

# 红提糖度和硬度的高光谱成像无损检测

高升<sup>1</sup>, 王巧华<sup>1,2\*</sup>, 付丹丹<sup>1</sup>, 李庆旭<sup>1</sup>

<sup>1</sup>华中农业大学工学院, 湖北 武汉 430070;

<sup>2</sup>农业部长江中下游农业装备重点实验室, 湖北 武汉 430070

**摘要** 红提的糖度和硬度是评价红提品质的重要指标, 探究了基于高光谱成像技术的红提糖度和硬度的无损检测方法及其最佳预测模型。在红提果粒的三种放置模式(横放、果柄侧朝下、果柄侧朝上)下, 分别采集 213 个样本在 400~1000 nm 波长范围内的高光谱图像, 对比分析光谱采集的最优模式; 然后在最优采集模式下对光谱进行预处理; 应用遗传算法(GA)、连续投影算法(SPA)、竞争性自适应重加权(CARS)算法和无信息变量消除法(UVE)针对原始光谱提取特征波长; 结合化学计量学方法分别建立基于全光谱和特征波长的偏最小二乘回归(PLSR)、最小二乘支持向量机(LSSVM)和随机森林(RF)的红提糖度、硬度的无损预测模型。结果表明: 基于 RF 建立的糖度和硬度模型的效果较优; 预测糖度的最优模型为遗传算法优化的随机森林(GA-RF), 其校正集相关系数( $R_c$ )、预测集相关系数( $R_p$ )分别为 0.969、0.928, 校正集均方根误差(RMSEC)、预测集均方根误差(RMSEP)分别为 0.266、0.254; 预测硬度的最优模型为基于移动窗口平滑结合连续投影算法优化的随机森林(MA-SPA-RF), 其  $R_c$ 、 $R_p$  分别为 0.961、0.932, RMSEC、RMSEP 分别 2.119、1.634。研究结果表明基于高光谱成像技术预测红提的糖度和硬度是可行的。

**关键词** 光谱学; 红提; 糖度; 硬度; 高光谱成像; 无损检测

中图分类号 TS255.7

文献标识码 A

doi: 10.3788/AOS201939.1030004

## Nondestructive Detection of Sugar Content and Firmness of Red Globe Grape by Hyperspectral Imaging

Gao Sheng<sup>1</sup>, Wang Qiaohua<sup>1,2\*</sup>, Fu Dandan<sup>1</sup>, Li Qingxu<sup>1</sup>

<sup>1</sup> College of Engineering, Huazhong Agricultural University, Wuhan, Hubei 430070, China;

<sup>2</sup> Key Laboratory of Agricultural Equipment in Mid-lower Yangtze River, Ministry of Agriculture and Rural Affairs, Wuhan, Hubei 430070, China

**Abstract** The sugar content and firmness of red globe grapes are important indicators for evaluating their quality. This study explores nondestructive detection methods and best prediction models for determining the sugar content and firmness of red globe grapes based on hyperspectral imaging technology. The hyperspectral images of 213 samples, in the wavelength range of 400-1000 nm, are collected in three placement orientations (horizontal, fruit stalk-side down, and fruit stalk-side up). The optimal orientation for spectral imaging is compared and analyzed, and subsequently the spectrum is preprocessed in the optimal orientation. Several preprocessing methods, i.e., genetic algorithm (GA), successive projections algorithm (SPA), competitive adaptive reweighted sampling (CARS) algorithm, and uninformative variable elimination algorithm (UVE), are applied to the images to extract characteristic wavelengths from the original spectra. Using chemometrics methods, combined with either partial least squares regression (PLSR), least squares support vector machine (LSSVM), and random forest (RF) analysis based on full spectra and characteristic wavelengths, several protocols are established to mathematically predict the sugar content and firmness of red globe grapes from the images. Results show that the sugar and firmness model based on RF performs the best. The optimal model for predicting sugar content proves to be RF optimized by GA (GA-RF), with corrected-set correlation coefficient ( $R_c$ ) and predicted-set correlation coefficient ( $R_p$ ) values of 0.969 and 0.928, respectively, and corrected-set root-mean-square error (RMSEC) and predicted-set root-mean-square error (RMSEP) values of 0.266 and 0.254, respectively. The optimal model for predicting firmness proves

收稿日期: 2019-03-26; 修回日期: 2019-06-25; 录用日期: 2019-07-08

基金项目: 国家自然科学基金(31871863)、湖北省自然科学基金(2012FKB02910)、湖北省研究与开发计划项目(2011BHB016)

\* E-mail: wqh@mail.hzau.edu.cn

to be RF optimized by moving-average method and SPA (MA-SPA-RF), with  $R_c$  and  $R_p$  values of 0.961 and 0.932, respectively, and RMSEC and RMSEP values of 2.119 and 1.634, respectively. These results prove the sugar content and firmness of red globe grapes can be nondestructively predicted via hyperspectral imaging.

**Key words** spectroscopy; red globe grape; sugar content; firmness; hyperspectral imaging; nondestructive detection

**OCIS codes** 300.6170; 300.1030; 300.6190

## 1 引 言

红提被誉为世界四大水果之首,因果肉质地坚实、香甜可口、富含多种维生素而受到人们的喜爱<sup>[1]</sup>。糖作为红提果实中最主要的营养物质,决定着果实的风味,是果实成熟度的衡量标准和重要指标。采后红提的果实质地会不断发生变化,内部组织变软,风味变差,特别是在长距离运输过程中,挤压易造成果粒发生损伤,严重影响红提的品质和价格<sup>[2]</sup>。而且,损伤的果粒更容易腐烂。水果的硬度是判断果蔬成熟度和贮运品质的一个重要指标,决定了水果的耐贮性和成熟度<sup>[3]</sup>,因此检测红提的糖度和硬度具有重要意义。

