

文章编号: 0253-2239(2010)01-0267-05

# 基于经验模态分解的近红外光谱预处理方法

蔡剑华<sup>1,2</sup> 王先春<sup>1</sup>

(<sup>1</sup> 湖南文理学院信息研究所, 湖南 常德 415000; <sup>2</sup> 中南大学信息物理工程学院, 湖南 长沙 410083)

**摘要** 针对微分法在有效消除光谱背景和基线漂移的同时会增加光谱噪声的问题,把最新发展的经验模态分解方法(EMD)引入到近红外光谱处理中来,以烟草的一阶导数近红外(NIR)光谱为研究对象,探讨经验模态分解在近红外光谱预处理中的应用,并与小波变换消噪效果进行了对比分析。结果表明,用基于经验模态分解去噪后的光谱进行分析,预测集的决定系数  $r^2$  由去噪前的 0.9705 提高到 0.9832,均方根误差(RMSEP)由去噪前的 0.5606 降为 0.3310,比基于小波变换的分析结果略高。因此,经验模态分解方法对消除光谱的噪声是有效的,有效地提高了光谱的分析精度和模型的稳定,为近红外光谱预处理提供了一种新方法。

**关键词** 光谱学;近红外光谱预处理;经验模态分解

中图分类号 O657.1 文献标识码 A doi: 10.3788/AOS20103001.0267

## Near-Infrared Spectrum Pretreatment Based on Empirical Mode Decomposition

Cai Jianhua<sup>1,2</sup> Wang Xianchun<sup>1</sup>

<sup>1</sup> Information Institute, Hunan University of Arts and Science, Changde, Hunan 415000, China

<sup>2</sup> School of Info-Physics and Geomatics Engineering, Central South University, Changsha, Hunan 410083, China

**Abstract** The derivative method can correct baseline effects but also add noise to it. The empirical mode decomposition (EMD) method was proposed to get the pretreatment of near-infrared (NIR) spectrum. Herein, the first derivative NIR spectrum of tobacco was served as the target and the application of EMD in NIR spectrum pretreatment was studied. Experimental results showed that the result after de-noising with EMD was satisfactory. The correlation ratio of the prediction set was improved from 0.9705 to 0.9832, and the RMSE reduced from 0.5606 to 0.3310. And this method obtained a better de-noising effect compared to wavelet transform method. It is concluded that EMD is a useful method to eliminate noise of NIR signals, which makes the final model more representative, stable and robust. EMD provides a new method for near-infrared spectrum pretreatment.

**Key words** spectroscopy; near-infrared spectrum pretreatment(NIR); empirical mode decomposition(EMD)

## 1 引言

在近红外(NIR)分析中,原始光谱中含有与样品组成无关的信息,如环境温度、电噪声、样品背景干扰等<sup>[1,2]</sup>。光谱预处理在近红外光谱分析技术中是相当重要的,预处理方法的选取对所建立光谱模型的预测精度和稳定性具有很重要的意义。光谱的一阶微分可以去除部分线性或接近线性的背景和噪声对目标光谱的影响<sup>[2,3]</sup>。但是,原始光谱经微分后,噪声增大,所以更需要有效的消噪处理。许多文献研究了小波变换在这领域的应用,并显示了小波

变换良好的应用效果<sup>[1~4]</sup>。经验模式分解(EMD)方法是 1998 年由美国工程院院士 Norden E. Huang<sup>[5]</sup>提出的,已成为众多学科关注的热点,在通讯、医学、地质等领域获得了极其广泛的应用<sup>[6~11]</sup>。本文以烟草的一阶导数近红外光谱为研究对象,把经验模态分解方法引入到近红外光谱处理中来,旨在探索近红外光谱预处理的新方法。

