第29卷，第7期
2009年7月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 29, No. 7, pp1768-1771
July, 2009

近红外光谱预测猕猴桃硬度模型的简化研究

吕 强，汤明杰，赵杰文，蔡健荣*，陈全胜

江苏大学食品与生物工程学院，江苏 镇江 212013

摘 要 为简化猕猴桃硬度的预测模型，利用标准正态变量变换对猕猴桃1 000～2 500 nm近红外光谱进行预处理，在优选建模波段和采用净分析物预处理(NAP)降低建模主因子数两个方面简化猕猴桃硬度偏最小二乘(PLS)模型。结果表明，优选5 189～5 370 cm^{-1}，4 549～4 620 cm^{-1}，6 049～6 230 cm^{-1}，6 999～7 730 cm^{-1}，6 249～6 614 cm^{-1}等5个波段进行建模，NAP/PLS模型性能最佳，主因子数为5，校正集相关系数R^2和均方根误差RMSECV分别为0.819 41和0.701 77，预测集相关系数R^2和均方根误差RMSEP为0.780 67和0.882 71。与简化前的PLS模型相比，模型不仅更加简洁，而且预测能力和精度均有所提高。

关键词 近红外光谱；猕猴桃硬度；净分析物预处理；偏最小二乘

中图分类号：S663.4 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2009)07-1768-04

引 言

猕猴桃鲜美清香，营养丰富，食药兼用，经济、医疗价值很高。猕猴桃的硬度是衡量其可食度及口感的重要指标之一，也是决定采摘时间的一个重要因素。猕猴桃的采摘时间直接影响到它的贮藏寿命和果品质量[1]。相对于其他水果的硬度，猕猴桃的硬度在理化指标中极为重要。

目前，猕猴桃硬度的测定通常采用硬度仪或触摸的方式，存在样品预处理操作繁琐，检测周期长和主观性强等问题。近红外光谱(NIR)检测技术近年来发展很快，能实现定量、快速、无损检测，便于在线检测[2]。McGlone等[3]应用近红外光谱漫反射技术建立猕猴桃硬度的偏最小二乘(partial least square，PLS)预测模型，相关系数达0.76。Terasaki等[4]采用激光多普勒振动计检测猕猴桃的硬度和可溶性固形物含量，对不同成熟度的猕猴桃进行了分类。Lammertyn，Hu，Fu等应用可见近红外光谱对苹果[5]、番茄[6]、桃子和枇杷[7]的硬度等指标进行研究，均得到了比较满意的结果。猕猴桃的硬度是随着细胞间果胶物质的溶解程度而变化的[8]，果胶物质中含有可以吸收近红外光的CH_x和OH等化学键，这就为近红外光谱定性和定量检测猕猴桃的硬度提供了依据。

目前，国外猕猴桃近红外检测主要采用光谱波长范围为300～1 100 nm，为了尽量全面地获取待测品质的近红外光谱信息，延伸检测光谱范围，尝试利用1 000～2 500 nm的近红外光谱对猕猴桃内部品质进行检测。在猕猴桃的近红外光谱无损检测中，常用PLS建立预测模型。但是由于猕猴桃是一种组成非常复杂的生物体，加上所采取的近红外区范围广、谱带复杂、重叠多，因此光谱中必定会夹杂很多与待测品质不相关的信息，对待测样品的预测结果产生干扰，会造成PLS模型的主因子数增多[9]，模型过于复杂。为了解决这一问题，在优化建模波段和降低建模主因子数两个方面简化猕猴桃硬度模型。

1 材料与方法

1.1 材料

试验选用市售陕西省周至县中华猕猴桃111个，逐个编号后置于4 ℃冰柜中贮藏。试验前，将猕猴桃从冰柜中取出在常温下放置8 h使之与环境温度一致。试验时环境温度控制在20 ℃左右(空调控制)。先将猕猴桃表面茸毛擦拭掉，以减小其对光谱采集的影响，并将样本随机分为校正集和预测集，其中校正集73个，预测集38个。

1.2 主要仪器与检测方法

1.2.1 光谱采集

光谱采集用美国热电公司生产的Antaris Ⅱ型傅立叶变换近红外(FT-NIR)光谱仪，采集条件：以仪器内置背景为参比，积分球漫反射，扫描范围为10 000～4 000 cm^{-1}(1 000～

