

光谱预处理对苹果可溶性固形物含量 VIS/NIR 预测模型的影响

刘 亚¹, 郭俊先¹, 木合塔尔·米吉提², 岳建魁¹, 王 艳¹, 刘明涛¹

(1. 新疆农业大学 机械交通学院, 新疆 乌鲁木齐 830052; 2. 阿克苏职业技术学院, 新疆 阿克苏 843000)

摘 要:可溶性固形物含量是苹果内部品质的重要指标,以“红富士”苹果为试材,测定了从果实膨大定型期到采收期整个成熟阶段,不同生长天数苹果的可见近红外反射光谱;对光谱进行了预处理,包括导数处理、标准正态变换、多元散射校正等方法;利用偏最小二乘法(PLS)建立了苹果可溶性固形物含量的预测模型。结果表明:使用一阶导数和二阶导数处理后的光谱进行预测的准确度高于原始光谱,相关系数分别为 0.915 9、0.934 4。

关键词:光谱预处理;苹果成熟期;可溶性固形物含量;可见近红外光谱;偏最小二乘法

中图分类号:S 129 **文献标识码:**A **文章编号:**1001-0009(2016)20-0001-04

苹果果实中含有丰富的糖类、有机酸、果胶、维生素等营养物质,这些营养物质可溶性强,易被人体吸收,俗有“活水”之称,故成为人们日常食品中不可或缺的水果。伴随着生活水平的提高,消费者对苹果的风味口感、糖酸度等内在品质的要求在不断提高。传统的破坏和化学试验的苹果内部品质检测方法,已经很难应用于大批量水果品质的检测。近年来可见近红外光谱技术使水果内部品质的无损快速检测成为可能,其具有快速、高效、简便、无损伤、准确性高的特点,可广泛应用于农产品和食品的品质检测。

国内外学者对苹果的品质进行了大量的研究。PEIRS 等^[1]利用近红外光谱研究了苹果的果梗周围、赤道面、花萼周围的糖含量与漫反射光谱数据之间的数学模型。FAN 等^[2]对苹果进行了透射的光谱试验,结合可溶性固形物含量和透射光谱进行了建模,建模效果受苹果的个体大小和摆放位置的影响。GEYER 等^[3]研究了不同天气下采集到的苹果光谱参数与其可溶性固形物含量的相关关系,结果表明天气因素对二者的相关性有一定的影响。赵杰文等^[4]使用近红外漫反射光谱技术,研究了 1 300~2 100 nm 波长范围内无损检测苹果糖度的可能性。王加华等^[5]在研究不同年份苹果糖度的近

红外无损预测时,构建了一种线性组合权重偏最小二乘法模型, Rp 系数达到 0.942,所建的 LCW-MWPLS 模型预测性能好。袁雷明等^[6]组建了采集苹果半透射光谱的试验平台,并比较了不同光谱采集方式对苹果可溶性固形物含量建模预测的影响。董一威等^[7]使用 CCD 近红外光谱分析技术,验证了在 630~1 030 nm 可以实现对苹果糖度酸度无损检测的可行性。王铭海等^[8]用近红外光谱技术研究了成熟期梨的可溶性固形物,得出 UVE-LSSV 模型可以用于成熟期梨可溶性固形物的无损检测。刘辉军^[9]研究了田间梨糖度的可见近红外光谱检测方法,得出了背景光校正方法可用于田间黄花梨可见/近红外光谱采集。

综上所述,已有的研究主要针对市场上销售的苹果,结果也表明近红外光谱技术应用于苹果固酸比检测的可行性。然而针对不同生长期的苹果果园内采摘后的试验研究较少。该试验以不同生长期的苹果为研究对象,对原始光谱及经过不同预处理后的光谱与苹果的可溶性固形物含量进行了偏最小二乘法建模。以期为今后基于近红外光谱技术在苹果成熟期多品质检测仪器的开发奠定基础,发挥可见/近红外光谱技术在水果成熟期大田管理、采摘期预测、可溶性固形物检测和成熟度等级判别上的作用。

1 材料与方法

1.1 试验材料

供试苹果采摘于新疆阿克苏地区红旗坡农场园艺分场的优质“红富士”苹果种植基地,试验时间为 2015 年 8 月 20 日至 10 月 30 日,该时间处在苹果果实膨大定形

第一作者简介:刘亚(1989-),男,硕士研究生,研究方向为农产品品质无损快速检测。E-mail:zztyly@163.com.

