

基于近红外高光谱成像技术的涩柿 SSC 含量无损检测

Research on Non-destructive methods for soluble solid content detection of
astringent persimmon based on near-infrared hyperspectral technology

魏 萱¹ 何金成¹ 叶大鹏¹ 介邓飞²

WEI Xuan¹ HE Jin-cheng¹ YE Da-peng¹ JIE Deng-fei²

(1. 福建农林大学机电工程学院, 福建 福州 350002; 2. 华中农业大学工学院, 湖北 武汉 430070)

(1. College of Mechanical and Electronic Engineering, Fujian Agriculture and Forestry University, Fuzhou, Fujian, 350002, China; 2. College of Engineering, Huazhong Agricultural University, Wuhan, Hubei 430070, China)

摘要:对 150 个涩柿采集 900~1 700 nm 波段的近红外高光谱图像信息,利用蒙特卡罗—无信息变量消除(MC-UVE)和连续投影算法(SPA)对感兴趣区域光谱进行波长优选。通过 MC-UVE-SPA 优选出 924.69, 928.05, 1 112.72, 1 270.91, 1 365.3, 1 402.42, 1 453.06, 1 547.69 nm 8 个特征波长,对应的光谱反射率作为柿子可溶性固形物含量(SSC)检测的偏最小二乘回归(PLSR)检测模型输入,其预测集相关系数 $r_{pre}=0.942$, 预测集均方根误差 RMSEP=1.009 °Brix。结果表明, MC-UVE-SPA 可以有效提取与柿子 SSC 含量相关的特征信息,从而保留较少的波长建立较好的预测模型。

关键词:近红外高光谱成像;可溶性固形物;柿子;无损检测

Abstract: This study collected the near infrared (NIR) hyperspectral images of 150 astringent persimmons, with the spectra are in 900~1700 nm. Monte Carlo-uninformative variable elimination (MC-UVE) algorithm and successive projections algorithm (SPA) were adopted to the optimization of wavelengths obtained from the region of interest (ROI). Eight wavelengths were selected by MC-UVE-SPA. These feature wavelengths were 924.69, 928.05, 1 112.72, 1 270.91, 1 365.3, 1 402.42, 1 453.06 and 1 547.69 nm, respectively. The spectral reflectance of the 8 feature wavelengths were applied to establish the detective model for the soluble solid content (SSC) of persimmon by partial least squares regression (PLSR) method. The correlation coefficient and root mean square error of prediction set are $r_{pre}=0.942$, RMSEP=1.009 °Brix. The results indicated that MC-UVE-SPA could effectively extract the char-

acteristic information related to the SSC and develop a better predictive model with fewer wavelengths. This work can provide technical support and research basis for the nondestructive detection, grading and processing equipment for persimmon quality.

Keywords: NIR hyperspectral imaging; SSC; persimmon; non-destructive detection

柿子具有较高的经济和营养价值,深受人们的喜爱^[1]。柿子在收获和消费阶段易发生衰老褐变,导致风味变劣,影响其食用品质,降低商品价值^[2-3]。目前,在中国鲜柿采后品质分级仍然依靠工人的经验直观判断或者抽样进行破坏性检测,人工分级易受主观判断影响,抽样破坏性检测制样繁琐且检测效率低下,无法满足大规模生产和市场需求。因此,研究柿子内部品质无损检测技术,可进一步提高柿果农产品附加值,增强市场竞争力。

国内外学者^[4-6]采用无损检测技术对柿子内部品质方面已经开展了不同程度的研究,目前使用最多的是可见/近红外光谱检测技术,已被应用于柿子硬度^[7]、单宁^[7-8]、涩度^[8]和可溶性固形物^[9]等内部品质检测。高光谱成像技术兼备图像与光谱两者的优点,在农产品品质和安全性评估方面具有巨大应用潜力,被研究学者广为青睐,在柿子内部品质检测方面也开展了一些研究^[10],但这些报道主要集中于可见—近红外高光谱技术^[11-12],对于近红外高光谱技术的研究报道较缺乏,柿子在近红外波段具有丰富的信息,有学者^[13]对比分析了近红外波段范围内不同模型对柿子涩度检测精度的影响,但并未对特征波长优选方法作进一步研究。

