Investigations on Key Issues of Seismic Data Processing Based on Ensemble Empirical Mode Decomposition

X. T. Yang
Graduate Student, Institute of Earthquake Engineering,
Faculty of Infrasturcture Engineering. Dalian University of Technology, 116023,China

D. S. Li, L. Ren & H. N. Li
Associate Professor, Institute of Earthquake Engineering,
Faculty of Infrasturcture Engineering. Dalian University of Technology, 116023,China

SUMMARY:
Seismic signal is a kind of nonlinear and non-stationary time series. The empirical mode decomposition (EMD) is adaptive and suitable for seismic data analysis. The ensemble empirical mode decomposition (EEMD) adds various white noises to seismic signal and removes original noise ingredients by using statistical characteristics of added noises. However, the added noise levels are still unsolved problems in applying EEMD. In the paper, noise-polluted data are projected into phase space by time delays and added noise levels in EEMD are determined through the singular values in the process of signal phase space reconstruction. In this paper, authors present numerical simulate signal to evaluate the effectiveness of this approach. Furthermore, the work also presents experimental validation of this approach using a transmission tower model to obtain its vibratory response signal and fundamental noise level. Results illustrate that this method is a powerful signal processing tool.

Keywords: Data processing, EEMD, noise level, phase space reconstruction

1. INTRODUCTION

Signal processing is a necessary part in engineering research and practical application. Seismic signal is a kind of nonlinear and non-stationary time series. Traditional Fourier analysis has some crucial restrictions, such as the data must be linear, strictly periodic or stationary. So it may give misleading results when the Fourier analysis is used to process such seismic data. The Wavelet Analysis can deal with non-stationary signals more effectively by using a set of non-stationary waveforms to represents a signal. Unfortunately we must define a mother wavelet beforehand for wavelet decomposition and face the energy leakage that generate a number of undesired spikes over the whole frequency range of time consuming continuous wavelet transform. To avoid these problems, Huang et al [1998] proposed the Empirical Mode Decomposition (EMD) method, which is a key constituent part of the Hilbert-Huang Transform. The EMD can be also used as an independent method. Through this method, any complicated data set can be decomposed into a finite of Intrinsic Mode Functions (IMFs). It is self-adaptive and highly efficient to deal with nonlinear and non-stationary processes. Hence, this method was immediately applied widely for different fields of engineering, such as structural health monitoring, vibration signal analysis, mechanical fault diagnosis, the signals instantaneous characteristic analysis and filtering techniques.

In the actual applications it is found that EMD method has some outstanding open problems, such as end effects, spline problems and stoppage criterion. One major disadvantage of EMD is mode mixing. For some kind of data, especially intermittent oscillations data, straightforward application of the original EMD may run into mode mixing. Hence, each IMF ceases to have any physical meaning by itself in the mode-mixing area. To overcome this mixing problem, a new noise-assisted data analysis method has been proposed: Ensemble Empirical Mode Decomposition (EEMD) [Wu and Huang, 2009]. The principle of the EEMD is to take advantage of the statistical characteristics of white noise and add white noise into the original signal with many trials. If the added white noise level is close to the actual noise level and the number of trials is sufficient, the EEMD will give a better decomposition result. The method can be also used as a noise attenuation method. However, there is still no clear selection standard for the added noise level and the number of trials. In this paper, the author projected the original signal into phase space by time delays and added noise levels in EEMD are determined through the singular values in the process of signal phase space reconstruction. Through numerical simulation signals and real tested signals, analysis shows that the method is effective. Thus, EEMD will be a more mature tool for non-linear and non-stationary time series analysis.

2. EMD AND EEMD METHOD

The EMD and EEMD are adaptive and highly efficient for nonlinear and non-stationary time series analysis. With the EMD method, any signal can be decomposed into a collection of (always finite) Intrinsic Mode Functions (IMFs), from high-frequency to low-frequency components. This technique depends on the original data without any priori knowledge.

