Advances in Adaptive Data Analysis
Vol. 2, No. 1 (2010) 1–24

World Scientific
www.worldscientific.com

A CRITERION FOR SELECTING RELEVANT INTRINSIC MODE FUNCTIONS IN EMPIRICAL MODE DECOMPOSITION

ALBERT AYENU-PRAH* and NII ATTOH-OKINE†
Department of Civil and Environmental Engineering
301 DuPont Hall, University of Delaware
Newark, DE 19716, USA
**ayenuprah@gmail.com*
†*okine@udel.edu*

Information extraction from time series has traditionally been done with Fourier analysis, which use stationary sines and cosines as basis functions. However, data that come from most natural phenomena are mostly nonstationary. A totally adaptive alternative method has been developed called the Hilbert–Huang transform (HHT), which involves generating basis functions called the intrinsic mode functions (IMFs) *via* the empirical mode decomposition (EMD). The EMD is a numerical procedure that is prone to numerical errors that may persist in the decomposition as extra IMFs. In this study, results of numerical experiments are presented, which would establish a stringent threshold by which relevant IMFs are distinguished from IMFs that may have been generated by numerical errors. The threshold is dependent on the correlation coefficient between the IMFs and the original signal. Finally, the threshold is applied to IMFs of earthquake signals from five accelerometers located in a building.

Keywords: Empirical mode decomposition; Hilbert-Huang transform; intrinsic mode functions; correlation coefficient; Fourier analysis; nonstationary data; nonlinear system.

1. Introduction

Extracting relevant information from data has traditionally been achieved with Fourier analysis and Fourier-based methods. Fourier methods use stationary sines and cosines as basis functions to decompose data, which data will simply be referred to as a "signal." However, data that come from natural phenomena are usually originating from nonlinear systems, and because there is usually no idea about the beginning and end of the natural process the resulting signal will also be assumed nonstationary since there is no information on whether certain standard wavelengths repeat or not. Defining nonlinearity for signals that originate from natural phenomena, though, is not trivial for there is no certain way of telling nonlinearity except, perhaps, observing by eye the signal for certain signs of nonlinearity such as wide troughs and sharp crests; or finding the probability density function,

which should remain approximately normal if the system is linear; however, nonlinear signals can be simulated. In civil infrastructure systems such as bridges, roads, and buildings, acquired data are also usually nonlinear and nonstationary. Example data are road roughness, road profiles, rutting, cracking, bridge settlement, dynamic response, traffic flow, alkali-silica reaction (ASR), as well as pollution. However, since it is difficult to define nonlinearity, this paper will simply treat the signal as nonstationary. Rather than assume that the data are stationary, the solution to the data analysis problem should be adapted from the data themselves without using *a priori* basis sets; the data should reveal their own solutions, not the solutions being imposed on the data. In order to get energy content information from the data, there is a transformation from the original domain of the data into an energy–frequency domain, which causes information in the original domain of the data to be lost, explained by the Heisenberg uncertainty principle. However, sometimes it is desirable and essential to be able to see the evolution of frequency and energy with the original domain of the data, as in the case of road profile analysis in which it becomes even necessary to observe the variation of energy with distance along the road. Various methods have been developed in an attempt to handle nonstationarity, which methods include wavelet transforms, short-time Fourier transforms, and Wigner–Ville distribution. A common thread running through these methods is the prescription of *a priori* basis sets to define the solution to the data analysis. An alternative method of analysis that is totally adaptive has been developed by Huang *et al.* (1998) called the Hilbert–Huang transform (HHT). Basis sets are derived from the intrinsic time-scales of the data through a sifting process called the empirical mode decomposition (EMD); the basis are called intrinsic mode functions (IMFs), which are almost orthogonal based on an index of orthogonality explained in Huang *et al.* (1998), and they form a complete set by virtue of the fact that adding up all the IMFs and the residual recovers the original signal. An IMF has two critical properties: the number of extrema is equal to the number of zero crossings, or they may differ by at most one, and the mean of the envelopes defined by the local maxima and the local minima is zero at every point; and the envelopes are constructed with cubic splines, which are found to be the most suitable [Huang *et al.* (1998)]. Therefore, an IMF is like a sinusoid that is allowed to have varying amplitudes. In addition, IMFs have well-behaved Hilbert transforms, from which meaningful instantaneous frequency and amplitude values can be determined. The instantaneous frequencies and amplitudes can be plotted with the original domain of the data (say, time) to form an energy–frequency–time spectrum called the Hilbert spectrum, which step is known as the Hilbert spectral analysis (HSA); the EMD and the HSA constitute the HHT. The fact that basis sets are derived a posteriori makes the HHT relatively more advantageous over Fourier-based methods in terms of analyzing nonstationary data. Several authors have compared the HHT with other traditional methods like the Fourier transform and wavelets [Huang *et al.* (1998); Peng *et al.* (2005); Hwang *et al.* (2005); Quek *et al.* (2005)]. Although Fourier analysis remains a very powerful method, HHT presents an equally effective alternative.

An important issue in EMD is interpreting the physical meaning of the IMFs. In order to achieve that, it would be helpful to determine which of the IMFs are relevant to the original signal since the EMD is a numerical procedure that is prone to numerical errors; the numerical errors may persist in the decomposition as extra IMFs, which should be excluded from further analysis. In this study, results of numerical experiments are presented, which would establish a stringent threshold for discriminating between relevant and irrelevant IMFs. The threshold is then applied to IMFs of signals from accelerometers located in a building that record seismic data.

2. Empirical Mode Decomposition

To begin the EMD, a function or signal is decomposed as follows:

Identify all the local extrema; then connect all the local maxima by cubic spline as the upper envelope. Repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should include all the data. If the mean of the upper and lower envelopes is designated as m_1 and the difference between the data and m_1 is the first component h_1, then

$$x(t) - m_1 = h_1. \tag{1}$$

The mean m_1 is given by

$$m_1 = \frac{L + U}{2}, \tag{2}$$

where U is the local maxima and L is the local minima.

Technically, h_1 is supposed to be an IMF, except some error might be introduced by the spline curve fitting process — in many cases there are overshoots and undershoots after the first round of processing; therefore, the sifting process has to be repeated many times. The sifting process serves two purposes [Huang *et al.* (1998)]: it eliminates riding waves (smaller waves that seem to "ride" bigger waves), and it makes the signal or profile more symmetric about the local zero-mean line. In the second round of sifting, h_1 is treated as the data or the first component; then a new mean is computed as before. If the new mean is m_{11}, then

$$h_1 - m_{11} = h_{11}. \tag{3}$$

After repeating the sifting process up to k times, h_{1k} becomes an IMF; that is:

$$h_{1(k-1)} - m_{1k} = h_{1k}. \tag{4}$$

Let $h_{1k} = c_1$, the first IMF from the data. c_1 should contain the finest scale or the shortest period component of the data. The process to generate one IMF may be considered as the inner loop. Now c_1 is separated from the original data as:

$$x(t) - c_1 = r_1, \tag{5}$$

where r_1 is the residue, and it contains information on longer period components; it is now treated as the new data and subjected to the same sifting process (this

is now the beginning of the outer loop, which will go on to the next inner loop for the next IMF). The procedure is repeated for all subsequent r_js resulting in

$$r_1 - c_2 = r_2, \ldots, r_{n-1} - c_n = r_n, \tag{6}$$

where c_2 to c_n are the subsequent IMFs of the data. A schematic of the inner and outer loops of the EMD is shown in Fig. 1.

