

蒋霞, 张 晓, 白铁成, 等. 近红外光谱技术结合 PLS 和 SPA 检测鲜冬枣表面农药残留量的方法[J]. 江苏农业科学, 2018, 46(2): 146–149.
doi:10.15889/j.issn.1002-1302.2018.02.039

近红外光谱技术结合 PLS 和 SPA 检测 鲜冬枣表面农药残留量的方法

蒋霞, 张 晓, 白铁成, 陈 杰, 张楠楠

(塔里木大学信息工程学院/中国农业科学院农业信息研究所新疆南疆农业信息化研究中心, 新疆阿拉尔 843300)

摘要:以喷洒不同浓度毒死蜱的鲜冬枣为研究对象, 研究近红外光谱技术结合偏最小二乘法(PLS)和连续投影算法(SPA)检测鲜冬枣表面农药残留的方法。运用 Antaris II 近红外光谱仪对喷洒不同浓度的毒死蜱的鲜冬枣样品进行扫描, 首先建立全波段 PLS 模型, 然后应用 SPA 提取特征波长, 作为 PLS 的输入变量, 建立 SPA-PLS 模型, 将全波段 PLS 模型和 SPA-PLS 模型进行比较。经 SPA 提取 5 个特征波长建立的模型, 使用变量数仅占全波段的 0.32%, 但建立的冬枣表面农药残留模型的准确度和精度优于全波段所建立的模型。近红外光谱技术结合 SPA 和 PLS 建立鲜冬枣表面不同浓度毒死蜱农药残留的模型是可行的, 同时 SPA 算法简化模型复杂度, 提高模型精度及稳定性。

关键词:近红外光谱; 农药残留量; SPA; 冬枣; PLS

中图分类号: O433.1; O657.33; S481+.8 **文献标志码:** A **文章编号:** 1002-1302(2018)02-0146-04

冬枣是新疆南疆地区的重要经济作物, 它以味道鲜美, 糖分含量高, 富含 19 种人体所需氨基酸和维生素、多种微量元素和较多的药用成分, 被公认为品质最好的鲜食枣品种之一^[1], 具有很高的食疗和多种保健功效, 被誉为“活维生素丸”^[2]。由于大面积种植以及气候等环境因素的影响, 冬枣树病虫害越来越严重, 严重影响了果农的利益, 果农不得不使用农药, 但不合理的滥用容易造成冬枣表面农药残留超标。随着人们生活水平的提高和健康环保意识的增强, 消费者越来越关注果蔬的质量安全问题, 其中果蔬表面的农药残留问题尤其备受关注^[3]。由于冬枣在采摘前最后 1 次喷洒农药的时间与上市期的间隔时间较短, 导致上市鲜冬枣会出现不同量的农药残留现象, 而且很多消费者食用前不清洗, 或者即使清洗也清洗不彻底, 因此, 导致很多人受害于农药残留而并不自知。及时对市场上销售的鲜冬枣进行农药残留限量检测显得极为重要。

近年来, 国内外研究学者采用液相色谱法^[4]、气相色谱法^[5]、气相色谱-串联质谱法^[6-7]、酶抑制法^[8]、超临界流体萃取法(SFE)^[9]、生物传感器法^[10]等技术对果蔬类农药残留进行了研究, 并取得了相关进展。然而, 这些检测方法尤其是化学方法虽然检测精度高, 但普遍存在耗时长、费用昂贵、对样品损坏严重、样品和试剂浪费严重等现象, 且对检测产品均

