中国光学（中英文）
Chinese Optics
Ei 核心期刊

基于高光谱成像技术的涌泉蜜桔糖度最优检测位置

李斌 万霞 刘爱伦 邹吉平 卢英俊 姚迟 刘燕德

Optimal position for suger content detection of Yongquan honey oranges based on hyperspectral imaging technology

LI Bin, WAN Xia, LIU Ai-lun, ZOU Ji-ping, LU Ying-jun, YAO Chi, LIU Yan-de

引用本文:
李斌, 万霞, 刘爱伦, 邹吉平, 卢英俊, 姚迟, 刘燕德. 基于高光谱成像技术的涌泉蜜桔糖度最优检测位置[J]. 中国光学, 2024, 17(1): 128−139. doi: 10.37188/CO.2023−0057
LI Bin, WAN Xia, LIU Ai-lun, ZOU Ji-ping, LU Ying-jun, YAO Chi, LIU Yan-de. Optimal position for suger content detection of Yongquan honey oranges based on hyperspectral imaging technology[J]. *Chinese Optics*, 2024, 17(1): 128-139. doi: 10.37188/CO.2023-0057

在线阅读 View online: https://doi.org/10.37188/CO.2023−0057

文章编号　2097-1842(2024)01-0128-12

基于高光谱成像技术的涌泉蜜桔糖度最优检测位置

李　斌, 万　霞, 刘爱伦, 邹吉平, 卢英俊, 姚　迟, 刘燕德*
（华东交通大学 智能机电装备创新研究院 水果智能光电检测技术与设备国家与地方联合工程研究中心, 南昌 330013）

摘要: 本文旨在探索涌泉蜜桔糖度的最优检测位置和最佳预测模型, 以便为蜜桔糖度检测分级提供理论依据。本文利用波长为 390.2～981.3 nm 的高光谱成像系统对涌泉蜜桔糖度最佳检测位置进行研究, 将涌泉蜜桔的花萼、果茎、赤道和全局的光谱信息与其对应部位的糖度结合, 建立其预测模型。使用标准正态变量变换(SNV)、多元散射校正(MSC)、基线校准(Baseline)和 SG 平滑(Savitzkv-Golay)4 种预处理方法对不同部位的原始光谱进行预处理, 用预处理后的光谱数据建立偏最小二乘回归(PLSR)和最小二乘支持向量机(LSSVM)模型。找出蜜桔不同部位的最佳预处理方式, 对经过最佳预处理后的光谱数据采用竞争性自适应重加权算法(CARS)和无信息变量消除法(UVE)进行特征波长筛选。最后, 用筛选后的光谱数据建立 PLSR 和 LSSVM 模型并进行分析比较。研究结果表明, 全局的 MSC-CARS-LSSVM 模型预测效果最佳, 其预测集相关系数 Rp=0.955, 均方根误差 RMSEP=0.395, 其次是蜜桔赤道部位的 SNV-PLSR 模型, 其预测集相关系数 Rp=0.936, 均方根误差 RMSEP=0.37。两者预测集相关系数相近, 因此可将赤道位置作为蜜桔糖度的最优检测位置。本研究表明根据蜜桔不同部位建立的糖度预测模型的预测效果有所差异, 研究最优检测位置和最佳预测模型可以为蜜桔进行糖度检测分级提供理论依据。

关　键　词: 涌泉蜜桔; 高光谱; 糖度; 偏最小二乘回归; 最小二乘支持向量机
中图分类号: O433.4　　**文献标志码:** A　　**doi:** 10.37188/CO.2023-0057

Optimal position for suger content detection of Yongquan honey oranges based on hyperspectral imaging technology

LI Bin, WAN Xia, LIU Ai-lun, ZOU Ji-ping, LU Ying-jun, YAO Chi, LIU Yan-de*
(*Intelligent Electromechanical Equipment Innovation Research Institute*, *East China Jiaotong University*, *National and Local Joint Engineering Research Center of Fruit Intelligent Photoelectric Detection Technology and Equipment*, *Nanchang* 330013, *China*)
* *Corresponding author*, *E-mail*: *liuyande@ecjtu.jx.cn*

Abstract: The objective of this study is to explore the optimal detection location and the best prediction model of the suger level of Yongquan honey oranges, which can provide a theoretical basis for the brix measurement and classification of honey oranges. With the wavelength range of 390.2−981.3 nm hyperspectral imaging system was used to study the best position for detecting the sugar content of Yongquan honey oranges, and the spectral information of the calyx, fruit stem, equator and global of Yongquan honey oranges

收稿日期: 2023-03-30; **修订日期:** 2023-04-19
基金项目: 青年科学基金项目(No.12103019)
Supported by Youth Science Fund Projects (No.12103019)

were combined with their sugar content of corresponding parts to establish its prediction model. The original spectra from the different locations were pre-processed by Standard Normal Variance (SNV) transformation, Multiple Scattering Correction (MSC), baseline calibration (Baseline) and SG smoothing, respectively, and the Partial Least Squares Regression (PLSR) and Least Squares Support Vector Machine (LSSVM) models were established based on the pre-processed spectral data. The best pre-processing methods for different parts of the honey oranges were found, and the optimal spectral data obtained by the best pre-processing methods were conducted to identify characteristic wavelengths using the Competitive Adaptive Re-weighting Sampling algorithm (CARS) and Uninformative Variable Elimination (UVE). Finally, the PLSR and LSSVM models were established and compared based on the selected spectral data. The results show that the global MSC-CARS-LSSVM model demonstrates the most accurate prediction performance, with a correlation coefficient of *Rp*=0.955 and an RMSEP value of 0.395. Alternatively, the SNV-PLSR model of the equatorial location of honey oranges was found to be the next more effective, with a correlation coefficient of *Rp*=0.936, and an RMSEP value of 0.37. The correlation coefficients of the two prediction models are similar, the equatorial location can be used as the optimal position for measuring the sugar content of honey oranges. This study demonstrates that the prediction models based on different parts of the orange have different effects. Identifying the optimal position and prediction model can provide a theoretical basis for classifying oranges for sugar content testing.

Key words: Yongquan honey orange; hyperspectral; sugar content; partial least-squares regression; least-squares support vector machine

1　引　言

涌泉蜜桔产于浙江东南沿海地带,以果肉细嫩、皮薄易剥、无核、甜度高等特点深受人们喜爱,更是有“天下一奇,吃桔带皮”的美誉。浙江临海市常年种植蜜桔面积可达 20 万亩,蜜桔年产量近 30 万吨,涌泉作为临海蜜桔的核心产区,年产量约占其整个市蜜桔总产量的 1/4。涌泉蜜桔曾多次获得省部级优秀奖、浙江省名牌产品和中国名牌产品称号。随着物流和电商行业的兴起,涌泉蜜桔也被销往全国各地[1]。近年来,随着涌泉蜜桔产量的增加以及人们消费水平的逐渐提高,人们对蜜桔的品质方面的要求越来越严格。蜜桔含糖量的高低直接影响其口感和价格,因此,糖度是衡量其品质的一个重要指标[2-3]。此外,在对采摘后的蜜桔进行无损检测和分级时,其糖度也是重要参考指标之一。如何快速、准确和稳定地对其糖度进行检测和分级是当前蜜桔产业待解决的问题之一。在蜜桔的生长过程中,受生长环境的影响,其不同部位的糖度分布不同,因此在检测过程中,若不考虑检测位置则会对蜜桔糖度无损检测产生较大的影响,导致蜜桔的品质分级不够准确,所以有必要寻找出蜜桔糖度的最优检测位置,从而提高蜜桔糖度的无损检测精度[4-5]。