红提糖度和硬度的常规检测方法是破坏性抽样检测,不仅繁琐费时,而且检测后的样品因组织遭到严重破坏而不能销售和食用,存在诸多弊端,因此迫切需要开发一种无损检测技术。高光谱成像技术被广泛应用于果实糖度和成熟度的无损检测研究中<sup>[4-9]</sup>,目前,国内外研究人员大都是针对哈密瓜<sup>[10]</sup>、苹果<sup>[11]</sup>、蓝莓<sup>[12]</sup>等进行硬度检测。Li等<sup>[13]</sup>利用高光谱技术测定了不同成熟期樱桃的可溶性固形物含量(SSC)与pH,确定将遗传算法优化多元线性回归模型(GA-MLR)作为最终的建模方法,预测集中可溶性固形物的标准差与均方根误差的比值为2.7,pH的标准差与均方根误差的比值为2.4,证明了利用近红外高光谱成像技术检测樱桃果实品质的可行性。Mo等<sup>[14]</sup>使用可见/近红外高光谱成像技术建立了苹果内部可溶性固形物含量的偏最小二乘回归(PLSR)模型,结果表明,高光谱成像技术可用于预测苹果内部可溶性固形物的含量。目前,采用高光谱技术研究红提糖度和硬度的综合研究还未见报道。

本文基于高光谱成像系统探讨红提在不同放置模式下的光谱数据对模型的影响规律,分别提取红提果粒在三种放置模式下(横放、果柄侧朝下、果柄侧朝上)的光谱数据,并将三种放置模式下的光谱数据进行平均,得到了整个果粒的平均光谱,然后根据相应光谱数据与化学计量方法得到的数据建立糖度

和硬度的预测模型,并进行验证,为红提糖度和硬度的无损检测提供了可靠的检测模型及方法。

## 2 材料与方 法

### 2.1 材 料

实验材料为新鲜红提,在每穗红提的穗外部、穗中部、穗顶部、穗尖分别挑选大小相近、颜色差异较大、完好无损的红提果粒作为实验样本,建模样本总数为213粒。将样本编号后放入恒温恒湿箱中保存12 h<sup>[15]</sup>,温度设置为(22±1)℃,相对湿度设置为65%。

### 2.2 仪器与设备

实验仪器包括TMS-PRO型高精度专业食品物性分析仪(质构仪)、恒温恒湿箱(上海新苗医疗器械制造有限公司)、申光WAY(2WAJ)阿贝尔折射仪、Zolix Hyper SIS-VNIR-CL高光谱成像系统。高光谱成像系统主要由高光谱成像光谱仪(芬兰Spectral Imaging Ltd.公司)、CCD相机(日本Hamamatsu公司)、4个50 W的卤素灯、1台丝杆式位移控制平台等主要部件组成。该系统采集的光谱波长范围为391~1043 nm(含有520个波长),分辨率为2.8 nm,整个采集系统置于暗箱内。

### 2.3 方 法

#### 2.3.1 高光谱图像信息的采集

高光谱成像系统预热30 min后进行实验。由于暗电流及CCD相机芯片不稳定的影响,图像会产生一定的噪声,因此在样本采集前需要对高光谱图像进行黑白校正。将标准白板(聚四氟乙烯长方形白板)放在采集平台上,获得白板数据 $I_w$ ;盖上相机盖,获得全黑标定图像 $I_D$ ;将红提放在采集平台上获得原始高光谱漫反射图像 $I_R$ ;根据公式 $R = (I_R - I_D) / (I_w - I_D) \times 100\%$ 得到校正后红提的图像 $R$ 。

实验时,在平台上放置自制的带孔载物平板,并分别采集不同样本放置模式下的高光谱图像。设置高光谱成像系统的参数如下:相机曝光时间为0.15 s,平台移动速度为1.7 mm/s,平台移动范围为0~245 mm,样本平台与镜头的距离为420 mm。三种放置模式下的高光谱图像如图1所示。

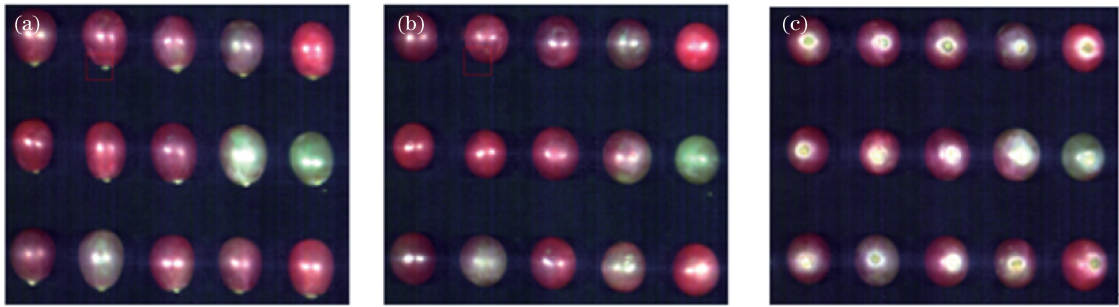


图 1 三种放置模式下红提果粒的高光谱图像。(a)横放;(b)果柄侧朝下;(c)果柄侧朝上

Fig. 1 Hyperspectral images in three placement orientations. (a) Horizontal;

(b) fruit stalk-side down; (c) fruit stalk-side up

### 2.3.2 红提糖度和硬度的测定方法

硬度的测定:采集完样本的高光谱图像后,利用 TMS-PRO 型质构仪先进行质地剖面 TPA (Texture Profile Analysis) 试验,采用 P100/R 探头,测前速率为 2.0 mm/s,测试速率为 1.0 mm/s,测后速率为 2 mm/s,测试时间间隔为 1 s,压缩百分比为 30%<sup>[16]</sup>。横向压缩,压缩部位为红提的中部位置。

