

多模式可调节光学机构的苹果可溶性固形物近红外光谱检测

刘燕德, 王军政, 姜小刚, 黎丽莎, 胡 宣, 崔惠桢

华东交通大学机电与车辆工程学院, 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 利用多模式可调节的光学机构采集了苹果漫透射、全透射和漫反射三种检测方式的光谱, 研究在不同的检测方式下苹果的光谱特征并采用 PLS 建立苹果可溶性固形物含量 SSC 预测的模型。首先分别采集每个样品赤道上四点的漫透射、全透射和漫反射光谱, 然后分别使用多元散射校正 MSC、基线偏移校正 BOC、归一化 Normalize 和高斯滤波平滑 GFS 等方法对平均后的 120 个光谱做预处理, 并结合竞争性自适应权重取样 CARS 法对漫反射光谱进行特征波长筛选, 最后采用偏最小二乘法 PLS 回归建立预测苹果 SSC 的模型, 并另购 30 个苹果验证模型性能。结果表明, 苹果在三种检测方式下采集的光谱吸收峰和波谷所处的波段大致相同, 但光谱强度有差异。三种光谱经 3 点高斯滤波平滑 GFS 预处理后建立预测苹果 SSC 含量的模型均取得很好的结果。漫透射的模型性能为 $R_{cal}=0.972$, $R_{pre}=0.967$ 和 $RMSEC=0.436\%$, $RMSEP=0.507\%$; 全透射的模型性能为 $R_{cal}=0.964$, $R_{pre}=0.957$ 和 $RMSEC=0.5\%$, $RMSEP=0.574\%$; 漫反射的模型性能为 $R_{cal}=0.963$, $R_{pre}=0.949$ 和 $RMSEC=0.522\%$, $RMSEP=0.536\%$; 三种光谱经归一化预处理后融合建模的模型性能为 $R_{cal}=0.894$, $R_{pre}=0.857$ 和 $RMSEC=0.836\%$, $RMSEP=0.966\%$ 。进一步将漫反射光谱结合 CARS 算法筛选特征波长, 使用 119 个变量建立模型的性能为 $R_{cal}=0.986$, $R_{pre}=0.977$ 和 $RMSEC=0.323\%$, $RMSEP=0.362\%$ 。最后将该模型导入新型多模式可调节的水果检测系统中, 使用 30 个未参与建模的冰糖心苹果检验该模型预测苹果 SSC 的性能。结果显示 30 个外部验证集的相关系数为 0.906, 验证均方根误差为 0.707%。进一步表明使用多模式可调节的水果内部品质检测系统采集的漫反射光谱结合光谱预处理、波段筛选算法和偏最小二乘回归方法可以建立较好的模型预测苹果可溶性固形物 SSC 含量。本研究为苹果的品质检测提供新的技术支持。

关键词 光学机构; 漫透射; 全透射; 漫反射; 可溶性固形物; CARS; 苹果

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)07-2064-07

引言

苹果不仅营养价值高而且产量多价格实惠, 深受广大人民的喜爱。苹果中含有许多酚类化合物, 而酚类化合物可以清除血液中的胆固醇, 进而保护人体的心脑血管。苹果中含有大量的纤维素可以促进人体的肠道蠕动, 有利于人体的消化和排毒。其中苹果的可溶性固体含量直接影响着苹果的口感, 它是苹果最重要的内部质量属性之一。也是人们可以直观感受到苹果的品质, 然后决定是否继续购买。

随着精密光谱检测仪器的研发和化学计量学的快速发展, 近红外光谱检测技术应用在水果品质检测方面已经非常成熟。应用近红外光谱技术检测水果的品质常见的为三种方式即漫反射、漫透射和全透射。其中漫反射是最早应用于农

产品检测, 漫反射检测结构比较简单, 但是漫反射容易受到杂散光、颗粒度的影响, 测量结果的误差也会比较大。使用光谱预处理和波段筛选算法可以减少杂散光和颗粒度的影响提高模型性能, 使漫反射检测技术得到广泛应用。如利用近红外漫反射光谱检测技术结合斜率和偏差的校正方法, 可以保证苹果可溶性固形物含量预测的模型具有长期稳定的性能^[1]; 利用漫反射光谱检测技术研究富士苹果六种不同的检测方向对预测 SSC 含量精度的影响^[2]; 利用漫反射光谱检测技术区分草莓储藏时间和预测草莓采后的糖度含量^[3]; 漫透射适用范围比较广, 可用于薄皮水果如苹果的品质检测^[4]; 也可用于厚皮水果如“脐橙的糖度检测”^[5]; 研究柚子的品种和糖度检测^[6]; 全透射多用于研究透光性比较好的水果内部病变和小型果的内部品质, 如苹果维管褐变的检测^[7]。全透射检测的缺点在于水果尺寸的影响较大, 有研究

