

大气开放光路傅里叶变换红外光谱的定性分析

余莲莲 邵利民*

(中国科学技术大学化学系,合肥 230026)

摘 要 通过目标波段熵最小化(BTEM)和目标因子分析(TFA)两种化学计量学方法研究了大气开放光路傅里叶变换红外(OP/FT-IR)光谱,从多组分测量数据重建特定目标分子的光谱特征,以实现大气这种复杂体系的定性分析。在牧场周围完成了5次连续OP/FT-IR光谱测量,将光谱按照测量时间顺序排列,获得5个数据矩阵作为研究对象。结果表明,BTEM和TFA两种方法均可从多组分光谱数据矩阵重建出目标分子的光谱特征,重建效果略有差异。TFA能够克服干扰,提取出目标分子的光谱特征,其重建光谱与相应的参考光谱高度相似。BTEM方法无需目标分子的参考光谱,目标函数最优解所对应的谱图就是纯组分光谱。本研究利用两种方法各自的优势,用化学计量学方法定性分析OP/FT-IR光谱,不仅将BTEM方法应用于OP/FT-IR光谱,而且比较这两种方法的应用,为实现多组分OP/FT-IR光谱定性分析提供了新思路。

关键词 开放光路傅里叶变换红外光谱;目标波段熵最小化;目标因子分析

1 引言

开放光路傅里叶变换红外(OP/FT-IR)光谱是大气测量的有效工具,已广泛用于大气分析和环境监测^[1~4]。OP/FT-IR具有仪器可靠、操作简单、无需采样、在线测量、以及无人值守监测等优点^[5~8]。在实际测量中,OP/FT-IR光谱受到水蒸汽和二氧化碳的振-转光谱以及基线缓慢变化的干扰,严重影响了痕量组分的分析。此外,光谱定性分析主要基于目视检查或光谱库搜索,要求操作者能有效识别目标化合物的特征,这些都给数据解析带来一定的困难。本研究将采用两种化学计量学方法克服上述困难。

目标因子分析(Target factor analysis, TFA)是一种自模因子分析方法,通过分析多个光谱组成的数据矩阵,能够提取出隐藏在单个光谱中的信息。TFA目前在化学中已广泛应用于色谱^[9],反应力学和动力学^[10,11],光谱分析^[12],医疗和医药研究^[13,14]以及食品^[15]等领域。

目标波段熵最小化(Band-target entropy minimization, BTEM)方法能够从多组分光谱中重建纯组分光谱,具有无需目标分子的先验信息,也不依赖于统计学检验等优点^[16]。BTEM已应用于各种液相和固相的反应体系^[17],已经成功重建FT-IR^[18], Raman^[19], NMR^[20]和MS^[21]等复杂光谱,但是对OP/FT-IR光谱的重建目前尚未见报道。

Widjaja等^[22]将BTEM与其它曲线分解程序(如OPA-ALS, SIMPLISMA, SIMPLISMA-ALS, IPCA)进行了对比,表明BTEM是可以重建出最接近纯组分参考光谱的方法,但是尚未见将BTEM与TFA这两种算法进行比较的报道。因此,本研究将用BTEM和TFA两种方法定性分析混合物光谱中NH₃, C₂H₅OH和O₃的光谱信息,并对结果进行比较。结果表明,TFA基于参考光谱,主要组分的重建效果很好,低信号强度的次要组分中存在干扰会影响重建的效果;BTEM中,无论是主要组分还是次要组分,都在不依靠参考光谱的前提下进行重建,而重建结果的好坏主要取决于目标函数的选择。本研究不仅将BTEM方法应用于OP/FT-IR光谱,而且将BTEM和TFA两种方法的结果进行了比较,为实现多组分OP/FT-IR光谱定性分析提供了一定的思路和参考。

2014-09-15 收稿; 2014-10-24 接受

本文系国家自然科学基金资助(No. 21175123)和教育部新世纪优秀人才支持计划(No. NCET-11-0878)资助项目

* E-mail: lshao@ustc.edu.cn

2 实验部分

2.1 仪器及光谱测量

OP/FT-IR 光谱测量在美国 Idaho 州南部的一个牧场进行,用于研究牧场周围的大气组成。OP/FT-IR 光谱仪为 MDA(Atlanta, GA) 产品; Bomem Michelson 100 干涉器(加拿大 Bomem 公司); 直径为 31.5 cm 的望远镜用于产生准直红外光束; 汞镉碲(MCT) 检测器; 准直反射器为立方角阵列反射器。

