10.1098/rspa.2003.1221

THE ROYAL SOCIETY

A study of the characteristics of white noise using the empirical mode decomposition method

By Zhaohua Wu[1] and Norden E. Huang[2]

[1]*Center for Ocean–Land–Atmosphere Studies, Suite 302, 4041 Powder Mill Road, Calverton, MD 20705, USA*
(zhwu@cola.iges.org)

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

Received 14 January 2003; accepted 8 August 2003; published online 9 March 2004

Based on numerical experiments on white noise using the empirical mode decomposition (EMD) method, we find empirically that the EMD is effectively a dyadic filter, the intrinsic mode function (IMF) components are all normally distributed, and the Fourier spectra of the IMF components are all identical and cover the same area on a semi-logarithmic period scale. Expanding from these empirical findings, we further deduce that the product of the energy density of IMF and its corresponding averaged period is a constant, and that the energy-density function is chi-squared distributed. Furthermore, we derive the energy-density spread function of the IMF components. Through these results, we establish a method of assigning statistical significance of information content for IMF components from any noisy data. Southern Oscillation Index data are used to illustrate the methodology developed here.

Keywords: empirical mode decomposition; intrinsic mode function; characteristics of white noise; energy-density function; energy-density spread function; statistical significance test

1. Introduction

Noise is an inevitable part of our existence. In scientific study, noise can come in many ways: it could be part of the natural processes generated by local and intermittent instabilities and sub-grid phenomena; it could be part of the concurrent phenomena in the environment where the investigations were conducted; and it could also be part of the sensors and recording systems. As a result, when we face data, we inevitably face an amalgamation of signal and noise,

$$x(t) = s(t) + n(t), \tag{1.1}$$

in which the $x(t)$ our our data, and $s(t)$ and $n(t)$ are our true signal and noise, respectively. Once the noise contaminates the data, it is not a trivial task to remove it. For the obvious cases, when the processes are linear and the noises have distinct time or frequency scales different from those of the true signal, Fourier filters can be employed to separate the noise from the real signal. But, filter methods will fail when the processes are either nonlinear or non-stationary. Then, even if the real signal and

the noises have distinct fundamental frequencies, the harmonics of the fundamental can still mix with the noises. This mixing of harmonics with noises will render the Fourier filter as an ineffective noise-separation method. Under such conditions, the empirical mode decomposition (EMD) method (Huang *et al*. 1998, 1999) can offer some help. EMD is an adaptive method to decompose any data into a set of intrinsic mode function (IMF) components, which become the basis representing the data. As the basis is adaptive, the basis usually offers a physically meaningful representation of the underlying processes. Also because of the adaptive nature of the basis, there is no need for harmonics; therefore, EMD is ideally suited for analysing data from non-stationary and nonlinear processes. Even with these nice properties, EMD still cannot resolve the most complicated cases, when the processes are nonlinear and the noises also have the same time-scale as the signal; their separation becomes impossible. Nevertheless, EMD offers a totally different approach to data decomposition, and we decided to apply it to study the characteristics of the white noise. We will show that, with this approach, we can offer some measure of the information content of signals buried under unknown noise.

The problem of separating noise and signal is complicated and difficult when we do not even know the level of noise in the data. The question then becomes more fundamental: are we looking simply at noise, or are we getting any information in the series of numbers known as data? In this case, knowing the characteristics of the noise is an essential first step before we can attach any significance to the signal eventually extracted from the data. In this paper, we present the results of a numerical experimental study on uniformly distributed white noise using EMD. It should be noted here that the same numerical experiments with the normally distributed white noise lead to the same results described in the following. Therefore, the characteristics of uniformly distributed white noise revealed in this study can also be extended to the normally distributed white noise.

Through this study, we find the following empirical facts: the EMD is effectively a dyadic filter capable of separating the white noise into IMF components having mean periods exactly twice the value of the previous component; the IMF components are all normally distributed; the Fourier spectra of the IMF components are identical in shape, and cover the same area on a semi-logarithmic period scale. Based on these empirical facts, we deduce the following results: first, the product of the energy density of IMF and its corresponding mean period must be a constant. Second, the energy-density function must be chi-squared distributed, since the IMFs are distributed normally. Furthermore, we also derive the statistical bounds for the energy-density spread function of the IMF components. The pertinent results we obtained from this study include an analytical expression of the relationship between the energy density and the mean period of the IMF components derived from the white noise through EMD, an analytic expression of the energy-density distribution and its spreading function. All the analytic expressions are tested against the results produced by Monte Carlo method on a numerically generated random noise. Through these results, we also establish a method to assign statistical significance of the information content for IMF components from any noisy data. Southern Oscillation Index (SOI) data are used to illustrate the methodology developed here.

In this paper, we will first present the numerical experiment and the empirical relationship between the energy density and the mean period. We will then present the empirical result of normally distributed IMF components and the deduced energy

spread function in § 3. This is followed by the statistical test on the information content of the SOI data. Finally, we will offer some discussions and our conclusions that, based on the characteristics of the white noise, the information content of IMF components from a noisy dataset derived from EMD can be objectively determined relative to the white-noise statistics.

2. The numerical experiment and the relation between energy and period

The results in this paper are on the statistical characteristics of the IMF components from the uniformly distributed white noise. For lack of analytical expression for the IMFs *ab initio*, we decided to study the problem empirically. In fact, most of the important results reported subsequently are all based on the empirically determined findings from the numerical experiment; therefore, we will describe the numerical experiment first, followed by the relationship deduced between the energy density and the mean period of the IMF components.

