

The ESMD Method for Climate Data Analysis

Jinliang Wang, Zongjun Li

College of Science, Qingdao Technological University, Qingdao

Email: wangjinliang0811@126.com

Received: Aug. 12th, 2013; revised: Sep. 15th, 2013; accepted: Sep. 23rd, 2013

Copyright © 2014 Jinliang Wang, Zongjun Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for Hans and the owner of the intellectual property Jinliang Wang, Zongjun Li. All Copyright © 2014 are guarded by law and by Hans as a guardian.

Abstract: The new ESMD method has been reported by the Science Net and the China Science Daily. It is the new development of the well-known Hilbert-Huang transform and can be used effectively in the science researches and engineering applications associated with data analysis from atmospheric and oceanic sciences, informatics and so on. The ESMD method has three superiorities in climate data analysis: 1) it is good at finding the global changing trend. It can not only extract the annual changing trend from the observation sequence of a few years, but also draw out the climate changing trend from that of hundreds of years which is helpful for exploring the problem of global warming; 2) it is good at abnormal diagnosis which is helpful for exploring the problem of climatic anomaly due to the ability of finding the abnormal time and frequency from the decomposed modes; 3) it is good at time-frequency analysis and energy variation analysis. With the abandon of Hilbert transform which has many defects, the data-based direct interpolating approach is developed to compute the total energy and instantaneous frequencies at any time scales. It follows from the above features that the ESMD method has a good application prospect in the research of climate change.

Keywords: ESMD Method; Hilbert-Huang Transform; Data Analysis; Climate Change; Time-Frequency Analysis; Air-Sea Flux

可用于气候数据分析的 ESMD 方法

王金良, 李宗军

青岛理工大学理学院, 青岛

Email: wangjinliang0811@126.com

收稿日期: 2013 年 8 月 12 日; 修回日期: 2013 年 9 月 15 日; 录用日期: 2013 年 9 月 23 日

摘要: 科学网和中国科学报曾先后对我们新研发的 ESMD 方法进行过报导。ESMD 方法是著名的 Hilbert-Huang 变换的新发展, 可用于海洋和大气科学、信息科学等领域所有涉及数据分析的科研和工程应用。ESMD 方法在气候数据分析方面有如下优势: 1) 优于寻找变化趋势, 不仅能够从数年的观测序列中分离出年际变化趋势, 也能够从数百年的长时间气候观测序列中分离出气候变化总趋势, 有助于探究全球气候变暖问题; 2) 优于异常诊断, 能够从分解模式中发现异常时段与频段, 有利于气候异常研究; 3) 优于时-频分析和能量变化分析, 存在诸多弊端的 Hilbert 变换被代之以先进的直接插值法, 能够直观地分析各时间尺度上的频率变化和总能量变化。由此可见, ESMD 方法在气候变化问题的研究中存在很大的应用前景。

关键词: ESMD 方法; Hilbert-Huang 变换; 数据分析; 气候变化; 时频分析; 海气通量

1. 引言

2013 年 5 月 27 日科学网以“青岛理工大学研发数据处理新方法”为题对我们的 ESMD 方法进行了报导(<http://news.sciencenet.cn/htmlnews/2013/5/278345.shtml>); 继而, 5 月 30 日我校校报又以“我校教师研发数据处理新方法受关注”为题做了后续报导(<http://news.qtech.edu.cn/article.php?articleid=24900>); 7 月 3 日报导正式见于《中国科学报》(2013-07-03 第 4 版国际), 题目为“青岛理工发明数据处理新方法”(<http://news.sciencenet.cn/sbhtmlnews/2013/7/274855.shtml>)。

ESMD 方法是“极点对称模态分解方法”的简称^[1], 是著名的 Hilbert-Huang 变换^[2]的新发展, 可用于海洋和大气科学、信息科学、经济学、生态学、医学和地震学等领域所有涉及数据分析的科研和工程应用。

该方法的研发历时两年之久, 完成于 2012 年 4 月, 公开于 2013 年 3 月。相关文稿可通过 arXiv 电子论文公开网站免费下载查阅(<http://arxiv.org/abs/1303.6540>), 正式论文^[1]发表于常规国际期刊 *Advances in Adaptive Data Analysis*。相关计算程序已获批二项计算机软件著作权^[3,4]。

