

DOI:10.6041/j.issn.1000-1298.2012.S0.042

班菲尔脐橙可溶性固形物近红外光谱特征谱区选择*

吕强¹ 何绍兰¹ 刘斌² 田喜¹ 易时来¹ 邓烈¹

(1. 中国农业科学院柑桔研究所, 重庆 400712; 2. 中国农业机械化科学研究院, 北京 100083)

【摘要】 为探讨快速无损检测班菲尔脐橙可溶性固形物(TSS)含量的方法,利用多元散射校正对脐橙1000~2500 nm近红外光谱进行了预处理,并用偏最小二乘法(PLS)、区间偏最小二乘法(iPLS)和联合区间偏最小二乘法(siPLS)分别建立预测模型。结果表明,采用siPLS将光谱划分为17个子区间,利用其中的第4(1267~1355 nm)、5(1356~1443 nm)、9(1708~1795 nm)、15(2236~2323 nm)号4个子区间联合建立的TSS模型效果最佳,其校正集决定系数和均方根误差分别为0.9109和0.3312。预测集决定系数和均方根误差分别为0.8789和0.4487,主因子数为6个。研究表明,近红外光谱技术结合siPLS可优选出表征班菲尔脐橙TSS含量信息的特征光谱区间简化预测模型,同时提高模型预测能力和精度。

关键词: 班菲尔脐橙 可溶性固形物 近红外光谱 特征谱区 联合区间偏最小二乘法

中图分类号: S666.4; O657.33 **文献标识码:** A **文章编号:** 1000-1298(2012)S0-0211-04

Selection of NIR Efficient Regions for Total Soluble Solids of *Bamfild* Navel Orange

Lü Qiang¹ He Shaolan¹ Liu Bin² Tian Xi¹ Yi Shilai¹ Deng Lie¹

(1. Citrus Research Institute, Chinese Academy of Agricultural Sciences, Chongqing 400712, China

2. Chinese Academy of Agricultural Mechanization Sciences, Beijing 100083, China)

Abstract

In order to explore a nondestructive approach to predict total soluble solids (TSS) of *Bamfild* navel orange, multiplicative scatter correction (MSC) was used to preprocess the NIR spectra (1000~2500 nm) of navel orange. Partial least square (PLS), interval partial least square (iPLS) and synergy interval partial least square (siPLS) methods were employed to establish the calibration models of TSS. The results showed that siPLS model was obtained with 17 intervals and the optimal combinations of intervals selected were No. 4 (1267~1355 nm), No. 5 (1356~1443 nm), No. 9 (1708~1795 nm) and No. 15 (2236~2323 nm). The coefficient of determination r was 0.9109 and RMSECV was 0.3312 in calibration set, and r was 0.8789 and RMSEP was 0.4487 in prediction set. It was achieved with six factors. This work demonstrated that NIR spectroscopy with siPLS model could not only select the efficient regions for TSS of orange and simplify the model, but also improve precision.

Key words *Bamfild* navel orange, Total soluble solids, NIR spectroscopy, Efficient spectral regions, Synergy interval partial least square

引言

柑桔是我国重要的经济作物,适时采收有利于

提高果实品质、延长贮藏时间和提高果实的市场价值。可溶性固形物(total soluble solids,简称TSS)含量是决定柑桔采摘时间的一个重要指标。为确定柑

收稿日期:2012-06-30 修回日期:2012-07-16

* 国家高技术研究发展计划(863计划)资助项目(2012AA101904)、重庆市国际科技合作计划资助项目(cstc2011gjh280001)和重庆市科技攻关资助项目(CSCT2012gg-yyjs80002)

作者简介:吕强,讲师,主要从事柑桔农业信息化研究,E-mail:lvqiang1111@gmail.com

通讯作者:邓烈,研究员,主要从事柑桔生理与农业信息化研究,E-mail:liedeng@163.com

桔果实的最佳采收时间,果园中柑桔 TSS 快速无损检测技术已成为当前研究的热点。

随着光谱分析技术、计算机技术和化学计量学的发展,近红外光谱分析技术以其快速、廉价、无损的优势已越来越广泛地应用在果品质量的无损检测中^[1-10]。但国内外学者在进行脐橙近红外数据处理时,常选取连续光谱建模,其数据量大,模型复杂,运算时间长。