收稿日期: 2020-07-16, 修订日期: 2020-12-07

基金项目: 国家自然科学基金项目(31760344), 水果光电检测技术能力提升项目(S2016-90), 研究生创新资金项目(YC 2019-S261)资助

作者简介: 刘燕德, 女, 1967年生, 华东交通大学机电与车辆工程学院教授 e-mail: jxliuyd@163.com

提出一种基于水果尺寸的光谱校正法^[6]用于减少苹果尺寸带来的影响,提高对患病苹果的识别;可以看到不同的检测方式可以满足各类水果不同品质的检测需求,很少有人研究苹果在三种检测方式下的光谱特征,并建立预测苹果糖度的模型。

本研究是利用多模式可调节的光学机构,分别采集苹果样品在漫透射、全透射和漫反射检测方式下的可见近红外光谱信息,分别使用多元散射校正(multiplicative scatter correction, MSC)、基线偏移校正(baseline offset correction, BOC)、归一化 Normalize 和高斯滤波平滑(gaussian filter smoothing, GFS)等方法对平均后的 120 个苹果的光谱做预处理,这些预处理可以对光谱进行修正和优化。并对预处理后的漫反射的光谱进行了 CARS 方法波段筛选,筛选了与苹果糖度相关的光谱变量建立预测苹果 SSC 含量的模型,另外还尝试将三种检测方式下的光谱混合建立苹果糖度预测模型。

1 实验部分

1.1 样本

从中国南昌当地的水果市场购买 120 个冰糖心苹果,苹果的赤道直径范围为 70~80 mm。在实验之前,首先对所有苹果进行单独清洗并擦干,其次在苹果的果梗部大约间隔 90°做 4 点标记。购买的苹果刚从冷藏室拿出温度低于室温,必须在实验室中放置 24 h 使苹果温度统一都达到室温,从而降低实验过程中因温度变化对预测精度的影响^[9]。实验样本一共 120 个,随机选择 90 个作为校正集建模,30 个作为预测集预测模型。为保证模型具有很强的适应性,校正集的糖度范围必须要广,手动将最大和最小的苹果糖度理化值和相应的光谱值放入校正集用于建模。

1.2 检测系统与光谱采集

检测使用 Ocean Optics 公司生产的 QE65PRO 光谱仪,它的波长范围为 350~1 150 nm。光源使用 6 盏 12 V、100 W 的卤钨灯,采用稳压电源供电。三种检测方式如图 1 所示,进行漫透射检测时,苹果两侧各布置 3 盏卤钨灯,调整两侧的光源与水平面呈 45°夹角,检测探头放置在苹果下端。调整水果托盘的高低可以确保光源均匀照射至苹果下半部;进行全透射检测时,苹果在光源和检测探头之间,而且三者保持在一条竖直线上;进行漫反射检测时,调整两侧的光源与

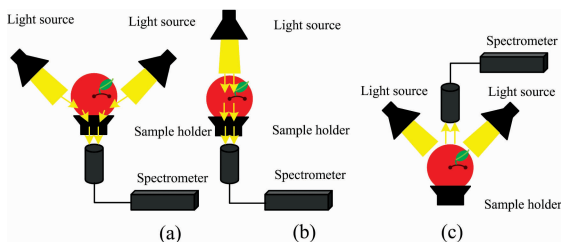


图 1 漫透射(a)、全透射(b)和漫反射(c)检测方式示意图

Fig. 1 Schematic diagram of detection methods of diffuse transmittance (a), total transmittance (b) and diffuse reflection (c)

水平面呈 45°夹角,检测探头放置在苹果上端。

采用聚四氟乙烯球为参比,预热 0.5 h 光源稳定后采集苹果光谱。光谱检测系统如图 2 所示,采集光谱时,将苹果采集面朝向检测探头,光源照射在苹果表面,穿透果皮进入苹果内部,经果肉散射、吸收和透射后进入检测探头。采用海洋光学的 FOP-UV 光纤探头连接光谱仪,使用海洋光学的光谱采集软件采集并保存光谱,依次对已标记的四个部位进行光谱采集和保存。

