The Royal Society

The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis

By Norden E. Huang[1], Zheng Shen[2], Steven R. Long[3], Manli C. Wu[4], Hsing H. Shih[5], Quanan Zheng[6], Nai-Chyuan Yen[7], Chi Chao Tung[8] and Henry H. Liu[9]

[1]*Laboratory for Hydrospheric Processes/Oceans and Ice Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA*

[2]*Department of Earth and Planetary Sciences, The John Hopkins University, Baltimore, MD 21218, USA*

[3]*Laboratory for Hydrospheric Processes/Observational Science Branch, NASA Wallops Flight Facility, Wallops Island, VA 23337, USA*

[4]*Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA*

[5]*NOAA National Ocean Service, Silver Spring, MD 20910, USA*

[6]*College of Marine Studies, University of Delaware, DE 19716, USA*

[7]*Naval Research Laboratory, Washington, DC, 20375-5000, USA*

[8]*Department of Civil Engineering, North Carolina State University, Raleigh, NC 27695-7908, USA*

[9]*Naval Surface Warfare Center, Carderock Division, Bethesda, MD 20084-5000, USA*

Received 3 June 1996; accepted 4 November 1996

Contents

		PAGE
1.	Introduction	904
2.	Review of non-stationary data processing methods	907
	(*a*) The spectrogram	907
	(*b*) The wavelet analysis	907
	(*c*) The Wigner–Ville distribution	908
	(*d*) Evolutionary spectrum	909
	(*e*) The empirical orthogonal function expansion (EOF)	909
	(*f*) Other miscellaneous methods	910
3.	Instantaneous frequency	911
4.	Intrinsic mode functions	915
5.	The empirical mode decomposition method: the sifting process	917
6.	Completeness and orthogonality	923
7.	The Hilbert spectrum	928
8.	Validation and calibration of the Hilbert spectrum	933
9.	Applications	948
	(*a*) Numerical results from classic nonlinear systems	949
	(*b*) Observational data from laboratory and field experiments	962

Proc. R. Soc. Lond. A (1998) **454**, 903–995
Printed in Great Britain

10. Discussion 987
11. Conclusions 991
References 993

A new method for analysing nonlinear and non-stationary data has been developed. The key part of the method is the 'empirical mode decomposition' method with which any complicated data set can be decomposed into a finite and often small number of 'intrinsic mode functions' that admit well-behaved Hilbert transforms. This decomposition method is adaptive, and, therefore, highly efficient. Since the decomposition is based on the local characteristic time scale of the data, it is applicable to nonlinear and non-stationary processes. With the Hilbert transform, the 'instrinic mode functions' yield instantaneous frequencies as functions of time that give sharp identifications of imbedded structures. The final presentation of the results is an energy–frequency–time distribution, designated as the Hilbert spectrum. In this method, the main conceptual innovations are the introduction of 'intrinsic mode functions' based on local properties of the signal, which makes the instantaneous frequency meaningful; and the introduction of the instantaneous frequencies for complicated data sets, which eliminate the need for spurious harmonics to represent nonlinear and non-stationary signals. Examples from the numerical results of the classical nonlinear equation systems and data representing natural phenomena are given to demonstrate the power of this new method. Classical nonlinear system data are especially interesting, for they serve to illustrate the roles played by the nonlinear and non-stationary effects in the energy–frequency–time distribution.

Keywords: non-stationary time series; nonlinear differential equations; frequency–time spectrum; Hilbert spectral analysis; intrinsic time scale; empirical mode decomposition

1. Introduction

Data analysis is a necessary part in pure research and practical applications. Imperfect as some data might be, they represent the reality sensed by us; consequently, data analysis serves two purposes: to determine the parameters needed to construct the necessary model, and to confirm the model we constructed to represent the phenomenon. Unfortunately, the data, whether from physical measurements or numerical modelling, most likely will have one or more of the following problems: (*a*) the total data span is too short; (*b*) the data are non-stationary; and (*c*) the data represent nonlinear processes. Although each of the above problems can be real by itself, the first two are related, for a data section shorter than the longest time scale of a stationary process can appear to be non-stationary. Facing such data, we have limited options to use in the analysis.

Historically, Fourier spectral analysis has provided a general method for examining the global energy–frequency distributions. As a result, the term 'spectrum' has become almost synonymous with the Fourier transform of the data. Partially because of its prowess and partially because of its simplicity, Fourier analysis has dominated the data analysis efforts since soon after its introduction, and has been applied to all kinds of data. Although the Fourier transform is valid under extremely general conditions (see, for example, Titchmarsh 1948), there are some crucial restrictions of

the Fourier spectral analysis: the system must be linear; and the data must be strictly periodic or stationary; otherwise, the resulting spectrum will make little physical sense.

The stationarity requirement is not particular to the Fourier spectral analysis; it is a general one for most of the available data analysis methods. Therefore, it behoves us to review the definitions of stationarity here. According to the traditional definition, a time series, $X(t)$, is stationary in the wide sense, if, for all t,

$$\left.\begin{aligned} E(|X(t)^2|) &< \infty, \\ E(X(t)) &= m, \\ C(X(t_1), X(t_2)) = C(X(t_1+\tau), X(t_2+\tau)) &= C(t_1 - t_2), \end{aligned}\right\} \tag{1.1}$$

in which $E(\cdot)$ is the expected value defined as the ensemble average of the quantity, and $C(\cdot)$ is the covariance function. Stationarity in the wide sense is also known as weak stationarity, covariance stationarity or second-order stationarity (see, for example, Brockwell & Davis 1991). A time series, $X(t)$, is strictly stationary, if the joint distribution of

$$[X(t_1), X(t_2), \ldots, X(t_n)] \quad \text{and} \quad [X(t_1+\tau), X(t_2+\tau), \ldots, X(t_n+\tau)] \tag{1.2}$$

are the same for all t_i and τ. Thus, a strictly stationary process with finite second moments is also weakly stationary, but the inverse is not true. Both definitions are rigorous but idealized. Other less rigorous definitions for stationarity have also been used; for example, piecewise stationarity is for any random variable that is stationary within a limited time span, and asymptotically stationary is for any random variable that is stationary when τ in equations (1.1) or (1.2) approaches infinity. In practice, we can only have data for finite time spans; therefore, even to check these definitions, we have to make approximations. Few of the data sets, from either natural phenomena or artificial sources, can satisfy these definitions. It may be argued that the difficulty of invoking stationarity as well as ergodicity is not on principle but on practicality: we just cannot have enough data to cover all possible points in the phase plane; therefore, most of the cases facing us are transient in nature. This is the reality; we are forced to face it.

Other than stationarity, Fourier spectral analysis also requires linearity. Although many natural phenomena can be approximated by linear systems, they also have the tendency to be nonlinear whenever their variations become finite in amplitude. Compounding these complications is the imperfection of our probes or numerical schemes; the interactions of the imperfect probes even with a perfect linear system can make the final data nonlinear. For the above reasons, the available data are usually of finite duration, non-stationary and from systems that are frequently nonlinear, either intrinsically or through interactions with the imperfect probes or numerical schemes. Under these conditions, Fourier spectral analysis is of limited use. For lack of alternatives, however, Fourier spectral analysis is still used to process such data. The uncritical use of Fourier spectral analysis and the insouciant adoption of the stationary and linear assumptions may give misleading results; some of those are described as follows.

First, the Fourier spectrum defines uniform harmonic components globally; therefore, it needs many additional harmonic components to simulate non-stationary data that are non-uniform globally. As a result, it spreads the energy over a wide frequency range. For example, using a delta function to represent a flash of light will give

a phase-locked wide white Fourier spectrum. Here, many Fourier components are added to simulate the non-stationary nature of the data in the time domain, but their existence diverts energy to a much wider frequency domain. Constrained by the energy conservation, these spurious harmonics and the wide frequency spectrum cannot faithfully represent the true energy density in the frequency space. More seriously, the Fourier representation also requires the existence of negative light intensity so that the components can cancel out one another to give the final delta function. Thus, the Fourier components might make mathematical sense, but do not really make physical sense at all. Although no physical process can be represented exactly by a delta function, some data such as the near-field strong earthquake records are of extremely short durations lasting only a few seconds to tens of seconds at most. Such records almost approach a delta function, and they always give artificially wide Fourier spectra.

Second, Fourier spectral analysis uses linear superposition of trigonometric functions; therefore, it needs additional harmonic components to simulate the deformed wave-profiles. Such deformations, as will be shown later, are the direct consequence of nonlinear effects. Whenever the form of the data deviates from a pure sine or cosine function, the Fourier spectrum will contain harmonics. As explained above, both non-stationarity and nonlinearity can induce spurious harmonic components that cause energy spreading. The consequence is the misleading energy–frequency distribution for nonlinear and non-stationary data.

In this paper, we will present a new data analysis method based on the empirical mode decomposition (EMD) method, which will generate a collection of intrinsic mode functions (IMF). The decomposition is based on the direct extraction of the energy associated with various intrinsic time scales, the most important parameters of the system. Expressed in IMFs, they have well-behaved Hilbert transforms, from which the instantaneous frequencies can be calculated. Thus, we can localize any event on the time as well as the frequency axis. The decomposition can also be viewed as an expansion of the data in terms of the IMFs. Then, these IMFs, based on and derived from the data, can serve as the basis of that expansion which can be linear or nonlinear as dictated by the data, and it is complete and almost orthogonal. Most important of all, it is adaptive. As will be shown later in more detail, locality and adaptivity are the necessary conditions for the basis for expanding nonlinear and non-stationary time series; orthogonality is not a necessary criterion for our basis selection for a nonlinear system. The principle of this basis construction is based on the physical time scales that characterize the oscillations of the phenomena. The local energy and the instantaneous frequency derived from the IMFs through the Hilbert transform can give us a full energy–frequency–time distribution of the data. Such a representation is designated as the Hilbert spectrum; it would be ideal for nonlinear and non-stationary data analysis.

We have obtained good results and new insights by applying the combination of the EMD and Hilbert spectral analysis methods to various data: from the numerical results of the classical nonlinear equation systems to data representing natural phenomena. The classical nonlinear systems serve to illustrate the roles played by the nonlinear effects in the energy–frequency–time distribution. With the low degrees of freedom, they can train our eyes for more complicated cases. Some limitations of this method will also be discussed and the conclusions presented. Before introducing the new method, we will first review the present available data analysis methods for non-stationary processes.

2. Review of non-stationary data processing methods

We will first give a brief survey of the methods available for processing non-stationary data. Since most of the methods still depend on Fourier analysis, they are limited to linear systems only. Of the few methods available, the adoption of any method is almost strictly determined according to the special field in which the application is made. The available methods are reviewed as follows.

(*a*) *The spectrogram*

The spectrogram is the most basic method, which is nothing but a limited time window-width Fourier spectral analysis. By successively sliding the window along the time axis, one can get a time–frequency distribution. Since it relies on the traditional Fourier spectral analysis, one has to assume the data to be piecewise stationary. This assumption is not always justified in non-stationary data. Even if the data are piecewise stationary how can we guarantee that the window size adopted always coincides with the stationary time scales? What can we learn about the variations longer than the local stationary time scale? Will the collection of the locally stationary pieces constitute some longer period phenomena? Furthermore, there are also practical difficulties in applying the method: in order to localize an event in time, the window width must be narrow, but, on the other hand, the frequency resolution requires longer time series. These conflicting requirements render this method of limited usage. It is, however, extremely easy to implement with the fast Fourier transform; thus, it has attracted a wide following. Most applications of this method are for qualitative display of speech pattern analysis (see, for example, Oppenheim & Schafer 1989).

(*b*) *The wavelet analysis*

The wavelet approach is essentially an adjustable window Fourier spectral analysis with the following general definition:

$$W(a,b;X,\psi) = |a|^{-1/2} \int_{-\infty}^{\infty} X(t)\psi^* \left(\frac{t-b}{a}\right) \mathrm{d}t, \tag{2.1}$$

in which $\psi^*(\cdot)$ is the basic wavelet function that satisfies certain very general conditions, a is the dilation factor and b is the translation of the origin. Although time and frequency do not appear explicitly in the transformed result, the variable $1/a$ gives the frequency scale and b, the temporal location of an event. An intuitive physical explanation of equation (2.1) is very simple: $W(a,b;X,\psi)$ is the 'energy' of X of scale a at $t=b$.

Because of this basic form of $at+b$ involved in the transformation, it is also known as affine wavelet analysis. For specific applications, the basic wavelet function, $\psi^*(\cdot)$, can be modified according to special needs, but the form has to be given before the analysis. In most common applications, however, the Morlet wavelet is defined as Gaussian enveloped sine and cosine wave groups with 5.5 waves (see, for example, Chan 1995). Generally, $\psi^*(\cdot)$ is not orthogonal for different a for continuous wavelets. Although one can make the wavelet orthogonal by selecting a discrete set of a, this discrete wavelet analysis will miss physical signals having scale different from the selected discrete set of a. Continuous or discrete, the wavelet analysis is basically a linear analysis. A very appealing feature of the wavelet analysis is that it provides a

uniform resolution for all the scales. Limited by the size of the basic wavelet function, the downside of the uniform resolution is uniformly poor resolution.

Although wavelet analysis has been available only in the last ten years or so, it has become extremely popular. Indeed, it is very useful in analysing data with gradual frequency changes. Since it has an analytic form for the result, it has attracted extensive attention of the applied mathematicians. Most of its applications have been in edge detection and image compression. Limited applications have also been made to the time–frequency distribution in time series (see, for example, Farge 1992; Long *et al.* 1993) and two-dimensional images (Spedding *et al.* 1993).

Versatile as the wavelet analysis is, the problem with the most commonly used Morlet wavelet is its leakage generated by the limited length of the basic wavelet function, which makes the quantitative definition of the energy–frequency–time distribution difficult. Sometimes, the interpretation of the wavelet can also be counter-intuitive. For example, to define a change occurring locally, one must look for the result in the high-frequency range, for the higher the frequency the more localized the basic wavelet will be. If a local event occurs only in the low-frequency range, one will still be forced to look for its effects in the high-frequency range. Such interpretation will be difficult if it is possible at all (see, for example, Huang *et al.* 1996). Another difficulty of the wavelet analysis is its non-adaptive nature. Once the basic wavelet is selected, one will have to use it to analyse all the data. Since the most commonly used Morlet wavelet is Fourier based, it also suffers the many shortcomings of Fourier spectral analysis: it can only give a physically meaningful interpretation to linear phenomena; it can resolve the interwave frequency modulation provided the frequency variation is gradual, but it cannot resolve the intrawave frequency modulation because the basic wavelet has a length of 5.5 waves. In spite of all these problems, wavelet analysis is still the best available non-stationary data analysis method so far; therefore, we will use it in this paper as a reference to establish the validity and the calibration of the Hilbert spectrum.

(*c*) *The Wigner–Ville distribution*

The Wigner–Ville distribution is sometimes also referred to as the Heisenberg wavelet. By definition, it is the Fourier transform of the central covariance function. For any time series, $X(t)$, we can define the central variance as

$$C_{\mathrm{c}}(\tau, t) = X(t - \tfrac{1}{2}\tau)X^*(t + \tfrac{1}{2}\tau). \tag{2.2}$$

Then the Wigner–Ville distribution is

$$V(\omega, t) = \int_{-\infty}^{\infty} C_{\mathrm{c}}(\tau, t)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau. \tag{2.3}$$

This transform has been treated extensively by Claasen & Mecklenbräuker (1980*a*, *b*, *c*) and by Cohen (1995). It has been extremely popular with the electrical engineering community.

The difficulty with this method is the severe cross terms as indicated by the existence of negative power for some frequency ranges. Although this shortcoming can be eliminated by using the Kernel method (see, for example, Cohen 1995), the result is, then, basically that of a windowed Fourier analysis; therefore, it suffers all the limitations of the Fourier analysis. An extension of this method has been made by Yen (1994), who used the Wigner–Ville distribution to define wave packets that reduce

a complicated data set to a finite number of simple components. This extension is very powerful and can be applied to a variety of problems. The applications to complicated data, however, require a great amount of judgement.

(*d*) *Evolutionary spectrum*

The evolutionary spectrum was first proposed by Priestley (1965). The basic idea is to extend the classic Fourier spectral analysis to a more generalized basis: from sine or cosine to a family of orthogonal functions $\{\phi(\omega,t)\}$ indexed by time, t, and defined for all real ω, the frequency. Then, any real random variable, $X(t)$, can be expressed as

$$X(t)=\int_{-\infty}^{\infty}\phi(\omega,t)\,\mathrm{d}A(\omega,t), \tag{2.4}$$

in which $\mathrm{d}A(\omega,t)$, the Stieltjes function for the amplitude, is related to the spectrum as

$$E(|\,\mathrm{d}A(\omega,t)|^2)=\mathrm{d}\mu(\omega,t)=S(\omega,t)\,\mathrm{d}\omega, \tag{2.5}$$

where $\mu(\omega,t)$ is the spectrum, and $S(\omega,t)$ is the spectral density at a specific time t, also designated as the evolutionary spectrum. If for each fixed ω, $\phi(\omega,t)$ has a Fourier transform

$$\phi(\omega,t)=a(\omega,t)\mathrm{e}^{\mathrm{i}\Omega(\omega)t}, \tag{2.6}$$

then the function of $a(\omega,t)$ is the envelope of $\phi(\omega,t)$, and $\Omega(\omega)$ is the frequency. If, further, we can treat $\Omega(\omega)$ as a single valued function of ω, then

$$\phi(\omega,t)=\alpha(\omega,t)\mathrm{e}^{\mathrm{i}\omega t}. \tag{2.7}$$

Thus, the original data can be expanded in a family of amplitude modulated trigonometric functions.

The evolutionary spectral analysis is very popular in the earthquake community (see, for example, Liu 1970, 1971, 1973; Lin & Cai 1995). The difficulty of its application is to find a method to define the basis, $\{\phi(\omega,t)\}$. In principle, for this method to work, the basis has to be defined *a posteriori*. So far, no systematic way has been offered; therefore, constructing an evolutionary spectrum from the given data is impossible. As a result, in the earthquake community, the applications of this method have changed the problem from data analysis to data simulation: an evolutionary spectrum will be assumed, then the signal will be reconstituted based on the assumed spectrum. Although there is some general resemblance to the simulated earthquake signal with the real data, it is not the data that generated the spectrum. Consequently, evolutionary spectrum analysis has never been very useful. As will be shown, the EMD can replace the evolutionary spectrum with a truly adaptive representation for the non-stationary processes.

(*e*) *The empirical orthogonal function expansion (EOF)*

The empirical orthogonal function expansion (EOF) is also known as the principal component analysis, or singular value decomposition method. The essence of EOF is briefly summarized as follows: for any real $z(x,t)$, the EOF will reduce it to

$$z(x,t)=\sum_{1}^{n}a_k(t)f_k(x), \tag{2.8}$$

in which

$$f_j \cdot f_k = \delta_{jk}. \tag{2.9}$$

The orthonormal basis, $\{f_k\}$, is the collection of the empirical eigenfunctions defined by

$$C \cdot f_k = \lambda_k f_k, \tag{2.10}$$

where C is the sum of the inner products of the variable.

EOF represents a radical departure from all the above methods, for the expansion basis is derived from the data; therefore, it is *a posteriori*, and highly efficient. The critical flaw of EOF is that it only gives a distribution of the variance in the modes defined by $\{f_k\}$, but this distribution by itself does not suggest scales or frequency content of the signal. Although it is tempting to interpret each mode as independent variations, this interpretation should be viewed with great care, for the EOF decomposition is not unique. A single component out of a non-unique decomposition, even if the basis is orthogonal, does not usually contain physical meaning. Recently, Vautard & Ghil (1989) proposed the singular spectral analysis method, which is the Fourier transform of the EOF. Here again, we have to be sure that each EOF component is stationary, otherwise the Fourier spectral analysis will make little sense on the EOF components. Unfortunately, there is no guarantee that EOF components from a nonlinear and non-stationary data set will all be linear and stationary. Consequently, singular spectral analysis is not a real improvement. Because of its adaptive nature, however, the EOF method has been very popular, especially in the oceanography and meteorology communities (see, for example, Simpson 1991).

(*f*) *Other miscellaneous methods*

Other than the above methods, there are also some miscellaneous methods such as least square estimation of the trend, smoothing by moving averaging, and differencing to generate stationary data. Methods like these, though useful, are too specialized to be of general use. They will not be discussed any further here. Additional details can be found in many standard data processing books (see, for example, Brockwell & Davis 1991).

All the above methods are designed to modify the global representation of the Fourier analysis, but they all failed in one way or the other. Having reviewed the methods, we can summarize the necessary conditions for the basis to represent a nonlinear and non-stationary time series: (*a*) complete; (*b*) orthogonal; (*c*) local; and (*d*) adaptive.

The first condition guarantees the degree of precision of the expansion; the second condition guarantees positivity of energy and avoids leakage. They are the standard requirements for all the linear expansion methods. For nonlinear expansions, the orthogonality condition needs to be modified. The details will be discussed later. But even these basic conditions are not satisfied by some of the above mentioned methods. The additional conditions are particular to the nonlinear and non-stationary data. The requirement for locality is the most crucial for non-stationarity, for in such data there is no time scale; therefore, all events have to be identified by the time of their occurences. Consequently, we require both the amplitude (or energy) and the frequency to be functions of time. The requirement for adaptivity is also crucial for both nonlinear and non-stationary data, for only by adapting to the local variations of the data can the decomposition fully account for the underlying physics

of the processes and not just to fulfil the mathematical requirements for fitting the data. This is especially important for the nonlinear phenomena, for a manifestation of nonlinearity is the 'harmonic distortion' in the Fourier analysis. The degree of distortion depends on the severity of nonlinearity; therefore, one cannot expect a predetermined basis to fit all the phenomena. An easy way to generate the necessary adaptive basis is to derive the basis from the data.

In this paper, we will introduce a general method which requires two steps in analysing the data as follows. The first step is to preprocess the data by the empirical mode decomposition method, with which the data are decomposed into a number of intrinsic mode function components. Thus, we will expand the data in a basis derived from the data. The second step is to apply the Hilbert transform to the decomposed IMFs and construct the energy–frequency–time distribution, designated as the Hilbert spectrum, from which the time localities of events will be preserved. In other words, we need the instantaneous frequency and energy rather than the global frequency and energy defined by the Fourier spectral analysis. Therefore, before going any further, we have to clarify the definition of the instantaneous frequency.

3. Instantaneous frequency

The notion of the instantaneous energy or the instantaneous envelope of the signal is well accepted; the notion of the instantaneous frequency, on the other hand, has been highly controversial. Existing opinions range from editing it out of existence (Shekel 1953) to accepting it but only for special 'monocomponent' signals (Boashash 1992; Cohen 1995).

There are two basic difficulties with accepting the idea of an instantaneous frequency as follows. The first one arises from the deeply entrenched influence of the Fourier spectral analysis. In the traditional Fourier analysis, the frequency is defined for the sine or cosine function spanning the whole data length with constant amplitude. As an extension of this definition, the instantaneous frequencies also have to relate to either a sine or a cosine function. Thus, we need at least one full oscillation of a sine or a cosine wave to define the local frequency value. According to this logic, nothing shorter than a full wave will do. Such a definition would not make sense for non-stationary data for which the frequency has to change values from time to time. The second difficulty arises from the non-unique way in defining the instantaneous frequency. Nevertheless, this difficulty is no longer serious since the introduction of the means to make the data analytical through the Hilbert transform. Difficulties, however, still exist as 'paradoxes' discussed by Cohen (1995). For an arbitrary time series, $X(t)$, we can always have its Hilbert Transform, $Y(t)$, as

$$Y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{X(t')}{t - t'} \, \mathrm{d}t', \tag{3.1}$$

where P indicates the Cauchy principal value. This transform exists for all functions of class L^p (see, for example, Titchmarsh 1948). With this definition, $X(t)$ and $Y(t)$ form the complex conjugate pair, so we can have an analytic signal, $Z(t)$, as

$$Z(t) = X(t) + \mathrm{i}Y(t) = a(t)\mathrm{e}^{\mathrm{i}\theta(t)}, \tag{3.2}$$

in which

$$a(t) = [X^2(t) + Y^2(t)]^{1/2}, \quad \theta(t) = \arctan\left(\frac{Y(t)}{X(t)}\right). \tag{3.3}$$

Theoretically, there are infinitely many ways of defining the imaginary part, but the Hilbert transform provides a unique way of defining the imaginary part so that the result is an analytic function. A brief tutorial on the Hilbert transform with the emphasis on its physical interpretation can be found in Bendat & Piersol (1986). Essentially equation (3.1) defines the Hilbert transform as the convolution of $X(t)$ with $1/t$; therefore, it emphasizes the local properties of $X(t)$. In equation (3.2), the polar coordinate expression further clarifies the local nature of this representation: it is the best local fit of an amplitude and phase varying trigonometric function to $X(t)$. Even with the Hilbert transform, there is still considerable controversy in defining the instantaneous frequency as

$$\omega = \frac{\mathrm{d}\theta(t)}{\mathrm{d}t}. \tag{3.4}$$

This leads Cohen (1995) to introduce the term, 'monocomponent function'. In principle, some limitations on the data are necessary, for the instantaneous frequency given in equation (3.4) is a single value function of time. At any given time, there is only one frequency value; therefore, it can only represent one component, hence 'monocomponent'. Unfortunately, no clear definition of the 'monocomponent' signal was given to judge whether a function is or is not 'monocomponent'. For lack of a precise definition, 'narrow band' was adopted as a limitation on the data for the instantaneous frequency to make sense (Schwartz *et al.* 1966).

There are two definitions for bandwidth. The first one is used in the study of the probability properties of the signals and waves, where the processes are assumed to be stationary and Gaussian. Then, the bandwidth can be defined in terms of spectral moments as follows. The expected number of zero crossings per unit time is given by

$$N_0 = \frac{1}{\pi}\left(\frac{m_2}{m_0}\right)^{1/2}, \tag{3.5}$$

while the expected number of extrema per unit time is given by

$$N_1 = \frac{1}{\pi}\left(\frac{m_4}{m_2}\right)^{1/2}, \tag{3.6}$$

in which m_i is the ith moment of the spectrum. Therefore, the parameter, ν, defined as

$$N_1^2 - N_0^2 = \frac{1}{\pi^2}\frac{m_4 m_0 - m_2^2}{m_2 m_0} = \frac{1}{\pi^2}\nu^2, \tag{3.7}$$

offers a standard bandwidth measure (see, for example, Rice 1944*a*, *b*, 1945*a*, *b*; Longuet-Higgins 1957). For a narrow band signal $\nu = 0$, the expected numbers of extrema and zero crossings have to equal.

The second definition is a more general one; it is again based on the moments of the spectrum, but in a different way. Let us take a complex valued function in polar coordinates as

$$z(t) = a(t)\mathrm{e}^{\mathrm{i}\theta(t)}, \tag{3.8}$$

with both $a(t)$ and $\theta(t)$ being functions of time. If this function has a spectrum, $S(\omega)$, then the mean frequency is given by

$$\langle\omega\rangle = \int \omega|S(\omega)|^2\,\mathrm{d}\omega, \tag{3.9}$$

which can be expressed in another way as

$$\begin{aligned}\langle\omega\rangle &= \int z^*(t)\frac{1}{\mathrm{i}}\frac{\mathrm{d}}{\mathrm{d}t}z(t)\,\mathrm{d}t \\ &= \int\left(\dot{\theta}(t)-\mathrm{i}\frac{\dot{a}(t)}{a(t)}\right)a^2(t)\,\mathrm{d}t \\ &= \int\dot{\theta}(t)a^2(t)\,\mathrm{d}t. \end{aligned} \tag{3.10}$$

Based on this expression, Cohen (1995) suggested that $\dot{\theta}$ be treated as the instantaneous frequency. With these notations, the bandwidth can be defined as

$$\begin{aligned}\nu^2 &= \frac{(\omega-\langle\omega\rangle)^2}{\langle\omega\rangle^2} = \frac{1}{\langle\omega\rangle^2}\int(\omega-\langle\omega\rangle)^2|S(\omega)|^2\,\mathrm{d}\omega \\ &= \frac{1}{\langle\omega\rangle^2}\int z^*(t)\left(\frac{1}{\mathrm{i}}\frac{\mathrm{d}}{\mathrm{d}t}-\langle\omega\rangle\right)^2 z(t)\,\mathrm{d}t \\ &= \frac{1}{\langle\omega\rangle^2}\left[\int\dot{a}^2(t)\,\mathrm{d}t+\int(\dot{\theta}(t)-\langle\omega\rangle)^2a^2(t)\,\mathrm{d}t\right]. \end{aligned} \tag{3.11}$$

For a narrow band signal, this value has to be small, then both a and θ have to be gradually varying functions. Unfortunately, both equations (3.7) and (3.11) defined the bandwidth in the global sense; they are both overly restrictive and lack precision at the same time. Consequently, the bandwidth limitation on the Hilbert transform to give a meaningful instantaneous frequency has never been firmly established. For example, Melville (1983) had faithfully filtered the data within the bandwidth requirement, but he still obtained many non-physical negative frequency values. It should be mentioned here that using filtering to obtain a narrow band signal is unsatisfactory for another reason: the filtered data have already been contaminated by the spurious harmonics caused by the nonlinearity and non-stationarity as discussed in the introduction.

In order to obtain meaningful instantaneous frequency, restrictive conditions have to be imposed on the data as discussed by Gabor (1946), Bedrosian (1963) and, more recently, Boashash (1992): for any function to have a meaningful instantaneous frequency, the real part of its Fourier transform has to have only positive frequency. This restriction can be proven mathematically as shown in Titchmarsh (1948) but it is still global. For data analysis, we have to translate this requirement into physically implementable steps to develop a simple method for applications. For this purpose, we have to modify the restriction condition from a global one to a local one, and the basis has to satisfy the necessary conditions listed in the last section.

Let us consider some simple examples to illustrate these restrictions physically, by examining the function,

$$x(t)=\sin t. \tag{3.12}$$

Its Hilbert transform is simply $\cos t$. The phase plot of x–y is a simple circle of unit radius as in figure 1*a*. The phase function is a straight line as shown in figure 1*b* and the instantaneous frequency, shown in figure 1*c*, is a constant as expected. If we move the mean off by an amount α, say, then,

$$x(t)=\alpha+\sin t. \tag{3.13}$$

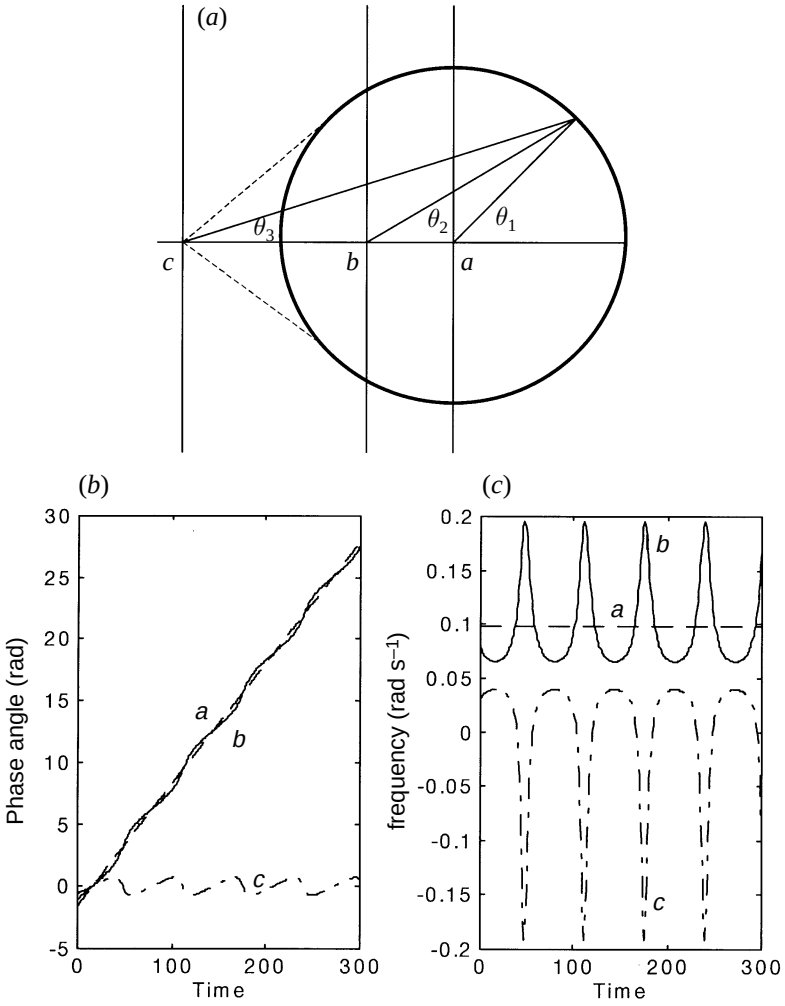

Figure 1. Physical interpretation of instantaneous frequency. (a) The phase plane for the model functions of $x(t) = \alpha + \sin t$. (a) $\alpha = 0$; (b) $\alpha < 1$; (c) $\alpha > 1$. (b) The unwrapped phase function of the model functions. (c) The instantaneous frequency computed according to equation (3.4).

The phase plot of x–y is still a simple circle independent of the value of α, but the centre of the circle will be displaced by the amount of α as illustrated in figure 1a. If $\alpha < 1$, the centre is still within the circle. Under this condition, the function has already violated a restriction, for its Fourier spectrum has a DC term; nevertheless, the mean zero-crossing frequency is still the same as in the case for $\alpha = 0$, but the phase function and the instantaneous frequency will be very different as shown in figures 1b, c. If $\alpha > 1$, the centre is outside the circle; thus, the function no longer satisfies the required conditions. Then both the phase function and the instantaneous frequency will assume negative values as shown in figures 1b, c, which are meaningless. These simple examples illustrate physically that, for a simple signal such as a sine function, the instantaneous frequency can be defined only if we restrict the function to be symmetric locally with respect to the zero mean level.

For general data, any riding waves would be equivalent to the case of $\alpha > 1$ locally;

any asymmetric wave form will be equivalent to the case of $\alpha < 1$, but non-zero, locally. In order to have a meaningful instantaneous frequency, this local restriction should be used in lieu of the global requirements given previously. Furthermore, this local restriction also suggests a method to decompose the data into components for which the instantaneous frequency can be defined. The examples presented above, however, actually lead us to the definition of a class of functions, based on its local properties, designated as intrinsic mode function (IMF) for which the instantaneous frequency can be defined everywhere. The limitation of interest here is not on the existence of the Hilbert transform which is general and global, but on the existence of a meaningful instantaneous frequency which is restrictive and local.

4. Intrinsic mode functions

The simple examples given above provide more physical interpretation of the restrictive conditions; they also suggest a practical way to decompose the data so that the components all satisfy the conditions imposed on them. Physically, the necessary conditions for us to define a meaningful instantaneous frequency are that the functions are symmetric with respect to the *local zero mean*, and have the same numbers of zero crossings and extrema. Based on these observations, we propose a class of functions designated as intrinsic mode functions here with the following formal definition.

An intrinsic mode function (IMF) is a function that satisfies two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The first condition is obvious; it is similar to the traditional narrow band requirements for a stationary Gaussian process. The second condition is a new idea; it modifies the classical global requirement to a local one; it is necessary so that the instantaneous frequency will not have the unwanted fluctuations induced by asymmetric wave forms. Ideally, the requirement should be 'the local mean of the data being zero'. For non-stationary data, the 'local mean' involves a 'local time scale' to compute the mean, which is impossible to define. As a surrogate, we use the local mean of the envelopes defined by the local maxima and the local minima to force the local symmetry instead. This is a necessary approximation to avoid the definition of a local averaging time scale. Although it will introduce an alias in the instantaneous frequency for nonlinearly deformed waves, the effects of nonlinearity are much weaker in comparison with non-stationarity as we will discuss later. With the physical approach and the approximation adopted here, the method does not always guarantee a perfect instantaneous frequency under all conditions. Nevertheless, we will show that, even under the worst conditions, the instantaneous frequency so defined is still consistent with the physics of the system studied.

The name 'intrinsic mode function' is adopted because it represents the oscillation mode imbedded in the data. With this definition, the IMF in each cycle, defined by the zero crossings, involves only one mode of oscillation, no complex riding waves are allowed. With this definition, an IMF is not restricted to a narrow band signal, and it can be both amplitude and frequency modulated. In fact, it can be non-stationary. As discussed above, purely frequency or amplitude modulated functions can be IMFs

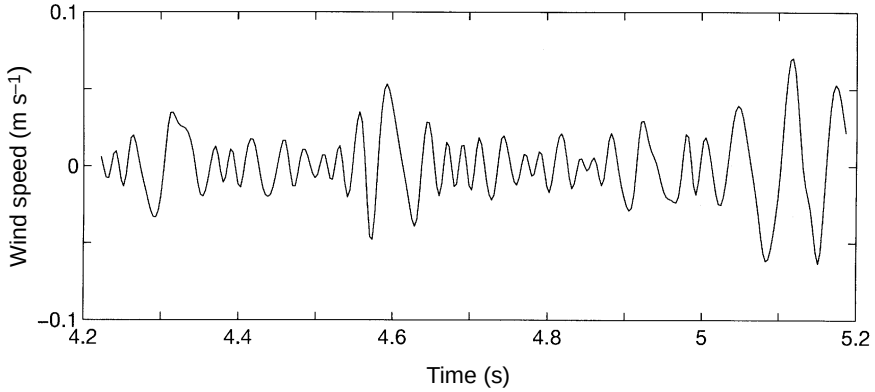

Figure 2. A typical intrinsic mode function with the same numbers of zero crossings and extrema, and symmetry of the upper and lower envelopes with respect to zero.

even though they have finite bandwidth according to the traditional definition. A typical IMF is shown in figure 2.

Having defined IMF, we will show that the definition given in equation (3.4) gives the best instantaneous frequency. An IMF after the Hilbert transform can be expressed as in equation (3.2). If we perform a Fourier transform on $Z(t)$, we have

$$W(\omega) = \int_{-\infty}^{\infty} a(t)\mathrm{e}^{\mathrm{i}\theta(t)}\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t, = \int_{-\infty}^{\infty} a(t)\mathrm{e}^{\mathrm{i}(\theta(t)-\omega t)}\,\mathrm{d}t. \tag{4.1}$$

Then by the stationary phase method (see, for example, Copson 1967), the maximum contribution to $W(\omega)$ is given by the frequency satisfying the condition

$$\frac{\mathrm{d}}{\mathrm{d}t}(\theta(t) - \omega t) = 0; \tag{4.2}$$

therefore, equation (3.4) follows. Although mathematically, the application of the stationary phase method requires a large parameter for the exponential function, the adoption here can be justified if the frequency, ω, is high compared with the inversed local time scale of the amplitude variation. Therefore, this definition fits the best for gradually changing amplitude. Even with this condition, this is still a much better definition for instantaneous frequency than the zero-crossing frequency; it is also better than the integral definition suggested by Cohen (1995) as given in equation (3.10). Furthermore, it agrees with the definition of frequency for the classic wave theory (see, for example, Whitham 1975).

As given in equation (4.2) and the simple analogy given in equations (3.2)–(3.4), the frequency defined through the stationary phase approximation agrees also with the best fit sinusoidal function locally; therefore, we do not need a whole oscillatory period to define a frequency value. We can define it for every point with the value changing from point to point. In this sense, even a monotonic function can be treated as part of an oscillatory function and have instantaneous frequency assigned according to equation (3.4). Any frequency variation is designated as frequency modulation. There are actually two types of frequency modulations: the interwave and the intrawave modulations. The first type is familiar to us; the frequency of the oscillation is gradually changing with the waves in a dispersive system. Technically, in the dispersive waves, the frequency is also changing within one wave, but that

was not emphasized either for convenience, or for lack of a more precise frequency definition. The second type is less familiar, but it is also a common phenomenon: if the frequency changes from time to time within a wave its profile can no longer be a simple sine or cosine function. Therefore, any wave-profile deformation from the simple sinusoidal form implies the intrawave frequency modulation. In the past such phenomena were treated as harmonic distortions. We will show in detail later that most such deformations are better viewed as intrawave frequency modulation, for the intrawave frequency modulation is more physical.

In order to use this unique definition of instantaneous frequency, we have to reduce an arbitrary data set into IMF components from which an instantaneous frequency value can be assigned to each IMF component. Consequently, for complicated data, we can have more than one instantaneous frequency at a time locally. We will introduce the empirical mode decomposition method to reduce the data into the needed IMFs.

5. The empirical mode decomposition method: the sifting process

Knowing the well-behaved Hilbert transforms of the IMF components is only the starting point. Unfortunately, most of the data are not IMFs. At any given time, the data may involve more than one oscillatory mode; that is why the simple Hilbert transform cannot provide the full description of the frequency content for the general data as reported by Long *et al.* (1995). We have to decompose the data into IMF components. Here, we will introduce a new method to deal with both non-stationary and nonlinear data by decomposing the signal first, and discuss the physical meaning of this decomposition later. Contrary to almost all the previous methods, this new method is intuitive, direct, *a posteriori* and adaptive, with the basis of the decomposition based on, and derived from, the data.

The decomposition is based on the assumptions: (1) the signal has at least two extrema—one maximum and one minimum; (2) the characteristic time scale is defined by the time lapse between the extrema; and (3) if the data were totally devoid of extrema but contained only inflection points, then it can be differentiated once or more times to reveal the extrema. Final results can be obtained by integration(s) of the components.

