

Rubberband 方法中分段数量的选择对基线校正效果的影响分析

赵安新^{1,2}, 汤晓君², 张钟华^{2,3}, 刘君华²

- (1. 西安科技大学 网络中心, 陕西 西安 710054;
2. 西安交通大学 电力设备电气绝缘国家重点实验室, 陕西 西安 710049;
3. 中国计量科学研究院, 北京 100013)

摘要: 针对混合气体红外光谱在应用现场基线漂移和畸变、基线校正方法 Rubberband 方法中分段数量的选择不确定等问题进行探讨, 通过遍历的形式筛选出针对烃类气体在中红外波段、波数分辨率为 4 cm^{-1} 的情况下, 优选出分段数量在 50~120 之间、基线点的曲线拟合方法选择分段三次 Hermite 插值方法时, 其基线估计和校正的效果较为理想。同时针对 Rubberband 方法易将吸收点误判为基线点的问题, 使用中位数和偏离度首先剔除分段内的明显吸收点, 然后将剩余谱线点采用偏最小二乘法设定一定的阈值范围进行基线点的筛选, 筛选出的基线点在整个波段采用分段三次 Hermite 插值方法进行基线的估计。从实测的光谱基线校正效果来看, 提出的算法使用简单, 并适合于在线连续分析的应用场合。

关键词: 傅里叶变换红外光谱; 基线校正; 样条插值; 在线监测

中图分类号: O433.4 **文献标志码:** A **文章编号:** 1007-2276(2015)04-1172-06

Influence analysis of Rubberband's piecewise quantity on baseline correction effect

Zhao Anxin^{1,2}, Tang Xiaojun², Zhang Zhonghua^{2,3}, Liu Junhua²

- (1. Network Center, Xi'an University of Science and Technology, Xi'an 710054, China;
2. State Key Laboratory of Electrical Insulation And Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China;
3. National Institute of Metrology, Beijing 100013, China)

Abstract: The problems of baseline drifting and distortion of mixed gases infrared spectra in the application field and the uncertain choice of the Rubberband's piecewise quantity were discussed. In the situation of the spectrometer resolution of 4 cm^{-1} and mid-infrared band for the hydrocarbon gases, the baseline estimation and correction would be more ideal in the case of the piecewise quantity was set in the 50-120 and piecewise cubic Hermite interpolation method was selected to fit the baseline points by the ergodic method. For the problem of absorption points, which was mistaken as the baseline points, first of all, the obvious absorption points within the segment were eliminated by the method of the median and deviation. And then, the baseline points were selected by the partial least square method and certain

收稿日期: 2014-08-07; 修订日期: 2014-09-21

基金项目: 国家重大科学仪器设备开发专项(2012YQ240127); 国家自然科学基金(51277144);
电力设备电气绝缘国家重点实验室基金(EIPE11307)

作者简介: 赵安新(1981-), 男, 博士, 讲师, 主要从事多传感数据融合及信息处理方面的研究。Email: zhaoranxin@126.com

threshold was set in the residual baseline points. Lastly, the selected baseline points was used to estimate the baseline by piecewise cubic Hermite interpolation method in the whole band. From the baseline correction result of the measured spectra, the provided method was simple and fast, and also suitable for on-line continuous analysis applications.

Key words: Fourier transform infrared spectroscopy; baseline correction; spline interpolation; in-situ monitoring

