Advances in Adaptive Data Analysis
Vol. 3, No. 4 (2011) 493–508

DOI: 10.1142/S179353691100088X

A NEW APPLICATION OF ENSEMBLE EMD AMELIORATING THE ERROR FROM INSUFFICIENT SAMPLING RATE

DISHAN HUANG* and YULIN XU

School of Mechatronics Engineering and Automation,
Shanghai University,
Shanghai 200072, P. R. China
**dishan@shu.edu.cn*

The objective of this paper is to apply an assisted noise method for ameliorating the empirical mode decomposition (EMD) error from insufficient sampling rate for a vibration signal. When the intrinsic mode functions (IMFs) are extracted from a signal mixed noise at a certain level on the sifting algorithm, an extraordinary phenomenon, where noise submerges the EMD error, is discovered. Thus, noise-assisted data is proposed to disturb the EMD error in the sifting process. In order to cancel out noise after serving its purpose, the IMFs are processed with an ensemble mean. As a result, the noise-assisted data ameliorates the EMD error from insufficient sampling rate, and the method treats the mean as the final true result. An EMD example of ball bearing vibration is presented to illustrate the validity of the approach. This paper recommends implementing the noise-assisted method in the EMD on vibration and acoustic signals with broad band.

Keywords: Empirical mode decomposition; sampling; intrinsic mode function; empirical mode error; cancellation.

1. Introduction

Huang *et al.* (1998) proposed the empirical mode decomposition (EMD) method to decompose nonlinear and nonstationary signals into a series of intrinsic mode functions (IMFs) that are optimized for instantaneous frequency estimation using Hilbert-based technique. Significant mathematical works (Chen *et al.*, 2006 Qin and Zhong, 2006) have been dedicated to the detailed analysis of the local EMD method. Now, EMD has been widely used in the area of signal detection, system identification, and fault diagnosis (Rai and Mohanty, 2007; Ramesh Babu *et al.*, 2008; Yan and Gao, 2006).

*Corresponding author.

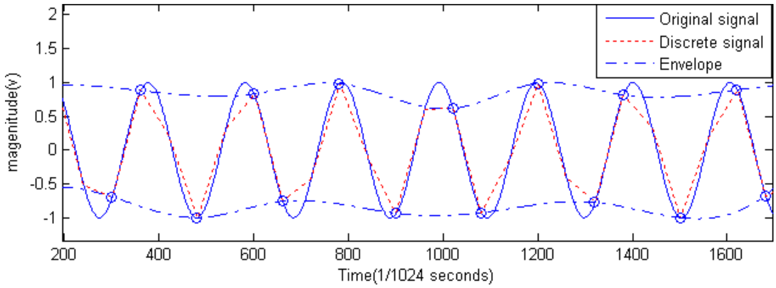


Fig. 1. Sketch of deviation in envelope computation caused by insufficient sampling rate.

Although the signal used in the sifting algorithm has implicitly been considered as time continuous, it is often discrete in time when used in practice. Local extrema are consequently shifted in the discrete signal and cannot be accurately found in computations. As shown in Fig. 1 where the solid line represents a continuous sine wave and the dotted line represents a discrete sine wave, the "envelope" of the continuous wave is obviously different from that of the discrete one, and the discrete sine wave is no longer a constant "envelope." Essentially, the EMD-processing results in the deviation of the envelope computation due to poor amplitude resolution in the discrete signal. The deviation magnifies, as the signal sampling rate decreases and incurs changes in the up-down envelope mean. As a result, when the up-down envelope mean error enters iteration in the sifting process, it influences the IMFs and causes the EMD error. Stevenson *et al.* (2005) presented the linear frequency-modulated (LFM) signal to derive the relationship between EMD error and signal frequency at a fixed sampling rate and proposed to use five times the Nyquist frequency as the signal sampling frequency. Rilling and Flandrin (2006) also evaluated the EMD error related to signal sampling. However, many detailed properties of the EMD error need to be studied.

This paper thoroughly investigates the EMD error caused by insufficient sampling rate (signal sampling obeys the law of Nyquist sampling, but the sampling rate is not high enough) and proposes to apply the noise-assisted data to ameliorate the EMD error. The study contains four aspects:

(1) It provides the basic concept of the EMD error related to insufficient sampling rate;
(2) It investigates the EMD characteristic of mixing signal with noise;
(3) It applies the ensemble EMD with the added noise to ameliorate the EMD error; and
(4) It processes ball-bearing vibration signal using the ensemble EMD with the added noise data to illustrate the advantage of the approach.

In the investigation, numerical computation was chosen over an analytical solution to analyze the EMD error because only an analytical mathematical expression is not enough to describe the EMD algorithm that mixes with logical operation.

2. EMD Error

For a discrete sine wave, ideal decomposition will be achieved if the sampling frequency is high enough such that the wave will keep its waveform as a unique IMF after the EMD operation. In contrast, if the sampling frequency is not high enough but is still subjected to the law of Nyquist sampling, the discrete sine wave will be decomposed into several IMF.

In order to illustrate the above-stated phenomenon, consider a sine wave with unit amplitude and 330 Hz of frequency. As shown in Fig. 2(a), while the sampling frequency is 1,024 Hz, the waveform of the discrete wave looks distorted and appears to be a case of "multiple harmonics." After the EMD processing, the one mode problem is changed to a case of multiple modes, and the sine wave is decomposed into two IMFs, as represented in Figs. 2(b) and 2(c), respectively.

