

# 噪声环境中的 EMD 改进算法\*

刘迎军<sup>1</sup>, 杨志景<sup>2</sup>, 董健卫<sup>3</sup>, 李淑龙<sup>1</sup>

(1. 南方医科大学生物医学工程学院, 广东 广州 510515;

2. 广东工业大学信息工程学院, 广东 广州 510006;

3. 广东药学院基础学院, 广东 广州 510006)

**摘要:** 经验模式分解 (Empirical Mode Decomposition, EMD) 是近年来出现的一种自适应的信号分解算法, 该方法受到了广泛的关注, 被成功应用于许多领域。然而, 当信号包含噪声时, 它存在过度分解的弊端, 容易受噪声的干扰, 因而严重地限制了该方法的推广。为了解决这一问题, 提出了一种改进的 EMD 方法: 在首轮分解时, 采用光滑样条拟合来代替原来的三次样条插值, 可避免对噪声成分过度分解, 从而极大地减少了噪声成分的干扰。仿真实验证实了新方法有显著的改进效果; 两个实际气候数据序列分解的例子进一步说明了新方法的有效性和优越性。

**关键词:** 经验模式分解; 噪声; 本征模函数; 光滑样条; 广义交叉验证

**中图分类号:** TP391   **文献标志码:** A   **文章编号:** 0529-6579 (2014) 04-0025-10

## Improved EMD Method for Noisy Signal

LIU Yingjun<sup>1</sup>, YANG Zhijing<sup>2</sup>, DONG Jianwei<sup>3</sup>, LI Shulong<sup>1</sup>

(1. School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China;

2. School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China;

3. School of Basic Courses, Guangdong Pharmaceutical University, Guangzhou 510006, China)

**Abstract:** Recently, an adaptive method called Empirical mode decomposition (EMD) is proposed for signal analysis. It has attracted great deal of attention and been used in many areas successfully since its advent. However, when the signal is contaminated by noise, EMD suffers from the drawback of over decomposition and likely is affected by noise, which severely restricts its applications. In order to solve this problem, an improved version of EMD is proposed. During the first decomposition circle, the original cubic spline interpolation is replaced by the smoothing spline fitting, which can avoid the over decomposition problem and then reduce the disturbance of noise component. Simulations validate the improvement of the new proposed method. Moreover, two real climate data examples show the effective and superiority of the new method for real signals.

**Key words:** empirical mode decomposition; noise; intrinsic mode function; smoothing spline; generalized cross validation

1998年, 美国工程院院士 Huang 提出了一种新的信号分解方法——经验模式分解 (Empirical

Mode Decomposition, EMD) 方法<sup>[1]</sup>。该方法适用于非线性和非平稳信号, 能将信号自适应地分解成

\* 收稿日期: 2013-09-13

基金项目: 国家自然科学基金资助项目 (11101437); 2012 广东高校优秀青年创新人才培养资助项目 (2012LYM-0036); 博士点新教师基金资助项目 (20110171120044)

作者简介: 刘迎军 (1981 生), 男; 研究方向: 信号处理、模式识别; 通讯作者: 杨志景; E-mail: yangzhijing521@yahoo.com.cn

若干个本征模函数 (Intrinsic Mode Function, IMF) 之和, 已经成功地被应用到各个领域<sup>[2-5]</sup>。然而, 该方法理论基础尚不完善, 还存在一些问题有待进一步研究, 比如边界延拓、停止条件、模式混叠等等。对于确定性的信号, EMD 的分解结果是合理的; 但是对于含有噪声的信号, 经过 EMD 分解之后, 原来的噪声成分也被强制分解成了多个 IMF 之和, 此时的每个 IMF 分量的物理意义不明确, 甚至分解得到的噪声分量个数也无法完全确定<sup>[6]</sup>, 使得分解结果变得不可靠。为了解决该问题, 本文提出一种改进的 EMD 方法: 在首轮分解时, 采用光滑样条拟合来代替三次样条插值。这样可以避免对噪声成分过度分解, 从而使得结果更加可信。

## 1 EMD 方法简介

### 1.1 IMF 的定义

EMD 将信号分解成一系列的 IMF 之和, 其中 IMF 的定义为满足下列两个条件的信号: ① 整个信号中零点与极值点最多相差为 1; ② 信号上任意一点, 由局部极大值点和局部极小值点确定的上、下包络线的均值为零。

### 1.2 EMD 算法

任给一个信号  $s(t)$ , 首先对其所有极大值点和极小值点分别用三次样条插值, 得到上下两条包络线。用  $m(t)$  记为上下包络线的平均值, 令  $h(t) = s(t) - m(t)$ , 将  $h(t)$  当作新的  $s(t)$ , 重复以上操作, 直到  $h(t)$  满足 IMF 的条件为止, 则得到原信号  $s(t)$  的第一个 IMF, 即

$$c_1(t) = h(t) \quad (1)$$

再令

$$r(t) = s(t) - c_1(t) \quad (2)$$

然后将  $r(t)$  视为新的  $s(t)$ , 重复以上过程, 可依次得第 2 个 IMF  $c_2(t)$ , 第 3 个 IMF  $c_3(t)$ , …, 第  $N$  个 IMF  $c_N(t)$ , 最终得分解式:

$$s(t) = \sum_{k=1}^N c_k(t) + r(t) \quad (3)$$

其中  $r(t)$  为余量信号, 代表信号的平均趋势。

## 2 改进的 EMD 方法

### 2.1 GCV - 光滑样条

设  $f(t)$  为未知的光滑函数且  $\varepsilon_i \sim N(0, \sigma^2)$ , 即  $\varepsilon_i$  为服从同一分布的高斯白噪声, 而数据点  $\{(t_i, y_i) \mid i = 1, 2, \dots, n\}$  由以下模型得到:

$$y_i = f(t_i) + \varepsilon_i \quad (4)$$

其中  $a < t_1 < \dots < t_n < b$ 。根据已知数据点  $(t_i, y_i)$

求  $f(t)$ , 显然可以采用拟合的方法。

光滑样条是很重要的一种数据拟合方法, 最常用的定义式为:

$$\hat{f}(t) = \arg \min_{f \in C^2[a, b]} \left( \sum_{i=1}^n (y_i - f(t_i))^2 + \lambda \int_a^b (f''(t))^2 dt \right) \quad (5)$$

其中,  $C^2[a, b]$  表示区间  $[a, b]$  上二阶连续可微的函数的集合,  $f''(t)$  为  $f(t)$  的二阶导数。  $\lambda$  是一个光滑参数, 用来权衡拟合程度与光滑程度。

