2012

Compact empirical mode decomposition an algorithm to reduce mode mixing, end effect, and detrend uncertainty

Fan, Chenwu; Huang, Norden; Chu, Peter C.

Chu, P.C., C.W. Fan, and N. Huang, 2012: Compact empirical mode decomposition an algorithm to reduce mode mixing, end effect, and detrend uncertainty. Advances in Adaptive Data Analysis, 4 (3) 1250017 (18 pages) doi: 10.1142/S1793536912500173 (paper download).

https://hdl.handle.net/10945/36089

Advances in Adaptive Data Analysis
Vol. 4, No. 3 (2012) 1250017 (18 pages)

DOI: 10.1142/S1793536912500173

World Scientific
www.worldscientific.com

COMPACT EMPIRICAL MODE DECOMPOSITION: AN ALGORITHM TO REDUCE MODE MIXING, END EFFECT, AND DETREND UNCERTAINTY

PETER C. CHU*, CHENWU FAN* and NORDEN HUANG†

*Naval Ocean Analysis and Prediction Laboratory
Department of Oceanography
Naval Postgraduate School, Monterey, California, USA

†Research Center for Adaptive Data Analysis
National Central University, Chungli, Taiwan

Received
Accepted
Published

A compact empirical mode decomposition (CEMD) is presented to reduce mode mixing, end effect, and detrend uncertainty in analysis of time series (with N data points). This new approach consists of two parts: (a) highest-frequency sampling (HFS) to generate pseudo extrema for effective identification of upper and lower envelopes, and (b) a set of $2N$ algebraic equations for determining the maximum (minimum) envelope at each decomposition step. Among the $2N$ algebraic equations, $2(N-2)$ equations are derived on the base of the compact difference concepts using the Hermitan polynomials with the values and first derivatives at the $(N-2)$ non-end points. At each end point, zero third derivative and determination of the first derivative from several (odd number) nearest original and pseudo extrema provide two extra algebraic equations for the value and first derivative at that end point. With this well-posed mathematical system, one can reduce the mode mixing, end effect, and detrend uncertainty drastically, and separate scales naturally without any *a priori* subjective criterion selection.

Keywords: Compact empirical mode decompositions (CEMD); empirical mode decomposition (EMD); highest-frequency sampling (HFS); pseudo extrema; compact difference; highest-frequency sampling (HFS); Hermitian polynomials; intrinsic mode function (IMF); end effect; detrend uncertainty; mode mixing.

1. Introduction

Most of the time series we encounter are nonlinear and nonstationary. Because of the limitations of available methodologies for analyzing data, the crucial phase of data analysis has in the past been relegated to "data processing", where data are routinely put through certain well-established algorithms to extract some standard parameters. These traditional methods usually assume linearity and stationarity of the time series. Since the seminal paper by Huang *et al.* (1998), adaptive empirical

model decomposition (EMD) has been developed to analyzing nonlinear and nonstationary data. The EMD decomposes a nonlinear and nonstationary signal into several intrinsic mode functions (IMFs) and a trend. Instantaneous frequency can be obtained by the Hilbert–Huang transform (HHT), and then the time–frequency–energy distribution characteristics. An IMF is a function that must satisfy two conditions according to the algorithm originally developed: (a) the difference between the number of local extrema and the number of zero-crossings must be zero or one; (b) the running mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. For such a signal, the interior extrema are easily identified. However, these extrema are not enough to determine two well-behaved fitting spline envelopes near the two ends for the sifting, especially in the cases when the total number of splines are small, for the extrapolation of a spline often leads to undesirable big error especially near the end points.

The easiest way is to treat the two end points as "frozen" points, i.e. the two end points are on both maximum and minimum envelopes. Such a treatment makes the trend varying from the first end point to the last end point. Other practice is to extend data points beyond the end points so as to carry out the spline envelope fitting over and even beyond the existing data range; such as the wave extension method [Huang *et al.* (1998)], local straight-line extension method [Wu and Huang (2009)], mirror or anti-mirror extension [Zhao and Huang (2001)], and self-similarity [Wang *et al.* (2007)], to name a few, have been used. While methods for extending data vary, the essence of all these methods is to predict data, a dauntingly difficult procedure even for linear and stationary processes. The problem that must be faced is how to make predictions for nonlinear and nonstationary stochastic processes. The original signal only has the extrema in the data series, so the extending points beyond the two end points are not ***real***. Therefore the data extension methods will not solve the problem no matter how much efforts have been spent. Besides, the end error may propagate from the ends to the interior of the data span that would cause severe deterioration of the IMFs obtained.

An ensemble EMD was proposed to solve the irregular distribution of local maxima (minima) problem [Wu and Huang (2009)]. This approach consists of sifting an ensemble of white noise-added signal (data) and treats the mean as the final true result. The reason to add finite, not infinitesimal, amplitude white noise is to obtain relatively uniform distribution of local maxima and minima such that the upper and lower envelope can be identified.

Questions arise: Can relatively uniform distribution of local maxima and minima be established at each EMD step without adding any white noise? Can the end effect be eliminated without using either extrapolation or interpolation (with extra point beyond the end point)? Can the upper and lower envelopes be determined in a systematic way? Can the trend be determined objectively? These problems will be answered in this study through identification of highest-frequency pseudo extrema and use of compact difference concepts [Chu and Fan (1998), (1999)]. With highest-frequency pseudo extrema, relatively uniform distributions of local

(original and pseudo) maxima and minima are found. Determination of either upper or lower envelope becomes to solve a set of $2N$ algebraic equations of the values and first derivatives for each envelope at all N data points (including the two end points) using the Hermitian polynomials. This method, called the compact EMD (CEMD), shows evident improvement in data analysis. The rest of the paper is organized as follows. Section 2 introduces the classical EMD. Section 3 depicts procedure of CEMD. Section 4 demonstrates the reduction of the end-point effect by CEMD. Section 5 shows the enhancement of detrend capability by CEMD. Section 6 presents the conclusions.

