

CHAPTER 1

INTRODUCTION TO THE HILBERT–HUANG TRANSFORM AND ITS RELATED MATHEMATICAL PROBLEMS

Norden E. Huang

The Hilbert–Huang transform (HHT) is an empirically based data-analysis method. Its basis of expansion is adaptive, so that it can produce physically meaningful representations of data from nonlinear and non-stationary processes. The advantage of being adaptive has a price: the difficulty of laying a firm theoretical foundation. This chapter is an introduction to the basic method, which is followed by brief descriptions of the recent developments relating to the normalized Hilbert transform, a confidence limit for the Hilbert spectrum, and a statistical significance test for the intrinsic mode function (IMF). The mathematical problems associated with the HHT are then discussed. These problems include (i) the general method of adaptive data-analysis, (ii) the identification methods of nonlinear systems, (iii) the prediction problems in nonstationary processes, which is intimately related to the end effects in the empirical mode decomposition (EMD), (iv) the spline problems, which center on finding the best spline implementation for the HHT, the convergence of EMD, and two-dimensional EMD, (v) the optimization problem or the best IMF selection and the uniqueness of the EMD decomposition, (vi) the approximation problems involving the fidelity of the Hilbert transform and the true quadrature of the data, and (vii) a list of miscellaneous mathematical questions concerning the HHT.

1.1. Introduction

Traditional data-analysis methods are all based on linear and stationary assumptions. Only in recent years have new methods been introduced to analyze nonstationary and nonlinear data. For example, wavelet analysis and the Wagner-Ville distribution (Flandrin 1999; Gröchenig 2001) were designed for linear but nonstationary data. Additionally, various nonlinear time-series-analysis methods (see, for example, Tong 1990; Kantz and Schreiber 1997; Diks 1999) were designed for nonlinear but stationary and deterministic systems. Unfortunately, in most real systems, either natural or even man-made ones, the data are most likely to be both nonlinear and nonstationary. Analyzing the data from such a system is a daunting problem. Even the universally accepted mathematical paradigm of data expansion in terms of an *a priori* established basis would need to be eschewed, for the convolution computation of the *a priori* basis creates more problems than solutions. A necessary condition to represent nonlinear and nonstationary data is to have an

adaptive basis. An *a priori* defined function cannot be relied on as a basis, no matter how sophisticated the basis function might be. A few adaptive methods are available for signal analysis, as summarized by Windrow and Stearns (1985). However, the methods given in their book are all designed for stationary processes. For nonstationary and nonlinear data, where adaptation is absolutely necessary, no available methods can be found. How can such a basis be defined? What are the mathematical properties and problems of the basis functions? How should the general topic of an adaptive method for data analysis be approached? Being adaptive means that the definition of the basis has to be data-dependent, an *a posteriori*-defined basis, an approach totally different from the established mathematical paradigm for data analysis. Therefore, the required definition presents a great challenge to the mathematical community. Even though challenging, new methods to examine data from the real world are certainly needed. A recently developed method, the Hilbert–Huang transform (HHT), by Huang et al. (1996, 1998, 1999) seems to be able to meet some of the challenges.

The HHT consists of two parts: empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA). This method is potentially viable for nonlinear and nonstationary data analysis, especially for time-frequency-energy representations. It has been tested and validated exhaustively, but only empirically. In all the cases studied, the HHT gave results much sharper than those from any of the traditional analysis methods in time-frequency-energy representations. Additionally, the HHT revealed true physical meanings in many of the data examined. Powerful as it is, the method is entirely empirical. In order to make the method more robust and rigorous, many outstanding mathematical problems related to the HHT method need to be resolved. In this section, some of the problems yet to be faced will be listed, in the hope of attracting the attention of the mathematical community to this interesting, challenging and critical research area. Some of the problems are easy and might be resolved in the next few years; others are more difficult and will probably require much more effort. In view of the history of Fourier analysis, which was invented in 1807 but not fully proven until 1933 (Plancherel 1933), it should be anticipated that significant time and effort will be required. Before discussing the mathematical problem, a brief introduction to the methodology of the HHT will first be given. Readers interested in the complete details should consult Huang et al. (1998, 1999).

1.2. The Hilbert–Huang transform

The development of the HHT was motivated by the need to describe nonlinear distorted waves in detail, along with the variations of these signals that naturally occur in nonstationary processes. As is well known, the natural physical processes are mostly nonlinear and nonstationary, yet the data analysis methods provide very limited options for examining data from such processes. The available methods are either for linear but nonstationary, or nonlinear but stationary and statistically de-

terministic processes, as stated above. To examine data from real-world nonlinear, nonstationary and stochastic processes, new approaches are urgently needed, for nonlinear processes need special treatment. The past approach of imposing a linear structure on a nonlinear system is just not adequate. Other than periodicity, the detailed dynamics in the processes from the data need to be determined because one of the typical characteristics of nonlinear processes is their intra-wave frequency modulation, which indicates the instantaneous frequency changes within one oscillation cycle. As an example, a very simple nonlinear system will be examined, given by the non-dissipative Duffing equation as

$$\frac{d^2x}{dt^2} + x + \epsilon x^3 = \gamma \cos(\omega t), \quad (1.1)$$

where ϵ is a parameter not necessarily small, and γ is the amplitude of a periodic forcing function with a frequency ω . In (1.1), if the parameter ϵ were zero, the system would be linear, and the solution would be easily found. However, if ϵ were non-zero, the system would be nonlinear. In the past, any system with such a parameter could be solved by using perturbation methods, provided that $\epsilon \ll 1$. However, if ϵ is not small compared to unity, then the system becomes highly nonlinear, and new phenomena such as bifurcations and chaos will result. Then perturbation methods are no longer an option; numerical solutions must be attempted. Either way, (1.1) represents one of the simplest nonlinear systems; it also contains all the complications of nonlinearity. By rewriting the equation in a slightly different form as

$$\frac{d^2x}{dt^2} + x(1 + \epsilon x^2) = \gamma \cos(\omega t), \quad (1.2)$$

its features can be better examined. Then the quantity within the parenthesis can be regarded as a variable spring constant, or a variable pendulum length. As the frequency (or period) of the simple pendulum depends on the length, it is obvious that the system given in (1.2) should change in frequency from location to location, and time to time, even within one oscillation cycle. As Huang et al. (1998) pointed out, this intra-frequency frequency variation is the hallmark of nonlinear systems. In the past, when the analysis was based on the linear Fourier analysis, this intra-wave frequency variation could not be depicted, except by resorting to harmonics. Thus, any nonlinear distorted waveform has been referred to as “harmonic distortions.” Harmonics distortions are a mathematical artifact resulting from imposing a linear structure on a nonlinear system. They may have mathematical meaning, but not a physical meaning (Huang et al. 1999). For example, in the case of water waves, such harmonic components do not have any of the real physical characteristics of a real wave. The physically meaningful way to describe the system is in terms of the instantaneous frequency, which will reveal the intra-wave frequency modulations.

The easiest way to compute the instantaneous frequency is by using the Hilbert transform, through which the complex conjugate $y(t)$ of any real valued function

$x(t)$ of L^p class can be determined (see, for example, Titchmarsh 1950) by

$$\mathcal{H}[x(t)] = \frac{1}{\pi} \text{PV} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau, \quad (1.3)$$

in which the PV indicates the principal value of the singular integral. With the Hilbert transform, the analytic signal is defined as

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)}, \quad (1.4)$$

where

$$a(t) = \sqrt{x^2 + y^2}, \quad \text{and} \quad \theta(t) = \arctan\left(\frac{y}{x}\right). \quad (1.5)$$

Here, $a(t)$ is the instantaneous amplitude, and θ is the phase function, and the instantaneous frequency is simply

$$\omega = \frac{d\theta}{dt}. \quad (1.6)$$

A description of the Hilbert transform with the emphasis on its many mathematical formalities can be found in Hahn (1996). Essentially, (1.3) defines the Hilbert transform as the convolution of $x(t)$ with $1/t$; therefore, (1.3) emphasizes the local properties of $x(t)$. In (1.4), 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, defining the instantaneous frequency still involves considerable controversy. In fact, a sensible instantaneous frequency cannot be found through this method for obtaining an arbitrary function. A straightforward application, as advocated by Hahn (1996), will only lead to the problem of having frequency values being equally likely to be positive and negative for any given dataset. As a result, the past applications of the Hilbert transform are all limited to the narrow band-passed signal, which is narrow-banded with the same number of extrema and zero-crossings. However, filtering in frequency space is a linear operation, and the filtered data will be stripped of their harmonics, and the result will be a distortion of the waveforms. The real advantage of the Hilbert transform became obvious only after Huang et al. (1998) introduced the empirical mode decomposition method.

