

苹果可溶性固形物含量的多元线性回归预测

孟庆龙, 尚静, 张艳
(贵阳学院, 贵阳 550005)

摘要: **目的** 可溶性固形物含量是评价苹果品质的重要指标, 为开发苹果品质快速检测设备提供理论基础。**方法** 采用高光谱图像采集系统采集“富士”苹果的高光谱图像, 并获取感兴趣区域的反射光谱; 应用连续投影算法对标准正态变换预处理后的光谱进行降维; 基于选取的特征光谱建立预测苹果可溶性固形物含量的多元线性回归模型。**结果** 采用连续投影算法从 256 个全光谱中提取了 12 个波长作为特征光谱, 明显提升了多元线性回归预测模型的运行效率; 基于特征光谱建立的多元线性回归预测模型具有较好的校正性能 ($R_C=0.804$, $R_{Cm}=0.665\%$) 和预测性能 ($R_p=0.859$, $R_{Pm}=0.413\%$)。**结论** 研究建立的苹果可溶性固形物含量预测模型性能较稳定, 可以满足实际应用需求。

关键词: 高光谱成像; 苹果; 可溶性固形物含量; 多元线性回归; 无损检测

中图分类号: TB255.7 文献标识码: A 文章编号: 1001-3563(2020)13-0026-05

DOI: 10.19554/j.cnki.1001-3563.2020.13.004

Prediction for Soluble Solids Content of Apples Based on Multi Linear Regression

MENG Qing-long, SHANG Jing, ZHANG Yan

(Guiyang University, Guiyang 550005, China)

ABSTRACT: Soluble solids content is an important index to evaluate apple quality. The works aims to provide theoretical basis for the development of detection equipment for rapidly predicting apple quality. The hyperspectral image acquisition system was used to collect hyperspectral images of "Fuji" apples and obtain the reflectance spectra in the regions of interest. The successive projection algorithm was used for the dimensionality reduction of the reflectance spectra subject to standard normal variation preprocessing. The multi linear regression model was established based on selected characteristic wavelengths to predict soluble solids content of apples. The results showed that 12 wavelengths as characteristic spectra were extracted by successive projection algorithm from 256 full spectra, and the working efficiency of multi linear regression prediction model was obviously improved. The multi linear regression model based on characteristic spectra had better calibration ability ($R_C=0.804$, $R_{Cm}=0.665\%$) and prediction ability ($R_p=0.859$, $R_{Pm}=0.413\%$). The prediction model established in this study for detection of soluble solids content of apples has stable properties and can meet the requirements of practical application.

KEY WORDS: hyperspectral imaging; apple; soluble solids content; multi linear regression; nondestructive detection

收稿日期: 2020-01-13

基金项目: 国家自然科学基金 (61505036); 贵州省科技计划 (黔科合基础[2020]1Y270); 贵州省普通高等学校工程研究中心项目 (黔教合 KY 字[2016]017)

作者简介: 孟庆龙 (1989—), 男, 博士, 贵阳学院副教授, 主要研究方向为农产品无损检测。

通信作者: 张艳 (1977—), 女, 博士, 贵阳学院教授, 主要研究方向为农产品无损检测。

苹果的可溶性固形物含量 (Soluble Solids Content, SSC) 是评价其内部品质的关键指标之一, 同时也是判别苹果成熟度的主要指标。由此可见, 检测苹果的 SSC 不仅有助于指导消费, 而且对于指导苹果的采收时间、采后贮藏及其加工都具有重要的作用。

传统的水果 SSC 检测方法主要利用折射仪来检测果汁的糖度^[1-2], 而这种方法最大的问题是会破坏样品, 且不能进行大批量检测。由此可见, 寻求一种快速且不破坏样品的水果 SSC 测定方法在水果内部品质检测领域中占有非常重要的地位。近年来, 基于光谱技术的水果品质无损检测以其快速、无损等诸多优势深受国内外科研工作者的青睐^[3-6]。国内外研究人员已利用光谱技术开展了关于草莓^[7]、猕猴桃^[1,8]、梨^[9-10]、脐橙^[11]、西瓜^[12]和蓝莓^[13-14]等水果 SSC 的无损检测研究, 并取得了较好的研究成果。采用高光谱技术结合多元线性回归 (Multi Linear Regression, MLR) 对苹果 SSC 的无损检测研究相对较少。