(*a*) *Numerical experiments*

Our main experiment is to study the characteristics of white noise using EMD (Huang *et al*. 1998, 1999). For this purpose, our data consist of numerically generated one-million-term white-noise time-series. In this study, four different random data generators have been tried, including the one installed in the Compaq-DEC Alpha workstations, and three other popular generators described in the *Numerical recipes* (Press *et al*. 1992). Through our experiments, we found that the results are not sensitive to the random number generator. Therefore, we decided to report only the results from the random number generator installed in the Compaq-DEC Alpha workstations, which gives us a uniformly distributed white noise.

For the experiment, the white-noise data generated from the above-mentioned random number generator are decomposed into IMFs by the EMD method (Huang *et al*. 1998, 1999). In the decomposition, we found that, after iteration of the sifting process five times, an IMF would generally satisfy the Cauchy condition proposed in Huang *et al*. (1998); continuing iteration would not change the IMF significantly. However, to guarantee the stability and convergence of the resulting IMFs, we decided to double the sifting process for each IMF to 10 iterations.

In order to study the statistical characteristics, the resulting IMFs of 10^6 data points are then divided into many segments of various lengths. By virtue of the white-noise generator, each segment is considered to be independent of the others. Before discussing the more sophisticated statistical properties, let us examine the obvious result of the mean periods of the IMFs.

(*b*) *Mean periods of IMFs*

Before examining any results, we should list the properties of an IMF as follows: an IMF is any function having symmetric envelopes defined by the local maxima and minima separately, and also having the same number of zero-crossings and extrema. Based on this definition, we can determine the mean period of the function by counting the number of peaks (local maxima) of the function. The results are listed in table 1. The second column of the table gives the total number of local maxima:

Table 1. *EMD as a dyadic filter*

(The sampled white-noise series of 10^6 data points is decomposed into IMFs 1–9. The second column is the number of peaks for each IMF; the third column is the mean period in terms of number of data points; and the fourth column is the corresponding period in years if the data are sampled monthly.)

IMFs	number of peaks	mean period	period in year
1	347 042	2.881	0.240
2	168 176	5.946	0.496
3	83 456	11.98	0.998
4	41 632	24.02	2.000
5	20 877	47.90	3.992
6	10 471	95.50	7.958
7	5 290	189.0	15.75
8	2 658	376.2	31.35
9	1 348	741.8	61.75

the peaks of each IMF in the million data points. The third column shows the mean period measured in terms of the number of data points. An interesting pattern now emerges: the mean period of any IMF component almost exactly doubles that of the previous one, suggesting that the EMD is a dyadic filter. This finding is consistent with the recent result by Flandrin *et al*. (2003). Next, we will examine the more detailed distribution of the energy with respect to the period in the form of spectral function. For this special application, we will consider the Fourier-based spectrum, as it is the one most familiar to most applications.

(*c*) *The Fourier spectra of IMFs*

Let us consider the general properties of the energy density as function of period for the data. For a normalized white-noise time-series, f_j, for $j = 1, \ldots, N$, we can express it either in terms of Fourier series components, or in terms of IMFs, i.e.

$$f_j = \mathrm{Re}\left[\sum_{k=1}^{N} F_k \exp\left(\mathrm{i}\frac{2\pi jk}{N}\right)\right] = \sum_n C_n(j), \tag{2.1}$$

where $\mathrm{i} = \sqrt{-1}$, $C_n(j)$ is the nth IMF and

$$F_k = \sum_{j=1}^{N} f_j \exp\left(-\mathrm{i}\frac{2\pi jk}{N}\right).$$

With this expression, we can define the energy density of the nth IMF as

$$E_n = \frac{1}{N}\sum_{j=1}^{N}[C_n(j)]^2. \tag{2.2}$$

Since the IMFs are nearly orthogonal to each other, with the only leakage begin from nonlinear distortion of the data, we have the total energy for the data, to a

high degree of approximation, as

$$\sum_{n=1}^{N} f_j^2 = \sum_{k=1}^{N} |F_k|^2 = N \sum_{n} E_n. \tag{2.3}$$

As the Fourier spectrum of a white-noise time-series is well known, it may be treated as a constant. Because a synthetically generated white-noise time-series is short in length, however, its Fourier spectrum may be a constant upon which many spikes have been superimposed. When the series is long enough, these spikes will be smoothed out, and the Fourier spectrum approaches a constant. Theoretically, the Fourier spectrum of a white-noise series should be a perfect constant, indicating that the contribution to the total spectrum energy comes from each Fourier component uniformly and equally. The Fourier spectra for the IMFs, however, will not yield a constant white spectrum, for the decomposition through EMD effectively subjects the data through a dyadic filter bank. Therefore, the IMFs are band passed as discussed by Flandrin *et al*. (2003).

Now, let us examine the shape of the Fourier spectra for each IMF. To this end, the Fourier spectra for 200 independent segments of 4000 data points each of all the IMFs are calculated; the individually averaged Fourier spectra for all the IMFs are plotted on a semi-logarithmic scale in figure 1. We can see that all the Fourier spectra except the first one have almost identical shapes in terms of the $\ln T$-axis, where T is the period of a Fourier component. From the figure, it is obvious that the ratios of the neighbouring spectra are almost identically equal to 2, which is consistent with the doubling of mean periods of neighbouring IMFs.

