(25℃)放置24h。每个样品在光谱采集前都经清理和标记,并测量其外形尺寸(最大横径和纵径)及重量。

1.2 近红外光谱采集

枇杷的近红外漫反射光谱是用美国热电尼高力仪器公司(Thermo Nicolet Corp, USA)的NEXUS智能型FTIR光谱仪及其所带的光纤漫反射附件采集的,仪器的光源是一个50W的石英卤素灯,扫描次数和分辨率分别设置为64次和16cm⁻¹,光谱范围为800~2500nm。

光谱测量时将水果平稳地放置在水果支架上,位于水果和光纤探头之间的橡皮垫圈在防止表面反射的同时又可作为一个柔性支承。在水果光谱采集之前,先测量标准白板的光谱作为背景参比,每个水果在最大横径位置取两点测量漫反射光谱。实验获取的光谱以每一波长下的吸光度值log(1/R)表示,R是指反射光强与入射光强之比。

1.3 可溶性固形物测量

采完光谱后,将每个水果去皮去核后榨汁,使用数字折光仪(Atago Co., Ltd., Tokyo, Japan)测量果汁的可溶性固形物含量。

1.4 光谱数据处理和分析

光谱数据的采集和保存通过计算机运行光谱仪自带的

收稿日期: 2006-06-18, 修订日期: 2006-08-20

基金项目: 教育部“新世纪优秀人才支持计划”基金(NCET-04-0524)和国家博士点基金(20030335060)资助

作者简介: 李建平, 1962年生, 浙江大学生物系统工程与食品科学学院副教授 * 通讯联系人

OMNIC6 01(Thermo Nicolet Com, USA) 软件实现, 软件还可实现光谱的平滑和微分等预处理。应用 TQ Analyst (Thermo Nicolet Com, USA) 软件进行数据的分析及逐步多元线性回归 SMLR (stepwise multiple linear regression)、主成分回归 PCR (principal component regression) 和偏最小二乘回归 PLSR (partial least squares regression) 等线性模型的建立。

预测模型通过可溶性固形物的实际值与预测值的相关系数 r 、交互验证均方根误差 RMSECV (root mean square error of cross-validation)、校正均方根误差 RMSEC (root mean square error of calibration) 及预测均方根误差 RMSEP (root mean square error of validation) 定量评价。一个好的模型应该是具有较高的 r 值, 较低的 RMSECV, RMSEC 和 RMSEP 值; RMSEC 和 RMSEP 差异也是越小越好。RMSEC, RMSEP 和 RMSECV 的计算方法如下所示

$$RMSEC = \sqrt{\frac{1}{N_c - 1} \sum_{i=1}^{I_c} (\hat{y}_i - y_i)^2}$$

$$RMSEP \text{ or } RMSECV = \sqrt{\frac{1}{N_p - 1} \sum_{i=1}^{I_p} (\hat{y}_i - y_i - bias)^2}$$

其中, \hat{y}_i = 模型预测值; y_i = 实际测量值; I_c = 校正集中样品的个数; I_p = 预测集中样品的个数; $bias = \frac{1}{I_p} \sum_{i=1}^{I_p} (\hat{y}_i - y_i)$ 。

2 结果与讨论

2.1 外形参数和可溶性固形物测量结果

表 1 所示是 3 个品种枇杷外形尺寸参数和可溶性固形物检测结果。塘栖产地的样品果形较淳安的大且较匀称, 软条白沙的可溶性固形物含量比白沙的含量略高, 图 1 所示是 3 个品种枇杷的可溶性固形物分布图。

Table 1 Morphological properties and soluble solid contents of loquat samples

品种		塘栖软条白沙	塘栖白沙	淳安白沙
数量(个)		100	100	70
横径/mm	平均值	34.13	37.62	32.20
	标准偏差	2.60	2.33	3.05
纵径/mm	平均值	31.62	36.03	30.93
	标准偏差	2.88	2.20	3.19
重量/g	平均值	20.75	28.13	17.39
	标准偏差	1.51	4.57	4.75
可溶性固形物 (° Brix)	最大值	23.00	18.40	20.70
	最小值	10.50	4.30	6.60
	平均值	16.69	12.43	12.86
	标准偏差	2.68	2.80	3.19

