

# 柑橘叶片叶绿素含量拉曼光谱定量分析方法研究

刘燕德, 程梦杰, 郝勇, 张宇, 侯兆国

华东交通大学机电与车辆工程学院, 江西 南昌 330013

**摘要** 柑橘叶片叶绿素含量的准确检测对柑橘营养状况和生长态势具有极其重要的意义。研究了快速无损诊断柑橘叶片中叶绿素含量的方法, 以期拉曼光谱检测技术用于柑橘叶片叶绿素含量检测提供参考。采集不同冠层高度和不同地理分布的柑橘叶片 120 片, 拭去叶片表面的灰尘, 用去离子水对其清洗、晾干装入密封袋中并用标签分类标注。然后对柑橘叶片进行拉曼光谱采集, 参数设置如下: 分辨率为  $3\text{ cm}^{-1}$ , 积分时间为 15 s; 激光功率为 50 mW。分别采用 BaselineWavelet、迭代限制最小二乘 (IRLS) 和不对称最小二乘 (ALS) 三种算法对柑橘叶片的拉曼光谱背景进行扣除, 使用偏最小二乘 (PLS) 方法建立定量模型; 四种光谱预处理方法归一化 (Normalization), Savitzky-Golay 卷积平滑 (SG smoothing, SG 平滑)、多元散射校正 (MSC) 和 Savitzky-Golay 一阶导数 (SG 1st Der) 对扣除背景后的光谱进行进一步的优化处理。结果表明: 采用原始光谱、BaselineWavelet、IRLS、ALS 背景扣除处理后的光谱建立 PLS 模型, 模型的相关系数  $r$  分别为 0.858, 0.828, 0.885 和 0.862, 交互验证均方根误差 (RMSECV) 分别为 5.392, 5.870, 4.934 和 5.336, 最佳因子数分别为 8, 3, 8 和 8; IRLS 背景扣除处理后的 PLS 模型的 RMSECV 最小, 相关系数最高, 建模效果最好。分别采用 SG 平滑、归一化、MSC 和 SG 1st Der 预处理方法对 IRLS 背景扣除后光谱进行预处理并建立 PLS 模型, 结果表明: IRLS 光谱及其结合 SG 平滑、归一化、MSC 和 SG 1st Der 四种预处理方法的 PLS 模型的  $R$  分别为 0.885, 0.897, 0.852, 0.863 和 0.888, RMSECV 分别为 4.934, 4.715, 5.595, 5.182 和 4.962; 最佳因子数分别为 8, 8, 8, 8 和 5; IRLS-SG 平滑后 PLS 模型的 RMSECV 最小, 模型效果最优。对 IRLS-SG 平滑预处理后的 PLS 模型展开验证, 预测相关系数  $r$  为 0.844, 预测均方根误差 (RMSEP) 为 5.29, 预测精确度较高。采用拉曼光谱结合三种光谱背景扣除方法和四种预处理方法对柑橘叶片叶绿素含量进行定量分析表明: 采用 IRLS 背景扣除结合 SG 平滑预处理后的 PLS 模型最优, 建模集  $r$  为 0.897, RMSECV 为 4.715; 预测集  $r$  为 0.844, RMSEP 为 5.29, 预测精度较高。拉曼光谱结合背景扣除方法可以为柑橘叶片叶绿素含量的定量分析提供一种快速简便的分析方法。

**关键词** BaselineWavelet; 迭代限制最小二乘; 不对称最小二乘; SG 平滑; 拉曼光谱; 叶绿素; 柑橘叶片  
**中图分类号**: O657.3   **文献标识码**: A   **DOI**: 10.3964/j.issn.1000-0593(2019)06-1768-05

## 引言

叶绿素是植物用来进行光合作用的绿色色素, 叶绿素有造血、提供维生素、解毒、抗病等多种用途<sup>[1]</sup>。柑橘是我国重要的经济水果之一, 但柑橘果树病虫害会导致柑橘大量落果。柑橘叶片中叶绿素含量不仅能反应整颗植株的营养状况, 并且能反映出植株早期是否有病害、冻害、虫害等的发生, 因此叶片叶绿素含量是判断植株健康程度的重要指标<sup>[2]</sup>。目前测定叶绿素含量的方法主要有活体叶绿素仪法、