2. EMD

Let $x(t)$ represent the time series with fluctuations on various time scales (Fig. 1). The EMD method is depicted as follows. First, the local minima and maxima of the signal $x(t)$ are identified. Second, the local maxima are connected together by a cubic spline interpolation (other interpolations are also possible), forming an upper envelope $e_{\max}(t)$. The same is done for local minima, providing a lower envelope $e_{\min}(t)$. Third, the mean of the two envelopes are calculated

$$m_1(t) = [e_{\max}(t) + e_{\min}(t)]/2. \tag{1}$$

Fourth, the mean is subtracted from the signal, providing the local detail

$$h_1(t) = x(t) - m_1(t), \tag{2}$$

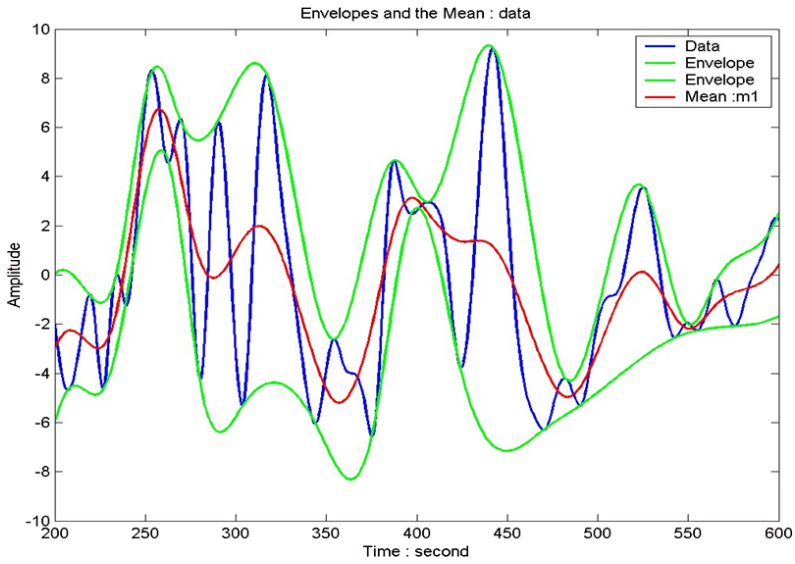

Fig. 1. Procedure of traditional EMD.

which is then considered to check if it satisfies the above two conditions to be an IMF. If yes, it is considered as the first IMF and denoted

$$c_1(t) = h_1(t). \tag{3}$$

It is subtracted from the original signal and the first residual,

$$r_1(t) = x(t) - c_1(t), \tag{4}$$

is taken as the new series in step 1. If $h_1(t)$ is not an IMF, a procedure called "sifting process" is applied as many times as necessary to obtain an IMF. In the sifting process, $h_1(t)$ is considered as the new data, and the same procedure applies. The IMFs are orthogonal, or almost orthogonal functions (mutually uncorrelated). This method does not require stationarity and linearity of the data and is especially suitable for nonstationary and nonlinear time series analysis.

By construction, the number of extrema decreases when going from one residual to the next; the above algorithm ends when the residual has only one extrema, or is constant, and in this case no more IMF can be extracted; the complete decomposition is then achieved in a finite number of steps. The signal $x(t)$ is finally written as the sum of mode time series $c_i(t)$ and the residual $r_m(t)$:

$$x(t) = \sum_{i=1}^{m} c_i(t) + r_m(t). \tag{5}$$

There is no any oscillation (i.e. nonexistence of both maximum and minimum envelopes) in the residue $r_m(t)$, which should represent the trend. Obviously, successfulness of the EMD depends on accurate determination of upper and lower envelopes. Three long recognized difficulties (end-point effect, mode mixing, and detrend uncertainty) are related to each other and caused by uncertain identification of upper and lower envelopes due to (a) uncertain maximum and minimum values at the end points and (b) irregular distribution of local maximum (minimum) points.

3. CEMD

A new CEMD method is developed on the base of (a) compact difference concept [e.g. Chu and Fan (1998), (1999)] using Hermitian polynomials and (b) identification of uniformly distributed pseudo maximum and minimum points using highest-frequency sampling. CEMD determines upper and lower envelopes objectively and accurately at all N data points including the two end points.

3.1. *Basic algebraic equations*

Let $\{x_i, i = 1, 2, \ldots, N\}$ be the original time series with corresponding time instances $\{t_i, i = 1, 2, \ldots, N\}$ and two end points (x_1, x_N). Let $(x_j^{(\max)}, j = 1, 2, \ldots, J)$ and $(x_k^{(\min)}, k = 1, 2, \ldots, K)$ be the local maxima and minima with $J = K$

or differing at most by one; and $\{u_i, i = 1, 2, \ldots, N\}$ and $\{l_i, i = 1, 2, \ldots, N\}$ be the corresponding upper and lower envelopes. Here, the upper envelope is taken as an example for illustration. On the upper envelope, the values are given at the local maxima ($x_j^{(\max)}$), and unknown at the other time steps (Fig. 2). For simplicity without loss of generality, a uniform time step $\Delta t = t_{i+1} - t_i\,(i = 1, 2, \ldots, N-1)$ is assumed.

Let (u_i, u_i') be the value and first derivative of the upper envelope at the time t_i. For an interval $[t_{i-1}, t_{i+1}]$, $(i = 2, 3, \ldots, N-1)$, a quartic Hermitian polynomial,

$$p_4(\xi) = a_0 + a_1\xi + a_2\xi^2 + a_3\xi^3 + a_4\xi^4, \quad \xi = t - t_i, \tag{6}$$

is used to determine (u_i, u_i') simultaneously (Fig. 3) with

$$\begin{gathered} p_4(-\Delta t) = u_{i-1}, \quad p_4(0) = u_i, \quad p_4(\Delta t) = u_{i+1}, \\ p_4'(-\Delta t) = u_{i-1}', \quad p_4'(\Delta t) = u_{i+1}'. \end{gathered} \tag{7}$$

Substitution of (6) into (7) leads to the expressions for the coefficients of the Hermitian polynomials (6),

$$\begin{aligned} a_0 &= u_i, \\ a_1 &= \frac{3}{2}\left(\frac{u_{i+1} - u_{i-1}}{2\Delta t}\right) - \frac{1}{4}(u_{i+1}' + u_{i-1}'), \\ a_2 &= \frac{u_{i-1} - 2u_i + u_{i+1}}{\Delta t^2} - \frac{1}{2}\left(\frac{u_{i+1}' - u_{i-1}'}{2\Delta t}\right), \end{aligned}$$

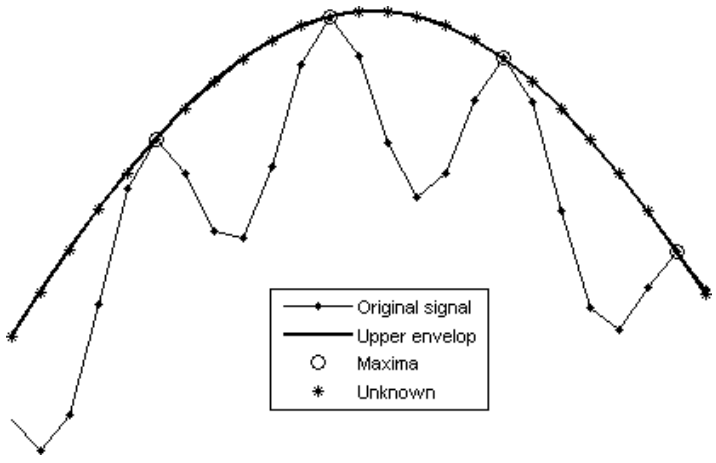

Fig. 2. Upper envelope with maxima and unknown points (to be determined).