2.2 模型预测结果

图 2 所示为枇杷可溶性固形物和吸光度在各波长下的相关系数分布曲线, 从图中可以发现, 无论是单个品种还是 3

个品种混合, 最大相关系数都分布在 1400~1500 nm 和 1900~2000 nm 左右, 因此可用 SMLR, PCR 和 PLSR 三种建模方法分别在全波段(800~2500 nm)和选择波段(1400~1500 nm 和 1900~2000 nm)建立 3 个品种枇杷的独立模型及 3 个品种的混合模型。对每个建模样本的数据集, 取 2/3 样本用于建模, 剩余 1/3 样本用于验证模型的预测性能, 所建模型还用交互验证进行评价, 结果如表 2 和表 3 所示。

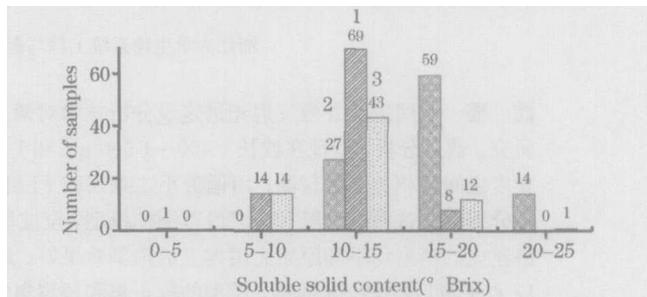


Fig 1 Soluble solid content distribution of loquat samples

1: 塘栖软条白沙; 2: 塘栖白沙; 3: 淳安白沙

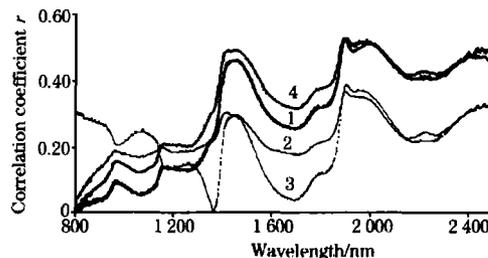


Fig 2 Correlation coefficient between SSC and $\log(1/R)$ at each wavelength

1: 塘栖白沙; 2: 塘栖软条白沙; 3: 淳安白沙; 4: 三品种

根据表 2 和表 3 所示结果, 比较 SMLR, PCR 和 PLSR 3 种方法所建模型的结果, PLSR 要明显优于 PCR 和 SMLR 方法, 这与许多研究的结论是一致的; 比较全波段和选择波段的模型结果则可以发现, 在全波段范围所建的模型的预测效果比较好; 对于 3 个品种的独立模型和 3 个品种混合的混合模型, 结果的好坏没有一定的规律。总的说来, 根据全波段范围的 PLSR 模型的交互验证结果, 三品种混合模型的实测值和光谱预测值之间的相关系数最高, 说明模型较稳定。

2.3 平滑和微分预处理对建模结果的影响

根据上面关于建模方法和建模波段选择的试验结果, 下面将采用全波段范围三品种混合的 PLSR 可溶性固形物预测模型考察光谱平滑和微分对建模的影响。对每个样品的近红外漫反射光谱用 Savit sky-Golay 方法分别进行平滑处理, 用平滑后的光谱进行建模, 比较不同平滑点数对模型预测效果的影响。图 3 所示是不同平滑点数对应的交互验证的相关系数和均方根误差。随着平滑点数的增加, 交互验证的相关系数略有上升、均方根误差略有下降, 在 17 点平滑时获得最大相关系数和最小均方根误差, 之后相关系数略有下降、均方根误差略有上升。从总体上来看, 相关系数和均方根误差随光谱平滑点数的变化还是比较小的。 <http://www.cnki.net>

Table 2 Modeling results of loquats using SMLR, PCR and PLSR methods (spectral region: 800~ 2 500 nm)

