

不同产地苹果糖度可见近红外光谱在线检测

刘燕德 徐海 孙旭东 姜小刚 饶宇 徐佳 王军政

On-line detection of soluble solids content of apples from different origins by visible and near-infrared spectroscopy

LIU Yan-de, XU Hai, SUN Xu-dong, JIANG Xiao-gang, RAO Yu, XU Jia, WANG Jun-zheng

引用本文:

刘燕德, 徐海, 孙旭东, 姜小刚, 饶宇, 徐佳, 王军政. 不同产地苹果糖度可见近红外光谱在线检测[J]. *中国光学*, 2020, 13(3): 482-491. doi: 10.3788/CO.2019-0128

LIU Yan-de, XU Hai, SUN Xu-dong, JIANG Xiao-gang, RAO Yu, XU Jia, WANG Jun-zheng. On-line detection of soluble solids content of apples from different origins by visible and near-infrared spectroscopy[J]. *Chinese Optics*, 2020, 13(3): 482-491. doi: 10.3788/CO.2019-0128

在线阅读 View online: <https://doi.org/10.3788/CO.2019-0128>

您可能感兴趣的其他文章

Articles you may be interested in

近红外光谱法检测乙醇柴油主要性能指标

Detection of key performance indicators of ethanol diesel by the infrared spectroscopy method

中国光学. 2017, 10(3): 363 <https://doi.org/10.3788/CO.20171003.0363>

中红外光谱技术对乙醇汽油乙醇含量的检测

Determination of the content of ethanol in ethanol gasoline using mid-infrared spectroscopy

中国光学. 2017, 10(6): 752 <https://doi.org/10.3788/CO.20171006.0752>

PbSe量子点近红外光源的CH₄气体检测

CH₄ detection based on near-infrared luminescence of PbSe quantum dots

中国光学. 2018, 11(4): 662 <https://doi.org/10.3788/CO.20181104.0662>

近红外光热转换纳米晶研究进展

Research progress of near-infrared photothermal conversion nanocrystals

中国光学. 2017, 10(5): 541 <https://doi.org/10.3788/CO.20171005.0541>

光程补偿近红外光透射反射干涉重构微结构内部形貌

Internal profile reconstruction of microstructures based on near-infrared light transmission reflection interferometry with optical path compensation

中国光学. 2019, 12(2): 395 <https://doi.org/10.3788/CO.20191202.0395>

Na₅[B₂P₃O₁₃]晶体的紫外-远红外光谱分析

Analysis of ultraviolet-far-infrared spectra of Na₅[B₂P₃O₁₃] crystal

中国光学. 2019, 12(5): 1118 <https://doi.org/10.3788/CO.20191205.1118>

不同产地苹果糖度可见近红外光谱在线检测

刘燕德^{1,2*}, 徐海^{1,2}, 孙旭东^{1,2}, 姜小刚^{1,2}, 饶宇^{1,2}, 徐佳^{1,2}, 王军政^{1,2}

(1. 华东交通大学机电与车辆工程学院, 江西南昌 330013;

2. 水果智能光电检测技术与装备国家地方联合工程研究中心, 江西南昌 330013)

摘要: 为了实现不同产地苹果糖度的快速在线无损检测, 减少产地差异对近红外光谱检测模型的影响, 建立了不同产地苹果糖度的在线检测通用模型。首先, 采用水果动态在线检测设备采集了包括栖霞、洛川与会宁 3 个产地的红富士苹果的漫透射光谱。其次, 采用偏最小二乘算法(PLS), 结合无信息变量消除(UVE)方法, 筛选出 58 个特征变量, 建立了苹果糖度的 UVE-PLS 通用模型, 该模型对个体产地预测集及总预测集的均方根误差分别为 0.50~0.74°Brix 与 0.63°Brix, 较原始个体模型分别提高了 23.2%~44.4% 与 35.7%。最后, 提出了一个新的外部验证样本集对模型性能进行评价, 其残留预测偏差为 2.33, 预测值在 $\pm 1.0^\circ\text{Brix}$ 和 $\pm 1.5^\circ\text{Brix}$ 误差范围内的占比分别为 85% 与 100%。实验结果表明: 建立多个产地苹果糖度的在线检测通用模型, 能够提高其他产地样本糖度的预测稳健性, 并且采用合适的波长筛选方法能够简化模型。开发不同产地水果内部品质通用模型在波长有限的光谱设备中具有良好的应用潜力。

关键词: 在线检测; 近红外光谱; 糖度; 偏最小二乘

中图分类号: O657.33 文献标志码: A doi: 10.3788/CO.2019-0128

On-line detection of soluble solids content of apples from different origins by visible and near-infrared spectroscopy

LIU Yan-de^{1,2*}, XU Hai^{1,2}, SUN Xu-dong^{1,2}, JIANG Xiao-gang^{1,2}, RAO Yu^{1,2}, XU Jia^{1,2}, WANG Jun-zheng^{1,2}

(1. School of Mechatronics & Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China;

2. National and Local Joint Engineering Research Center of Fruit Intelligent Photoelectric

Detection Technology and Equipment, Nanchang 330013, China)

* Corresponding author, E-mail: jxliuyd@163.com

Abstract: In order to realize fast, on-line, non-destructive testing of the Soluble Solids Content (SSC) of apples from different origins, and to reduce the effect of origin variability on NIR models, a universal model for predicting the SSC of apples from different origins is established. Firstly, the diffuse transmission spectra of Fuji apples from Qixia, Luochuan and Huining are collected with fruit dynamic online detection equipment. Then, 58 characteristic variables are selected and a UVE-PLS universal model for predicting the SSC

