

DOI: 10.3969/j.issn.1673-6141.2019.04.005

环境大气 VOCs 色谱峰数据处理方法的设计与实现

刘永超^{1,2}, 陆钊¹, 刘伟^{1,2}, 胡荣春^{1,2}

(1 中国科学院安徽光学精密机械研究所中国科学院环境光学与技术重点实验室 安徽 合肥 230031;

2 中国科学技术大学, 安徽 合肥 230026

摘要: 针对大气挥发性有机化合物 (Volatile organic compounds, VOCs) 色谱谱图基线漂移严重、色谱峰峰形复杂的特点, 设计了一种色谱数据处理方法。采用移动平均法和小波变换法对高频噪声和基线漂移进行处理, 保证了谱峰解析的准确性。采用一、二阶导数结合法对色谱峰进行识别, 高斯曲线拟合法进行峰面积计算, 提高了峰面积计算精度。实验结果表明, 该算法运行速度快, 噪声去除和基线校正效果明显, 能够准确识别重叠峰、肩峰等不规则峰形, 峰面积计算精度高, 满足环境大气 VOCs 色谱数据处理的要求。

关键词: 挥发性有机物; 基线校正; 小波变换; 谱峰识别; 曲线拟合法

中图分类号: TP79

文献标识码: A

文章编号: 1673-6141(2019)04-0272-07

Design and Realization of VOCs Peak Data Processing Method in Atmospheric Environment

LIU Yongchao^{1,2}, LU Fan¹, LIU Wei^{1,2}, HU Rongchun^{1,2}

(1 Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China;

2 University of Science and Technology of China, Hefei 230026, China)

Abstract: A data processing method is designed to deal with the severe baseline drift and complex peak shape of the volatile organic compounds (VOCs) chromatography. Methods of moving average and wavelet transform are applied to deal with the high frequency noise and baseline drift, which ensure the accuracy of peak recognition and area calculation. The combination of first order and two order derivative method is used to identify the chromatographic peaks. The Gauss curve fitting method is used to calculate the peak area, which improves the calculation accuracy of peak area. The experimental results show that the algorithm is very fast and the effect of the denoising and baseline correction are obvious. The algorithm can accurately

基金项目: Supported by Young Scientists Fund of the National Natural Science Foundation of China(国家自然科学基金青年科学基金, 41605020), Major Research Plan of the National Natural Science Foundation of China(国家自然科学基金重大研究计划, 91544232)

作者简介: 刘永超 (1993 -), 女, 山东日照人, 研究生, 主要从事环境大气 VOCs 数据处理方法等方面的研究。

E-mail: ycliu@aiofm.ac.cn

导师简介: 陆钊 (1976 -), 男, 安徽合肥人, 博士, 副研究员, 主要从事颗粒物测量仪器研发等方面的研究。

E-mail: lufan@aiofm.ac.cn

收稿日期: 2018-03-29; **修改日期:** 2018-04-11

identify overlapping peaks, acromion peaks and can get high precision of peak area calculation, which are able to meet the requirements of VOCs data processing in ambient air.

Key words: volatile organic compounds; baseline correction; wavelet transform; peak recognition; curve fitting

1 引言

挥发性有机化合物 (Volatile organic compounds, VOCs) 是我国大气关键污染物臭氧和 PM_{2.5} 的重要前体物, 现行国家标准均采用气相色谱法对其进行检测^[1]。由于待测物质在流动相和固定相间亲和作用不同, 故各组分在色谱柱中的运行速度不同, 经过一定的柱长后, 流动相的各组分先后从色谱柱末端流出, 从而得到色谱图。不同物质在色谱图中通过不同形式的峰表现出来, 因此对色谱图中谱峰的处理直接影响物质信息的判定, 数据处理方法的好坏直接影响结果的精确度。

色谱数据处理主要包括基线处理和谱峰解析两个方面。基线处理主要包括高频噪声滤除和基线漂移校正。目前常用的基线校正方法主要有多项式拟合法、群峰基线法和小波变换法。其中多项式拟合法运算效率低^[2], 群峰基线法得到的是非平滑基线, 误差较大^[3], 而小波变换法具有比多项式拟合法计算快, 比群峰基线法误差小的特点, 被广泛应用于色谱基线处理中^[4]。常用的谱峰解析方法主要有时间窗法、小波变换法和导数法。其中, 时间窗法需要已知色谱峰的保留时间, 对于未知色谱峰不能识别^[5]; 小波变换法的小波基函数和分解层数难以确定, 而导数法具有计算简单、处理迅速、谱峰识别率高等优点^[6-7]。

