

多品种洋梨糖度近红外普适性模型的建立

王加华¹, 王一方^{2,3}, 韩东海^{2*}

(1. 许昌学院食品科学与工程学院, 许昌 461000; 2. 中国农业大学食品科学与营养工程学院, 北京 100083;
3. 许昌市食品药品监督管理局, 许昌 461000)

摘要: **目的** 开发多品种洋梨糖度的普适性模型。**方法** 采用主成分得分空间距离将5个品种洋梨分为两组: 阿巴特+康佛伦斯+五九香(组1), 凯斯凯德+红考密斯(组2)。分别建立多品种洋梨SSC的普适性模型, 以 Q 值来评价模型综合性能。**结果** 组1和组2洋梨SSC的普适性模型具有较好性能, 其 $Q = 0.849$ 、 0.735 (PLS)和 0.875 、 0.749 (MLR)。**结论** 多品种洋梨品质的MLR普适性模型可用于便携仪器, 实现现场洋梨SSC精确定量检测。

关键词: 可见/近红外光谱; 主成分得分; 普适性模型; 洋梨; 糖度

Development of general soluble solids content model of European pear (*Pyrus communis* L.) by Vis/NIR transmittance technique

WANG Jia-Hua¹, WANG Yi-Fang^{2,3}, HAN Dong-Hai^{2*}

(1. School of Food Science & Engineering, Xuchang University, Xuchang 461000, China;
2. College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China;
3. Xuchang Food and Drug Administration, Xuchang 461000, China)

ABSTRACT: Objective To develop a general soluble solids content(SSC) model of various European pears. **Methods** According to the space distribution distance of the first three principal components score, five cultivar pears were divided into 'Abate'+ 'Conference'+ 'Wujiuxiang' (group 1) and 'Red Comice'+ 'Cascade' (group 2). **Results** The general modeling for SSC of group 1 and group 2 had perfect performance, with $Q = 0.849$, 0.735 (PLS) and 0.875 , 0.749 (MLR), respectively. **Conclusion** The results highlight the potential of portable Vis/NIR instruments with MLR model for assessing internal quality indices of more cultivars on-site. **KEY WORDS:** Vis/NIR spectroscopy; PCs score; general model; European pear; soluble solids content

1 引言

近红外结合相关化学计量学算法已广泛用于单品种、单采收年份、同一产区水果的品质评价, 对于SSC、酸度、硬度实现了准确的快速检测, 在综述文献中有较多报道^[1-3]。水果品种繁多, 产地分布广, 导致光谱差异性大, 模型适用性受到限制。对于单品种、单地域分别建模, 则模型成本高, 不利于推广使

用。开发多品种或者多地域的普适性模型, 对近红外技术的推广应用具有积极作用。

文献报道, 只有少数学者研究了不同品种水果普适性模型。对布兰基亚和康佛伦斯混合洋梨的品质检测, Paz等^[4]采用3种仪器(FNS-6500, 400~2500 nm; LabSpec Pro A108310, 350~2500 nm; NIR-512, 900~1700 nm)漫反射技术建立SSC和硬度普适性模型, 选

基金项目: 国家自然科学基金项目(31071555)、"十一五"国家科技支撑计划项目(2006BAD05A06-Z012)

*通讯作者: 韩东海, 博士, 教授, 主要从事食品无损检测研究。E-mail: handh@cau.edu.cn

择不同信息区间建立最优模型,其结果分别为:RPD=2.04、1.82和1.38(SSC),RPD=1.65、1.87和1.89。潘璐等^[5]采用漫反射光谱(4000~12000 cm^{-1})建立砂梨系的丰水、圆黄、黄金三个品种 SSC 普适性模型,其结果为RMSEP=0.641% Brix。Camps等^[6]采用漫反射模式采集 Bergarouge、Harostar 和 Kioto 三个品种杏光谱,建立 SSC、滴定酸和硬度的普适性模型,其结果为RMSECV=1.1 %Brix、4.39 g/100 mL 和 10.9 N。