有一定的破坏性, 无法实现经济成本低廉的快速无损检测^[11]。因此, 迫切需要一种实时、无损、在线农药残留快速筛查技术, 以适应特色林果业的健康快速发展需求。

近红外光谱技术可利用近红外光谱区域的全谱或部分波段光谱对被测物进行快速定量或定性分析^[12-13]。该技术已用于农产品品质检测、种类鉴别、农药残留分析等研究^[14-17]。但近红外光谱通常包含数以千计的波长变量, 利用全波段数据建模时, 并非每个波长都能提供有用信息, 大量冗余数据会增加建模工作量。开展波长变量优选, 不仅可以剔除无关变量, 还能提高模型的预测精度以及稳健性。同时, 优选的特征波长可用于构建在线光谱快速测量系统, 从而降低生产成本^[18]。连续投影算法(SPA)利用向量的投影分析, 寻找含有最少冗余信息的变量组合, 使变量间的共线性小, 该方法在光谱特征波长优选中应用广泛^[19-20]。偏最小二乘法(PLS)是一种常见的多元统计数据分析方法, 被广泛应用于近红外、高光谱、拉曼、核磁和质谱等波谱定量模型的建立, 几乎成为光谱分析中建立定量校正模型的通用方法。

基于以上考虑, 本研究以南疆鲜冬枣为对象, 通过近红外光谱技术得到冬枣样品表面吸光度信息, 以毒死蜱农药为例, 建立喷洒不同浓度农药的鲜冬枣样品的全波段偏最小二乘法(PLS)模型, 同时应用 SPA 提取特征波长, 作为 PLS 的输入变量, 建立了 SPA-PLS 模型, 比较 2 种模型的预测精度, 以期对鲜冬枣农药残留的快速检测提供一种新的思路和方法, 并对其他果蔬类农药残留光谱检测提供借鉴意义。

1 材料与方法

1.1 材料与试验设计

选取新疆南疆某农场购买的鲜冬枣为研究对象, 随机选取 180 个成熟度良好、颜色均匀的冬枣样本, 所用的农药为毒死蜱, 购买于国家物质中心, 剂型为乳油, 质量百分含量为 40%。把农药配置成 1:10、1:50、1:100、1:1 000 不同浓

收稿日期: 2017-04-20

基金项目: 国家自然科学基金应急管理项目(编号: 61640413); 中国农业科学院农业信息研究所新疆南疆农业信息化研究中心开放课题(编号: ZX2015005, ZX2015004)。

作者简介: 蒋霞(1982—), 女, 甘肃甘谷人, 硕士, 讲师, 主要从事农产品品质无损检测和光谱图像方面的研究。E-mail: 49429533@qq.com。

通信作者: 张 晓, 硕士, 讲师, 主要从事农产品品质无损检测和光谱图像方面的研究。E-mail: zhangxiaoscnu@163.com。

度梯度的溶液和空白对照溶液(只含蒸馏水)。

首先对 180 个冬枣样本进行清洗并自然风干,然后立即按照试验要求,随机分为 5 组,每组 36 个,将配好的 4 种浓度的农药溶液及空白对照组分别喷洒 1 组冬枣,其中每种浓度随机抽取 30 个冬枣作为建模集,剩余的 6 个冬枣作为预测集。利用喷雾器喷洒,保证喷洒均匀。

将喷洒过药液的冬枣放在通风处 12 h 后,进行近红外光谱图像的采集。整个过程均在实验室进行,实验室环境条件为温度 20 ~ 25 °C,相对湿度为 30% ~ 40%。

1.2 光谱采集仪器及方法

实验设备为 Antaris II 傅立叶变换近红外(FT-NIR)光谱仪(Thermo Scientific, USA),无需另外提供采样背景,以仪器内部空气为背景,测量波长范围 4 000 ~ 10 000 cm^{-1} ,分辨率为 4 cm^{-1} ,共 1 557 个采样点数,每张光谱扫描 32 次。测样方式:开机预热 30 min 后,用近红外光谱仪对红枣的可区别 3 个部位分别扫描 32 次,求其平均值作为原始光谱。应用 OMNIC 软件和与之配套的标准白板采集归一化后的光谱数据并转换格式,利用 MATLAB2010b(美国 Mathworks)软件进行数据分析和处理。本次采用的近红外光谱采集设备系统见图 1。