2.2 光谱处理

通过干涉谱计算得到分辨率 1 cm^{-1} 的吸收光谱, 填零因子为 8, 切趾函数 Norton-Beer(Medium), 并且对 MCT 检测器的非线性响应进行了校正^[23]。所有的光谱预处理和数据计算采用 MATLAB(美国 MathWorks 公司)。

进行 5 次连续测量, 每一次连续测量的单个光谱按照测量时间的顺序进行排列, 组成 5 个混合物光谱数据集矩阵, 5 个数据集中混合物光谱数分别为 1177, 600, 811, 332 和 1184, 光谱波数范围为 $750 \sim 1250 \text{ cm}^{-1}$ 。

3 计算部分

OP/FT-IR 光谱连续测量获得 m 个吸收光谱, 每个光谱有 n 个数据点, 通过逐行排列的方式排列光谱, 得到一个 m 行 n 列的光谱 $D_{m \times n}$, 由朗伯-比尔定律可知 $D_{m \times n}$ 是浓度矩阵 $C_{m \times p}$ 和光谱矩阵 $S_{n \times p}$ 的线性组合。

$$D_{m \times n} = C_{m \times p} S_{n \times p}^T \quad (1)$$

其中 p 是可以重建出纯组分的数量, 上标 T 表示矩阵转置。

对 $D_{m \times n}$ 进行 PCA

$$D_{m \times n} = U_{m \times q} V_{n \times q}^T + R_{m \times n} \quad (2)$$

其中 q 是主因子的个数, 在理论上等于式(1)中的 p ; U 和 V 分别是主成分向量和特征向量组成的矩阵; R 是残余矩阵。忽略残余矩阵 R , 得到:

$$C_{m \times p} S_{n \times p}^T = U_{m \times q} V_{n \times q}^T \quad (3)$$

式(3)表明, V 包含了 S 中所有的真实光谱信息, 只是形式不同, 因此特征向量也称为抽象光谱。特征向量与真实光谱向量张成同一线性空间, 因此 S 中任何一个光谱向量 $s_{n \times 1}$ 都可以由特征向量 V 线性表示:

$$s_{n \times 1} = V_{n \times q} T_{q \times 1} \quad (4)$$

其中 $T_{q \times 1}$ 表示旋转向量。

3.1 目标因子分析

TFA 是 Malinowski 提出的一种化学因子分析方法^[24], 检验测量数据矩阵中是否包含目标物的化学信息, 进而判断样品中是否存在目标物。

TFA 通过式(4)获得目标向量, 其中旋转向量 T 通过最小二乘方法获得:

$$T_{q \times 1} = (V_{n \times q}^T V_{n \times q})^{-1} V_{n \times q}^T s_{n \times 1} \quad (5)$$

TFA 中, 首先选择目标物的参考光谱, 然后将参考光谱代入式(5), 计算得到旋转向量 T , 再将 T 代入式(4)即可得到重建光谱。最后, 比较参考光谱和重建光谱的相似程度, 以此来判断目标物是否存在于被测样品中。

3.2 目标波段熵最小化方法

目标波段熵最小化方法^[22]是一种将光谱的目标波段和信息熵最小化的概念相结合的曲线分解方法。BTEM 在特征向量中选择具有目标分子光谱特征的波段作为目标波段, 设计了目标函数, 当目标函数达到最小时, 这些光谱特征能够被保留, 而且其它干扰最小。通过模拟退火算法(Simulated Annealing, SA)搜索最优旋转向量 T , 使目标函数最小。最后, 将最优旋转向量 T 代入(4)式以重建目标分子的纯光谱。在模拟退火算法中, 旋转向量 T 的初始值随机产生, 因此多次运行结果存在一定差异。

BTEM 的目标函数具有以下形式^[22]:

$$G = H(T) + P(T) \quad (6)$$

其中 $H(T)$ 为 Shannon 信息熵函数, $P(T)$ 为惩罚函数。Shannon 在研究信号处理时通过最小化信息熵来简化光谱^[25]。信息熵反映了频谱范围内光谱信息的离散程度; 信息熵最小能够使光谱特征在光谱基线平滑时得以集中, 从而最大限度地简化光谱。惩罚函数 P 的作用是使目标物的光谱和浓度非负, 越接近于 0 目标函数越能获得最优解^[26]。