漫反射模式穿透能力弱,光谱受样品表层或表皮影响大。洋梨果皮厚、颜色差和果肉质地差异大,漫透射模式更具优势。选定五个品种洋梨(阿巴特、康佛伦斯、凯斯凯德、红考密斯和五九香),采用光谱相似度方法(主成分得分空间距离)将洋梨分组,分别建立多品种近红外品质评价的普适性模型。模型均采用综合评价参数—Q 值来评价^[7]。

2 材料与方法

2.1 仪器设备

K-BA100R 型便携式近红外光谱仪(日本 Kubota 公司),配备光纤采集附件,采用 CCD 检测器;PAL-1 型手持数字式糖度仪(日本 Atago 公司),读数结果为 %Brix,具备自动温度校正功能。

2.2 实验材料

洋梨(*Pyrus communis* L.)供试样品产自北京大兴区有机梨栽培园,无外部缺陷,着色理想。洋梨包含5个品种:阿巴特‘Abate’,凯斯凯德‘Cascade’,康佛伦斯‘Conference’,红考密斯‘Red Comice’和五九香‘Wujiuxiang’。洋梨采收后,在实验前贮藏于商业气调库(2.5% O_2 , 0.7% CO_2 , $-1\text{ }^\circ\text{C}$, 94~96%RH)。

为了获取一定范围 SSC 含量值,所有的洋梨置于货架上($20\text{ }^\circ\text{C}\pm 2\text{ }^\circ\text{C}$, 70~75%RH),整个试验阶段都保持该贮藏状态。将每个品种随机分为4组,分别第0、2、5、8天进行光谱采集和 SSC 测定,移出冷库当天作为第0天。每个样品在最大直径圈上等距离标定3个测试区域(直径约为40 mm),每个测试区域避免明显缺陷,在同一区域进行光谱采集和 SSC 测定,每个观测点作为一个独立样本。

2.3 基础数据采集

将梨手动放置于光纤头上,标定的采集区域对准光线头,采集漫透射光谱,以内置分光作为背景。

样品和背景积分时间分别为200 ms和50 ms,采集光谱范围为500~1010 nm,间隔为2 nm,共有256个数据点。在每个梨的标示位置测定,累积5次平均作为输出光谱。

参照食品卫生检验方法理化部分总则(GB/T 5009.1—2003),在对应光谱采集面上取直径约40 mm 果肉挤汁测定 SSC 标准值。

3 结果与分析

3.1 光谱预处理

为了放大目标光谱信息,消除高频噪声,原始吸光度光谱采用卷积平滑(Savitsky Golay)二阶导数处理。图1为二阶导数和卷积平滑后的光谱,平滑窗口大小为25,处理后消除了基线漂移,近红外区信息被放大。

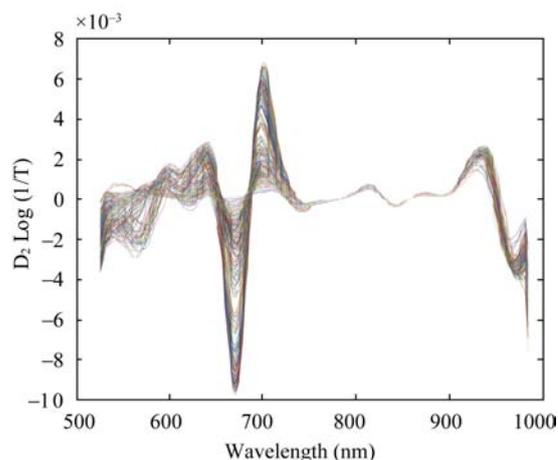


图1 五个品种洋梨经二阶导数和卷积平滑处理后光谱
Fig. 1 Spectra of five cultivars obtained from Savitsky-Golay 2nd order derivation processing data