There are stopping criteria for the sifting process for IMFs since allowing sifting to go beyond a certain point may remove important signal variations and features that arise from the natural dynamics of the system — the IMF components need to retain enough physical sense of both amplitude and frequency modulations. This can be achieved by limiting the value of the sum of the difference (SD), computed

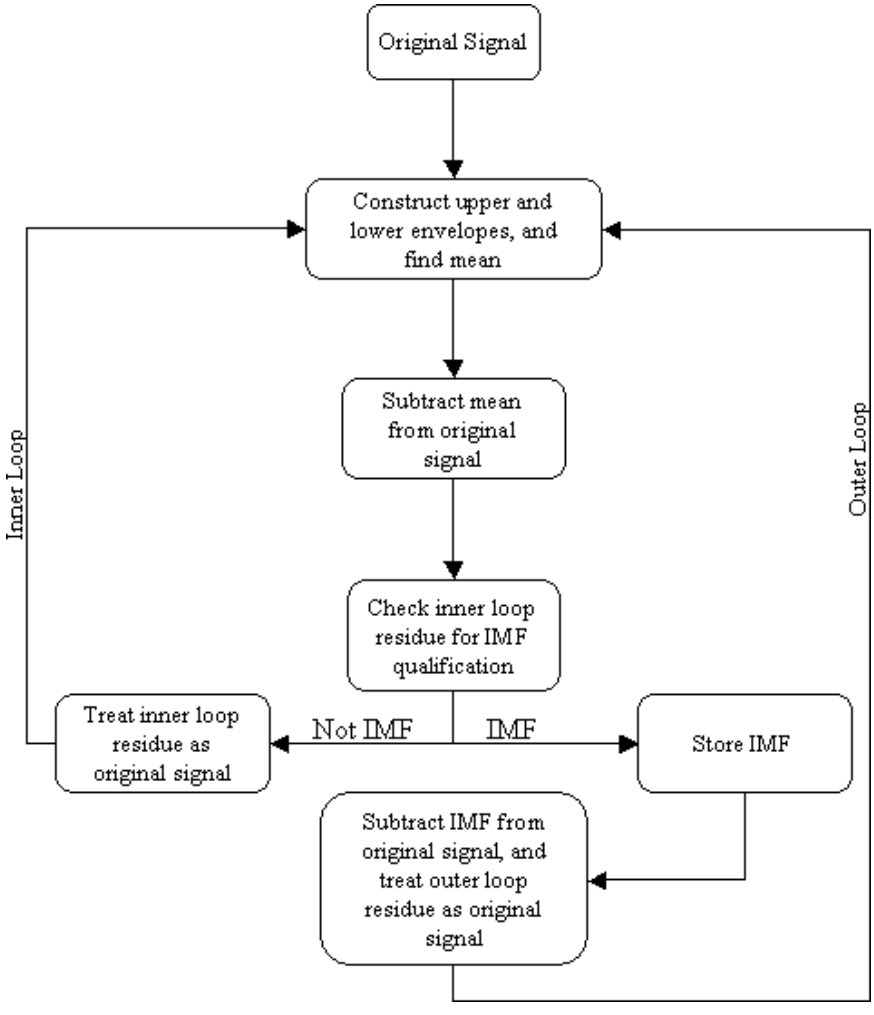


Fig. 1. Pictorial representation of EMD.

from two consecutive sifting results as:

$$SD = \frac{\sum_{t=0}^{T} |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^{T} h_{k-1}^2(t)}. \tag{7}$$

A value of SD between 0.2 and 0.3 is usually preferable based on experimental analyses performed by Huang *et al.* (1998). To check that the number of zero crossings is equal to, or differs by at most one from the number of extrema, an alternate stopping criterion is proposed by Huang *et al.* (2003). Sifting is stopped when the number of zero crossings is equal to, or differs by at most one from the number of extrema for S successive sifting steps; the optimum value for S was found to be between 4 and 8.

A number of issues have come up concerning empirical mode decomposition, including the following:

- finding mathematical and physical meaning for IMFs, since EMD is essentially algorithmic in nature and lacking mathematical rigor;
- determining the most appropriate interpolation scheme; and
- handling of boundary or end effects during data interpolation.

HHT has had wide applications in ocean engineering in which the method has been used to analyze the properties and behavior of ocean waves [Datig and Schlurmann (2004); Veltcheva and Soares (2004); Gloersen and Huang (2003)], in biomedical applications such as analysis of signals from heartbeats and investigation of obstructive sleep apnea [Echeverria *et al.* (2001); Salisbury and Sun (2007)], and in the financial industry where HHT has been applied to financial time series to examine volatility of markets and the correlation of foreign exchange rates in currency markets [Wu (2007); Huang *et al.* (2003)]. However, one issue still persists in all these advancements: the physical significance of IMFs derived from the original data series or signal. A thorough understanding of the physical processes that generate data is required before any form of scientific explanation can be attributed to any particular IMF or group of IMFs. Even with this kind of thorough knowledge there is still a level of ambiguity when trying to extract information from the IMFs that are directly relevant to the original signal and the physics of the underlying system. Before getting to the point where essential information can be extracted from the IMFs there is a need to determine which IMFs are really relevant to the decomposition process and which carry the necessary information required to understanding the underlying system, for EMD is a numerical procedure with possible numerical errors in the results. Therefore, a repeatable method is needed to discriminate between relevant and not so relevant IMFs. The uncertainties about which IMFs are relevant in any EMD process contribute to the overall difficulty in trying to physically interpret the IMFs of signals; the question always arises about the physical meaning of each IMF when there is no prior knowledge of the system producing the signals. A way to increase the level of confidence in the IMFs would be to somehow narrow down the IMFs to only the relevant ones by some theoretical

and repeatable means. In this paper an expression is presented that attempts to discriminate between relevant and irrelevant IMFs.

Wu and Huang (2004; 2005) studied the statistical characteristics of uniformly distributed white noise using the EMD method and numerical experiments. Based on results obtained from the numerical experiments, the authors put forward a method of assigning statistical significance to information content of IMFs from noisy signals. The results presented were that the product of the energy density of an IMF and the averaged period of the IMF is a constant, and the energy-density distribution of an IMF sample follows the Chi-squared distribution. A spread function for the energy distribution of an IMF was also obtained. The test method for IMF information content was consistent with the Monte Carlo test. The test method begins with first decomposing the target data set into IMFs using the EMD method. Now, construct a long artificial normalized white-noise data set, divide it into sections with identical lengths as the target data set, and determine the spread function of various percentiles. Select a confidence limit, say 99%, and determine upper and lower spread lines. The final step is to compare the energy density of the IMFs from the target data set with the spread functions. At the chosen confidence level, the IMFs with energy located above the upper spread line and below the lower spread line are considered to contain information. Essentially, Wu and Huang presented a method of determining information content of IMFs from the statistical characteristics of uniformly distributed white noise.

EMD is a numerical procedure and prone to numerical errors that may persist in the decomposition results. Before assigning statistical significance to information content in the IMFs, the IMFs themselves need to be relevant to the decomposition; in other words, IMFs need to be sought that are less likely to have been generated solely by the numerical errors that persisted. Therefore, while Wu and Huang (2004) presented sound and effective analytical expressions for statistical significance of information content, an additional step that could be potentially important was disqualifying any spurious IMF that could have been the result of numerical errors. Peng *et al.* (2005) proposed an expression to discriminate between relevant and irrelevant IMFs in order to improve the overall EMD process. Since the IMFs are supposed to be almost orthogonal components of the original signal, each IMF would have a relatively good correlation with the original signal; this presupposes that the irrelevant components would have relatively poor correlation with the original signal. Therefore, a threshold, λ, is introduced, which is given by

$$\lambda = \frac{\max(\mu_i)}{10}, \quad i = 1, 2, \ldots, n, \tag{8}$$

where μ_i is the correlation coefficient of the ith IMF with the original signal, and n is the total number of IMFs; $\max(\mu_i)$ is the maximum correlation coefficient observed. The selection criterion for IMFs is given as follows:

If $\mu_i \geq \lambda$, then keep the ith IMF,
else eliminate the ith IMF and add it to the residue.

When Eq. (8) is applied to a signal composed of two sinusoids of different frequencies, the relevant IMFs are successfully separated. Adding the eliminated IMFs to the residue clearly brings out the large swings at the signal ends. Although Eq. (8) by Peng *et al.* (2005) seems to be successful at discriminating between relevant and irrelevant IMFs, no analysis is mentioned that considers the effect of noise on the efficacy of the equation. Additionally, Eq. (8) was tried out by the authors of the present study on various forms of signals composed of sinusoids of differing frequencies, but it was not able to distinguish between the real IMFs and the irrelevant ones; it tended to accept most of the IMFs, which meant that the criterion was possibly not stringent enough. Furthermore, in the presence of uniformly distributed white noise, there was no apparent improvement.