责任作者:郭俊先(1975-),男,博士,副教授,研究方向为农产品品质无损快速检测。E-mail:junxianguo@163.com.

基金项目:国家自然科学基金资助项目(61367001)。

收稿日期:2016-07-18

期至果实成熟的采收期。

在苹果种植基地选择面积为 5 336 m² 左右的果园作为试验的样本果园,该果园种植有 264 株果树,果树年龄为 9 年,选择枝干形状近似、挂果量相同,生长健康状况基本一致的 23 株果树做为试验果树,从 8 月 23 日开始,间隔 3 d 采摘 23 株果树中 1 株上的 24 个苹果。苹果选择原则:选择果树枝叶上、中、下 3 层,东、西、南、北 4 个方向,树冠外围和内膛的 24 个方位做为采摘点,徒手进行样本的采摘,以保证苹果果梗没有断裂、无外部缺陷,采摘后装入塑料自封袋内。

1.2 试验方法

使用美国海洋光学公司 Maya 2000 Pro 可见近红外光纤光谱仪(波长范围 380~1 100 nm),采集苹果的可见与短波近红外反射光谱。采集条件:室温(22±2)℃,湿度(25±3)%,用标准聚四氟乙烯白板为背景进行光谱白板校正和遮光方式进行暗校正;开启电子暗噪声校正和杂散光校正,光谱曲线平滑度设置为 10,采集次数为 3 次。光谱采集位置:果实赤道上均匀分布 3 点,相邻测点之间间隔 120°,每个点采集 3 次光谱数据,取 3 点平均光谱作为单个样品的的光谱。光谱数据由光谱仪配套 Ocean Options Spectra Suite 软件进行采集和保存,同时,此软件还可以进行光谱图的叠加、删除、光谱数据的格式转换,可以方便的使用近红外图谱分析软件进行处理。

1.3 项目测定

1.3.1 可溶性固形物含量的测定 苹果样品光谱采集后,分别剥取 3 个采集位置处体积约 15×15×15 mm³ 的果肉,通过压蒜器榨汁,滴在 ATAGO PR-101 型数字折光仪样液池中(Tokyo, Japan, 精度±0.1°Brix),测定苹果的可溶性固形物含量,连续测量 3 次,取平均值作为测点的苹果可溶性固形物含量值。整个试验期间,测定的 552 个可溶性固形物含量范围为 8.8%~19.4%。

1.3.2 光谱预处理方法 光谱数据在采集的过程中,由于光谱仪器自身的温湿度响应特性、样品表面的不规则性及操作方法等其它不可控因素的存在,导致光谱除含有样品自身的化学信息外,还含有其它无关信息和噪声,如电噪声、样品背景、杂散光、基线漂移等^[10]。因此,为了消除这些因素对预测模型准确性的影响,需对原始光谱进行预处理。现在常用的光谱预处理方法很多,该研究中使用的预处理方法有一阶导数(first derivation, FD)和二阶导数(second derivation, SD)、多元散射校正(multiplicative scattering correction, MSC)、标准正态变换(standard normal variation, SNV)^[11-13]。对应像苹果这样的固体农产品,在整个近红外波段都存在基线漂移问题。对原始光谱进行求导处理,可以消除基线漂移、强

化谱带特征、克服谱带重叠。一阶导数处理可以消除同波长无关的漂移,二阶导数处理可以消除同波长线性相关的漂移^[14-15]。多元散射校正通常消除近红外检测中光在没有受到样品吸收或未与样品发生相互作用时经样品表面的细小微粒表面漫反射的影响,在固体漫反射光谱中应用较为广泛^[16]。标准正态变换可以用来校正因光的散射和光程的微小变化给样品光谱带来的误差,这主要是由于样品的不均匀性造成光通过、透过或从样品反射回来时的发生散射和光程的变化所导致的。同时结合去趋势算法,用来消除漫反射光谱的基线漂移,在改善光谱信号质量等方面很有效果^[17-18]。

1.4 数据分析

该研究借鉴化学计量学理中的样本选择和回归分析方法来建立成熟期苹果可溶性固形物含量预测模型,使 Matlab R2014a(Mathworks, Inc., USA)软件设计程序进行建模。