收稿日期: 2019-06-21; 修订日期: 2019-08-20

基金项目: 国家自然科学基金(No.31760344); 江西省创新能力建设项目(No.S2016-90)

Supported by National Natural Science Foundation of China(No.31760344); Jiangxi Provincial Project for Innovation Capacity Construction(No.S2016-90)

of apples is established using the Partial Least Squares (PLS) algorithm combined with Uninformative Variable Elimination (UVE). The root mean square errors of single-origin prediction sets and the total-origin prediction set are $0.50\sim 0.74^\circ$ Brix and 0.63° Brix, respectively, which increase by 23.2%~44.4% and 35.7% respectively compared to the original individual model. Finally, a new external sample set is used to assess the performance of the model, showing a residual prediction deviation of 2.33 and ratios of the predicted values within the error range of $\pm 1.0^\circ$ Brix and $\pm 1.5^\circ$ Brix of 85% and 100%, respectively. Experimental results indicate that the establishment of a universal model for on-line detection of the SSC of apples, including those from multiple origins can improve the robustness of predicting the SSC of the samples from other origins. The results also show that an appropriate wavelength screening method can simplify the model. The development of a common model for the internal quality of fruit from different origins has strong potential for applications in wavelength-limited spectroscopy equipment.

Key words: online detection; near infrared spectroscopy; soluble solids; partial least squares

1 引言

苹果是一种被广泛生产及消费的水果品类,其富含抗氧化成分,如抗坏血酸和多酚类化合物,可以对各种退化性疾病起到预防作用^[1-2]。苹果的外观和内部品质直接决定消费者是否购买。可溶性固形物含量(Soluble Solids Content, SSC),一般也称为糖度,是评价苹果内部品质的关键参数,也是集可溶性糖、酸、纤维素等成分在内的综合指标^[3]。因此,开发一种快速可靠的SSC检测方法至关重要。近年来,具有快速、准确、易于实现等优点的近红外(Near Infrared, NIR)光谱无损检测技术在水果品质评定方面得到了广泛应用^[4-8]。

由于土壤养分、光照条件、气候以及果园生产管理方式等各不相同,导致水果的成分和品质也不同^[9-10]。此外,这种生物变异性也会影响光学传播特性及其相互作用,从而降低水果内、外部品质检测的准确性^[11]。经过调研发现,关于消除产地差异对SSC检测模型准确性影响的报道很少。Fan^[12]等通过混合4个产地苹果的校正集建立了混合模型,结果表明,混合产地模型在提升苹果SSC预测模型稳健性方面具有较好的效果。Li等^[13]建立了富士苹果SSC的多产地回归测量模型,结果表明,与单一产地模型相比,结合产地判别法和模型搜索方法建立的多产地模型对苹

果SSC的预测具有更好的预测结果及鲁棒性。然而,这些研究都是基于傅立叶变换近红外光谱仪进行的,光谱采集都为静态采集。对于水果实际分选环节,建立基于动态在线检测设备的SSC检测模型尤为关键。因此,本研究以产自山东栖霞、陕西洛川与甘肃会宁的红富士苹果为实验对象,利用团队自主研发的水果动态在线检测设备采集样品光谱,建立单一产地及混合产地苹果糖度的在线检测模型,并采用合适的波长筛选方法筛选出特征波长变量,尝试在保持模型预测精度及稳定性的基础上简化模型。本文研究结果将对苹果SSC的准确检测具有重要应用价值。

2 材料与方 法

2.1 实验样本

实验样品于2018年11月购自3个不同的富士苹果商业果园,经过手工分拣后共选出132个无明显缺陷与损伤的苹果样品。其中44个来自山东栖霞,45个来自陕西洛川,43个来自甘肃会宁。将所有样品单独洗涤,干燥,编号,然后在赤道周围标记取样点,各点相距 90° 并垂直于茎轴,且每个观测点均作为一个独立样本。在光谱采集前,将苹果置于实验室环境(温度: 22°C ;相对湿度: 60%)下储存24小时以使样品达到室温,降低样品温度对预测精度的影响^[14]。每个批次使用15~20个果实,并进行定期评估。

2.2 水果动态在线检测设备与光谱采集

本文所用漫透射检测机构如图 1 所示。考虑光源穿透性及水果糖度的光谱范围,以保证光谱仪接收到足够的光谱信息,考虑光照强度以确保光源不会对果皮造成灼伤,故设备采用 10 盏 12 V, 100 W 的卤素灯,以圆弧状配置方式对称排列在待测样品两侧,每盏灯与样品垂直呈 45°放置,卤素灯发出的光均匀照射样品表面,透过样品内部的光由果杯下方的光纤探头收集后传回光谱仪,这样可以得到更多的样品果肉信息,从而避免了果实表面反射光及外部光的干扰,同时能有效消除由于照射果实不同部位导致的糖、酸等成分的测定误差。采用 QE65Pro 光纤光谱仪(Ocean optics, INC., USA)测量,波长范围为 350~1 100 nm。