The essence of the method is to identify the intrinsic oscillatory modes by their characteristic time scales in the data empirically, and then decompose the data accordingly. According to Drazin (1992), the first step of data analysis is to examine the data by eye. From this examination, one can immediately identify the different scales directly in two ways: by the time lapse between the successive alternations of local maxima and minima; and by the time lapse between the successive zero crossings. The interlaced local extrema and zero crossings give us the complicated data: one undulation is riding on top of another, and they, in turn, are riding on still other undulations, and so on. Each of these undulations defines a characteristic scale of the data; it is intrinsic to the process. We have decided to adopt the time lapse between successive extrema as the definition of the time scale for the intrinsic oscillatory mode, because it not only gives a much finer resolution of the oscillatory modes, but also can be applied to data with non-zero mean, either all positive or all negative values, without zero crossings. A systematic way to extract them, designated as the sifting process, is described as follows.

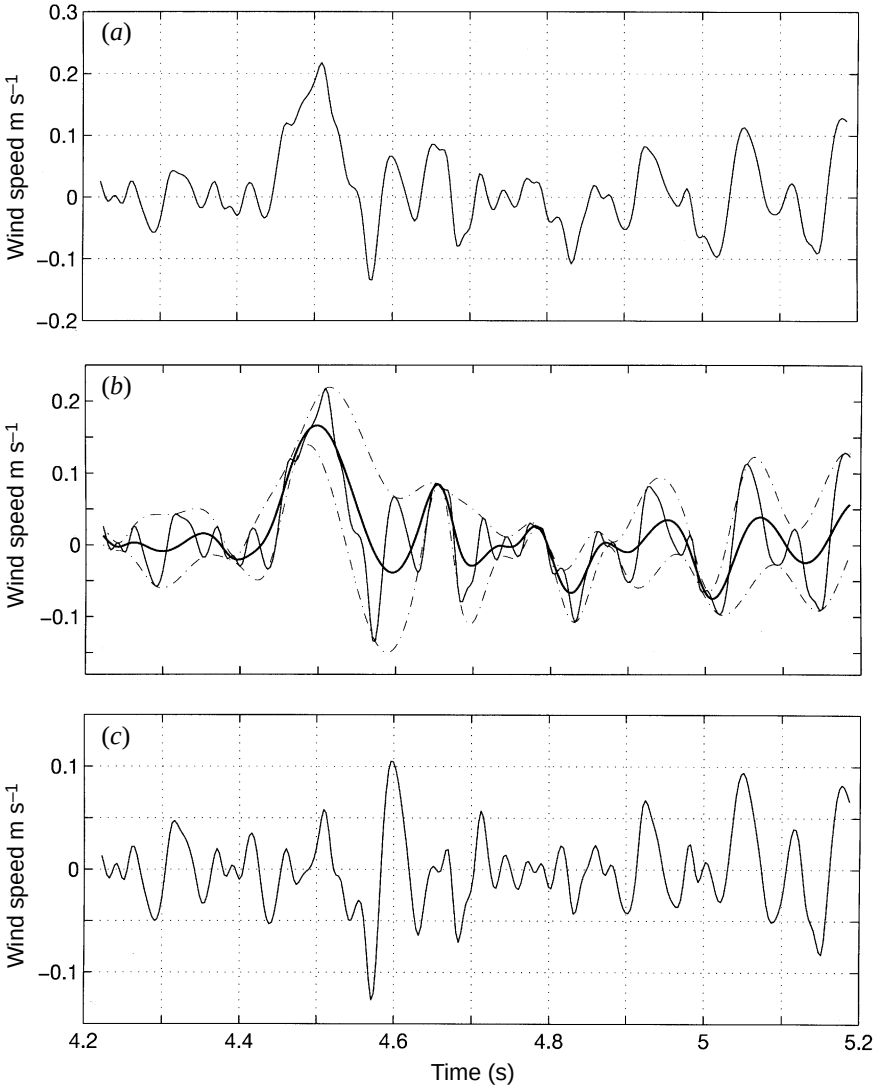

Figure 3. Illustration of the sifting processes: (*a*) the original data; (*b*) the data in thin solid line, with the upper and lower envelopes in dot-dashed lines and the mean in thick solid line; (*c*) the difference between the data and m_1. This is still not an IMF, for there are negative local maxima and positive minima suggesting riding waves.

By virtue of the IMF definition, the decomposition method can simply use the envelopes defined by the local maxima and minima separately. Once the extrema are identified, all the local maxima are connected by a cubic spline line as the upper envelope. Repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them. Their mean is designated as m_1, and the difference between the data and m_1 is the first component, h_1, i.e.

$$X(t) - m_1 = h_1. \tag{5.1}$$

The procedure is illustrated in figures 3*a–c* (figure 3*a* gives the data; figure 3*b* gives

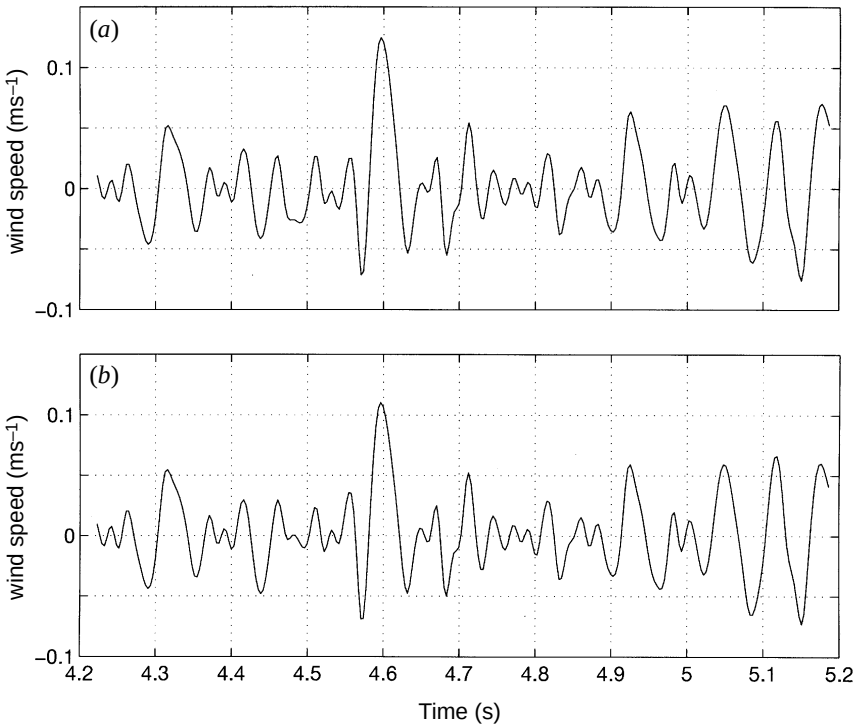

Figure 4. Illustration of the effects of repeated sifting process: (*a*) after one more sifting of the result in figure 3*c*, the result is still asymmetric and still not a IMF; (*b*) after three siftings, the result is much improved, but more sifting needed to eliminate the asymmetry. The final IMF is shown in figure 2 after nine siftings.

the data in the thin solid line, the upper and the lower envelopes in the dot-dashed lines, and their mean in the thick solid line, which bisects the data very well; and figure 3*c* gives the difference between the data and the local mean as in equation (5.1)).

Ideally, h_1 should be an IMF, for the construction of h_1 described above seems to have been made to satisfy all the requirements of IMF. In reality, however, overshoots and undershoots are common, which can also generate new extrema, and shift or exaggerate the existing ones. The imperfection of the overshoots and undershoots can be found at the 4.6 and 4.7 s points in figure 3*b*. Their effects, however, are not direct, for it is the mean, not the envelopes, that will enter the sifting process. Nevertheless, the problem is real. Even if the fitting is perfect, a gentle hump on a slope can be amplified to become a local extremum in changing the local zero from a rectangular to a curvilinear coordinate system. An example can be found for the hump between the 4.5 and 4.6 s range in the data in figure 3*a*. After the first round of sifting, the hump becomes a local maximum at the same time location as in figure 3*c*. New extrema generated in this way actually recover the proper modes lost in the initial examination. In fact, the sifting process can recover low-amplitude riding waves with repeated siftings.

Still another complication is that the envelope mean may be different from the true local mean for nonlinear data; consequently, some asymmetric wave forms can still exist no matter how many times the data are sifted. We have to accept this approximation as discussed before.

Other than these theoretical difficulties, on the practical side, serious problems of the spline fitting can occur near the ends, where the cubic spline fitting can have large swings. Left by themselves, the end swings can eventually propagate inward and corrupt the whole data span especially in the low-frequency components. We have devised a numerical method to eliminate the end effects; details will be given later. At any rate, improving the spline fitting is absolutely necessary. Even with these problems, the sifting process can still extract the essential scales from the data.

The sifting process serves two purposes: to eliminate riding waves; and to make the wave-profiles more symmetric. Toward this end, the sifting process has to be repeated more times. In the second sifting process, h_1 is treated as the data, then

$$h_1 - m_{11} = h_{11}. \tag{5.2}$$

Figure 4*a* shows the much improved result after the second sifting, but there are still local maxima below the zero line. After another sifting, the result is given in figure 4*b*. Now all the local maxima are positive, and all the local minima are negative, but many waves are still asymmetric. We can repeat this sifting procedure k times, until h_{1k} is an IMF, that is

$$h_{1(k-1)} - m_{1k} = h_{1k}, \tag{5.3}$$

the result is shown in figure 2 after nine siftings. Then, it is designated as

$$c_1 = h_{1k}, \tag{5.4}$$

the first IMF component from the data.

As described above, the process is indeed like sifting: to separate the finest local mode from the data first based only on the characteristic time scale. The sifting process, however, has two effects: (*a*) to eliminate riding waves; and (*b*) to smooth uneven amplitudes.

While the first condition is absolutely necessary for the instantaneous frequency to be meaningful, the second condition is also necessary in case the neighbouring wave amplitudes have too large a disparity. Unfortunately, the second effect, when carried to the extreme, could obliterate the physically meaningful amplitude fluctuations. Therefore, the sifting process should be applied with care, for carrying the process to an extreme could make the resulting IMF a pure frequency modulated signal of constant amplitude. To guarantee that the IMF components retain enough physical sense of both amplitude and frequency modulations, we have to determine a criterion for the sifting process to stop. This can be accomplished by limiting the size of the standard deviation, SD, computed from the two consecutive sifting results as

$$\mathrm{SD} = \sum_{t=0}^{T} \left[\frac{|(h_{1(k-1)}(t) - h_{1k}(t))|^2}{h_{1(k-1)}^2(t)} \right]. \tag{5.5}$$

A typical value for SD can be set between 0.2 and 0.3. As a comparison, the two Fourier spectra, computed by shifting only five out of 1024 points from the same data, can have an equivalent SD of 0.2–0.3 calculated point-by-point. Therefore, a SD value of 0.2–0.3 for the sifting procedure is a very rigorous limitation for the difference between siftings.

Overall, c_1 should contain the finest scale or the shortest period component of the signal. We can separate c_1 from the rest of the data by

$$X(t) - c_1 = r_1. \tag{5.6}$$

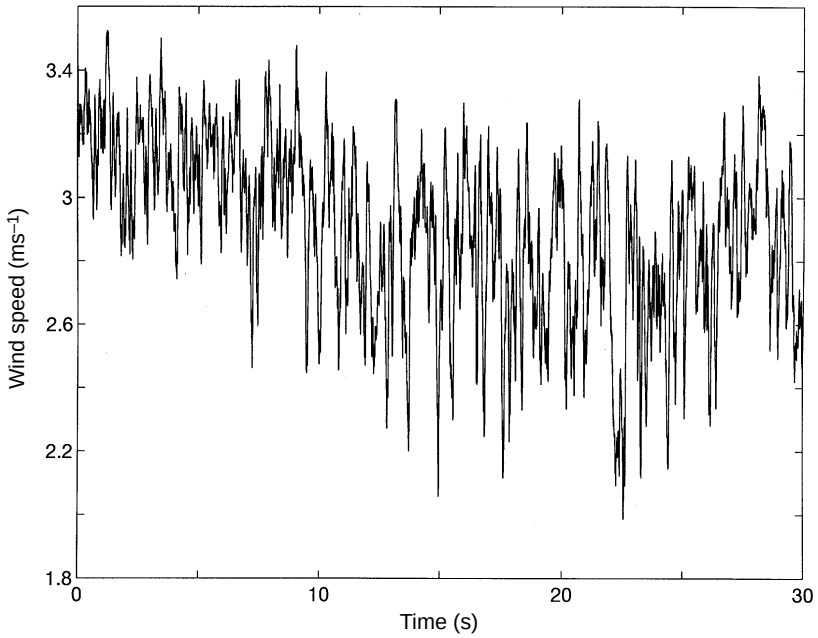

Figure 5. Calibrated wind data from a wind-wave tunnel.

Since the residue, r_1, still contains information of longer period components, it is treated as the new data and subjected to the same sifting process as described above. This procedure can be repeated on all the subsequent r_js, and the result is

$$r_1 - c_2 = r_2, \ldots, r_{n-1} - c_n = r_n. \tag{5.7}$$

The sifting process can be stopped by any of the following predetermined criteria: either when the component, c_n, or the residue, r_n, becomes so small that it is less than the predetermined value of substantial consequence, or when the residue, r_n, becomes a monotonic function from which no more IMF can be extracted. Even for data with zero mean, the final residue can still be different from zero; for data with a trend, then the final residue should be that trend. By summing up equations (5.6) and (5.7), we finally obtain

$$X(t) = \sum_{i=1}^{n} c_i + r_n. \tag{5.8}$$

Thus, we achieved a decomposition of the data into n-empirical modes, and a residue, r_n, which can be either the mean trend or a constant. As discussed here, to apply the EMD method, a mean or zero reference is not required; EMD only needs the locations of the local extrema. The zero references for each component will be generated by the sifting process. Without the need of the zero reference, EMD eliminates the troublesome step of removing the mean values for the large DC term in data with non-zero mean, an unexpected benefit.

To illustrate the sifting process, we will use a set of wind data collected in a laboratory wind-wave tunnel (Huang & Long 1980) with a high-frequency response Pitot tube located 10 cm above the mean water level. The wind speed was recorded under the condition of the initial onset of water waves from a calm surface. Calibrated wind data are given in figure 5. Clearly, the data are quite complicated with many

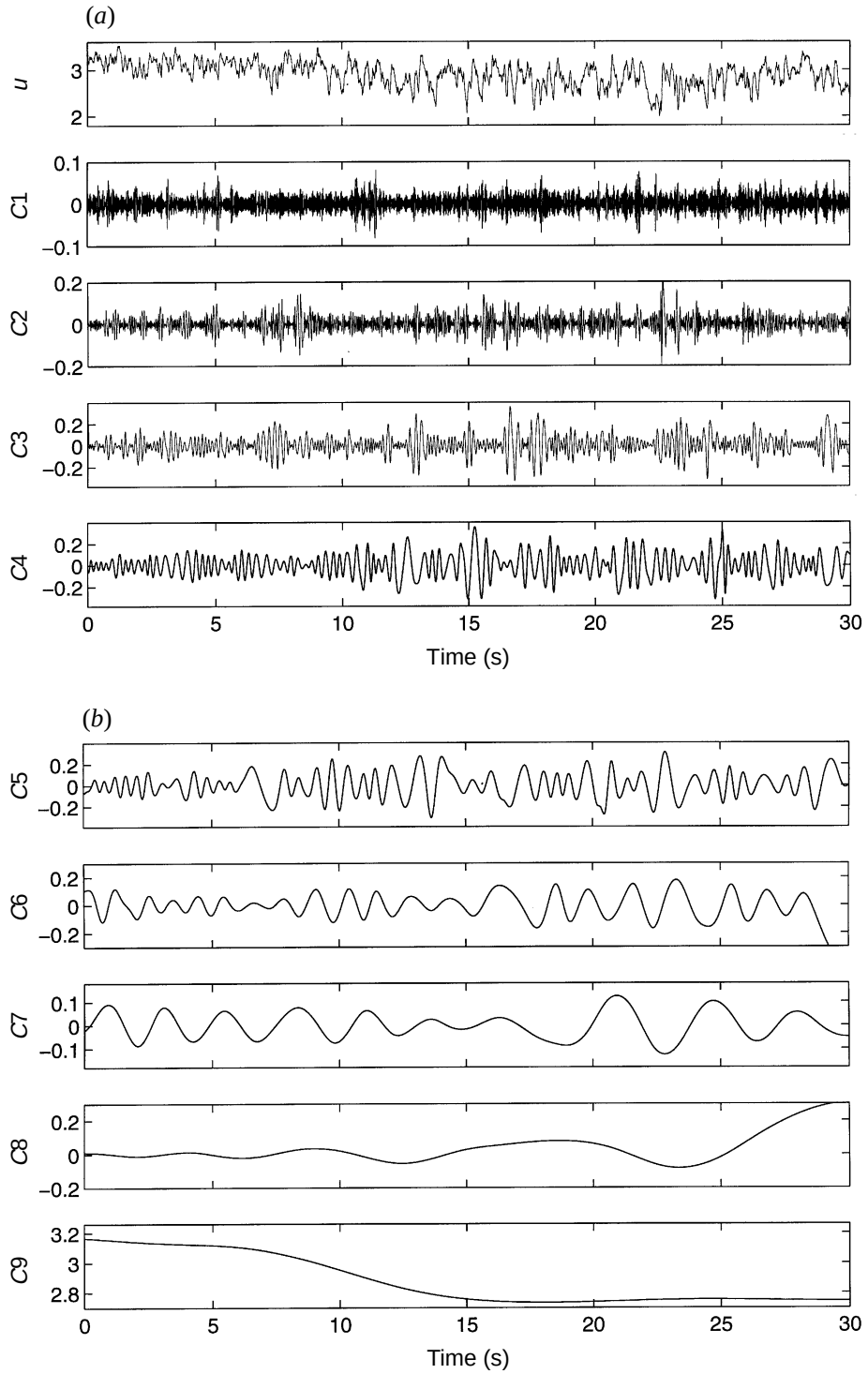

Figure 6. The resulting empirical mode decomposition components from the wind data: (*a*) the original data and the components c_1–c_4; (*b*) the components c_5–c_9. Notice the last component, c_9, is not an IMF; it is the trend.

local extrema but no zero crossings, for the time series represents all positive numbers. Although the mean can be treated as a zero reference, defining it is hard, for the whole process is transient. This example illustrates the advantage of adopting the successive extrema for defining the time scale; it also illustrates the difficulties of dealing with non-stationary data: even a meaningful mean is impossible to define, but for EMD this difficulty is eliminated. Figures 6*a*, *b* summarize all the IMF obtained from this repeated sifting processes. We have a total of nine components. Comparing this with the traditional Fourier expansion, one can immediately see the efficiency of the EMD: the expansion of a turbulence data set with only nine terms. From the result, one can see a general separation of the data into locally non-overlapping time scale components. In some components, such as c_1 and c_3, the signals are intermittent, then the neighbouring components might contain oscillations of the same scale, but signals of the same time scale would never occur at the same locations in two different IMF components.

Finally, let us examine the physical meaning of each IMF component. The components of the EMD are usually physical, for the characteristic scales are physical. Nevertheless, this is not strictly true, for there are cases when a certain scale of a phenomenon is intermittent. Then, the decomposed component could contain two scales in one IMF component. Therefore, the physical meaning of the decomposition comes only in the totality of the decomposed components in the Hilbert spectrum. Even with the entire set of decomposed components, sound physical interpretation is still not guaranteed for other decompositions such as Fourier expansion. Further discussions will be given later in detail. Having established the decomposition, we should check the completeness and orthogonality of this decomposition. Because this decomposition is *a posteriori*, the check should also be *a posteriori*.

6. Completeness and orthogonality

By virtue of the decomposition, completeness is given, for equation (5.8) is an identity. As a check of the completeness for the wind data numerically, we can reconstruct the data from the IMF components starting from the longest to the shortest periods in the sequence from figure 7*a–j*. Figure 7*a* gives the data and the longest period component, c_9, which is the residue trend, not an IMF. By itself, the fitting of the trend is quite impressive, and it is very physical: the gradual decrease of the mean wind speed indicates the lack of drag from the calm surface initially and the increase of drag after the generation of wind waves. As the mean wind speed decreases, the amplitude of the fluctuation increases, another indication of wind–wave interactions. If we add the next longest period component, c_8, the trend of the sum, $c_9 + c_8$, takes a remarkable turn, and the fitting to the data looks greatly improved as shown in figure 7*b*. By successively adding more components with increasing frequency, we have the results in the series of figures 7*c–i*. The gradual change from the monotonic trend to the final reconstruction is illustrative by itself. By the time we reach the sum of IMF components up to c_3 in figure 7*g*, we essentially have recovered all the energy containing eddies already. The components with the highest frequencies add little more energy, but they make the data look more complicated. In fact, the highest frequency component is probably not physical, for the digitizing rate of the Pitot tube is too slow to capture the high-frequency variations. As a result, the data are jagged artificially by the digitalizing steps at this frequency.

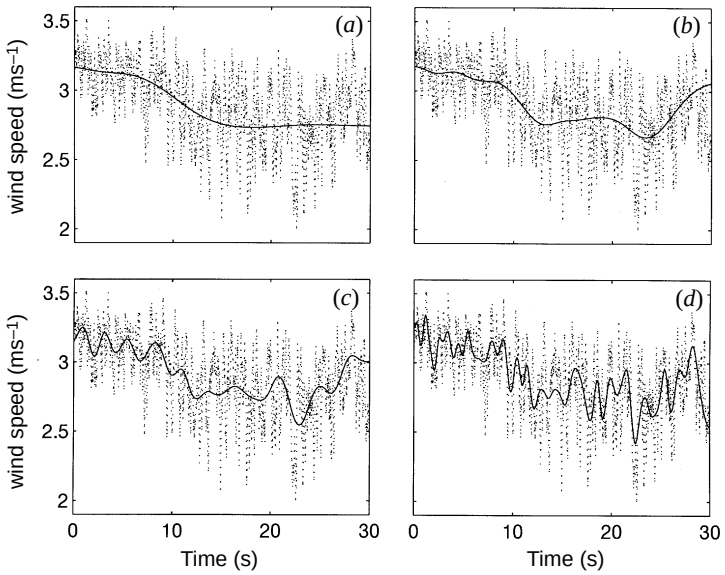

Figure 7. Numerical proof of the completeness of the EMD through reconstruction of the original data from the IMF components. (*a*) Data (in the dotted line) and the c_9 component (in the solid line). c_9 serves as a running mean, but it is not obtained from either mean or filtering. (*b*) Data (in the dotted line) and the sum of c_9–c_8 components (in the solid line). (*c*) Data (in the dotted line) and the sum of c_9–c_7 components (in the solid line). (*d*) Data (in the dotted line) and the sum of c_9–c_6 components (in the solid line).

The difference between the reconstructed data from the sum of all the IMFs and the original data is shown in figure 7*j*, in which the maximum amplitude is less than 5×10^{-15}, the roundoff error from the precision of the computer. Thus, the completeness is established both theoretically by equation (5.8), and numerically by figure 7*j*.

The orthogonality is satisfied in all practical sense, but it is not guaranteed theoretically. Let us discuss the practical aspect first. By virtue of the decomposition, the elements should all be locally orthogonal to each other, for each element is obtained from the difference between the signal and its local mean through the maximal and minimal envelopes; therefore,

$$\overline{(x(t) - \overline{x(t)}) \cdot \overline{x(t)}} = 0. \tag{6.1}$$

Nevertheless, equation (6.1) is not strictly true, because the mean is computed via the envelopes, hence it is not the true mean. Furthermore, each successive IMF component is only part of the signal constituting $\overline{x(t)}$. Because of these approximations, leakage is unavoidable. Any leakage, however, should be small.

The orthogonality of the EMD components should also be checked *a posteriori* numerically as follows: let us first write equation (5.8) as

$$X(t) = \sum_{j=1}^{n+1} C_j(t), \tag{6.2}$$

in which we have included r_n as an additional element. To check the orthogonality of the IMFs from EMD, we define an index based on the most intuitive way: first,

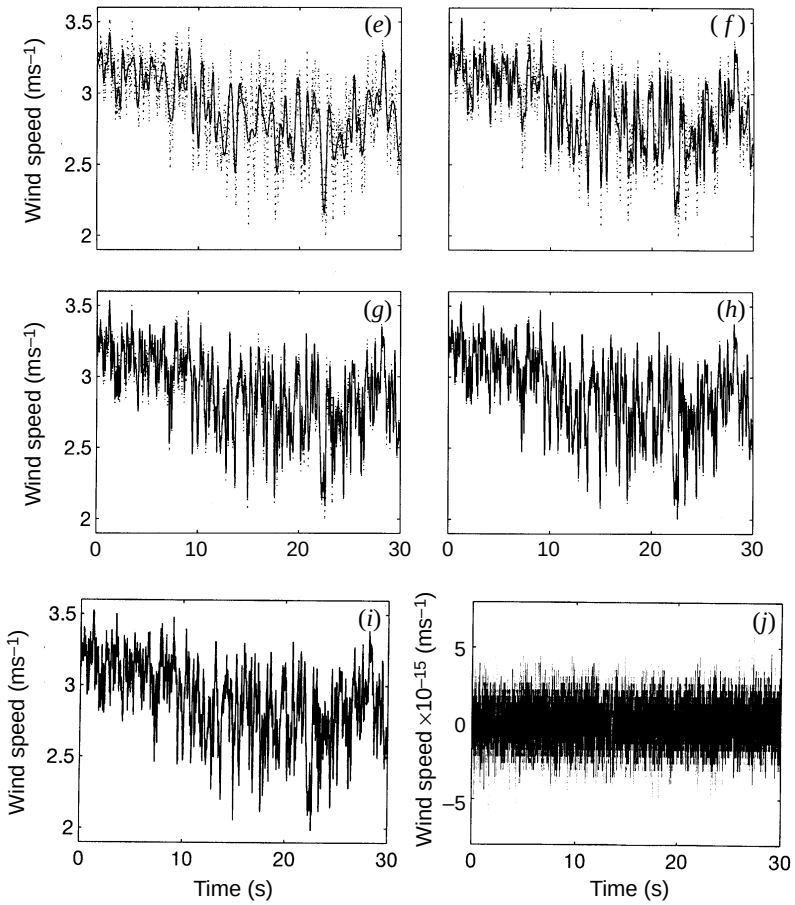

Figure 7. *Cont.* (*e*) Data (in the dotted line) and the sum of c_9–c_5 components (in the solid line). (*f*) Data (in the dotted line) and the sum of c_9–c_4 components (in the solid line). (*g*) Data (in the dotted line) and the sum of c_9–c_3 components (in the solid line). By now, we seem to have recovered all the energy containing eddies. (*h*) Data (in the dotted line) and the sum of c_9–c_2 components (in the solid line). (*i*) Data (in the dotted line) and the sum of c_9–c_1 components (in the solid line). This is the final reconstruction of the data from the IMFs. It appears no different from the original data. (*j*) The difference between the original data and the reconstructed one; the difference is the limit of the computational precision of the personal computer (PC) used.

let us form the square of the signal as

$$X^2(t) = \sum_{j=1}^{n+1} C_j^2(t) + 2\sum_{j=1}^{n+1}\sum_{k=1}^{n+1} C_j(t)C_k(t). \tag{6.3}$$

If the decomposition is orthogonal, then the cross terms given in the second part on the right-hand side should be zero. With this expression, an overall index of orthogonality, IO, is defined as

$$\mathrm{IO} = \sum_{t=0}^{T}\left(\sum_{j=1}^{n+1}\sum_{k=1}^{n+1} C_j(t)C_k(t)/X^2(t)\right). \tag{6.4}$$

For the wind data given above, the IO value is only 0.0067. Orthogonality can also

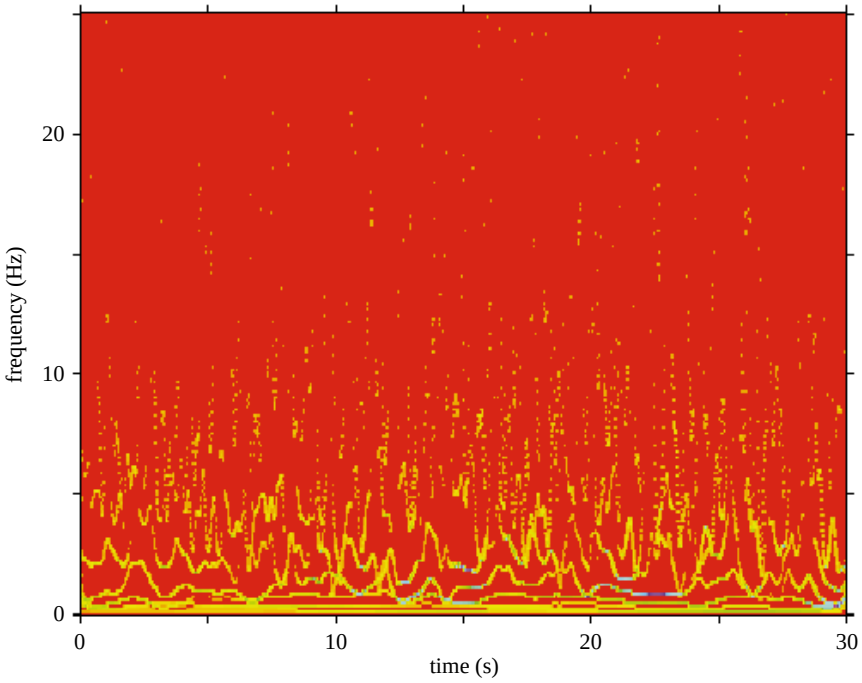

Figure 8. The Hilbert spectrum for the wind data with 200 frequency cells. The wind energy appears in skeleton lines representing each IMF.

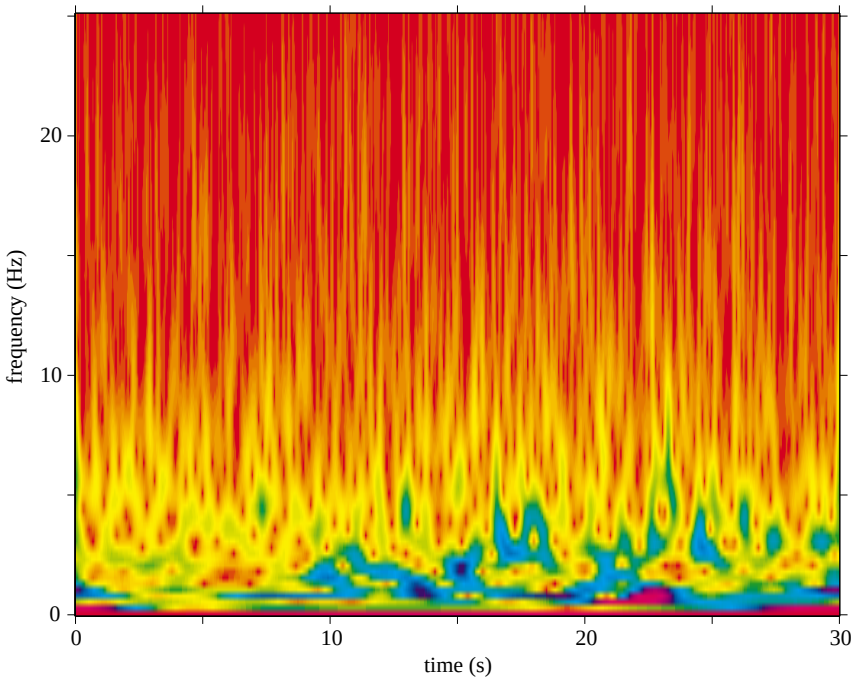

Figure 9. The Morlet wavelet spectrum for the wind data with the same number of frequency cells. Wind energy appears in smoothed contours with a rich energy distribution in the high harmonics.

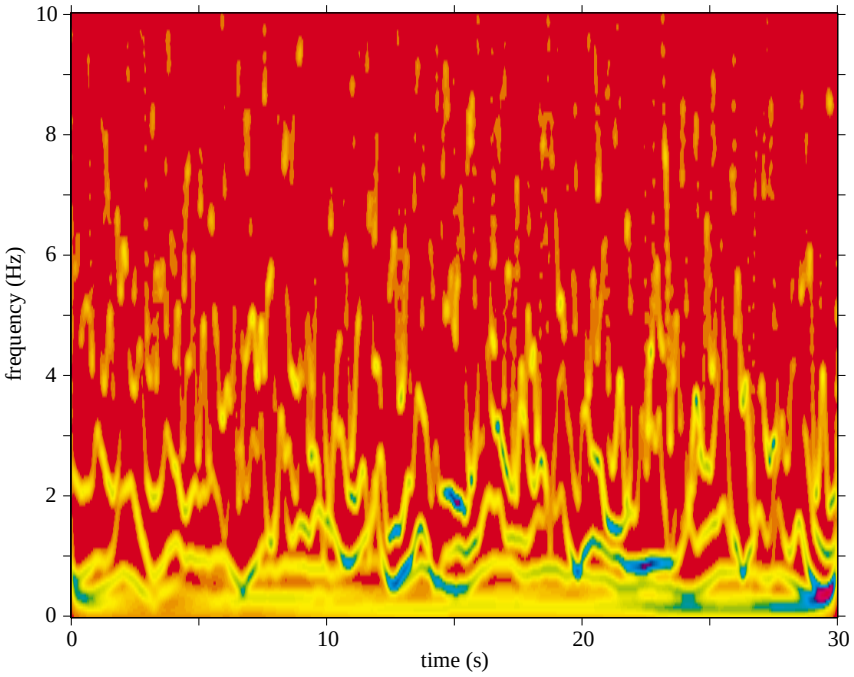

Figure 10. A 15×15 weighted Gaussian filtered Hilbert spectrum. The appearance gives a better comparison with the wavelet result, but the information content has been degraded.

be defined for any two components, C_f and C_g. The measure of orthogonality will, then, be

$$\mathrm{IO}_{fg} = \sum_t \frac{C_f C_g}{C_f^2 + C_g^2}. \tag{6.5}$$

It should be noted that, although the definition of orthogonality seems to be global, the real meaning here applies only locally. For some special data, the neighbouring components could certainly have sections of data carrying the same frequency at different time durations. But locally, any two components should be orthogonal for all practical purposes. The amount of leakage usually depends on the length of data as well as the decomposition results. In reality, because of the finite data length, even pure sinusoidal components with different frequencies are not exactly orthogonal. That is why the continuous wavelet in the most commonly used Morlet form suffers severe leakage. For the EMD method, we found the leakage to be typically less than 1%. For extremely short data, the leakage could be as high as 5%, which is comparable to that from a set of pure sinusoidal waves of the same data length.

Theoretically, the EMD method guarantees orthogonality only on the strength of equation (6.1). Orthogonality actually depends on the decomposition method. Let us consider two Stokian waves each having many harmonics. Separately, they are IMFs, of course, and each represents a free travelling wave. If the frequency of one Stokian wave coincides with the frequency of a harmonic of the other, then the two waves are no longer orthogonal in the Fourier sense. EMD, however, can still separate the two Stokian wave trains. Such a decomposition is physical, but the separated IMF components are not orthogonal. Therefore, orthogonality is a requirement only

for linear decomposition systems; it would not make physical sense for a nonlinear decomposition as in EMD. Fortunately, in most cases encountered, the leakage is small.

7. The Hilbert spectrum

Having obtained the intrinsic mode function components, we will have no difficulties in applying the Hilbert transform to each component, and computing the instantaneous frequency according to equation (3.4). After performing the Hilbert transform on each IMF component, we can express the data in the following form:

$$X(t) = \sum_{j=1}^{n} a_j(t) \exp\left(\mathrm{i} \int \omega_j(t)\, \mathrm{d}t\right). \tag{7.1}$$

Here we have left out the residue, r_n, on purpose, for it is either a monotonic function, or a constant. Although the Hilbert transform can treat the monotonic trend as part of a longer oscillation, the energy involved in the residual trend could be overpowering. In consideration of the uncertainty of the longer trend, and in the interest of information contained in the other low-energy and higher-frequency components, the final non-IMF component should be left out. It, however, could be included, if physical considerations justify its inclusion.

Equation (7.1) gives both the amplitude and the frequency of each component as functions of time. The same data if expanded in Fourier representation would be

$$X(t) = \sum_{j=1}^{\infty} a_j \mathrm{e}^{\mathrm{i}\omega_j t}, \tag{7.2}$$

with both a_j and ω_j constants. The contrast between equations (7.1) and (7.2) is clear: the IMF represents a generalized Fourier expansion. The variable amplitude and the instantaneous frequency have not only greatly improved the efficiency of the expansion, but also enabled the expansion to accommodate non-stationary data. With IMF expansion, the amplitude and the frequency modulations are also clearly separated. Thus, we have broken through the restriction of the constant amplitude and fixed-frequency Fourier expansion, and arrived at a variable amplitude and frequency representation. This expression is numerical. If a function is more desired, an empirical polynomial expression can be easily derived from the IMFs.

Equation (7.1) also enables us to represent the amplitude and the instantaneous frequency as functions of time in a three-dimensional plot, in which the amplitude can be contoured on the frequency–time plane. This frequency–time distribution of the amplitude is designated as the Hilbert amplitude spectrum, $H(\omega, t)$, or simply Hilbert spectrum. If amplitude squared is more desirable commonly to represent energy density, then the squared values of amplitude can be substituted to produce the Hilbert energy spectrum just as well.

Various forms of Hilbert spectra presentations can be made: the colour coded maps and the contour maps all with or without smoothing. The Hilbert spectrum in the colour map format for the wind data is given in figure 8. This spectrum gives a very different appearance in comparison with the corresponding wavelet spectrum shown in figure 9. While the Hilbert spectrum appears only in the skeleton form with emphasis on the frequency variations of each IMF, the Wavelet analysis result

gives a smoothed energy contour map with a rich distribution of higher harmonics. If a more continuous form is preferred, two smoothing methods can be applied. The first one is to use a weighted spatial filter with which to average over a range of cells. However, such a smoothing degrades both frequency and time resolutions. If we applied this approach with large enough spatial averaging, we would get a result similar to what the wavelet analysis would give. Even here, we still will not be encumbered by the many spurious harmonics as in the Fourier-based wavelet analysis. For example, a 15×15 weighted Gaussian filter will give the smoothed spectrum as in figure 10, which gives a better resemblance to the wavelet result, though not necessarily a better physical interpretation than the original, for the information content has been degraded. In the smoothed form, however, the energy density and its trends of evolution as functions of frequency and time are easier to identify. In general, if more quantitative results are desired, the original skeleton presentation is better; if more qualitative results are desired, the smoothed presentation is better. As a guide, the first look of the data is better in the smoothed format.

The alternative to the spatial smoothing is to select a lower frequency resolution and leave the time axis undisturbed. The advantage of this approach is the preservation of events' locations, but gives a more continuous frequency variation. Furthermore, with low-frequency resolution, we also save some computation time. The optimal frequency resolution in the Hilbert spectrum can be computed as follows.

Let the total data length be T, and the digitizing rate be Δt. Then the lowest frequency one can extract from the data is $1/T$ Hz, which is also the limit of frequency resolution for the data. The highest frequency one can extract from the data is $1/(n\Delta t)$ Hz, in which n represents the minimum number of Δt needed to define the frequency accurately. Since the Hilbert transform defines instantaneous frequency by differentiation, we do need more data points to define an oscillation. The absolute minimum number of data points is five for a whole sine wave. We do not need a whole sine wave to define its frequency, but we do need many points within any part of the wave to get a stable derivative. Therefore, the maximum number of the frequency cells, N, of the Hilbert spectrum should be

$$N = \frac{(1/n\Delta t)}{(1/T)} = \frac{T}{n\Delta t}. \tag{7.3}$$

In order to make the derivative stable, we have averaged over three adjacent cell values for the final presentation. The wind data, digitized at a rate of 0.01 s, has a total length of 30 s. Therefore, the highest frequency we can extract is 25 Hz. The total cell size could be 600 but they have been averaged to 200 as used in figure 8.

