

基于透射光谱技术的温州蜜柑含水率检测

余怀鑫¹, 潘新星², 朱旭¹, 叶子凡¹, 祝志慧¹, 刘继红³

1. 华中农业大学工学院, 武汉 430070; 2. 浠水县长福水稻专业合作社, 浠水 438200;

3. 华中农业大学园艺林学学院, 武汉 430070

摘要 利用可见/近红外光谱透射技术检测温州蜜柑含水率。采用微分处理(differential processing, SD)、多元散射校正(multivariate scattering correction, MSC)、标准正态变换(standard normal variate, SNV)、SG卷积平滑以及标准化等预处理方法比较建立的偏最小二乘回归模型(partial least squares regression, PLS)的拟合准确度, 并确定最佳预处理方法, 同时采用竞争性自适应重加权采样算法(competitive adaptive reweighted sampling algorithm, CARS)提取特征波长, 以此建立基于柑橘含水率的PLS模型、BP神经网络模型和最小二乘支持向量机模型(least squares support vector machine, LSSVM)。结果显示, 使用经过SNV预处理后的光谱进行CARS筛选得到的359个波长建立的LSSVM模型预测效果最佳, 校正集的相关系数和均方根误差分别为0.937 5和0.008 6, 验证集相关系数和均方根误差分别为0.831 6和0.012 0, 表明可见/近红外光谱技术用于温州蜜柑的含水率检测是可行的。

关键词 可见/近红外光谱; 温州蜜柑; 无损检测; 含水率; 透射光谱技术; 品质分级

中图分类号 S 513 **文献标识码** A **文章编号** 1000-2421(2021)01-0086-07

含水率是衡量柑橘品质的重要指标之一, 水分丰富的柑橘更受市场青睐。温州蜜柑是我国柑橘栽培的重要鲜食品种, 至今已有2 400多年的栽培历史, 果肉及果汁具有解热生津、开胃、利尿、祛痰止咳的功效, 橘皮及络可做中药, 具有很大的经济价值。目前对温州蜜柑等水果的外部品质检测方法主要有光电分选法, 通常用于外部品质的分级, 通过分析水果外部的特征以大小、颜色、缺陷等作为分选的依据。而内部品质的检测方法还不成熟, 主要有人工试吃和通过化学实验进行分析两种, 前者存在很强的主观性, 后者检测时间长、成本高且为有损检测^[1], 无法满足温州蜜柑含水率无损检测和实现品质的快速评定分级的需求, 给消费者带来了极大的不便。因此, 对温州蜜柑含水率进行快速无损检测具有重要意义。

近些年来, 光谱技术的出现为无损检测提供了良好的手段, 可以进行定量和定性检测^[2-3]。许多学者利用可见/近红外光谱对柑橘品质进行了无损检测, 主要的研究指标为柑橘的可溶性固形物、糖度、

酸度以及维生素C含量等^[4-8], 对柑橘水分研究甚少。但也有学者通过高光谱/近红外光谱对梨^[9]、黑宝石李^[10]和南疆红枣^[11]等进行了水分无损检测, 均取得了较好的效果。光谱技术作为一种新兴的无损检测技术, 具有效率高、成本低且易于实现自动化生产的特点^[12]。本研究利用光谱技术对温州蜜柑的含水率进行研究, 旨在考察基于光谱技术检测温州蜜柑含水率的可行性。

1 材料与方法

1.1 试验材料

2019年11月份从湖北省武汉市农业科学院林业果树研究所采摘的新鲜温州蜜柑样本40个, 样本的最大横向直径45~60 mm、质量为60~120 g/个。将样本用抹布清理掉表面污渍, 置于常温下、通风处储存, 让其自然风干。每2 d采集1次透射光谱, 并用精度为1 mg的电子天平称质量, 共采集10次, 最后用苏珀电热恒温鼓风干燥箱进行脱水, 根据GB 5009.3—2016方法测定40个样本在试验