图1 近红外光谱采集设备系统

1.3 连续投影算法

连续投影算法(successive projections algorithm, SPA)^[18]在光谱分析建模中可以提取光谱数据的特征波长,减少数据维度的同时能够有效避免信息重叠,最大限度提取解释信息,从而提高预测精度和速度,使模型能够更有效地达到实际利用目的。

设 $X(n \times p)$ 为鲜冬枣表面不同浓度农药残留的吸光度矩阵,其中 n 为样本容量, p 为全谱波的个数。设 x_k 为初始迭代量, M 为提取波长变量数的范围,SPA 的具体算法为:

预测集的光谱矩阵 $X(n \times m)$ 给出需要选择的波长数 n , SPA 算法如下:

(1) 在开始迭代前($m=1$)把校正集光谱矩阵的任意第 k 列($k=1 \cdots p, p$ 为波长总数)赋值给 x_k ;

(2) 把没有被选中的其他波长点位置集合记为 S :

$$S = \{k, 1 \leq k \leq p, k \notin \{\text{var}(0), \cdots, \text{var}(m-1)\}\};$$

(3) 计算 x_k 对剩余列向量 $X_{\text{var}(m-1)}$ 的正交投影:

$$P_{x_k} = x_k - [x_k^T x_{\text{var}(m-1)}] x_{\text{var}(m-1)} [x_{\text{var}(m-1)}^T x_{\text{var}(m-1)}]^{-1}, k \in S;$$

(4) 记

$$\text{var}(m) = \arg[\max(\|P_{x_j}\|), k \in S];$$

(5) 记

$$x_k = P_{x_k}, k \in S;$$

(6) $m = m + 1$,如果 $m < M$ 返回步骤(2)循环计算;

(7) 最后提取波长位置 $\{\text{var}(m) = 0, \cdots, M-1\}$;

对应于每一对 $\text{var}(m)$ 和 M ,循环 1 次后进行多元线性回归分析(MLR),得到验证集的标准偏差(RMSEP),最小 RMSEP 值对应的 $\text{var}(m)$ 和 M 则为最优值。

1.4 建模方法和模型验证指标

偏最小二乘法(partial least square regression, PLS)是通过建立光谱数据与农药浓度分类值之间的多元统计回归模型进行分析。除了线性回归分析,PLS 在建模过程中集合了包括主成分分析、典型相关分析等方法的功能特点,因此在分析结果中,不仅可以建立更优化的回归模型,而且在光谱分析中得到广泛应用^[19]。

将喷洒毒死蜱农药浓度为 1 : 10 的样本赋值为 1,喷洒毒死蜱农药浓度为 1 : 50 的样本赋值为 2,喷洒毒死蜱农药浓度为 1 : 100 的样本赋值为 3,喷洒毒死蜱农药浓度为 1 : 1 000 的样本赋值为 4,喷洒蒸馏水的样本赋值为 5。PLS 数值间的间距为 1。判别依据是以各数值为基准,上下 0.5 个单位之内的为该数值代表的农药浓度,结果见表 1。

表1 PLS 模型判别标准

| 农药浓度 | 鉴别标准范围 |
|-----------|-------------|
| 1 : 10 | 0.5 ~ 1.5 |
| 1 : 50 | > 1.5 ~ 2.5 |
| 1 : 100 | > 2.5 ~ 3.5 |
| 1 : 1 000 | > 3.5 ~ 4.5 |
| 0 | > 4.5 ~ 5.5 |

为有效评价模型精度,本研究选取相关系数(R)、建模集中交互验证预测均方差(SECV)、验证集中预测标准差(SEP)、准确率(precision)进行模型分析检验,其计算公式见表 2。其中, R 越接近 1,回归(或预测)结果越好;SECV 越小,说明该模型的预测能力越高;SEP 值越小,则表示模型对外部样品的预测能力越高;同一批次样本,SECV 和 SEP 值越小,说明模型的精度越高,二者值越接近说明模型稳定性越好;Precision 用来验证模型的正确程度,用来反应模型的正确性。