## 2 经验模态分解去噪原理

EMD 方法即 Huang 变换,它依据信号本身的

收稿日期: 2009-02-10; 收到修改稿日期: 2009-03-30

基金项目: 国家 863 计划(2006AA06Z105)和湖南省“十一五”重点建设学科光学基金资助课题。

作者简介: 蔡剑华(1979—),男,讲师,主要从事光电信息处理方面的研究。E-mail: cjh1021cjh@163.com

时间尺度特征,将信号分解为含有不同时间尺度且满足以下两个定义条件的一组本征模态函数(IMF)<sup>[5,6]</sup>: 1) 对于一系列数据,极值点和过零点数目必须相等或至多相差一点;2) 在任意点,由局部极大点和极小点构成的两条包络线的平均值为零。每个 IMF 可以被认为是信号中固有的一个模态函数。

由 EMD 分解得到 IMF 的算法可说明如下<sup>[5,7,8]</sup>: 1) 找出信号  $x(t)$  的所有极大点和极小点,将其用三次样条函数分别拟合为原数据序列的上、下包络线,上、下包络线的均值为平均包络线  $m_1$ ,将原数据序列减去  $m_1$  可得到一个去掉低频的新数据序列  $h_1$ 。一般  $h_1$  不是一个平稳数据序列,为此重复以上过程  $n$  次,使所得到的平均包络线趋于零,此时  $h_{1n}$  就是第一个  $f_{\text{IMF}}(c_1)$ ,它表示信号数据中的最高频率成分。2) 用  $x(t)$  减去  $c_1$  得到一个去掉高频成分的新数据序列,重复步骤 1),得到一系列  $c_n$  和最后一个序列  $r_n$ ,  $r_n$  代表  $x(t)$  的均值或趋势项。那么原序列  $x(t)$  表示为 IMF 分量和一个残余项的和:

$$x(t) = \sum_{i=1}^n c_i + r_n, \quad (1)$$

其中分解出的  $n$  个分量  $c_i$  分别包含了信号从高频到低频的不同频率段成分。信号经过 EMD 分解后的各阶模态函数能够完全重构,几乎没有能量损失。这样 EMD 就可用来减少和消除信号中混杂的噪声。当噪声为一个或多个 IMF 分量时,可直接利用时空滤波器进行消噪,如一个包含  $n$  个 IMF 的低通时空滤波器可表示为<sup>[9~11]</sup>

$$x_{lk}(t) = \sum_{i=k}^n c_i(t) + r_n(t), \quad (2)$$

高通时空滤波器为

$$x_{hk}(t) = \sum_{i=1}^k c_i(t). \quad (3)$$

若噪声和信号混叠,可设置硬(或软)门限去噪,其过程类似小波变换中的方法<sup>[12~14]</sup>。本文采用的阈值方法为:给定信号  $x(t)$  经 EMD 分解后得到  $N$  个 IMF,对每一层 IMF 选取一合适的阈值,并用此阈值对  $c_i$  进行截断为  $\hat{c}_i$ ,然后再进行 EMD 的重构

$$x(t) = \sum_{i=1}^n \hat{c}_i(t) + r_n, \quad (4)$$

根据 Donoho 等<sup>[15]</sup>的理论给出消除噪声的阈值为

$$\begin{aligned} \tau_i &= \sigma_i \sqrt{2 \lg n}, \\ \sigma_i &= (x_{\text{MAD}})_i / 0.6745, \end{aligned} \quad (5)$$

其中  $\sigma_i$  为第  $i$  层 IMF 的噪声水平;  $(x_{\text{MAD}})_i$  为第  $i$  层 IMF 的绝对中值偏差,且定义为

$$(x_{\text{MAD}})_i = Q_{\frac{1}{2}}(|c_i(t) - Q_{\frac{1}{2}}(c_i)|), \quad (6)$$

其中  $Q_{\frac{1}{2}}$  为中位数(Median)。第  $i$  层 IMF 的重构值  $\hat{c}_i$  为

$$\hat{c}_i(t) = \begin{cases} \text{sign}[c_i(t)] [|c_i(t) - \tau_i|], & |c_i| \geq \tau_i \\ 0, & |c_i| < \tau_i \end{cases} \quad (7)$$

## 3 实验部分

### 3.1 样品和仪器

样品:青烟粉末,共计 87 个样品。其中 67 个作为校正集,20 个为验证集。由田间采摘的新鲜烟叶切丝,直接微波干燥,使样品的水分含量基本保持在 7% 左右,再用 XF-98B 型旋风精密粉碎机充分粉碎、研磨均匀,得到 80 目的粉末样品,进行近红外测试和总糖含量的测定。