糖度测定:完成硬度测定后立即参照 NY/T 2637—2014《水果、蔬菜制品可溶性固形物含量的测定——折射仪法》进行糖度测定。

### 2.3.3 感兴趣区域的提取及平均光谱数据的处理

高光谱图像中包含无用的背景图像,因此,对上述三种放置模式下的高光谱图像进行分割,分别将分割后的单粒图像作为感兴趣区域(RIOs)进行提取。在 MATLAB R2014a 软件上对高光谱图像进行处理,选择图像中整粒红提的图像作为感兴趣区域,提取感兴趣区域的平均光谱作为原始光谱。分析时去除两端噪声较大的波长,本文选取 450~1000 nm(含有 438 个波长)范围的波长进行建模。在进行感兴趣区域的提取时,由图 2 中背景与红提果粒区域的反射率可知,在 600~900 nm 范围内,

背景的反射率较低而红提果粒的反射率较高,在 726.6 nm 时两者反射率的差值最大。因此选取 726.6 nm 处的灰度图像进行果粒区域的提取,结果如图 3(a)所示。首先采用 Otsu 阈值分割方法获得二值图像,然后利用中值滤波和腐蚀运算得到背景区域为 0、果粒区域为 1 的二值图像,并将得到的二值图像作为掩模板,如图 3(b)所示,最后依次将每个果粒分割出来,并提取高光谱图像中单个样本红提果粒区域,掩模后红提区域的图像如图 3(c)所示。

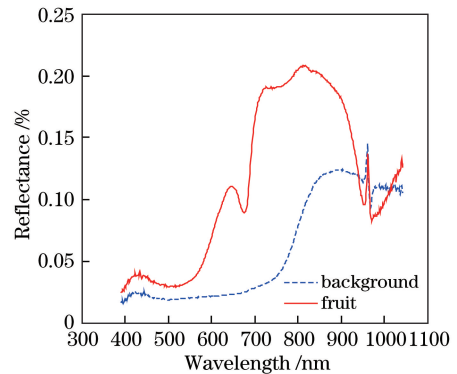


图 2 高光谱图像中背景与红提区域的反射率

Fig. 2 Reflectivity of background and red globe

grape area in hyperspectral images

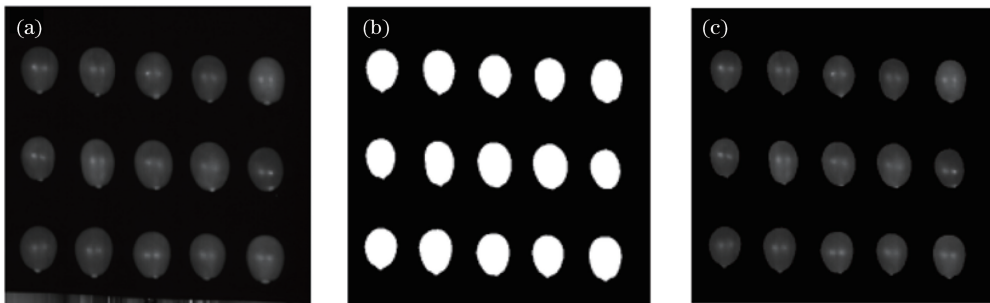


图 3 红提果粒的高光谱图像处理。(a) 726.6 nm 处高光谱的图像;(b)掩模板图像;(c)掩模后红提区域的图像

Fig. 3 Hyperspectral image processing of red globe grapes. (a) Hyperspectral image at 726.6 nm;

(b) mask template image; (c) masked image of red globe grape area

遗传算法(GA)<sup>[17]</sup>模仿自然界生物的进化机制,使用选择、交叉和变异进行编码,并通过不断的迭代来优胜劣汰,逼近最优解。

连续投影算法(SPA)<sup>[18]</sup>可以挑选出具有最少冗余信息的变量组,有效解决信息重叠和不共线的问题。

竞争性自适应重加权(CARS)<sup>[19]</sup>算法是一种基于蒙特卡罗采样法对模型进行变量选择的方法,利用PLS(Partial Least Squares)建模并去掉权重较小的变量,将RMSECV(Root Mean Square Error of Cross)最小的变量集作为最终选定的最优变量子集。

无信息变量消除法(UVE)<sup>[20]</sup>的原理为在光谱中加入一个随机矩阵,建立偏最小二乘回归交互验证模型,将回归系数均值与标准差的商作为衡量指标,挑选出需要的光谱变量。

### 2.3.4 模型的建立及评价方法

本文分别建立基于原始光谱和上述3种特征波长提取方法后的特征波长建立红提糖度和硬度的PLSR模型、LS-SVM模型和RF模型。

偏最小二乘回归(PLSR)是一种经典的结合多元线性回归、相关分析和主成分优点的算法,通过最小化偏差平方和对曲线进行线性拟合。

最小二乘支持向量机(LSSVM)通过求解一次线性方程组来代替传统支持向量机中复杂的二次优化问题,在简化模型的同时提高了模型的运算速度。

随机森林(RF)算法是基于统计学习理论建立的,具有较高的预测准确率,建模过程较简单且有效解决了模型的过拟合问题。

模型的准确性和稳定性采用校正集的相关系数( $R_c$ )、均方根误差(RMSEC),以及预测集的相关系

数( $R_p$ )、均方根误差(RMSEP)进行评价<sup>[21]</sup>。相关系数越接近1,均方根误差越接近0,模型的预测性能及稳定性越好。

## 3 结果与分析

### 3.1 光谱数据的采集及预处理

从图4中可以看出,所有红提样本的光谱都呈现出相同的变化趋势:在450~550 nm区间内,曲线平滑,吸光度变化较小;550 nm之后,反射强度快速升高,在725~820 nm区间,光强度较高但变化较小。

