

## 基于连续小波变换的 FTIR 光谱拟合算法

姜安<sup>1,2</sup>, 彭江涛<sup>1,2</sup>, 王怀松<sup>1,2</sup>, 彭思龙<sup>3</sup>, 谢启伟<sup>3\*</sup>

1. 南京财经大学管理科学与工程学院, 江苏 南京 210046
2. 江苏省质量安全工程研究院, 江苏 南京 210046
3. 中国科学院自动化研究所, 北京 100190

**摘要** 提出了一种基于连续小波变换的 FTIR 光谱拟合算法。在计算差减尺度因子时, 同时考虑原始谱图及其连续小波变换谱图, 从光谱最小二乘拟合的角度求解, 克服了常规差谱算法中的参考峰及差减因子的人工选择问题。采用六种不同的小波进行光谱拟合, 用计算得到的差减因子来定量酒精度, 误差绝对值的平均值仅为  $0.047^\circ \sim 0.072^\circ$ , 误差标准差仅为  $0.056^\circ \sim 0.091^\circ$ 。实验结果表明, 连续小波变换结合最小二乘拟合的光谱拟合模型能为 FTIR 差谱提供一种准确可靠的新方法。

**关键词** 连续小波变换; 红外光谱; 光谱拟合; 差减因子

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2012)09-2385-04

### 引言

差谱方法是分析未知复杂混合物的重要手段之一<sup>[1,2]</sup>。该方法从混合物光谱中减去已知物质成分的光谱, 然后再对得到的差谱做进一步分析。差谱方法能够简化光谱并分离重叠谱带, 有助于提取目标光谱并突出特征峰, 因此对差谱方法的研究十分必要。差谱分析方法的理论依据是朗伯-比尔定律, 谱峰高度与物质的浓度成正比, 且谱峰位置能反映分子结构的特点, 因此可以通过减去混合物中某一特定位置(已知成分)的已知峰高的谱峰实现待测成分的含量解析。

常用的差谱方法是吸光度光谱差减法, 它在混合物光谱中寻找一个吸收峰作为参考峰, 人工调节差减因子, 直到将参考峰全部减掉为止。吸光度光谱差减法多用于混合溶液的去水实验<sup>[3-6]</sup>, 如蛋白质光谱去水实验, 盐溶液中的去水问题<sup>[3]</sup>等。吸光度光谱差减法经验性太强, 通常需要人工选择参考峰, 并手工调节差减因子。实际应用中会受到谱峰变形、谱峰重叠、基线漂移等因素的影响<sup>[7,8]</sup>, 难以得到理想的谱峰以及合适的差减因子, 差谱结果通常不稳定、不可靠。

鉴于传统差谱方法中差减因子难以求取, Din 等<sup>[9,10]</sup>提出了一种基于连续小波变换的差谱方法, 将混合物光谱表示为不同成分光谱的线性组合, 对原始光谱进行连续小波变换, 借助小波变换后的谱图的过零点, 求取线性组合系数,

该系数相当于差谱方法中的差减因子。但是该方法的精度依赖于小波变换后过零点计算的准确性, 原始信号的质量会对小波变换结果产生影响, 影响过零点计算。同时该方法主要应用于成分比较简单的两组分混合物的定量分析, 当成分复杂时, 组分之间相互作用会影响过零点计算。

本工作将原光谱图的连续小波变换谱图引入差减因子的计算过程, 提出基于连续小波变换的 FTIR 光谱拟合算法, 将差谱问题转换为光谱拟合问题, 通过对原始谱图以及连续小波变换谱图同时进行最小二乘拟合<sup>[11]</sup>, 自动求取差谱算法中的差减因子。若对计算得到的差减因子进行归一化, 可近似得到复杂混合物中各组分对应的含量, 从而还可将该算法用于复杂混合物组分的定量分析。