近年来,高光谱成像技术作为新一代光电检测技术,它集成像技术与光谱技术两者于一身,可以同时获取被检测对象的光谱和空间信息[6-7]。与其他研究方法相比,利用高光谱成像技术可以更全面、更深入且更具体地了解被检测对象[8-11]。该技术在农产品品质检测领域有着很大的潜力,现如今已被广泛应用于柑橘[12]、苹果[13]、猕猴桃[14]、葡萄[15]、茶叶[16] 等农产品的内部品质检测。Liu[17] 等在赤道部位建立了脐橙可溶性固化物(Soluble Solids Content, SSC)含量的预测模型,其预测集相关系数 Rp 为 0.90。许丽佳[14] 等利用高光谱成像技术对猕猴桃糖度进行无损检测。选择猕猴桃赤道位置进行糖度检测,其最优预测模型的预测集相关系数 Rp 为 0.839。Yang[18] 等采用传统的破坏性方法,将番茄切块榨汁后测量其 SSC 含量,经过高度和重量两个生理特征进行补

偿后的最优 SSC 预测模型的 Rp 为 0.91。以上研究并未考虑糖度检测位置对模型精度的影响，故所建立模型的精度不高。与此同时，也有部分学者就水果不同部位的糖度预测模型进行了研究。介邓飞[2]等利用高光谱技术对柑橘花萼、果梗和赤道部位进行检测，建立了不同部位糖度的预测模型，研究发现花萼部位对应的糖度预测模型效果最好，其预测集相关系数达到 0.950。袁琳[19]等利用近红外反射光谱对网纹瓜的不同部位进行光谱采集，并与对应部位的糖度结合建立了三个局部和全局 PLSR 预测模型，研究发现全局模型预测效果最佳，其预测集相关系数 Rp 为 0.889 5。以上两位学者均采用 PLSR 模型作为预测模型，未进行多种模型对比分析。

考虑到目前针对涌泉蜜桔不同部位的糖度预测模型的研究鲜有报道。本文利用高光谱成像技术对涌泉蜜桔糖度最优检测位置和最优糖度预测模型进行研究，分别测定蜜桔花萼、赤道、果茎 3 个部位的糖度，并结合这 3 个部位相应的光谱信息建立其局部糖度预测模型。在此基础上，将 3 个部位的平均光谱与平均糖度信息相结合，建立其全局模型。采用 4 种预处理方式（SNV、MSC、Baseline、SG）和两种建模方法（PLSR、LSSVM），分析比较找出蜜桔不同部位的最佳预处理方式。然后，通过特征波长筛选（CARS、UVE）进一步优化模型，从而提高其预测精度。最后，比较蜜桔各部位优化后的最佳糖度预测模型，找出蜜桔糖度的最优检测位置和预测模型。该研究结果不仅可以为蜜桔分级处理提供理论依据，而且可以为涌泉蜜桔检测分级的加工设备研制提供一定的研究基础。

2 实验材料与方法

2.1 实验材料

本研究选用涌泉蜜桔作为实验样本，购买于南昌水果市场，所选涌泉蜜桔大小接近，形状规则且外观完好。将买来的涌泉蜜桔进行检查和筛选，选取 120 个涌泉蜜桔洗净并逐个编号，在室温为 20 °C，相对湿度为 60% 的环境下储存 24 小时后采集蜜桔的光谱数据，避免温度对结果的影响。蜜桔花萼、赤道和果茎图像如图 1 所示。

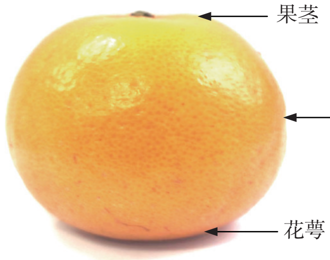


图 1 涌泉蜜桔不同部位图像

Fig. 1 Images of different parts of Yongquan honey oranges

2.2 实验装置与光谱数据采集

本次实验装置为高光谱成像光谱仪，装置示意图如图 2 所示。该系统主要由相机（C8484-05G 型，Hamamastu，日本）、光谱仪（ImSpector V10E 型，Specim，芬兰）、镜头、卤素灯（DE-COSTAR51 MR16 型，OSRAM，德国）、位移平台和计算机组成。本研究采集的光谱波长范围为 390.2 ~ 981.3 nm，分辨率为 3.3 nm。光谱采集装置需要提前预热 30 min，在实验正式开始前，还要对试验装置的各项参数进行调整，本次实验曝光时间设置为 6 ms，检测速度为 20 mm/s。每个实验样品均需采集花萼、果茎和赤道部位的光谱数据。

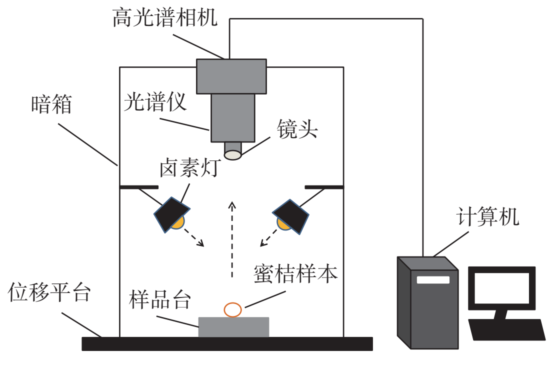


图 2 高光谱成像装置示意图

Fig. 2 Schematic diagram of the hyperspectral imaging device

由于高光谱相机存在暗电流和光源亮度分布不均匀现象，导致采集到的蜜桔不同部位的高光谱图像易受其影响，因此，在对光谱数据进行处理之前，需要对其原始光谱图像的反射率进行黑白校正。首先将镜头遮住，进行扫描，将得到的全黑图像作为黑色参考图像，再取下镜头遮盖物，对白色校准板进行扫描得到白色参考图像。利用两种参考图像进行校准，校准公式如下：

$$I = \frac{I_r - I_d}{I_w - I_d} \qquad (1)$$

式中: I_r为原始光谱图像; I_w为全白参考光谱图像; I_d为全黑参考光谱图像; I为校准后的光谱图像。利用 SpectraVIEW 软件对蜜桔的高光谱图像进行黑白校正, 然后通过 ENVI4.5 软件提取蜜桔不同部位的光谱数据, 在蜜桔不同部位选择一个矩形感兴趣区域(ROI), 计算出 ROI 区域的平均光谱值。

2.3　不同部位糖度的测量

将蜜桔的糖度值作为预测模型的真值, 对蜜桔样本采集光谱后, 进行糖度理化指标的测定。通过温度补偿糖度计(型号 PAL-1, Atago Co, Tokyo, Japan)测量其糖度值。该仪器的糖度测量范围为 0−53% $^{\mathrm{o}}$Brix。使用前用蒸馏水对糖度计进行零点校正。在蜜桔不同部位上切取果肉进行榨汁, 将果汁滴入糖度计中进行测量, 每测完一个部位后都要用蒸馏水进行清洗, 重复 3 次测量取平均值, 作为该部位的最终糖度值, 取 3 个部位糖度值的均值作为全局的糖度值。