建模前先对光谱进行预处理,以有效消除仪器噪声、暗电流等因素的影响<sup>[22]</sup>,因此本研究采用标准正态变量变换(SNV)、SavitZky-Golay 卷积平滑处理法(S\_G)、多元散射校正(MSC)、移动窗口平滑(MA)、中心化(MC)等方法进行预处理。

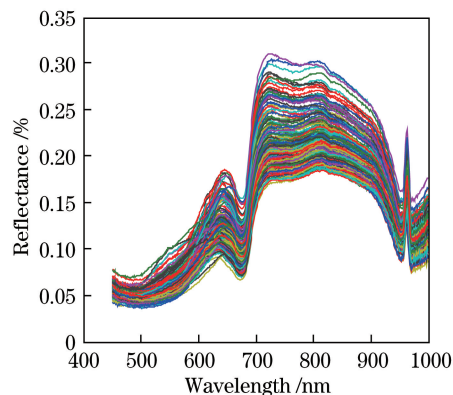


图4 红提样本的原始光谱

Fig. 4 Original spectra of red globe grape samples

表1中的LVs为提取的主因子数目。由表1可知:对红提糖度的光谱数据进行预处理后所建模型的相关系数和均方根误差都有所降低,模型变得

表1 采用不同预处理方法得到的全波长PLSR预测模型

Table 1 Full-band PLSR prediction model using different preprocessing methods

Index	Pretreatment	LVs	Calibration set		Prediction set	
			$R_c$	RMSEC	$R_p$	RMSEP
Sugar content	Raw	19	0.827	0.564	0.726	0.474
	SNV	14	0.808	0.597	0.712	0.493
	S_G	2	0.602	0.769	0.483	0.619
	MSC	8	0.811	0.595	0.665	0.515
	MA	8	0.479	0.782	0.373	0.881
	MC	18	0.825	0.617	0.701	0.503
Firmness	RAW	8	0.696	4.743	0.675	4.575
	SNV	6	0.605	5.015	0.684	4.365
	MSC	9	0.685	4.613	0.569	4.789
	MA	16	0.730	4.226	0.808	3.821
	MC	10	0.646	4.652	0.691	4.830

更加不稳定,所以利用原始光谱直接建模;采用 MA 对红提硬度的光谱数据进行预处理后所建模型的相关系数和均方根误差都有明显改善,因此,利用 MA 预处理后再建模。

### 3.2 样本集的划分

光谱理化值共生距离(SPXY)法<sup>[23]</sup>的优点是能够有效地覆盖多维向量空间,且同时考虑样本的光谱数据与所测指标,使划分的样本集更合理,已被广泛应用于光谱定量模型的建立。本文共采集了

213 份红提样本。

从表 2 中可以看出:糖度分布范围为 13.875~18.625 °Brix,校正集和预测集的标准差值分别为 0.971、0.653,个体样本数据相对集中;硬度的分布范围为 1.200~27.000 N,校正集和预测集的标准差值分别为 6.213、6.246,个体样本数据相对离散。通过 SPXY 法划分的糖度和硬度校正集的分布范围比预测集的分布范围更广,说明挑选出来的校正集样本具有代表性。

表 2 利用 SPXY 算法划分样本集的数据统计

Table 2 Data statistics of partitioning sample sets by SPXY algorithm

Number of samples	Index	Minimum	Maximum	Mean	Standard deviation
Calibration set (126 samples)	Sugar content / (° Brix)	13.875	18.625	16.109	0.971
	Firmness / N	1.200	27.000	13.711	6.213
Prediction set (42 samples)	Sugar content / (° Brix)	15.000	17.5000	15.858	0.653
	Firmness / N	2.700	23.500	12.774	6.264

### 3.3 放置模式对 PLSR 模型的影响

根据横放、果柄侧朝下、果柄侧朝上、整个果粒平均得到的高光谱数据建立偏最小二乘回归模型<sup>[24]</sup>,以确定最优的光谱采集方式。

由表 3 可知:基于果柄侧朝下模式建立的糖度模型的效果比果粒侧朝上、横放模式的效果更好,这个可能与果粒的成熟机理有关,该结论与李瑞

等<sup>[12]</sup>研究蓝莓的结果一致;基于果柄侧朝上放置模式建立的硬度模型效果比果粒侧朝上、横放模式的效果更好;利用整个果粒的平均光谱建立的糖度和硬度模型的效果最好,糖度和硬度的预测集相关系数  $R_p$  分别为 0.726、0.808,预测集的均方根误差分别为 0.474、3.821,模型的稳定性和精度都显著提高。

表 3 不同放置位置下得到的全波长 PLSR 预测模型

Table 3 Full-band PLSR prediction model with different placement orientations

Placement position	Index	LVs	Calibration set		Prediction set	
			$R_c$	RMSEC	$R_p$	RMSEP
Fruit stalk-side down	Sugar content	17	0.807	0.628	0.712	0.488
	Firmness	3	0.546	5.290	0.534	5.017
Fruit stalk-side up	Sugar content	13	0.792	0.659	0.705	0.492
	Firmness	6	0.646	4.854	0.606	4.780
Horizontal	Sugar content	19	0.805	0.631	0.690	0.497
	Firmness	6	0.558	5.245	0.602	4.785
Whole fruit	Sugar content	19	0.827	0.564	0.726	0.474
	Firmness	19	0.730	4.226	0.808	3.821

### 3.4 特征波长的提取

根据上文分析的结论,采用整个果粒的平均光谱作为原始光谱进行研究。因原始光谱得到的相关系数较低,均方根误差较大,模型精度和稳定性较低,故研究在不同特征波长提取方法下建立的不同的 PLSR 模型,以得到最好的红提糖度和硬度模型<sup>[25]</sup>。下文以糖度为例进行特征波长的提取。