Spectral analysis is carried out for all signals shown in Fig. 2 to understand the characteristics of the above-mentioned decomposition result (corresponding spectrums are shown in Fig. 3). The decomposition result IMF_1 is no longer an ideal mono-component, and it mixes with an extra component referred to as the mode error that is defined as the difference between the ideal mode and the decomposed mode IMF_1. However, there are several IMFs produced in the sifting process, but only one is correlated to the original sine wave, whereas the others are insignificant IMFs. In given example, another decomposition result, IMF_2, is an insignificant

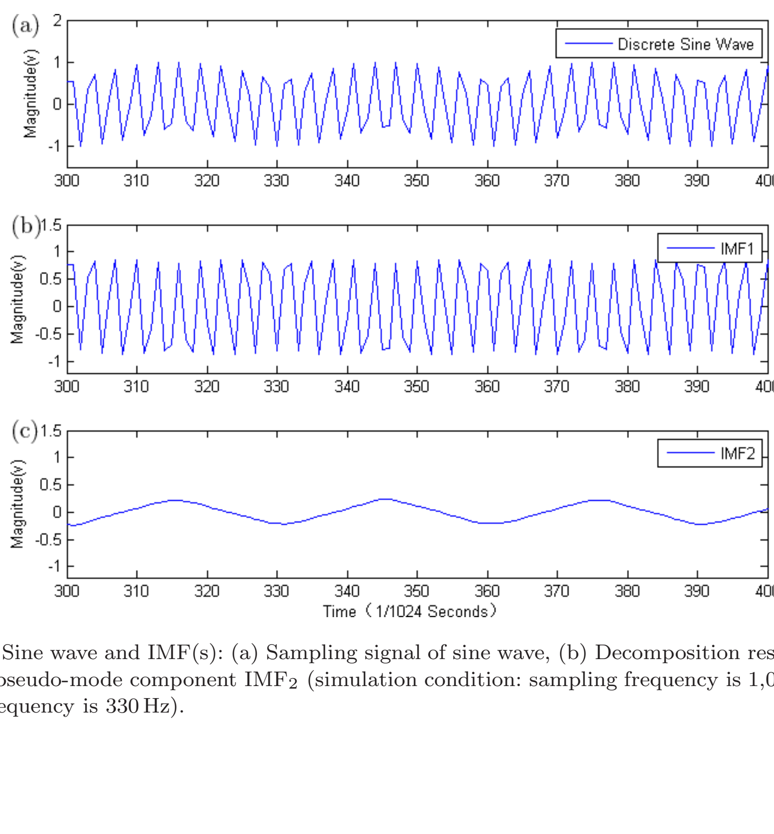


Fig. 2. Sine wave and IMF(s): (a) Sampling signal of sine wave, (b) Decomposition results IMF_1 and (c) pseudo-mode component IMF_2 (simulation condition: sampling frequency is 1,024 Hz and signal frequency is 330 Hz).

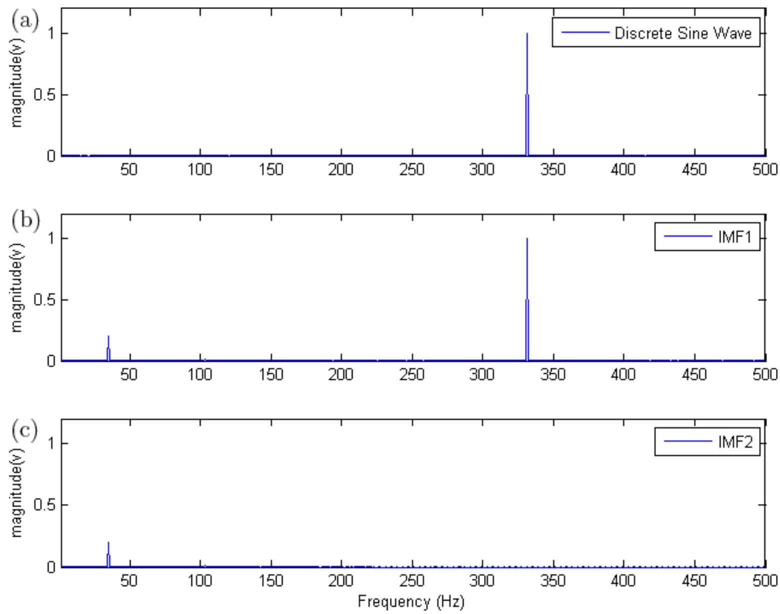


Fig. 3. Spectrum of sine wave and IMF(s): (a) Sampling signal of sine wave, (b) decomposition result IMF_1 and (c) pseudo-mode component IMF_2.

IMF and in this paper it is referred to as a pseudo-mode component. The mode error and the pseudo-mode components compose the concerned EMD error in the case of insufficient sampling rate.

The problem can also be examined in terms of energy conservation. The energy of the original sine wave $\hat{E}_0$ was calculated to be 0.5005, and the energy of the first-order IMF $\hat{E}_1$ and the second-order IMF $\hat{E}_2$ are 0.5212 and 0.0207, respectively. It is obvious that the EMD decomposition is not subjected to the law of energy conservation while the mode error and the pseudo-mode component are present in the EMD decomposition. In this case, their energy relationship in the original EMD decomposition can be summarized as follows:

$$\hat{E}_0 < \hat{E}_1 + \hat{E}_2. \tag{1}$$

Since energy is not conserved if there is the EMD error, the energy relationship (Equation (1)) between the original signal and all IMF(s) can be employed to inspect the existence of the EMD error in the decomposition result and assess the EMD decomposition accuracy.

In addition, the pseudo-mode component has the same magnitude as the mode error in IMF_1, but opposite in phase. The pseudo-mode component comes from the same source as the mode error, but it distributes in the higher-order IMF. As a result, a correlation coefficient between the first-order IMF and the higher-order IMF(s) will be negative, and a correlation coefficient between the sine wave and IMF(s) will be changed. In fact, the correlation coefficient C_{12} is -0.1993 and C_{s1} is reduced from 1 to 0.9789 in the given example. Therefore, the correlation measure can be used as an auxiliary tool to inspect the existence of the EMD error in the decomposition result.