2.1. Intrinsic Mode Function

According to Huang [1998], this decomposition method assumed that at any given time the signal may have many coexisting simple oscillatory modes of different frequencies, one superimposed on the other. Each component is an intrinsic mode function (IMF) that must satisfies two conditions: (1) In the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; and (2) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. The name IMF means that it represents the oscillation mode imbedded in the original data.

2.2. Sifting Process

By the above definition of an IMF, we can extract the IMF from a give data set. The sifting process is described as follows:

(1) Identify all the maxima and minima of the signal $x(t)$.

(2) Generate its upper and lower envelopes, $e_{\max}(t)$ and $e_{\min}(t)$, by the cubic spline interpolation of the extrema point developed in step (1). The upper and lower envelopes should cover all the data between them.

(3) Calculate the mean of the upper and lower envelopes $m_1(t)$.

$$m_1(t) = \frac{e_{\max}(t) + e_{\min}(t)}{2} \tag{2.1}$$

(4) Subtracting the mean from the original signal generates a new time series $h_1(t)$.

$$h_1(t) = x(t) - m_1(t) \tag{2.2}$$

(5) Check the properties of $h_1(t)$: If $h_1(t)$ satisfy the definition of an IMF, then the iteration will stop and $h_1(t)$ is an IMF1, named $c_1(t)$; otherwise , replace $x(t)$ with $h_1(t)$ and go to step (1) , repeat the procedure from steps (1) to (4) k times, until $h_{1k}(t)$ is an IMF, that is

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1k} \tag{2.3}$$

Then, it is designated as

$$c_1(t) = h_k \tag{2.4}$$

(6) Calculate the residue $r_1(t)$.

$$r_1(t) = x(t) - c_1(t) \tag{2.5}$$

(7) Replace $x(t)$ with $r_1(t)$ and repeat steps (1) to (5) n times to obtain IMF2-IMFn, named $c_2(t)$ - $c_n(t)$. When the final residual $r_n(t)$ is a monotonic function the process will be stopped.

When performing the sifting process one must pay particular attention to cubic spline interpolation near the boundaries. To guarantee that each IMF retains enough physical meaning, we must determine a measure for terminating the sifting process by limiting the size of the standard deviation (SD),

$$SD = \sum_{0}^{T} \left[\frac{\left(h_{1(k-1)}(t) - h_{1k}(t) \right)^2}{h_{1(k-1)}^2(t)} \right] \tag{2.6}$$

Typical values for SD can be set between 0.2 and 0.3. At the end of this process, the original signal $x(t)$ can be expressed as follows:

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t) \tag{2.7}$$

2.3. Mode Mixing and EEMD

It is well known that EMD is defined by empirical algorithm and has no analytical formulation. Huang [2009] point out that for some signal (especially intermittent), the EMD algorithm will lead to mode mixing. The mode mixing problem is the major obstacle to use EMD on many signals. Mode mixing is defined as a signal of a similar scale coexisting in different IMFs or a single IMF consisting of oscillations of dramatically disparate scales. When mode mixing occurs, it could not only cause serious aliasing in the time-frequency distribution, but also make an IMF cease to have physical meaning. An example of the mode mixing is illustrated in Figure 2.1, in which the decomposition of a low-frequency sinusoidal wave exists in high-frequency intermittent oscillations. As the Figure 2.1 shown, the input data is a low-frequency sinusoidal wave but at the three middle crests exist high-frequency intermittent oscillations whose amplitude is 0.1. We can find that the IMF1 occur mode mixing especially at the end of IMF1.