令  $h_i = t_{i+1} - t_i$ ,  $y = (y_1, \dots, y_n)^T$ ,  $f = (f(t_1), \dots, f(t_n))^T$  且  $\hat{f}_\lambda = (\hat{f}_\lambda(t_1), \dots, \hat{f}_\lambda(t_n))^T$ 。设  $Q = (q_{ij})$  为  $n \times (n-2)$  的矩阵, 其中  $i = 1, \dots, n, j = 2, \dots, n-1$ , 而元素  $q_{ij}$  满足:  $q_{j+1, j} = h_j^{-1}$ ,  $q_{j-1, j} = h_{j-1}^{-1}$ ,  $q_{j, j} = -h_{j-1}^{-1} - h_j^{-1}$ , 其它均为零; 又设  $R = (r_{ij})$  为  $(n-2) \times (n-2)$  的对称矩阵, 其元素  $r_{ij}$  满足:  $r_{i+1, i} = r_{i, i+1} = \frac{1}{6} h_i$ ,  $i = 2, \dots, n-2$ ,  $r_{i, i} = \frac{1}{3} (h_{i-1} + h_i)$ ,  $i = 2, \dots, n-1$ , 其它均为零。

再令  $K = QR^{-1}Q^T$ ,  $S_\lambda = (I + \lambda K)^{-1}$ , 其中  $I$  为单位矩阵。于是, 根据公式 (5) 的求解, 有  $\hat{f}_\lambda = S_\lambda y$  (更详细内容参见文献 [7])。

如何选取光滑参数  $\lambda$  是个关键问题, 已经有一些方法, 比如交叉验证 (Cross Validation, CV)、广义交叉验证 (Generalized Cross Validation, GCV)、Akaike 信息准则等方法来选取合适的光滑参数。其中 GCV 是普遍使用的一种方法, 计算公式如下

$$\text{GCV}(\lambda) = \frac{1}{n} \cdot \frac{\sum_{i=1}^n (y_i - \hat{f}(t_i))^2}{(1 - n^{-1} \text{tr}(S_\lambda))^2} \quad (6)$$

通过最小化公式 (6) 中的 GCV 函数就能得到最优光滑参数  $\lambda$ , 从而得到相应的 GCV - 光滑样条。

在模型 (4) 中, 噪声  $\varepsilon_i$  假设为高斯白噪声。实际上, GCV - 光滑样条不仅适用于高斯白噪声的情形, 而且还对于一定程度内的高斯白噪声和非白噪声的混合情形也是适用的, 而源自实际的数据序列经常可看作这种情形<sup>[7]</sup>, 因此它被广泛应用到许多领域。

### 2.2 噪声 IMF

如第 1 节所述, 一个信号要成为 IMF, 必须满足两个条件, 以保证 IMF 具有时间轴上的局部对称性<sup>[8]</sup>。这与通常的周期信号、准周期信号, 甚

至复杂周期信号的特点是一致的；但是对于噪声信号而言，这样的条件就显得过于苛刻。因为噪声信号与确定性的信号不同，描述噪声的 IMF 应该从概率分布的角度出发。于是，在考虑噪声信号时，不适合采用样条插值来得到包络，而应该考虑拟合的手段，故可定义噪声 IMF 满足如下两个条件：

- 1) 噪声 IMF 中极大值序列的均值曲线大于零，而极小值序列的均值曲线小于零；
- 2) 噪声 IMF 的极大值和极小值序列的均值曲线之和为零，即两者关于 X 轴对称。

如图 1 所示，给定一个高斯白噪声序列，过其极大值和极小值点序列作三次样条插值，就得到了上下包络（图 1 (a)）。对上下包络求平均，则得

到了均值曲线（图 1 (c)）。显然，对于纯高斯白噪声序列，这样的均值曲线没有什么物理意义，反而会影分解的最终结果。这是传统 EMD 方法的一个缺陷。如果使用 GCV - 光滑样条分别对极大和极小值点序列进行拟合，则得到近似的水平直线（图 1 (b)）。再将两者平均，得到的还是近似水平直线（图 1 (d)）。这与高斯白噪声的零均值是一致的。

图 1 的例子表明，就纯噪声信号而言，对极值点序列采用拟合比插值更合适。在 GCV - 光滑样条拟合的方法下，纯噪声序列不会被强制进行分解，而是作为一个独立的分量而存在，即为噪声 IMF。

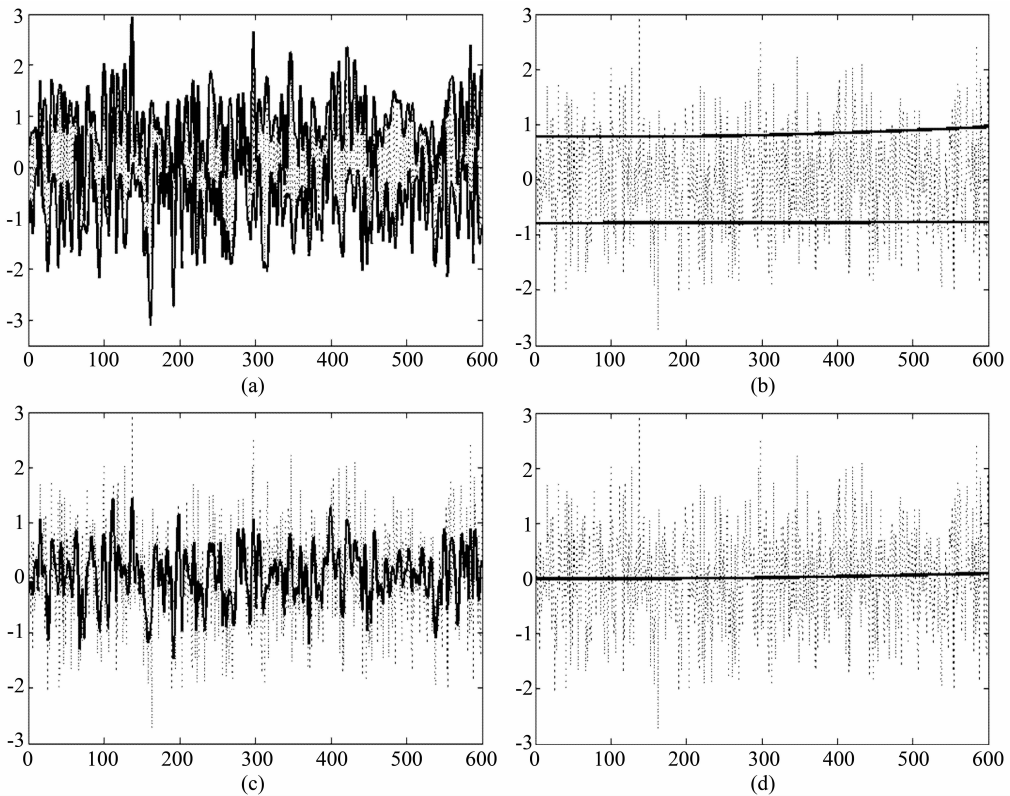


图 1 图中虚线为仿真产生的纯高斯白噪声

Fig. 1 The dotted curve is a pure Gaussian noise sequence

(a) 实线分别为过极大值和极小值点序列的三次样条插值所求得的上包络；(b) 实线分别为拟合极大值和极小值点序列的 GCV - 光滑样条；(c) 实曲线为 (a) 中上下包络的均值；(d) 实线为 (b) 上下 GCV - 光滑样条的均值