Based on the fact that the spectral shape and area coverage for each spectrum are identical, we can have an integral expression to represent, to the first order of approximation, the functional form of Fourier spectrum for any IMF (except the first one) as

$$\int S_{\ln T,n} \, \mathrm{d} \ln T = \text{const.}, \tag{2.4}$$

where $S_{\ln T,n}$ is the Fourier spectrum of the nth IMF as a function of $\ln T$ and the subscript n representing the nth IMF. Notice that this expression is not scaled properly with respect to energy; it only expresses the fact that the spectra are similar, and that the spectra cover a similar area on the semi-logarithmic period–energy space. To scale the spectra properly with respect to energy, we have to resort to the expression for the total energy. Then, the energy of the nth IMF can be written as

$$NE_n = \int S_{\omega,n} \, \mathrm{d}\omega,$$

where $S_{\omega,n}$ is the Fourier spectrum of the nth IMF in terms of frequency, ω. Through a series of variable changes, from frequency to period and from period to the logarithmic scale, we have

$$NE_n = \int S_{\omega,n} \, \mathrm{d}\omega = \int S_{T,n} \frac{\mathrm{d}T}{T^2} = \int S_{\ln T,n} \frac{\mathrm{d} \ln T}{T} = \frac{\int S_{\ln T,n} \, \mathrm{d} \ln T}{\bar{T}_n}, \tag{2.5}$$

where $S_{T,n}$ is the Fourier spectrum function of period T for the nth IMF, in which

$$\bar{T}_n = \int S_{\ln T,n} \, \mathrm{d} \ln T \bigg(\int S_{\ln T,n} \frac{\mathrm{d} \ln T}{T} \bigg)^{-1}. \tag{2.6}$$

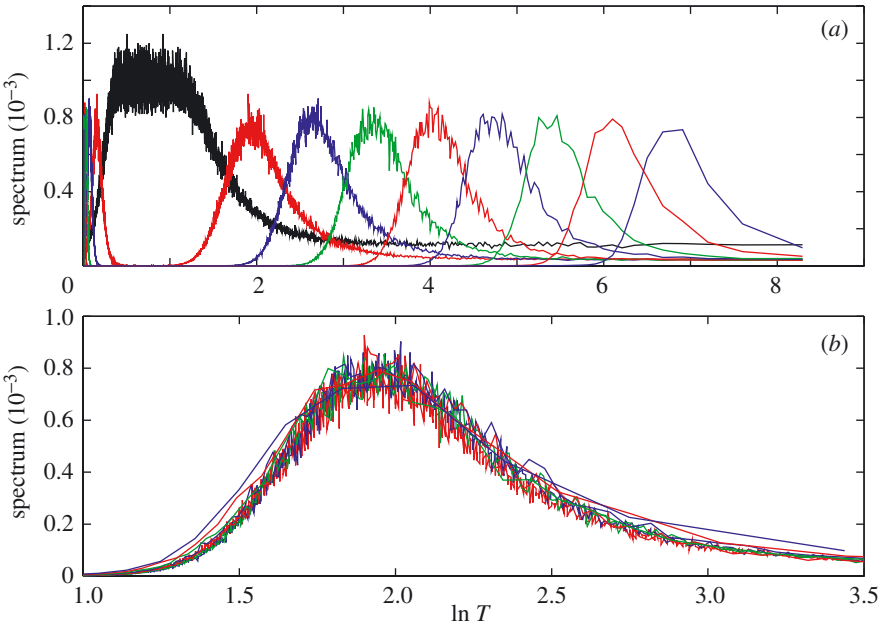

Figure 1. The Fourier spectra of IMFs as a function of the logarithm of the period. In (a), the black line is the Fourier spectrum of the first IMF, and the Fourier spectra of IMFs 2–9 are located away from that of the first IMF. In (b), the Fourier spectra of IMFs 2–9 are redrawn and the Fourier spectrum of the nth IMF is shifted to the left by $(n-2)\ln 2$ for IMFs 3–9.

In this form, $\bar{T}_n$ is the definition of the averaged period calculated from any given spectrum. This value is almost identical to the one derived from counting the zero-crossings. By substituting equation (2.4) into equation (2.5) we obtain a simple equation that relates the energy density, E_n, and the averaged period $\bar{T}_n$, i.e.

$$E_n \bar{T}_n = \text{const.}, \tag{2.7a}$$

$$\ln E_n + \ln \bar{T}_n = \text{const.} \tag{2.7b}$$

This simple relation has already been stated in Wu *et al.* (2001).

If a white-noise series is normalized, one can determine the constant on the right-hand side of equations (2.7). For a normalized white-noise series, there is no loss in generality by assuming that the constant in equation (2.7a) is unity. Therefore, the constant is expected to be zero, i.e.

$$\ln \bar{E}_n + \ln \bar{T}_n = 0, \tag{2.8}$$

where $\bar{E}_n$ is the mean of E_n when N approaches infinity.

The Monte Carlo verification of equation (2.8) is given in figure 2, where the 10^6 points of the numerically generated random dataset are divided into 1000 independent samples with an identical length of 1000 data points each, and each sample is decomposed into IMFs. The energy of each sample is plotted against its averaged period. Groups of points of the same colour are the scattered distributions of the paired values of the nth IMFs of these 1000 samples. The straight black line is the expectation line derived from equation (2.8). Clearly, equation (2.8) offers an excellent fit to these scattered points. Thus, we showed in this example that the energy