仪器为美国 BRIMROSE 公司产的 Luminar 5030 型 Mini-AOTF/(NIR)便携式近红外光谱仪。

### 3.2 光谱的采集

将样品置于样品池中,用与之配套的砝码压实样品,以保证样品装载的紧密度、厚度相同。所用检测仪器为砷化镓(InGaAs)光检测器,波长范围为 1100~2300 nm,1 nm 的波长增量,分辨率 2 nm,扫描次数为 600。

## 4 结果与讨论

图 1 为 87 个烟草样品原始光谱图和其对应的一阶导数光谱图。从图 1(a)可以看出,原始光谱图非常光滑,噪声很低信噪比高,但是背景光谱非常强,基线漂移较严重,因此分析偏差大。经过一阶微分处理后,基本上扣除了原始谱图的背景光谱,说明微分对于扣除基线漂移具有很好的效果,但微分谱图光滑性较原始光谱差,其信噪比降低显著。因此,需要采用合适的方法进行去噪处理。

### 4.1 基于 EMD 的一阶导数光谱去噪处理

所有数据预处理均在 Matlab 软件平台上编程完成。图 2 为 2 号烟草样品的一阶导数光谱图,可以看出,求导使随机误差也被放大。采用 EMD 方法对光谱进行消噪处理,光谱经 EMD 分解后自适应的得到 4 阶模态函数(图 3),每一个 IMF 分量都有不同的振幅和频率,分解顺序是按频率从高至低自适应进行的。 $c_1$  表现为信号包含的白噪声, $c_2, c_3$  为信号的高频分量, $r$  为光谱的优势频率子频带。根据(4)~(7)式,先对 EMD 分解后导数光谱的高频部分  $c_1, c_2$  和  $c_3$  计算各分量的阈值,分别为 0.0002,

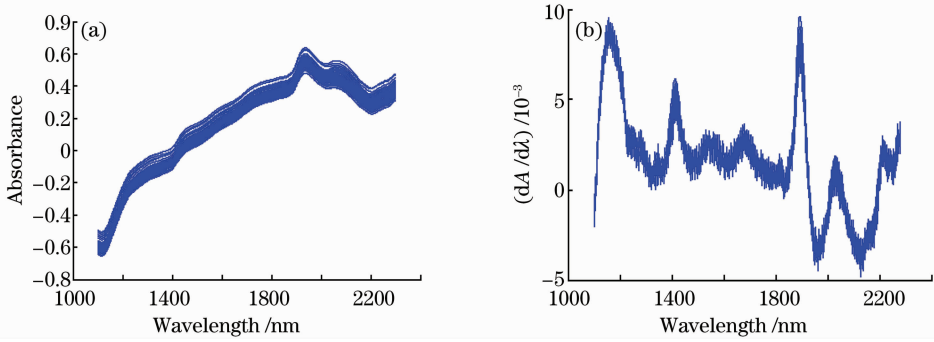


图 1 烟草样品的近红外光谱:(a) 原始光谱, (b) 一阶导数谱

Fig. 1 NIR spectrum of tobacco: (a) raw spectrum, (b) first derivative spectrum

0.0003 和 0.0003, 再以此阈值对分解系数进行修正, 最后重构得到消噪后的光谱(图 4)。对比图 2 和图 4, 重构后导数光谱的噪声基本得到消除, 凸出了被噪声掩盖的峰值信息。为作比较, 图 5 给出了 87 个样品一阶导数光谱经 EMD 去噪和小波软阈值去噪后的的光谱图。小波去噪中母小波采用扩展性较好的 db8 小波, 阈值的确定采用 Rigsure 方式, 分解水平为 4。显然, 两种方法都有很强的去噪能力, 与图 1(b) 相比, 去噪后谱图噪音明显降低, 光谱的平滑性增强, 且对光谱的峰形没有影响。

