

An Ensemble Empirical Mode Decomposition Approach to Wear Particle Detection in Lubricating Oil Subject to Particle Overlap

Zhendang Li

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the M.Sc. degree in Systems Science

Systems Science
University of Ottawa

©Zhendan Li, Ottawa, Canada, 2011

Abstract

With the development of mechanical fault diagnosis technology, complex mechanical systems do not need to be shut down periodically for the maintenance. The working condition of the mechanical systems can be monitored by analyzing the wear metal particles in the systems' lubricating oil. By counting the number of particles in the oil and comparing the number with a preset threshold, the health information of the oil-wetted components can be acquired without shutting down the whole system. Some oil sensors are installed in the lubricating oil system. When the metal particles flow through the sensor, the sensor will generate signals corresponding to the size and type of the particles.

There are some difficulties for decomposing the output signals of the oil sensor. First, the output signal is non-stationary. The traditional Fourier transform does not fit for non-stationary signals. Second, when multiple particles flow through the sensor in a short interval, the particle signals will superimpose with each other and form a long signal with a random shape. This makes the method which decomposes a signal by matching each component's shape unpractical. Also, the typical signal analysis methods will transform the signal from time domain to frequency domain. However, frequency information is not of much help for this problem.

The goal of this thesis is to find a working method to decompose the oil sensor signal, into a form that the number of particles can be counted. After analyzing the character of the oil signal, a method named Empirical Mode Decomposition (EMD) was chosen to decompose the oil signal in time domain. However, during the experiment, EMD was found to be sensitive to the noise change in the original signal. To resolve this problem, a method named Ensemble Empirical Mode Decomposition (EEMD) was introduced. EEMD improved EMD method by introducing the Noise-Assisted Data Analysis (NADA) method during the EMD processing. The optimized preset parameter values for EEMD were also discussed. Although EEMD alleviated some of EMD shortcomings, it also brought some side effects into the decomposition result because of the NADA method. The side effects were analyzed, and a post processing method was implemented to get a better result.

Acknowledgements

I would like to thank my thesis supervisor, Professor Tet Hin Yeap at the School of Information Technology and Engineering (SITE), for his help in my study.

I also wish to thank my thesis co-supervisor, Professor Ming Liang at the Department of Mechanical Engineering, for the suggestions for the thesis.

Finally I would like to thank Iman Soltani Bozchalooi, for providing the test signal data for this thesis.

Table of Contents

Abstract.....	i
Acknowledgements.....	iii
List of Figures	vi
Acronym List.....	vi
Chapter 1 Introduction	1
1.1 Motivation.....	1
1.2 Thesis Objectives.....	3
1.3 Thesis Contributions	4
1.4 Thesis Organization	5
Chapter 2 Background.....	6
2.1 Oil Particle Sensors and Implementation	6
2.2 Existing Method.....	8
2.3 Conventional Signal Analytical Methods	10
2.3.1 Analysis method for stationary signal	10
2.3.2 Some analysis methods for nonstationary signal	11
2.4 The EMD Method for Particle Detection.....	14
Chapter 3 Empirical Mode Decomposition	16
3.1 Introduction	16
3.1.1 The instantaneous frequency.....	17
3.1.2 The Intrinsic Mode Function (IMF)	18
3.1.3 Empirical Mode Decomposition (EMD).....	20
3.1.4 Completeness and orthogonality	26
3.2 Some limitations of EMD	28
3.2.1 Lack of theoretical background	28
3.2.2 Stopping Criteria for sifting	29
3.2.3 End effect.....	31
3.2.4 Mode Mixing.....	33
Chapter 4 Ensemble Empirical Mode Decomposition.....	35
4.1 Introduction	35
4.1.1 Noise-Assisted Data Analysis	36

4.1.2	The mode mixing problem.....	37
4.2	Ensemble Empirical Mode Decomposition.....	40
4.2.1	The number of ensemble for EEMD.....	43
4.2.2	The selection of the amplitude of added noise.....	44
Chapter 5	The implementation of the EMD and EEMD methods in particle detection.....	45
5.1	Introduction.....	45
5.2	The experiments.....	45
5.2.1	EMD and the noise problem.....	45
5.2.2	EEMD method.....	53
5.2.3	Post processing.....	58
5.2.4	Other experiments.....	60
Chapter 6	Conclusions and future work.....	63
6.1	Conclusions.....	63
6.2	Future work.....	65
References	67

List of Figures

Figure 2-1 The structure of a MetalSCAN oil debris sensor.....6

Figure 2-2 Typical particle signals.....7

Figure 2-3 A typical MetalSCAN System8

Figure 3-1 Envelopes and Mean21

Figure 3-2 Rilling's mirror method33

Figure 4-1 An example of signal intermittency.38

Figure 4-2 Decomposed IMFs of the signal c39

Figure 4-3 EEMD decomposition with ensemble number of 5042

Figure 5-1 Constructed signal and its components47

Figure 5-2 EMD Decomposition result of the signal in Figure 5-148

Figure 5-3 EMD decomposition result for the signal with intermittency.....51

Figure 5-4 EEMD result of the signal in Figure 5-355

Figure 5-5 Post processing of the EEMD output in Figure 5-459

Acronym List

A/D Converter	Analog to Digital Converter
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
HHT	Hilbert-Huang transform
IMF	Intrinsic Mode Function
NADA	Noise-Assisted Data Analysis
STFT	Short Time Fourier Transform