#### 3.4.1 GA 提取特征波长

以采用 GA 提取特征波长为例,在 GA 运算过程中,设定初始群体为 30,交叉率为 50%,变异率为 1%,迭代次数为 100,以最小的 RMSECV 值为标准,筛选出在迭代过程中出现频次较多的波长点为最优波长点;经过 20 次随机搜索后,最终选定特征波长点为 26 个,如图 5(b)所示,占原始光谱信息的

5.94%, GA 所选特征波长如图 5(a)所示。

### 3.4.2 SPA 提取特征波长

以采用 SPA 提取特征波长为例,设定波长选择变量数范围为 5~30,步长为 1,根据均方根误差

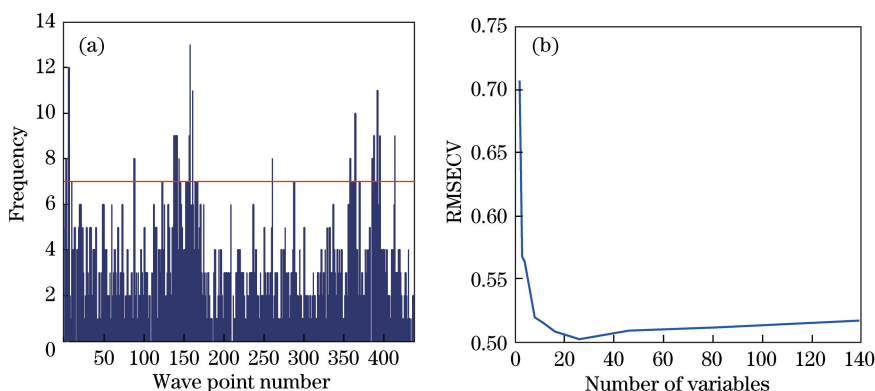


图 5 红提糖度的 GA 特征波长选取图。(a)GA 筛选图;(b)RMSECV 变化图

Fig. 5 GA characteristic wavelength extraction of sugar content of red globe grape. (a) GA-screened image; (b) change of RMSECV

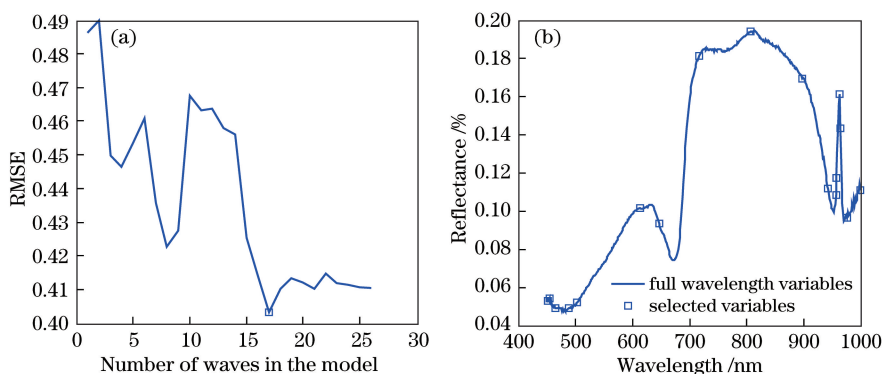


图 6 红提糖度的 SPA 特征波长选取图。(a)RMSE 变化图;(b)SPA 选取的变量

Fig. 6 SPA characteristic wavelength extraction of sugar content of red globe grape. (a) Change of RMSE; (b) selected variables of SPA

### 3.4.3 CARS 提取特征波长

以采用 CARS 提取特征波长为例,本研究设定蒙特卡罗采样为 50 次,采用 5 折交叉验证法。由图 7(b)可知,取采样 50 次建立的 PLSR 模型所对应的最小 RMSECV 作为最优结果,由图 7 可知,当 RMSECV 值达到最小值时,各变量的回归系数位于图 7(c)中竖直线位置,采样运行 27 次。

### 3.4.4 UVE 提取特征波长

以采用 UVE 提取特征波长为例,采用 UVE 提取光谱数据中的有效信息,设定噪声矩阵处最大稳定性绝对值的 99%作为剔除阈值,图中左侧曲线代表光谱变量的稳定性值,右侧曲线代表噪声变量的稳定性值,两水平虚线为变量的选择阈值( $\pm 28.74$ ),虚线内部为无用信息,外部为有用信息,选取结果如图 8 所示,对应的波长为选择的特征波长。采用

(RMSE)的变化确定选择特征变量的个数。如图 6(a)所示,当变量个数为 17 时,RMSE 最小为 0.40,占原始光谱信息的 3.88%,在原始光谱中所选特征波长的位置如图 6(b)所示。

UVE 共选择了 47 个特征波长。

## 3.5 模型的建立及结果

### 3.5.1 模型建立

将各算法提取的特征波长作为模型的输入量,以通过实验方法获得的红提糖度和硬度作为模型的结果,建立 PLSR、LSSVM 和 RF 模型,模型的结果如表 4 所示。

由表 4 可知:与原始光谱相比,采用 GA、UVE、SPA、CARS 方法提取出的特征波长的数量明显减少,特征波长提取后所建模型的相关系数与原始光谱所建模型相比明显增大,均方根误差明显减小,上述 4 种特征波长提取方法都有效提高了预测模型的准确性和可靠性;与其他特征波长提取方法相比,采用 UVE 提取的特征波长数量明显多于其他方法,而模型的稳定性却没有明显提高,UVE 提取特征波

