

文章编号: 0253-2239(2010)09-2742-06

生物气溶胶红外光谱信号预处理算法研究

兰天鸽 熊 伟 方勇华 李大成 袁越明

(中国科学院安徽光学精密机械研究所, 通用光学定标与表征技术重点实验室, 安徽 合肥 230031)

摘要 被动傅里叶变换红外(FTIR)遥感是一种具有应用潜力的生物气溶胶远程探测技术。红外遥感测量中目标光谱特征上往往存在噪声信号和基线漂移。而生物气溶胶的光谱特征相对较宽,传统的基线校正方法都不适用。由于生物气溶胶红外光谱和不同形式的基线漂移都是非高斯信号,把非高斯性作为独立性度量,基于独立成分分析(ICA)技术设计了生物气溶胶红外光谱信号的预处理算法。试验结果表明,该算法可以把未知干扰成分、基线漂移等作为独立分量分离出来,从而不影响进一步的定性、定量分析。

关键词 红外光谱信号;生物气溶胶;被动红外探测;独立成份分析;基线校正

中图分类号 TP722.5 文献标识码 A doi: 10.3788/AOS20103009.2742

Research on Preprocessing Algorithm for Infrared Spectral Signals of Biological Aerosols

Lan Tiange Xiong Wei Fang Yonghua Li Dacheng Yuan Yueming

(Key Laboratory of Optical Calibration and Characterization, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Science, Hefei, Anhui 230031, China)

Abstract It has the application potentials of standoff detection of biological aerosol by passive infrared remote sensing based on Fourier-transform infrared (FTIR) spectroscopic technique. There are often noise signals and baseline drift on the target spectral signature in infrared remote sensing measurement. The spectrum of biological aerosol is relatively broad and the traditional methods of baseline correction are inapplicable. Due to infrared spectra of biological aerosol and different baseline drift are non-Gaussian signals, an algorithm for preprocessing infrared spectra of biological aerosol is devised based on independent component analysis (ICA), where non-Gaussian is used as independent measure. The results of experiments show that this algorithm can separate unknown interference and baseline drift as independent component, and with no effects on the further qualitative and quantitative analysis.

Key words infrared spectral signal; biological aerosol; passive infrared detection; independent component analysis; baseline correction

1 引 言

气溶胶是指悬浮在空气中的固体或液体微粒,在大气环境、空气污染等方面有重要影响,得到了广泛和深入的研究^[1~4]。气溶胶中有生命活性的部分被称为生物气溶胶,包括空气中的细菌、真菌、病毒、尘螨、花粉、孢子和动植物碎裂分解体等具有生命活性的微小粒子。生物气溶胶在空气中的扩散、传输会引起人类急、慢性疾病以及动植物疾病的流行传播。生物气溶胶研究涉及环境保护、卫生防疫、畜牧

业和工农业生产等领域,有重要的价值和意义。

近几年来,基于傅里叶变换红外(FTIR)遥感技术遥测生物气溶胶的研究受到关注。初步试验结果证明了被动红外方法遥测生物气溶胶的可行性^[5~10],但针对生物气溶胶红外光谱的信号处理算法研究则鲜见报道。由于外部环境及仪器状态变化等各种复杂因素的影响,实际测量得到的生物气溶胶红外光谱信号往往叠加在一个低频趋势项上,即所谓的基线漂移。基线漂移、随机噪声等干扰的存

收稿日期: 2009-08-19; 收到修改稿日期: 2009-11-13

作者简介: 兰天鸽(1981—),男,博士,助理研究员,主要从事红外遥感探测和遥感信息分析与处理等方面的研究。

E-mail: tingerlan@aiofm.ac.cn

在会对光谱信号的定性、定量分析产生不利影响,需要进行消除噪声,基线校正等预处理。生物气溶胶红外光谱信号特征相对较宽,在频谱上和干扰成份有较多重叠,传统上基于目标信号特征和干扰成分频谱差异的预处理算法都不适用^[11,12]。独立成分分析(ICA)是主成分分析(PCA)从二阶统计分析向高阶统计分析的拓展,经 ICA 分解出的各信号分量之间则是相互独立的,可用于混合光谱数据的分离及识别^[13,14]。