收稿日期: 2020-07-15

基金项目: 湖北省技术创新专项(2017ABA158)

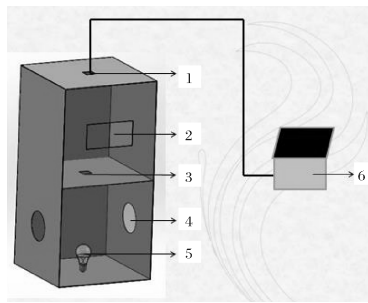
余怀鑫, E-mail: 1337362346@qq.com

通信作者: 祝志慧, E-mail: zzh@mail.hzau.edu.cn

过程中的含水率,共得到 400 组数据。测量过程如下:用精度为 1 mg 的电子天平对试验当天的样本称质量并记录,待到 19 d 试验结束后,将全部样本放入烘干箱中以 105 ℃ 烘干 72 h,再记录烘干后样本的质量。含水率=(初始质量-烘干后质量)/初始质量×100%。

1.2 光谱采集

采用如图 1 所示的光谱采集暗箱,暗箱上端 1 处安装光谱采集探头,与电脑相连,下部中间位置安装光源,由于需要采集透射光谱,而柑橘果皮厚度不一,故要选用穿透性强、能量强度高的光源。本次光源选用的是 MR16 卤素灯,卤素灯与水平面垂直朝向中央采集窗口。由于 MR16 卤素灯功率大、温度高,导致暗箱内部散热差,易使设备老化,故在暗箱下端两侧安装散热风扇进行散热。光谱仪采用的是海洋光学的 Maya2000pro 光谱仪,光谱仪参数设置:扫描范围为 200~1 100 nm,平滑次数为 5,积分时间为 100 ms,得到的每条光谱含有 2 068 个变量。



1.光纤探头 Optical fiber probe; 2.取物口 Sampling port; 3.样本放置口 Sample placing port; 4.散热风扇 Cooling fan; 5.光源 Light source; 6.电脑 Computer.

图 1 光谱采集暗箱

Fig.1 Spectral acquisition darkbox

1.3 原始光谱特性分析

图 2 为使用 Maya2000pro 光谱仪采集同一个样本的果肉和完整果实的透射光谱。由图 2 可知,果肉的透过率比完整果实透过率高,是因为果皮的存在导致的。果肉和完整果实光谱图在 680、720、780 nm 处均出现峰值,光谱相似度较高。随着柑橘的逐渐成熟,叶绿素会逐渐转化为类胡萝卜素与类黄酮,这也是柑橘表皮颜色出现变化的原因,在 400~700 nm 这一部分主要是受到叶绿素的影响,而在 700~800 nm 之间出现高吸收峰则是由类胡萝卜素造成的,这也是完整果实的透射光谱吸收强

度远比果肉大的原因。由于需要进行无损检测,因此相比其果肉,选择完整果实的透射光谱进行建模是可行的。

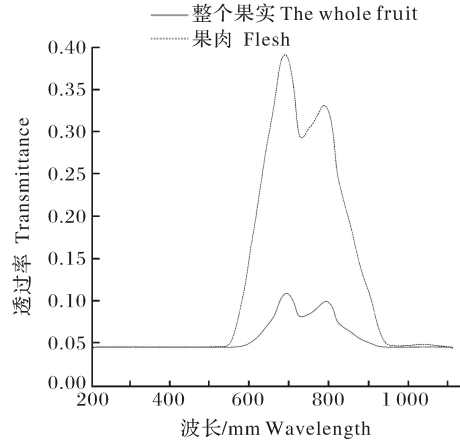


图 2 样本完整果实和果肉的透射光谱图

Fig.2 The transmission spectrum of intact fruit and pulp of the same sample

1.4 水分变化情况

在未进行试验前,温州蜜柑的含水率无法得知,由于是同一批采摘的蜜柑,其含水率差异不大,为了扩大试验样本的含水率范围,采用常温保存让其自动失水的方法。每隔 2 d 测量温州蜜柑的质量,得到其含水率均值变化趋势(图 3)。由图 3 可知,随着时间的变化,温州蜜柑的水分含量不断减少,且 40 个样本的含水率变化趋势大致相同,表明在该试验条件下该批柑橘的变化趋势一致。当进行第 10 次采集时,即试验进行至第 19 天时,柑橘表皮已经出现了明显的硬化,表明柑橘在试验期间散失了一