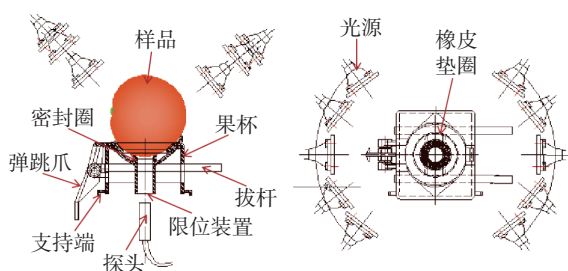


图 1 漫透射检测机构

Fig. 1 Mechanism of diffuse transmission detection



图 2 光谱采集触发装置

Fig. 2 Trigger device of spectral acquisition

实验前,设备预热 30 min,设置运行速度为 5 个/s,以保证样品平稳运行。以聚四氟乙烯球作为参比,校正参比后采集光谱。采用配套的 SpectraSuite 软件采集样品光谱,在光谱采集时设置积分时间为 100 ms,并采用硬件触发方式采集光谱,装置实物图如图 2 所示。齿轮每隔 4 个齿安装一个果杯,果杯和分度盘齿一一对应。分度盘下方 2 mm 处安装有光电开关,每转一齿,即传动

一个果杯的行程,触发传感器,使后续电路发出信号,从而触发光谱仪采集并保存一条光谱信息。

2.3 SSC 测定

光谱采集完成后,使用温度补偿糖度计(型号 PAL-1; Atago Co., Tokyo, Japan)测定样品的 SSC。从每个样品相应测量位置取出一个圆形楔形物(最大直径为 50 mm),通过粗棉布压制挤汁滴于糖度计镜面上,并记录读数。对于同一测量部位,重复测量 3 次,取其平均值作为 SSC 参考值。

2.4 偏最小二乘回归

偏最小二乘(PLS)回归是一种被广泛使用的无监督统计方法,用于在 NIR 分析中建立校正模型^[15]。PLS 分析的目的是通过将光谱矩阵 X 与浓度矩阵 Y 投影到新的潜在空间,使其中的潜在变量(Latent Variables, LVs)之间的协方差最大化,最终执行潜变量线性回归模型,其基本公式如下:

$$Y = Xb + e, \quad (1)$$

式中, b 表示回归系数的向量, e 表示模型残差。

在校正与验证过程中,将 LVs 作为新的特征向量来解释方差和降低原始光谱维数^[3]。在本研究中,PLS 回归用于在光谱数据和苹果 SSC 属性之间建立定量关系,其中, LVs 的最佳数量由基于校正样本的 10 倍交叉验证最低均方根误差确定。

PLS 广泛用于建立近红外光谱检测模型,具有独特的优越性^[16]:

(1)可灵活使用光谱矩阵,根据需要全部或者部分数据达到很好的建模效果;

(2)同时进行数据分解与回归,使特征矢量与被测成分有效融合;

(3)适用于处理样品数量少,但变量相对多的问题;

(4)PLS 是多元线性回归与主成分分析的完美结合,显著提高了校正模型的预测能力。

通过校正集的决定系数(R_c^2)与均方根误差(RMSEC)来评价 SSC 与光谱数据之间的校正方程。通过预测集的决定系数(R_p^2)与均方根误差(RMSEP)来评价校正模型的预测精度。利用残留预测偏差(Residual Prediction Deviation, RPD)来判断模型的质量。上述统计指标的计算公式

如下:

$$R_c^2, R_p^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_m)^2}, \quad (2)$$

$$RMSEC, RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (3)$$

$$RPD = \frac{SD}{RMSEP} = \frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - y_m)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}, \quad (4)$$

式(2)~式(4)中, n 表示校正集或预测集中的样本数, y_i 和 \hat{y}_i 分别表示校正集和预测集中第 i 个样本的测量值和预测值, y_m 表示样本集所有测量值的平均值。

RPD 值越大表明模型质量越好, Chang 等^[17] 定义了 3 个质量类别以考虑模型的可靠性, 它们分别是: 优秀模型, RPD 大于 2; 可靠模型, RPD

为 1.4~2; 不可靠模型, RPD 小于 1.4。

3 结果与讨论

3.1 样本集的划分

对所建模型进行校正之前, 对样本进行肖维勒准则测试。其中, 对于栖霞, 洛川与会宁 3 个产地的样本, 各剔除异常样本 4、4 和 5 个, 共计 13 个。将栖霞、洛川与会宁依次定义为产地 1、2 与 3, 每个产地的剩余样本通过 Kennard-Stone (K-S) 法按约 3 : 1 的比例分成 2 组, 分别作为校正集与预测集。

全部剩余样本的校正集与预测集分别由各个产地的校正子集和预测子集组成。校正集仅用于构建校正模型, 预测集仅用于决定模型参数, 以评估校正模型。样本集 SSC 统计结果见表 1。如表 1 所示, 校正集与预测集的 SSC 范围分别为 8.5~18.2°Brix 与 8.9~17.6°Brix, 标准差分别为 1.61°Brix 与 1.64°Brix, 样本集覆盖的范围足够大, 且校正集大于每个产地的预测集, 这些特征有利于建立稳健的近红外模型。

表 1 样本集糖度含量统计

Tab. 1 Statistical values of the SSC (°Brix) for sample sets

产地	校正集				预测集			
	数量	范围	平均值	标准差	数量	范围	平均值	标准差
1	129	8.8~16.6	12.93	1.43	43	9.3~15.5	12.47	1.24
2	135	8.5~16.4	13	1.28	41	8.9~15.1	12.93	1.23
3	127	10.1~18.2	14.97	1.21	40	11.7~17.6	15.03	1.17
总	391	8.5~18.2	13.62	1.61	124	8.9~17.6	13.45	1.64