本文基于独立成分分析技术研究生物气溶胶红外光谱信号的预处理算法。这种算法可以把未知干扰成分、基线噪声等作为独立分量分离出来,而不影响目标气溶胶的识别结果。

2 理论基础

2.1 独立成份分析

若观测信号是相互统计独立的非高斯信号经线性混合而产生的,利用独立成分分析方法可以从观测信号中提取出各独立信号分量。设有一个 N 维观测混合信号 $\mathbf{X}(\nu)$ 的形式为

$$\mathbf{X}(\nu) = \mathbf{A}\mathbf{S}(\nu) + \mathbf{n}(\nu), \quad (1)$$

式中 \mathbf{A} 为源信号的混合矩阵; $\mathbf{S}(\nu)$ 为 $M \times N$ ($M \leq N$) 维源信号矩阵; $\mathbf{n}(\nu)$ 为观测噪声矢量。

ICA 的目的就是计算出 \mathbf{A} 的逆矩阵,即去混合矩阵 \mathbf{W} ,通过它对观测混合信号 $\mathbf{X}(\nu)$ 去混合,恢复相互独立的源信号 $\mathbf{S}(\nu)$,即

$$\tilde{\mathbf{S}}(\nu) = \mathbf{W}\mathbf{X}(\nu), \quad (2)$$

由概率论中的中心极限定理可知,多个随机变量的混合信号趋近于高斯分布。因此,在 ICA 模型中,若干个独立源信号组成的混合信号比任何一个源信号都更接近高斯分布,于是可以使用分离信号的非高斯性作为分离信号之间独立性的度量。需要注意的是,由于高斯信号的线性混合仍然是一个高斯信号,因此在源信号 \mathbf{S} 中最多只允许有一个高斯信号。

实现独立成分分析的算法很多,文中利用 FastICA 算法来完成对数据的独立成分分析^[13]。

2.2 基于 ICA 的生物气溶胶红外光谱信号预处理算法

生物气溶胶红外光谱信号特征相对较宽,属于低频信号,无法在频谱上和基线干扰信号有效区分。注意到生物气溶胶红外光谱和各种不同形式的基线干扰都是非高斯信号,把非高斯性作为它们之间独

立性的度量,利用独立成分分析方法则可能把特征光谱信号和基线分离开。也就是说,对一组生物气溶胶红外光谱信号作 ICA,可以把纯成分目标光谱信号以及各种未知干扰成分等分别作为独立分量分离出来。至于随机噪声信号则可以通过平滑滤波等方法进行处理。

在 FastICA 算法中,初始化去混合矩阵 \mathbf{W} 是随机产生的,那么对同一组数据作多次 ICA,每次分离出的独立分量的次序一般不相同,也就是说首先分离出的独立分量可能并不是目标光谱,而是各种干扰噪声。为了解决这一问题,针对目标气溶胶红外光谱特征设计一组去混合向量代替随机产生的去混合向量,以确保分离出的是目标光谱特征对应的独立分量,而其他干扰成分对应的独立分量顺序仍然是随机的。

综上所述,基于 ICA 的气溶胶红外光谱信号预处理算法可以归纳如下:

1) 对观测混合信号进行中心化、白化。白化是指对数据去相关,即主成分分析。具体做法就是把原始数据 X 去均值,再白化,得 Z ;

2) 假设 $\mathbf{S}_i = [s_{i1}, s_{i2}, \dots, s_{in}]$ 为纯成分目标光谱信号,其对应的去混向量可以根据公式 $\mathbf{W}_n = \mathbf{Z}\mathbf{S}_i$ 计算。

3) 如果纯成分目标光谱信号的数目 n 小于要提取的独立分量的数目 m ,再随机生成 $m-n$ 个初始去混向量,利用 FastICA 算法计算去混合向量 \mathbf{W}_{m-n} 。