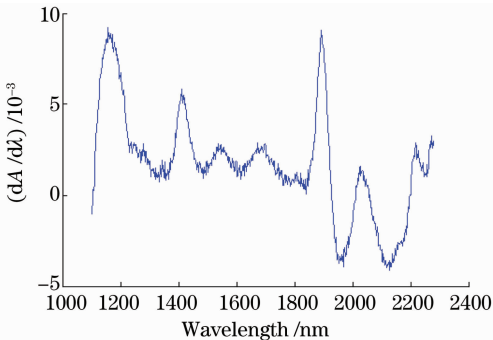


图 2 2号烟草样品的一阶近红外光谱

Fig. 2 First derivative NIR spectrum of No. 2 tobacco

### 4.2 处理结果对比

将经过处理后的光谱数据与烟草中的总糖等化学成分的基础数据进行关联, 采用偏最小二乘法(PLS1), 交叉验证法(cross validation), 用 The Unscrambler 定量分析软件建立模型。表 1 列出了分别利用 9 点一阶微分、9 点一阶微分加小波去噪、9 点一阶微分加 EMD 去噪 3 种预处理方式烟草近红外光谱分析总糖含量结果与化学值的比较。化学检测方法采用流动注射分析法(YC/T 159-2002)。结果表明, EMD 去噪后分析结果的准确性明显得到改善, 预测集的决定系数  $r^2$  由原来的 0.9705 提高到 0.9832, 均方根误差(RMSE)由原来的 0.5606

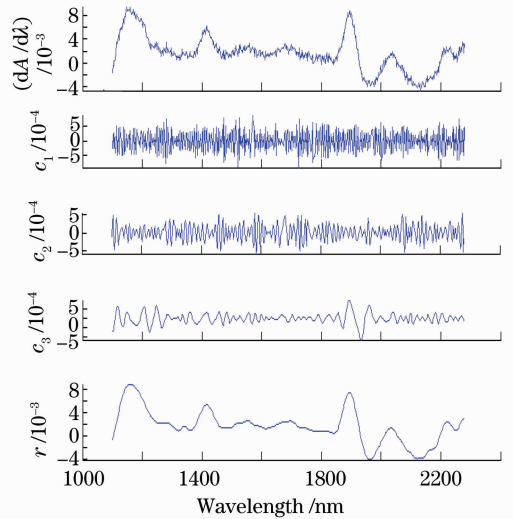


图 3 2号烟草样品光谱的 EMD 分解

Fig. 3 EMD decomposition of No. 2 tobacco spectrum

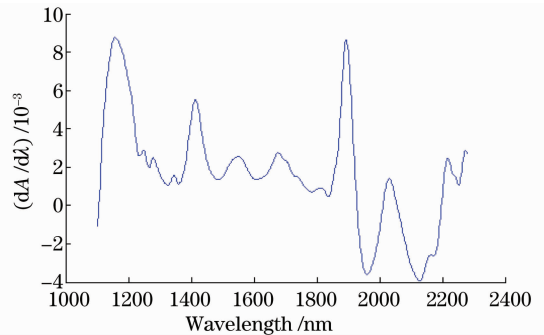


图 4 EMD 去噪后的 2号烟草样品光谱

Fig. 4 De-noised spectrum of No. 2 tobacco by EMD

降为 0.3310, 有效的提高了光谱的分析精度和模型的稳健性。与小波变换去噪后的分析精度比较表明, 两种预处理方式所得结果的分析精度都很高, EMD 方法略高。可见, EMD 预处理方法是有效的, 且与小波变换方法相比, EMD 分解是完全自适应的, 这就避免了选择母小波的麻烦, 增加了分析结果的可靠性。