1.2.1. The empirical mode decomposition method (the sifting process)

As discussed by Huang et al. (1996, 1998, 1999), the empirical mode decomposition method is necessary to deal with data from nonstationary and nonlinear processes. In contrast to almost all of the previous methods, this new method is intuitive, direct, and adaptive, with an *a posteriori*-defined basis, from the decomposition method, based on and derived from the data. The decomposition is based on the

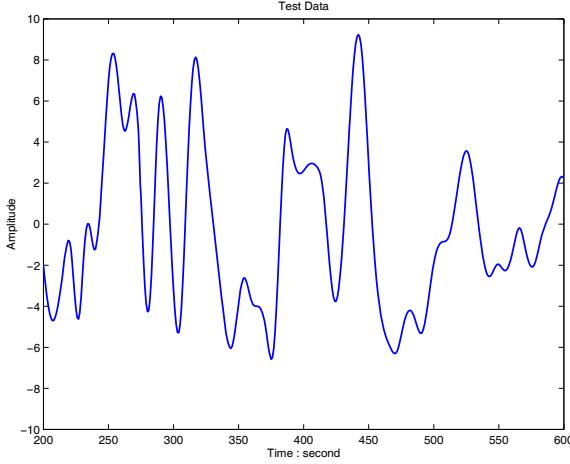


Figure 1.1: The test data.

simple assumption that any data consists of different simple intrinsic modes of oscillations. Each intrinsic mode, linear or nonlinear, represents a simple oscillation, which will have the same number of extrema and zero-crossings. Furthermore, the oscillation will also be symmetric with respect to the “local mean.” At any given time, the data may have many different coexisting modes of oscillation, one superimposing on the others. The result is the final complicated data. Each of these oscillatory modes is represented by an intrinsic mode function (IMF) with the following definition:

- (1) in the whole dataset, 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.

An IMF represents a simple oscillatory mode as a counterpart to the simple harmonic function, but it is much more general: instead of constant amplitude and frequency, as in a simple harmonic component, the IMF can have a variable amplitude and frequency as functions of time. With the above definition for the IMF, one can then decompose any function as follows: take the test data as given in Fig. 1.1; identify all the local extrema, then connect all the local maxima by a cubic spline line as shown in 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, as shown in Fig. 1.2. Their mean is designated as m_1 , also shown in Fig. 1.2, and the difference between the data and m_1 is the first component h_1 shown in Fig. 1.3; i. e.,

$$h_1 = x(t) - m_1 . \quad (1.7)$$

The procedure is illustrated in Huang et al. (1998).

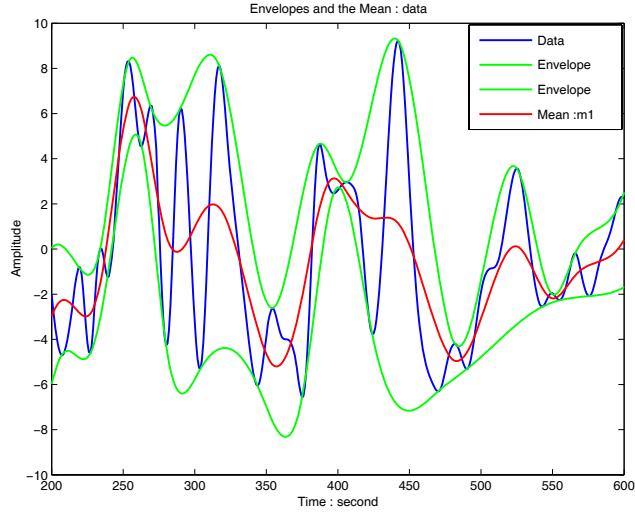


Figure 1.2: The data (blue) upper and lower envelopes (green) defined by the local maxima and minima, respectively, and the mean value of the upper and lower envelopes given in red.

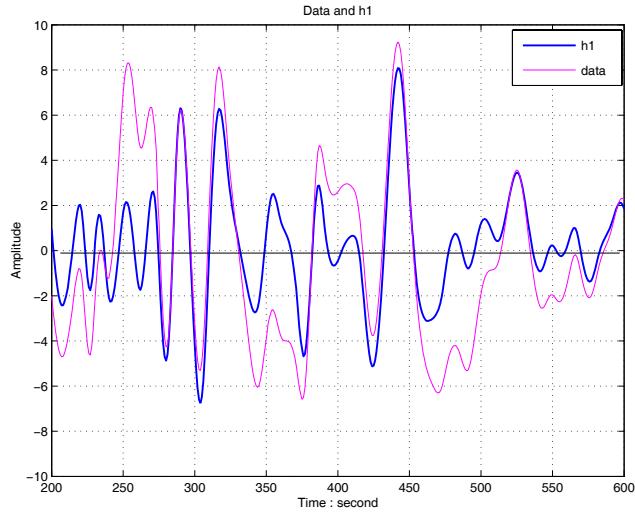


Figure 1.3: The data (red) and h_1 (blue).

Ideally, h_1 should satisfy the definition of an IMF, for the construction of h_1 described above should have made it symmetric and have all maxima positive and all minima negative. However, 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. After the first round of sifting, the hump may become a local maximum. New extrema generated in this way actually reveal

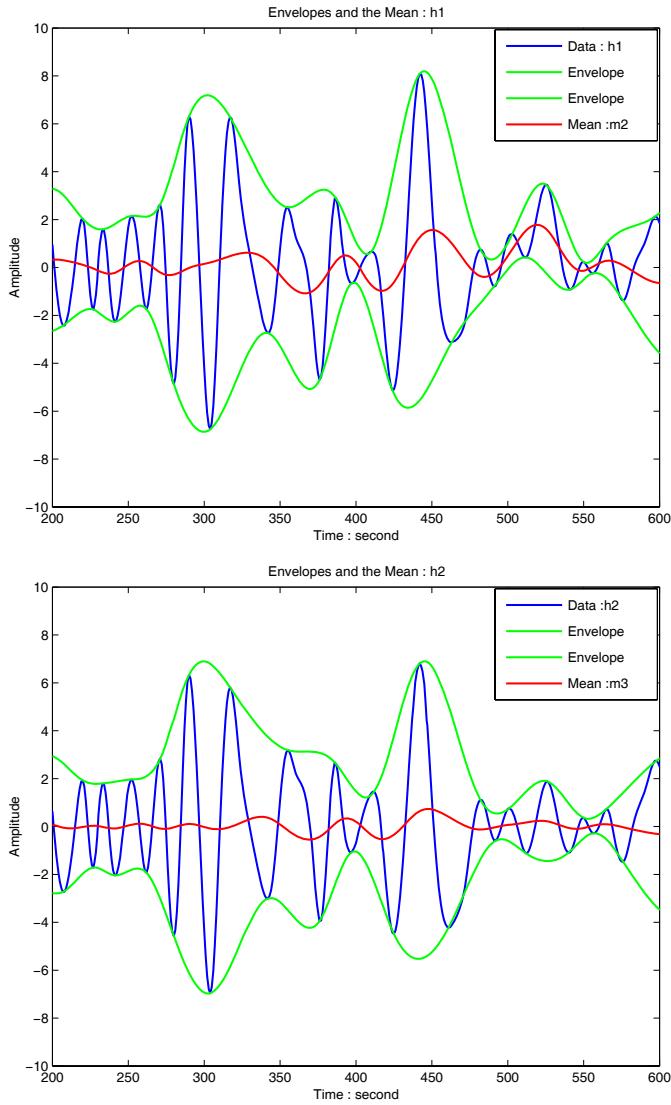


Figure 1.4: (a, top) Repeated sifting steps with h_1 and m_2 . (b, bottom) Repeated sifting steps with h_2 and m_3 .

the proper modes lost in the initial examination. In fact, with repeated siftings, the sifting process can recover signals representing low-amplitude riding waves.