高效液相色谱法, 荧光分析法, 当下运用最多的是分光光度法<sup>[2-3]</sup>, 但此方法需要对叶片进行破坏。

拉曼光谱法是一种快速、无损的分析方法, 但是拉曼仪器在测量光谱中, 因为受到激光光强的改变, 暗电流噪声以及样品放置等因素的干扰, 获得的拉曼光谱通常存在不小的噪声<sup>[4]</sup>。所以若想取得较为理想的预测效果, 一定要对拉曼光谱展开充分的背景扣除。当下扣除荧光背景大体存在两个途径, 第一个途径为, 增强实验仪器的性能以及相应的操作条件, 但增强仪器性能需要支付的成本昂贵, 而调整实验条件的方法具有的通用性又低。第二个途径是, 运用化学计量

收稿日期: 2018-04-27, 修订日期: 2018-09-05

基金项目: 国家自然科学基金项目(31760344), 南方山地果园智能化管理技术与装备协同创新中心项目(赣教高字[2014]60号), 江西省教育厅科学技术研究项目(GJJ160517)资助

作者简介: 刘燕德, 女, 1967年生, 华东交通大学机电与车辆工程学院教授   e-mail: sduhys@163.com

学方法针对获取的光谱予以背景扣除,运用较多的方法有:导数法、不对称最小二乘(asymmetric least squares, ALS)、迭代限制最小二乘(iterative restricted least squares, IRLS)、BaselineWavelet 背景扣除算法、全自动背景扣除算法等<sup>[5]</sup>。目前在光谱的建模过程中,光谱的预处理也是不可缺少的,合适的光谱预处理方法,可以有效的降低光谱中的噪声信息,提高模型的稳健性<sup>[6]</sup>。

本文以柑橘叶片为研究对象,采用拉曼光谱技术对所采集的柑橘叶片叶绿素含量进行定量分析检测,对比分析三种不同的荧光背景扣除方法、应用简单偏最小二乘回归分析建立数学模型,探讨不同背景扣除算法拉曼光谱的处理效果;并对最佳预测模型进行不同方法预处理对比评价,为拉曼光谱的荧光背景、噪声扣除提供一种可靠的方法。

## 1 实验部分

### 1.1 仪器

光谱采集仪器为 SENTERRA 型激光共焦拉曼光谱仪,激发波长是 780 nm,高灵敏度 CCD 检测器,运用 OLYMPUS 光学显微镜,波数精度可以达到  $0.1 \text{ cm}^{-1}$ 。拉曼光谱采集范围为  $3\ 500 \sim 100 \text{ cm}^{-1}$ ,共 6 821 个数据点,采集的光谱转换为光谱仪自带 OPUS(verion. 6. 5; Bruker Optik GmbH, Germany)软件。

### 1.2 材料

实验所用柑橘果树叶片采于赣州市某果园,叶片共采集 120 片,分别采自不同地理位置的 24 棵果树,环绕每棵果树一周均匀采集 5 片无破损叶片。采摘后,拭去叶片表面的灰尘,用去离子水对其清洗、晾干装入密封袋中并用标签分类标注。对 120 个样本叶片按照 3:1 比例划分为建模集与预测集,叶绿素含量真实值最大和最小的样品划入建模集,保证建模集中叶绿素含量范围大于预测集,划为 91 个校正集样本,29 个预测集样本。

### 1.3 拉曼光谱采集

采集叶片光谱时,实验室环境温度控制在  $25 \text{ }^\circ\text{C}$  左右,相对湿度控制在 80% 以下,分辨率为  $3 \text{ cm}^{-1}$ ,积分时间为 15 s;激光功率为 50 mW。每个柑橘叶片样品避开叶脉位置,每片叶片随机采集三条拉曼光谱,并取三条光谱数据平均值用于后续数据分析,共采集 360 条光谱数据。因为原始光谱数据两端拥有许多无关紧要的信息,造成信噪比提升<sup>[7]</sup>,为了防止该无效信号干扰实验结果,运用光谱在  $1\ 700 \sim 180 \text{ cm}^{-1}$  之间的拉曼光谱展开接下来的分析,总计截取 3 041 个数据点,得到的光谱图如图 1 所示。