Chapter 1 Introduction

1.1 Motivation

Many mechanical systems require constantly maintenance during their life time. The machines and structures in the system sometimes fail during their operation. There are many failure modes, include wear, corrosion, fatigue etc. Most of them can be detected at the early stage. The traditional way of maintaining those systems is periodically shutting down and checking the system. However, for some systems, such as some complex equipment, the shutting down cost will be very high and disassembling the equipment for diagnostic may be impractical. For these situations, it is better to use a monitoring system to track the systems' conditions[1], only when some warning signals are detected, the systems will be shut down.

In a mechanical system, usually some heavy duty components like gear systems are submerged into lubricating oil. If those components in the gear system are worn, there will be some metallic debris from the components suspended in the lubricating oil. For some systems such as aircraft, a monitoring device can be designed to detect the metallic particles in lubricating oil. If the number of debris is higher than a setting value, a conclusion can be made that the components in the gear-box are worn and need maintenance. For those monitoring device, usually oil debris sensors are used to provide the input signals. These oil debris

sensors are connected with the lubrication system to detect the metallic debris when lubricating oil flows through them.

A typical oil debris sensor is a through-flow device. When the lubricating oil which contains metallic particles flows through the sensor, the sensor will detect the particles and generate a signal. By analyzing the output signals from the oil debris sensors, the monitoring system can count the number of particles and determine the status of the mechanical system.

For the output signal from the oil debris sensor, there are two kinds of scenarios. If each time there is only one particle pass through the sensor, the sensor will generate a sinusoidal shape output. This situation usually happened in some systems which have a strict requirement about the number of worn particles in the lubricating oil, for example, the airplane engine. The engine system will need maintenance when a few particles are detected. For this kind of scenario, particle signals can be detected and counted directly. However, because of the working environment of these systems, noise usually mixed with the useful signals in the sensor's output. Sometimes the useful signals are totally covered by the noise. How to remove the noise and extract the useful signals will be the first problem of this scenario. Effective algorithms are necessary to separate the useful signals from the noise.

In another scenario, some other mechanical systems can work well even more worn particles are detected in the lubrication system, which means that sometimes there will be more than one particle pass through the sensor in a very short time interval. The output signal will be the mixture of those particle signals superimposed with each other. There is relatively small noise in the output signal. In this situation, a decomposition algorithm will be needed to decompose the mixed signal into separate particle signals before the monitoring system counts the number of particles. Because those particle signals are similar with each other, the algorithm should be sensitive and adaptive to the output signal of the oil sensor.

1.2 Thesis Objectives

This thesis focuses on the second scenario. It tries to find a working algorithm to decompose the particle signal from the original signal when more than one particle passes through the oil debris sensor in a very short time interval.

First, the character of the particle signals from the oil sensor should be analyzed. Based on the type of the particle signals, an appropriate method should be chosen for the decomposition.

Second, the chosen method will be implemented to decompose the particle. The result will be investigated. If there are any problems during the implementation of the chosen method, after analyzing the problems, a solution will be proposed and

tested. There will be a detailed discussion about the solution and the experiments.

Finally, a conclusion will be made for the decomposition method, and the future works will be listed.

1.3 Thesis Contributions

The goal of this thesis is to find a working method to decompose the oil sensor signal, into a form that the number of particles can be counted. After analyzing the character of the oil signal, a method named Empirical Mode Decomposition (EMD) was chosen to decompose the oil signal in time domain. However, during the experiment, EMD was found to be sensitive to the noise change in the original signal. To resolve this problem, a method named Ensemble Empirical Mode Decomposition (EEMD) was introduced. EEMD improved EMD method by introducing the Noise-Assisted Data Analysis (NADA) method during the EMD processing. The optimized preset parameter values for EEMD were also discussed. Although EEMD alleviated some of EMD shortcomings, it also brought some side effects into the decomposition result because of the NADA method. The side effects were analyzed, and a post processing method was implemented to get a better result.

1.4 Thesis Organization

In chapter 2 the background information of oil particle detection will be introduced. There is a brief review for the typical signal analysis methods, and then the Empirical Mode Decomposition (EMD) is introduced. Chapter 3 will describe the principle of the EMD method. Some limitations of the EMD will also be discussed. Chapter 4 will introduce the Ensemble Empirical Mode Decomposition (EEMD), an improvement for the EMD method. Chapter 5 will discuss the detail of the implementation of EMD and EEMD to the particle signal, the problems encountered and the solution. Chapter 6 will conclude the research and list the future works.

Chapter 2 Background

2.1 Oil Particle Sensors and Implementation

In this project, a MetalSCAN oil debris sensor from GasTOPS Company will be used to detect the metallic particles in the lubricating oil. The sensor is based on an inductive measurement technique. When a metallic particle above a minimum size threshold passes through the sensor, the sensor will generate a signal according to the particle's size and type (ferromagnetic and non-ferromagnetic).