3. Numerical Experiments and Discussion

Numerical experiments are run in order to establish a more reliable framework for retaining only relevant IMFs; various simulated signals with different frequencies, including noisy signals, are analyzed with EMD. The method consists of finding correlation coefficients between each IMF and the original signal, and determining a more stringent threshold coefficient that can be used to discriminate between relevant and irrelevant IMFs, especially considering signals that have noise. After decomposing several trial signals, a threshold is determined as a function of the maximum correlation coefficient; only IMFs with correlation coefficients greater than the prescribed threshold are retained. Eight signals are presented in Table 1 and shown in Figs. 2–9 together with their respective IMFs; the white noise in Signals 4, 6, and 8 is normally distributed white noise with mean 0 and standard deviation 1, which was generated using MATLAB® version 7. Tables 2–8 give the correlation coefficients between each IMF and the original signal. Equation (9) gives the expression for the threshold.

$$\mu_{TH} = \frac{\max(\mu_i)}{10 \times \max(\mu_i) - 3}, \quad i = 1, 2, \ldots, n, \tag{9}$$

where μ_{TH} is the threshold, μ_i is the correlation coefficient of the ith IMF with the original signal, and n is the total number of IMFs; $\max(\mu_i)$ is the

Table 1. Signal equations.

Signal	Equation
1	$y = \sin(3\pi t) + \sin(\pi t)$
2	$y = \cos(2\pi t) + \sin(5\pi t)$
3	$y = \sin(3\pi t) + \sin(\pi t) + \cos(2\pi t) + \sin(5\pi t)$
4	$y = \sin(3\pi t) + \sin(\pi t) + white\,noise$
5	$y = \sin(11\pi t) + \cos(3\pi t) + \sin(2.5\pi t)$
6	$y = \sin(9\pi t) + \cos(24\pi t) + white\,noise$
7	$y = 3\cos(\pi t) + 2\cos(13\pi t) + 2e^{-t}\cos(2\pi t)$
8	$y = \cos(2\pi t + 0.6\sin(2\pi t))e^{-0.25t} + white\,noise$

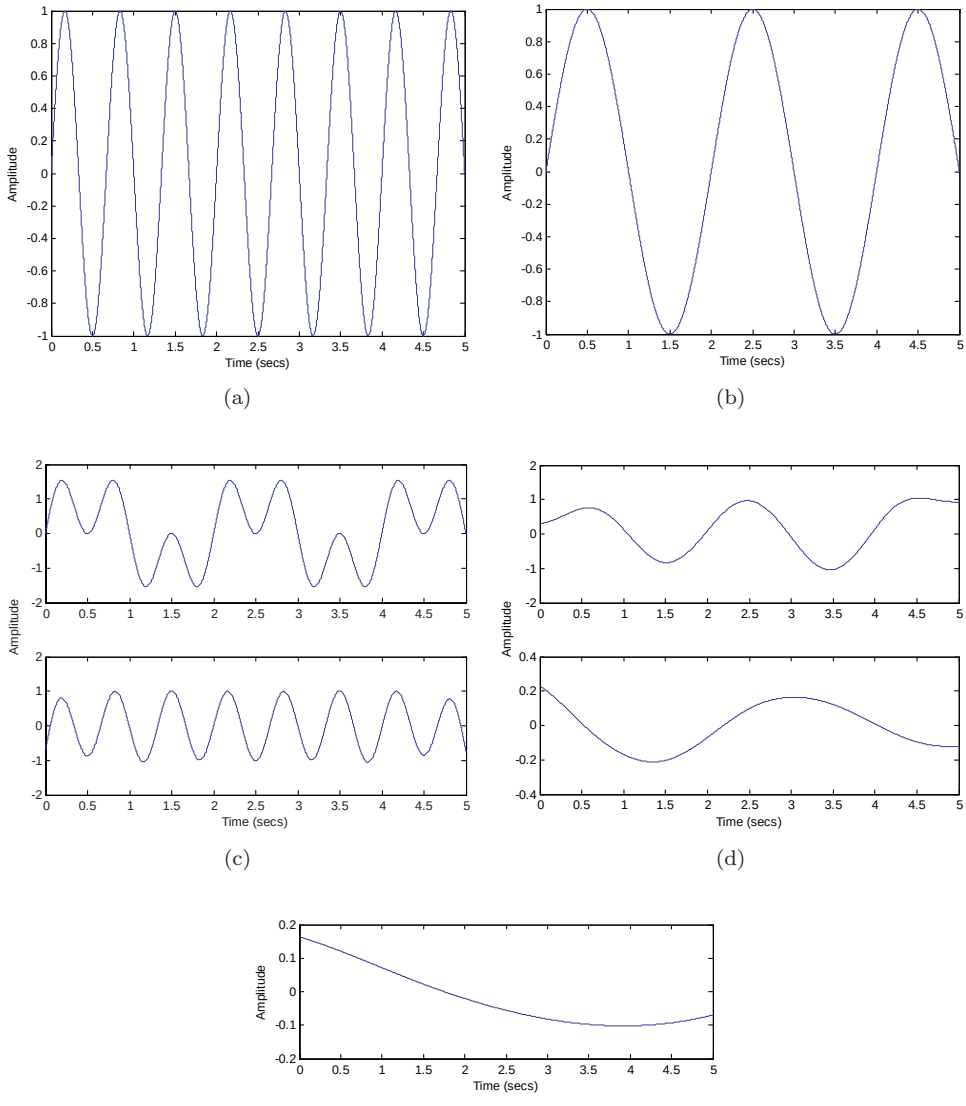


Fig. 2. (a) Signal 1: Component 1; (b) Signal 1: Component 2; (c) Signal 1: Original combined signal (top) and IMF 1 (below); (d) Signal 1: IMF 2 (top) and IMF 3 (below). (e) Signal 1: Residue.

maximum correlation coefficient observed. This new equation has been developed using both clean and noisy signals, including nonlinear signals. Comparing Figs. 2–9 and Table 9 it is observed from the time-scales of the retained IMFs that the threshold given by Eq. (9) is able to eliminate most of the spurious IMFs from the EMD. In the case of signals with noise, the first IMF is not considered since it has been observed in the literature that the first IMF usually carries

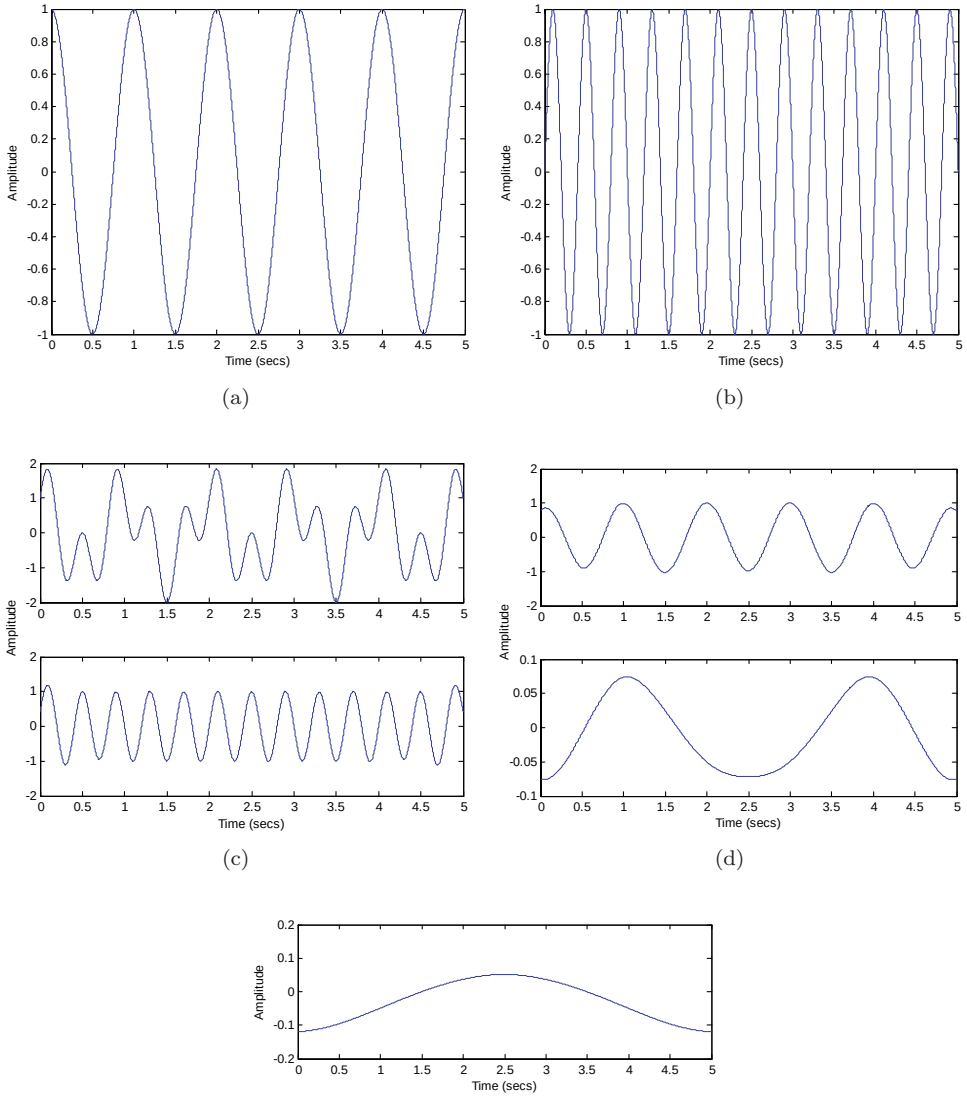


Fig. 3. (a) Signal 2: Component 1; (b) Signal 2: Component 2; (c) Signal 2: Original combined signal (top) and IMF 1 (below); (d) Signal 2: IMF 2 (top) and IMF 3 (below); (e) Signal 2: Residue.