### 2.3 改进的 EMD 算法

改进的 EMD 算法具体描述如下：

- 1) 任给信号  $s(t)$ ，对其进行滑动平均得  $s_a(t)$ （滑动窗宽取为 5），然后计算

$$T = \frac{\sum_{i=1}^n [s(t_i) - s_a(t_i)]^2}{\sum_{i=1}^n s(t_i)^2} \quad (7)$$

如果  $T > 0.05$ ，说明是信号含有噪声，转②；否则对  $s(t)$  进行原始 EMD 分解，程序终止；

- 2) 求出  $s(t)$  的极值点序列，分别对其极大值和极小值点序列用 GCV - 光滑样条拟合，得到上下两条极值点序列分布的均值曲线  $e_{k1}(t), e_{k2}(t)$ ；

3) 令  $m_k(t) := [e_{k1}(t) + e_{k2}(t)] / 2$ ,  $h_k(t) := s(t) - m_k(t)$ ，若相邻两次的  $h_k(t)$

和  $h_{k-1}(t)$  满足停止条件 ( $\rho$  经验取值为 0.05)

$$SD = \frac{\sum_{i=1}^n [h_k(t_i) - h_{k-1}(t_i)]^2}{\sum_{i=1}^n h_{k-1}(t_i)^2} \leq \rho \quad (8)$$

就得到原信号  $s(t)$  的第一个 IMF, 也就是噪声 IMF  $c_1(t)$ ; 否则, 将  $h_k(t)$  当作新的  $s(t)$ , 转②;

4) 令  $s(t) := s(t) - c_1(t)$ , 接着对  $s(t)$  进行原始 EMD 分解。

从上述算法不难看出, 改进的 EMD 方法与原始 EMD 的区别在于对于含有噪声的信号的第 1 个 IMF 的求法不同。

## 3 仿真实验

### 3.1 仿真数据

基于模型 (4), 每个仿真数据由高斯白噪声和不同趋势叠加而成, 其中高斯白噪声的均值为 0, 方差为 1, 数据长度取  $n = 600$ 。正弦趋势分别

取: 一种为单频率的正弦信号  $y = \cos(18\pi t/n)$ ,  $t = 1, 2, \dots, n$ ; 另一种为两个频率的正弦信号的叠加, 即  $y = \cos(18\pi t/n) + \sin(10\pi t/n)$ ,  $t = 1, 2, \dots, n$ 。

### 3.2 实验结果

当原信号为正弦信号  $y = \cos(18\pi t/n)$  与高斯白噪声叠加而成时, EMD 和改进的 EMD 分解结果分别显示在图 2 和图 3。在图 2 中, 前面 4 个 IMF 成分是噪声成分, 而图 3 只有 IMF1 才是噪声成分。图 2 中的 IMF5 信号与图 3 中的 IMF2 频率相似, 它们都是对原始信号中正弦趋势的刻画。但不难发现, 图 2 中的 IMF5 在时间坐标区间  $[100, 200]$  之间已经严重偏离原始的正弦信号, 而图 3 中 IMF2 则基本上保留了原始的正弦信号的特征。由此可以得出, 改进的 EMD 算法避免了对噪声成分的过度分解, 一方面使得结果变得简洁, 另一方面极大地减弱了噪声成分对分解造成的干扰。