### 1 基于连续小波变换的光谱拟合方法

连续小波变换(continuous wavelet transform, CWT)

$$W_{\Psi} f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \Psi^* \left( \frac{t-b}{a} \right) dt \quad (1)$$

其中, 尺度因子  $a \in R$  且  $a \neq 0$ ,  $b$  为平移因子。 $\Psi$  为小波基函数。

根据朗伯-比尔定律, 红外光谱的吸收具有加和性, 即混合物在某波长处的吸收是其所含组分在该波长处吸收的叠加。假定混合物中含有  $n$  种组分, 将混合物光谱记为  $S$ , 各

收稿日期: 2012-02-28, 修订日期: 2012-06-10

基金项目: 国家自然科学基金项目(61101219, 61032007)资助

作者简介: 姜安, 1979年生, 江苏省质量安全工程研究院博士研究生 e-mail: an.jiang@ia.ac.cn

\*通讯联系人 e-mail: qw\_xie@yahoo.com.cn

组分的光谱分别记为  $S_1, S_2, \dots, S_n$ , 则混合物光谱减去各组分光谱所得的差谱为

$$f = S - \lambda_1 S_1 - \lambda_2 S_2 - \dots - \lambda_n S_n \quad (2)$$

式中,  $\lambda_1, \lambda_2, \dots, \lambda_n$  为各组分的差减因子。

考虑对原始谱图信号进行连续小波变换, 将得到不同尺度下的连续小波变换谱图, 其中小尺度对应的是原始谱图窄峰等细节部分, 大尺度对应的是原始谱图宽峰及趋势项部分。基于连续小波变换的光谱拟合算法在谱图拟合过程中, 综合考虑原始光谱信号以及不同尺度下的连续小波变换谱图来求取差减因子。其原理如下: 选取适当的小波基函数, 对混合物光谱  $S$  及其各组分光谱  $S_1, S_2, \dots, S_n$  分别做连续小波变换, 得到  $C, C_1, C_2, \dots, C_n$ 。算法可用下式表达

$$\arg \min \{ \| S - \lambda_1 S_1 - \dots - \lambda_n S_n \|^2 + \sum_{i=1}^p \omega_i \| C^{(i)} - \lambda_1 C_1^{(i)} - \dots - \lambda_n C_n^{(i)} \|^2 \} \quad (3)$$

式中,  $C^{(i)}$  表示  $C$  的第  $i$  个尺度谱图,  $\omega_i$  为不同尺度下连续小波变换谱图的权重系数, 起到折中原始谱图和连续小波变换谱图的作用。

求解上式, 得差减因子  $\lambda$  的最优解为

$$\hat{\lambda} = A^{-1} B \quad (4)$$

其中,

$$A = (A_{r,k})_{r,k=1}^n, A_{r,k} = \langle S_r, S_k \rangle + \sum_{i=1}^p \omega_i \langle C_r^{(i)}, C_k^{(i)} \rangle \quad (5)$$

$$B = (B_1, \dots, B_n), B_k = \langle S, S_k \rangle + \sum_{i=1}^p \omega_i \langle C, C_k^{(i)} \rangle$$