文中采用高光谱图像采集系统来采集“富士”苹果的高光谱图像, 并提取感兴趣区域的反射光谱, 然后利用连续投影算法对预处理后的光谱进行降维, 最终基于选取的特征光谱构建苹果 SSC 的多元线性回归模型, 以期能为开发水果品质快速检测设备提供理论基础。

1 实验

1.1 材料

实验所用“富士”苹果采购于当地的沃尔玛超市, 从买回来的苹果样本中选取表面无损伤苹果 120 个, 对其依次进行编号后贮藏在室温 (22±2)°C 条件下, 取样间隔为 1 d, 共取样 4 次。

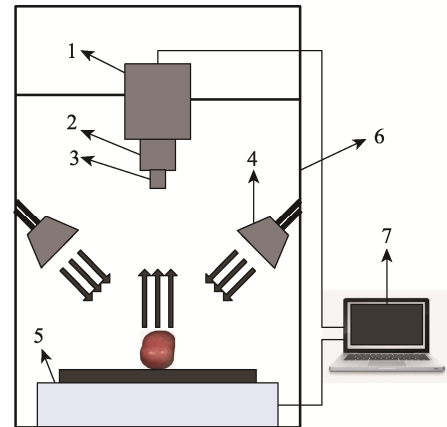
1.2 方法

1.2.1 高光谱图像采集与校正

实验采用四川双利合谱科技有限公司提供的高光谱图像采集系统 (GaiaField-F-V10) 来获取苹果样本的高光谱图像, 该系统主要包括 CCD 相机、成像光谱仪、镜头、溴钨灯光源、电动平移台、暗箱和高性能计算机等, 系统结构见图 1。其中, 成像光谱仪的图像空间分辨率为 696×700, 光谱采集范围为 392~1030 nm。为了确保采集的图像清晰且不失真, 相机的曝光时间为 9.5 ms, 样本距离镜头约 40 cm, 然后采集全白以及全黑的标定图像进行反射率校正。

1.2.2 SSC 的测定

实验采用日本 Atago 公司提供的 ATAGO PAL- α 数字手持袖珍折射仪来测定苹果的 SSC, 仪器测量范围为 0.0~85%, 测量精度为±0.2%。先将苹果样本



1. CCD 相机 2. 成像光谱仪 3. 镜头 4. 溴钨灯光源
5. 电动平移台 6. 暗箱 7. 高性能计算机

图 1 高光谱图像采集系统

Fig.1 Hyperspectral image acquisition system

榨汁后再离心 (离心机的转速为 3000 r/min, 离心时间为 5 min), 再将苹果果汁涂抹在折射仪的测试镜面上, 每个样本分别测量多次, 取其均值为实测值。

1.3 数据处理方法

1.3.1 光谱预处理及数据降维

通过前期比较多种光谱预处理方法 (一阶微分 (Derivative)、二阶微分 (Second Derivative)、标准正态变换 (Standard Normal Variation, SNV) 以及多元散射校正 (Multi-scatter Calibration, MSC) 等) 对原始光谱数据预处理的建模效果, 得出 SNV 是最优的光谱预处理方法^[2]。采用连续投影算法 (Successive Projection Algorithm, SPA) 对预处理后的光谱数据进行降维, 以提升模型的运行效率。SPA 是一种使矢量空间共线性最小化的前向变量选择算法, 其优势在于能从全波段中提取几个特征波长, 进而消除原始光谱中的冗余信息。光谱预处理及其降维过程在 Matlab R2016b 软件中执行。SNV 预处理方法的具体计算见式 (1)。

$$\begin{cases} Z_{ij} = \frac{x_{ij} - \bar{x}_i}{S_i} \\ \bar{x}_i = \frac{1}{p} \sum_{j=1}^p x_{ij} \\ S_i = \sqrt{\frac{1}{p-1} \sum_{j=1}^p (x_{ij} - \bar{x}_i)^2} \end{cases} \quad (1)$$