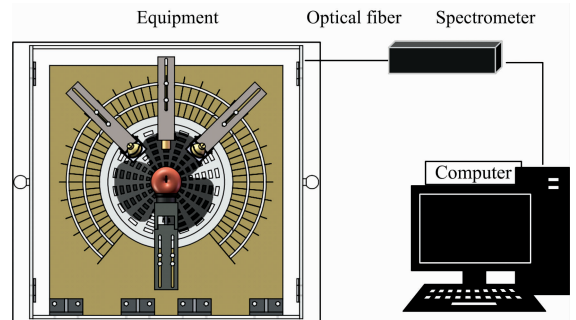


图 2 光谱检测系统示意图

Fig. 2 Schematic diagram of the spectrum detection system

1.3 可溶性固形物测定

在完成光谱采集后,榨取苹果采集面的果汁,并用滤纸过滤掉果肉、果皮残渣。使用一次性吸管将过滤好的果汁吸取并滴到日本 ATAGO 公司生产的 PAL-1 型便携式数显折射计的测定窗口,进行苹果的可溶性固形物 SSC 的测定。每次必须进行 3 次重复测量,因为刚滴定的果汁还不够稳定,多测几次取平均可以减少误差。榨汁的工具和折射计的测定窗口每测完一次都要用蒸馏水清洗,然后用未使用的吸水纸擦干净,减少人为因素的干扰,进而为下一次测量做好准备。该便携式数显折射计的测量精度很高为 $\text{Brix} \pm 0.2\%$,而且具有自动温度补偿功能,实验室的温度为 20°满足其温度使用范围 10°~40°。

1.4 化学计量学方法

1.4.1 数据处理与分析

从光谱仪采集到的苹果样本的原始光谱中除包含有效信息外,还可能具有背景信息和噪声。采用预处理往往可以校正光谱信息和消除噪音提高模型的性能。本研究使用 Unscrambler v9.7 软件中自带的几种预处理方法,如多元散射校正 MSC、基线偏移校正 BOC、归一化 Normalize 和高斯滤波平滑 GFS 等来优化原始光谱数据。多元散射校正 MSC 可以消除水果内部不均导致的散射效应,多用于水果漫反射和漫透射光谱的优化;基线偏移校正 BOC 通过估计基线并随后从测量光谱中减去估计值来校正导致信号具有垂直偏移或斜率的基线效应;归一化 Normalize 常用来校正由微小光程差异引起的光谱变化;高斯滤波 GFS 可以消除光谱中的随机噪声。

1.4.2 回归分析

偏最小二乘法不仅可以考虑可变矩阵 X (光谱矩阵),还可以考虑可变矩阵 Y (浓度矩阵)。PLS 将光谱矩阵和浓度矩阵同时进行分解,并在分解时考虑两者之间的关系。即使在

强共线性和大量变量的情况下,该技术也可以将变量与线性模型构建中的少量样本相关联^[10]。一般来说,偏最小二乘法用于从大量高度相关和共线的原始光谱数据中提取不超过前 20 个潜在变量(LVs)^[11]。LVs 可以解释方差,降低原始光谱的维数。本研究利用偏最小二乘法建立苹果的光谱和可溶性固形物之间的相关性模型。

1.4.3 变量选择方法 CARS

为使模型的性能更加优化,常采用有效波长选择方法来提高预测精度,简化模型以降低测量成本,并在一定程度上促进模型解释^[12]。竞争性自适应权重取样(CARS)法是一种创新且有效的变量选择算法^[13],它能够从全光谱中筛选出少量的有效变量并与偏最小二乘回归结合建模。CARS 根据每个变量重要程度,以迭代和补偿的方式从 N 个蒙特卡罗抽样运行中依次选择 N 个变量子集。在每次采样运行中,首先以固定比率随机选择一些样本,以构建模型。其次,采用指数递减函数和自适应权重取样方法,根据回归系数选择关键变量。最后,选择交叉验证均方根误差(RMSECV)值最低的子集建立模型。

1.4.4 模型性能的评估

模型的建立和评估模型的性能会使用一些统计参数,一般使用校正相关系数(R_c)、预测相关系数(R_p)、校正均方根误差(RMSEC)和预测相关系数(RMSEP)。它们的计算公式如式(1)~式(4)

$$R_{cal} = \sqrt{\frac{\sum_{i=1}^{n_c} (Y_{pi} - Y_{mi})^2}{\sum_{i=1}^{n_c} (Y_{pi} - Y_{mean})^2}} \quad (1)$$

$$R_{pre} = \sqrt{\frac{\sum_{i=1}^{n_p} (Y_{pi} - Y_{mi})^2}{\sum_{i=1}^{n_p} (Y_{pi} - Y_{mean})^2}} \quad (2)$$

$$RMSEC = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (Y_{pi} - Y_{mi})^2} \quad (3)$$

$$RMSEP = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (Y_{pi} - Y_{mi})^2} \quad (4)$$

其中 Y_{pi} 是第 i 个苹果的 SSC 的预测值, Y_{mi} 是第 i 个苹果的 SSC 的测量值, Y_{mean} 是校正集或者预测集中苹果 SSC 的平均值,校正集和预测集中冰糖心苹果的数量分别表示为 n_c 和 n_p 。一般来说,比较好的模型不仅要有较高的校正相关系数 R_{cal} 值和预测相关系数 R_{pre} 值以及较低的 RMSEC 和 RMSEP 值,还要选择 RMSEC 和 RMSEP 之间的差异比较小^[14]。