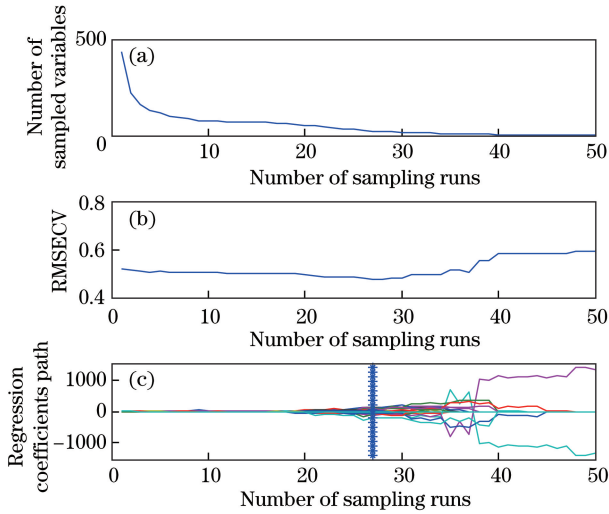


图 7 红提糖度的 CARS 特征波长选取图。(a)采样变量数；(b) RMSECV；(c)回归系数路径

Fig. 7 CARS characteristic wavelength extraction of sugar content of red globe grape. (a) Number of sampled variables; (b) RMSECV; (c) paths of regression coefficients

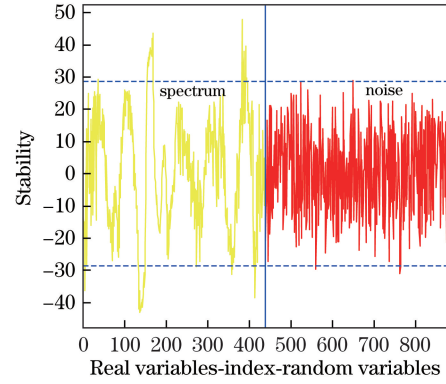


图 8 红提糖度的 UVE 特征波长选取图

Fig. 8 UVE characteristic wavelength extraction of sugar content of red globe grape

长可能夹杂着更多无用的信息,导致模型的精度不高;红提糖度的最佳建模方法为 GA-RF,提取的特征波长数量为 26 个,占原始光谱数据的 5.94%;红提硬度的最佳建模方法为 MA-SPA-RF,提取的特征波长数量为 24 个,占原始光谱数据的 5.48%,模型运行时间缩短。具体所提特征波长如表 5 所示。

表 4 基于红提特征波长建立的糖度和硬度预测模型的结果

Table 4 Results of prediction model for sugar content and firmness based on characteristic wavelengths of red globe grape

Index	Modeling method	Extraction method	No. of wavelength	Calibration set		Prediction set	
				$R_c$	RMSEC	$R_p$	RMSEP
Sugar content	PLSR	Raw	438	0.827	0.564	0.726	0.474
		GA	26	0.875	0.469	0.728	0.443
		SPA	17	0.862	0.492	0.745	0.429
		CARS	24	0.879	0.461	0.753	0.422
		UVE	47	0.863	0.490	0.729	0.444
	LSSVM	Raw	438	0.825	0.568	0.486	0.675
		GA	26	0.870	0.479	0.759	0.415
		SPA	17	0.864	0.489	0.752	0.426
		CARS	24	0.866	0.486	0.810	0.376
		UVE	47	0.875	0.470	0.749	0.426
	RF	Raw	438	0.954	0.260	0.873	0.402
		GA	26	0.969	0.266	0.928	0.254
		SPA	17	0.962	0.268	0.895	0.411
		CARS	24	0.946	0.296	0.890	0.406
		UVE	47	0.961	0.267	0.917	0.297
Firmness	PLSR	MA-Raw	438	0.730	4.226	0.808	3.821
		MA-GA	60	0.802	3.696	0.898	3.273
		MA-SPA	24	0.802	3.699	0.903	2.888
		MA-CARS	22	0.731	4.224	0.886	3.215
		MA-UVE	139	0.804	3.687	0.887	3.114
	LSSVM	MA-Raw	438	0.738	4.224	0.754	4.021
		MA-GA	60	0.795	3.788	0.901	3.023
		MA-SPA	24	0.741	4.183	0.870	3.578
		MA-CARS	22	0.746	4.163	0.893	3.288
		MA-UVE	139	0.833	3.444	0.921	2.674
	RF	MA-Raw	438	0.960	2.195	0.905	3.049
		MA-GA	60	0.950	2.132	0.918	2.031
		MA-SPA	24	0.961	2.119	0.932	1.634
		MA-CARS	22	0.948	2.199	0.911	2.053
		MA-UVE	139	0.959	2.120	0.921	1.893

表 5 糖度和硬度最优模型的特征波长列表

Table 5 Characteristic wave points of optimal model for sugar content and firmness

Index	Modeling method	Selected variables (wavelength) /nm
Sugar content (26 points)	GA-RF	452.76, 456.53, 461.55, 600.98, 626.10, 627.36, 628.62, 631.13, 633.64, 639.92, 644.95, 646.20, 647.46, 648.71, 651.23, 655.00, 859.75, 894.92, 918.78, 922.55, 927.58, 936.37, 941.40, 943.91, 945.16, 969.03
Firmness (24 points)	MA-SPA-RF	450.24, 451.50, 454.01, 464.06, 476.62, 489.19, 505.51, 557.02, 677.61, 688.91, 706.50, 825.83, 938.88, 947.68, 952.70, 958.98, 961.49, 962.75, 965.26, 969.03, 977.82, 990.38, 996.67, 997.92