4 结果与讨论

4.1 主因子数的确定

对数据集 1 的矩阵 $D_{1177 \times 2075}$ 进行 PCA, 得到的前 10 个特征向量按照特征值的大小排列。由图 1 可见, 前 8 个特征向量显示了明确的光谱特征, 其余主要是仪器噪声。因此保留前 8 个主成分进行下一步分析。对其它 4 个数据集的矩阵进行 PCA, 通过类似的方法得到了相应的主因子数。这 5 个数据集的主因子数分别为 8, 6, 10, 6 和 9。

4.2 TFA 与 BTEM 的结果

对数据集 1, 图 1 中 EV1 ~ EV2 主要为水蒸汽的吸收, EV3 ~ EV6 在 $900 \sim 1000 \text{ cm}^{-1}$ 有两个特征峰, 将 EV3 中较明显的两个特征峰分别标记为 i 和 j , EV7 ~ EV8 在 $1000 \sim 1100 \text{ cm}^{-1}$ 有一个特征峰, 将 EV8 中此特征峰标记为 k (各特征峰的波数范围见图 1 标注)。用 TFA 和 BTEM 两种方法进一步分析, 通过重建光谱与参考光谱的相似程度以及二者的相关系数数值作为组分定性分析的依据。

分别以图 1 中的 i, j 作为目标波段, 通过 BTEM 方法进行光谱重建。结果表明, 重建光谱与 NH_3 参考光谱十分相似, 重建光谱与参考光谱之间相关系数分别为 0.9394 和 0.9408。由此可以确定数据集 1 中存在 NH_3 的光谱信息, 尽管在 $1150 \sim 1250 \text{ cm}^{-1}$ 有水蒸汽的干扰。综上所述, 对 NH_3 光谱进行重建时, 无论选择 i 还是 j 作为目标波段, 重建出的光谱图相似, 相关系数数值大小相近。因此, 从其它数据集重建 NH_3 光谱时均选择 j 作为目标波段。

为确定数据集 1 中其它组分, TFA 方法分别通过 $\text{C}_2\text{H}_5\text{OH}$ 和 O_3 参考光谱进行光谱重建, 这两种组分在 $1000 \sim 1100 \text{ cm}^{-1}$ 均有特征吸收峰。BTEM 方法则以 k 为目标波段进行光谱重建, 光谱重建结果见图 2B 和 2C。结果表明, TFA 方法无论是依靠 $\text{C}_2\text{H}_5\text{OH}$ 还是 O_3 参考光谱进行光谱重建, 重建的光谱都与 $\text{C}_2\text{H}_5\text{OH}$ 参考光谱高度相似; BTEM 方法重建的光谱也与 $\text{C}_2\text{H}_5\text{OH}$ 参考光谱十分相似。此外, 两种方法的重建光谱与 $\text{C}_2\text{H}_5\text{OH}$ 参考光谱相关系数分别为 0.9476 和 0.9058, 与 O_3 参考光谱相关系数分别为 0.8328 和 0.7682。综上所述, 可以确定数据集 1 中存在 $\text{C}_2\text{H}_5\text{OH}$ 。

尽管可以确定 $\text{C}_2\text{H}_5\text{OH}$ 的存在, 两种方法的 $\text{C}_2\text{H}_5\text{OH}$ 重建光谱还是存在明显干扰, 包括噪声, 以及在 $750 \sim 760 \text{ cm}^{-1}$ 和 $1150 \sim 1250 \text{ cm}^{-1}$ 的水蒸汽吸收。无法完全消除干扰的原因是数据集中 $\text{C}_2\text{H}_5\text{OH}$ 光谱信息较弱, 致使干扰相对较强。对于 BTEM 方法重建的 $\text{C}_2\text{H}_5\text{OH}$ 光谱, $930 \sim 980 \text{ cm}^{-1}$ 出现了两个微小的 NH_3 吸收峰, 这一干扰并没有出现在 TFA 方法的 $\text{C}_2\text{H}_5\text{OH}$ 重建光谱。主要原因是, TFA 方法依靠 $\text{C}_2\text{H}_5\text{OH}$ 参考光谱进行重建, BTEM 方法则没有任何可参考的信息, 因此重建的光谱不可避免地受到吸收较强的 NH_3 干扰。