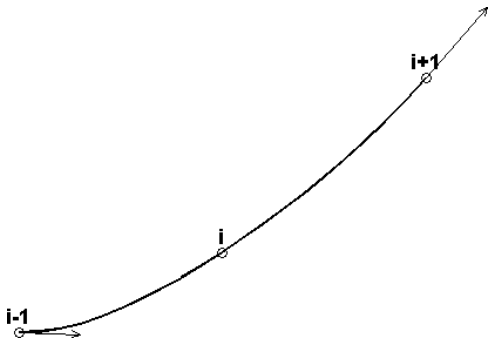

Fig. 3. Along the upper envelope, the coefficients of the quartic Hermitian polynomial at the point i, is calculated from u_{i-1}, u_i, u_{i+1}, u'_{i-1}, and u'_{i+1}.

$$a_3 = \frac{1}{4\Delta t^2}\left(u'_{i-1} - \frac{u_{i+1}-u_{i-1}}{\Delta t} + u'_{i+1}\right),$$
$$a_4 = \frac{1}{2\Delta t^2}\left(\frac{u'_{i+1}-u'_{i-1}}{2\Delta t} - \frac{u_{i-1}-2u_i+u_{i+1}}{\Delta t^2}\right). \tag{8}$$

First derivative of (6) at time t_i (i.e. $\xi = 0$) gives

$$p'_4(0) = a_1 = u'_i. \tag{9}$$

Substitution of a_1 from (8) into (9) leads to

$$u'_{i-1} + 4u'_i + u'_{i+1} = 6\frac{u_{i+1}-u_{i-1}}{2\Delta t}. \tag{10}$$

Since u_i is located at the upper envelope, it may be or may not be a local maximum. If u_i is a local maximum, it has a given value,

$$u_i = x_{j_i}^{(\max)}. \tag{11a}$$

If u_i is not a local maximum, an additional relation is needed for u_i. This relation can be obtained by reducing the order of the Hermitian polynomial from quartic to cubic (i.e. the fourth derivative equals zero), or letting $a_4 = 0$ in (6), which leads to

$$\frac{u_{i-1}-2u_i+u_{i+1}}{\Delta t} = \frac{u'_{i+1}-u'_{i-1}}{2}. \tag{11b}$$

Here, Eq. (8) is used. The values and first derivatives (u_i, u'_i) for all the interior points $(i = 2, 3, \ldots, N-1)$ are obtained by solving $2 \times (N-2)$ algebraic equations expressions (10), and (11a) or (11b) with the given values of local maxima $x_j^{(\max)}$. Two more algebraic equations are needed for each end point ($i = 1$, or N). The

first equation is obtained from the assumption that the third derivative vanishes at the two end points,

$$u_1''' = 0, \quad u_N''' = 0. \tag{12}$$

Setting the quartic Hermitian polynomial (6) at t_2 (i.e. $i = 2$), the third derivative at t_1 is given by

$$\frac{d^3 p_4(-\Delta t)}{dt^3} = 6a_3 + 24a_4 \times (-\Delta t) = 0. \tag{13}$$

Substitution of (8) with $i = 2$ into (13) leads to

$$\frac{9u_1 - 16u_2 + 7u_3}{\Delta t} = 3u_3' - 5u_1'. \tag{14}$$

Similarly, at t_N we have

$$\frac{7u_{N-2} - 16u_{N-1} + 9u_N}{\Delta t} = 5u_N' - 3u_{N-2}'. \tag{15}$$

The second equation for each end point is obtained from the assumption that the first derivatives (G_1, G_N) can be calculated at the two end points,

$$u_1' = G_1, \quad u_N' = G_N. \tag{16}$$

3.2. *Determination of G_1, G_N*

An odd number (L) of extrema points ($x_l^{(\mathrm{ext})}, l = 1, 2, \ldots, L$) close to each end point are used to determine G_1 and G_N with a weighted least squares (WLS) regression. Taking G_1 as an example, the WLS regression is represented by

$$\hat{x}_l^{(\mathrm{ext})} = C + G_1 t_l. \tag{17}$$

In the WLS regression, cases with greater weights contribute more to the fit of the regression line. The result is that the estimated coefficients are usually very close to what they would be in ordinary least square regression, but under WLS regression their standard errors are smaller. Figure 4 shows the example with 5 extrema points for determining G_1. The weight is 0.5 at the first and last extrema points and 1.0 otherwise. Thus, G_1 is a function of ($x_l^{(\mathrm{ext})}, l = 1, 2, \ldots, L$),

$$G_1 = F_1(x_1^{(\mathrm{ext})}, x_2^{(\mathrm{ext})}, \ldots, x_L^{(\mathrm{ext})}). \tag{18}$$

Similarly, G_N is a function of ($x_l^{(\mathrm{ext})}, l = M, M-1, \ldots, M-L+1$),

$$G_N = F_N(x_M^{(\mathrm{ext})}, x_{M-1}^{(\mathrm{ext})}, \ldots, x_{M-K+1}^{(\mathrm{ext})}). \tag{19}$$

Solutions of the set of $2N$ algebraic equations (10), (11a), (11b), (14), (15), and (16) with (G_1, G_N) are calculated by (18) and (19) give $2N$ values of (u_i, u_i') for all the points ($i = 1, 2, \ldots, N$).

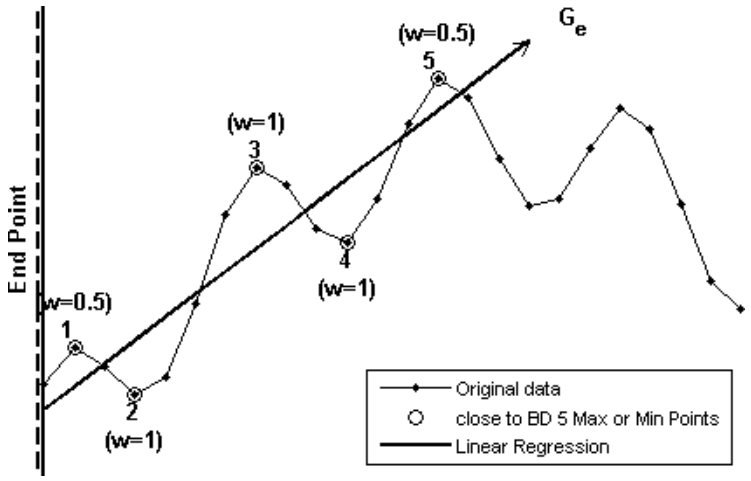

Fig. 4. Determination of G_1 using the WLS regression.