most of the noise. However, it is observed in this study that the uniformly distributed white noise is actually decomposed into several IMFs before the other components of the original signal are revealed in the later IMFs; this is particularly evident in Signal 4 in which the original modes begin to show up as late as in the eighth IMF (Fig. 5(h)), meaning the first seven IMFs are from the white noise in the original signal. Equation (9) is more stringent than Eq. (8)

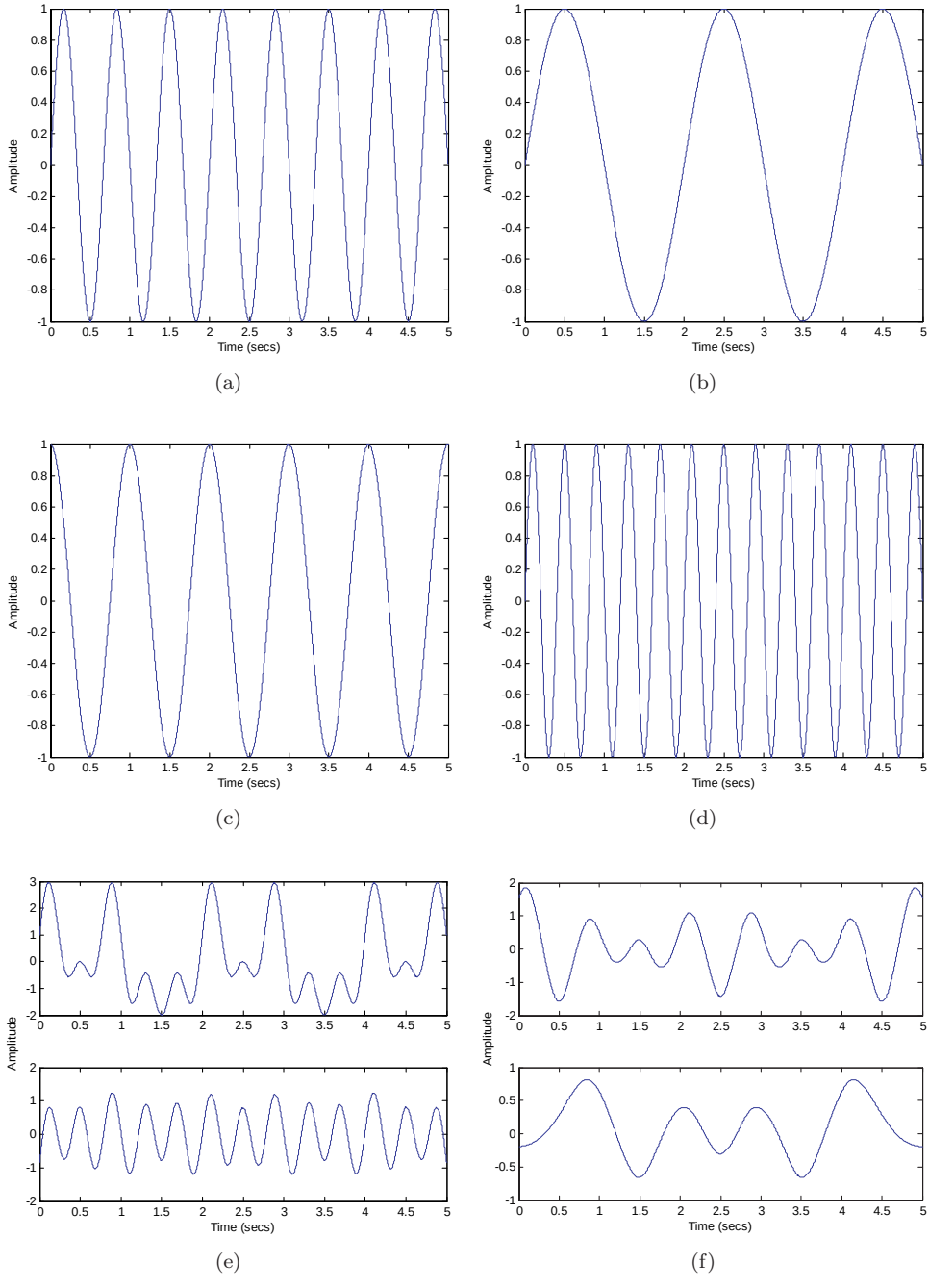
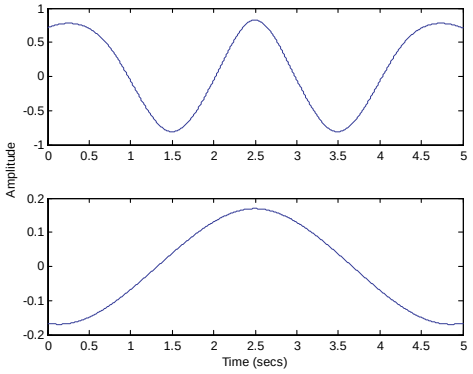
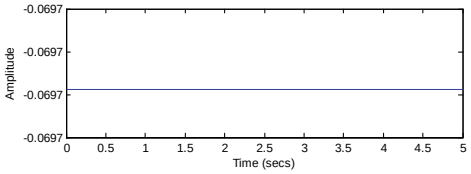


Fig. 4. (a) Signal 3: Component 1; (b) Signal 3: Component 2; (c) Signal 3: Component 3; (d) Signal 3: Component 4; (e) Signal 3: Original combined signal (top) and IMF 1 (below); (f) Signal 3: IMF 2 (top) and IMF 3 (below); (g) Signal 3: IMF 4 (top) and IMF 5 (below); (h) Signal 3: Residue.

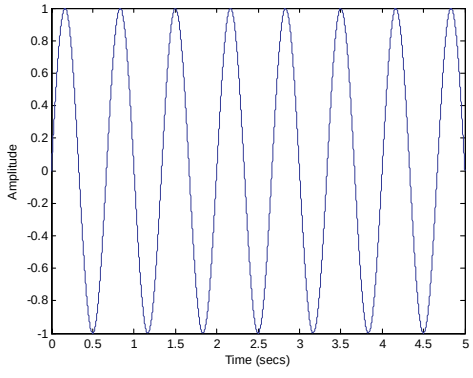


(g)

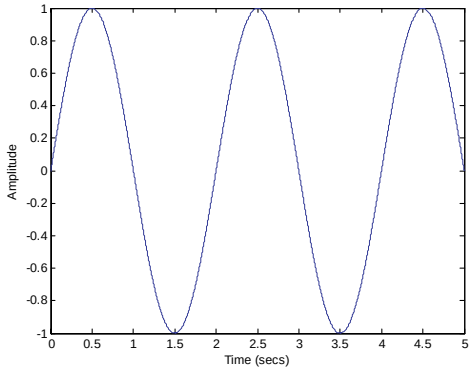


(h)

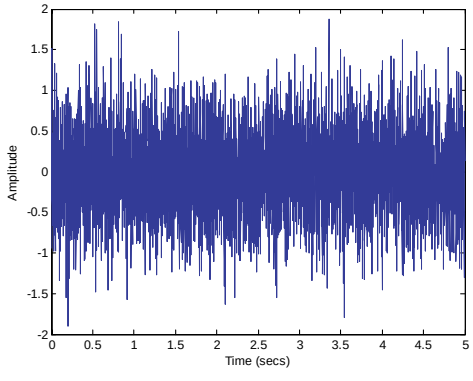
Fig. 4. (*Continued*)



