

ENTROPIC INTERPRETATION OF EMPIRICAL MODE DECOMPOSITION AND ITS APPLICATIONS IN SIGNAL PROCESSING

CHIH-YUAN TSENG* and HC LEE†

*Graduate Institute of Systems Biology and Bioinformatics,
National Central University, 300 Jhongda Road,
Jhongli, Taiwan 320*

**chih-yuan.tseng@ualberta.ca*

†hlee@sansan.phy.ncu.edu.tw

The Hilbert-Huang transform (HHT) method, which is designed to analyze nonstationary and nonlinear time-dependent data, is attracting lots of attention. The HHT first applies the empirical mode decomposition (EMD) to decompose data into intrinsic mode functions (IMF). The Hilbert transform then is applied to the IMFs to reveal its instantaneous frequency spectrum. However, because the EMD lacks analytical interpretation, the meaning of IMFs is unclear. This work proposes an entropic analysis strategy to provide an information-based interpretation. Based on this strategy, three applications in data analysis are demonstrated: (1) studies of characteristic of white noise, (2) determination of minimum sampling rates to generate sufficient numbers of realizations, and (3) a low pass noise filter design.

Keywords: Hilbert-Huang transform; empirical mode decomposition; intrinsic mode function; maximum entropy; Bayesian interpretation.

1. Introduction

Data analysis is a crucial step in interpreting data and in revealing corresponding underlying functions in all disciplines of science and engineering. The Fourier transform is particularly favored by many disciplines to study time-dependent signals. However, when signals are generated from either nonstationary, nonlinear, or both processes, the Fourier transform is inadequate. A wavelet transform method based on the Fourier type transform is thus proposed for resolving this problem [Donoho and Johnstone (1994, 1995); Donoho *et al.*, 1995]. Yet it still has three shortcomings: (1) the limited length of the basic wavelet function makes definition of the energy-frequency-time distribution difficult; (2) a nonadaptive nature; and (3) it can only analyze nonstationary data [Huang *et al.* (1998); Kizhner *et al.* (2004)]. Even though several approaches have been proposed to

*Corresponding author. C.-Y. Tseng is currently with Department of Oncology, University of Alberta, Edmonton, AB T6G 1Z2, Canada.

improve it [Polygiannakis *et al.* (2003); Lang *et al.* (1996); Sternickel *et al.* (2001)] and shown promising results, their applications in nonlinear and nonstationary data analyses are still limited by the assumption of the Fourier transform, which treats signals to be represented by combinations of sinusoids and cosinusoids.

The Hilbert-Huang transform (HHT) method is proposed by Huang and his colleagues to surmount these limitations [Huang *et al.* (1998)]. First, the HHT decomposes the data through the empirical mode decomposition (EMD) method into the intrinsic mode functions (IMFs) with different instantaneous frequencies and the last component, the residue or the trend. An IMF satisfies two conditions: (1) that the number of extrema and the number of zero crossings must either equal or differ at most by one in the whole data set and (2) that the mean value of the envelope defined by the local maximum and the envelope defined by the local minima at any point is zero [Huang *et al.* (1998)]. Because the IMFs are adaptive and locally determined, they have physical representation of the underlying processes [Huang *et al.* (1998); Kizhner *et al.* (2004)]. In addition, IMFs form an orthogonal set, thus they can be used as the basis to represent the data. After obtaining the IMFs, one can perform instantaneous frequency analysis on IMFs using either the Hilbert transform, Fourier transform based methods, or others [Huang *et al.* (2009)]. Particularly [Huang *et al.* (2009)], show normalized Hilbert transform and direct quadrature method outperform Hilbert transform, Wigner-Ville distribution, the generalized zero crossing, and the Teager energy operator methods to obtain instantaneous frequency. However, since our goal is to investigate the meaning of IMFs, the computation of instantaneous frequency is not an issue in this work.

Because the EMD method lacks analytical interpretation, it is unclear what IMFs mean and also which IMF carries information relevant to underlying functions of the raw data. Wu and Huang propose a hypothesis test-based strategy with consideration of mean energy and periods of IMFs (hereafter denoted as the WH method) to investigate the properties of IMFs [Wu and Huang (2004, 2005)]. However, the WH method becomes inadequate for low frequency IMFs. This inadequacy may be resolved based on the studies of Xu *et al.* (2009), in which they introduce Fourier interpolation to better identify extrema of signals with low sampling rates. Because the WH method only accounts the macroscopic properties of signals, mean energy and mean period, the detailed information of IMFs is not spelled out completely and the interpretation is limited. Despite these, the HHT has yielded promising applications in many fields such as denoising [Flandrin *et al.* (2004a); Khaldi *et al.* (2008); Kopsinis and McLaughlin (2008a, 2008b); Boudraa and Cexus (2006)], speech recognition [Huang and Pan (2006)], ground moving target indication [Cai *et al.* (2006)], geophysics [Huang and Wu (2008)], financial problems [Huang *et al.* (2003); Wu *et al.* (2006)], and biological problems [Wu and Hu (2006); Liang *et al.* (2005); Ai *et al.* (2008)].

In this paper, our first goal is to develop an information theory-based approach to provide a comprehensive interpretation of IMFs. Second goal is to demonstrate three applications of this approach in data processing. Basically, the approach follows the Bayesian interpretation to treat probability as the state of knowledge about the systems of interest rather than a frequency. Once we can assign each IMF a probability distribution function (PDF), the extent of the information relevant to the systems of interest codified in the PDF can then be measured by entropy. Therefore, we can determine to what extent an IMF contains information. Because a PDF contains complete information in the corresponding IMF, one can expect the proposed approach to reveal more properties of IMFs than the WH method that only accounts the macroscopic property of signals does. Furthermore, because the proposed approach does not involve with mean period estimate, the inadequate problem in the WH is dismissed.

The rest of this paper is structured as follows. Sections 2 and 3 will briefly discuss the EMD method and the energy interpretation from the WH method, respectively. Section 4 presents an entropic analysis strategy and a statistical study of applying such a strategy to reveal properties of IMFs. Section 5 then discusses three applications of the entropic strategy in data analysis: studies of characteristic of IMFs of white noise, the determination of minimum sampling rate, and a low pass noise filter. Finally, a conclusion is given.

2. Empirical Mode Decomposition

This section particularly focuses on the EMD of the HHT and briefly discusses its rationale (please refer to Huang *et al.* (1998) for detailed derivations). The essence of the EMD is to empirically identify the intrinsic oscillatory modes, which satisfies two conditions introduced in the Introduction, by their characteristic time scales in the data, and decompose the data accordingly. A sifting process is used to eliminate riding waves and to make wave-profiles more symmetric to systematically extract intrinsic oscillatory modes.

Consider a real time series signal $S_{\text{real}}(t)$ that is contaminated by noise $n(t)$ and generates

$$S_{\text{raw}}(t) = S_{\text{real}}(t) + n(t), \quad (1)$$

the raw data to be measured. The process first determines a mean of upper and lower envelopes defined by local minimum and maximum in $S_{\text{raw}}(t)$ respectively, m_1 . Subtracting m_1 from $S_{\text{raw}}(t)$,

$$S_{\text{raw}}(t) - m_1 = h_1 \quad (2)$$

should yields h_1 as the first IMF. In reality, however, there may be overshoots and undershoots in creating the two envelopes. Furthermore, the extrapolation used in connecting two extremes will cause serious problems in the end point. Thus, the

sifting process needs to be repeated several times

$$h_{1(k-1)} - m_{1k} = h_{1k}, \quad (3)$$

where k denotes the number of repetitions until $h_{1k} = C_1(t)$ is an IMF. This is the first IMF with the highest instantaneous frequency from the data. Thereafter, one can separate $C_1(t)$ from the rest of data by

$$S_{\text{raw}}(t) - C_1(t) = r_1. \quad (4)$$

One then continues the sifting process on the residue r_1 to extract second IMF and so on until either when the residue becomes so small that it is less than the predetermined value of substantial consequence, or when the residue becomes a monotonic function. Superposition of these IMFs and the residue $C_{N_c}(t)$ will recover the raw data, $S_{\text{raw}}(t) = \sum_{i=1}^{N_c} C_i(t)$.