3. EMD of Noise Signal

The EMD of a white noise signal is essentially similar to binary wavelet decomposition (Wu and Huang, 2004; Flandrin *et al.*, 2004), and it has the same effect as a filter group.

The following numeric computation example provides explanation for the EMD of white noise. In the computation, a B-spline interpolation algorithm is first used in the EMD operation. Next, the SD is designated 0.2 as a reasonable recursion criterion. EMD of white noise signal $n(t)$ consisting of 10,240 data points is independently carried on 5,000 times to obtain the first seven IMFs (IMF_1–IMF_7), which are correspondingly normalized (zero-dimension frequency). Finally, the frequency spectrum of each $IMF_i(t)$ is compiled on one chart (Fig. 4). Based on numerical computation experiments on uniformly distributed white noise using the EMD method, the function of EMD is found to be a dyadic filter, where the IMF components are all normally distributed.

4. EMD of Signal Mixed with Noise

When the harmonic wave is mixed with the white noise, the EMD operation serves as the dyadic filter for the data. The implication of this operation is that a signal of a similar scale in the data set could possibly be contained in one IMF component. As shown in the numerical experiment, the EMD error from insufficient sampling rate is still present in the decomposition result when the harmonic wave is mixed with a small portion of noise (Fig. 5 shows the decomposition result of the EMD operation on the sine wave of magnitude 1 mixed with the white noise of magnitude 0.1). The mode error could be found in the first-order IMF, but the pseudo-mode components are now in the form of harmonics and are additional components in the decomposition. As shown in Fig. 5, the mode error can be found in IMF_1, similar to that in Fig. 3, but the pseudo-mode harmonic occurs in IMF_3.

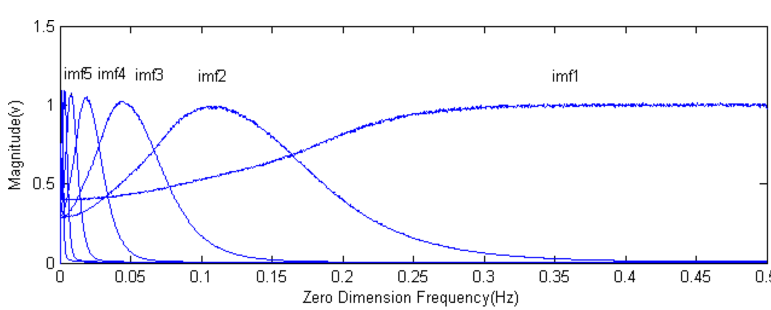


Fig. 4. Filter bank frequency domain characteristic of EMD decomposition.

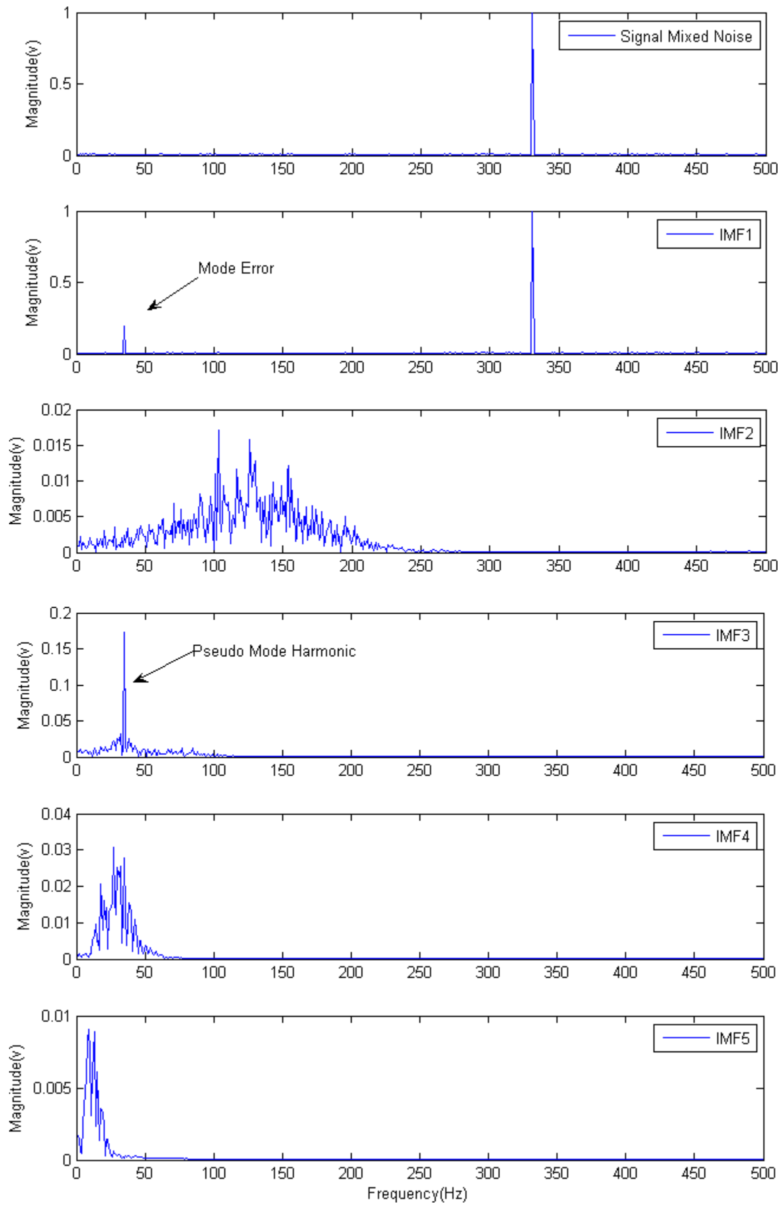


Fig. 5. EMD decomposition on a sine wave mixed with white noise: (a) (the ratio of white noise to sine wave of 0.1, sine wave with frequency 330 Hz, and sampling frequency of 1,024 Hz).

