



## A MODIFICATION OF ENSEMBLE EMPIRICAL MODE DECOMPOSITION METHOD

Chu-Kuan Lin

*Department of Harbor and River Engineering, National Taiwan Ocean University, Taiwan, R.O.C.*

Wei-Po Huang

*Department of Harbor and River Engineering, National Taiwan Ocean University, Taiwan, R.O.C.*

Grzegorz Różyński

*Institute of Hydroengineering, Polish Acad. Sci. (IBW PAN), Poland.*

Jaw-Guei Lin

*Department of Harbor and River Engineering, National Taiwan Ocean University, Taiwan, R.O.C.,  
jglin@mail.ntou.edu.tw*

Follow this and additional works at: <https://jmstt.ntou.edu.tw/journal>



Part of the [Engineering Commons](#)

### Recommended Citation

Lin, Chu-Kuan; Huang, Wei-Po; Różyński, Grzegorz; and Lin, Jaw-Guei (2017) "A MODIFICATION OF ENSEMBLE EMPIRICAL MODE DECOMPOSITION METHOD," *Journal of Marine Science and Technology*. Vol. 25: Iss. 1, Article 11.

DOI: 10.6119/JMST-016-1017-1

Available at: <https://jmstt.ntou.edu.tw/journal/vol25/iss1/11>

This Research Article is brought to you for free and open access by Journal of Marine Science and Technology. It has been accepted for inclusion in Journal of Marine Science and Technology by an authorized editor of Journal of Marine Science and Technology.

# A MODIFICATION OF ENSEMBLE EMPIRICAL MODE DECOMPOSITION METHOD

Chu-Kuan Lin<sup>1</sup>, Wei-Po Huang<sup>1</sup>, Grzegorz Różyński<sup>2</sup>, and Jaw-Guei Lin<sup>1</sup>

Key words: EEMD, AEEMD, Noise-Assisted data analysis method.

## ABSTRACT

An adjustment about the added white noise in Ensemble Empirical Mode Decomposition (EEMD) of Wu and Huang (2008) was presented in this paper. The EEMD establishes an ensemble by adding time series of finite but not infinitesimal amplitude white noise into a time series of the signal to solve the mode-mixing problem occurred in the conventional EMD method. The adding ensembles of noise are supposed to be exhausted from all possible solutions from the sifting process. However, in the Matlab script of the theory, it was found that the added noise could not be averaged out through the whole process. The residue added noise thus causes some extra signals exist in the sifting results, even the number of trials of the ensemble was up to 5000. In the adjusted method, the added noises were randomly selected without repetition from a noise set which satisfies the normal distribution at each sampled node. With this approach, the added noises can be entirely averaged out without any residue at each node and on the entire time series. The experiments show that number of trials can be reduced to 50 sets. It not only avoid the time consuming problem but also retains the benefit of EEMD.

## I. INTRODUCTION

Empirical mode decomposition (Huang et al., 1998) is an intuitive, direct, and adaptive time-frequency data analysis method for extracting signals from data with noisy, nonlinear or non-stationary properties. It has been broadly applied to many scientific or technologic fields. The EMD method implicitly assumes that, at any given time, the data may have many coexisting simple oscillatory modes of significantly different frequencies, one (riding waves) superimposed on the other (carrier waves). Each component is defined as an intrinsic mode function (IMF) which satisfies two conditions. Firstly, the number of extrema

and the number of zero crossings must either be equal or differ at most by one in the whole data set, and secondly, the mean value of the upper envelope defined using the local maxima and the lower envelope defined using the local minima is zero everywhere. By using the sifting process, the original signal can then be decomposed into many IMFs and one residual. The siftings will be ended when a signal becomes a monotonic function or a function with only one extreme from which no more IMF can be extracted. Moreover, due to the number of extrema needed in cubic spline process, when the total number of extrema is less than 3, the sifting process should stop and apply a suitable process to extract the final IMF.

The EMD method still leaves some unresolved problems, however, like the frequency appearance of mode-mixing problem. The mode-mixing phenomenon is defined as a single IMF either consisting of signals of widely disparate scales, or a signal of a similar scale residing in different IMF components (Huang and Shen, 2005).

To overcome the problem, Wu and Huang (2008) proposed the EEMD method by adding a finite but not infinitesimal amplitude white Gaussian noise set with sufficient number of trials into the original signal to create an ensemble and performing the sifting on the ensemble. The addition of white Gaussian noise should force the noise-added signal to exhaust all possible solutions in the sifting process and making the different scale signals to collate in the proper IMF dictated by the dyadic filter banks. The added white noise can be averaged out with sufficient number of trials and only the component of the original signal survives from the averaging process. The final "true" IMFs are then taken from the mean of the sifted ensemble. The experiment with the algorithm and Matlab script provided by Wu and Huang (2008) shows the process of EEMD is much easier than EMD. For example, the number of IMFs can be predicted, no shifting criteria need to be checked, and the mode-mixing problem can be solved. With this ensemble mean, one can separate scales naturally without any a priori subjective criterion selection as in the intermittence test for the original EMD algorithm. However, it also causes some new problems that the reconstructed signal contains residual noises and different realizations of signal with noise may have different number of modes (Torres et al., 2011).

When analyzing the field wave data, the situation must be taken into account is the amount of sample data is huge. For

Paper submitted 10/21/15; revised 05/11/16; accepted 10/17/16. Author for Correspondence: Jaw-Guei Lin (jglin@mail.ntou.edu.tw).

<sup>1</sup> Department of Harbor and River Engineering, National Taiwan Ocean University, Taiwan, R.O.C.

<sup>2</sup> Institute of Hydroengineering, Polish Acad. Sci. (IBW PAN), Poland.

example, when ocean waves are sampled for every 2 hours interval, there will be 4380 records per year for just one wave gauge. The data process is always done in batch mode without any manual check, and the time consuming problem must be precaution. In order to obtain the complete information from data, the analysis should guarantee that no information is lost from and no new information is added into the original signal. Under such circumstance, by referring the procedure of EEMD, this study has carried out a series of numerical experiments to find the suitability for using the EEMD to analyze the wave data. The added noise was firstly checked independently to find the suitable number of trials which can ensure the noise be averaged out from the whole time series. A series of tests for both Gaussian-type and uniform-type white noise shows that even though the number of trials exceeds 5000 sets, the mission is still failed. In shifting process of the ensemble, by comparing with the original signal, the summation of the true IMFs was found retain some residual added noise. It is therefore the process of added noise was re-evaluated and an adjusted EEMD method (AEEMD) was proposed. The new method considers the added noises in the ensemble at each sampled point of the signal should satisfy the Gaussian/uniform distribution to ensure the zero mean of the noises of the ensemble at each node and in the entire signal. Some tests about the new method were done in the study.

In the following, the EEMD process was briefly reviewed first, and then the new method is introduced. The experiments about the added noise and EEMD, and the proposed new method are discussed in sequence. Finally, there are conclusions. Ten field wave data sets sampled from Guei-San Island water zone on the north-east coast of Taiwan was used in the experiments.

