

## 基于 SERS 与 PCA-SLR 实现乙基对氧磷定量检测

翁士状<sup>1,2</sup>, 陈 晟<sup>1,2</sup>, 曾新华<sup>2</sup>, 李 盼<sup>1,2</sup>,  
郑守国<sup>2</sup>, 尤聚军<sup>1,2</sup>, 李 森<sup>2\*</sup>, 朱泽德<sup>1,2</sup>

1. 中国科学技术大学, 安徽 合肥 230026  
2. 中国科学院合肥智能机械研究所, 安徽 合肥 230031

**摘 要** 利用表面增强拉曼光谱(SERS), 结合主成分分析(PCA)与分段线性回归(SLR)算法实现乙基对氧磷的定量检测。首先采集  $820\sim 1\ 630\text{ cm}^{-1}$  乙基对氧磷溶液 SERS, 并对  $820\sim 1\ 630\text{ cm}^{-1}$  (全范围)与  $845\sim 875\text{ cm}^{-1}$  (特征范围)光谱分别进行标准正态变换(SNV)、多元散射校正(MSC)、一阶导数绝对值、二阶导数等预处理; 然后经 PCA 降维后利用 SLR 建立乙基对氧磷溶液浓度预测模型。通过对比不同模型的预测准确度, 发现特征范围光谱采用 MSC 预处理后所建立的模型为最优, 总体预测均方误差值(RMSEP)为 0.33, 满足乙基对氧磷定量检测的需要。

**关键词** 乙基对氧磷; 表面增强拉曼光谱; 主成分分析; 分段线性回归

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2013)09-2438-05

### 引 言

有机磷农药, 因其合成简单, 广谱高效, 被广泛应用于农作物病虫害的防治, 但其在作物、土壤、水体中的残留对人体产生较大的危害。研究表明, 有机磷农药与胆碱酯酶结合, 形成磷酸化胆碱酯酶, 使胆碱酯酶失去催化乙酰胆碱水解的作用, 进而可以引起迟发反应, 导致迟发性周围神经疾病。因此进行有机磷农药检测具有重要意义。

目前有机磷农药主要检测方法为色谱法, 如气相色谱法、液相色谱法、超临界流体色谱法、薄层色谱法与色谱-质谱联用法等<sup>[1-3]</sup>。上述方法检测精度高, 但需大型仪器与繁杂的前处理过程, 导致检测成本高、时间长。而使用光谱来检测农药具有前处理简单, 检测费用低, 单次检测时间短等优点, 能满足当前农药检测中快速简便的应用需求。拉曼光谱作为一种分子振动光谱, 能反映分子内部各种简正振动频率及振动能级的情况, 从而可以用来鉴定分子中存在的官能团<sup>[4]</sup>, 被广泛地用于物质的定性分析与定量分析。El-Abassy 利用拉曼光谱与 PLR 定量检测液态牛奶中的脂肪含量<sup>[5]</sup>。Elena 将多种数据处理方法应用到橄榄拉曼光谱分析中, 实现对其品质的鉴定<sup>[6]</sup>。普通散射拉曼光谱检测限高, 不能满足有机磷农药的国家检测标准, 而 SERS 信号强度较普通拉曼散

射光谱可提高  $10^3\sim 10^5$  倍, 达到检测标准。Zhai 利用 SERS 实现对猪尿液中的莱克多巴胺(瘦肉精)的定量分析, 取得  $0.4\ \mu\text{g}\cdot\text{mL}^{-1}$  检测限<sup>[7]</sup>。关于 SERS 结合多元数据处理方法检测乙基对氧磷, 相关工作未见报道。本研究提出一种基于 SERS 与 PCA-SLR 的乙基对氧磷定量检测方法。

实验先分别获取  $2\times 10^{-4}$ ,  $2\times 10^{-5}$ ,  $2\times 10^{-6}\text{ mol}\cdot\text{L}^{-1}$  乙基对氧磷溶液的 SERS, 然后选取全范围与特征范围 SERS, 并对其分别进行 SNV、MSC、一阶导数绝对值与二阶导数预处理, 最后取累积贡献率为 95% 时 PCA 的主成分得分经 SLR 建立浓度预测模型。文中分析不同范围光谱、不同预处理方法对模型预测性能的影响, 以获得建立最优模型的方法。同时通过讨论主成分数目变化对预测模型准确度的影响, 论证主成分得分选取方法的合理性。

