ISSN 1009-3095 Journal of Zhejiang University (SCIENCE) V.2, No.3, P.247-252, July-Sep., 2001
http://www.chinainfo.gov.cn/periodical; http://www.zju.edu.cn;
http://lib.zju.edu.cn/English; http://www.zjupress.com

MIRROR EXTENDING AND CIRCULAR SPLINE FUNCTION FOR EMPIRICAL MODE DECOMPOSITION METHOD*

ZHAO Jin-ping(赵进平), HUANG Da-ji(黄大吉)

(*Second Institute of Oceanography, State Oceanic Administration, Hangzhou, 310012, China*)

E-mail: jpzhaog@yahoo.com

Received Mar. 14, 2001; revision accepted May 20, 2001

Abstract: The Mirror Extending (ME) approach is proposed in this paper for solving the end extending issue in the Empirical Mode Decomposition (EMD) method. By this approach, the data is extended into a closed circuit without end. The derivatives on ends are not necessary any more for Spline fitting. The approach eliminates the possible problems in reliability and uniqueness in the original extending approach of the EMD method. In the ME approach only one extending is necessary before the data analysis. A theoretical criterion is proposed here for checking the extending approach. ME approach has been proved to satisfy the theoretical criterion automatically and permanently. This approach makes the EMD method reliable and easy to follow.

Key words: empirical mode decomposition, mirror extension, circular Spline function, time series analysis, time-frequency domain

Document code: A **CLC number:** O212.4

INTRODUCTION

The Empirical Mode Decomposition (EMD) method and its Hilbert Spectrum technique is a newly developed method for time series analysis (Huang et al., 1998). This method is suitable for analyzing nonlinear and non-stationary process based on an empirical approach and provides a precise definition in time-frequency domain. With EMD method, the signal is decomposed into several intrinsic mode functions (IMF) with the features of completeness and orthogonality. Compared with Fourier Analysis (FA), the IMFs of EMD are signals coming entirely from the data and usually display obvious physical significance in different time scale, whereas the infinite harmonics decomposed by FA are theoretical ones (Jaeger and Starfield, 1974). FA is a good method for analyzing stationary and linear data, but EMD can be used in whatever processes and situations included in data. EMD method has similar feature with Wavelet Analysis (WA) in analyzing nonlinear and non-stationary processes and providing time-frequency distribution results (Chan, 1995), but the spectrum of EMD is of higher resolution than that of WA. Therefore, EMD will be a very useful method in distinguishing processes with different time scale (Salvino, 2001), and so is being widely used in many scientific domains (Zhu et al., 1997; Xie et al., 2000; Vincent et al., 1999).

The decomposition process of the EMD method separates the original data into both mean and the first IMF. Then the mean is considered as new data, which is decomposed into new mean and the second IMF. The decomposing process is repeated until the last IMF is obtained, whose mean is a constant. Two envelops of the original data are defined by local maxima and minima respectively, by which the mean curve is determined by averaging both envelops. The envelops are easy to be determined by cubic Spline fitting between extrema. At both ends of data, first or second order derivatives are required for Spline fitting (Esch, 1974). However, as the data curve does not provide any information on the envelops at the ends, the derivatives cannot be given unless the data is extended at two ends. Huang et al. (1998) gave an approach to extend

* Project supported by NSFC (49634140, 49876008). Contribution No. 2001B001 from the Key Lab of Ocean Dynamic Processes and Satellite Oceanography, SOA.

data by adding several characteristic waves at both ends (hereafter referred to as CW approach). The characteristic waves can be used to calculate the derivatives of data out of ends.