0 引言

傅里叶变换红外光谱技术(FTIR)作为一种有效的定性和定量分析技术手段,被广泛应用于石油、化工、能源、环境、信息技术等各个领域^[1]。以往应用场合仅限于比较理想的实验室环境,随着野外探测及实时监测需求,目前该类仪器也逐步从理想的实验室环境部署到环境恶劣的工业应用场,比如石油录井、电力设备监测终端、矿藏勘探、大气监测等现场,然而由于光谱仪属于光学精密仪器,其涉及的光学部件较多,易受环境因素的影响,例如光源温度的变化、干涉仪的震动、测量物质的变换等因素^[2],从而会导致光谱出现漂移或者畸变,使分析结果出现较大偏差,严重制约了光谱技术在工业现场的应用。目前常用的基线校正方法可以主要分成三类技术:以 OPUS 软件的 Rubberband 方法为代表的分段处理方法^[3]、以经典微分法为代表的差分估计方法^[4]和以小波变换为代表的智能处理算法方法^[5]等。分段处理方法有 Rubberband 算法^[3,6]、分段极值辨别与估计^[7]、分段比线性校正^[2,8]极具代表性,这些算法基本都延续了基线校正过程中的分段处理的思路。微分方法由于其基线校正的效果不是很理想,而实际应用较少。智能化处理算法应用于基线校正主要有回归估计算法^[9]、小波变换^[10]、偏最小二乘以及基于偏最小二乘基础之上的改进算法^[11]、多元校正方法^[12]、移动平均^[13]等,这些智能算法的应用提高了基线校正的效果,同时由于牵涉一些复杂运算的和阈值的判定,使得这些算法善于在实验室手动或者半自动场合进行应用。在上述分段处理算法中由于其简单的运算原理使得在工业场合尤其实时处理领域很受欢迎,然而分段处理算法中如何分段以及分段多少合适一直是工程应用中困扰科研人员的问题,一般在使用分段处理算法时常常事先选定一定的分段数量进行

测试,找出合适的分段数量和数据拟合处理方法,然后才进行应用,如果遇到新的特征成分,一般需要再进行重复测试。文中针对在利用傅里叶变换红外光谱实时在线分析烃类气体成分的场合中如何选定 Rubberband 基线校正方法的分段数量进行探讨,并对在一定分段数量的情况下如何根据估计的基线点进行基线的估计方法进行了研究。

1 Rubberband 校正方法基本原理

Rubberband 方法是由 M. Pirzer 和 K. Sawatski 提出的,集成在 Bruker 公司的 Opus 软件包中供用户使用的一种基线校正方法。其基本实现过程可以分为如下步骤:(1) 首先,将整张光谱按照一定的分段数量将光谱等间隔分成相等的分段,分段数量由用户自主设定;(2) 其次,将每个分段内谱线的吸光度值或者透射率按照大小顺序进行排序,找出每分段内的极小值,将每个分段内的极小值认为是基线点,其中对于透射率谱线,每个分段内的极大值认为是基线点;(3) 最后,利用初步选定的基线点,在整个谱图区间利用数据拟合方法进行基线估计,利用估计的基线完成基线的校正。从 Rubberband 的基本原理可知:(1) 其整个谱图的分段数量的选择尤其关键,选择的好坏直接影响基线估计的效果,如果分段数量太少,可以很好地找出基线点但是在进行数据拟合时将很不平滑,同时也不准确,如果分段数量太多,将会把吸收点也误判为基线点,从而造成基线估计的误差;(2) 基线点的判别和利用这些基线点进行数据拟合方法也会在一定程度上影响基线估计效果。

2 实验材料和方法

2.1 实验材料及仪器

实验仪器采用德国 Bruker 公司的 alpha 型傅里叶变换红外光谱仪,采用氘代硫酸三肽 DTGS 检测

器,设定扫描范围为 $400\sim 4000\text{ cm}^{-1}$, 单次扫描次数为 16, 切趾函数选择 boxcar 矩形函数, 气压维持在 101.3 kPa (即一个标准大气压强)和 $25\text{ }^{\circ}\text{C}$ 温度下, 波数分辨率为 4 cm^{-1} , 谱线值为吸光度光谱, 每张光谱共有 2542 条谱线。