The sifting process serves two purposes: to eliminate riding waves, and to make the wave profiles more symmetric. While the first purpose must be achieved for the Hilbert transform to give a meaningful instantaneous frequency, the second purpose must also be achieved in case the neighboring wave amplitudes have too large a disparity. Toward these ends, the sifting process has to be repeated as many

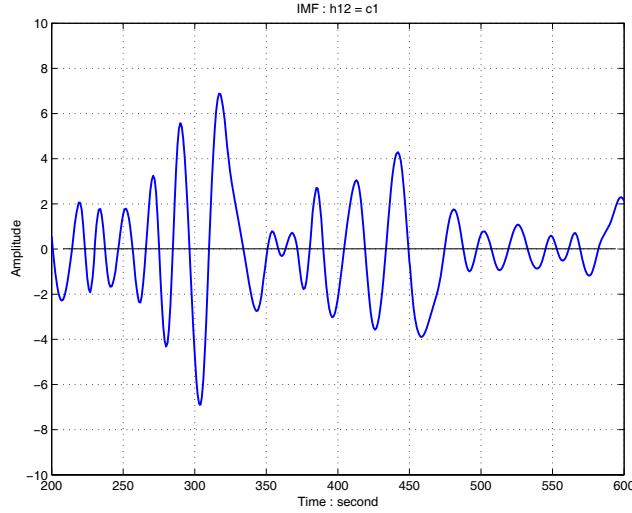


Figure 1.5: The first IMF component c_1 after 12 steps.

times as is required to reduce the extracted signal to an IMF. In the subsequent sifting processes, h_1 can be treated only as a proto-IMF. In the next step, it is treated as the data; then,

$$h_{11} = h_1 - m_{11}. \quad (1.8)$$

After repeated siftings in this manner, shown in Fig. 1.4a,b, up to k times, h_{1k} becomes an IMF; that is,

$$h_{1k} = h_{1(k-1)} - m_{1k}; \quad (1.9)$$

then, it is designated as

$$c_1 = h_{1k}, \quad (1.10)$$

the first IMF component from the data shown in Fig. 1.5. Here, a critical decision must be made: the stoppage criterion. Historically, two different criteria have been used: The first one was used in Huang et al. (1998). This stoppage criterion is determined by using a Cauchy type of convergence test. Specifically, the test requires the normalized squared difference between two successive sifting operations defined as

$$SD_k = \frac{\sum_{t=0}^T |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^T h_{k-1}^2} \quad (1.11)$$

to be small. If this squared difference SD_k is smaller than a predetermined value, the sifting process will be stopped. This definition seems to be rigorous, but it is very difficult to implement in practice. Two critical questions need to be resolved: first,

the question of how small is small enough needs an answer. Second, this criterion does not depend on the definition of the IMFs. The squared difference might be small, but nothing guarantees that the function will have the same numbers of zero-crossings and extrema, for example. These shortcomings prompted Huang et al. (1999, 2003) to propose a second criterion based on the agreement of the number of zero-crossings and extrema. Specifically, a S -number is pre-selected. The sifting process will stop only after S consecutive times, when the numbers of zero-crossings and extrema stay the same and are equal or differ at most by one. This second choice has its own difficulty: how to select the S number. Obviously, any selection is *ad hoc*, and a rigorous justification is needed.

In a recent study of this open-ended sifting, Huang et al. (2003) used the many possible choices of S -numbers to form an ensemble of IMF sets, from which an ensemble mean and confidence were derived. Furthermore, through comparisons of the individual sets with the mean, Huang et al. established an empirical guide. For the optimal siftings, the range of S -numbers should be set between 4 and 8. More details will be given later.

Now assume that a stoppage criterion was selected, and that the first IMF c_1 was found. Overall, c_1 should contain the finest scale or the shortest period component of the signal. It follows that c_1 can be separated from the rest of the data by

$$r_1 = x(t) - c_1. \quad (1.12)$$

Since the residue r_1 still contains longer period variations in the data, as shown in Fig. 1.6, it is treated as the new data and subjected to the same sifting process as described above. This procedure can be repeated with all the subsequent r_j 's, and

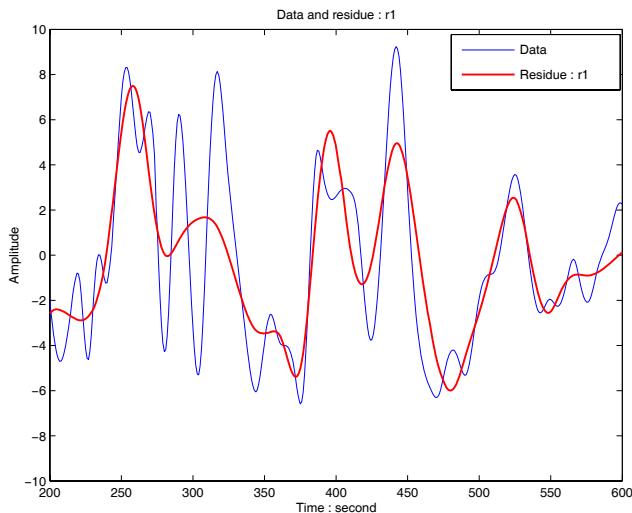


Figure 1.6: The original data (blue) and the residue r_1 .

the result is

$$\begin{aligned} r_2 &= r_1 - c_1 \\ &\vdots \\ r_2 &= r_1 - c_1 \end{aligned} \tag{1.13}$$

The sifting process can be stopped finally 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 IMFs can be extracted. Even for data with zero mean, the final residue still can be different from zero. If the data have a trend, the final residue should be that trend. By summing up (1.12) and (1.13), we finally obtain

$$x(t) = \sum_{j=1}^n c_j + r_n. \tag{1.14}$$

Thus, a decomposition of the data into n -empirical modes is achieved, and a residue r_n obtained 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; the EMD technique needs only the locations of the local extrema. The zero reference for each component will be generated by the sifting process. Without the need for the zero reference, EMD has the unexpected benefit of avoiding the troublesome step of removing the mean values for the large DC term in data with a non-zero mean.

The components of the EMD are usually physically meaningful, for the characteristic scales are defined by the physical data. To understand this point, consider the length-of-day data shown in Fig. 1.7, which measure the deviation of the

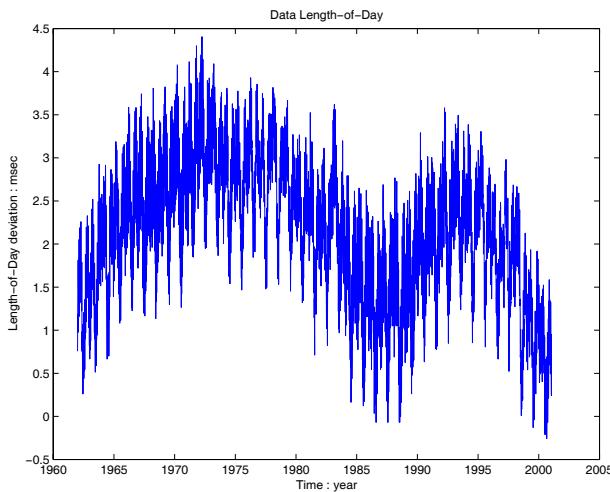


Figure 1.7: The length-of-day data.