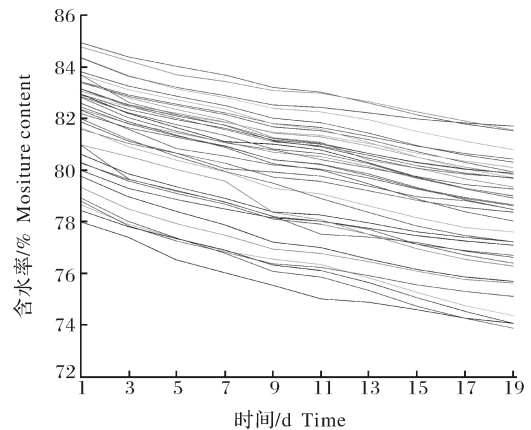


图 3 温州蜜柑水分随时间变化

Fig.3 Satsuma mandarin content change diagram of reed with time

定水分,从试验结果来看,水分平均散失了 4~5 个百分点,水分梯度也进一步扩大,表明通过常温储存的方法扩大含水率梯度是可行的。在蜜柑贮存过程中存在呼吸作用等影响内部变化的因素,会造成蜜柑质量的变化,但对质量影响最大的还是水分,其他物质如糖、酸、类黄酮等含量变化较小,因此不考虑其对含水率测量结果的影响。

1.5 光谱处理

1) 光谱预处理。由于仪器对于外界环境的变化比较敏感,采集到的光谱中会含有部分的噪声信息,采用合适的预处理方法能够有效地消除噪声及特定物理因素的干扰,确保光谱信息的准确性^[13]。因此,将无预处理和 SD、MSC/SNV、SG 平滑、标准化 5 种预处理方法根据不同预处理的作用按照表 1 进行排列组合(其中 0 表示不进行预处理),共得到 $2 \times 3 \times 2 \times 2 = 24$ 种预处理方式,用数字 1~24 对所有组合进行编号,最后根据预处理后建立的 PLS 模型的建模效果,得到较优的预处理方法。

表 1 预处理方法

Table 1 Pretreatment methods

基线校正 Baseline correction	散射校正 Scatter correction	平滑 Smooth	尺度缩放 Scale scaling
0	0	0	0
SD	SNV MSC	SG	标准化 Standardization

2) 光谱特征波长选择。采集的光谱波段为 200~1 100 nm,光谱范围较宽、波段数目较多(2 068 维),部分波段存在较强的相关性,原始光谱除了包含反映内部成分差异的信息外还含有大量冗余信息,分类时可能会产生“维数灾难”现象^[14],且会影响分类的精度。因此降维就显得十分必要,本研究采用 CARS 来筛选特征波长以达到降维目的。竞争性自适应重加权采样算法(CARS)采用“适者生存”法则,利用蒙特卡罗采样(MCS)法采样 N 次,对比每次产生的 PLS 回归模型的交互验证均方差值(root mean square error of cross validation, RMSECV),RMSECV 值最小的那个模型所对应的变量子集被选为最优变量子集^[15]。

1.6 模型的建立及评价指标

PLS 分析是一种新型的多元数据分析方法,通过投影将预测变量和观测变量投影到一个新的空间之中以寻找一个线性回归的模型,集主成分分析、典

型相关分析以及多元线性回归于一身,因此被广泛应用于多种分析统计领域。BP 神经网络主要由输入层、隐含层和输出层组成,输入层的每个节点对应一个自变量,输出层的每个节点对应 1 个应变量,而各层的节点数则代表了该模型的训练效果。当节点越多,得到的信息会越全面,但训练时间也会随之延长。LSSVM 采用最小二乘线性系统作为损失函数,替代传统的支持向量机所用的二次规划方法,简化了计算的复杂性,是一种可以同时处理线性和非线性信息的建模方法^[16]。