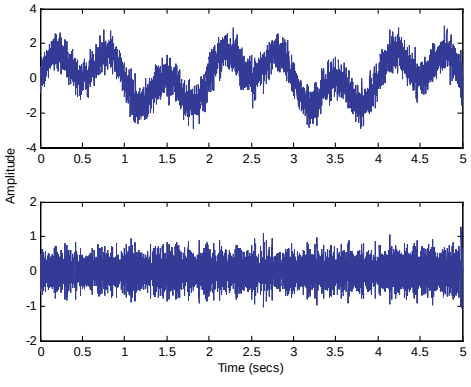
(a)



(b)

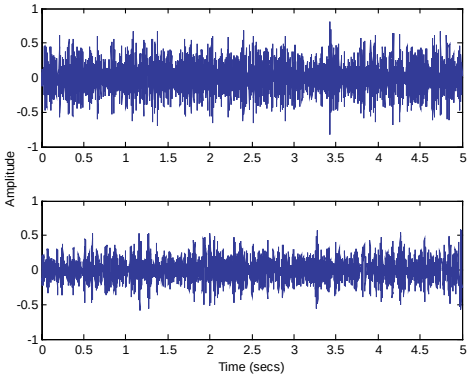


(c)

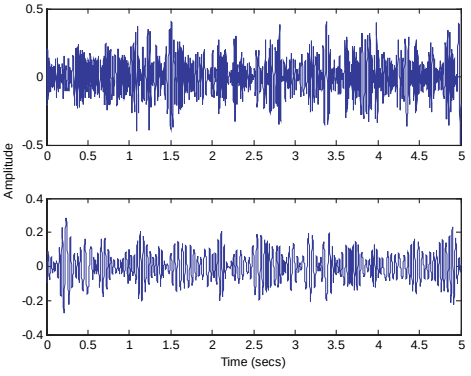


(d)

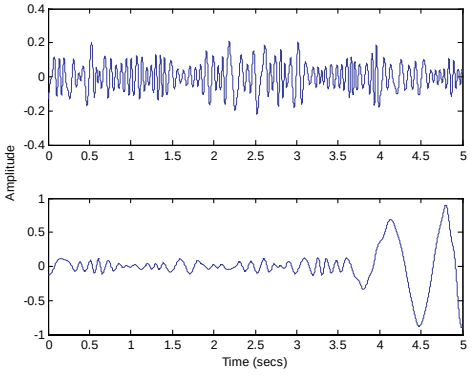
Fig. 5. (a) Signal 4: Component 1; (b) Signal 4: Component 2; (c) Signal 4: Component 3 (uniformly distributed white noise); (d) Signal 4: Original combined signal (top) and IMF 1 (below); (e) Signal 4: IMF 2 (top) and IMF 3 (below); (f) Signal 4: IMF 4 (top) and IMF 5 (below); (g) Signal 4: IMF 6 (top) and IMF 7 (below); (h) Signal 4: IMF 8 (top) and IMF 9 (below); (i) Signal 4: Residue.



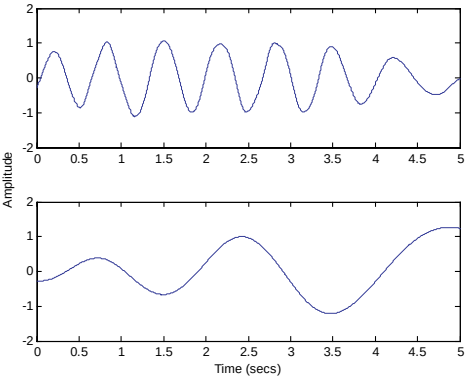
(e)



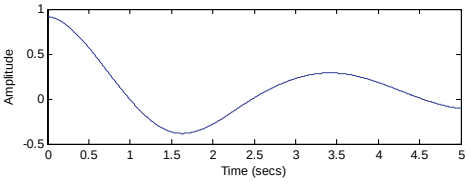
(f)



(g)



(h)



(i)

Fig. 5. (*Continued*)

in separating real and spurious IMFs especially for noisy signals. Using the proposed equation, it seems long-period, low-frequency components that make up the last few IMFs are mostly left out as irrelevant. This is not unusual since even in the original HHT proposed by Huang *et al.* (1998), the residue is left out of the subsequent Hilbert spectrum for fear that the energy contained in it might be overpowering because of uncertainties about the longer trend of which it is part. Furthermore, the IMFs that were not retained did not come from any of the

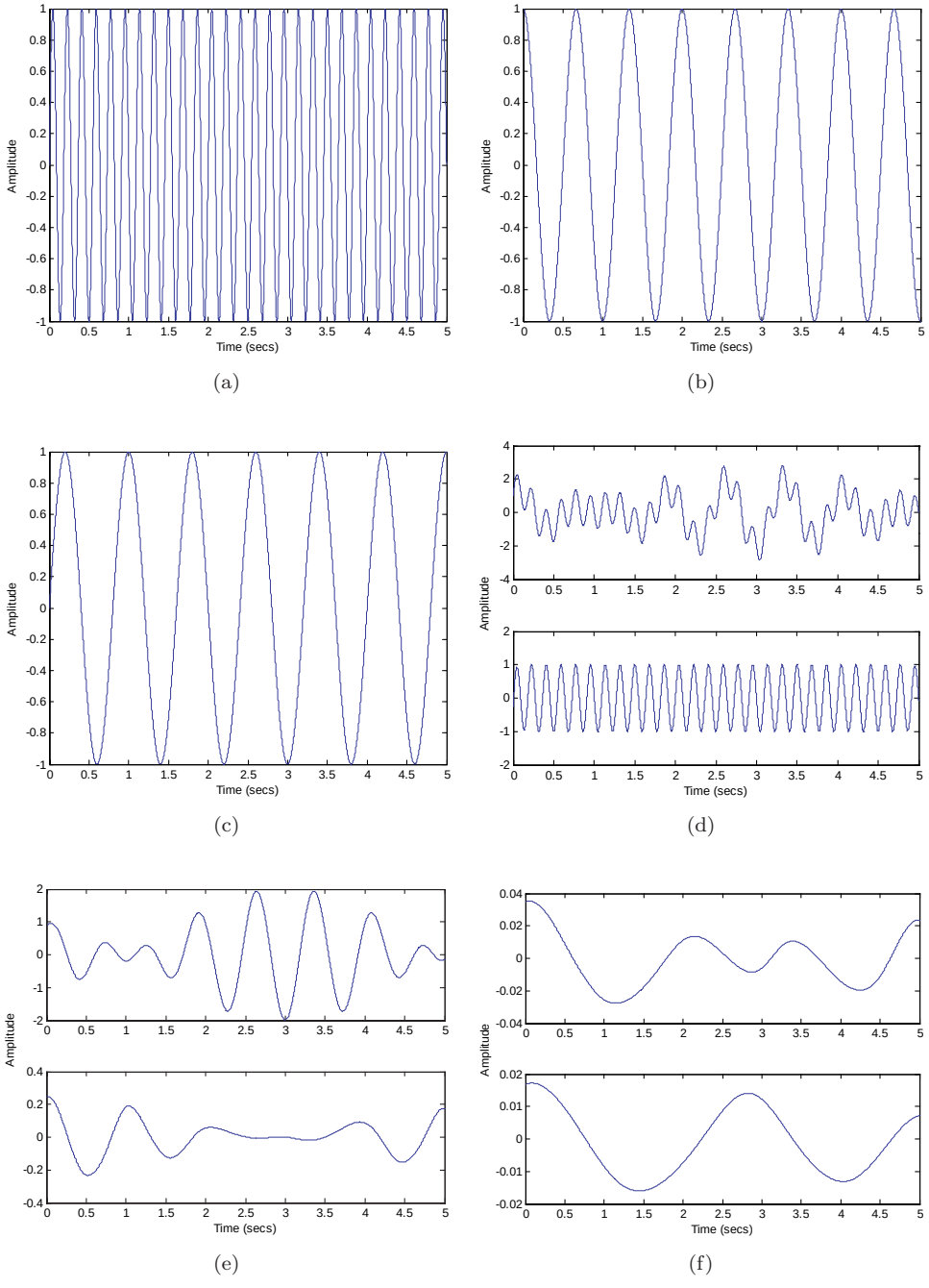
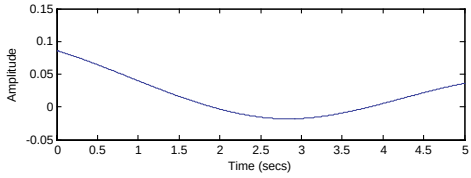