以班菲尔脐橙(*Citrus sinensis* (L) cv. *Bamfield navel orange*)为研究对象,利用多元散射校正法(multiplicative scatter correction,简称 MSC)对脐橙近红外光谱(1 000~2 500 nm)进行预处理,尝试利用多个谱区联合建模(联合区间偏最小二乘法, synergy interval partial least square,简称 siPLS),在保证 TSS 模型预测精度的前提下,优化光谱谱区选择,减少建模所用的变量数,并缩短运算时间。

1 材料与与方法

1.1 材料

试验样本班菲尔脐橙采自重庆市奉节县晚熟柑桔示范果园。剔出果皮外表有伤痕或病斑果实后留样 49 个,置于环境温度可控的实验室(试验环境温度控制为 20℃)中 8 h,使脐橙整体温度达到与环境温度一致。将样本随机分为校正集和预测集,其中校正集 30 个,预测集 19 个。单果标号后进行果面光谱采集,随后用常规方法单果榨汁测定果汁可溶性固形物含量。

1.2 检测方法

1.2.1 光谱采集

试验采用中国农业科学院柑桔研究所与聚光科技杭州股份有限公司联合研发的 SupNIR-1500D 型便携式近红外扫描光谱仪采集样本光谱,波长范围 1 000~2 500 nm,光谱采样间隔:1 nm,扫描次数:10 次,光源采用 12 V、45 W 卤钨灯。光谱采集前用标准白板进行参比扫描。对每个果实在赤道线 4 个方向做 4 次光谱扫描,取其平均光谱。班菲尔脐橙近红外光谱如图 1 所示。

1.2.2 TSS 测定

对采集完光谱的脐橙单果榨汁,双层纱布过滤并静置 1 h 后,取上清液,用 PAL-1 型手持数字折射计(日本 ATAGO 公司)测定每个单果的果汁 TSS,每个样本测定 3 次,取平均值。表 1 列出了脐橙 TSS 实测值的变化范围、平均值、标准偏差及变异系数。

1.3 数据处理

由于样品表面的不均匀性和光谱仪电噪声等影响,脐橙近红外光谱含有较大噪声。多元散射校正

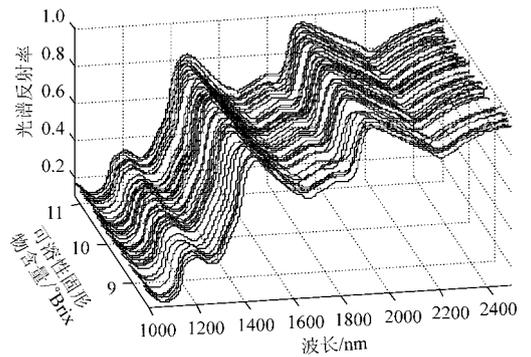


图 1 班菲尔脐橙近红外光谱

Fig. 1 NIR spectra of *Bamfield navel orange*

表 1 TSS 实测值统计

Tab. 1 Statistic of TSS

样本集	样本数	TSS/°Brix				变异系数/%
		平均值	最大值	最小值	标准偏差	
校正集	30	9.98	11.2	8.4	0.810 1	8.12
预测集	19	9.90	11.1	8.5	0.857 0	8.66

(MSC)可以有效消除表面散射以及光程变化对 NIR 漫反射光谱的影响^[11]。研究采用 MSC 对光谱进行预处理,以消除噪声影响。

对去噪后的近红外光谱分别采用最小二乘法、区间偏最小二乘法(iPLS)和联合区间偏最小二乘法(siPLS)建立脐橙 TSS 预测模型,以优选出反映 TSS 的特征光谱区间、简化模型。研究采用校正集决定系数(r)、校正均方根误差(RMSECV)、预测集决定系数和预测均方根误差(RMSEP)等作为 TSS 预测模型评价指标。