测试变压器油中溶解气体在线监测和油气探井常用的烃类气体成分 CH_4 、 C_2H_6 、 C_3H_8 、*iso*- C_4H_{10} 、*n*- C_4H_{10} 、*iso*- C_5H_{12} 和 *n*- C_5H_{12} 等。由于是对基线校正方法的试验, 因此试样样本气体成分只设定为单组分气体成分, 其配气的样本气体成分的浓度分别为 0.01% 、 0.02% 、 0.05% 、 0.1% 、 0.2% 、 0.5% 、 1% 。在进行连续多次样本扫描过程中, 发现部分基线发生了倾斜或者漂移。

2.2 基线估计的改进方法及评价指标

依据标准的 Rubberband 方法的基本原理, 文中采用基线点的均值、方差和标准差衡量不同分段数量对分段效果的影响。同时在基线估计过程中通过如下两点进行原理方法的改进:(1) 每个分段内不以极小值作为基线点的判别, 而是根据分段内的数据点先进行中位数和偏离度的计算, 首先剔除该分段内比较明显的吸收点, 再在剩余数据点内利用偏最小二乘设定阈值进行基线点的判别;(2) 对选定的基线点分别采用线性插值、三次样条插值、分段三次 Hermite 插值和最近邻点插值法等曲线拟合方法进行基线估计。

3 结果与分析

3.1 不同分段数量对目标气体傅里叶变换红外光谱基线估计的影响

此部分采用基线发生漂移的 1% 浓度的甲烷傅里叶变换红外光谱分析不同分段数量对基线估计的效果。其中甲烷单张光谱共有 2452 条谱线, 因此设定分段数量为 [8 13 18 23 28 33 38 43 48 53 58 63 68 73 78 83 88 93 98 103 108 113 118 123 128 133 138 142 148 153 158 163 168 173 178 183 188 193 198 203 208 213 228 223 228 233 243 248 253 258 268 273 278 283 303 308 313 318], 共 58 个分段值。在设定每个分段数量和线性插值基线点拟合方法的情况下, 基线估计的均值、方差和标准差如图 1、2、3 所示。从图 4 中 CH_4 的原始光谱可以看出, 其谱图发生了倾斜, 并向下漂移, 因此大部分基线位于零以下

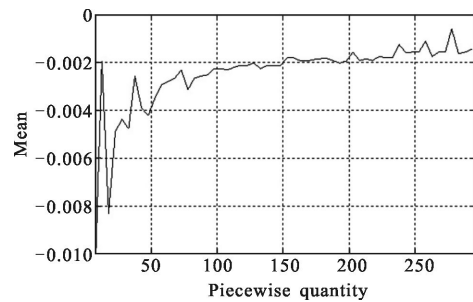


图 1 浓度 1% 的甲烷光谱基线估计的均值与分段数量的关系
Fig.1 Relation between mean of baseline estimation and piecewise quantity for $1\% \text{ CH}_4$

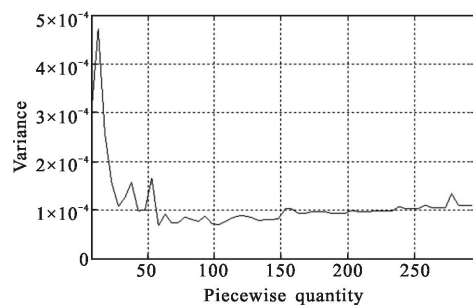


图 2 浓度 1% 的甲烷光谱基线估计的方差与分段数量的关系
Fig.2 Relation between variance of baseline estimation and piecewise quantity for $1\% \text{ CH}_4$

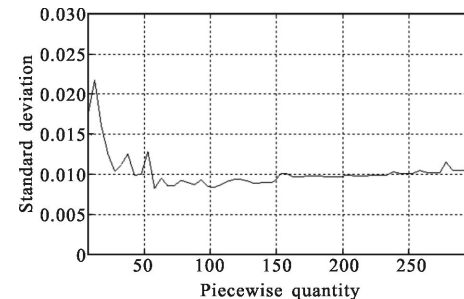
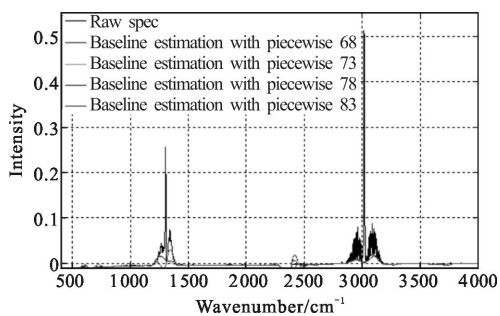