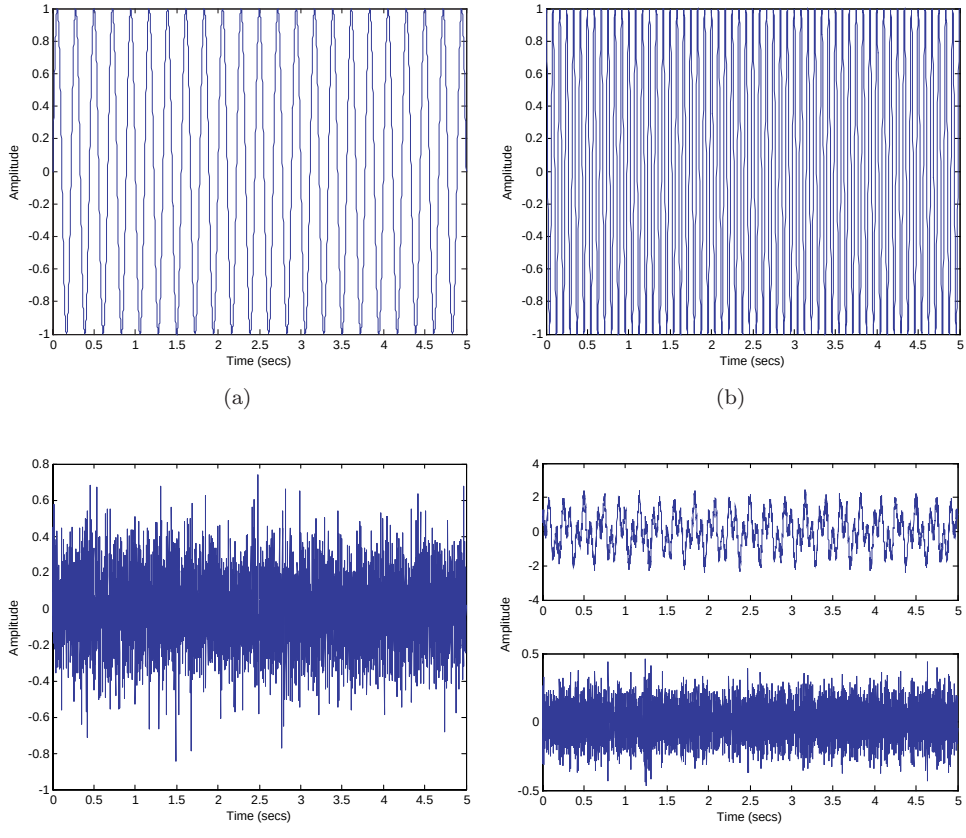
Fig. 6. (a) Signal 5: Component 1; (b) Signal 5: Component 2; (c) Signal 5: Component 3; (d) Signal 5: Original combined signal (top) and IMF 1 (below); (e) Signal 5: IMF 2 (top) and IMF 3 (below); (f) Signal 5: IMF 4 (top) and IMF 5 (below); (g) Signal 5: Residue.



(g)

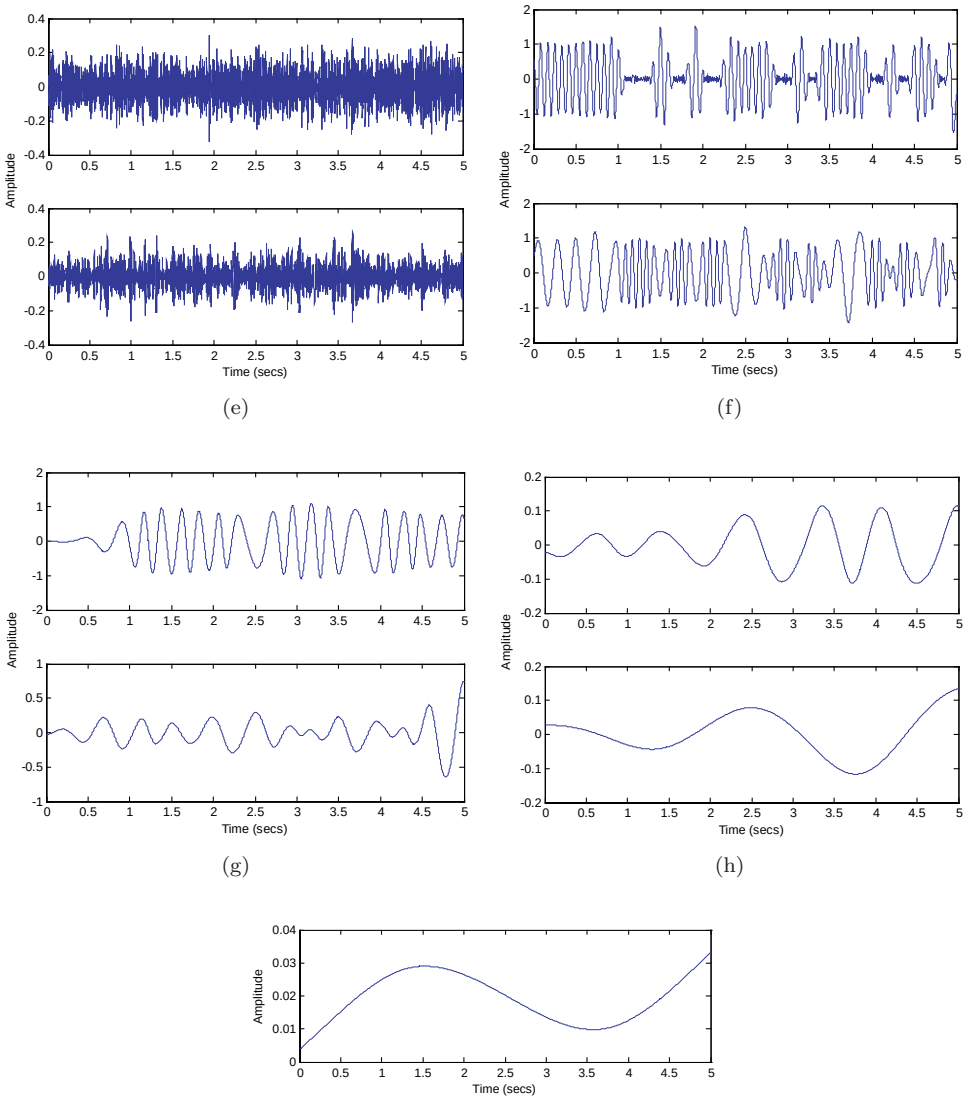
Fig. 6. (*Continued*)

original components used to create the original signal. Equation (9) improves on the equation proposed by Peng *et al.* (2005) by being able to deal better with noisy signals; this is also evident in Signal 4 in Table 5 in which Eq. (8) retains almost all the IMFs whereas Eq. (9) retains only three — IMFs 7–9. However, it should



(a) (b) (c) (d)

Fig. 7. (a) Signal 6: Component 1; (b) Signal 6: Component 2; (c) Signal 6: Component 3 (uniformly distributed white noise); (d) Signal 6: Original combined signal (top) and IMF 1 (below); (e) Signal 6: IMF 2 (top) and IMF 3 (below); (f) Signal 6: IMF 4 (top) and IMF 5 (below); (g) Signal 6: IMF 6 (top) and IMF 7 (below); (h) Signal 6: IMF 8 (top) and IMF 9 (below); (i) Signal 6: Residue.



(e) (f)

(g) (h)

(i)

Fig. 7. (*Continued*)

be noted that the presence of noise tends to mess up the results in that the first few IMFs could actually be noise IMFs, but Eq. (9) would eliminate most of these IMFs.