校正集和预测集的划分依据样本间欧氏距离方法,该方法优点是保证校正库中样本按照空间距离均匀分布。样本划分权重系数为 0.8,结果是 442 个样品作为校正集,110 个样品作为预测集。

偏最小二乘回归集中了多元线性回归、典型相关和主成分回归分析 3 种分析方法的优点,能够利用所有有效的数据构建模型,提取出反映数据变异的最大信息,具有良好的预测功能^[19-20]。因此利用该研究中使用 PLS 方法构建原始光谱及预处理后的光谱与苹果可溶性固形物含量的预测模型,并用独立的样本进行验证^[21-22]。

2 结果与分析

2.1 光谱预处理方法

2.1.1 标准正态变换 图 1 为不同可溶性固形物含量的苹果样本的反射光谱,整体上来看,光谱曲线的变化趋势基本一致,在 380~1 100 nm 范围内,反射率差异都比较大。这说明试验中所采集的光谱数据是准确的。在 400~700 nm 波段,光谱反射率较低,并呈现低-高-低的走势,即蓝光波段 430~500 nm 低,黄绿波段 520~590 nm 高,橙红波段 600~680 nm 低^[23]。在 675~730 nm 出现了一个急速的上升沿;在 740~930 nm 波段出现较高的反射平台;在 930~1 040 nm 波段出现吸收谷;在 1 040~1 100 nm 波段再次出现较高的反射平台。在 450 nm 和 670 nm 出现了叶绿素的吸收峰,550 nm 出现了花青素的吸收峰,980 nm 出现了水分的吸收峰。在可见光波段,光谱曲线比较平滑,噪声较少,在短波近红外波段噪声点数较高。通过对原始光谱进行标准正态变换,结果如图 2 所示,其服从标准正态分布,提高了样本光谱的集中性,吸收峰的位置更加的明显。在 500~