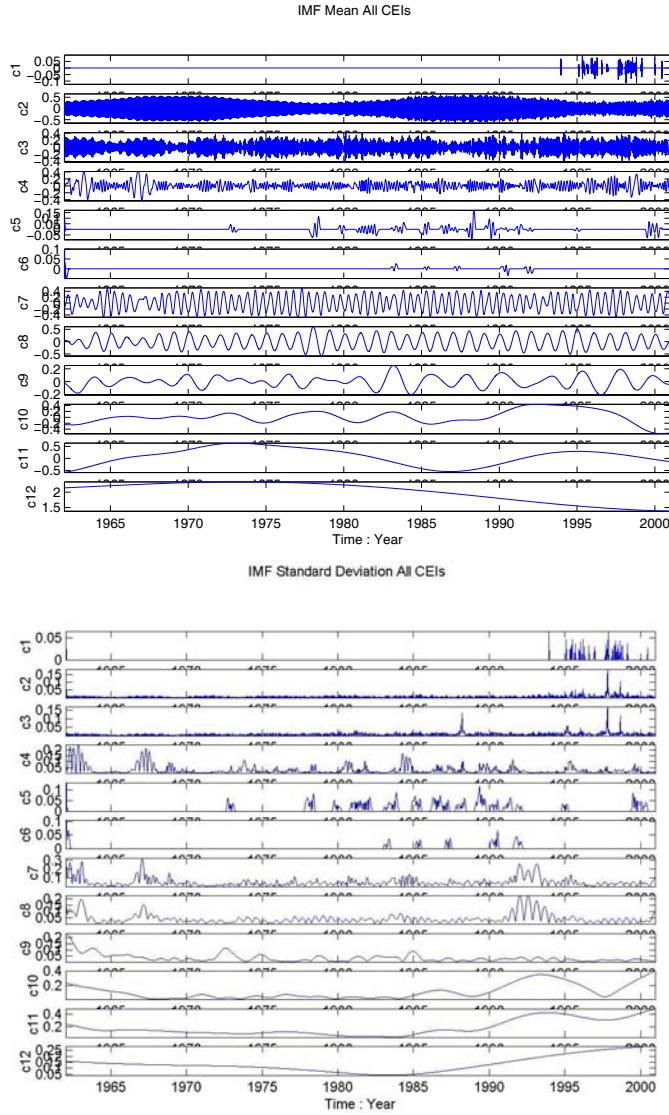


Figure 1.8: (a, top) The mean IMF for nine different siftings. (b, bottom) The standard deviation of the IMF for nine different siftings.

rotational period from the fixed cycle of 24 h. The mean and the standard deviation of the IMFs, given in Fig. 1.8a,b, were obtained after using a different S -number for sifting. The sifting results are quite robust with respect to the selection of a stoppage criteria, as indicated by the low standard deviation values; thus, these IMF results are physically meaningful. The first component represents the very short period of perturbation caused by large-scale storms to the earth's rotational speed;

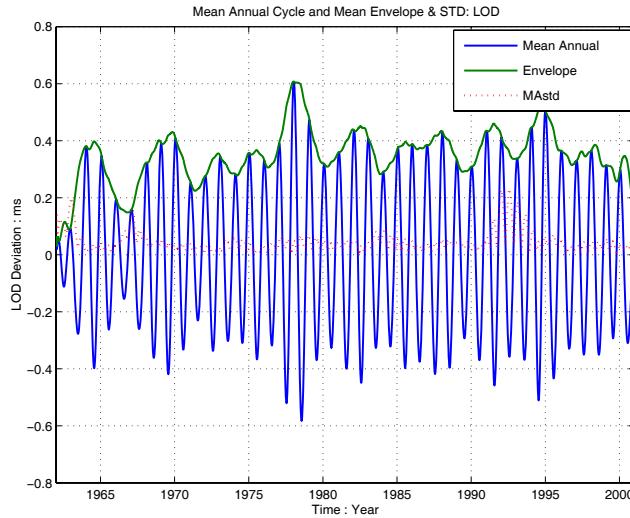


Figure 1.9: The mean annual cycle and its envelope. Each peak of the envelope coincides with an El Niño event.

this perturbation could be measured only after the 1990s by using many Global Positioning Satellites (GPS). The second component represents the half-monthly tides; the eighth component, the annual tidal variations. In fact, a plot of the annual variation by itself in Fig. 1.9 shows that the inter-annual variations are actually associated with the El Niño events. During an El Niño event, the equatorial water in the Pacific Ocean is warmed up, and this warming imparts more energy into the atmosphere. This result, in turn, causes the atmosphere to be more energetic. The resulting increase in angular momentum makes the rotational speed of the earth slow down. Even more surprisingly, the standard deviation values of the different siftings were unusually large during 1965 to 1970, and 1990 to 1995, the periods identified by NOAA as the anomaly periods for El Niño events.

This example established the physical meaning of the IMF components beyond any real doubt. Even more fundamental, the recent studies by Flandrin et al. (2004), Flandrin and Gonçalvès (2004) and Wu and Huang (2004) further established the statistical significance of the IMF components. Thus, whether a given IMF contains significant information or just represents noise can now be tested.

1.2.2. The Hilbert spectral analysis

Having obtained the intrinsic mode function components, one will have no difficulty in applying the Hilbert transform to each IMF component, and in computing the instantaneous frequency according to (1.2)–(1.6). After performing the Hilbert transform on each IMF component, the original data can be expressed as the real

part \Re in the following form:

$$x(t) = \Re \left\{ \sum_{j=1}^n a_j(t) \exp \left[i \int \omega_j(t) dt \right] \right\}. \quad (1.15)$$

Here, the residue r_n has been left out 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 representing a mean offset could be overpowering. In consideration of the uncertainty of the longer trend, and in the interest of obtaining the information contained in the other low-energy but clearly oscillatory components, the final non-IMF component should be left out. However, it could be included if physical considerations justify its inclusion.

Equation (1.15) gives both the amplitude and frequency of each component as functions of time. The same data expanded in a Fourier representation would be

$$x(t) = \Re \left[\sum_{j=1}^n a_j e^{i\omega_j(t)t} \right], \quad (1.16)$$

with both a_j and ω_j as constants. The contrast between (1.15) and (1.16) 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 nonlinear and nonstationary data. With the IMF expansion, the amplitude and the frequency modulations are also clearly separated. Thus, the restriction of the constant amplitude and fixed frequency of the Fourier expansion has been overcome, with a variable amplitude and frequency representation. 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 the more preferred method to represent energy density, then the squared values of the amplitude can be substituted to produce the Hilbert energy spectrum just as well.

The skeleton Hilbert spectrum presentation is more desirable, for it gives more quantitative results. Actually, Bacry et al. (1991) and Carmona et al. (1998) have tried to extract the wavelet skeleton as the local maximum of the continuous wavelet coefficient. Even that approach is still encumbered by the harmonics. If more qualitative results are desired, a fuzzy representation can also be derived from the skeleton Hilbert spectrum presentation by using two-dimensional smoothing. The result is a smoother presentation of time-frequency distribution, but the spurious harmonics are still not needed.

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

$$h(\omega) = \int_0^T H(\omega, t) dt. \quad (1.17)$$

The marginal spectrum offers a measure of the total amplitude (or energy) contribution from each frequency value. This spectrum represents the accumulated amplitude over the entire data span in a probabilistic sense.