2 结果与分析

2.1 原始光谱曲线与特征分析

为采集喷洒 4 种不同浓度毒死蜱农药的样品组与对照组的原始光谱,从图 2 可以看出,喷洒不同浓度农药样本及对照组样本的近红外光谱在 4 000 ~ 10 000 cm^{-1} 曲线趋势基本相似,并没有太明显的差异,但 6 800、5 200 cm^{-1} 附近有明显的吸收峰存在,喷洒不同浓度农药的样品在此 2 波峰处的吸光强度不同。在所测范围 4 000 ~ 10 000 cm^{-1} 内,其吸光强度与毒死蜱农药的浓度呈负相关关系,毒死蜱浓度越高,在该范围内,其峰值越低,本结果与前人研究结论一致。

2.2 预处理结果

分别采用一阶导数(savitzky golay first derivative, SGFD)、二阶导数(savitzky golay second derivative, SGSD)、多元散射校正(multiplicative scatter correction, MSC)等 3 种方法对光谱进行预处理,光谱预处理结果见图 3。然后结合 PLS 对不同浓度农药进行建模。采用 3 种光谱预处理方法和原光谱所建立的 PLS 模型对冬枣中农药残留量分类的建模精度和预测能

表 2 偏最小二乘法模型的检验指标

| 验证指标 | 公式 |
|----------------------------|--|
| 相关系数(<i>R</i>) | $R = \sqrt{1 - \frac{\sum_{i=1}^n (y_{\text{实际}} - y_{\text{预测}})^2}{\sum_{i=1}^n (y_{\text{实际}} - \bar{y}_{\text{实际}})^2}}$ |
| 交互验证预测均方差(<i>RMSECV</i>) | $RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_{\text{实际}} - y_{\text{预测}})^2}{n}}$ |
| 预测标准差(<i>SEP</i>) | $SEP = \sqrt{\frac{\sum_{i=1}^n (y_{\text{实际}} - y_{\text{预测}})^2}{n}}$ |
| 准确率(<i>Precision</i>) | $Precision = 1 - \frac{ y_{\text{实际}} - y_{\text{预测}} }{y_{\text{预测}}}$ |

注: $y_{\text{实际}}$ 、 $\bar{y}_{\text{实际}}$ 、 $y_{\text{预测}}$ 分别为实测值、实测值均值、预测值; n 为建模集(或验证集)的冬枣样品数量。

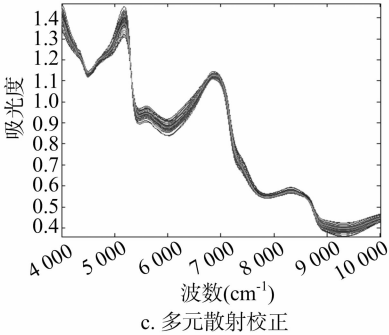
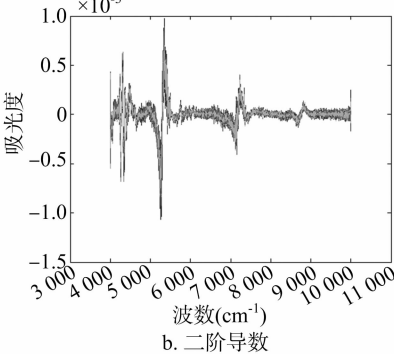
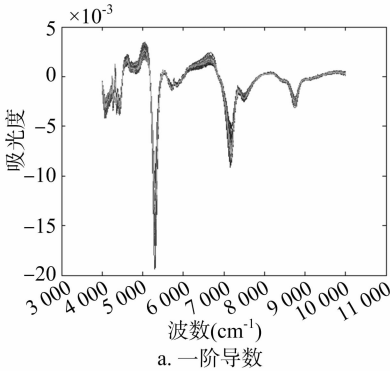
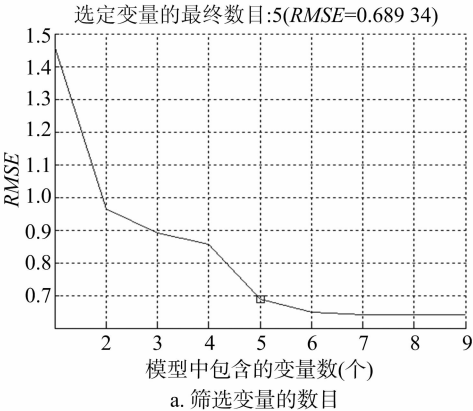