本研究将柑橘的含水率测量值作为因变量,光谱的波长值作为自变量,建立 PLS、BP 以及 LSSVM 模型。建立模型的好坏由训练集相关系数(R_c)、训练集均方根误差(RMSEC)、测试集相关系数(R_p)以及测试集均方根误差(RMSEP)决定,其中相关系数越接近 1,均方根误差越小,模型的效果就越好,上述计算均由 MATLAB 2014a 完成。

2 结果与分析

2.1 光谱预处理

将试验得到的 400 个温州蜜柑样本的初始可见/近红外光谱通过马氏距离法剔除异常样本,如图 4 所示,位于绿色实线上方的样本被当做异常样本剔除,共 23 个。为了将差异性大的样本划入训练集中,保证训练集的样本具有较强的代表性,采用 Kennard-Stone 算法^[17]以 3:1 划分校正集和验证集,并将 24 种排列组合预处理后的光谱建立基于含水率的 PLS 模型,得到的建模结果如图 5 所示。

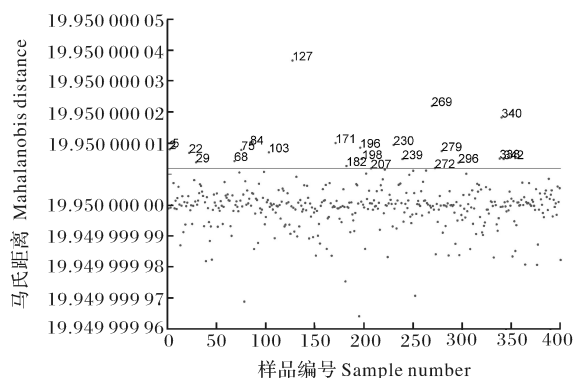


图 4 马氏距离法剔除异常样本

Fig.4 Eliminating abnormal samples by Mahalanobis distance method

从图 5 可知,当采用 FD(编号 13~24)后,训练集相关系数较高,测试集相关系数较低,出现了过拟合现象。通过对比相关系数(越大越好)和均方根误差(越小越好)可以发现,相较于第 1 组(无预处理),2~12 组建模结果均有所提升,表明有效的预处理可以去除外界干扰信息,提高建模效果,其中建模效果最好的是第 5 组,采用的预处理方式为 SNV。

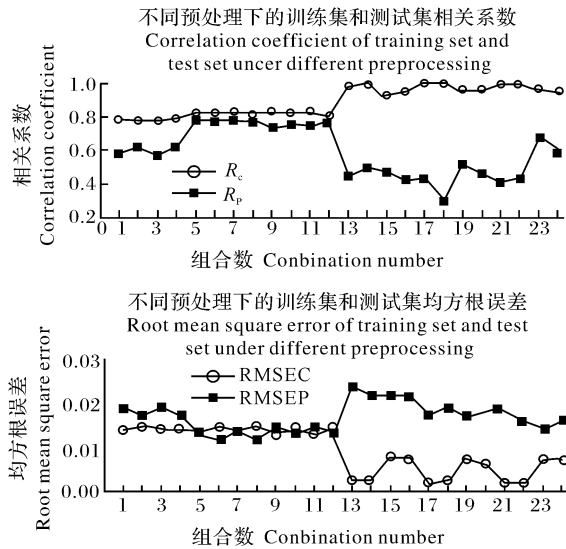


图 5 24 种不同预处理组合方式下的 PLS 建模结果
Fig.5 Results of PLS modeling under 24 different pretreatment combinations