Nevertheless, the CW approach is not sufficiently reliable, as it implies some indefiniteness in the result. Firstly, the data only gives information on the extrema inside the data series. Any extending out of the ends is not reliable. So, to determine the extrema at the ends requires a more strict definition in order to ensure the reliability of results. Secondly, A uniqueness result is necessary for using the EMD method, but the CW approach is not a strict method. By adding different characteristic waves, the results could be quite different. It means the differences among results from different users might exist. Thirdly, the data is usually divided into limited modes and, for each mode tens to hundreds of iterations are necessary. The characteristic waves should be added for each iterations, by which it is easy to bring error into the interior of data. So, Huang et al. (1998) suggested that the CW approach still needs to be improved.

Logically, there might be many end extending approaches. For example, Deng et al. (2001) proposed a new extending approach by artificial neural networks. In this paper, a Mirror Extending approach is given. Then, a criterion for the optimal extension is proposed and the optimal characteristics of the ME approach is verified. The approach eliminates some disadvantages of the CW approach, and is approved to be reliable in the EMD method. In fact, the EMD method up to now is hard to follow, because of the restriction of end extending. By this approach, EMD method can be widely used in many scientific domains.

BRIEF INTRODUCTION TO THE EMD METHOD

The EMD method consists of two steps. Time series data are first decomposed into several IMFs by the above mentioned approach after which, the time-dependent amplitudes and frequencies of IMFs are defined by using Hilbert Transform.

For the data $u(t)$, the first IMF is defined as

$$c_1(t) = u(t) - m(t) \tag{1}$$

where $m(t)$ is the mean defined by the average of upper and lower envelops. Many iterations are usually necessary to get the mean until the mean of $c_1(t)$ tends to zero at a given precision. An accurate mean is absolutely necessary because of the orthogonality requirement. Then, considering $m(t)$ as a new data series and after continually obtaining other IMFs, c_2, c_3, $\cdots$, c_n, the data can finally be expressed by the sum of IMFs as follows:

$$u(t) = \sum_{i=1}^{n} c_i(t) \tag{2}$$

where n is the number of the last mode and is usually a finite and small number. As the mean of the last mode is a constant, it can be combined into c_n. Fig. 1 gives an example of the structure of IMFs decomposed by Huang et al. (1998), which shows each IMF for given data. In each mode the local frequency varies with the frequency of a latter mode being lower than that of a former mode.

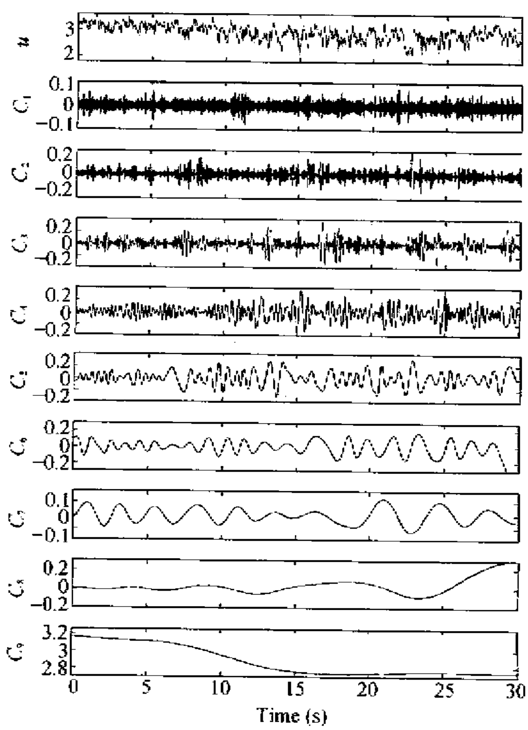

Fig. 1 An example of EMD method (Huang et al., 1998)

Let, $x_i(t) = c_i(t)$, the Hilbert Transform of $x_i(t)$ is $y_i(t)$ (Bedrosian, 1963)

$$y_i(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x_i(\tau)}{t-\tau} d\tau \quad (3)$$

where P denotes to take Cauchy principal value. As $x_i(t)$ and $y_i(t)$ comprise a complex conjugate pair

$$x_i(t) + jy_i(t) = A_i(t)\exp[j\theta_i(t)]. \quad (4)$$

The corresponding amplitude and frequency for each of the IMFs are

$$A_i(t) = \sqrt{x_i^2(t) + y_i^2(t)} \quad (5)$$

$$\omega_i(t) = \frac{d\theta_i(t)}{dt} \quad (6)$$

Eqs. (5) and (6) give a spectrum with variable frequency, called Hilbert-Huang spectrum, from which a three dimensional figure in time-frequency-amplitude space can be plotted.