针对环境大气 VOCs 色谱图中基线漂移严重、色谱峰峰形复杂的特点, 本文设计了一种 VOCs 色谱数据处理算法。采用移动平均法和小波变换法处理高频噪声和基线漂移, 采用导数法结合曲线拟合法进行谱峰解析。结果表明, 该方法处理迅速, 效果明显, 完全满足环境大气 VOCs 色谱数据处理的需求。

2 VOCs 色谱图数据处理过程

对于 VOCs 色谱数据的处理包括基线处理和谱峰解析两个过程。首先, 采用移动平均滤波和小波变换法, 有效去除高频噪声和基线漂移, 以提高谱峰解析的准确性。然后, 采用导数法进行谱峰识别, 高斯曲线拟合法进行峰面积计算。图 1 为数据处理过程中存在的问题及相应处理方法。

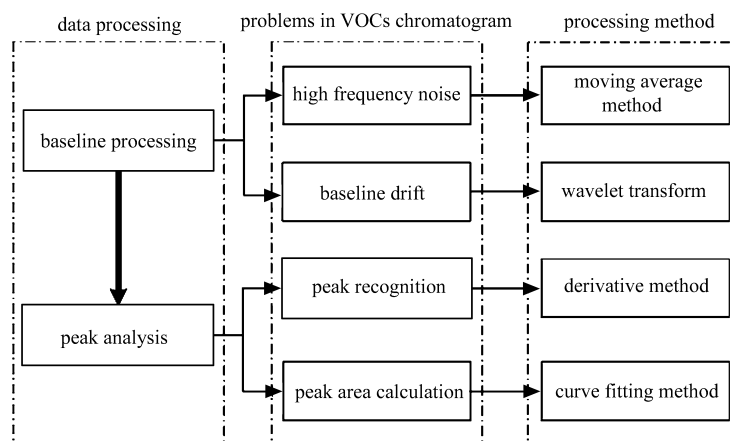


图 1 VOCs 色谱数据处理过程

Fig.1 Data processing of VOCs chromatography

2.1 基线处理

由于受到检测环境、样品质量以及色谱柱条件等诸多因素的影响, 色谱基线上会叠加不同程度的高频

噪声和低频漂移, 直接影响谱峰解析的准确性。在使用移动平均法滤除基线高频噪声后, 采用小波变换法校正基线低频漂移。

小波变换是一种信号的时间—频率分析方法, 具有多分辨率分析特性, 而且在时域和频域都具有表征信号局部特征的能力。在低频部分具有较低的时间分辨率和较高的频率分辨率, 在高频部分具有较高的时间分辨率和较低的频率分辨率, 适用于分析非平稳信号和提取信号的局部特征。

小波变换法校正基线漂移的原理是使用小波函数将原始信号分解, 分解过程如图 2 所示, 得到小波细节 cD (高频信号) 和小波逼近 cA (低频信号), 其中小波细节包含色谱峰信息, 小波逼近包含基线信息。将包含低频基线信息的小波逼近系数置零后重构, 得到仅含色谱峰信息的色谱信号, 达到去除基线的目的^[8]。