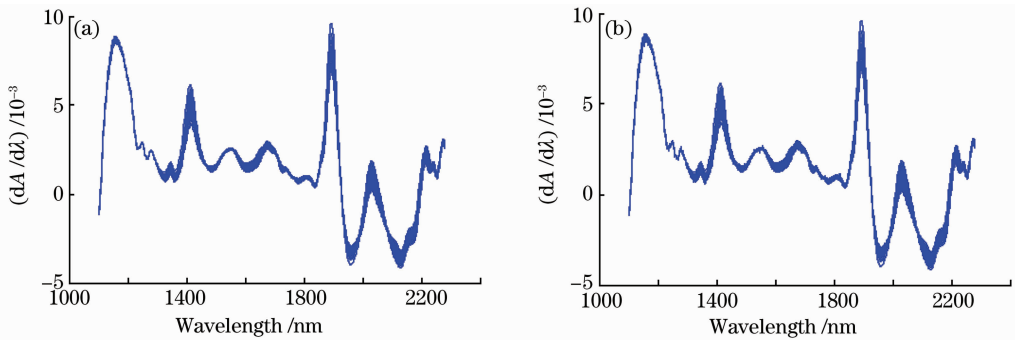


图 5 两种方法去噪后的光谱比较。(a) EMD 去噪后的光谱, (b)小波方法去噪后的光谱

Fig. 5 Comparison of de-noised spectrum with EMD and wavelet transform. (a) De-noised spectrum by EMD, (b) de-noised spectrum by wavelet transform

表 1 几种预处理方式烟草近红外分析总糖结果与化学值的比较

Table 1 Comparison of the prediction results of total sugar and chemical value by several de-noising methods

Sample number	Reference value /%	Nine points first derivative		Nine points first derivative + wavelet de-nosing		Nine points first derivative + EMD de-nosing	
		Predicted value /%	Bias /%	Predicted value /%	Bias /%	Predicted value /%	Bias /%
1	22.83	23.41	0.58	22.51	-0.32	23.11	0.28
2	30.74	31.67	0.93	31.28	0.54	31.24	0.55
3	32.24	31.99	-0.25	32.24	0.00	32.1	-0.14
4	32.62	32.19	-0.43	32.3	-0.32	32.31	-0.31
5	27.07	27.71	0.64	26.89	-0.18	27.36	0.29
6	20.08	19.53	-0.55	19.7	-0.48	19.59	-0.49
7	21.55	21.09	-0.46	21.87	0.32	21.27	-0.18
8	22.40	22.14	-0.26	22.57	0.17	22.4	0.00
9	25.93	26.85	0.92	26.42	0.49	26.38	0.45
10	27.24	27.8	0.56	26.96	-0.28	27.45	0.21
11	26.97	26.59	-0.38	26.74	-0.23	26.75	-0.22
12	26.28	25.99	-0.29	26.6	0.32	25.97	-0.31
13	24.11	23.84	-0.27	24.4	0.39	23.94	-0.17
14	26.23	27.1	0.87	26.7	0.47	26.73	0.50
15	27.37	28.08	0.71	27.85	0.58	27.83	0.46
16	24.83	24.01	-0.82	24.32	-0.61	24.34	-0.49
17	20.47	20.47	0.00	20.33	-0.14	20.57	0.10
18	30.74	30.14	-0.60	30.9	0.16	30.39	-0.35
19	21.75	21.57	-0.18	21.5	-0.25	21.88	0.13
20	19.69	19.93	0.24	19.38	-0.31	19.42	-0.21
$r^2$		0.9705		0.9830		0.9832	
RMSE		0.5606		0.3375		0.3310	

## 5 结 论

EMD 方法是最近发展起来的信号处理方法,由上面的分析可知在光谱预处理中引进 EMD 方法去除光谱噪声得到了很好的处理效果,为近红外光谱的分析精度和模型的稳定性奠定了良好的基础。EMD 与广泛使用的小波变换相比具有分解模态少

和不用选择基函数的优点,为近红外光谱预处理提供了新的途径。当然,EMD 分解方法提出的时间还不长,数学背景是科研人员面临的一个困难,随着模态混叠,端点效益等问题的进一步优化解决,EMD 分解技术必将得到更加广泛的应用。