2. 数据分析研究现状

高科技的核心是“数学技术”, 而“数学技术”的主要手段是数值模拟和数据分析。有成熟数学模型的问题适用数值模拟, 没有数学模型的问题只能依靠数据分析了。特别地, 对于物理机制不明确的过程, 研究有赖于观测实验。探索的方式往往是将杂乱无章的随机观测数据分解成不同频率的模态, 从中寻找可能的变化规律。

在随机数据分析方面, 经典的方法是基于线性叠加原理的 Fourier 变换。它将一个观测时间序列映射到频率-能谱空间, 其每一个模态都是振幅不变频率也不变的正弦函数。其缺点是只适用于针对线性变化的平稳信号; Wavelet 变换是目前比较盛行的方法, 它通过取定一系列的频率窗口对信号进行分解, 在一定程度上弥补了 Fourier 变换的缺陷, 能够表达出频率的时变性。但其理论基础还是线性叠加原理, 只适用于针对线性变化的非平稳信号; 以经验模分解

(EMD)为基础的 Hilbert-Huang 变换是目前比较热门的方法, 它是一种数据自适应处理方法, 不需要预先取定基函数或窗口长度, 其分解模态不但频率可变振幅也可变, 适用于非线性非平稳信号。存在问题主要有: 筛选次数难以确定^[5], 分解出的趋势函数太粗略, Hilbert 谱分析存在局限等^[1]。

随机数据分析的根本问题在于其非平稳性, 一是趋势有变化, 二是振幅和频率有时变性。对于存在大的趋势变化时, 如何抽出全局均线是最要紧的问题。Fourier 变换在一开始就认为全局均线为零, “最小二乘法”必须有先验的函数形式, “滑动平均法”在时间窗口和权函数选取上缺少依据, 小波变换其实也是一种滑动平均。只有把全局均线比较好地滤除了剩下的才能看成脉动量。

3. ESMD 方法

所提出的 ESMD 方法借鉴了 EMD 的思想, 将外部包络线插值改为内部极点对称插值, 借用“最小二乘”的想法来优化最后剩余模态使其成为整个数据的“自适应全局均线”, 并由此来确定最佳筛选次数。考虑到包括 Hilbert 变换在内的所有积分变换在分析时-频变化方面都存在固有缺陷, 我们抛弃了频谱分析依靠积分变换的传统观念, 创造性地提出了针对数据的“直接插值(DI)法”。借此不但可以直观地体现各模态的振幅与频率的时变性, 还可明确地获知总能量变化(此处“能量”是广义的, 对气温而言指温度脉动强度)。其实, 软件程序所输出的基于“直接插值法”的时变频谱图比 Hilbert 谱更直观也更合理, 因为不但频率是变化的总能量也是变化的刻意将能量视为恒量并将其映射到一系列固定频率上是牵强的。

3.1. 模态分解示例

例 1: 对由正弦函数、加权周期函数^[6,7]和抛物线合成的信号:

$$-\sin(8\pi t) + 1.5e^{-0.2t} \sin(1.9\pi t + \pi/20) + (t-2)^2$$

进行模态分解试验。由图 1 可见最佳筛选次数为 29, 它对应着最小的方差比率(表明最后的剩余模态 R 是数据的最佳自适应全局均线)。此时所对应的分解也是最佳的, 三条函数曲线得到了明确分离(见图 2)。

例 2: 对由美国气候数据中心所提供的 2008.05.10~

2011.11.03 间实测气温数据进行模态分解试验。此时最佳筛选次数为 30，相应的分解见图 3。其中，剩余模态 R 为最佳自适应全局均线对应于年际气温变化(图 4 说明 R 能很好地拟合数据)，第 5、4、3 模态分别对应平均周期大约为 66 天、35 天、17 天的气温波动。特别地，我们可以通过各模态振幅的变化情况来判定温度异常发生的频段和时间(图 3，图 5)。此例中，模态 5 振幅变化小而模态 4 振幅变化大这表明气温异常主要发生在周期为 35 天的时间尺度上，而异常时间主要集中在 2009 年的 1 至 3 月份。另外，软件程序还可输出