3. Energy Interpretation of Intrinsic Mode Functions

The crux of the WH method in interpretation of IMFs is to identify whether or not an IMF carries information relevant to underlying functions of real signals based on energy perspective [Wu and Huang (2004, 2005)]. A null-hypothesis test method is introduced in the WH method, in which the null hypothesis is that an IMF contains no information. Because white Gaussian noise contains no information, Wu and Huang investigate properties of IMFs decomposed from noise to determine a statistical distribution that associates with information-free data. By analyzing the mean energy and periods of IMFs of white noise, Wu and Huang found the energy distribution of each IMF can be described by the Chi-square distribution. From this, one can determine the energy spread function of white noise with an upper and lower boundary. The boundaries represent the confidence level to treat data as a noise component. When the mean energy of a data's IMF lies above the upper boundary of the energy spread function, the null hypothesis is rejected. Therefore, one can conclude that this IMF carries information relevant to underlying functions of the data with the selected confidence level.

Although Wu and Huang have demonstrated the applicability of the WH method [Wu and Huang (2004, 2005)], it still has two shortcomings. First, as mentioned in Wu and Huang's work [Wu and Huang (2004, 2005)], the WH method proves inadequate in analyzing low-frequency IMFs because of the inappropriate estimation of the corresponding mean periods. Wu and Huang proposed using the extrema-counting and Hilbert transform to resolve this problem [Wu and Huang (2004, 2005)]. However, the expectation values of the mean periods and the energy density are still likely to deviate from the theoretical values. This still does not correctly identify an IMF as a component of noise.

Second, statistical physics shows that mean energy of signal, a macroscopic quantity, only provides partial information of the corresponding systems. For example, the microscopic properties of a thermal system such as equations of motion of

particles cannot be unveiled by thermodynamics because it only studies macroscopic properties. Therefore, one can expect a limited interpretation of IMFs from the WH method.

4. Entropic Interpretation

An alternative interpretation of IMF, which hinges on a conceptual breakthrough, is proposed in this section. This breakthrough resulted from studies of statistical mechanics and information theory. First, the development of statistical mechanics shows that detailed information of thermal systems can be codified in a probability distribution function PDF $p(q)$, where parameters q characterize the system. The canonical ensemble is an example. Not only one can study thermodynamical quantities but one also can tackle problems such as irreversible thermodynamics by utilizing concept of PDFs. This concept has become prominent after Shannon's work in communication theory [Shannon (1948)], later is recognized as information theory. Based on the Bayesian interpretation, a PDF represents a state of knowledge about systems of interest rather than merely a frequency. Furthermore, Cox demonstrates that the state of knowledge can be manipulated through Boolean algebra [Cox (1961)]. Shannon, therefore, can show that the extent of information carried by a signal can be quantified by entropy $S = -\sum_q p(q) \log p(q)$ [Shannon (1948)] using axiomatic approach. This conceptual advance provides an information theory-based approach to study statistical systems.

Based on this information treatment, we present a two-step strategy to investigate properties of IMFs.

4.1. Entropic analysis strategy

Step 1: Signal clustering. After we apply the EMD to decompose the raw data $S_{\text{raw}}(t)$ into IMFs, $C_i(t)$, where $i = 1 \cdots N_c - 1$, we cluster the IMFs into two sets in a straightforward way. One set is the superposition of the first k IMFs $\mathcal{N}_k(t) = \sum_{i=1}^k C_i(t)$, which has high instantaneous frequencies, where k will be incrementally increased from 1 to $N_c - 1$. The superposition of the rest of the IMFs, $\mathcal{S}_k(t) = \sum_{i=k+1}^{N_c} C_i(t)$ forms the other set. Note that we only consider a simplest way of clustering signals here. There is no restriction of introducing other clustering methods based on problems of interest.

Step 2: Information-rich components identification. Next, we ask to what extent information relevant to underlying functions of $S_{\text{raw}}(t)$ is carried in either the $\mathcal{N}_k(t)$ or $\mathcal{S}_k(t)$ set.

It has been shown that the relative entropy or negative Kullback-Leibler distance of two PDFs P_i and P'_i of observing a system at state i , $S[P, P'] = -\sum_i P_i \log P_i / P'_i \leq 0$, quantifies the difference of extent of information carried in both PDFs [Kullback (1997)]. Tseng and his colleagues extend the use of relative entropy for model [Tseng (2006)] and variable [Chen *et al.* (2007)] selection problems

in physics and geology. They further consider P_i to be the test probability models and P'_i as a uniform reference distribution, which represents complete ignorance of the systems. In this consideration, the maximum relative entropy indicate the test model is identical to the uniform reference. The test model also represents complete ignorance of the systems. Namely, it carries no information relevant to the system. On the contrary, the minimum relative entropy of a test model and the uniform reference indicate that the test model contains information the most. Therefore, the test model will be selected as the preferred model. Notes that the entropic analysis provides a complete ranking scale for all possible probability models.

Therefore, according to this Bayesian logic, suppose one can assign probability models to $\mathcal{N}_k(t)$, the extent of information codified in $\mathcal{N}_k(t)$ can be measured by relative entropy of such probability models and the information-free reference probability distribution. The question then becomes what are probability models for $\mathcal{N}_k(t)$ or $\mathcal{S}_k(t)$ and what is an appropriate information-free reference model.

Because an important property of a time-series data is its oscillation amplitude x , we consider the PDF for the $\mathcal{N}_k(t)$ to be function of amplitude x , $\hat{P}_{\mathcal{N}_k}(x)$. In this work, the exact probability $\hat{P}_{\mathcal{N}_k}(x)$ is approximated by the normalized histogram distribution of the amplitudes x of $\mathcal{N}_k(t)$, $P_{\mathcal{N}_k}(x)$. Notes that because we only consider a large amount of data, this approximation should be statistically legitimate to represent the real probability distribution.

Next, we consider white Gaussian noise to represent the information-free reference. Specifically, the reference is set to $\mathcal{N}_k^{\text{noise}}(t)$, the superposition of first k IMFs decomposed from a pure noise that is generated from the same noise sources in the raw data $S_{\text{raw}}(t)$. Therefore, one also can obtain the probability distribution $P_{\mathcal{N}_k^{\text{noise}}}(x)$ from noise amplitude histograms for the $\mathcal{N}_k^{\text{noise}}(t)$. Although our choice of noise as reference is the same as the WH method does, the way of utilizing it is different. Here, we take the complete amplitude distribution of noise into account while the WH method only considers the tail beyond certain percentile of the energy distribution of noise.

Finally, to quantify the extent of information codified in $P_{\mathcal{N}_k}(x)$, we calculate relative entropy $S[P, P']$, where P' is set to $P_{\mathcal{N}_k}(x)$ and P is set to information-free $P_{\mathcal{N}_k^{\text{noise}}}(x)$. Namely,

$$S[P_{\mathcal{N}_k^{\text{noise}}}, P_{\mathcal{N}_k}] = - \sum_{x=-\infty}^{\infty} P_{\mathcal{N}_k^{\text{noise}}}(x) \log \frac{P_{\mathcal{N}_k^{\text{noise}}}(x)}{P_{\mathcal{N}_k}(x)} \leq 0. \quad (5)$$