3.2 漫透射光谱响应特征

由于光谱下部 (350~594 nm) 与上部 (900~1100 nm) 的光谱能量值较低, 主要包含的是无效信息, 故将有效波长范围定义为 594~900 nm (400 个数据点)。由图 3 (彩图见期刊电子版) 可知, 3 个产地的平均原始光谱曲线走势相似, 但也存在一定的差异: 在 600~700 nm 波段, 光谱连续上升至 645 nm 处对应的波峰, 接着下降至 678 nm 处对应的波谷, 这可能与叶绿素吸收有关^[18]; 在

700~900 nm 波段, 光谱急剧上升至 705 nm 处对应的波峰, 接着又下降至 740~780 nm 对应的宽阔区域, 该区域光谱的吸收变化主要受 O-H 的三倍频伸缩振动与 C-H 的四倍频伸缩振动影响^[19], 其中会宁样本光谱能量值较高, 其原因可能与其果型较小有关, 曲线最终从 810 nm 附近继续下降直至结束。经过比较分析可知, 使用多元散射校正 (MSC) 的光谱预处理方法可以减小因颗粒分布不均及颗粒大小不同产生的散射影响。

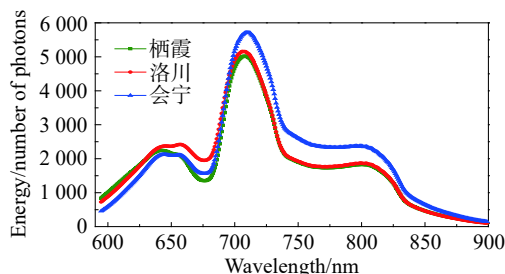


图 3 样品原始光谱

Fig. 3 Samples' original spectra

MSC 主要是通过一组样本的光谱, 基于统计的方法来修正光谱因散射所产生的线性变化。该方法假定散射系数在所有波长处均相同, 将化学物质的吸收信息和光谱中的散射光信号进行有效分离, 其校正过程^[16]如下:

首先计算所需校正光谱的平均光谱, 如式(5)所示:

$$\bar{X}_{i,j} = \frac{\sum_{i=1}^n X_{i,j}}{n}, \quad (5)$$

对平均光谱进行回归, 如式(6)所示:

$$X_i = m_i \bar{X} + b_i, \quad (6)$$

最后对每条光谱作多元散射校正, 如式(7)所示:

$$X_{i(MSC)} = \frac{(X_i - b_i)}{m_i}, \quad (7)$$

式(5)~式(7)中, X 为 $n \times p$ 维光谱矩阵, n 表示样本数, p 表示波长点数, \bar{X} 表示所有样品的原始近红外光谱在各个波长点处求平均所得的平均光谱, X_i 为 $1 \times p$ 维向量, 表示单个样品光谱矢量, m_i 表示一元线性回归得到的相对偏移系数, b_i 表示偏移量。

对处理后的光谱进行主成分分析(PCA)以判断 3 个产地光谱的差异性, 通过光谱的前 3 个主成分得分计算样本的空间距离(前 3 个主成分已表征了 95% 的原始信息), 最终的三维得分分布图如图 4 所示。由图 4 可知, 3 个产地样本的主成分空间分布具有较大的重合区, 光谱相似性较高, 对不同产地具有通用性。

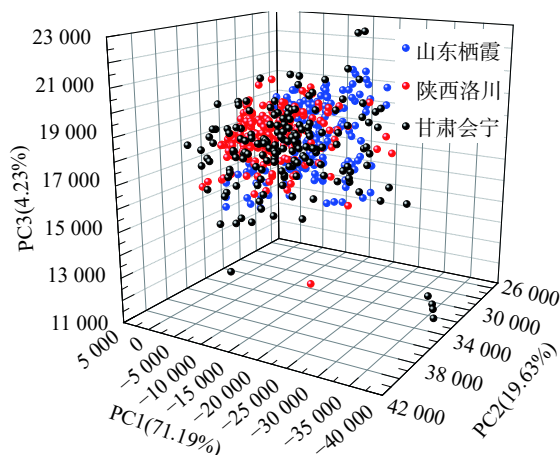


图 4 经过 MSC 的苹果光谱的前 3 个主成分得分图

Fig. 4 The first three PC score plot of apple spectra after MSC processing

3.3 不同产地的 PLS 模型

对 3 个产地分别进行 PLS 建模, 并对其进行交互预测, 即用单个产地的校正集去预测其余产地的预测集。以栖霞为例, 将栖霞的校正集作为校正模型, 分别去预测洛川, 会宁以及包含所有产地的预测集, 其他产地预测方式类似, 最终结果如表 2、表 3 所示。

表 2 单个产地的 PLS 建模结果

Tab. 2 Results of PLS modeling for single origin

产地	LVs	校正集		预测集		RPD
		R_c^2	RMSEC(°Brix)	R_p^2	RMSEP(°Brix)	
1	9	0.92	0.40	0.90	0.41	3.02
2	10	0.89	0.42	0.85	0.47	2.62
3	11	0.86	0.46	0.80	0.51	2.29