### 1.4 柑橘叶片拉曼光谱背景消除方法

BaselineWavelet 是基于连续小波变换和惩罚最小二乘的背景扣除算法,能够方便快速地解决拉曼光谱中荧光背景所带来的问题,该算法主要由确定峰的位置,确定峰的宽度以及背景拟合三部分组成<sup>[8]</sup>。其中峰位的确定又可以分为利用连续小波变换确定脊线,确定脊线处的峰值,确定参数估计峰三部分。利用 Haar 作为母函数求导的方法确定峰的宽度,最后结合惩罚最小二乘算法进行背景拟合。

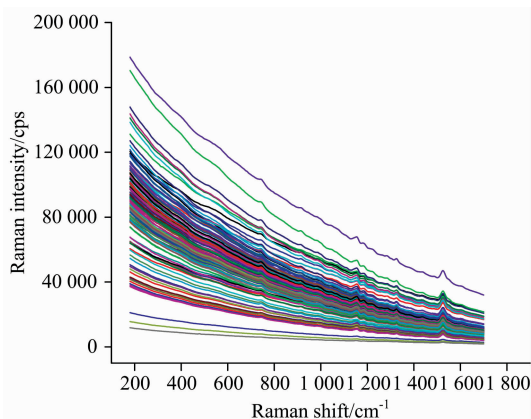


图 1 柑橘叶片原始光谱截取后光谱图

Fig. 1 Citrus leaves original spectra after interception spectra

迭代限制最小二乘(IRLS)算法是迭代算法从加权最小二乘的解析解中迭代重新加权以收敛到最优  $l_p$  近似解<sup>[9]</sup>,IRLS 用于解决目标函数的优化问题,实际上是用 2 范数来近似替代  $p$  范数<sup>[10]</sup>。

不对称最小二乘(ALS)是由惩罚最小二乘与不对称加权所结合成的一种背景扣除算法,此方法用于光谱背景扣除速度非常快,且能保留完整的有用信号,其原理是

$$S = \sum_i \omega_i (y_i - z_i)^2 + \lambda \sum_i (\Delta^2 z_i)^2 \quad (1)$$

其中  $w$  为权重, $\lambda$  为正则化参数, $y$  为光谱信号, $z$  为拟合出的背景, $\Delta$  为二阶差分算子,式中第一项用于衡量拟合函数  $z$  与原始数据  $y$  的非对称拟合程度;第二项是平滑程度的惩罚,系数  $\lambda$  起到平衡非对称逼近程度和光滑性的作用<sup>[11]</sup>。

### 1.5 模型的建立及评价

偏最小二乘法是一种多元因子回归方法,是拉曼光谱分析技术中比较成熟的建模方法,根据偏最小二乘(partial least squares, PLS)预测的公式

$$y = \sum_{i=1}^N \beta_i \lambda_i + b \quad (2)$$

式(2)中, $y$ ,  $N$ ,  $\beta$ ,  $\lambda$  和  $b$  依次表示模型的叶绿素预测含量值,参与建模的光谱变量数,能量谱强度,回归系数以及模型的截距。光谱变量同回归系数的加权求和,然后同截距相加,得到的便是 PLS 模型叶绿素含量的预测值。

采用相关系数( $r$ )、预测集均方根误差(root mean square error of prediction, RMSEP)、交互验证均方根误差(root mean square error of cross validation, RMSECV)评价模型的性能。模型建立过程中,相关系数  $r$  越接近 1,回归或预测的效果越好;RMSECV 越小,表明模型回归的越好;RMSEP 越小,表明所建模型的预测能力越强<sup>[12]</sup>。

## 2 结果与讨论

### 2.1 柑橘叶片拉曼光谱特征分析

选取一条截取后原始光谱进行特征分析,如图 2 所示:光谱线在  $1\ 155$  和  $1\ 527 \text{ cm}^{-1}$  处都存在有明显的特征峰。

1 155 和 1 527  $\text{cm}^{-1}$  处的特征峰是叶绿素外环中的 C—O—C 和 C—C 双键振动引起的, 是叶绿素 a 和叶绿素 b 的特征谱线。因此 1 155 和 1 527  $\text{cm}^{-1}$  两条特征谱线的相对强度可以作为柑橘叶片内叶绿素含量高低的判断指标。