With the Hilbert spectrum defined, we can also define the marginal spectrum, $h(\omega)$, as

$$h(\omega) = \int_0^T H(\omega, t)\,\mathrm{d}t. \tag{7.4}$$

In figure 11 the solid line gives the corresponding marginal spectrum of the Hilbert spectrum given in figure 8. The marginal spectrum offers a measure of total amplitude (or energy) contribution from each frequency value. It represents the cumulated amplitude over the entire data span in a probabilistic sense. As pointed out by Huang *et al.* (1996), the frequency in either $H(\omega, t)$ or $h(\omega)$ has a totally different meaning from the Fourier spectral analysis. In the Fourier representation, the existence of energy at a frequency, ω, means a component of a sine or a cosine wave persisted

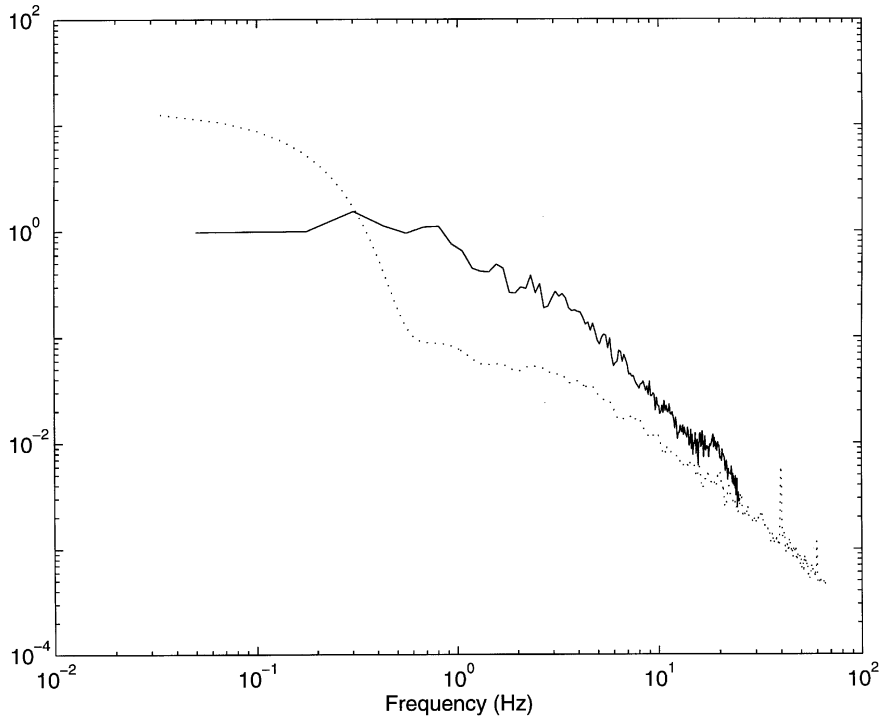

Figure 11. The comparison of the marginal Hilbert spectrum (the solid line) and the Fourier spectrum (the dotted line) all from the same wind data. The Fourier spectrum is dominated by the DC term due to the non-zero mean wind; it is also meaningless because of the non-stationarity of the data.

through the time span of the data. Here, the existence of energy at the frequency, ω, means only that, in the whole time span of the data, there is a higher likelihood for such a wave to have appeared locally. In fact, the Hilbert spectrum is a weighted non-normalized joint amplitude–frequency–time distribution. The weight assigned to each time–frequency cell is the local amplitude. Consequently, the frequency in the marginal spectrum indicates only the likelihood that an oscillation with such a frequency exists. The exact occurrence time of that oscillation is given in the full Hilbert spectrum. The corresponding Fourier spectrum of the wind data is also given in figure 11 in a dotted line. There is little similarity between the Fourier spectrum and the marginal spectrum. While the Fourier spectrum is dominated by the DC term because of the non-zero mean wind speed, the marginal spectrum gives a nearly continuous distribution of energy. The Fourier spectrum is meaningless physically, for the data is not stationary.

In addition to the marginal spectrum, we can also define the instantaneous energy density level, IE, as

$$\mathrm{IE}(t) = \int_{\omega} H^2(\omega, t)\,\mathrm{d}\omega. \tag{7.5}$$

Obviously, this IE also depends on time; it can be used to check the energy fluctuation.

With the energy–frequency–time distribution given, we can discuss the definition of stationarity quantitatively now. As we stated in the introduction, the classic defi-

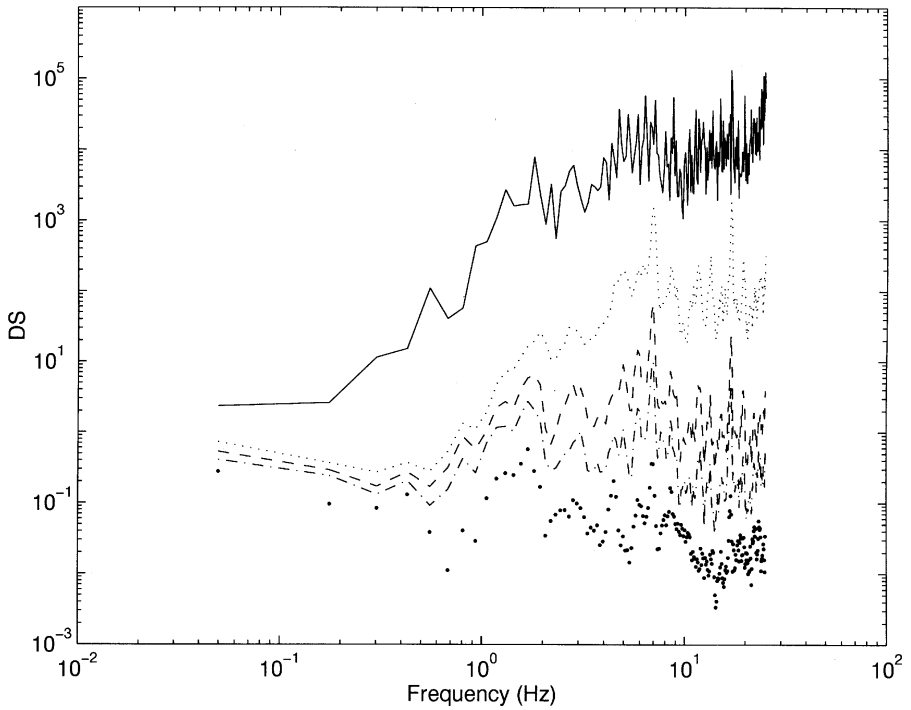

Figure 12. The variation of the degree of stationarity and the degree of statistical stationarity. In computing the DSS, time averages of 10, 50, 100 and 300 have been used. The DSS indeed decreases with the length of averaging especially in the high-frequency range.

nitions of stationarity given in equations (1.1) and (1.2) are dichotomous: a process is either stationary or non-stationary, the description is only qualitative. Such definitions are both stringent and useless at the same time: few data sets can satisfy the rigour of these definitions; consequently, no one even bothers using them for checking the stationarity. As a result, data as non-stationary as earthquake and seismological signals are routinely treated as stationary (see, for example, Hu *et al.* 1996). Sometimes, for some obvious non-stationary data, two less stringent definitions are invoked: the piecewise stationary; and the asymptotically stationary. These definitions are still dichotomous. To quantify the statistical processes further, an index is needed to give a quantitative measure of how far the process deviates from stationarity; a prerequisite for such a definition is a method to present the data in the frequency–time space. Now, having established the Hilbert spectrum, we can introduce the index of stationarity as follows.

The first step to define the degree of stationarity, $DS(\omega)$, is to find the mean marginal spectrum, $n(\omega)$, as

$$n(\omega) = \frac{1}{T} h(\omega). \tag{7.6}$$

Then, the degree of stationarity is defined as

$$DS(\omega) = \frac{1}{T} \int_0^T \left(1 - \frac{H(\omega, t)}{n(\omega)}\right)^2 \mathrm{d}t, \tag{7.7}$$

in which the integrand is similar to the intermittency defined in the wavelet analysis. We decided to adopt the integration representation because it gives a quantitative

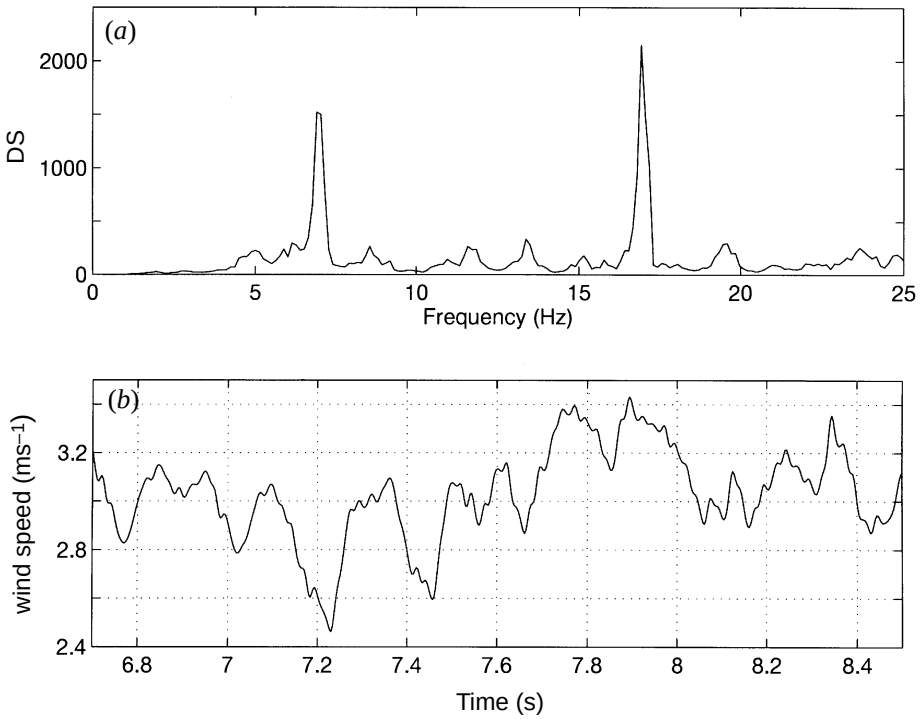

Figure 13. The intermittence and the DSS. (*a*) A linear plot of the DSS with 100 time step average. Energy at 7 and 17 Hz are highly energetic and intermittent. (*b*) A section of wind data demonstrating the presence of energetic and intermittent oscillations at 7 and 17 Hz.

measure of the whole data set. Obviously, for a stationary process, the Hilbert spectrum cannot be a function of time. Then, the Hilbert spectrum will consist of only horizontal contour lines; $DS(\omega)$ will then be identically zero. Only under this condition will the marginal spectrum be identical to the Fourier spectrum, then the Fourier spectrum will also make physical sense. If the Hilbert spectrum depends on time, the index will not be zero, then the Fourier spectrum will cease to make physical sense. The higher the index value, the more non-stationary is the process. The DS for the wind data is shown in figure 12. As the index shows, the data are highly non-stationary especially for the high-frequency components.

Equation (7.7) defines the stationarity as a function of frequency. This is necessary, for certain frequency components can be non-stationary while other components remain stationary. An example is the sporadic riding wind waves on otherwise uniform swell: the low-frequency swell component is stationary, while the high-frequency wind waves are intermittent, and hence non-stationary.

The degree of stationarity can also be a function of time implicitly, for the definition depends on the time length of integration in equations (7.6) and (7.7). Therefore, a process can be piecewise stationary. On the other hand, for a singular outburst in an otherwise stationary signal, the process can be regarded as almost stationary if we take a long time integral, but non-stationary if we look at the immediate neighbourhood of the outburst. Stationarity can be a complicated property of the process: for any data shorter than a typical long wave period, the process may look transient, yet as the data length gets longer, the process can have many longer wave periods and becomes stationary. On the other hand the data can be locally stationary while

in a long time sense non-stationary. An index is therefore not only useful but also necessary to quantify the process and give a measure of the stationarity.

The degree of stationarity defined in equation (7.7) can be modified slightly to include the statistically stationary signals, for which the degree of statistic stationarity, $DSS(\omega, \Delta T)$, is defined as

$$DSS(\omega, \Delta T) = \frac{1}{T}\int_0^T \left(1 - \frac{\overline{H(\omega, t)}}{n(\omega)}\right)^2 \mathrm{d}t, \tag{7.8}$$

where the overline indicates averaging over a definite but shorter time span, ΔT, than the overall time duration of the data, T. For periodic motions, the ΔT can be the period. Such a time scale, however, is hard to define precisely for high-dimensional non-stationary dynamic systems. Even with this difficulty, the definition for DSS could be more useful in characterizing random variables from natural phenomena. Obviously, DSS will depend on both frequency and the averaging time span. For the wind data taken as an example, the DSS with ΔT as a parameter is given in figure 12 with $\Delta T = 10$, 50, 100 and 300 time steps, respectively. The results show that while the high-frequency components are non-stationary, they can still be statistically stationary. Two frequency bands at 7 and 17 Hz are highly non-stationary as the DSS averaged at 100 time steps shows in figure 13*a*. These components are intermittent as can be seen in the IMF components and the marginal spectrum. A section of the original wind data is also plotted in figure 13*b* to demonstrate that there are indeed prominent 7 and 17 Hz time scale oscillations.

8. Validation and calibration of the Hilbert spectrum

Through the empirical mode decomposition and the associated Hilbert spectral analysis we obtained the probabilistic Hilbert spectrum representation of the non-stationary data. Now, we will validate the approach and the results, and calibrate its fidelity compared with the best existing method, the wavelet analysis. Before proceeding further, let us first examine the physical meaning of EMD.

Historically, there are two important methods to expand a function in series: the Taylor and the Fourier expansions, both powerful but based on totally different approaches. The Taylor series for a function, $f(\mathrm{t})$, is expanded near a point, t_0, as

$$f(t+t_0) = f(t_0) + f'(t_0)(t-t_0) + f''(t_0)\tfrac{1}{2}(t-t_0)^2 + \cdots + f^{(n)}(t_0)((t-t_0)^n/n!) + \cdots, \tag{8.1}$$

in which the primes and (n) indicate the order of differentiation. The Fourier expansion for the same function is expanded globally as,

$$f(t) = \sum_{j=0}^{n} a_j \mathrm{e}^{\mathrm{i}\omega_j t}, \tag{8.2}$$

in which the coefficients, a_j, are given by

$$a_j = \frac{1}{2\pi}\int_t f(t)\mathrm{e}^{-\mathrm{i}\omega_j t}\,\mathrm{d}t. \tag{8.3}$$

From equations (8.1) and (8.2), the different mathematical principles in deriving the expansions are quite clear. While the Taylor expansion is based on the properties of

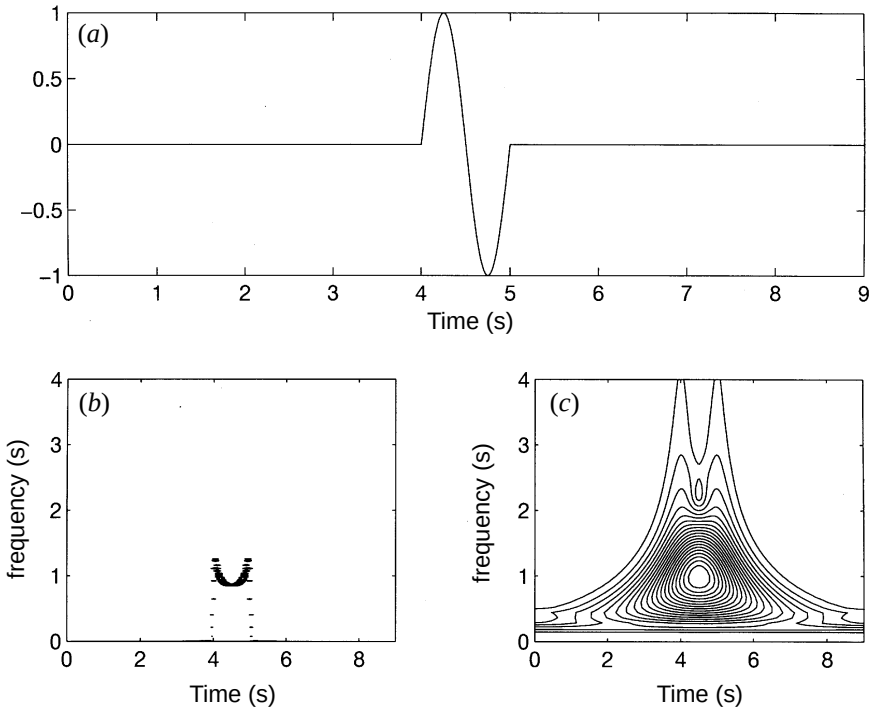

Figure 14. A calibration of time localization of the Hilbert spectrum analysis. (*a*) The calibration data, a single sine wave. (*b*) The Hilbert spectrum for the calibration signal: the energy is highly localized in time and frequency, though there are some end effects. (*c*) The Morlet wavelet spectrum for the calibration signal: the calibration signal is localized by the high-frequency components, yet the energy distribution in the frequency space spreads widely in comparison with the Hilbert spectrum.

the function at a point through differentiation, the Fourier expansion is based on the properties of the function over the whole time span through integration. Although both expansions are valid within the range of convergency, there are different requirements on each function: one of the necessary conditions for the Taylor expansion to exist is that the function has to be analytic to the nth order. Such a restriction is not necessary for the Fourier expansion, which can be applied to functions with denumerable finite jumps. The Fourier expansion introduces a powerful idea: to expand a function with any orthogonal and complete basis. Therefore, there are infinitely many ways to decompose a signal. For example, a signal can be decomposed into simple harmonic components with the Fourier transform. By the same token, the same signal can also be decomposed into Hermite polynomials, or some other special functions such as Legendre or Laguerre polynomials. In fact any collection of orthogonal functions can serve as the basis for a decomposition for a linear system. Once we have a complete and orthogonal basis, the expansion is mathematically correct even for a transient nonlinear system, as testified by the results produced by the classical perturbation analysis for nonlinear differential equations. Whether this mathematically correct expansion also makes physical sense is entirely a separate, but crucial, problem. In general, we believe the answer provided by any linear expansion for a nonlinear system, such as the answer obtained by the perturbation method, should not make physical sense. This conclusion can be argued from two different direc-

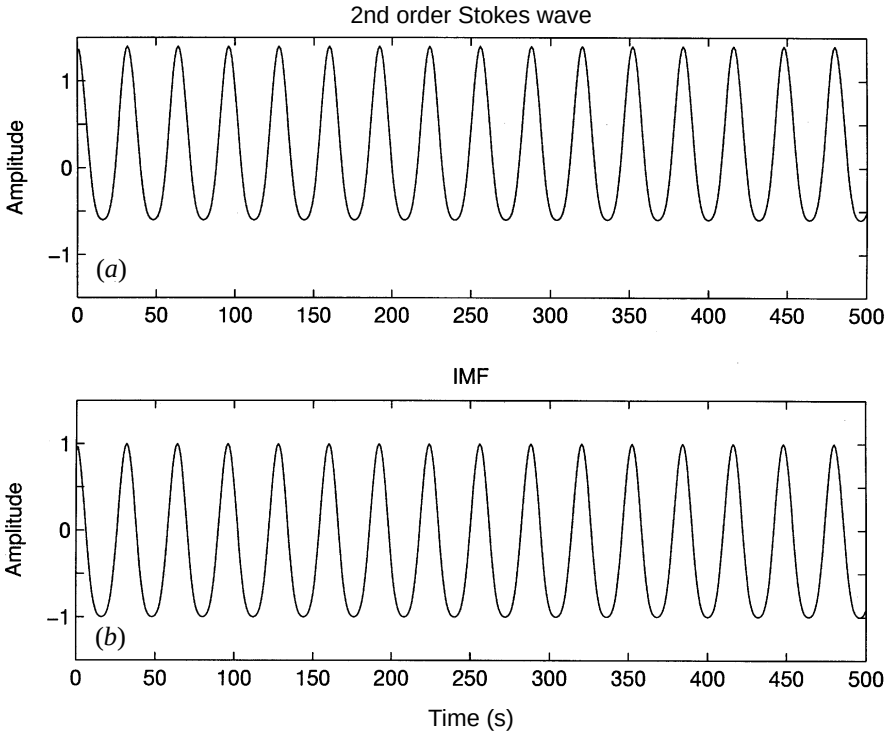

Figure 15. Validation of the intrawave frequency modulation with Stokian waves. (*a*) The profile of a second-order Stokes wave in deep water with sharp crests and rounded-off troughs in comparison with the pure cosine waves. (*b*) The IMF generated by the Stokes wave, there is only one component; the constant off-set is not shown.

tions: first, a nonlinear system does not admit an answer by superposition; second, the perturbation method is forcing a linear approximation on a nonlinear system. Both of the above arguments are true; therefore, it is fortuitous to have the sum of the linear system to approximate the full nonlinear system. It is entirely different to also ask each individual component in the linear system to have physical meaning related to the full nonlinear system.

EMD expansion is at the same time both different from and similar to the above expansions. It is based on both integration, as in the Hilbert transform given in equation (3.1), and differentiation, as in the instantaneous frequency given in equation (3.4). The integration in the Hilbert transform is not exactly global, for it is the convolution of the function with $1/t$, which makes the result extremely local. Essentially, the Hilbert transform gives the best fit with a sinusoidal function to the data weighted by $1/t$.

Let us compare how local the Hilbert transform can be with the result from wavelet analysis by considering an isolated sine wave given in figure 14*a*. With wavelet analysis we get the spectrum in figure 14*c*, in which the event is well defined on the time axis by the high-frequency components, even though the event is a low-frequency wave. In the result, neither the energy density nor the frequency is well localized; they give a counter-intuitive interpretation of the wavelet spectrum: to look for definition of a low-frequency event in the high-frequency range. When the same data are treated by the Hilbert spectral analysis, we have the result in figure 14*b*, in which the

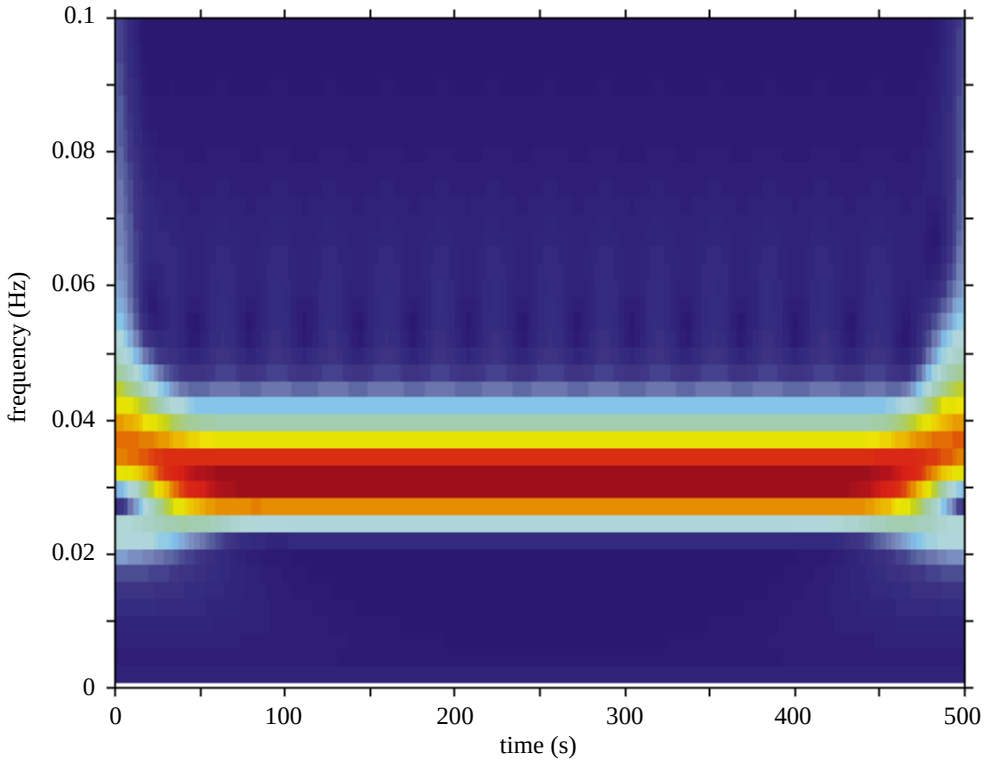

Figure 16. Morlet wavelet spectrum for the Stokes wave showing two bands of energy for the fundamental and the harmonics around 0.03 and 0.06 Hz as expected from the traditional view. The end effect of the wavelet analysis is clearly visible.

energy is well localized in both frequency and time domains. This simple example illustrates the unique property of the Hilbert spectrum in elimination of the spurious harmonic components to represent the non-stationary data.

Next, let us examine the validity and the implication of the instantaneous frequency for nonlinear data through the idealized Stokes wave in deep water, which is a classic example of using harmonic components to represent nonlinear wave form distortion (see, for example, Lamb 1932). It is also the first successful application of the perturbation method to solve a nonlinear analytic equation system for a natural phenomenon. To second order, the profile is given by

$$X(t) = \tfrac{1}{2}a^2k + a\cos\omega t + \tfrac{1}{2}a^2k\cos 2\omega t + \cdots, \tag{8.4}$$

in which a is the amplitude and k is the wave number. For $a = 1$, $ak = 0.2$, and $\omega = \frac{1}{32}$ Hz, the wave-profile is shown in figure 15*a*. Because of the harmonic distortion, the wave form shows sharpened crests and rounded-off troughs, which make the crests and troughs asymmetric with respect to the mean surface. Processed with EMD, these data yield only one IMF component, shown also in figure 15*b*, with a constant offset component not shown. The Stokes wave is deformed due to the harmonic distortion; it is asymmetric with respect to the mean, while the IMF is symmetric. Although this wave has only one characteristic scale, the wavelet analysis result shown in figure 16 gives two bands of energy. The IMF data can be processed immediately to give the Hilbert spectrum as shown in figure 17, which has only one frequency band centred around 0.03 Hz, the fundamental frequency of the wave train,

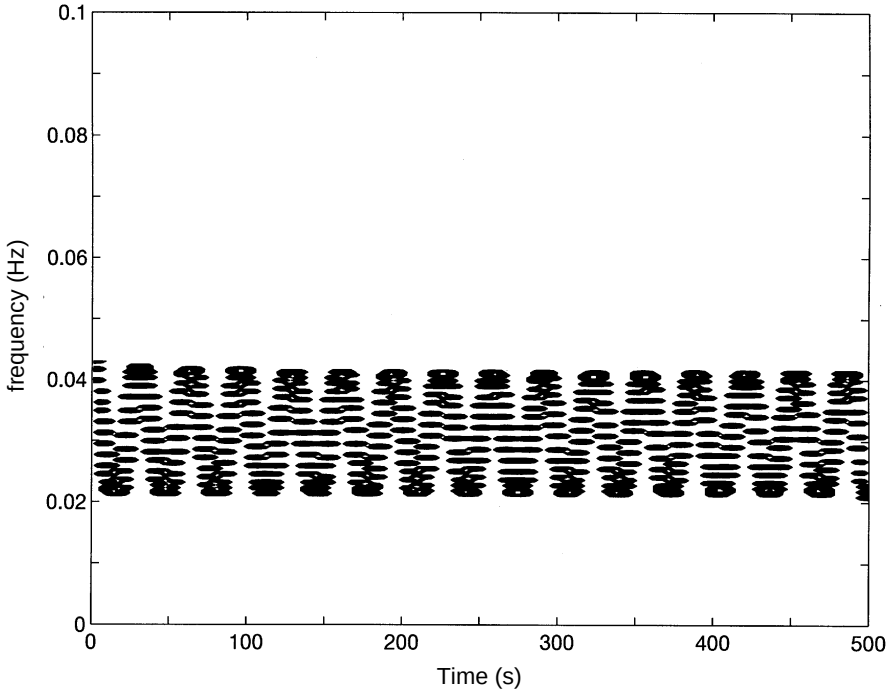

Figure 17. The Hilbert spectrum from the IMF with forced symmetry of the profile in which the intrawave frequency modulation interpretation of the Stokes wave is clear. In the Hilbert spectrum the frequency variation is bounded in a narrow range around 0.03 Hz with no harmonics.

but there is an intrawave frequency modulation with a magnitude range of 0.02–0.04 Hz. This intrawave frequency modulation has been totally ignored in the past, for the traditional definition of frequency is based on the reciprocal of periodicity. Consequently, the notion of intrawave frequency modulation cannot be represented or discussed. Is the intrawave frequency modulation meaningful? What is its physical significance? To answer these questions, let us consider a simple model wave given by

$$X(t) = \cos(\omega t + \epsilon \sin \omega t). \tag{8.5}$$

Figure 18*a* gives the profile of this wave together with the cosine wave given by the dotted line. This wave shows the harmonic distortion similar to the Stokes wave given in figure 15*a*: sharpened crests and rounded-off troughs. In the past, this kind of wave form has been treated as a typical case of harmonic distortion. Indeed, the second-order approximation of equation (8.5) is

$$\begin{aligned} X(t) &= \cos(\omega t + \epsilon \sin \omega t) \\ &= \cos \omega t \cos(\epsilon \sin \omega t) - \sin \omega t \sin(\epsilon \sin \omega t) \\ &\approx \cos \omega t - \epsilon \sin^2 \omega t \approx (1 - \tfrac{1}{2}\epsilon) \cos \omega t + \tfrac{1}{2}\epsilon \cos 2\omega t. \end{aligned} \tag{8.6}$$

This expression is very similar to the one given in equation (8.4).

Equation (8.6), however, is only an approximation to the full expression of equa-

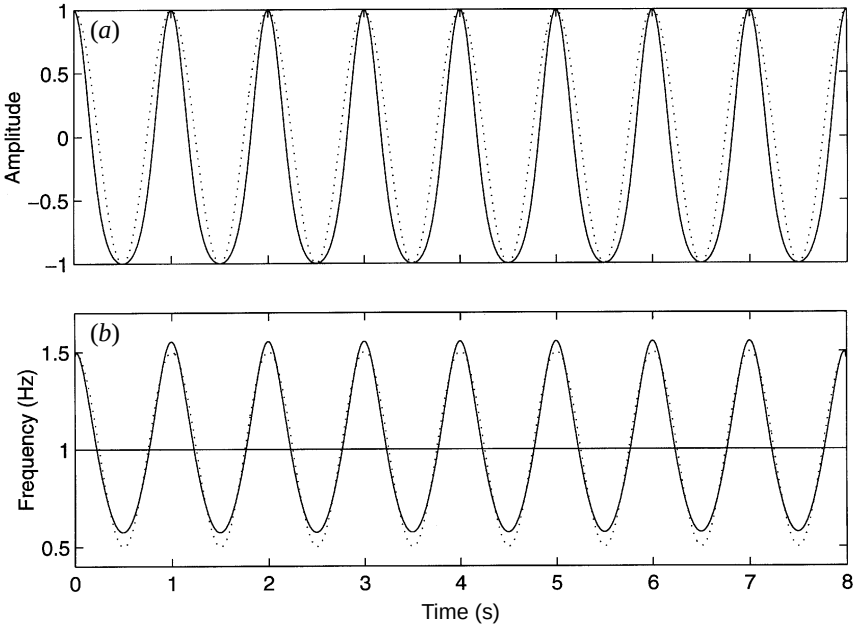

Figure 18. Model wave with intrawave frequency modulation to simulate the Stokes wave. (*a*) Profile of the model wave showing the same characteristics as the Stokes wave: sharp crest and rounded-off trough. A pure cosine wave is shown by a dotted line for reference. (*b*) Frequency of the model wave computed from Hilbert transform (solid line) and equation (8.8) (dotted line). The agreement is excellent for this case.

tion (8.5), which satisfies the following highly nonlinear differential equation:†

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + (\omega + \epsilon\omega\cos\omega t)^2 x - \epsilon\omega^2 \sin\omega t(1-x^2)^{1/2} = 0. \tag{8.7}$$

Although this equation bears no relationship to the system generating the Stokes wave, the example illustrates that harmonic components are the results of using a linear system to simulate the nonlinear one. Although the final solution form is similar to the real one, the harmonic components have no physical meaning other than providing better mathematical approximation.

As given in equation (8.5), the function is clearly a frequency modulated wave with the modulating frequency equal to that of the main oscillating frequency. The frequency given by the classic wave theory for this model wave is

$$\Omega = \frac{\mathrm{d}\theta}{\mathrm{d}t} = \omega(1 + \epsilon\cos\omega t). \tag{8.8}$$

With $\omega = 1$ Hz and $\epsilon = 0.5$, the frequency values are plotted in figure 18*b*, with the solid line representing the values given by the Hilbert transform and equation (3.4), and the dotted line representing the values given in equation (8.8). Both methods give similar variation of the frequency with only minor deviation at the extreme values.

The agreement between the two methods confirms that a deformed sinusoidal wave

† Note added in proof: It can also be shown that a general form of equation (8.5), $X(t) = \cos(\omega t + \epsilon f(t))$, satisfies a linear differential equation with variable coefficients.

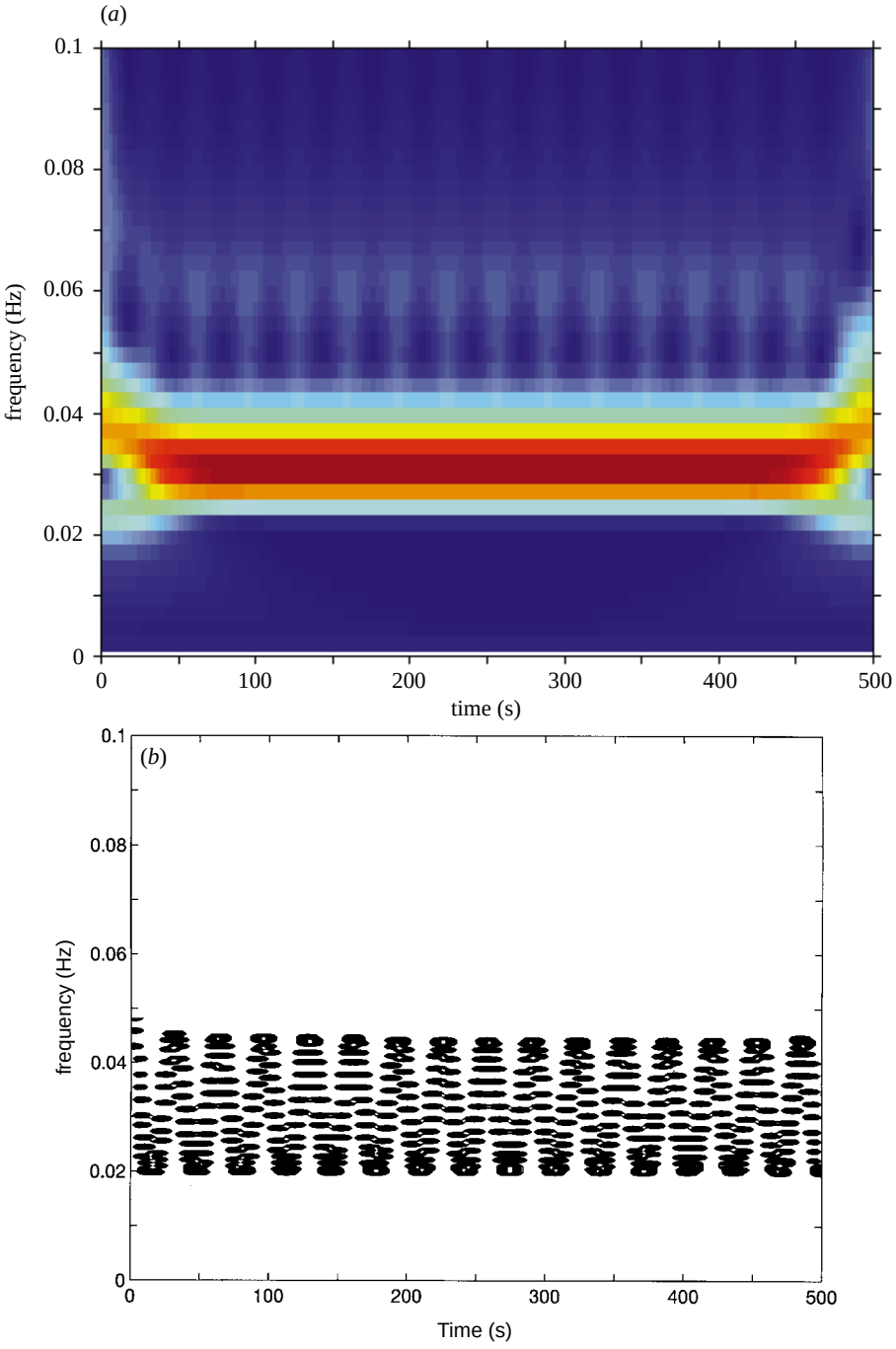

Figure 19. Comparison of the wavelet spectrum with the Hilbert spectrum for the model intrawave frequency modulated wave. (*a*) The Morlet wavelet spectrum of the model wave showing a rich complement of harmonics to simulate the wave-profile deformation due to the intrawave frequency modulation: the main energy is located around 0.03 Hz, with the harmonic at 0.06 Hz as expected. (*b*) The Hilbert spectrum for the same data showing the intrawave frequency modulation as the model wave simulated. These results are almost identical to those shown in figures 16 and 17.

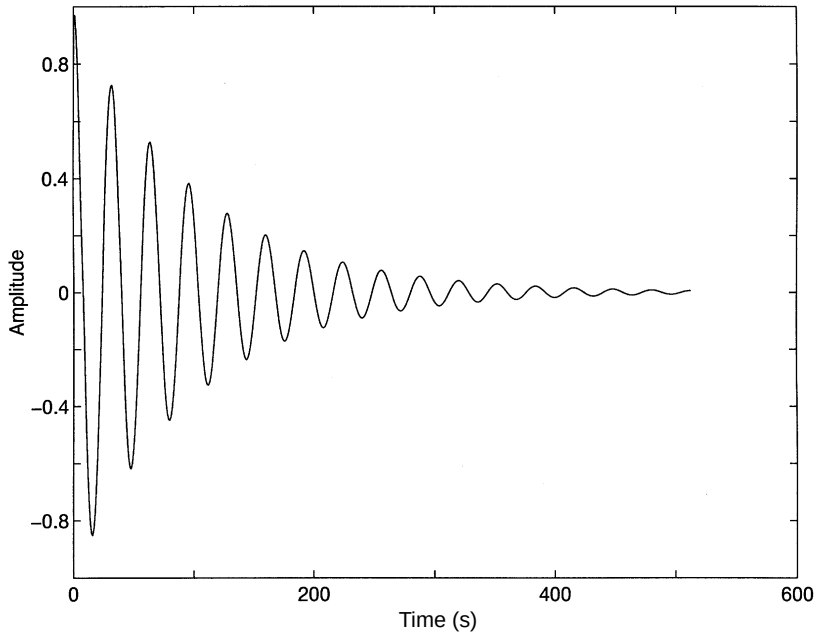

Figure 20. Data for an amplitude modulated wave: a single carrier with exponentially decaying envelope.

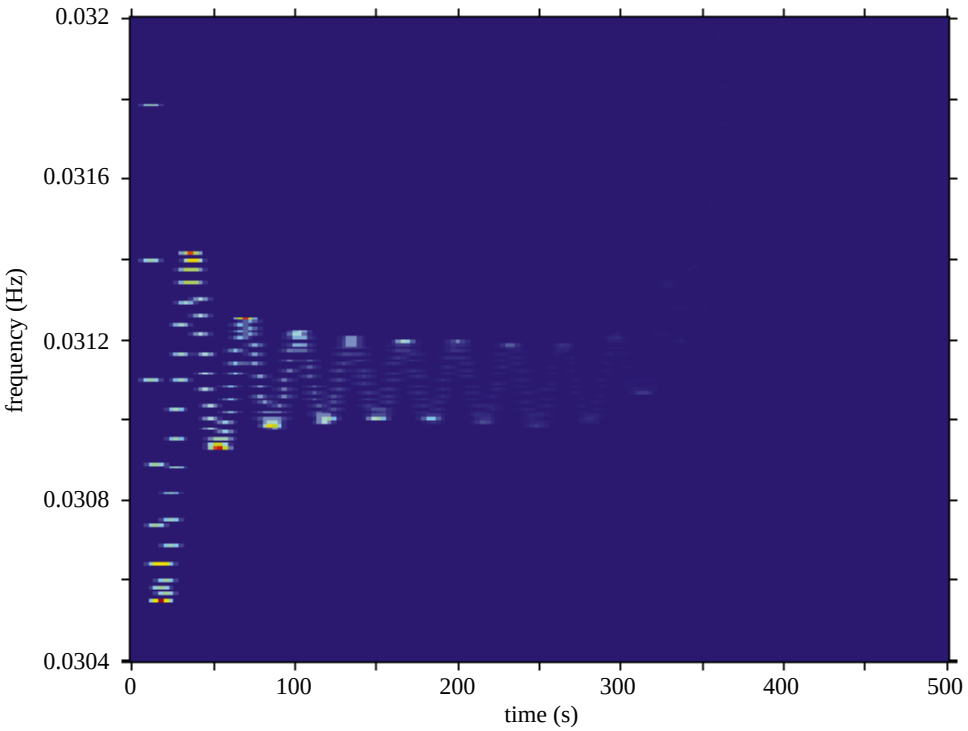

Figure 21. The Hilbert spectrum of the data shown in figure 20. This example shows that the amplitude modulation can also generate intrawave frequency modulation, but the range of the modulation is only ±1.5%.

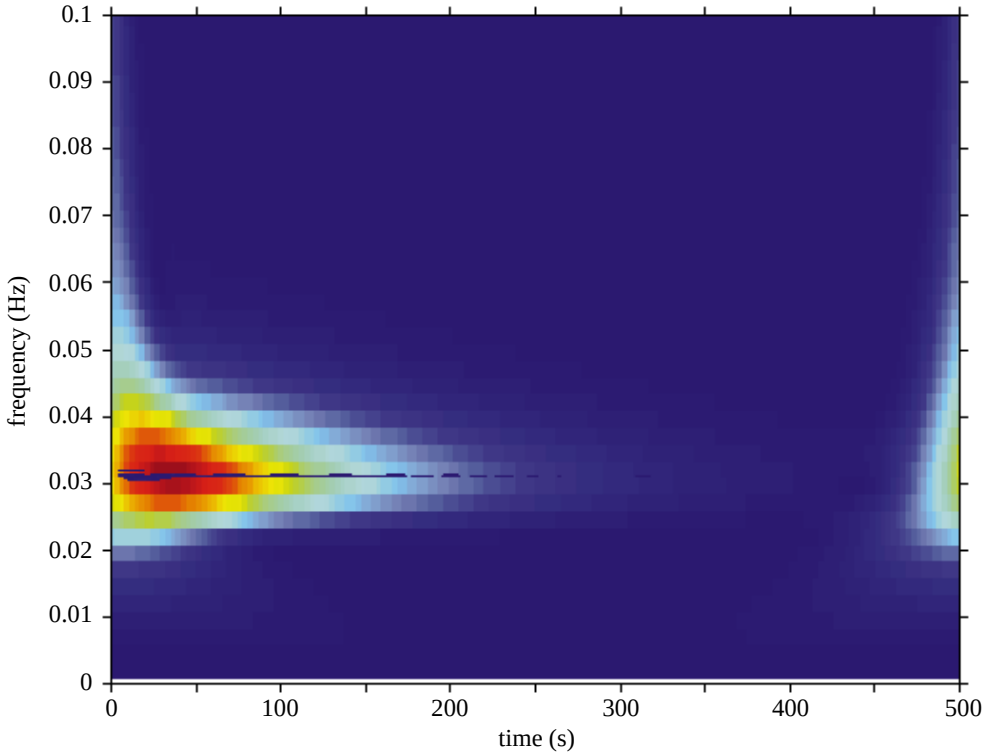

Figure 22. The wavelet spectrum from the same data shown in figure 20. In this representation, the end effect is very prominent; the frequency is widely spread. As a comparison of wavelet and Hilbert spectral analysis, the Hilbert spectrum is also plotted in this figure in contour lines, which shows up as a thin line around 0.03 Hz. This example illustrates two points: the Hilbert spectrum has much better frequency definition, and the amplitude-variation-induced frequency modulation is small.

should be interpreted as an intrawave frequency modulated wave; it is more physical. To illustrate the non-physical representation using harmonics, let us consider the water-wave problem again. Traditionally, the gravity water waves have always been represented by the harmonics to account for the nonlinear distortion. The inclusion of the harmonics, however, introduced a difficulty: not all components follow the same dispersion relationship. As a result, we are forced to divide the components into the 'free' and the 'forced' waves. The 'free' waves are the fundamentals that follow the dispersion relationship; the 'forced' waves are the harmonics that are bound to the fundamentals. In a wave spectrum, one cannot clearly separate the 'free' from the 'forced' unless the wave is monochromatic. This division is a consequence of using a linear system to approximate the nonlinear one; it is totally artificial. We will return to this point later with more examples.