However, as the white noise proportion increased to a certain level in the data, the mode error and the pseudo-mode harmonic disappeared. As shown in Fig. 6 where the amplitude of the sine wave remains at 1 and the magnitude of the mixed noise increases to 0.6, the mode error and the pseudo-mode harmonic can no longer

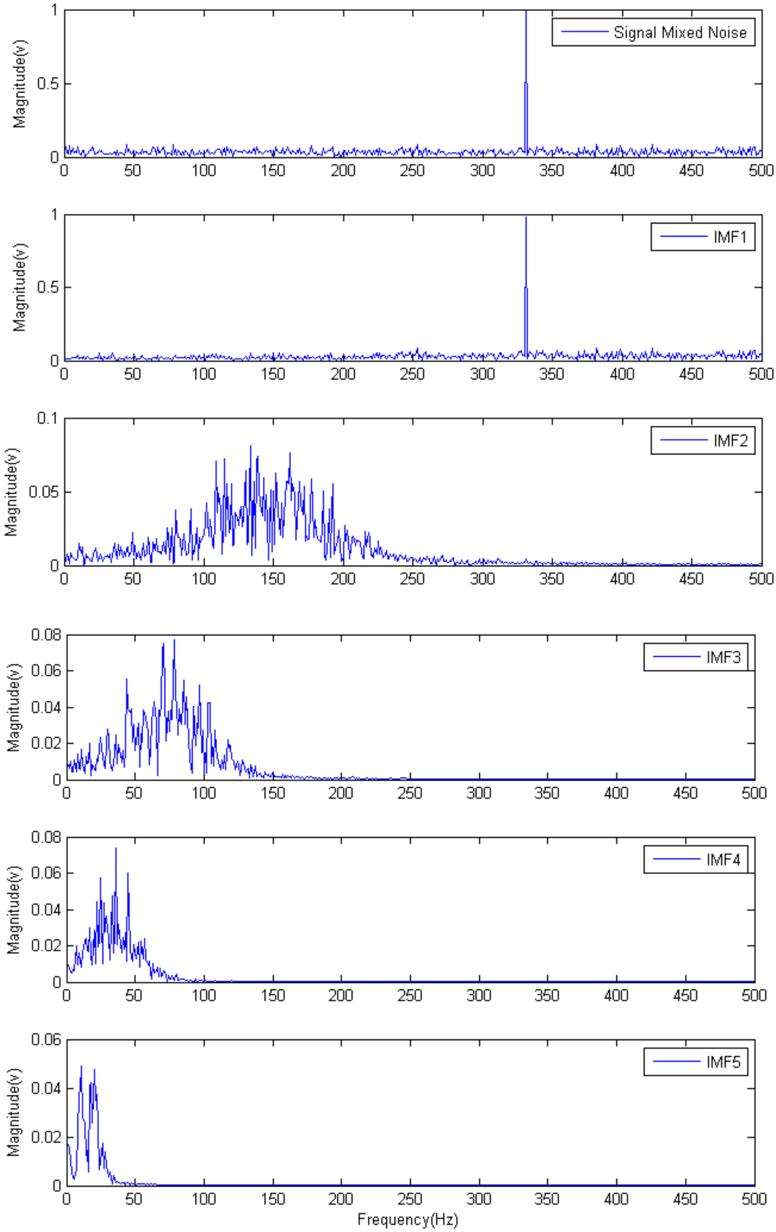


Fig. 6. EMD decomposition on a sine wave mixed with white noise — no mode error: and pseudo-mode harmonics caused by insufficient sampling rate (the ratio of white noise to sine wave of 0.6, sine wave with frequency 330 Hz and sampling frequency of 1,024 Hz).

be seen. An explanation for the phenomenon is that the white noise disturbs the deviation in the envelope computation in the sifting process, resulting in the submerging of the pseudo-envelope formation. In effect, the mixed noise thoroughly submerses the mode error and pseudo-mode harmonics from insufficient sampling rate.

5. Noise-Added Data Analysis

As stated in the above section, added noise with finite rather than infinitesimal amplitude to data indeed creates a noisy data set. Therefore, the added noise, having uniformly filled all the scale space, can help to solve the problem caused by insufficient sampling rate. Of course, a single trial will produce very noisy decomposition results because each of the noise-added decomposition consists of the signal and the added white noise. Since the noise in each trial is different in separate trials, it can be canceled out in the ensemble mean with enough trails. In the end, the ensemble mean is treated as the true answer, where the signal of the decomposition result remains unchanged, as more trials are added into the ensemble. Based on the above results, this paper proposes the concept of ameliorating the EMD error using noise-assisted data and gives a new application of ensemble EMD (EEMD).

The decomposition result of noise-assisted data is defined by the number of trials approaching infinity as following:

$$\hat{c}_j(t) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} [c_j(t) + \alpha n_i(t)], \tag{2}$$

where i represents the number of trials and j the order of IMF in the noise-added signal, and α is the magnitude of the added noise. The number of trials in the ensemble, N, has to be large. The difference between the true decomposition result of the signal and the result of the ensemble is governed by the well-known statistical rule:

$$\varepsilon_n = \frac{\varepsilon}{\sqrt{N}}, \tag{3}$$

where, ε is the amplitude of the added noise and ε_n is the final standard deviation of the error, which is defined as the difference between the input signal component and the corresponding IMF(s). The equation shows that as N increases, the standard deviation, ε_n, decreases.

Based on the known properties of the EMD, the signal processing is described as follows:

(1) Add a white noise series to the targeted data;
(2) Decompose the data with added white noise into IMF(s);
(3) Repeat step 1 and step 2 N times but using a different series of white noise in each trial; and
(4) Obtain the (ensemble) means of corresponding IMF(s) as the final result.