4. Application of New Threshold to Earthquake Data

For comparison purposes, the new threshold from Eq. (9) and the threshold from Eq. (8) are applied to seismic response data from five accelerometers located in a

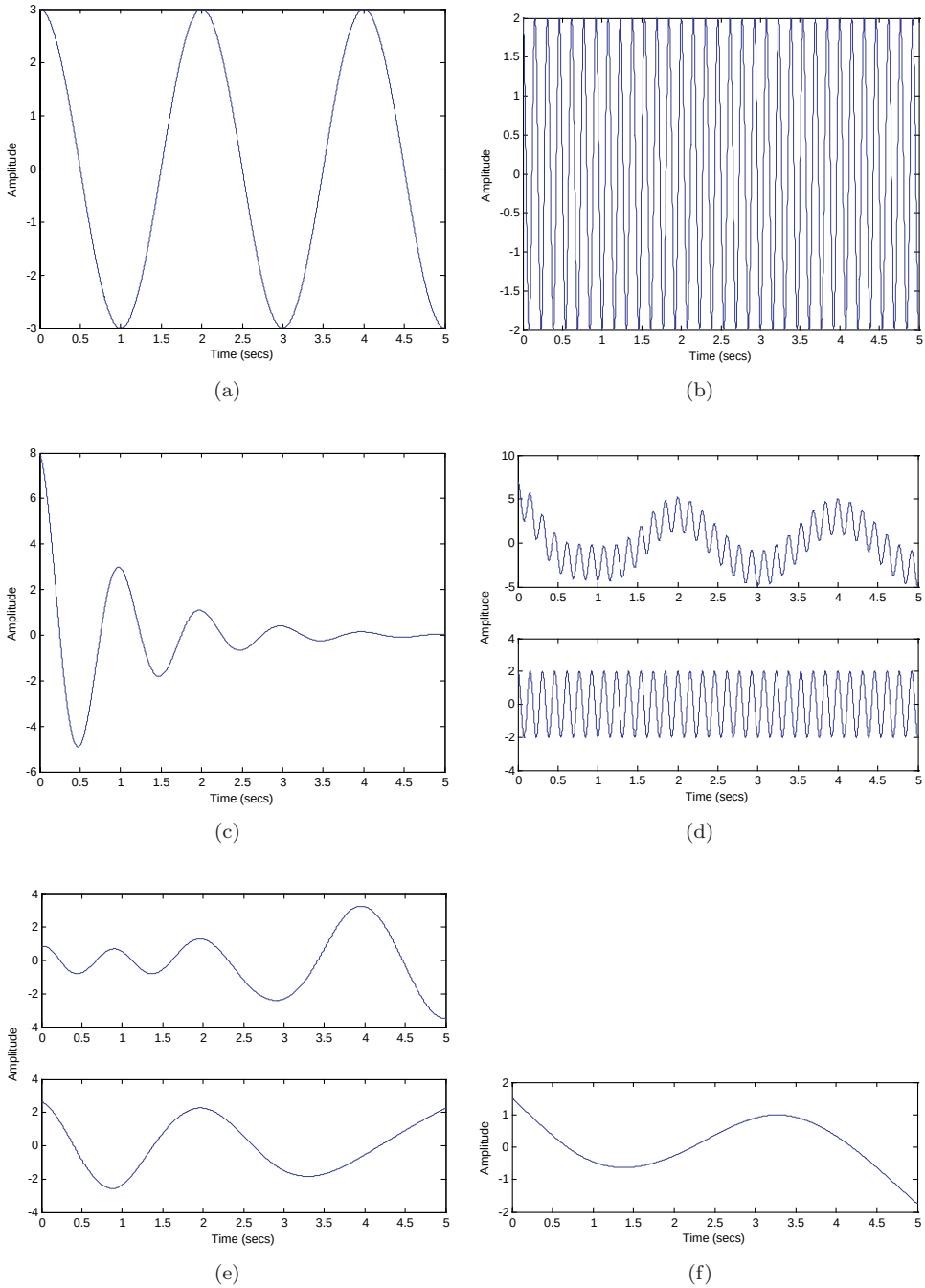


Fig. 8. (a) Signal 7: Component 1; (b) Signal 7: Component 2; (c) Signal 7: Component 3; (d) Signal 7: Original combined signal (top) and IMF 1 (below); (e) Signal 7: IMF 2 (top) and IMF 3 (below); (f) Signal 7: Residue.

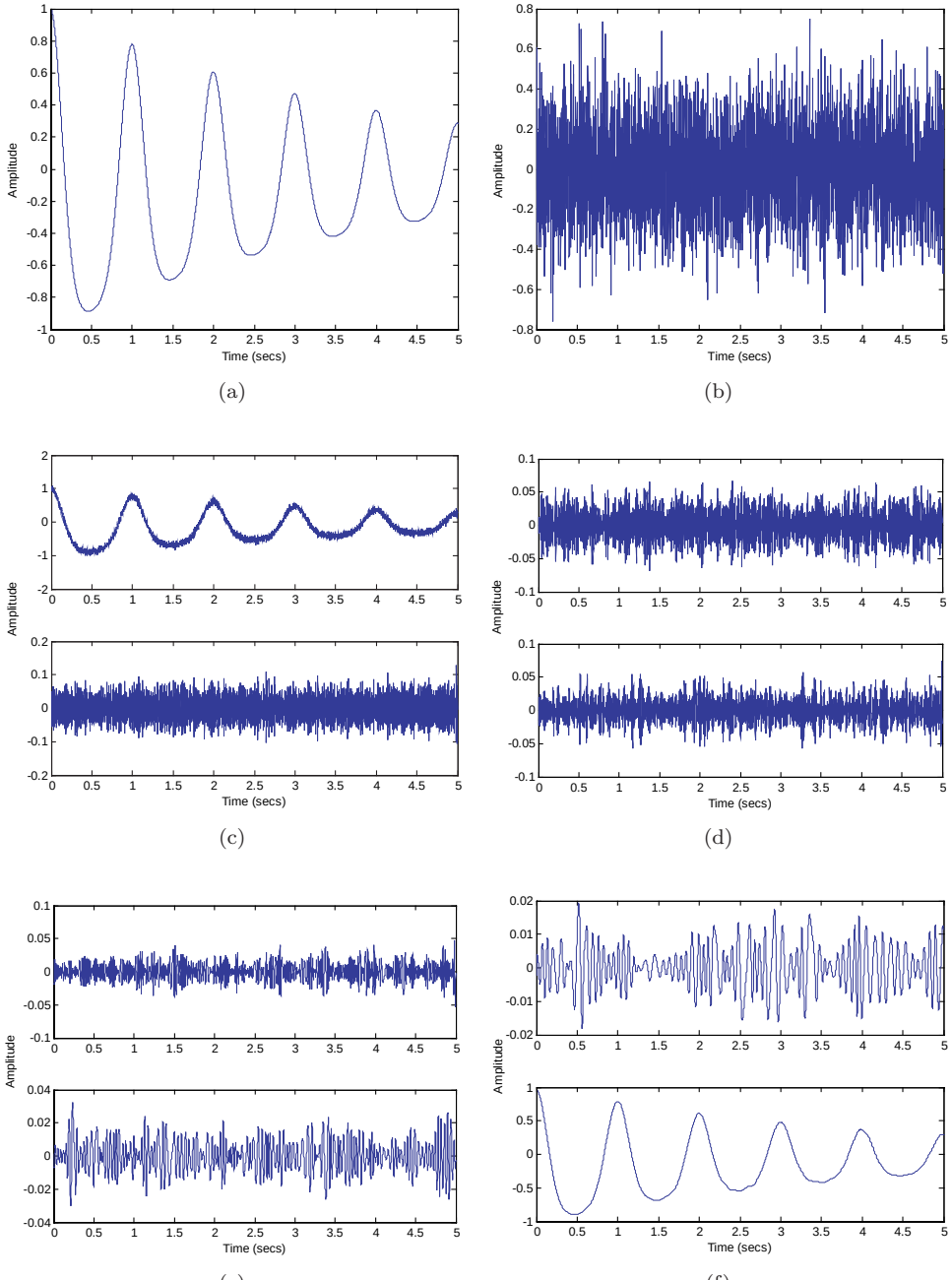
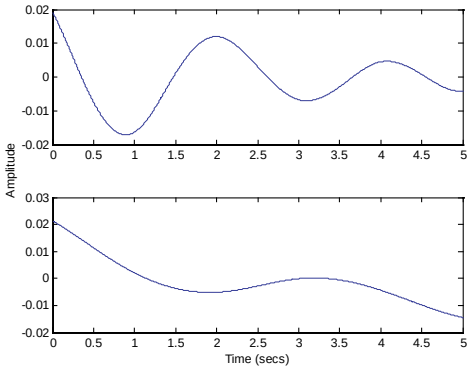


Fig. 9. (a) Signal 8: Component 1; (b) Signal 8: Component 2 (uniformly distributed white noise); (c) Signal 8: Original combined signal (top) and IMF 1 (below); (d) Signal 8: IMF 2 (top) and IMF 3 (below); (e) Signal 8: IMF 4 (top) and IMF 5 (below); (f) Signal 8: IMF 6 (top) and IMF 7 (below); (g) Signal 8: IMF 8 (top) and Residue (below).