3.2 空间分布距离

对于多品种、不同地域水果样品,在建立普适性模型时,需要将样品进行归类。洋梨品种多,采用将样品光谱相似度高的归为一组的方法,建立多品种的普适性模型。

本文提出以光谱前3个主成分得分计算空间距离,按照距离大小将5个品种洋梨分为2~3组。采用TQ V7.2分析软件(美国 Thermo Nicolet 公司)分别计算阿巴特、凯斯凯德、康佛伦斯、红考密斯和五九香梨的主成分得分,5种洋梨在前3个主成分的3维分布如图2所示。

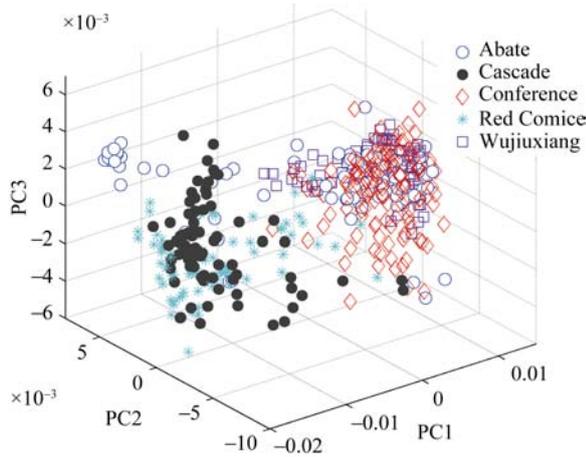


图 2 五个品种洋梨的前 3 个主成分得分分布

Fig. 2 The first three PCs score plots of spectra of all pears

对每个品种洋梨, 主成分得分空间分布距离公式:

$$d = \sqrt{\sum (\hat{X}_{(a)j} - \hat{X}_{(b)j})^2} \quad (j = 1, 2, \dots, K) \quad (1)$$

$$\hat{X}_{(a)} = \frac{1}{m} \sum (X_{(a)i}) \quad (i = 1, 2, \dots, m) \quad (2)$$

其中, $\hat{X}_{(a)j}$ 是品种 a 在第 j 个主成分的平均得分值; $\hat{X}_{(b)j}$ 是品种 b 在第 j 个主成分的平均得分值; K 是主成分数; $X_{(a)i}$ 是 a 品种第 i 个样品的主成分得分; m 是样品数。

5 个品种洋梨的空间分布距离如表 1 所示, 阿巴特和康佛伦斯距离为 0.0055, 康佛伦斯与五九香距离为 0.0016, 阿巴特和五九香距离为 0.0060, 凯斯凯德和红考密斯距离为 0.0073, 而阿巴特、康佛伦斯、

五九香与凯斯凯德、红考密斯两两之间的距离为 0.0153~0.0266。因此, 选定空间阈值(0.01)将 5 种洋梨分为两组: 组 1(阿巴特+康佛伦斯+五九香)和组 2(凯斯凯德+红考密斯)。

3.3 样本分集

在建模前, 在 95% 的置信度下采用 Chauvenet 检验来识别异常光谱。通常以杠杆值和学生残差值平均值的 3 倍作为异常判定阈值。对于组 1(阿巴特+康佛伦斯+五九香)、组 2(凯斯凯德+红考密斯)分别有 5、7 个样本异常, 剔除异常后, 剩余样本按照 K-S 分类法, 将样本分为校正集和预测集, 样本的 SSC 统计结果如表 2 所示。

3.4 多品种洋梨 SSC 普适性模型

3.4.1 组 1(阿巴特+康佛伦斯+五九香)SSC 普适性模型

将组 1(阿巴特+康佛伦斯+五九香)导数处理光谱矩阵和 SSC 矩阵导入 MATLAB V7.0(美国 The mathworks 公司), 依照文献方法^[8]独立进行 GAPLS 程序, 每次输出 0-1 二进制编码字符串, 计算每个波长标识为‘1’的概率。以 RMSECV 为目标函数, 当满足 F 检验($P < 0.1$)时为最佳结果。