## 2.2 柑橘叶片拉曼光谱的背景校正方法分析

将 OPUS 软件中的光谱数据文件转换为 R 语言可以读取的 CSV 文件, 利用 R 语言自带的 BaselineWavelet, IRLS 和 ALS 程序包对光谱数据进行背景扣除, 经三种不同算法扣除后与原始光谱图像进行比较, 如图 3 所示: (a), (b), (c), (d) 分别表示原始光谱, BaselineWavelet, IRLS 和 ALS 背景扣除后光谱。图 3(a) 可以看到荧光背景信号表现为一个典型的倾斜背景, 影响数据的后续处理。经过背景扣除后的图 3(b), (c) 和 (d), 倾斜的荧光背景明显削弱, 保留下有用的拉曼光谱信号。与图 3(a) 相比, 其他三个图形中拉曼位移 1 155 与 1 527  $\text{cm}^{-1}$  处的特征峰非常明显, 并且图 3(c) 比 (b) 和 (d) 的峰形更丰富。可以看出经过算法处理过的光谱数据

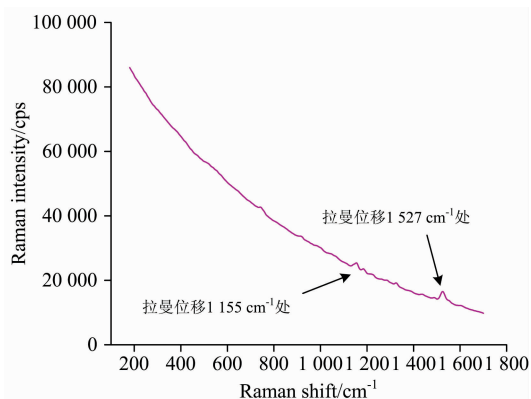


图 2 光谱特征

Fig. 2 Spectral characteristics

背景有不同程度的减少, 并且该方法能够在保持拉曼光谱的峰形, 即有用信息的情况下将背景扣除。

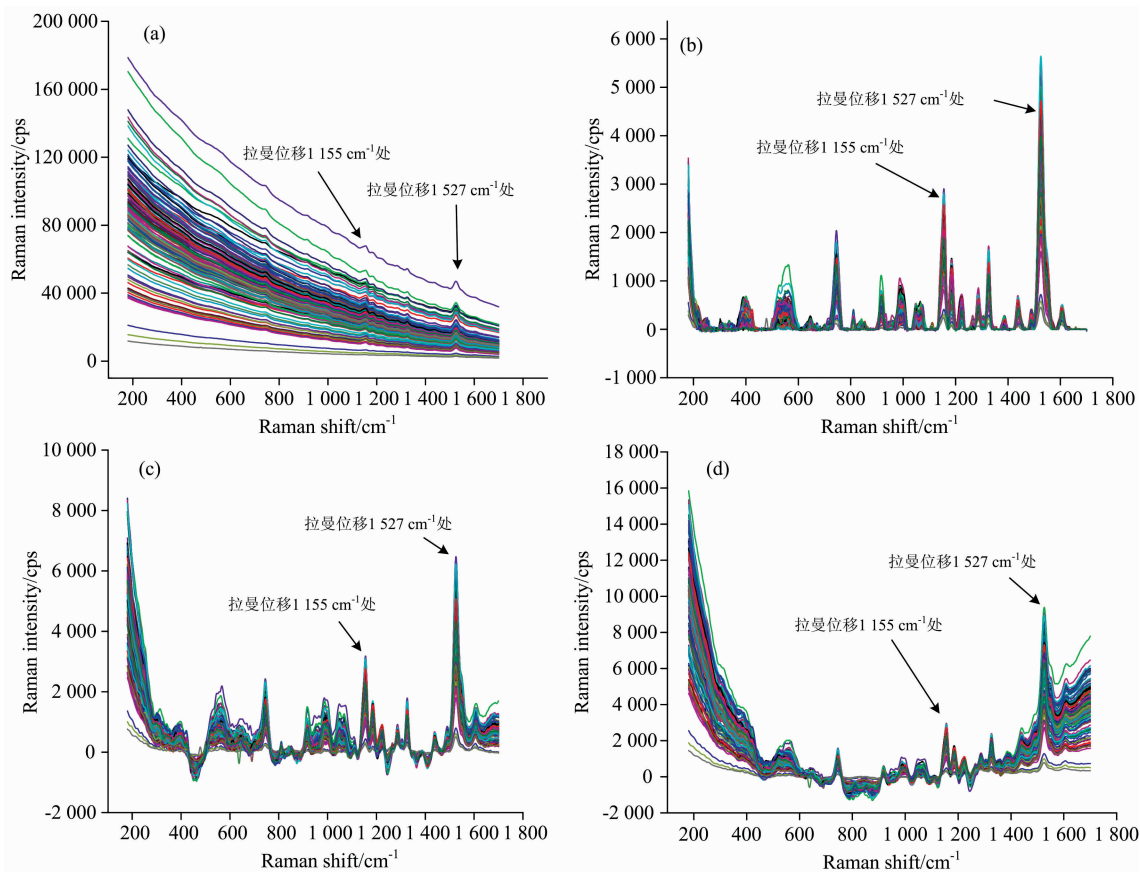


图 3 原始光谱与不同背景扣除后拉曼光谱图

(a): 原始光谱; (b): BaselineWavelet 背景扣除后光谱; (c): IRLS 背景扣除后光谱; (d): ALS 背景扣除后光谱

Fig. 3 Original spectra and different background correction Raman spectra

(a): Original spectrum; (b): BaselineWavelet background correction spectrum;

(c): IRLS background correction spectrum; (d): ALS background correction spectrum

对不同校正算法背景扣除后光谱数据建立数学模型, 采用交互验证法建立偏最小二乘 (PLS) 模型, 根据  $r$  和 RM-SECV 对模型的质量进行评价, 选取出最优的建模结果。不

同背景扣除方法的 PLS 建模结果显示在表 1 中。

可以看出, 三种背景扣除方法中, BaselineWavelet 与原始光谱建模结果相比  $r$  减小, RMSECV 增大, 且最佳因子数