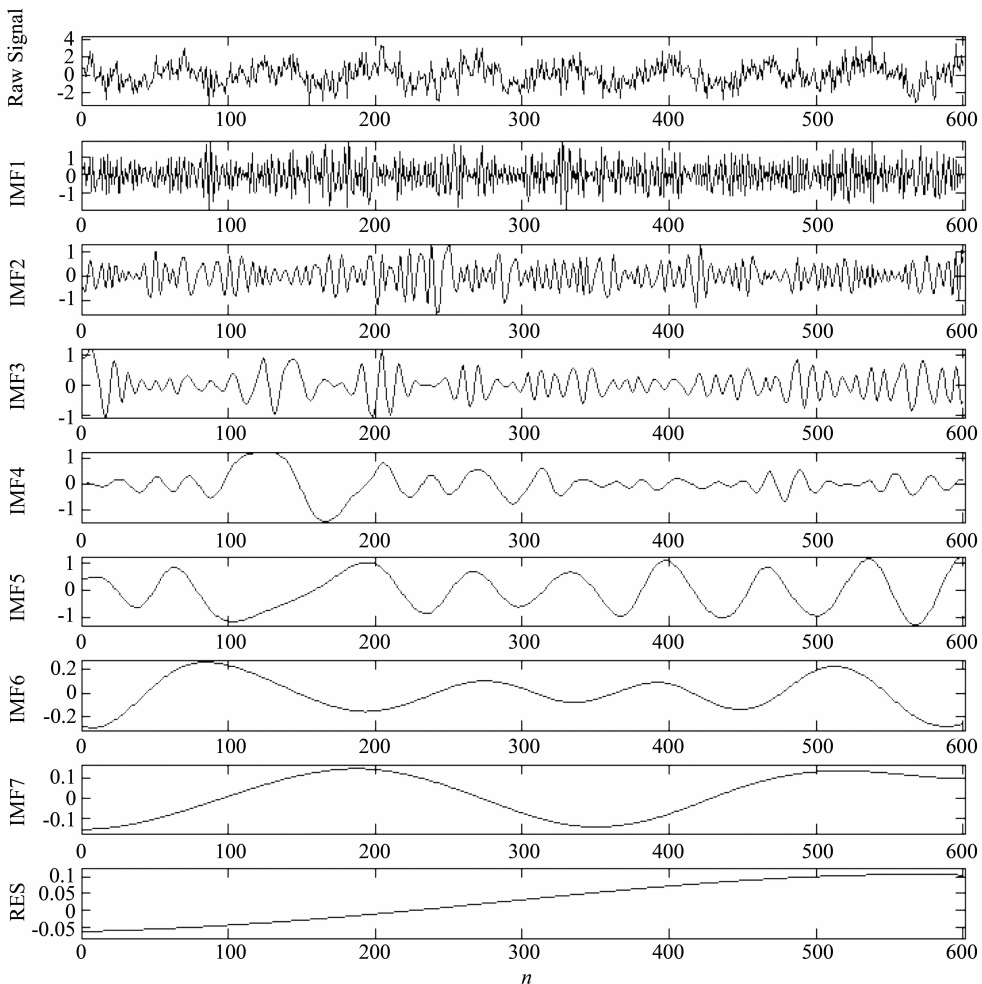


图 2 EMD 分解结果

Fig. 2 The results of EMD

最上方为原信号, 由正弦信号  $y = \cos(18\pi t/n)$  与高斯白噪声叠加而成。下面依次是分解产生的各个 IMF, 最后一个余量

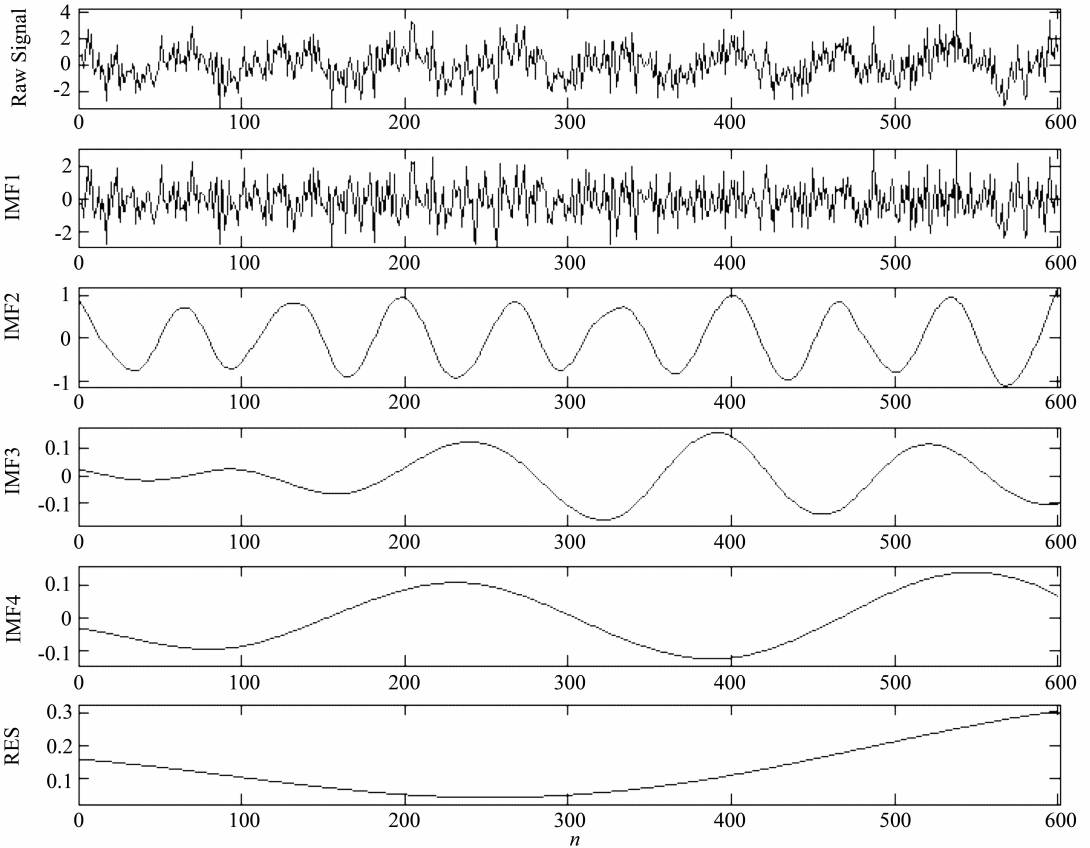


图 3 改进的 EMD 分解结果

Fig. 3 The results of the proposed improved EMD

最上方为原信号，由正弦信号  $y = \cos(18\pi t/n)$  与高斯白噪声叠加而成。下面依次是分解产生的各个 IMF，最后一个是余量

图 4 和图 5 则给出了原信号为正弦信号  $y = \cos(18\pi t/n) + \sin(10\pi t/n)$  与高斯白噪声叠加而成时，EMD 和改进的 EMD 分解结果。图 4 中的 IMF5 和 IMF6 分别与图 5 中的 IMF2 和 IMF3 对应，它们都是对原始信号中高低频率的两个正弦趋势的刻画。仔细观察，容易发现，图 4 中的 IMF5 在末端处有衰减，而 IMF6 在起始处也有衰减，其原因可能是受到了噪声成分的干扰；而图 5 中 IMF2 和 IMF3 与原始的两个正弦信号保持很强的一致性。该仿真实验进一步证实了改进的 EMD 算法比原 EMD 算法的更为可靠。

以上仿真实验结果表明，本文提出的改进 EMD 算法，对含有噪声的信号分解结果有很大的改进。

## 4 两个实际数据例子

EMD 方法很早就被应用于气候时间序列的分析（如文献 [1, 9]）。然而，由于影响气候的因素过于复杂，在数据分析处理时需要十分谨慎。为

了验证本文方法对 EMD 的改进效果，我们两种方法同时应用到两个实际的气候数据例子上。一个数据例子是 1979 - 2012 年纬向平均的每月的 500 mb 的温度异常值序列，另一个则是相同时间范围内的 200 mb 赤道纬向风 ( $165^\circ\text{W} - 110^\circ\text{W}$ ) 的原始数据。它们都源自网站 <http://www.cpc.ncep.noaa.gov/data/indices/>，具体分解结果如图 6 至图 9 所示。

对比图 6 和图 7，可知改进的 EMD 方法结果更简洁，所有噪声成分都集中在噪声 IMF 中（图 7 中 IMF1），而 EMD 方法结果中前两个 IMF 都是噪声成分。通过进一步观察，不难发现图 7 中 IMF3 - 6 与图 6 中的 IMF4 - 7 大致相对应，但是周期性规律更为显著。图 6 中的 IMF5 不同周期的长度变化剧烈，尤其在 1985 年到 1995 年之间的单个周期跨度太大，明显不合理，很可能是受到了噪声的严重干扰；而图 7 中的 IMF4 表现出合理的周期节律，而且幅度值更大，结果更为可信。