### 1 实验部分

#### 1.1 材料与设备

硝酸银、柠檬酸钠、乙基对氧磷纯品(99%)、共聚焦拉曼光谱仪(LabRAM HR800, Horiba Jobin Yvon)、球形冷凝管、三口烧瓶、加热套、搅拌台、搅拌子、硅片、载玻片。

乙基对氧磷纯品加入双馏水, 配制成  $2\times 10^{-4}$ ,  $2\times 10^{-5}$ ,  $2\times 10^{-6}\text{ mol}\cdot\text{L}^{-1}$  三种浓度溶液, 备用。

收稿日期: 2013-01-15, 修订日期: 2013-03-12

基金项目: 国家高技术研究发展计划(“863 计划”)项目(SS2013AA100302)资助

作者简介: 翁士状, 1989 年生, 中国科学技术大学博士研究生 e-mail: weng1989@mail.ustc.edu.cn

\* 通讯联系人 e-mail: mli@iim.ac.cn

## 1.2 表面增强拉曼光谱的测量

采用银纳米颗粒作为拉曼光谱的增强基底<sup>[8]</sup>。取 100 mL 的 1 mmol 硝酸银溶液剧烈搅拌并加热至沸腾, 迅速加入 4 mL 的 1% 柠檬酸钠, 溶液由无色变为墨绿色, 保持沸腾 30 分钟停止加热, 自然冷却至室温, 离心浓缩成溶胶状, 银纳米颗粒即制备完成。

将银纳米颗粒溶胶滴在硅片上, 待干后即可滴加乙基对氧磷溶液, 将其置于拉曼光谱仪下, 在 532 nm 光激发条件下采集光谱。光谱采集的波数范围从 820~1 630  $\text{cm}^{-1}$ , 分辨率为 1  $\text{cm}^{-1}$ , 扫描时间为 1 s。每采集 20 条光谱平均后作为实验中的 1 条原始光谱。不同浓度样本一次采集 10 条原始光谱, 每隔 5 天采样一次, 每次测量时需要重新配制溶液, 共采集 50 条光谱。

## 1.3 光谱预处理方法

SERS 测量过程中易受基底的平整度、宇宙射线及样本差异等因素的影响, 会出现较强的基线偏移与荧光背景, 给后续分析带来困难, 因此需要先对光谱进行预处理。本文采用 SNV, MSC 及求导法(一阶导数绝对值及二阶导数)进行预处理。其中, SNV 和 MSC 可消除固体颗粒大小、表面散射以及光程变化引起的基线漂移及荧光背景<sup>[9, 10]</sup>, 求导法对原始光谱求导后只保留光谱变化信息可有效消除基线偏移<sup>[11]</sup>。

## 1.4 PCA-SLR

文中 SERS 数据维数大, 若直接进行线性回归, 存在运算复杂度大、模型建立时间长等缺点, 须先进行数据降维, 本文采用 PCA 降维。PCA 是一种以特征量为基础的多元统计方法, 可有效地简化数据。它通过对协方差矩阵进行特征分解, 将数据矩阵转换成一组线性不相关的变量, 以得出数据的主成分和对应的得分。通过保留对数据集的方差较大的分量, 删除对数据集方差较小的分量, 在减少数据维数的同时, 尽可能保持数据集的对方差贡献最大的特征, 从而保留住数据的主要信息。

不同浓度溶液 SERS 之间并非线性关系, 本文使用 SLR 构建模型。SLR 是一种对不具有线性关系的数据集进行线性回归的方法。通过引入一或多个折点, 将数据集分割成具有线性关系的多部分, 并分别建立独立的线性回归模型<sup>[12, 13]</sup>。因此, SLR 可理解为对非线性关系的线性分割。

实验中, 每个浓度溶液的 SERS 共采集 50 条, 每次随机抽取 20 条建立预测模型, 其他 30 条用于模型测试, 以上过程重复进行十次。模型的性能用 RMSEP 评价, 计算公式如下