(g)

Fig. 9. (*Continued*)

Table 2. Signal 1 analysis.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.7094	0.0709	1	0.1733	1
2	0.7073		2		2
3	0.1216		3		
4	0.1363		4		

Table 3. Signal 2 analysis.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.7447	0.0744	1	0.1675	1
2	0.7147		2		2
3	−0.0255				
4	−0.0418				

Table 4. Signal 3 analysis.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.5937	0.0685	1	0.1779	1
2	0.6853		2		2
3	0.4585		3		3
4	0.5007		4		4
5	−0.1951				
6	0.0000				

Table 5. Signal 4 analysis.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.3383				
2	0.1995		2		
3	0.1375		3		
4	0.0995		4		
5	0.0729		5		
6	0.0491	0.0565		0.2131	
7	0.2606		7		7
8	0.5653		8		8
9	0.5486		9		9
10	0.2092		10		

Table 6. Signal 5 analysis.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.6051		1		1
2	0.7945		2		2
3	0.2185		3		3
4	0.1062	0.0795	4	0.1607	
5	0.0497				
6	0.0702				

Table 7. Signal 6 analysis.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.1440				
2	0.0903		2		
3	0.0728		3		
4	0.5359		4		4
5	0.5912		5		5
6	0.4685	0.0591	6	0.2030	6
7	0.0452				
8	0.0122				
9	0.0179				
10	−0.0022				

Table 8. Signal 7 analysis.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.5415	0.0634	1	0.1897	1
2	0.6344		2		2
3	0.4511		3		3
4	0.1889		4		

Table 9. Signal 8 analysis.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.0892	0.0993		0.1433	
2	0.0538				
3	0.0378				
4	0.0165				
5	−0.0001				
6	0.0030				
7	0.9931		7		7
8	0.1015		8		
9	−0.0771				

Table 10. Accelerometer 1 retained IMFs.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.4425	0.0530		0.2307	
2	0.4923		2		2
3	0.5295		3		3
4	0.4151		4		4
5	0.3425		5		5
6	0.1908		6		
7	0.0936		7		
8	0.0751		8		
9	0.0377				
10	−0.0036				
11	−0.0012				
12	−0.0003				
13	−0.0002				
14	0.0003				

Table 11. Accelerometer 2 retained IMFs.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.0514				
2	0.1268		2		
3	0.3592		3		3
4	0.7466		4		4
5	0.4620		5		5
6	0.0869		6		
7	0.0314	0.0747		0.1672	
8	0.0057				
9	−0.0001				
10	0.0000				
11	0.0000				
12	0.0000				

Table 12. Accelerometer 3 retained IMFs.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.1002				
2	0.2041		2		
3	0.4608		3		3
4	0.5144		4		4
5	0.5142		5		5
6	0.2429		6		6
7	0.1757	0.0514	7	0.2399	
8	0.1367		8		
9	0.0005				
10	−0.0001				
11	0.0001				
12	0.0002				

tall building in an attempt to eliminate irrelevant IMFs; the results are shown in Tables 10–14. Although the results are comparable, usually Eq. (9) eliminates more IMFs than Eq. (8), sometimes as many as three more IMFs; the first IMF in each signal is ignored since it may likely contain noise in the signal [Huang *et al.* (1998)]. In the subsequent Hilbert spectrum only the retained IMFs are to be used, but the Hilbert spectra are not plotted here since the main objective is to attempt to distinguish between relevant and irrelevant IMFs.

5. Conclusion

Among the challenges of the EMD procedure are establishing mathematical rigor and physical meaning for the IMFs. Regarding the latter, attributing physical

Table 13. Accelerometer 4 retained IMFs.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.0255				
2	0.1442		2		
3	0.6474		3		3
4	0.6651		4		4
5	0.0997		5		
6	0.0421				
7	0.0762	0.0665	7	0.1822	
8	0.0253				
9	0.0000				
10	0.0000				
11	0.0001				
12	−0.0001				

Table 14. Accelerometer 5 retained IMFs.

IMF	Correlation coefficient with original signal, μ_i	Equation (8)		Equation (9)	
		Threshold	Retained IMFs after invoking threshold	Threshold	Retained IMFs after invoking threshold
1	0.0418				
2	0.1407		2		
3	0.7285		3		3
4	0.6728		4		4
5	0.1458		5		
6	0.2295		6		6
7	0.0359	0.0729		0.1700	
8	0.0024				
9	−0.0002				
10	0.0001				
11	0.0000				
12	0.0000				

meaning to a set of IMFs would depend critically on the context within which the IMFs are generated, that is, the source of the original signal and the objectives of the EMD analysis. This presupposes that a thorough understanding of the field in which the results of the analysis would be used is essential; as an example, explaining the physical meaning of the IMFs for the accelerometer output signals in Tables 1–5 above would require a sound understanding of earthquake engineering and the seismic response of tall buildings. However, before getting to the point of explaining what the IMFs mean in the decomposition it is important to know which IMFs are actually most relevant. Therefore, a quick, simple, and repeatable scientific framework is needed to discriminate between relevant and spurious IMFs;

this framework is presented in the form of a stringent threshold expressed as a function of the correlation coefficient between the original signal and its IMFs, given by Equation (9), which tends to work well even in the presence of noise. For civil infrastructure systems, the ability to do on-demand structural health monitoring is essential in keeping infrastructure in acceptable condition for longer service lives. EMD is an effective and efficient method for analyzing civil infrastructure condition data; therefore, proper interpretation of the IMFs from EMD would be helpful for sound interpretation of the health of the infrastructure.

Acknowledgments

This study was supported by the Bloc Davis Fellowship from the Department of Civil & Environmental Engineering at the University of Delaware, Newark, Delaware, USA.

References

Datig, M. and Schlurmann, T. (2004). Performance and limitations of the Hilbert–Huang transformation (HHT) with an application to irregular water waves. *Ocean Eng.*, **31**: 1783–1834.

Echeverria, J. C., Crowe, J. A., Woolfson, M. S. and Hayes-Gill, B. R. (2001). Application of empirical mode decomposition to heart rate variability analysis. *Med. Biol. Eng. Comput.*, **39**: 471–479.

Gloersen, P. and Huang, N. (2003). Comparison of interannual intrinsic modes in hemispheric sea ice covers and other geophysical parameters. *IEEE Trans. Geosci. Remote Sens.* **41**(5).

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 non-stationary time series analysis. *Proc. R. Soc. Lond. A* **454**: 903–995.

Huang, N. E., Wu, M. C., Long, S. R., Shen, S. S., Qu, W., Gloersen, P. and Fan, K. L. (2003). A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc. R. Soc. Lond. A* **459**: 2317–2345.

Huang, N. E., Wu, M.-L., Qu, W., Long, S. R. and Shen, S. S. P. (2003). Applications of Hilbert–Huang transform to non-stationary financial time series analysis. *Appl. Stochastic Models Bus. Indus.*, **19**: 245–268.

Hwang, P. A., Wang, D. W. and Kaihatu, J. M. (2005). A comparison of the energy flux computation of shoaling waves using Hilbert and wavelet spectral analysis techniques. *The Hilbert–Huang Transform in Engineering*, eds. Norden E. Huang and Nii O. Attoh-Okine, pp. 83–95.

Peng, Z. K., Tse Peter, W. and Chu, F. L. (2005). A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mech. Syst. Signal Process.*, **19**: 974–988.

Quek, S.-T., Tua, P.-S. and Wang, Q. (2005). Comparison of Hilbert–Huang, wavelet, and Fourier transforms for selected applications. *The Hilbert–Huang Transform in Engineering*, eds. Norden E. Huang and Nii O. Attoh-Okine, pp. 213–244.

Salisbury, J. I. and Sun, Y. (2007). Rapid screening test for sleep apnea using a nonlinear and nonstationary signal processing technique. *Med. Eng. Phys.*, **29**: 336–343.

Veltcheva, A. D. and Soares, C. G. (2004). Identification of the components of wave spectra by the Hilbert–Huang transform method. *Appl. Ocean Res.*, **26**: 1–12.

Wu, M.-C. (2007). Phase correlation of foreign exchange time series. *Physica A*, **375**: 633–642.

Wu, Z. and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. R. Soc. Lond. A*, **460**: 1597–1611.

Wu, Z. and Huang, N. E. (2005). Statistical significance test of intrinsic mode functions. *Hilbert–Huang Transform and Its Applications,* eds. N. E. Huang and S. S. P. Shen, World Scientific, New Jersey, pp. 107–127.