Note that literature (Wu and Huang, 2009) presented the ensemble EMD with added noise to overcome the drawbacks of mode mixing in the original EMD and successfully decomposed a nonstationary signal, such as the digitalized sound "Hello." In this paper, the ensemble EMD with added noise is used to ameliorate the EMD error from insufficient sampling rate. This approach takes advantage of the statistical characteristics of white noise in order to perturb the signal in its true solution neighborhood and then to cancel itself out after serving its purpose. Therefore, the method represents a substantial improvement over the original EMD and is truly a noise-assisted data analysis.

6. Ball Bearing Vibration

The envelope analysis method is a practical analytic approach for the fault diagnosis of ball bearing. Recently, the envelope analysis method and the EMD method have been combined and applied to the fault diagnosis of ball bearings to obtain better analytic results (Yu *et al.*, 2009; Du and Yang, 2007). In the signal processing, the first step is to decompose the vibration signal into different frequency bands (IMF(s)) with time scale by the EMD decomposition and then to apply the Hilbert transform to one of the mode components. After obtaining the envelope spectrum, the defect characteristics of ball bearing can also be identified.

However, the harmonic components of bearing vibration can be found in the range of low to high frequency. In fact, this phenomenon influences the signal processing of EMD, where although the EMD could be used to decompose the ball bearing vibration signal, the higher harmonic components would be prone to mode error in the first-order IMF and pseudo-mode components in the higher-order IMF (as mentioned in Secs. 2–4). As a result, the accuracy of the EMD decomposition of the bearing vibration signal would also be unreliable. Since the high-frequency component in bearing vibration produces the EMD decomposition error, the problem can be solved using the ensemble EMD approach.

The bearing vibration signal in Figs. 7(a) and 7(b) is an engineering example (where B&K instrument PULSE3560_B is used for sampling data and the data sampling rate is 65,434 Hz), and it is used to assess the accuracy of EMD decomposition. The signal was detected in the bearing quality inspection, and the energy of the original vibration signal $\hat{E}_0 = 119.7$. The IMF(s) obtained from the original EMD are plotted using a uniform scale in Figs. 8(a) and 8(b). Here, the total energy estimated from the 1st IMF to 8th IMF (figures of the 6th to 8th IMF are omitted) is $\sum_i^8 \hat{E}_i = 163.8$, and the energy relationship is

$$\hat{E}_0 < \sum_{i=1}^{8} \hat{E}_i. \tag{4}$$

Clearly, the original EMD decomposition is not subjected to the law of energy conservation because $\hat{E}_0$ must be greater than or equal to the total energy of the

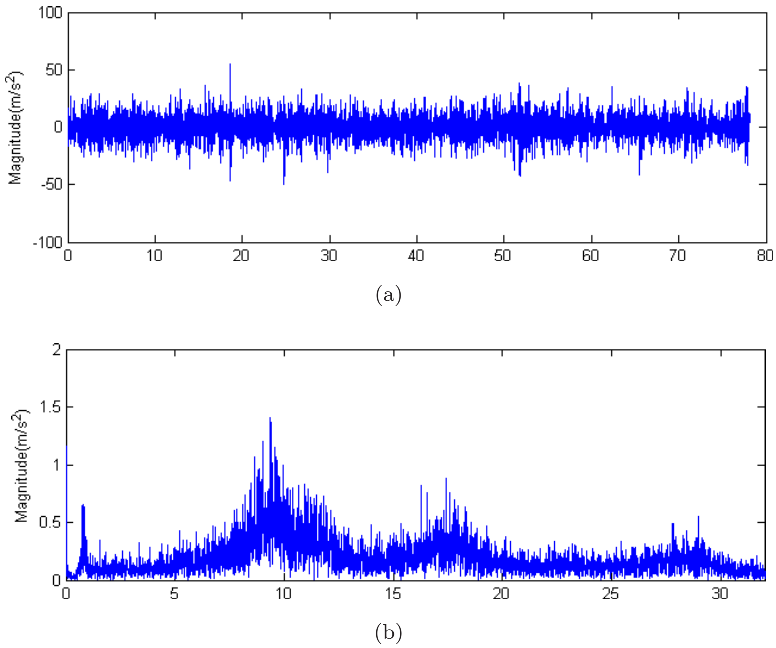


Fig. 7. Vibration acceleration and its spectrum: (a) Bearing vibration acceleration signal and (b) spectrum of bearing vibration acceleration.

IMF(s). The difference in energy reveals the loss in accuracy of the signal decomposition. Therefore, the original EMD decomposition is unreliable for the given problem of bearing vibration signal with broad band.

To improve the accuracy of EMD decomposition, a noise is introduced to the data. The same bearing vibration signal is then processed with the EMD with an added noise of magnitude 4, and 256 trials are carried out in the ensemble. The EMD decomposition results are shown in Fig. 9.

Comparing Fig. 9(b) with Fig. 8(b), the maximum magnitude in the range of 0–15 kHz of the 1st IMF has been reduced from 0.9 to 0.6. The effect of the pseudo-mode component has also been ameliorated, which can be observed from the reduction in the peak-to-peak value of the higher-order IMF components.

In terms of energy, the total energy from the 1st IMF to the 8th IMF in Fig. 9 is estimated to be $\sum_{i=1}^{8} \hat{E}_i = 103.9$, and then the energy relationship is