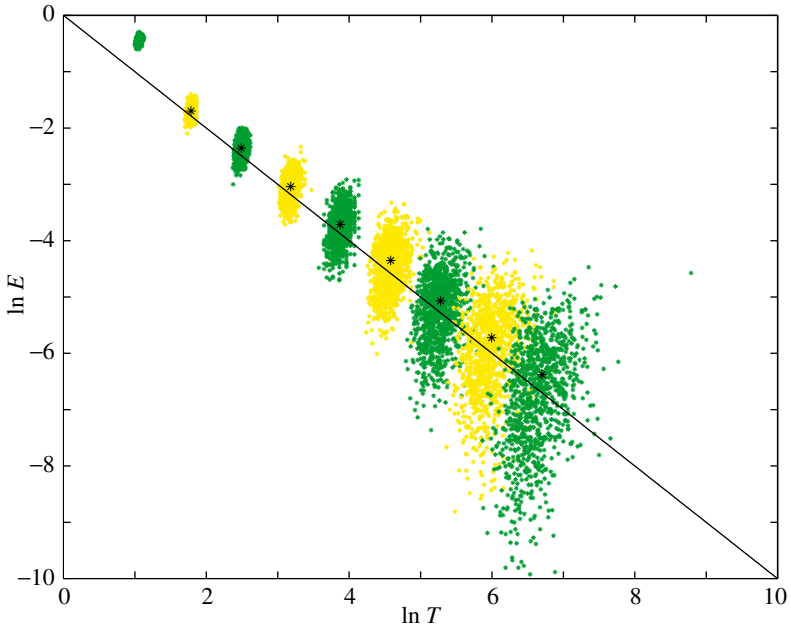

Figure 2. The Monte Carlo verification of the relation between the energy density and the averaged period. The groups of the dots from upper left to the lower right are the energy density as a function of the averaged period for IMFs 1–9 for all 1000 samples with an identical length of 1000 data points. The superimposed black dots are the energy density as a function of the averaged period for IMFs 2–9 for a single sample with 10^6 data points.

density of the IMF and its averaged period follows a hyperbolic function, a result displayed in Wu *et al*. (2001). From figure 2 we can see that the energy density and the averaged period for each IMF of an individual sample could deviate from the expectation line given by equation (2.8), but this deviation is rather small.

3. The energy distribution and spread functions

Having established the relationship between energy density and the averaged period, we will next establish the energy distribution function for each IMF and the spread function of the energy distribution for various percentiles.

(*a*) *Energy distribution function*

The next level of quantification is naturally the spread function. To achieve this quantification, we first examine the probability distribution of an individual IMF. Figure 3 plots the probability density of each IMF for a sample of 50 000 data points. The probability density function for each IMF is approximately normally distributed, which is evident from the superimposed fitted normal distribution function (the red line). This fit is to be expected from the central limit theorem. Indeed, the deviation from the normal distribution function grows as the mode number increases. This is because, in the higher modes, the IMFs contain a smaller number of oscillations; therefore, the number of events decreases and the distribution becomes less smooth. Presumably, when a sample of longer length is used, the IMFs of the higher modes

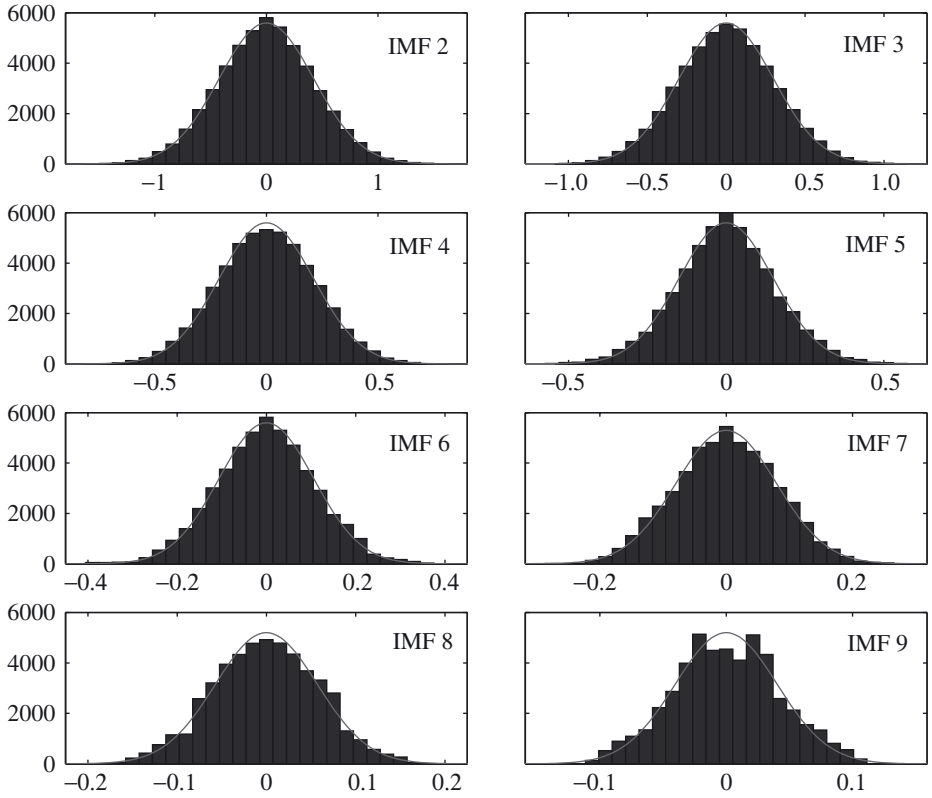

Figure 3. Histograms of IMFs (modes) 2–9 for a white-noise sample with 50 000 data points. The superimposed grey lines are the Gaussian fits for each IMF.

will have more oscillations and the distribution will converge to a normal distribution according to the central limit theorem.

According to the probability density function theory (see, for example, Papoulis 1986), for a time-series that has a normal distribution, its energy, defined by equation (2.2), should have a χ^2 distribution, with the degrees of freedom of the χ^2 distribution equal to the mean of the energy.