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (y_r - y_p)^2}{n}}$$

其中,  $y_r$  为测试样本真实浓度值,  $y_p$  为测试样本预测浓度,  $n$  为测试样本数, 实验结果中给出的 RMSEP 为 10 次的平均值。

所有的数据处理在实验室电脑上完成, MS Windows XP, 双核 i3 处理器, 3.5G 内存。SNV、MSC、一阶导数绝对值、二阶导数及 PCA-SLR 算法在 Matlab(Matlab 7.1, USA)

上实现。

## 2 结果与讨论

### 2.1 特征范围光谱的选择

全范围 SERS 中许多峰位是其他物质产生, 会干扰到乙基对氧磷溶液的定量分析, 因此实验中利用拉曼光谱的指纹特性, 选取特征峰位附近的光谱作为特征范围光谱。本文以粉末状乙基对氧磷的特征峰作为参考峰, 来确定乙基对氧磷溶液 SERS 的特征峰。图 1 给出的是粉末状乙基对氧磷的拉曼光谱和不同浓度溶液 SERS 的对比, 其中实竖线标注的是粉末状乙基对氧磷拉曼光谱的峰位, 虚竖线为乙基对氧磷溶液 SERS 的峰位, 波数 859  $\text{cm}^{-1}$  上所有样本峰位重合和 1 592  $\text{cm}^{-1}$  周围溶液 SERS 特征峰未出现, 这两处采用实线标注。由图 1 可知粉末状乙基对氧磷光谱峰位集中在 859, 1 110, 1 348, 1 592  $\text{cm}^{-1}$  四个波数上, 由此确定溶液 SERS 特征峰位处于 859, 1 110, 1 348  $\text{cm}^{-1}$  波数上, 但由于 1 110 和 1 348  $\text{cm}^{-1}$  位置上谱图存在重叠交叉, 所以实验选取 859  $\text{cm}^{-1}$  附近光谱(845~875  $\text{cm}^{-1}$ )作为特征范围光谱。

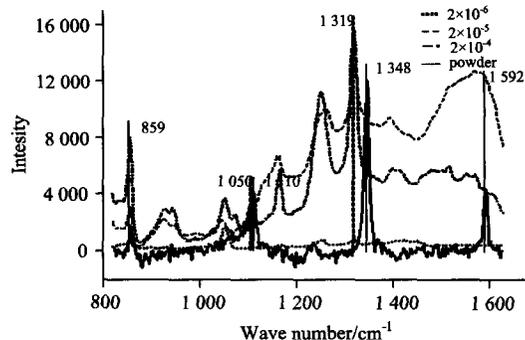


Fig. 1 Comparison of peak of different SERS

### 2.2 光谱预处理

拉曼光谱实现物质定量检测是基于光谱强度与溶液浓度成特定对应关系。SERS 测量中, 易受到基底的平整度、宇宙射线及样本差异等因素影响, 会出现较强的基线偏移与荧光背景。因此在分析之前, 分别采用 SNV, MSC, 求导法(一阶导数绝对值、二阶导数)对 SERS 进行预处理(图 2)。其中, (a)~(e) 为全范围光谱, (f)~(j) 为特征范围光谱, (a) 和 (f) 为原始光谱, (b) 和 (g) 为一阶导数绝对值后的光谱, (c) 和 (h) 为二阶导数光谱, (d) 和 (i) 为 MSC 处理后的光谱, (e) 和 (j) 为 SNV 处理后的光谱。

针对全范围光谱[图 2(a)~(e)]: 一阶导数绝对值能有效去除基线漂移; 二阶导数处理后光谱中重叠交叉现象增多; SNV 处理后  $2 \times 10^{-6}$  溶液光谱在 1 050  $\text{cm}^{-1}$  位置上出现一个峰位, 不同浓度光谱间间隔较原始光谱减小。因此, 两种方法会导致部分预测结果恶化。还应注意(a)~(e)中虚线和线点谱图总有重叠交叉, 这将干扰到模型对乙基对氧磷溶液相关浓度的预测, 因此全范围光谱所建立模型的预测准确度将会受到较大限制。针对特征范围光谱[图 1(f)~(j)], 不同方法预处理对其的效果与全范围光谱相近, 但原始光谱与