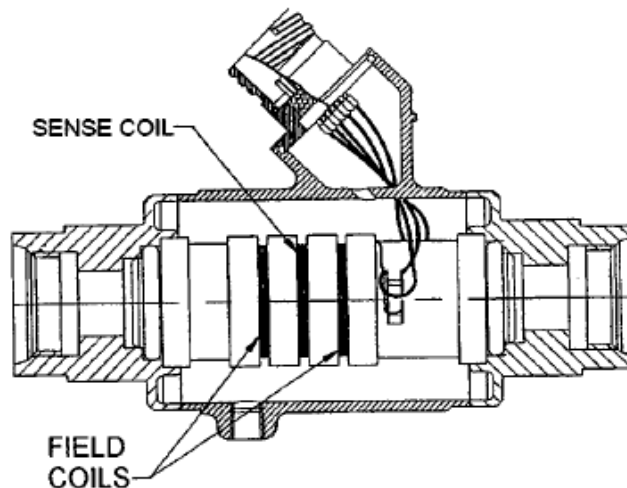


Figure 2-1 The structure of a MetalSCAN oil debris sensor

Figure 2-1 is the structure of the sensor. The sensor is a throughflow device. When the lubricating oil which contains metallic particles flows through the sense coil assembly, the metallic particles will alternate the electromagnetic field generated by the field coils assembly. The coil assembly will detect the electromagnetic field change and output a signal. The amplitude of the output

signal is proportional to the mass of the ferromagnetic metal particles, and for the conductive non-ferromagnetic metal particles, is proportional to their surface area. The phase of the output signal for ferromagnetic particles is opposite to that of non-ferromagnetic particles, by which a distinction can be made between the different types of particle metal [2]. Figure 2-2 is the typical output signals of a ferromagnetic particle and a non-ferromagnetic particle.

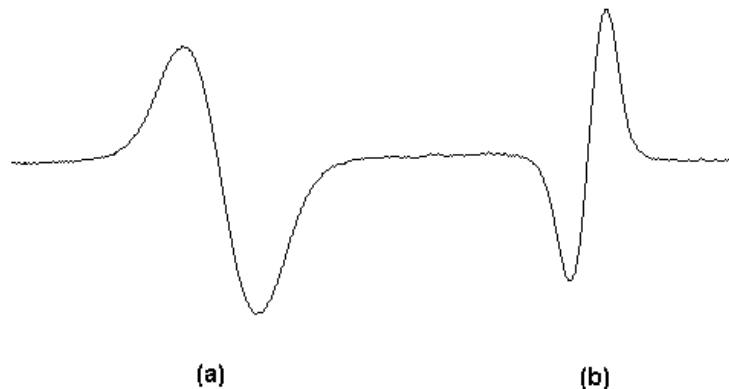


Figure 2-2 (a) is a ferromagnetic particle signal, (b) is a non-ferromagnetic particle signal

For this project, a nonmetal container filled with test lubricating oil will be passed through the oil sensor. The oil sensor will output a signal which corresponded to the metal particles in the lubricating oil. The output signal is the combination of the multi-particle signal and noise, and the output data has limited length. Then the signal was amplified and sent to an A/D converter. The A/D converter convert the analogy signal into digital form and sent it to the decomposition program running on the computer to decompose the signal and count the number of particles.

Figure 2-3 is the block diagram of the typical implementation of the oil sensors and the signal processing circuits. The goal of this project is to develop an effective method to detect the number of metallic particles in the lubricating oil that pass through the oil sensor.