To determine the exact number of degrees of freedom for the χ^2 distribution of IMFs decomposed from a white-noise series of length N, we can argue as follows: we use the Fourier spectrum of a white-noise series of the same lengths, N. For such a white-noise series, its number of degrees of freedom is N. When such a white-noise series is decomposed in terms of Fourier components according to equation (2.1), we have N Fourier components which form a complete set. Since each component has a unit degree of freedom, the number of degrees of freedom of an IMF is essentially the sum of the Fourier components it contains. As the energy in a white-noise series is evenly distributed to each Fourier component, we propose that the fraction of energy contained in an IMF is the same as the fraction of the number of degrees of freedom. For a normalized white-noise time-series with unit total energy, the number of degrees of freedom of the nth IMF should be the energy of that particular IMF; thus, $r_n = N\bar{E}_n$. Therefore, the probability distribution of NE_n is the χ^2 distribution

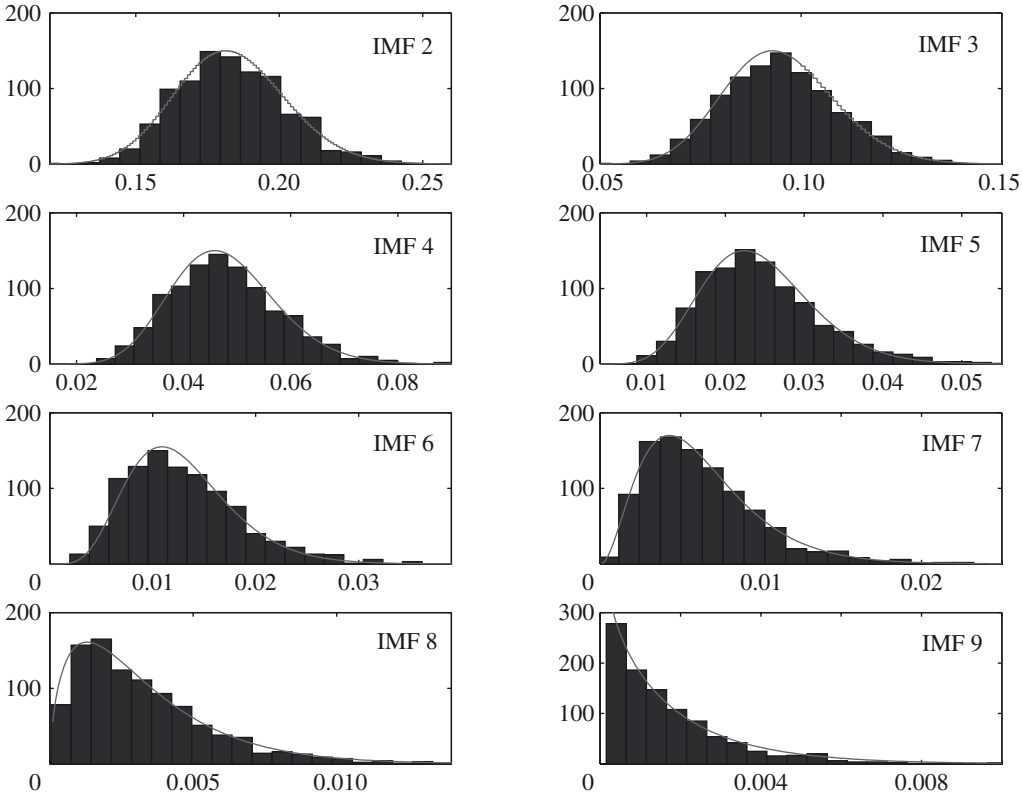

Figure 4. Histograms of the energy density for IMFs (modes) 2–9 for a white-noise sample with 50 000 data points. The superimposed grey lines are the χ^2 fits for each IMF.

with $N\bar{E}_n$ degrees of freedom, i.e.

$$\rho(NE_n) = (NE_n)^{N\bar{E}_n/2-1}\mathrm{e}^{-NE_n/2}. \tag{3.1}$$

Therefore, the probability distribution of E_n is given by

$$\rho(E_n) = N(NE_n)^{N\bar{E}_n/2-1}\mathrm{e}^{-NE_n/2}. \tag{3.2}$$

The Monte Carlo test confirms our conjecture. Figure 4 shows the histogram of the distribution of energy for each IMF for 1000 samples of white-noise series, each of the length of 1000 data points. The red lines are the corresponding χ^2 distributions based on equation (3.2). Clearly, the theoretical lines and histograms are in excellent agreement with each other.

Having determined the distribution function of the energy, we will discuss the spreading of this distribution in the next section.

(*b*) *Spread of energy*

Now we are ready to derive the spread of the energy of white-noise samples of a certain length N. First, we introduce a new variable, y, as

$$y = \ln E. \tag{3.3}$$

For simplicity, the subscript n is omitted. The distribution of y is therefore

$$\begin{aligned}\rho(y) &= N(N\mathrm{e}^{y})^{N\bar{E}/2-1}\mathrm{e}^{-NE/2}\mathrm{e}^{y}\\ &= C\exp(\tfrac{1}{2}yN\bar{E}-\tfrac{1}{2}NE)\\ &= C\exp\left[-\frac{N\bar{E}}{2}\left(\frac{E}{\bar{E}-y}\right)\right],\end{aligned} \tag{3.4}$$

where $C = N^{N\bar{E}/2}$. Since $E = \mathrm{e}^{y}$,

$$\frac{E}{\bar{E}} = \mathrm{e}^{y-\bar{y}} = 1 + y - \bar{y} + \frac{(y-\bar{y})^2}{2!} + \frac{(y-\bar{y})^3}{3!} + \cdots. \tag{3.5}$$

Substituting equation (3.5) into equation (3.4), we have

$$\begin{aligned}\rho(y) &= C\exp\left\{-\frac{N\bar{E}}{2}\left[1-\bar{y}+\frac{(y-\bar{y})^2}{2!}+\frac{(y-\bar{y})^3}{3!}+\cdots\right]\right\}\\ &= C'\exp\left\{\frac{-N\bar{E}}{2}\left[\frac{(y-\bar{y})^2}{2!}+\frac{(y-\bar{y})^3}{3!}+\cdots\right]\right\},\end{aligned} \tag{3.6}$$

where $C' = C\exp[-\frac{1}{2}N\bar{E}(1-\bar{y})]$.