区间偏最小二乘波长筛选法和联合区间偏最小二乘法均是由 Nørgaard L^[12]于 2000 年提出的一种波长筛选法。iPLS 首先将整个光谱划分为 n 个等宽子区间。在每个子区间上进行偏最小二乘回归,可以得到 n 个局部回归模型。然后以 RMSECV 值为各模型的精度衡量标准,在 RMSECV 值最小的子区间上建立待测品质的偏最小二乘模型。

siPLS 是建立在 iPLS 基础上的一种方法,它将同一次区间划分中精度较高的几个局部模型所在的子区间联合起来,以联合模型的 RMSECV 值为各模型的精度衡量标准,在所有模型中选出性能最佳者(RMSECV 最小),其所对应的子区间组合即为最佳组合。在最佳组合区间上所建立的联合区间偏最小二乘模型预测能力最强^[13]。

2 结果与讨论

2.1 PLS 建模

用全光谱(1 000~2 500 nm)数据与脐橙 TSS 进

行 PLS 建模,当 PLS 主因子数为 9 时,模型的均方根误差(RMSECV)达到最小,模型最优:校正集决定系数为 0.865 7,RMSECV 为 0.401 0;预测集决定系数为 0.820 6,RMSEP 为 0.524 7。

2.2 iPLS 建模

将班菲尔脐橙近红外全光谱划分为 10 ~ 20 个子区间,分别进行 iPLS 建模,结果如表 2 所示。当整个光谱划分为 14 个子区间时,第 8 个子区间(1 751 ~ 1 857 nm)主因子为 1 时的 iPLS 模型最优:校正集决定系数为 0.785 4,RMSECV 为 0.493 4,预测集决定系数为 0.778 6,RMSEP 为 0.526 4。这一精度低于全光谱模型的精度,其原因主要是由于脐橙 TSS 是一项反映多种物质的综合作用指标,单一小区间不能提供预测 TSS 所需要的足够信息。

2.3 siPLS 建模

试验将光谱分别划分为 10 ~ 20 个子区间后,各最优联合区间建立脐橙 TSS 预测模型的结果如表 3

表 2 区间数的统计结果

Tab.2 Results of different iPLS models

区间数	建模区间	变量数	主因子数	校正集		预测集	
				RMSECV	r	RMSEP	r
10	5	150	3	0.458 0	0.820 9	0.663 2	0.736 7
11	3	137	4	0.462 0	0.816 3	0.641 9	0.749 8
12	4	125	4	0.366 2	0.891 8	0.669 8	0.693 0
13	2	116	5	0.434 8	0.838 5	0.605 4	0.758 3
14	8	107	1	0.493 4	0.785 4	0.526 4	0.778 6
15	3	100	4	0.468 1	0.814 7	0.703 5	0.675 6
16	3	94	4	0.479 6	0.800 2	0.686 6	0.647 1
17	3	89	4	0.514 2	0.767 4	0.667 1	0.624 6
18	4	84	1	0.526 0	0.762 9	0.697 0	0.676 7
19	11	79	1	0.496 6	0.782 0	0.547 2	0.757 0
20	11	75	1	0.521 7	0.756 5	0.548 2	0.706 5

所示。从表中可以看出,参加建立联合最优模型的波数点数目都减少了一半以上。采用多个子区间联合建模后,TSS 模型的校正性能较建立于单一子区

表 3 选择不同光谱区域的 siPLS 模型的统计结果

Tab.3 Results of siPLS model with different spectral regions selected

区间数	建模区间	变量数	主因子数	校正集		预测集	
				RMSECV	r	RMSEP	r
10	3,6,8,9	600	4	0.359 4	0.896 2	0.490 6	0.843 9
11	1,3,6	410	3	0.353 4	0.896 2	0.490 5	0.852 0
12	3,4,6,7	500	2	0.343 4	0.902 9	0.545 2	0.820 9
13	2,4,7	347	2	0.381 9	0.877 9	0.511 0	0.844 0
14	4,8,11	321	3	0.344 8	0.901 5	0.477 5	0.860 7
15	5,7,8	300	6	0.341 3	0.904 3	0.593 9	0.781 0
16	4,5,9	282	4	0.351 0	0.898 1	0.487 5	0.852 0
17	4,5,9,15	353	6	0.331 2	0.910 9	0.448 7	0.878 9
18	4,5,10	251	3	0.346 9	0.900 6	0.499 1	0.845 7
19	5,11,17	237	4	0.337 4	0.907 1	0.538 4	0.836 2
20	10,12,15	225	2	0.346 9	0.900 2	0.543 3	0.828 9