Now let us examine the data analysis results from different methods: The wavelet analysis result for the model data given in equation (8.5) is shown in figure 19*a*, while the Hilbert spectrum is shown in figure 19*b*. Comparison between the wavelet analysis and Hilbert spectra pairs in figures 19*a* versus 19*b* shows that the intrawave frequency modulation is a new and better physical interpretation for the energy–frequency–time distribution for the Stokian waves. The difference is again clear: the Hilbert spectrum clearly resolves the intrawave frequency modulation, while the wavelet result gives

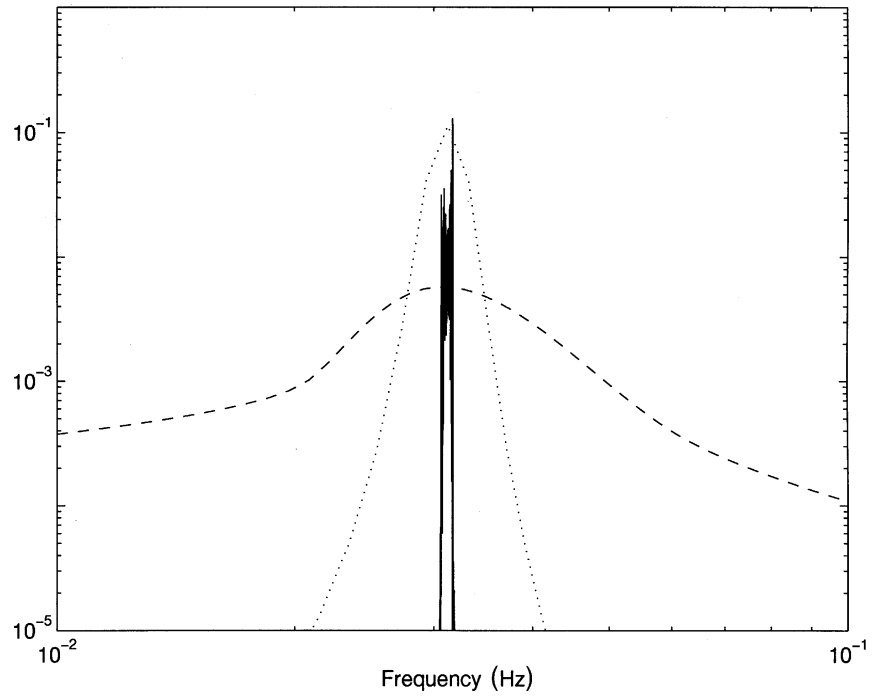

Figure 23. The comparisons among the Fourier (dotted line), marginal Hilbert (solid line) and wavelet spectra (dashed line): it is obvious that the end effects and the leakage of the wavelet analysis renders the marginal spectrum almost useless. Even the Fourier spectrum is sharper than the marginal wavelet spectrum, but it totally failed to show the time variation of the signal. The marginal Hilbert is still sharp to define the carrier frequency.

a smeared average frequency range over which the main wave energy resides. There is a total lack of detail beyond this smoothed mean, which is to be expected, for the wavelet determines the frequency based on the Fourier transform in a window with 5.5 waves. Any change shorter than this window will be obscured. Other than the lack of details, Fourier-based analysis also spreads energy to the high-frequency range as the harmonics. Traditionally, these harmonics are viewed as a matter of fact, but the new method here reveals that the Fourier expansion is a mathematical approximation to a nonlinear physical process, in which the true physical meaning is beyond the reach of Fourier-based analysis.

The above examples have not only validated the EMD and the Hilbert spectrum representation, but also clarified the conditions under which the spurious harmonics are generated in the Fourier-based analysis: nonlinearity and non-stationarity. The nonlinear case is illustrated by the Stokes waves, and the non-stationary case by the single sine wave. By comparison, this nonlinear distortion of the wave form causes much fewer serious non-physical harmonic terms in Fourier analysis than the non-stationary data. In the past, this crucial problem has not been addressed. Typically, the perturbation analysis gives a solution in series expansion. Each term is the solution of a linear equation. Although the infinite series expansion is so powerful that it can approximate some transient phenomena with uniform amplitude components, their physical meaning has never been examined critically. The mathematical success has obscured our physical insights.

To illustrate the cause of intrawave frequency modulation further, let us consider an example of an amplitude modulated wave given by

$$x = \exp(-0.01t)\cos\tfrac{2}{32}\pi t, \tag{8.9}$$

for t from 1–512 s. The wave form is given in figure 20. Since this wave is already an IMF, there is no need for invoking EMD. The Hilbert spectrum of the data is shown in figure 21. Clearly, the amplitude has introduced a 1.5% instantaneous frequency modulation around the mean of the carrier frequency. This frequency modulation should not be a surprise, equation (3.11) has already indicated that amplitude variations could cause a frequency fluctuation, but not a change of the mean. The uneven range of the frequency modulation is caused by the Gibbs phenomenon through the Hilbert transform. To get a realistic appraisal of the fluctuation range, the same data are processed by the Morlet wavelet analysis with the result given in figure 22, in which the Hilbert spectrum is superimposed in contour lines for comparison. On this scale, the frequency modulation in the Hilbert spectrum appears as a single thin line. The wavelet results not only spread the frequency over a wide range, but also contain severe end effects caused by the Gibbs phenomenon. To explore the differences among the various spectral representations further, the marginal Hilbert and wavelet spectra are plotted together with the Fourier spectrum in figure 23. One can immediately see the sharpness of the Hilbert marginal spectrum in its definition of the mean frequency. In this case, the Fourier (dotted line) is actually sharper than the marginal wavelet spectrum, though the Fourier spectrum has totally missed the amplitude variation as a function of time. The spreading of energy in the wavelet analysis is due both to the leakage and the end effects. In all fairness, it should be noted that the marginal spectrum for a non-stationary time series does not make sense. This example only intends to make a reference to the Fourier analysis. More importantly, this example also demonstrates that, while amplitude variation can cause intrawave modulation, the effects of amplitude-induced intrawave modulation are negligible compared with the nonlinear distortion.

Finally, let us examine the resolution power of the EMD and Hilbert spectral analysis by the following examples. First, let us consider the case of a simple cosine wave with one frequency suddenly switching to another frequency as shown in figure 24*a*, the classic example used in many reports to illustrate the power of wavelet analysis in frequency–time distributions. The standard Morlet wavelet analysis clearly identifies the local frequency before and after the frequency switch as well as the location of the frequency switch as in figure 24*b*. At the same time, the result also shows the leakage of the energy to the neighbouring modes, and the smearing of the precise time location of the frequency switch event in the lower-frequency range. Figure 24*c* shows the result of Hilbert spectral analysis from the same data. Here we can see much sharper frequency definitions and the time location of the frequency switch than those shown in figure 24*b*. Although the finite data length and the sharp frequency change both contribute to some oscillation of the frequency in the Hilbert spectrum due to the Gibbs phenomena, the range of variation is insignificant compared with the leakage in the wavelet analysis result.

Second, let us consider the case of the linear sum of two cosine waves as

$$x(t) = \cos\tfrac{2}{30}\pi t + \cos\tfrac{2}{34}\pi t, \tag{8.10}$$

again over t from 1–512 s, with the wave form given in figure 25. The slight asymmetric wave form suggests that the EMD method has to be used. If the sifting process

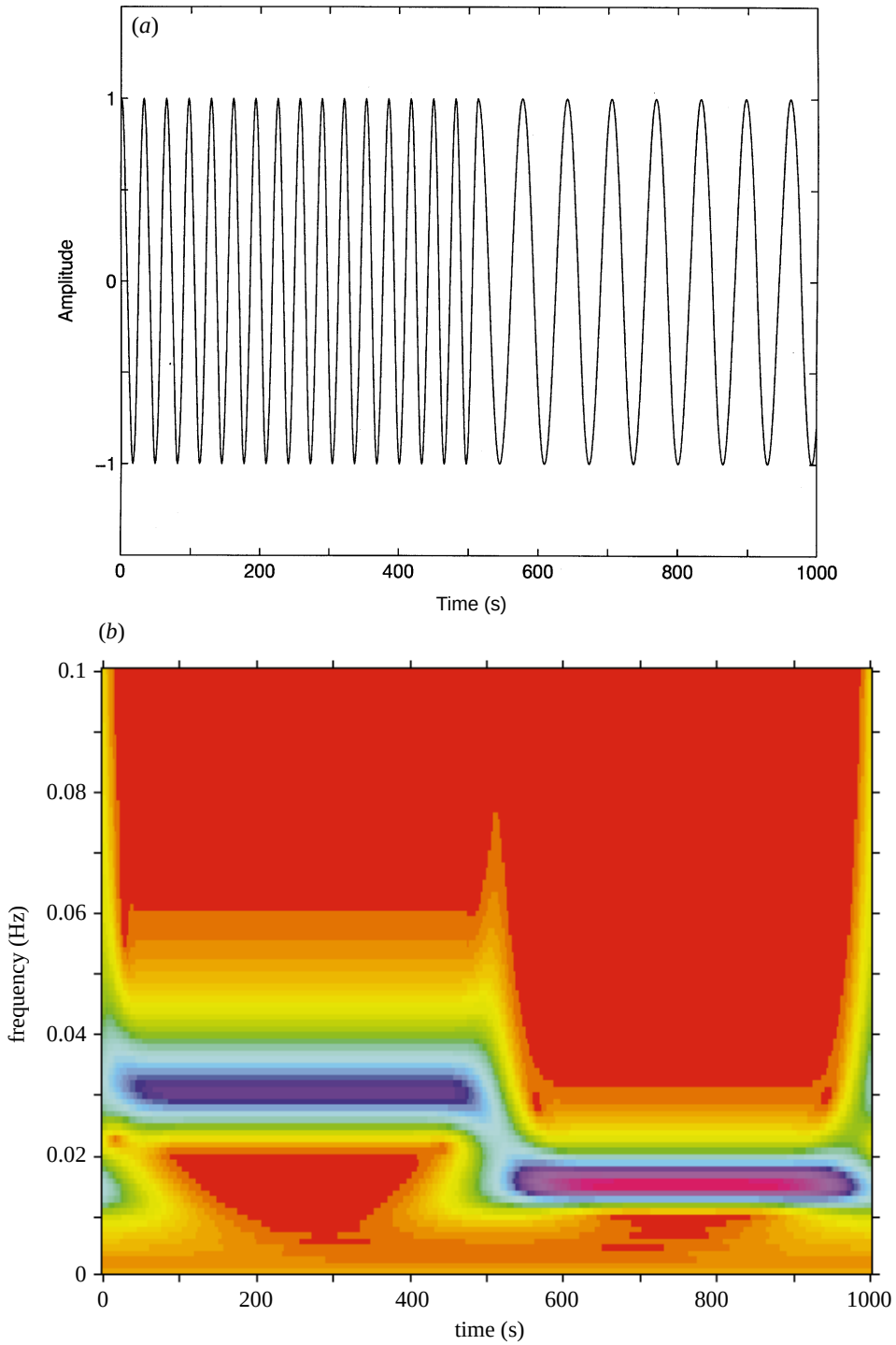

Figure 24. Comparison between the wavelet spectrum and the Hilbert spectrum for the classic frequency shift cosine waves: (*a*) the data from the frequency shift cosine waves; (*b*) the Morlet wavelet spectrum for the same data showing poor frequency and time localizations for the frequency shift event.

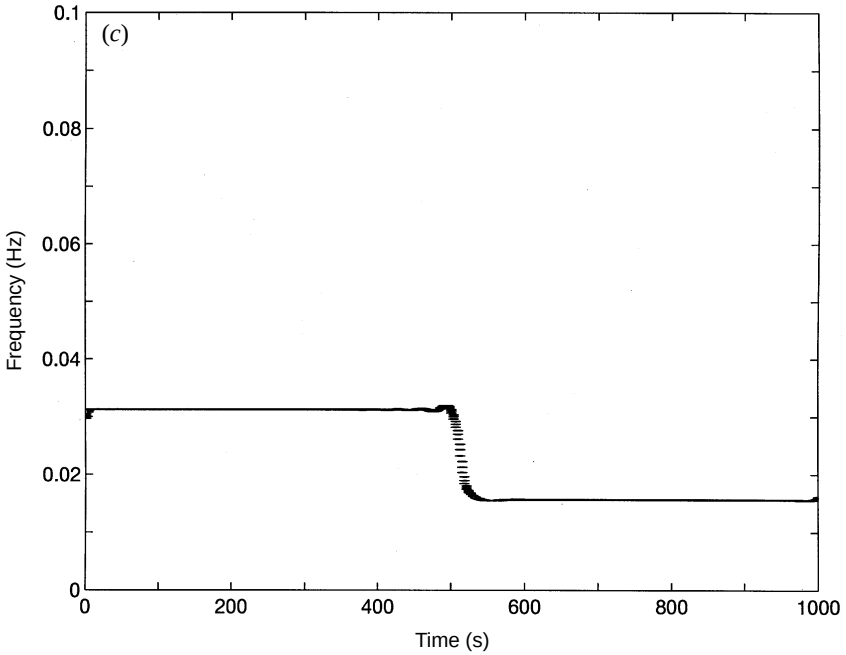

Figure 24. *Cont.* (*c*) The Hilbert spectrum for the data showing sharp frequency and time localizations. Some frequency oscillations are visible due to the Gibbs phenomena, but the magnitude of these frequency oscillations is insignificant compared to the wavelet spectrum.

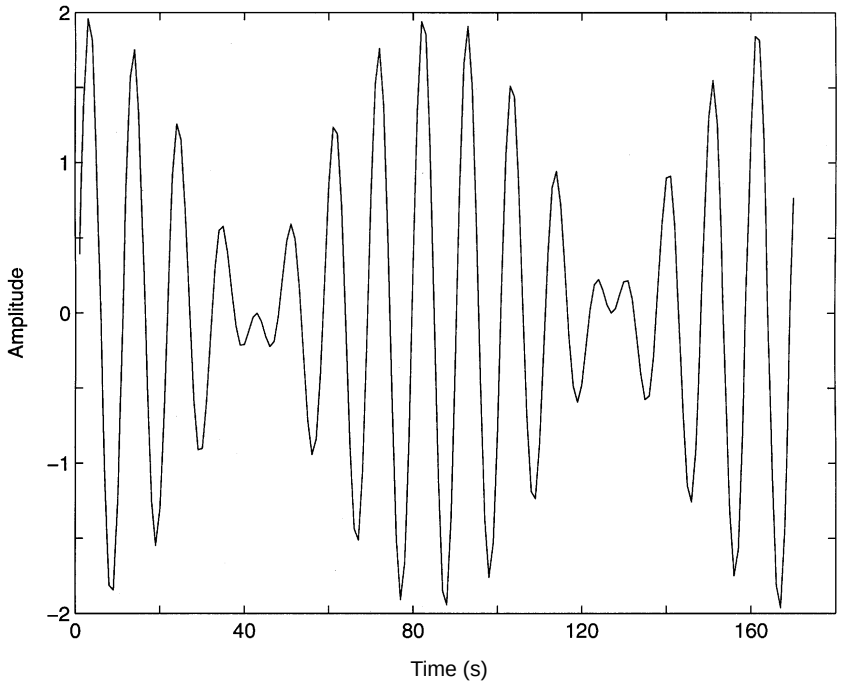

Figure 25. Data of the linear superimposed cosine waves given in equation (8.10): $\cos(\frac{2}{30}\pi t)+\cos(\frac{2}{34}\pi t)$. The data show the expected beats. Although the wave form is asymmetric, the envelopes should be symmetric.

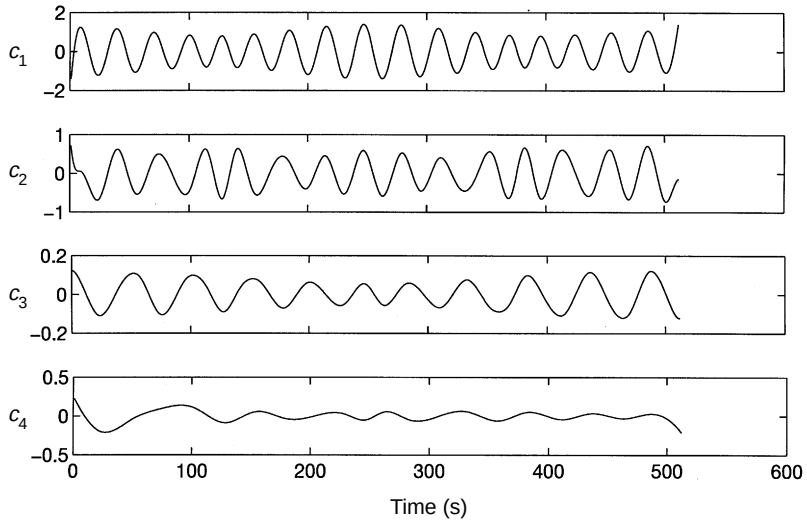

Figure 26. The first four IMF components after stringent application of EMD. There are up to 3000 siftings for some components. The first components, c_1, indeed shows $512/30 = 17$ waves; while the second components, c_2, shows $512/34 = 15$ waves. The other components are of negligible energy. While the EMD succeeded in extracting the intrinsic time scale of the processes, the components failed to give the exact constant amplitudes and frequencies, a flaw due to the accumulation of errors from the imperfect spline fittings in the sifting processes.

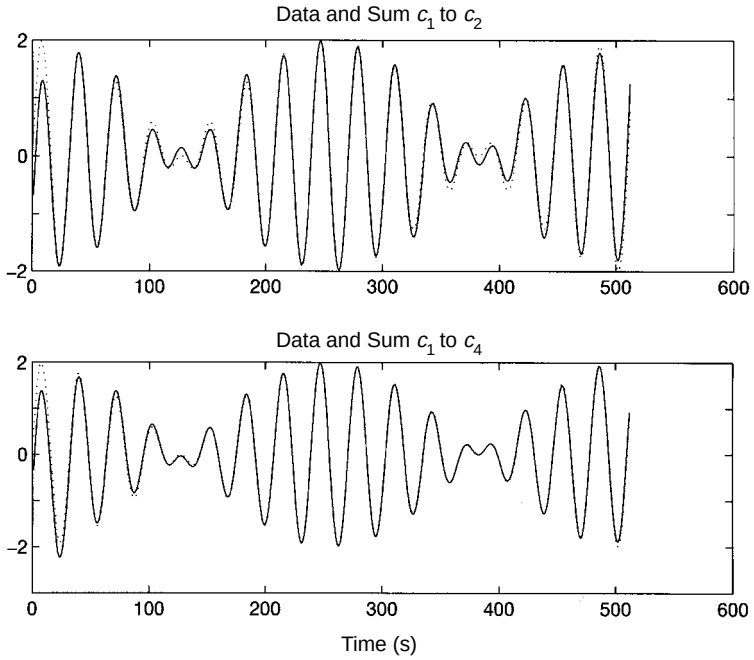

Figure 27. The reconstruction of the signal from the IMFs for the linear superimposed cosine waves given in equation (8.10) and figure 25. Data shown in dotted lines, and the reconstructions shown in solid lines. The upper panel shows the sum of the first two terms. In this reconstruction, the difference between the data and the reconstruction is small. The lower panel shows the sum of the first four terms. In this reconstruction, the difference between the data and the reconstruction is negligible except at the begining due to end effects.

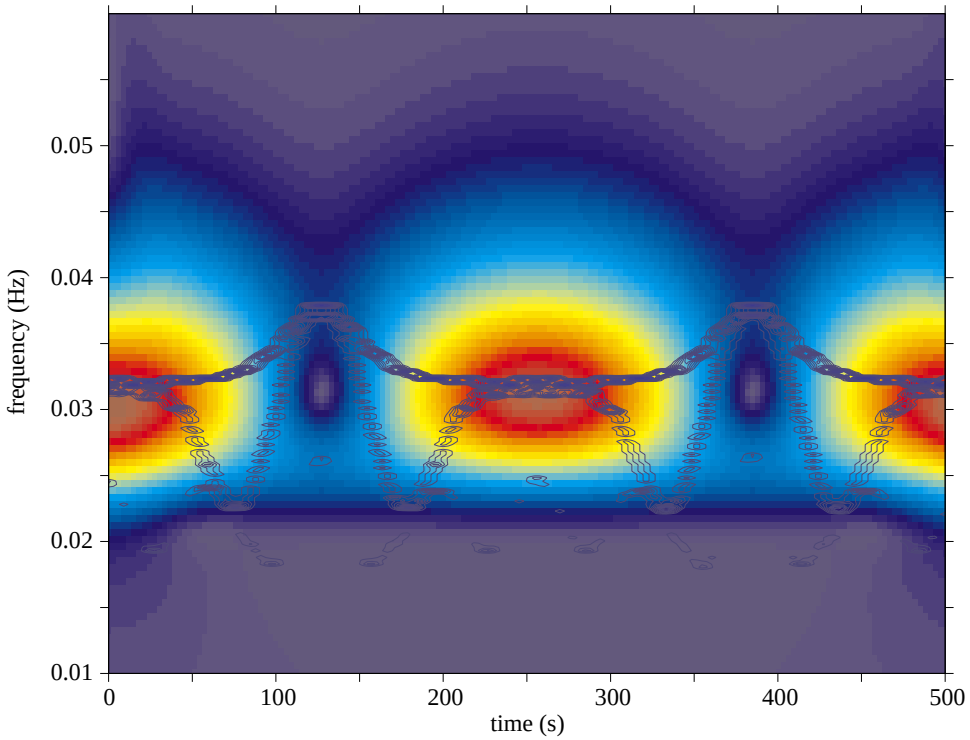

Figure 28. The comparison between the wavelet spectrum (in colour) and the Hilbert spectrum (in contour lines). Both methods show a similar energy frequency distribution, and both methods treat the data as a time varying signal.

is used with limited number of siftings as in the regular problems, no IMF will be produced, for the data are already an IMF to the first order of approximation. To force the EMD to extract different components, an extremely stringent criterion is invoked here. After up to 3000 siftings, EMD produces eight components from the data. The first four are shown in figure 26. As one can see, the first two components carry most of the energy. They represent the $\frac{1}{30}$ Hz (17 waves in the given time span) and $\frac{1}{32}$ Hz (15 waves) components, respectively. Comparisons of the sums of the first two and the first four components with the full data are given in figure 27. The difference between the two-term sum and the data is already small; the difference between the four-term sum and the data is negligible except near the ends.

From the IMFs, the Hilbert spectrum is constructed and presented in contour lines in figure 28, in which the Morlet wavelet spectrum from the same data is also given in colour. The agreement between the two methods is obvious. Again, the wavelet result shows severe leakage. Neither method shows two distinct linear components; both methods treated the beats as an amplitude modulation wave. Figure 29 gives the marginal spectra from both Hilbert and Morlet wavelet spectra, together with the Fourier spectrum. From this figure, it is clear that none of the three methods can separate the peaks of the two linear components. For this linear and stationary case the wavelet performs the worst. Although EMD succeeded in extracting the two components, the IMFs have been severely contaminated through the thousands of spline fitting in the siftings. As a result, the instantaneous frequencies of the

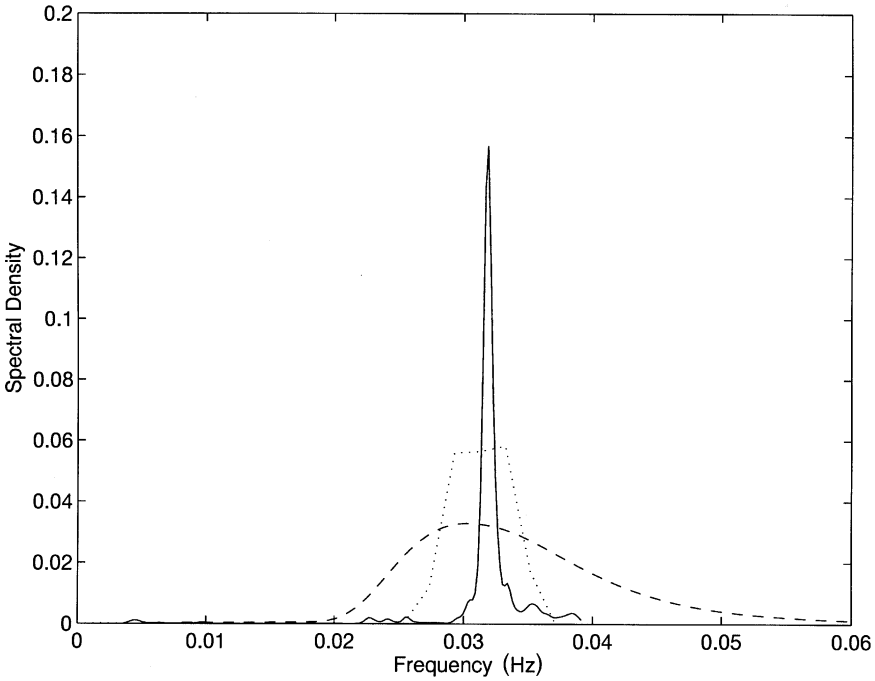

Figure 29. The comparisons between the Fourier (dotted line), marginal Hilbert (solid line) and wavelet spectra (dashed line): the Fourier spectrum should give the true frequency distribution, but it also failed to separate the frequency of the two components. It does, however, give a clear definition of the range of their frequencies. It is obvious that the leakage of the wavelet analysis renders the marginal spectrum almost useless. The marginal Hilbert is still sharp to define the mean frequency, but also fails to separate the components.

components become overlapped to form a single peak located near the mean value. The Fourier spectrum, though, also fails to separate the components for this short data span that we have here, it can clearly define the frequencies by repeated folding of the data to increase the total data span. Longer data will certainly help the EMD by limiting the end effects, but the key problem is not the data length but spline fitting. Unless and until a better spline fitting is implemented, EMD will not be able to improve the instantaneous frequency definition in this case. Using the higher-order spline with a condition on higher-order continuity will certainly improve the result. Before then, the Fourier analysis, but not the Wavelet, should be used for the known stationary time series.

With these idealized examples, we have established the validity and the limitations of the Hilbert spectral analysis. Next, we will present some applications in numerically computed results from low-dimensional nonlinear equations, and some data from observations.

9. Applications

In this section, we will demonstrate applications through a number of examples roughly divided into two categories: the first category contains numerical solutions from some low-dimensional classical nonlinear equations; the second category contains the applications to specific natural phenomena and laboratory data that include

wave data from both the laboratory and the field, the equatorial region ocean surface elevation measured by altimeter, earthquake records and turbulence. In all the cases, the EMD and Hilbert spectral analysis will offer physical insight.

(*a*) *Numerical results from classic nonlinear systems*

We will first present some examples on classic nonlinear systems. The advantage of studying these systems is their simplicity, yet they contain all the essentials of the possible nonlinear effects. All these systems have been studied extensively; therefore, most of their dynamic characteristics are familiar, though the detailed physics may not necessarily be well appreciated. Using them as examples with the new method, we can contrast the different views and gain new insights, which will also help us in interpreting more complicated cases in high-degrees-of-freedom natural phenomena.

(i) *Duffing equation*

The Duffing equation is one of the simplest nonlinear equations with the form,

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + x + \epsilon x^3 = \gamma \cos \omega t, \tag{9.1}$$

in which ϵ is a small parameter, γ is the amplitude of the oscillatory driving force with a basic frequency ω. The homogeneous part of the equation has the unperturbed Hamiltonian,

$$H_0 = \tfrac{1}{2}\dot{x}^2 + \tfrac{1}{2}x^2 + \tfrac{1}{4}\epsilon x^4, \tag{9.2}$$

in which $\dot{x}$ is the time derivative of x. For $\epsilon = -1$; the intrinsic frequency can be shown to be 0.124 Hz. Traditionally, Duffing's equation has been solved by the perturbation method (see, for example, Drazin 1992) with the solution expressed as a series of the basic frequency and all its superharmonics. Such a solution is no doubt correct mathematically, for the perturbation method generates the superharmonics to approximate the highly deformed wave form due to the nonlinearity as

$$x(t) = \cos \omega t + \epsilon \cos 3\omega t + \epsilon^2 \cos 6\omega t + \cdots. \tag{9.3}$$

This solution from the perturbation method is a typical example of approximating a nonlinear system with a collection of linear ones. Collectively, they show the correct wave-profile, but neither collectively nor individually will the components give any physically meaningful representation of the full nonlinear system. Equation (9.1), however, can be rewritten as

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + x(1 + \epsilon x^2) = \gamma \cos \omega t. \tag{9.4}$$

In this form, it offers a clearer physical interpretation of the equation: it can be viewed as representing a simple spring–mass system with a forcing function. The spring constant, however, is not a constant, but varies with position of the mass. The equation can equally be viewed as representing a pendulum with a forcing function. The pendulum length, however, is not constant, but varies with the angle of the swing. Either of these views suggests that the oscillation is not one of a constant frequency within a period. And the changing of spring constant or the length of the pendulum calls for local change of 'instantaneous frequency', an intrawave frequency modulation. This conception of intrawave frequency modulation clearly makes physical sense from equation (9.4). It offers more detailed information than

the Hamiltonian system, in which the averaged frequency is defined as

$$\frac{\partial J}{\partial t} = 0 \quad \text{and} \quad \omega = \frac{\partial \theta}{\partial t} = \frac{\partial H(J)}{\partial J}, \tag{9.5}$$

with J as the averaged action density defined by

$$J = \frac{1}{2\pi} \iint \mathrm{d}p\mathrm{d}q, \tag{9.6}$$

in which θ is the angular variable, p and q are the generalized momentum and the coordinate respectively, and H is the total Hamiltonian in terms of the action density variable. In these canonical expressions, the most important parameter is the averaged period or frequency, based on which the Poincaré section and the modern topological view of the dynamic system are built. With such a view, the shape of the phase plane is not as important as the time needed to trace a full cycle. As long as they are closed curves, they are topologically equivalent. Yet the different shapes of the phase curves represent the different details of the oscillations. Such details can only be represented by the instantaneous frequency, as will be shown presently. The motion described by equation (9.4) clearly requires the instantaneous frequency variation to specify its full physical details.

Now let us examine a specific example by taking $\gamma = 0.1$; $\epsilon = -1$; $\omega = \frac{1}{25}$ Hz; and the initial conditions: $\{x(0); x'(0)\} = \{1, 1\}$. Numerically integrating equations (9.1) and (9.4) give the solutions for $x(t)$ and $x'(t)$ in figure 30 with the phase plane presentation in figure 31. Two features stand out from the phase diagram. First, the phase diagram shows continuous winding, an indication of no fixed period for the oscillation. Second, the shape of the phase diagram is strongly deformed from a simple circle, which gives both the sharp peaks and troughs in the time series data. The sharpness of the peaks and the troughs requires the Fourier expansion of the solution to contain many harmonics.

When subjected to the EMD, this numerical result yields only three prominent IMF components as shown in figure 32: a uniform amplitude high-frequency component representing the intrinsic frequency of the system; a uniform intermediate frequency component representing the forcing function; and a low-frequency component representing the very low-intensity subharmonics. The first, and most energetic, component is highly deformed from the simple sinusoidal profile, a clear indication of the nonlinear nature of the oscillator. Yet, this component represents a physical entity, and is treated as such by the EMD, a clear indication that the EMD can indeed analyse nonlinear data. In the Hilbert spectrum given in figure 33, the intrinsic frequency shows strong intrawave frequency modulation, which is presented as variable frequency oscillating between 0.06 and 0.18 Hz, with a weighted mean around 0.1 Hz, the averaged frequency as predicted by the Hamiltonian method. The detailed variations of the intrinsic frequency indicate that it contains both inter- and intrawave frequency modulations. The forcing function is also clearly shown with the expected frequency. The only unexpected component is the low-frequency and low-amplitude signal. It represents the slow aperiodic wobbling of the phase; therefore, it is real. Although an explanation is needed for its generation and its physical significance, its existence has been reported by Kevorkian (1966).

When the same data are treated with the Morlet wavelet analysis with the same total number of frequency resolutions, the results are given in figure 34. The richness of the harmonics is generated by the sharpness of the wave crests and troughs.

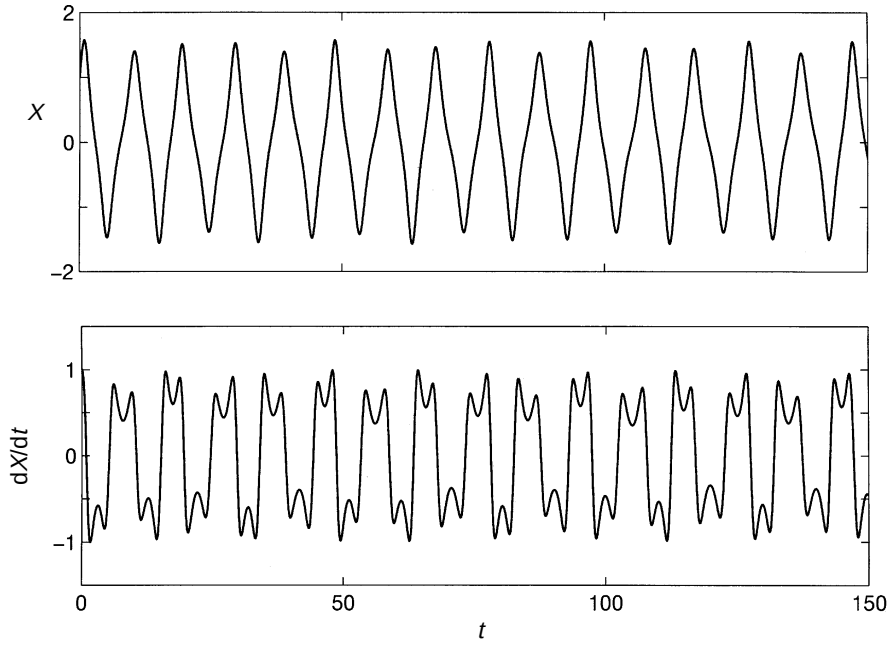

Figure 30. Numerical solution for the Duffing equation. The oscillation is highly nonlinear as indicated by the deformed wave-profile with pointed crests and troughs.

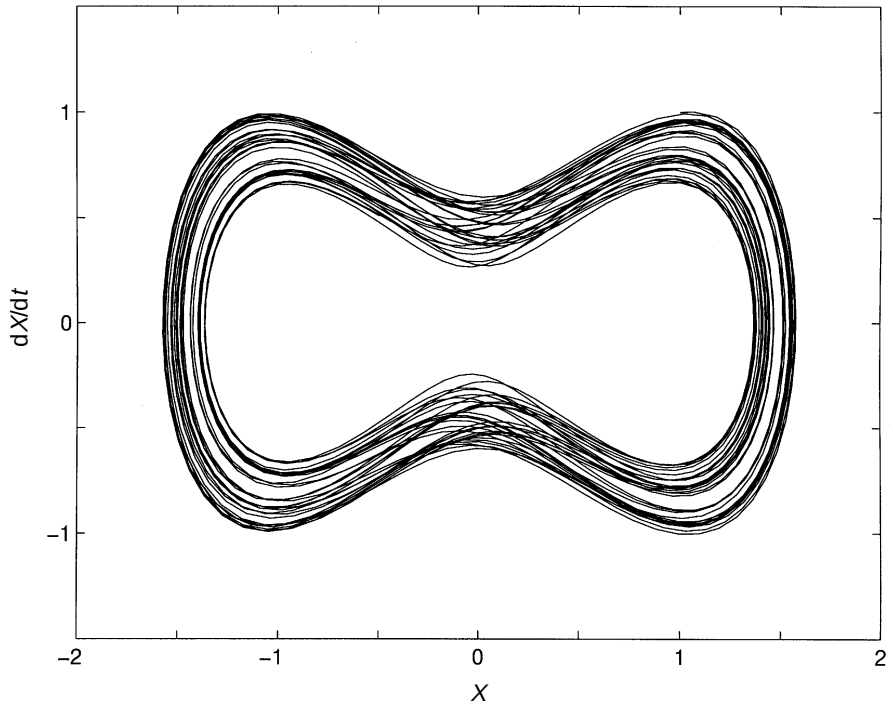

Figure 31. The phase diagram of the Duffing equation solutions with continuous winding indicating no fixed period of the oscillations. The strongly deformed shape from a simple circle gives the pointed crests and troughs of the time series data.

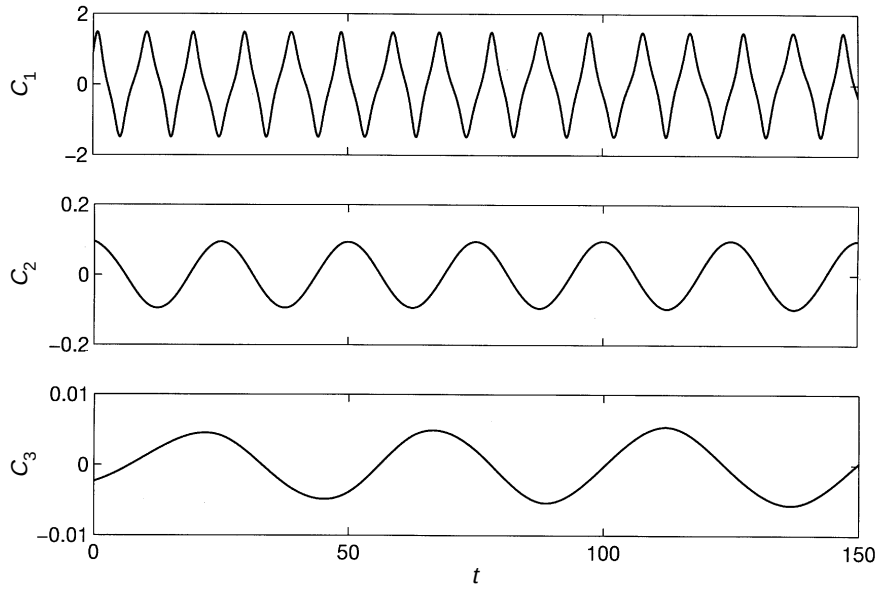

Figure 32. The IMF components of the Duffing equation from EMD: the first component is nonlinear, but it is a physical mode. This example shows that EMD is a nonlinear decomposition method.

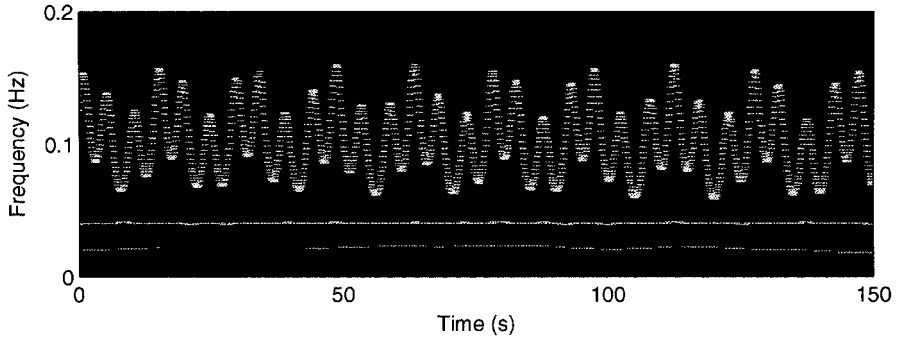

Figure 33. The Hilbert spectrum for the Duffing equation solution showing a strongly intrawave frequency modulated oscillation and two low-frequency oscillations. The intermediate frequency component is the forcing function.

Although the forcing function and the intrinsic frequency are still discernible as wide belts of energy concentration, all the details of the intrinsic frequency variations are totally lost.