When $|y-\bar{y}| \ll 1$,

$$\rho(y) = C\exp\left\{-\frac{N\bar{E}}{2}\left[1-\bar{y}+\frac{(y-\bar{y})^2}{2}\right]\right\} = C'\exp\left[\frac{-N\bar{E}(y-\bar{y})^2}{4}\right]. \tag{3.7}$$

From equations (3.6) and (3.7), one can determine the spread of different confidence levels. The spread can be identified as a function of $\bar{E}_n$. For the case of $|y-\bar{y}| \ll 1$, the distribution of E_n is approximately a Gaussian with a standard deviation

$$\sigma^2 = \frac{2}{N\bar{E}_n} = \frac{2\bar{T}_n}{N}. \tag{3.8}$$

For such a case, the spread lines can be defined as

$$y = -x \pm k\sqrt{\frac{2}{N}}\mathrm{e}^{x/2}, \tag{3.9}$$

where $x = \ln\bar{T}_n$ and k is a constant determined by percentiles of a standard normal distribution. For example, we will have k equal to -2.326, -0.675, -0.0 and 0.675 for the first, 25th, 50th, 75th and 99th percentiles, respectively.

Figure 5 plots the spread lines for the first and 99th percentiles based on equations (3.6) and (3.7). The thinly dashed lines are calculated from equation (3.9). Clearly, the simplified calculations based on equation (3.9) agree well with the theoretical lines (bold blue dashed) that are based on equations (3.6), which provides more details of abnormal distribution that skews toward the lower-energy side.

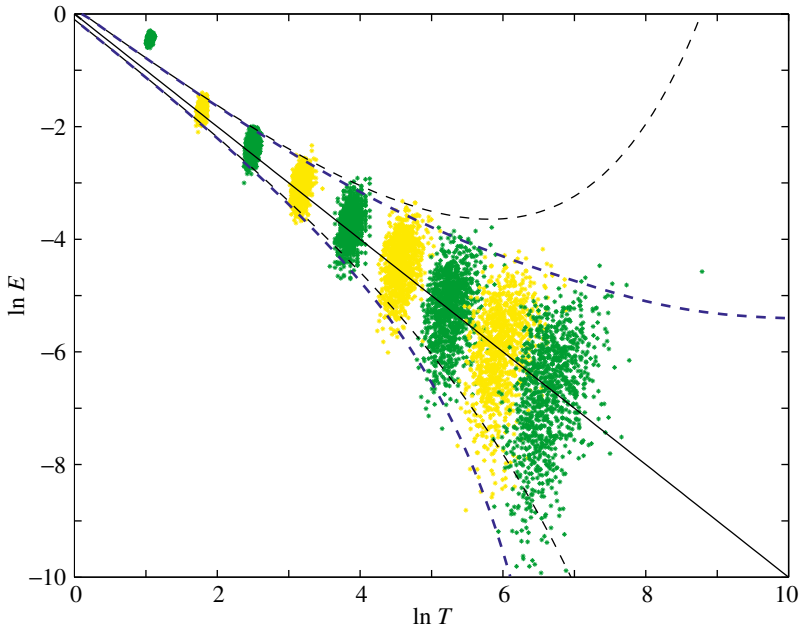

Figure 5. The spread function. The grouped dots and the black line are the same as in figure 2. The bold blue dashed lines are the first and 99th percentiles calculated from equation (3.6). The thin black dashed lines are the first and 99th percentiles calculated from equation (3.7).

4. A test for the statistical information content of noisy data

Having derived the statistical characteristics of white noise that include the expressions for the relationship between the energy density and the average period, the energy distribution and its spread function (which are confirmed by the Monte Carlo test), we propose a method, based on these results, to test the information content of a dataset with unknown noise level. Specifically, we want to determine which IMFs from a noisy dataset contain information, and which IMFs are purely noise. This test can be accomplished by using the above results on the statistical characteristics of white noise, which, by definition, contains no information. The specific steps for the test are as follows. First, decomposes the targeted noisy dataset (normalized) into IMFs. Second, construct a long artificial white-noise record as a reference, and divide these reference data into sections of identical length with the target dataset. Apply the EMD method to decompose each section of the reference white-noise data into IMFs, since the statistical characteristics of this artificial reference white-noise sections should follow the results derived in §§ 2 and 3. Therefore, one may skip this process by directly using the statistical characteristics of white noise derived in §§ 2 and 3, especially equation (3.6) to calculate the spread function of various percentiles. Third, select the confidence-limit level (e.g. 99%) and determine the upper and lower spread lines. Finally, compare the energy density for the IMFs from the data with the spread functions; the IMFs that have their energy located above the upper bound and below the lower bound should be considered to be containing information at that selected confidence level.