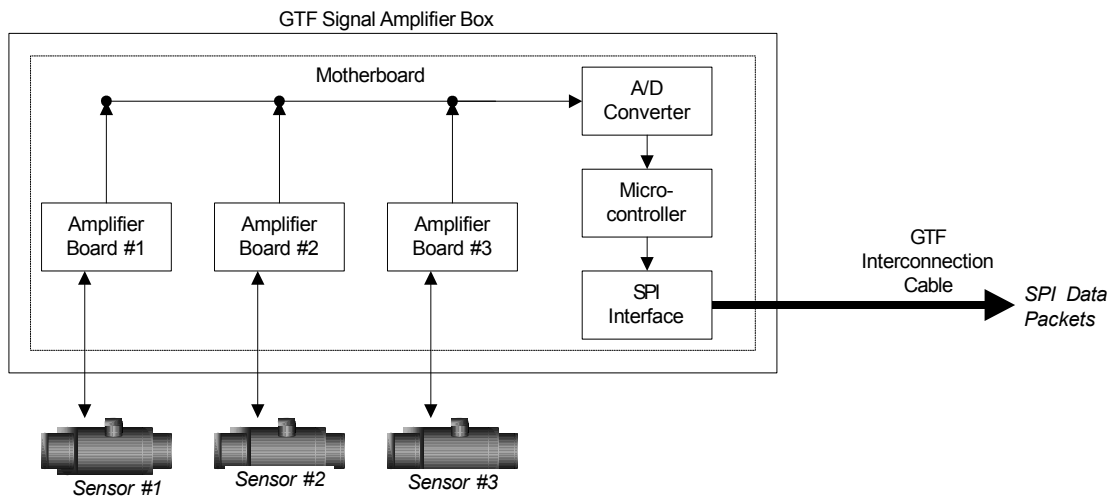


Figure 2-3 A typical MetalSCAN System

2.2 Existing Method

As described in section 1.1, there are two kinds of scenarios for the output signal from the oil debris sensor. The first scenario is for the systems which have strict requirement for the components' wear like aircraft engines. Those systems are working in a very noisy environment and only few particles are allowed in the lubricating system. The particles will generate separate signals when they pass through the oil sensor, like the signals in Figure 2-2. The main problem is to remove the noise in the output of oil sensor and extract the separated particle signals. The second scenario is for the systems which can tolerate more worn

particles in the lubricating oil. They do not require a real time on line monitoring system. The noise in the output of the oil sensor can be relatively small. However, because there will be more than one particle pass through the sensor in a very short time interval, the particle signals from the sensor will superimposed with each other. The main problem for this scenario is to find a decomposition algorithm to decompose the output signal into separate particle signals before the monitoring system counts the number of particles.

In 2007, Xianfeng Fan et al[3] researched the first scenario and developed a method to extract the particle signals from the oil sensor output. They tried to detect very weak particle signals buried in strong background noise and eliminate vibration-like spurious signals.

Fan uses time-invariant wavelet transform for noise reduction, and then uses the kurtosis analysis to extract the signals of metal particles. Kurtosis as a statistical indicator is sensitive to the shape of a signal. It can be used to detect impulses in time domain[4]. Kurtosis is effective to detect the metal particle signal in the de-noised sensor output. The method has been proved to be practical for the first scenario, i.e. each time there is only one particle pass through the oil debris sensor, which means the shape of the useful metal particle signal can be known in advance.

However, for the second scenario, because the output of the oil sensor will be the mixture of overlapped particle signals and the noise. The shape of the sensor signal cannot be predicted, and the kurtosis analysis does not work.

2.3 Conventional Signal Analytical Methods

2.3.1 Analysis method for stationary signal

For a given signal, there are many ways to describe it. Time and frequency are two most fundamental parameters to describe a signal. Stationary signal means its statistical parameters are constant over time. To analyze stationary signals, the most common and most important method is Fourier transform. Fourier transform is built on the ideal signal model. It assumes the signal is linear and stationary. For a given signal $x(t)$, its Fourier transform is:

$$X(j\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t} dt \quad (2.1)$$

The traditional signal analyzing and processing theory is based on Fourier transform. It is used to get the information of stationary signals by transforming signals from time domain to frequency domain. Because it is a global representation of a signal, it cannot describe the local time-frequency properties of nonlinear and nonstationary (time variant) signals. However, in the real world, most of the signals are non-stationary signals. To analyze those non-stationary signals, Fourier transform was modified, and some new theory was developed based on Fourier transform for nonstationary signals.

2.3.2 Some analysis methods for nonstationary signal

To analyze nonstationary signals, many methods were developed in the past years. For example, short time Fourier transform, wavelet transform, Wigner-Ville distribution etc. Many of those methods are based on Fourier transform. Therefore, some of them also inherited parts of the limitations of Fourier transform. There is no such a method that will work with all kinds of non-stationary signals. Appropriate method much be chosen based on the character of the researched signal.

2.3.2.1 Short Time Fourier Transform (STFT)

Dennis Gabor first proposed Gabor transform in 1946[5], which brought the concept of window function into Fourier transform. Then it was developed into Short Time Fourier Transform (STFT). STFT is widely used in nonstationary signal analysis. Its basic idea is to divide the signal into small time intervals (like windows), then apply Fourier transform to get the frequency information in each time interval. The total frequency information of all the time intervals will present how the signal frequency changes over time. By introducing a window function $g(t)$, and constantly sliding the window through the signal, we can get short time Fourier transform of signal $x(t)$:

$$STFT_x(\tau, \omega) = \int_{-\infty}^{+\infty} x(t)g(t - \tau)e^{-j\omega t} dt \quad (2.2)$$

In STFT, $g(t)$ is the function of time. By the introduction of $g(t)$, the signal which is converted to the frequency domain has a connection with time. Because STFT

still needs Fourier transform to analyze the signal in each window, it must assume that the signal in each window is stationary, which is difficult to guarantee. The time window in STFT cannot change its size during the processing. However, for the nonstationary signal, its frequency is changing over time. To get high time resolution, the time window should be small enough. On the other hand, if high frequency resolution is also necessary at the same time, the time window needs to be big enough, which means for STFT the time resolution and frequency resolution cannot be acquired at the same time. Although STFT has some drawbacks, because it can be easily implemented with fast Fourier transform, STFT is still used in many fields.

2.3.2.2 The wavelet transform

Wavelet transform is developed based on Fourier transform. The basic idea is to build a window function which can adjust itself with the changing of signal. In another word, wavelet transform is one kind of adjustable window Fourier transforms.