高光谱成像仪具有较高的光谱分辨率,获得信息非常丰富,但原始高光谱数据中也包含有冗余或不相关的信息,这

基金项目:福建省自然科学基金(编号:2017J05041);福建农林大学现代农林装备及其自动化创新平台(编号:612014017)

作者简介:魏萱(1987—),女,福建农林大学讲师,博士。

E-mail: dfjie@mail.hzau.edu.cn

收稿日期:2017—08—03

些信息会削弱光谱模型的预测性和稳定性^[14]。因此,挑选出具有代表性的波长能够提高模型性能,目前已经有许多特征波长选取方法。连续投影算法(SPA)能够有效消除众多波长变量之间的共线性影响,降低模型的复杂度,以其简便、快速的特点在可见/近红外光谱分析中广泛应用,也取得了较好的波长优选效果^[15]。蒙特卡罗(Monte Carlo, MC)方法,是一种基于“随机数”和概率统计来考察问题的计算方法。MC 方法在分析计算复杂多变量问题时有很强的优势,目前已在诸多领域得到广泛应用^[16]。无信息变量消除方法是基于 PLS 回归系数提出的一种波长优选方法,它是通过一定的变量优选标准,利用变量的稳定性值来评价模型中每个变量的可靠性,从而决定变量的取舍^[17-18]。本研究探索适用于柿子近红外高光谱数据的变量选择方法和优选出与内部品质相关的特征波长,寻找最优的预测组合模型,可为开发柿子在线检测分级设备提供参考。

1 材料与方法

1.1 材料与仪器

1.1.1 试验材料

柿子:采摘于湖北省武汉市江夏区某果园,所摘样本大小接近,外观完好。采摘当天运回实验室后,擦拭表面去污,选取试验样本 150 个,并依次编号,测定样本的重量、高度、周长等物理指标,获取样本的近红外高光谱图像。

1.1.2 主要仪器

高光谱摄像机:ImSpector N17E 型,芬兰 Specim 公司;
数字式糖度计:PR-101 型,日本 Atago 公司。

1.2 试验方法

1.2.1 光谱采集 柿子试验样本的高光谱图像信息通过漫反射的方式获取。近红外高光谱成像系统(见图 1)主要包括高光谱摄像机、150 W 卤素灯线光源、水平移动控制台、计算机及暗箱等。调试试验装置的电动位移平台速度、镜头高度、焦距、光圈和曝光时间等参数以获得较高信噪比高光谱图像。采集柿子花萼部位的光谱图像信息,光谱范围为 900~1 700 nm,曝光时间为 5 000 μ s,电动位移平台移动速度为 25 mm/s,镜头高度为 40 cm。为了消除镜头中存在的

暗电流以及光源强度分布不均匀所产生的噪声,根据式(1)对高光谱图像进行黑白校正^[19]。

$$I_c = \frac{I_r - I_d}{I_w - I_d}, \tag{1}$$

式中:

- I_c ——校正后的图像;
- I_w ——扫描标准白色校正板得到的标定图像;
- I_d ——关闭光源和镜头盖得到标定图像;
- I_r ——原始的高光谱图像。

1.2.2 理化指标的测定 高光谱图像采集完毕之后,进行糖度理化指标测定。柿子样本糖度的测量使用数字式糖度计,检测精度为 0.1 °Brix,测量范围为 0~45 °Brix。去除柿子样本非可食部分,将柿子的花萼部位的可食部分切碎、榨汁后,用滤纸过滤得到柿子汁液,取其汁液进行 SSC 含量测定,并对 150 个柿子样本均进行重复试验。