式中: $i=1, 2, \dots, n$; $j=1, 2, \dots, p$; n 为样本数量; p 为光谱点数量。

1.3.2 建模方法和模型评价

在建立苹果 SSC 预测模型时, 当有 m 个自变量 X (光谱数据), 即 $X_1, X_2, X_3, \dots, X_m$, 则自变量 X 和因变量 Y (SSC 实测值) 之间的线性回归方程为:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m + \varepsilon \quad (2)$$

设观察值为 $\{(Y_i, X_{i1}, \dots, X_{im}), i=1, 2, \dots, n\}$ ，则多元线性回归 (Multi Linear Regression, MLR) 模型可以表示成如下方程：

$$\left. \begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \dots + \beta_m X_{1m} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \dots + \beta_m X_{2m} + \varepsilon_2 \\ &\dots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \dots + \beta_m X_{nm} + \varepsilon_n \end{aligned} \right\} \quad (3)$$

以校正集相关系数 (R_C) 及其均方根误差 (Root Mean Squares Errors for Calibration, RMSEC, 用 R_{Cm} 表示), 预测集相关系数 (R_P) 及其均方根误差 (Root Mean Squares Errors for Prediction, RMSEP, 用 R_{Pm} 表示) 作为判别模型性能的关键指标, 其中:

$$R_C = \sqrt{\frac{\sum_{i=1}^{n_c} [y_{act}(i) - y_{cal}(i)]^2}{\sum_{i=1}^{n_c} [y_{act}(i) - y_{mean}(i)]^2}} \quad (4)$$

$$R_P = \sqrt{\frac{\sum_{i=1}^{n_p} [y_{act}(i) - y_{pre}(i)]^2}{\sum_{i=1}^{n_p} [y_{act}(i) - y_{mean}(i)]^2}} \quad (5)$$

$$R_{Cm} = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} [y_{act}(i) - y_{cal}(i)]^2} \quad (6)$$

$$R_{Pm} = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} [y_{act}(i) - y_{pre}(i)]^2} \quad (7)$$

式中: n_c 和 n_p 为校正集和预测集中的样本数量; y_{act} 和 y_{mean} 为样本 SSC 的实测值和平均值; y_{cal} 和 y_{pre} 为校正集和预测集中样本 SSC 的预测值。

通常性能好的检测模型具有较大的 R_C 和 R_P 值, 以及较小的 R_{Cm} 和 R_{Pm} 值^[15]。

2 结果与分析

2.1 苹果 SSC 统计分析

采用 SPXY (Sample Set Partitioning Based on Joint x-y Distances) 方法^[16]按照 3 : 1 的比例将 120 个苹果样本划分为 90 个校正样本集和 30 个预测样本集。校正样本集和预测样本集中苹果样本 SSC 的频率分布见图 2, 统计分析结果见表 1。从图 2 和表 1 可以看出, 校正样本集中苹果的 SSC 值具有较宽的范围, 这样有利于建立良好的检测模型。

2.2 苹果样本的光谱及预处理

校正样本集和预测样本集中苹果的原始反射光谱 (见图 3a) 以及经过 SNV (见图 3b) 预处理后的反射光谱曲线见图 3 (感兴趣区域为整个苹果样本, 见图 3a 中插图所示)。从图 3a 可以看出, 在波长 675 nm 附近的吸收峰由表面叶绿素的吸收而产生, 体现了苹果样本表面颜色信息, 波长 980 nm 附近的吸收峰则由水分的吸收而产生, 反映了苹果的水分含量信息^[17]。

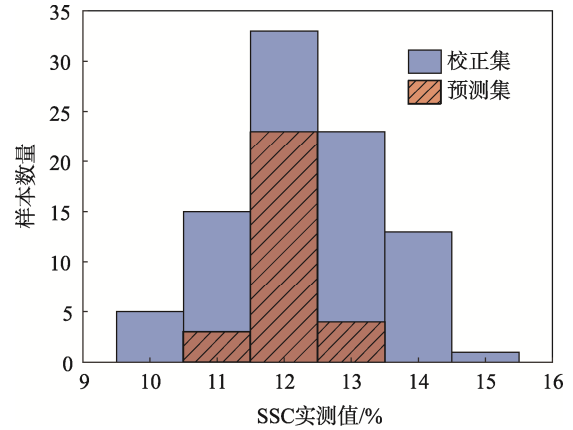


图 2 苹果 SSC 频率分布

Fig.2 Probability distribution of SSC of apples

表 1 苹果 SSC 的统计结果

Tab.1 Statistics results of SSC of apples

样本集	样本数量	可溶性固形物含量 (质量分数) /%			
		最小值	最大值	平均值	标准差
校正集	90	9.5	14.5	12.3	1.1
预测集	30	10.8	13.3	12.1	0.5