为了进一步比较 TFA 和 BTEM, 分别将两种方法得到的旋转向量代入 BTEM 的目标函数, 得到目标函数中的各项值, 见表 1。

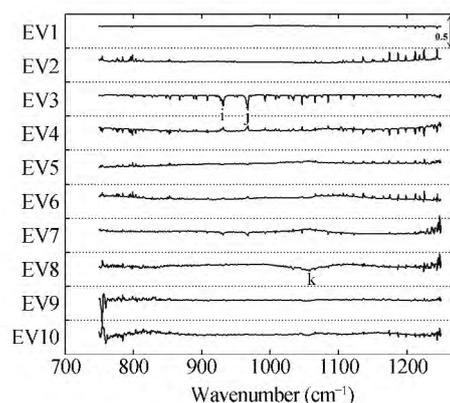


图 1 数据集 1 的前 10 个特征向量图

Fig. 1 The first ten eigenvectors calculated for data set #1. i, j, k 分别代表 BTEM 进行光谱重建时所使用的目标波段, 波数范围分别是 $931 \sim 933 \text{ cm}^{-1}$, $967 \sim 968 \text{ cm}^{-1}$, $1100 \sim 1102 \text{ cm}^{-1}$. i, j , and k represent target bands used in band-target entropy minimization (BTEM) while reconstructing spectra, respectively. The wavenumber ranges are $931 \sim 933 \text{ cm}^{-1}$, $967 \sim 968 \text{ cm}^{-1}$, $1100 \sim 1102 \text{ cm}^{-1}$ respectively.

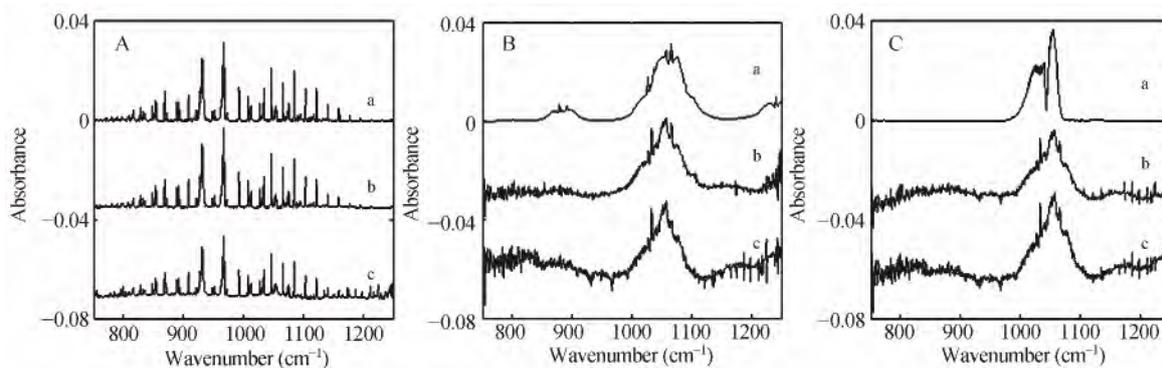


图2 数据集1处理结果

Fig.2 The result of data set 1

A, B, C 分别代表数据集1中重建光谱的谱图; a, b, c 分别对应参考光谱, TFA 方法和 BTEM 方法的重建光谱。

A, B and C represent the reconstructed spectra in the data set #1; a, b and c correspond to the reference spectra, the reconstructed spectra by target factor analysis (TFA) and BTEM, respectively.

表1 TFA 与 BTEM 方法的旋转向量在 BTEM 目标函数中各项值

Table 1 Values in the objective function of BTEM with rotation vectors obtained by TFA and BTEM

数值 Value	氨 Ammonia		乙醇 Ethanol	
	目标因子分析 Target factor analysis	目标波段熵最小化 Band-target entropy minimization	目标因子分析 Target factor analysis	目标波段熵最小化 Band-target entropy minimization
P	0.0191	0.0029	0.0064	0.00014
H	1.7154	1.9622	2.0159	2.0309
G	1.7345	1.9651	2.0223	2.0310

表1 数据显示, TFA 的目标函数值 G 均小于 BTEM 的目标函数值 G , 其中主要原因是 TFA 的 Shannon 信息熵值 H 较小, 而 H 较小是因为 TFA 在依靠纯组分参考光谱的前提下进行光谱重建, 但是 BTEM 方法无任何可参考信息, 所以 BTEM 受到的干扰相对较大。此结论与图 2A 和 2B 中部分光谱图的结果一致。此外, 两种方法的惩罚函数 P 均接近于 0, 而且对目标函数值 G 的影响很小, 表明光谱重建中目标物光谱和浓度非负的影响较小。