1.2.3 分析方法和软件 利用 ENVI 4.7(美国 Exelis Visual Information Solutions 公司)、Unscrambler V10.1(挪威 CAMO AS 公司)和 MATLAB 2010a(美国 MathWorks 公司)软件对采集到的柿子高光谱信息进行数据处理。通过相关系数(r)、模型的校正均方根误差(Root mean square error of calibration, RMSEC)、模型的预测均方根误差(Root mean square error of prediction, RMSEP)对模型的稳定性和准确性进行评价,模型的相关系数 r (包括校正集相关系数 r_{cal} 和预测集相关系数 r_{pre})越大越好, RMSEC 和 RMSEP 越小越好,所建的模型效果越好^[20-21]。

2 结果与分析

2.1 光谱分析

本研究以范围尽可能大和大小一致的原则选取样本感兴趣区域。最终选取柿子图像中心部分 200×200 像素方形 ROI 的光谱信息并取得平均光谱。去除光谱首尾明显噪声后,150 个样本于 910~1 600 nm(205 个波段)的光谱曲线见图 2。由图 2 可知,由于低反射率意味着高吸收率,图中有 3 处较明显的波谷分别位于 980, 1 150, 1 420 nm 附近。SSC 由包含 C—H、O—H、C—O 和 C—C 等化学键的物质组成。980 nm 属于 O—H 二级倍频吸收峰,1 150 nm 属于 C—H 二级倍频伸缩,1 420 nm 属于水分子中 O—H 一级倍频吸收带^[22-23]。为了消除样品表面散射以及光程变化等对光谱的影响,获取的 ROI 光谱经过标准化正态变量变换(SNV)预处理之后进行后续分析与处理。

2.2 样本理化值测定结果

将样品按照 2:1 的比例划分为建模集和预测集,建模集样品的糖度值包含预测集样品的糖度值。最终建模集和预测集分别有 100 个和 50 个样本。其理化指标测定结果见表 1。校正集范围较好地涵盖了预测集,有助于建立一个较好的预测模型。变量相关性分析显示重量、高度、赤道周长与 SSC 含量变化均没有显著相关性(P 值分别为 0.608, 0.318, 0.652)。

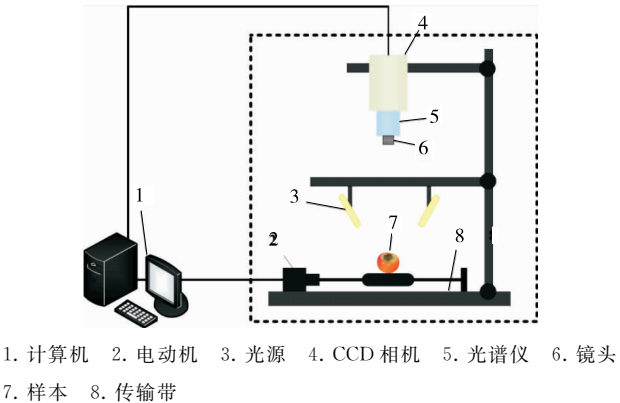


图 1 近红外高光谱成像系统

Figure 1 Schematic diagram of NIR hyperspectral imaging system

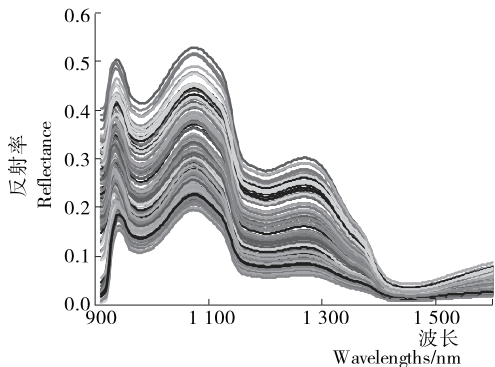


图 2 样本近红外高光谱图像感兴趣区域光谱曲线
Figure 2 The spectra of the samples obtained from the ROI in the NIR hyperspectral images

表 1 样本理化及可溶性固形物含量测量结果

Table 1 The measurement of the SSC and physicochemical parameters

样本	指标	重量/g	高度/mm	周长/cm	糖度/°Brix
建模集	最小值	118.9	40.90	20.90	12.30
	最大值	188.5	51.50	24.60	25.50
	平均值	146.9	45.60	22.60	17.40
	标准方差	14.9	2.80	0.78	2.92
预测集	最小值	122.7	41.40	20.90	12.50
	最大值	186.1	52.90	24.80	25.50
	平均值	149.2	46.00	22.60	18.70
	标准方差	15.8	2.60	0.90	3.01