### 3.5.2 最优模型的比较

分别利用最优特征波长组合建立红提糖度和硬度的 RF 模型,校正集和预测集样本的预测值和化学测量值之间的散点图如图 9~10 所示。

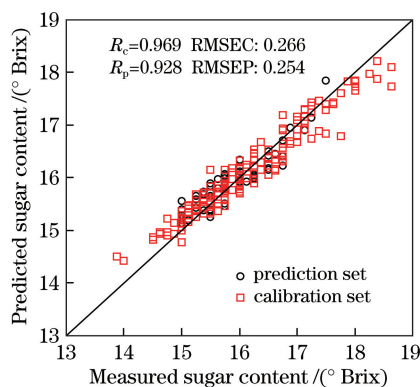


图 9 基于 GA-RF 的红提糖度的最优模型

Fig. 9 Optimal model for sugar content of red globe grape based on GA-RF

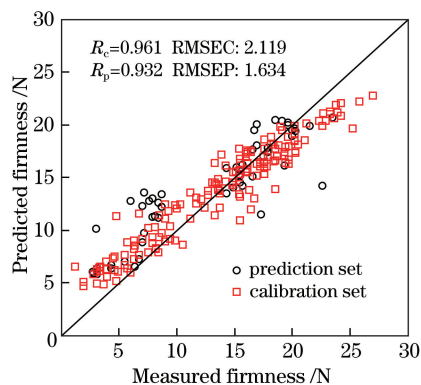


图 10 基于 MA-SPA-RF 的红提硬度的最优模型

Fig. 10 Optimal model for firmness of red globe grape based on MA-SPA-RF

由图 9~10 可知:训练集和预测集的数据都比较集中,模型的预测效果较好;所建立的红提糖度和硬度的最优 RF 模型的校正集相关系数  $R_c$  分别为 0.969、0.961,校正集均方根误差分别为 0.266、2.119;预测集相关系数  $R_p$  分别为 0.928、0.932,预测集均方根误差分别为 0.254、1.634,模型的稳定性和

准确性有了较大提高。

## 4 结 论

利用果柄侧朝下建立的糖度模型比果粒侧朝上、横放建立的糖度模型的效果更好,利用果柄侧朝上建立的硬度模型效果比果粒侧朝上、横放建立的硬度模型的效果更好,利用整个果粒的平均光谱建立的糖度、硬度模型效果最好。采用 GA、SPA、CARS 和 UVE 算法对原始光谱提取特征波长,波点数量明显减少,波点提取后所建模型的相关系数与基于原始光谱所建模型相比明显增大,均方根误差明显减小,4 种特征波长提取方法都有效地提高了预测模型的准确性和可靠性;基于 RF 建立的糖度、硬度模型的效果较优;预测糖度的最优模型为 GA-RF,其  $R_c$ 、 $R_p$  分别为 0.969、0.928,校正集均方根误差、预测集均方根误差分别为 0.266、0.254;预测硬度的最优模型为 MA-SPA-RF,其  $R_c$ 、 $R_p$  分别为 0.961、0.932,校正集均方根误差、预测集均方根误差分别为 2.119、1.634。研究结果表明,基于高光谱成像技术预测红提的糖度和硬度是可行的。

## 参 考 文 献

- [1] Xu F, Fu D D, Wang Q H, *et al.* Nondestructive detection of sugar content and acidity in red globe table grapes using visible near infrared spectroscopy based on Monte-Carlo cross validation-competitive adaptive reweighted sampling-random forest (MCCV-CARS-RF) [J]. Food Science, 2018, 39(8): 149-154.  
许锋,付丹丹,王巧华,等. 基于 MCCV-CARS-RF 建立红提糖度和酸度的可见-近红外光谱无损检测方法[J]. 食品科学, 2018, 39(8): 149-154.
- [2] Tian H L, Zhang P, Nong S Z, *et al.* Analysis on texture properties of treated grape fruit with 1-MCP based on TPA Test [J]. Food & Machinery, 2011, 27(3): 104-107.  
田海龙,张平,农绍庄,等. 基于 TPA 测试法对 1-