It measures average differences of $\log P_{\mathcal{N}_k}(x)$ and $\log P_{\mathcal{N}_k^{\text{noise}}}(x)$ over all $P_{\mathcal{N}_k^{\text{noise}}}(x)$. Maximum relative entropy indicates $P_{\mathcal{N}_k^{\text{noise}}}(x)$ is identical to $P_{\mathcal{N}_k}(x)$ and $P_{\mathcal{N}_k}(x)$ is completely information free. On the other hand, minimum relative entropy, $\min_k S[P_{\mathcal{N}_k^{\text{noise}}}, P_{\mathcal{N}_k}]$, indicates the maximum information differences between $P_{\mathcal{N}_k^{\text{noise}}}(x)$ and $P_{\mathcal{N}_k}(x)$. The ratio $S[P_{\mathcal{N}_k^{\text{noise}}}, P_{\mathcal{N}_k}] / \min_k S[P_{\mathcal{N}_k^{\text{noise}}}, P_{\mathcal{N}_k}]$ denotes to what extent $P_{\mathcal{N}_k}$ is close to the complete information-free distribution among all $P_{\mathcal{N}_k}(x)$ s. Therefore, we can define a normalized information scale to measure the

information extent codified in probability distribution $P_{\mathcal{S}_k}(x)$ for the corresponding $\mathcal{S}_k(t)$ set as

$$P_{\text{info}}(k) \stackrel{\text{def}}{=} 1 - S[P_{\mathcal{N}_k^{\text{noise}}}, P_{\mathcal{N}_k}] / \min_k S[P_{\mathcal{N}_k^{\text{noise}}}, P_{\mathcal{N}_k}] \leq 1. \quad (6)$$

When $P_{\text{info}}(x) = 0$ indicates $\mathcal{N}_k(t)$ is completely information free. Because the EMD does not guarantee the complete separation of information-rich and -free components from raw data, an IMF may contain partial information. Namely, $\mathcal{N}_k(t)$ may contain partial information. For $P_{\text{info}}(x) = 0.5$, it indicates that the corresponding $\mathcal{N}_k(t)$ only contains half extent of information. Therefore, in order to not to dismiss any information, we consider an $\mathcal{N}_k(t)$ to be an information-rich component when the corresponding $P_{\mathcal{N}_k}(x)$ is larger than 0.5.

One can expect this information interpretation to be consistent with energy interpretation. Furthermore, because a PDF contains complete information of an IMF, it shall reveal more properties of IMFs than the WH method does.

4.2. Numerical investigations of the applicability of entropic strategy

Here, we demonstrate the application of the proposed strategy in interpreting IMFs by studying two contaminated test signals. We first consider a pure sine wave function to illustrate the use of the entropic strategy. Afterward, we consider a more general case, a test signal that has nonstationary amplitudes and multiple oscillation frequencies, to show the applicability of the approach.

Pure sine wave. The first one-second long time-dependent test signal is given by $S_{\text{real}}(t) = \sin(2\pi f_{si}t)$ with signal frequency $f_{si} = 2\text{ Hz}$ and is sampled at the rate $f_s = 1000\text{ Hz}$, where t denotes time. The test signal is contaminated by white Gaussian noises using the MATLAB in-house tool (`awgn`) with different SNRs = 10, 5, 0, -5, -50 dB as raw test datum $S_{\text{raw}}(t)$. Additionally, a white Gaussian noise with the same length, arbitrary power strength, and sampled at the same rate of $f_s = 1000\text{ Hz}$ is generated using the MATLAB tool (`wgn`) as the reference. Totally, 100 samples for each six raw test data are generated for the statistical analysis.

Figure 1 shows an example of six IMFs and one residue of the first raw test signal with SNR = 10 dB. We cluster these IMFs into two groups, $(\mathcal{N}_k(t), \mathcal{S}_k(t))$. Similarly, we obtain two groups for noise, $(\mathcal{N}_k^{\text{noise}}(t), \mathcal{S}_k^{\text{noise}}(t))$. The probabilities $P_{\mathcal{N}_k}(x)$ and $P_{\mathcal{N}_k^{\text{noise}}}(x)$ of oscillation amplitude x in $\mathcal{N}_k(t)$ and $\mathcal{N}_k^{\text{noise}}(t)$ are assigned by amplitude histograms, as shown in Fig. 2, in which the dark hollow squares denote $P_{\mathcal{N}_k}(x)$ and gray hollow squares denote $P_{\mathcal{N}_k^{\text{noise}}}(x)$. The figure shows two remarks. First, as expect, $P_{\mathcal{N}_k^{\text{noise}}}(x)$ is almost normally distributed for different k values. Second, it indicates that when the k value is less than six, both probability distributions share similarities in mean values and standard deviations. The differences between the distributions arise when the k value is larger than six. This qualitative comparison suggests that the $\mathcal{N}_k(t)$ for $k \leq 5$ is highly likely to be the information-free component.

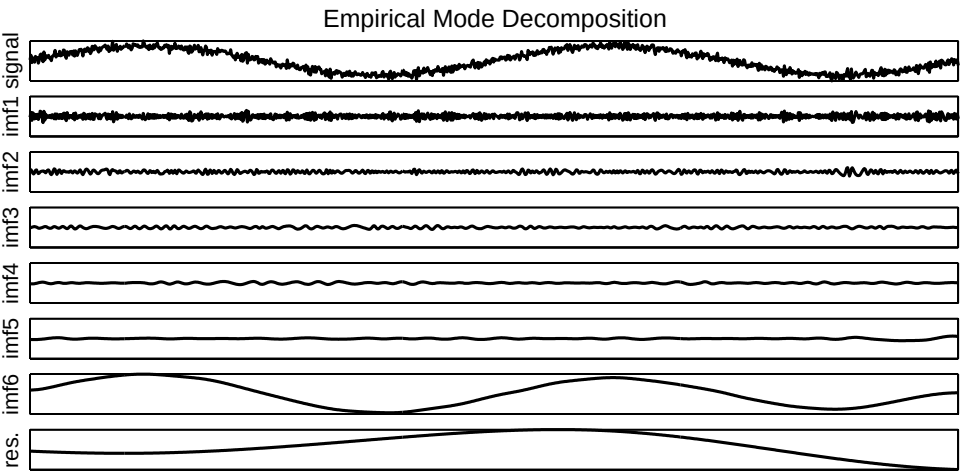


Fig. 1. An example of the EMD method. The top row shows the raw data, which is the linear superposition of the test signal $S_{\text{real}}(t) = \sin(2\pi f_{si}t)$ and a white Gaussian noise with $\text{SNR} = 10\text{ dB}$. Six IMFs and one residue are then decomposed from it using the EMD.

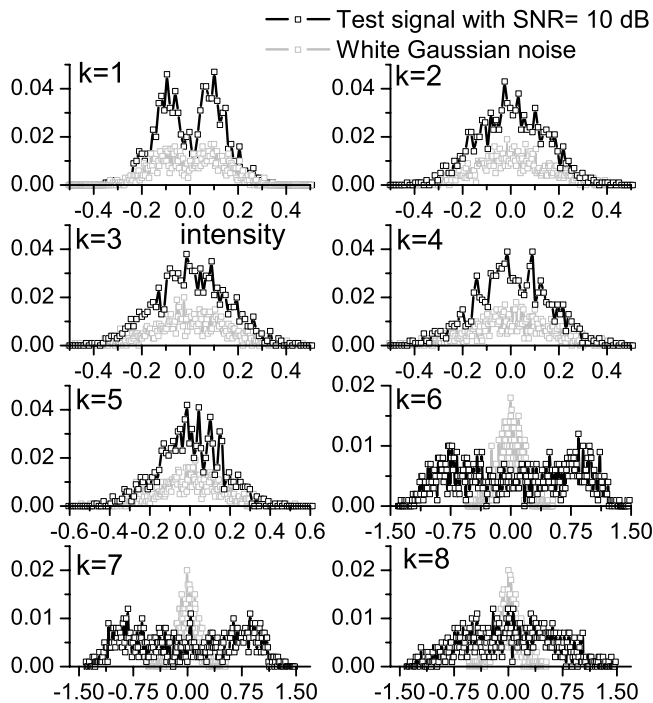


Fig. 2. An example of probability distributions of intensities in the $\mathcal{N}_k(t)$ and the $\mathcal{N}_k^{\text{noise}}(t)$ for the case of $\text{SNR} = 10\text{ dB}$. Notes that the calculated probability values of intensities are denoted by hollow symbols.