2.3 特征波长优选

分别对经预处理后 205 个波长采用 SPA 和 MC-UVE-SPA 进行特征波长优选。图 3 是对 205 个波长通过 SPA 选择的不同变量数的 RMSE 分布。由图 3 可知,选取 3 个波长时, RMSE 快速下降,选择 6 个波长时, RMSE 又有一个快速下降过程,到提取出 10 个有效波长时 RMSE 到达一个低点,以后 RMSE 下降速率趋于平缓,由此确定从 205 个波长中选择出其中的 10 个波长,分别是 917.99, 924.69, 928.05, 931.4, 978.37, 1 334.95, 1 402.42, 1 453.06, 1 483.46, 1 598.43 nm。

图 4 给出了波长 910~1 600 nm 时,柿子 SSC 通过 MC-

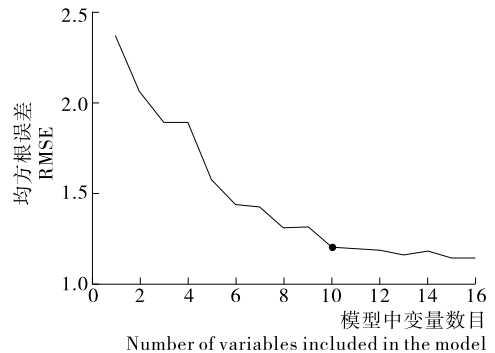


图 3 SPA 选择的不同变量数的 RMSE 分布
Figure 3 The distribution of the RMSE with different variables selected by SPA

UVE 方法得到的每个变量的稳定性值。在 MC-UVE 方法中,保留变量的数目决定着模型的预测稳定性和精确度。如果保留的变量个数过少,可能会造成有用信息变量的丢失;相反,如果保留的变量个数过多,多余的无用信息变量会使得模型变差。图 4 中,虚线表示变量筛选的阈值。位于两个虚线之外的稳定性值所对应的变量将用于 SPA 特征波长优选,此时共有 120 个变量被保留。此后采用 SPA 进一步对 120 个波长进行特征波长优选,与直接应用 SPA 类似,选取 3 个波长和 6 个波长时, RMSE 快速下降,到提取出 8 个有效波长时 RMSE 到达一个低点,此后趋于平稳。由此确定 8 个特征波长,分别是 924.69, 928.05, 1 112.72, 1 270.91, 1 365.3, 1 402.42, 1 453.06, 1 547.69 nm。

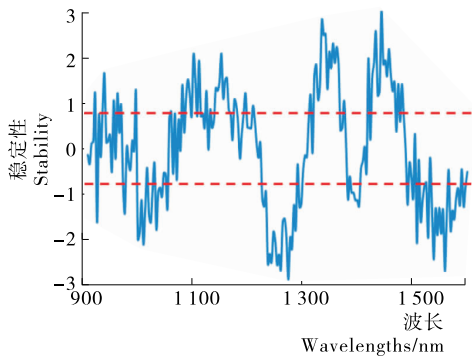


图 4 光谱变量经 MC-UVE 方法处理的稳定性分布图
Figure 4 The stability distribution of variables processed by MC-UVE

对比两次优选的特征波长分布,通过 MC-UVE 实施之后通过 SPA 选择的特征波长在光谱曲线中分布更为均匀,直接实施 SPA 优选出的特征波长主要集中在 900~1 000 nm 和 1 300~1 600 nm。仅有 3 个波长在两次选择中都被选择出来,分别是 924.69, 928.05, 1 453.06 nm。