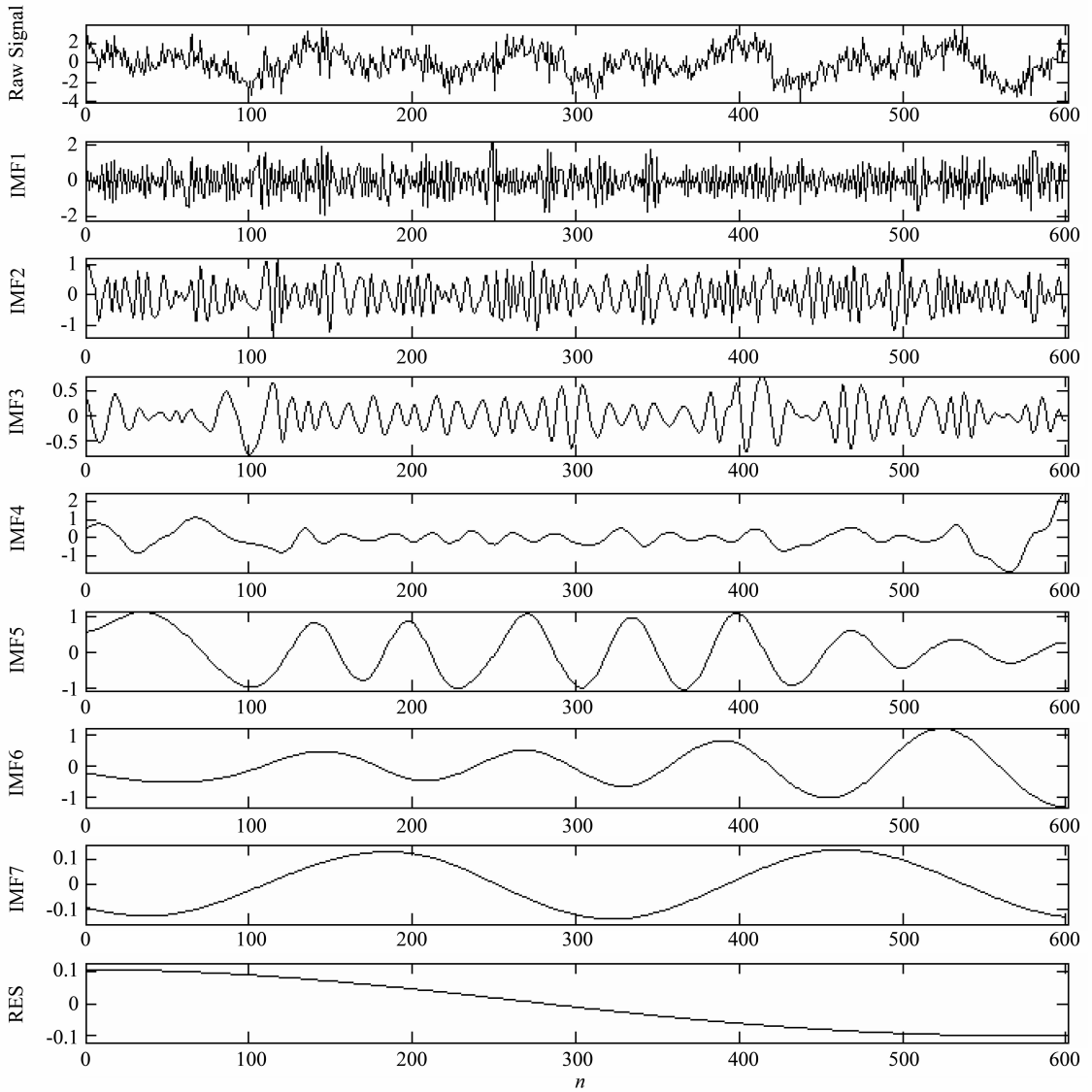
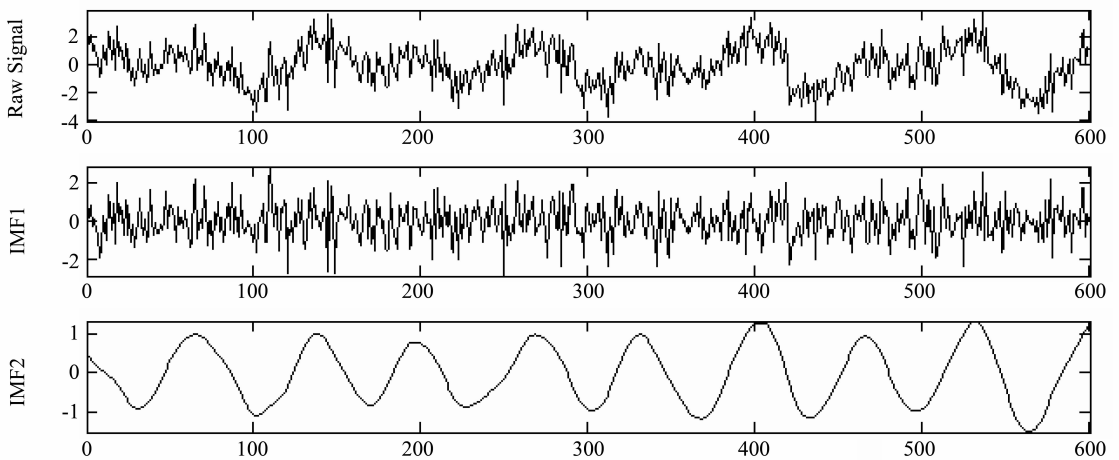


图 4 EMD 分解结果

Fig. 4 The results of EMD

最上方为原信号，由正弦信号  $y = \cos(18\pi t/n) + \sin(10\pi t/n)$  与高斯白噪声叠加而成；

下面依次是分解产生的各个 IMF，最后一个是余量



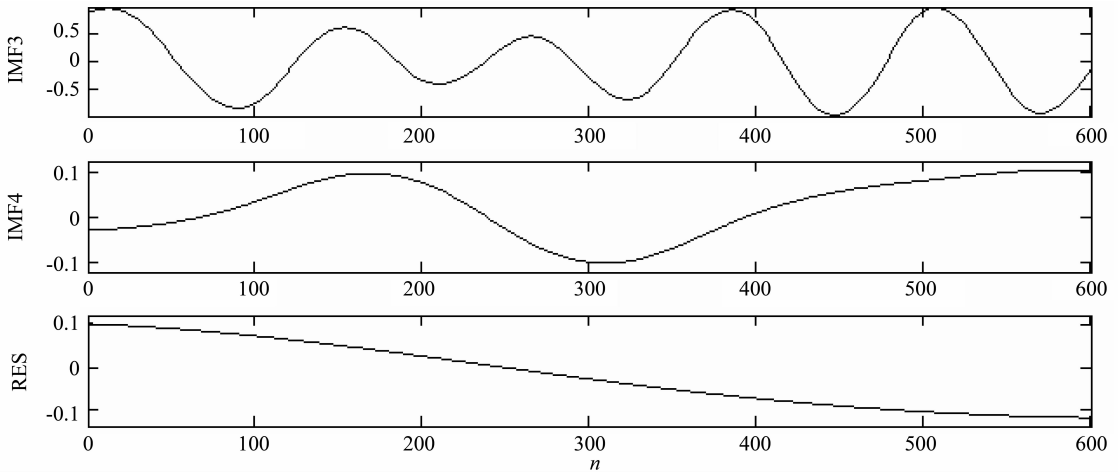


图 5 改进的 EMD 分解结果

Fig. 5 Results of the proposed improved EMD

最上方为原信号，由正弦信号  $y = \cos(18\pi t/n) + \sin(10\pi t/n)$  与高斯白噪声叠加而成；  
下面依次是分解产生的各个 IMF，最后一个是余量