收稿日期：2008-05-10，**修订日期**：2008-08-20

基金项目：国家高技术研究与发展计划项目(2006AA10Z263)和国家自然科学基金项目(30771243)资助

作者简介：吕 强，1981年生，江苏大学食品与生物工程学院博士研究生 e-mail：lvqiang1111@gmail.com

*通讯联系人 e-mail：jrcai@ujs.edu.cn

2 500 nm)，扫描次数 32 次，分辨率 4 cm^{-1}。采集的猕猴桃近红外光谱如图 1(a)所示。

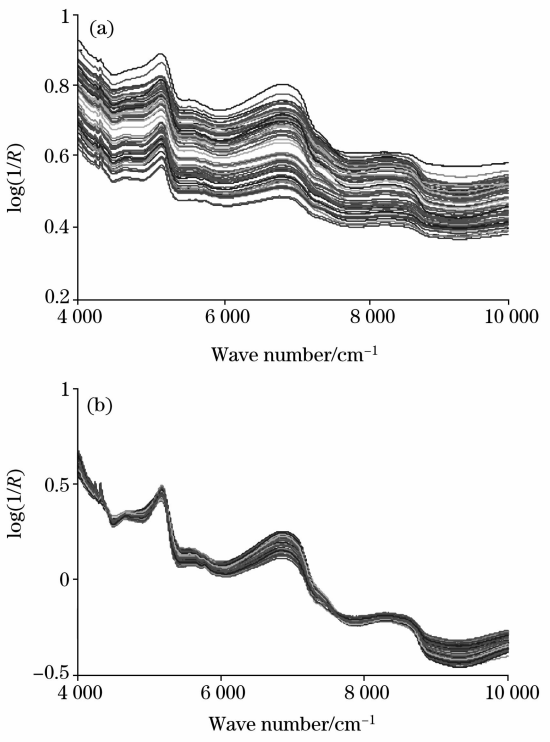

Fig. 1　NIR spectra of kiwifruit

(a): Original spectra; (b): SNV preprocessing spectra

1.2.2　硬度测定

光谱采集结束后将猕猴桃去除表皮进行硬度检测。硬度测定用英国 Stable Micro Systems 公司生产的 TA-XT2i 物性分析仪，选用直径 5 mm 的圆柱形探头 P5，对猕猴桃果实进行穿刺测试。测试参数：测前速率 1 mm·s^{-1}，测后返回速率 4 mm·s^{-1}，压缩变形量为 5 mm。穿刺测试在 3～4 mm 变形处果肉产生屈服和破裂现象，在破裂处探头的受力载荷达到最大，破裂后探头载荷明显下降。测试过程中记录最大载荷 F_{max} 表征猕猴桃果肉的硬度(N)。猕猴桃硬度实测值的变化范围、平均值、标准偏差及变异系数如表 1 所示。

Table 1　Statistic of kiwifruit firmness

	样本数	硬度(N)				变异系数/%
		平均值	最大值	最小值	标准偏差	
校正集	73	5.463	10.362	1.927	1.831	33.532
预测集	38	5.439	10.073	2.351	1.884	34.649

2　数据处理与分析

2.1　光谱预处理

采集的近红外光谱包含了噪声，噪声主要包括：随机噪声、基线漂移、信号本底、样品不均匀、光散射等。为此要对原始光谱数据进行预处理，降低噪声影响，提高检测精度[10, 11]。分别采用了均一化(mean centering，MC)、多元散射校正(multiplicative scatter correction，MSC)、标准正态变量变换(standard normal vitiate transformation，SNV)、极小/极大归一化(min/max normalization，Min/Max)、一阶导数(first derivative，1st Der)、二阶导数(second derivative，2nd Der)六种方法对猕猴桃原始光谱数据进行预处理，同时采用 PLS 建模。通过六种预处理方法的 PLS 模型的结果对比表明，SNV 预处理后的模型效果最好，校正集相关系数为 0.770 92，均方根误差(RMSECV)为 0.790 38。经过标准正态变量变换预处理后的猕猴桃近红外光谱如图 1(b)所示。