3.3. *Pseudo extrema identification*

The data with uneven distribution of local maximum (minimum) points is illustrated in Fig. 5. The data has its fundamental part as a low-frequency sinusoidal wave with unit amplitude. At the two middle crests and first three troughes of the low-frequency wave, high-frequency intermittent oscillations with amplitude of 0.1 are riding on the fundamental [Fig. 5(a)]. Local maximum and minimum points are distributed unevenly. The high and low frequency components are shown in Figs. 5(b) and 5(c). This signal has no trend [Fig. 5(d)]. To solve such a problem, Wu and Huang (2009) proposed an ensemble empirical mode decomposition (EEMD) with sifting an ensemble of white noise-added signal (data) and treating the mean as the final true result. This method creates extra local maxima and minima in such a way that a uniform reference frame in the time–frequency space is provided.

Different from EEMD, unevenly distributed local maximum and minimum points can be solved by identification of pseudo maximum and minimum points at each step of the EMD analysis. Let δ be the minimum value of the two smallest half time period of two neighboring maximum points (δ_1) and minimum points (δ_2),

$$\delta = \min(\delta_1, \delta_2), \tag{20}$$

which represents highest frequency fluctuation. For a neighboring local maximum–minimum pair away from the end points with Δ the time period in between (Fig. 6), the integer p is calculated by

$$p = \frac{1}{2}\left[\frac{\Delta}{\delta} - 1\right], \tag{21}$$

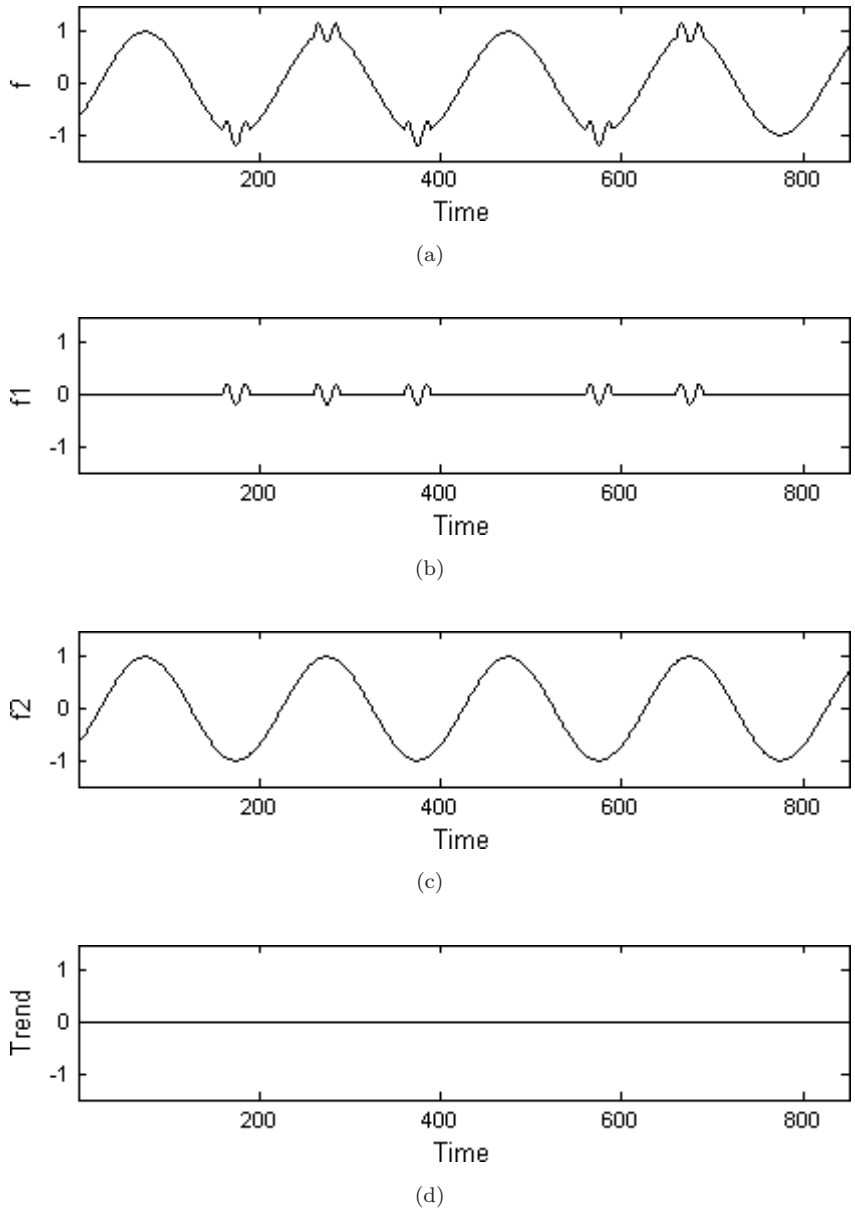

Fig. 5. Time series containing a high frequency component and a low frequency component without trend: (a) signal, (b) high frequency component, (c) low frequency component, and (d) zero trend.

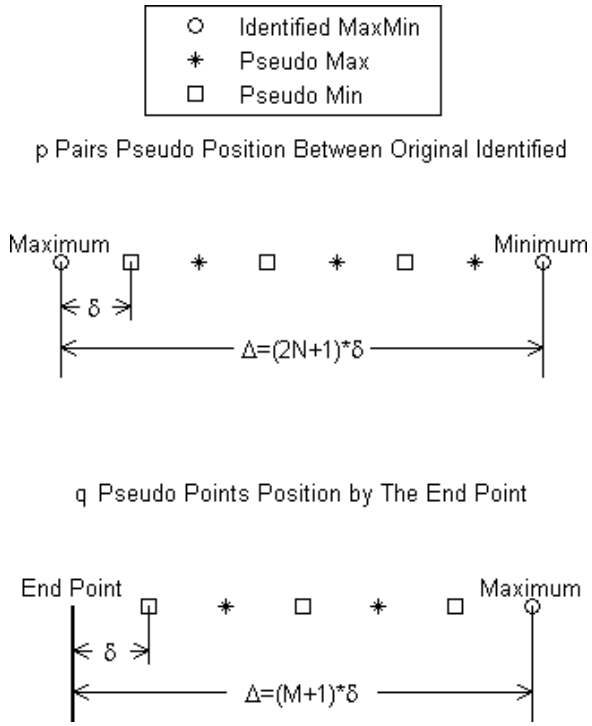

Fig. 6. Pseudo maxima and minima generated by highest-frequency sampling for the upper envelope: (a) local maximum away from the end point, and (b) local maximum next to the end point. It is noted that similar procedure is adopted if the local minimum on the left side of the pair.