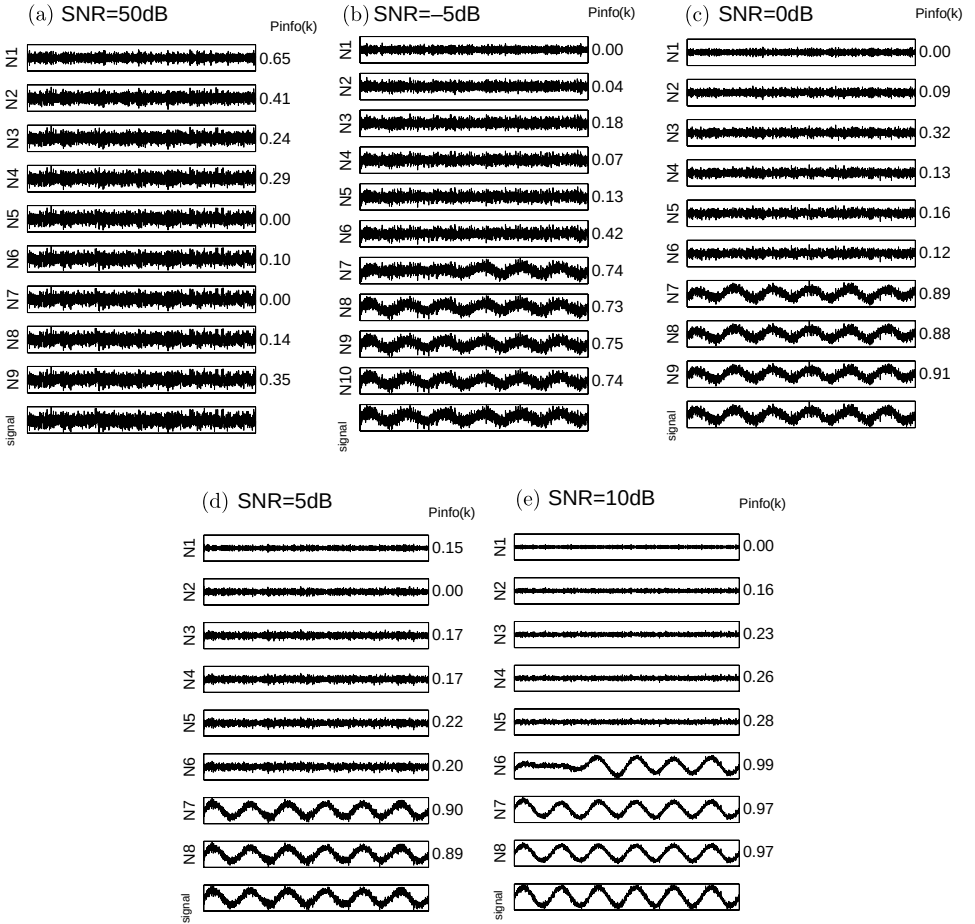


Fig. 3. This figure plots $\mathcal{N}_k(t)$ for the test signal contaminated by noise with five various strengths. It also shows the corresponding $P_{\text{info}}(k)$, Eq. (6), on the right.

Next, we quantitatively determine the extent of information in IMFs according to the scale $P_{\text{info}}(k)$ of $\mathcal{N}_k(x)$ that carries information, Eq. (6). Figure 3 shows $\mathcal{N}_k(t)$ and the corresponding $P_{\text{info}}(k)$ values on the right for all five cases. In panel (A), the test signal is severely contaminated by white noise. As expected, the scale $P_{\text{info}}(k)$ s in this case are less than 0.5 for all $\mathcal{N}_k(t)$ except $\mathcal{N}_1(t)$. Notes that because the relative entropies of $\mathcal{N}_k(x)$ for the case of $\text{SNR} = -50$ dB are all within 0 and -1 , the $\mathcal{N}_k(t)$ is still likely to be noise dominant even though the corresponding $P_{\text{info}}(k)$ is larger than 0.5. Thus, all $\mathcal{N}_k(t)$ s are noise dominant. When the SNR value is increased as shown in panel (B)–(E) consecutively, the real signal, $\sin(2\pi f_{st})$, is observed in $\mathcal{N}_k(t)$ s with k less than 6 or 7. The corresponding $P_{\text{info}}(k)$ are larger than 0.7. Furthermore, the $P_{\text{info}}(k)$ is raised from around 0.75 in (B), 0.9 in (C) and (D) to 0.97 in (E) as the SNR is increased.

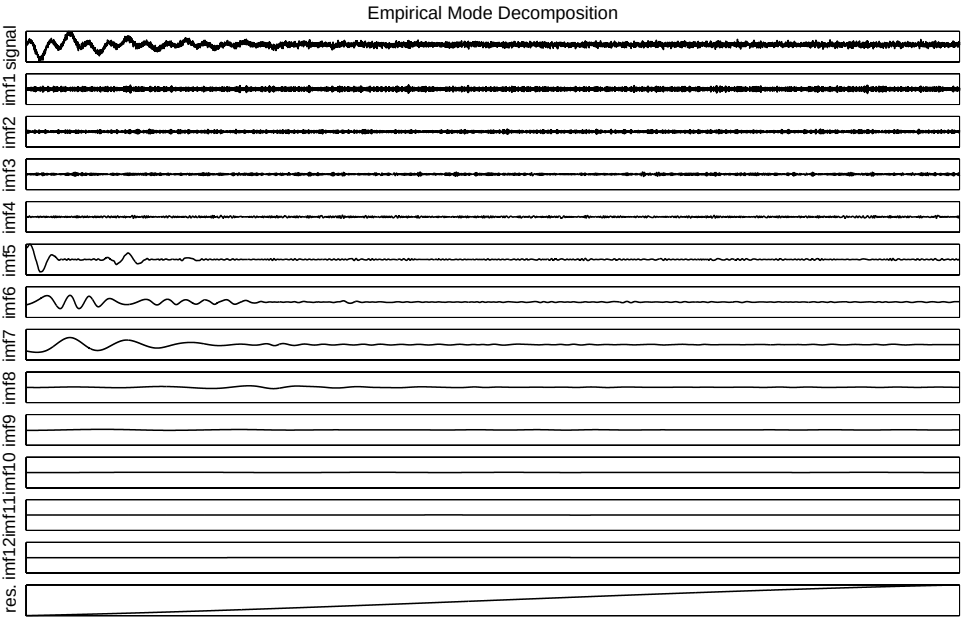


Fig. 4. The second test signal that is contaminated by white Gaussian noises with SNR = 0 dB and its 12 IMFs.

A more complicated signal. The second test signal is 8 seconds long and given by $S'_{\text{real}}(t) = 1000 * \exp(-t/1000) \sin(2\pi f_{si}t) \cos(2\pi f'_{si}t)$ with signal frequency $f_{si} = 2$ Hz and $f'_{si} = 4$ Hz. The sampling rate is $f_s = 8000$ Hz. Furthermore, it is also contaminated by white Gaussian noise with SNR = 0 dB. The first row of Fig. 4 is the generated raw data and shows a gradually decreased amplitude and two distinct frequencies. The next 12 rows show the IMFs obtained from EMD. The results indicate that the EMD primarily decompose the signal that has large amplitude into IMFs 5, 6, and 7. The IMFs 8 to 12 are the rests of components left in the signal. First four IMFs are likely the components of white Gaussian noise.

Following the same approach as before, we show 12 $\mathcal{N}_k(t)$ components in Fig. 5 with the corresponding $P_{\text{info}}(k)$ values list on the right. It shows when we include IMF 5 in the $\mathcal{N}_4(t)$ to get $\mathcal{N}_5(t)$ the information scale of $\mathcal{N}_5(t)$, $P_{\text{info}}(5)$ jumps from 0.26 to 0.85. Furthermore, when IMFs 6 and 7 are included, $P_{\text{info}}(6)$ and $P_{\text{info}}(7)$ are 0.8 and 0.93, respectively.