4) 获取独立分量及对应的混合系数,从观测光谱信号中扣除基线干扰成分。

3 试验与讨论

3.1 试 验

在该部分以某种生物孢子气溶胶和高岭土气溶胶的红外光谱为研究对象,对提出的预处理算法进行检验。使用德国 Bruker 公司的 OPAG33 型 FTIR 遥感光谱仪进行气溶胶红外光谱测量。FTIR 遥感光谱仪输出的干涉图经过傅里叶变换得到光谱,测量范围为 $700 \sim 1300 \text{ cm}^{-1}$,光谱分辨率约为 2 cm^{-1} 。试验中,扣除实际测量红外光谱中的背景信号,得到气溶胶光谱特征,再把实测气溶胶特征光谱归一化和一组参考光谱信号(如图 1 所示)一起做独立成分分析,其中孢子和高岭土气溶胶的参考光谱信号分别为 3 条。

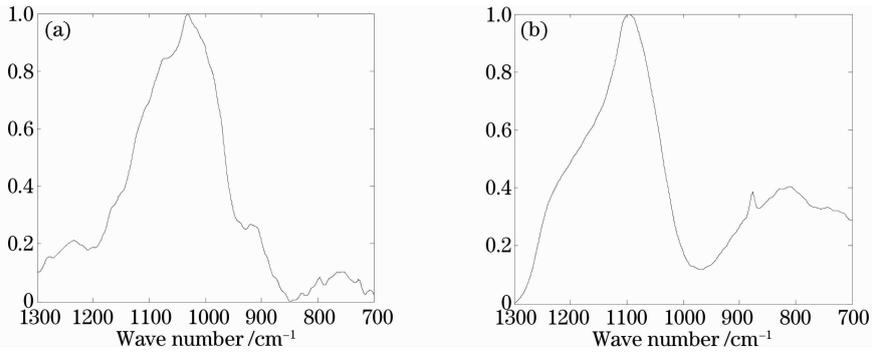


图 1 气溶胶的参考光谱特征。(a)生物孢子,(b)高岭土粉末

Fig.1 Reference spectral signatures of aerosols. (a) biological spore, (b) kaolin powder

3.2 结果分析

如图 2 所示,是实际测量的生物孢子气溶胶红外光谱 ΔL 表示光谱辐射量。可以看到,目标光谱特征叠加在基线干扰成分之上。对其进行归一化处理,按照前文所述方法做独立成分分析。

图 3 所示是分离出的 4 个独立分量。第一和第二独立分量分别是孢子、高岭土气溶胶对应的红外光谱特征,而第三和第四独立分量则表征随机噪声、基线漂移信号。表 1 是 ICA 得到的混合系数 A ,它每一行的 4 个数值表示该条光谱中对应 4 个独立分量的系数。第一行中,第一个数绝对值最大,其余三个都很小,结合分离出来的独立分量,可以得出这是孢子气溶胶光谱的结论,而且噪声也很小,不存在基线漂移,这与第一条光谱是标准参考信号的情况相符。其他 5 条参考光谱信号的分析结果也是正确的。

对于待处理的气溶胶特征光谱,分析表中第三行,第一、第四个系数绝对值都较大,第二和第三个系数较小,说明光谱信号中存在孢子气溶胶光谱特征和第四个独立分量所代表的基线。

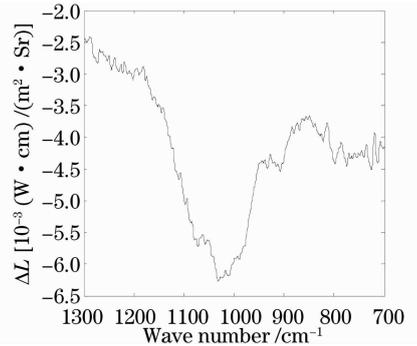


图 2 试验测量的生物孢子气溶胶红外光谱特征
Fig.2 Infrared spectral signature of biological spore aerosol in measurements