	品种	交互验证		校正和预测		
		<i>r</i>	RMSECV(° Brix)	<i>r</i>	RMSEC(° Brix)	RMSEP(° Brix)
SMLR	塘栖软条白沙	0.61	2.17	0.66	2.07	1.82
	塘栖白沙	0.64	2.18	0.70	2.03	1.72
	淳安白沙	0.63	2.54	0.68	2.39	1.76
	三品种混合模型	0.72	2.39	0.74	2.35	2.35
PCR	塘栖软条白沙	0.68	2.03	0.76	1.79	1.75
	塘栖白沙	0.90	1.35	0.92	1.10	1.20
	淳安白沙	0.68	2.43	0.79	2.02	1.67
	三品种混合模型	0.91	1.42	0.92	1.32	1.44
PLSR	塘栖软条白沙	0.87	1.36	0.98	0.57	0.66
	塘栖白沙	0.90	1.26	0.96	0.80	0.97
	淳安白沙	0.82	1.87	0.93	1.20	1.22
	三品种混合模型	0.93	1.32	0.96	0.93	1.11

Table 3 Modeling results of loquats using SMLR, PCR and PLSR methods (spectral region: 1 400~ 1 500 nm and 1 900~ 2 000 nm)

	品种	交互验证		校正和预测		
		<i>r</i>	RMSECV(° Brix)	<i>r</i>	RMSEC(° Brix)	RMSEP(° Brix)
SMLR	塘栖软条白沙	0.31	4.02	0.33	3.90	3.23
	塘栖白沙	0.47	5.56	0.49	5.38	5.11
	淳安白沙	0.35	3.59	0.41	3.36	3.09
	三品种混合模型	0.51	3.25	0.52	3.21	3.58
PCR	塘栖软条白沙	0.78	1.71	0.82	1.57	1.73
	塘栖白沙	0.31	2.76	0.54	2.39	2.23
	淳安白沙	0.61	2.60	0.71	2.31	1.55
	三品种混合模型	0.64	2.67	0.67	2.56	2.79
PLSR	塘栖软条白沙	0.82	1.57	0.89	1.26	1.56
	塘栖白沙	0.87	1.40	0.90	1.24	1.18
	淳安白沙	0.80	1.98	0.85	1.73	1.68
	三品种混合模型	0.86	1.75	0.88	1.63	1.80

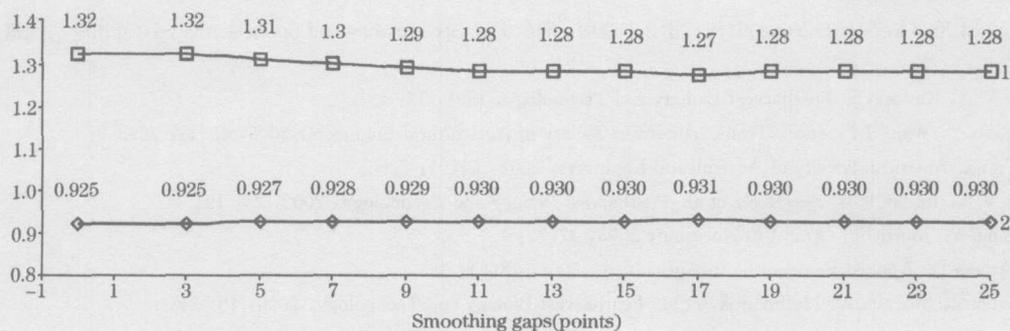


Fig 3 Effect of spectra smoothing on modeling results

1: RMSECV; 2: *r*

接下来将对三品种混合的样品原始光谱经 17 点平滑后的平滑光谱作一阶和二阶微分处理。用微分光谱建立 PLSR 模型，一阶和二阶微分光谱对应模型的交互验证的相关系数分别为 0.88 和 0.85，均方根误差分别为 1.67 和 1.84，比较交互验证结果可以发现微分光谱建立的模型不如原始光谱建立的模型。图 4 所示是用三品种混合样品原始光谱经 17 点平滑后所建立的 PLSR 模型的校正和预测结果。