运行 GAPLS 程序, 输出结果的 RMSECV 最小为 0.493% Brix, 最佳因子数为 12, 对应的选波段为 654~682 nm、690~764 nm 和 854~892 nm。采用上述因子数和信息区间建立 PLS 模型, 模型对未知样本预测结果如表 3 所示, 其 $R_p = 0.921$ 。

表 1 五种洋梨 3 个主成分下的空间分布距离(组间计算)

Table 1 Space distribution distance of the mean scores of the first three PCs between five cultivars

品种	阿巴特	凯斯凯德	康佛伦斯	红考密斯	五九香
阿巴特	0				
凯斯凯德	0.0211	0			
康佛伦斯	0.0055	0.0257	0		
红考密斯	0.0153	0.0073	0.0203	0	
五九香	0.0060	0.0266	0.0016	0.0211	0

表 2 分组洋梨 SSC 统计结果

Table 2 Statistic results of SSC in different groups

组别	数据集	样品数	范围	平均值	标准偏差	变异系数 (%)
组 1	校正集	212	11.3~18.5	14.86	1.34	9.02
	预测集	71	11.7~18.0	14.15	1.15	8.13
组 2	校正集	143	11.3~17.0	14.16	1.07	7.56
	预测集	42	11.7~16.7	14.56	1.09	7.49

将组 1 导数处理光谱矩阵和 SSC 矩阵导入 TQ Analyst V7.2 软件进行 MLR 变量筛选, 建立 MLR 模型。当入选 6 个变量时, 模型具有最小 RMSECV 值为 0.515 %Brix, 其 MLR 回归方程如下:

$$y = 18.0931 - 245.507x_{528} + 8843.619x_{904} + 128838.8x_{886} - 292.335x_{684} - 91011.3x_{898} - 18905.9x_{806} \quad (3)$$

MLR 模型对预测集样本进行预测结果如表 3 所示, 其 $R_p = 0.948$, $RMSEP = 0.411\%Brix$, $RPD_p = 2.80$ 。

根据认定可接受糖度测试绝对误差为 $\pm 0.8\%Brix$, 分别计算 PLS 和 MLR 模型 Q 值, 其 $Q = 0.849, 0.875$, 均大于 0.80, 可以用于准确定量检测。

3.4.2 组 2(凯斯凯德+红考密斯)SSC 普适性模型

将组 2(凯斯凯德+红考密斯)导数处理光谱矩阵和 SSC 矩阵导入 MATLAB V7.0, 依照文献方法^[8]独立进行 GAPLS 程序, 每次输出 0-1 二进制编码字符串, 计算每个波长标识为‘1’的概率。以 RMSECV 为目标函数, 当满足 F 检验($P < 0.1$)时为最佳结果。

运行 GAPLS 程序, 输出结果的 RMSECV 最小为 0.624%Brix, 最佳因子数为 8, 对应的选波段为 630 nm、652~680 nm、708~760 nm 和 780~892 nm。采用上述因子数和信息区间建立 PLS 模型, 模型对未知样本预测结果如表 3 所示, 其 $R_p = 0.821$, $RMSEP = 0.643\%Brix$, $RPD_p = 1.70$ 。

将组 2 导数处理光谱矩阵和 SSC 矩阵导入 TQ Analyst V7.2 软件进行 MLR 变量筛选, 建立 MLR 模型。当入选 4 个变量时, 模型具有最小 RMSECV 值为 0.597%Brix, 其 MLR 回归方程如下:

$$y = 16.58475 - 24746.3x_{832} + 158020.4x_{882} - 57402.2x_{802} - 84983.7x_{888} \quad (4)$$

MLR 模型对预测集样本进行预测结果如表 3 所示, 其 $R_p = 0.823$, $RMSEP = 0.592\%Brix$, $RPD_p = 1.84$ 。