## II. THE ENSEMBLE EMPIRICAL MODE DECOMPOSITION (EEMD)

The EEMD generates an ensemble by adding different realizations of white Gaussian noise of finite but not infinitesimal amplitude into the signal, and defines the “true” IMF components as the mean of the corresponding IMFs obtained via EMD sifting processes over the ensemble of trials. The EEMD algorithm can be described as follow:

1. Given an original signal  $x_0(t)$ ;
2. Loops for the ensemble: for  $i = 1 \sim m$ , where  $m$  is the number of trials in the ensemble;
  - (1) In each trial  $i$ , generate a noise series  $w_i(t)$  with the same length of  $x_0(t)$  by using Gaussian random number function;
  - (2) Establish  $x_i(t) = x_0(t) + w_i(t)$  as a trial;
  - (3) Let  $j = 0$  and  $x_j(t) = x_i(t)$ ;
  - (4) Shifting iterations while  $x_j(t)$  is not an IMF:
    - (a) Let  $k = 0$ ;

- (b) Let  $h_j^{(k)}(t) = x_j(t)$ ;
  - (c) Identify all extrema (maxima/minima) of  $h_j^{(k)}(t)$ ;
  - (d) Connect all the local maxima (minima) by a cubic spline line to form the upper (lower) envelope, named as  $e_{\max}(t)$  ( $e_{\min}(t)$ );
  - (e) Calculate the mean  $a(t) = [e_{\max}(t) + e_{\min}(t)]/2$ ;
  - (f) Find the difference between  $h_j^{(k)}(t)$  and  $a(t)$ , i.e.,  

$$h_j^{(k+1)}(t) = h_j^{(k)}(t) - a(t)$$
;
  - (g) Check if  $h_j^{(k+1)}(t)$  is not an IMF, let  $k = k + 1$  and repeat steps 4.(b) to 4.(f);
- (5) When  $h_j^{(k+1)}(t)$  is an IMF, then let  $j = j + 1$ ,  
 $H_j(t) = h_j^{(k+1)}(t)$ , and  $x_j(t) = x_{j-1}(t) - h_{j-1}^{(k+1)}(t)$ ;
  - (6) Check if  $x_j(t)$  is neither a monotonic function nor a function with only one extrema for which no more IMF can be extracted, then repeat step (4); otherwise, let  $r_i = x_j(t)$  be the residual and the sifting end;
  - (7) Let  $n + 1$  be the final  $j$ , then  $x_i(t) = \sum_{j=1}^n H_{ij}(t) + r_i$ ,

where,  $n$  is the number of IMFs

- (8) Repeat step 3 for next  $i$ .
3. The final “true” IMFs is the average of the corresponding IMFs of all trials in the ensemble, i.e.,

$$\bar{H}_j(t) = \frac{\sum_{i=1}^m H_{ij}(t)}{m}$$

where

$$j = 1 \sim n, \text{ and } \bar{H}_{n+1}(t) = \frac{\sum_{i=1}^m r_i(t)}{m}.$$

In Wu and Huang (2008), the white noise is Gaussian and is generated using the random function of normal distribution (*randn*) and multiplied by a given ratio (0.2) of the standard deviation (*YSTD*) of the original signal with the same length, i.e.,

$$w_i(t) = 0.2 * YSTD * randn(t) \quad (1)$$

Since the random functions in computer language are all based on some specific formula, it is necessary to confirm the satisfaction of true randomness.

## III. EXPERIMENTS ON ADDED NOISE USED IN EEMD

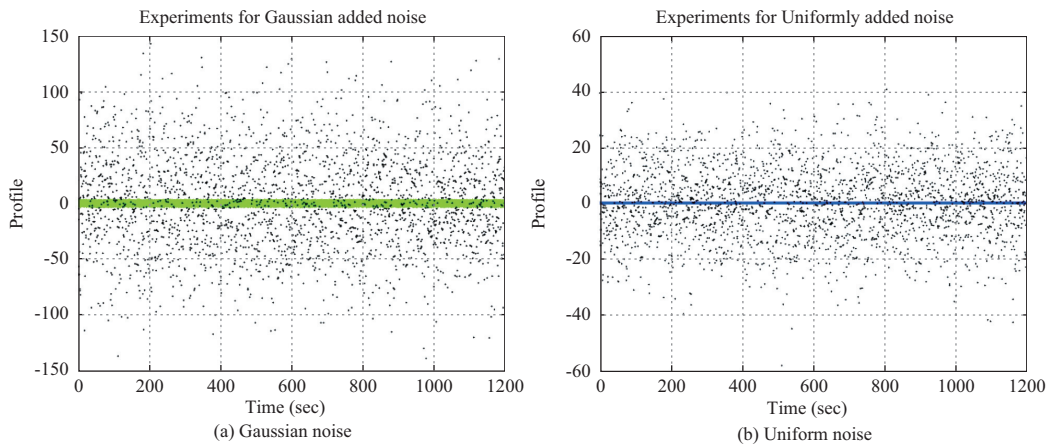
In this section, the properties of added white noise were checked first. As stated previously, the noise are generated by random function of normal distribution multiplied by a given

**Table 1. Statistics of Gaussian noise in different trials.**

Gaussian Noise No. sets	Statistics of all noise data in the whole trial			Ensemble Statistics of sum of noise at each node		
	Mean	Std.	Conf.	Mean	Std.	Conf.
50	0.0022	1.0016	0.0057	0.1084	7.1095	1.9706
100	0.0036	0.9995	0.0040	0.3645	9.9349	1.9472
500	-0.0004	1.0004	0.0018	-0.1768	22.2477	1.9501
1000	-0.0001	0.9997	0.0013	-0.0679	31.6184	1.9597
2000	0.0003	1.0002	0.0009	0.5194	43.8169	1.9204
5000	-0.0001	0.9999	0.0006	-0.2961	70.2947	1.9480
10000	0.0001	1.0002	0.0004	0.8269	99.1095	1.9425

**Table 2. Statistics of uniform noise in different trials.**

Uniform Noise No. sets	Statistics of all noise data in the whole trial			Ensemble Statistics of sum of noise at each node		
	Mean	Std.	Conf.	Mean	Std.	Conf.
50	-0.0006	0.2886	0.0016	-0.0277	2.0663	0.5728
100	0.0005	0.2882	0.0012	0.0460	2.8112	0.5510
500	-0.0002	0.2887	0.0005	-0.0891	6.4753	0.5676
1000	-0.0001	0.2886	0.0004	-0.0552	9.0996	0.5640
2000	0.0000	0.2886	0.0003	0.0511	13.0177	0.5705
5000	0.0000	0.2887	0.0002	0.2125	19.9927	0.5542
10000	-0.0001	0.2886	0.0001	-0.7607	28.9585	0.5676



**Fig. 1. Ensemble noise and its nodal sum for 2000 trials of random noise.**

ratio (0.2) of the standard deviation (*YSTD*) of the original signal with the same length (Eq. (1)). In Matlab, there are two random number functions, *rand()* for uniform distribution and *randn()* for Gaussian distribution which were all tested in this paper. The range of random number generated by *rand()* was adjusted from [0, 1] to [-0.5,0.5], and to [-2.5, 2.5] by *randn()*.

In general, the field wave data are mostly sampled with 2400 points with sampling rate of 2 Hz, therefore in our experiments the original signal is set as 2400 data points. Various numbers of trials in the ensemble were tested, and the statistics properties of the ensemble noise were discussed.