The wavelet concept was introduced in seismic signal analysis by S. G. Mallat in 1984. When he analyzed the local properties of seismic wave, he needed an algorithm which could change the time frequency resolution during the decomposition. Mallat and A. Grossman proposed the wavelet transform for nonstationary signal analysis. At first their inversion formula was not accepted by mathematicians. Until 1986, Y. Meyer[6] constructed the first real wavelet base.

Wavelet transform started its fast development. In 1989, Mallat proposed the multi-resolution signal decomposition (MRSD) algorithm[7]. Daubechies[8] used the approach to construct orthonormal bases of compactly supported wavelets, and provided a general way for constructing orthogonal wavelet bases. All their works led to the development of fast wavelet transform algorithm[9].

Wavelet transform is effective in nonstationary signal analysis, especially for the data with gradual frequency changes. It is very popular in many fields like image processing, signal filtering and even in physics and astronomy.

Although wavelet transform is versatile in many fields, it also has some limitations. For the most commonly used Morlet wavelet, there is energy leakage during the processing. Another limitation for the wavelet transform is, it is not a self-adaptive algorithm. The wavelet base is set before the processing. It must be used to analyze all the data and cannot be changed during the processing. Because the Morlet wavelet is based on Fourier transform, it also has some shortcomings of Fourier transform. It is only fit for the linear signal processing. However, even with all those problems, wavelet transform is still a widely used nonstationary data analysis methods.

2.3.2.3 The Winger-Ville distribution

The Wigner-Ville distribution is widely used in the electrical engineering field. It is the Fourier transform of the central covariance function. For a signal $X(t)$, its central covariance is:

$$C_c(\tau, t) = X\left(t - \frac{1}{2}\tau\right)X^*\left(t + \frac{1}{2}\tau\right) \quad (2.3)$$

The Wigner-Ville distribution is

$$V(\omega, t) = \int_{-\infty}^{\infty} C_c(\tau, t)e^{-i\omega\tau} d\tau \quad (2.4)$$

This method has the difficulty of the interference due to the cross-terms, because it introduces negative energy into the processing. An improvement of this method was made by Yen[10]. Yen defined wave packets by Wigner-Ville distribution, and then use the wave packet to decompose the complicated data into simple components. This method is effective in some situations. It still needs more proof for the applications to complicated data.

2.4 The EMD Method for Particle Detection

Because the output signal of the oil sensor is non-stationary, an appropriate signal decomposition algorithm must be chosen. In 1998, Norden E. Huang et al[11] developed a new technique to decompose a signal without leaving the time domain. They called this method Empirical Mode Decomposition (EMD). It is an adaptive data analysis method designed for analyzing non-linear and non-stationary signals. With EMD method a complicated signal can be decomposed

into finite number of components with different frequencies. Those components are called intrinsic mode functions (IMF). The combination of EMD and the Hilbert spectral analysis is called Hilbert-Huang transform (HHT). As a new decomposition method, EMD has some characters which made it suitable to process the output signal from the oil sensor:

1. EMD is designed to process the non-linear and non-stationary signals, which is not fit for traditional Fourier Transform. By combining with Hilbert transform, this method can reveal the accurate time-frequency relationship of the original signal.
2. Because EMD method is data-adaptive, it has good local resolution when processing non-stationary signals. The signal is decomposed in time domain, which makes it possible to count the number of particles directly from the decomposition result.
3. EMD method can process short data series, and usually has better results than other decomposition method.

Although EMD method has advantages in signal decomposition, it still has limitations, like the end effect and the mode mixing problem. As its name, EMD method is still an empirical method. It has been proved to be very effective in many fields, but it still needs theoretical support.

Chapter 3 Empirical Mode Decomposition

3.1 Introduction

The Empirical Mode Decomposition (EMD) technique was invented by Norden E. Huang in 1998, when Huang was working for NASA[12]. The invention of EMD is the result of studying the Hilbert transform. Hilbert transform is a very useful signal analysis method. However, because of the preconditions of Hilbert transform, its application is limited. The decomposition result of EMD can provide data that will fulfill the preconditions of Hilbert transform. In many case EMD is combined with Hilbert transform, and named Hilbert-Huang transform (HHT). EMD is used to decompose the signal into finite number of components, and then the components are converted by Hilbert transform to analyze their amplitude-frequency-time characters. HHT is an adaptive time-frequency data analysis method designed to analyze nonlinear and non-stationary data, and was proved to be effective in wide range of applications.

After studying the concept of instantaneous frequency in Hilbert transform, Norden E. Huang et al [11, 13] proposed a new concept of Intrinsic Mode Function (IMF), which is the basic component of any complicated data set. Then they developed a new method named Empirical Mode Decomposition to decompose a signal into a set of Intrinsic Mode Functions. Different from traditional decomposition method like Fourier transform and wavelet, this new

method is data-adaptive. It does not use any predetermined filters or transforms, and it does not need to assume data is linear and stationary like Fourier transform. Therefore, it works well for data that are nonlinear and non-stationary.