2 结果与讨论

2.1 光谱特征和光谱分析

如图 3 所示,在苹果茎-萼轴水平方向上获得的范围从 500~1 100 nm 的漫透射(a)、全透射(b)和漫反射(c)平均的原始光谱。由图可知,漫反射光谱的能量最强,处于漫透射和全透射光谱的上方;其次是漫透射光谱处于漫反射光谱和全透射光谱之间;全透射因为光程比较长被散射和吸收的光子比较多,所以光谱能量较低。另外这三种检测方式所采集的光谱的吸收峰和波谷所处的波段大致相同,只是强度不

同。其中大约在 675 和 740 nm 附近存在波谷,而约在 650、710 和 820 nm 附近存在波峰。由于光谱的两端存在一些噪声信号的影响,选取 500~1 100 nm 的漫透射光谱进行建模。在 675 nm 附近,可能由于叶绿素的强吸收引起的光谱吸收峰。在 750 nm 的光谱吸收峰与 H—O 的第三倍频和 C—H 的第四倍频相关。上述的光谱波段都与苹果中 SSC 的含量具有相关性。

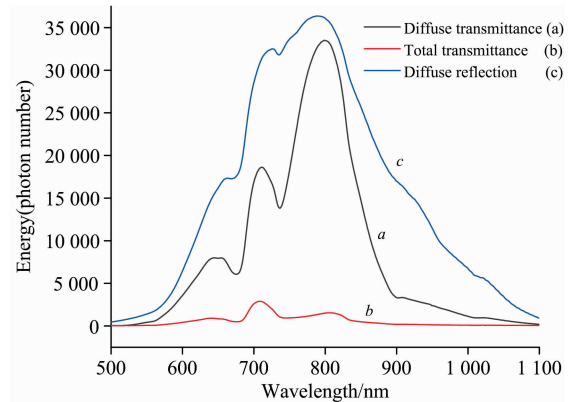


图 3 苹果的漫透射(a)、全透射(b)和漫反射(c)的原始光谱图

Fig. 3 The original spectra of Apple's diffuse transmittance (a), total transmittance (b) and diffuse reflection (c)

2.2 可溶性固形物统计

表 1 为三种检测方式的校正集和预测集可溶性固形物的范围,标准偏差和平均值。从表 1 可知,三种检测方式的校正集和预测集糖度范围并不相同。但是可以看到校正集可溶性固形物含量的范围大于预测集,这是手动调节最大值和最小值放入校正集的结果,好处在于所建模型的适应性更好。

表 1 校正集和预测集中可溶性固形物 SSC 含量的范围、标准偏差和平均值

Table 1 The range, standard deviation and average value of SSC in the calibration and prediction sets

	Parameter	Data set	Samples	Mean	S. D	Range
漫透射	SSC/Brix°	Calibration	90	13.17	1.86	8.9~17.3
		Prediction	30	13.07	1.99	9.1~17.0
全透射	SSC/Brix°	Calibration	90	13.10	1.89	8.9~17.3
		Prediction	30	13.27	1.88	9.2~17.0
漫反射	SSC/Brix°	Calibration	90	13.10	1.94	8.9~17.3
		Prediction	30	13.40	1.65	9.2~17.2

2.3 不同预处理方法对比分析

不同的预处理方法处理原始光谱会达到有不同的效果。选择几种方法进行对比,从而选择模型效果最好的。表 2 分别给出了原始光谱和使用不同预处理方法的光谱结合偏最小二乘法 PLS 建立模型预测苹果糖度的结果。结果表明,用 3 点高斯滤波平滑预处理后的光谱进行偏最小二乘法建立模型的预测结果优于原始光谱和其他预处理方法的光谱进行偏最小二乘法建立模型的预测结果。混合三种光谱建立模型使用

归一化 Normalize 的预处理方法较好。

表 2 原始光谱和使用不同预处理方法的光谱结合偏最小二乘法 PLS 建立模型预测苹果糖度的结果

Table 2 The original spectrum and the spectrum using different pretreatment methods combined with PLS are used to build a model to predict the results of apple SSC