最后拟合出的复杂混合物光谱及拟合误差分别为

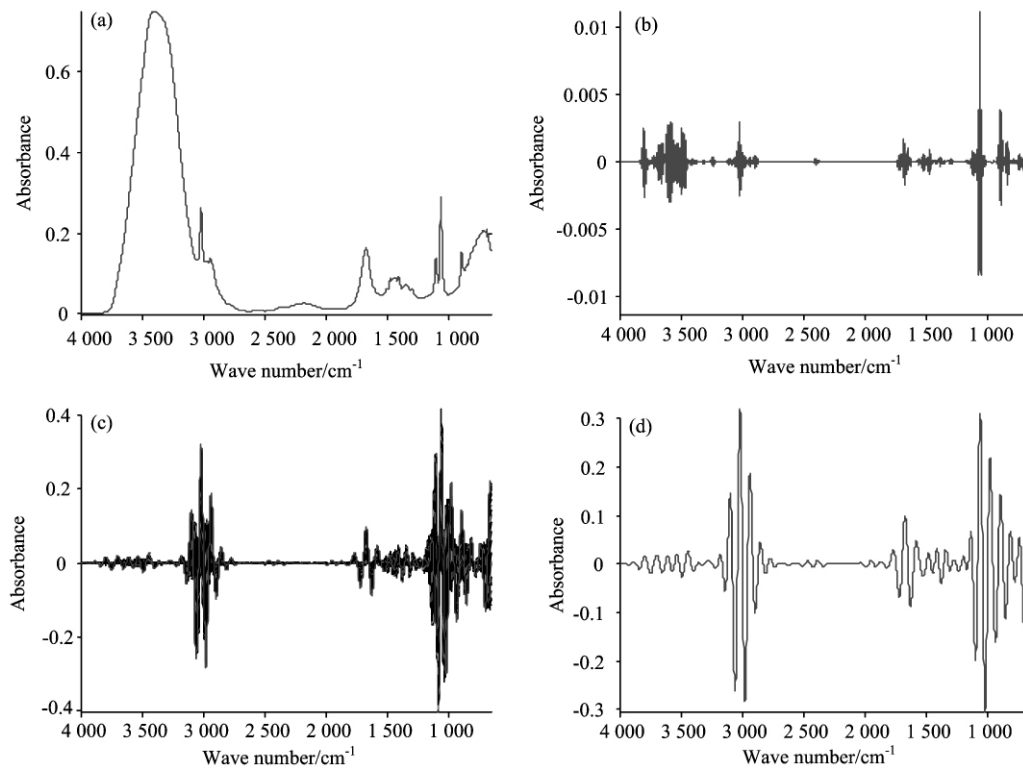


Fig 2 The infrared spectrum of the mixture and its continuous wavelet transform spectra

- (a): Spectrum of the mixture; (b): CWT spectrum with scale of 8;  
(c): CWT spectra with all scales; (d): CWT spectrum with scale of 64

$$\hat{S} = \sum_{i=1}^n \hat{\lambda}_i S_i, \hat{f} = S - \hat{S} = S - \sum_{i=1}^n \hat{\lambda}_i S_i \quad (6)$$

## 2 仿真实验分析

以水、乙醇、酒水混合物为例(酒水混合物由水和乙醇配制而成, 如图 1 所示), 来说明基于连续小波变换的光谱拟合算法的优点。算法优点如下:

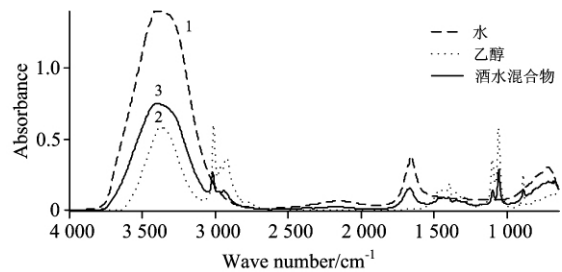


Fig 1 Spectra of water, alcohol and their mixture

(1) 算法对水的吸收不敏感。光谱图中水的既强且宽的吸收峰在连续小波变换谱图中的幅度远小于其他吸收峰, 因此在光谱拟合时受水吸收峰的影响会更小, 拟合的效果也更好。

选用 Morlet 小波, 酒水混合物及其不同尺度的小波变换谱如图 2, 其中(a)是酒水混合物的光谱图, (b)是其尺度为 8 的 CWT 谱图, (c)是其在尺度为 8, 10, 12, ..., 60, 62, 64 总共 29 个尺度上的 CWT 谱图叠加显示的效果, 图(d)是