To show that the highest frequency IMF is indeed intrawave modulated, let us consider the following mathematical model:

$$X(t) = \cos(\omega t + \epsilon \sin 2\omega t). \tag{9.7}$$

According to the classic wave theory, this expression is a clear case of intrawave frequency modulation similar to that of the Stokes waves. Using the same expansion as in equation (8.6), one can show that equation (9.7) can be reduced to

$$X(t) \approx (1 - \tfrac{1}{2}\epsilon)\cos\omega t + \tfrac{1}{2}\epsilon \cos 3\omega t + \cdots, \tag{9.8}$$

which is the second-order approximation of the series solution for the Duffing equa-

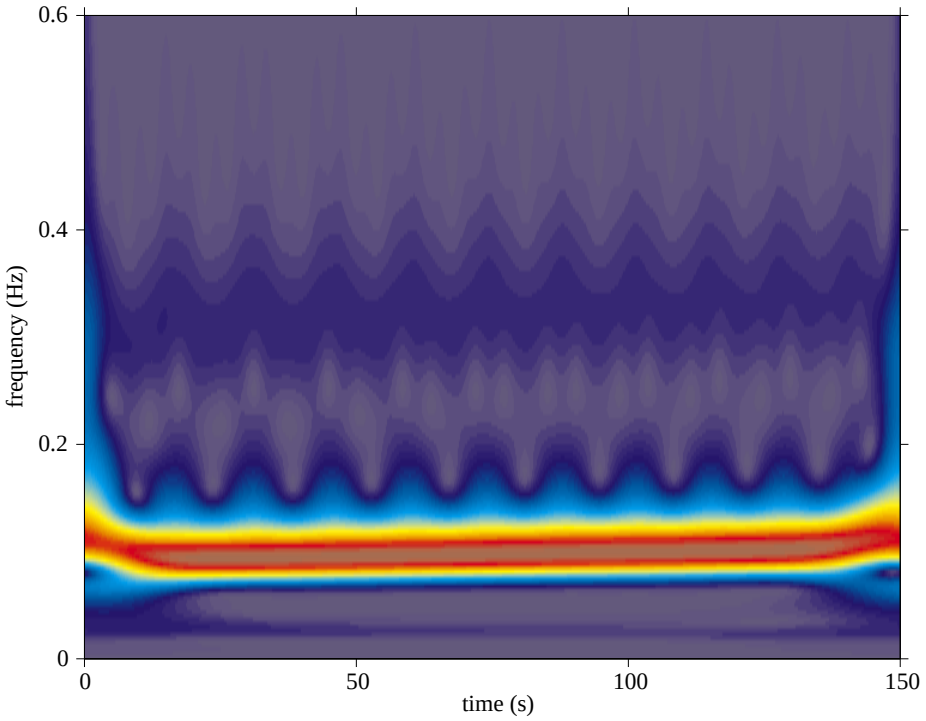

Figure 34. The Morlet wavelet spectrum for the Duffing equation solution showing a rich distribution of harmonics but no details of the intra- or interwave frequency modulations. Notice the scale of this figure. If the Hilbert spectrum is plotted on this scale, all the information is confined to the region between 0.05 and 0.15 Hz.

tion. For $\epsilon = 0.3$ and $\omega = \frac{1}{64}$ Hz, the wave-profile represented by equation (9.7) is given in figure 35 with the frequency computed from either the Hilbert transform through equation (3.4) and the classic wave theory. The differences are caused by the leakage of energy into the negative frequency range, but the overall agreements are still clear. Here again we have two views for the same mathematical expression. The one based on the classic wave theory is the more physical one. It calls for an intrawave frequency modulation.

Now, let us use the Hilbert spectrum and the wavelet analysis methods on this expression. The results are shown in figures 36 and 37, respectively. The Hilbert spectrum clearly shows the intrawave modulation, in agreement with the classic wave theory, while the wavelet result is encumbered by the harmonics. Thus, we prove again that intrawave frequency modulation is physical. This simple example again illustrates that the instantaneous frequency, with intrawave frequency modulation defined by the EMD and Hilbert spectrum, does make physical sense. In fact, such an instantaneous frequency presentation actually reveals more details of the system: it reveals the variation of the frequency within one period, a view never seen before. Again, the wavelet result spreads energy over a much wider frequency range, but fails to give any true detailed frequency variations.

Based on the new idea of intrawave frequency modulation, one can easily pick out the minute variations of the waves in the IMF component and in the original data computed from the Duffing equation. With the Hilbert spectrum as a guide,

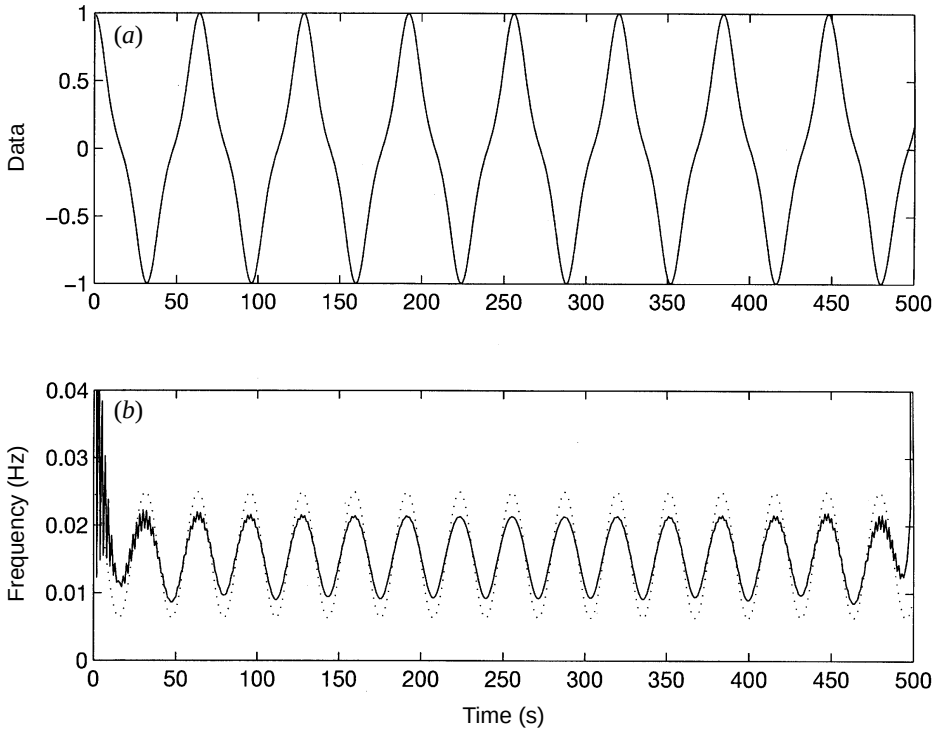

Figure 35. An intrawave frequency modulated Duffing wave: (*a*) the wave-profile for the model based on equation (9.7) showing the distinct Duffing wave form; (*b*) the instantaneous frequency of the Duffing wave model based on Hilbert transform (solid line) and based on the classical wave theory and equation (3.4) (dotted line).

the unevenness of the intrawave frequency variation can be shown to have one-to-one correspondence with the variations of the data. This observation is consistent with the non-repeating phase function as in figure 31. In figure 33, a much lower frequency of the forcing function at 0.04 Hz is also clearly shown. Additionally, we can also see a faint suggestion of still lower frequency subharmonics reported by Kevorkian (1966).

(ii) *Lorenz equation*

The famous Lorenz equation (Lorenz 1963) was proposed initially to study deterministic non-periodic flow. It can be regarded as the starting point of the modern chaos investigation. The standard form of the Lorenz equation is

$$\dot{x} = -\sigma x + \sigma y, \quad \dot{y} = rx - y - xz, \quad \dot{z} = -bz + xy, \tag{9.9}$$

in which σ, r and b are positive constant parameters. The Lorenz system is dissipative; therefore, it is not a Hamiltonian system. The divergence of the velocity is

$$\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \frac{\partial \dot{z}}{\partial z} = -(\sigma + b + 1) < 0, \tag{9.10}$$

therefore, the volume of the fluid will eventually tend to zero, suggesting that the system has an attractor or attractors.

With $\sigma = 10$, $b = 3$, $r = 20$, this system has three stationary points: $(0, 0, 0)$, $(\sqrt{57}, \sqrt{57}, 19)$ and $(-\sqrt{57}, -\sqrt{57}, 19)$. From the initial position of $(10, 0, 0)$, the

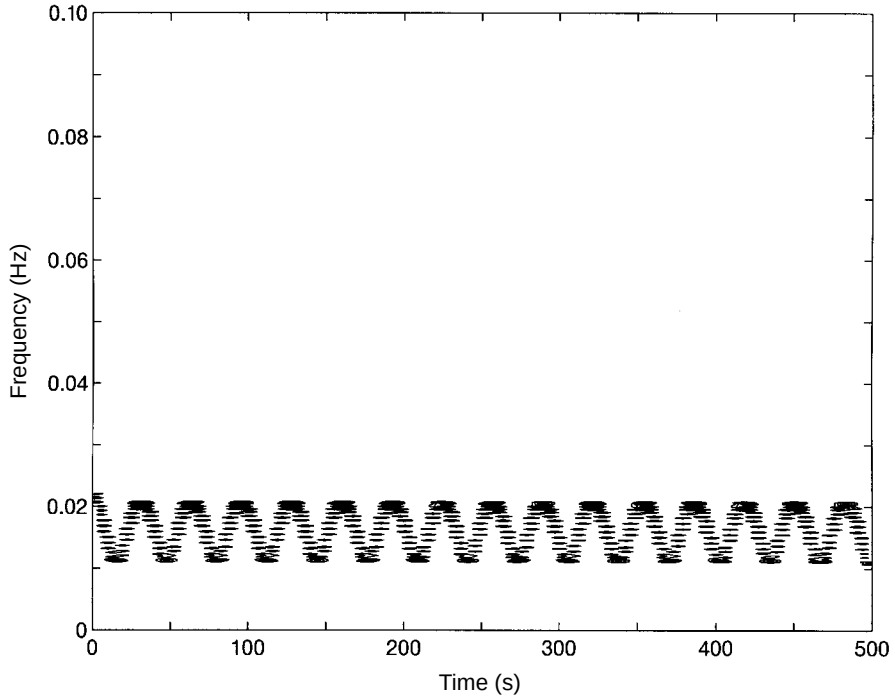

Figure 36. The Hilbert spectrum for the model intrawave frequency modulated wave, equation (9.7).

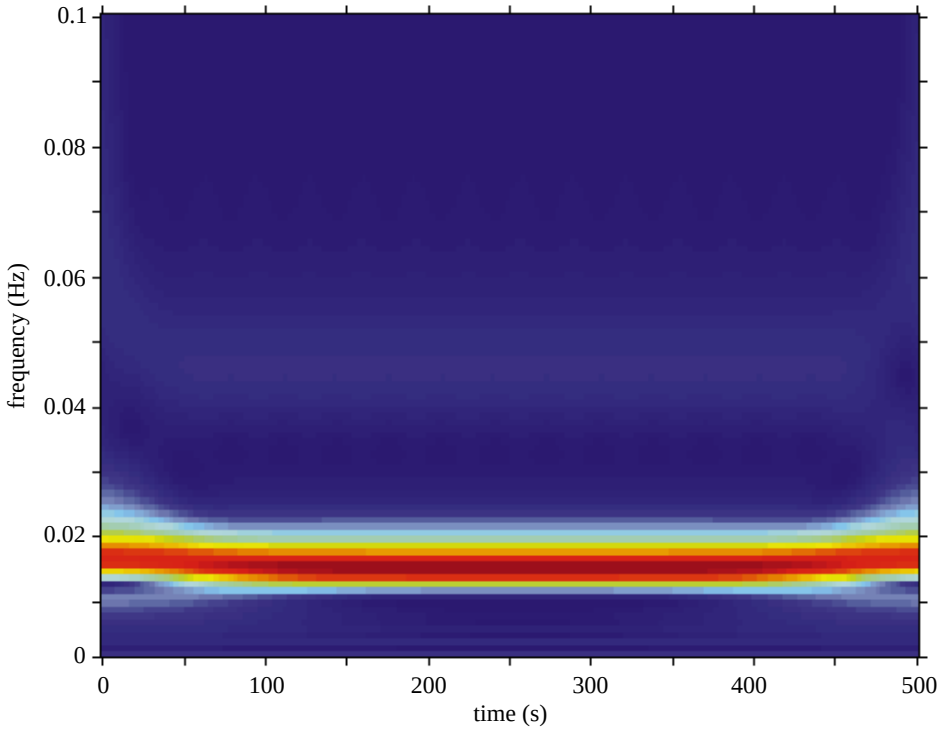

Figure 37. The Morlet spectrum for the model wave based on equation (9.7), in which the intrawave modulation is represented by the non-physical harmonics located at 0.05 Hz.

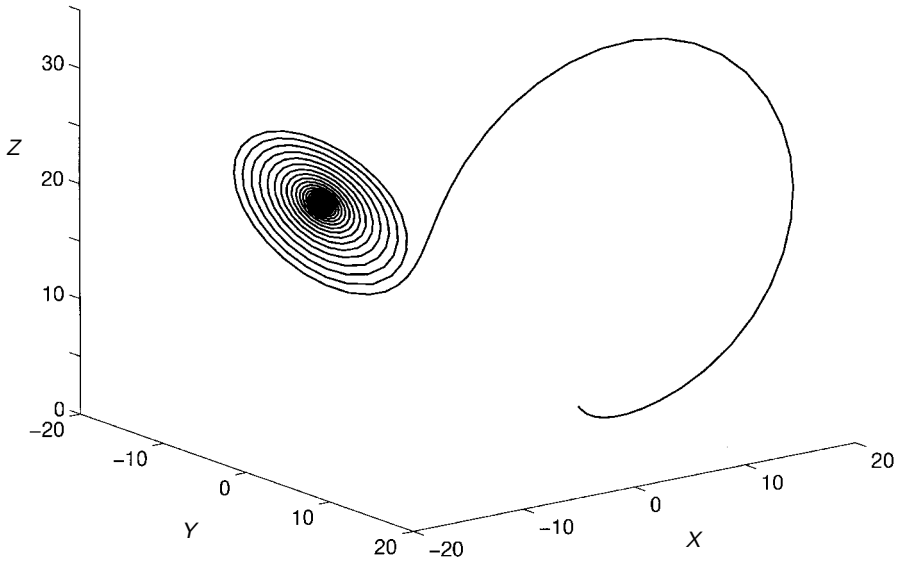

Figure 38. The three-dimensional representation of the numerical solution for the Lorenz equation with $\sigma = 10$, $b = 3$, $r = 20$ and the initial position of $(10, 0, 0)$.

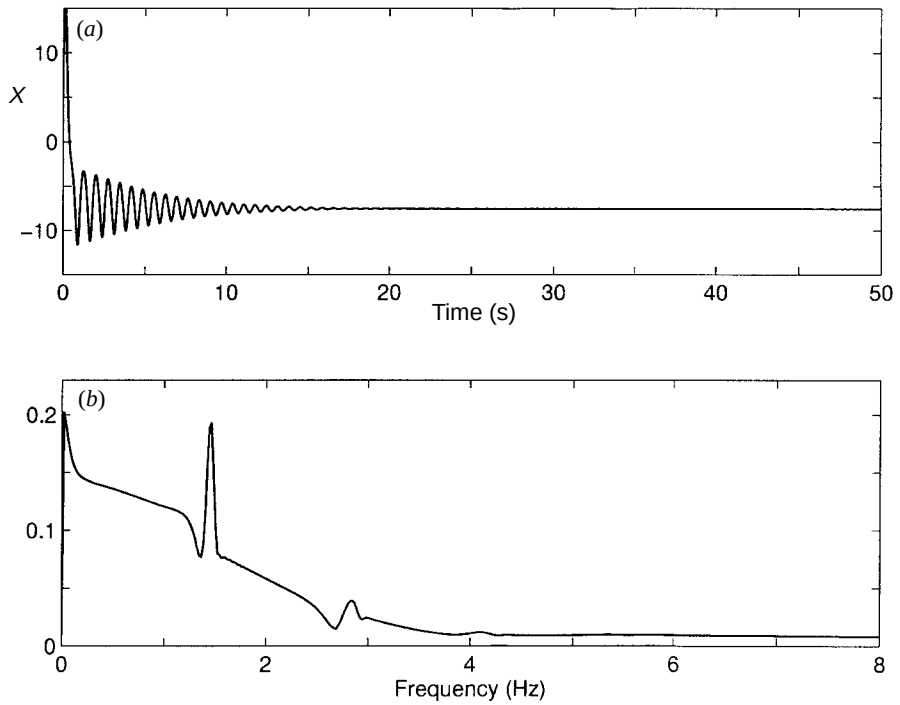

Figure 39. The numerical solution of the Lorenz equation: (a) the x-component of the numerical solution; (b) the Fourier spectrum for the x-component. Since the data are transient, the wide spectrum does not make physical sense.

solution depicts a spiral motion converging to one of the point attractors at $(-\sqrt{57}, -\sqrt{57}, 19)$ as shown in figure 38. The x-component and its Fourier spectrum are shown in figure 39. The Fourier spectrum shows a sharp peak at 1.4 Hz and

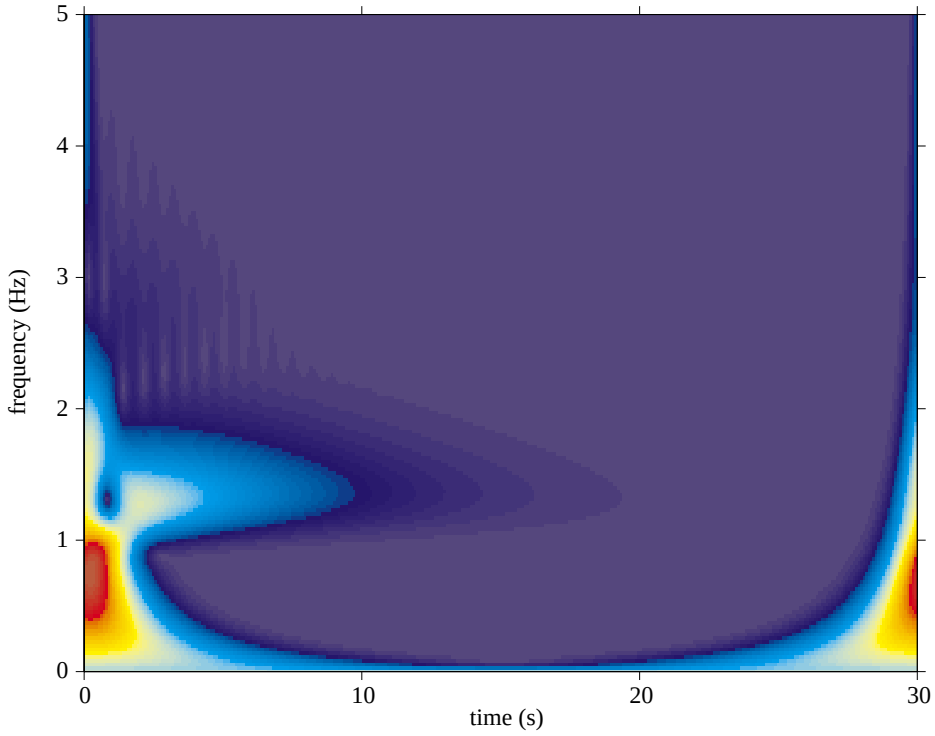

Figure 40. The Morlet wavelet spectrum for the Lorenz equation. The transient property, the harmonic, and the edge effect of the wavelet analysis are all clearly shown.

its harmonics at 2.8 Hz. Additionally, there are broad energy distributions in order to simulate the transient property of the signal. A Morlet wavelet analysis result of the same signal, shown in figure 40, reveals the transient property of the signal, the harmonic components and the edge effects. As discussed previously, the transient property of the data indicates that the Fourier spectrum is useless here.

The EMD, when applied to the signal, yields two IMF components: both transient, but with totally different frequencies and damping characteristics as shown in figure 41. The Hilbert spectrum is given in figure 42, in which we can see an oscillation with a frequency centred around 1.4 Hz and a low-frequency relaxation, again a surprise as in the Duffing system. All these variations can be seen through the linearized equation near $(-\sqrt{57}, -\sqrt{57}, 19)$, which gives the following eigenvalues: $-13.56 - 0.22 + 9.17\mathrm{i}$, and $-0.22 - 9.17\mathrm{i}$. Notice that $9.17/2\pi = 1.46$ Hz, which is the main oscillating frequency. The main frequency exhibits similar intrawave frequency modulation as in other nonlinear systems: the further from this equilibrium point, the stronger the nonlinearity. Although the amplitude variations can also cause some intrawave frequency modulation as shown in figure 21, the magnitude of such effects should be small in comparison to the nonlinear effects. In view of physical analogy with the Duffing equation as in equation (9.4), and the discussions on all the previous cases, we believe the frequency–time distribution in the Hilbert spectrum is more physical. The intrawave frequency modulations not only present a clear physical picture of the motion, but also eliminate the need for the spurious harmonics. Again the low-frequency component needs a better interpretation.

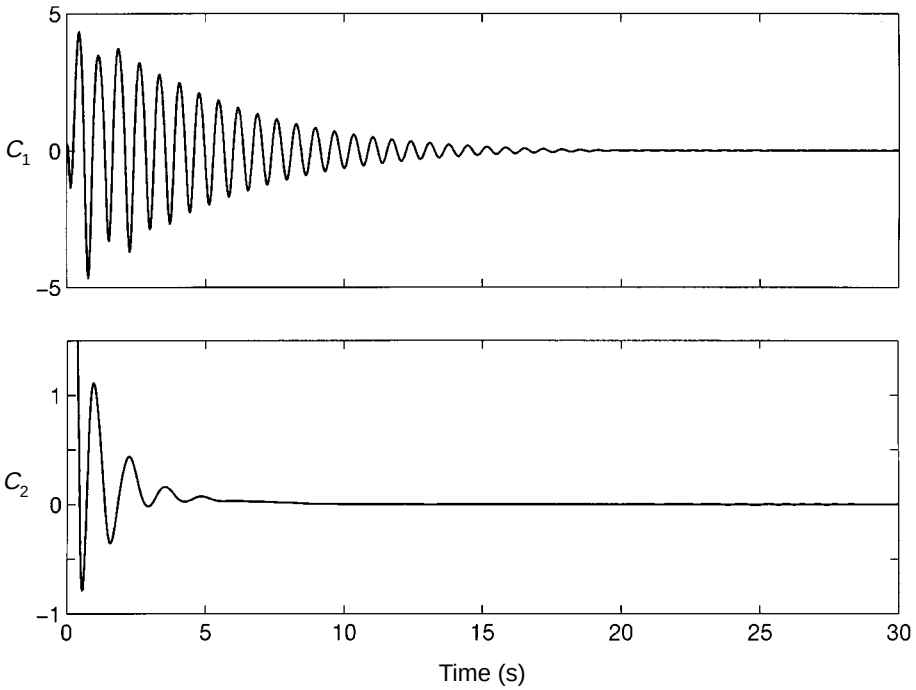

Figure 41. The two IMF components for the Lorenz equation solution. The solution, though, looks symmetric. It can actually be decomposed into two components with different frequency ranges.

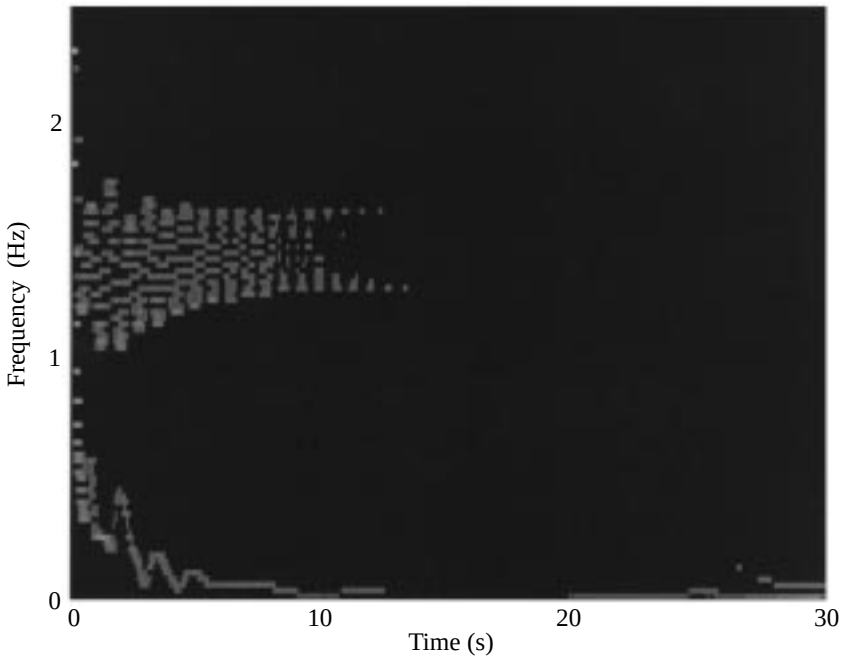

Figure 42. The Hilbert spectrum for the Lorenz equation solution. The transient nature of both components are clear. The main component is also intrawave modulated, a clear indication of the nonlinear effect of the oscillation.

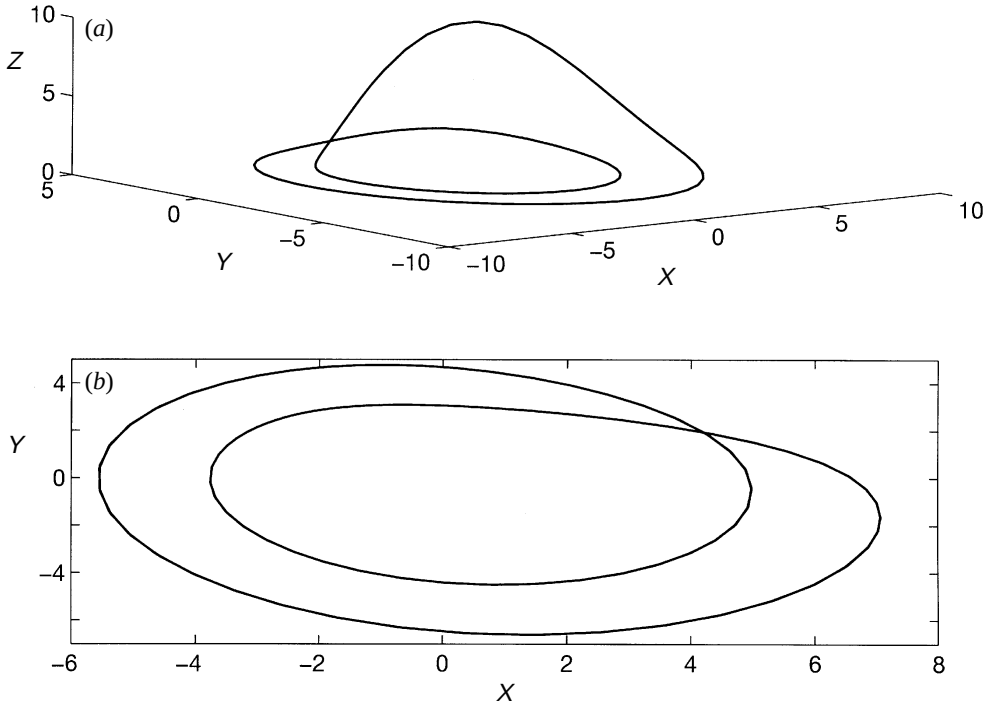

Figure 43. The numerical solution of the Rössler equation given in equation (9.11), for the period-doubling case: (*a*) the three-dimensional representation; (*b*) the two-dimensional representation.

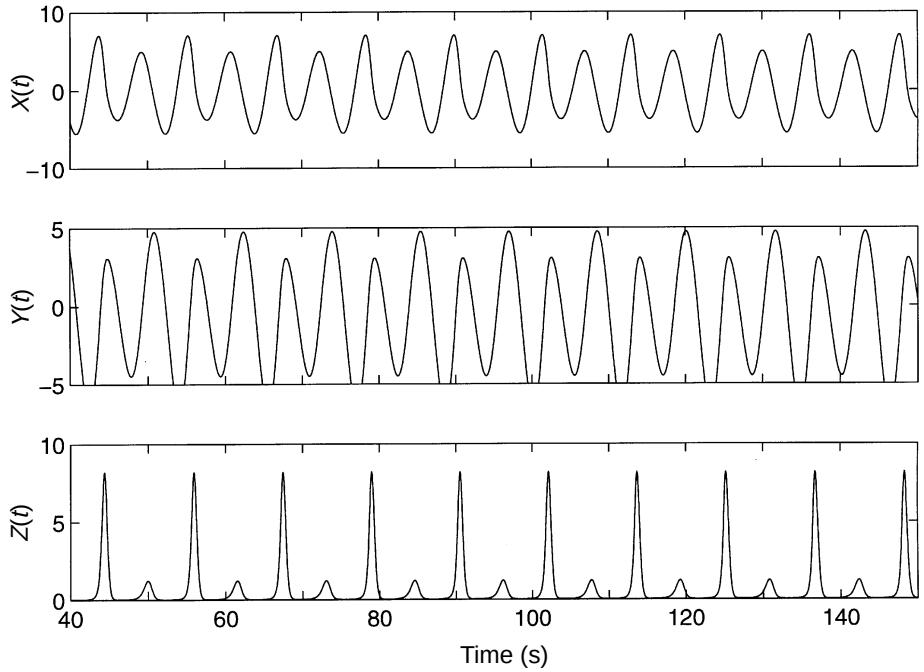

Figure 44. The time series for the numerical solution of the Rössler equation given in equation (9.11).

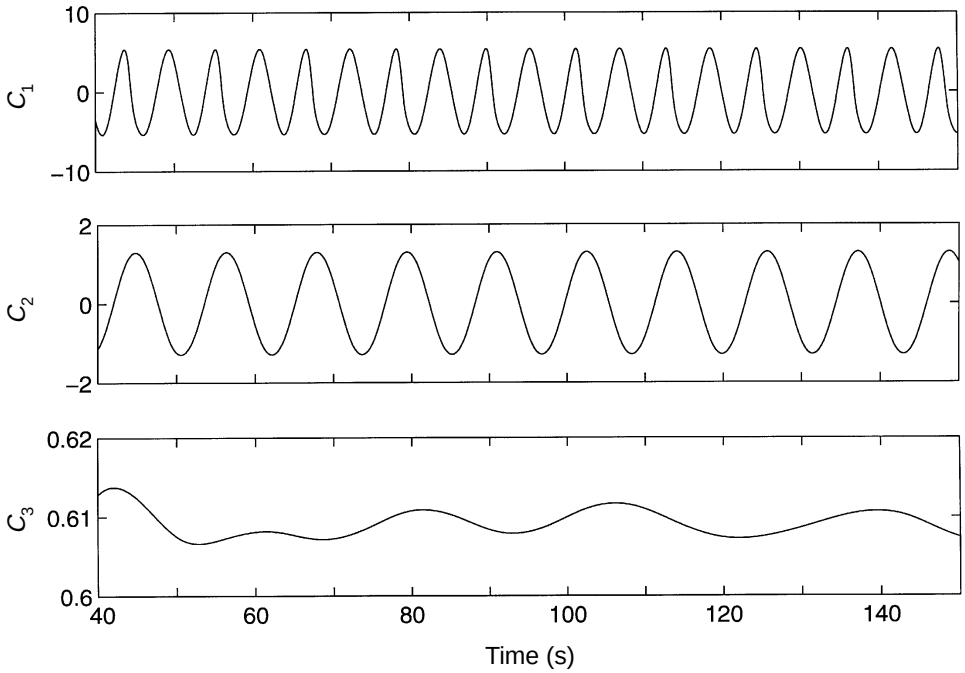

Figure 45. The IMF components of the x-components of the Rössler equation given in equation (9.11): component c_1 is non-uniform but the most energetic; c_2 is uniform; and the last component contains negligible energy.

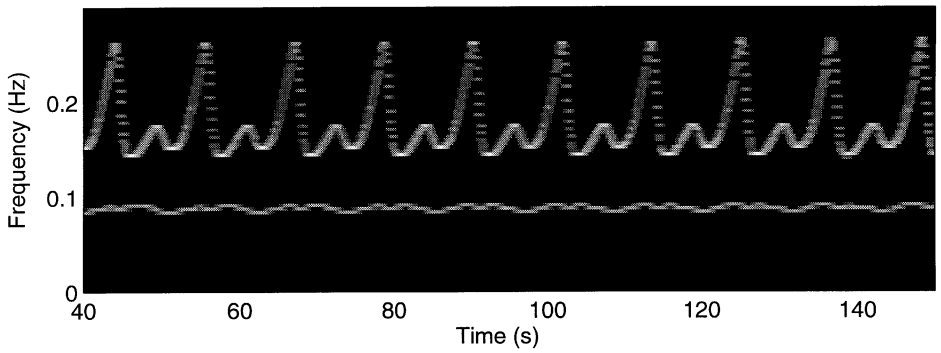

Figure 46. The Hilbert spectrum for the x-component. The most energetic component shows regular intra- and interwave frequency modulations. The frequency variations bear one-to-one correspondence with the time series data or the IMF component.

(iii) *Rössler equation*

The Rössler equation is given by the system,

$$\dot{x} = -(y + z), \quad \dot{y} = x + \tfrac{1}{5}y, \quad \dot{z} = \tfrac{1}{5} + z(x - \mu), \tag{9.11}$$

in which μ is a constant parameter. We will just study the regular case with $\mu = 3.5$, which represents the famous period-doubling event. Figure 43 presents the phase diagram; figure 44, the time series of the components. The wave form of the x-component is regular with an exact period of twice the high-frequency oscillation. The Fourier spectrum of x gives many harmonics, the characteristics of the Rössler system. Applying the EMD method to the data, we have only three IMF components as

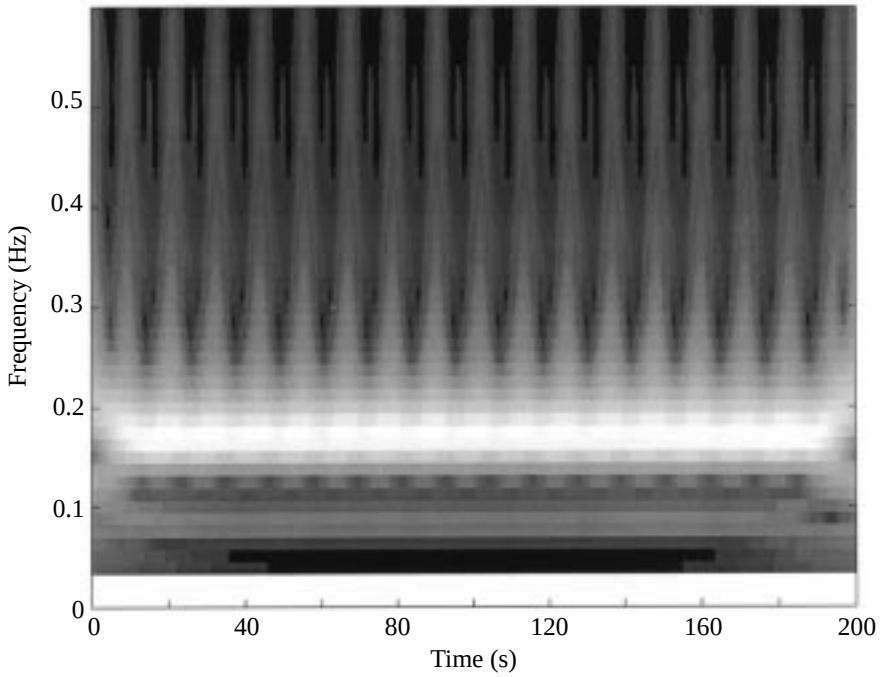

Figure 47. The Morlet wavelet spectrum shows only harmonics and some indication of interwave frequency modulation. There are no details for the intrawave frequency modulations.

shown in figure 45: a seemingly uniform high-frequency component; an intermediate frequency component; and a low-intensity low-frequency component with non-zero mean. As this system contains exactly two time scales, this extra low-frequency component must be generated by numerical error. With the energy level only 10^{-6} times that of the fundamental oscillation, its existence here has no dynamic consequence. The remaining two periods are exactly once and twice that of the fundamental one.

The corresponding Hilbert spectrum, shown in figure 46, contains only two prominent components: the high-frequency component fluctuating widely between 0.15 and 0.26 Hz, and the steady low-frequency band at 0.1 Hz. As shown in figure 46, the fluctuating high-frequency component is strictly periodic with two basic wave periods as the period doubling requires. Still, within the two wave periods, the frequency undergoes intrawave frequency modulation that corresponds to the alternative long and more rounded waves interlaced with the short and much sharper ones. With this as a guide, one can see this frequency change clearly from the first IMF component. Here, we have both intrawave and interwave frequency modulations. To represent such a case with either Fourier spectrum or wavelet analysis, one would need many harmonics as shown in figure 47. Notice again here the difference of the scales on the frequency axis in figures 46 and 47. All the interesting details pertinent to the frequency doubling can be seen clearly in the Hilbert spectrum, while the wavelet only gives a profuse display of harmonics with little physically significant detail. Thus, for the first time, the details of frequency variations have been determined quantitatively. No other methods can match the resolution power displayed here. The example also illustrates the advantage of not requiring the IMF components to be orthogonal: here the two components in the frequency-doubling event share many of their harmonics if expanded in the linear Fourier modes; then, the components are not

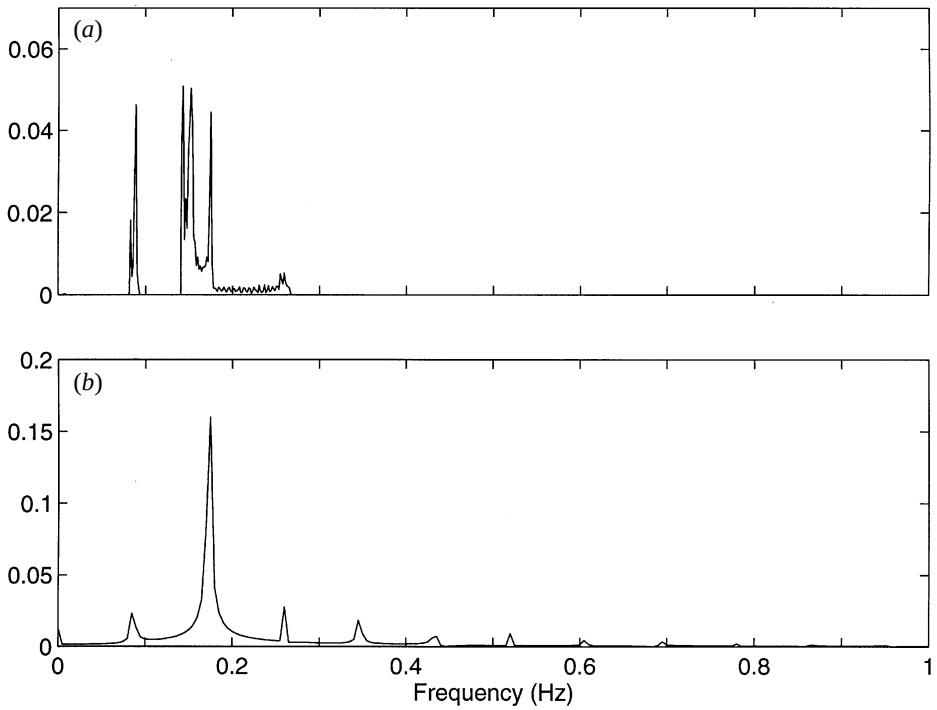

Figure 48. Comparison between the marginal spectrum and the Fourier spectrum for the Rössler equation: (*a*) the marginal spectrum showing the low-frequency oscillation near 0.1 Hz, and the high bimodal frequency distribution around 0.2 Hz generated by the intrawave frequency modulation; (*b*) the Fourier spectrum contains only harmonics with no indication of the two time scales involved in the period doubling. The main peak is the weighted mean of the physical time scales.

orthogonal. Yet, each component is physically distinct from the other with different characteristic time scales. Since EMD is based on the time scale and not on the insistence of orthogonality, it has overcome the shortcomings of the orthogonal Fourier expansion, and successfully decomposed the data into two nonlinear components.

Another interesting comparison of the frequency contents from different representations is in the marginal spectrum and the Fourier spectrum given in figures 48*a*, *b*, respectively. The marginal spectrum gives a low-frequency peak and a broad bimodal distribution near the main peak frequency, suggesting a typical distribution of a periodic variable. None of the peaks in the marginal spectrum agree with the main peak in the Fourier spectrum, at 0.18 Hz and its harmonics. The 0.18 Hz oscillation given by the Fourier spectrum is neither the observed nor the physical frequency; it is the weighted mean of the broad band of interlaced long and short waves. This example illustrates that even the main Fourier peak could be non-physical, an unexpected but shocking revelation.

(*b*) *Observational data from laboratory and field experiments*

The main applications of the present method are to the time series analysis. In this part, we will present a variety of data to illustrate the usefulness of the combined EMD and Hilbert spectrum applied to controlled and natural phenomena.