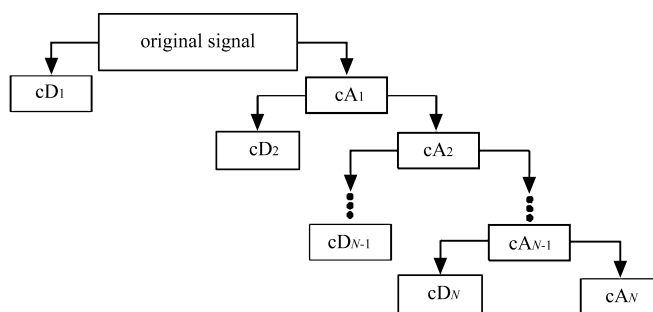


图 2 小波变换信号分解过程示意图

Fig.2 Schematic diagram of signal decomposition process in wavelet transform

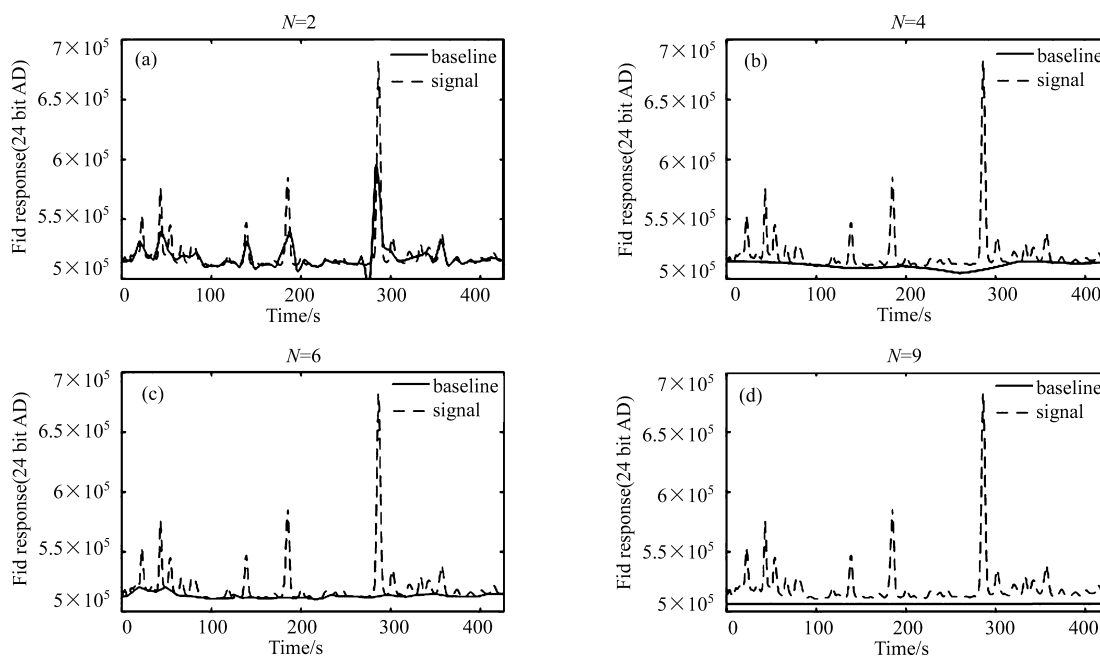


图 3 不同分解层数下拟合基线情况

Fig.3 Fitting baseline under different decomposition levels

小波分解层数对于基线校正效果影响较大。当分解层数过少时, 小波逼近信号频率接近色谱峰信号频率, 导致重构信号中损失色谱峰信息; 当分解层数过多时, 小波逼近信号频率远低于基线漂移信号频率, 导致重构信号中仍包含基线漂移信息。确定小波最优分解层数过程如下: 首先, 使用小波函数对原始信号进行分解, 将得到的小波逼近系数置零后重构, 得到重构信号; 然后, 将原始信号减去重构信号得到基线

信息; 最后, 对比基线信号选取最优分解层数。不同分解层数下得到的基线信号如图 3 所示, 可以看出, 当分解层数 $N=6$ 时, 得到的基线与原始信号基线最为接近, 所以最终选取 6 为最优分解层数。

采用 db6 小波函数对原始信号进行 6 层分解后, 将得到的小波逼近系数置零并重构得到扣除基线的色谱信号如图 4 所示。

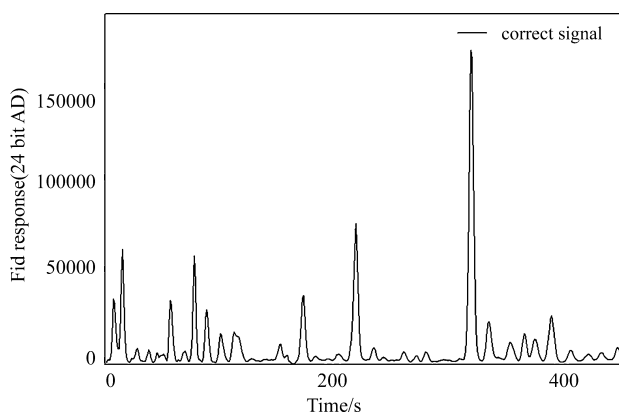


图 4 基线校正后色谱图

Fig.4 Chromatogram after baseline correction

2.2 谱峰解析

色谱峰解析包含谱峰识别和峰面积计算两个过程。采用一阶导数和二阶导数结合的方法进行色谱峰识别, 该方法能够准确识别重叠型色谱峰, 且抗干扰能力强, 运行速度快。完成色谱峰识别后, 采用高斯曲线拟合法^[9]进行色谱峰峰面积计算。