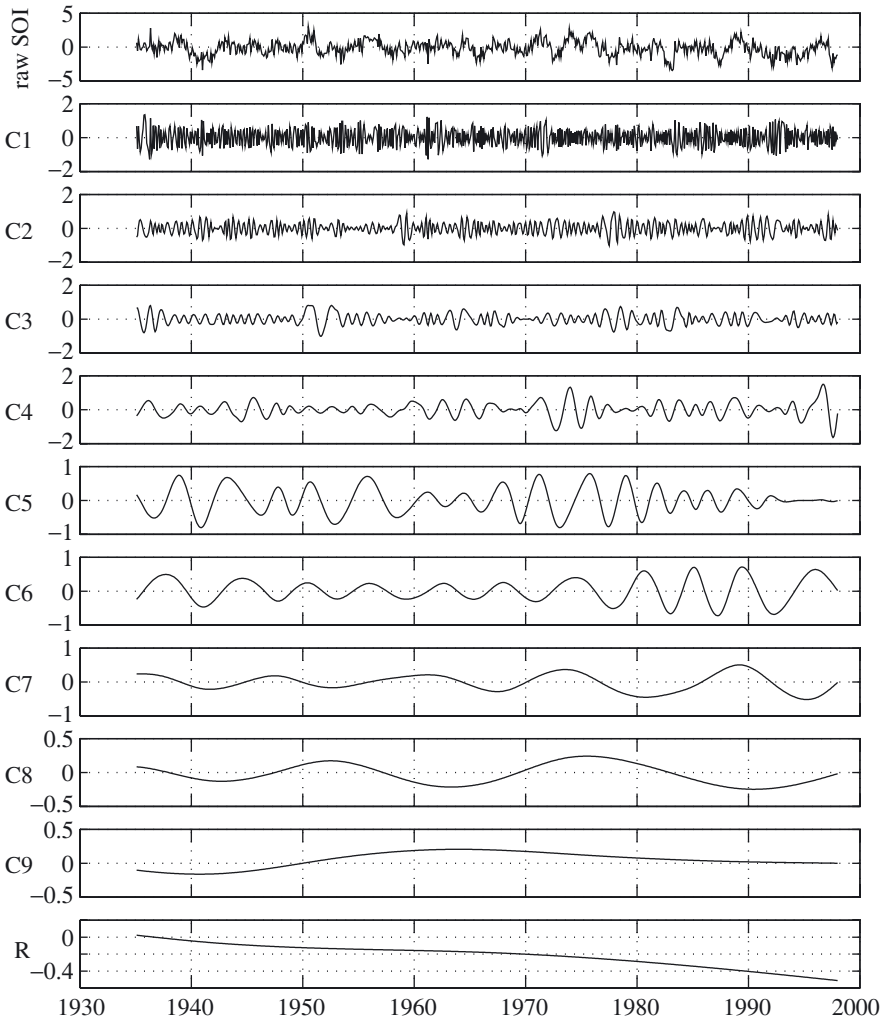

Figure 6. The raw SOI index (top panel), the corresponding intrinsic mode functions (C1–9), and the monotonic trend (R).

We will test this method with a real dataset in the SOI from January 1935 to December 1997 (Trenberth 1984). The SOI is a normalized monthly sea-level pressure index that reflects primarily the large-scale dynamically coupled system of atmosphere and ocean in the tropical Pacific. A large negative (positive) peak of SOI, which often happens with a two- to seven-year period, corresponds to a strong El Niño (La Niña) event. With its rich statistical properties and scientific importance, the SOI is one of the most prominent time-series in the geophysical research community and has been well studied. Many time-series-analysis tools analyse this time-series to display their capability of revealing useful scientific information (e.g. Wu *et al*. 2001; Ghil *et al*. 2002). For the reason, we selected this time-series to illustrate the usefulness of our proposed statistical method. The SOI and its IMFs are displayed in figure 6.

Here again, we use the same synthetically generated white-noise series of 10^6 points as the reference data, and decompose them into IMFs. The IMFs are then divided into more than a thousand samples, each sample containing 756 data points, the same as the length of SOI index examined previously by Wu *et al.* (2001). The averaged period (in terms of the number of data points within two consecutive peaks of an IMF) of each IMF and its energy density (inner product) are then plotted as the scatter points in green and yellow in figure 7, serving as independent samples of the reference white-noise data in a Monte Carlo test.

The black solid line is the theoretical expectation of the pair of averaged period and energy. The dashed blue lines are the theoretical spread lines of the first and 99th percentiles. Therefore, after a targeted time-series (e.g. SOI) is decomposed into IMFs, the averaged period and corresponding energy can be compared with the reference white-noise samples to determine whether a specific IMF contains any statistically significant information.

The red '⊛' points are the averaged period and corresponding energy for the first nine IMFs of SOI. It is clear that there are four '⊛' points that stay above the 99% confidence level. The averaged periods for these IMFs are 2.0, 3.1, 5.9 and 11.9 years, respectively.

The *a priori* test result is consistent with the statistical test using a regular spectral analysis method, which also shows that SOI has inter-annual peaks that are statistically significant. However, the EMD results give us the following additional information. First, EMD identifies the significant IMFs. As IMFs serve as adaptive bases, they can represent the underlying processes more effectively. Furthermore, the IMFs isolate physical processes of various time-scales, and also give the temporal variation with the processes in their entirety without resorting to the linear assumption as in the Fourier-based decomposition. Not being encumbered by the harmonics, the IMFs can show the nonlinear distortion of the waveform locally as discussed by Huang *et al.* (1998) and Wu *et al.* (2001). Finally, the IMFs can be used to construct the time–frequency distribution in the form of a Hilbert spectrum, which offers minute details of the time variation of the underlying processes.