The combination of the empirical mode decomposition and the Hilbert spectral analysis is also known as the “Hilbert–Huang transform” (HHT) for short. Empirically, all tests indicate that HHT is a superior tool for time-frequency analysis of nonlinear and nonstationary data. It is based on an adaptive basis, and the frequency is defined through the Hilbert transform. Consequently, there is no need for the spurious harmonics to represent nonlinear waveform deformations as in any of the *a priori* basis methods, and there is no uncertainty principle limitation on time or frequency resolution from the convolution pairs based also on *a priori* basis. A comparative summary of Fourier, wavelet and HHT analyses is given in the following table:

	Fourier	Wavelet	Hilbert
Basis	<i>a priori</i>	<i>a priori</i>	adaptive
Frequency	convolution: global uncertainty	convolution: regional uncertainty	differentiation: local, certainty
Presentation	energy- frequency	energy-time- frequency	energy-time- frequency
Nonlinear	no	no	yes
Nonstationary	no	yes	yes
Feature Extraction	no	discrete: no; continuous: yes	yes
Theoretical base	theory complete	theory complete	empirical

This table shows that the HHT is indeed a powerful method for analyzing data from nonlinear and nonstationary processes: it is based on an adaptive basis; the frequency is derived by differentiation rather than convolution; therefore, it is not limited by the uncertainty principle; it is applicable to nonlinear and nonstationary data and presents the results in time-frequency-energy space for feature extraction.

1.3. Recent developments

Some recent developments in the following areas will be discussed in some detail:

- (1) normalized Hilbert transform
- (2) confidence limit
- (3) statistical significance of the IMFs.

1.3.1. Normalized Hilbert transform

It is well known that although the Hilbert transform exists for any function of L^p class, the phase function of the transformed function will not always yield a physically meaningful instantaneous frequency, as discussed above. Reducing the function into IMFs has improved the chance of getting a meaningful instantaneous frequency, but obtaining IMFs satisfies only the necessary condition; additional limitations have been summarized succinctly in two additional theorems:

First, the Bedrosian theorem (1963) states that the Hilbert transform for the product of two functions $f(t)$ and $h(t)$ can be written as

$$\mathcal{H}[f(t)h(t)] = f(t)\mathcal{H}[h(t)], \quad (1.18)$$

only if the Fourier spectra for $f(t)$ and $h(t)$ are totally disjoint in frequency space, and the frequency range of the spectrum for $h(t)$ is higher than that of $f(t)$. This limitation is critical: if the instantaneous frequency is to be computed from the phase function as defined in (1.3)–(1.6), the data can be expressed in the IMF form as

$$x(t) = a(t) \cos[\theta(t)]; \quad (1.19)$$

then, the Hilbert transform will give us the conjugate part as

$$\mathcal{H}\{a(t) \cos[\theta(t)]\} = a(t)\mathcal{H}\{\cos[\theta(t)]\}. \quad (1.20)$$

However, according to the Bedrosian theorem, (1.20) can be true only if the amplitude is varying so slowly that the frequency spectra of the envelope and the carrier waves are disjoint. This condition has made the application of the Hilbert transform problematic. To satisfy this requirement, Huang and Long (2003) proposed that the IMFs be normalized as follows: start from the data that is already an IMF. First, find all the maxima of the IMFs; then, define the envelope by a spline through all the maxima, and designate the envelope as $E(t)$. Now, normalize the IMF by dividing it by $E(t)$ as

$$Co(t) = \frac{x(t)}{E(t)}, \quad (1.21)$$

where $Co(t)$ should be the carrier function with all local maxima equal to unity. The normalized function of the above example is given in Fig. 1.10d.

This construction should give an amplitude always equal to unity, but anomalies clearly exist, and complications can arise from the spline fitting, which mainly occurs at the point where the amplitude fluctuation is large. Then the spline line could go under the data momentarily and cause the normalized function to have an amplitude greater than unity. Though these conditions are rare, they can occur. Whenever they do, error will certainly occur, which will be discussed next. Even with a perfect normalization, not all of the problems have been solved. The next difficulty is given by the Nuttall theorem.

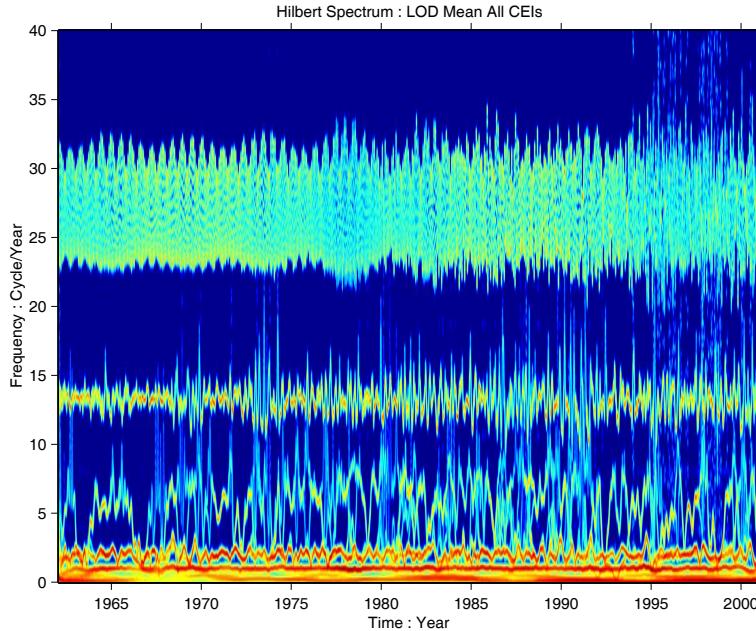


Figure 1.10: The mean Hilbert spectrum.

The Nuttall (1996) theorem states that the Hilbert transform of a cosine is not necessarily a simple 90° phase shift, resulting in the sine function with the same phase function for an arbitrary phase function. Nuttall gave an error bound ΔE , defined as the difference between the Hilbert transform C_h and the quadrature C_q (with a phase shift of exactly 90°) of the function as

$$\Delta E = \int_0^T |C_q(t) - C_h(t)|^2 dt = \int_{-\infty}^0 S_q(\omega) d\omega, \quad (1.22)$$

in which S_q is the Fourier spectrum of the quadrature function. Though the proof of this theorem is rigorous, the result is hardly useful: first, it is expressed in terms of the Fourier spectrum of a still unknown quadrature; and second, it gives a constant error bound over the whole data range. For a nonstationary time series, such a constant bound will not reveal the location of the error on the time axis. With the normalization, Huang and Long (2003) have proposed a variable error bound as follows: the error will be the difference between the squared amplitude of the normalized signal and unity.

The proof of this concept is simple: if the Hilbert transform is exactly the quadrature, by definition the square amplitude will be unity, and the difference zero. If the squared amplitude is not exactly unity, then the Hilbert transform cannot be exactly the quadrature. Two possible complications can contribute to the errors: First, the normalization processes are not clean, as discussed above, so the normal-

ized amplitude could exceed unity, and the error would not be zero. The second complication could come from a highly complicated phase function, as discussed in Huang et al. (1998); then the phase plan will not be a perfect circle. Any deviation from the circle will result in the amplitude being different from unity. Huang and Long (2003) and Huang et al. (2005) conducted detailed comparisons and found the result quite satisfactory. Alternatively, Huang et al. (2005) suggested that the phase function can be found by computing the inverse cosine of the normalized function. The results obtained in this manner were also found to be satisfactory. Two problems, however, still plagued this approach: first, any imperfect normalization will occasionally give the values of the normalized function greater than unity, as discussed above. Under that condition, the inverse cosine will break down. Second, the computation precision requirement is too high near the phase angle 0° and 180° . One can show that the problem of the normalized Hilbert transform always occurs at the location where the amplitude either changes drastically or is very low.