2.2　建立模型

经过 SNV 预处理后的全光谱(4 000～10 000 cm^{-1})数据与猕猴桃硬度数据进行 PLS 建模。当 PLS 主因子数为 16 时，模型的 RMSECV 达到最小，校正集相关系数 R^2 和 RMSECV 分别为 0.770 92 和 0.790 38，预测集相关系数 R^2 和均方根误差(RMSEP)为 0.744 13 和 0.940 76。

2.3　模型简化

2.3.1　建模波段的优选

本文采用的优化建模波段方法是对 Nogaard[12] 于 2000 年提出的一种波长筛选法的改进。其优选步骤如下。

(1)将整个光谱区域划分为 n 个等宽区间。

(2)在每个区间的光谱轴上以步长 m 移动，每次截取 m，$2m$，$3m$…个光谱点数据。

(3)对每次截取的波段进行 PLS 建模；将每个区间上最佳 PLS 模型(预测残差平方和(PRESS)最小)采用的波段作为该区间的优选波段，共 n 个。

(4)比较 n 个优选波段的 PRESS 值，取 PRESS 最小的波段为联合模型的第一入选波段。

(5)将余下的(n－1)个优选波段逐一与第一入选波段联合，产生(n－1)组联合波段，并在每一联合波段上进行 PLS 回归，得到(n－1)个联合模型，选择 RMSECV 值最低的模型作为该级最优联合模型；重复进行，直至余下所有优选波段都有序进入各级联合模型，并选出各级最优联合模型，共 n 个。

(6)比较第(5)步中 n 个最优联合模型的 RMSECV 值，找 RMSECV 值最小者，其对应的波段组合即为最佳组合。在最佳组合波段上所建立的 PLS 模型预测能力最强。

经过多次试验研究和比较，同时考虑程序运行时间，最终将预处理后的全光谱数据(4 000～10 000 cm^{-1})划分为 8 个等宽区间(n=8)，步长 m 取值 19。表 2 列出了 8 个光谱区间上的优选波段和 PRESS 值。从表中可以看出，精度最高的第一入选波段为第 2 区间的优选波段(5 189～5 370 cm^{-1})。在 PLS 的处理过程中，随着联合波段数的增加，所建立的 PLS 模型的 RMSECV 逐渐变小，得到最小值；如果继续增加，RMSECV 又逐渐变大。实验证明，入选波段数为 5 时，模型的 RMSECV 最小，为 0.703 28，依次入选的波段为：5 189～5 370 cm^{-1}，4 549～4 620 cm^{-1}，6 049～6 230 cm^{-1}，6 999～7 730 cm^{-1}，6 249～6 614 cm^{-1}。

将光谱划分为 8 个等宽区间，利用实验所得的 5 个优选波段联合建立硬度偏最小二乘模型，主因子数为 16，与全光

谱模型相同。而模型的校正集和预测集的预测能力却好于全光谱模型，其校正集相关系数 R^2 和 RMSECV 分别为 0.818 63和 0.703 28，预测集相关系数 R^2 和 RMSEP 分别为 0.791 68 和 0.860 25；模型得到了初步简化，实际采用波数点为 798 个(全光谱模型为 3 112 个)运算量大幅减少。

Table 2 Results of PLS calibration model selected different spectral regions

光谱区间/cm^{-1}	优选波段/cm^{-1}	波数点个数	PRESS
4 000～4 748	4 549～4 620	38	130.004 2
4 749～5 498	5 189～5 370	95	122.638 0
5 499～6 248	6 049～6 230	95	153.851 2
6 249～6 998	6 249～6 614	190	170.542 1
6 999～7 748	6 999～7 730	380	166.758 1
7 749～8 498	7 749～8 187	228	201.457 6
8 499～9 248	8 573～9 230	342	198.755 1
9 249～100 00	9 249～9 575	171	257.728 2

2.3.2 净分析物预处理

净分析物预处理(net analyte preprocessing，NAP)是由 Goicoechea 等于 2001 年首先提出，是一种基于净分析物信号[13, 14](net analyte signal，NAS)理论的光谱预处理方法，主要用于混合物体系中某一组分的光谱计量分析，其基本思想是利用数学空间正交的方法，将原始光谱矩阵中待测组分的净分析物信号提取出来，从而达到滤除无用信息的目的。