尺度为 64 的 CWT 谱图。从图中可以看出, 相对于别的吸收峰, 3 400 cm<sup>-1</sup> 附近水的大吸收峰对连续小波变换谱图的影响是非常有限的。

(2) 对基线漂移不敏感。小波具有消失矩特性, 如墨西哥草帽小波就是高斯函数的二阶导数, 而导数的方法可以很好地去除光谱中的基线漂移, 这样光谱图中基线的线性漂移不会影响到光谱信号连续小波变换的结果。

(3) 拟合精度更高。基于最小二乘的光谱拟合算法只是在原始谱图上去做拟合, 其过程可以理解为一维信号的匹配; 而基于连续小波变换的 FTIR 光谱拟合算法可以理解为一个二维信号的匹配过程, 不但要求在原始谱图上有好的拟合效果, 而且要求在不同尺度下连续小波变换得到的谱图上也要有好的拟合效果, 因此理论上会比原始光谱信号上的匹配更精确。

### 3 结果与讨论

#### 3.1 实验数据

用乙醇和水配成已知酒精度的酒水混合物样品 10 个, 度数范围 44°~53°不等, 采集其红外谱图。实验采用的仪器为 Perkin Elmer 公司的 Spectrum GX FTIR 红外光谱仪, DTGS 检测器, ATR 附件(ZnSe cell); 谱图采集条件: 扫描信号累加 16 次, 光谱分辨率 4 cm<sup>-1</sup>, 测量范围 4 000~650 cm<sup>-1</sup>, ATR 反射 3 次。

#### 3.2 实验参数选择

基于连续小波变换的光谱拟合算法中, 小波权重系数  $w_i$  起到折中原始谱图和连续小波变换谱图的作用, 实验中取所有的  $w_i$  相等。小波尺度不能选择过小, 太大的尺度对应的连续小波变换谱图上会有大量噪声; 小波尺度也不能选择过大, 过大的尺度对应的连续小波变换谱图上呈现出的基本上是一些低频缓变信号, 而在这些缓变信号上做拟合很多情况下没有太多实际意义。通过实验, 选取 8, 10, 12, ..., 60, 62, 64 作为小波变换的尺度。小波基函数可以根据具体问题灵活选择。目前常用的小波基函数有 Haar, Morlet, Meyer, Daubechies, Gaussian, Symlets 小波等。

#### 3.3 实验评价指标

根据光谱拟合得到的差减尺度因子, 计算酒水混合物中

的乙醇含量  $d$  (即酒精度) 作为算法评价指标。乙醇含量  $d$  可表示为

$$d = 100 \times \hat{\lambda}_1 / (\hat{\lambda}_1 + \hat{\lambda}_2) \quad (7)$$

其中  $\hat{\lambda}_1$  和  $\hat{\lambda}_2$  分别为乙醇和水对应的差减因子。

#### 3.4 不同小波基函数的拟合结果

比较 gaus1, db5, sym5, Haar, Meyer, Morlet 等 6 种不同小波对定量分析酒水混合物中乙醇含量(酒精度)的影响, 结果如图 3 所示。不同小波基函数对应的拟合效果有所不同。这六种小波的误差绝对值的平均值范围为 0.047°~0.072°, 误差标准差的范围为 0.056°~0.091°。

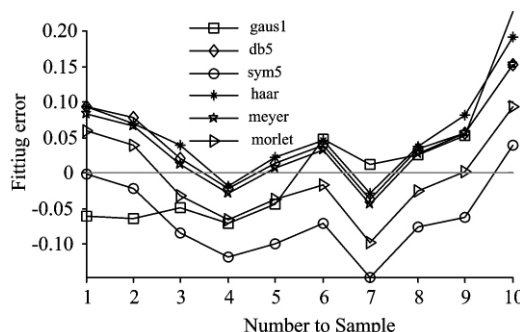


Fig 3 Fitting error of different wavelet

若只以平均绝对误差这一个参数来衡量不同小波基函数的拟合效果, 按拟合效果从好到差, 分别是 Morlet, Meyer, db5, Haar, gaus1 和 sym5 小波。拟合效果较好的 Morlet 小波, 误差绝对值的平均值为 0.047°, 最大值也不超过 0.1°, 最小值仅为 0.002°; 拟合效果较差的 sym5 小波, 误差绝对值的平均值为 0.072°, 最大值为 -0.146°, 最小值为 -0.001°。