3 结 论

通过对两个不同产地共 3 个品种枇杷的可溶性固形物进行近红外光谱检测试验研究，可得到以下结论。

(1) 通过分析 3 个品种枇杷的独立样品集与混合样品集各波长下每个样品的可溶性固形物与光谱吸光度之间的相关系数可知，在 1 400~ 1 500 nm 和 1 900~ 2 000 nm 两段波段

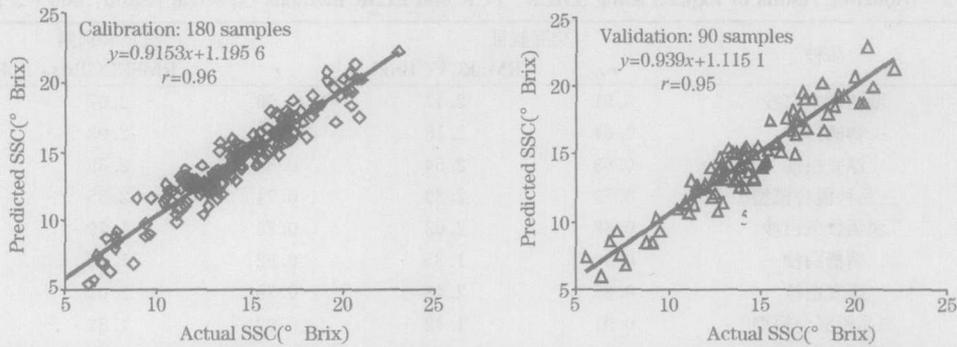


Fig 4 Calibration and validation results of PLSR model based on 17 points smoothing spectra of all three loquat varieties

围左右的相关系数较高,但在这两个波段建立的模型的效果仍不如全波段建立的模型;

(2) 通过比较 3 种建模方法的模型结果可知 PLSR 模型比 SMLR 和 PCR 模型的效果好,但 3 个品种枇杷的独立模型和混合模型的预测效果好差没有一定的规律;

(3) 研究平滑和微分预处理对建模结果的影响可知,模型交互验证相关系数和均方根误差随光谱平滑点数的变化较小,在 17 点平滑时获得最大相关系数和最小均方根误差,而一阶和二阶微分光谱建立的模型均不如原始光谱建立的模型

的效果好。

根据以上分析最终建立 3 个品种枇杷样品的原始光谱在全波段范围经 17 点平滑后的 PLSR 模型,模型的校正集和预测集的相关系数分别为 0.96 和 0.95。综上所述,近红外光谱检测技术可用于枇杷可溶性固形物含量的定量分析;同时进一步的研究还是有必要的。由于枇杷相对于苹果和梨等水果而言果形较小,光谱采集时较易漏光,对光谱采集装置的设计要特别注意。

参 考 文 献

- [1] DUAN Min-xiao, ZHAO Jiu-ran, GUO Jing-lun, et al(段民孝,赵久然,郭景伦,等). Beijing Agricultural Science(北京农业科学), 2002, 1: 11.
- [2] LIU Yan-de, YING Yi-bin, FU Xia-ping(刘燕德,应义斌,傅霞萍). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(11): 1793.
- [3] SUN Su-qin, TANG Jun-ming, YUAN Zi-min, et al(孙素琴,汤俊明,袁子民,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(2): 258.
- [4] HE Yong, LI Xiao-li, SHAO Yong-ni(何勇,李晓丽,邵咏妮). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(5): 850.
- [5] McGlone V A, Kawano S. Postharvest Biology and Technology, 1998, 13: 131.
- [6] Ying Y, Liu Y, Wang J P, et al. Trans. American Society of Agricultural Engineers, 2005, 48(1): 229.
- [7] Lu R. Trans. American Society of Agricultural Engineers, 2001, 44(5): 1265.
- [8] McGlone V A, Jordan R B, Seelye R, et al. Postharvest Biology and Technology, 2002, 26: 191.
- [9] Liu Y, Ying Y. Journal of Optical Engineering, 2005, 47(7): 1.
- [10] LU R, Ariana D. Applied Engineering in Agriculture, 2002, 18(5): 585.
- [11] Schmilovitch Z, Mizrach A, Hoffman A, et al. Postharvest Biology and Technology, 2000, 19: 245.