2.4　光谱预处理

在采集原始光谱信息时, 易受到外界干扰, 从而出现随机噪声和表面散射等现象, 导致提取的光谱数据中存在许多干扰信息, 对模型的精度和稳定性有显著影响[20]。对光谱进行预处理能够消除这些不利影响, 从而提高其信噪比。本研究中采用了 SNV、MSC、Baseline 和 SG 四种预处理方式, 针对蜜桔不同部位的光谱信息挑选出最适合的预处理方式。

2.5　特长波长筛选

由于全光谱波长数据量很大, 并且存在着大量的无信息波长, 导致数据处理极其缓慢。因此, 有必要通过算法提取出最具有代表性的波长, 以构建更具稳定性和鲁棒性的糖度含量预测模型, 并简化建模过程。本研究采用的波长筛选方式为竞争性自适应重加权算法(CARS)和无信息变量消除法(UVE)。

其中, CARS 算法主要是利用自适应重加权采样技术和指数衰减函数从构建的 PLS(Partial Least Squares)子集模型中选出回归系数绝对值较大的变量, 然后通过交叉验证选取 RMSECV (Root Mean Square Error of Cross Validation)最小的子集中的变量作为特征波长[21-22]。UVE 是一种基于 PLS 回归系数稳定性分析的变量选择方法。它用于消除无用信息变量或冗余光谱变量[23]。其基本思想是将偏最小二乘回归系数作为波长重要性的衡量指标。

2.6　预测模型的建立及其评价

利用 Kennard-Stone(KS)方法将 120 个涌泉蜜桔样本进行分类, 其中校正集为 90 个, 预测集为 30 个。分别建立蜜桔花萼、果茎、赤道和全局的 PLSR 和 LSSVM 模型。偏最小二乘回归(PLSR)是一种线性回归方法。该方法在普通多元回归的基础上融合了主成分分析和典型相关的分析方法, 可解决变量之间的多重共线性问题[21]。最小二乘支持向量机(LSSVM)是一种非线性回归算法。该算法解决了经典 SVM(Support Vector Machines)中复杂的二次优化问题, 计算的复杂程度有所降低。非线性 LSSVM 预测模型对光谱和 SSC 中可能存在的非线性扰动有较好的鲁棒性[20]。模型性能主要是通过建模集相关系数 Rc、预测集相关系数 Rp、建模集均方根误差 RMSEC 和预测集均方根误差 RMSEP 4 个指标来评价, 其中 Rp 和 Rc 越接近 1 且 RMSEC 和 RMSEP 越小, 则表明该模型既精度高又稳定[24]。

3　结果与分析

3.1　涌泉蜜桔不同部位的光谱分析

本研究选择 390.2 ~ 981.3 nm 范围内的光谱进行分析。取蜜桔 3 个局部位置光谱的平均值作为其全局光谱。蜜桔赤道、果茎、花萼和全局的原始光谱如图 3(a)(彩图见期刊电子版)所示。由图 3(a)可以看出所有光谱曲线的变化趋势十分相似, 在 650 ~ 900 nm 波段范围内, 吸收峰不太明显, 波峰和波谷之间没有剧烈起伏。为了更加直观地观察和比较蜜桔不同部位的光谱信息, 计算得到蜜桔不同部位的平均光谱曲线, 如图 3(b)(彩图见期刊电子版)所示。由图 3(b)可知, 4 条光谱的光谱强度之间存在明显差异, 赤道位置的反射率高于其他部位。这可能与光的穿透深度和蜜桔内部糖度分布不均匀有关。从蜜桔赤道到花萼再到果茎, 果皮厚度依次增大, 光的穿透深度逐渐减小。同时这也表明光谱的检测位置会对光谱值产生影响。这一结论与先前研究者在对哈密瓜、西瓜和苹果不同部位进行光谱分析时给出的结论一致[20]。

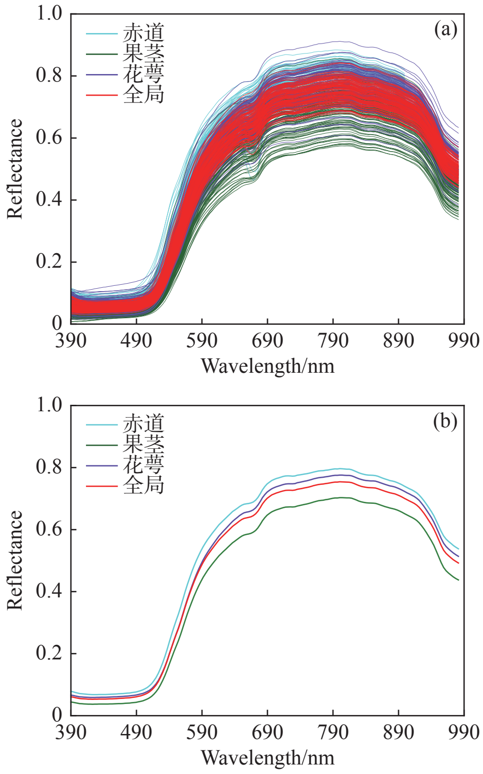


图 3 涌泉蜜桔光谱曲线。(a)不同部位的原始光谱曲线;(b)不同部位的平均光谱曲线

Fig. 3 Spectral curves of Yongquan honey orange. (a) Original spectral curves of different parts; (b) average spectral curves of different parts

3.2 涌泉蜜桔不同部位的糖度测量结果

120 个涌泉蜜桔样本不同部位糖度值的分布情况如表 1 所示。由表 1 可知，花萼、果茎、赤道、全局的糖度平均值分别为 15.2、14.2、14.5 和 14.6 ^{o}Brix，标准差分别为 1.39、1.52、1.37 和 1.34 ^{o}Brix。其中，蜜桔花萼部位糖度值大于果茎和赤道部位的糖度值，果茎部位的糖度值最低，说明蜜桔内部糖度分布是不均匀的。这种现象可能是由于不同部位中各种糖(蔗糖、葡萄糖、果糖)的含量不同导致的[25]。

表 1 涌泉蜜桔不同部位的糖度统计分析结果

Tab. 1 Statistical analysis of the sugar content of different parts of Yongquan honey orange

蜜桔部位	样本数	最大值/^{o}Brix	最小值/^{o}Brix	平均值/^{o}Brix	标准差/^{o}Brix
花萼	120	19.8	10.8	15.2	1.39
果茎	120	17.9	10.1	14.2	1.52
赤道	120	18.2	11.3	14.5	1.37
全局	120	17.8	11.2	14.6	1.34

3.3 涌泉蜜桔不同部位的全变量模型比较

为了比较不同建模方法之间的模型性能，此次研究采用了 PLSR 和 LSSVM 两种建模方法分别建立了蜜桔花萼、果茎和赤道部位的局部糖度预测模。为了更进一步评估其局部模型的性能，同时建立一个全局糖度预测模型作为对比。采用 SNV、MSC、Baseline 和 SG 对光谱进行预处理。基于不同模型和预处理方法的局部糖度预测模型和全局糖度预测模型的建模效果如表 2 和表 3 所示。