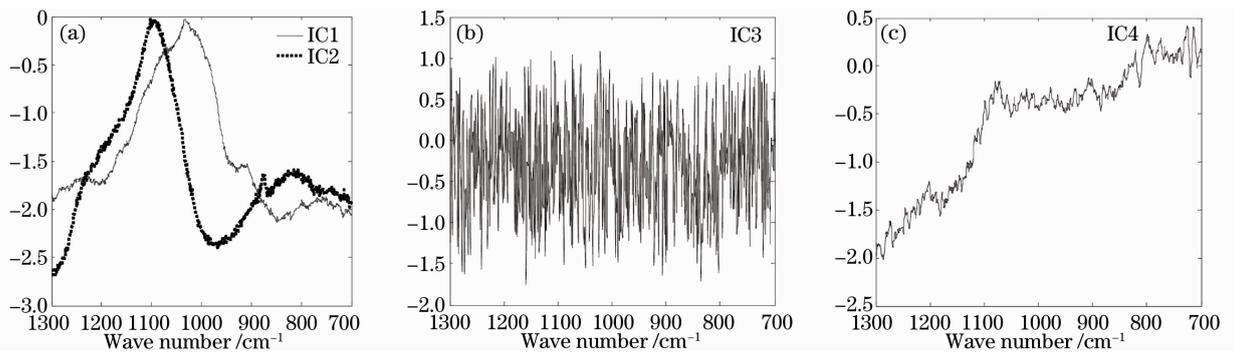


图 3 ICA 分析的结果。(a)孢子和高岭土气溶胶对应的红外光谱特征,(b)随机噪声,(c)基线漂移信号

Fig.3 Independent components achieved from ICA. (a) spore and kaolin powders spectral signatures,

(b) stochastic noise; (c) base line drift signal

表 1 对孢子气溶胶光谱作 ICA 得到的独立分量对应的部分混合系数

Table 1 Mixing coefficients of independent components by ICA for spore spectral signature

	IC1	IC2	IC3	IC4
S_1	-0.46396	-0.00087	-0.00111	0.00077
S_4	0.00152	-0.3644	-0.01646	-0.00072
S_7	-0.33724	0.05697	0.00479	-0.25952

利用独立分量和表 1 中第三行的混合系数,可以得到重构信号图 4(a)以及重构信号和原始信号的误差,即重构误差图 4(b)。从图中可以看到,重构误差非常小,重构信号和原始信号几乎完全重合,这说明上述分析和结论是可靠的。如图 5 所示,进

一步分别重构了纯成分光谱信号和噪声干扰信号。为了便于比较,各谱线的位置作了上下平移。可以看到,重构的噪声干扰信号和原始光谱数据中存在的基线漂移趋势一致,而纯成分光谱信号不再有随机噪声和基线漂移的干扰。

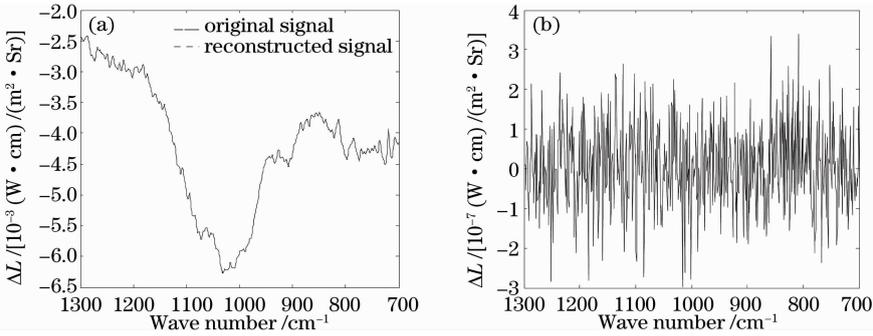


图 4 重构分析。(a)原始信号和重构信号,(b)重构误差

Fig. 4 Reconstructed analysis. (a) original signal and reconstructed signal, (b) error of reconstruction

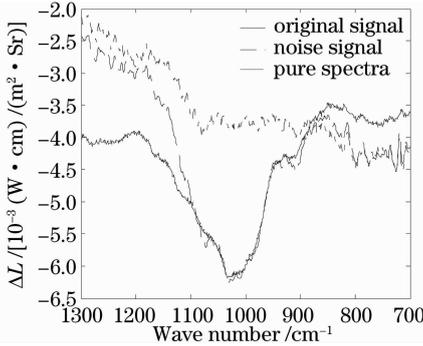


图 5 重构的孢子气溶胶纯成分光谱信号和噪声信号

Fig. 5 Reconstructed pure component spectral signal and noise signal of spore aerosol