Remark. The above numerical studies show that the scale $P_{\text{info}}(k)$ correctly quantifies the extent of information that is codified in $\mathcal{N}_k(x)$. Based on this scale, the IMFs that are information-free components can be identified. This way of identification is completely different from the WH method, in which one simply consider energy perspective to identify whether or not an IMF contains information with certain confidence levels. The WH method cannot further determine to what extent information is codified in IMFs. These studies suggest three applications in data

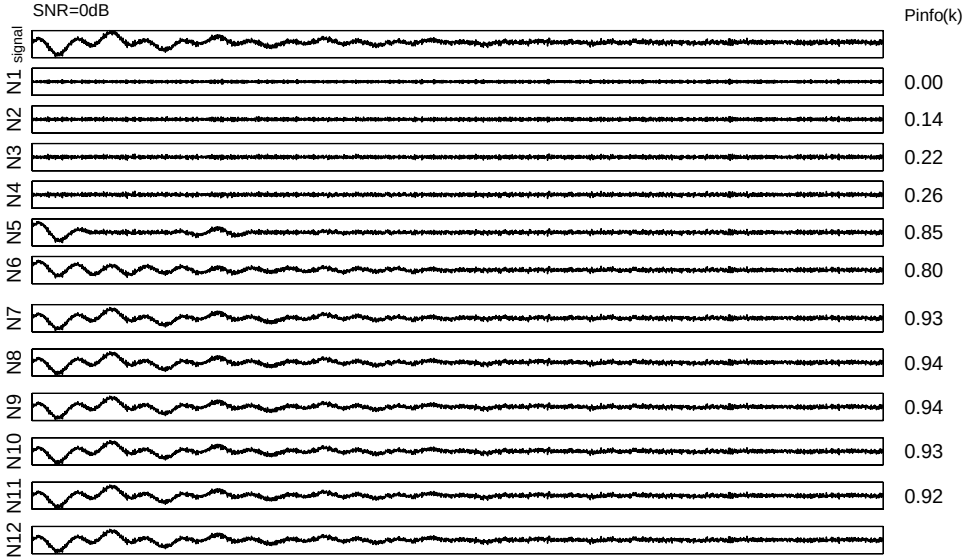


Fig. 5. This figure plots $\mathcal{N}_k(t)$ of the first four seconds of the second contaminated test signal. The corresponding $P_{\text{info}}(k)$, Eq. (6), are listed on the right.

analysis: studies of characteristic of IMFs of white noise, determination of minimum sampling rates, and design of a low pass filter. We will discuss them in next section.

5. Three Applications of the Entropic Strategy

5.1. Characteristic of IMFs of white noise

The first application is to study the characteristics of IMFs of white noise based on the step 1 signal clustering of the proposed strategy. Because IMFs are approximately orthogonal to each other, one can show that the energy density of the $\mathcal{N}_k(t)$ set is

$$E_{\mathcal{N}_k} = \frac{1}{N} \sum_{t=1}^N [\mathcal{N}_k(t)]^2 \cong \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^k [C_i(t)]^2 \quad (7)$$

and the energy density of the $\mathcal{S}_k(t)$ set is

$$E_{\mathcal{S}_k} = \frac{1}{N} \sum_{t=1}^N [\mathcal{S}_k(t)]^2 \cong \frac{1}{N} \sum_{t=1}^N \sum_{i=k+1}^{N_c-1} [C_i(t)]^2. \quad (8)$$

Furthermore, we can compare the energy of the two sets by defining the energy ratio,

$$R_k = 10 \log_{10} \frac{E_{\mathcal{S}_k}}{E_{\mathcal{N}_k}}. \quad (9)$$

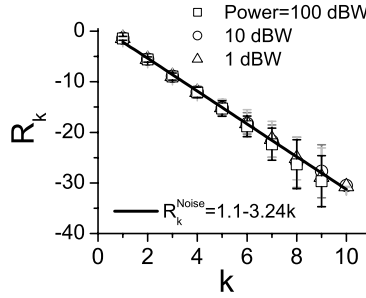


Fig. 6. The R_k distribution of white Gaussian noise with three different intensities. Black line is a linear equation fitting to the mean R_k .

Based on this definition, a characteristic of IMFs of white noise other than the one discovered by Wu and Huang (2004, 2005) is revealed through the following numerical studies.

Consider 100 samples of three white Gaussian noises with different powers of 1, 10, and 100 dBW, each with 1000 data points, are generated through the MATLAB tool (wgn). After applying, the EMD decomposes these 100 samples into IMFs; the ratio R_k of each sample is calculated for $k = 1 \cdots N_c - 1$. The mean values and standard deviations of these R_k are computed and shown in Fig. 6. The figure shows that the R_k distribution of a white Gaussian noise is almost independent of the noise power.

Furthermore, the mean values of these R_k are inversely proportional to the k values and can be fitted by a linear equation

$$R_k^{\text{noise}} = 1.1 - 3.24k, \quad (10)$$

where the standard error for the intersection is 0.4 and for the slope is 0.06. The standard deviation of R_k values, however, is increased when k increases.

One can attribute this characteristic to the fact that the EMD is a dyadic filter bank [Flandrin *et al.* (2004b)]. Based on the dyadic filter bank model, the definition of Eq. (9) can be approximated by $10 \log_{10}(2^{-k}/(1 - 2^{-k}))$. When k is large, this approximation will close to $-3k$. This also suggests that the linear equation of R_k^{noise} is independent of noise sources. The results, not shown here, using different white noise generators confirm it. This linear relation indicates that the act of choosing white Gaussian noise as the reference in both the entropic strategy and the WH method is sufficient and appropriate to represent information-free distributions even though the sources of white noise in data of interest are unknown.

Furthermore, we investigate the effects of information-rich IMFs on R_k distribution by considering the test signals $S_{\text{raw}}(t)$ with different SNR values. The results are shown in Fig. 7 with the hollow symbols representing the test signals and the dark symbols representing noise. When the test signal is slightly distorted by noise (SNR = 30 dB), the R_k distribution completely differs from Eq. (10) shown as the black line. The R_k distributions of the less contaminated test signals

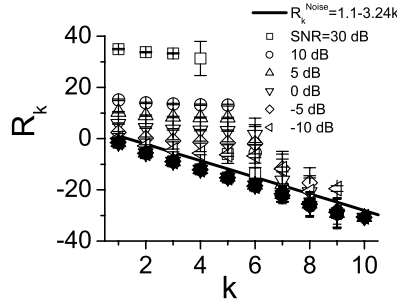


Fig. 7. The R_k distribution of data, which is contaminated by white Gaussian noise with five different SNR denoted by hollow symbols. Dark symbols denoted the R_k distribution of noise.

are likely to agree with the distributions of noise when the k value is larger than a certain value k_t . Conversely, when noise severely contaminates the test signal ($\text{SNR} \leq -10$ dB), the R_k distributions almost agree with the distributions of pure noise or the Eq. (10). The small standard deviations of R_k for small k suggest the EMD almost obtains the same decomposition for each sample.

5.2. Sampling issue of the EMD

A proper sampling rate decides whether or not can we correctly investigate underlying functions from the sampled data. Furthermore, as addressed in Rilling and Flandrin's work [Rilling and Flandrin (2009)], because the EMD is designed for continuous signals, the sampling issue in practically applying EMD on discrete signals is raised. They investigate the influence of the sampling on EMD and show the existence of bounds on possible errors. It will lead to a quantitative approach to obtain a principle of applying EMD properly.

Here, an alternative quantitative analysis to study the sampling issue in applying EMD is presented. Specifically, the characteristic of IMFs of white noise, Eq. (10), provides an EMD-based approach to determine the proper sampling rate. The approach contains two steps. The first step systematically adjusts sampling rates and the second step applies the entropic strategy on each sampled data and evaluates the corresponding R_k distribution. The sampled data that has R_k distribution close to Eq. (10) is likely to be information-free. Namely, the corresponding sampling rate does not generate sufficient number of realizations.