表 2 显示了单个产地校正模型预测出的相同产地果实 SSC 的结果。在校正集与预测集的产地相同时, R_p^2 及 RMSEP 的范围分别为 0.80~0.90 与 0.41~0.51°Brix, 且 RPD 为 2.29~3.02。结果表明: 使用原始个体模型在预测同一产地果实糖度时表现优秀。而由表 3 可知, 当使用原始单个产地校正模型预测其他个体产地果实糖度时, 其结果并不理想, 其 RMSEP 较高 (0.82~1.44°Brix)。而对于预测包含 3 个产地的总预测集, 个体校正模型的预测结果同样不佳, RMSEP 为 0.98~1.30°Brix。这可能是由于当使用个

体模型学习校正集数据中的细节和噪声时, 模型对其他产地数据的预测性能产生了负面影响, 从

而导致过度拟合。显然, 只用原始个体产地建立的模型并不能很好地预测其他产地的 SSC 值。

表 3 不同产地红富士苹果的预测结果

Tab. 3 Prediction results of Fuji apples from different origins

产地	预测集							
	1		2		3		总	
	R_p^2	RMSEP(°Brix)	R_p^2	RMSEP(°Brix)	R_p^2	RMSEP(°Brix)	R_p^2	RMSEP(°Brix)
1	/	/	0.54	0.82	0.73	1.24	0.42	1.30
2	0.54	0.90	/	/	0.67	1.34	0.67	0.98
3	0.73	1.44	0.72	1.25	/	/	0.68	1.27

3.4 通用模型的建立及简化

为了提升对其他产地数据的预测效果, 采用 PLS 回归算法建立混合 3 个产地的通用模型, 用于预测所有产地的 SSC 值。为了消除不包含有效信息的变量, 采用基于 PLS 回归系数的无信息变量消除法 (Uninformative Variable Elimination, UVE) 筛选出特征变量。其基本思想是把变量对应的回归系数平均值与其标准差的比值作为变量选择的衡量依据^[20], 其表达式如下:

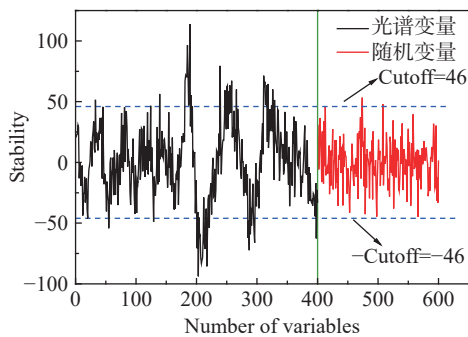
$$C_i = \frac{\text{mean}(b_i)}{\text{Std}(b_i)}, \quad (8)$$

式中, C 为回归系数平均值与标准差之比, 表示变

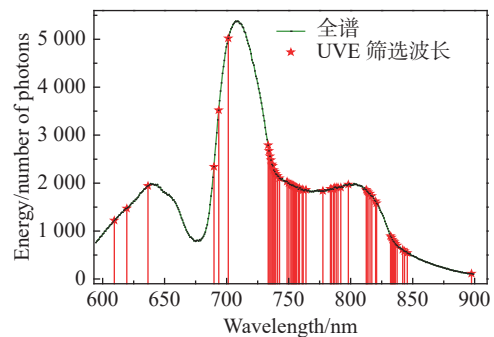
量的稳定性。

最终的判断方法是将一定数量的随机变量矩阵加入光谱矩阵中, 然后通过交叉验证建立 PLS 模型, 进而得到回归系数矩阵, 由 $|C_i|$ 的大小确定是否将第 i 列变量用于 PLS 建模, 当 $|C_i|$ 低于随机变量稳定性的 99% 时, 该列变量被剔除。

采用 UVE 对经过预处理的全谱变量 (400 个数据点) 进行筛选, 控制随机变量的大小, 使其为原变量的 10^{-10} , 以保证扩展矩阵的特征值基本不变, 从而最大程度地降低随机矩阵对原变量回归系数的影响。随机变量个数设为 200, 最终的处理结果如图 5 所示。



(a) UVE 筛选变量的稳定性分布
(a) Stability distribution of variables selected by UVE



(b) UVE 选中的 58 个波长分布图
(b) Distribution plot of 58 wavelengths selected by UVE

图 5 UVE 选择变量结果

Fig. 5 Results of variables selected by UVE

图 5(a) 中, 竖线为光谱变量与随机变量的分界线, 其左侧曲线表示光谱变量的稳定性分布情况, 右侧曲线表示随机变量的稳定性分布情况, 2 条水平虚线表示算法所得阈值的上下限。虚线

之间的变量为被剔除的变量, 而虚线外的变量即为选中的光谱变量。最终经过 UVE 筛选出的剩余光谱变量个数为 58, 图 5(b) 显示了经过 UVE 处理后选中变量在全谱变量中的位置分布。不难

看出,选中变量主要集中在由含氢基团伸缩振动所引起的吸收光谱区间,模型的预测结果如表 4 所示。

表 4 苹果糖度通用模型预测结果

Tab. 4 Results of SSC of apples predicted by universal modeling

模型	变量数	LVs	R_c^2	RMSEC($^{\circ}$ Brix)	R_p^2	RMSEP($^{\circ}$ Brix)	RPD
Ori-PLS	400	12	0.84	0.64	0.85	0.63	2.60
UVE-PLS	58	8	0.82	0.68	0.85	0.63	2.60