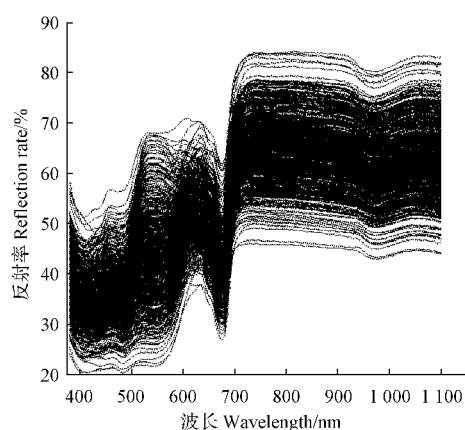


图1 原始光谱图

Fig. 1 The original spectra

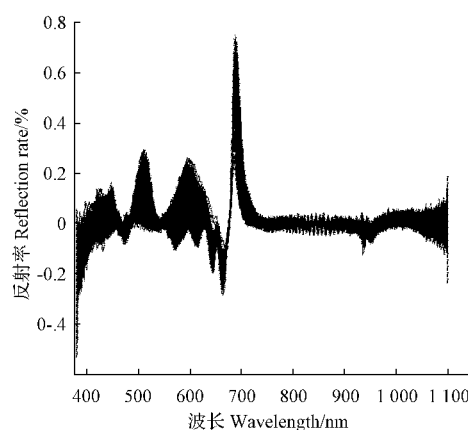


图3 一阶导数处理后的光谱

Fig. 3 Preprocessing spectra by FD

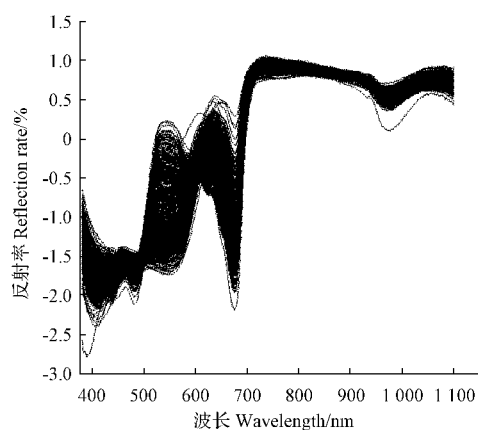


图2 标准正态变换光谱图

Fig. 2 Preprocessing spectra by SNV

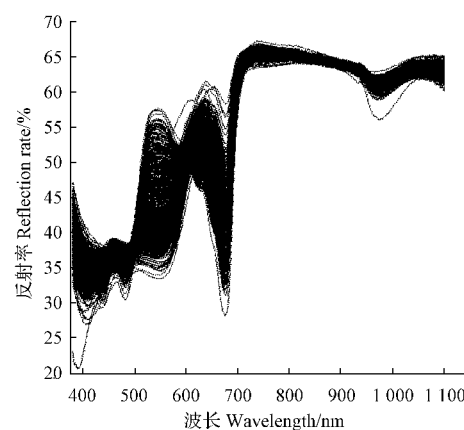


图4 多元散射校正光谱图

Fig. 4 Preprocessing spectra by MSC

600 nm 波段范围内,同时存在反射峰和反射谷,可以通过分析这一波段的特征来对不同成熟度的苹果进行分级。

2.1.2 导数处理 导数处理后的光谱图 3 和原始光谱图 1 相比,吸收峰的个数明显增多。在可见光波段,波峰与波谷之间的差别非常明显,吸收峰的宽度差异水平比较大。在短波近红外波段,吸收峰的个数明显增多,光谱曲线基本上没有平滑连续的特点。

2.1.3 多元散射校正 图 4 为多元散射校正后的光谱图,与原始光谱图相比,明显的抑制了基线的漂移,并且使得光谱的吸收峰位更为突出。

2.2 苹果可溶性固形物含量和可见近红外光谱的模型创建

苹果可溶性固形物含量建模的结果如表 1 所示。经过一阶导数处理和二阶导数处理后进行 PLS 建模得到的校正集相关系数、模型的预测能力都要优于原始光谱直接进行分析的结果,但是经过多元散射校正和标准正态变换后的模型的性能却要比原始光谱低,这说明在

光谱数据的采集过程中启动软件的去除暗噪声、电子暗噪声校正和杂散光校正,对原始光谱曲线的漂移、噪声干扰还是有一定的抑制作用,使用二阶导数光谱进行建模的效果比较好。

表 1 不同预处理方法苹果可溶性固形物含量 PLS 建模结果

Table 1 PLS calibration models for prediction of soluble solids content in apples using different pretreatment methods

预处理方法	相关系数	校正集均方根误差	预测集均方根误差
Preprocessing method	R^2	RMSEC	RMSEP
原始光谱 Original spectrum	0.908 0	0.621 9	0.574 3
多元散射校正 Multi-scatter calibration	0.895 6	0.639 2	0.775 8
标准正态变换 Stand-normal variation	0.898 9	0.630 7	0.737 0
一阶导数 First derivation	0.915 9	0.592 1	0.685 1
二阶导数 Second derivation	0.934 4	0.527 0	0.920 6

3 结论

对比苹果光谱的一阶导数处理、二阶导数处理、标准正态变换、多元散射校正等预处理方法在成熟期苹果可溶性固形物含量预测中的作用。结果表明,这些预处

理方法减弱噪声、突出更多的吸收峰。当使用一阶导数光谱和二阶导数光谱,采用偏最小二乘预测苹果可溶性固形物含量的效果要好于原始光谱所建模型,同时好于标准正态变换光谱和多元散射校正的光谱。该研究为今后果园近红外光谱技术在苹果成熟期多品质检测仪器的开发上奠定一些基础,发挥可见/近红外光谱技术在水果成熟期田间管理、采摘期预测和不同成熟度等级苹果评价的作用。