图 3 浓度 1% 的甲烷光谱估计基线的标准差与分段数量的关系
Fig.3 Relation between standard deviation of estimation baseline and piecewise quantity for $1\% \text{ CH}_4$

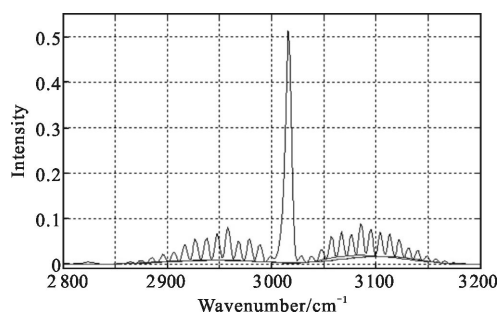
的区域, 理想的基线校正将是光谱中基线位置修正到零位置附近。图 1 是根据不同分段数量对光谱进行估计基线的均值, 可以看出: 随着分段数量的增加, 其估计基线的均值逐步增加, 这也初步验证了随着分段数量的增加, 其估计的基线越来越接近理想基线位置, 然而如果分段数量过多, 会把一部分吸收点误判为基线点, 因而造成估计的基线偏离理想位置, 如图中分段数量超过 150 的分段, 其基线估计存在此问题。图 2 和图 3 为不同分段数量情况下估计基线的方差和标准差与分段数量之间的变化关系,

在分段数量低于 50 的区域,其方差和标准差变化较大,从而说明估计的基线偏差较大,从图 2、图 3 可以看出:针对文中分析目标光谱,其在分段数量选择 50~150 区间时方差和标准差基本较稳定而且相对较小,结合图 1,因此在采用 Rubberband 进行基线校正时应选择分段数量在 50~120 之间进行变化其结果较为理想。图 4 和图 5 是在设定分段数量在 [68 73 78 83],其基线估计的效果和光谱基线校正的效果,可以看出:其估计的基线和基线校正效果相对都较为理想,从 2 800~3 200 cm^{-1} 区间主吸收峰区域的基线估计效果和校正效果都相对较好,从 1 600~2 000 cm^{-1} 区间非吸收区域的基线估计基本接近于理想基线位置,同时其经过校正后的光谱在此区间也分布在理想基线位置。因此,针对文中的中红外气体光谱在采用 Rubberband 基线校正方法时,可以设定的分段数量在 [68 73 78 83] 之间进行选择,建议可以优先选择分段数量为 73。从上述结果中可以得出:分段数量的选择与具体的分析目标没有特定的关系,算法会根据预先设定的选择范围和校正后估计的均值、方差和标准差等因素来选定分段数量,从而使得该方法可以应用在其他光谱基线校正的场合。



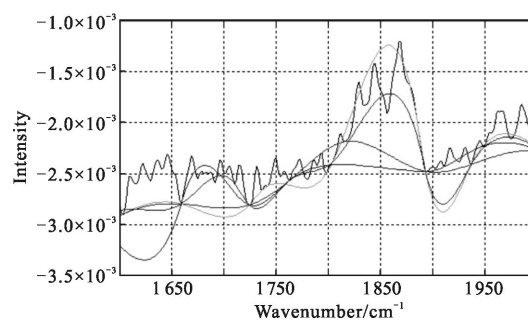
(a) 全波段

(a) Full wave band



(b) 主吸收波段

(b) Main absorption wave band



(c) 非吸收波段

(c) Second absorb wave band

图 4 浓度 1% 的甲烷原始光谱和在不同分段数量下估计的基线

Fig.4 Estimated baselines with different piecewise quantities and raw spectrum for 1% CH_4