Application of NIR Quantitative Analysis to Nondestructive Detection of Loquat Soluble Solid Content

LI Jian-ping, FU Xia-ping, ZHOU Ying, YING Yi-bin*, XIE Li-juan, NIU Xiao-ying, YAN Zhan-ke, YU Hai-yang
College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China

Abstract The objective of the present research was to evaluate the potential of NIR diffuse reflectance spectroscopy as a way for nondestructive measurement of loquat soluble solid content of three varieties harvested from two orchards. According to the analysis, it was shown that the correlation coefficient of curves was relatively high in the two regions of 1400-1500 nm and 1900-

2 000 nm. Statistical models were developed using partial least square regression (PLSR), stepwise multiple linear regression (SMLR) and principal component regression (PCR) methods in the full spectral region from 800 to 2 500 nm and in the combined region of 1 400-1 500 nm and 1 900-2 000 nm. The results of PLSR model in the full spectral region were better than those of other models. The modeling results based on derivative spectra were not as good as those based on original spectra. Finally, a model was established based on the original spectra with 17 points smoothing for all the three varieties in the full spectral region by PLSR method. The correlation coefficients of calibration and validation were 0.96 and 0.95, respectively. The results indicate that it is feasible to use NIR spectroscopy technique for quantitative analysis of loquat soluble solid content.

Keywords Near infrared; Spectroscopy; Quantitative analysis; Soluble solid content; Loquat

(Received Jun. 18, 2006; accepted Aug. 20, 2006)

* Corresponding author

《光谱学与光谱分析》2007年征订启事

欢迎投稿 欢迎订阅

《光谱学与光谱分析》1981年创刊，国内统一刊号：CN 11-2200/O4，国际标准刊号：ISSN 1000-0593，CODEN码：GYGFED，国内外公开发行人，大16开本，208页，月刊；是中国科协主管，中国光学学会主办，钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的学术性刊物。北京大学出版社出版，每期售价30.00元，全年360元；国内邮发代码82-68，国外发行代码M905。刊登主要内容：激光光谱测量、红外、拉曼、紫外、可见光谱、发射光谱、吸收光谱、X射线荧光光谱、激光显微光谱、光谱化学分析、国内外光谱化学分析领域内的最新研究成果、开创性研究论文、学科发展前沿和最新进展、综合评述、研究简报、问题讨论、书刊评述。

《光谱学与光谱分析》适用于冶金、地质、机械、环境保护、国防、天文、医药、农林、化学化工、商检等各领域的科学研究单位、高等院校、制造厂家、从事光谱学与光谱分析的研究人员、高校有关专业的师生、管理干部。

《光谱学与光谱分析》为我国首批自然科学核心期刊，中国科协优秀科技期刊，中国科协择优支持基础性、高科技学术期刊，中国科技论文统计源刊，“中国科学引文数据库”，“中国物理文摘”，“中国学术期刊文摘”，同时被国内外的CSCI, SCL, AA, CA, EI, MEDLINE等文献机构收录。根据国家科技部信息研究所发布信息，中国科技期刊物理类影响因子及引文量《光谱学与光谱分析》都居前几位。

《光谱学与光谱分析》的主编为黄本立院士。

欢迎新老客户到全国各地邮局订阅，若有漏订者可直接与光谱学与光谱分析期刊社联系。

联系地址：北京市海淀区学院南路76号，光谱学与光谱分析期刊社

邮政编码：100081

联系电话：010-62181070, 62182998

电子信箱：chnghpxy@vip.sina.com; 修改稿专用邮箱：mengzh1018@vip.sina.com

网 址：http://ghpxy.gpfx.periodicals.net.cn