	Pretreatment	LVs	Calibration		Prediction	
			R_c	RMSEC	R_p	RMSEP
漫透射	Raw	18	0.980	0.370	0.966	0.516
	MSC	16	0.979	0.379	0.960	0.559
	Normalize	15	0.964	0.488	0.951	0.619
	BOC	18	0.982	0.348	0.962	0.550
	GFS	18	0.972	0.436	0.967	0.507
全透射	Raw	16	0.974	0.431	0.958	0.562
	MSC	12	0.969	0.464	0.913	0.761
	Normalize	12	0.966	0.484	0.942	0.639
	BOC	16	0.971	0.447	0.949	0.610
	GFS	16	0.964	0.500	0.957	0.574
漫反射	Raw	14	0.968	0.490	0.947	0.546
	MSC	13	0.970	0.473	0.889	0.782
	Normalize	14	0.984	0.350	0.896	0.762
	BOC	13	0.963	0.523	0.947	0.545
	GFS	14	0.963	0.522	0.949	0.536
混合	Raw	20	0.837	1.022	0.823	1.069
	MSC	19	0.888	0.861	0.856	0.983
三种光谱	Normalize	20	0.894	0.836	0.857	0.966
	BOC	20	0.838	1.020	0.823	1.068
	GFS	20	0.832	1.038	0.823	1.070

通过对苹果不同的检测方式和不同的预处理方法的比较,表明:漫透射预测相关系数和预测均方根误差分别为 $R_p = 0.967$ 和 $RMSEP = 0.507\%$ 优于漫反射的 $R_p = 0.949$ 和 $RMSEP = 0.536$ 。这是由于采集的漫反射光谱具有大量杂散光的影响,而漫透射光谱因为采集方式的差异会减少杂散光的影响。全透射的建模结果为 $R_p = 0.957$ 和 $RMSEP = 0.574\%$,这是因为苹果不规则的果核会散射和吸收掉一部分有效信息,苹果的尺寸也会导致很大的影响,所以建模结果相对漫透射和漫反射要差些。尝试将三种光谱混合进行建模,经过归一化预处理 PLS 建模结果为 $R_p = 0.857$ 和 $RMSEP = 0.966\%$;建立苹果在不同检测方式下的通用模型有很大的实际应用价值,虽然建模效果比不上单一检测方式的模型,但适用性广。进一步对混合三种光谱进行波段筛选。

观察表 2 可以发现,相对于漫透射光谱和全透射光谱,不论是原始光谱建模还是预处理之后的光谱建模,漫反射光谱的潜在变量最少,可以减小模型的维度。漫透射校正和预测均方根误差之间的差值为 0.071% ,全透射校正和预测均方根误差之间的差值为 0.074% ,漫反射校正和预测均方根误差之间的差值为 0.014% ,漫反射光谱所建模型的校正均方根误差 RMSEC 和预测均方根误差 RMSEP 两者之间的差值最小。可以得出结论,通过对漫反射光谱进行 3 点高通滤波平滑预处理可以提高模型的性能,消除杂散光的影响。因

此,使用波段筛选算法进一步的分析基于 3 点高斯平滑预处理后的漫反射光谱。

2.4 三种检测方式优缺点分析

从图 1 所示的漫反射和漫透射光学结构示意图。可知,漫透射底部收集光谱,需要给果杯加上橡胶垫进行遮光处理。有些漫反射结构也会在采集光纤上做一些喇叭状结构的遮光,通过预处理和波段筛选也可以消除杂散光影响。光源和采集光纤在同一侧的漫反射结构比较简单,还可以将光源和光纤集成在一起,这是其他检测方式不能实现的。全透射光源和光纤在一条直线上,这种检测方式可以得到整个果的信息。

从图 3 所示的三种原始光谱图,我们可以了解到相同的光源强度下,光谱强度顺序为漫反射大于漫透射大于全透射,这是由采集光谱的光学结构不同,导致光子在苹果内部的光程不同,内部光程越长散射吸收的光子越多,采集的光谱强度越低。对于全透射方式采集的光谱强度太低,这会丢失有效信息,可以通过增强光源或调节积分时间解决。但是对于薄皮水果太强的光源容易损伤水果表皮,造成灼伤。对于动态在线检测可以通过散热和提高传送带速度,但是太快的速度会导致收集光谱有效信息的丢失。一般情况下薄皮水果多采用漫反射和漫透射检测,不需要很强的光源也可以取得很好的效果。

从表 2 所示的模型效果,我们可以了解到全透射的糖度模型预测效果最差。这是因为全透射光子在苹果的内部走过的光程最长,苹果的结构对这种检测方式的影响很大,每个苹果内部不规则果核都有差异,苹果的整体尺寸也不能保证一致。还有我们并未将整个苹果果肉全部榨汁,仅仅切取采集光谱部位的一定厚度的果肉。

总的来说不同的检测方式都有其优缺点,应该根据不同水果的不同检测需求合理选择检测方式。如漫反射检测薄皮果和透光性较差果;漫透射检测薄皮果和厚皮果;全透射的检测方式多用于透光性好的水果内部病变,对于定性判别内部损伤果的效果很好。

2.5 CARS 变量优选

波长优选可以剔除光谱信息中与被测物质非线性相关的波长,还可以减少波长变量的个数简化模型,广泛应用于可见/近红外光谱的模型开发。运用竞争性自适应权重取样(CARS)法,从 3 点高斯滤波平滑预处理后的光谱中选择特征波长,建立预测苹果 SSC 含量的模型。