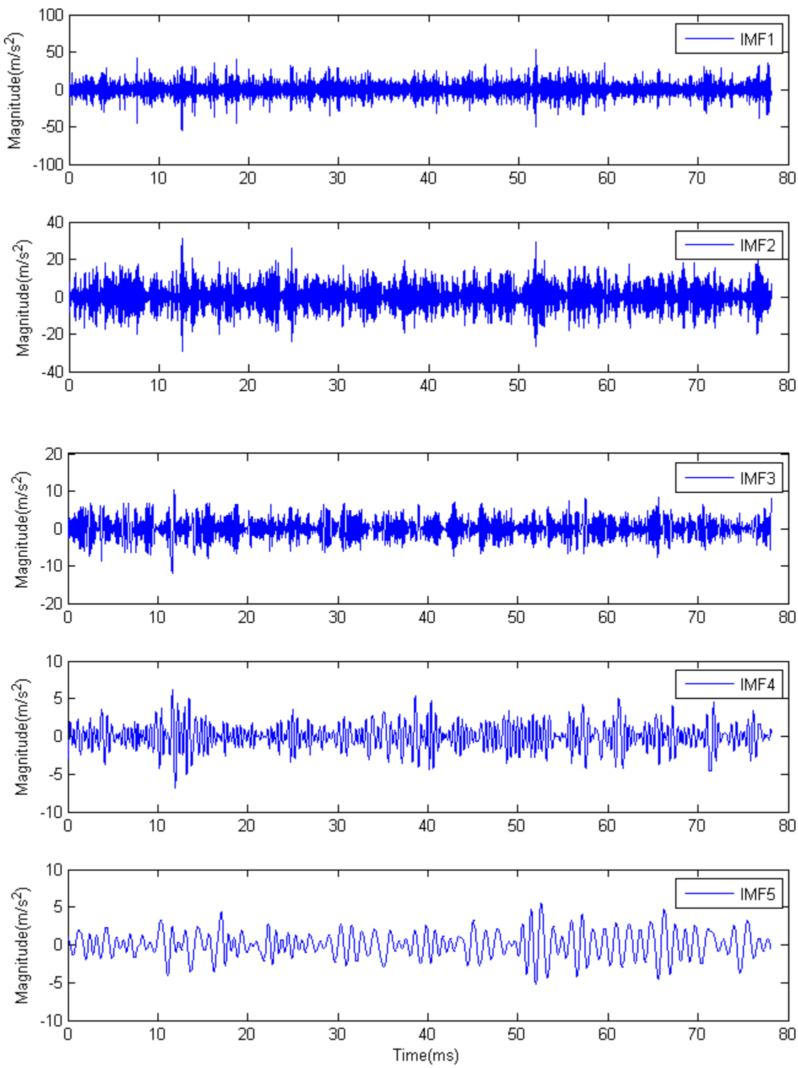
$$E_0 > \sum_{i=1}^{8} \hat{E}_i. \tag{5}$$

Fortunately, the EEMD with the added noise is now subjected to the law of energy conservation, consequently improving the results of the EMD decomposition.

In addition, the approach changes the correlation coefficients. As shown in Table 1, the correlation coefficients C_{vj} between the bearing vibration and its

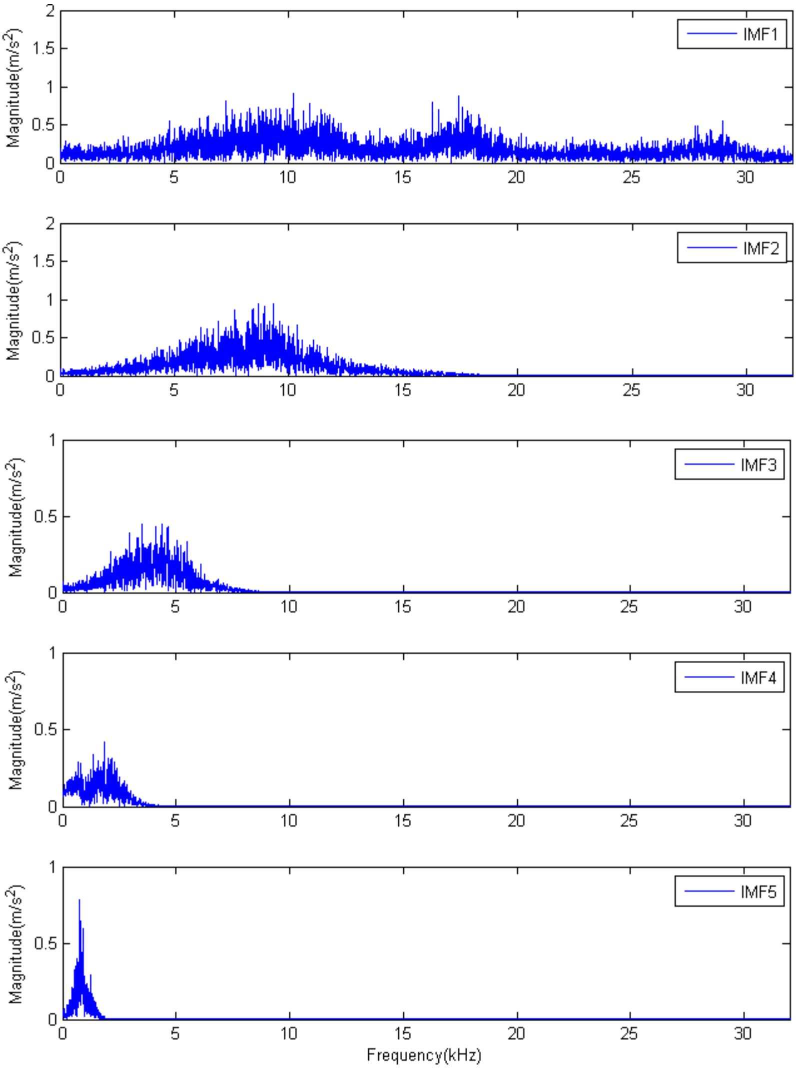
IMF(s) are investigated by using the original EMD and the EEMD, and it is obviously seen that all of correlation coefficients C_{vj} are increased, as the EEMD is employed.

As mentioned in Sec. 3, the EMD operation serves as the binary wavelet decomposition for the data, and the correlation coefficients between two neighbor intrinsic modes should be positive. Since the original EMD would incur the mode error and the pseudo-mode components in the case of insufficient sampling rate, the



(a)

Fig. 8. Original EMD of bearing vibration signal: (a) IMFs of bearing vibration acceleration without additional noise and (b) spectrum of each IMF.