过小, 因此建模结果变差; IRLS 和 ALS 两种方法相对原始光谱模型  $R$  增大, RMSECV 减小, 因此建模效果均有不同程度的提高。其中, IRLS 建模集的相关系数  $r=0.885$  比 ALS 的  $r=0.862$  更接近于 1, 交互验证均方根误差 RMSECV=4.934 比 ALS 的 RMSECV=5.336 更小, 因此 IRLS 建立的模型最优。综上比较, IRLS 背景扣除后的光谱数据所得建模效果最好。

表 1 不同背景扣除方法 PLS 建模结果  
Table 1 The PLS modeling results of different background correction methods

背景扣除方法	最佳因子数	相关系数( $r$ )	RMSECV
原始	8	0.858	5.392
BaselineWavelet	3	0.828	5.870
IRLS	8	0.885	4.934
ALS	8	0.862	5.336

### 2.3 不同光谱预处理方法对柑橘叶片叶绿素含量定量模型的影响

经过 IRLS 背景扣除后, 荧光背景的干扰大大减少, 但拉曼光谱数据不仅包括荧光背景, 还有噪声、杂散光等的干扰。对 IRLS 背景扣除后光谱数据进一步的进行 SG 平滑、归一化、MSC、一阶导数等不同方法的预处理, 光谱预处理能进一步消除无关信息和噪声的影响。建立全交互 PLS 数学模型, 不同光谱预处理方法的建模结果显示在表 2 中: 归一化、MSC 与原始光谱建模结果相比  $r$  有不同程度的减小, RMSECV 有不同程度的增大, 因此效果稍微差些。SG 平滑与一阶导数与原始光谱建模结果相比  $r$  有不同程度的增加, RMSECV 有不同程度的减小, 因此建模效果均有提高。其中 SG 平滑建模集的相关系数  $r$  为 0.897, 最接近与 1, RMSECV 为 4.715 最小, 并且最佳因子数为 8, 即不会因为太大出现过拟合现象, 也不会因为太小出现过拟合现象, 因此 SG 平滑预处理后建模效果最优。

表 2 预处理方法 PLS 建模结果  
Table 2 The PLS modeling results of different pretreatment method

预处理方法	最佳因子数	相关系数( $r$ )	RMSECV
原始	8	0.885	4.934
SG 平滑	8	0.897	4.715
归一化	8	0.852	5.595
MSC	8	0.863	5.182
SG 一阶导数	5	0.888	4.962