MIRROR EXTENDING AND CIRCULAR SPLINE FITTING

As mentioned above, the envelops of data are determined by cubic Spline fitting, in which the first or second order derivatives at ends are needed to close the equations. The CW approach to extend the data is not a strict method and easy to lead to indefiniteness. Up to now, there is still no a theoretical principle on how to extend the data. Here, a new approach for extending data is proposed.

If a mirror is put at each end of data, the image of data in the mirror is in opposite direction and symmetrical to the mirror. The image of data becomes a connecting curve between the two mirrors and, with the original data together, presents a closed circular pattern. A time series is extended into a closed circuit as shown in Fig.2. We call this Mirror Extending (ME) approach. As shown in Fig. 2, the data over the upper plane (solid line) is extended to the lower plane (dashed line). But only the output from the upper plane is used as the result of the EMD method.

As a whole, IMFs are processed for the data circle as shown in Fig.3 from first mode to last mode without any other extension.

Owing to no ends in the extended circular data set, the Spline fitting should be different from the CW approach.

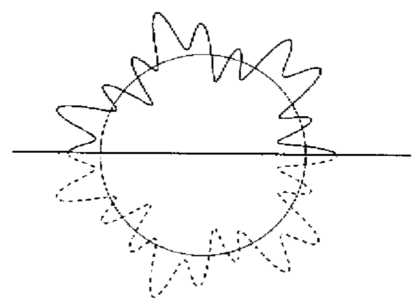

Fig.2 Sketch for mirror extending of data
Straight line is the mirror put at both ends of data; Solid line is the original data and dashed line is the symmetric reflection of data in the mirror.

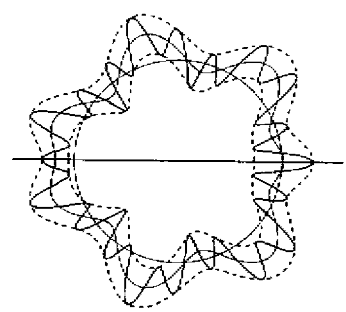

Fig.3 Sketch for circular Spline fitting
The two dashed lines are for upper and lower envelops respectively; the solid line is the mean

Given a discrete set of $y = y(x)$ as (x_0, y_0), (x_1, y_1), …, (x_n, y_n), which has been extended by the ME appoach. For a circular data set, $y_n = y_0$ when $x = x_n$. The cubic Spline function is defined as

$$S_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + y_i \quad (x_i < x < x_{i+1}) \quad (7)$$

for $i = 0, 1, \cdots, n-1$. Let the function $S_i(x)$ and its first and second order derivatives be continuous at each data points. We have

$$\begin{aligned} a_{i-1}h_{i-1}^3 + b_{i-1}h_{i-1}^2 + c_{i-1}h_{i-1} &= y_i - y_{i-1} \\ 3a_{i-1}h_{i-1}^2 + 2b_{i-1}h_{i-1} + c_{i-1} &= c_i \\ 6a_{i-1}h_{i-1} + 2b_{i-1} &= 2b_i \end{aligned} \quad (8)$$

where the a_i, b_i, and c_i can be expressed by

$$a_i = \frac{b_{i+1} - b_i}{3h_i}$$

$$b_i = \frac{1}{h_{i-1}}\left(2c_i + c_{i-1} - 3\,\frac{y_i - y_{i-1}}{h_{i-1}}\right) \qquad (9)$$