在运行 CARS 之前,蒙特卡罗采样运行的次数设置为 50,所选择的变量数由 5 倍交互验证均方差 RMSECV 确定。在图 4(a)中可以看出,由于采用指数递减函数(EDF),去掉回归系数绝对值相对较小的波长点。可以看到样本的波长变量个数在开始时快速减少,在采样结束时缓慢变化。快速选择和精细选择的结合是 CARS 算法的重要功能,有助于提高波长变量选择的速度。如图 4(b)中所示,随着蒙特卡罗采样次数的增加,交互验证均方差 RMSECV 值也在不断变化。首先由于剔除无效变量,RMSECV 值以平缓的方式减小,然后随着有效变量的丢失 RMSECV 值迅速增加。如图 4(c)所示,随着采样次数的增加,每条线代表每个波长的回归系数

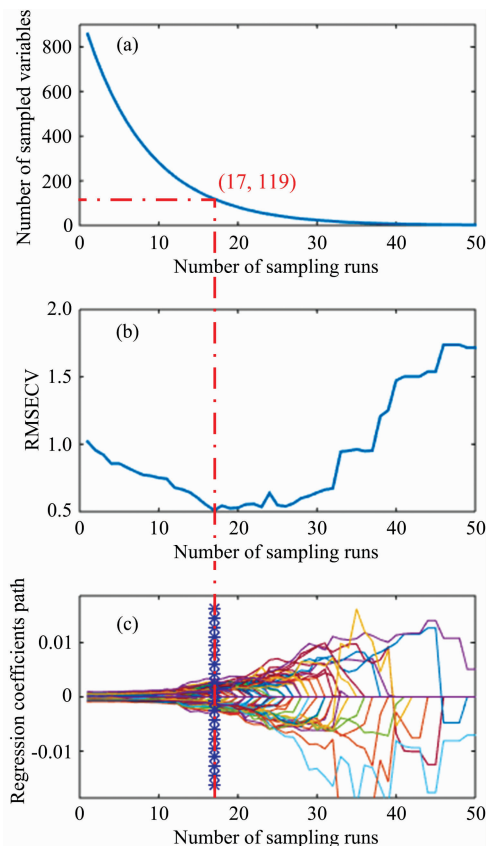


图 4 预处理后的漫反射光谱采用 CARS 方法的波长变量筛选过程

Fig. 4 Wavelength screening using CARS method for pre-processed diffuse reflectance spectrum

变化, 一般来说具有较大回归系数绝对值的波长点更容易被选择为特征波长用于建模。最终从漫反射光谱的 858 个波长中选择 RMSECV 值最小的 119 个波长作为预测苹果糖度的特征变量。

2.6 PLS 模型比较

将 CARS 方法所选择的 119 个波长变量采用 PLS 建模。从表 3 可知, 与全光谱的 PLS 模型相比, 由 CARS-PLS 构建的模型获得更好的预测精度, 漫反射波长变量数由 858 个减少为 119 个, 且 RMSEP 值由 0.536% 下降为 0.362%。虽然漫反射具有杂散光的影响但是采用 CARS 筛选经 3 点高斯滤波平滑之后的漫反射光谱依然取得很好的结果。所以预处理和波段筛选在优化光谱、降低变量个数和提高模型性能方面是非常有效的。另外对混合三种检测方式的光谱进行 CARS 波段筛选, 波长变量数由 1044 个减少为 154 个, 且 RMSEP 值由 0.966% 下降为 0.901%。

如图 5 所示, 漫反射光谱通过 CARS-PLS 方法获得的测量真值和预测值的散点图。可以看到这些点分布在直线 $y=x$ 的两侧, 且距离理想的回归线很近, 说明模型效果很好。

2.7 外部验证实验

首先将该模型导入新型多模式可调节的水果检测系统中, 为了验证模型的预测糖度的性能, 从原来的水果批发市场再一次购买了 30 个冰糖心苹果。新型水果无损检测系统采集新购买的冰糖心苹果漫反射光谱, 将采集的光谱数据代入到苹果糖度定量预测模型中, 最终显示出苹果的糖度预测结果。结果如图 6 所示, 新型水果无损检测系统对 30 个苹果样品的糖度含量预测值与标准理化值相关系数为 0.906, 验证均方根误差为 0.707%。进一步表明使用多模式可调节的水果内部品质检测系统采集的漫反射光谱结合光谱预处理、波段筛选算法和偏最小二乘回归方法可以建立较好的模型预测苹果可溶性固形物 SSC 含量。