先对 SNV 预处理后猕猴桃光谱进行优选，再进行净分析物预处理，同时采用 PLS 建立硬度的预测模型(NAP/PLS 模型)。猕猴桃光谱经净分析物预处理前后的试验结果如表 3 所示。

从表 3 可以看出，对猕猴桃光谱进行净分析物预处理后，NAP/PLS 模型采纳的主因子数随着预处理过程中所用 NAP 因子的增加而逐渐减少。从预测能力及简洁性两方面综合考虑，当采用 11 个 NAP 因子时，模型性能最佳，此时主因子数为 5。校正集相关系数 R^2 和 RMSECV 分别为 0.819 41 和 0.701 77，预测集相关系数 R^2 和 RMSEP 为 0.780 67 和 0.882 71(见图 2)。与 NAP 预处理前模型相比，其预测能力虽没有得到显著提高，但模型却更简洁(主因子数减少了 11 个)。这主要是由于 NAP 法在提取硬度信号的过程中，最大程度地剔除了与硬度不相关的所有信息，只含有硬度信息和少量干扰信息。

Table 3 PLS results calculated from kiwifruit spectra before and after being preprocessed by NAP

NAP 因子数	PLS 主因子数	校正集		预测集	
		R^2	RMSECV	R^2	RMSECV
0	16	0.818 63	0.703 28	0.791 68	0.860 25
1	15	0.818 65	0.703 23	0.791 60	0.860 41
2	14	0.818 69	0.703 16	0.791 30	0.861 05
3	13	0.818 75	0.703 05	0.779 92	0.884 21
4	12	0.818 76	0.703 03	0.779 81	0.884 42
5	11	0.818 79	0.702 96	0.778 18	0.887 70
6	10	0.818 79	0.702 98	0.778 18	0.887 70
7	9	0.818 81	0.702 94	0.785 58	0.872 75
8	8	0.818 84	0.702 87	0.783 41	0.877 17
9	7	0.818 91	0.702 73	0.780 03	0.883 98
10	6	0.819 09	0.702 39	0.779 77	0.884 50
11	5	0.819 41	0.701 77	0.780 67	0.882 71
12	4	0.820 06	0.700 50	0.776 02	0.892 00
13	3	0.821 29	0.698 11	0.774 17	0.894 07
14	3	0.819 72	0.701 17	0.773 28	0.897 43
15	2	0.818 01	0.704 47	0.773 26	0.897 69
16	2	0.819 35	0.701 87	0.775 76	0.892 52

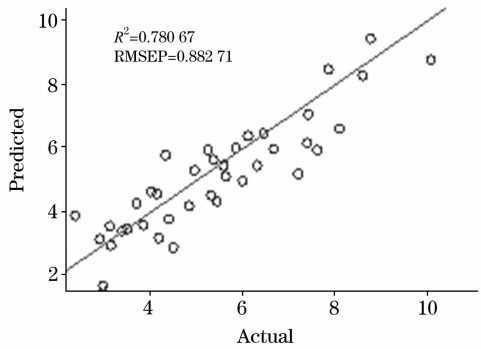

Fig. 2 Predicted *vs* actual (NAP/PLS)

利用全光谱进行 PLS 预测猕猴桃硬度时，数据量大，精度低。先进行 SNV 预处理，再优选建模波段和采用 NAP 降低建模主因子数后，PLS 模型得到简化，检测精度也有所提高(见表 4)。

Table 4 Results for different PLS models

模型	主因子数	R^2(校正集)	RMSECV	R^2(预测集)	RMSEP
全光谱 PLS	16	0.770 92	0.790 38	0.744 13	0.940 76
优化波段 PLS	16	0.818 63	0.703 28	0.791 68	0.860 25
NAP/PLS	5	0.819 41	0.701 77	0.780 67	0.882 71

3 结 论

研究采用 1 000～2 500 nm 的近红外光谱对猕猴桃硬度进行检测。通过六种预处理方法的比较，采用标准正态变量变换(SNV)对猕猴桃近红外光谱进行了预处理，滤除了近红外光谱中的部分噪声信息，通过优选建模波段和降低建模主因子数简化了猕猴桃硬度预测模型。研究结果表明，优选波段后的 PLS 模型，减少了建模所用变量数，提高了模型的预测精度；净分析物预处理(NAP)能有效地减少建模主因子