3.1.1 The instantaneous frequency

For the stationary signal processing, in Fourier transform, signal frequencies do not change over time. However, for non-stationary signal analyzing, the signal frequencies vary from time to time. Fourier transform is not fit to be used as an analysis tool because it cannot reveal the time-frequency relationship for the non-stationary signal. In this case, the term "instantaneous frequency" was introduced to describe the non-stationary signal's frequency at each given time point[14].

The instantaneous frequency can be computed through the Hilbert Transform. For any time series $x(t)$, its Hilbert transform $y(t)$ is

$$y(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau \quad (3.1)$$

The analytic function

$$z(t) = x(t) + iy(t) = a(t)e^{i\Phi(t)} \quad (3.2)$$

Where

$$a(t) = \sqrt{x(t)^2 + y(t)^2} \quad (3.3)$$

$$\Phi (t) = \tan^{-1} \frac{y (t)}{x (t)} \quad (3.4)$$

$a(t)$ is the instantaneous amplitude, $\Phi(t)$ is the phase function.

The instantaneous frequency is

$$\omega (t) = \frac{d\Phi (t)}{dt} \quad (3.5)$$

However, if we calculate the instantaneous frequency in this way, in some situation we will get a negative value, which has no physical meaning in real world.

After investigating the character of the instantaneous frequency, Norden E. Huang et al found that the Hilbert Transform can only produce physically meaningful results for 'mono-component' signals[11]. They summarized the requirement for those 'mono-component' signals and proposed the concept of Intrinsic Mode Function (IMF).

3.1.2 The Intrinsic Mode Function (IMF)

The two requirements for being an Intrinsic Mode Function:

1. In the whole data set, the number of extrema N_e (include all the minima and maxima) and the number of zero-crossings N_z must either equal or differ at most by one, namely,

$$(N_z - 1) \leq N_e \leq (N_z + 1) \quad (3.6)$$

2. At any time point t_i , the mean value of the upper envelope $f_{max}(t)$ defined by the local maxima and the lower envelope $f_{min}(t)$ defined by the local minima is zero, namely,

$$\frac{[f_{\max(t_i)} + f_{\min(t_i)}]}{2} = 0 \quad (3.7)$$

In another word, the second requirement is used to make sure the local mean of an IMF signal is zero. For a non-stationary signal, calculating local mean needs to define a local time first, but choosing the appropriate time scale for local time is not easy. In this case the mean of the upper and lower envelopes are used to approximate the same result of local mean, to force the IMF signal to be locally symmetrical.

An IMF function which satisfies those two requirements has only one extremum between two zero-crossings. There is no other signals overlap with an IMF. Therefore, Huang et al. presume that IMFs are the fundamental components of the signals in real world.

3.1.3 Empirical Mode Decomposition (EMD)

Although IMFs are the fundamental components of signals, they are different from the simple harmonic functions: their amplitude and frequency are not required to be constant; an IMF can have variable amplitude and frequency.

Usually, signals in real world are not IMFs. They are the result of some IMFs superimposed with each other. Then the next question is, how can we decompose those signals into IMFs? Since the instantaneous frequency generated from IMFs' Hilbert Transform has physical meaning, Huang et al. developed a sifting process to separate each IMF in a signal. They called this sifting method Empirical Mode Decomposition (EMD).

The EMD decomposing process is as follows:

1. For the original signal $x(t)$, treat $x(t)$ as the data.
 - a. Identify all the extrema (include all the minima and maxima) of $x(t)$.
 - b. Then interpolate between minima by a cubic spline, get the lower envelope $e_{\min}(t)$.
 - c. Repeat the same processes to the maxima to get the upper envelope $e_{\max}(t)$. All the data of $x(t)$ should be covered by upper and lower envelopes as given in Figure 3-1.