表 3 全光谱变量和 CARS 筛选变量的建模结果

Table 3 Modeling results of full spectrum variables and CARS selection variables

Method	Number of variables	LVs	Calibration		Prediction		
			R_c	RMSEC	R_p	RMSEP	
漫反射	Full-spectra PLS	858	14	0.963	0.522	0.949	0.536
	CARS-PLS	119	12	0.986	0.323	0.977	0.362
混合三种光谱	Full-spectra PLS	1 044	20	0.894	0.836	0.857	0.966
	CARS-PLS	154	15	0.880	0.889	0.876	0.901

3 结论

利用多模式可调节的光学机构采集苹果漫透射、全透射和漫反射三种检测方式的光谱, 采用高斯滤波平滑 GFS 对光谱进行去噪声预处理, 结合 PLS 建立苹果糖度预测模型均取得很好的结果。漫透射的模型性能为 $R_{cal}=0.972$, $R_{pre}=0.967$ 和 $RMSEC=0.436\%$, $RMSEP=0.507\%$ 。全透射的模型性能为 $R_{cal}=0.964$, $R_{pre}=0.957$ 和 $RMSEC=0.5\%$, $RMSEP=0.574\%$ 。漫反射的模型性能为 $R_{cal}=0.963$,

$R_{pre}=0.949$ 和 $RMSEC=0.522\%$, $RMSEP=0.536\%$ 。进一步将漫反射光谱结合 CARS 算法筛选特征波长, 使用 119 个变量建立模型的性能为 $R_{cal}=0.986$, $R_{pre}=0.977$ 和 $RMSEC=0.323\%$, $RMSEP=0.362\%$ 。利用 30 个未参与建模的样品进行外部验证, 苹果可溶性固形物验证集的相关系数为 0.906, 验证均方根误差为 0.707%。结果表明经过高斯滤波平滑预处理和 CARS 筛选波段后的漫反射光谱消除杂散光和随机噪声, 显著提高预测苹果糖度模型的性能。还尝试将混合三种检测方式的光谱进行了归一化预处理、CARS 波段筛选和 PLS 建模, 模型性能为 $R_{cal}=0.880$, $R_{pre}=0.876$ 和

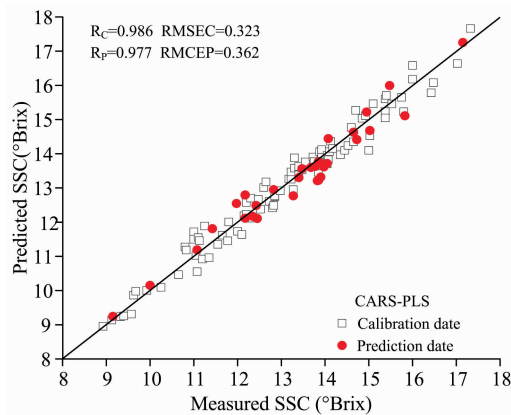


图 5 结合 CARS-PLS 方法测量与预测苹果可溶性固形物的散点图

Fig. 5 Scatter plot of measured versus predicted SSC by the PLS calibration models combining CARS method