间上的模型都有所改善, RMSECV 值均在 0.36 以下,比最优 iPLS 模型的 RMSECV 值有较大幅度下降,决定系数也有了明显提高。从模型简洁性、预测能力等方面综合评价,将全光谱(1 000 ~ 2 500 nm)划分为 17 个子区间时,建立于第 4、5、9、15 号 4 个子区间(1 267 ~ 1 355 nm、1 356 ~ 1 443 nm、1 708 ~ 1 795 nm、2 236 ~ 2 323 nm)之上的模型最佳,校正集决定系数为 0.910 9,RMSECV 为 0.331 2(图 2);预测集决定系数为 0.878 9, RMSEP 为 0.448 7(图 3)。其预测能力不仅超过了全光谱模型,更重要的是,该模型得到了很大的简化,其实际采用的波数点数为 353 个,比全光谱模型少了 76.47%。这 353 个波数点最终组合为 6 个主因子,比全光谱模型少了

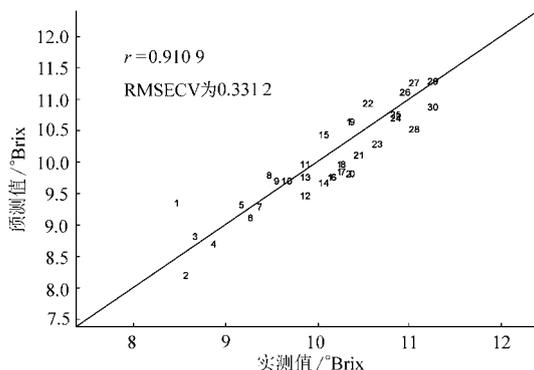


图 2 校正集 siPLS 模型 TSS 预测值与实测值

Fig.2 Predicted vs measured in calibration set (siPLS)

2 个。从上面的结果可以看出,将精度较高的几个局部模型所在子区间联合起来建立班菲尔脐橙 TSS

预测模型是可行的。

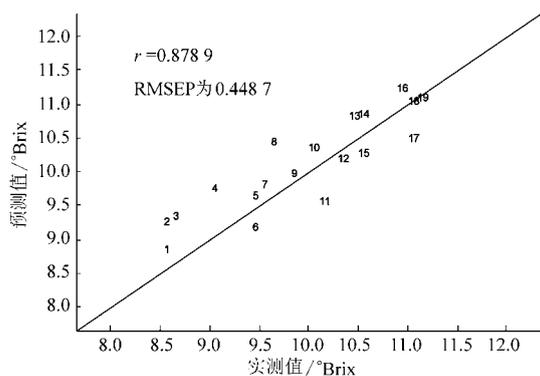


图3 预测集 siPLS 模型 TSS 预测值与实测值

Fig.3 Predicted vs measured in prediction set (siPLS)

3 结束语

利用多元散射校正法(MSC)对脐橙近红外光谱进行了预处理,然后分别采用偏最小二乘法、区间偏最小二乘法和联合区间偏最小二乘法(siPLS)建立了可溶性固形物(TSS)预测模型。结果表明,通过联合区间偏最小二乘法优选了班菲尔脐橙 TSS 的特征光谱区间(1 267 ~ 1 355 nm、1 356 ~ 1 443 nm、1 708 ~ 1 795 nm、2 236 ~ 2 323 nm)进行建模,不仅能有效地剔除噪声过大谱区和冗余信息,减少建模所用的变量数,缩短运算时间,而且提高了脐橙 TSS 近红外光谱检测模型的预测能力和精度。