As a demonstration, again, we consider the two test signals used in Sec. 4.2. The first case, $S_{\text{real}}(t) = \sin(2\pi f_{si}t)$, has signal frequency f_{si} fixed to 2 Hz and $\text{SNR} = 0$ dB. The sampling rate f_s is gradually decreased from 2,000 to 14 Hz. The R_k distributions for all cases are then plotted in panel (a) of Fig. 8. The R_k distribution approaches the linear equation (10) of pure noise R_k^{noise} , shown as the dark line, when the sampling rate f_s is about 125 Hz. It leads to a high root mean square (RMS) value of superposition of all information-rich components and real signal (Panel (b)). This result suggests the EMD does not properly separate

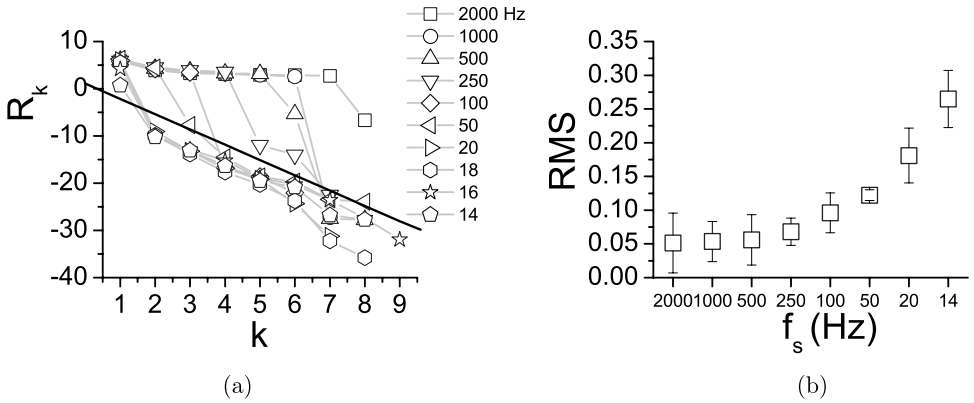


Fig. 8. Determination of minimum sampling rate for $S_{\text{real}}(t)$. Panel (a) shows the R_k distribution of data with various ratios of sampling rate and signal frequency; Panel (b) shows RMS of smoothed data and real signal vs. the frequency ratio.

signal and noise components. It is just the consequence of insufficient number of realizations. Therefore, one can conclude that the minimum sampling rate for this data is 125 HZ.

Similarly, for the second test signal, $S'_{\text{real}}(t) = 1000 * \exp(-t/1000) \sin(2\pi f_{si}t) \cos(2\pi f'_{si}t)$, the signal frequencies f_{si} and f'_{si} are 2 and 4 Hz, respectively and the signal is also contaminated by white Gaussian noise with SNR = 0 dB. The panel (a) of Fig. 9 plots the R_k distribution of data with sampling rates ranging from 8000 to 800 Hz. The dark line shows the linear equation (10) of pure noise R_k^{noise} . The results show that when the sampling rate is lower than 1333 Hz, the R_k distribution of data is within the vicinity of the pure noise R_k^{noise} distribution. Namely, the minimum acceptable sampling rate for this signal is around 1333

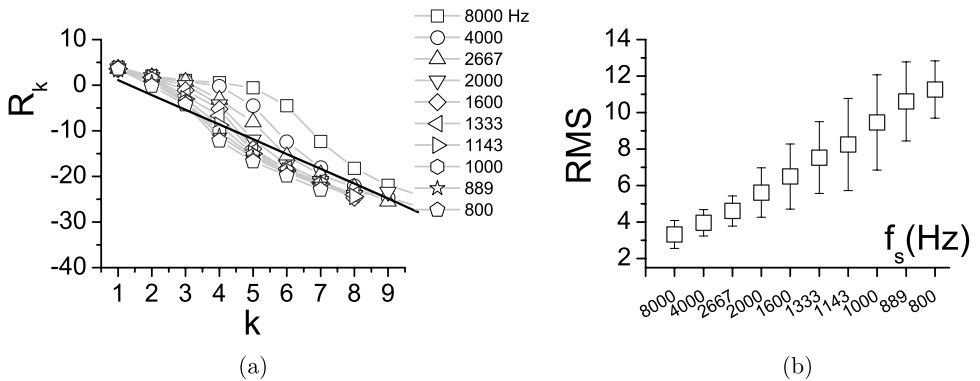


Fig. 9. Determination of minimum sampling rate for $S'_{\text{real}}(t)$. Panel (a) shows the R_k distribution of data with different ratio of sampling rate and signal frequency and Panel (b) shows RMS of smoothed data and real signal vs. the frequency ratio.

Hz. The RMS values of superposition of all information-rich components obtained using sampling rates larger than 1333 Hz and real signal (panel (b)) are all larger than 8.

5.3. The EMD-based low pass filter

There has been many studies, such as Flandrin *et al.* (2004a) and Khaldi *et al.* (2008) proposed to apply EMD-based methods for denoising. Basically, both methods are developed still based on the energy of IMFs. As argued in Sec. 3, the energy interpretation of IMFs may just spell out partial information of IMFs. One can expect that the energy-based approach will face the same issue.

On the contrary, the entropic strategy provides an alternative as a low pass filter. The EMD-based filtering method is straightforward. If the $\mathcal{N}_k(t)$, which contains mostly high frequency IMFs, is identified as information-free components, the $\mathcal{S}_k(t)$ that contains the rest of IMFs and residue will then be recognized as components relevant to real signal. Instead of demonstrating this application by studying the simple test signal used previously, we consider a contaminated male sound wave signal. The results will be compared to three other filters, the WH, five-point mean filter (MFS5), and wavelet denoise approaches.

A one-second long male voice signal with a weak background noise is recorded at a sampling rate of 44.1 kHz, which will serve as the real signal $S'_{\text{real}}(t)$. The waveform of this voice signal is shown as the dark line in top row of Fig. 10. Then, this real signal is linearly superimposed with a white Gaussian noise with SNR = 0 dB and will be treated as raw data $S'_{\text{raw}}(t)$, shown as a gray line in the same figure. The goal is to remove high-frequency white noise from the raw data.

For the EMD-based low pass filter, the EMD is applied to decompose $S'_{\text{raw}}(t)$, which is shown in Fig. 10. There are nine IMFs and one residue. The $\mathcal{N}_k(t)$ and the corresponding probability of $P_{\text{info}}(k)$ containing information, Eq. (6), are then calculated and shown in Fig. 11. When k is larger than 5 and $P_{\text{info}}(k)$ is larger than 0.67. $\mathcal{N}_4(t)$ is an information-free dominant component. The smoothed signal, superposition of rest of the IMFs, and residue are shown as light gray line in panel (b) of Fig. 12. Notes that we present the one-second-long results in two columns; left column, which records 0 to 0.5 second, contains mostly low-frequency parts of signal and the right column, which records 0.5 to 1 second, contains mostly high-frequency parts, to manifestly present performance of the results.

For the WH method, according to the energy distribution of IMFs as shown in Fig. 13, it indicates the IMFs 5 to 8 are above the 99% percentile range of noise energy spread function. Namely, these four IMFs contain information relevant to the real signal with 99 confidence.

The smoothed signal (superposition of IMFs 5 to 8 and residue) is plotted in Panel (c) of Fig. 12. The only difference between the proposed filter and the WH method is that the ninth IMF is identified as a noise term by the WH method.

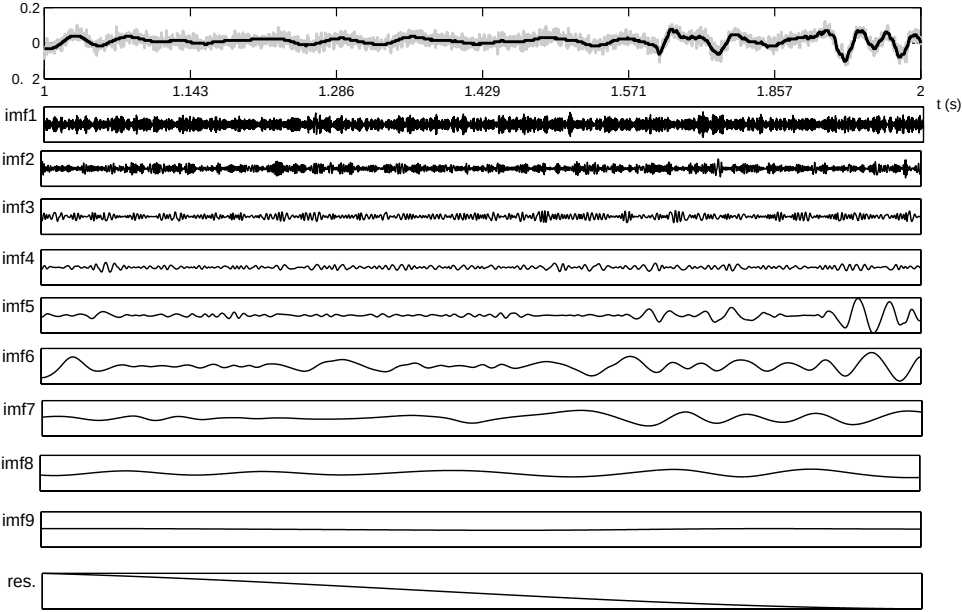


Fig. 10. The IMFs of the contaminated male voice signal with SNR = 0 dB. The top row shows the raw data. The dark line denotes the real signal $S'_{\text{real}}(t)$ and the gray line plots the raw signal $S'_{\text{raw}}(t)$. The EMD results in nine IMFs and one residue.