Tables 1 and 2 show the statistic properties of Gaussian/uniform distributed noise sequentially in different number of trials, varied from 50 to 10000. In column 2 to 4 of each table there are the mean value, standard deviation and confidence interval, sequentially, of all noise data in the whole trial. It was found that the added noise can almost be averaged out no matter how many trials in the ensemble. The mean value approaches

to zero, the standard deviation is nearly unity, and the confidence interval is little than 0.01 and decreases as the number of trials becomes larger. Within these two random functions in Tables 1 and 2, however, the convergence of uniform distributed noise is better than Gaussian distributed noise. Columns 5 to 7 of each table show mean value, standard deviation and confidence interval of the ensemble sum at each data point, sequentially. That is, the noises of the whole trials at each node were summed up first to get the time series of the ensemble sum, and calculated the mean value, standard deviation and confidence interval of whole time series. It was found that, in both random functions, the mean value of ensemble sum is not equal to zero which means that the ensemble noise at each data point cannot be averaged out. The standard deviations were accumulated as the number of trials became larger. With these experiments, due to the essential of random number functions in Matlab, the added noise cannot be entirely averaged out from the ensemble. Fig. 1 shows the time series of experimental results of ensemble noise for

**Table 3. Statistics of adjusted Gaussian noise in different trials.**

Gaussian Noise No. sets	Statistics of all noise data in the whole trial			Ensemble Statistics of sum of noise at each node		
	Mean	Std.	Conf.	Mean	Std.	Conf.
50	0.0000	0.9229	0.0053	0.0000	0.0000	0.0000
100	0.0000	0.9474	0.0038	0.0000	0.0000	0.0000
500	0.0000	0.9801	0.0018	0.0000	0.0000	0.0000
1000	0.0000	0.9932	0.0013	0.0000	0.0000	0.0000
2000	0.0000	0.9946	0.0009	0.0000	0.0000	0.0000
5000	0.0000	0.9975	0.0006	0.0000	0.0000	0.0000
10000	0.0000	0.9995	0.0004	0.0000	0.0000	0.0000

**Table 4. Statistics of adjusted uniform noise in different trials.**

Uniform Noise Nofsets	Statistics of all noise data in the whole trial			Ensemble Statistics of sum of noise at each node		
	Mean	Std.	Conf.	Mean	Std.	Conf.
50	0.0000	0.5888	0.0033	0.0000	0.0000	0.0000
100	0.0000	0.5888	0.0023	0.0000	0.0000	0.0000
500	0.0000	0.5888	0.0010	0.0000	0.0000	0.0000
1000	0.0000	0.5888	0.0007	0.0000	0.0000	0.0000
2000	0.0000	0.5888	0.0005	0.0000	0.0000	0.0000
5000	0.0000	0.5888	0.0003	0.0000	0.0000	0.0000
10000	0.0000	0.5888	0.0002	0.0000	0.0000	0.0000

the 2000 trials of Gaussian noise and of uniform noise. In the figure, the solid line is the collection of the ensemble noise of all trials, and the dotted points show the ensemble sum of the noise at each node along the entire time series. The random noise (solid line) are found uniformly distributed within a specific range, but the ensemble sum (dotted points) are found dispersive and non-zero.

In EEMD, the added white Gaussian noise was defined on the whole time series and was independent between each pair of trials in the ensemble. Therefore, the ensemble might find a zero sum of the noise of the entire signal, but cannot guarantee the ensemble sum of noise at each data point to be zero. The experimental results in this section show that no matter how large the number of trials of the ensemble is, the noises in time appear uniformly distributed, but the ensemble sum at each data point is not equal to zero. It is not only inconsistency with the assumption of EEMD but also some extra signal might be added into the signal.

#### IV. NEW DEFINITION OF ADDED NOISE

In Eq. (1), the amplitude of the added noise is set as 0.2 of the standard deviation (*YSTD*) of the original signal. In this section, different ratio (*YSTD*) were tested. The added noise is defined as follow:

$$w_i(t) = NSTD * YSTD * randn(t) \quad (2)$$

In order to adjust the non-zero problem in EEMD, this paper

introduces a new way to generate the added noise by forcing the ensemble noise at each data point to satisfy a specific distribution. At first, after assigning the number of trials and the *NSTD* value, a uniformly distributed data set with specified (Gaussian/uniform) distribution with the same number of trials was generated. Such data set were deployed at each data point of the time series with random permutation to establish the added noise for each trial in the ensemble. With such process, the ensemble sum of the noise at each data point and also on the entire time can be ensured to be zero which means the added noise can be averaged out in the final “true” IMF.

With different number of trials varied from 50 to 10000, Tables 3 and 4 show the statistic properties of Gaussian/uniform distributed noise sequentially for the new added noise process. In columns 2 to 4 of each table, the mean value of all noise data in the ensemble is zero for all experiments, the standard deviation still tends to unity, and confidence interval are small and uniform distributed noise has better convergence than Gaussian distributed noise. In columns 5 to 7 of each table, the mean value of the ensemble sum at each data point, and also its standard deviation and confidence interval are all equal to zero which means no residue added noise left in the time series. Still the convergence of uniform distributed noise is better than Gaussian distributed noise.

Comparing the parameters in Tables 1 to 4, the new added noise has better performance than the original added noise. The statistics of all noise data in the ensemble of both methods (columns 5 to 7 in each table) have similar tendency with almost zero mean, small confidence interval, and non-zero but

unity standard deviations, which proves that the random functions have their essential characteristics. However, when observing the results at each node (columns 5 to 7 in each table), the original added noise method shows its shortcoming that the added noise at each node cannot be averaged out from the ensemble mean. From the experiments, the new adjusted process for added noise is proved that it can solve the residual added noise problem either in Gaussian noise or in uniform noise. Even though the number of trials is only 50 sets can still perform good convergence. On the other hand, the standard deviation of original process of added noise will be piled up as the number of trials increased.

## V. THE ADJUSTED ENSEMBLE EMPIRICAL MODE DECOMPOSITION (AEEMD)

From the above discussions, a new EEMD was proposed in this paper. The added white noise signal is generated from a Gaussian random number set where the number of data equals the number of trials in the ensemble. And, at each trial, randomly picked up values without repetition for each data point of the time series are used to create a noise signal. With this procedure, we can ensure that the noise of the ensemble be averaged out at each data point and from the whole original signal. The algorithm of the new EEMD method (named as AEEMD) can be described as follow:

1. Generate a finite number data set  $z_i$ , ( $i = 1 \sim m$ ) which satisfies Gaussian distribution with zero mean and unit standard deviation, where  $m$  is the number of trials in the ensemble. The data set will be applied at each data point in the original signal as a source of random number;
2. Given an original signal  $x_0(t)$ ;
3. Loops for ensemble: for  $i = 1 \sim m$ 
  - (1) Randomly select a value without repetition from data set  $z$  at each data point of the signal, and create the noise signal  $w_i(t)$ ;
  - (2) Establish  $x_i(t) = x_0(t) + w_i(t)$  as a trial;
  - (3) Let  $j = 0$  and  $x_j(t) = x_i(t)$ ;
  - (4) Shifting iterations while  $x_j(t)$  is not an IMF:
    - (a) Let  $k = 1$ ;
    - (b) Let  $h_j^{(k)}(t) = x_j(t)$ ;
    - (c) Identify all extrema (maxima/minima) of  $h_j^{(k)}(t)$ ;
    - (d) Connect all the local maxima (minima) by a cubic spline line to form the upper (lower) envelope, named as  $e_{\max}(t)$  ( $e_{\min}(t)$ );
    - (e) Compute the mean  $a(t) = [e_{\max}(t) + e_{\min}(t)]/2$ ;
    - (f) Obtain the difference between the  $h_j^{(k)}(t)$  and  $a(t)$ , i.e.,  $h_j^{(k+1)}(t) = h_j^{(k)}(t) - a(t)$ ;
  - (4) If  $h_j^{(k+1)}(t)$  is an IMF, then let  $j = j + 1$ ;  
 $x_j(t) = x_{j-1}(t) - h_{j-1}^{(k+1)}(t)$ , and  $H_j(t) = h_{j-1}^{(k+1)}(t)$ ; oth-

erwise, let  $k = k + 1$  and repeat steps 4.(b) to 4.(f);

- (5) If  $x_j(t)$  is neither a monotonic function nor a function with only one extrema from which no more IMF can be extracted, repeat step (4); otherwise,  $r = x_j(t)$  is the residual and sifting end;

- (6) Let  $n + 1$  be the final  $j$ , then  $x_i(t) = \sum_{j=1}^n H_{ij}(t) + r_i$ ,

where,  $n$  is the number of IMFs.

- (7) Repeat step 3 for next  $i$ .

4. The final 'true' IMFs can be obtained as the average of the corresponding IMFs of all trials in the ensemble, i.e.,

$$\bar{H}_j(t) = \frac{\sum_{i=1}^m H_{ij}(t)}{m}$$

where

$$j = 1 \sim n, \text{ and } \bar{H}_{n+1}(t) = \frac{\sum_{i=1}^m r_i(t)}{m}.$$

## VI. EXPERIMENTS ON FIELD WAVE RECORDS

In order to evaluate the applicability of EEMD and AEEMD on water wave analysis, 10 wave records taken from Guei-San Island at north-east coast of Taiwan measured in 1998 were used in experiments. Each record contains 2400 data points with the sampling rate of 2 Hz. The unit of water elevation is meter, and time is second, respectively. Based on the added white Gaussian noise was suggested by Wu and Huang (2008), the following discussions focus only on Gaussian noise. In the experiments, three ensembles were established with 50, 100 and 500 trials, respectively. Followed from Wu and Huang (2008), the ratio of standard deviation of Gaussian added noise and original signal was set at 0.2, and the number of IMFs was equal to  $\text{fix}(\log(2400)) = 11$ , plus a residual. The total square error (TSE) between the sum of ensemble IMFs (including the residual) and the original signal is defined as

$$TSE = \sum_{i=1}^m \left[ x_0(t) - \sum_{j=1}^{n+1} \bar{H}_j(t) \right]^2 \quad (3)$$

### 1. Experiments on NSTD Value

In order to find the suitable NSTD value for AEEMD, an experiment was carried out first for NSTD varied from 0.05 to 0.5 with an increment of 0.05. Fig. 2 shows some experimental results, including the mean value and standard deviation of the ensemble IMFs for different NSTD values (0.05, 0.1, 0.15, 0.2, and 0.25), the results of EEMD are also included. The parameters  $c_1, c_2, \dots, c_{10}$  represent the IMF components from sifting process. Observing the ensembles IMFs of different NSTD (not shown here), and the ensemble mean and standard deviation of IMFs in Figs. 2(a) and 2(b) respectively,

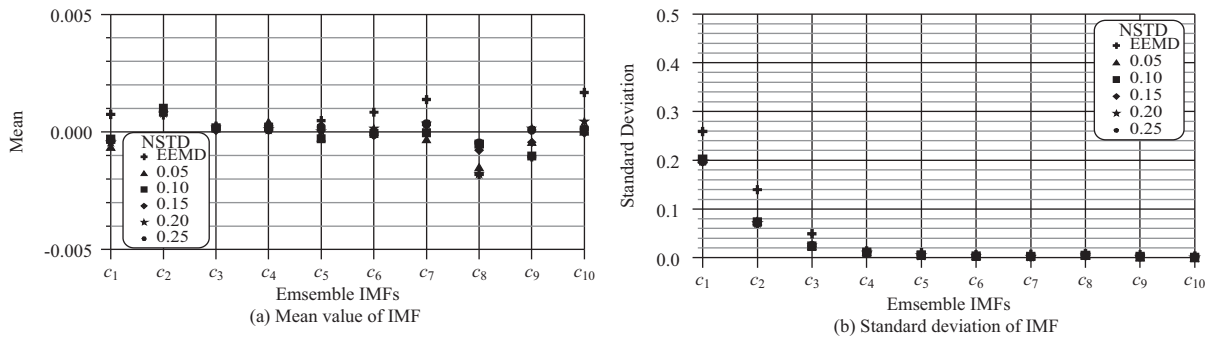


Fig. 2. Comparisons of ensemble IMFs for different *NSTD* values.