## References

- [1] XIAO Yi, ZHU Xin-guang(肖 怡, 朱新广). Plant Physiology Journal(植物生理学报), 2016, (11): 1663.
- [2] YUE Xue-jun, QUAN Dong-ping, HONG Tian-sheng(岳学军, 全东平, 洪添胜), 等. Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(1): 294.
- [3] Peng F, Liu S B, Xu H Y, et al. Advanced Materials Research, 2013, 726-731: 1411.
- [4] LI Xiang, LÜ Yong(李 响, 吕 勇). Journal of Beijing Information Science & Technology University(北京信息科技大学学报自然科学版), 2013, (2): 27.

### 2.4 柑橘叶片叶绿素含量 PLS 模型预测

在模型建立完成后, 仍然需要验证集对模型的准确性、重复性、稳健性等性能进行验证, 只有通过验证的模型方可使用。因为经过 IRLS 背景扣除结合 SG 平滑预处理后的建模结果最佳, 因此对其建立预测模型验证。PLS 模型预测散点图如图 4 所示, 预测模型  $r$  为 0.844, RMSEP 为 5.29, 预测准确性较高。

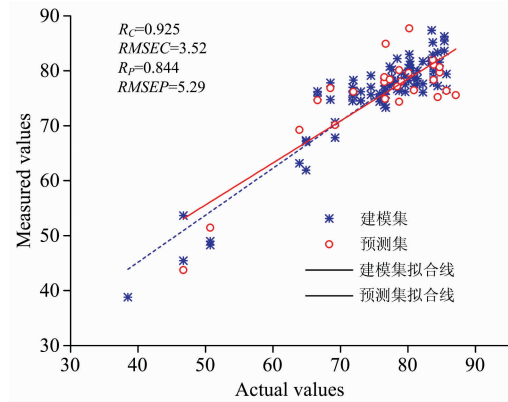


图 4 PLS 回归建模和预测散点图

Fig. 4 PLS regression modeling and prediction scatter plot

## 3 结论

拉曼光谱信号总是存在荧光背景、噪声等的影响, 对其进行背景扣除及预处理是必不可少的。采用拉曼光谱对柑橘叶片中的叶绿素含量进行快速定量分析, 利用 R 语言中 BaselineWavelet, IRLS 和 ALS 三种不同的背景扣除程序包对原始光谱进行背景扣除, 并建立全交互 PLS 模型。结果表明: BaselineWavelet 背景扣除与原始光谱二者建立模型相比效果变差, IRLS, ALS 背景扣除与原始光谱建立模型相比效果有了明显的提高, 其中 IRLS 算法扣除后的光谱建立的模型效果最优, 建模集相关系数  $r$  为 0.885, 交互验证均方根误差 RMSECV=4.934。进一步的对最优模型数据预处理, SG 平滑预处理效果最好, 建模集相关系数  $r$  为 0.897, 交互验证均方根误差 RMSECV=4.715。最后对 IRLS 背景扣除结合 SG 平滑预处理后模型进行验证, 预测集相关系数  $r$  为 0.844, 预测集均方根误差 RMSEP=5.29, 预测精度准确。在 IRLS 的背景扣除算法的基础上进一步进行 SG 平滑预处理使得拉曼光谱荧光背景、噪声信号得到有效的减少。研究表明拉曼光谱结合背景扣除方法可以为柑橘叶片叶绿素含量的定量分析提供一种快速简便的分析方法。

- [ 5 ] FANG Cheng-xuan, LI Jian-hua, LIANG Yi-zeng(房承宣, 李建华, 梁逸曾). Journal of Instrumental Analysis(分析测试学报), 2012, 31(5): 541.
- [ 6 ] Bian X, Chen D, Cai W, et al. Chinese Journal of Chemistry, 2011, 29(11): 2525.
- [ 7 ] HAO Yong, CHEN Bin(郝 勇, 陈 斌). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2015, 35(9): 2563.
- [ 8 ] CHEN Shan(陈 珊). Central South University(中南大学), 2011.
- [ 9 ] Lee K, Kwon H, You K. TDOA International Conference on Intelligent Informatics and Biomedical Sciences. IEEE, 2017: 92.
- [10] HUANG Shi-hao, PAN Tie-zhu(黄时浩, 潘铁柱). Science and Technology and Innovation(科技与创新), 2016, (16): 81.
- [11] JIANG An, PENG Jiang-tao, XIE Qi-wei, et al(姜 安, 彭江涛, 谢启伟, 等). Computers and Applied Chemistry(计算机与应用化学), 2012, 29(5): 537.
- [12] LIU Yan-de, XIE Qing-hua, WANG Hai-yang, et al(刘燕德, 谢庆华, 王海阳, 等). Chinese Journal of Luminescence(发光学报), 2016, 37(9): 1135.