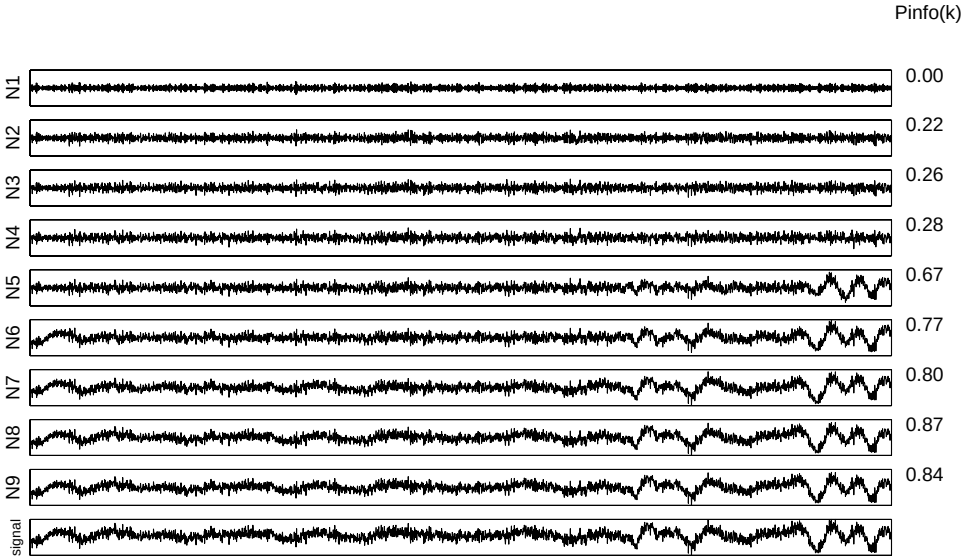


Fig. 11. The figure plots $\mathcal{N}_k(t)$ of the contaminate male voice signal. Furthermore, the corresponding probabilities $P_{\text{info}(k)}$, Eq. (6), are shown on the right of each row.

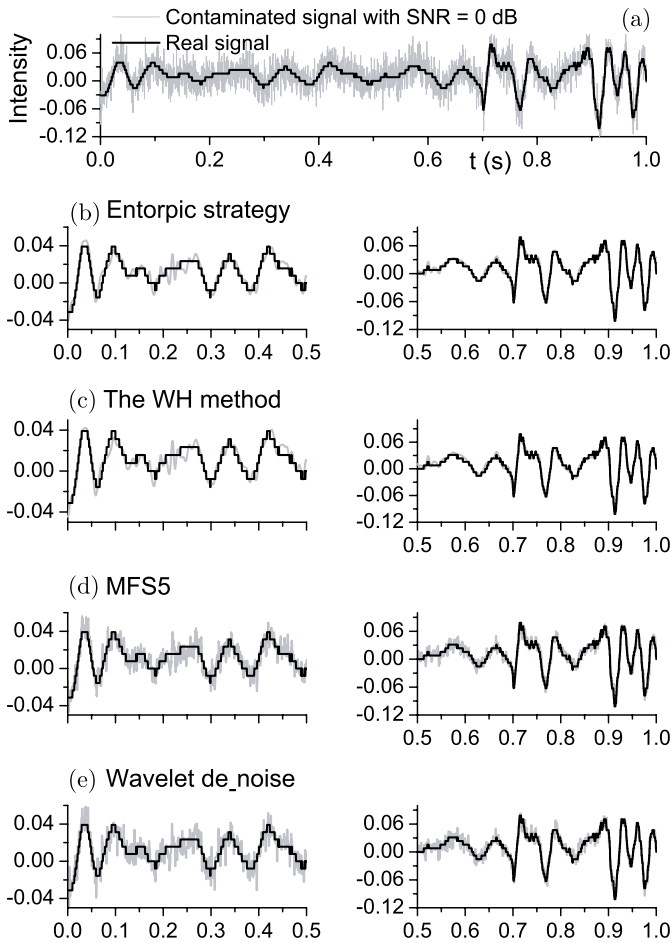


Fig. 12. Demonstration of four filters. Panel (a) shows the clean male voice signal in dark line and the noisy signal with $\text{SNR} = 0\text{ dB}$ shown in light gray line. Gray lines in Panels (b), (c), (d), and (e) plot smoothed signals obtained from the entropic strategy, the WH method, the MFS5, and the wavelet denoise method, respectively.

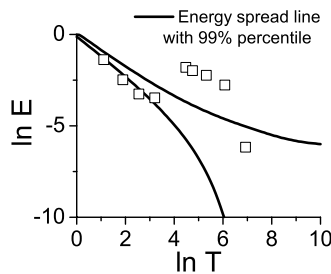


Fig. 13. The figure plots logarithm of energy density vs. mean period of each IMF (hollow square) and energy spread function with 99% percentile of white Gaussian noise based on the WH method.

Table 1. Comparisons of mean RMS and standard deviation of smoothed and real signal. The smoothed signals are obtained from four filters.

SNR (dB)	Entropic strategy	WH	MFS5	Wavelet
20	0.003 ± 0.001	0.003 ± 0.0003	$0.004 \pm 3.24\text{e-}5$	$0.026 \pm 4.05\text{e-}5$
10	0.004 ± 0.001	0.004 ± 0.0002	$0.005 \pm 9.44\text{e-}5$	0.027 ± 0.0001
0	0.008 ± 0.002	0.009 ± 0.002	0.01 ± 0.0003	0.028 ± 0.0004
-10	0.03 ± 0.012	0.068 ± 0.001	0.031 ± 0.001	0.043 ± 0.001
-20	0.08 ± 0.066	0.215 ± 0.002	0.097 ± 0.002	0.11 ± 0.002

The smoothed results of using the MFS5 and wavelet denoise method are shown in Panels (d) and (e), respectively. Note that the soft thresholding with a default value 0.089 generated by the MATLAB tool (ddencmp, which is developed based on Donoho *et al.* approaches [Donoho and Johnstone (1994, 1995); Donoho *et al.* (1995)]) is used in the wavelet denoise method.

Although the figure shows these filters have similar performance in the high-frequency part of the raw data (right column), one can still qualitatively identify the proposed filter and the WH performs the best in this part and is followed by the MFS5 and wavelet denoise method. The same trend is manifest for the low-frequency part as shown in the left column.

Furthermore, 100 trials, which are contaminated by $SNR = 20, 10, 0, -10$, and -20 dB, are generated for investigating differences of these filters statistically. The mean RMS values and standard deviations of the smoothed signal produced by the proposed method, WH, wavelet and MFS5, and real signal are listed in Table 1. In general, the proposed filter outperforms the other three in all five cases. Particularly, it performs the best in the data that is severely contaminated by noise. It is interesting to find that the simplest filter, the five-point mean filter, only slightly outperformed by the proposed filter. Because both the entropic strategy and the WH method show similar performances in $SNR = 20, 10$, and 0 dB case, the energy interpretation is sufficient to quantify the extent of information carried in IMFs. However, when the noise level is increased (-10 and -20 dB), it shows the consideration of the detailed information in IMFs by the entropic strategy is crucial to quantify appropriately the extent of information.