一阶导数绝对值、MSC 处理后的光谱无重叠交叉现象出现, 则其对应的模型预测性能将会有较大提升。

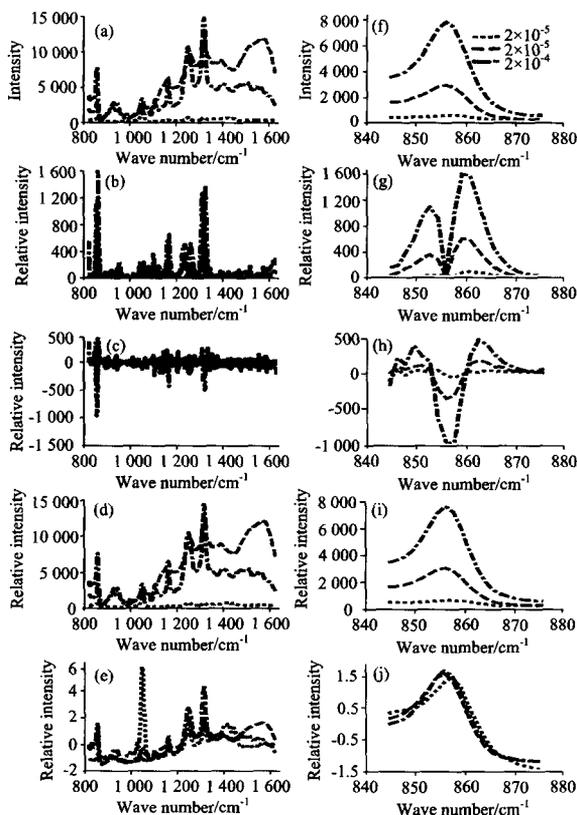


Fig. 2 Effect of different preprocessing methods

### 2.3 预测模型建立

不同光谱在累积贡献率为 95% 下, PCA 的主成分数目如表 1, 以此作为数据降维的标准。表 1 主成分分析中主成分数目(累积贡献率为: 95%)。

Table 1 Number of principal components in PCA (Accumulating contribution rate: 95%)

预处理方法	全范围光谱	特征范围光谱
原始光谱	2	1
一阶导数绝对值	3	1
二阶导数	75	3
MSC	2	1
SNV	3	2

实验中, 先直接对降维后的三种浓度乙基对氧磷溶液 SERS 进行线性回归建立模型, 但预测结果很差, 为分析原因, 图 3 给出了特征范围原始光谱经线性回归建立的模型预测值分布情况与其主成分得分散点图。由图 3(a)可知, 除了  $2 \times 10^{-4} \text{ mol} \cdot \text{L}^{-1}$  乙基对氧磷溶液预测值较准确外, 其他预测结果都出现了较大偏差,  $2 \times 10^{-6} \text{ mol} \cdot \text{L}^{-1}$  的预测值在  $4 \times 10^{-6} \text{ mol} \cdot \text{L}^{-1}$  附近,  $2 \times 10^{-5} \text{ mol} \cdot \text{L}^{-1}$  的预测值集中在  $5 \times 10^{-5} \text{ mol} \cdot \text{L}^{-1}$  附近, 但每个浓度溶液的预测值集中, 也说明 SERS 可用于定量分析。图 3(b)给出此类情况下主成分得

分散点图, 相同浓度溶液光谱聚集在空间的特定范围, 不同浓度样本间间距较大且无交叉, 需注意三类样本间关系并非线性, 但两两间却可建立线性关系, SLR 正是用来解决此类问题。