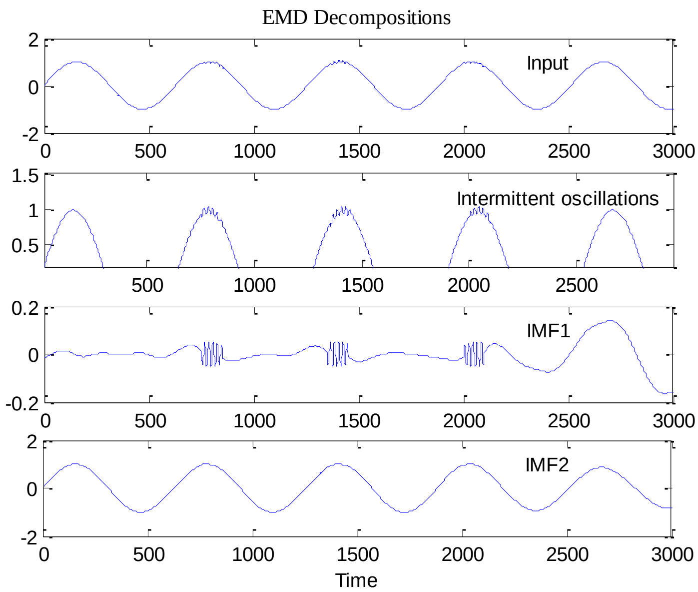

Figure 2.1. Mode Mixing Using the EMD

To overcome the mode mixing problem, a new noise-added data analysis method has been proposed: EEMD [2009]. This EEMD method utilized many statistical characteristics of white noise. The added white noise would populate the whole time scales of original signal, and it will automatically project onto proper scales. Of course, each individual added-noise will produce very noisy results. But the white noise in each trial is different, when the trials are enough each individual added-noise effects will be cancelled out by ensemble mean.

According to Wu and Huang [2009], the EEMD algorithm contains the following steps:

(1) Add a white noise series $w_i(t)$ to the targeted signal $y(t)$.

$$x_i(t) = y(t) + w_i(t) \tag{2.9}$$

(2) Decompose the noise-added signal $x_i(t)$ by EMD method into IMFs.

(3) Repeat steps (1) and (2) again and again, each time with different added white noise series. Then a new IMF assembly $C_{jk}(t)$ is obtained.

(4) Estimate the ensemble means of the final IMF of the decompositions as the final output:

$$c_j(t) = \frac{1}{N}\sum_{1}^{N} c_{jk}(t) \tag{2.10}$$

in which

$$c_{jk}(t) = c_j(t) + r_{jk}(t) \tag{2.11}$$

Where N is the ensemble number of the trials, j is the iteration number and k is the IMF scale.

Base on the EEMD algorithm, the number of ensemble and the noise amplitude are the two parameters that need to be prescribed. Appropriate decomposition parameters will get better results. If the added noise amplitudes are close to the noise level of the original signal, the EEMD will be eliminating the effect of noise. In next section, the use of phase space reconstruction approach for understanding and estimating the noise levels of polluted signal is investigated.

3. PHASE SPACE RECONSTRUCTION AND NOISE LEVEL ESTIMATE

It is well known that the dynamics of multiple degree of freedom system is given by the second order differential equation as followings [Anil K.Chopra, 2001],

$$M\ddot{u} + C\dot{u} + Ku = F \tag{3.1}$$

where, M, C and K are mass, damping and stiffness matrices, respectively, u is a vector of the responses of each degree of freedom in the systems and the dot operator above the u represents the first derivative ($\dot{u}$) and the second derivative ($\ddot{u}$). F is the input force vector. This dynamics system can also be expressed as following model,

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bq \\ y &= Gx(t) \end{aligned} \tag{3.2}$$

where, $x(t) = \begin{Bmatrix} u \\ \dot{u} \end{Bmatrix}$ is the state vector, q and y are vector representing the input and observed output, respectively. $A = \begin{bmatrix} 0 & I \\ M^{-1}K & -M^{-1}C \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ M^{-1} \end{bmatrix}$ and $G = \begin{bmatrix} I & 0 \end{bmatrix}$ are matrices which determine the relationships between the input, output, and state variables. I is an identity matrix.