d. Calculate the mean $m_1(t) = (e_{\min}(t) + e_{\max}(t))/2$

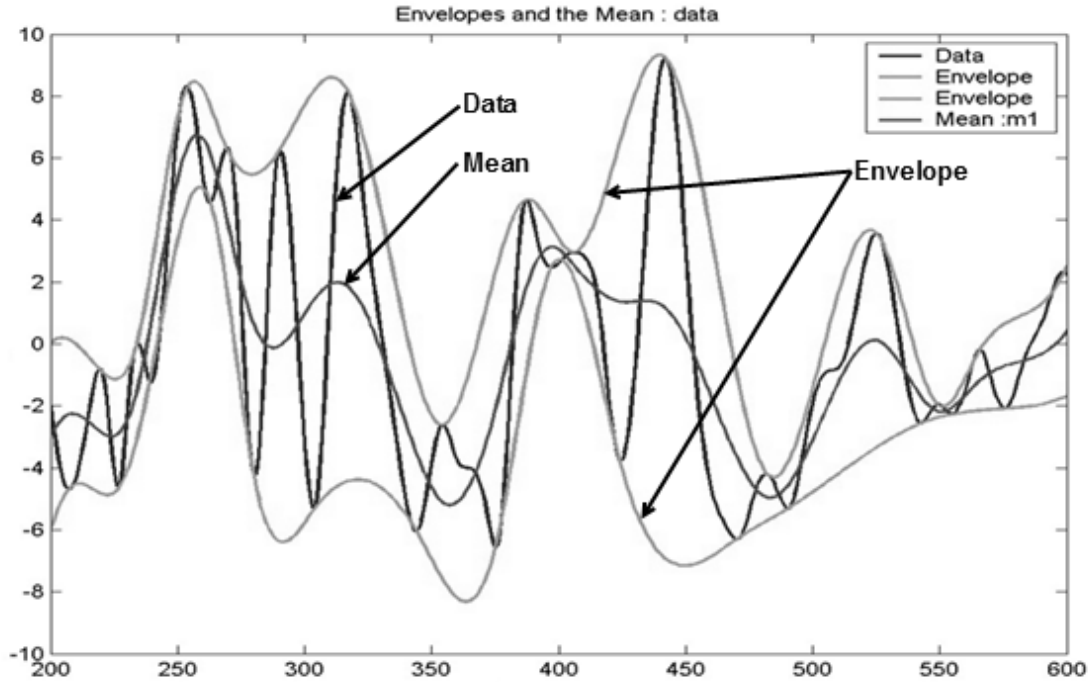


Figure 3-1 Envelopes and Mean[15]

2.

a. Get the first component $h_1(t)$ by subtracting $m_1(t)$ from original signal $x(t)$:

$$h_1(t) = x(t) - m_1(t) \quad (3.8)$$

Test if $h_1(t)$ satisfies the two IMF requirements. If $h_1(t)$ is not IMF, treat $h_1(t)$ as data, repeat all the processes in step 1 and get new component $h_2(t)$,

$$h_2(t) = h_1(t) - m_2(t) \quad (3.9)$$

Test $h_2(t)$, repeat processes in step 1 up to k times, until $h_k(t)$ is an IMF.

- b. Record $h_k(t)$ as the first IMF $c_1(t)$:

$$c_1(t) = h_k(t)$$

3. Get the residue of the data $r_1(t)$ by subtracting $c_1(t)$ from $x(t)$,

$$r_1(t) = x(t) - c_1(t)$$

Treat $r_1(t)$ as new data.

4. Iterate step 1, 2, 3, find the new IMFs $c_2(t)$, $c_3(t)$, ..., $c_n(t)$, until $r_n(t)$ fulfills at least one of these 3 conditions:
- a. $r_n(t)$ is a monotonic function.
 - b. $r_n(t)$ has only one extremum.
 - c. $r_n(t)$ is a constant.

Then the sifting is stopped. We get all the IMFs and the last residue of data $r_n(t)$.

Finally, the original signal $x(t)$ can be expressed as the sum of n IMFs and $r_n(t)$:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (3.10)$$

Through the sifting process, the number of extrema is decreasing during the iteration and the time interval among extrema is increasing. Each decomposition process will generate a detail signal (IMF) and a residue whose frequency is lower than the detail signal. In another word, during the iteration usually IMFs which contain higher frequencies are sifted out first, then those lower frequency IMFs.

In step 2, theoretically, the two IMF requirements can be used as criteria to test if $h_k(t)$ is an IMF. However, for the second requirement, practically it is difficult to make the calculated result of local mean to be exact zero. An alternative criterion should be chosen to determine the number of sifting steps of producing an IMF.

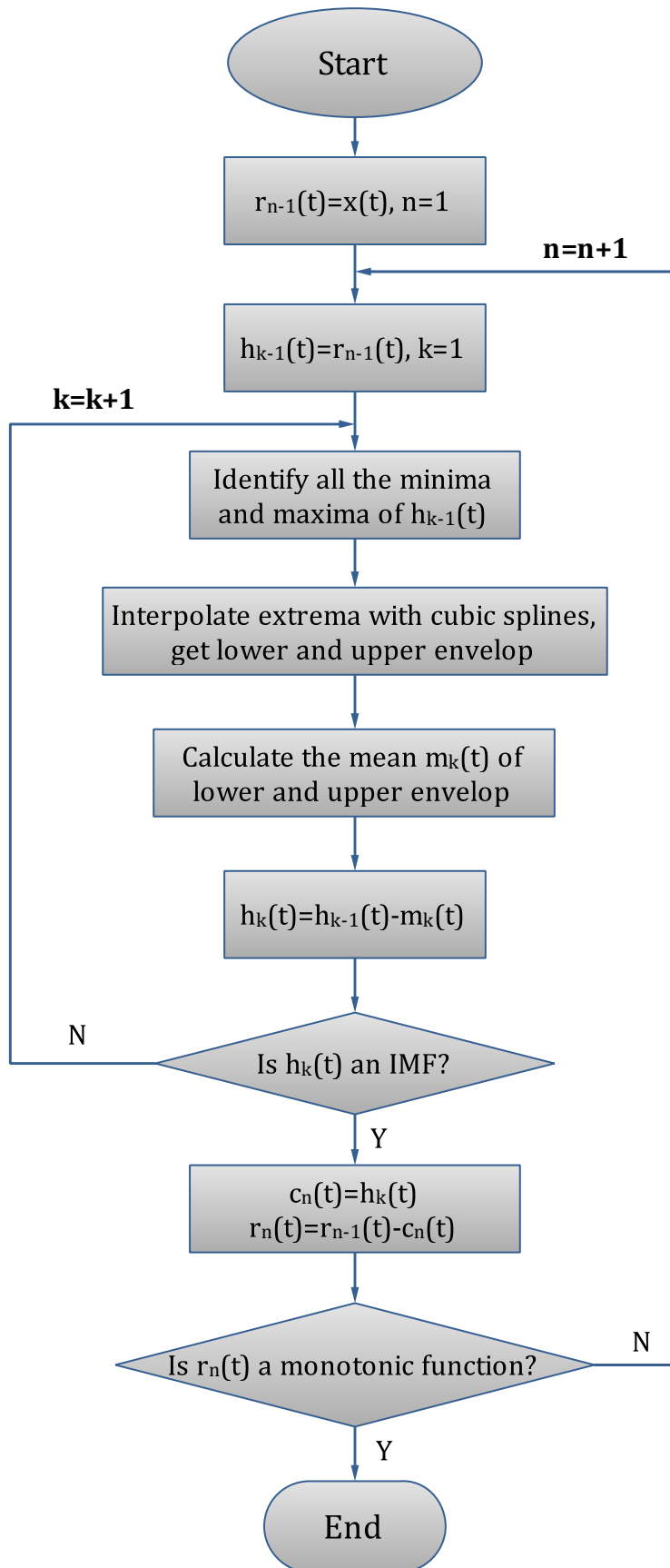
There are two kinds of stoppage criteria can be chosen. The first criterion is based on the agreement of the numbers of extrema and zero-crossings. A value S is preset, if the sifted signal $h_k(t)$ satisfies the first requirement of IMF, and the numbers of extrema and zero-crossings do not change after S consecutive times process, the iteration is stopped.