$$c_i = \frac{y_{i+1} - y_i}{h_i} - \frac{1}{3}(b_{i+1} + 2b_i)h_i$$

Eq. (7) becomes

$$b_{i-1}h_{i-1} + 2b_i(h_{i-1} + h_i) + b_{i+1}h_i = q_i \qquad (10)$$

where

$$h_i = x_{i+1} - x_i$$

$$q_i = 3\left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}\right)$$

As the data circuit has no end, the equations become

$$\begin{aligned} &b_0h_0 + 2b_1(h_0 + h_1) + b_2h_1 = q_1 \\ &b_1h_1 + 2b_2(h_1 + h_2) + b_3h_2 = q_2 \\ &\square \\ &b_{n-1}h_{n-1} + 2b_n(h_{n-1} + h_n) + b_1h_n = q_n \\ &b_nh_n + 2b_1(h_n + h_1) + b_2h_1 = q_0 \end{aligned} \qquad (11)$$

Eq. (11) in matrix form becomes

$$\begin{bmatrix} 2(h_n + h_0) & h_1 & 0 & \cdots & 0 & 0 & h_n \\ h_0 & 2(h_0 + h_1) & h_1 & \cdots & 0 & 0 & 0 \\ \square & \square & \square & \ddots & \square & \square & \square \\ 0 & 0 & 0 & \cdots & h_{n-2} & 2(h_{n-2} + h_{n-1}) & h_{n-1} \\ h_n & 0 & 0 & \cdots & 0 & h_{n-1} & 2(h_{n-1} + h_n) \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \square \\ b_{n-1} \\ b_n \end{bmatrix} = \begin{bmatrix} q_0 \\ q_1 \\ \square \\ q_{n-1} \\ q_n \end{bmatrix} \qquad (12)$$

After solving b_i by Eq. (12), Eq. (9) can be used to determine the a_i and c_i and get Spline fitting by Eq. (7). It can be seen in Eq. (12) that the only difference between circular Spline fitting and standard Spline fitting is in the coefficient matrix. The coefficient matrix of standard Spline fitting is triple diagonal and can be solved by a simple elimination approach, a faster approach to solving linear equations with only data in a narrow band beside the diagnal. However, there are two elements appearing in the upper-right corner and lower-left corner of the coefficient Matrix in circular Spline fitting. Eq. (12) must be solved by an ordinary method for linear equations, such as Gaussian elimination, which will take much more time than that needed by using method for triple diagonal equations, especially for longer data set.

CRITERION FOR DATA EXTENDING AND SUPERIORITY OF ME APPROACH

As indicated above, any data extension approach is not for providing correct values out of ends. Data extension is aimed to provide such derivatives at ends that do not exert impact into interior of IMF during the sifting process. Although many end extending approaches can be proposed, they should have the following feature in theory as a criterion.

If an extending does not exert impact on the IMF interior, it must be ensured that the Spline coefficients in Eq. (7) of the segment near the end should be entirely determined by interior data, not by any data from the extended data set. This theoretical criterion can be used to check the reliability of an approach before it is used.

Let subscript 0 denote the point on the end or out of the end, and 1 denote the first point inside the data. For convenience, let us choose the first end at x_0 as the example and suppose the value at the end is a maximum. In this case, subscript 0 expresses the point on the end and 1 expresses the first point inside data for the upper envelop. For the lower envelop, subscript 1 refers to first minimum and 0 is the point extended. If a_1, b_1, and c_1 in Eq. (7) are independent of the data point with subscript 0, the extended data will not affect the IMF interior during iteration. From Eq. (9), the coefficients for point 1 are expressed by

$$\begin{aligned} &a_1 = \frac{b_2 - b_1}{3h_1} \\ &h_0b_1 - 2c_1 = c_0 - 3\,\frac{y_1 - y_0}{h_0} \\ &\frac{2}{3}h_1b_1 + c_1 = \frac{y_2 - y_1}{h_1} - \frac{1}{3}h_1b_2 \end{aligned} \qquad (13)$$

It can be seen from Eq. (13) that only parameter with subscript 0 is $c_0 - 3(y_1 - y_0)/h_0$, where c_0 is the first derivative of data on the end as shown in Eq. (7) and $(y_1 - y_0)/h_0$ is also an estimation for the first derivative on the end by the values at point 0 and 1. If $c_0 - 3(y_1 - y_0)/h_0$ is equal to zero, or to a value determined by original data, a_1, b_1, and c_1 will not be affected by extending.