The above method can be further refined through a re-scaling step. With this approach, we can determine the noise level of a time-series by assuming that the targeted time-series contains two parts: the signal and the noise. This can be accomplished as follows. If we can ascertain that any specific IMF contains little useful information, then we can assume that the energy of that IMF comes solely from noise, and assign it on the 99% line. We will then use the energy level of that IMF to re-scale the rest of IMFs. If the energy level of any IMF lies above the theoretical reference white-noise line, we can safely assume that it contains statistically significant information. If the rescaled energy level lies below the theoretical white-noise line, then we can safely assume that it contains little useful information. For the SOI data, the first IMF consists of a broadband spectrum with its peak corresponding to no perceivable physical process; therefore, it can be safely assumed to be pure noise. We can use it to re-scale energy density of other IMFs. The results are plotted as the blue '◊' points in figure 7. It turns out, under such a criterion, that all the IMFs seem to contain some useful information, i.e. they are bordering on the statistically significant at 99% confidence level. The latter *a posteriori* test is verified by examining the global SST variability associated with each mode which will be reported elsewhere.

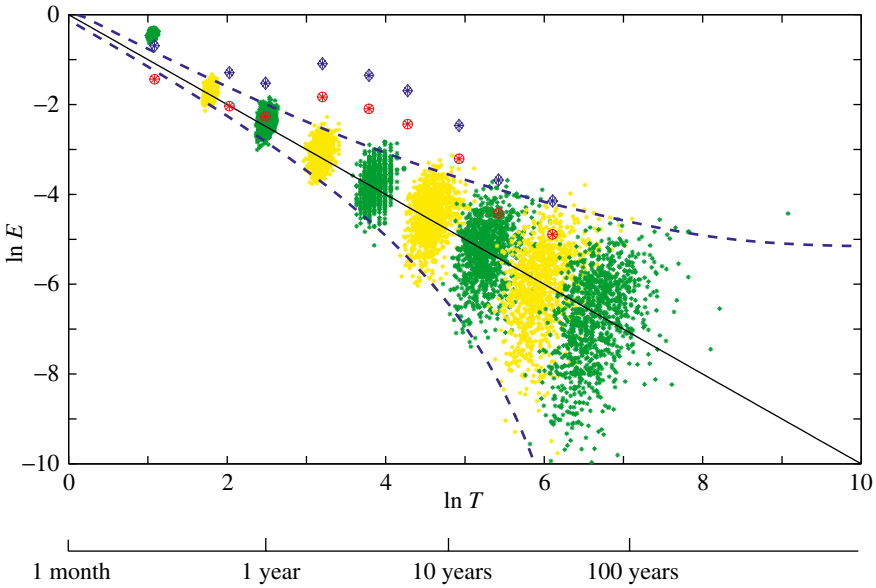

Figure 7. *A priori*, *a posteriori* and Monte Carlo tests of SOI. The SOI is sampled monthly. The corresponding real-time axis is added at the bottom.

5. Summary and discussions

The properties of the white noise under the EMD are studied. We carried out numerical experiments of decomposing uniformly distributed white noise into IMFs. The empirical findings are almost identical to those reported by Flandrin *et al.* (2003). Based on the empirical findings, we have deduced additional analytical expressions for the various statistical properties of white noise. From the properties of the Fourier spectra of IMFs, we derived an equation that states that the product of the energy density of IMFs and the corresponding averaged period of IMFs is a constant. The numerical experiments also show that all the IMFs of white noise have normal distribution and, therefore, the energy-density distribution of an IMF sample satisfies the χ^2 distribution. Furthermore, from the properties of Fourier spectra of white noise, we derived the spread function of the IMF components. These results allow us to establish a methodology for assigning statistical significance of IMF components for any noisy data. The statistical test method shows excellent consistency with the Monte Carlo test.

The *a priori* test method based directly on the above results was applied to the SOI. The results are consistent with traditional spectral analysis: the components of inter-annual time-scale are statistically significant. An *a posteriori* test method is also proposed by taking advantage of that fact that the IMFs are more effective in isolating physical processes of various time-scales than a usual Fourier component. These results demonstrated that our test methods are valid and effective.

The authors thank the seminal comments made by Professor Samuel Shen, of the University of Alberta, suggesting we study the Fourier spectra of the IMFs from the white noise. Through out the study, he has also offered valuable discussions and encouragements. We also thank Dr T. DelSole, of the Center for Ocean–Land–Atmosphere Studies, for his stimulating questions and discussions, and Professor Patrick Flandrin, of Ecole Normale Supèrieure de Lyon, for allowing

us to read a yet-to-be published manuscript. Z.W. is supported by NSF grant ATM-9907915, and N.E.H. is supported in part through a grant from the NASA RTOP Oceanic Processes Program, an ONR Processes and Prediction Program grant no. N00014-98-F-0412, and an NOAA Climate Center grant no. NEEF4100-3-00269.

References

Flandrin, P., Rilling, G. & Gonçalvès, P. 2004 Empirical mode decomposition as a filterbank. *IEEE Signal Process. Lett.* **11**, 112–114.

Ghil, M. (and 10 others) 2002 Advanced spectral methods for climatic time series. *Rev. Geophys.* **40**, 1003. (DOI: 10.1029/2000RG000092.)

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, E. H., Zheng, Q., Yen, N.-C., Tung, C. C. & Liu, H. H. 1998 The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis. *Proc. R. Soc. Lond.* A **454**, 903–995,

Huang, N. E., Shen, Z. & Long, S. R. 1999 A new view of nonlinear water waves: the Hilbert spectrum. *A. Rev. Fluid Mech.* **31**, 417–457.

Papoulis, A. 1986 *Probability, random variables, and stochastic processes*, 2nd edn. McGraw-Hill.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992 *Numerical recipes in C*, 2nd edn. Cambridge University Press.

Trenberth, K. E. 1984 Signal versus noise in the Southern Oscillation. *Mon. Weather Rev.* **112**, 326–332.

Wu, Z., Schneider, E. K., Hu, Z.-Z. & Cao, L. 2001 The impact of global warming on ENSO variability in climate records. Technical report 110, Center for Ocean–Land–Atmosphere Studies, Calverton, MD, USA.