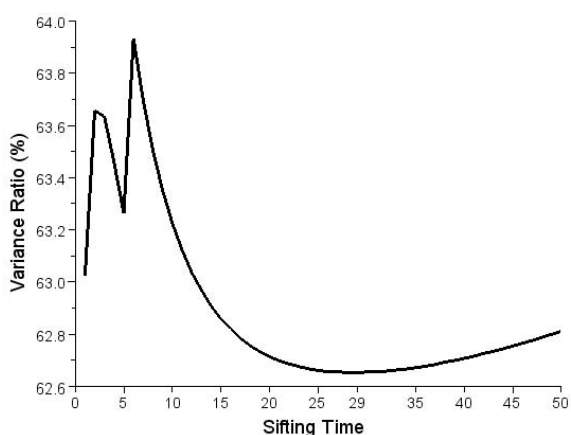


Figure 1. The variation of the variance ratio along the sifting times (the optimal times is 29) [The copyright belongs to World Scientific Publishing Company]

图 1. 方差比率随筛选次数的变化(29 次最佳)[图片版权隶属 World Scientific 出版公司]

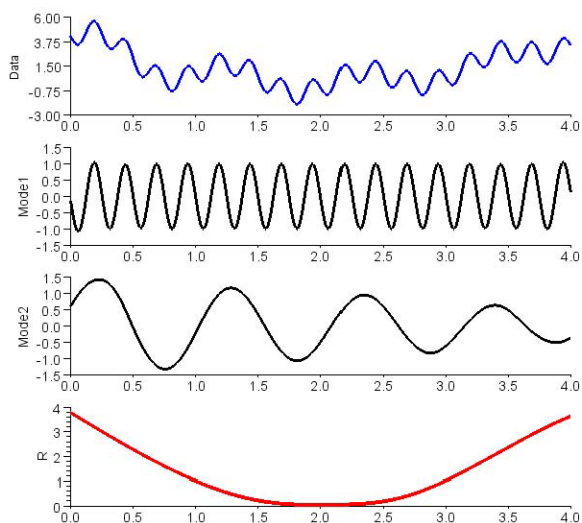


Figure 2. The ESMD decomposition results with 29 sifting times (the first subfigure is the synthesized signal) [The copyright belongs to World Scientific Publishing Company]

图 2. ESMD 对应于筛选次数为 29 的分解结果(其中第一子图合成信号)[图片版权隶属 World Scientific 出版公司]

基于“直接插值法”的时变频谱图(图 6)，比 Hilbert 谱更直观也更合理，因为不但频率是变化的总能量也是变化的(图 7)不能将能量视为恒量并将其映射到一系列固定频率上。最后的图 8 反映了滤除均线后的脉动量，比“最小二乘”法和“滑动平均”法得到的要好。

3.2. ESMD 方法的优势

由前例可见，ESMD 方法在数据分析方面有如下优势：

- 1) 优于寻找变化趋势，不仅能够从数年的观测

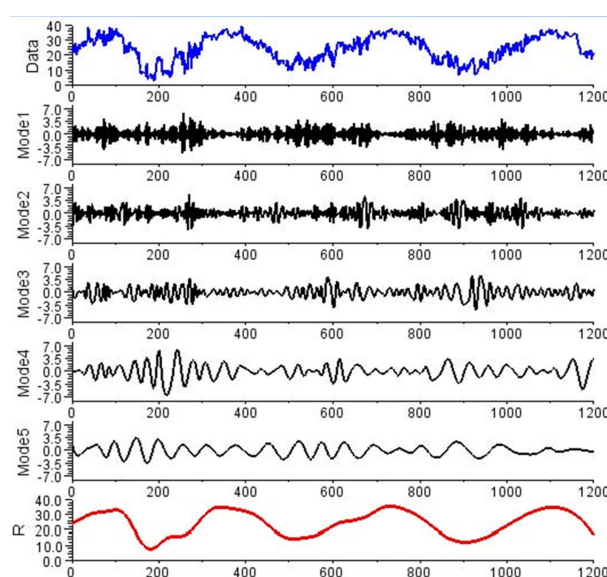


Figure 3. The ESMD decomposition results with 30 sifting times (the horizontal axis stands for the time with unit day) [The copyright belongs to World Scientific Publishing Company]