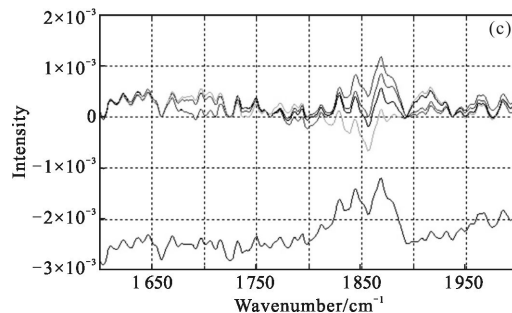
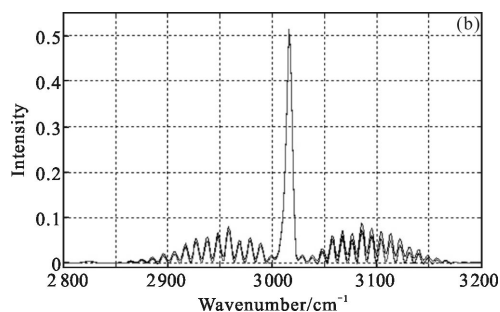
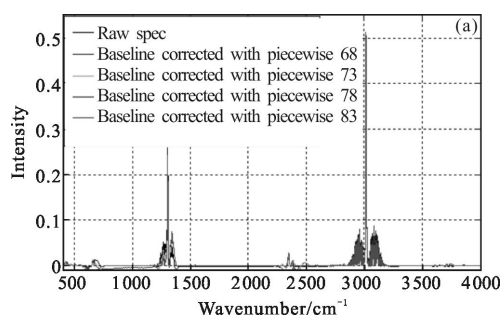


图 5 浓度 1% 的甲烷原始光谱和在不同的分段数量下估计基线校正光谱

Fig.5 Corrected spectrum with different piecewise quantities and raw spectrum for 1% CH_4

3.2 曲线拟合方法对基线估计效果的影响

目前常用的插值曲线拟合方法有线性插值、三次样条插值、分段三次 Hermite 插值和最近邻点插值法等曲线拟合方法,文中选用以上 4 种曲线拟合方法在 Rubberband 中分段数量为 73 的情况下对估计的基线点进行基线估计。其测试光谱选择上文中使用的 1%浓度的基线发生漂移的甲烷光谱,在不同数值插值方法下其估计基线的均值、方差和标准差如表 1 所示。从表 1 可以得出:在选用分段三次 Hermite 插值时,其进行曲线拟合的效果较为理想。同时从图 6 和图 7 其估计的基线和基线校正的效果也可以验证表 1 得出的结论。图 6 为在分段数量选择 73 的情况下,其四种曲线插值拟合方法对基线估计的结果,图 7 为根据估计的基线进行光谱基线校正的结果。从图 6 和图 7 看出:通过放大,在 $2800\sim 3200\text{ cm}^{-1}$ 区间主吸收峰区域的基线估计效果和校正效果都相对较好,其中线性插值和分段三次 Hermite 插值方法表现较优, $1600\sim 2000\text{ cm}^{-1}$ 区间非吸收区域的基线估计基本接近于理想基线位置,同时其经过校正后的光谱在此区间也分布在理想基线位置。因此,针对文中的中红外气体光谱在采用 Rubberband 基线校正方法时,在选定光谱分段数量的情况下,根据估计的基线点,在对这些点进行曲线拟合时采用分段三次 Hermite 插值方法时,其效果较为理想。选择何种曲线拟合方法对估计的基线点进行拟合,需要考虑 rubberband 校正过程中分段数量和分析的目标光谱,因此这里没有统一的方法,针对文中分析的目标,选择分段三次 Hermite 插值方法是适合的。

表 1 各种方法估计基线的相关均值、方差和标准差

Tab.1 Mean, variance, standard deviation of estimation baselines using different methods