对图 6 所示的高岭土气溶胶光谱信号进行相同处理。图 7 是分离出的独立分量。表 2 则是得到的混合系数 A。分析表中第三行,第二、第四个系数绝对值都较大,第一和第三个系数较小,说明信号中存在第二个和第四个独立分量所代表的高岭土气溶胶

光谱特征和基线成分。

图 8 所示是重构的纯成分光谱信号和噪声干扰信号。为了便于比较,各谱线的位置作了上下平移。可以看到,重构的噪声干扰信号和原始光谱数据中的基线漂移趋势也一致,而纯成分光谱信号不再有随机噪声和基线漂移的干扰。

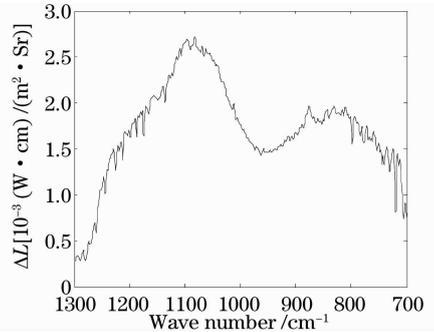


图 6 试验测量的高岭土粉末气溶胶红外光谱特征
Fig. 6 Infrared spectral signature of kaolin powder aerosol in measurements

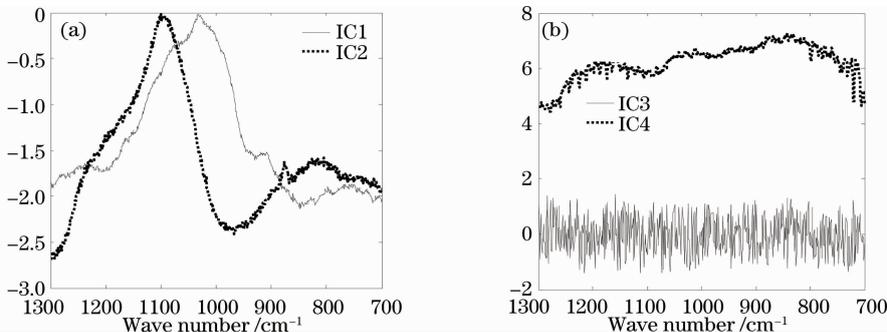


图 7 对图 6 所示的 ICA 分析得到的独立分量。(a)红线光谱特征,(b)噪声和基线漂移

Fig. 7 Independent components of Fig. 6 achieved from ICA analysis. (a) spectral signatures of signal, (b) stochastic noise and baseline drifts

表 2 对高岭土气溶胶光谱作 ICA 得到的独立分量对应的部分混合系数

Table 2 Mixing coefficients of independent components by ICA for Kaolin spectral signature

	IC1	IC2	IC3	IC4
S_1	-0.46338	-5.04×10^{-5}	-0.00011	0.0011
S_4	0.00171	-0.36413	0.018767	0.00112
S_7	0.095022	0.24135	-0.003253	0.17849

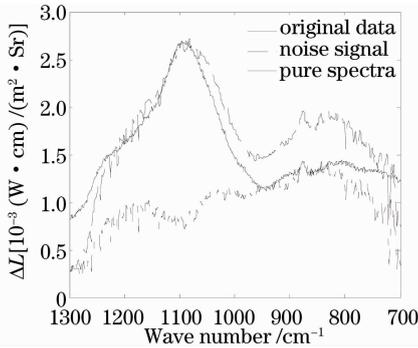


图 8 重构的高岭土纯成分光谱信号和噪声信号

Fig. 8 Reconstructed pure component spectral signal and noise signal of kaolin aerosol

气溶胶一般都是多组分体系,测量得到的光谱信号可能是包含多种成分特征的混合光谱。如图 9 所示,是获取的一条生物孢子和高岭土混合粉末气溶胶的红外光谱特征信号。对其进行归一化,并按照本文提出的算法进行处理。