图 3. ESMD 对应于筛选次数为 30 的分解结果(横坐标代表时间/天)[图片版权隶属 World Scientific 出版公司]

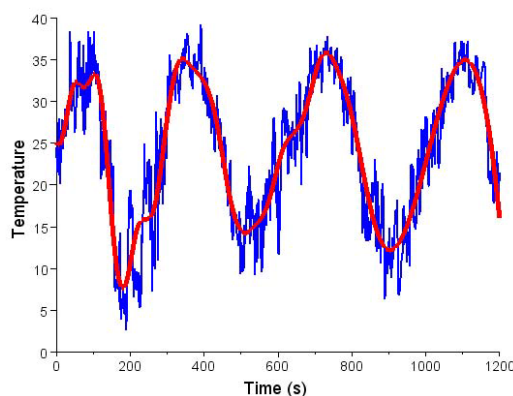


Figure 4. The fitting of the optimal adaptive global mean curve R to the data [The copyright belongs to World Scientific Publishing Company]

图 4. 最佳自适应全局均线 R 对数据的拟合情况[图片版权隶属 World Scientific 出版公司]

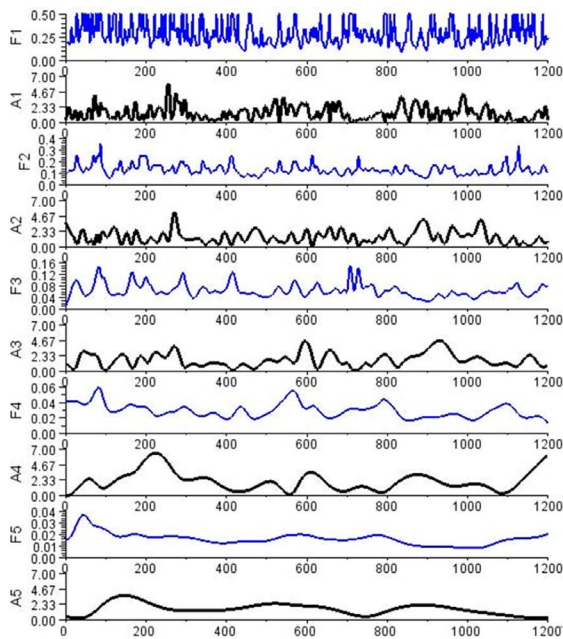


Figure 5. The time variation of the frequencies and amplitudes of the modes (F and A stand for the frequency and amplitude separately)
 图 5. 模态频率和振幅的时变图(F 和 A 分别代表频率和振幅)

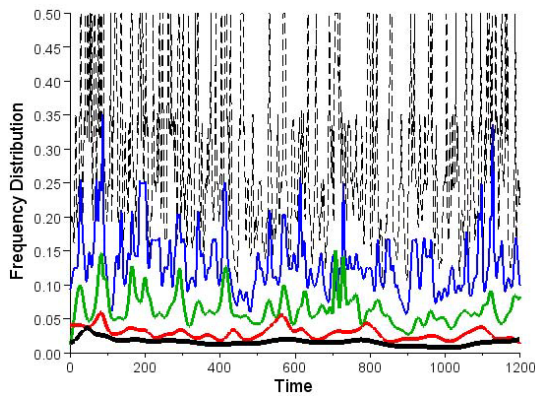


Figure 6. The time-varying frequency spectrum (each line stands for a frequency variation of a mode)
 图 6. 时变频谱图(一条线代表一个模态的频率变化)