2.2 特征波长提取

本研究采用的特征波长的提取方法为 CARS,采样次数设定为 50 次,交叉验证设定为 5 次,以建立 LSSVM 为例,特征波长提取的过程如图 6 所示,其中横坐标均为采样次数。从图 6A 可以看出,选用的变量数先急剧减少后趋于稳定,这是指数衰减函数造成的;图 6B 表示 RMSECV 值的变化情况,在采样第 7 次时为最小值,表明采样第 7 次时剔除后剩余的波长点与温州蜜柑含水率相关性较大;

图 6C 中各线是随着采样次数的增加回归系数的变化,图 6 中的“*”线为 RMSECV 最小值处,与图 6B 相对应,利用 CARS 算法进行多次筛选后最终确定第 7 次筛选出的 359 个特征波长变量作为最终入选的特征值。经过 CARS 处理后可以极大地减少参与建模的波数,提高建模效率。

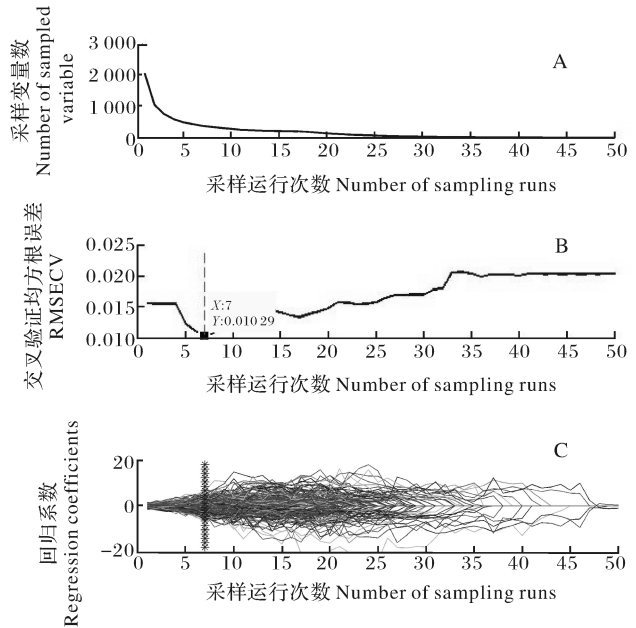


图 6 CARS 筛选特征波长过程

Fig.6 Selection of characteristic wavelengths by car

2.3 建模与验证

将经过 SNV 预处理后的光谱利用 CARS 算法提取特征波长,分别建立 PLS 模型、BP 神经网络模型以及 LSSVM 模型。BP 神经网络模型选用 LM 函数作为训练函数,并确定隐含层节点个数为 10; LSSVM 模型中选用径向基函数(RBF)作为核函数,因其能够实现非线性映射,且相比于多项式核函数它所需参数较少,模型复杂程度较低,计算速度较快。核函数参数为 100,惩罚因子为 30,3 个模型的建模结果如表 2 所示。

表 2 基于 CARS 的不同模型建模结果

Table 2 Modeling results of different models based on CARS

模型 Model	特征数 Feature number	训练集 Training set		测试集 Testing set	
		R_c	RMSEC	R_p	RMSEP
PLS	277	0.838 8	0.012 9	0.823 7	0.012 9
BP	169	0.883 9	0.039 2	0.720 1	0.085 5
LSSVM	359	0.937 5	0.008 6	0.831 6	0.012 0

由表 2 可知,效果最佳的是经过 CARS 提取特征波长后建立的 LSSVM 模型,经过 CARS 筛选后参与建模的波数由 2 068 个减少为 359 个,仅占原始波数的 17.36%,提高了模型的运算效率,该模型的校正集相关系数和均方根误差分别为 0.937 5 和

0.008 6,验证集相关系数和均方根误差分别为 0.831 6和 0.012 0。图 7 是 LSSVM 含水率预测模型效果图,表明利用 SNV+CARS+LSSVM 检测温州蜜柑水分是可行的。