图 10 和表 3 分别是分离出的 4 个独立分量及其对应的混合系数。很明显第一和第二独立分量分别是孢子、高岭土气溶胶对应的红外光谱特征,而第三和第四独立分量则表征随机噪声、基线漂移信号。

表 3 中第三行 S_7 是混合粉末气溶胶红外光谱信号对应各独立分量的混合系数,可以看到,第一独立分量的对应的混合系数最大,可以确定目标气溶胶中存在该生物孢子成分。注意到第二独立分量对应的混合系数也较大,为 0.13044,说明目标气溶胶中也可能存在高岭土成分,这与实际情况是相符的。结果说明,提出的算法可对混合光谱信号去混合,从混合光谱中分离出多种单一目标特征光谱,而不相互影响。

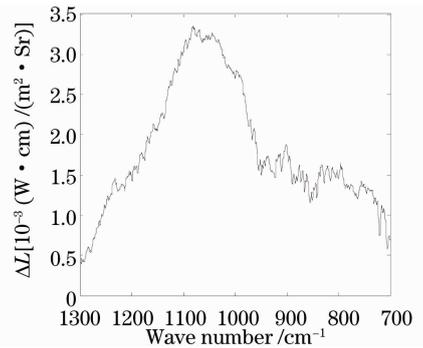


图 9 生物孢子和高岭土混合粉末气溶胶的红外光谱特征

Fig. 9 Infrared spectral signature of mixed powder aerosol (biological spore and kaolin)

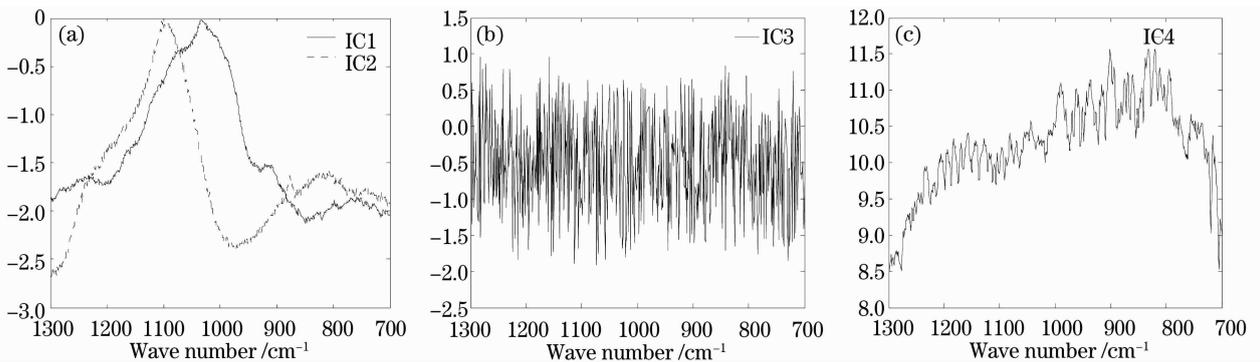


图 10 对图 9 所示的 ICA 分析得到的独立分量(a),随机噪声信号(b),基线漂移信号(c)

Fig. 10 Independent components of Fig. 9 from ICA analysis (a), spectral signatures (b), baseline drift (c)

表 3 对孢子高岭土混合粉末气溶胶光谱作 ICA 得到的独立分量对应的部分混合系数

Table 3 Mixing coefficients of independent components by ICA for mixed powder aerosol spectral signature

	IC1	IC2	IC3	IC4
S_1	-0.46333	1.7687×10^{-5}	-0.00014026	0.00089243
S_4	0.0016279	-0.36424	0.018715	0.0014755
S_7	0.31355	0.13044	-0.0021212	0.11079

4 结 论

基于独立成分分析技术设计了生物气溶胶红外光谱信号的预处理算法。由于生物气溶胶红外光谱和不同形式的基线漂移都是非高斯信号,本算法把非高斯性作为独立性度量,通过独立成分分析,把未知干扰成分、基线漂移等作为独立分量分离出来,而不影响进一步的定性、定量分析结果。另外,本算法可对混合光谱信号去混合,从混合光谱中分离出多种单一目标特征光谱,而不相互影响。