where the bracket [ ] shows the integer part of the real number inside the bracket. Pseudo maximum–minimum pairs (p-pairs) are identified with the time period,

$$\delta = \frac{\Delta}{2p+1}, \tag{22}$$

between the two neighboring pseudo maximum and minimum. For the local maximum or minimum next to the end points (local maximum point as an example in Fig. 6), the integer (q) is calculated by

$$q = \left[\frac{\Delta}{\delta}\right] - 1.$$

Pseudo maximum and minimum points are identified between local maximum (or minimum) and the end point with the time period,

$$\delta = \frac{\Delta}{q+1}, \tag{23}$$

between the two neighboring pseudo maximum and minimum.

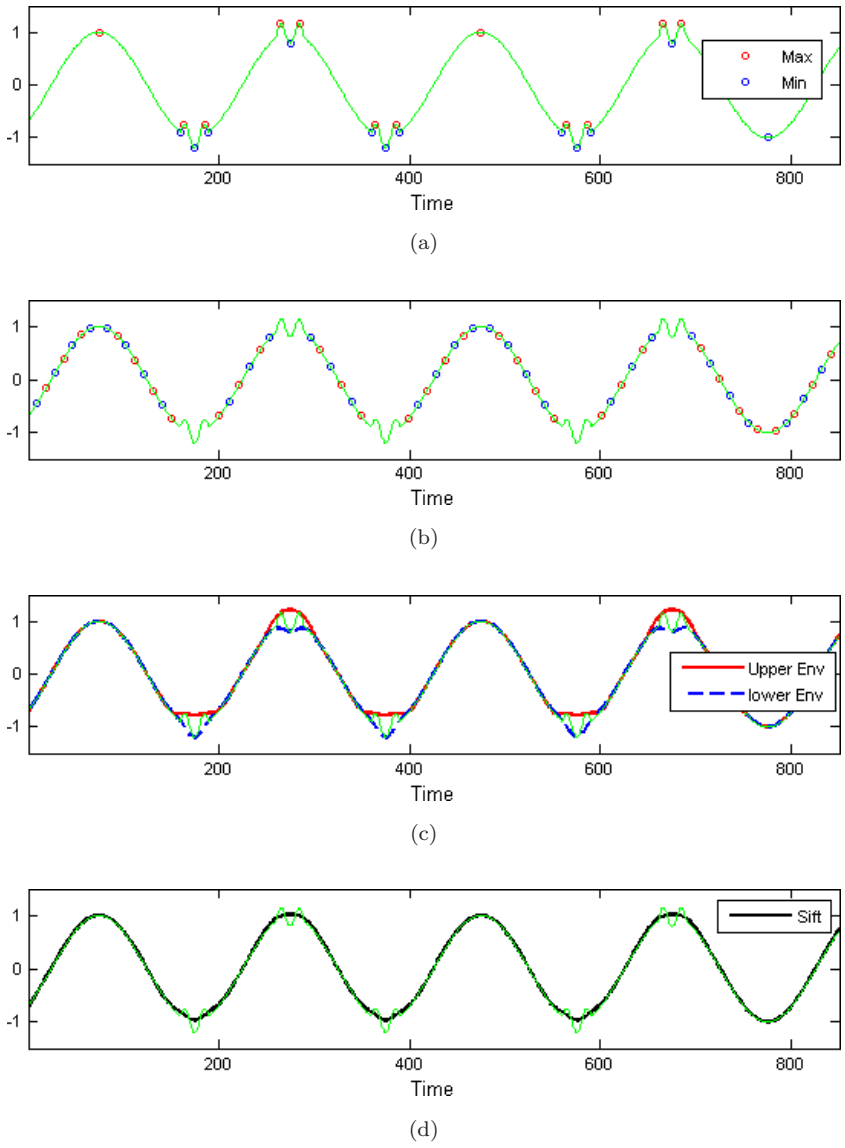

Fig. 7. Identification of pseudo maximum and minimum points by highest frequency sampling: (a) original signal [same as in Fig. 6(a)], (b) distribution of pseudo maxima (red circles) and minima (blue circles), (c) upper (red) and lower (blue) envelopes, and (d) sifting.

The identified pseudo extrema have the following features: (a) the frequency of the pseudo maximum (or minimum) points is nearly the same and not higher than the highest frequency of original signal; (b) the total (i.e. original and pseudo) maximum and minimum points must be alternatively distributed [Fig. 7(a)]; (3) the pseudo maximum and minimum still keep the original values; and (4) the total

(original and pseudo) extrema will be distributed uniformly [Fig. 7(b)]. With the uniform distributed total extrema, the upper and lower envelopes can be objectively identified [Fig. 7(c)]. After sifting, the high-frequency component is easily obtained [Fig. 7(d)].

3.4. *Elimination of mode mixing*

The CEMD is conducted on the data shown in Fig. 5(a). Two IMFs and a residue (zero line) are obtained (Fig. 8). Comparison between Figs. 8 and 5 shows the capability of the CEMD to analyze the original time series, $f(t) = f_1(t) + f_2(t)$, shown in Fig. 5. During the comparison, the correlation coefficient (CC) and the relative root mean square error (RRMSE),

$$\text{RRMSE} = \frac{\sqrt{\sum_{i=1}^{N} [T_i^{(e)} - T_i^{(o)}]^2}}{\sqrt{\sum_{i=1}^{N} [T_i^{(o)} - \bar{T}^{(o)}]^2}}, \tag{24}$$

are used with $T_i^{(o)}$and $T_i^{(e)}$ being the time series of components (trend and components/IMFs) of original and analyzed data, N the number of data points, and $\bar{T}^{(o)}$ the mean of $\{T_i^{(o)}\}$.