序列中分离出年际变化趋势，也能够从数百年的长时间气候观测序列中分离出气候变化总趋势，有助于探究全球气候变暖问题；

2) 优于异常诊断，能够从分解模态中发现异常时段与频段，有利于气候异常研究；

3) 优于时-频分析和能量变化分析，存在诸多弊端的 Hilbert 变换被代之以先进的直接插值法，能够直观地分析时间尺度上的频率变化和总能量变化。

3.3. 关于频率交叉问题

“频率交叉”(或称“模态混叠”)问题被认为是

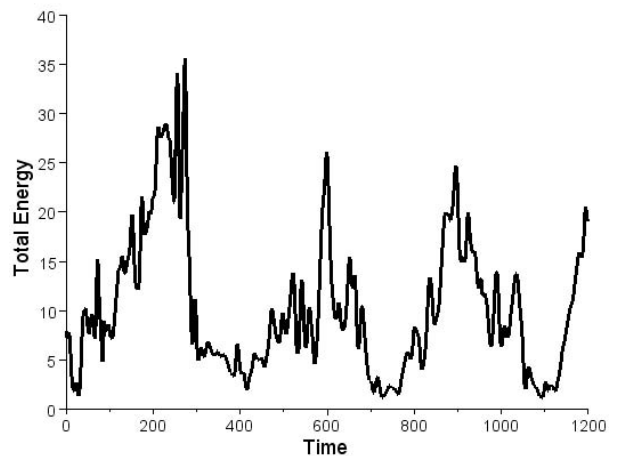


Figure 7. The time-varying total energy of the modes
 图 7. 模态总能量的时变图

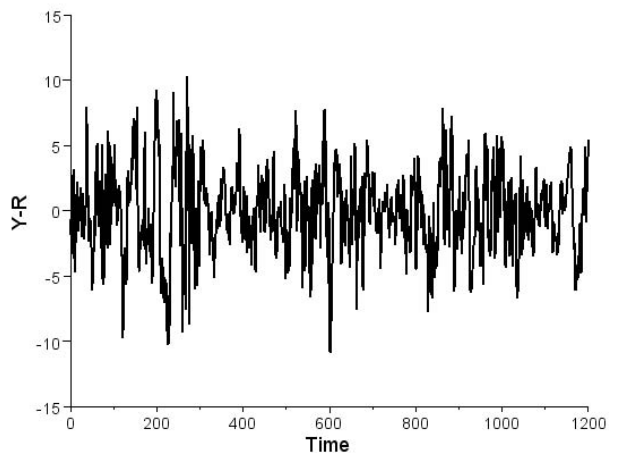


Figure 8. The fluctuation part of the data with getting rid of the non-stationary adaptive global mean curve R
 图 8. 滤除非平稳自适应全局均线 R 后的脉动量

困扰 EMD 方法使用者的一个重要问题，通常的理解是所分解出的模态彼此之间不能满足数学上的“正交性”，相邻模态之间存在频率混叠现象。再者，分解结果也会因筛选次数不同而异。所以分解结果具有不稳定性从而导致其模态有效性受质疑。EEMD 方法^[8]就是为试图解决这一问题发展起来的。其基本想法是：将原信号多次添加白噪声进行 EMD 分解再取平均。模态由于是多次加噪平均的结果，对抗扰不敏感，被认为是分解稳定的并以此来理解物理现象。就我们的认识而言，加噪处理后的结果难以令人信服，其可靠性是令人怀疑的，毕竟白噪声在各个频段上都有能量分布其污染效应会进入各个模态而不仅仅作用于高频部分。因此，加噪处理不纯是有利的，其副作用也很大，上百次的加噪处理很可能使信号面目全非。

溯本求源，问题还在于解决所谓的“频率交叉”问题。下面我们在 ESMD 方法的框架下理解这一点。就 ESMD 方法的第二组成部分“直接插值法”所给出的瞬时频率变化图(图 6)而言，有这么一个现象：虽然从总体来看相邻模态之间会有同频率出现，但却不发生在同一时刻，而相同时刻下的频率一般不重叠。因而，只要筛选次数适当，“频率交叉”问题在时变意义下将不再是困惑的事，而分解出的模态也因瞬时频率变化不同而异。另外，由于最佳筛选次数是通过优化自适应全局均线在稳定区间内获取的，而自适应全局均线是数据的最佳拟合曲线(图 4)，剩余信号看成多个振动子信号的叠加是合理的。由此而论，所得有限个模态各自都具独立的代表性，能够构成随机空间的一组自适应基。另外，由于每个自适应的基函数都是频率变化振幅也变化的“广义周期函数”，其表现力很强，一个相当于 Fourier 意义下的很多个正弦函数，有限个叠加足可以替代无穷个频率成倍的正弦函数。一些研究者能够用优化迭代的办法分离出这样的函数也说明经验模分解在数学上的确有迹可寻。当然，要建立一套完善的理论基础还有待时日。