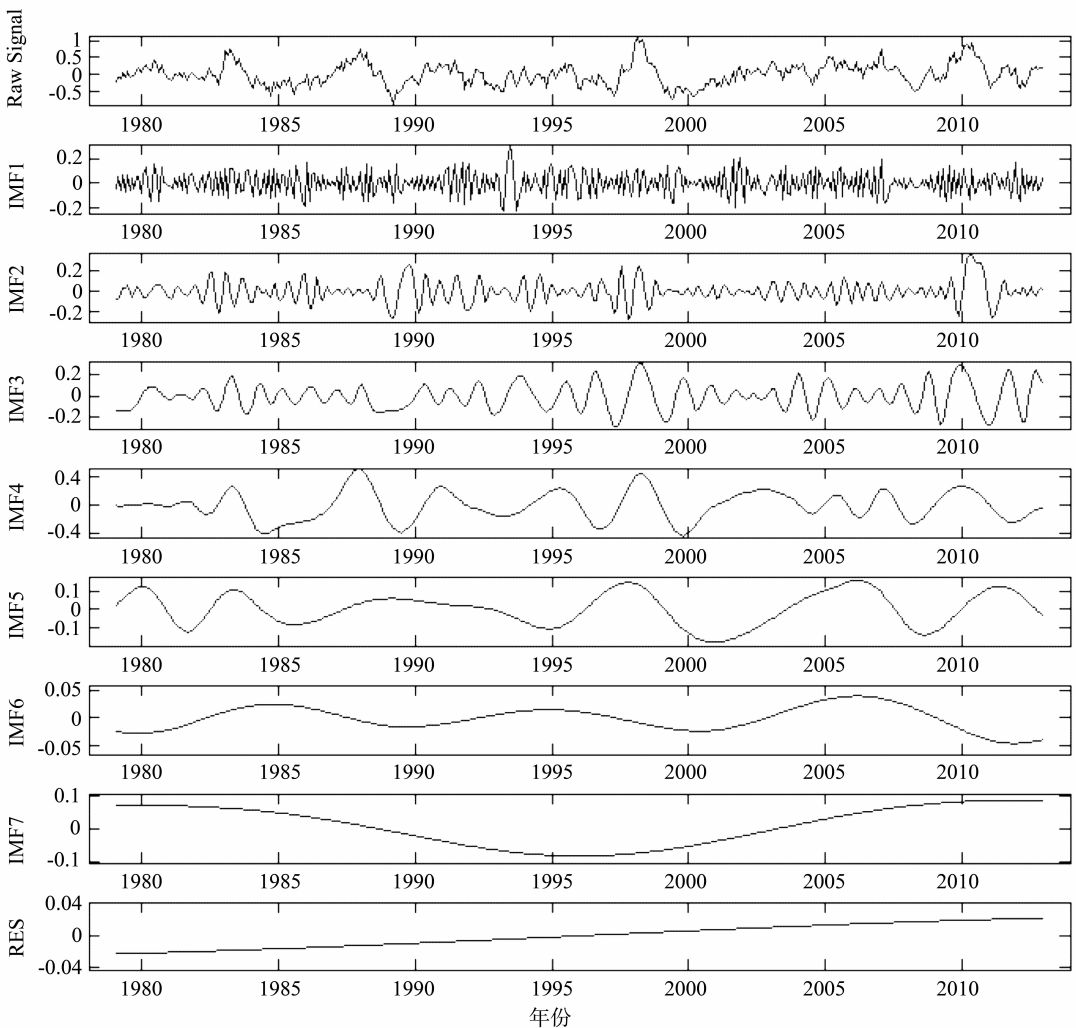


图 6 纬向平均的 500 mb 的温度异常值序列的 EMD 分解结果图

Fig. 6 The EMD decomposition result of the zonally average 500 mb temperature anomalies signal  
最上方是原始信号序列，下面依次是 EMD 分解得到的一系列 IMF 分量，最后一个是余量