It seems the wavelet denoise method is the worst. However, it should be noted that one always can design a better wavelet denoise filter with a more appropriate soft threshold value based on properties of data rather than with the default value used here [Kopsinis and McLaughlin (2008a, 2008b); Boudraa and Cexus (2006)]. Because our purpose here is only to demonstrate a promising application of the entropic strategy in filter design, we are not pursuing to develop a best filter. The proposed approach sketches a general strategy for designing filters for different purposes. It depends on the clustering method in the first step. Yet it is out of our current scope.

6. Conclusion

We discuss an adaptive signal analysis method, HHT, which has been shown a promising tool to outperform Fourier transform-based analysis methods such as wavelet approach in analyzing nonlinear and nonstationary signals. Particularly, in this study, we focus on the first step of HHT, the empirical mode decomposition. Because the EMD lacks analytical interpretation, we introduce a two-step entropic strategy to investigate properties of IMFs. The entropic approach considers a Bayesian interpretation of probability. Probabilities represent our state of knowledge about the systems of interest. Thus, it brings directly an information-based treatment to interpret IMFs. Furthermore, because the work of Xu *et al.* improves performance of the EMD under low sampling rate [Xu *et al.* (2009)], we can expect a wide applicability of the EMD based on the combination of their methodology and the entropic strategy.

To demonstrate the use of entropic strategy, the properties of IMFs of two test signals are numerically investigated. The extent of information in IMFs can be quantitatively determined through the calculation relative entropy and probability of $\mathcal{N}_k(t)$ containing information.

Three applications of the entropic strategy are demonstrated. First, a characteristic of IMFs of white noise is revealed by the energy ratio of $\mathcal{N}_k(t)$ and $\mathcal{S}_k(t)$ of white noise. The energy ratio is found to be inversely proportional to k value. Second, one can utilize the R_k distribution to determine minimum sampling rate in order to obtain sufficient number of realizations. The third application provides a low pass filter design. As a demonstration, the proposed low pass filter is applied to smooth a contaminated male sound wave signal and we compare it to three other well-known filters. We show that the proposed filter is consistent with the WH method for high SNR cases. It outperforms the WH method in the cases of signal being severely contaminated. Furthermore, both the filters are competitive to the mean filter and wavelet-based filter.

These results not only support the applicability of the WH method but also, mostly, show that the detailed information carried in IMFs can be unveiled by the entropic interpretation.

Acknowledgment

This work is partially supported by grants NSC 95-2811-M-008-016 to CYT from the National Science Council, Taiwan, ROC. The authors are grateful for many discussions with CW Su.

References

- Ai, L., Wang, J. and Wang, X. (2008). Multi-features fusion diagnosis of tremor based on artificial neural network and D = S evidence theory. *Signal Process.*, **88**: 2927–2935.

- Boudraa, A. O. and Cexus, J. C. (2006). Denoising via empirical mode decomposition. *Proceedings of the IEEE International Symposium on Control, Communications and Signal Processing (ISCCSP '06)*, pp. 4–7.
- Cai, C., Liu, W., Fu, J. and Lu, Y. (2006). A new approach for ground moving target indication in foliage environment. *Signal Process.*, **86**: 84–97.
- Chen, C.-C., Tseng, C.-Y. and Dong, J.-J. (2007). New entropy based method for variables selection and its application to the debris-flow hazard assessment. *Eng. Geolo.*, **94**: 19–26.
- Cox, R. T. (1961). *The Algebra of Probable Inference*, Johns Hopkins, Baltimore, MD.
- Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**: 425–455.
- Donoho, D. L. and Johnstone, I. M. (1995). De-noising by soft threshold. *IEEE Trans. Inf. Theory*, **41**: 613–627.
- Donoho, D. L., Johnstone, I. M., Kerkycharian, G. and Picard D. (1995). Wavelet shrinkage: asymptopia. *J. R. Stat. Soc. Series B*, **57**: 301–369.
- Flandrin, P., Goncalvés, P. and Rilling, G. (2004a). Detrending and denoising with empirical mode decompositions. *12th European Signal Processing Conference (EUSIPCO)*, pp. 1581–1584.
- Flandrin, P., Rilling, G. and Goncalvés, P. (2004b). Empirical mode decomposition as a filter bank. *IEEE Signal Proc. Lett.*, **11**: 112–115.
- Huang, H. and Pan, J. (2006). Speech pitch determination based on Hilbert-Huang transform. *Signal Processing*, **86**: 792–803.
- Huang, N. E. and Wu, Z. (2008). A review on Hilbert-Huang transform: method and its applications to geophysical studies. *Rev. Geophys.*, **46**: RG2006.
- Huang, N. E., Chen, Z., Long, S. R., Wu, M. L., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C. and Liu, H. H. (1998). The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. R. Soc. Lond. A*, **454**: 903–995.
- Huang, N. E., Wu, M.-L., Qu, W., Long, S. R. and Shen, S. S. P. (2003). Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Appl. Stochastic Models Bus. Ind.*, **19**: 245–268.
- Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Blank, K. and Liu T. W. (2009). On instantaneous frequency. *Adv. Adapt. Data Anal.*, **2**: 177–229.
- Khalidi, K., Turki-Hadj, M. and Boudraa, A.-O. (2008). A new EMD denoising approach dedicated to voiced speech signals. *2nd Int. Conf. Circuits Systems (SCS 2008)*, pp. 1–5.
- Kizhner, S., Flatley, T. P., Huang, N. E., Blank, K. and Conwell, E. (2004). On the Hilbert-Huang transform data processing system development. *2004 IEEE Aerospace Conf. Proc.*, Big Sky, MT, 6–13 March 2004, pp. 1961–1979.
- Kopsinis, Y. and McLaughlin, S. (2008a). Empirical mode decomposition based soft-thresholding. *16th Eur. Signal Process. Conf. (EUSIPCO)*.
- Kopsinis, Y. and McLaughlin, S. (2008b). Empirical mode decomposition based denoising techniques. *1st IAPR Workshop on Cognitive Information Processing (CIP)*, pp. 42–47.
- Kullback, S. (1997). *Information Theory and Statistics*, Dover, New York.
- Lang, M., Guo, H., Odegard, J. E., Burrus, C. S. and Wells Jr., R. O. (1996). Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Process. Lett.*, **3**: 10–12.
- Liang, H., Bressler, S. L. and Buffalo, E. A. (2005). Empirical mode decomposition of field potentials from Macaque V4 in visual spatial attention. *Biol. Cybern.*, **92**: 380–392.

- Polygiannakis, J., Preka-Papadema, P. and Moussas, X. (2003). On signal-noise decomposition of time-series using the continuous wavelet transform: Application to sunspot index. *Mon. Not. R. Astron. Soc.*, **343**: 725–734.
- Rilling G. and Flandrin P. (2009). Sampling effects on the empirical mode decomposition. *Adv. Adapt. Data Anal.*, **1**: 43–59.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Tech. J.*, **27** 379–423 and 623–656.
- Sternickel, K., Effer, A., Lehnertz, K., Schreiber, T. and David, P. (2001). Non-linear noise reduction using reference data. *Phys. Rev. E*, **63**: 0362091–0362094.
- Tseng, C.-Y. (2006). Entropic criterion for model selection. *Physica A*, **370**: 530–538.
- Wu, M.-C. and Hu, C.-K. (2006). Empirical mode decomposition and synchrogram approach to cardiorespiratory synchronization. *Phys. Rev. E*, **73**: 0519171–05191711.
- Wu, Z. and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. R. Soc. Lond. A*, **460**: 1597–1611.
- Wu, Z. and Huang, N. E. (2005). Statistical significance test of intrinsic mode functions. *Hilbert-Huang Transform and Its Applications*, Ed. N. E. Huang, World Scientific Inc., Singapore.
- Wu, M.-C., Huang, M.-C., Yu, H.-C. and Chiang T. C. (2006). Phase distribution and phase correlation of financial time series. *Phys. Rev. E*, **73**: 0161181–01611816.
- Xu, Z., Huang, B. and Zhang, F. (2009). Improvement of empirical mode decomposition under low sampling rate. *Signal Process.*, **89**: 2296–2303.