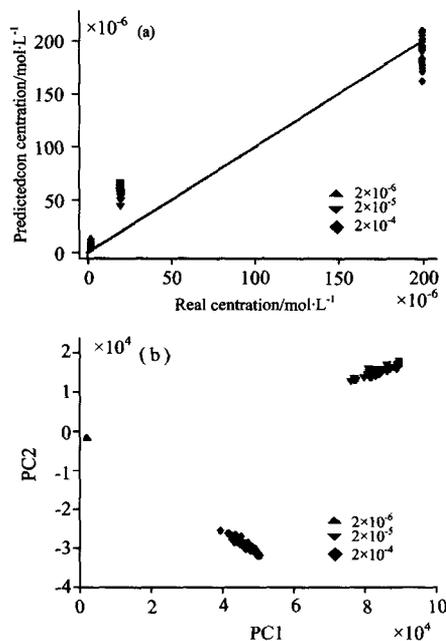


Fig. 3 Experimental result of raw SERS in characteristic range

利用 SLR 构建模型, 先建立  $2 \times 10^{-6}$  和  $2 \times 10^{-5} \text{ mol} \cdot \text{L}^{-1}$  乙基对氧磷溶液浓度预测模型, 再建立  $2 \times 10^{-5}$  和  $2 \times 10^{-4} \text{ mol} \cdot \text{L}^{-1}$  乙基对氧磷溶液浓度预测模型,  $2 \times 10^{-5} \text{ mol} \cdot \text{L}^{-1}$  溶液预测值取总体较优值。实验结果如表 2 所示, 表中每项包括  $2 \times 10^{-6}$ ,  $2 \times 10^{-5}$ ,  $2 \times 10^{-4} \text{ mol} \cdot \text{L}^{-1}$  乙基对氧磷溶液 RMSEP 与总体 RMSEP。由表 2 可知, 全范围光谱除了求导法外, 其他方法处理后光谱中  $2 \times 10^{-5}$  和  $2 \times 10^{-4} \text{ mol} \cdot \text{L}^{-1}$  溶液谱图的多处重叠交叉依然存在, 因此其建立的模型对  $2 \times 10^{-4} \text{ mol} \cdot \text{L}^{-1}$  溶液预测结果出现较大偏差; 二阶导数、SNV 后光谱强度与浓度值不合比例的情况增多, 则其对应模型对  $2 \times 10^{-5}$ ,  $2 \times 10^{-4} \text{ mol} \cdot \text{L}^{-1}$  溶液浓度预测准确度下降; 而一阶导数绝对值光谱除在小范围内出现重叠交叉外, 其他范围光谱强度均与浓度成比例, 故模型预测准确度整体提高。对于特征范围光谱, MSC 明显提升了模型预测准确度, 其他方法预处理效果与前者相近, 为了解 MSC 具体作用, 实验分析了特征范围原始光谱与经 MSC 处理后的光谱主成分得分散点图(图 4)。对比图 4 中(a)和(b), 可知经 MSC 处理后的同浓度光谱散点图聚集程度更高, 即消除了同浓度样本个体间的偏差, 进而提高预测准确度。

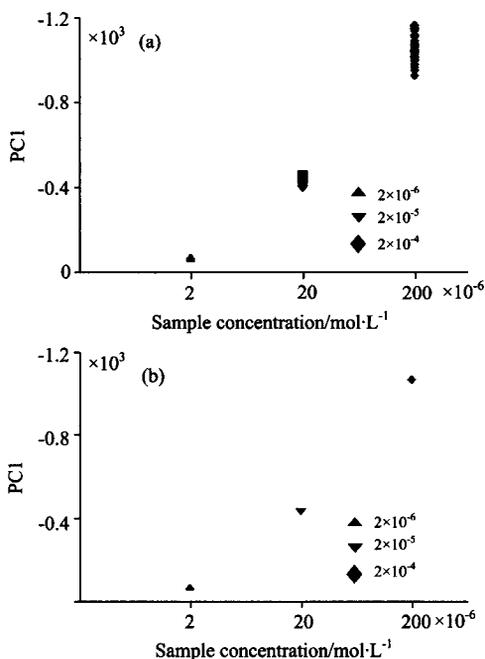
总之, 一阶导数绝对值与 MSC 预处理效果较好, 提升了对应模型预测性能; 使用特征范围 SERS 建立的模型预测准确度明显提高。由 MSC 处理后特征范围 SERS 所构建的模型取得预测结果的最好, 总体 RMSEP 为  $0.33 \times 10^{-6} \text{ mol} \cdot \text{L}^{-1}$ , 预测精度达乙基对氧磷国家检测标准, 满足其定量检测中的应用需求。实验中同时引入偏最小二乘法回归