#### 3.5 不同算法的对比结果

将所提出的算法与现有参考峰方法以及 FTIR 光谱拟合算法进行对比。参考峰法是现有商业软件常用的方法。此处的做法是人工选定 3 010 cm<sup>-1</sup> 附近乙醇的吸收峰作为参考峰, 根据该参考峰在乙醇以及酒水混合物光谱图中的强度来计算酒水混合物的酒精度。同时, 利用 FTIR 光谱拟合算法计算酒水混合物的酒精度。三种算法定量效果对比如表 1 所示。

Table 1 Quantitative comparison of different algorithms

样本编号	真实酒精度/(°)	参考峰法		最小二乘法		本文算法(采用 Morlet 小波)	
		计算值/(°)	误差	计算值/(°)	误差	计算值/(°)	误差
1	44.143	44.617	0.474	43.756	-0.387	44.202	0.059
2	45.086	45.542	0.456	44.875	-0.212	45.127	0.040
3	46.170	46.604	0.434	46.187	0.017	46.139	-0.031
4	47.140	46.313	-0.826	47.092	-0.047	47.073	-0.066
5	48.129	46.896	-1.234	48.072	-0.057	48.093	-0.037
6	49.185	46.665	-2.520	49.153	-0.031	49.169	-0.016
7	50.062	48.906	-1.156	50.248	0.186	49.964	-0.098
8	51.185	49.633	-1.552	51.184	-0.001	51.160	-0.025
9	52.186	48.241	-3.944	52.281	0.096	52.187	0.002
10	53.221	48.673	-4.548	53.187	-0.035	53.315	0.093
平均绝对误差		1.714		0.107		0.047	
误差标准差		1.776		0.158		0.058	

从表 1 可以看出, 参考峰法计算出的酒精度误差较大, 共有 6 个样本的误差绝对值大于  $1^\circ$ 。误差绝对值的平均值达到了  $1.714^\circ$ , 误差标准差达到了  $1.776^\circ$ 。相对于参考峰法, 最小二乘法的定量误差降了一个数量级。误差绝对值大于  $0.1^\circ$  的样本有 3 个, 最大的误差来自 1 号样本, 为  $-0.387^\circ$ , 最小的误差来自 8 号样本, 为  $-0.001^\circ$ 。误差绝对值的平均值为  $0.107^\circ$ , 误差标准差为  $0.158^\circ$ 。

本文所提算法, 误差绝对值最大值不超过  $0.1^\circ$ , 为 10 号样本的  $0.093^\circ$ ; 误差绝对值最小的是 9 号样本, 仅为  $0.002^\circ$ 。误差绝对值的平均值仅为  $0.047^\circ$ , 误差标准差仅为  $0.058$ , 拟合效果明显优于参考峰法和最小二乘法。即便选择拟合效果较差的 sym5 小波, 其误差绝对值的平均值也仅为  $0.072^\circ$ , 误差标准差仅为  $0.056^\circ$ , 其拟合效果依然优于参考峰法和最小二乘法。

本算法与最小二乘法对酒水混合物进行光谱拟合, 得到的拟合误差图见图 4。从图中可以看出, 本算法误差(虚线所示)比最小二乘拟合算法误差减小了很多。这也进一步说明, 基于连续小波变换的光谱拟合算法综合考虑了原始谱图和其 CWT 谱图, 算法精度更高。