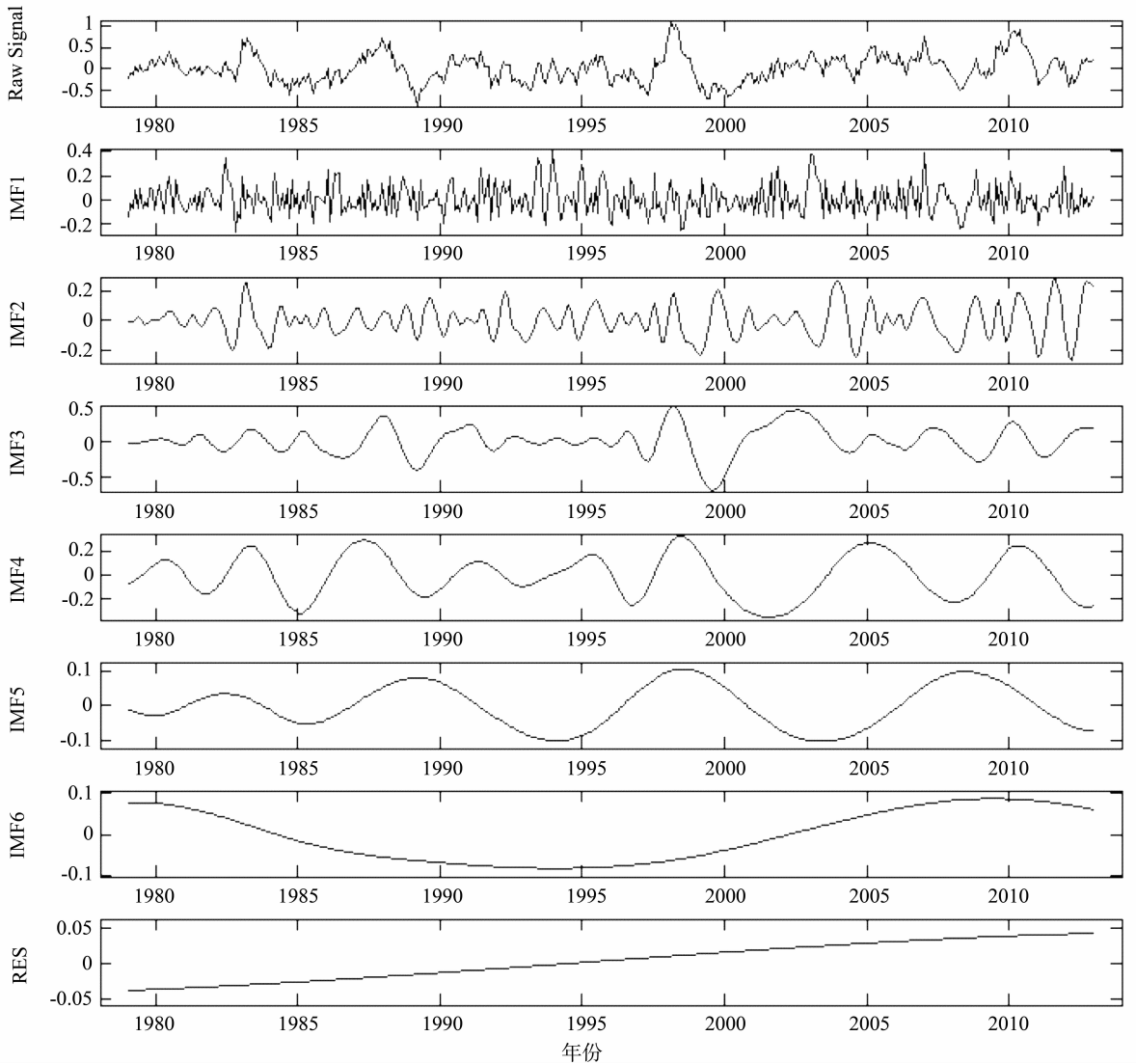


图 7 纬向平均的 500 mb 的温度异常值序列的改进的 EMD 分解结果图

Fig. 7 The decomposition of the zonally average 500 mb temperature anomalies signal by the proposed improved EMD  
最上面是原始信号序列，下面依次是分解得到的一系列 IMF 分量，最后一个是余量

图 8 和图 9 分别显示了不同方法对 200 mb 赤道纬向风的原始数据分解的结果。从图 8 最上方的原始数据来看，有明显的 1 年的周期性规律，EMD 的分解结果中 IMF2 与该规律最为相近，但差异较大，主要是其中夹杂了更快的小幅度震荡波。显而易见，这个现象是典型的“模式混叠”<sup>[10]</sup>；相比之下，图 9 中改进的 EMD 方法中的 IMF2 忠实地反映

了周期为 1 年的基本规律。

仔细对比图 8 和图 9 中其他 IMF 的表现，不难发现图 9 中 IMF 周期性明显更清晰：以连续两个极大或极小值之间的长度近似为一个周期，可计算从 IMF3 到 IMF7 平均周期分别近似为 2.5 年，4.8 年，11 年和 23 年。

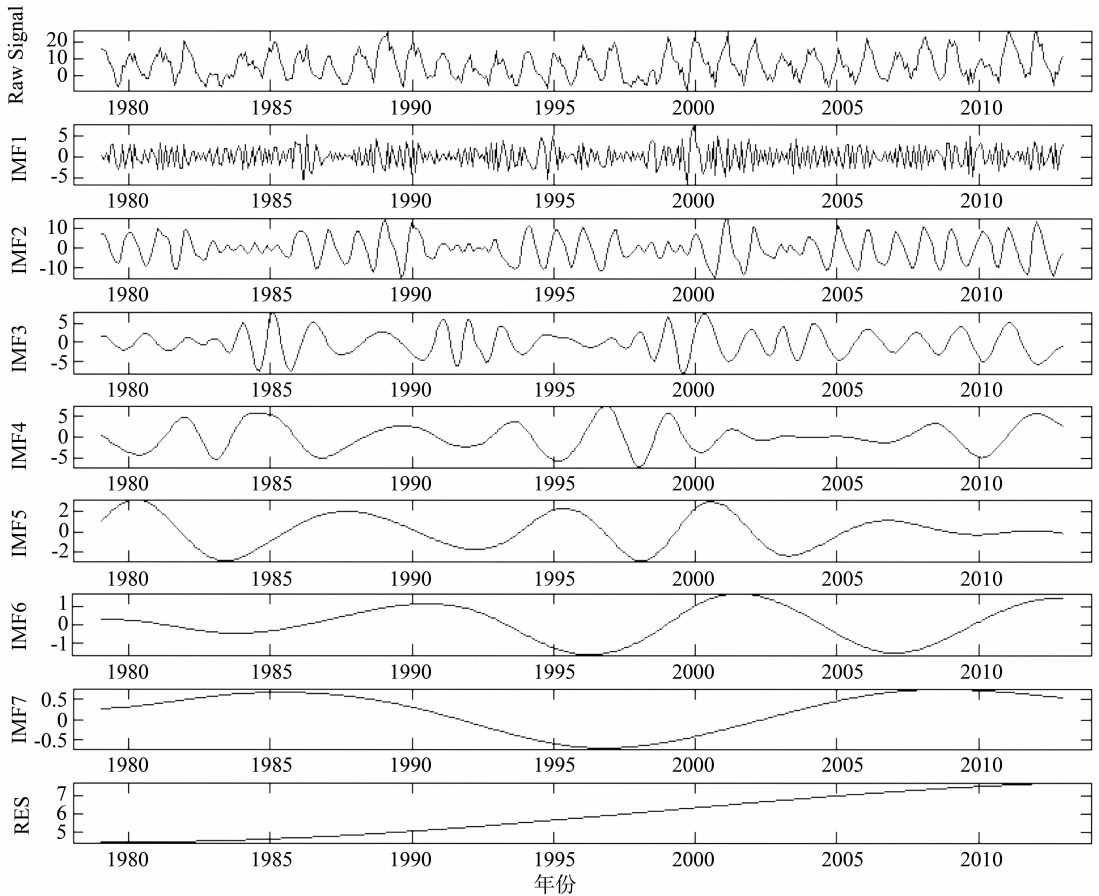


图8 200 mb 赤道纬向风的原始数据的 EMD 分解结果图

Fig. 8 The EMD decomposition result of the 200 mb Zonal Winds Equator ( $165^{\circ}\text{W} - 110^{\circ}\text{W}$ )