3.1. Phase Space Reconstruction

Phase space is another way to characterize the dynamic system which is described above. Phase space is also defined as phase space diagram, which is a coordinate system, whose coordinates are all the variables that describe the evolution of the system. The principle of phase space reconstruction is simple: In the system, the evolution of every component is interacting with other components. Therefore, the information of the relative components is implicit in any component in the process of development. According to this principle, we can know that a point in the phase-space represents the state of the system at a given time. When dealing with the real world systems, it is difficult to collect information about all the variables of the system. Fortunately, we can also describe the system by phase space reconstruction to build a similar model directly from a time series of a single variable. That is reconstruction of a single time series in a multi dimensional phase space.
The most popular method to reconstruct the phase space is the embedding theorem of Takens [1981]. Through the concept of time delay, a phase space can be reconstructed from an observed time series. A time series $\{x_1, x_2, \ldots\ldots, x_N\}$ can be reconstructed a multi dimensional phase space in the form of delayed versions of the time series. After the reconstruction, the m-dimensional space matrix as follow,

$$X_{m\times n} = \begin{bmatrix} x_1 & x_{1+t} & \cdots & x_{1+(n-1)t} \\ x_2 & x_{2+t} & \cdots & x_{2+(n-1)t} \\ \vdots & \vdots & \cdots & \vdots \\ x_m & x_{m+t} & \cdots & x_N \end{bmatrix} \tag{3.3}$$

where, $m = N-(n-1)t$, n is the embedding dimension and t is the delay time. Typical value for the delay time can be 1, the optimal embedding dimension for reconstruction is not known a priori. In this study the embedding dimension is N/2 that is proper and stabile.

3.2. Noise Level Estimate

For a real ($m\times n$) matrix, $X_{m\times n}$ which had been reconstructed by phase space, has the singular value decomposition (SVD),

$$X = USV^T \tag{3.4}$$

where, U is an $m\times m$ orthonormal matrix, V is an $n\times n$ orthonormal matrix, and S is a $m\times n$ diagonal matrix with positive or zero elements, called the singular values. Using the decomposition above, we can identify the singular values as following,

$$\Lambda = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}, S = diag(\sigma_1, \sigma_2, \cdots, \sigma_r) \tag{3.5}$$

where, $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_m = 0$. According to the principal component analysis we find the large singular values corresponding to the characteristic components of the signal, smaller singular values corresponding to the noise components of the signal.

For a noise polluted signal $x(t)$, we can estimate its noise level through the following steps: (1) reconstruct the original signal by phase space into a real matrix, $X_{m\times n}$; (2) decompose the matrix by SVD and preserve the first K singular value and the other singular values are set to 0; (3) use the inverse process of SVD to reconstruct the matrix, $X'_{m\times n}$; (4) inverse transform the matrix $X'_{m\times n}$, and obtain a new signal $x'(t)$ according to the phase space reconstruction approach; (5) subtract the new signal $x'(t)$ from the original signal $x(t)$ generates a residual series $R(t)$, which is the noise component of the original signal; (6) calculate the residual's standard deviation, that is the noise level we desired.