1.3.2. Confidence limit

In data analysis, a confidence limit is always necessary; it provides a measure of assurance about the legitimacy of the results. Therefore, the confidence limit for the Fourier spectral analysis is routinely computed, but the computation is based on ergodic theory, where the data are subdivided into N sections, with spectra from each section being computed. The confidence limit is determined from the statistical spread of the N different spectra. When all the conditions of ergodic theory are satisfied, the temporal average is treated as the ensemble average. Unfortunately, the ergodic condition is satisfied only if the processes are stationary; otherwise, averaging them will not make sense. Huang et al. (2003) have proposed a different approach by utilizing the existence of infinitely many ways to decompose one given function into difference components. Even using EMD, many different sets of IMFs may be obtained by varying the stoppage criteria. For example, Huang et al. (2003) explored the stoppage criterion by changing the S -number. Using the length-of-day data, they varied the S -number from 1 to 20 and found the mean and the standard deviation for the Hilbert spectrum. The confidence limit so derived does not depend on the ergodic theory. If the same data length is used, the spectral resolution is not downgraded in frequency space through sub-dividing of the data into sections. Additionally, Huang et al. have also invoked the intermittence criterion and forced the number of IMFs to be the same for different S -numbers. As a result, Huang et al. were able to find the mean for the specific IMFs shown in Fig. 1.9. Of particular interest are the periods of high standard deviations, from 1965 to 1970, and 1990–1995. These periods are the anomaly periods of the El Niño phenomenon, when the sea-surface-temperature values in the equatorial region were consistently high based on observations, indicating a prolonged heating of the ocean, rather than the changes from warm to cool during the El Niño to La Niña changes.

Finally, from the confidence-limit study, an unexpected result is the determination of the optimal S -number. Huang et al. (2003) computed the difference between the individual cases and the overall mean and found that a range always exists where the differences reach a local minimum. Based on their limited experience from using different datasets, Huang et al. concluded that a S -number in the range of 4 to 8 performed well. Logic also dictates that the S -number should not be high enough to drain all the physical meaning out of the IMF, nor low enough to leave some riding waves remaining in the resulting IMFs.

1.3.3. Statistical significance of IMFs

EMD is a method for separating data into different components according to their scales. The question of the IMFs' statistical significance is always an issue. In data containing noise, how can the noise be separated confidently from the information? These questions were addressed by Flandrin et al. (2004), Flandrin and Gonçalvès (2004), and Wu and Huang (2004) through a study of noise only.

Flandrin et al. (2004) and Flandrin and Gonçalvès (2004) studied the fractional Gaussian noises and found that the EMD is a dyadic filter. These researchers also found that when one plotted the root-mean-squared (RMS) values of the IMFs as a function of the mean period derived from the fractional Gaussian noise on a log-log scale, the results formed a straight line. The slope of the straight line for white noise is -1 ; however, the values change regularly with the different Hurst indices. Based on these results, Flandrin et al. (2004) and Flandrin and Gonçalvès (2004) suggested that the EMD results could be used to determine what kind of noise one was encountering.

Instead of fractional Gaussian noise, Wu and Huang (2004) studied the Gaussian white noise only. They also found the same relationship between the RMS values of the IMFs as a function of the mean period. Additionally, they also studied the statistical properties of the scattering of the data and found the bounds for the noise data distribution analytically. From the scattering, they deduced a 95% bound for the white noise. Therefore, they concluded that when a dataset is analyzed by using EMD, if the RMS-mean period values exist within the noise bounds, the components most likely represent noise. On the other hand, if the mean period-RMS values exceed the noise bounds, then those IMFs must represent statistically significant information. Thus, with the study of noise, Wu and Huang have found a way to discriminate noise from information. They applied this method to the Southern Oscillation Index (SOI) and concluded that the phenomena with the mean periods of 2.0, 3.1, 5.9 and 11.9 years are statistically significant signals.

1.4. Mathematical problems related to the HHT

Over the past few years, the HHT method has gained some recognition. Unfortunately, the full theoretical base has not been fully established. Up to this time,

most of the progress with the HHT has been in its application, while the underlying mathematical problems have been mostly left untreated. All the results have come from case-by-case comparisons conducted empirically. The work with the HHT is presently at the stage corresponding historically to that of wavelet analysis in the earlier 1980s, producing great results but waiting for mathematical foundations on which to rest its case. The work is waiting for someone like Daubechies (1992) to lay the mathematical foundation for the HHT as was done for wavelets. The outstanding mathematical problems at the forefront at the present time are as follows:

- (1) Adaptive data analysis methodology in general
- (2) Nonlinear system identification methods
- (3) Prediction problem for nonstationary processes (end effects)
- (4) Spline problems (best spline implementation for the HHT, convergence and 2-D)
- (5) Optimization problems (the best IMF selection and uniqueness)
- (6) Approximation problems (Hilbert transform and quadrature)
- (7) Miscellaneous questions concerning the HHT.

1.4.1. Adaptive data-analysis methodology

Most data-analysis methods are not adaptive. The established approach is to define a basis (such as trigonometric functions in Fourier analysis, for example). Once the basis is determined, the analysis is reduced to a convolution computation. This well established paradigm is specious, for we have no *a priori* reason to believe that the basis selected truly represents the underlying processes. Therefore, the results produced will not be informative. This paradigm does, however, provide a definitive quantification with respect to a known metric for certain properties of the data based on the basis selected.

If one gives up this paradigm, no solid foundation remains, yet data-analysis methods need to be adaptive, for their goal is to find out the underlying processes. Only adaptive methods can let the data reveal their underlying processes without any undue influence from the basis. Unfortunately, no mathematical model or precedent exists for such an approach. Recently, adaptive data processing has gained some attention. Some adaptive methods are being developed (see Windrows and Stearns 1985). Unfortunately, most of the methods available depend on feedback; therefore, they are limited to stationary processes. Generalizing these available methods to nonstationary conditions is a difficult task.

1.4.2. Nonlinear system identification

System-identification methods are usually based on having both input and output data. For an ideally controlled system, such datasets are possible, yet for most of the cases studied, natural or man-made, no such luxury involving data is available. All that might be available is a set of measured results. The question is whether the nonlinear characteristics can be identified from the data. This problem might

be ill-posed, for this is very different from the traditional input vs. output comparison. Whether the system can be identified through data only is an open question. Unfortunately, in most natural systems, control of the input is not possible. Additionally, the input and even the system itself are usually unknown. The only data available usually correspond to the output from an unknown system. Can the system be identified? Or short of identification, can anything be learned about the system? The only thing that might be available is some general knowledge of the underlying controlling processes connected with the data. For example, the atmosphere and ocean are all controlled by the generalized equations for fluid dynamics and thermodynamics, which are nonlinear. The man-made structures, though linear under design conditions, will approach nonlinearity under extreme loading conditions. Such *a priori* knowledge could guide the search for the characteristics or the signatures of nonlinearity. The task, however, is still daunting.

So far, most of the definitions or tests for nonlinearity from any data are only necessary conditions: for example, various probability distributions, higher-order spectral analysis, harmonic analysis, and instantaneous frequency (see, for example, Bendat 1990; Priestly 1988; Tong 1990; Kantz and Schreiber 1997). Certain difficulties are involved in making such identifications from observed data only. This difficulty has made some scientists talk about only “nonlinear systems” rather than “nonlinear data.” Such reservations are understandable, but this choice of terms still does not resolve the basic problem: How to identify the system nonlinearity from its output alone. Is doing so possible? Or, is there a definite way to define a nonlinear system from the data (system output) at all? This problem is made even more difficult when the process is also stochastic and nonstationary. With a nonstationary process, the various probabilities and the Fourier-based spectral analyses are all problematic, for those methods are based on global properties, with linear and stationary assumptions.

Through the study of instantaneous frequency, intra-wave frequency modulation has been proposed as an indicator for nonlinearity. More recently, Huang (2003) identified the Teager energy operator (Kaiser 1990; Maragos et al. 1993a,b) as an extremely local and sharp test for harmonic distortions within any IMF derived from data. The combination of these local methods offers some hope for system identification, but the problem is not solved, for this approach is based on the assumption that the input is linear. Furthermore, all these local methods also depend on local harmonic distortion; they cannot distinguish a quasi-linear system from a truly nonlinear system. A test or definition for nonlinear-system identification based on only observed output is urgently needed.