表 2 基于不同预处理方法的涌泉蜜桔糖度检测 PLSR 模型比较

Tab. 2 Comparison of PLSR models for detecting the sugar content of Yongquan honey orange based on different pretreatments

预测模型	预处理方法	建模集		预测集	
		R_C	RMSEC/^{o}Brix	R_P	RMSEP/^{o}Brix
花萼模型	Raw	0.946	0.384	0.893	0.457
	SNV	0.847	0.58	0.806	0.688
	MSC	0.832	0.622	0.766	0.564
	Baseline	0.921	0.409	0.890	0.518
	SG	**0.932**	**0.427**	**0.898**	**0.436**
果茎模型	Raw	0.949	0.428	0.859	0.587
	SNV	0.902	0.593	0.882	0.669
	MSC	0.889	0.599	0.864	0.587
	Baseline	**0.931**	**0.498**	**0.913**	**0.468**
	SG	0.943	0.455	0.868	0.569
赤道模型	Raw	0.932	0.471	0.861	0.553
	SNV	**0.946**	**0.408**	**0.936**	**0.370**
	MSC	0.960	0.365	0.878	0.458
	Baseline	0.964	0.349	0.933	0.384
	SG	0.924	0.497	0.861	0.555
全局模型	Raw	0.971	0.305	0.920	0.388
	SNV	0.945	0.403	0.901	0.435
	MSC	**0.953**	**0.374**	**0.934**	**0.435**
	Baseline	0.926	0.469	0.855	0.495
	SG	0.927	0.476	0.923	0.384

由表 2 可知，对于花萼部位模型来说，SG-PLSR 模型预测效果更佳，其模型 Rp 为 0.898，RMSEP 为 0.436 ^{o}Brix。对于果茎部位模型来说，预测效果最好的模型是 Baseline-PLSR 模

型, 其预测集相关系数 *Rp* 为 0.913, RMSEP 为 0.468 ^{o}Brix, 与花萼部位的模型相比, 其预测效果更好。对于赤道部位, SNV-PLSR 模型的预测效果最好, 预测集相关系数 *Rp* 为 0.936, RMSEP 为 0.37 ^{o}Brix。比较花萼、果茎和赤道的最优 PLSR 模型可以发现, 赤道部位最优模型的 *Rp* 最高, 预测效果最好。为了更加全面地探索出涌泉蜜桔糖度的最优检测位置, 将蜜桔花萼、果茎和赤道部位的光谱信息取平均值, 并与其对应部位的平均糖度相结合, 建立其全局模型。从全局 PLSR 模型分析结果看, 经过 MSC 预处理后的 PLSR 模型预测效果最好, 其 *Rp* 为 0.934, RMSEP 为 0.435 ^{o}Brix。其与赤道部位的 SNV-PLSR 模型预测集的相关系数相近。表明两个模型的预测效果差不多。

表 3　基于不同预处理的涌泉蜜桔糖度 LSSVM 模型比较

Tab. 3　Comparison of LSSVM models for detecting the sugar content of Yongquan honey orange based on different pretreatments

预测模型	预处理方法	建模集		预测集	
		R_C	RMSEC/^{o}Brix	R_P	RMSEP/^{o}Brix
花萼模型	Raw	0.921	0.470	0.860	0.513
	SNV	0.938	0.383	0.789	0.700
	MSC	0.959	0.323	0.788	0.539
	Baseline	0.942	0.360	0.869	0.585
	SG	**0.923**	**0.459**	**0.876**	**0.477**
果茎模型	Raw	0.979	0.286	0.782	0.750
	SNV	0.908	0.594	0.834	0.710
	MSC	**0.955**	**0.404**	**0.884**	**0.596**
	Baseline	0.924	0.527	0.642	0.854
	SG	0.953	0.419	0.827	0.650
赤道模型	Raw	0.965	0.355	0.829	0.594
	SNV	**0.954**	**0.388**	**0.906**	**0.405**
	MSC	0.973	0.315	0.827	0.530
	Baseline	0.979	0.281	0.867	0.544
	SG	0.956	0.388	0.845	0.575
全局模型	Raw	0.962	0.355	0.892	0.443
	SNV	0.972	0.296	0.897	0.456
	MSC	**0.980**	**0.253**	**0.946**	**0.400**
	Baseline	0.973	0.293	0.811	0.590
	SG	0.961	0.356	0.909	0.414

基于不同预处理方式建立蜜桔不同部位和全局 LSSVM 预测模型的预测结果如表 3 所示。由表 3 可知: 对于蜜桔花萼部位, 在其 LSSVM 模型中, 经 SG 预处理后的模型最佳, 其 *Rp* 为 0.876, RMSEP 为 0.477 ^{o}Brix; 对于蜜桔果茎部位模型来说, 经 MSC 预处理的 LSSVM 模型预测效果最佳, 其 *Rp* 为 0.884, RMSEP 为 0.596 ^{o}Brix; 对于蜜桔赤道部位的 LSSVM 模型来说, 最佳预处理方式为 SNV, 其最佳糖度预测模型的 *Rp* 为 0.906, RMSEP 为 0.405 ^{o}Brix。与其他两个部位最佳预测模型相比, 依旧是赤道部位预测模型的预测效果更佳。这可能是因为赤道位置比其他位置的日照时间长且温度高, 使得该部位的糖度比较高。对全局的 LSSVM 模型进行分析可知, MSC-LSSVM 模型是最优模型, 其 *Rp* 为 0.946, RMSEP 为 0.400 ^{o}Brix。

3.4　基于涌泉蜜桔不同部位特征变量的模型比较

由于全光谱中存在着大量冗余信息且波长之间存在相互干扰, 数据量大导致处理速度缓慢且建模过程复杂。利用算法提取出包含更多有效信息的特征波长, 可以降低光谱数据的维数, 加快建模速度及提高模型精度。本文选择的特征波长算法有 CARS 和 UVE。

3.4.1　CARS 特征波长筛选

以蜜桔果茎部位为例, 波长筛选过程和结果见图 4 和图 5(彩图见期刊电子版)。图 4 为特征波长的选择流程, 设置 MC 采样次数为 100。图 4(a) 表明随着采样次数的增加, 选择波长变量数逐渐减少, 减少速度为先快后慢。图 4(b) 表明, 随着采样次数的增加, RMSECV 值先减小后增加,当采样次数为 38 次时, RMSECV 达到最低值; 当采样次数小于 38时, RMSECV 值缓慢减小, 表明原始光谱中所含的冗余信息被剔除; 当采样次数大于 38 次时, RMSECV 值开始上升, 则表明光谱中有效特征波长被剔除, 模型性能变差。因此, 选择经过 38 次采样得到的变量作为建立果茎部位糖度预测模型的特征变量[26]。根据 3.3 的结论, 在蜜桔果茎部位模型中, 两种模型最佳预处理方法分别为 Baseline 和 MSC 预处理, 利用 CASR 算法对两种预处理后的光谱数据中的变量进行筛选, 筛选后的特征波长位置和数量如图 5 所示。分别筛选出 32 和 37 个特征变量, 分别占全波段的 18.9% 和 21.9%。