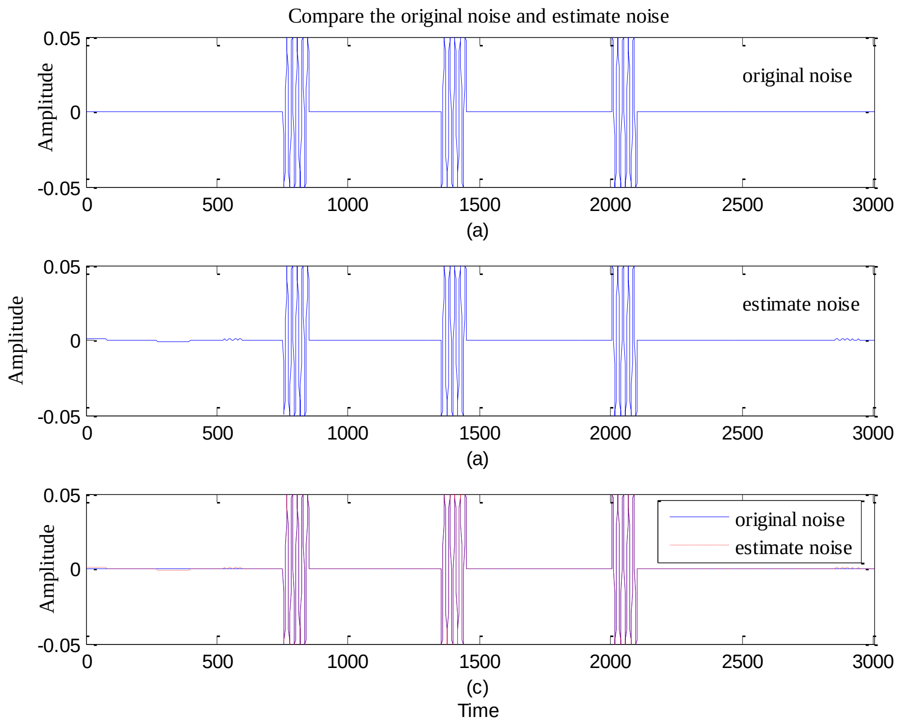

Figure 3.1. Intermittent Noise Estimate Result

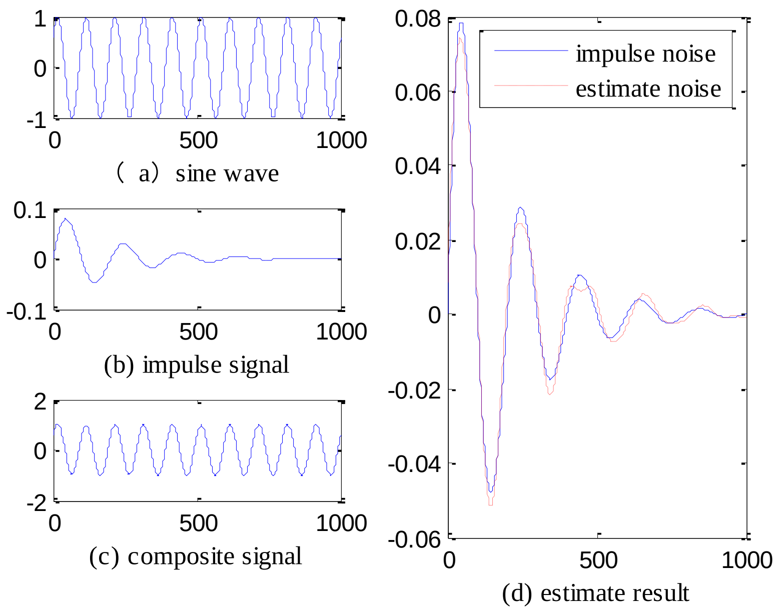

Figure 3.2. Impulse Noise Signal and Its Estimate Result

A simulated signal is illustrated in section 2, which is contaminated by the intermittent oscillations which its amplitude and noise level are 0.1 and 0.0112. Through the above noise level estimate method, we can obtain its noise series. The standard deviation of the noise series is 0.0112 which equals to the original noise level. The estimated result shown in Figure 3.1, in which Figure 3.1a is the original noise, Figure 3.1b is the estimated noise, Figure 3.1c is the comparison chart of the above two noise. Another simulated signal is shown in Figure 3.2c. It was superposed by a sine signal (Figure 3.2a) and an impulse signal (Figure 3.2b). From the simulated signal we cannot observe the impulse signal directly. The impulse noise estimate result shown in Figure 3.2d, original impulse signal shown in solid line, and the estimate impulse noise shown in dotted line.

4. NUMERICAL DATA AND EXPERIMENTAL DATA STUDY

After obtaining the encouraging results from the last section with two simple simulated signal, the application of the noise level estimate method was examined with numerical data and experimental data. We have already known that the smaller singular values corresponding to the noise components of the signal. However, it is not always easy to determine for complicated signal or real world signal. In this study, the noise components could also be determined from the characteristics of the singular values. When the singular values stabilize, the remaining components are usually noise. It is still difficult to determine when the singular values have reached a steady state. The authors define that, the steady state is the second derivative of the singular values close to 0. In the Figure 4.1 below, an example of singular values and its second derivative values is shown. From arrow showed position and up, the second derivative values of the singular value are almost zero, indicating that the singular value is already reach steady state.