由表 4 可知,将仅经过预处理的原始混合产地模型及经过 UVE 筛选变量后的通用模型分别记为 Ori-PLS 与 UVE-PLS。其中, Ori-PLS 模

型的 RMSEC 与 RMSEP 差值不大,分别为 0.64° Brix 与 0.63° Brix,建模所用变量数及 LVs 数分别高达 400 与 12。而 UVE-PLS 通用模型所用建模变量数及 LVs 数仅为 58 与 8,这极大地提升了模型的运算速率,降低了模型的复杂程度。但是 UVE-PLS 模型的 R_c^2 及 RMSEC 均略差于 Ori-PLS 模型,这可能与算法在计算过程中遗漏了某些关键变量有关。

UVE-PLS 通用模型能够预测包含 3 个产地预测集的 SSC 值(图 6(a)),其中 RMSEP 为 0.63° Brix,偏差为 0.03° Brix。而图 6(b)~6(d)显示了每个产地果实 SSC 值预测的散点图。对于每个产地的预测集,模型的糖度预测值位于目标线附近,且偏差较小。

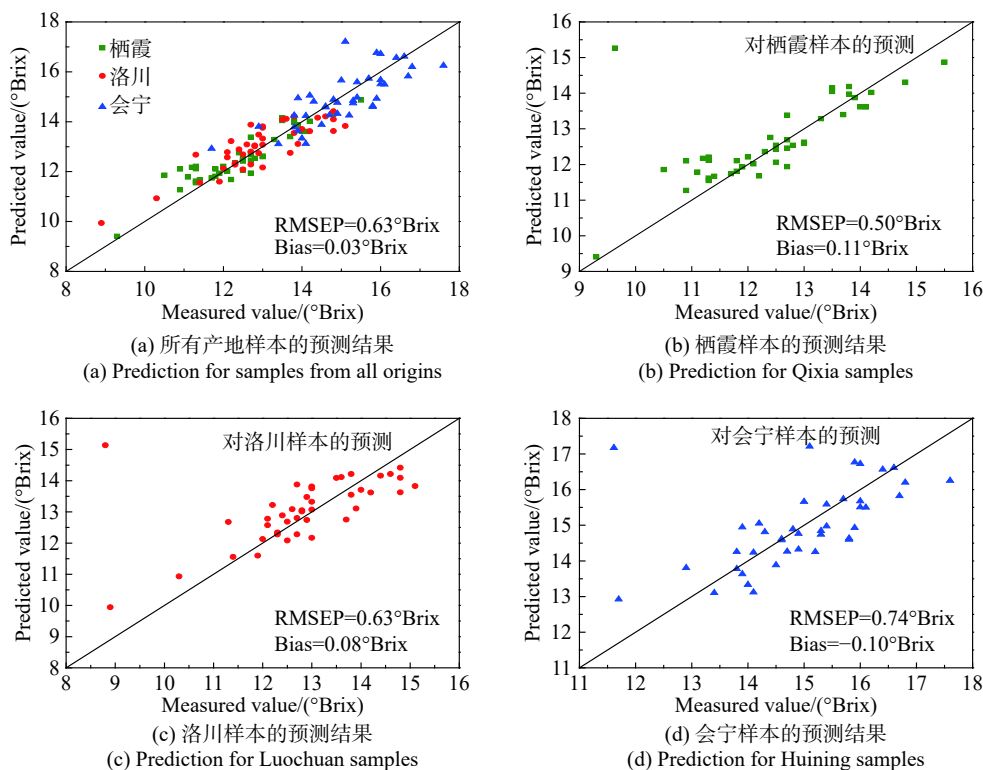


图 6 UVE-PLS 通用模型糖度测量值与预测值的散点图

Fig. 6 Scatter plots of measured values versus predicted values for SSC by using UVE-PLS universal model

对比表 3 相关数据可知,在预测其他产地预测集样本时,相较于原始单个产地校正模型,UVE-PLS 通用模型的 RMSEP 均有所降低,与交互预测中的较好结果 ($0.90, 0.82, 1.24$ 与 0.98° Brix) 相比,预测精度分别提高了 44.4%, 23.2%, 40.3% 与 35.7%。

3.5 外部验证

本文提出了一个新的样本集用来验证 UVE-PLS 方法所获得的 SSC 模型的实际性能。从当地超市购买了 30 个苹果,每个产地各 10 个。在每个苹果上选择两个测量点,在样品的赤道处随机标记,间隔约 180° 。对新样本采集光谱及测得

SSC 真值后,放入模型中进行验证,每个产地与总样本的预测误差列于表5。

表5 UVE-PLS 糖度通用模型的实际性能

Tab.5 The practical performance of UVE-PLS universal model for SSC

	RMSEP/(°Brix)			RPD	
	栖霞	洛川	会宁	总	总
	0.67	0.51	0.72	0.64	2.33

由表5可知,对新样本集中每个产地的SSC进行预测,其RMSEP值为0.51~0.72°Brix,而图7(彩图见期刊电子版)显示了SSC预测值与测量值之间的绝对误差值的分布情况。

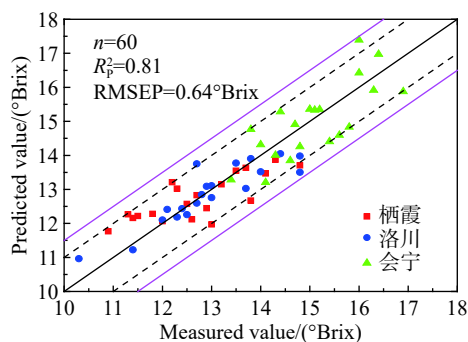


图7 糖度测量值与预测值的散点图

Fig.7 Scatter plot of measured value versus predicted value of SSC for a new external sample set