## 参 考 文 献

- 1 Zhao Jiewen, Zhang Haidong, Liu Muhua. Preprocessing methods of near-infrared spectra for simplifying prediction model of sugar content of apples[J]. *Acta Optica Sinica*, 2006, **26**(1): 136~140  
赵杰文, 张海东, 刘木华. 简化苹果糖度预测模型的近红外光谱预处理方法[J]. *光学学报*, 2006, **26**(1): 136~140
- 2 Tian Gaoyou, Yuan Hongfu, Chu Xiaoli *et al.*. Near infrared spectra (NIR) analysis of octane number by wavelet denoising derivative method [J]. *Spectroscopy and Spectral Analysis*, 2005, **25**(4): 516~520  
田高友, 袁洪福, 褚小立 等. 结合小波变换与微分法改善近红外光谱分析精度[J]. *光谱学与光谱分析*, 2005, **25**(4): 516~520
- 3 Li Suwen, Xie Pinhua, Li Yujin *et al.*. Wavelet transform based differential optical absorption spectroscopy data processing [J]. *Acta Optica Sinica*, 2006, **26**(11): 1601~1604  
李素文, 谢品华, 李玉金 等. 基于小波变换的差分吸收光谱数据处理方法[J]. *光学学报*, 2006, **26**(11): 1601~1604
- 4 Hao Yong, Chen Bin, Zhu Rui. Analysis of several methods for wavelet denoising used in near infrared spectrum pretreatment [J]. *Spectroscopy and Spectral Analysis*, 2006, **26**(10): 1838~1842  
郝勇, 陈斌, 朱锐. 近红外光谱预处理中几种小波消噪方法的分析[J]. *光谱学与光谱分析*, 2006, **26**(10): 1838~1842
- 5 N. E. Huang, Z. Shen, S. R. Long *et al.*. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. *Proc. R. Soc. Lond. Ser. A*, 1998, **454**, 903~995
- 6 Olivier Adam. The use of the Hilbert-Huang transform to analyze transient signals emitted by sperm whales [J]. *Applied Acoustics*, 2006, **67**: 1134~1143
- 7 Jordan B. Camp, John K. Cannizzo, Kenji Numata. Application of the Hilbert-Huang transform to the search for gravitational waves[J]. *Phys. Rev. D*, 2007, **75**(6): 061101-1~061101-5
- 8 Hang Dong, Zhang Shuangnan. Comparison between windowed FFT and Hilbert-Huang transform for analyzing time series with poissonian fluctuations: a case study[J]. *Chinese J. Astronomy and Astrophysics*, 2006, **6**(4): 503~512
- 9 Tang Jingtian, Hua Xirui, Cao Zheming *et al.*. Hilbert-Huang transformation and noise suppression of magnetotelluric sounding data[J]. *Chinese J. Geophysics*, 2008, **51**(2): 603~610  
汤井田, 化希瑞, 曹哲民 等. Hilbert-Huang 变换与大地电磁噪声压制[J]. *地球物理学报*, 2008, **51**(2): 603~610
- 10 Jiang Rong, Yan Hong. Studies of spectral properties of short genes using the wavelet subspace Hilbert-Huang transform (WSHHT)[J]. *Physica A*, 2008, **387**, 4223~4247
- 11 M. B. Bradley, K. Camelia. Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data [J]. *Geophysics*, 2008, **72**, H29~H37
- 12 Ma Jing, Tan Liying, Ran Qiwen. The elementary theory of optical wavelet filtering[J]. *Chinese J. Lasers*, 1999, **A26**(4): 343~347  
马晶, 谭立英, 冉启文. 光学小波滤波理论初探[J]. *中国激光*, 1999, **A26**(4): 343~347
- 13 Zhou Zhenyu, Yang Hongyu, Gong Hui *et al.*. Brain signal analysis of functional near-infrared imaging based on Hilbert-Huang transform [J]. *Acta Optica Sinica*, 2007, **27**(2): 307~312  
周振宇, 杨宏宇, 龚辉 等. 基于希尔伯特-黄变换的近红外脑功能成像信号分析[J]. *光学学报*, 2007, **27**(2): 307~312
- 14 Shi Chunxiang, Luo Qifeng. Hilbert-Huang transform and wavelet analysis of time history signal [J]. *Acta Seismologica Sinica*, 2003, **16**(4): 422~429
- 15 Daniel O. Trad, Jandyr M. *et al.*. Wavelet filtering of magnetotelluric data[J]. *Geophysics*, 2000, **65**(2): 482~491