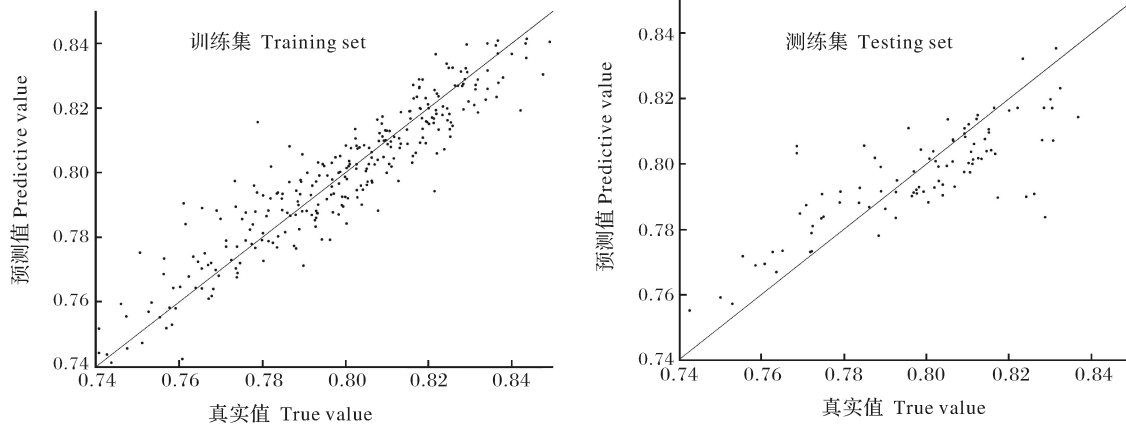


图 7 LSSVM 含水率预测模型

Fig.7 Moisture perdition model based on LSSVM

3 讨论

水分作为柑橘品质的一项重要指标,对采收后柑橘的贮存和加工具有重要的意义。前人对柑橘品质检测的研究重点集中在糖度、酸度、维生素 C 等营养物质方面,对水分的研究大都集中在柑橘种植期的水分供给方面。本试验对温州蜜柑进行了水分检测,得到的测试集相关系数为 0.831 6,与文献[18]预测砂糖橘水分结果相近,但相比其 0.82~0.87 的水分梯度^[18],本试验通过自然贮存的方式将含水率梯度扩大到了 0.73~0.85,提高了模型的适应性,且能够满足蜜柑水分的检测要求,为便携式仪器的开发提供了理论基础。后续研究可以将不同品种、种植区域的蜜柑进行试验,并比较不同成熟期蜜柑的水分差异性,实现对不同成熟期柑橘的预测,探究通过控制烘干时长来进一步扩大含水率范围,并改进光谱特征提取方法以及建模方法来提高模型的准确性和适应性。同时可以探究不同贮藏期水分含量的变化趋势,便于根据水分信息合理安排柑橘的贮藏条件与时间,有利于提高整体经济效益。

参考文献 References

[1] 张静.温州蜜柑和几种晚熟柑橘理化品质及功能成分研究[D].重庆:西南大学,2019. ZHANG J.Study on physical and chemi-

cal quality and functional components of Wenzhou tangerine and several late-maturing citrus[D].Chongqing:Southwest University,2019(in Chinese with English abstract).

- [2] 周娇娇,吴潇扬,陈周,等.近红外光谱技术快速预测团头鲂新鲜度[J].华中农业大学学报,2019,38(4):120-126.ZHOU J J, WU X Y, CHEN Z, et al.Rapid prediction of *Megalobrama amblycephala* freshness by near infrared spectroscopy[J]. Journal of Huazhong Agricultural University, 2019, 38(4): 120-126(in Chinese with English abstract).
- [3] 周娇娇,徐文杰,许竞,等.基于近红外光谱与 KPCA-SVM 鉴别淡水鱼肉[J].华中农业大学学报,2019,38(5):98-104. ZHOU J J, XU W J, XU J, et al.Identification of freshwater fish meat based on near infrared spectroscopy and KPCA-SVM[J].Journal of Huazhong Agricultural University,2019,38(5): 98-104(in Chinese with English abstract).
- [4] 许文丽.基于可见/近红外光谱的江西特色柑橘可溶性固形物检测研究[D].南昌:江西农业大学,2014. XU W L.Research on the soluble solids detection of Jiangxi characteristic citrus based on visible/near infrared spectroscopy[D]. Nanchang: Jiangxi Agricultural University,2014(in Chinese with English abstract).
- [5] 夏俊芳,李小昱,李培武,等.基于小波变换的柑橘维生素 C 含量近红外光谱无损检测方法[J].农业工程学报,2007(6):170-174.XIA J F, LI X Y, LI P W, et al.Near-infrared spectroscopy non-destructive detection method for citrus vitamin C content based on wavelet transform[J].Transactions of the Chinese society of agricultural engineering,2007(6):170-174(in Chinese with English abstract).
- [6] 袁雷,刘辉军,余桂英,等.柑橘品质的近红外光谱无损检测