处理其它 4 个数据集, 光谱重建结果见图 3。图 3A 表明, 除了数据集 2, 从其它数据集均能重建出 NH_3 光谱。这是因为数据集 2 是远离牧场测量的, NH_3 的浓度非常低, 无法从测量光谱中提取出 NH_3 的信息。

图 3B 表明, 从数据集 2 和 4 中能够成功重建 $\text{C}_2\text{H}_5\text{OH}$ 光谱。在数据集 2 中, 虽然 NH_3 浓度较小重建不出 NH_3 光谱, 但是在重建 $\text{C}_2\text{H}_5\text{OH}$ 光谱时, 还是会受到微弱 NH_3 和水蒸气的干扰。而在数据集 4 中重建的 $\text{C}_2\text{H}_5\text{OH}$ 光谱只受到水蒸气的干扰, 这是因为在这个数据集中乙醇的浓度较大而且对红外的吸收较强, NH_3 的干扰相对较小。对数据集 3, TFA 方法以乙醇的参考光谱进行重建, 重建出具有臭氧特征峰的光谱, 这表明在数据集 3 中微量组分臭氧的浓度较大, 尽管在原始实验光谱中无法看到臭氧的特征吸收。对于数据集 5, 在 $1000 \sim 1100 \text{ cm}^{-1}$ 有一个类似于 $\text{C}_2\text{H}_5\text{OH}$ 的特征峰, 但是不明显。计算出两种方法的重建光谱与 $\text{C}_2\text{H}_5\text{OH}$ 参考光谱的相关系数分别为 0.8108 和 0.7535, 进一步确认不存在 $\text{C}_2\text{H}_5\text{OH}$ 。

图 3C 表明, 从数据集 3、5 中能重建出 O_3 光谱。对于数据集 3 和 5, 两种方法的重建光谱尽管受到水蒸汽红外吸收的干扰, 但可以很明显地看到 O_3 的特征峰。对于数据集 2 和 4, 两种方法的重建光谱均为 $\text{C}_2\text{H}_5\text{OH}$ 光谱, 表明在这两个数据集中不存在 O_3 的光谱信息。

图 3B-c 和图 3C-c 分别是 BTEM 处理同一数据、使用相同波段重建的光谱, 二者的光谱特征基本一致, 然而存在细微差别, 这与 BTEM 中旋转向量 T 的初始值有关。 T 的初始值通过随机产生, 每次计算时的初值并不相同, 因此多次运行结果略有差异。尽管如此, 重建光谱的光谱特征基本一致, 保证了 BTEM 方法定性分析结果的可靠性。