Interpolation method	Mean	Variance($1.0\text{e}-04$)	Standard deviation
Linear	-0.003 2	0.678 6	0.008 2
Cubic spline	-0.002 8	0.916 6	0.009 6
Piecewise cubic Hermite	-0.003 1	0.696 6	0.008 3
Nearest neighbor	-0.003 2	0.788 2	0.008 9

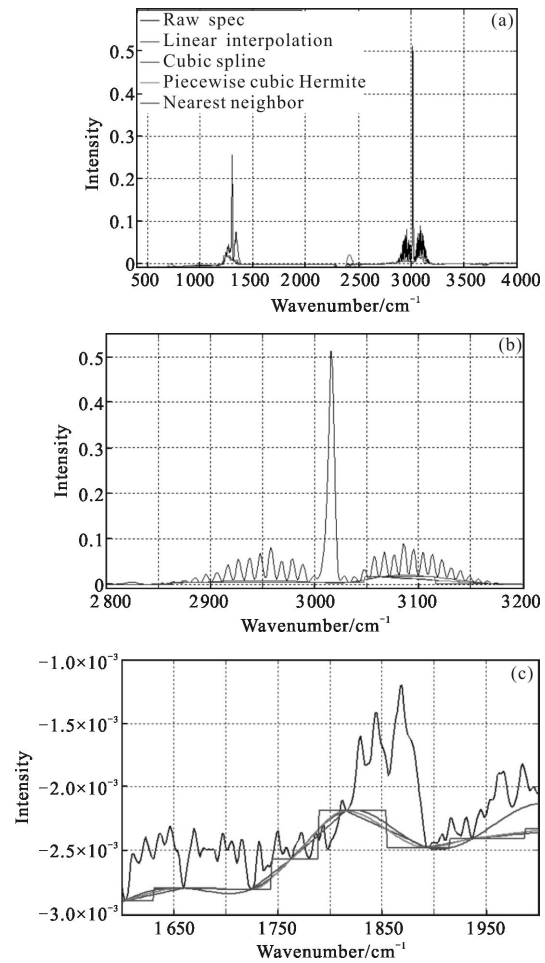
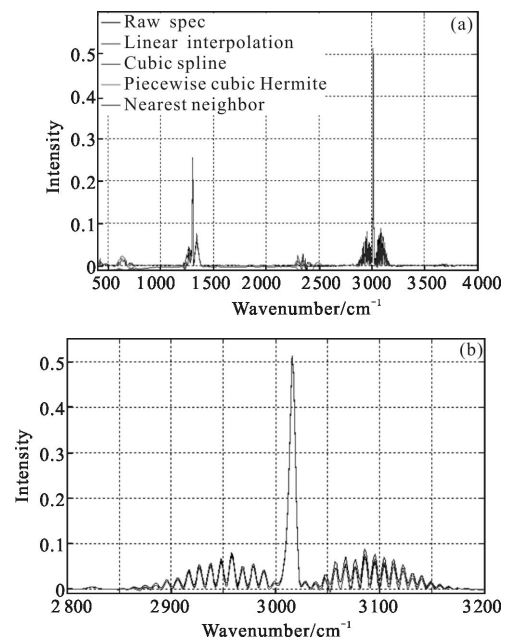


图 6 浓度 1% 的甲烷原始光谱和不同数据拟合方法拟合的基线

Fig.6 Estimation baselines with different interpolation methods and raw spectrum for 1% CH_4



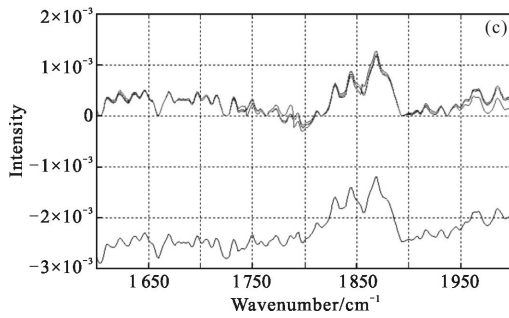


图7 浓度1%的甲烷原始光谱和不同基线校正方法校正的光谱
Fig.7 Corrected spectrum with different baseline correction methods and raw spectrum for 1% CH₄