2.4 SSC 预测模型

分别将优选后的波长作为 PLS 模型输入,建立 SSC 的检测模型,结果见表 2。其中,结果最好的为 MC-UVE-PLS 模型, r_{pre} 最高为 0.950,同时 RMSEP 最低为 0.934。说明 MC-UVE 能够有效去除无信息变量,提高检测精度。MC-UVE-SPA-PLS 模型 r_{pre} 为 0.942,较 MC-UVE-PLS 模型有少量降低,同时 RMSEP 有少许升高,为 1.009 °Brix,该模型预测效果降低程度微小,但是波长数目从 120 个减少到 8 个,显著降低,这可为在线检测设备的研究提供依据。MC-UVE-SPA-PLS 模型预测值与真实值的 PLSR 模型拟合曲线见图 5。

表 2 不同波长优选方法建模结果

Table 2 The results of different wavelength optimization methods

模型	特征波长数目	建模集		校正集	
		r_{cal}	RMSEP	r_{pre}	RMSEP
Raw-PLS	205	0.976	0.628	0.946	0.965
SPA-PLS	10	0.932	1.059	0.914	1.211
MC-UVE-PLS	120	0.973	0.662	0.950	0.934
MC-UVE-SPA-PLS	8	0.952	0.885	0.942	1.009

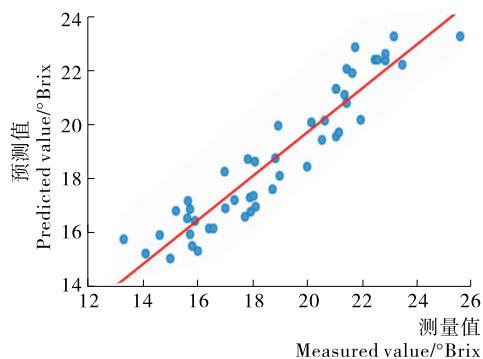


图 5 MC-UVE-SPA-PLS 模型预测集预测值与真实值相关系数图

Figure 5 Scatter plot of predicted versus measured SSC obtained by MC-UVE-SPA-PLS

3 结论

本试验应用近红外高光谱成像技术对柿子内部 SSC 含量进行无损检测研究,探讨了 MC-UVE 和 SPA 方法对特征波长优选效果。结果显示,通过 MC-UVE-SPA 从 205 个波长中优选出 8 个特征波长,建立的 SSC 检测 PLS 模型 $r_{pre} = 0.942$, $RSMEP = 1.009$ °Brix, 优于直接使用 SPA 优选出的 10 个特征波长,但略低于利用 MC-UVE 优选出的 120 个波长作为输入变量的 SSC 检测 PLS 模型。综合考虑波长数目与 SSC 检测结果,MC-UVE-SPA 联用更适用于特征波长优选。本研究的不足之处是所选样本数相对较少,所建立的模型只适合于该品种涩柿 SSC 含量的预测,为了提高模型的稳定性和通用性,后续研究还需增大样本量。

参考文献

- [1] BALTACIOGLU H, ARTIK N. Study of postharvest changes in the chemical composition of persimmon by HPLC[J]. Turkish Journal of Agriculture and Forestry, 2013, 37(5): 568-574.
- [2] SAFIYARI H, RAHMANIAN H, SALMANIZADE F, et al. Some physical and mechanical properties of persimmon fruit[J]. Mathematische Nachrichten, 2014, 284(13): 1 701-1 714.
- [3] HIRAI S, ROKUHARA S, SHIMIZU S. Changes of invertase activity in Japanese persimmon fruits during maturation, storage and processing[J]. Journal of Japanese Society of Food Science & Technology, 2010, 33(6): 369-374.
- [4] SUZUKI T, NIKAWA T, SAKURAI N. Evaluation of flesh texture and prediction of optimum ripeness in 'Fuyu' persimmon after harvest[J]. Horticultural Research, 2011, 10(3): 421-427.
- [5] MOHAMMADI V, KHEIRALIPOUR K, GHASEMI-VARNAMKHAHASTI M. Detecting maturity of persimmon fruit based on image processing technique[J]. Scientia Horticulturae, 2015, 184: 123-128.
- [6] ASHTIANI S H M, SALARIKIA A, GOLZARIAN M R, et al. Non-destructive estimation of mechanical and chemical properties of persimmons by ultrasonic spectroscopy[J]. International Journal of Food Properties, 2016, 19(7): 1 522-1 534.
- [7] ZHANG Peng, XUE Ying, LI Jiang-kuo, et al. Research on

nondestructive measurement of firmness and soluble tannin content of 'mopanshi' persimmon using Vis/NIR diffuse reflection spectroscopy[J]. Acta Horticulturae, 2013, 996: 447-452.