混合物中组分重建光谱与参考光谱之间相关系数值见表 2。结果表明, 相关系数基本上反映了

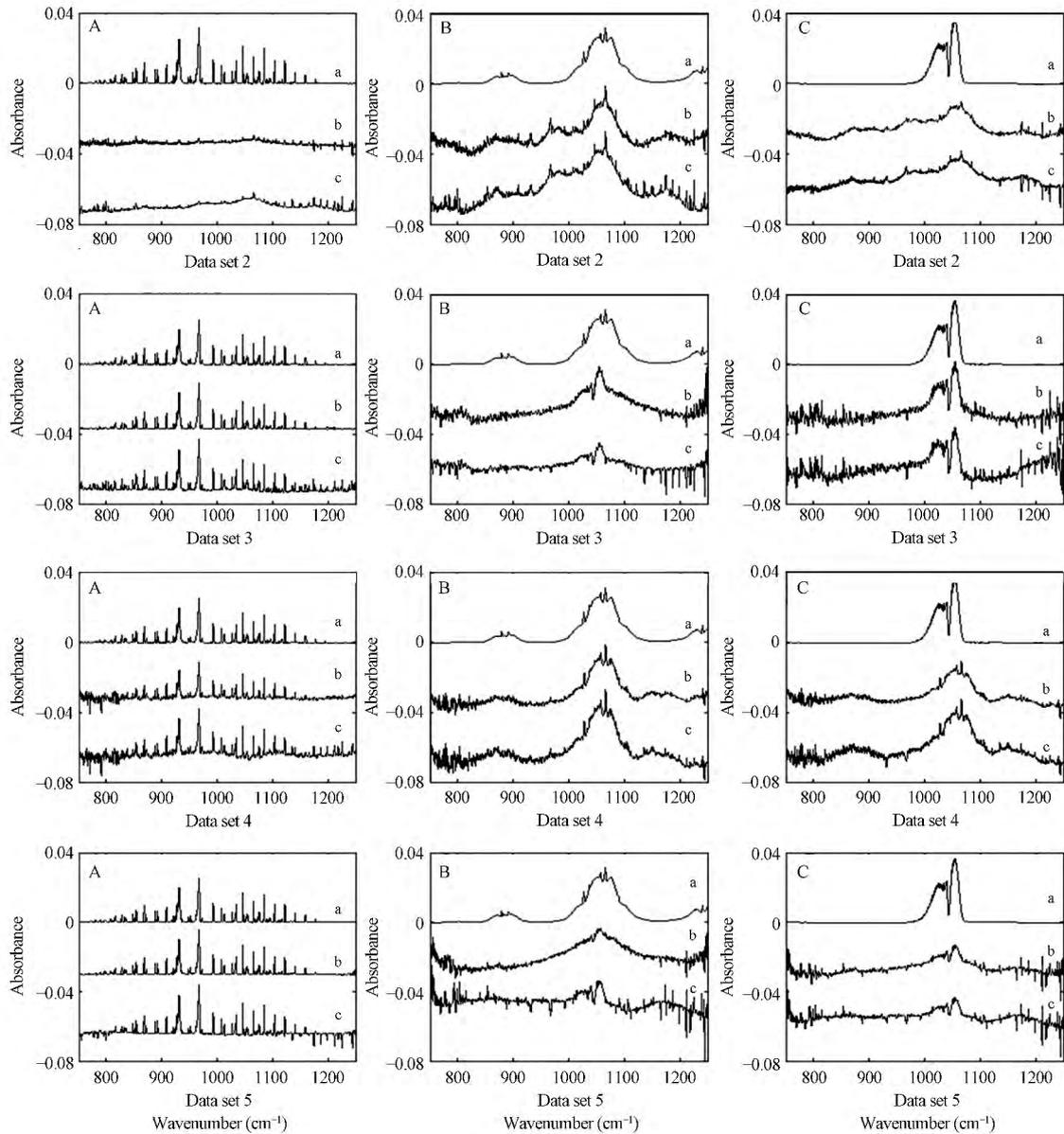


图 3 数据集 2~5 处理结果

Fig. 3 The results of data sets #2 to #5

A, B, C 分别代表数据集 2~5 中重建光谱的谱图。图中 a, b, c 分别对应参考光谱、TFA 方法的重建光谱、BTEM 方法的重建光谱。

A, B and C represent the reconstructed spectra in the data sets 2-5; a, b and c correspond to the reference spectra, the reconstructed spectra by TFA and BTEM, respectively.

表 2 混合物光谱的组分判定结果

Table 2 Identifications of target molecules from the spectra mixtures

数据集 Data set	臭氧 Ozone		乙醇 Ethanol		氨 Ammonia	
	目标因子分析 Target factor analysis	目标波段熵最小化 Band-target entropy Minimization	目标因子分析 Target factor analysis	目标波段熵最小化 Band-target entropy minimization	目标因子分析 Target factor analysis	目标波段熵最小化 Band-target entropy minimization
1	0.8328	0.7682	0.9476	0.9058	0.9859	0.9408
2	0.6640	0.6324	0.8666	0.8531	0.3349	0.3271
3	0.9129	0.8995	0.8437	0.7689	0.9956	0.9775
4	0.7086	0.6727	0.9541	0.8835	0.8853	0.8832
5	0.7861	0.7075	0.8108	0.7535	0.9933	0.9805

参考光谱与重建光谱的相似程度,可以作为判断依据。但对于数据集5,表2中O₃的相关系数只有0.7861,但在图3C中却可以看到O₃的特征峰,这是因为光谱中噪声很大,并且NH₃和水蒸气干扰较强,所以重建光谱中虽然能看到O₃的光谱特征,但是相关系数值却比0.85小。

5 结 论

TFA方法是大气OP/FT-IR光谱定性分析的重要工具,但是BTEM方法应用于OP/FT-IR光谱定性分析还未见报道。将BTEM方法引入OP/FT-IR光谱中,相对于TFA方法,BTEM方法具有无需目标分子的先验信息的特点,为实现多组分OP/FT-IR光谱定性分析提供了新的辨别方法,为OP/FT-IR光谱的定性分析发展拓展了空间。