参 考 文 献

- Nicolaï B M, Beullens K, Bobelyn E, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review[J]. *Postharvest Biology and Technology*, 2007, 46(2): 99 ~ 118.
- Lin Hongjian, Ying Yibin. Theory and application of near infrared spectroscopy in assessment of fruit quality: a review[J]. *Sensing and Instrumentation for Food Quality and Safety*, 2009, 3(2): 130 ~ 141.
- 徐惠荣,陈晓伟,应义斌. 基于多元校正法的香梨糖度可见/近红外光谱检测[J]. *农业机械学报*, 2010, 41(12): 126 ~ 129, 147.
Xu Huirong, Chen Xiaowei, Ying Yibin. Multivariate approach to determinate sugar content of fragrant pears with temperature variation by visible/NIR spectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2010, 41(12): 126 ~ 129, 147. (in Chinese)
- 史波林,赵镭,刘文,等. 苹果内部品质近红外光谱检测的异常样本分析[J]. *农业机械学报*, 2010, 41(2): 132 ~ 137.
Shi Bolin, Zhao Lei, Liu Wen, et al. Outlier sample analysis on near infrared spectroscopy determination for apple interior quality[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2010, 41(2): 132 ~ 137. (in Chinese)
- 吕强,汤明杰,赵杰文,等. 近红外光谱预测猕猴桃硬度模型的简化研究[J]. *光谱学与光谱分析*, 2009, 29(7): 1 768 ~ 1 771.
Lü Qiang, Tang Mingjie, Zhao Jiewen, et al. Study of simplification of prediction model for kiwifruit firmness using near infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2009, 29(7): 1 768 ~ 1 771. (in Chinese)
- Lü Qiang, Tang Mingjie, Cai Jianrong, et al. Long-term prediction of kiwifruit dry matter by near infrared spectroscopy[J]. *Science Asia*, 2010, 36(3): 210 ~ 215.
- Francesca Antonucci, Federico Pallottino, Graziella Paglia, et al. Non-destructive estimation of mandarin maturity status through portable VIS-NIR spectrophotometer[J]. *Food and Bioprocess Technology*, 2011, 4(5): 809 ~ 813.
- 刘燕德,罗吉,欧阳爱国. 可见光光谱检测赣南脐橙糖度的研究[J]. *光谱学与光谱分析*, 2007, 27(3): 569 ~ 572.
Liu Yande, Luo Ji, Ouyang Aiguo. Visible transmittance techniques for rapid analysis of sugar content of fresh navel orange juices[J]. *Spectroscopy and Spectral Analysis*, 2007, 27(3): 569 ~ 572. (in Chinese)
- 孙旭东,章海亮,欧阳爱国,等. 南丰蜜桔可溶性固形物近红外特征波段选[J]. *农业机械学报*, 2009, 40(7): 129 ~ 132.
Sun Xudong, Zhang Hailiang, Ouyang Aiguo, et al. Selection of NIR characteristic wavelength bands for soluble solids content in Nanfeng mandarin fruit[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(7): 129 ~ 132. (in Chinese)
- 邓烈,何绍兰,易时来,等. 与哈姆林甜橙内在品质主要参数相关的特征光谱研究[J]. *光谱学与光谱分析*, 2010, 30(4): 1 049 ~ 1 052.
Deng Lie, He Shaolan, Yi Shilai, et al. Study on synchronous correlation between fruit characteristic spectrum and the parameter of internal quality for Hamlin sweet orange fruit[J]. *Spectroscopy and Spectral Analysis*, 2010, 30(4): 1 049 ~ 1 052. (in Chinese)
- Chen J Y, Iyo C, Teradab F J. Effect of multiplicative scatter correction on wavelength selection for near infrared calibration to determine fat content in raw milk[J]. *Journal of Near Infrared Spectroscopy*, 2002, 10(4): 301 ~ 307.
- Nørgaard L, Saudland A, Wagner J, et al. Interval partial least squares regression(iPLS): a comparative chemometric study with an example from near-infrared spectroscopy[J]. *Applied Spectroscopy*, 2000, 54(3): 413 ~ 419.
- Chen Q, Zhao J, Liu M, et al. Determination of total polyphenols content in green tea using FT - NIR spectroscopy and different PLS algorithms[J]. *Journal of Pharmaceutical and Biomedical Analysis*, 2008, 46(3): 568 ~ 573.