1.4.3. *The prediction problem for nonstationary processes (the end effects of EMD)*

End effects have plagued data analysis from the beginning of any known method. The accepted and timid way to deal with these effects is by using various kinds of

windowing, as is done routinely in Fourier analysis. Although sound in theory, such practices inevitably sacrifice some precious data near the ends. Furthermore, the use of windows becomes a serious hindrance when the data are short. In the HHT approach, the extension of data beyond the existing range is necessary, for a spline through the extrema is used to determine the IMF. Therefore, a method is needed to determine the spline curve between the last available extremum and the end of the data range. Instead of windowing, Huang et al. (1998) introduced the idea of using a “window frame,” a way to extend the data beyond the existing range in order to extract some information from all the data available.

The extension of data, or data prediction, is a risky procedure even for linear and stationary processes. The problem that must be faced is how to make predictions for nonlinear and nonstationary stochastic processes. Here the age-old cozy shelter of the linear, stationary, low-dimension and deterministic assumptions must be abandoned, and the complicated real world must be faced. The data are mostly from high-dimensional nonlinear and nonstationary stochastic systems. Are these systems predictable? What conditions must be imposed on the problem to make it predictable? How well can the accuracy of the predictions be quantified? In principle, data prediction cannot be made based on past data alone. The underlying processes have to be involved. Can the available data be used to extract enough information to make a prediction? This issue is an open question at present.

However, EMD has an advantage to assist the analysis: the whole data span need not be predicted, but only the IMF, which has a much narrower bandwidth, for all the IMFs should have the same number of extrema and zero-crossings. Furthermore, all that is needed is the value and location of the next extrema, not all the data. Such a limited goal notwithstanding, the task is still challenging.

1.4.4. Spline problems (the best spline implementation for HHT, convergence and 2-D)

EMD is a “Reynolds type” decomposition: it is used to extract variations from the data by separating the mean, in this case the local mean, from the fluctuations by using spline fits. Although this approach is totally adaptive, several unresolved problems arise from this approach.

First, among all the spline methods, which one is the best? The answer to this question is critical, for it can be shown easily that all the IMFs other than the first are a summation of spline functions, for from (1.5) to (1.8), it follows that

$$c_1 = x(t) - (m_{1k} + m_{1(k-1)} + \cdots + m_{11} + m_1), \quad (1.23)$$

in which all m functions are generated by splines. Therefore, from equation (1.10),

$$r_1 = x(t) - c_1 = m_{1k} + m_{1(k-1)} + \cdots + m_{11} + m_1 \quad (1.24)$$

is totally determined by splines. Consequently, according to (1.11), all the rest of the IMFs are also totally determined by spline functions. What kind of spline is the best

fit for the EMD? How can one quantify the selection of one spline vs. another? Based on experience, it was found that the higher-order spline functions needed additional subjectively determined parameters, yet the requirement violates the adaptive spirit of the approach. Furthermore, higher-order spline functions could also introduce additional length scales, and they are also more time-consuming in computations. Such shortcomings are why only the cubic spline was selected. However, the possible advantages and disadvantages of higher-order splines and even a taut spline have not been definitively established and quantified.

Finally, the convergence of the EMD method is also a critical issue: is there a guarantee that in finite steps, a function can always be reduced into a finite number of IMFs? All intuitive reasoning and experience suggest that the procedure is converging. Under rather restrictive assumptions, the convergence can even be proved rigorously. The restricted and simplified case studied involved sifting with middle-points only. With further restriction of the middle-point sifting to linearly connected extrema, the convergence proof can be established by *reductio ad absurdum*, and it can be shown that the number of extrema of the residue function has to be less than or equal to that in the original function. The case of equality exists only when the oscillation amplitudes in the data are either monotonically increasing or decreasing. In this case, the sifting may never converge and forever have the same number in the original data and the IMF extracted. The proof is not complete in another aspect: can one prove the convergence once the linear connection is replaced by the cubic spline? Therefore, this approach to the proof is not complete.

Recently, Chen et al. (2004) used a B-spline to implement the sifting. If one uses the B-spline as the base for sifting, then one can invoke the variation-diminishing property of the B-spline and show that the spline curve will have less extrema. The details of this proof still have to be established.

1.4.5. *The optimization problem (the best IMF selection and uniqueness mode mixing)*

Does the EMD generate a unique set of IMFs, or is the EMD method a tool to generate infinite sets of IMFs? From a theoretical point of view, infinitely many ways to decompose a given dataset are available. Experience indicates that the EMD process can generate many different IMF sets by varying the adjustable parameters in the sifting procedure. How are these different sets of IMF related? What is the criterion or criteria to guide the sifting? What is the statistical distribution and significance of the different IMF sets? Therefore, a critical question involves how to optimize the sifting procedure to produce the best IMF set. The difficulty is that it must not sift too many times and drain all the physical meaning out of each IMF component, and, at the same time, one must not sift too few times and fail to get clean IMFs. Recently, Huang et al. (2003) studied the problem of different sifting parameters and established a confidence limit for the resulting IMFs and Hilbert spectrum, but the study was empirical and limited to cubic splines only. Optimization of the sifting process is still an open question.

This question of the uniqueness of the IMF can be traced to this more fundamental one: how to define the IMF more rigorously? The definition given by Huang et al. (1998, 1999) is hard to quantify. Fortunately, the results are quite forgiving: even with the vague definition, the results produced are similar enough. Is it possible to give a rigorous mathematical definition and also find an algorithm that can be implemented automatically?

Finally, there is the problem of IMF mode rectifications. Straightforward implementation of the sifting procedure will produce mode mixing (Huang et al. 1999, 2003), which will introduce aliasing in the IMFs. This mode mixing can be avoided if an “intermittence” test is invoked (see Huang et al. 2003). At this time, one can implement the intermittence test only through interactive steps. An automatic mode rectification program should be able to collect all the relevant segments together and avoid the unnecessary aliasing in the mode mixing. This step is not critical to the HHT, but would be a highly desirable feature of the method.

1.4.6. Approximation problems (the Hilbert transform and quadrature)

One of the conceptual breakthroughs involving the HHT has been the ability to define the instantaneous frequency through the Hilbert transform. Traditionally, two well-known theorems, the Bedrosian theorem (Bedrosian 1963) and the Nuttall theorem (Nuttall 1966), have considered the Hilbert transform to be unusable. The Bedrosian theorem states that the Hilbert transform for the product functions can be expressed only in terms of the product of the low-frequency function and the Hilbert transform of the high-frequency one, if the spectra of the two functions are disjointed. This condition guarantees that the Hilbert transform of $a(t) \cos[\theta(t)]$ is given by $a(t) \sin[\theta(t)]$. The Nuttall theorem (Nuttall 1966), further stipulates that the Hilbert transform of $\cos[\theta(t)]$ is not necessarily $\sin[\theta(t)]$ for an arbitrary function $\theta(t)$. In other words, a discrepancy exists between the Hilbert transform and the perfect quadrature of an arbitrary function $\theta(t)$. Unfortunately, the error bound given by Nuttall (1966) is expressed in terms of the integral of the spectrum of the quadrature, an unknown quantity. Therefore, the single valued error bound cannot be evaluated.

Through research, the restriction of the Bedrosian theorem has been overcome through the EMD process and the normalization of the resulting IMFs (Huang 2003). With this new approach, the error bound given by Nuttall has been improved by expressing the error bound as a function of time in terms of instantaneous energy. These developments are major breakthroughs for the Hilbert transform and its applications. However, the influence of the normalization procedure must be quantified. As the normalization procedure depends on a nonlinear amplification of the data, what is the influence of this amplification on the final results? Even if the normalization is accepted, for an arbitrary $\theta(t)$ function, the instantaneous frequency is only an approximation. How can this approximation be improved?

Also related to the normalization scheme, are other questions concerning the Hilbert transform: for example, what is the functional form of $\theta(t)$ for the Hilbert transform to be the perfect quadrature and also be analytic? If the quadrature is not identical to the Hilbert transform, what is the error bound in the phase function (not in terms of energy as it has been achieved now)?