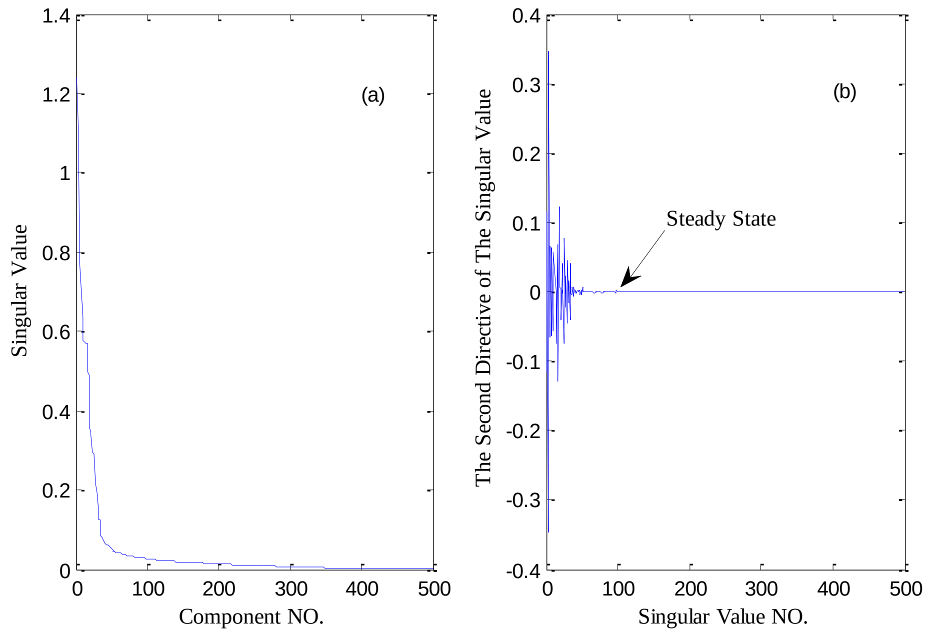

Figure 4.1. Characteristics of The Singular Values

4.1. Numerical Data Study

In order to investigate the effectiveness of this noise level estimate (NLE) method, a set of random noise with different level added into the clean complicated signal. And then the NLE method is used to estimate these noise polluted singals. The clean complicated signal is shown in Figure 4.2. The added noise levels and estimated results are listed in Table 4.1. From the results of Table 4.1, NLE method performed effective with different noise levels.

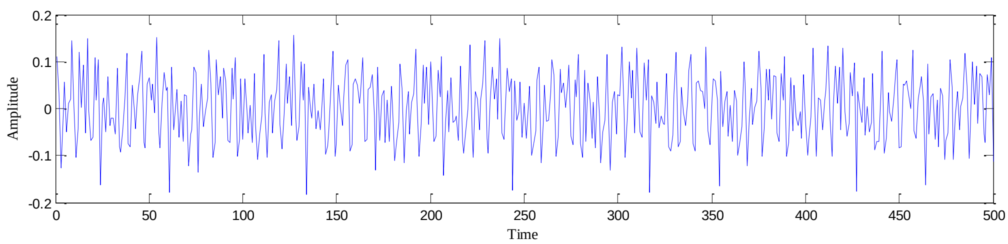

Figure 4.2. Illustration of complicated signal

Table 4.1. Added noise levels and estimated results

Noise NO.	1	2	3	4	5
Added noise levels	0.010	0.012	0.015	0.018	0.020
Estimated rusults	0.009974	0.0122	0.0151	0.0179	0.0198

4.2. Experimental Data Study

The experimental data are directly from the dynamic response of the transmission tower model, and ten fiber Brag grating (FBG) acceleration sensors were bonded to the structure. Figure 4.3 depicts the experimental model and the locations of sensors. To observe the vibration response, the structure was excited with a vibration exciter. This exciter produces a constant force over the frequency range of interest. The response of the FBG sensors were simultaneously collected through a data acquistion system which consisted of an analog input module. During all experiments the sampling frequency was kept at 500 Hz.