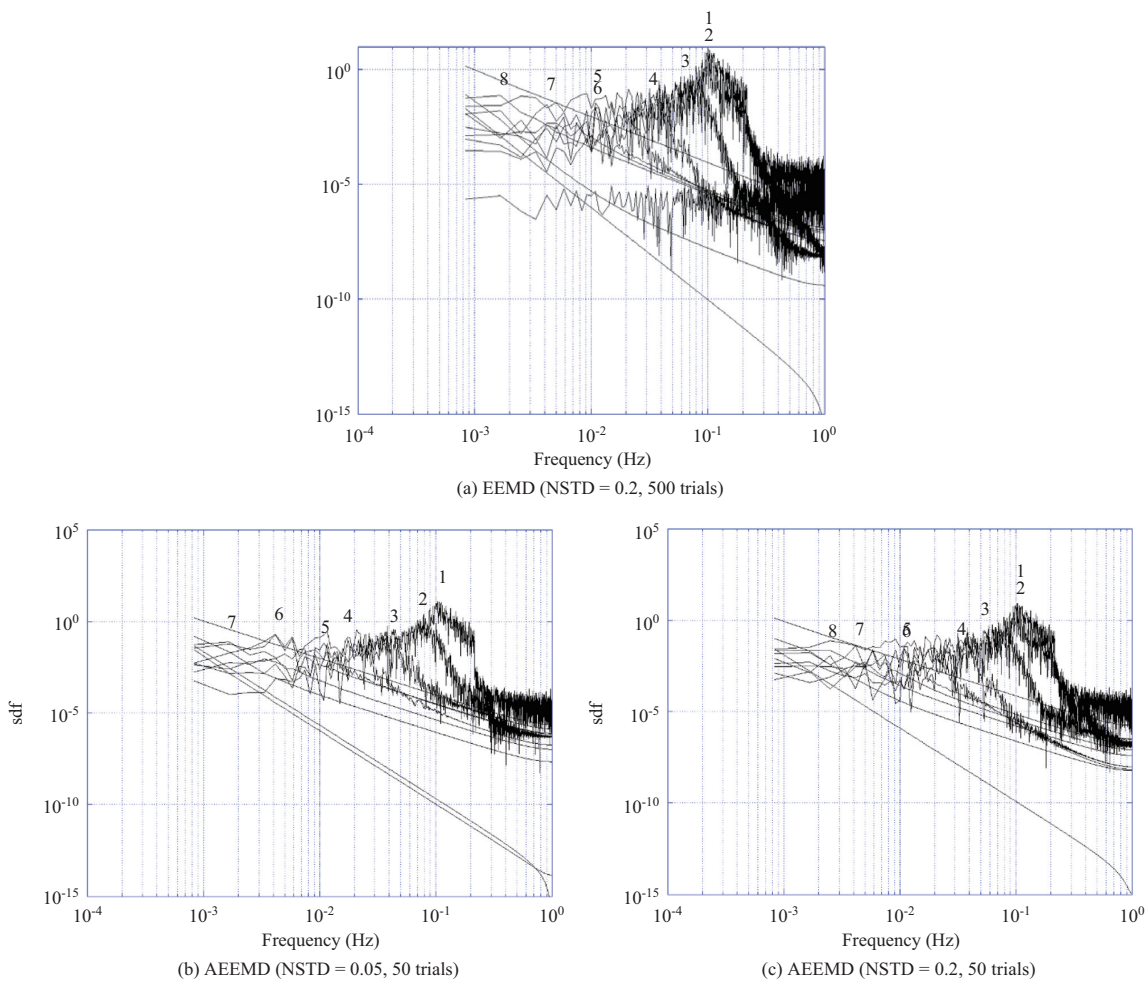


Fig. 3 Samples of the spectra of ensemble IMFs.

no obvious difference was found in the results of different *NSTD* in AEEMD. However, such differences were identified for the results of EEMD, especially within the  $c_1$ ,  $c_2$  and  $c_3$  which belong to higher frequency components.

Examining the results of the spectra of ‘true’ IMFs of either EEMD or AEEMD with different *NSTD* values show that the mode-mixing problem can be better solved as *NSTD* is smaller than 0.1 by AEEMD. Taken as an example, Fig. 3 shows the

spectra of IMFs of one wave record decomposed by EEMD with *NSTD* = 0.2 for 500 trials, and AEEMD with *NSTD* = 0.05 and 0.2 for 50 trials. The numbers in the figures are the order of ensemble IMFs. For *NSTD* = 0.2 in Fig. 3(a) for EEMD and in Fig. 3(c) for AEEMD, the spectra of the first two IMFs are too close to be identified, while in Fig. 3(b) is not (*NSTD* = 0.05 for AEEMD). The results show that the AEEMD with *NSTD* = 0.05 has better mode decomposition

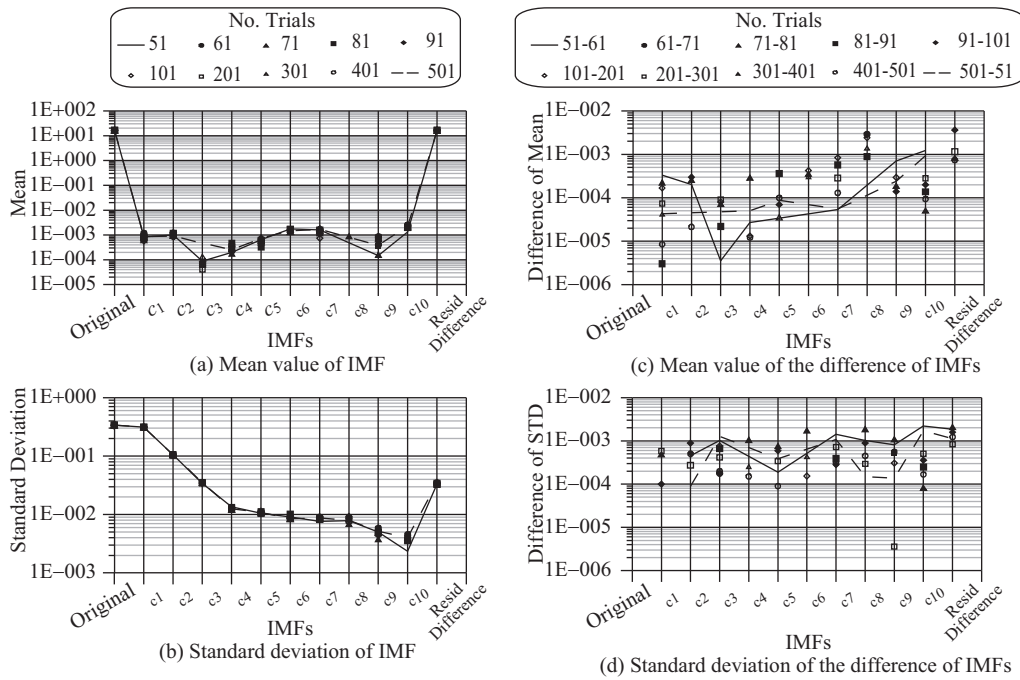


Fig. 4. Comparisons of ensemble IMFs between different trials.

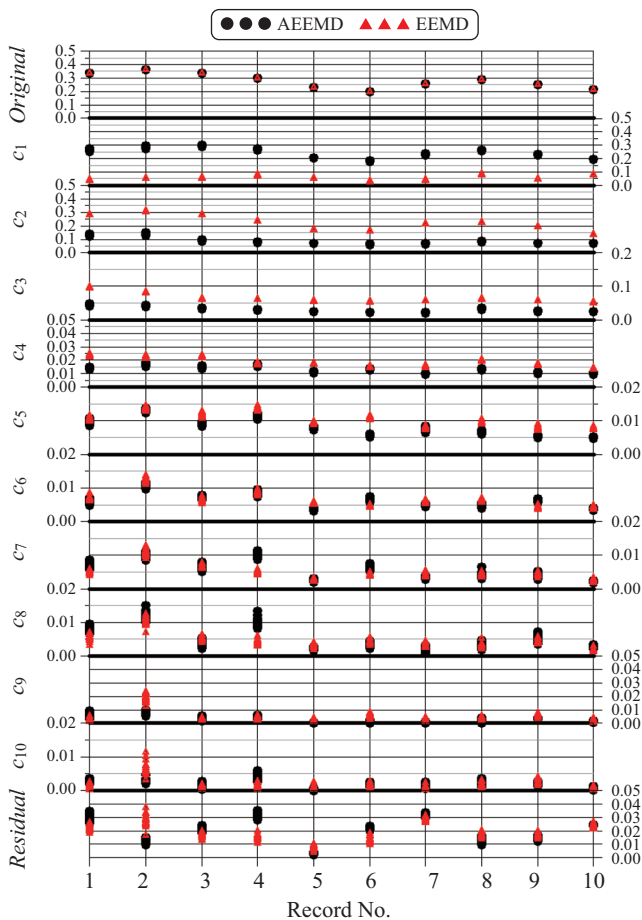


Fig. 5. Comparisons of Standard deviations of all IMFs for all records between EEMD and AEEMD.

results than that of EEMD and AEEMD with  $NSTD = 0.2$  in Fig. 3. The  $NSTD$  was suggested to be set smaller than 0.1.