如图7所示,对于总样本组的SSC预测,其 R_p^2 和RMSEP分别为0.81和0.64°Brix。中间的黑色实线为目标线,即 $y=x$,两条黑色虚线表示预

参考文献:

- [1] GIOVANELLI G, SINELLI N, BEGHI R, *et al.*. NIR spectroscopy for the optimization of postharvest apple management[J]. *Postharvest Biology and Technology*, 2014, 87: 13-20.
 - [2] FAN SH X, ZHANG B H, LI J B, *et al.*. Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple[J]. *Biosystems Engineering*, 2016, 143: 9-19.
 - [3] MENDOZA F, LU R F, ARIANA D, *et al.*. Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content[J]. *Postharvest Biology and Technology*, 2011, 62(2): 149-160.
 - [4] 高升, 王巧华, 李庆旭, 等. 基于近红外光谱的红提维生素C含量、糖度及总酸含量无损检测方法[J]. *分析化学*, 2019, 47(6): 941-949.
- GAO SH, WANG Q H, LI Q X, *et al.*. Non-destructive detection of vitamin c, sugar content and total acidity of red globe grape based on near-infrared spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2019, 47(6): 941-

测偏差为 $\pm 1.0^\circ\text{Brix}$ 的分界线,而两条品红色实线表示预测偏差为 $\pm 1.5^\circ\text{Brix}$ 的分界线。经过计算,在 $\pm 1.0^\circ\text{Brix}$ 和 $\pm 1.5^\circ\text{Brix}$ 偏差范围内的预测值占比分别为85%与100%,而RPD为2.33,因此,使用UVE-PLS方法建立的通用模型更适用于SSC的定量检测。

4 结论

本研究采用多元散射校正方法,有效减小了由产地变异性引起的光谱差异。单个产地糖度校正模型在预测同一产地数据时显示出较好的预测结果,RMSEP为0.41~0.51°Brix,但当预测集包含其他产地时,相关性则不显著,RMSEP较高,为0.98~1.30°Brix。采用PLS回归算法建立了不同产地苹果的糖度通用模型,结合UVE方法筛选出58个特征波长变量,在保证预测精度的基础上,简化了模型,其RMSEP为0.63°Brix。与单一产地校正模型相比,UVE-PLS通用模型在预测每个产地糖度时,预测精度大大提升,表明随着更多的生物变异性被考虑进来,预测精度对外部因素的未知变化变得不那么敏感,使得校正模型对未来的变化更加稳健。外部验证结果表明,UVE-PLS通用模型对3种产地苹果糖度均具有出色的预测能力。实验结果表明,水果动态在线检测设备可以用于开发不同产地苹果的糖度模型,本研究简化了水果商品化加工过程,降低了建立糖度分选模型的维护成本,对水果采后商品化处理具有指导意义。

949. (in Chinese)
- [5] 史云颖, 李敬岩, 褚小立. 多元校正模型传递方法的进展与应用[J]. 分析化学, 2019, 47(4): 479-487.
SHI Y Y, LI J Y, CHU X L. Progress and applications of multivariate calibration model transfer methods[J]. *Chinese Journal of Analytical Chemistry*, 2019, 47(4): 479-487. (in Chinese)
- [6] 王凡, 李永玉, 彭彦昆, 等. 基于可见/近红外透射光谱的番茄红素含量无损检测方法研究[J]. 分析化学, 2018, 46(9): 1424-1431.
WANG F, LI Y Y, PENG Y K, *et al.*. Nondestructive determination of lycopene content based on visible/near infrared transmission spectrum[J]. *Chinese Journal of Analytical Chemistry*, 2018, 46(9): 1424-1431. (in Chinese)
- [7] 路皓翔, 徐明昌, 张卫东, 等. 基于压缩自编码融合极限学习机的柑橘黄龙病鉴别方法[J]. 分析化学, 2019, 47(5): 652-660.
LU H X, XU M CH, ZHANG W D, *et al.*. Identification of citrus huanglongbing based on contractive auto-encoder combined extreme learning machine[J]. *Chinese Journal of Analytical Chemistry*, 2019, 47(5): 652-660. (in Chinese)
- [8] 郭文川, 王铭海, 谷静思, 等. 近红外光谱结合极限学习机识别贮藏期的损伤猕猴桃[J]. 光学精密工程, 2013, 21(10): 2720-2727.
GUO W CH, WANG M H, GU J S, *et al.*. Identification of bruised kiwifruits during storage by near infrared spectroscopy and extreme learning machine[J]. *Optics and Precision Engineering*, 2013, 21(10): 2720-2727. (in Chinese)
- [9] 郭志明, 黄文倩, 彭彦昆, 等. 自适应蚁群优化算法的近红外光谱特征波长选择方法[J]. 分析化学, 2014, 42(4): 513-518.
GUO ZH M, HUANG W Q, PENG Y K, *et al.*. Adaptive ant colony optimization approach to characteristic wavelength selection of NIR spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2014, 42(4): 513-518. (in Chinese)
- [10] ZHANG B H, HUANG W Q, GONG L, *et al.*. Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier[J]. *Journal of Food Engineering*, 2015, 146: 143-151.
- [11] ZHANG B H, DAI D J, HUANG J CH, *et al.*. Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: a review[J]. *Critical Reviews in Food Science and Nutrition*, 2018, 58(12): 2099-2118.
- [12] 樊书祥, 黄文倩, 郭志明, 等. 苹果产地差异对可溶性固形物近红外光谱检测模型影响的研究[J]. 分析化学, 2015, 43(2): 239-244.
FAN SH X, HUANG W Q, GUO ZH M, *et al.*. Assessment of influence of origin variability on robustness of near infrared models for soluble solid content of apples[J]. *Chinese Journal of Analytical Chemistry*, 2015, 43(2): 239-244. (in Chinese)
- [13] LI X N, HUANG J CH, XIONG Y J, *et al.*. Determination of soluble solid content in multi-origin 'Fuji' apples by using FT-NIR spectroscopy and an origin discriminant strategy[J]. *Computers and Electronics in Agriculture*, 2018, 155: 23-31.
- [14] JANNOK P, KAMITANI Y, HIRONAKA K, *et al.*. Development of a near infrared calibration model with temperature compensation using common temperature-difference spectra for determining the Brix value of intact fruits[J]. *Journal of Near Infrared Spectroscopy*, 2017, 25(1): 26-35.
- [15] 王拓, 戴连奎, 马万武. 拉曼光谱结合后向间隔偏最小二乘法用于调和汽油辛烷值定量分析[J]. 分析化学, 2018, 46(4): 623-629.
WANG T, DAI L K, MA W W. Quantitative analysis of blended gasoline octane number using raman spectroscopy with backward interval partial least squares method[J]. *Chinese Journal of Analytical Chemistry*, 2018, 46(4): 623-629. (in Chinese)
- [16] 刘翠玲, 吴静珠, 孙晓荣. 近红外光谱技术在食品品质检测方法中的研究[M]. 北京: 机械工业出版社, 2016.
LIU C L, WU J ZH, SUN X R. *Study on Near Infrared Spectroscopy in Food Quality Testing Methods*[M]. Beijing: China Machine Press, 2016. (in Chinese)
- [17] CHANG CH W, LAIRD D A, MAUSBACH M J, *et al.*. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties[J]. *Soil Science Society of America Journal*, 2001, 65(2): 480-490.