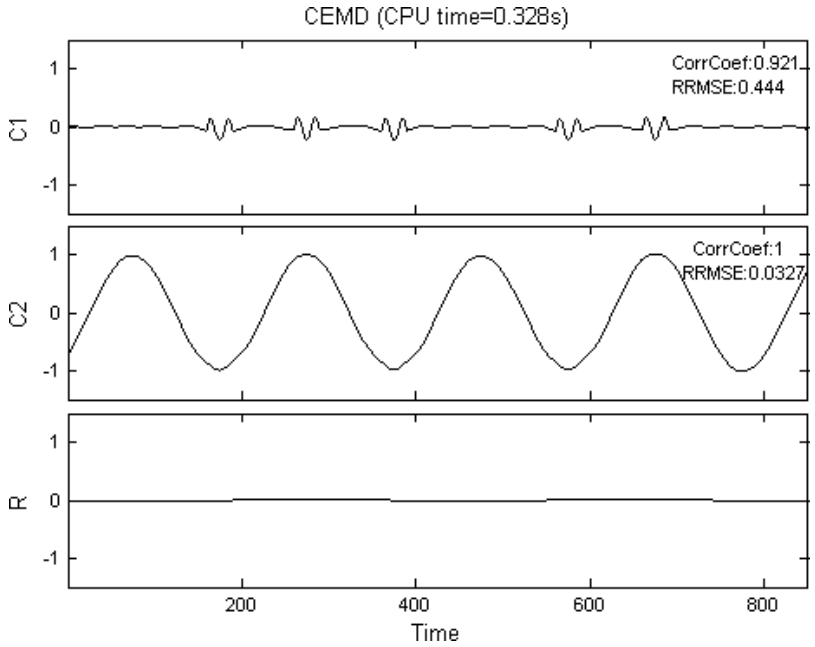

Fig. 8. CEMD analysis on the time series $\{x_i\}$ given in Fig. 5(a) with IMF-1 resembling the high-frequency component $f_1(t)$ in Fig. 5(b), IMF-2 resembling the low-frequency component $f_2(t)$, and no trend and IMF-2 resembling. It clearly shows the capability of eliminating mode mixing, end effect, and detrend uncertainty.

The original two components (f_1 and f_2) in Fig. 5 are fully recovered by the CEMD. The IFM-1 [i.e. $c_1(t)$] resembles $f_1(t)$ with CC of 0.92 and RRMSE of 0.44. The IFM-2 [i.e. $c_2(t)$] resembles $f_2(t)$ in Fig. 5 with CC of 1.0 and RRMSE of 0.03. Moreover, the CEMD provides the one-to-one correspondence of the components between original analyzed data: $c_1(t)$ to $f_1(t)$, $c_2(t)$ to $f_2(t)$, zero trend to zero trend.

4. Trend Determination

4.1. *Signal with a trend*

A time series consisting of a quadratic trend and four harmonics,

$$\begin{aligned} x(t_i) &= f_0(t_i) + \sum_{k=1}^{3} f_k(t_i), \quad f_0(t_i) = A_0 t_i^2, \\ f_k(t_i) &= A_k \sin(\omega_k t_i + \varphi_k), \quad k = 1, 2, 3 \\ x_i &= x(t_i), \quad t_i = (i-1)\Delta t, \quad t_1 = 0, \\ t_N &= 0.9\,\text{s}, \quad \Delta t = 0.0018\,\text{s}, \quad N = 501, \end{aligned} \tag{25}$$

is used to demonstrate the end effect in EMD. The parameters in (25) are given in Table 1. Figure 9 shows the time series of $\{x_i\}$ with the two end points with the values of

$$x_1 = 0.1096, \quad x_N = 0.4495. \tag{26}$$

Table 1. Values of parameters used in Eq. (25).

k	0	3	2	1
A_k	0.5	1.0	0.5	0.20
ω_k		6π (3 Hz)	40π (20 Hz)	100π (50 Hz)
φ_k		0.01	0	0.005

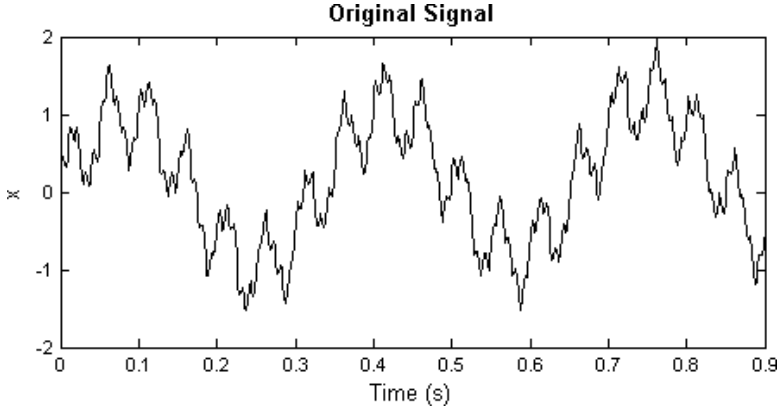

Fig. 9. Time series $\{x_i\}$ represented by Eq. (25).

Figure 10 shows the trend and harmonics of the data shown in Fig. 9: $f_{0i} = f_0(t_i)$, $f_{1i} = f_1(t_i)$, $f_{2i} = f_2(t_i)$, and $f_{3i} = f_3(t_i)$. Obviously, only $f_0(t_i)$ represents the trend of the signal $\{x_i\}$. The trend of $\{x_i\}$ varies from f_{01} to f_{0N},

$$f_{01} = 0, \quad f_{0N} = 0.405. \tag{27}$$

The CEMD is conducted on the time series $\{x_i\}$ (shown in Fig. 9) to obtain three IMFs and a trend (Fig. 11). Obviously, IMFs well correspond to the harmonics, $c_i(t)$ versus $f_i(t)$, with high CCs (0.9992 between c_1 and f_1, 0.9836 between c_2 and f_2, 0.9992 between c_3 and f_3, 0.9890 between two trends) and low RRMSEs (0.1834 between c_1 and f_1, 0.082 between c_2 and f_2, 0.0441 between c_3 and f_3, 0.172 between two trends).