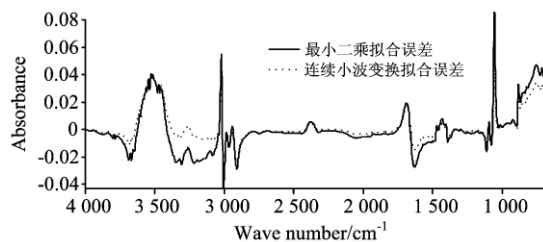


Fig 4 Fitting error comparison of the two algorithms

## 4 结 论

针对现有差谱算法的不足, 提出了基于连续小波变换的 FTIR 光谱拟合算法。分析了算法的原理, 论证了算法的优点, 并在实验中详细讨论了小波权重、尺度、阶数以及基函数选择等关键技术问题。该算法从光谱拟合的角度巧妙地求出差谱算法中的差减因子, 有效地解决了商业软件中差减因子人工选择的问题。同时, 该算法具有对水的吸收和基线漂移不敏感的优点, 且拟合精度较高。

## References

- [1] Kozlov D, Besov A. *Applied Spectroscopy*, 2011, 65(8): 918.
- [2] YUAN Yue-ming, XIONG Wei, FANG Yong-hua, et al(袁越明, 熊伟, 方勇华, 等). *Infrared and Laser Engineering(红外与激光工程)*, 2011, 40(3): 408.
- [3] Max J, Trudel M, Chapados C. *Applied Spectroscopy*, 1998, 52(2): 234.
- [4] Yang J, Zhang D, Li Y. *Beilstein J Nanotechnol*, 2011, 2: 653.
- [5] Spexard M, Immeln D, Thöing C. *Vibrational Spectroscopy*, 2011, 57(2): 282.
- [6] Simpson J V, Oshokoya O, Wagner N. *Analyst*, 2011, 136: 1239.
- [7] LI Cui-ping, HAN Jiu-qiang, HUANG Qi-bin, et al(李翠萍, 韩九强, 黄启斌, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2011, 31(11): 3050.
- [8] Liland K H, Rukke E O, Olsen E F. *Chemometrics and Intelligent Laboratory Systems*, 2011, 109: 51.
- [9] Din E, Baleanu D. *Journal of the Brazilian Chemical Society*, 2007, 18: 962.
- [10] Din E, Baleanu D. *Mathematical Methods in Engineering*, 2007, 4: 265.
- [11] Yamamoto T, Nanbu F, Tanaka T. *Anal. Chem.*, 2011, 83(5): 1681.

# FTIR Spectral Fitting Algorithm Based on Continuous Wavelet Transform

JIANG An<sup>1,2</sup>, PENG Jiang-tao<sup>1,2</sup>, WANG Huai-song<sup>1,2</sup>, PENG Si-long<sup>3</sup>, XIE Qi-wei<sup>3\*</sup>

1. School of Management Science & Engineering, Nanjing University of Finance & Economics, Nanjing 210046, China

2. Jiangsu Province Institute of Quality and Safety Engineering, Nanjing 210046, China

3. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

**Abstract** An FTIR spectrum fitting algorithm based on continuous wavelet transform is proposed. In calculating the factor of difference spectrum, the algorithm takes into account both the original spectrum and its continuous wavelet transformed spectra, which effectively overcomes the problem of reference peak selection and manual factor selection in most commercial software. The detailed discussions on wavelet scale, order and basis are included. The spectral fitting is performed on six wavelet basis functions and the obtained scale factor is used to quantify the content of liquor, and the corresponding mean absolute error ranges from 0.047 to 0.072, and the standard deviation ranges from 0.056 to 0.091. Experimental results show that the CWT combined with least squares fitting provides an accurate and reliable new method for FTIR spectral subtraction.

**Keywords** Continuous wavelet transform; Infrared spectroscopy; Spectral fitting; Scale factor

\* Corresponding author

(Received Feb. 28, 2012; accepted Jun. 10, 2012)