数。研究采用5 189～5 370 cm^{-1}，4 549～4 620 cm^{-1}，6 049～6 230 cm^{-1}，6 999～7 730 cm^{-1}，6 249～6 614 cm^{-1} 5 个波段，共 798 个波数点进行建模，NAP/PLS 模型性能最佳，主因子数为 5，校正集相关系数 R^2 和 RMSECV 分别为 0.819 41和 0.701 77，预测集相关系数 R^2 和 RMSEP 为 0.780 67 和 0.882 71。与简化前的 PLS 模型相比，不仅提高了预测能力，模型也更加简洁。

参　考　文　献

[1] Slaughter David C, Crisosto Carlos H. Seminars in Food Analysis, 1998, 3: 131.

[2] XU Guang-tong, YUAN Hong-fu, LU Wan-zhen(徐广通，袁洪福，陆婉珍). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2000, 20(2): 134.

[3] McGlone V Andrew, Kawano Sumio. Postharvest Biology and Technology, 1998, 13: 131.

[4] Terasaki S, Wada N, Sakurai N, et al. Transactions of the American Society of Agricultural Enginees, 2001, 44(1): 81.

[5] Lammertyn J, Nicolai B, Ooms K, et al. Transactions of the American Society of Agricultural Enginees, 1998. 41(4): 1089.

[6] Hu Xingyue, He Yong, Pereira Annia García, et al. Engineering in Medicine and Biology 27th Annual Conference. Shanghai. China. September, 2005.

[7] FU Xia-ping, YING Yi-bin, LIU Yan-de, et al(傅霞萍，应义斌，刘燕德，等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(6): 1038.

[8] Redgwell R J, Melton L D, Brasch D J. Plant Physiology, 1992, 98: 71.

[9] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立，袁洪福，陆婉珍). Progress in Chemistry(化学进展), 2004, 16(4): 528.

[10] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al(陆婉珍，袁洪福，徐广通，等). Modern NIR Spectral Analytical Techniques(现代近红外光谱分析技术). Beijing: China Petrochemistry Press(北京：中国石化出版社), 2000.

[11] Busch Kenneth W, Soyemi Olusola, Rabbe Dennis. Applied Spectroscopy, 2000, 54(9): 1321.

[12] Nogaard L, Saudland A, Wagner J, et al. Applied Spectroscopy, 2000, 54(3): 413.

[13] Goicoechea Héctor C, Olivieri Alejandro C. Chemometrics and Intelligent Laboratory Systems, 2001, 56: 73.

[14] Lorber Avraham. Analytical Chemistry, 1986, 58: 1167.

Study of Simplification of Prediction Model for Kiwifruit Firmness Using Near Infrared Spectroscopy

LÜ Qiang, TANG Ming-jie, ZHAO Jie-wen, CAI Jian-rong*, CHEN Quan-sheng

School of Food & Biological Engineering, Jiangsu University, Zhenjiang 212013, China

Abstract To simplify the prediction model of kiwifruit firmness, SNV was used to preprocess the near infrared(NIR) spectra (1 000-2 500 nm) of kiwifruit. PLS model simplification by optimizing spectral intervals and decreasing the number of factors through net analyte preprocessing(NAP) was carried out. Results showed that the performance of NAP/PLS model is the best. It was achieved with 5 factors in five wavenumber ranges(5 189-5 370, 4 549-4 620, 6 049-6 230, 6 999-7 730, and 6 249-6 614 cm^{-1}). The optimal model was achieved with R^2=0.819 41 and RMSECV=0.701 77 in the calibration set and R^2=0.780 67 and RMSEP=0.882 71 in the prediction set. This indicates that the model not only may efficiently simplify PLS model, but also may improve precision and predictive ability.

Keywords Near infrared(NIR) spectroscopy; Kiwifruit firmness; Net analyte preprocessing (NAP); Partial least square(PLS)

(Received May 10, 2008; accepted Aug. 20, 2008)

* Corresponding author