4 结论

在使用 Rubberband 基线校正方法时, 需要根据分析的目标光谱, 选择不同的分段数量。文中根据在变压器油溶解气体和油气录井场合的在线监测分析, 其获取的主要是烃类气体成分, 且分析的光谱范围在中红外区域的特点, 结合仪器设定的波数分辨率, 探索了何种分段数量和基线点曲线拟合方法对估计基线效果最优两方面的问题, 经过实测光谱的测试分析, 给出了针对文中分析的目标可以选定分段数量为 50~150 之间, 曲线拟合方法可以选择分段三次 Hermite 插值。

参考文献:

- [1] Schmitt J, Flemming H C. FTIR-spectroscopy in microbial and material analysis [J]. *International Biodeterioration & Biodegradation*, 1998, 41(1): 1-11.
- [2] Tang Xiaojun, Wang Jin, Zhang Lei, et al. Spectral baseline correction by piecewise dividing in Fourier transform infrared gasanalysis [J]. *Spectroscopy and Spectral Analysis*, 2012, 33(2): 334-339. (in Chinese)
汤晓君, 王进, 张蕾, 等. 气体光谱分析应用中傅里叶变换红外光谱基线漂移分段比较校正方法 [J]. *光谱学与光谱分析*, 2013, 33(2): 334-339.
- [3] Rowlands C, Elliott S. Automated algorithm for baseline subtraction in spectra [J]. *Journal of Raman Spectroscopy*, 2011, 42(3): 363-369.
- [4] Krishnan S, Vogels J T W E, Coulier L, et al. Instrument and process independent binning and baseline correction methods for liquid chromatography-high resolution-mass spectrometry deconvolution [J]. *Analytica Chimica Acta*, 2012, 740: 12-19.
- [5] Hu Y, Jiang T, Shen A, et al. A background elimination method based on wavelet transform for Raman spectra [J]. *Chemometrics and Intelligent Laboratory Systems*, 2007, 85(1): 94-101.
- [6] Dietrich W, Rudel C H, Neumann M. Fast and precise automatic base-line correction of one-dimensional and 2-dimensional and 2-dimensional NMR-spectra [J]. *Journal of Magnetic Resonance*, 1991, 91(1): 1-11.
- [7] Weakley A T, Griffiths P R, Aston D E. Automatic baseline subtraction of vibrational spectra using minima identification and discrimination via adaptive, least-squares thresholding [J]. *Appl Spectrosc*, 2012, 66(5): 519-529.
- [8] Boucheham B, Ferdi Y, Batouche M C. Piecewise linear correction of ECG baseline wander: a curve simplification approach [J]. *Computer Methods and Programs in Biomedicine*, 2005, 78(1): 1-10.
- [9] Ruckstuhl A F, Jacobson M P, Field R W, et al. Baseline subtraction using robust local regression estimation [J]. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 2001, 68(2): 179-193.
- [10] Shao L, Griffiths P R. Automatic baseline correction by wavelet transform for quantitative open-path fourier transform infrared spectroscopy [J]. *Environmental Science & Technology*, 2007, 41(20): 7054-7059.
- [11] Peng J, Peng S, Jiang A, et al. Asymmetric least squares for multiple spectra baseline correction [J]. *Analytica Chimica Acta*, 2010, 683(1): 63-68.
- [12] Liland K H, Almoy T, Mevik B. Optimal choice of baseline correction for multivariate calibration of spectra [J]. *Applied Spectroscopy*, 2010, 64(9): 1007-1016.
- [13] Prakash B D, Wei Y C. A fully automated iterative moving averaging (AIMA) technique for baseline correction [J]. *Analyst*, 2011, 136(15): 3130-3135.