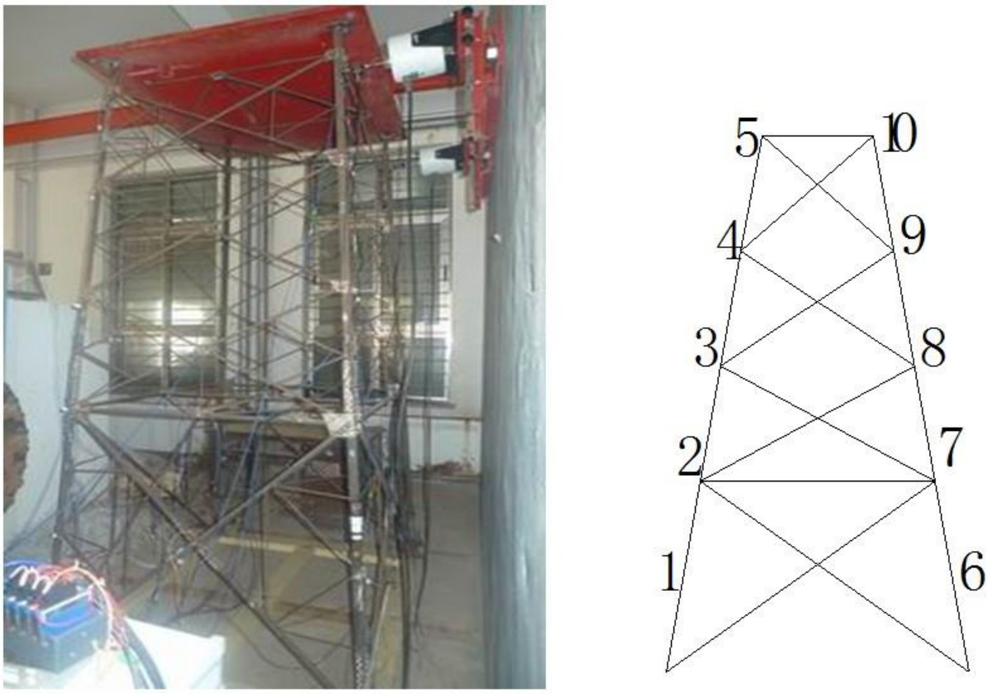

Figure 4.3. Experiment model and sensor position

Switch off the vibration exciter to keep the structure static, and switch on the data acquisition system to collect the static data (commonly base noise) of the structure. Then switch on the exciter and collect the response signal of the structure. In fact, these signals contain the noise components. After performing the vibration test on each exciting frequency, the exciting frequency was adjusted higher to simulate different vibrational state of the structure. The detailed exciting frequency and structure state are listed in Table 4.2.

Table 4.2. Different experimental case

Case NO.	Exciting frequency(HZ)	Structure state description
1	0(Switch off)	Static
2	14	Slight vibration
3	16	Vibration bigger
4	18	Violent vibration, almost resonance
5	20	Slight vibration

By analyzing the data of experiment 1, we can get the basic noise level. The experimental results were lised in Table 4.3.

Table 4.3. Basic noise levels for different sensor

Sensor NO.	1	2	3	4	5	6	7	8	9	10	Mean value
Noise level ($\times 10^{-5}$)	3.41	3.24	1.50	3.21	1.77	4.68	3.86	2.30	3.21	2.19	2.94

The estimated results for experimental data are shown in Figure 4.4. The figure illustrates the relationship between the noise level and the amplitude of the test signal. In Figure 4 a and b, the X label is the mean of the upper and the effective value of the signal, respectively, Y label is the estimated noise level. The dashed line is the cubic polynomial fitting curve of experimental results. The figure indicates the base noise levels are 4.635e-5 and 2.919e-5, that are close to the actual basic noise levels. Furthermore, the figure also indicates that the noise of the measurement signal can be divided into two parts, one is the basic noise (internal noise) of the test system, another is the external noise (ambient noise) generated by the interaction of environment and test systems. The ambient noise is related to the amplitude of test signal, and increases with the increase of the measurement signal.