图 3 光谱预处理结果图像

表 3 不同预处理方法 PLS 建模结果

| 模型 | 预处理方法 | SEP | RMSECV | <i>R</i> | Precision |
|-----|-------|-----------|-----------|-----------|-----------|
| PLS | 原始光谱 | 0.653 683 | 0.573 286 | 0.886 780 | 0.737 027 |
| | 一阶求导 | 0.739 075 | 0.005 520 | 0.852 575 | 0.694 676 |
| | 二阶求导 | 1.232 677 | 0.689 179 | 0.490 157 | 0.478 865 |
| MSC | | 1.117 406 | 1.609 444 | 0.612 946 | 0.485 502 |

集样品划分处理,最后使用 SPA 算法进行光谱变量压缩。指定提取波长变量数 M 的范围为 2 ~ 20,采用均方根误差(RMSE)确定选定的最终变量数目。从图 4 可以看出,当最小均方根误差的最小值为 0.689 34 时,从原始光谱中共选出



5 个特征波长,分别为 3 999.64、4 578.18、5 179.861、5 283.999、7 162.325 cm^{-1} ,其重要性依次减弱。从所选变量的波长可以看出,冬枣表面毒死蜱农药在 4 000 ~ 5 300 cm^{-1} 波段存在相关性较高的光谱特征,可用于冬枣表面毒死蜱农药残留量的快速检测,其中 5 179.861、5 283.999 cm^{-1} 波段贡献值较大,本结果与已有研究中 5 200 cm^{-1} 附近是农药浓度分类敏感波段结论相吻合。

2.4 基于 SPA 特征波长的 PLS 模型

对从原始光谱做 SPA 后选出的 5 个特征波长,仍采用相同的 3 种光谱预处理方法和原始光谱建立 SPA - PLS 模型,

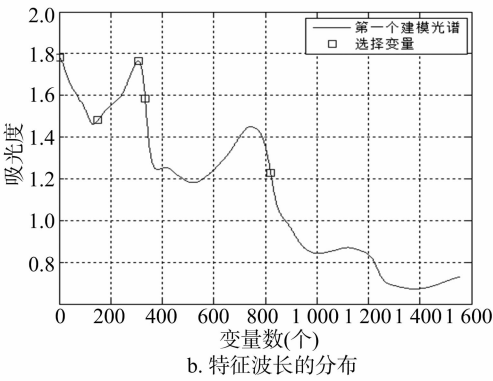


图 4 SPA 提取特征波长

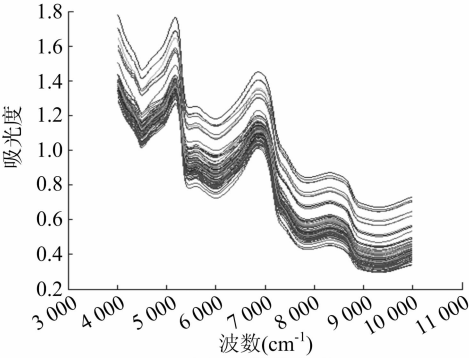


图 2 原始光谱

力见表 3。通过比较,采用原始光谱所建立的模型最优(图 4),故后面的连续投影算法以原始光谱为基础进行。

2.3 SPA 选择特定波长

采用连续投影算法在全谱范围内对原始光谱使用 SPXY (Sample set partitioning based on joint x - y distance) 进行校正