## Quantitative Analysis of Chlorophyll Content in Citrus Leaves by Raman Spectroscopy

LIU Yan-de, CHENG Meng-jie, HAO Yong, ZHANG Yu, HOU Zhao-guo

School of Mechatronics Engineering, East China Jiaotong University, Nanchang 330013, China

**Abstract** The accurate detection of the content of chlorophyll in citrus leaves is of great significance to the nutritional status and the growth trend of citrus. A rapid and non-destructive method for diagnosing chlorophyll content in citrus leaves was studied in order to provide a reference for the detection of chlorophyll content in citrus leaves by Raman spectroscopy. A hundred and twenty slices of citrus leaves with different canopy heights and different geographical distributions were collected. The dust on the surface of the leaves was wiped off. The deionized water was used in the laboratory to clean it, dried in a sealed bag, and labeled with a label. The Raman spectra of citrus leaves were then collected. The parameters were set as follows: resolution  $3\text{ cm}^{-1}$ , integration time 15 s; laser power 50 mW. Three methods were used, such as baseline wavelet, iterative restricted least squares (IRLS) and asymmetric least squares (ALS), for background correction of Raman spectroscopy. After that, Partial least squares (PLS) method was used to establish the quantitative model. Subsequently, four methods of spectral pretreatment, like Savitzky-Golay convolution smoothing (SG smoothing), normalization, multiplicative scatter correction (MSC) and the Savitzky-Golay 1st derivative, were used to further optimize the spectra which had been treated by the background correction. The research process showed that the PLS model was established by the spectra of the original spectrum, Baseline Wavelet, IRLS, and ALS preprocessing. The correlation coefficients of the models were 0.858, 0.828, 0.885, and 0.862, respectively. The root mean square error cross validation, RMSECV were 5.392, 5.870, 4.934, and 5.336, respectively. The best principal component factors were 8, 3, 8 and 8 respectively. The RMSECV of the pre-processed PLS model deducted from the IRLS background was the smallest, the correlation coefficient was the highest, and the modeling effect was the best. SG smoothing, normalization, MSC and SG 1st Der preprocessing methods were used to preprocess IRLS background correction spectrum and establish PLS model. The results showed that: IRLS spectrum and its combination of SG smoothing, normalization, MSC and SG 1st Der The PLS of the four pretreatment methods of  $r$  were 0.885, 0.897, 0.852, 0.863, and 0.888, respectively. The RMSECV were 4.934, 4.715, 5.595, 5.182, and 4.962, respectively. The best principal component factors were 8, 8, 8, 8 and 5, respectively; the RMSECV of the PLS model after IRLS-SG smoothing was the smallest, and the model had the best effect. After verifying the PLS model preprocessed by IRLS-SG, the predictive correlation coefficient  $r$  of the prediction set was 0.844, the root mean square error of prediction (RMSEP) was 5.29, and the prediction accuracy was high. Three kinds of background correction methods combined with four kinds of spectral pretreatment methods were used to quantitatively model the Raman spectra of citrus leaves. It can be concluded that the experimental results after IRLS background correction combined with the SG smoothing are optimal. The modeling set  $r$  is 0.897, the RMSECV is 4.715, the prediction set  $r$  is 0.844, and the RMSEP is 5.29, and the prediction accuracy is high. Studies have shown that Raman spectroscopy combined with background correction methods can provide a quick and easy analytical method for quantitative analysis of chlorophyll content in citrus leaves.

**Keywords** BaselineWavelet; Iterative restricted least squares; Asymmetric least squares; SG smoothing; Raman spectrum; Chlorophyll; Citrus leaves

(Received Apr. 27, 2018; accepted Sep. 5, 2018)