常见的 VOCs 色谱峰类型及其一、二阶导数曲线如图 5 所示。

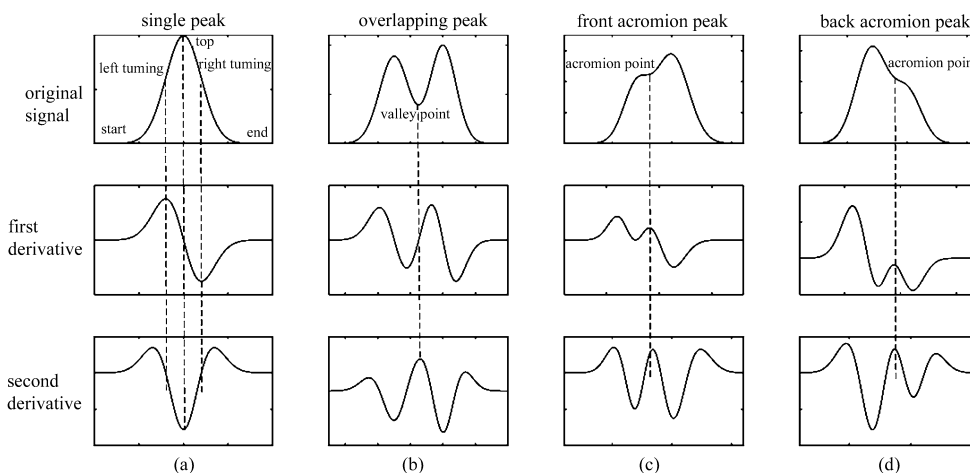


图 5 常见 VOCs 色谱峰类型及其一、二阶导数曲线图

Fig.5 Type of chromatographic peak and the first and second order derivative curve

基于导数法进行色谱峰识别实现过程如下, 设原始色谱信号为 H , 求导后得到一阶导数数据 F , 二阶导数 S , 色谱峰特征点判别条件如下:

- (1) 峰起点前: 曲线处于平静状态, $F(t)$ 和 $S(t)$ 均趋近于 0。
- (2) 峰起点检测: 当 $F(t) > 0, S(t) > 0$ 时, 判断为峰起点, 如图 5(a) 所示。
- (3) 左拐点检测: 当 $F(t) > F(t-1), F(t) > F(t+1)$ 且 $S(t) = 0$ 时, 判断为左拐点, 如图 5(a) 所示。

(4) 峰顶点检测: 当 $H(t) > H(t-1), H(t) > H(t+1)$ 且 $F(t) = 0$ 时, 判断为峰顶点, 如图 5(a) 所示。

(5) 右拐点检测: 当 $F(t) < F(t-1), F(t) < F(t+1)$ 且 $S(t) = 0$ 时, 判断为右拐点, 如图 5(a) 所示。

(6) 峰终点检测条件: 分 3 种情况考虑:

1) $F(t)$ 和 $S(t)$ 再次趋近 0 时, 判断为峰终点 (单峰), 如图 5(a) 所示。

2) $F(t)$ 出现符号跳变, $S(t) > 0$, 判断为重叠峰峰谷点, 如图 5(b) 所示。

3) $H(t)$ 和 $F(t)$ 均出现下降沿, $S(t) > 0$ 肩峰峰终点出现, 如图 5(c) 和 (d) 所示。

经过上述特征点的判断, 可以实现对单峰、肩峰、重叠峰的识别, 识别到一个完整的色谱峰后, 采用高斯曲线拟合法进行峰面积计算, 色谱峰解析流程如图 6 所示。

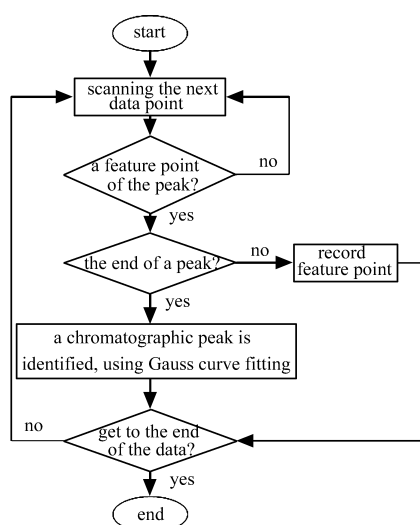


图 6 色谱峰解析流程图

Fig.6 Chromatogram peak analytic flowchat

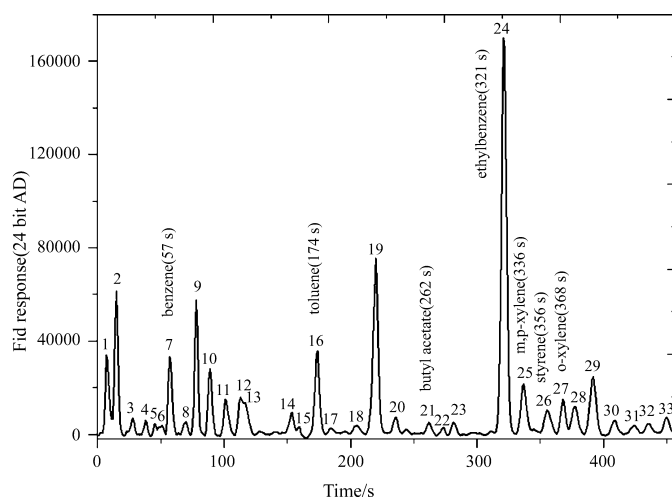


图 7 空气 VOCs 采样结果数据处理后谱图

Fig.7 Chromatogram of air VOCs sampling data processing

3 实验部分

3.1 实验方法

使用数据处理方法对吸附管采样, 气相色谱法分析得到的环境大气 VOCs 色谱图进行处理, 结果如图 7 所示。

3.2 实验条件

(1) 色谱柱选型: OV-1 毛细管柱 (30 m×0.32 mm×1 μm, 中国科学院兰州化学物理研究所); (2) 检测器: 氢火焰离子检测器, 温度设为 280 ℃; (3) 色谱分析条件: 初始柱温 40 ℃, 保持 2 min, 以 5 ℃/min 的速度升温至 220 ℃, 保持 5 min; 恒流模式, 柱流量为 2.5 mL/min。

3.3 结果与讨论

3.3.1 小波函数选取

目前小波函数的选取方式主要是通过对比不同小波函数处理后的结果来判断选择该小波函数是否合适。采用利用小波系数的信息熵^[10]选取合适小波函数的方法。小波熵的大小反应了小波细节系数矩阵的疏密程度, 即信号的不确定度和复杂度, 小波熵 W 为

$$W = - \sum_j p_j \ln(p_j),$$

式中 p_j 为小波系数能量概率分布。小波熵越小, 表明信息的不确定性越大, 信号能量分布越集中, 信号特征

越明显。所以, 通过考察常用小波函数在最优分解层数下的小波熵, 选小波熵最小者为最优小波函数^[11]。各小波函数最优分解层数及小波熵如表1所示。根据表1给出的数据可知, db6小波具有最小信息熵, 所以选用db6小波对VOCs谱图进行基线校正, 最优分解层数为6。

表1 各小波函数最佳分解层数及小波熵

Table 1 Best decomposition layer number and wavelet entropy of each wavelet

Function	Level	Entropy	Function	Level	Entropy
sym2	6	2.525687	db8	7	2.320967
sym4	7	2.786703	db10	6	2.209674
sym6	7	2.674032	coif1	7	2.515096
sym8	7	2.434028	coif2	8	2.499087
db4	7	2.250964	coif3	8	2.492094
db6	6	2.060956	coif4	8	2.370946

3.3.2 色谱峰解析结果讨论:

采用色谱峰识别算法, 对空气采样得到的VOCs色谱峰进行识别, 结果如图7所示。该算法对于图中所有单峰、重叠峰(1和2,5和6,14和15,27和28)以及肩峰(12和13)均可以准确识别。色谱峰识别后, 采用高斯曲线拟合法进行峰面积计算结果, 如表2所示, 同时, 使用苯系物混合标气对图中部分VOCs实现了定性。

表2 部分色谱峰峰面积计算结果

Table 2 Calculation of peak area of partial chromatographs

Peak number	Peak type	Peak name	Peak area
5	Overlap peak	undefined	23031
6	Overlap peak	undefined	18539
7	Independent peak	benzene	144208
12	Acromion peak	undefined	79908
13	Acromion peak	undefined	58192
14	Overlap peak	undefined	35359
24	Independent peak	ethylbenzene	611098
27	Overlap peak	o-xylene	69023

4 结 论

针对环境大气VOCs色谱谱图特点, 设计了适应于环境大气VOCs监测系统的数据处理算法, 该算法具有以下特点:

(1) 采用移动平均法和小波变换法处理高频噪声和基线漂移。基线校正过程首先得到常用小波函数的最优分解层数, 并计算此时小波熵, 最终选取了具有最小信息熵的db6小波函数对基线进行了校正, 能够达到消除基线漂移的目的, 提高了谱峰解析的准确性。