4.2. *Existence of noises*

Comparison between trends from the original data [i.e. the quadratic trend $f_0(t_i)$], the CEMD (Fig. 11) clearly shows perfect detrend using CEMD and false detrend

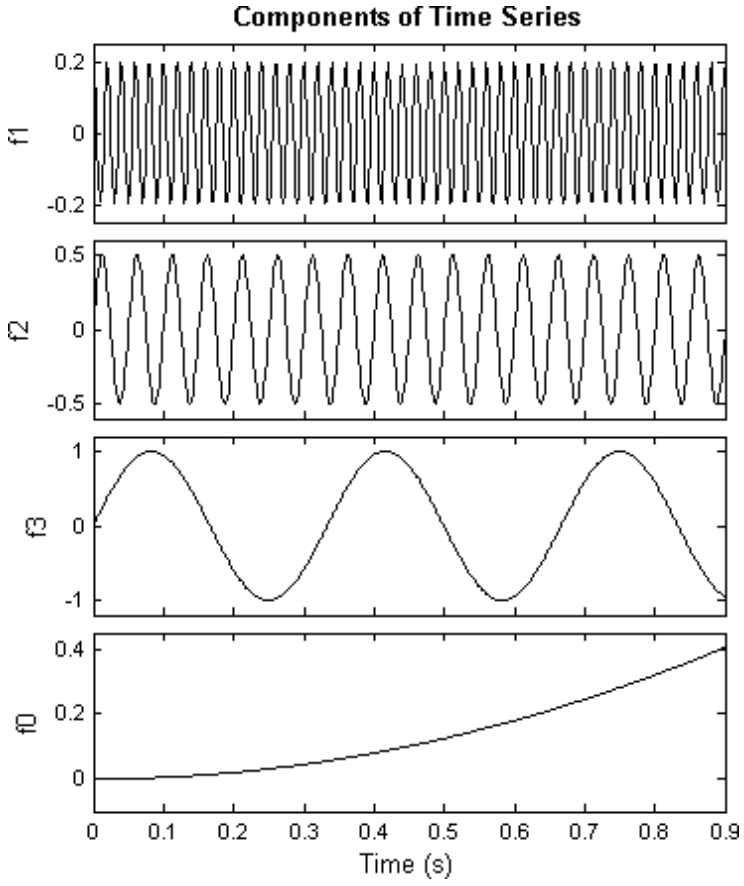

Fig. 10. Components of time series $\{x_i\}$: a quadratic trend and four harmonics given by Eq. (25).

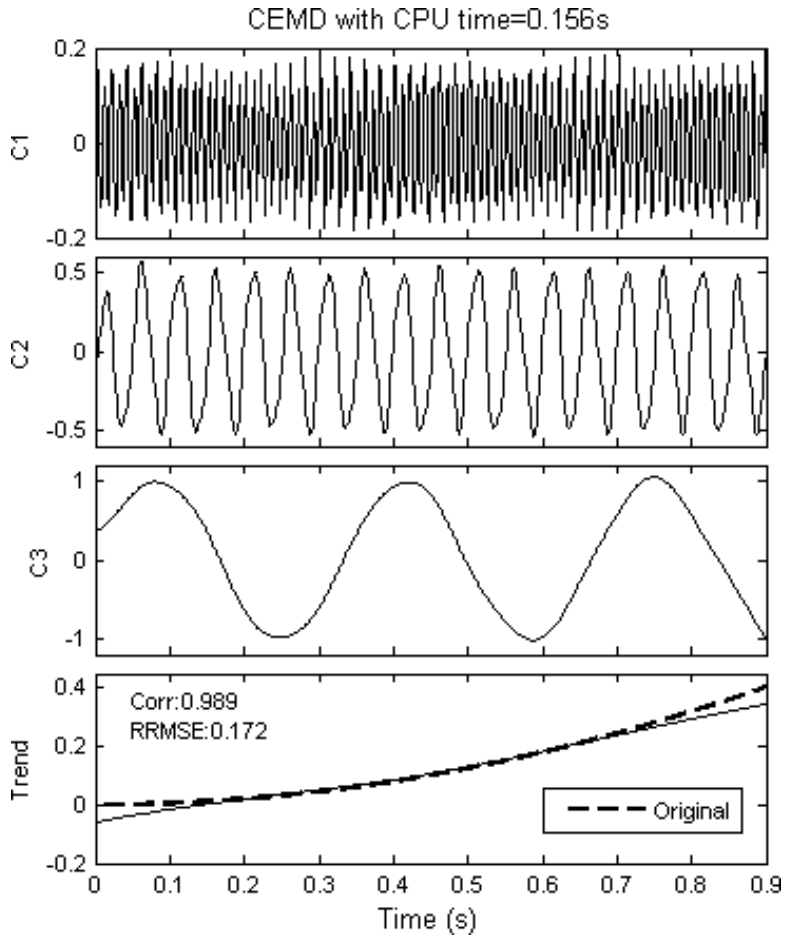

Fig. 11. CEMD on the signal shown in Fig. 9: (a) original time series, (b) IMF-1, (c) IMF-2, (d) IMF-3, and (e) trend. Comparison between Figs. 11 and 10 shows the capability to reduce the mode mixing, end effect, and detrend uncertainty.

using the traditional EMD and EEMD. Since the observational data contain errors, four time series $s_m(t_i)\,(m = 1, 2, 3)$ are constructed each by a signal [components of (25)] and random noise $n(t_i)$ produced by a pseudorandom number generator with the amplitude of 0.1,

$$\begin{aligned} s_1(t_i) &= f_0(t_i) + f_1(t_i) + f_2(t_i) + n(t_i), \\ s_2(t_i) &= f_0(t_i) + f_1(t_i) + f_2(t_i) + f_3(t_i) + n(t_i), \\ s_3(t_i) &= f_0(t_i) + f_1(t_i) + f_2(t_i) + f_3(t_i) + f_4(t_i) + n(t_i). \end{aligned} \tag{28}$$

Here, the time step and the total number of data (N) is the same as shown in (25). Figure 12 shows the three time series, $s_m(t_i)$.