### 2. Experiments on the Number of Trials

In order to find the suitable number of trials of the ensemble for the application of AEEMD for water wave analysis, a series of experiments on establishing the ensemble with different number of trials was carried out. By altering the number of trials in the ensemble, varied from 51 to 501, Fig. 4 shows the comparisons of IMFs of different number of trials on one wave record, including the mean and standard deviation of IMFs, and the mean and standard deviations of the difference of ensemble IMFs between different trials. The standard deviation of original signal is 0.338746 m. Observing the results in Fig. 4 and comparing them with the standard deviation of the original signal, we may find that only the first two IMFs have larger variations which should be affected by the added noise, but the differences of IMFs between each pair of trials are acceptable (within  $\pm 0.002$ ), and one can proceed a similar experiments for the number of trials for a specific signal to determine a suitable number of trials in the ensemble. For the current wave records, 50 trials can offer a satisfactory solution.

### 3. Comparisons Between EEMD and AEEMD

Due to the random character of the observed water waves and the white noise, Fig. 5 show the comparisons of the standard deviations of ten repetitions of ensemble IMFs for all ten wave records between EEMD and AEEMD. The differences between EEMD and AEEMD are small for  $c_4$  to  $c_{10}$ . The large differences at  $c_1$  to  $c_3$  are caused by the residual white noise in EEMD. The added white noise might affect the decomposition



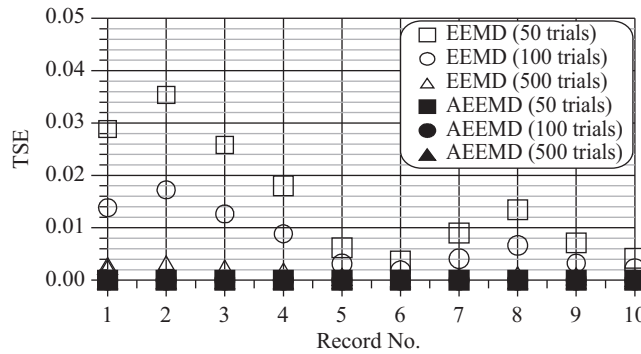


Fig. 6. Total square error (TSE) of the ensembles between 50, 100 and 500 trials ( $NSTD = 0.2$ ).

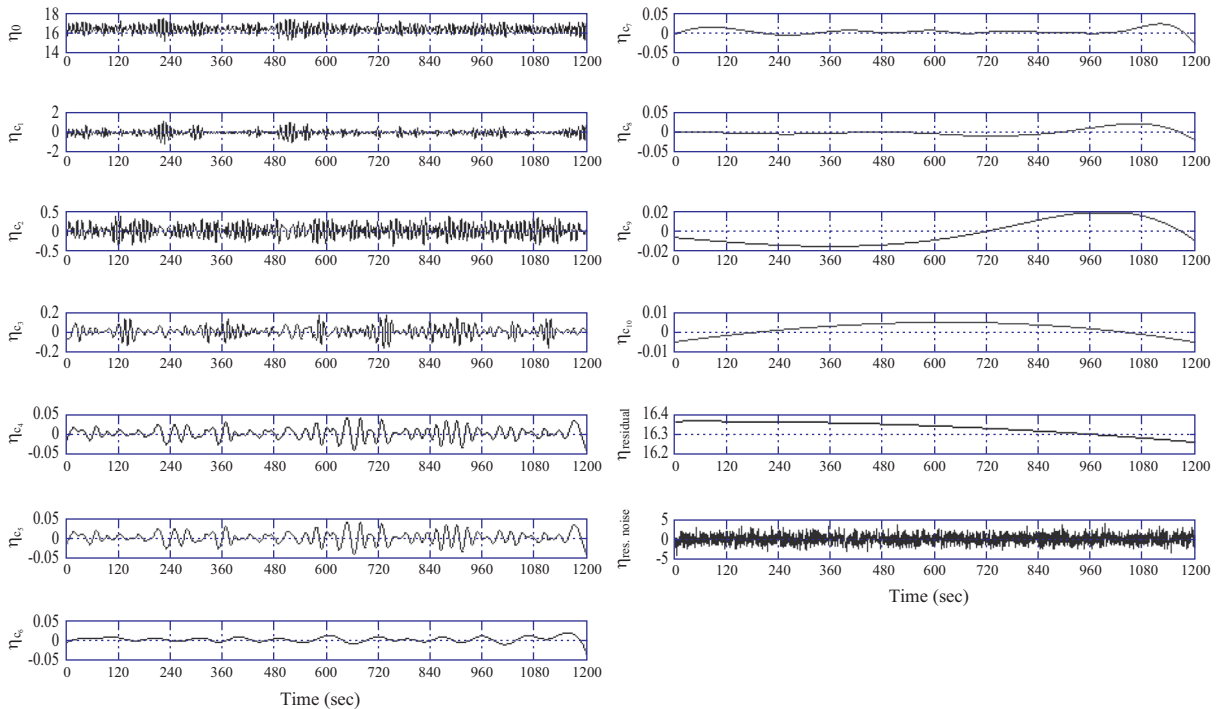


Fig. 7. Samples of the ensemble IMFs with 500 trials by EEMD.

of higher frequency portion of the wave data. The difference between the repetitions analyzed by EEMD is larger than that of AEEMD. It is because in AEEMD, the added white noise are randomly picked up from a limit set of Gaussian number generator instead of fully random as in EEMD.

Fig. 6 show the total square error (TSE) of the ensemble by EEMD and AEEMD with 50, 100 or 500 trials. The TSE of EEMD decreased as number of trials increased; on the contrary, the TSE of AEEMD is always zero.

Taken as an example, Fig. 7 shows the ensemble results of IMF of one wave record with 500 trials by EEMD method, similar results can also be found for the results with 1000 trials. In this figure, the related IMFs are plotted from left to right, and from top to bottom where the first plot is the original signal, the 2<sup>nd</sup> to 11<sup>th</sup> plots are ensemble IMF components, the

12<sup>th</sup> plot at the right column is the residual, and the final plot on the right column is the residual noise obtained from the subtraction all IMFs from original signal. The related spectra of each ensemble IMF are shown in Fig. 3(a). The last plots in Fig. 7 shows that since the added noise cannot be averaged out, there are some extra signal elements that were added into the original wave signal during the analysis, and its spectrum appears to be a white noise.