References

- 1 LIU Wen-Qing, CUI Zhi-Cheng, DONG Feng-Zhong. *Optoelectronic Technol. Inform.*, **2002**, 15: 1-12
刘文清,崔志成,董凤忠. *光电子技术与信息*, **2002**, 15: 1-12
- 2 SUN Xiu-Yun, LI Yan, WANG Jun-De. *Spectroscopy and Spectral Analysis*, **2003**, 23(4): 739-741
孙秀云,李燕,王俊德. *光谱学与光谱分析*, **2003**, 23(4): 739-741
- 3 XU Liang, LIU Jian-Guo, LU Yi-Huai, LIU Wen-Qing. *Spectroscopy and Spectral Analysis*, **2007**, 27(5): 889-891
徐亮,刘建国,陆亦怀,刘文清. *光谱学与光谱分析*, **2007**, 27(5): 889-891
- 4 WANG Yan, ZHANG Yan-Chao, LI Ning, ZHAO Xue-Hong. *Opto-Electronic Engineering*, **2008**, 35: 73-81
汪曦,张艳超,李宁,赵学珺. *光电工程*, **2008**, 35: 73-81
- 5 Newman A R. *Anal. Chem.*, **1997**, 69: 43A-47A
- 6 Walter W T. *Proc. SPIE.*, **2004**, 5270: 144-150 (Environmental Monitoring and Remediation III)
- 7 Childers J W, Thompson E L, Harris D B, Kirchgessner D A, Clayton M, Natschke D F, Phillips W J. *Atmos. Environ.*, **2001**, 35: 1923-1936
- 8 Wiseman J M, Ifa D R, Song Q Y, Cooks R G. *Angew. Chem. Int. Ed.*, **2006**, 45(43): 7188-7192
- 9 van Zomeren P V, Metting H J, Coenegracht P M J, de Jong G J. *J. Chromatogr. A*, **2005**, 1096: 165-176
- 10 Carvalho A R, Wattoom J, Zhu L, Brereton R G. *Analyst*, **2006**, 131: 90-97
- 11 Day J P R, Campbell R A, Russell O P, Bain C D. *J. Phys. Chem. C*, **2007**, 111(25): 8757-8774
- 12 Kuzmanovski I, Trpkovska M, Soptrajanov B, Stefov V. *Vib. Spectrosc.*, **1999**, 19: 249-253
- 13 Tetteh J, Mader K T, Andanson J M, McAuley W J, Lane M E, Hadgraft J, Kazarian S G, Mitchell J C. *Anal. Chim. Acta*, **2009**, 642(1-2): 246-256
- 14 Kauffmana J F, Dellibovi M, Cunninghamb C R. *J. Pharm. Biomed. Anal.*, **2007**, 43: 39-48
- 15 Monfreda M, Gobbi L, Grippa A. *Food. Chem.*, **2014**, 145: 584-592
- 16 Widjaja E, Li C Z, Garland M. *J. Comb. Chem.*, **2009**, 11(2): 261-266
- 17 Widjaja E, Lee W L, Loo S C J. *Anal. Chem.*, **2010**, 82(4): 1277-1282
- 18 Tjahjono A M, Feng G, Hermanto M W. *RSC Adv.*, **2014**, 4: 22194-22204
- 19 Cheng S, Li C, Guo L. *Vib. Spectrosc.*, **2014**, 70: 110-114
- 20 Guo L F, Wiesmath A, Sprenger P, Garland M. *Anal. Chem.*, **2005**, 77(6): 1655-1662
- 21 Zhang H J, Garland M, Zeng Y Z, Wu P. *J. Am. Soc. Mass. Spectrom.*, **2003**, 14(11): 1295-1305
- 22 Widjaja E, Li C Z, Chew W, Garland M. *Anal. Chem.*, **2003**, 75(17): 4499-4507
- 23 Shao L, Griffiths P R. *Anal. Chem.*, **2008**, 80(13): 5219-5224
- 24 Malinowski E R. *Factor Analysis in Chemistry*, 3rd ed.; John Wiley and Sons. New York, **2002**
- 25 Shannon C E. *E. Bell. Sys. Tech. J.*, **1948**, 27: 379-623
- 26 Widjaja E, Garland M. *J. Comput. Chem.*, **2002**, 23(9): 911-919