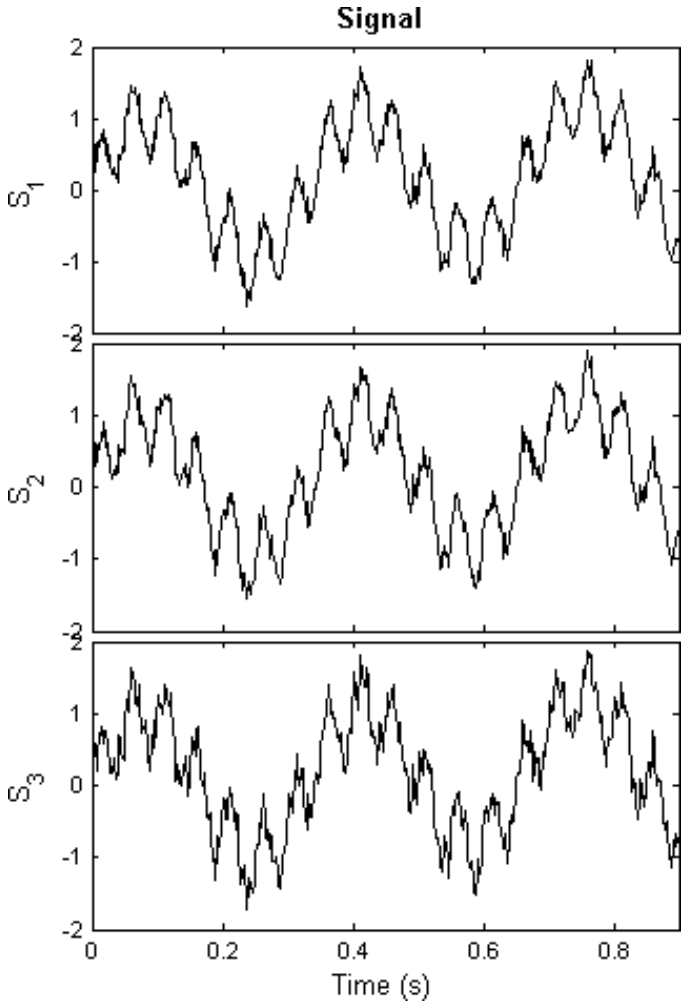

Fig. 12. Time series of three noisy signals (s_1, s_2, s_3).

Figure 13 also shows the capability of CEMD in detrend with high positive CCs,

$$\mathrm{CC}_1 = 0.984, \quad \mathrm{CC}_2 = 0.989, \quad \mathrm{CC}_3 = 0.996, \tag{29}$$

and low RRMSEs,

$$\mathrm{RRMSE}_1 = 0.287, \quad \mathrm{RRMSE}_2 = 0.359, \quad \mathrm{RRMSE}_3 = 0.502. \tag{30}$$

The ratio of CPU with EEMD-1k versus CEMD is 310 (77.5 s versus 0.25 s) for analyzing $s_1(t)$, 270 (80 s versus 0.296 s) for analyzing $s_2(t)$, and 220 (79.1 s versus 0.359 s) for analyzing $s_3(t)$. The ratio of CPU with EEMD-5k versus CEMD is 1548 (387 s versus 0.25 s) for analyzing $s_1(t)$, 1365 (404 s versus 0.296 s) for analyzing $s_2(t)$, and 1103 (396 s versus 0.359 s) for analyzing $s_3(t)$.

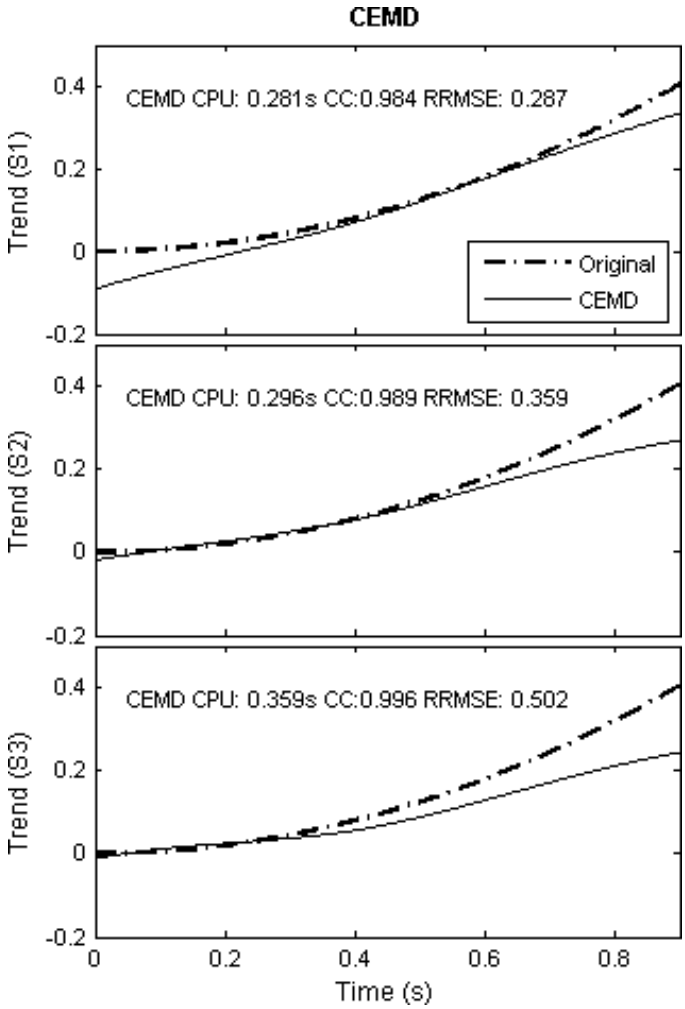

Fig. 13. Detrend of three noisy signals (s_1, s_2, s_3) by CEMD. Here, in each panel, the bold dashed-dotted curve is the trend of signal, $f_0(t)$. The solid curve is the trend identified using CEMD.

5. Conclusions

(1) Mode mixing, end effect, and related detrend uncertainty in the traditional EMD and EEMD can be eliminated using the new method (i.e. CEMD), which has been established on the base of (a) compact difference scheme concepts [Chu and Fan (1998), (1999)] and (b) identification of uniformly distributed pseudo maximum and minimum points using highest-frequency sampling. Determination of either upper or lower envelope becomes a mathematical problem to solve a set of $2N$ algebraic equations of the values and first derivatives for each envelope at all N data points (including the two end points).

(2) The CEMD uses hybrid Hermitain polynomials rather than the cubic spline. A set of $2N$ algebraic equations (u_i, u_i') for the upper envelope (same for the lower envelope) are obtained from more realistic conditions such as zero third derivative and use of weighted regressed first derivative at the end points. No extending points are needed.

(3) Capability of the CEMD for eliminating mode mixing, end effect, and detrend uncertainty is demonstrated using a time series consisting of a quadratic trend and four harmonics. The CEMD can recover the original signal really well with low RRMSEs. Furthermore, the capability of the CEMD is not affected by the existence of random noises.

References

Chu, P. C. and Fan, C. W. (1998). A three-point combined compact difference scheme. *J. Comput. Phys.*, **140**: 370–399.

Chu, P. C. and Fan, C. W. (1999). A three-point non-uniform combined compact difference scheme. *J. Comput. Phys.*, **148**: 663–674.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C. and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. R. Soc. London, Ser. A*, **454**: 903–993.

Wang, J., Peng, Y. and Peng, X. (2007). Similarity searching based boundary effect processing method for empirical mode decomposition. *Elect. Lett.*, **43**(1): 58–59.

Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Analys.*, **1**: 1–41.

Zhao, J. and Huang, D. (2001). Mirror extending and circular spline function for empirical mode decomposition method. *J. Zhejiang Univ. Sci.*, **2**: 247–252.