Let us check the Mirror Extending approach by using the above two criteria. For the upper envelop (with maximum at end), we have $c_0 = 0$ by mirror effect and $(y_1 - y_0)/h_0$ is determined by given data, so the criterion is satisfied. For lower envelop, we have, from the definition of ME approach that

$$c_0 = 0 \text{ and } y_1 - y_0 = 0 \tag{14}$$

because $c_0 \equiv 0$ and $y_1 - y_0 \equiv 0$ also by mirror effect. Therefore, the ME approach satisfies the theoretical criterion. The CW approach can also satisfy the condition given by Eq. (14) if the waves added are chosen properly in every iteration and for every IMF. However, by ME approach, the data is extended only once, and then the criterion will be satisfied in all iterations and for all IMFs automatically and permanently.

Owing to this feature of the ME approach, in fact, the data in upper plane and the image of data in lower plane in Fig.2 do not interchange although the data are connected. The extended data only provide suitable derivative condition, but do not affect interior data. From this point of view, ME is a good approach in data extending.

We used the ME approach to the examples given by Huang et al. (1998). The results were quite consistent. It implies that the CW approach satisfies to a great extent the criterion proposed above. However, there still are little difference between the results of CW and ME approaches. As an example, a superimposed function is given as (shown in Fig.4)

$$x(t) = \cos\left(\frac{2}{30}\pi t\right) + \cos\left(\frac{2}{34}\pi t\right) \tag{15}$$

As this function is an amplitude-changed periodic function. The standard result has only one mode as indicated by Huang et al. (1998). But if we want to separate the two components in Eq. (15), the possibility depends on the iteration times and data precision in computer. Huang et al. (1998) made the sifting up to 3000 times and got a result as shown in Fig.5 for the first four IMF components. This result is not very successful in separating two waves because the frequency difference of them is too small, but it does give a possible way to separate waves with such a small frequency difference. By the ME approach in this paper, the same sifting up to 3000 times is carried out, and the result is shown in Fig.6. It can be seen that both results in Figs. 5 and 6 are quite similar. Little difference occurs in left end of IMFs. It looks the result in Fig.6 are more reasonable and stable.

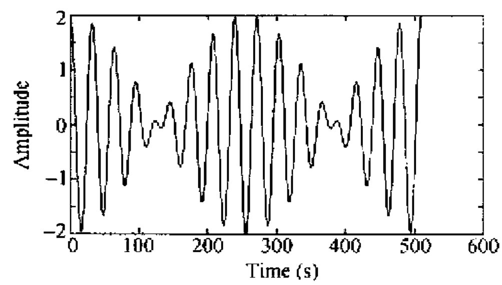

Fig.4 Data of linear superimposed waves given by Eq. (15)

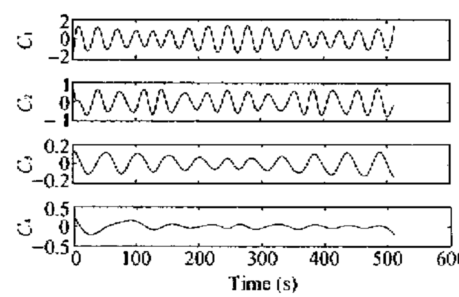

Fig.5 The first four IMF components after stringent application of EMD up to 3000 times sifting by Huang et al. (1998)