- [18] ZHANG D Y, XU L, WANG Q Y, *et al.*. The optimal local model selection for robust and fast evaluation of soluble solid content in melon with thick peel and large size by Vis-NIR spectroscopy[J]. *Food Analytical Methods*, 2019, 12(1): 136-147.
- [19] YUAN L M, CAI J R, SUN L, *et al.*. Nondestructive measurement of soluble solids content in apples by a portable fruit analyzer[J]. *Food Analytical Methods*, 2016, 9(3): 785-794.
- [20] YUN Y H, LI H D, DENG B CH, *et al.*. An overview of variable selection methods in multivariate analysis of near-infrared spectra[J]. *TrAC Trends in Analytical Chemistry*, 2019, 113: 102-115.

作者简介:



刘燕德(1967—),女,江西泰和人,博士,教授,博士生导师,1990年、2001年于江西农业大学分别获得学士、硕士学位,2006年于浙江大学获得博士学位,主要从事光电检测技术与装备方面的研究。E-mail: jxliuyd@163.com

《发光学报》

—EI 核心期刊(物理学类;无线电电子学、电信技术类)

《发光学报》是中国物理学会发光分会与中国科学院长春光学精密机械与物理研究所共同主办的中国物理学会发光分会的学术会刊。该刊是以发光学、凝聚态物质中的激发过程为专业方向的综合性学术刊物。

《发光学报》于1980年创刊,曾于1992年,1996年,2000年和2004年连续四次被《中文核心期刊要目总览》评为“物理学类核心期刊”,并于2000年同时被评为“无线电电子学、电信技术类核心期刊”。2000年获中国科学院优秀期刊二等奖。现已被《中国学术期刊(光盘版)》、《中国期刊网》和“万方数据资源系统”等列为源期刊。英国《科学文摘》(SA)自1999年;美国《化学文摘》(CA)和俄罗斯《文摘杂志》(AJ)自2000年;美国《剑桥科学文摘社网站》自2002年;日本《科技文献速报》(CBST, JICST)自2003年已定期收录检索该刊论文;2008年被荷兰“Elsevier Bibliographic Databases”确定为源期刊;2010年被美国“EI”确定为源期刊。2001年在国家科技部组织的“中国期刊方阵”的评定中,《发光学报》被评为“双效期刊”。2002年获中国科学院2001~2002年度科学出版基金“择重”资助。2004年被选入《中国知识资源总库·中国科技精品库》。本刊内容丰富、信息量大,主要反映本学科专业领域的科研和技术成就,及时报道国内外的学术动态,开展学术讨论和交流,为提高我国该学科的学术水平服务。

《发光学报》自2011年改为月刊,A4开本,144页,国内外公开发行。国内定价:40元,全年480元,全国各地邮局均可订阅。《发光学报》欢迎广大作者、读者广为利用,踊跃投稿。

地 址:长春市东南湖大路3888号
《发光学报》编辑部
邮 编:130033
电 话:(0431)86176862, 84613407
E-mail: fgxbt@126.com

国内统一刊号: CN 22-1116/O4
国际标准刊号: ISSN 1000-7032
国内邮发代号: 12-312
国外发行代号: 4863BM
http://www.fgxb.org