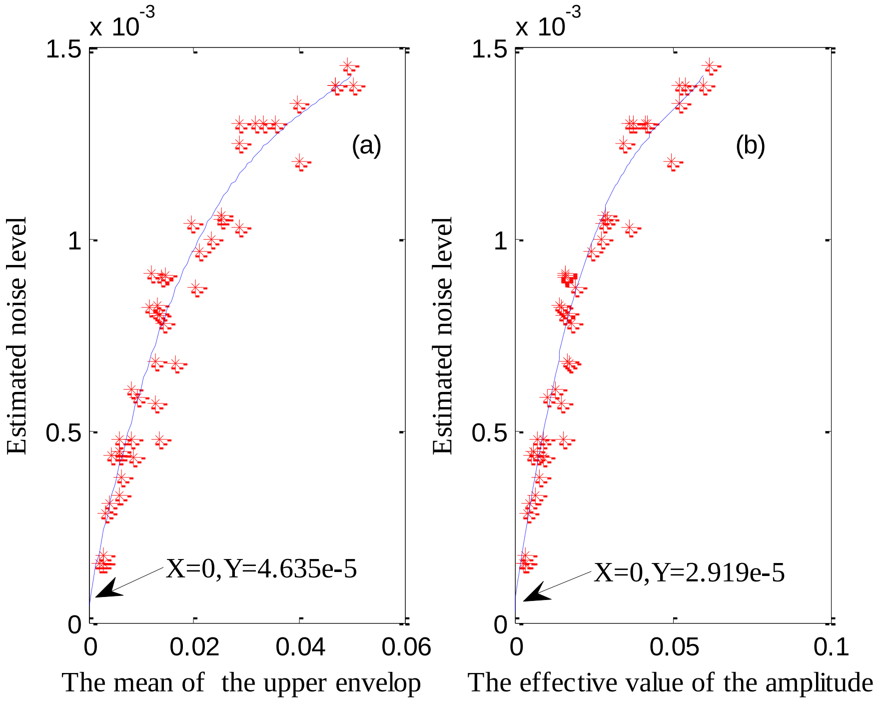

Figure 4.4. Estimated results for experimental data

5. CONCLUSION AND DISCUSSION

Mode mixing reduction is the main advantage of EEMD over traditional EMD. But the decomposition parameter, added noise level, is still not resolved. In the present work, a noise level estimate method base on phase space reconstruction is numerically and experimentally investigated. In the numerical study, simple sine wave with intermittent noise or impulse noise is simulated. Another numerical example is a set of random noise with different levels added into the clean complicated signal by artificial. In the experimental study, the authors using a transmission tower model to collect the basic noise and the response signal of the structure. Results from both the numerical and experimental investigation showed that in all cases, the proposed noise estimate method could effectively determine the noise level of the signal. According to this study, EEMD will be a more mature tool for non-linear and non-stationary time series analysis. This method also provides a new solution for random noise attenuation of seismic data.

For the measurement signal, the noise component can be divided into two parts: internal noise of the test system and ambient noise. The internal noise is determined by the characteristics of the test system and usually it has no relevance with the measurement signal. The ambient noise is generated by the interaction of environment and test systems. Its noise level is related to the amplitude of test signal, and increases with the increase of the measurement signal. However, there is no criterion to predict the specific relationship. This will be a challenging topic to be investigated in future work.

ACKNOWLEDGEMENTS
This research work has been supported by the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (Grant No. 51121005),by the Fundamental Research Funds for the Central Universities (Grant No. DUT11LK35) and State Key Laboratory of Coastal & Offshore Engineering, Dalian Univ. of Technology under the project (SL2010-9).

REFERENCES

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. and Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society Series **454A**, 903-995.

Wu, Z. and Huang, N.E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in Adaptive Data Analysis **1:1,** 1-41.

Huang, N.E. and Wu, Z. (2008). A review on Hilbert-Huang Transform: Method and its applications to geophysical studies. Reviews of Geophysics **46**,1-23.

Chopra,A. K. (2001). Dynamics of Structures: Theory and Applications to Earthquake Engineering, Prentice Hall, U.S.A.

Sivakumar,B., Jayawardena, A.W. and Fernando,T.M.K.G. (2002). River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *Journal of Hydrology* **265**, 225-245

Figueiredo, E., Todd,M.D., Farrar , C.R. and Flynn, E. (2010). Autoregressive modelling with state-space embedding vectors for damage detection under operational variability. International Journal of Engineering Science **48**, 822-834.