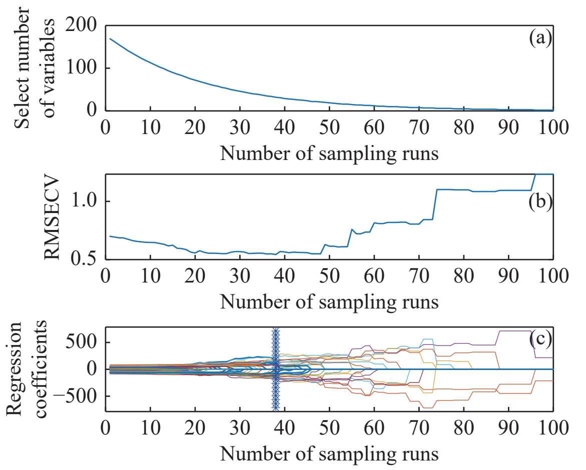


图 4　CARS 果茎部位特征波长选择过程。(a)变量数变化;(b)交叉验证均方根变化;(c)回归系数变化

Fig. 4　Selecting process of the characteristic wavelength of the fruit stem part by CARS. (a) Changes in number of variables; (b) changes in the RMSECV; (c) changes in regression coefficient

利用 CARS 算法对蜜桔花萼部位、赤道部位和全局光谱数据进行特征波长筛选,波长筛选过程与果茎部位类似。根据 3.3 结论可知,对于花萼部位、赤道部位和全局模型,最佳预处理方式分别为 SG、SNV 和 MSC。对预处理后的花萼部位光谱、赤道部位光谱和全局光谱分别进行特征波长筛选。筛选出的特征波长位置和数量如图 6 所示。从图 6 可以看出,不同部位对应的特征波长位置和数量均不相同。这表明不同部位糖度的光谱特征信息是不同的。分别筛选出了 48、24 和 34 个特征变量建立蜜桔糖度预测模型。它们分别占全波段的 28.4%、14.2% 和 20%,大部分特征波长位于 650 ~ 900 nm 之间。

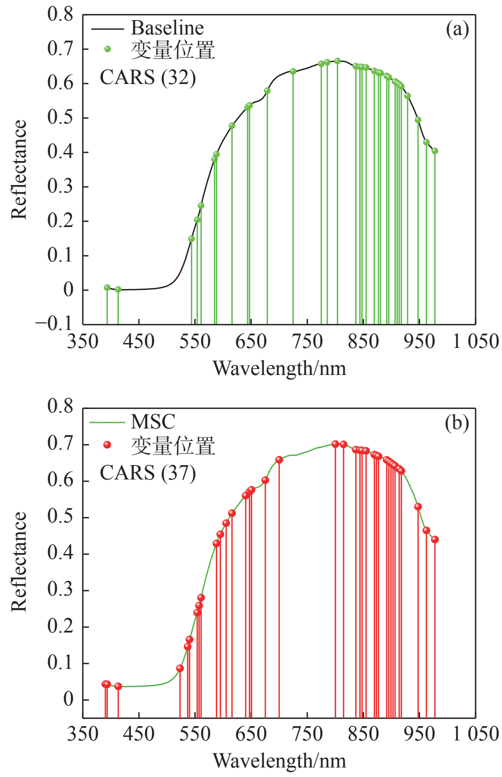


图 5　两种预处理方法基于 CARS 算法果茎部位特征波长位置图。(a)Baseline;(b)MSC

Fig. 5　Location map of the characteristic wavelengths in the fruit stem part based on the CARS algorithm corresponding to the pretreatments (a) Baseline and (b) MSC

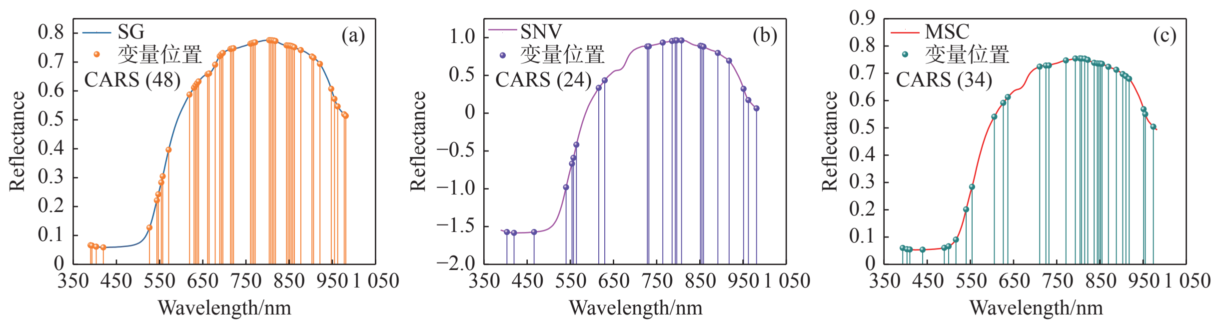


图 6　各部位基于 CARS 算法特征波长位置图。(a)花萼;(b)赤道;(c)全局

Fig. 6　Location maps of the characteristic wavelengths based on CARS algorithm. (a) Calyx; (b) equator and (c) global

3.4.2　UVE 特征波长筛选

以果茎部位特征波长筛选过程为例,图 7(彩图见期刊电子版)为 UVE 筛选后的稳定性值图。图中蓝竖线的左侧为 169 个原始波长,右侧引入相同数量的随机变量。上下两条虚线分别代表最大和最小截止阈值。两截止阈值中间的变量需剔除,两线之外的变量则为特征变量。果茎部位的两个最优模型分别为 Baseline-PLSR 和 MSC-LSSVM 模型。利用 UVE 算法对果茎部位两个最优模型中的全变量进行筛选,筛选出的特征波

长位置和数量如图 8(彩图见期刊电子版)所示,由图 8 可知, UVE 算法从果茎部位两个最佳模型的全变量中分别筛选出了 41 个和 69 个特征变量, 分别占全波段的 24.3% 和 40.8%。

用同样的方法分别对蜜桔花萼部位、赤道部位和全局最佳模型中的全变量进行特征波长筛选。筛选后的特征波长位置和数量如图 9(彩图见期刊电子版)所示。由图 9 可知, 利用 UVE 算法进行特征波长筛选后, 分别筛出了 42、58 和 29 个特征变量, 分别占全波段的 24.9%、34.3% 和 17.2%, 大多数波长位于 700 ~ 950 nm 之间。