- [8] NOYPITAK S, TERDWONGWORAKUL A, KRISANAPOOK K, et al. Evaluation of astringency and tannin content in 'Xichu' persimmons using near infrared spectroscopy[J]. International Journal of Food Properties, 2015, 18(5): 1 014-1 028.
- [9] 吴方龙, 沈黄通, 吴晨凯, 等. 基于近红外漫反射光谱无损检测梨可溶性固形物的光强影响探究[J]. 光谱学与光谱分析, 2013, 33(10): 2 671-2 674.
- [10] 刘燕德, 张光伟. 高光谱成像技术在农产品检测中的应用[J]. 食品与机械, 2012, 28(5): 223-226.
- [11] WEI Xuan, LIU Fei, QIU Zheng-jun et al. Ripeness classification of astringent persimmon using hyperspectral imaging technique[J]. Food & Bioprocess Technology, 2014, 7(5): 1 371-1 380.
- [12] MUNERA S, BESADA C, ALEIXOS N, et al. Nondestructive assessment of the internal quality of intact persimmon using colour and VIS/NIR hyperspectral imaging[J]. LWT-Food Science and Technology, 2017, 77: 241-248.
- [13] CORTES V, RODRIGUEZ A, BLASCO J, et al. Prediction of the level of astringency in persimmon using visible and near-infrared spectroscopy[J]. Journal of Food Engineering, 2017, 204: 27-37.
- [14] 思振华, 何建国, 刘贵珊, 等. 基于高光谱图像技术羊肉表面污染无损检测[J]. 食品与机械, 2013, 29(5): 75-79.
- [15] ARAUJOM C U, SALDANHAT C B, GALVAO R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis[J]. Chemometrics & Intelligent Laboratory Systems, 2001, 57(2): 65-73.
- [16] 郝勇, 孙旭东, 潘圆媛, 等. 蒙特卡罗无信息变量消除方法用于近红外光谱预测果品硬度和表面色泽的研究[J]. 光谱学与光谱分析, 2011, 31(5): 1 225-1 229.
- [17] XU Qing-song, LIANG Yi-zeng, DU Yi-Ping. Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration[J]. Journal of Chemometrics, 2004, 18(2): 112-120.
- [18] FELDICK A, MODEST M F. An improved wavelength selection scheme for Monte Carlo solvers applied to hypersonic plasmas[J]. Journal of Quantitative Spectroscopy & Radiative Transfer, 2011, 112(8): 1 394-1 401.
- [19] 徐爽, 何建国, 易东, 等. 基于高光谱图像技术的长枣糖度无损检测[J]. 食品与机械, 2012, 28(6): 168-170.
- [20] 介邓飞, 杨杰, 彭雅欣, 等. 基于高光谱技术的柑橘不同部位糖度预测模型研究[J]. 食品与机械, 2017, 33(3): 51-54.
- [21] 赵文英, 花锦, 张梨花, 等. 近红外光谱测定不同鲜肉类糜中蛋白质含量[J]. 食品与机械, 2017, 33(1): 48-50, 118.
- [22] LI Jiang-bo, TIAN Xi, HUANG Wen-qian, et al. Application of long-wave near infrared hyperspectral imaging for measurement of soluble solid content (SSC) in pear[J]. Food Analytical Methods, 2016, 9(11): 1-12.
- [23] GUO Wen-chuan, ZHAO Fan, DONG Jin-lei. Nondestructive measurement of soluble solids content of Kiwifruits using near-infrared hyperspectral imaging[J]. Food Analytical Methods, 2016, 9(1): 38-47.