- [J].光谱实验室, 2006(4): 820-822. YUAN L, LIU H J, YU G Y, et al. Near infrared spectroscopy nondestructive testing of citrus quality[J]. Spectroscopy laboratory, 2006(4): 820-822 (in Chinese with English abstract).
- [7] 但松健. 基于 NIR 光谱分析的柑橘产地鉴别及品质检测技术研究[D]. 重庆: 重庆大学, 2017. DAN S J. Research on citrus origin identification and quality detection technology based on NIR spectral analysis[D]. Chongqing: Chongqing University, 2017 (in Chinese with English abstract).
- [8] ANTONUCCI F, PALLOTTINO F, PAGLIA G, et al. Non-destructive estimation of mandarin maturity status through portable VIS-NIR spectro-photometer[J]. Food and bioprocess technology, 2011, 4(5): 809-813.
- [9] 洪添胜, 乔军, MICHAEL O N, 等. 基于高光谱图像技术的雪花梨品质无损检测[J]. 农业工程学报, 2007(2): 151-155. HONG T S, QIAO J, MICHAEL O N, et al. Non-destructive testing of snow pear quality based on hyperspectral image technology[J]. Journal of agricultural engineering, 2007(2): 151-155 (in Chinese with English abstract).
- [10] 蔡宋宋, 王宝刚, 冯晓元, 等. 黑宝石李品质参数的近红外漫反射无损检测模型的建立[J]. 食品工业科技, 2010, 31(1): 369-372. CAI S S, WANG B G, FENG X Y, et al. Establishment of a model for the near-infrared diffuse reflection non-destructive testing of the quality parameters of black gem plum[J]. Food industry science and technology, 2010, 31(1): 369-372 (in Chinese with English abstract).
- [11] 彭云发, 彭海根, 詹映, 等. 近红外光谱对南疆红枣水分无损检测的研究[J]. 食品科技, 2013, 38(11): 260-263. PENG Y F, PENG H G, ZHAN Y, et al. Near infrared spectroscopy study on non-destructive testing of red Jujube moisture in southern Xinjiang[J]. Food science and technology, 2013, 38(11): 260-263 (in Chinese with English abstract).
- [12] 刘燕德, 程梦杰, 郝勇. 光谱诊断技术及其在农产品质量检测中的应用[J]. 华东交通大学学报, 2018, 35(4): 1-7. LIU Y D, CHENG M J, HAO Y. Spectral diagnostic technology and its application in the quality detection of agricultural products[J]. Journal of East China Jiaotong University, 2018, 35(4): 1-7 (in Chinese with English abstract).
- [13] 第五鹏瑶, 卞希慧, 王姿方, 等. 光谱预处理方法选择研究[J]. 光谱学与光谱分析, 2019(9): 2800-2806. DIWU P Y, BIAN X H, WANG Z F, et al. Study on the selection of spectral preprocessing methods[J]. Spectroscopy and spectral analysis, 2019(9): 2800-2806 (in Chinese with English abstract).
- [14] 刘浩然, 赵翠香, 李轩, 等. 一种基于改进遗传算法的神经网络优化算法研究[J]. 仪器仪表学报, 2016, 37(7): 1573-1580. LIU H R, ZHAO C X, LI X, et al. Research on a neural network optimization algorithm based on improved genetic algorithm[J]. Journal of instrumental science, 2016, 37(7): 1573-1580 (in Chinese with English abstract).
- [15] 吴静珠, 徐云. 基于 CARS-PLS 的食用油脂脂肪酸近红外定量分析模型优化[J]. 农业机械学报, 2011, 42(10): 162-166. WU J Z, XU Y. Optimization of the near-infrared quantitative analysis model of fatty acid in edible oil based on CARS-PLS[J]. Journal of agricultural machinery, 2011, 42(10): 162-166 (in Chinese with English abstract).
- [16] 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: 136-147.
- [17] ZHAN X R, ZHU X R, SHI X Y, et al. Determination of hesperidin in tangerine leaf by near-infrared spectroscopy with spxy algorithm for sample subset partitioning and monte carlo cross validation[J]. Spectroscopy and spectral analysis, 2009, 29: 964-968.
- [18] 代芬, 黄冠勇, 洪添胜. 基于光谱技术的砂糖橘品质无损检测方法研究[C]//中国农业工程学会(CSAE). 中国农业工程学会 2011 年学术年会论文集. 北京: 中国农业工程学会(CSAE), 2011: 1646-1652. DAI F, HUANG G Y, HONG T S. Study on nondestructive testing method of sugar orange quality based on spectral technology[C]//Chinese Society of Agricultural Engineering (CSAE). Proceedings of 2011 academic annual meeting of Chinese Society of Agricultural Engineering. Beijing: Chinese Society of Agricultural Engineering (CSAE), 2011: 1646-1652.