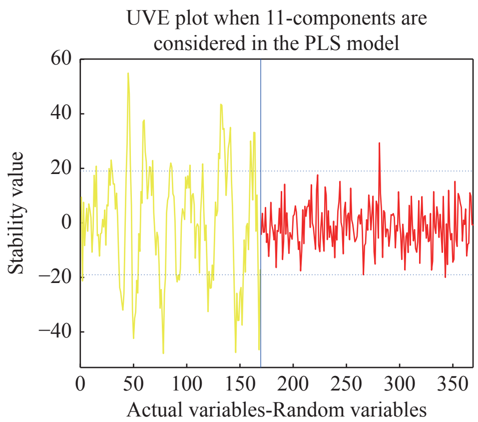


图 7　UVE 筛选后果茎部位的稳定性值图

Fig. 7　Stability values of the fruit stem part after UVE screening

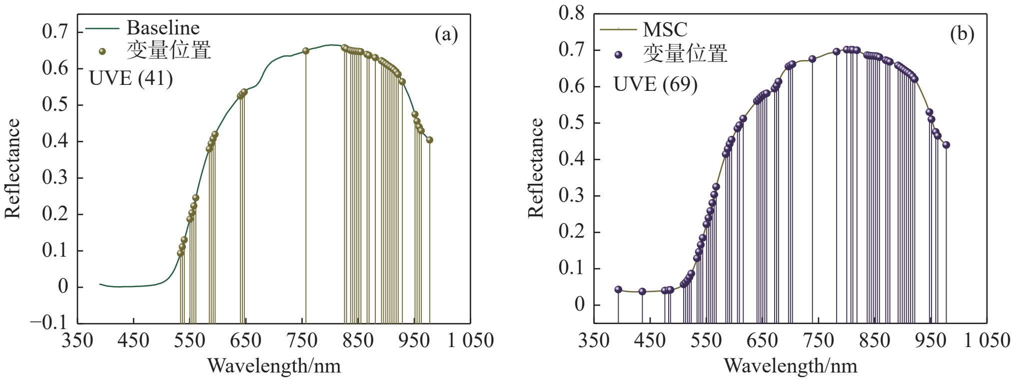


图 8　两种预处理方法下, 基于 UVE 算法果茎部位特征波长位置图。(a)Baseline; (b)MSC

Fig. 8　Location maps of characteristic wavelengths in the fruit stem part based on the UVE algorithm corresponding to (a) Baseline and (b) MSC

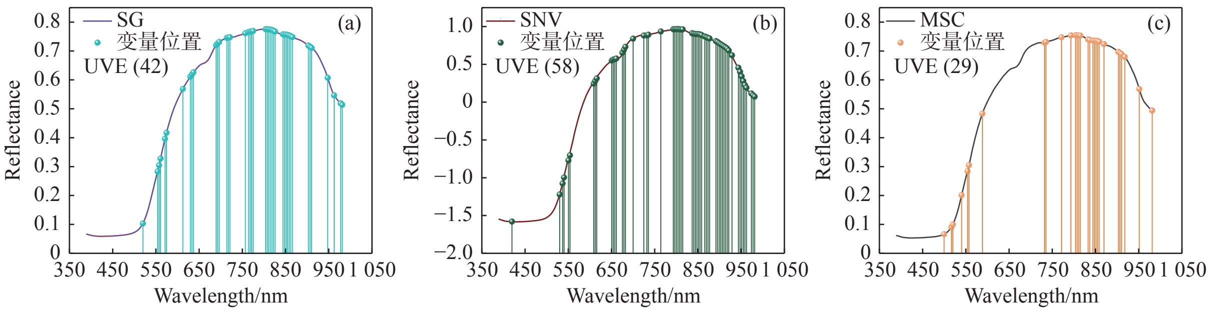


图 9　基于 UVE 算法特征波长位置图。(a)花萼; (b)赤道; (c)全局

Fig. 9　Location maps of the characteristic wavelengths of the UVE-based algorithm. (a) Calyx; (b) equator and (c) global

3.5　基于 CARS 和 UVE 筛选的不同部位特征波长的模型比较

根据 3.4.1 中的特征波长筛选结果, 分别建立蜜桔花萼、果茎、赤道和全局的 PLSR 和 LSSVM 糖度预测模型, 预测效果如表 4 所示。由表 4 可知, 进行特征波长筛选后, 大部分模型的预测效果都有所提升。对于蜜桔花萼部位模型, SG-CARS-PLSR 模型预测效果最佳, 其 *Rp* 为 0.918, RMSEP 为 0.400 °Brix; 对于蜜桔果茎部位模型, Baseline-CARS-PLRS 模型预测效果最好, 其 *Rp* 为 0.922, RMSEP 为 0.424 °Brix; 对于全局模型, 其最优模型为 MSC-CARS-LSSVM 模型, 其 *Rp* 为 0.955, RMSEP 为 0.395 °Brix。以上模型经过特征波长筛选后与之前模型相比, 预测精度都有

所提高。表明 CASR 算法剔除了原始光谱中的干扰信息，基于筛选后的特征波长建立的模型预测效果更好。然而，对于蜜桔赤道部位模型在经过特征波长筛选后，其最佳模型为 SNV-CARS-PLSR 模型，其 Rp 为 0.914，RMSEP 为 0.400 °Brix，与筛选前的 SNV-PLSR 模型比较，其预测精度有所下降，预测集相关系数从 0.936 下降到 0.914。说明 CARS 算法将蜜桔赤道部位光谱数据中跟糖度有关的信息剔除了。

表 4 基于 CARS 特征波长筛选后蜜桔不同部位的 PLSR 和 LSSVM 模型比较

Tab. 4 Comparison of PLSR and LSSVM models for different parts of honey oranges after CARS characteristic wavelengths screening

预测模型	不同部位	建模集		预测集	
		R_C	RMSEC/°Brix	R_P	RMSEP/°Brix
PLSR	花萼	0.926	0.447	0.918	0.400
	果茎	0.928	0.507	0.922	0.424
	赤道	0.933	0.452	0.914	0.400
	全局	**0.948**	**0.394**	**0.942**	**0.399**
LSSVM	花萼	0.927	0.445	0.914	0.408
	果茎	0.951	0.412	0.904	0.546
	赤道	0.960	0.352	0.901	0.423
	全局	**0.975**	**0.274**	**0.955**	**0.395**

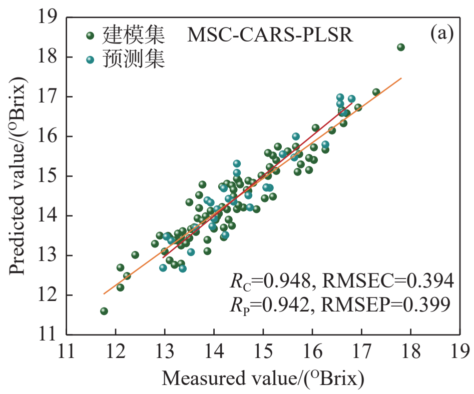
根据 3.4.2 中的特征波长筛选结果，分别建立蜜桔花萼、果茎、赤道和全局的 PLSR 和 LSSVM 糖度预测模型，预测结果如表 5 所示。从表 5 可以看出，对于花萼部位的两个最优模型来说，筛选后建立的模型精度都有一定程度的下降。这表明 UVE 算法去除了与糖度有关的变量。对于果茎部位模型，筛选后的最佳预测模型为 MSC-UVE-LSSVM 模型，其 Rp 为 0.896，RMSEP 为 0.575 °Brix，与其筛选前的 MSC-LSSVM 模型相比，预测精度有所上升，但是其 PLSR 模型筛选后预测精度有所下降。相比于前面两个部位的模型，赤道部位和全局的最佳预测模型筛选前后的预测效果相差无几。表明 UVE 算法没能有效的去除赤道和全局模型中的无效变量。由此可知，UVE 算法并不适合蜜桔糖度预测模型的优化。