- MCP 处理后葡萄果实质构性能的分析[J]. 食品与机械, 2011, 27(3): 104-107.
- [3] Deng Y, Wu Y, Li Y F. Effects of temperature and relative humidity on postharvest grapes berry firmness[J]. Food Science, 2007, 28(3): 46-49.  
邓云, 吴颖, 李云飞. 温度和相对湿度对采后葡萄浆果硬度的影响[J]. 食品科学, 2007, 28(3): 46-49.
- [4] Dong J L, Guo W C. Nondestructive detection of soluble solid content of postharvest kiwifruits based on hyperspectral imaging technology[J]. Food Science, 2015, 36(16): 101-106.  
董金磊, 郭文川. 采后猕猴桃可溶性固形物含量的高光谱无损检测[J]. 食品科学, 2015, 36(16): 101-106.
- [5] Guan X M, Du J, Zhang L R, *et al.* Optimization and visualization for the prediction of apple sugar content based on hyperspectral imaging technology [J]. Journal of Optoelectronics•Laser, 2018, 29(2): 173-180.  
管晓梅, 杜军, 张立人, 等. 基于高光谱技术的果糖检测优化算法和可视化方法 [J]. 光电子•激光, 2018, 29(2): 173-180.
- [6] Wang Z W, Chi Q, Guo W C, *et al.* Internal quality detection of apples during late developmental period based on near-infrared spectral technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(5): 348-354.  
王转卫, 迟茜, 郭文川, 等. 基于近红外光谱技术的发育后期苹果内部品质检测 [J]. 农业机械学报, 2018, 49(5): 348-354.
- [7] Li B, Cobo-Medina M, Lecourt J, *et al.* Application of hyperspectral imaging for nondestructive measurement of plum quality attributes[J]. Postharvest Biology and Technology, 2018, 141: 8-15.
- [8] Lu R. Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images[J]. Sensing and Instrumentation for Food Quality and Safety, 2007, 1(1): 19-27.
- [9] Liu Y D, Chen X M, Ouyang A G. Non-destructive measurement of soluble solid content in Gannan navel oranges by visible/near-infrared spectroscopy[J]. Acta Optica Sinica, 2008, 28(3): 478-481.  
刘燕德, 陈兴苗, 欧阳爱国. 可见/近红外光谱法无损检测赣南脐橙可溶性固形物[J]. 光学学报, 2008, 28(3): 478-481.
- [10] Sun J T, Ma B X, Dong J, *et al.* Study on maturity discrimination of Hami melon with hyperspectral imaging technology combined with characteristic wavelengths selection methods and SVM[J]. Spectroscopy and Spectral Analysis, 2017, 37(7): 2184-2191.  
孙静涛, 马本学, 董娟, 等. 高光谱技术结合特征波长筛选和支持向量机的哈密瓜成熟度判别研究[J]. 光谱学与光谱分析, 2017, 37(7): 2184-2191.
- [11] Feng D, Ji J W, Zhang L, *et al.* Optimal wavelengths extraction of apple brix and firmness based on hyperspectral imaging[J]. Chinese Journal of Luminescence, 2017, 38(6): 799-806.  
冯迪, 纪建伟, 张莉, 等. 基于高光谱成像提取苹果糖度与硬度最佳波长 [J]. 发光学报, 2017, 38(6): 799-806.
- [12] Li R, Fu L S. Nondestructive measurement of firmness and sugar content of blueberries based on hyperspectral imaging[J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(z1): 362-366.  
李瑞, 傅隆生. 基于高光谱图像的蓝莓糖度和硬度无损测量 [J]. 农业工程学报, 2017, 33(z1): 362-366.
- [13] Li X L, Wei Y Z, Xu J, *et al.* SSC and pH for sweet assessment and maturity classification of harvested cherry fruit based on NIR hyperspectral imaging technology[J]. Postharvest Biology and Technology, 2018, 143: 112-118.
- [14] Mo C, Kim M S, Kim G, *et al.* Spatial assessment of soluble solid contents on apple slices using hyperspectral imaging[J]. Biosystems Engineering, 2017, 159: 10-21.
- [15] Fan S X, Huang W Q, Guo Z 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.  
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.
- [16] Zhu D S, Zhang Q M, Cao X H, *et al.* Effect of relative humidity on the changes in water and texture of kyoho grape during storage[J]. Food Science, 2014, 35(22): 340-345.  
朱丹实, 张巧曼, 曹雪慧, 等. 湿度条件对巨峰葡萄贮藏过程中水分及质构变化的影响 [J]. 食品科学, 2014, 35(22): 340-345.
- [17] Jiang X S, Zhou H P. Research progress on non-destructive testing of agricultural products quality based on genetic algorithm [J]. Jiangsu Agricultural Sciences, 2013, 41(12): 17-19, 20.  
蒋雪松, 周宏平. 基于遗传算法的农产品品质无损检测研究进展 [J]. 江苏农业科学, 2013, 41(12): 17-19, 20.
- [18] Ye S F, Wang D, Min S G. Successive projections algorithm combined with uninformative variable elimination for spectral variable selection[J]. Chemometrics and Intelligent Laboratory Systems,

- 2008, 91(2): 194-199.
- [19] Li J B, Peng Y K, Chen L P, *et al.* Near-infrared hyperspectral imaging combined with CARS algorithm to quantitatively determine soluble solids content in “Ya” pear[J]. Spectroscopy and Spectral Analysis, 2014, 34(5): 1264-1269.  
李江波, 彭彦昆, 陈立平, 等. 近红外高光谱图像结合 CARS 算法对鸭梨 SSC 含量定量测定[J]. 光谱学与光谱分析, 2014, 34(5): 1264-1269.
- [20] Liu Y D, Xiao H C, Sun X D, *et al.* Spectral feature selection and discriminant model building for citrus leaf Huanglongbing[J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(3): 180-187.  
刘燕德, 肖怀春, 孙旭东, 等. 柑桔叶片黄龙病光谱特征选择及检测模型[J]. 农业工程学报, 2018, 34(3): 180-187.
- [21] Liu Y D, Ye L Y, Sun X D, *et al.* Maturity evaluation model of tangerine based on spectral index [J]. Chinese Optics, 2018, 11(1): 83-91.  
刘燕德, 叶灵玉, 孙旭东, 等. 基于光谱指数的蜜橘成熟度评价模型研究[J]. 中国光学, 2018, 11(1): 83-91.
- [22] Jiang X Q, Ye Q, Lin Y, *et al.* Inverting study on soil water content based on harmonic analysis and hyperspectral remote sensing[J]. Acta Optica Sinica, 2017, 37(10): 1028001.  
姜雪芹, 叶勤, 林怡, 等. 基于谐波分析和高光谱遥感的土壤含水量反演研究[J]. 光学学报, 2017, 37(10): 1028001.
- [23] Mao B H, Sun H, Liu H J, *et al.* Prediction of winter wheat chlorophyll content based on gram-Schmidt and SPXY algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(S1): 160-165.  
毛博慧, 孙红, 刘豪杰, 等. 基于正交变换与 SPXY 样本划分的冬小麦叶绿素诊断[J]. 农业机械学报, 2017, 48(S1): 160-165.
- [24] Peng Y, Shi C J, Zhu Y M, *et al.* Qualitative and quantitative analysis algorithms based on terahertz spectroscopy for biomedical detection [J]. Chinese Journal of Lasers, 2019, 46(6): 0614002.  
彭滢, 施辰君, 朱亦鸣, 等. 太赫兹光谱技术在生物医学检测中的定性定量分析算法[J]. 中国激光, 2019, 46(6): 0614002.
- [25] Ma S X, Liu C T, Li H C, *et al.* Feature extraction based on linear embedding and tensor manifold for hyperspectral image[J]. Acta Optica Sinica, 2019, 39(4): 0412001.  
马世欣, 刘春桐, 李洪才, 等. 基于线性嵌入和张量流形的高光谱特征提取[J]. 光学学报, 2019, 39(4): 0412001.