2.3 光谱数据降维

采用 SPA 对光谱数据进行降维时, 通常根据最小的 RMSEP 值来确定最佳的变量数量。RMSEP 随 SPA 中有效波长数的变化规律见图 4a。由图 4a 可见, RMSEP 值随着有效波长数的增加而减小, 当有效波长数大于 12 时, RMSEP 变化不明显, 因此将这 12 个波长作为特征光谱, 按重要性从高到低依次为 584.89, 411.14, 680.62, 434.52, 646.03, 964.16, 422.82, 870.41, 519.80, 546.20, 611.69, 727.99 nm。所有苹果样本的原始光谱反射率在前 3 个特征波长 584.89, 411.14, 680.62 nm 处的反射率变化分别为: 0.078~0.445, 0.056~0.152, 0.181~0.512。采用 SPA 选取的特征光谱分布见图 4b。

2.4 多元线性回归建模结果

将 SPA 选取的特征光谱作为多元线性回归模型的自变量, 校正样本集和预测样本集中苹果 SSC 实测值作为因变量, 构建的苹果 SSC 预测模型具有较好的校正性能 ($R_C=0.804$, $R_{Cm}=0.665\%$) 和预测性能 ($R_P=0.859$, $R_{Pm}=0.413\%$)。采用 SPA 特征波长选择方法从 256 个全光谱波段中提取了 12 个特征变量作为多元线性回归预测模型的光谱输入, 明显提升了预测模型的运行效率。由此表明, 基于 SPA 提取的特征光谱建立的多元线性回归模型对苹果 SSC 的无损检测具有良好的预测能力, 可应用在苹果 SSC 的无损检测中。其中, 苹果 SSC 的实测值和预测值见图 5。