(PLR)方法建立模型,实验结果表明经 MSC 处理的特征范围 SERS 所建立模型预测准确度最高(总体 RMSEP:  $0.33 \times$

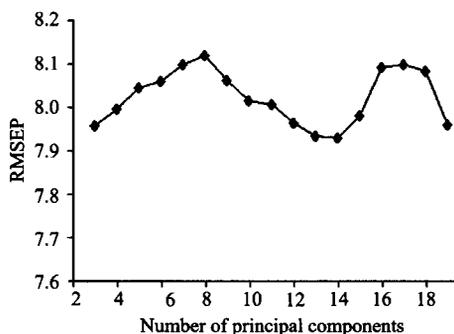
$10^{-6} \text{ mol} \cdot \text{L}^{-1}$ , 见表 2), 通过对比可知, 本文提出的方法较优。

**Table 2** Experimental result obtained by various range of spectra and various preprocessing methods( $10^{-6} \text{ mol} \cdot \text{L}^{-1}$ )

光谱范围	预处理方法	RMSEP( $2 \times 10^{-6}$ )	RMSEP( $2 \times 10^{-5}$ )	RMSEP( $2 \times 10^{-4}$ )	RMSEP(总体)
全范围	无预处理	0.22	1.04	176.47	100.61
	一阶导数绝对值	0.35	0.72	13.76	7.96
	二阶导数	7.21	16.01	26.79	20.79
	MSC	0.14	0.06	165.19	94.65
	SNV	8.77	5.99	105.61	62.68
特征范围	无预处理	1.02	1.46	16.33	8.64
	一阶导数绝对值	1.496	0.976	13.64	7.946
	二阶导数	5.709	2.349	17.84	11.57
	MSC	0.38	0.17	0.32	0.33
	SNV	8.52	5.80	25.15	19.07
特征范围(PLR)	MSC	0.41	1.56	2.31	1.63



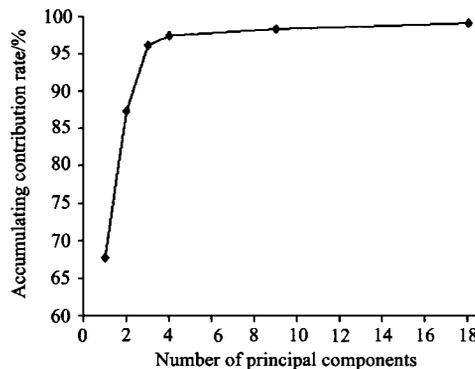
**Fig. 4** Scatter plot of principal component score of raw spectra and spectra preprocessed by MSC in characteristic range



**Fig. 5** Influence of variation of number of principal components on RMSEP

**2.4 主成分数目变化对模型预测准确度的影响**

数据降维中主成分得分取累积贡献率为 95% 时 PCA 的结果。为了论证主成分得分选取方法的合理性, 实验中分析主成分数目变化对回归模型预测准确度的影响(图 5, 图 6)。在图 5 中, 伴随着主成分数目增加, 回归模型总体 RMSEP 呈现无规则波动且波动值较小, 这是由于主成分数目为 3 时累积贡献率达到 95%(图 6), 但继续增加主成分数目引入的信息将是无关噪声, 对模型性能的提升无意义; 若主成分数目减小, 累积贡献率大幅降低(图 6), 信息量发生缺失, 其所建立的模型预测结果波动增大。由此可知, 降维中主成分得分的选取方法是合理的。



**Fig. 6** Accumulative contribution rate corresponding to the number of principal components

**2.5 方法的可靠性**

结合 SERS 与 PCA-SLR 建立模型并对未知乙基对氧磷溶模型进行预测, 此方法的可靠性取决于 SERS 数据的稳定性, 而其稳定性主要由基底与实验环境(如温湿度、光线等)决定, 实验中需保证合成基底的平整度与测量环境的一致性, 同时在测量光谱数据时, 选取多条光谱数据的平均值, 获得光谱整体趋势, 使得 SERS 波动能处于较小范围。总而言之, 作为一种新兴光谱技术, SERS 的稳定性方面存在一定的缺陷, 也是此技术现阶段实际应用中所面临的难题。但