RMSEC=0.889%, RMSEP=0.901%。本研究为苹果的品质检测提供新的技术支持。然而,本文仅仅研究了苹果的一

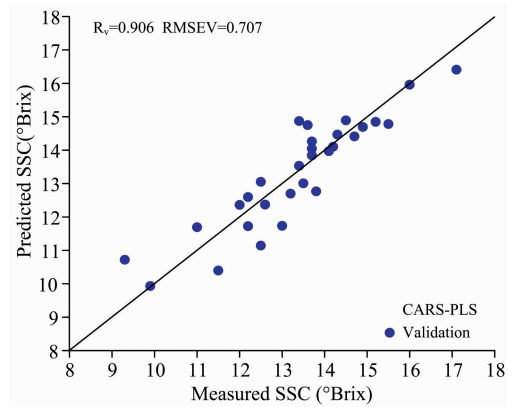


图 6 外部验证结果散点图

Fig. 6 Scatter plot of external verification results

种内部品质可溶性固形物含量,下一步使用多模式可调节光学机构采集不同检测方式下内部病变苹果的光谱,对不同检测方式下的光谱进行光谱特征分析并作出定性的判别预测。

References

- [1] Fan Shuxiang, Li Jiangbo, Xia Yu, et al. Postharvest Biology and Technology, 2019, 151: 79.
- [2] Xia Yu, Huang Wenqian, Fan Shuxiang, et al. Infrared Physics and Technology, 2019, 97: 467.
- [3] Shen Fei, Zhang Bin, Cao Chongjiang, et al. Journal of Food Process Engineering, 2018, 41(7): 12866.
- [4] Zhang Lurui, Xu Huirong, Gu Minyan, et al. Journal of Food Engineering, 2014, 139: 19.
- [5] LIU Yan-de, HU Jun, OUYANG Yu-ping, et al(刘燕德, 胡 军, 欧阳玉平, 等). Guangdong Agricultural Sciences(广东农业科学), 2016, 43(9): 105.
- [6] LI Xiong, LIU Yan-de, OUYANG Ai-guo(李 雄, 刘燕德, 欧阳爱国). Chinese Journal of Luminescence(发光学报), 2019, 40(6): 808.
- [7] Sun J, Kunemeyer R, McGlone A, et al. Computers and Electronics in Agriculture, 2018, 155: 32.
- [8] TIAN Shi-jie, ZHANG Jun-hua, ZHANG Zhong-xiong, et al. Infrared Physics and Technology, 2019, 100: 117.
- [9] Fan Shuxiang, Huang Wenqian, Guo Zhiming, et al. Food Analytical Methods, 2015, 8(8): 1936.
- [10] Tahir Mehmood, Kristian Hovde Liland, Lars Snipen, et al. Chemometrics and Intelligent Laboratory Systems, 2012, 118: 62.
- [11] Leiva-Valenzuela Gabriel A, Lu Renfu, Aguilera José Miguel. Journal of Food Engineering, 2013, 115(1): 91.
- [12] Cheng Junhu, Sun Dawen, Pu Hongbin, et al. Food Chemistry, 2015, 171: 258.
- [13] Li H, Liang Y, Xu Q, et al. Analytica Chimica Acta, 2009, 648(1): 77.
- [14] Li Jiangbo, Huang Wenqian, Chen Liping, et al. Food Analytical Methods, 2014, 7(9): 1891.

Research on Vis/NIR Detection of Apple's SSC Based on Multi-Mode Adjustable Optical Mechanism

LIU Yan-de, WANG Jun-zheng, JIANG Xiao-gang, LI Li-sha, HU Xuan, CUI Hui-zhen

School of Mechatronics & Vehicle Engineering, East China Jiaotong University, National and Local Joint Engineering Research Center of Fruit Intelligent Photoelectric Detection Technology and Equipment, Nanchang 330013, China

Abstract A multi-mode adjustable optical mechanism was used to collect the spectra of apples in three detection methods: diffuse transmittance, total transmittance and diffuse reflection. The spectrum characteristics of apples under different detection methods were studied and PLS established the apple soluble solid content prediction model. First, the system will collect the spectra of four points on the equator of each sample under the way of diffuse transmittance, total transmittance and diffuse reflection respectively, and then MSC (Multiplicative Scatter Correction), BOC (Baseline offset correction), Normalize, Gaussian filter smoothing and other methods will be applied to preprocess the 120 averaged spectra combined with the CARS method to filter the characteristic wavelength of the diffuse reflection spectrum. Finally, PLS will be used to establish a model for predicting apple's SC, and another 30 apples will be purchased to verify the performance of the model. The results show that the spectra collected by the self-designed fruit quality detection system under the three detection methods have good result in predict SSC content in apples after 3-point Gaussian filtering smoothing pretreatment. The performance of diffuse transmittance model is $R_{\text{cal}}=0.972$, $R_{\text{pre}}=0.967$ and $\text{RMSEC}=0.436\%$, $\text{RMSEP}=0.507\%$. The performance of total transmittance model is $R_{\text{cal}}=0.964$, $R_{\text{pre}}=0.957$ and $\text{RMSEC}=0.5\%$, $\text{RMSEP}=0.574\%$. The performance of diffuse reflection model is $R_{\text{cal}}=0.963$, $R_{\text{pre}}=0.949$ and $\text{RMSEC}=0.522\%$, $\text{RMSEP}=0.536\%$. The fusion modeling performance of the three spectra after normalization pretreatment is $R_{\text{cal}}=0.894$, $R_{\text{pre}}=0.857$ and $\text{RMSEC}=0.836\%$, $\text{RMSEP}=0.966\%$. Further, the diffuse reflection spectrum is combined with the CARS algorithm to filter the characteristic wavelengths. The performance of the model established with 119 variables is $R_{\text{cal}}=0.986$, $R_{\text{pre}}=0.977$ and $\text{RMSEC}=0.323\%$, $\text{RMSEP}=0.362\%$. Finally, the model is imported into this new multi-mode adjustable fruit detection system, and 30 non-model apples are used to test the model to predict the performance of apple's SSC. The results show that the correlation coefficient of the 30 external validation sets is 0.906, and the root means square error of validation is 0.707%. It further shows that the diffuse reflection spectrum collected by the multi-mode adjustable fruit internal quality detection system, which is combined with spectral pretreatment, CARS and PLS can establish a better model to predict the solid soluble content of apple. This research provides new technical support for Apple's internal quality testing.

Keywords Optical mechanism; Diffuse transmittance; Total transmittance; Diffuse reflection; Soluble solids content; CARS; Apple

(Received Jul. 16, 2020; accepted Dec. 7, 2020)