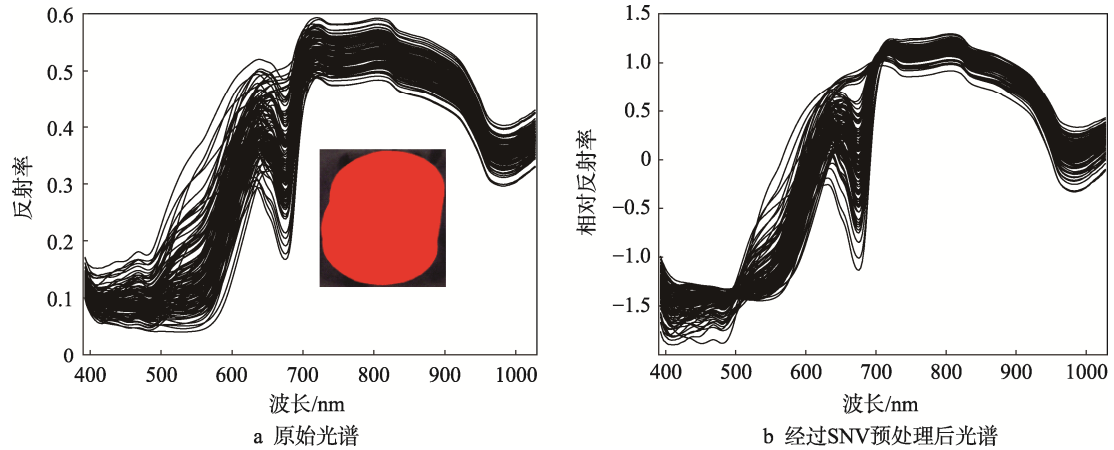


图 3 苹果样本的反射光谱
Fig.3 Reflectance spectra of apple samples

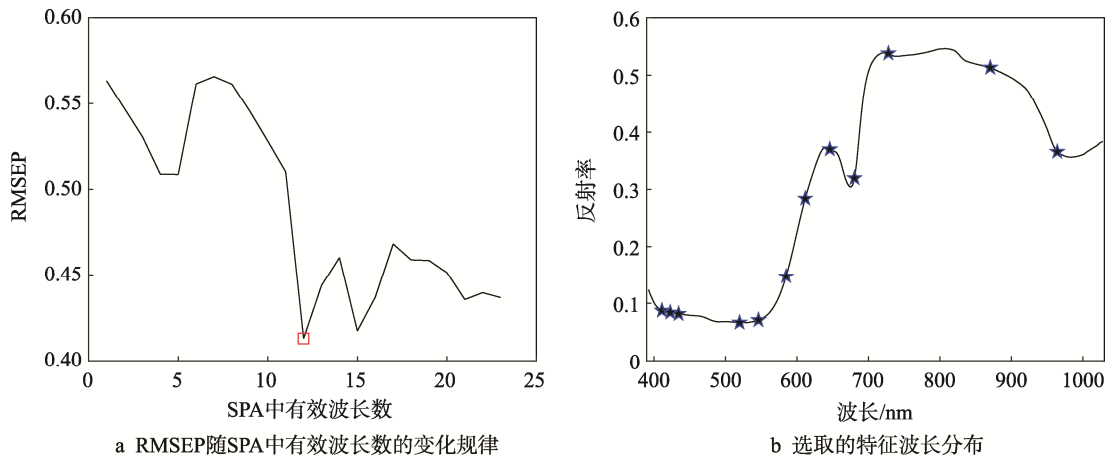


图 4 基于 SPA 算法从全光谱中提取特征波长的结果
Fig.4 Results of characteristic wavelengths extracted from full spectra based on SPA algorithm

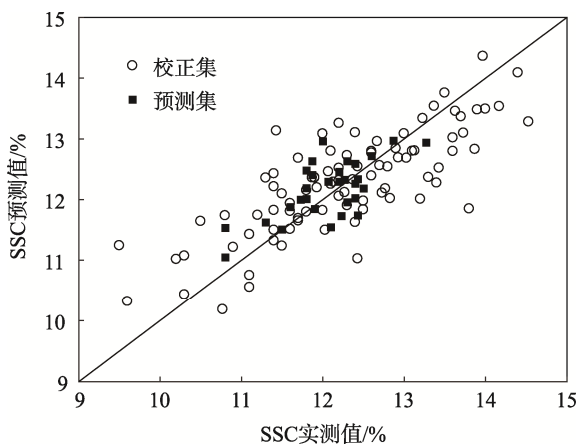


图 5 苹果 SSC 的预测结果
Fig.5 Prediction results of SSC of apples

3 结语

针对苹果内部品质无损检测的应用需求,基于高光谱成像技术和多元线性回归分析方法开展了苹果

可溶性固形物含量的快速无损检测。将高光谱成像技术应用在水果品质的无损检测中存在的主要问题是数据量大,进而导致模型运算效率低。文中研究应用连续投影算法从全光谱的 256 个波长中选取了 12 个特征波长,这样明显提升了预测模型的运行效率。基于选取的特征光谱建立的多元线性回归预测模型对苹果 SSC 的无损检测具有良好的预测能力 ($R_p=0.859, R_{Pm}=0.413\%$),可以满足苹果 SSC 检测的应用需求。结果表明,建立的苹果可溶性固形物含量多元线性回归预测模型性能较稳定,可应用于苹果可溶性固形物含量的无损检测。

参考文献:

[1] 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.
DONG Jin-lei, GUO Wen-chuan. Nondestructive Detection of Soluble Solid Content of Postharvest Kiwifruits Based on Hyperspectral Imaging Technology[J]. Food Science, 2015, 36(16): 101—106.