表 5 基于 UVE 特征波长筛选后蜜桔不同部位的 PLSR 和 LSSVM 模型比较

Tab. 5 Comparison of PLSR and LSSVM models for different parts of honey oranges after UVE characteristic wavelengths screening

预测模型	不同部位	建模集		预测集	
		R_C	RMSEC/°Brix	R_P	RMSEP/°Brix
PLSR	花萼	0.890	0.538	0.850	0.519
	果茎	0.885	0.633	0.812	0.655
	赤道	0.943	0.419	0.933	0.364
	全局	**0.949**	**0.393**	**0.937**	**0.434**
LSSVM	花萼	0.901	0.514	0.838	0.537
	果茎	0.950	0.416	0.896	0.575
	赤道	0.950	0.400	0.900	0.423
	全局	**0.956**	**0.368**	**0.943**	**0.414**

综合比较蜜桔花萼部位、果茎部位、赤道部位和全局的 PLSR 模型和 LSSVM 模型，对于 PLSR 模型，预测效果最好的是赤道 SNV-PLSR 模型和全局 MSC-CARS-PLSR 模型，两者对比可以发现全局模型预测效果更好。同样对于 LSSVM 模型，最优模型是花萼部位 SG-CARS-LSSVM 模型和全局 MSC-CARS-LSSVM 模型，对比之下依旧是全局模型预测效果更好。图 10（彩图见期刊电子版）是两个全局最优预测模型 MSC-CARS-PLSR 和 MSC-CARS-LSSVM 对涌泉蜜桔糖度含量的预测结果。从图 10 可知，MSC-CARS-LSSVM 模型中样本点比 MSC-CARS-PLSR 模型中样本点更加均匀和集中分布在拟回归线周围，表明全局 MSC-CARS-LSSVM 模型的预测效果更佳。综上所述，基于全局数据建立的非线性糖度预测模型更加稳定且预测更准确。



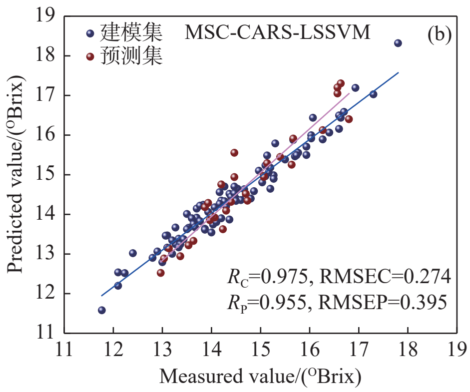


图 10　涌泉蜜桔糖度含量预测模型(a)MSC-CARS-PLSR 和(b)MSC-CARS-LSSVM 的散点图

Fig. 10　Scatter plots of the Yongquan honey oranges sugar content prediction models (a) MSC-CARS-PLSR and (b) MSC-CARS-LSSVM

4　结　论

为了得到涌泉蜜桔糖度最佳检测位置和最佳糖度预测模型,利用高光谱技术分别采集了涌泉蜜桔花萼、果茎、赤道 3 个部位的光谱数据,并分别测定对应部位的糖度,然后,分析比较了蜜桔 3 个局部和全局的糖度预测模型。采用 4 种预处理方式(SNV、MSC、Baseline、SG)和两种建模方法(PLSR、LSSVM)进行建模。对比发现蜜桔不同部位所建的预测模型之间有较明显的差异,花萼部位的最优模型预测性能明显低于果茎和赤道部位的最优模型预测性能,赤道部位 SNV-PLSR 模型和全局 MSC-LSSVM 模型对整个蜜桔糖度预测效果比较好,最佳模型的预测集相关系数 *Rp* 可达 0.946。上述结果表明不同的预处理方法和建模方法对其糖度预测效果有影响。为了进一步提高模型的预测性能同时消除冗余变量,利用 CARS 和 UVE 算法对蜜桔 3 个局部和全局的光谱数据进行特征波长筛选,并分别对筛选后的光谱数据建立 PLSR 和 LSSVM 模型。对用特征波长建立的模型进行对比分析后发现,最佳糖度预测模型为全局 MSC-CARS-LSSVM 模型,其 *Rp* 为 0.955, RMSEP 为 0.395 °Brix。其次是赤道部位的 SNV-PLSR 模型,其 *Rp* 为 0.936, RMSEP 为 0.37 °Brix。两个模型的预测精度均很高且接近,因此可以将赤道部位作为涌泉蜜桔糖度最佳检测位置。上述研究结果同时表明,对于大多数模型来说,经过 UVE 算法筛选后的模型预测精度不如用 CARS 算法筛选后的模型预测精度,但是对于涌泉蜜桔赤道部位的模型来说,采用 CASR 算法进行波长筛选后,模型预测效果会变差,采用 UVE 算法进行特征波长筛选后模型与筛选前模型预测效果相差无几。因此,在接下来的工作中将重点研究不同的波长筛选算法对蜜桔不同部位的糖度预测模型的影响。除此之外,本次研究中仅使用 3 个局部信息仍不能代表整个样本信息,所以下一步将通过获取整个样本区域更加全面的信息来评估本研究结果是否可靠,进一步优化模型。

参考文献:

[1]　王玲娇, 蒋守渭. 临海涌泉蜜桔网络营销策略研究[J]. 现代商业, 2015(35): 36-37.
WANG L J, JIANG SH W. Research on network marketing strategy of Linhai Yongquan honey tangerine[J]. *Modern Business*, 2015(35): 36-37. (in Chinese)

[2]　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.
JIE D F, YANG J, PENG Y X, *et al*.. Research on the detection model of sugar content in different position of citrus based on the hyperspectral technology[J]. *Food & Machinery*, 2017, 33(3): 51-54. (in Chinese)

[3]　王玥, 樊柳荫. 猪肉可视化新鲜度智能指示薄膜研究[J]. 分析化学, 2023, 51(1): 139-145.
WANG Y, FAN L Y. Intelligent indicator film for visual meat freshness monitoring[J]. *Chinese Journal of Analytical Chemistry*, 2023, 51(1): 139-145. (in Chinese)

[4]　田喜, 陈立平, 王庆艳, 等. 全透射近红外光谱的苹果整果糖度在线检测模型优化[J]. 光谱学与光谱分析, 2022, 42(6): 1907-1914.
TIAN X, CHEN L P, WANG Q Y, *et al*.. Optimization of online determination model for sugar in a whole apple using full transmittance spectrum[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(6): 1907-1914. (in Chinese)

[5]　陈玥瑶, 夏静静, 韦芸, 等. 近红外光谱法无损检测平谷产大桃品质方法研究[J]. 分析化学, 2023, 51(3): 454-462.

CHEN Y Y, XIA J J, WEI Y, *et al*.. Research on nondestructive quality test of Pinggu peach by near-infrared spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2023, 51(3): 454-462. (in Chinese)

[6] LU Y ZH, HUANG Y P, LU R F. Innovative hyperspectral imaging-based techniques for quality evaluation of fruits and vegetables: A review[J]. *Applied Sciences*, 2017, 7(2): 189.