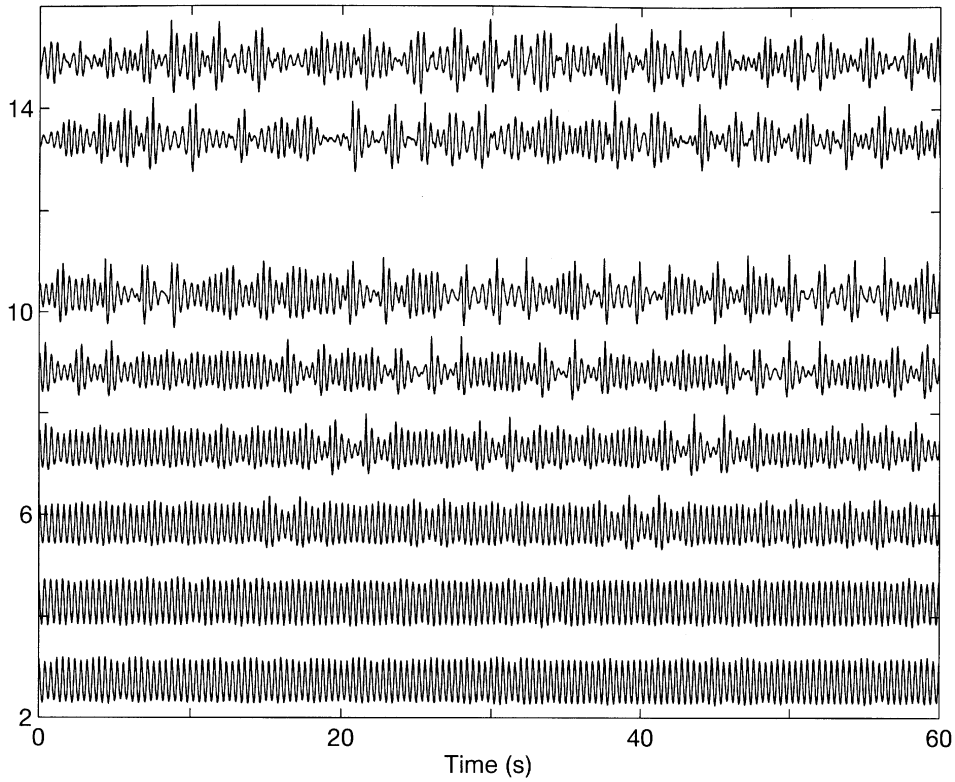

Figure 49. The surface elevation data from a mechanically generated regular sinusoidal wave train with initial frequency 2.5 Hz. The horizontal axis is time, the vertical axis is the distance from the wave maker in metres. As the data show, the initial regular wave soon becomes highly modulated as the wave moves down the tank.

(i) *Laboratory wave data*

Laboratory data from mechanically generated waves were used by Huang *et al.* (1996) to study the nonlinear evolution processes. Under the action of weak nonlinear interactions, the frequency of the waves will downshift as they propagate, a process necessary for the waves to become longer and grow higher under the wind. In the narrow band wave field, the downshift has been shown as the consequence of the well-known Benjamin–Fier instability (Benjamin & Fier 1967). The process is assumed to be gradual and continuous. With the resolution power of the present method, we want to show that the wave evolution process is not continuous and gradual but local and discrete; that the well-known frequency downshift is an accumulation of discrete fusion of n to $(n-1)$ waves, also known as 'crest pairing' (Ramamonjiarisoa & Mollo-Christensen 1979), and 'lost crest' (Lake & Yuan 1978). We will present some results reported by Huang *et al.* (1996) to illustrate the application of the Hilbert spectral analysis.

The data were collected in the wind-wave tank at NASA Wallops Flight Facility (Huang & Long 1980; Huang *et al.* 1996). Surface elevations were recorded at eight stations along the tank for a mechanically generated sinusoidal wave. Huang *et al.* (1996) used the Hilbert transform and examined the change of the phase of the wave trains. The waves are mechanically generated by a wave maker driven at 2.5 Hz with the raw data of the test for all eight stations given in figure 49. After performing the

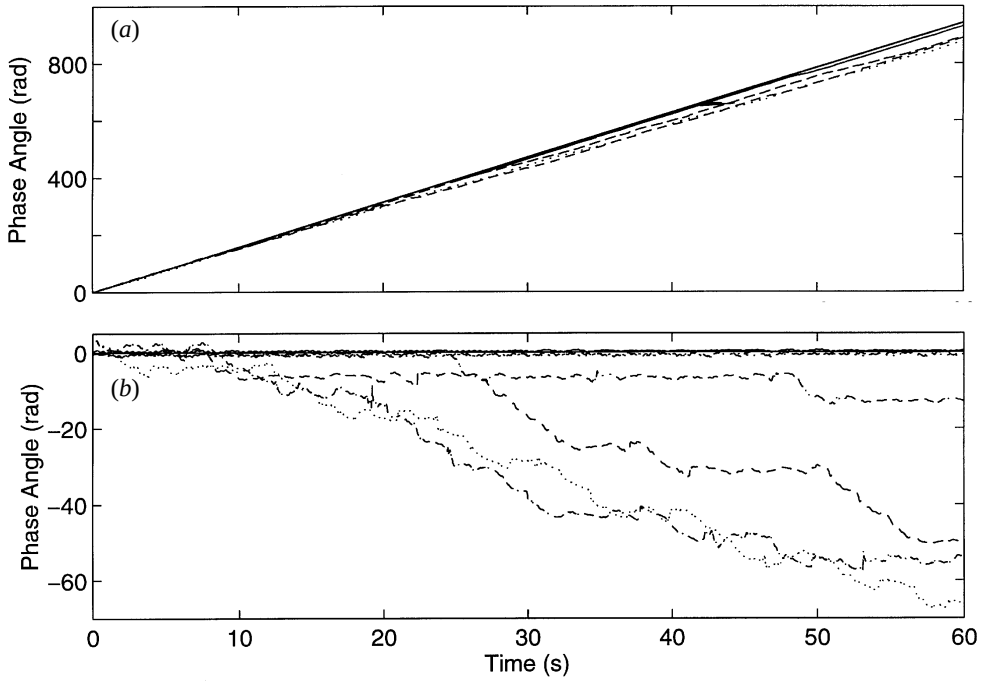

Figure 50. The phase function of the wave records from all eight stations. (*a*) The unwrapped phase function showing the phase function seems to have evolved gradually down the tank. The highest values are for the first four stations; the first time the phase deviates from the value of the first four stations is at station 5. (*b*) The phase function change relative to the values at station 1. All the changes are shown to be concentrated at a certain narrow region, while the phase function values remain unchanged at all other points. The amount of the changes stay at a discrete value, 2π.

Hilbert transform on the data, the unwrapped phase functions for all eight stations are shown in figure 50*a*, which indeed suggests the gradual change of the phase function. The decrease in the slope of the phase function indicates the decrease of the mean frequency. The slopes of the first four stations, however, stay almost identical to that of the first station. The first phase function that shows any deviation from this initially closely clustered group is at station 5. To this scale, the frequency change seems small and gradual. To examine the variation in more detail, we take the difference of the phase functions of all stations with respect to the first station, which is our initial condition. The difference is shown in figure 50*b*, from which we can see that the changes of the phase functions are the results of a series of steps with the sharp jumps confined in very short time spans.

To quantify these phase variations, we plot the phase change with respect to that of the first station in phase–amplitude diagrams in figure 51 for stations 2, 4, 5 and 6. The phase–amplitude diagram for station 5 shows that the phase starts to jump, which reflects the sudden shift of the phase function shown in figure 50*b*. These phase changes are very similar to the phase dislocations. As the number of jumps increases, the phase function seems to become a continuous and almost smoothly sloped line. But even for those cases, the amount of phase change involved in each jump, as revealed by the amplitude–phase diagram, is still constant for all the phase jumps: 2π. The 2π jump means a loss of exactly one wave in the process.

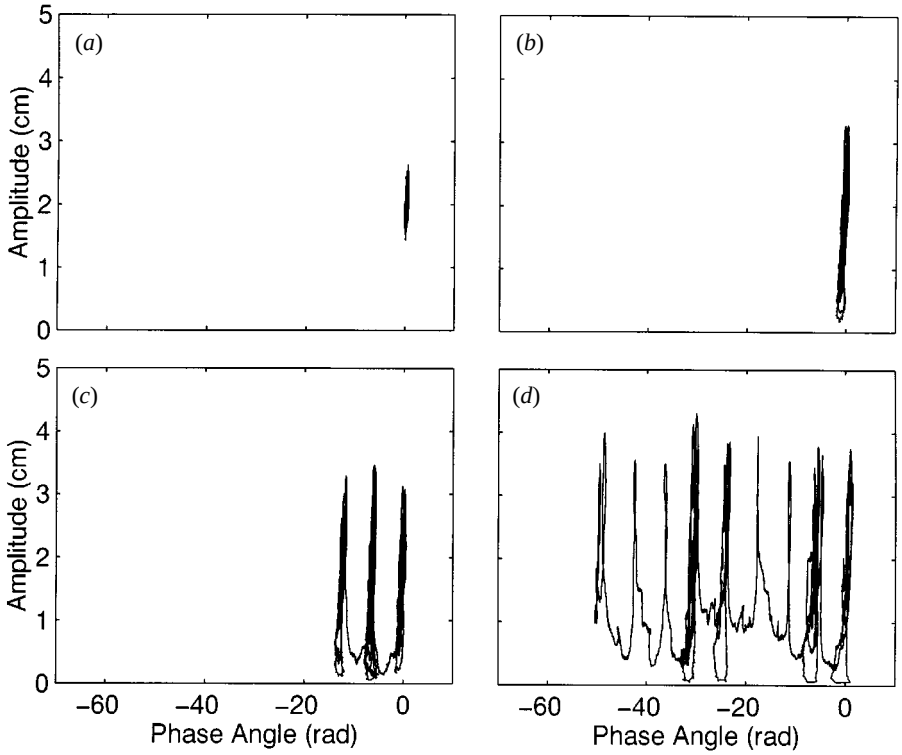

Figure 51. The phase-amplitude diagrams for stations 2, 4, 5 and 6. (*a*) Station 2, either the amplitude or the phase have large variations in their values. (*b*) Station 4, the amplitude variation becomes large, yet the phase stays constant. (*c*) Station 5, both amplitude and phase undergo large changes. The phase, however, stays only at discrete values, the multiples of 2π. (*d*) Station 6, the phase values continue to expand, but always stay at the discrete jumps.

Let us now examine the wave elevation data in detail at the jump points. The time series data are expanded and displayed in figure 52 in a two-way comparison. When we superimpose the data from station 5 with its phase difference from station 1, we find the jump in the phase is easily associated with a single wave within one wave period. Then we superimpose the raw elevation data from station 5 on those from station 1 also shown in figure 52. One can immediately see that all the wave peaks line up except at the location of the jump, where two waves become one; there is a loss of one wave in each phase jump event. All the changes are local, abrupt and discrete.

The Hilbert spectrum is shown in figure 53, in which one can identify very local events of abrupt change of frequency at the 25 and 48 s locations. The Hilbert spectrum reveals the typical intrawave frequency modulations of the Stokian water waves and the more striking local shifts of frequency to a lower value at the locations coinciding with the phase jumps shown in figure 50*b*. Another interesting phenomenon shows up in the marginal spectrum given in figure 54*a*, where there is a much wider frequency distribution than expected caused by the intrawave frequency modulations. Furthermore, there is also a slight tendency for energy being shifted to the subharmonics frequency. Whether this energy flux is related to the subharmonic instability studied by Longuet-Higgins (1978) deserves further investigation. The Fourier spectrum, given in figure 54*b*, shows only sidebands and harmonics without

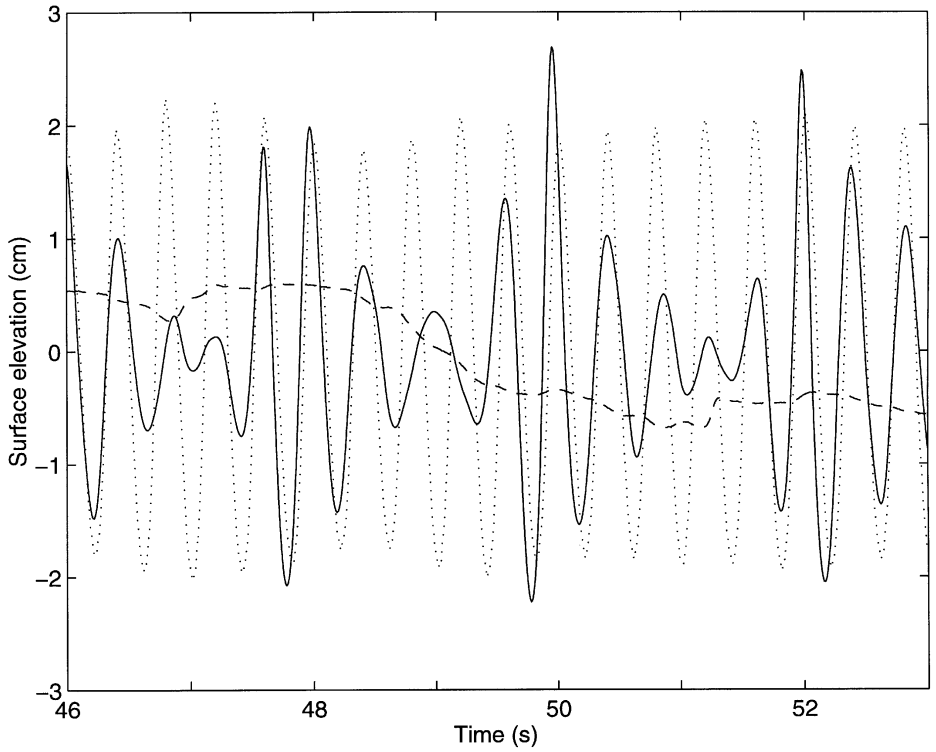

Figure 52. A two way comparison of the wave-profile change at the location of a phase jump. Surface elevation data at station 5 is given in the solid line; the data at station 1, in the dotted line serving as a time scale; and the phase change for station 5, in the dashed line. At around the 49 s location where the phase value jumps, the wave at station 5 has a fusion of two waves into one.

any suggestion of the local nature of the wave evolution. Again, the Morlet wavelet analysis from the same data, shown in figure 55, fails to resolve any of the local changes.

Using the Hilbert transform, we can examine the evolution of the weakly nonlinear wave trains in detail. Although the processes have always been assumed to be globally stationary, the Hilbert transform reveals that the processes are actually abrupt and local. Such variations are locally inhomogeneous; thus, they cannot be analysed adequately by Fourier analysis. Furthermore, the local variations are abrupt and discrete, having the typical characteristics of particles. The frequency downshift is a cumulative effect of fusions of two waves into one, or three waves into two, etc. The fusion processes are irreversible. This asymmetry is due to the preferred direction of downward energy flow. Our results also raise concerns over the past assumptions on gradual and continuous variations of the wave number and frequency of a wave train. With the discrete jumps and the particle-like nature of the individual wave, there may be a need for a new paradigm to represent the abrupt and discrete changes analytically. It will be a new challenge for nonlinear wave study.

Before leaving this section, let us also examine the data from station 1 to illustrate the effects of the nonlinear distortion of the water-wave surface. Because the emphasis is on the nonlinear distortion, we have first smoothed the data by a weighted running average so that the small irregularity will not cause mode mixing. A selected section

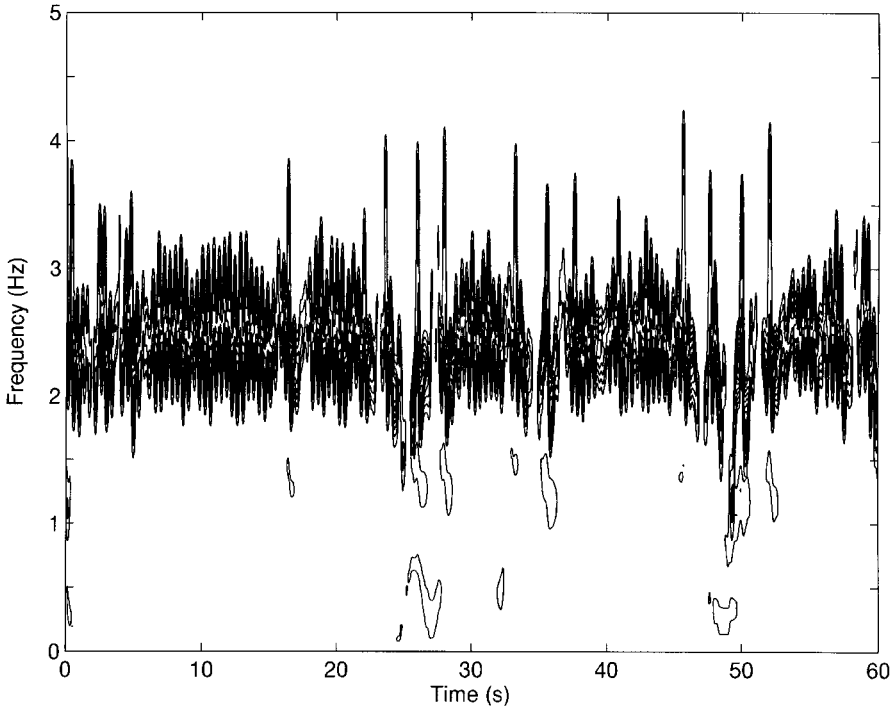

Figure 53. The Hilbert spectrum for the wave data at station 5. Other than the two locations of phase jumps, the frequency of the wave train stays at a very narrow range of intrawave modulation, and could be considered as having gradual and continuous frequency changes. But at the phase jump locations, the frequency of the wave train undergoes so much change as to make the assumption of a gradual and continuous function invalid.

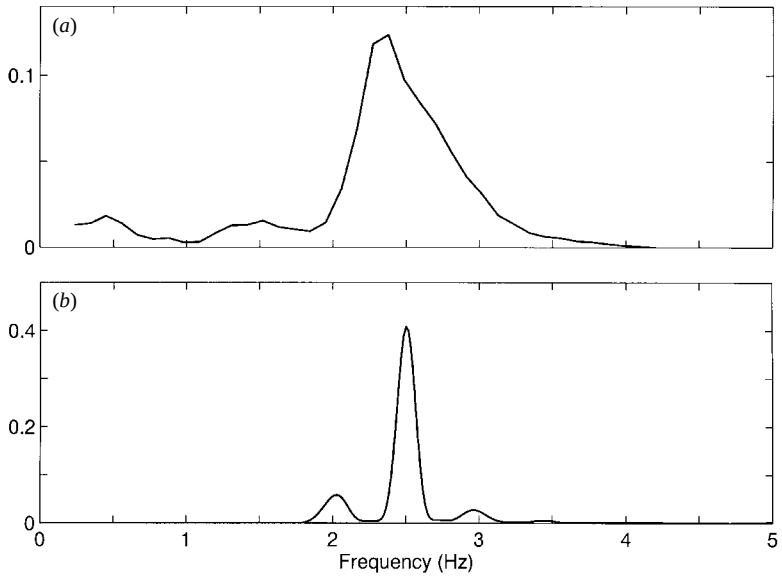

Figure 54. Comparison of the marginal spectrum and the Fourier spectrum at station 5. (*a*) The marginal spectrum for station 5 showing a relative wide distribution of frequency, and a slight tendency for energy to move into the subharmonics. (*b*) The Fourier spectrum for station 5 showing only the narrow fundamental frequency and the sidebands.

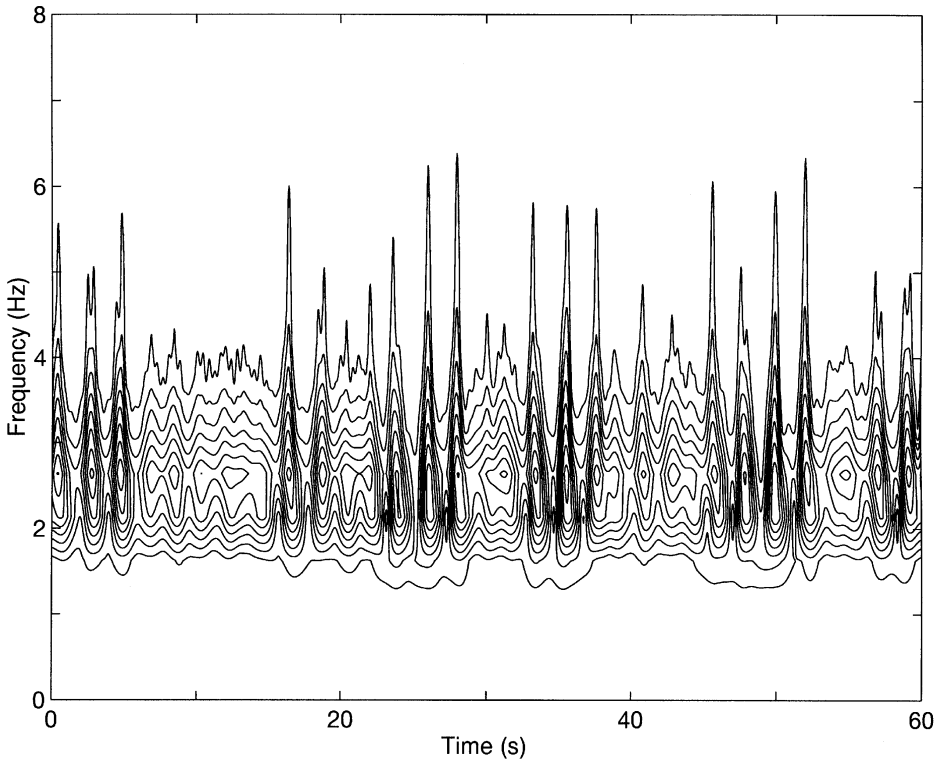

Figure 55. The Morlet wavelet spectrum for the data from station 5. The locations of the frequency changes have to be inferred from the high-frequency ranges, yet the changes are towards lower frequency by fusion. This counter-intuitive interpretation and the quality of the results make the wavelet spectrum hard to interpret.

of the data (5 s in length) and their Hilbert and wavelet spectra are given in figures 56 and 57, respectively. In figure 56, one can see the intrawave frequency modulation as shown in the model equation (8.5). The frequency fluctuates within a very narrow range around 2.5 Hz, the frequency of the wave maker. In general, the high local frequency values line up with the peaks, and the low local frequency values line up with the troughs as expected of the Stokes waves. In fact, it has been shown by Huang *et al.* (1990*a*, *b*) that the gravity water wave can be approximated well by the Stokes model. On more detailed examination, however, the alignment is not perfect: there is a slight but systematic shift of the phase toward the wave front, an indication of the front–back asymmetry of the wave-profile. In wavelet results shown in figure 57, one can no longer see any intrawave frequency modulation. Instead, there are harmonics again that line up with the wave front, those alignments confirm the Hilbert spectrum result: the waves have front and back asymmetry. In the wavelet spectrum, the energy containing frequency range is much wider than that in the Hilbert spectrum, an indication of energy leakage of the wavelet analysis. This comparison illustrates that the nonlinear wave distortion can indeed be explained by the much more physical interpretation as the intrawave frequency modulation.

Before ending this section, a few words on the Fourier representation of the wave field are necessary. Because the Fourier expansion of the wave field is so natural, it has been deeply associated with wave studies, so much so that thinking in terms of

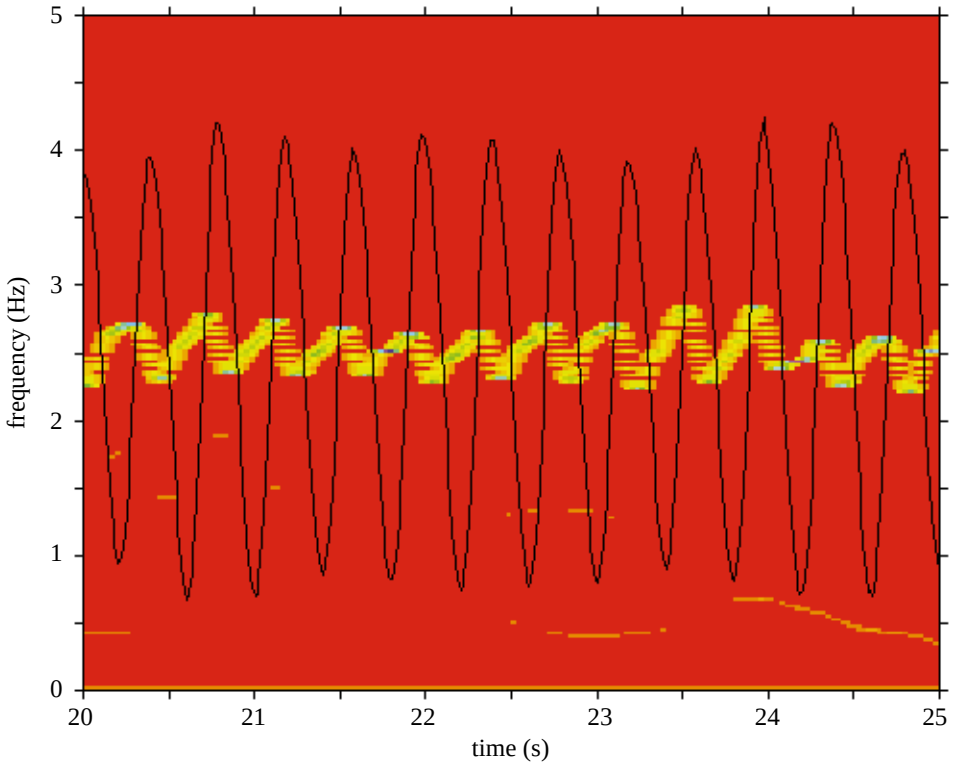

Figure 56. The Hilbert spectrum for a selected 5 s at station 1 with the surface elevation data plotted on an arbitrary scale. The Hilbert spectrum shows the intrawave frequency modulation of the water waves, an indication of the nonlinear surface distortion similar to the Stokes wave. But detailed comparison shows that there is a phase shift in the real data: the front of the wave is distorted more than the back even without wind, a physical process failed to be modelled by the Stokes wave.

Fourier components becomes a matter of fact. But these harmonics are not physical, they are not the solutions to any nonlinear equations. Furthermore, it is well known that a monochromatic wave is unstable; one cannot assume a collection of such components to exist for all the time everywhere to synthesize the whole wave field to explain the physical phenomena. A new paradigm is needed.

(ii) *Field wave data*

The field data of waves were collected by the NOAA New Tidal Gauge, located at Duck North Carolina, at the high data rate of 1 Hz. Figure 58 shows the raw data for 60 min. This is typical ocean wave data from the field stations: random and almost 'statistically stationary'. In the past, these types of data were treated with Fourier analysis. In fact, the studies of the wave spectra from Fourier analysis have been a main subject of the wave research (see, for example, Huang *et al.* 1990*a*). These data, when subject to the EMD method, yield eight components as shown in figure 59, with the last component indicating the tidal variation. The Hilbert spectrum is given in figure 60 and the corresponding wavelet spectrum in figure 61. While both spectra show a energy concentration around 0.1 Hz, the wavelet spectrum gives a much more continuous distribution in time, and much wider spread of energy consisting primarily of the harmonics in the frequency axis.

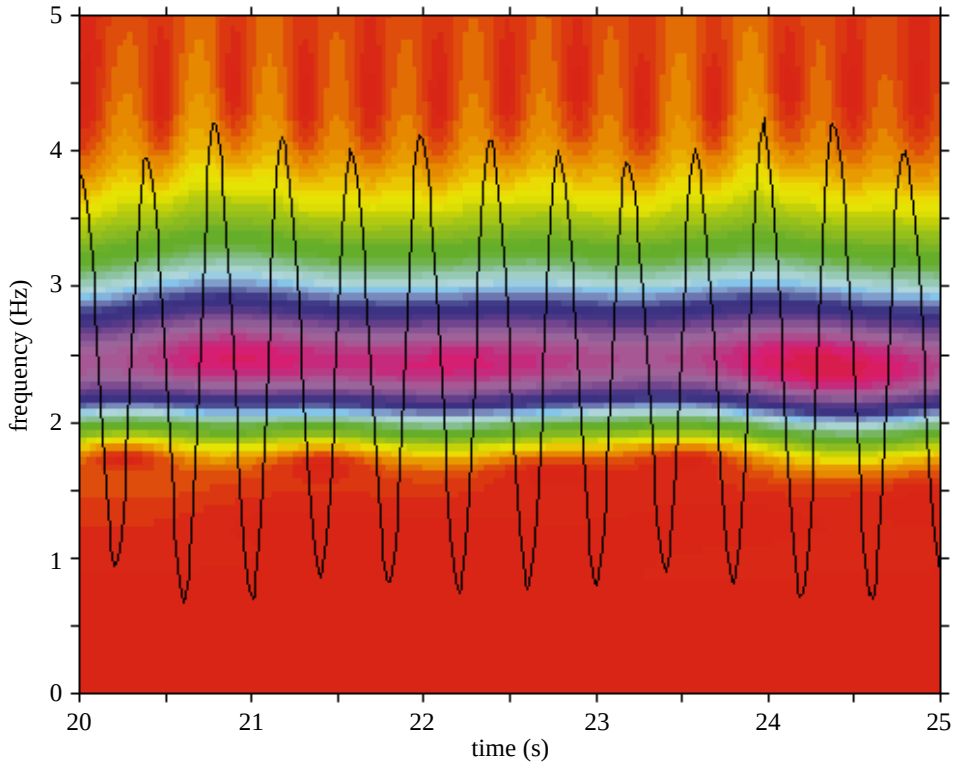

Figure 57. The Morlet wavelet spectrum of the same selected section as in figure 56. The wavelet spectrum fails to show the intrawave frequency modulation, but the leakage smears the energy over a much wider frequency range. The location of the harmonics, however, show the same phase shift towards the wave front as in the Hilbert spectrum.

To examine the results in detail, a short section of the data covering only 5 min is plotted in three ways. First, the Hilbert and the wavelet spectra are plotted separately with the corresponding surface elevation of the wave data in figures 62 and 63, respectively. In this expanded form, the Hilbert spectrum is very different from the wavelet spectrum. Then, the two different spectra are also plotted together with the Hilbert spectrum in contour lines superimposed on the wavelet spectrum in figure 64. In these presentations, they both show similar locations of energy concentration in time and frequency axes, but the Hilbert spectrum gives a sharper and more refined definition of the energy contour in figures 62 and 64. Take the data and the spectra near 57 min location for example: the data show a packet of high-amplitude low-frequency waves with frequency increasing and amplitude decreasing both before and after the packet. Both of these trends are vividly portrayed by the Hilbert spectrum, but they are only vaguely suggested in the wavelet spectrum. These variations suggested that the ocean waves are non-stationary, a conclusion supported by Huang *et al.* (1996).

Finally, let us also compare the marginal spectra from both the Hilbert and wavelet spectra with that of the Fourier spectrum in figure 65, in which the values of the spectra are staggered for clarity: the top line is the wavelet result; the middle one is the Hilbert spectrum; and the bottom is the Fourier spectrum. The contrast is similar to the cases discussed in the calibration and validation section: due to the

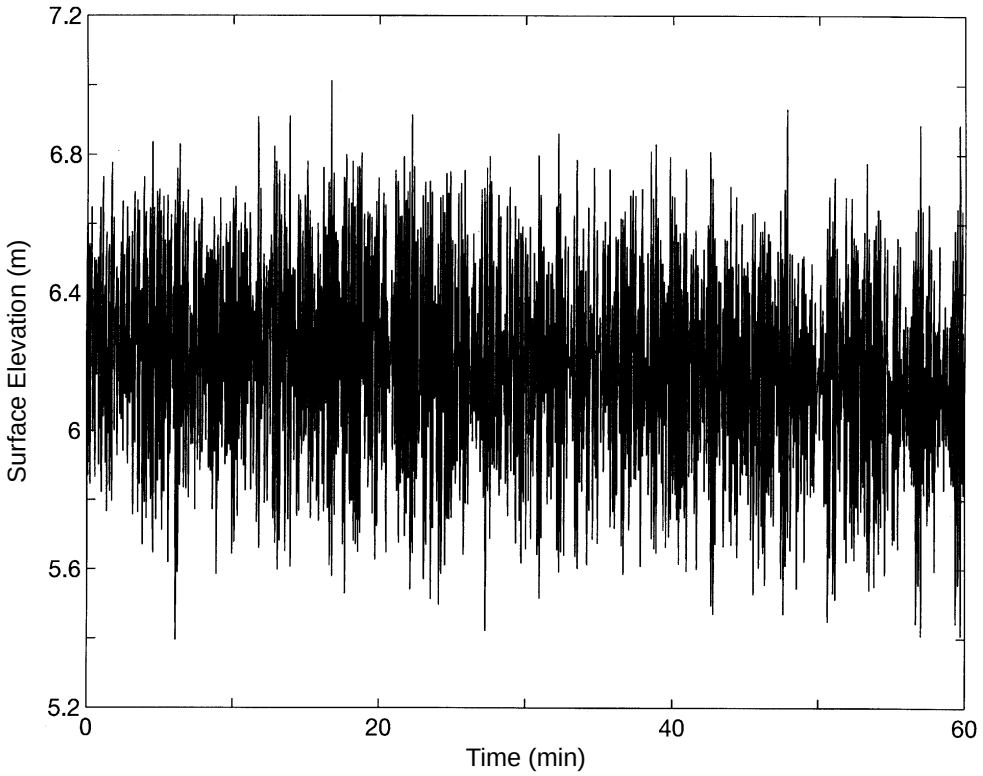

Figure 58. Field wave data measured at a tidal station covering a time span of 1 h. Because of the tides, the mean is not zero.

leakage, the wavelet spectrum is totally devoid of detail. While the Hilbert and direct Fourier spectra both show rich frequency contents, the lack of high harmonics due to either the nonlinear or the non-stationary effects in the Hilbert spectrum suggest that it can portray the energy–frequency presentation more precisely.

(iii) *Tide and tsunami data*

The field wave data selected here were also collected by the NOAA New Tidal Gauge. This gauge, located inside Kahului Harbour, Maui, is capable of recording data at 1 min steps. Figure 66 shows the raw data from the tidal gauge for five days from October 4–9, 1994. On October 5, tsunami-induced waves arrived at the site and created water level changes of a magnitude comparable to that of the tidal signal. Although the tidal data are traditionally analysed with Fourier expansion the added tsunami waves are transient. The combination, therefore, makes the whole time series non-stationary. Filtering cannot remove the tsunami signal cleanly, for the transient data and the tide will have many harmonic components in the same frequency range.

The EMD method yields eight IMF components from the data as shown in figure 67. Because the time scales of the tide and the tsunami waves are so different, the eight IMF components can be easily divided into two groups: the high-frequency signal representing the tsunami-induced waves; and the last three low-frequency components representing the tide. After the IMFs were used to reconstitute the two

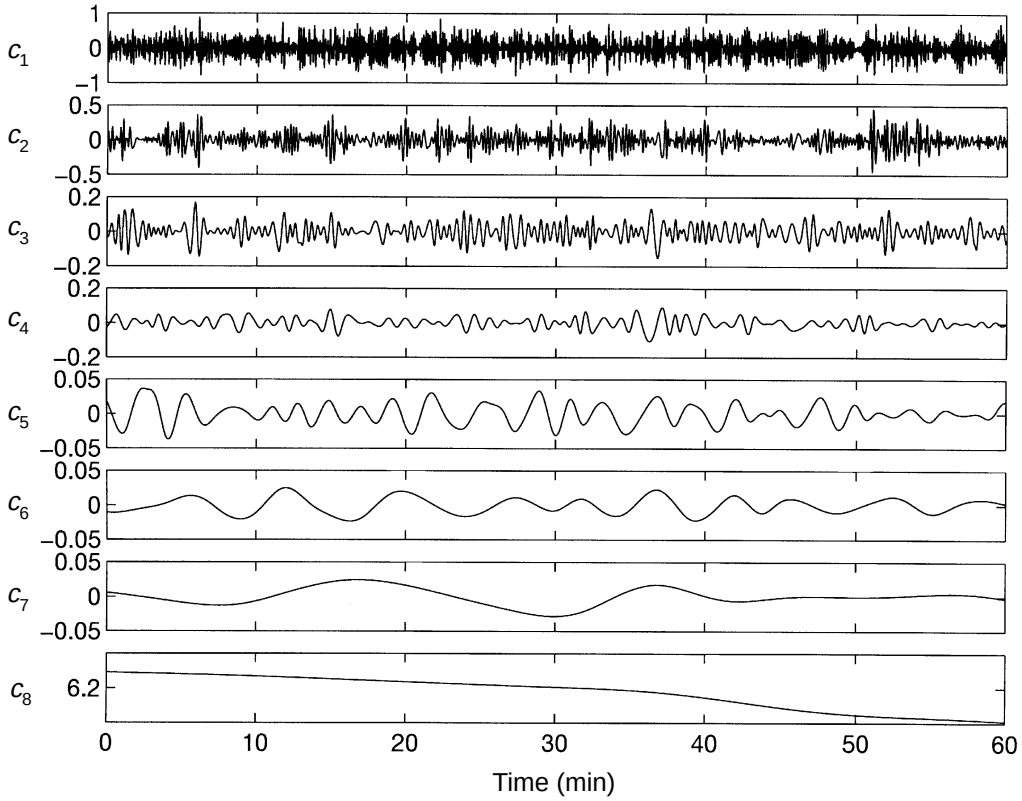

Figure 59. The IMF components derived from the data shown in figure 58: there are eight components with the last one showing the tidal range.

separate wave motions, the raw data, the tidal component and the tsunami-induced waves are plotted together in figure 68. Here, the EMD serves as a filter to separate the tide and the transient tsunami without any ambiguity.

Figure 69 gives the Hilbert spectrum from which the arrival time and the frequency change of the tsunami waves are clearly shown in this energy–frequency–time distribution. Besides the clear dispersion properties of the tsunami waves, there are two more interesting new observations: first, the variations of the tsunami wave frequency are phase locked with the tidal cycle; second, the tsunami waves in the harbour lasted many tidal cycles, with a frequency of half a cycle per hour. Such a period would not fit into the limited space inside the harbour of Kahului. The more likely explanation is for the tsunami waves being trapped in the bay, which has a much larger area for the wave to propagate. Details of the oscillation will be discussed in a separate paper.

In the above presentation, the tidal components are not emphasized. As a demonstration of the versatility of the Hilbert spectral analysis, the tidal components can be selected to construct the Hilbert spectrum. The result is given in figure 70, in which both the diurnal and semi-diurnal tides are clearly shown. In the marginal spectrum of the tidal components in figure 71, however, the frequency for the semi-diurnal tide is much more variable, especially at the time when the tsunami had just arrived, and towards the end of the measurement period. This variability seen in the marginal spectrum raises an interesting problem: whether there is interaction

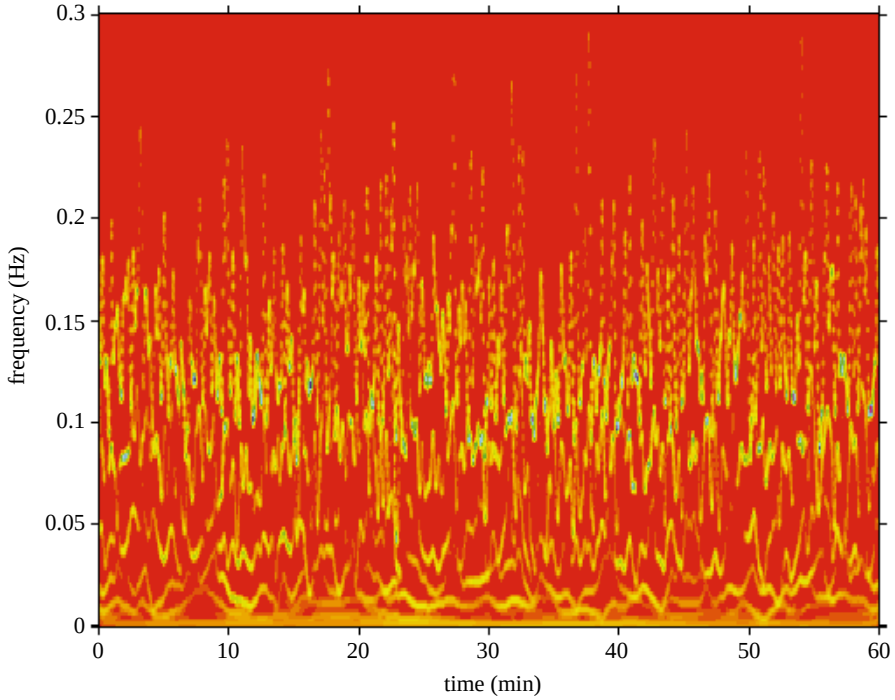

Figure 60. The 9×9 smoothed Hilbert spectrum for the data given in figure 58. The spectrum is extremely nodular, an indication that the wave is not stationary.

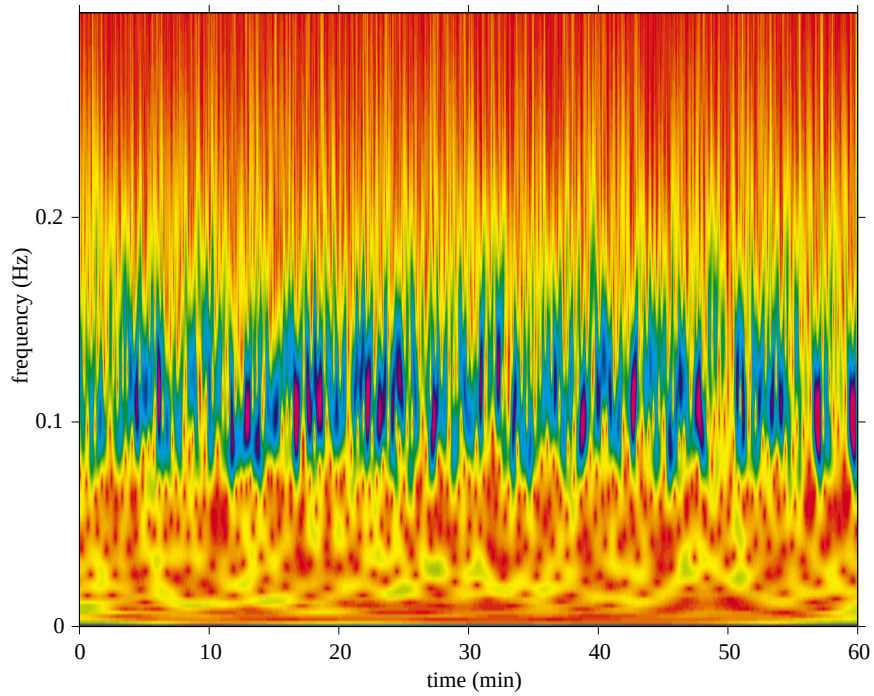

Figure 61. The Morlet wavelet spectrum of the data given in figure 58. Though the spectrum is more continuous, the energy distribution is still a function of time. Furthermore, there is a wider energy smearing in frequency space by the leakage and the harmonics.