- [2] 尚静, 孟庆龙, 张艳, 等. 紫外/可见光谱技术无损检测李子可溶性固形物含量[J]. 食品工业科技, 2020, 41(3): 228—231.
SHANG Jing, MENG Qing-long, ZHANG Yan, et al. Nondestructive Detection of Soluble Solids Content of Plums Based on UV/VIS Spectroscopy Technology[J]. Science and Technology of Food Industry, 2020, 41(3): 228—231.
- [3] HOU J, ZHANG Y, SUN Y, et al. Prediction of Firmness and pH for "Golden Delicious" Apple Based on Elasticity Index from Modal Analysis[J]. Journal of Food Science, 2018, 83: 1—9.
- [4] MA T, LI X, INAGAKI T, et al. Noncontact Evaluation of Soluble Solids Content in Apples by Near-infrared Hyper spectral Imaging[J]. Journal of Food Engineering, 2018, 224: 53—61.
- [5] ZHANG D, XU L, WANG Q, 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, 2018, 12: 136—147.
- [6] FAN S, ZHANG B, LI J, et al. Prediction of Soluble Solids Content of Apple Using the Combination of Spectra and Textural Features of Hyperspectral Reflectance Imaging Data[J]. Postharvest Biology and Technology, 2016, 121: 51—61.
- [7] 邵园园, 王永贤, 玄冠涛, 等. 高光谱成像快速检测壳聚糖涂膜草莓可溶性固形物[J]. 农业工程学报, 2019, 35(18): 245—254.
SHAO Yuan-yuan, WANG Yong-xian, XUAN Guan-tao, et al. Rapid Detection of Soluble Solids Content in Strawberry Coated with Chitosan Based on Hyperspectral Imaging[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(18): 245—254.
- [8] GUO W, ZHAO F, DONG J. Nondestructive Measurement of Soluble Solids Content of Kiwifruits Using Near-infrared Hyperspectral Imaging[J]. Food Analytical Methods, 2016, 9: 38—47.
- [9] FAN S, HUANG W, GUO Z, et al. Prediction of Soluble Solids Content and Firmness of Pears Using Hyper Spectral Reflectance Imaging[J]. Food Analytical Methods, 2015, 8: 1936—1946.
- [10] YU X, LU H, WU D. Development of Deep Learning Method for Predicting Firmness and Soluble Solid Content of Postharvest Korla Fragrant Pear Using VIS/NIR Hyperspectral Reflectance Imaging[J]. Postharvest Biology and Technology, 2018, 141: 39—49.
- [11] 宋杰, 李光林, 杨晓东, 等. 基于四方对称光源透射光谱的脐橙可溶性固形物检测[J]. 农业工程学报, 2019, 35(10): 267—273.
SONG Jie, LI Guang-lin, YANG Xiao-dong, et al. Detecting Soluble Solids Content of Navel Orange Based on Transmission Spectrum of Tetragonal Symmetric Light Source[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(10): 267—273.
- [12] 王世芳, 韩平, 崔广禄, 等. SPXY 算法的西瓜可溶性固形物近红外光谱检测[J]. 光谱学与光谱分析, 2019, 39(3): 738—742.
WANG Shi-fang, HAN Ping, CUI Guang-lu, et al. The NIR Detection Research of Soluble Solid Content in Watermelon Based on SPXY Algorithm[J]. Spectroscopy and Spectral Analysis, 2019, 39(3): 738—742.
- [13] 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(1): 362—366.
LI Rui, FU Long-sheng. Nondestructive Measurement of Firmness and Sugar Content of Blueberries Based on Hyperspectral Imaging[J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(1): 362—366.
- [14] LEIVA-VALENZUELA G A, LU R, AGUILERA J M. Assessment of Internal Quality of Blueberries Using Hyperspectral Transmittance and Reflectance Images with Whole Spectra or Selected Wavelengths[J]. Innovative Food Science & Emerging Technologies, 2014, 24: 2—13.
- [15] TIAN X, FAN S, LI J, et al. Comparison and Optimization of Models for SSC On-line Determination of Intact Apple Using Efficient Spectrum Optimization and Variable Selection Algorithm[J]. Infrared Physics & Technology, 2019, 102(102979): 1—11.
- [16] GALVÃO R K H, ARAUJO M C U, JOSÉ G E, et al. A Method for Calibration and Validation Subset Partitioning[J]. Talanta, 2005, 67(4): 736—740.
- [17] 黄文倩, 陈立平, 李江波, 等. 基于高光谱成像的苹果轻微损伤检测有效波长选取[J]. 农业工程学报, 2013, 29(1): 272—277.
HUANG Wen-qian, CHEN Li-ping, LI Jiang-bo, et al. Effective Wavelengths Determination for Detection of Slight Bruises on Apples Based on Hyperspectral Imaging[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(1): 272—277.