[7] MESA A R, CHIANG J Y. Multi-input deep learning model with RGB and hyperspectral imaging for banana grading[J]. *Agriculture*, 2021, 11(8): 687.

[8] 徐婧, 郑红, 谢丽芳, 等. 鸡肉样品中痕量喹诺酮类抗生素的表面增强拉曼光谱检测研究[J]. 分析化学, 2023, 51(3): 397-404.

XU J, ZHENG H, XIE L F, *et al*.. Fast detection of trace enrofloxacin and ciprofloxacin in chicken meat by surface-enhanced Raman spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2023, 51(3): 397-404. (in Chinese)

[9] 沈彦龙, 程立业, 孟祥茹, 等. 人参连作土壤对不同生育期人参生长发育及抗氧化系统的影响[J]. 应用化学, 2023, 40(1): 109-115.

SHEN Y L, CHENG L Y, MENG X R, *et al*.. Effects of ginseng continuous soil crop on growth development and antioxidant system of ginseng at different fertility stages[J]. *Chinese Journal of Applied Chemistry*, 2023, 40(1): 109-115. (in Chinese)

[10] 黄蕊, 叶长青, 李亚军, 等. 线粒体靶向的近红外HClO/ClO$^-$荧光探针的研究进展[J]. 应用化学, 2022, 39(3): 407-424.

HUANG R, YE CH Q, LI Y J, *et al*.. Progress of mitochondria-targeted near-infrared HClO/ClO$^-$ fluorescent probes[J]. *Chinese Journal of Applied Chemistry*, 2022, 39(3): 407-424. (in Chinese)

[11] 王瑞, 孟祥茹, 李琼, 等. 人参属中药腐解化感作用研究进展[J]. 应用化学, 2023, 40(1): 1-8.

WANG R, MENG X R, LI Q, *et al*.. Research progress on the decomposed Allelopathy of *Panax* genus[J]. *Chinese Journal of Applied Chemistry*, 2023, 40(1): 1-8. (in Chinese)

[12] KIM D, BURKS T F, RITENOUR M A, *et al*.. Citrus black spot detection using hyperspectral imaging[J]. *International Journal of Agricultural and Biological Engineering*, 2014, 7(6): 20-27.

[13] 孟田源, 王转卫, 迟茜, 等. 基于高光谱成像技术生长发育后期苹果糖度的无损检测[J]. 西北农林科技大学学报(自然科学版), 2016, 44(6): 228-234.

MENG T Y, WANG ZH W, CHI X, *et al*.. Hyperspectral imaging based non-destructive prediction of soluble solids content in apples at late development period[J]. *Journal of Northwest A& F University* (*Natural Science Edition*), 2016, 44(6): 228-234. (in Chinese)

[14] 许丽佳, 陈铭, 王玉超, 等. 高光谱成像的猕猴桃糖度无损检测方法[J]. 光谱学与光谱分析, 2021, 41(7): 2188-2195.

XU L J, CHEN M, WANG Y CH, *et al*.. Study on non-destructive detection method of kiwifruit sugar content based on hyperspectral imaging technology[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(7): 2188-2195. (in Chinese)

[15] 杨杰, 马本学, 王运祥, 等. 葡萄可溶性固形物的高光谱无损检测技术[J]. 江苏农业科学, 2016, 44(6): 401-403.

YANG J, MA B X, WANG Y X, *et al*.. Hyperspectral nondestructive detection technique of soluble solids in grapes[J]. *Jiangsu Agricultural Science*, 2016, 44(6): 401-403. (in Chinese)

[16] SUN X D, XU C, LUO CH G, *et al*.. Non-destructive detection of tea stalk and insect foreign bodies based on THz-TDS combination of electromagnetic vibration feeder[J]. *Food Quality and Safety*, 2023, 7: fyad004.

[17] LIU Y D, SUN X D, OUYANG A G. Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN[J]. *LWT - Food Science and Technology*, 2010, 43(4): 602-607.

[18] YANG Y, ZHAO CH J, HUANG W Q, *et al*.. Optimization and compensation of models on tomato soluble solids content assessment with online Vis/NIRS diffuse transmission system[J]. *Infrared Physics & Technology*, 2022, 121: 104050.

[19] 袁琳, 徐怀德, 李钰金. 近红外漫反射光谱检测网纹瓜可溶性固形物含量的研究[J]. 中国食品学报, 2010, 10(4): 272-277.

YUAN L, XU H D, LI Y J. Studies on the rapid measurements of soluble solids content in nutmeg melon by near infrared diffuse reflectance spectroscopy[J]. *Journal of Chinese Institute of Food Science and Technology*, 2010, 10(4): 272-277. (in Chinese)

[20] 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.

[21] 孙博康, 刘贵珊. 基于高光谱技术检测香水梨硬度的研究[J]. 食品安全导刊, 2021(19): 118-120.
SUN B K, LIU G SH. Study on the detection of hardness of perfumed pears based on hyperspectral technology[J]. *Journal of Food Safety*, 2021(19): 118-120. (in Chinese)

[22] ZHANG F, LI B, YIN H, *et al*.. Study on the quantitative assessment of impact damage of yellow peaches using the combined hyperspectral technology and mechanical parameters[J]. *Journal of Spectroscopy*, 2022, 2022: 7526826.

[23] WANG ZH L, CHEN J, FAN Y F, *et al*.. Evaluating photosynthetic pigment contents of maize using UVE-PLS based on continuous wavelet transform[J]. *Computers and Electronics in Agriculture*, 2020, 169: 105160.

[24] 廉小亲, 汤燊森, 吴静珠, 等. 基于近红外的兰州百合品质定量建模方法研究[J]. 食品科技, 2020, 45(7): 298-302.
LIAN X Q, TANG S M, WU J ZH, *et al*.. Research on quantitative model of Lilium Lanzhou quality by near infrared spectroscopy[J]. *Food Science and Technology*, 2020, 45(7): 298-302. (in Chinese)

[25] YANG X Y, LIU G SH, HE J G, *et al*.. Determination of sugar content in Lingwu jujube by NIR–hyperspectral imaging[J]. *Journal of Food Science*, 2021, 86(4): 1201-1214.

[26] 李斌, 邹吉平, 张烽, 等. 基于高光谱成像技术和力学参数对贡梨冲击损伤的定量研究[J]. 中国农业大学学报, 2023, 28(2): 186-197.
LI B, ZOU J P, ZHANG F, *et al*.. Quantitative study on impact damage of Gongli based on hyperspectral imaging technology and mechanical parameters[J]. *Journal of China Agricultural University*, 2023, 28(2): 186-197. (in Chinese)

作者简介:

李　斌(1989—), 男, 江西鹰潭人, 博士, 副教授, 华东交通大学机电学院教师, 2012 年于武汉大学获得学士学位, 2017 年于中国科学院光电技术研究所获得博士学位, 主要研究方向: 无损智能检测。E-mail: libingioe@126.com

刘燕德(1967—), 女, 江西吉安人, 博士, 二级教授, 智能机电装备创新研究院院长, 2001 年于江西农业大学获得硕士学位, 2006 年于浙江大学获得博士学位, 主要研究方向: 光电无损检测技术、光机电一体化技术与装备。E-mail: liuyande@ecjtu.jx.cn