One possible alternative is to abandon the Hilbert transform and to compute the phase function by using the inverse cosine of the normalized data. Two complications arise from this approach: the first one is the high precision needed for computing the phase function when its value is near $n\pi/2$. The second one is that the normalization scheme is only an approximation; therefore, the normalized functional value can occasionally exceed unity. Either way, some approximations are needed.

1.4.7. Miscellaneous statistical questions concerning HHT

The first question concerns the confidence limit of the HHT results. Traditionally, all spectral analysis results are bracketed by a confidence limit, which gives either a true or false measure of comfort. The traditional confidence limit is established from the ergodicity assumption; therefore, the processes are necessarily linear and stationary. If the ergodic assumptions are abandoned, can a confidence limit still exist without resorting to true ensemble averaging, which is practically impossible for most natural phenomena? The answer seems to be affirmative for Fourier analysis. For HHT, however, a confidence limit has been tentatively established, based on the exploitation of repeated applications of the EMD process with various adjustable parameters, which produces an ensemble of IMF sets. How representative are these different IMFs? How can the definition be made more rigorous? How can the statistical measure for such a confidence limit be quantified?

The second question concerns the degree of nonstationarity. This question has led to another conceptual breakthrough, for the qualitative definition of stationarity has been changed to a quantitative definition of the degree of nonstationarity. In Huang et al. (1998), in addition to a degree of nonstationarity, a degree of statistical nonstationarity was also given. For the degree of statistical nonstationarity, an averaging procedure is required. What is the time scale needed for the averaging?

1.5. Conclusion

Some of the problems encountered in the present state of the research have been discussed. Even though these issues have not been settled, the HHT method is still a very useful tool, but when they are settled, the HHT process will become much more rigorous, and the tool more robust. The author is using the HHT method routinely now, for as Heaviside famously said, when encountering the purist's objections on his operational calculus: "Shall I refuse my dinner because I do not fully understand the process of digestion." For us, the "the process of digestion" consists of fully addressing the questions that we have raised in this chapter. The path is clear; work must now begin.

Finally, the need for a unified framework for nonlinear and nonstationary data analysis is urgent and real. Currently, the field is fragmented, with partisans belonging to one camp or another. For example, researchers engaged in wavelet analysis will not mention the Wagner-Ville distribution method, as if it does not exist (see, for example, any wavelet book). On the other hand, researchers engaged with the Wagner-Ville distribution method will not mention wavelets (see, for example, Cohen 1995). Such a position is unscientific, and unhealthy for the data-analysis community. The time is right for some support from everyone to unify the field and push it forward. A concerted effort should be mounted to attack the problems of nonlinear and nonstationary time series analysis. One logical suggestion is to organize an activity group within SIAM to address all the mathematical and application problems, as well as all the scientific issues related to nonlinear and nonstationary data analysis. This task is worthy of the effort.

References

- Bacry, E., A. Arnéodo, U. Frisch, Y. Gagne, and E. Hopfinger, 1991: Wavelet analysis of fully developed turbulence data and measurement of scaling exponents. *Proc. Turbulence89: Organized Structures and Turbulence in Fluid Mechanics*, M. Lesieur, O. Métais, Eds., Kluwer, 203–215.
- Bedrosian, E., 1963: A product theorem for Hilbert transform. *Proc. IEEE*, **51**, 868–869.
- Bendat, J. S., 1990: *Nonlinear System Analysis and Identification from Random Data*. Wiley Interscience, 267 pp.
- Carmona, R., W. L. Hwang, and B. Torresani, 1998: *Practical Time-Frequency Analysis: Gabor and Wavelet Transform with an Implementation in S*. Academic Press, 490 pp.
- Chen, Q., N. E. Huang, S. Riemenschneider, and Y. Xu, 2005: A B-spline approach for empirical mode decomposition. *Adv. Comput. Math.*, in press.
- Cohen, L., 1995: *Time-Frequency Analysis*. Prentice Hall, 299 pp.
- Daubechies, I., 1992: *Ten Lectures on Wavelets*. CBMS-NSF Series in Applied Mathematics. Vol. 61, SIAM, 357 pp.
- Diks, C., 1999: *Nonlinear Time Series Analysis: Methods and Applications*. World Scientific Press, 180 pp.
- Flandrin, P., 1999: *Time-Frequency/Time-Scale Analysis*. Academic Press, 386 pp.
- Flandrin, P., G. Rilling, and P. Gonçalvès, 2004: Empirical mode decomposition as a filter bank. *IEEE Signal Process. Lett.*, **11**, 112–114.
- Flandrin, P., and P. Gonçalvès, 2004: Empirical mode decompositions as data-driven wavelet-like expansions. *Int. J. Wavelets Multiresolut. Inform. Process.*, **2**, 477–496.
- Gröchenig, K., 2001: *Foundations of Time-Frequency Analysis*. Birkhäuser, 359 pp.
- Hahn, S., 1995: *Hilbert Transforms in Signal Processing*. Artech House, 442 pp.
- Huang, N. E., S. R. Long, and Z. Shen, 1996: The mechanism for frequency downshift in nonlinear wave evolution. *Adv. Appl. Mech.*, **32**, 59–111.

- Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London, Ser. A*, **454**, 903–995.
- Huang, N. E., Z. Shen, and S. R. Long, 1999: A new view of water waves – The Hilbert spectrum. *Annu. Rev. Fluid Mech.*, **31**, 417–457.
- Huang, N. E., 2003: Empirical mode decomposition for analyzing acoustic signal. US Patent 10-073857, pending.
- Huang, N. E., and S. R. Long, 2003: Normalized Hilbert transform and instantaneous frequency. NASA Patent Pending GSC 14,673-1.
- Huang, N. E., M. C. Wu, S. R. Long, S. S. P. Shen, W. Qu, P. Gloersen, and K. L. Fan, 2003: A confidence limit for empirical mode decomposition and Hilbert spectral analysis. *Proc. R. Soc. London, Ser. A*, **459**, 2317–2345.
- Kaiser, J. F., 1990: On Teager’s energy algorithm and its generalization to continuous signals. *Proc. 4th IEEE Signal Process. Workshop*, Sept. 16–19, Mohonk, NY, IEEE, 230 pp.
- Kantz, H., and T. Schreiber, 1999: *Nonlinear Time Series Analysis*. Cambridge University Press, 304 pp.
- Maragos, P., J. F. Kaiser, and T. F. Quatieri, 1993a: On amplitude and frequency demodulation using energy operators. *IEEE Trans. Signal Process.*, **41**, 1532–1550.
- Maragos, P., J. F. Kaiser, and T. F. Quatieri, 1993b: Energy separation in signal modulation with application to speech analysis. *IEEE Trans. Signal Process.*, **41**, 3024–3051.
- Nuttall, A. H., 1966: On the quadrature approximation to the Hilbert transform of modulated signals. *Proc. IEEE*, **54**, 1458–1459.
- Plancherel, M., 1933: Sur les formules de réciprocité du type de Fourier. *J. London Math. Soc., Ser. 1*, **8**, 220–226.
- Priestley, M. B., 1988: *Nonlinear and Nonstationary Time Series Analysis*. Academic Press, 237 pp.
- Titchmarsh, E. C., 1950: *Introduction to the Theory of Fourier Integral*. Oxford University Press, 394 pp.
- Tong, H., 1990: *Nonlinear Time Series Analysis*. Oxford University Press, 564 pp.
- Windrows, B., and S. D. Stearns, 1985: *Adaptive Signal Processing*. Prentice Hall, 474 pp.
- Wu, Z., and N. E. Huang, 2004: A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. R. Soc. London, Ser. A*, **460**, 1597–1611.

Norden E. Huang

Goddard Institute for Data Analysis, Code 614.2, NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA
norden.e.huang@nasa.gov