Fig. 8 shows the AEEMD ensemble results of the same wave record as in Fig. 7 with 50 trials, the arrangement of plots is the same as in Fig. 7, and similar results can also be found for the results with 100 and 500 trials. The last plot shows that the difference between AEEMD process and the original signal is only a small value lying within  $\pm 2 \times 10^{-14}$  which might be caused by the truncation/cutoff error during the sifting and can

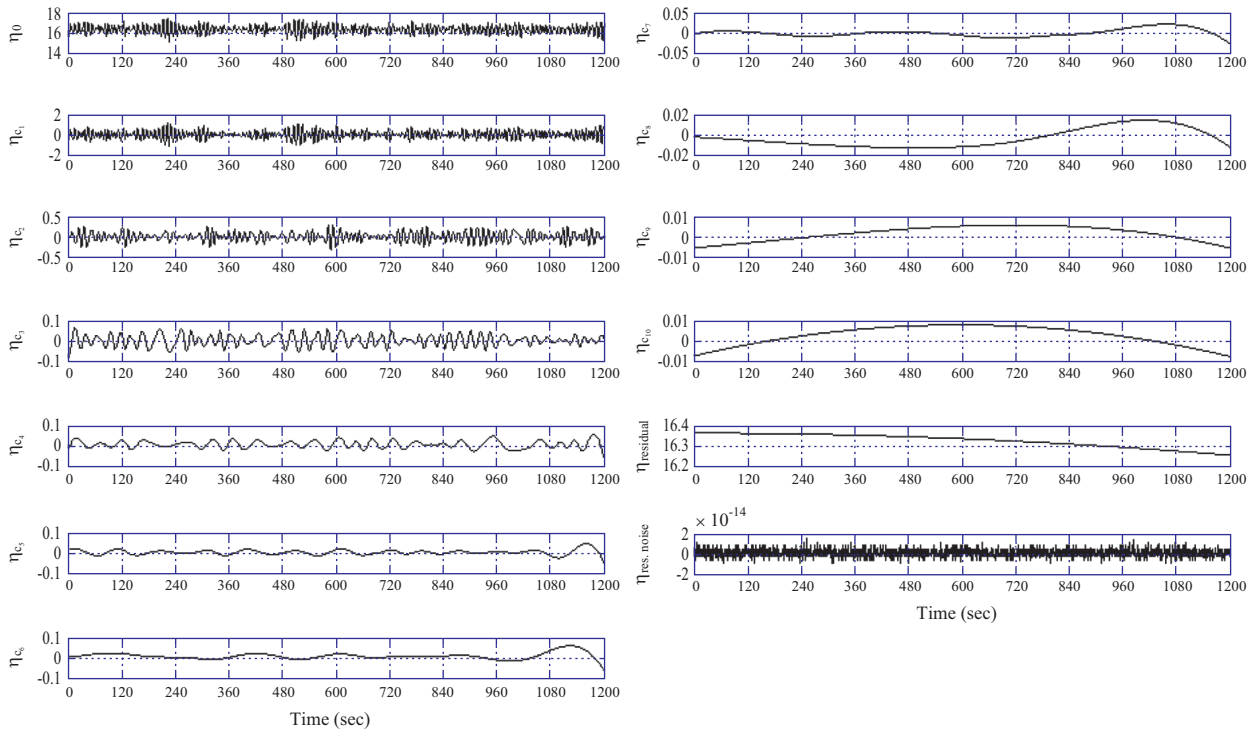


Fig. 8. Samples of the ensemble IMFs with 50 trials by AEEMD.

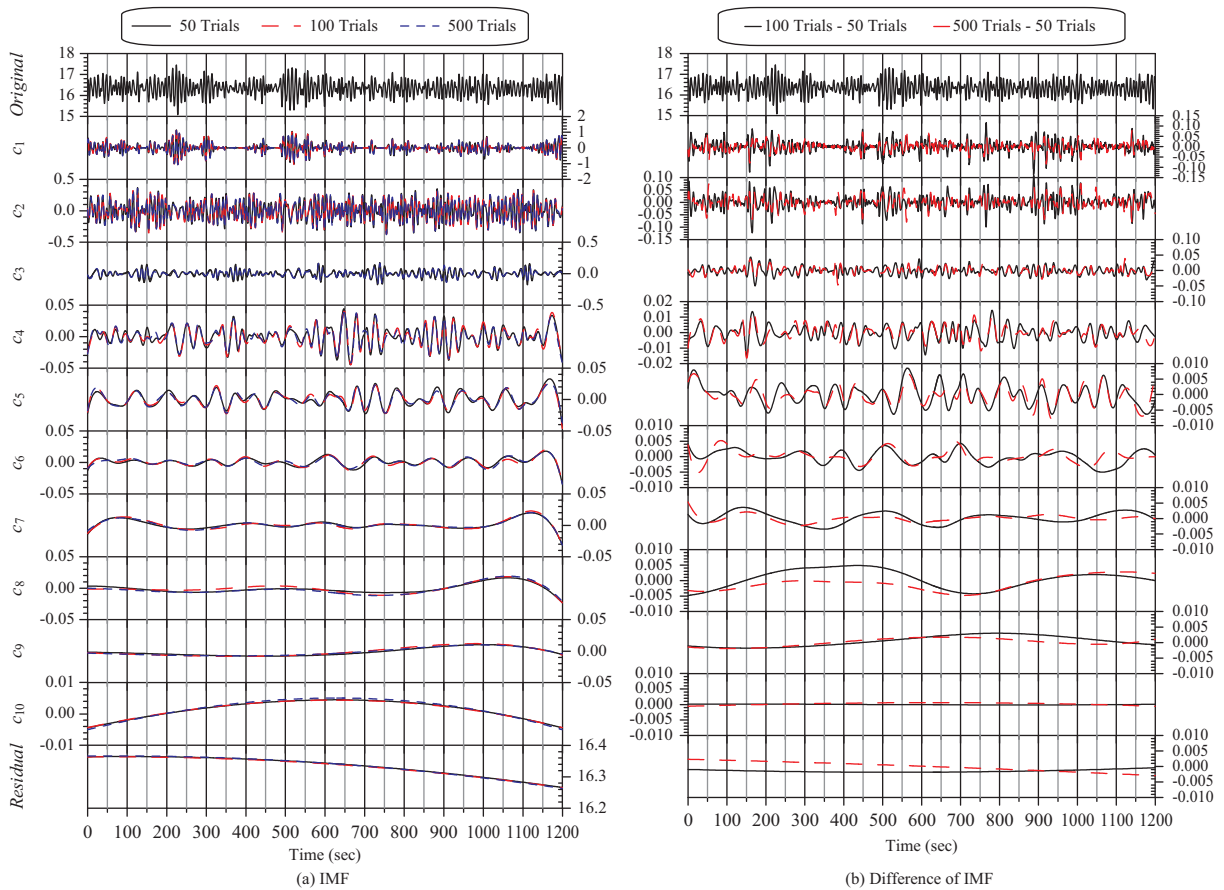


Fig. 9. Ensemble IMFs for different trials by EEMD.