该模型对冬枣中农药残留量分类的建模精度和预测能力见表 4。与 PLS 模型的建模精度和预测能力比较可知,原始光谱的 SEP 由 0. 653 683 增加至 0. 691 782, RMSECV 由 0. 573 286 增加至 0. 782 928, *R* 由 0. 886 780 降低至 0. 872 192, Precision 由 0. 737 027 提高至 0. 795 669;一阶求导的 SEP 从 0. 739 075 提高至 0. 771 944, RMSECV 从 0. 005 520 增加至 0. 365 338, *R* 从 0. 852 575 降低至 0. 837 885, Precision 从 0. 694 676 增加至 0. 761 169;二阶求导的 SEP 从 1. 232 677 降低至 0. 708 189, RMSECV 从 0. 689 179 降低至 0. 033 439, *R* 从 0. 490 157 提高至 0. 865 583, Precision 从 0. 478 865 提高至 0. 668 338;MSC 的 SEP 由 1. 117 406 降低至 0. 916 856, RMSECV 由 1. 609 444 降低至 1. 000 208, *R* 由 0. 612 946 提高至 0. 761 372, Precision 由 0. 485 502 提高至 0. 558 819。试验结果表明,SPA - PLS 只用了 5 个特征波长来建模,所用变量数仅占全波段的 0. 32%,建立的冬枣表面农药残留模型的准确度和精度优于在全谱区所建立的 PLS 模型。因为近红外光谱的谱峰重叠严重,冗余信息多,同时在全谱区还包含大量与毒死蜱成分无关的信息,将全光谱的所有信息参与建模,在一定程度上降低了模型的性能^[20]。而经 SPA 变量从全光谱中优选出有效特征变量建立的模型,最大程度地保留了有用光谱信息,剔除了大量无用和冗余信息,提高了模型精度和稳定性。

表 4 SPA - PLS 建模结果

| 预处理方法 | SEP | RMSECV | <i>R</i> | Precision |
|-------|------------|------------|------------|------------|
| 原始光谱 | 0. 691 782 | 0. 782 928 | 0. 872 192 | 0. 795 669 |
| 一阶求导 | 0. 771 944 | 0. 365 338 | 0. 837 885 | 0. 761 169 |
| 二阶求导 | 0. 708 189 | 0. 033 439 | 0. 865 583 | 0. 668 338 |
| MSC | 0. 916 856 | 1. 000 208 | 0. 761 372 | 0. 558 819 |

3 结论

采用连续投影算法(SPA)分别对测定的 180 组光谱数据进行变量选择,从原始光谱中筛选了 3 999. 64、4 578. 18、5 179. 861、5 283. 999、7 162. 325 cm⁻¹ 共 5 个波长,降低了 PLS 模型的计算量,且连续投影算法高度概括了绝大多数样品光谱的信息,避免了信息重叠,同时去除了冗余信息,简化了模型,提高了模型精度和稳定性。同时,将得到的特征波长用于开发便携式农药残留检测仪,可以大大降低对硬件的要求,为研制水果表面农药残留便携带式光谱仪提供依据和理论基础。

研究结果表明,可见近红外光谱技术联合 PLS 和 SPA 是一种有效区分冬枣表面不同浓度农药残留量的方法。在今后的研究中,将引入多种农药种类,增加农药梯度设置,进一步验证该方法的有效性。但所建模型的精度和准确度相对较低,提高模型精度和准确度的方法还需进一步开展研究。

参考文献:

[1] 王亚萍,王贵禧,李艳菊. 冬枣贮藏保鲜技术研究进展[J]. 中国农学通报,2006,22(3):82-87.
[2] 曲泽洲,王永惠. 中国果树志·枣卷[M]. 北京:中国林业出版社,1993:229-230.
[3] 薛 龙,黎 静,刘木华. 基于高光谱图像技术的水果表面农药残留检测试验研究[J]. 光学学报,2008,28(12):2277-2280.