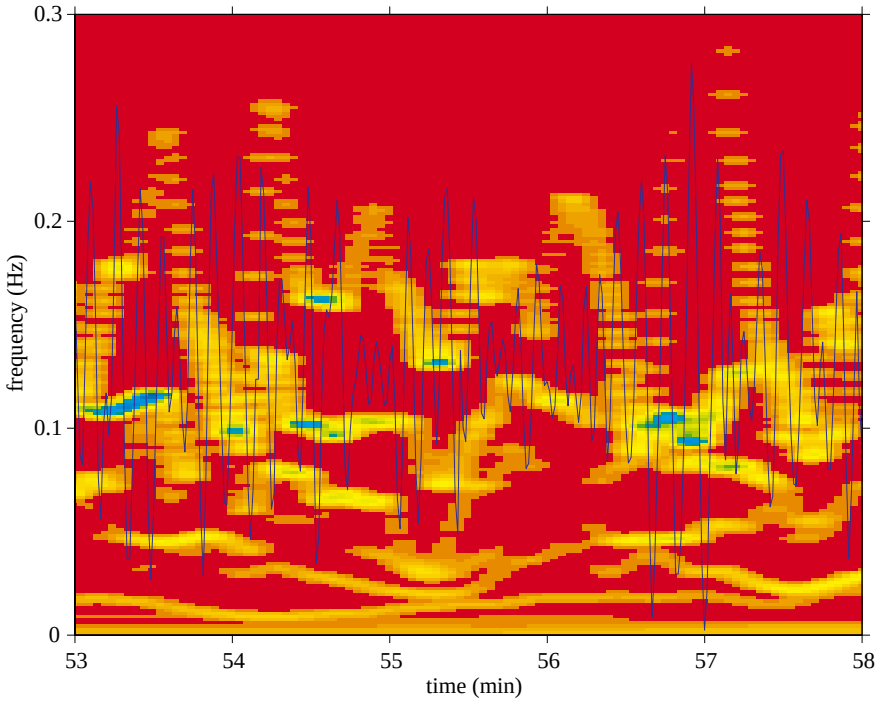

Figure 62. The Hilbert spectrum of a 5 min selected section of the data plotted with the surface elevation superimposed in an arbitrary scale. This figure illustrates the closeness of tracking energy–frequency variation in the Hilbert spectral presentation: notice the wave packet near the 57th minute. The wave energy decreases both before and after the packet, while the frequency increases. These and other variation are clearly represented.

between the tide and the tsunami. Compared to the traditional Fourier spectrum, the marginal spectrum again shows the absence of the harmonics. The reason seems to be clear: the tidal waves measured at a coastal station should be nonlinear because the shallow water propagation is governed by a nonlinear equation. The absence of the harmonics raises a question concerning the best way to present the tidal data. Should one use harmonic analysis? Or should one accept the intrawave frequency modulation to represent the nonlinearity? This is another interesting question for further study. Based on the discussion so far, it seems that the intrawave frequency modulation representation would be more physical.

(iv) *Altimeter data from the equatorial ocean*

Satellite altimetry has been a powerful technique for large scale ocean circulation studies (Huang *et al.* 1978; Robinson *et al.* 1983). Because of the importance of the equatorial region in determining the global climate pattern, the altimeter data have been used extensively to study the dynamics of this area (Miller *et al.* 1988; Miller & Cheney 1990; Zheng *et al.* 1994, 1995). The accepted view of the equatorial dynamics is the propagation of Kelvin waves forced by variable wind stress (Byod 1980; Zheng *et al.* 1995). In this model, the wave propagation will leave a surface elevation signature of the order of 10 cm, which can be measured by the satellite altimeter. Such continuous available data now are derived from both crossover and collinear differences (Miller & Cheney 1990), covering the period from April 1985

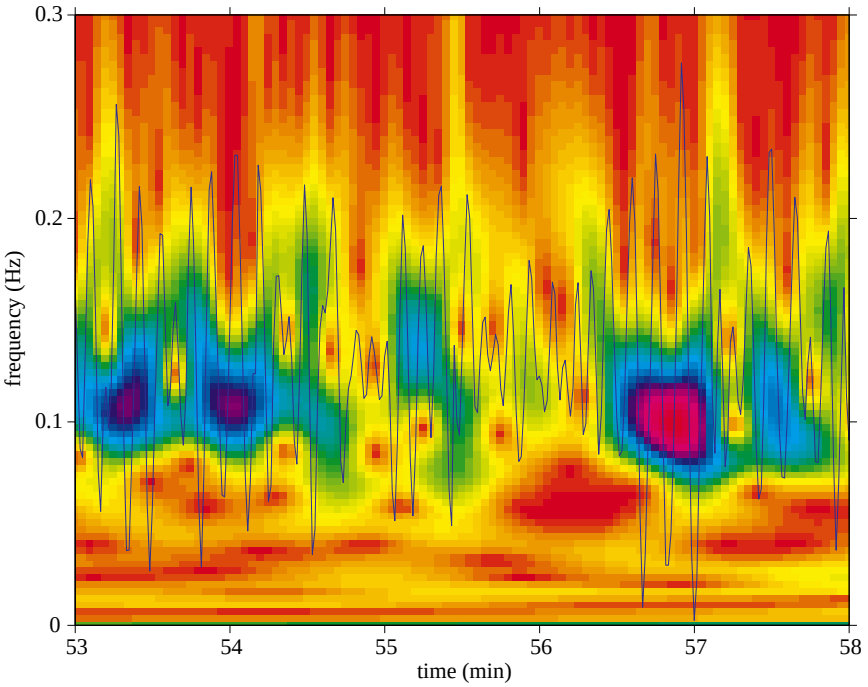

Figure 63. The same representation as in figure 62 with the Morlet wavelet spectrum. Due to the leakage, the tracking of the energy–frequency variation is much less tight.

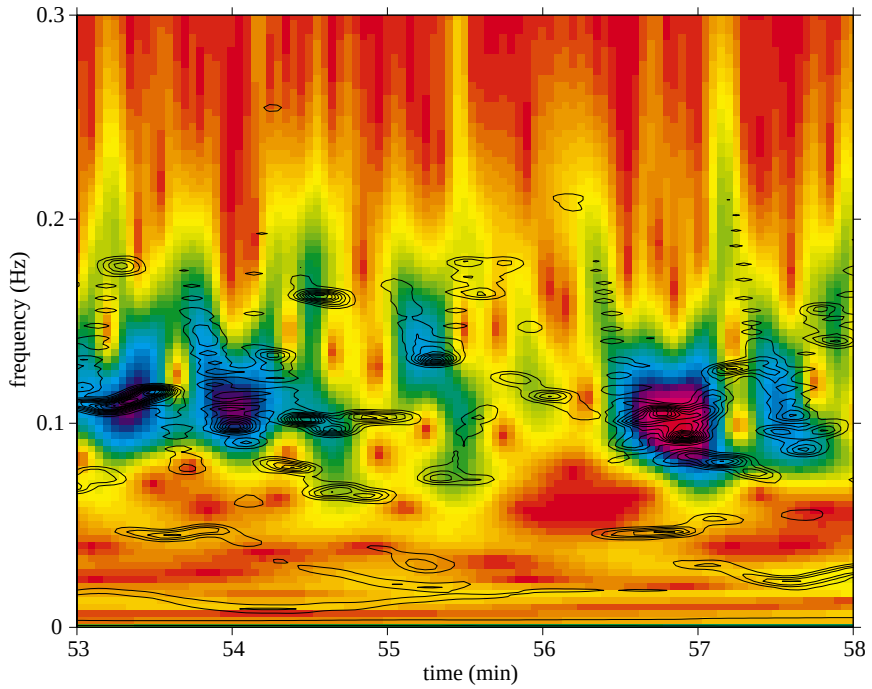

Figure 64. Comparison between the Hilbert spectrum (in contour lines) and the Morlet wavelet spectrum (in colour). While they both show similar energy–frequency distributions, the Hilbert spectrum gives a much sharper definition of the energy.

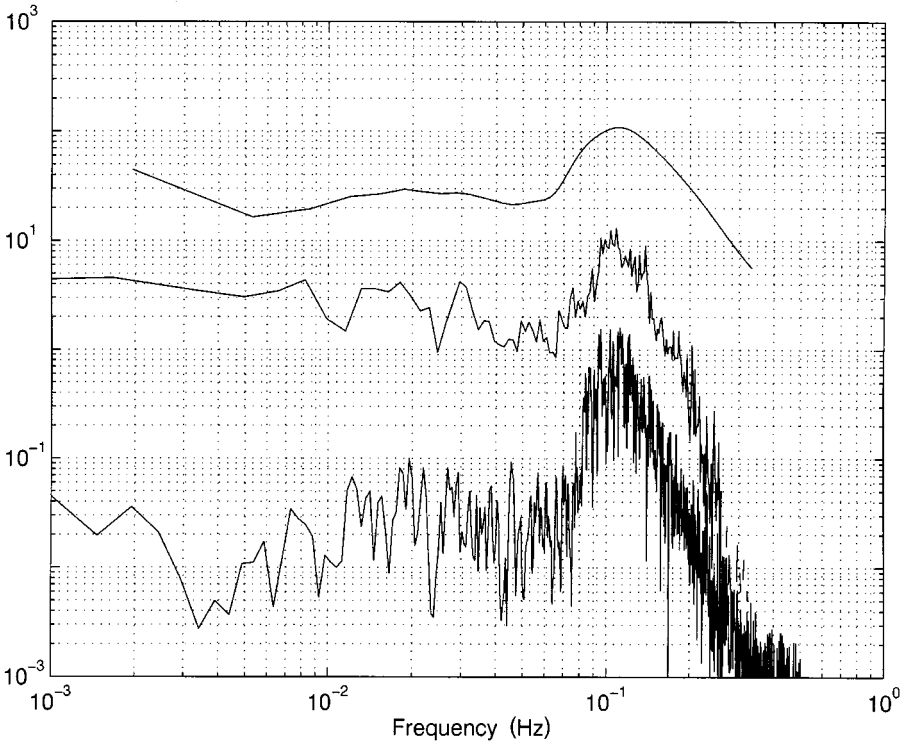

Figure 65. The comparison of the Fourier (the bottom line), marginal Hilbert (the middle line) and wavelet (the top line) spectra. While the wavelet spectrum lacks details, the Fourier spectrum gives spurious harmonics to represent the nonlinear and non-stationary nature of the data. The true spectrum of energy–frequency should be the Hilbert spectrum.

to September 1989. The final data have a spatial resolution of 8° longitude by 1° latitude. A typical time series on the equatorial sea surface elevation data at 174° E is given in figure 72.

Limited by the data length and complicated by the dynamics, all the past investigators have problems in processing this obviously non-stationary data. We have analysed the data using two methods: the wavelet analysis and the Hilbert spectral analysis. Although the wavelet results, given in figure 73 show the fluctuation of energy as a function of time and frequency, the energy distribution as a function of time and frequency is too diffused to yield any quantitative information.

The EMD method on the same sets of data yields four IMF components and a residual trend as shown in figure 74. The corresponding Hilbert spectrum is given in figure 75. The sharpness of the frequency–time resolution in the Hilbert spectrum is obvious. The most important properties of the wave are the strong intrawave frequency modulations. Each wave group will have a frequency change over 1.2–9.6 cycles per year within a period of 50–200 days. When we started the investigation of data using Hilbert spectral analysis, this was the first time we encountered the strong intrawave frequency modulation. The curious frequency variations led us to the re-examination of the classical nonlinear systems, which eventually clarified the role of intrawave frequency modulations. In the present example, one can argue that, with the ocean depth as a small fraction of the horizontal extent, the low-frequency geophysical wave motions are in shallow water, then the wave motions must be

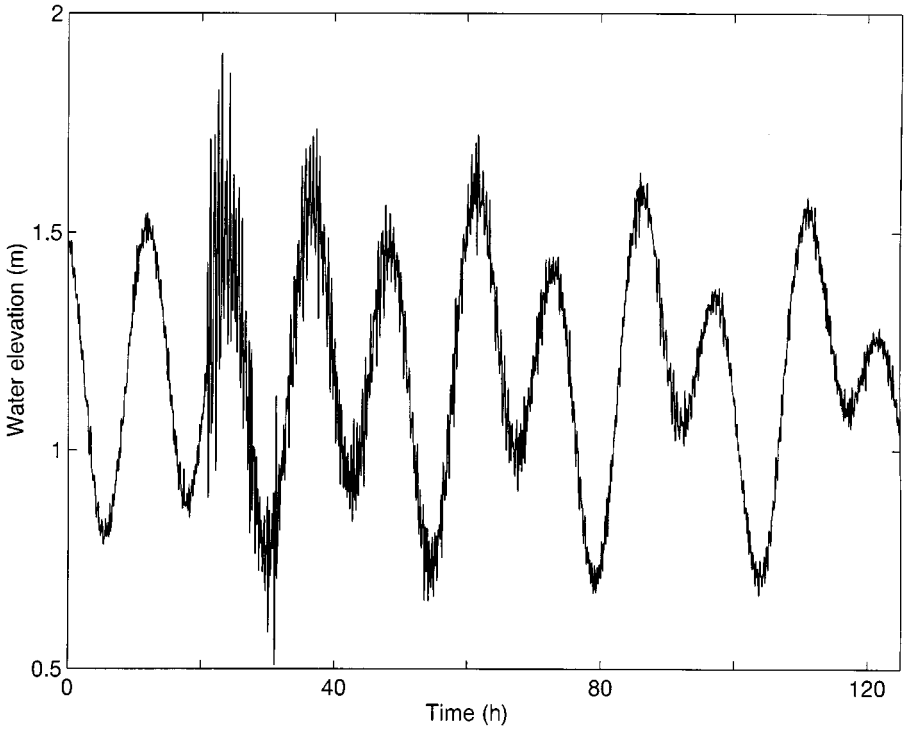

Figure 66. The tidal data collected inside the Kahului Harbour, Maui, from October 4–9, 1994. On October 5, tsunami waves reached the tidal station with a signal of comparable strength to that of the tides. Though the tides can be regular, the tsunami waves are transient.

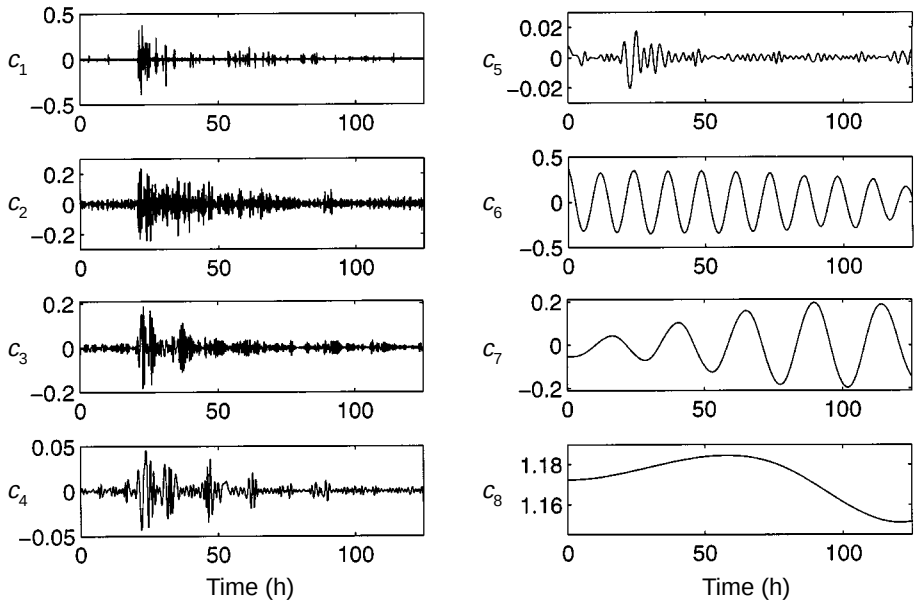

Figure 67. The eight IMF components obtained by the EMD method. The time scales of the tides and the tsunami waves are clearly separable.

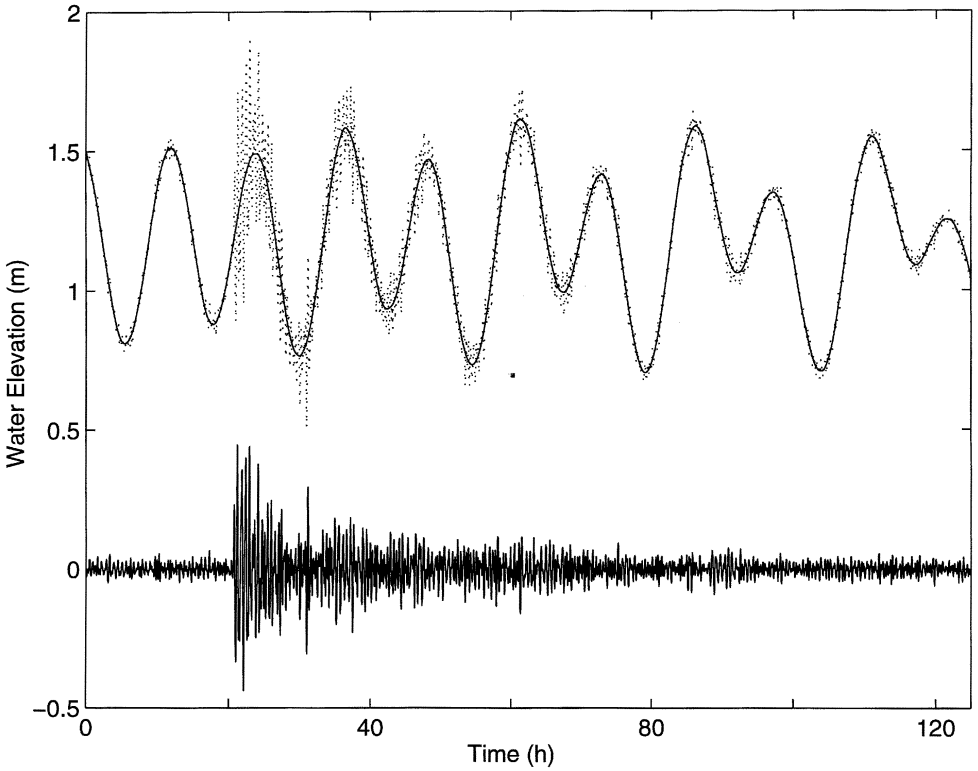

Figure 68. Separation of the tides and the tsunami waves by the IMF components: the raw data are shown in the dotted line; the tides represented by the sum of the last three IMFs shown in the solid line; and the tsunami waves represented by the sum of the first five IMFs shown in the thick solid line at the bottom. This example illustrates the capability of using EMD to separate transient noise from data.

governed by nonlinear equations. Although frequency modulation phenomena are known to exist in wavelet analysis, the strength and the clearness of the modulations as shown here are seldom seen in other natural phenomena by other time–frequency distribution methods.

The dynamics of the equatorial Kelvin waves have been discussed extensively by Zheng *et al.* (1995). They concluded that steady wind cannot excite sustained Kelvin waves, but the variable wind, especially the westerly wind burst, can cause resonant interactions of the excited Kelvin waves with an intrinsic frequency of $(1.0\pm 0.2)\times 10^{-2}$ cycles per day, or 0.22–0.34 cycles per year. A prevailing band of energy can be easily seen in figure 75. At around 400 days after the starting date of the data, a combination of strong low-frequency (less than 0.5 cycles per year) and high-frequency (4–5 cycles per year) events occurred, which coincided with the onset of the 1986–1987 El Niño. Although the Hilbert spectrum shows some energy distribution anomaly, we need more data to deduce the characteristics of such special events as El Niño. Further discussion will be given separately. Time series data at different locations can be synthesized to produce a four-dimensional Hilbert spectrum to show the energy–frequency–time–space evolution. The details of this will be discussed later in a separate paper.

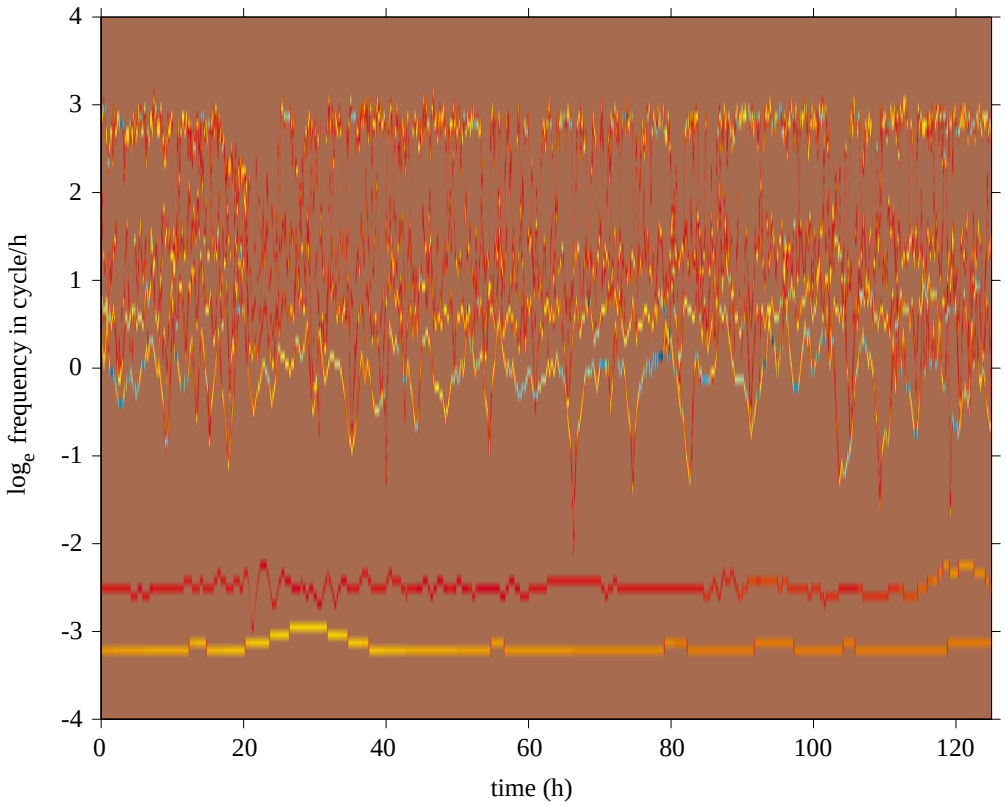

Figure 69. The Hilbert spectrum for the tides and the tsunami waves: the frequency of the tsunami wave is phase locked with the tidal cycle. A surprising aspect of the tsunami waves is the length of time they lasted. Both the diurnal and the semi-diurnal tidal cycles are also resolved.

(v) *Earthquake data*

All the earthquakes are transient (Anderson 1991); the data are, therefore, non-stationary. For lack of alternatives, in seismology as well as earthquake engineering, most data are still processed with the Fourier analysis (Geller & Ohminato 1994; Scherbaum 1994). The difficulties with the Fourier spectral analysis are associated with the nonlinear and the non-stationary nature of the data. Such methods cannot reveal the detailed information in the dispersion properties, the wave form deformation, and the energy–frequency distribution. All these properties are needed for understanding of the crest structure geophysics in general, and deducing the site specific ground motions in particular.

Most of the near field strong earthquake ground motions are of extremely short duration, lasting only a few ten to a few seconds at most. Such records always give artificially wide Fourier spectra because of the non-stationarity. The wide frequency distribution will dilute the energy content everywhere on the frequency axis and mask the true energy–frequency distribution, and result in under estimation of the energy density at the critical resonant frequency for specific structures.

Although the earthquake wave motion has always been treated as a linear process, the strong ground motion may not be linear as revealed by the distorted wave forms

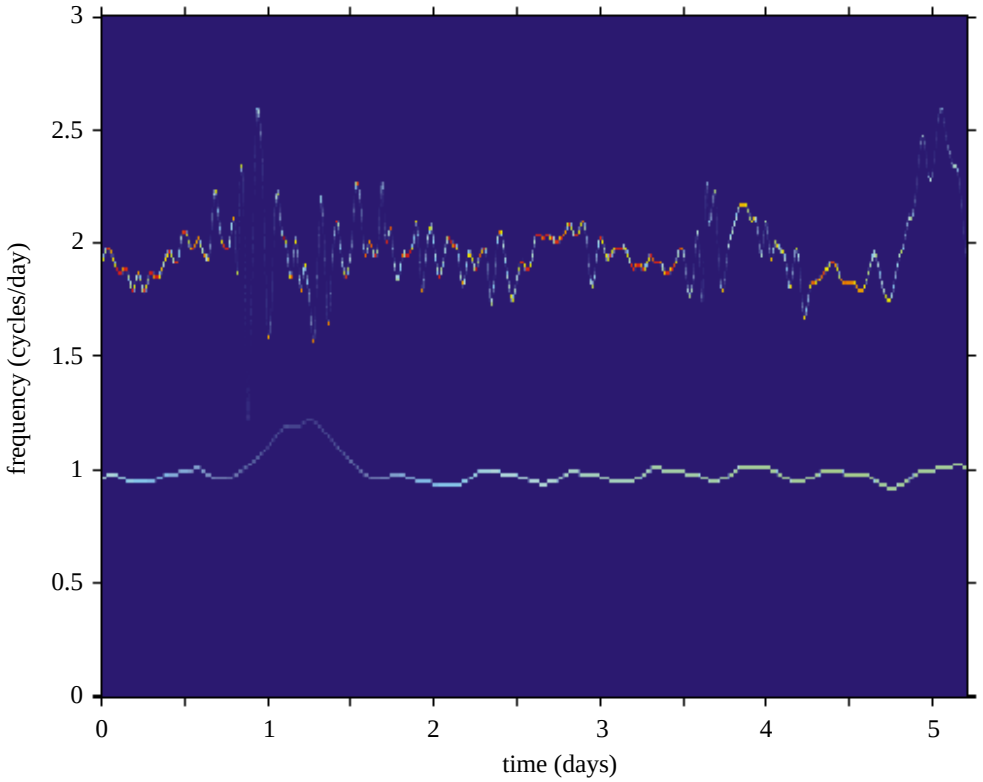

Figure 70. The Hilbert spectrum from the tidal components: in this detailed representation, neither the diurnal nor the semi-diurnal tides are of constant frequencies. The variation of the semi-diurnal tide is of particular interest; it shows possible interaction with the tsunami waves.

in the data (Newmark & Rosenblueth 1971). The rich higher harmonics, a typical consequence of Fourier analysis when applied to nonlinearly deformed signals, have also contributed to clouding the characteristics of the real signal.

Finally, to apply the Fourier analysis, some kind of window must be used to eliminate the end effects and to facilitate computation. For a truly non-stationary process, there is no time scale to guide the choice of the window size. A long time window (10 s, say) is necessary for resolving low frequency (0.1 Hz) which are critical for modern high-rise structures, but too long a window will be plagued by the non-stationarity. Such a difficulty has never been fully resolved when Fourier spectral analysis is applied to earthquake data.

As explained above, both non-stationarity and nonlinearity can induce artificial frequency smearing and reduce the true energy density. We will use a few sets of data to illustrate EMD and the Hilbert spectral analysis technique with the following emphases: to illustrate the gross misrepresentation of the energy–frequency distribution by the Fourier spectrum. We will use the well-tested data from the El Centro earthquake to illustrate our point here.

The original data are shown in figure 76. These data, when decomposed by EMD, give ten components as shown in figure 77, with the corresponding Hilbert spectrum given in figure 78. From the Hilbert spectrum, we can see the diffused energy in the high-frequency range, while the persistent energy resides along horizontal belts below 1 Hz. To assure the Hilbert spectrum is a valid representation of the energy–

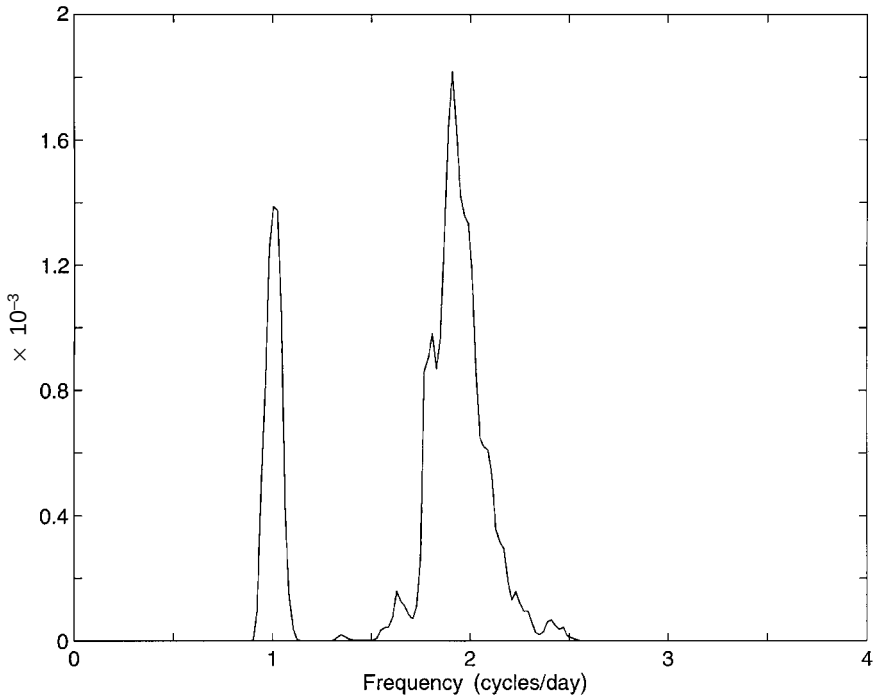

Figure 71. The marginal Hilbert spectrum of the tidal components only. The wider distribution of the semi-diurnal frequency indicates the possible interactions with other environmental conditions.

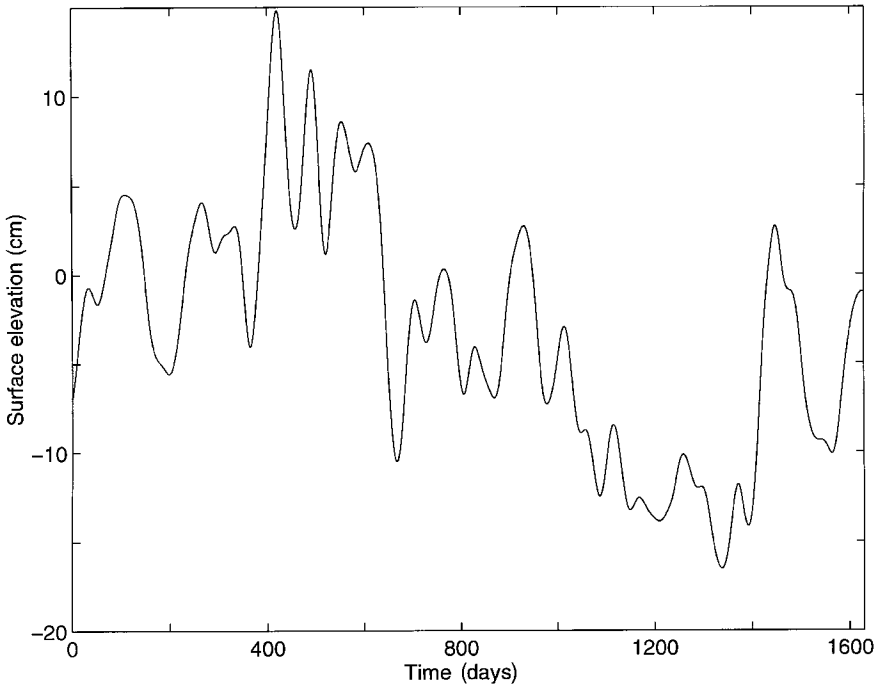

Figure 72. Altimeter data over the equatorial sea at 174° E on the Equator. The starting date of the data was 1 April 1985.

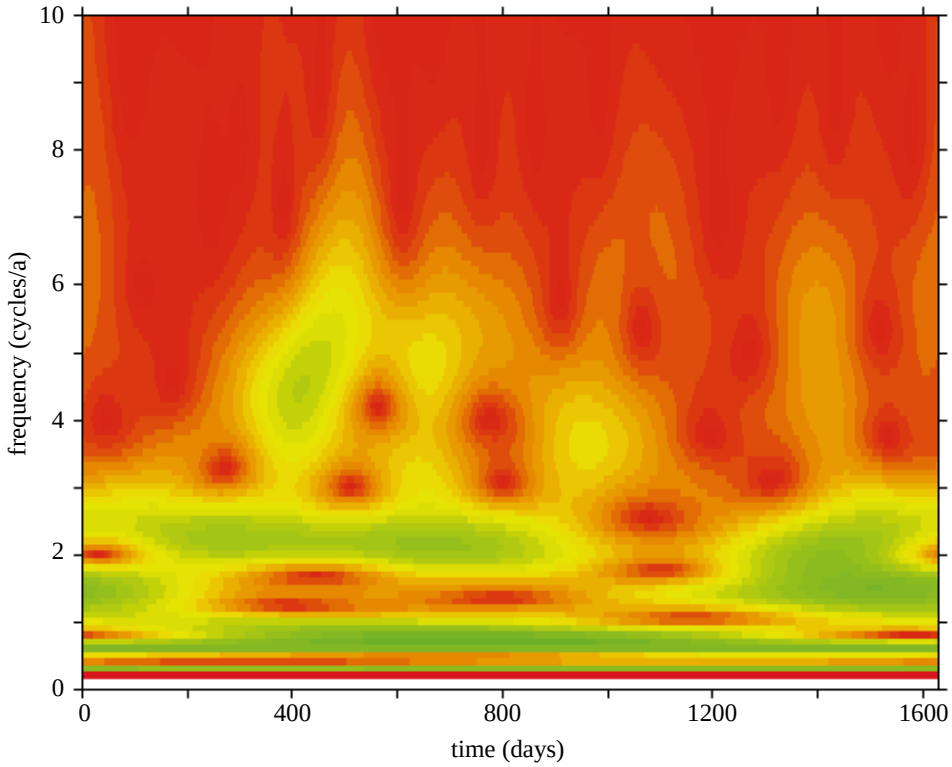

Figure 73. The Morlet wavelet spectrum for the altimeter data showing interwave frequency modulations.

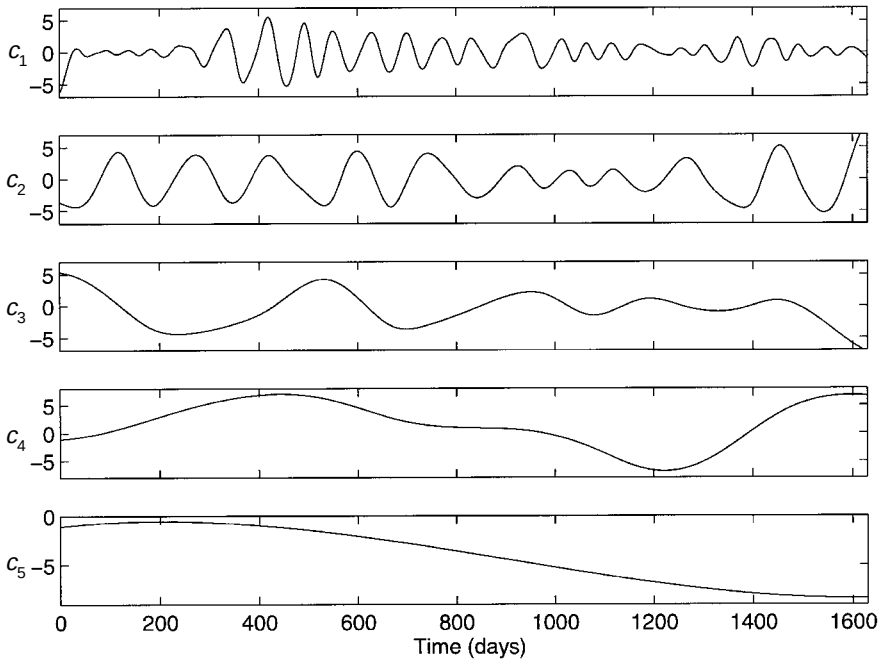

Figure 74. The IMF components from data through the EMD method.

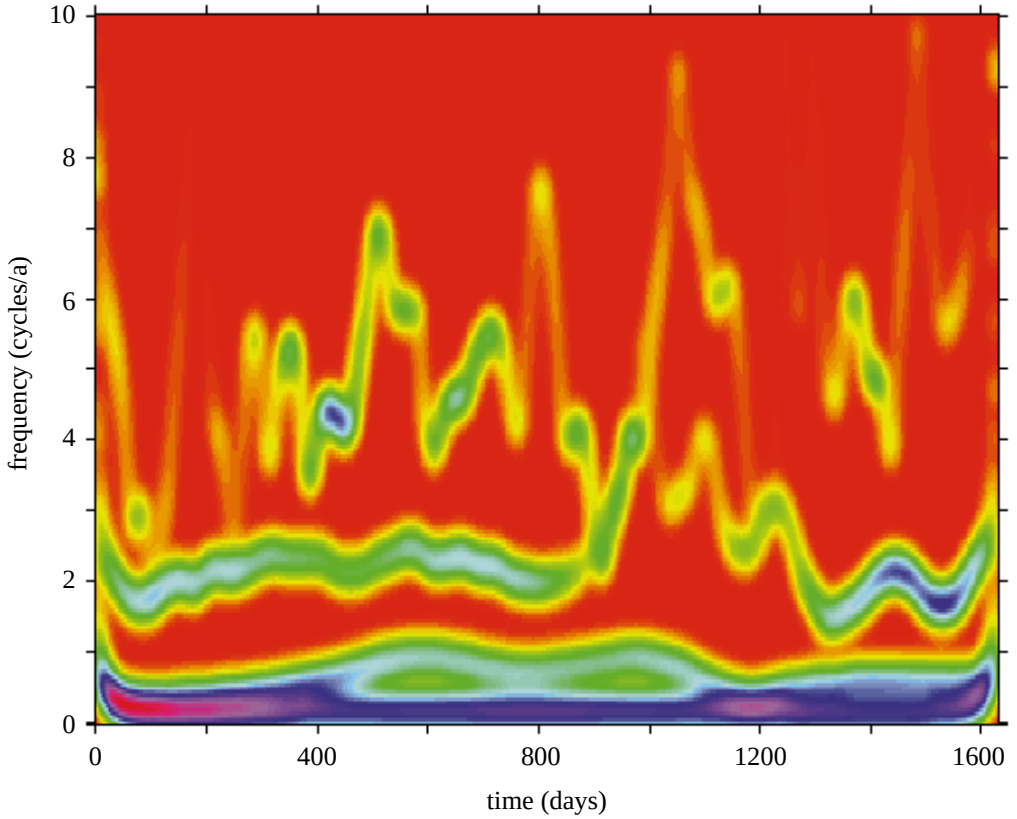

Figure 75. The Hilbert spectrum for the altimeter data showing the much sharper frequency modulations of both inter- and intrawave types. A major energy concentration near 400–500 days after the initiation of the data coincides with the 1986-87 El Niño.

frequency–time distribution we also used the standard Morlet wavelet analysis on the same data. The result is given in figure 79. Direct comparison between the Hilbert spectrum and the wavelet analysis is not easy, for the Hilbert spectrum gives too many details. A 15×15-point smoothing gives the smoothed spectrum in figure 80. Now the similarity is reassuring: they show a similar pattern of energy concentration in the low-frequency range. But the difference really calls for alternatives: being Fourier based, the wavelet result gives much more pronounced high-frequency components than does the Hilbert spectrum, a built-in deficiency in the Fourier analysis discussed in the last section.

Other than the Hilbert spectrum, the marginal spectrum and the corresponding Fourier spectrum are presented in figures 81*a–d*. Although our results are presented with linear scales and in terms of frequency rather than period, the energy–frequency distribution with further smoothing is consistent with the previous analysis of the data shown in Newmark & Rosenblueth (1971). The Fourier spectrum has a much wider energy distribution; consequently, the energy content for any frequency is also much lower than the corresponding values in the Hilbert spectrum. Furthermore, the two spectra have drastically different energy distribution patterns caused by the tendency for energy smearing of the Fourier spectrum. While both spectra give a decreasing trend in energy density as a function of frequency, the Fourier spectrum shows a conspicuous lack of energy in the low-frequency range (less than 1 Hz), as

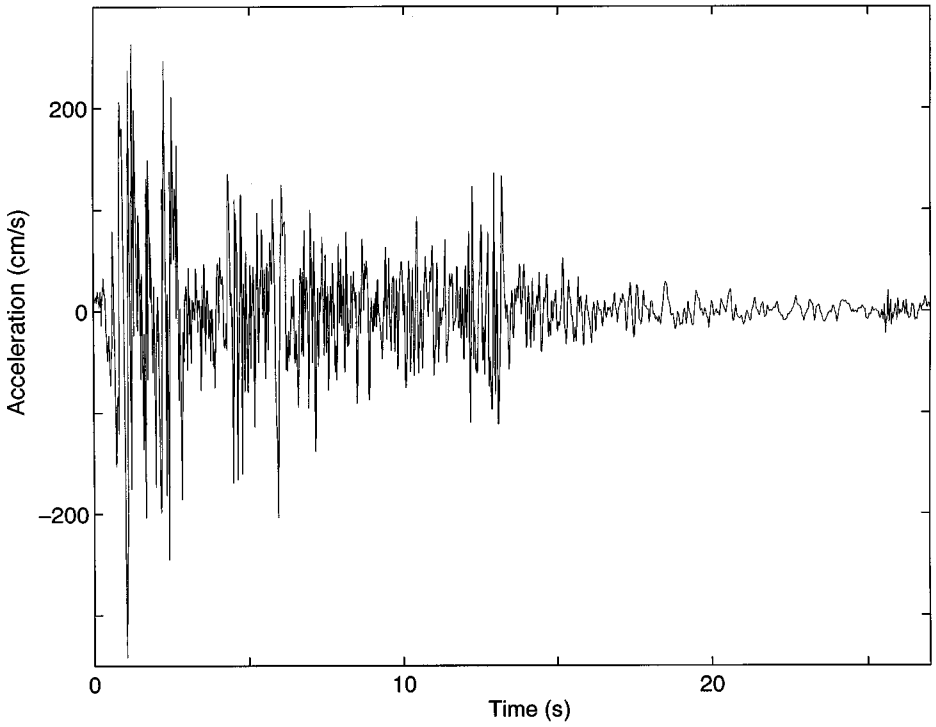

Figure 76. The earthquake data of El Centro, vertical acceleration in cm s^{-1}.