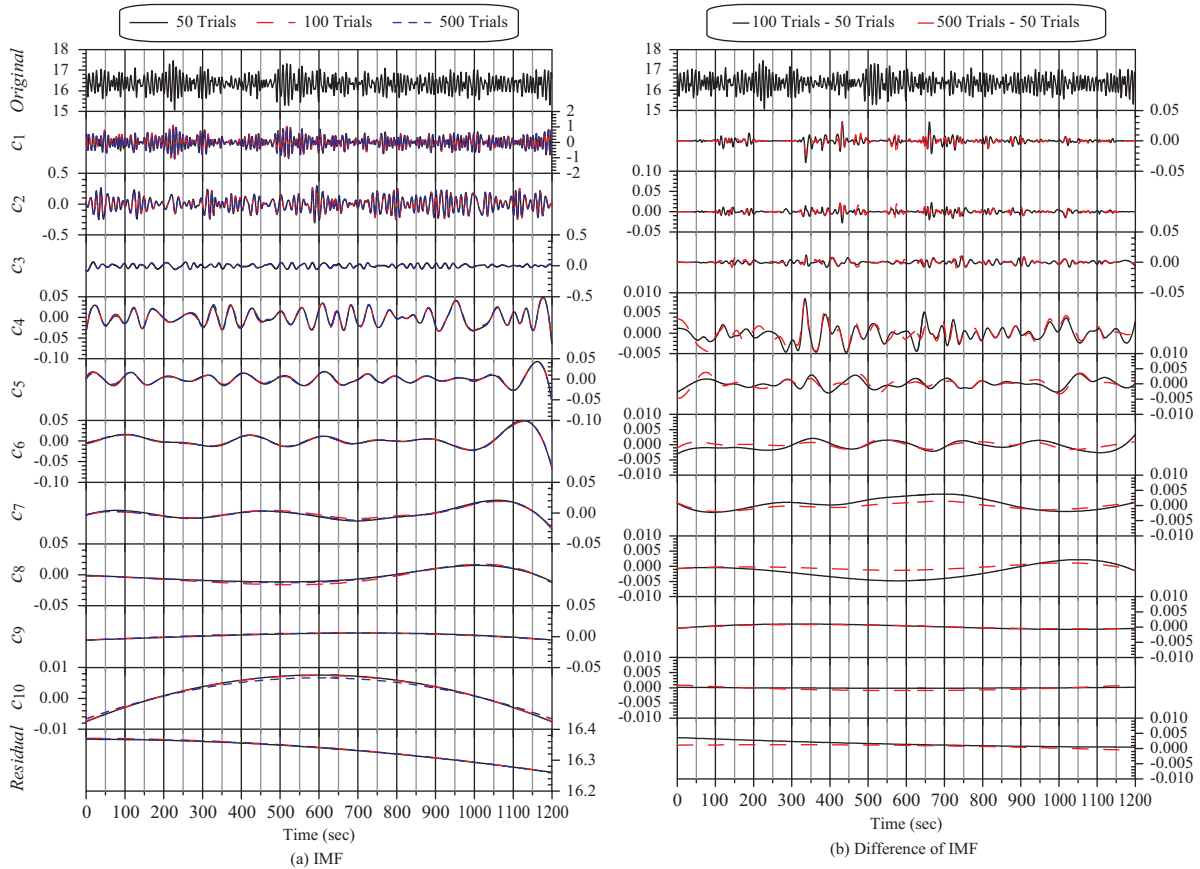


Fig. 10. Samples of the ensemble IMFs for different trials by AEEMD.

be neglected. This shows that the AEEMD has better results on keeping the original signal from residual noise. The related spectra of each ensemble IMF are shown in Fig. 3(b).

By using the same wave record as in Fig. 7, and setup of three ensembles with 50, 100 and 500 trials, Fig. 9 collects the sifting results from EEMD. Fig. 9(a) is the ensemble IMFs, only small difference of IMFs between them can be found. Fig. 9(b), however, shows the differences of IMFs between different ensembles with different number of trials are small which means, in EEMD, more trials in the ensemble can get more stable ensemble results. Fig. 10 collects the results from AEEMD. Only small difference of IMFs between them was found which means the results from 50 trials should be acceptable. In AEEMD, the increase of the number of trials in the ensemble is less advantageous as to the accuracy of the sifting results.

All these comparisons show that the AEEMD has the same performance on sifting the signals as EEMD. However, since EEMD employs a totally random noise, theoretically they should be averaged out with enough number of trials in the ensemble, but the random number function provided by any computer programming language doesn't guarantee this requirement. A controlled added noise was then employed in the new method and proved to be more reasonable without introducing any new component into the signal. The AEEMD was proved to have higher accuracy and more efficiency than EEMD.

## VII. DISCUSSIONS

1. In this study, we evaluated the convergence of EEMD by Wu and Huang (2008) and found that the added noise cannot be averaged out from the ensemble mean. This might be caused by the random number function in Matlab failing to guarantee the noise satisfy the Gaussian distribution even with the number of trials up to 10000, and the EEMD assigns the added noise with Gaussian distribution along the time axis instead of doing so at each node and thus causes some residue noise be introduced into the process. The residual added noise exists at each node and along the entire time series.
2. The adjusted method (AEEMD) firstly created a finite number data set which satisfies Gaussian distribution with zero mean and unit standard deviation and was applied at each node as a pool of random number, the added noise of each trial was randomly chosen from the pool without repetition at each node of the original signal independently. Such process ensures added noise in the ensemble can be averaged out at each node on time axis, and also be averaged out from the ensemble mean. From the experiments, the benefit of the original EEMD can still be retained.
3. Due to the added noise in the ensemble can be averaged from each node and from the entire time series, the number

of trials in AEEMD can be limited to 50 sets which avoid the time consuming problem when analyzing time series with large number of elements, like field wave observations.

4. For decomposition of water wave signals, the amplitude of the added noise (*NSTD*) can be set less than 0.1 of the standard deviation of the original signal. For other types of time series, however, an initial experiment is recommended.
5. In order to do the ensemble average with same number of IMFs, Wu and Huang (2008) suggested the number of IMFs equals to  $fix(\log(no.data))$ , plus a residual. As the number of data becomes large, the number of IMFs increased even the sampled data were taken from the same time series with longer duration or with higher sampling rate to the same duration. Such might distort the signal decompositions, and confused the sifting results.

### VIII. CONCLUSIONS

1. The AEEMD has better convergence than the original EEMD, the added noise of the ensemble can be averaged out at each node and from the whole time series which ensures no extra signal was added into the original signal. With a limited number of trials but similar approach as EEMD, the AEEMD has better performance on solving the mode-mixing problem in water wave analysis. Apart from that, EEMD and AEEMD have the same performance on decomposition of the original signal.
2. Using 50 sets of noise-added signal in the ensemble and choosing *NSTD* smaller than 0.1 in AEEMD process can solve the

mode-mixing problem and offer good performance to avoid time consuming problem. Such process is suitable for the analyses of large amount of field data.

3. The number of IMF set equal to  $fix(\log(no.data))$  plus a residual might distort the sifting results and increase unnecessary IMFs especially the decompositions of the trend. Further studies on choosing a suitable number of IMFs are needed.
4. End effect still plays an important role on the sifting and suitable end conditions in cubic spline calculations should be evaluated.

### REFERENCES

- Huang, N. E., A. Shen, S.-R. Long, M.-C. Wu, E.-H. Shih, Q. Zheng, C.-C. Tung and H.-H. Liu (1998). The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis. *Proc. Roy. Soc. London*, 454A, 903-995.
- Huang, N. E. and S. S.-P. Shen (ed.) (2005). *Hilbert-Huang Transform and Its Applications*. World Scientific, Singapore, 360.
- Huang, N. E. and Z. Wu (2008). A review on Hilbert-Huang transform: the method and its applications on geophysical studies. *Rev. Geophys.* 46, RG2006, 23.
- Torres, M. E., M. A. Colominas, G. Schlotthauer and P. Flandrin (2011). A complete Ensemble Empirical Mode decomposition with adaptive noise. *IEEE Int. Conf. on Acoust., Speech and Signal Proc.*, Prague (CZ). ICASSP-11, 4144-4147.
- Wu, Z. and N. E. Huang (2008). Ensemble Empirical Mode Decomposition: A Noise Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, World Scientific Publishing Company 1(1), 1-41.