[4] Yu C S, Fan S F, Li Y J, et al. Representative commodity for five brassica vegetables based on the determination and dissipation of six pesticide residues [J]. International Journal of Environmental Analytical Chemistry, 2015, 95(5): 419-433.
[5] 吴海燕,陈建军. 蔬菜中多种有机磷农药残留气相色谱分析法[J]. 食品与机械, 2010, 26(4): 76-77, 83.
[6] Lozano A, Rajski L, Belmonte - Valles N, et al. Pesticide analysis in teas and chamomile by liquid chromatography and gas chromatography tandem mass spectrometry using a modified QuEChERS method: validation and pilot survey in real samples [J]. Journal of Chromatography A, 2012, 1268: 109-122.
[7] 苏建峰,钟茂生,陈 晶,等. 气相色谱-质谱法与气相色谱法测定茶叶及茶叶加工品中 295 种农药多残留[J]. 分析测试学报, 2015, 34(6): 625-638.
[8] 刘永杰,张金振,曹明章,等. 酶抑制法快速检测农产品农药残留的研究与应用[J]. 现代农药, 2004, 3(2): 25-27, 42.
[9] Rissato S R, Galhiane M S, Knoll F R, et al. Supercritical fluid extraction for pesticide multiresidue analysis in honey: determination by gas chromatography with electron - capture and mass spectrometry detection[J]. Journal of Chromatography A, 2004, 1048(2): 153-159.
[10] Cesarino I, Moraes F C, Lanza M R, et al. Electrochemical detection of carbamate pesticides in fruit and vegetables with a biosensor based on acetylcholinesterase immobilised on a composite of polyaniline - carbon nanotubes[J]. Food Chemistry, 2012, 135(3): 873-879.
[11] 岳贤田,高桂枝. 蔬菜中农药残留检测方法研究进展[J]. 现代农业科技, 2008(12): 111-113, 115.
[12] Wu X H, Sun J, Wu B, et al. Qualitative nnalysis model of near infrared spectra of pork based on synergy interval partial least squares discriminant analysis[J]. Laser & Optoelectronics Progress, 2015, 52(4): 242-247.
[13] Gong Z Y, Li Y F, Liu Y D, et al. Study on influence of irradiation angle on detection of sugar content of apple by near infrared spectroscopy[J]. Laser & Optoelectronics Progress, 2016, 53(2): 235-240.
[14] 钟艳萍,钟振声,陈兰珍,等. 近红外光谱技术定性鉴别蜂蜜品种及真伪的研究[J]. 现代食品科技, 2010, 26(11): 1280-1282, 1233.
[15] Gao J F, Li X L, Zhu F L, et al. Application of hyperspectral imaging technology to discriminate different geographical origins of *Jatropha curcas* L. seeds [J]. Computers and Electronics in Agriculture, 2013, 99(6): 186-193.
[16] 温珍才,孙 通,耿 响,等. 可见/近红外联合 UVE - PLS - LDA 鉴别压榨和浸出山茶油[J]. 光谱学与光谱分析, 2013, 33(9): 2354-2358.
[17] 汤丹明,孙 斌,刘辉军. 近红外漫反射光谱鉴别鸡蛋种类[J]. 光谱实验室, 2012, 29(5): 2699-2702.
[18] 周 竹,尹建新,周素茵,等. 基于近红外光谱与连续投影算法的针叶材表面节子缺陷识别[J]. 激光与光电子学进展, 2017, 54(2): 305-313.
[19] 刘 飞,冯 雷,柴荣耀,等. 基于直接正交信号校正的水稻冠层叶瘟光谱诊断[J]. 光学学报, 2010, 30(2): 585-589.
[20] 罗 霞,洪添胜,罗 阔,等. 高光谱技术在无损检测火龙果可溶性固形物中的应用[J]. 激光与光电子学进展, 2015, 52(8): 309-317.