由于 SERS 具有低检测限以及不受水体干扰等优点,且不断优化的增强基底与拉曼光谱仪器致使稳定性逐步提升,从而使其越来越广泛地应用于物质的定性定量分析。

### 3 结 论

利用 SERS 结合 PCA-SLR 成功实现乙基对氧磷溶液浓度的定量检测。结果表明一阶导数能去除 SERS 中的基线偏移,而 MSC 则能消除同类样本 SERS 间的差异;特征范围

(845~875  $\text{cm}^{-1}$ )SERS 经 MSC 预处理后所建立模型取得最好结果,总体 RMSEP 为  $0.33 \times 10^{-6} \text{ mol} \cdot \text{L}^{-1}$ ,可应用于乙基对氧磷的定量检测。实验还通过讨论主成分数目变化对模型预测准确度的影响,表明主成分得分取累积贡献率为 95% 时 PCA 结果较为合理。在下一步研究中,我们计划将此方法扩展到其他农药定量检测中,实现多组分农药同时定量检测。

**致谢:**感谢中国科学技术大学李盼硕士在拉曼光谱测量过程中给予的帮助。

### References

- [1] Domotorova M, Matisova E. J. Chromatogr. A, 2008, 1207(1-2): 1.
- [2] He Y, Lee H K. J. Chromatogr. A, 2006, 1122(1-2): 7.
- [3] Ferrer C, Gomez M J, Reyes J F, et al. J. Chromatogr. A, 2005, 1069(2): 183.
- [4] Martina S, Nick S, Jennifer S, et al. J. Raman Spectrosc., 2011, (42): 897.
- [5] El-Abassy R M, Eravuchira P J, Donfack P, et al. Vib. Spectrosc., 2011, (56): 3.
- [6] Elena G, Vincent B, Juan A, et al. Talanta, 2012, (93): 94.
- [7] Zhai F, Huang Y, Li C, et al. J. Agric. Food. Chem., 2011, (59): 10023.
- [8] Xiao X H, Yasutaka K, Tamitake I, et al. Anal. Chem., 2009, (81): 3350.
- [9] Kyriakos K, Doris E B, Ulrich J G. J. Pharmaceut Biomed., 2007, (43): 407.
- [10] Juan X, Veerle V L, Michael V, et al. Food Control, 2005, (16): 357.
- [11] Watari M, Du Y P, Ozaki Y. Appl. Spectrosc., 2007, (61): 397.
- [12] Lamberson W R, Firman J D. Poultry Sci., 2002, (81): 481.
- [13] Yu J R, Tzeng G H, Li H L. Fuzzy Set. Syst., 2001, (119): 247.

## Quantitative Detection of Ethyl Paraoxon Based on SERS and PCA-SLR

WENG Shi-zhuang<sup>1,2</sup>, CHEN Sheng<sup>1,2</sup>, ZENG Xin-hua<sup>2</sup>, LI Pan<sup>1,2</sup>, ZHENG Shou-guo<sup>2</sup>, YOU Ju-jun<sup>1,2</sup>, LI Miao<sup>2\*</sup>, ZHU Ze-de<sup>1,2</sup>

1. University of Science and Technology of China, Hefei 230026, China

2. Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China

**Abstract** In the present paper, the surface-enhanced Raman spectroscopy (SERS) was used to build the model for the quantitative detection of ethyl paraoxon by the principal component analysis and segmented linear regression (PCA-SLR). Firstly, SERS in 820~1 630  $\text{cm}^{-1}$  of ethyl paraoxon solution were measured and the spectra in 820~1 630  $\text{cm}^{-1}$  (complete range) and 845~875  $\text{cm}^{-1}$  (characteristic range) of ethyl paraoxon solution were preprocessed by standard normal transformation (SNV), multiplicative scatter correction (MSC), the absolute values of first derivative and the second derivative respectively. Additionally, the number of dimensions of the spectra was reduced by PCA. Finally, the models were established by SLR. It was found that the model developed with MSC preprocessed spectroscopy of characteristic range performed best (RMSEP: 0.33) by comparing the predictive accuracy of the different models. The result could meet with the needs in the quantitative detection of ethyl paraoxon.

**Keywords** Ethyl paraoxon; SERS; PCA; SLR

(Received Jan. 15, 2013; accepted Mar. 12, 2013)

\* Corresponding author