参考文献

- [1] PEIRS A, LAMMERTYN J, OOMS K, et al. Prediction of the optimal date of different apple cultivars by means of VIS/NIR-spectroscopy[J]. Post-harvest Biology and Technology, 2001, 21(2): 189-194.
- [2] FAN G Q, ZHA J W, DU R, et al. Determination of soluble solids and firmness of apples by VIS/NIR transmittance[J]. J Food Eng, 2009, 93(4): 416-420.
- [3] GEYER M, GUESALAGA A R, AGOSIN E. Non-Destructive evaluation of apple fruit maturity on the tree[J]. Vegetable Crops Research Bulletin, 2007, 66(1): 161-169.
- [4] 赵杰文, 张海东, 刘木华. 利用近红外漫反射光谱技术进行苹果糖度无损检测的研究[J]. 农业工程学报, 2005, 21(3): 162-165.
- [5] 王加华, 汤智辉, 韩东海. 多年份苹果糖度近红外预测模型建立[J]. 食品安全质量检测学报, 2014, 5(3): 742-747.
- [6] 袁雷明, 高海宁, 吕松等. 可见/近红外光谱半透射法测苹果中可溶性固形物含量[J]. 食品安全质量检测学报, 2012, 3(5): 448-452.
- [7] 董一威, 籍保平, 史波林, 等. 苹果中糖酸度的 CCD 近红外光谱分析[J]. 食品科学, 2007, 28(8): 376-380.
- [8] 王铭海, 郭文川, 谷静思, 等. 成熟期梨可溶性固形物含量的近红外漫反射光谱无损检测[J]. 西北农林科技大学学报(自然科学版), 2013, 41(12): 114-119.
- [9] 刘辉军. 田间黄花梨糖度的可见/近红外光谱检测方法[J]. 光谱学与光谱分析, 2015, 35(11): 3078-3081.
- [10] 宫元娟, 周铁, 屈亚堃, 等. 寒富苹果品质无损检测光谱信息在线分析[J]. 沈阳农业大学学报, 2014, 45(6): 708-713.
- [11] 周扬, 戴曙光, 吕进, 等. 光谱预处理对近红外光谱快速检测黄酒酒精度的影响[J]. 光电工程, 2011, 38(4): 54-58.
- [12] 褚小立, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. 化学进展, 2004, 16(4): 528-542.
- [13] 王伟明, 董大明, 郑文刚, 等. 梨果糖浓度近红外漫反射光谱检测的预处理方法研究[J]. 光谱学与光谱分析, 2013, 33(2): 359-360.
- [14] 保罗·戈培林. 化学计量学使用指南[M]. 北京: 科学技术出版社, 2012.
- [15] 李民赞. 光谱分析技术及其应用[M]. 北京: 科学技术出版社, 2006.
- [16] 吴静姝, 李慧. 光谱预处理在农产品近红外模型优化中的应用研究[J]. 农机化研究, 2011, 33(3): 178-181.
- [17] 周丽萍, 胡耀华, 陈达, 等. 苹果可溶性固形物含量的检测方法-基于可见近红外光谱技术[J]. 农机化研究, 2009, 31(4): 104-106.
- [18] 尼珍, 胡昌勤. 近红外光谱分析中光谱预处理方法的作用及其发展[J]. 药物分析杂志, 2008, 28(5): 824-829.
- [19] 唐启义, 冯明光. DPS 处理系统[M]. 北京: 科学技术出版社, 2007.
- [20] 王学民. 应用多元分析[M]. 2 版. 上海: 上海财经大学出版社, 2004.
- [21] 刘秀英, 常庆瑞. 基于可见/近红外光谱的牡丹花青素含量预测[J]. 农业机械学报, 2015, 46(9): 320-323.
- [22] 牛晓颖, 贡东军, 王艳伟, 等. 基于近红外光谱和化学计量学的李果实成熟度鉴别方法研究[J]. 现代食品科技, 2014, 30(12): 230-234.
- [23] 肖冰, 潘存德, 王世伟, 等. 新温 185 号核桃叶片光谱特征及其对施肥的响应[J]. 新疆农业科学, 2015, 51(7): 1205-1212.

Effect of Spectral Data Pretreatment on Visible/Near Infrared Spectroscopy Model of Soluble Solids Content of Apples

LIU Ya¹, GUO Junxian¹, Muhetaer • MIJITI², YUE Jiankui¹, WANG Yan¹, LIU Mingtao¹

(1. Mechanical and Traffic College, Xinjiang Agricultural University, Urumqi, Xinjiang 830052; 2. Vocational and Technical College of Akesu, Akesu, Xinjiang 843000)

Abstract: The soluble solids content is an important index for apple internal quality. 'Fuji' apple was used as test material, visible near infrared reflectance spectra of different growth days of apple was measured, during the apple's ripening stage from enlargement period to harvest period. The original spectrum had pretreated by derivation, standard normal variation transformation and multiplicative scatter correction. Partial least square method was applied to construct the model between soluble solids content and spectrums. The results indicated that the forecast accuracy of the first order derivation and the second order derivation spectrums was higher than that of the original spectrums, the results of the correlation coefficient were 0.915 9, 0.934 4.

Keywords: spectral pretreatment; apple maturity; soluble solids content; visible/near spectroscopy; partial least squares