(2) 采用导数法进行色谱峰识别, 对噪声的抗干扰能力强, 并且有效地增强了对重叠峰、肩峰的识别, 同时, 采用高斯曲线拟合法以保证峰面积计算精度。

目前, 使用了该算法的数据处理软件在环境大气VOCs监测系统中已投入运行, 现场运行效果良好。

参考文献：

- [1] Lu Qian. *Study on the Method of Chromatographic Detection and Composition of VOC in Atmospheric Environment* [D]. Wuhan: Master's Thesis of Wuhan University of Technology, 2013(in Chinese).
鲁 茜. 大气环境中 VOC 的组成及其气相色谱检测方法研究 [D]. 武汉: 武汉理工大学硕士论文, 2013.
- [2] Du Guorong, Dong Yiqing, Li Pao. Chemometrics methods in automated analysis of complex chromatographic signals [J]. *Journal of Instrumental Analysis*, 2018, **37**(1): 112-117(in Chinese).
杜国荣, 董怡青, 李 跑. 复杂色谱信号自动解析中的化学计量学方法 [J]. 分析测试学报, 2018, **37**(1): 112-117.
- [3] Ye Guoyang, Xu Kejun. Development of method and software for data processing in gas chromatograph [J]. *Acta Metrologica Sinica*, 2017, **38**(5): 532-537(in Chinese).
叶国阳, 徐科军. 气相色谱仪数据处理方法研究和软件研制 [J]. 计量学报, 2017, **38**(5): 532-537.
- [4] Fang Yonghua, Kong Chao, Lan Tiange, et al. Denoising and baseline correction of spectrum by wavelet transform [J]. *Optics and Precision Engineering*, 2006, **14**(6): 1088-1092(in Chinese).
方勇华, 孔 超, 兰天鸽, 等. 应用小波变换实现光谱的噪声去除和基线校正 [J]. 光学精密工程, 2006, **14**(6): 1088-1092.
- [5] Liu Xiao. A new method for distinguishing chromatographic peaks [J]. *Analytical Instrumentation*, 2005(3): 54-57(in Chinese).
刘 晓. 识别色谱峰的一种方法 [J]. 分析仪器, 2005(3): 54-57.
- [6] Luo Rongmo. *Study on Data Processing in Chromatography Workstation* [D]. Hangzhou: Master's Thesis of Zhejiang University, 2003(in Chinese).
罗荣模. 色谱数据工作站的研究 [D]. 杭州: 浙江大学硕士论文, 2003.
- [7] Liu Mingming. *The Application of Automatic Resolution Method of Chromatographic Signals on Tobacco Analysis* [D]. Hefei: Master's Thesis of University of Science and Technology of China, 2009(in Chinese).
刘明明. 复杂色谱谱图自动解析技术在烟草分析中的应用 [D]. 合肥: 中国科学技术大学硕士论文, 2009.
- [8] Ye Song, Gan Yongying, Xiong Wei, et al. Baseline correction of spatial heterodyne spectrometer using wavelet transform [J]. *Infrared and Laser Engineering*, 2016, **45**(11): 121-125(in Chinese).
叶 松, 甘永莹, 熊 伟, 等. 采用小波变换的空间外差光谱仪基线校正 [J]. 红外与激光工程, 2016, **45**(11): 121-125.
- [9] Zhao Yaoqiang, Li Rong. Discussion on the application of curve fitting in chromatographic analysis [J]. *Mud Logging Engineering*, 2006, **17**(1): 47-49(in Chinese).
赵要强, 李 荣. 曲线拟合技术在色谱分析中的应用探讨 [J]. 录井工程, 2006, **17**(1): 47-49.
- [10] Cui Zhi, Li Jiasheng. Determination algorithm of optimal decomposition level via wavelet entropy [J]. *Instrument Technique and Sensor*, 2015(6): 127-130(in Chinese).
崔 志, 李加升. 基于小波熵自适应最佳分解层数确定算法 [J]. 仪器技术与传感器, 2015(6): 127-130.
- [11] Lu Hao. *Research on Spectrum Processing and Analysis Algorithms of Chromatography Workstation* [D]. Nanjing: Master's Thesis of Southeast University, 2016(in Chinese).
陆 浩. 色谱工作站谱图处理分析算法研究 [D]. 南京: 东南大学硕士论文, 2016.