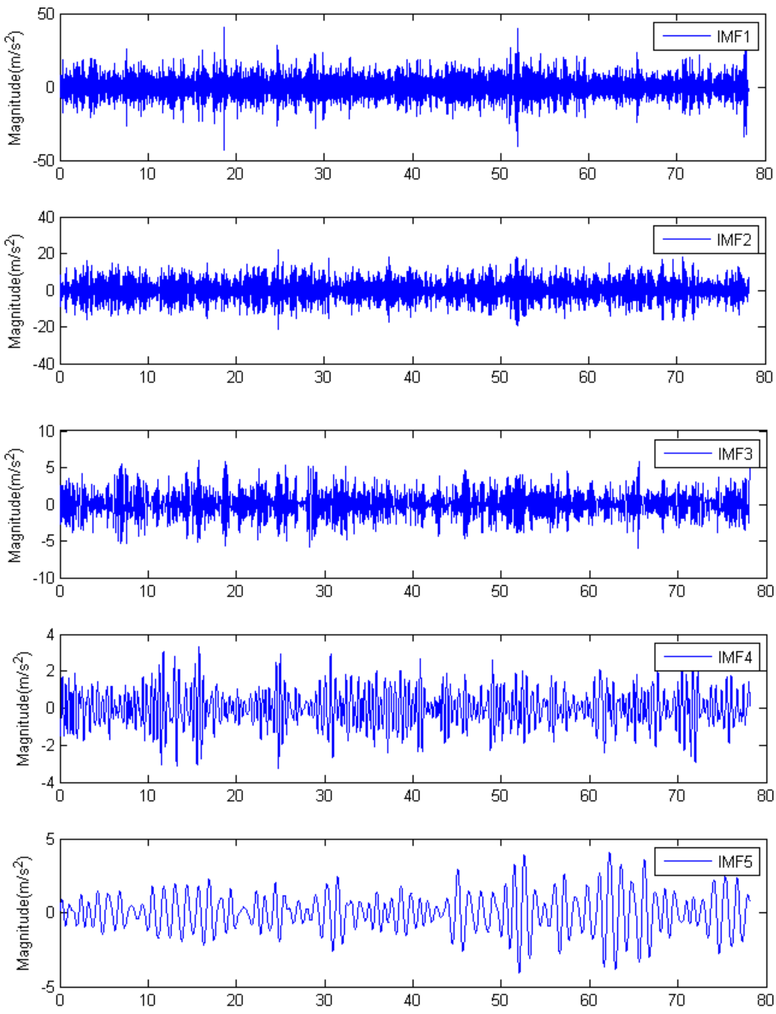
(b)

Fig. 8. (*Continued*)

correlation coefficients between two neighbor intrinsic modes would be negative. Table 2 shows that all of correlation coefficients between two neighbor intrinsic modes are negative in the original EMD decomposition. However, all of the correlation coefficients are positive in the case of EEMD decomposition.

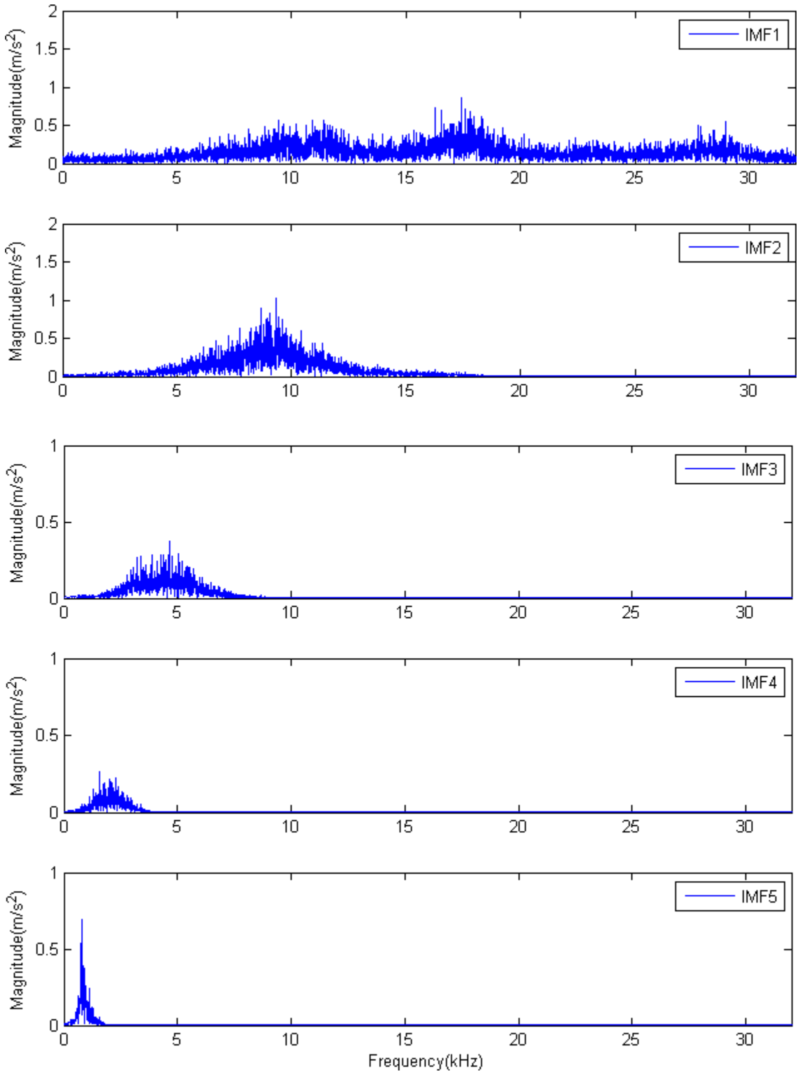
Besides, the absolute value of correlation coefficients C_{1j} between the first IMF and the higher IMF(s) decreases when the EEMD is employed. As listed in Table 3, the absolute value of coefficients C_{13}, C_{14}, and C_{15} are significantly reduced. Therefore, the EEMD with the added noise also improves the correlation measure of the EMD decomposition results IMF(s).