根据认定可接受糖度测试绝对误差为 $\pm 0.8\%Brix$, 分别计算 PLS 和 MLR 模型 Q 值, 其 $Q = 0.735$ 和 0.749, 根据评价标准, 可以用于定量检测或者定性分级。

表 3 洋梨 SSC(%Brix)普适性模型及性能评价
Table 3 Results of general modeling for SSC (%Brix) of European pear

品种	模型	信息变量(nm)	R_c	RMSEC	RMSEP	Q 值
阿巴特+康佛伦斯+五九香	GAPLS	654~682, 690~764, 854~892, 904	0.940	0.449	0.478	0.849
	MLR	528, 684, 806, 886, 898, 904	0.931	0.469	0.411	0.875
凯斯凯德+红考密斯	GAPLS	630, 652~680, 708~760, 780~892	0.854	0.540	0.643	0.735
	MLR	802, 832, 882, 888	0.857	0.564	0.592	0.749

4 结 论

主成分得分空间距离可用于多品种水果归类分组, 以建立普适性模型。洋梨 SSC 普适性模型有较好的预测性能, MLR 普适性模型可用于便携仪器, 实现 SSC 精确定量检测。

参考文献

[1] Nicolai BM, Beullens K, Bobelyn E, *et al.* Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review [J]. *Postharvest Biol Technol*, 2007, 46: 99-118.

[2] Saranwong S, Kawano S. Applications to agricultural and marine products: fruits and vegetables [M]//Ozaki Y, McClure W F, Christy, A A. *Near- Infrared Spectroscopy in Food Science and Technology*. Hoboken: John Wiley & Sons, 2007: 219-242.

[3] 韩东海, 王加华. 水果内部品质近红外光谱无损检测研究进展[J]. *中国激光*, 2008, 35(8): 1123-1131.

[4] Paz P, Sánchez MT, Pérez-Marín D, *et al.* Instantaneous quantitative and qualitative assessment of pear quality using near infrared spectroscopy [J]. *Comput Electron Agr*, 2009, 69: 24-32.

[5] 潘璐, 王加华, 李鹏飞, 等. 砂梨糖度近红外光谱波段遗传算法优化[J]. *光谱学与光谱分析*, 2009, 29(05): 1246-1250.

- [6] Camps C, Christen D. Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy [J]. LWT-Food Sci Technol, 2009, 42(6): 1125-1131.
- [7] Bossart R, Grabinski J. Automatic Calibration Method: US, 6,480,795 B1 [P]. 2002-11-12.
- [8] 王加华, 陈卓, 李振茹, 等. 洋梨硬度的便携式可见/近红外漫透射检测技术[J]. 农业机械学报, 2010, 41(11): 129-133.

(责任编辑: 张宏梁)

作者简介



王加华, 博士, 主要从事食品无损检测研究。

E-mail: w.jiahua@163.com



韩东海, 博士, 教授, 主要从事食品无损检测研究。

E-mail: handh@cau.edu.cn

“食品包装新材料”专题约稿

当前食品生产与包装研发的联系越来越紧密, 已成为食品安全的重要组成部分, 包装材料与包装技术的应用和研究将越来越重要。各种新型食品包装材料如雨后春笋般涌现出来, 如可食性包装材料、隔热隔水包装材料、可降解包装材料、功能性包装材料等。这些新型包装材料纷纷把安全环保、经济实用当作卖点, 掀起了食品包装的新一轮创新。

鉴于此, 本刊特别策划了“食品包装新材料”专题, 围绕食品包装新材料的研发、制造、应用、安全质量控制等的相关技术和方法等问题展开讨论, 计划在 2013 年出版。编辑部特向各位专家诚征惠稿, 综述、研究论文均可, 以期进一步提升该专题的学术质量和影响力。请在 2012 年 12 月 31 日前通过网站或 Email 投稿。我们将快速处理并优先发表专题论文。

投稿方式:

网站: www.chinafoodj.com

Email: tougao@chinafoodj.com

《食品安全质量检测学报》编辑部