4. ESMD 方法的应用前景

海洋和大气科学是以观测为主的学科，必然涉及数据分析。我们的 ESMD 方法在这一领域存在很大的应用前景。下面以海-气通量研究为例来说明。

从我们的初步尝试结果^[9]来看，ESMD 方法适用于海-气通量研究，优于观测风速的湍流与非湍流成份的分离。分离出风湍流成份是通量观测最核心的工作，毕竟海上观测实验是昂贵的，总是希望用一点观测值代表整个海区。在这个前提下，水平方向上存在的非湍流浓度梯度尽管对观测点处的通量有所贡献，但必须予以滤除。在这方面我们曾在文[10,11]中尝试发展了 Fourier 谱分离方法，但后来研究发现该方法对分界点比较敏感，所计算出的通量可信度不高。最近王娟娟等人在文[12]中发展了基于 Hilbert-Huang 变换的分离方法，但其中涉及加噪处理，也缺少说服力。文^[9]所给出的基于 ESMD 方法的分离手段能够很好地分离风速的湍流与非湍流成份，有待推广。另外，海上

观测的平台晃动校正问题^[13]也是通量计算必然涉及的问题，对于晃动幅度不太大的情况也可直接尝试 ESMD 方法。

气候变化问题没有成熟模型，总希望借助观测数据探究其变化规律，这是 ESMD 方法所擅长的，期待尝试。

致谢

感谢山东省自然科学基金项目“海气通量计算模型研究”(No. ZR2012DM004)；国家海洋局海洋遥测工程技术研究中心开放基金项目(No. 2013005)和国家自然科学基金项目(No. 41376030)的资助。

参考文献 (References)

- [1] Wang, J.-L. and Li, Z.-J. (2013) Extreme-point symmetric mode decomposition method for data analysis. *Advances in Adaptive Data Analysis*, **5**, 1350015 (36 pages).
- [2] Huang, N.E., Shen, Z., Long, S.R., et al. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A*, **454**, 903C995.
- [3] 王金良, 李宗军 (2012) 关于非线性信号处理的极点对称模态分解方法计算软件(简称 ESMD 计算软件). 中华人民共和国国家版权局计算机软件著作权登记, No. 2012SR052512.
- [4] 王金良, 李慧凤 (2012) 基于极点对称模态分解的频率直接插值法计算软件(简称 DI 频率计算软件). 中华人民共和国国家版权局计算机软件著作权登记, No. 2012SR102181.
- [5] Wang, J.-L. and Li, Z.-J. (2012) What about the asymptotic behavior of the intrinsic mode functions as the sifting times tend to infinity. *Advances in Adaptive Data Analysis*, **4**, 1250008.
- [6] Wang, J.-L. and Zhang, G. (2006) Asymptotic weighted periodicity for delay differential equations. *Dynamic Systems and Applications*, **15**, 479-500.
- [7] Wang, J.-L. and Li, H.-F. (2007) Asymptotic weighted-periodicity of the impulsive parabolic equation with time delay. *Acta Mathematicae Applicatae Sinica*, **23**, 1-8.
- [8] Wu, Z.-H. and Huang, N.E. (2009) Ensemble empirical mode decomposition: A noise assisted data analysis method. *Advances in Adaptive Data Analysis*, **1**, 1-41.
- [9] Li, H.-F., Wang, J.-L. and Li, Z.-J. (2013) Application of ESMD method to air-sea flux investigation. *International Journal of Geosciences*, **4**, 8-11.
- [10] 王金良 (2008) 关于海气通量的观测及其交换机制的研究. 中国科学院海洋研究所博士后出站报告.
- [11] 王金良, 宋金宝 (2009) 关于涡相关海气通量计算的资料处理技术. *海洋科学*, **11**, 1-5.
- [12] Wang, J.-J., Song, J.-B., Huang, Y.-S., et al. (2013) Application of the Hilbert-Huang transform to the estimation of air-sea turbulent fluxes. *Boundary-Layer Meteorology*, **147**, 553-568.
- [13] 王金良, 宋金宝 (2011) 晃动平台上海-气通量观测误差校正模型. *海洋科学*, **12**, 106-112.