The numeric computation results clearly demonstrate that the EEMD with the added noise has the capability of cancelling the mode error and the pseudo-mode components that disturb the underlying physics.



(a)

Fig. 9. Result of EEMD on ball bearing vibration signal with an added noise (magnitude 2, 256 trials): (a) IMFs of bearing vibration acceleration without additional noise and (b) spectrum of each IMF.



(b)

Fig. 9. (*Continued*)

Table 1. Correlation coefficient C_{vj} between bearing vibration and its IMF(s).

	C_{v1}	C_{v2}	C_{v3}	C_{v4}	C_{v5}	C_{v6}	C_{v7}	C_{v8}
Original EMD	0.7248	0.4534	0.1017	0.0579	0.0845	0.0102	0.0004	0.0009
EEMD	0.8017	0.6736	0.1772	0.1073	0.1213	0.0338	0.0063	0.0116

Table 2. Correlation coefficient $C_{j(j+1)}$ between two neighbor different order IMF(s).

	C_{12}	C_{23}	C_{34}	C_{45}	C_{56}	C_{67}	C_{78}
Original EMD	−0.1902	−0.0457	−0.0200	−0.0956	−0.0475	−0.1202	−0.0514
EEMD	0.1729	0.0837	0.1713	0.1619	0.1988	0.1780	0.1630

Table 3. Correlation coefficient C_{1j} between the 1st IMF and the higher IMF(s).

	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}
Original EMD	−0.1354	−0.0663	−0.0358	−0.0383	−0.029	−0.0096
EEMD	−0.069	−0.0347	−0.0183	−0.0350	−0.026	−0.0089

7. Conclusion and Discussion

Using the sine wave and the white noise as the study objects, this paper explained how the mode error and the pseudo-mode component could be eliminated when a certain level of noise is added the data set. The explanation provided the basis of implementing noise to ameliorate the EMD error from insufficient sampling rate. Furthermore, the example of the EEMD on bearing vibration signal demonstrated the validity of the approach.

The basic principle of the noise-added method for the EMD is simple, yet the power of this approach is obvious from the example of ball bearing vibration. The EEMD indeed can separate signals of different scales without special sampling requirement. Specifically, the noise-added method for the EMD presents a correction on the original EMD error from inefficient sampling rate and can be used without any subjective intervention. Thus, it retains the characteristics of an adaptive data analysis. In addition to canceling the mode error and the pseudo-mode harmonics, the method also produces a set of IMF(s) that bears physical meaning and decomposition results in a time frequency domain that obeys the law of energy conservation. Therefore, with the added noise and the ensemble approach, the EMD has become a more mature tool for nonlinear and nonstationary time series analyses.

Despite the advantages of the approach, the noise-added method still has a drawback: no existing method can consistently determine the level of noise that is required to be added. The level of added noise is of critical importance. The study employed trial and error to find the specific level of noise and found that if the level of added noise is too small it would not be able to submerge the mode error and the pseudo-mode harmonics, and if the level is too high, more trials would be needed in the ensemble approach. However, the cause of the mode error is known to be the high-frequency components in the data, for instance, in the range of 0.25–0.5 sampling frequency. As a result, a possible solution is to add noise at a level proportional to the sum of energy of the high-frequency components in the data. The next step would be to carry out an energy statistic and determine the required level of added noise before the using the EEMD.

References

Chen, Q. H., Huang, N. E., Riemenschneider, S. *et al.* (2006). A B-spline approach for empirical mode decompositions, *Adv. Comput. Math.*, **24**: 171–195.

Du, Q. and Yang, S. (2007). Application of the EMD method in the vibration analysis of ball bearings, *Mech. Syst. Signal Pr.*, **1**: 2634–2644.

Flandrin, P., Rilling, G., Gonc, P. E. *et al.* (2004). Empirical mode decomposition as a filter bank, *IEEE Sig. Proc. Lett.*, **11**(2): 112–114.

Huang, N. E. Shen, Z., Long, S. R. *et al.* (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. A*, **454**: 903–995.

Huang, D. and Xu, Y. (2011). Problem of insufficient sampling rate in EMD, *Instrumentation and Measurement Technology Conference (I2MTC)*, Hangzhou, pp. 244–249.

Qin, S. and Zhong Y. (2006). A new algorithm of Hilbert-Huang transform, *Mech. Syst. Signal Pr.*, **20**: 1941–1952.

Rai, V. K. and Mohanty, A. R. (2007). Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform, *Mech. Syst. Signal Pr.*, **21**: 2607–2615.

Ramesh Babu, T., Srikanth, S. and Sekhar, A. S. (2008). Hilbert-Huang transform for detection and monitoring of crack in a transient rotor, *Mech. Syst. Signal Pr.*, **22**: 905–914.

Rilling, G. and Flandrin, P. (2006). On the influence of sampling on the empirical mode decomposition, *Proceedings of ICASSP 2006*, Toulouse, Vol. 3, pp. 444–447.

Stevenson, N., Mesbah, M. and Boashash, B. (2005). A sampling limit for the empirical mode decomposition, *Proceedings — 8th International Symposium on Signal Processing and its Applications*, ISSPA 2005, Sydney, pp. 647–650.

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

Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise assisted data analysis method, *Adv. Adapt. Data Anal.*, **1**(1): 1–41.

Yan, R. and Robert, X. G. (2006). Hilbert-Huang transform-based vibration signal analysis for machine health monitoring, *IEEE Trans. Instrum. Meas.*, **55**: 2320–2329.

Yu, D., Cheng, J. and Yang, Y. (2005). Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, *Mech. Syst. Signal Pr.*, **19**: 59–270.