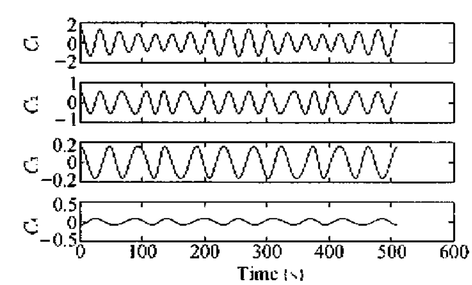

Fig.6 The first four IMF components after stringent application of EMD up to 3000 times sifting by ME approach in this paper

CONCLUSIONS

The Mirror Extending approach is proposed in this paper. By the ME approach, data is extended into a closed circuit without end. The derivatives on the ends are not necessary anymore for Spline fitting. The approach eliminates the possible problems of the CW approach in reliability and uniqueness. In the ME approach, only one time extending is necessary before data analysis.

A theoretical criterion is proposed for checking any data extending approach. ME approach was proved to satisfy the theoretical criterion automatically and permanently. This characteristic will always exist for any kind of data. Therefore, ME is a reasonable and reliable approach to extend data in the EMD method. ME approach makes the EMD method easy to apply. The only difference is in the coefficient matrix. The method for triple diagonal equations for standard Spline fitting cannot be used, and the ordinary methods should be used in solving the equations.

It is emphasized that the data is not required to be symmetric near ends, though the ME approach makes the extended data is symmetric to original one. It is because that the data extending at the ends is not for giving real data out of the ends, but for providing reasonable derivatives to ensure that there is no impact is exerted on the interior part of data.

Otherwise, the last mode in the EMD method by Huang et al. (1998) has one maximum and one minimum. So, the last mode is usually an ascending or descending curve, which usually represents the long-term trend. The last mode of circular data extending can have two maxima and two minima if the last extrema does not appear at the mirror position. Using half of the circular data, might obtain a peak or a trough in the long-term trend. Sometimes, the trend is not a monotonous ascent or descent, so the ME approach provides a relatively better way to express the trend of a process.

The only issue to pay attention is that the mirror should be put on an extrema. If it is not sure if the end data is an extremum or not, it is better to cut a segment of data to put the mirror on a true extremum. Even if a non-extremum is used as a datum on an end, the calculation is still stable and convergent. This stress on using a true extrema is aimed to achieve best possible reflection of the real processes in each of the IMFs.

References

Chan, Y. T., 1995. Wavelet basics. Kluwer, Boston.

Deng Yongjun, Wang Wei, Qian Chengchun, et al., 2001. An approach for ends issue in EMD method and Hilbert Transform. *Chinese Science Bulletin*, **46**(3): 257 – 263(in Chinese).

Esch, R., 1974. Functional approximation. *In*: Handbook of Applied Mathematics, (Ed. by C. E. Pearson), Van Nostrand Reinhold Company.

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. *Proc. R. Soc. Lond*, **454**: 903 – 995.

Jaeger, J. G., A. M. Starfield, 1974. An Introduction to Applied Mathematics. Second edition, Oxford University Press, Ely House, London, 504p.

Salvino, L., Navsea, C. Empirical Mode Decomposition and time-frequency analysis. (http://saviac.xservices.com/sv71abs/salvino.html)

Vincent, B., Hu, J., abd Hou, Z., 1999. Damage Detection Using Empirical Mode Decomposition Method and a Comparison with Wavelet Analysis, Proceedings of the Second International Workshop on Structural Health Monitoring. Stanford, p.891 – 900.

Xie, L., Kejian Wu, Pietrafesa, L. J., 2000. Empirical modes of landfalling tropical cyclones in North Carolina. The 24th Conference on Hurricanes and Tropical Meteorology, p.168.

Zhu, X., Shen Z., Eckermann S.D., et al., 1997. Gravity wave characteristics in the middle atmosphere derived from the empirical mode decomposition method. *J Geophys Res*, **102**(D14): 16; 545.