参 考 文 献

- Zhenyi Chen, Wenqing Liu, Yujun Zhang *et al.*. Measurements of aerosol distribution by an elastic-backscatter lidar in summer 2008 in Beijing[J]. *Chin. Opt. Lett.*, 2009, **7**(9): 753~755
- Dong Yunsheng, Liu Wenqing, Liu Jianguo *et al.*. Lidar study of the aerosol characteristic in Beijing during traffic controlled[J]. *Acta Optica Sinica*, 2009, **29**(2): 292~296
董云升, 刘文清, 刘建国等. 北京城区限行期间气溶胶特征激光雷达观测研究[J]. *光学学报*, 2009, **29**(2): 292~296
- Wang Xianhua, Qiao Yanli, Philippe Goloub *et al.*. Radiometric calibration of sunphotometer system applied to aerosol robotic network[J]. *Acta Optica Sinica*, 2008, **28**(1): 87~91
王先华, 乔延利, Philippe Goloub等. 应用于全球气溶胶测量网的太阳辐射计辐射定标系统[J]. *光学学报*, 2008, **28**(1): 87~91
- Li Xuebin, Xu Qingshan, Wei Heli *et al.*. Study on relationship between extinction coefficient and mass concentration[J]. *Acta Optica Sinica*, 2008, **28**(9): 1655~1658
李学彬, 徐青山, 魏合理等. 气溶胶消光系数与质量浓度的相关性研究[J]. *光学学报*, 2008, **28**(9): 1655~1658
- K. P. Gurton, D. A. Ligon, R. Kvavilashvili. Measured infrared spectral extinction for aerosolized *Bacillus subtilis* var. niger endospores from 3 to 13 μm [J]. *Appl. Opt.*, 2001, **40**(25): 4443~4448
- A. Bendavid, H. Ren. Detection, identification and estimation of biological aerosols and vapors with Fourier transform infrared spectrometer[J]. *Appl. Opt.*, 2003, **42**(24): 4887~4900
- A. Bendavid. Remote detection of biological aerosols at a distance of 3 km with a passive Fourier transform infrared (FTIR) sensor [J]. *Opt. Express*, 2003, **11**(5): 418~429
- J. M. Therault, E. Puckrin, J. O. Jensen. Passive standoff detection of BG aerosol; method and field trial results[C]. *SPIE*, 2004, **5268**, 163~172
- A. Ifarraguerri, A. Bendavid, R. G. Vanderbeek. Estimating the limit of bio-aerosol detection with passive infrared spectroscopy[J]. *Intern. J. High Speed Electron. and Systems*, 2008, **18**(3): 701~711
- Lan Tiange, Xiong Wei, Fang Yonghua *et al.*. Study on passive detection of biological aerosol with fourier-transform infrared spectroscopic technique[J]. *Acta Optica Sinica*, 2010, **30**(6): 1656~1661
兰天鸽, 熊伟, 方勇华等. 应用被动傅里叶变换红外光谱技术探测生物气溶胶研究[J]. *光学学报*, 2010, **30**(6): 1656~1661
- Tiange Lan, Yonghua Fang, Wei Xiong *et al.*. Automatic baseline correction of infrared spectra [J]. *Chin. Opt. Lett.*, 2007, **5**(10): 613~616
- Fang Yonghua, Kong Chao, Lan Tiange *et al.*. Denoising and baseline correction of spectrum by wavelet transform[J]. *Opt. Precision Eng.*, 2006, **14**(6): 1088~1092
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
- A. Hyvarinen, E. Oja. A fast fixed-point algorithm for independent component analysis[J]. *Neural Computation*, 1997, **9**(7): 1483~1492
- Nikos Pasadakis, Andreas A. Kardamakis. Identifying constituents in commercial gasoline using Fourier transform-infrared spectroscopy and independent component analysis[J]. *Analytica Chimica Acta*, 2006, **578**(2): 250~255