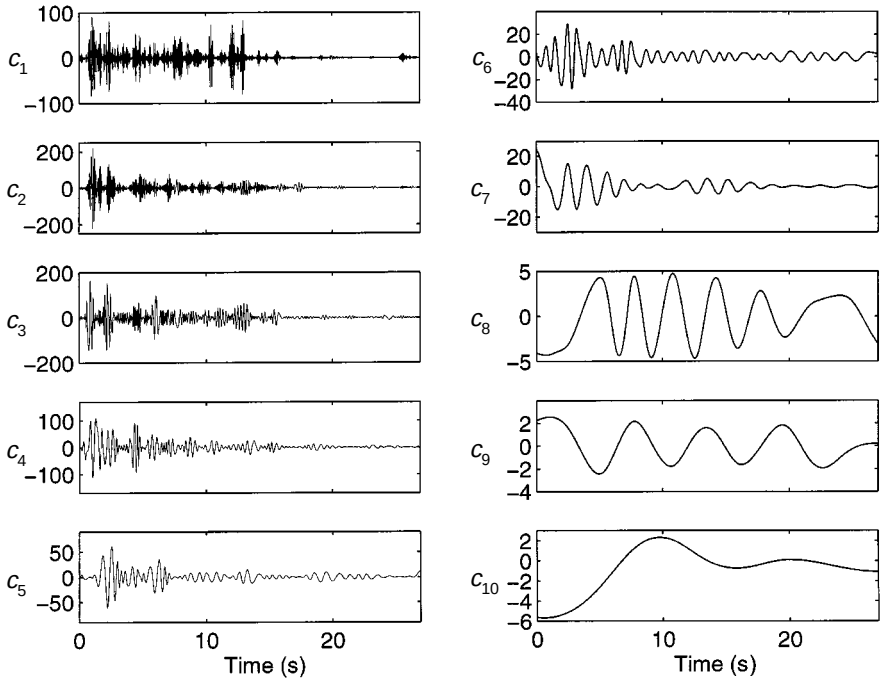

Figure 77. The ten IMF components obtained from the data shown in figure 76 through the EMD method.

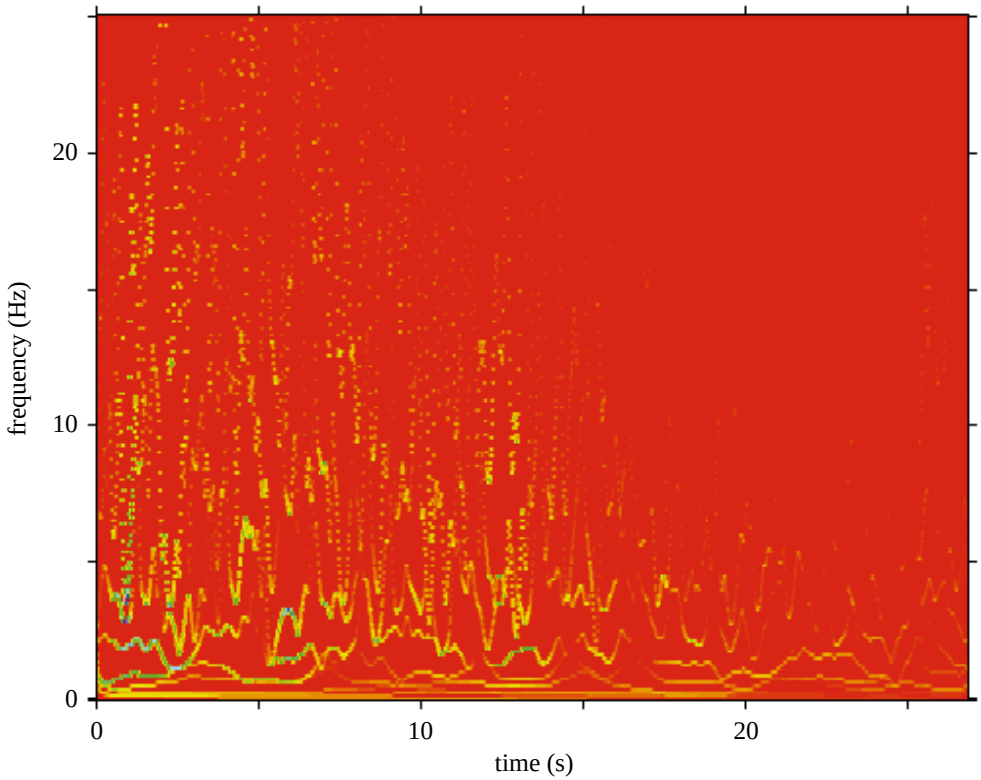

Figure 78. The Hilbert spectrum for the earthquake data. There is considerable frequency variation during the period of the strong ground motion; most of the energy, however, is concentrated in the low-frequency range with the frequency less than 5 Hz.

shown in the detailed spectra covering only the 0–5 Hz frequency range. The low-frequency range critical to the high-rise structures is severely under represented.

This peakiness of energy located at a very narrow frequency range can cause resonance oscillation of buildings, and has been observed to have caused great destruction at Mexico City during the 1985 earthquake. Hadley *et al.* (1991) studied such a case and attributed the peakiness to soil-induced amplification of the ground motion. He further concluded that the oscillation could not be duplicated by any linear model. The new Hilbert spectral analysis has clearly identified the new oscillation mode similar to the case studied by Hadley *et al.* (1991) but from a different cause. Such oscillation modes would have been totally obscured by Fourier spectral analysis. From the Fourier spectra, there are no clear-cut corner frequency or frequencies for one to construct the envelope spectrum (Aki & Richards 1980). There are also no discernible stationary segments as postulated by the time history envelope method. We also have to notice that even the criteria used to establish such segments based on the earthquake energy are still under debate currently (Boore & Atlinson 1987). In summary, there is no alternative to the Hilbert spectral analysis in earthquake or seismology studies.

(vi) *The laboratory wind data*

Finally, let us return to the laboratory wind data used at the beginning to illustrate the sifting and Hilbert spectral analysis procedures. In the smoothed form,

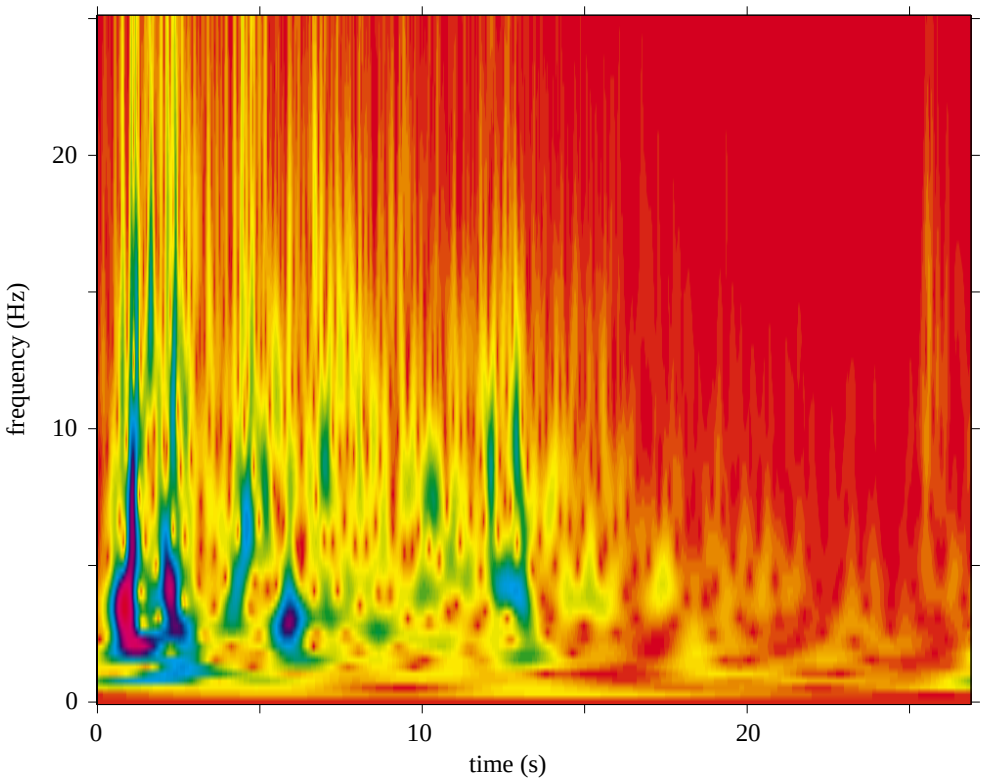

Figure 79. The Morlet wavelet spectrum for the same data as in figure 78. The overall energy distributions are similar to the Hilbert spectrum but the distribution is much diffused, and there is a lack of low-frequency components and an abundance of higher harmonics.

the Hilbert spectrum looks similar to the wavelet analysis result, yet they are still different. Two prominent differences stand out: the first is the absence of the higher harmonics in the Hilbert spectrum compared with the wavelet result; the second is the appearance of frequency modulated ridges of energy in the Hilbert spectrum. Both characteristics of the Hilbert spectrum are new and significant.

In the Hilbert spectrum, we can see the increase in energy density with time, especially after 10 s following the initiation of the data. Most of the new energy is in the 2–5 Hz range with strong continuous frequency modulation rather than the richness of high-harmonic components in the wavelet analysis. As the frequency modulation is the signature of nonlinear mechanisms, the Hilbert spectrum clearly shows the nonlinear properties of the turbulent flows.

From the IMF components, we can see that c_2 through c_6 are of the same magnitude. Collectively, they represent the bulk of the turbulent energy. All these components are highly intermittent, as indicated by the stationarity index study. Even after averaging, the DSS still indicates strong peaks at 7 and 17 Hz, which correspond to the first two peaks of the marginal spectrum in figure 11. These frequency ranges are represented by the components c_2 through c_6. In order to study the energy in this frequency range in detail, a special Hilbert spectrum for the components c_2 through c_6 is constructed as in figure 82. We cannot detect any change in the overall energy level, but we can certainly see the intermittency. According to Liu &

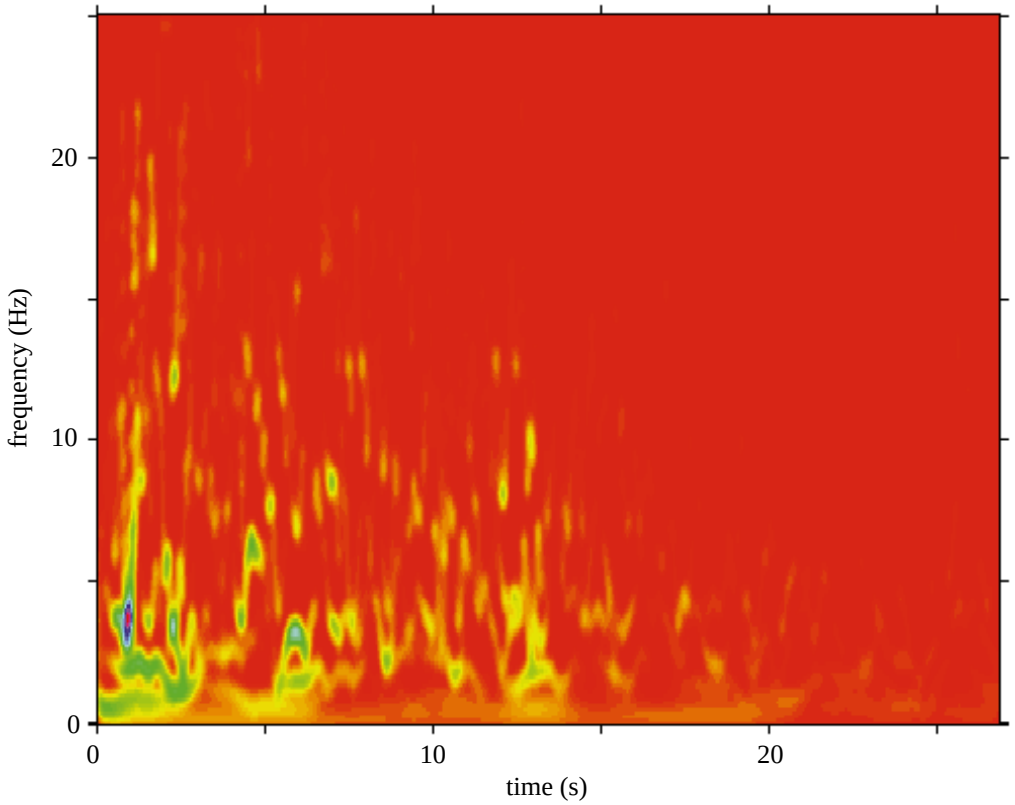

Figure 80. A 15×15 Gaussian filter smoothed Hilbert spectrum. It still shows better frequency and time localization than the wavelet spectrum given in figure 79.

Liu (1995), nonlinearity and intermittency are the common properties of the turbulent flows. Let us examine the instantaneous energy density as shown in figure 83. As expected, the fluctuating components of the wind contain relatively low energy compared with the mean wind, or the trend, as shown by the dotted line. If the energy of the trend is added to the fluctuation, the overall energy is quite levelled, but the density is too high to be shown with the fluctuating components. In figure 83, the overall energy density has been divided by 100 in order to plot the two curves together for comparison. Because the total energy is dominated by the mean wind, we have recommended its omission in construction of the Hilbert spectrum. It can easily overwhelm all the other components; therefore, it should not be included unless specially justified. Construction of the special Hilbert spectrum with selected IMF components is an equivalent of bandpass filtering through EMD, yet it is different from the Fourier bandpass filtering, for we are not restricted by the stationarity requirement as in the Fourier analysis.

10. Discussion

In this paper, we have presented the empirical mode decomposition method and the associated Hilbert spectral analysis. The combination has proven to be versatile and robust in analysis of nonlinear and non-stationary data. Although the new idea

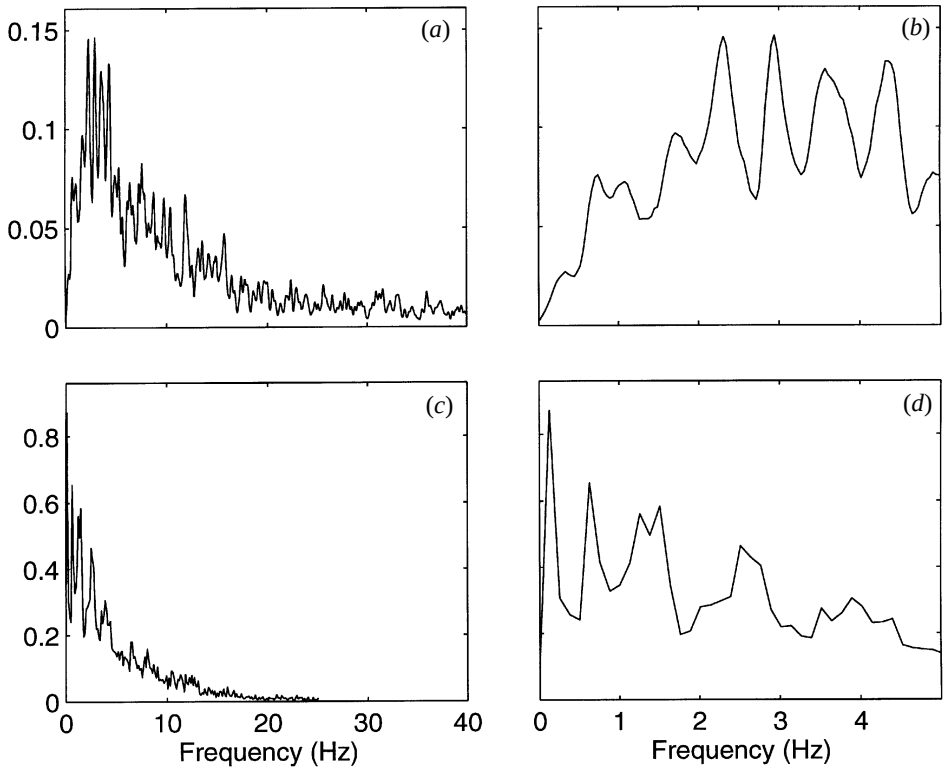

Figure 81. The comparison of the marginal spectrum and the Fourier spectrum for the earthquake data. (*a*) Fourier spectrum with full frequency range. Most of the energy is concentrated in the frequency range below 5 Hz. (*b*) Fourier spectrum for the detailed 5 Hz range. The energy fills over the whole range, with low density below 2 Hz. (*c*) Marginal spectrum with full frequency range. Most of the energy is concentrated in an even narrower range than the Fourier spectrum. Because of this narrow distribution in the frequency space, the energy density is at least a factor of three higher than the Fourier spectrum. (*d*) Marginal spectrum for the detailed 5 Hz range. The energy is still unevenly distributed with the concentration on the lower than 2 Hz range. The high density and low frequency should cause major concerns for earthquake engineering, for this is exactly the range of resonance for the high-rise buildings.

of an adaptive basis makes the method highly effective and physically sound, there are, however, areas needing future attention.

First the spline fitting needs improvements. As the description of the sifting process indicates, the spline fitting is the essential step in generating the intrinsic mode function, the basis for the Hilbert spectral analysis. Although the present envelope-mean method works well in most cases, problems still exist. The cubic spline fitting adopted here has both overshoot and undershoot problems. These problems can be alleviated by using more sophisticated spline methods, such as the taut spline in which the tension of the spline curve can be adjusted. But our trials of taut spline showed only marginal improvement. The alternative is the higher-order spline programs which would be more time consuming. Exploration of the higher-order spline is underway. Even with this problem, we believe that we have detected most of the dynamic characteristics of the data examined. Any improvements would probably be marginal.

As the spline procedure is time consuming, more efficient methods can be devised

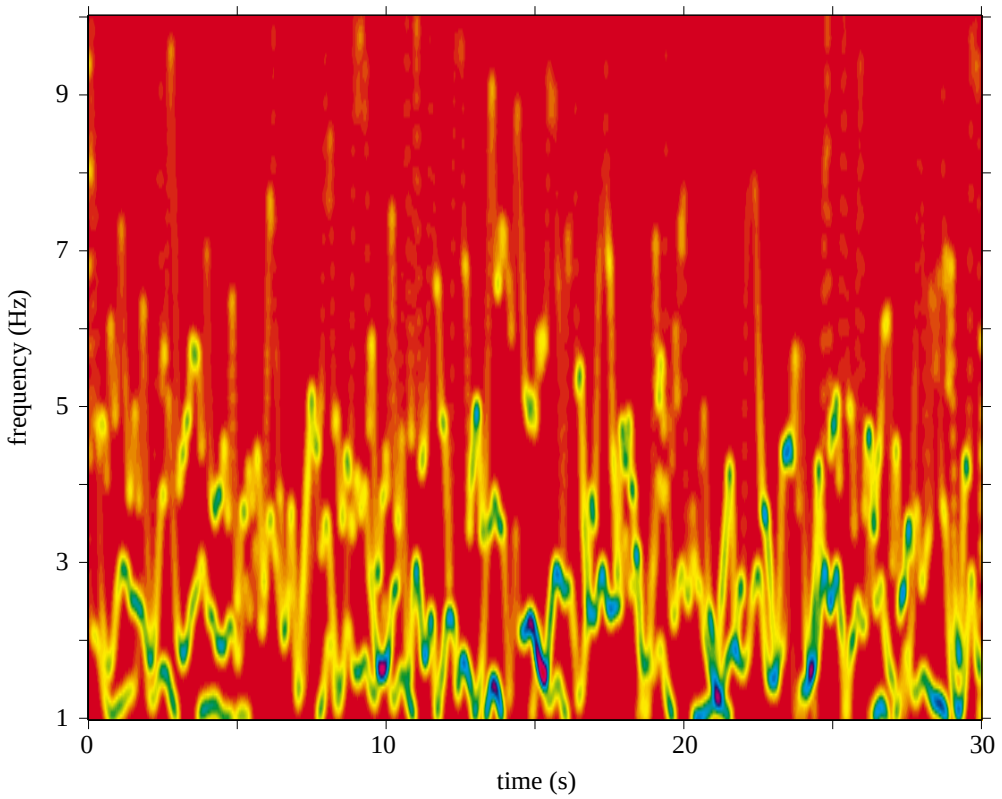

Figure 82. The special Hilbert spectrum with only the energy-containing IMFs (c_2–c_6). The turbulent flow is dominated by the intermittence and frequency modulations, a sure sign that nonlinear mechanisms are at work.

by using simple means of successive extrema instead from the envelope mean. In this way, only one spline fitting is required rather than two. It is easier to implement, but the shortcomings are the more severe amplitude averaging effects when the neighbouring extrema are of different magnitudes. The successive-mean method will have a stronger forcing to reach uniform amplitudes, in which the true physics associated with amplitude will be destroyed. Only for problems where the amplitudes of the components are nearly constant should this method be used.

Either the envelope-mean or the successive-mean method when applied with the requirement of absolute symmetry will produce the absurd result of uniform amplitude IMFs. Therefore, the criteria in the sifting process should be chosen judiciously. One should avoid the too stringent criterion that we would get uniform amplitude IMFs; on the other hand, one should also avoid too loose a criterion that would produce components deviating too much from IMFs. The results as reported in the sifting section were arrived at by our somewhat limited experience. With this new approach, further studies and quantitative criteria perhaps are worthwhile to explore.

Second, the end effects need more improvements, too. There are two types of end effects: in the spline fitting; and in the Hilbert transform. It is well known that the most serious problem of the spline fitting is at the ends, where cubic splines can have wide swings if left unattended. We have adopted a method of adding characteristic waves at the ends which are defined by the two consecutive extrema for both their

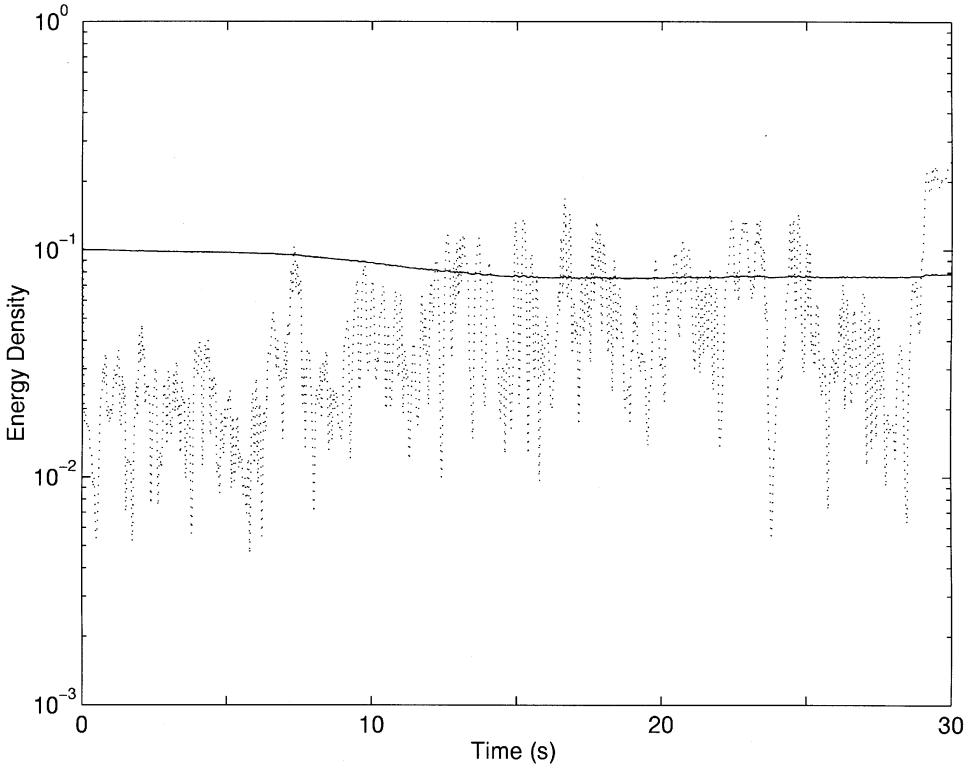

Figure 83. The instantaneous energy density of the fluctuating components only (dotted line) and the overall energy density with the trend included (solid line). The component representing the trend can easily overwhelm the fluctuating components; therefore, it should not be included unless specially justified.

frequency and amplitude of the added waves. This method has confined the large swings successfully. Further studies are needed to guarantee that the ends would not corrupt the data and propagate to the interior.

Other than the spline fitting, the Hilbert transform also has end effects, for the numerical method to implement the Hilbert transformation is based on the Fourier transform. Since the first and the last points of the data are usually of different values, the Fourier transform will induce additional components to bridge over the difference resulting in the well-known Gibbs phenomena. To eliminate it in the Fourier transform, various windows have been adopted (see, for example, Brigham 1974). In our applications, instead of a window which will eliminate some useful data at the end, we have again added two characteristic waves at either end. These attached waves begin at the slightly enlarged data set with zero, and likewise end it with a zero level. Thus, the annoying Gibbs phenomena are greatly reduced.

Third, we have to be careful in the case of weak signals imbedded in stronger ones. Both the envelope-mean and successive-mean methods depend on the existence of extrema. In fact, our method is built on the identification of scale from successive extrema. If weak signals are imbedded in other strong ones, the extrema might not even be visible to the eyes; however, the sifting process can still pick them up. If the weak signals are phase locked with, and occur only at, the maximum slope regions of the strong signals, then we would have difficulty picking them up. In this case, the

weak signals appear as intrawave frequency modulations. If it is necessary to separate them, the data can be differentiated once before processing. Since differentiation is a linear operation, it will neither create nor annihilate scales. Therefore, such an operation could be used if needed. It is also observed that, for strictly linear and stationary data, Fourier spectral analysis can produce superior results.

Fourth, the Hilbert transform needs over-sampled data to define the instantaneous frequency precisely. In Fourier analysis, the Nyquist frequency is defined by two points per wave. But the frequency is defined for a wave covering the whole data span. In our analysis, the instantaneous frequency is defined through a differentiation process, and thus more data points will help to define the frequency more accurately. Based on our experience, a minimum number of data points to define a frequency is five (or $4\Delta t$). The lack of fine time steps can be alleviated by interpolating more points between the available data. As a spline interpretation would neither create nor annihilate scales, it also can be used for the interpolation when the data are very jagged from being under-sampled. The smoothed data, though, have a longer length and are sometimes easier to process. The interpretation serves to give better frequency definition.

Fifth, we should reiterate the strength of the EMD method. EMD is built on the idea of identifying the various scales in the data, the quantities of great physical significance; therefore, in the sifting procedure, orthogonality is not a consideration, but scales are. Since orthogonal decomposition is a characteristic for linear systems, violating this restriction is not a shortcoming but a breakthrough. Therefore, the decomposed IMFs may or may not be orthogonal. As such, this method can be applied to nonlinear data. Though the IMFs in most cases are practically orthogonal, it is a coincidence rather than a requirement of the EMD. The dependence on the existence of scale for mode definition has one limitation: the decomposition method cannot separate signals when their frequencies are too close. In this case, there would not be any characteristic scale; therefore, physically they are identical. This may be the most severe limitation of this method, if we count it as one. Details of the restrictions will be studied later. Another advantage of the method is the effective use of all the data. In the sifting process, the longest scale is defined by the full length of the data. As a result, EMD can define many long period oscillations. As is well known, the Hilbert transform tends to identify the highest frequency (Boashash 1992), the extraction of the long-period components is indeed an unexpected bonus of the EMD.

Finally, though the EMD method will give IMF components, the individual component does not guarantee a well-defined physical meaning. This is true for all decompositions, especially for the methods with *a priori* basis. In most cases, however, the IMFs do carry physical significance. Great caution should be exercised in making such attempts. The rule for interpreting the physical significance of the IMFs is that the scales should be clearly separated. Together with the Hilbert spectrum, the totality of the presentation should give a much more detailed representation of the physical processes.

11. Conclusions

The combination of the empirical mode decomposition (EMD) method and the associated Hilbert spectral analysis has offered a powerful method for nonlinear non-

stationary data analysis. Central to the present approach is the sifting process to produce the IMFs, which enables complicated data to be reduced into such a form that the instantaneous frequencies can be defined. These IMFs forming the basis of the decomposition are complete and practically orthogonal. The expansion in terms of the IMF basis has the appearance of a generalized Fourier analysis with variable amplitudes and frequencies. It is the first local and adaptive method in frequency–time analysis.

A great advantage of EMD and Hilbert spectral analysis is the effective use of the data. In EMD, we have used all the data in defining the longest period component. Furthermore, we do not need a whole wave to define the local frequency, for the Hilbert transform gives the best fit local sine or cosine form to the local data; therefore, the frequency resolution for any point is uniformly defined by the stationary phase method or local derivative of the phase. This advantage is especially effective in extracting the low-frequency oscillations. Unlike the wavelet analysis, the instantaneous frequency can still be localized in time even for the longest period component without spreading energy over wide frequency and time ranges. Still, another advantage of EMD and Hilbert spectral analysis is its application to transient data without zero or mean references; the trend or the DC term is automatically eliminated. The introduction of the Hilbert spectrum also provided a quantitative measure of the degree of stationarity. Various examples presented here testify to the usefulness of this new method.

Other than the practical methodology, the most important conceptual innovations of the present study are the physical significance assigned to the instantaneous frequency for each mode of a complicated data set, and the introduction of the IMF. By adopting the instantaneous frequency, we can clearly define both the inter- and intrawave frequency modulations in a wave train. Such frequency modulations are totally lost in Fourier spectral analysis, and only the interwave frequency modulation can be vaguely depicted in the wavelet analysis. Yet, both the interwave and the intrawave frequency modulations are critical in interpretation of oscillatory phenomena: the former explains the wave form deformation by nonlinear effects, which traditionally has been taken as the harmonic distortion; the latter explains the dispersive propagation of waves. The intrawave frequency modulation offers new insight into nonlinear oscillation systems in more detail than the modern topological treatment. By adopting the instantaneous frequency, we have not only eliminated the need for higher harmonics to simulate the nonlinearly deformed waves, but also the spurious harmonics to simulate the non-stationary data. We believe this new method can give us new physical insight in all other nonlinear and non-stationary phenomena. The instantaneous frequency can only be defined for IMF, which is defined here based on the local properties of the data rather than the global restrictions proposed before.

We have only begun to explore the full physical interpretations of the Hilbert spectra for complicated data. Associated properties of the marginal spectra and various definitions of stationarity also need to be explored in the future.

We express our deep appreciation to Professor O. M. Phillips of the Johns Hopkins University, and Professor T. Y. Wu of the California Institute of Technology for their encouragement, valuable suggestions and critical comments without which the research could not have proceeded to its present form. We therefore dedicate this publication to them. We also express our thanks to Dr S. C. Liu of the National Science Foundation, for his encouragement and the inspiration from his pioneering papers on the time-varying spectral analysis; to Dr E. Mollo-Christensen of

Lexington, MA, Professor F. Browand and Professor G. Spedding of USC; Professor C. Friehe of UC Irvine; Dr M. Shlesinger, Dr L. Goodman and Dr T. Kinder of ONR; Dr R. Cawley, Dr Y. Liu and Dr Q. Tran of NSWC for their helpful discussions. This research has been supported in part by ONR Physical Oceanography and Coastal Programs, in part by the supporting service grant from NSWC Carderock Division and in part by the NASA RTOP Oceanic Processes Program. NASA has filed for a patent for the algorithms of the EMD and Hilbert spectral analysis methods.

References

Aki, K. & Richards, P. G. 1980 *Quantitative seismology.* San Francisco, CA: Freeman.

Anderson, J. G. 1991 Strong motion seismology. *Rev. Geophys. Suppl.* **29**, 700–720.

Bedrosian, E. 1963 A product theorem for Hilbert transform. *Proc. IEEE* **51**, 868–869.

Bendat, J. S. & Piersol, A. G. 1986 *Random data: analysis and measurement procedures*, 2nd edn. New York: Wiley.

Benjamin, T. B. & Feir, J. E. 1967. The disintegration of wavetrains on deep water. I. Theory. *J. Fluid Mech.* **27**, 417–430.

Boashash, B. 1992 Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. *Proc. IEEE* **80**, 520–538.

Boore, D. M. & Atlinson, G. M. 1987 Stochastic prediction of ground motion and spectral response parameters at hard-rock sites in Easter North America. *Bull. Seis. Soc. Am.* **77**, 440–467.

Brigham, E. O. 1974 *The fast Fourier transform.* Englewood Cliffs, NJ: Prentice-Hall.

Brockwell, P. J. & Davis, R. A. 1991 *Time series: theory and methods.* New York: Springer.

Byod, J. P. 1980 The nonlinear equatorial Kelvin waves. *J. Phys. Oceanogr.* **10**, 1–11.

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

Claasen, T. A. C. M. & Mecklenbräuker, W. F. G. 1980*a* The Wigner distribution—a tool for time-frequency signal analysis. I. Continuous time signals. *Philips Jl Research.* **35**, 217–250.

Claasen, T. A. C. M. & Mecklenbräuker, W. F. G. 1980*b* The Wigner distribution—a tool for time-frequency signal analysis. II. Discrete time signals. *Philips Jl Research.* **35**, 276–300.

Claasen, T. A. C. M. & Mecklenbräuker, W. F. G. 1980*c* The Wigner distribution—a tool for time-frequency signal analysis. III. Relations with other time-frequency signal transformations. *Philips Jl Research.* **35**, 372–389.

Cohen, L. 1995 *Time-frequency analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Copson, E. T. 1967 *Asymptotic expansions.* Cambridge University Press.

Drazin, P. G. 1992 *Nonlinear systems.* Cambridge University Press.

Farge, M. 1992 Wavelet transforms and their applications to turbulence. *Ann. Rev. Fluid Mech.* **24**, 395–457.

Gabor, D. 1946 Theory of communication. *Proc. IEE* **93**, 429–457.

Geller, R. J. & Ohminato, T. 1994 Computation of synthesis seismograms and their partial derivatives for heterogeneous media with arbitrary neutral boundary conditions using the direct solution method. *Geophys. J. Int.* **116**, 421–446.

Hadley, P. K., Askar, A. & Cakmak, A. S. 1991 Subsoil geology and soil amplification in Mexico Valley. *Soil Dyn. Earthq. Engng* **10**, 101–109.

Hu, Y. X., Liu, S. C. & Dong, W. 1996 *Earthquake engineering.* London: Chapman & Hall.

Huang, N. E. & Long, S. R. 1980 An experimental study of surface elevation probability distribution and statistics of wind generated waves. *J. Fluid Mech.* **101**, 179–200.

Huang, N. E., Leitao, C. D. & Parra, C. G. 1978 Ocean surface measurement using elevation from GEOS-3 altimeter. *J. Geophys. Res.* **83**, 673–674, 682.

Huang, N. E., Tung, C. C. & Long, S. R. 1990*a* Wave spectra. *The Sea* **9**, 197–237.

Huang, N. E., Tung, C. C. & Long, S. R. 1990*b* The probability structure of the ocean surface. *The Sea* **9**, 335–366.

Huang, N. E., Long, S. R. & Shen, Z. 1996 The mechanism for frequency downshift in nonlinear wave evolution. *Adv. Appl. Mech.* **32**, 59–111.

Kevorkian, J. 1966 *Lectures in applied mathematics*, vol. 7, Space Mathematics III, pp. 206–275. Providence, RI: American Mathematical Society.

Lake, B. M. & Yuan, H. C. 1978 A new model for nonlinear gravity waves. I. Physical model and experimental evidence. *J. Fluid Mech.* **88**, 33–62.

Lamb, H. 1932 *Hydrodynamics*. Cambridge University Press.

Lin, Y. K. & Cai, G. Q. 1995 *Probabilistic structural dynamics: advanced theory and applications*. New York: McGraw-Hill.

Liu, S. C. 1970 Evolutionary power spectral density of strong earthquake. *Bull. Seism. Soc. Am.* **60**, 891–900.

Liu, S. C. 1971 Time varying spectra and linear transformation. *Bell System Tech. Jl* **50**, 2365–2374.

Liu, S. C. 1973 An approach to time-varying spectral analysis. *J. EM. Div. ASCE* **98**, 245–253.

Liu, S. D. & Liu, S. K. 1995 *Solitary waves and turbulence*. Shanghai Scientific and Technological Education.

Long, S. R., Lai, R. J., Huang, N. E. & Spedding, G. R. 1993 Blocking and trapping of waves in an inhomogeneous flow. *Dyn. Atmos. Oceans* **20**, 79–106.

Long, S. R., Huang, N. E., Tung, C. C., Wu, M. L., Lin, R. Q., Mollo-Christensen, E. & Yuan, Y. 1995 The Hilbert techniques: an alternate approach for non-steady time series analysis. *IEEE Geoscience Remote Sensing Soc. Lett.* **3**, 6–11.

Longuet-Higgins, M. S. 1957 The statistical analysis of random moving surface. *Phil. Trans. R. Soc. Lond.* A **249**, 321–387.

Longuet-Higgins, M. S. 1978 The instabilities of gravity waves of finite amplitude in deep water. II. Subharmonics. *Proc. R. Soc. Lond.* A **360**, 489–505.

Lorenz, E. N. 1963 Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.

Melville, W. K. 1983 Wave modulation and breakdown. *J. Fluid Mech.* **128**, 489–506.

Miller, L. & Cheney, R. E. 1990 Large-scale meridional transport in the tropic Pacific Ocean during the 1986–87 El Niño from GEOSAT. *J. Geophys. Res.* **95**, 17 905–17 919.

Miller, L., Cheney, R. E. & Douglas, B. C. 1988 GEOSAT altimeter observation of Kelvin waves and the 1986–1987 El Niño. *Science* **239**, 52–54.

Newmark, N. M. & Rosenblueth, E. 1971 *Fundamentals of earthquake engineering*. Englewood Cliffs, NJ: Prentice-Hall.

Oppenheim, A. V. & Schafer, R. W. 1989 *Digital signal processing*. Englewood Cliffs, NJ: Prentice-Hall.

Priestley, M. B. 1965 Evolutionary spectra and non-stationary processes. *J. R. Statist. Soc.* B **27**, 204–237.

Ramamonjiarisoa, A. & Mollo-Christensen, E. 1979 Modulation characteristics of sea surface waves. *J. Geophys. Res.* **84**, 7 769–7 775.

Rice, S. O. 1944*a* Mathematical analysis of random noise. *Bell Sys. Tech. Jl* **23**, 282–310.

Rice, S. O. 1944*b* Mathematical analysis of random noise. II. Power spectrum and correlation functions. *Bell Sys. Tech. Jl* **23**, 310–332.

Rice, S. O. 1945*a* Mathematical analysis of random noise. III. Statistical properties of random noise currents. *Bell Sys. Tech. Jl* **24**, 46–108.

Rice, S. O. 1945*b* Mathematical analysis of random noise. IV. Noise through nonlinear devices. *Bell Sys. Tech. Jl* **24**, 109–156.

Robinson, A. R., Huang, N. E., Leitao, C. D. & Parra, C. G. 1983 A study of the variability of ocean currents in the Northwestern Atlantic using satellite altimetry. *J. Phys. Oceanogr.* **13**, 565–585.

Scherbaum, F. 1994 *Basic concepts in digital processing for seismologists*. Berlin: Springer.

Schwartz, M., Bennett, W. R. & Stein, S. 1966 *Communications systems and techniques*. New York: McGraw-Hill.

Shekel, J. 1953 'Instantaneous' frequency. *Proc. IRE* **41**, 548.

Simpson, J. J. 1991 Oceanographic and atmospheric applications of spatial statistics and digital image analysis. In *Spatial statistics and digital image analysis*. Washington, DC: National Academy.

Spedding, G. R., Browand, F. K., Huang, N. E. & Long, S. R. 1993 A 2D complex wavelet analysis of an unsteady wind-generated surface wave field. *Dyn. Atmos. Oceans* **20**, 55–77.

Titchmarsh, E. C. 1948 *Introduction to the theory of Fourier integrals*. Oxford University Press.

Vautard, R. & Ghil, M. 1989 Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica* D **35**, 395–424.

Whitham, G. B. 1975 *Linear and nonlinear waves*. New York: Wiley.

Yen, N. C. 1994 Wave packet decomposition. *J. Acoust. Soc. Am.* **95**, 889–896.

Zheng, Q., Yan, X. H., Ho, C. R. & Tai, C. K. 1994 The effects of shear flow on propagation of Rossby waves in the equatorial oceans. *J. Phys. Oceanogr.* **24**, 1680–1686.

Zheng, Q., Yan, X. H., Ho, C. R. & Tai, C. K. 1995 Observation of equatorially trapped waves in the Pacific using GEOSAT altimeter data. *Deep Sea Res.* **42**, 797–817.