Detecting water content of Satsuma mandarin based on transmission spectroscopy

YU Huaixin¹, PAN Xinxing², ZHU Xu¹, YE Zifan¹, ZHU Zhihui¹, LIU Jihong³

1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China;

2. Hubei Changfu Rice Professional Cooperative Agency, Xishui 438200, China;

3. College of Horticulture and Forestry, Huazhong Agricultural University, Wuhan 430070, China

Abstract Content of citrus is one of the important factors affecting storage and processing of citrus.

Real-time detection of water content of citrus can guarantee the quality of citrus. The visible/near infrared spectroscopy technology, as an increasingly mature non-destructive testing method, can effectively detect the water content of citrus. The Satsuma mandarin picked from the Institute of Forest and Fruit at Wuhan Academy of Agricultural Sciences were placed naturally at room temperature for reducing the water content of the Satsuma mandarin and increasing the water content gradient of the citrus. A spectrum collection device was built with Maya2000pro as a carrier. The score is obtained by dehydrating in a super electric heating constant temperature blast drying oven and calculated according to GB 5009.3—2016. The fitting accuracy of partial least squares regression (PLS) model established was compared through the use of differential processing, multivariate scattering correction, standard normal variate, SG convolution smoothing, standardization and other pretreatment methods. Results showed that SNV was the most effective preprocessing method. At the same time, a competitive adaptive reweighted sampling algorithm is used to extract characteristic wavelengths to establish a partial least square regression model, a BP neural network model and a least square support vector machine model based on water content of citrus. The results showed that the LSSVM model established with 359 wavelengths obtained by CARS screening using the spectrum after SNV preprocessing had the best prediction effect. The correlation coefficient and root mean square error of the calibration set are 0.937 5 and 0.008 6, respectively. The correlation coefficient of the verification set and the root mean square error are 0.831 6 and 0.012 0. The natural placement method at room temperature improves the water content gradient of Satsuma mandarin, thereby ensuring the adaptability of the model. It is indicated that the visible/near infrared spectroscopy technology is feasible for detecting the water content of Satsuma mandarin.

Keywords visible/near infrared spectroscopy; Satsuma mandarin; nondestructive testing; water content; transmission spectroscopy; quality classification

(责任编辑:陆文昌)