最上面是原始信号序列，下面依次是 EMD 分解一系列 IMF 分量，最后一个余量

## 5 结论

EMD 是一种适用于非平稳和非线性的信号分析方法，但它容易受噪声成分的干扰，常常表现出不稳定，从而导致分解结果不可靠。虽然文献 [11] 提出了一种利用 IMF 的能量密度与其平均周期乘积的分布来判别具体的 IMF 是噪声还是信号成分，但是 EMD 本身无法保证将噪声成分与信号成分很好的分离，比如图 8 中 IMF2 就属于这种情况，“模式混叠”是 EMD 分解中经常碰到的一个问题。如果通过其它手段抑制噪声成分，则会使得 EMD 分解更为有效。正是基于这一想法，本文提出了结合 GCV - 光滑样条的改进的 EMD 方法。该方法能克服原 EMD 方法对信号中噪声成分的过度分解的弊端，不仅使得分解结果更加简洁，而且极大地抑制了噪声对信号分解的干扰，从而使得结果更加可信。仿真实验和实际气候数据例子都证实了新提出的方法的有效性和优越性。

## 参考文献：

- [1] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1998, 454: 903 - 995.
- [2] PINES D, SALVINO L. Structural health monitoring using empirical mode decomposition and the Hilbert phase [J]. Journal of sound and vibration, 2006, 294(1): 97 - 124.
- [3] YUAN J S, ZHAO Z B, GAO Q, et al. Multimodal image registration based on empirical mode decomposition and mutual information [J]. Chinese Journal of Scientific Instrument, 2009, 10: 012.
- [4] 邵骏, 袁鹏, 颜志衡, 等. 基于 HHT 的雅鲁藏布江径流变化周期及趋势分析 [J]. 中山大学学报: 自然科学版, 2010, 49(1): 125 - 130.

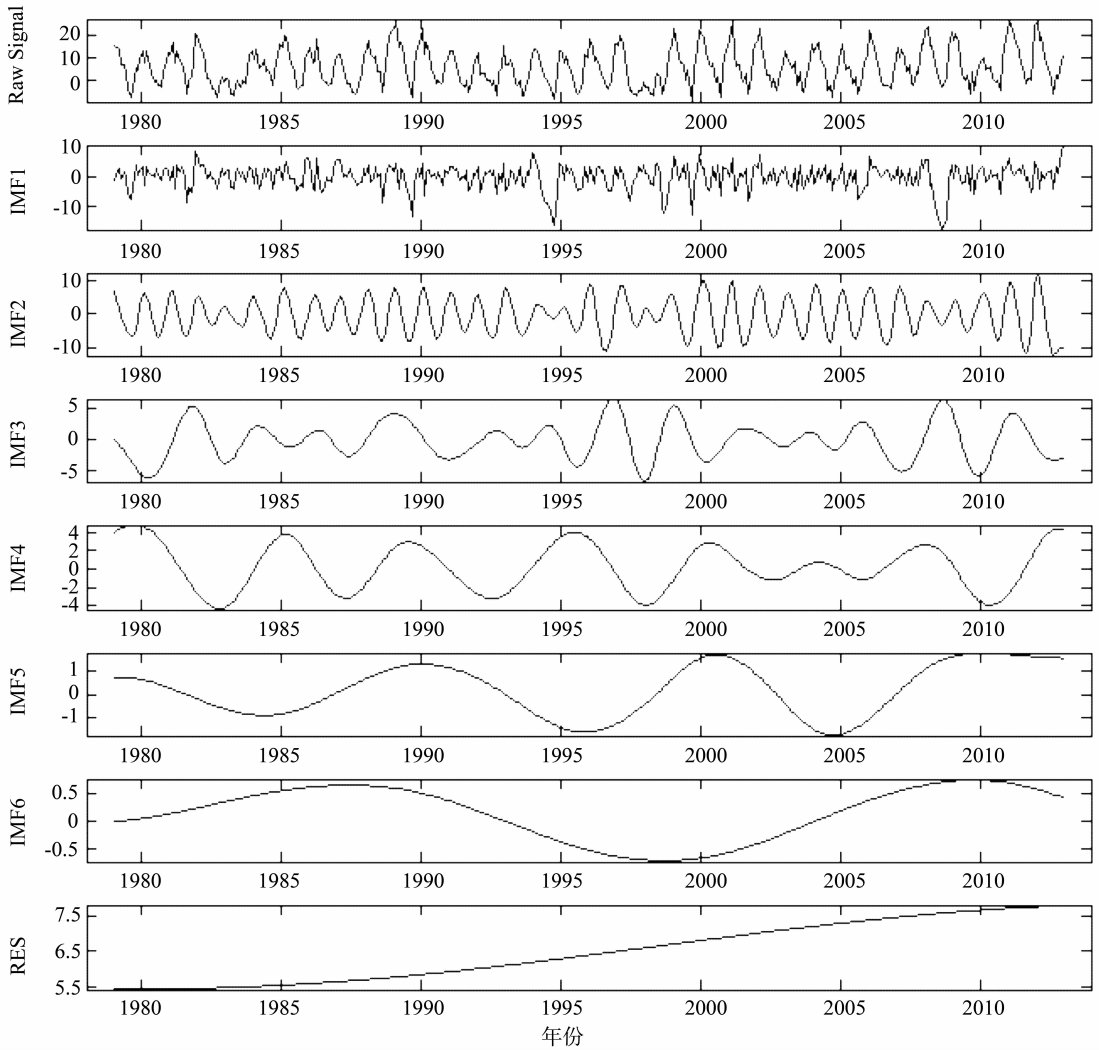


图9 200 mb 赤道纬向风的原始数据改进的 EMD 方法分解结果

Fig. 9 The decomposition of the 200 mb Zonal Winds Equator ( $165^{\circ}\text{W} - 110^{\circ}\text{W}$ ) by the proposed improved EMD  
最上面是原始信号序列,下面依次是 IMF 分量,最后一个是余量

- [5] WEI Y, CHEN M C. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks [J]. *Transportation Research Part C: Emerging Technologies*, 2012, 21(1): 148 - 162.
- [6] FLANDRIN P, GONÇALVES P, RILLING G. Detrending and denoising with empirical mode decomposition [C]// *Proceedings of the 12th European Signal Processing Conference (EUSIPCO'04)*, Vienna, Austria, September 2004.
- [7] WAHBA G. *Spline models for observational data* [M]. Philadelphia: SIAM, 1990.
- [8] 钟佑明, 秦树人. 希尔伯特 - 黄变换的统一理论依据研究[J]. *振动与冲击*, 2006, 25(3): 40 - 43.
- [9] COUGHLIN K T, TUNG K K. 11 - year solar cycle in the stratosphere extracted by the empirical mode decomposition method [J]. *Advances in space research*, 2004, 34(2): 323 - 329.
- [10] HU X Y, PENG S L, HWANG W L. EMD revisited: A new understanding of the envelope and resolving the mode-mixing problem in AM-FM signals [J]. *IEEE Transactions on Signal Processing*, 2012, 60(3): 1075 - 1086.
- [11] WU Z H, HUANG N E. A study of the characteristics of white noise using the empirical mode decomposition method [J]. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 2004, 460: 1597 - 1611.