The second criterion, which was used in Huang's program [11], is to set the standard deviation value S_d , which is calculated from two consecutive sifting results:

$$S_d = \sum_{t=0}^T \frac{|(h_{k-1}(t) - h_k(t))|^2}{h_{k-1}^2(t)} \quad (3.11)$$

Usually the typical value for S_d is chosen between 0.2 and 0.3[11]. This criterion compares the difference between two consecutive sifting results. When the difference is less than the setting value, the sifting for current IMF is stopped.

The flowchart of EMD algorithm is shown as follows:



3.1.4 Completeness and orthogonality

The completeness of a decomposition method refers to the character that the original data can be reconstructed by adding all the decomposed components and the residue. From equation(3.10), the completeness has been confirmed. In practical experiment, the order of magnitude of the reconstruction error between the reconstructed signal and original signal is usually $10^{-15} \sim 10^{-16}$. The errors are mainly the approximation errors during computing.

The orthogonality for a decomposition method is necessary to make sure the decomposed components are not overlapping with each other. In mathematics, if two function $x_1(t)$ and $x_2(t)$ satisfy

$$\int_{t_1}^{t_2} x_1(t)x_2(t) dt = 0, \quad t_1 < t < t_2$$

Then $x_1(t)$ and $x_2(t)$ are orthogonal.

The orthogonality of EMD method still cannot be theoretically proved. Huang et al.[11] had done many experiments in different research areas, tried to find the character of the EMD's orthogonality by investigating the decomposed data. For the decomposed data, in equation(3.10), if $r_n(t)$ is treated as an IMF, it can be recorded as the $n+1$ $c_i(t)$. The equation can be simplified:

$$x(t) = \sum_{i=1}^{n+1} c_i(t)$$

The square of the signal is:

$$x^2(t) = \sum_{i=1}^{n+1} c_i^2(t) + 2 \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} c_i(t)c_j(t) \quad i \neq j$$

If decomposition is orthogonal, the second part of the right side of the equation should be zero. We can define the overall index of orthogonality (IO) as

$$IO = \sum_{t=0}^T \left(\sum_{i=1}^{n+1} \sum_{j=1}^{n+1} c_i(t)c_j(t) / x^2(t) \right) \quad i \neq j$$

According to Huang et al. 1998, in one of the experiment of analyzing decomposed wind data, the IO value is only 0.0067. Orthogonality can also be defined for any two IMF components, C_f and C_g , their orthogonality will be

$$IO_{fg} = \sum_t \frac{C_f C_g}{C_f^2 + C_g^2}$$

According to the experiments done by Huang and other researchers, practically, it can be conclude that the EMD method is orthogonal. It should be noted that the orthogonality discussed here refers to local orthogonality, not global. For some special data, it is possible that the adjacent components have sections of data which contain the same frequency at different time durations. But in any

situations any two components should be locally orthogonal. Huang concluded that for the typical data, the IO value (the energy leakage) should be less than 1%, and the leakage could be 5% for some extremely short data[11].

3.2 Some limitations of EMD

As a new algorithm, EMD has been proved to be very effective in some research fields. However, like all the new algorithms, EMD also has some limitations need to be overcome.

3.2.1 Lack of theoretical background

Ten years after the first introduction of EMD algorithm, EMD still do not have a theoretical background. It also lack of good mathematical (analytical) properties. The research of EMD is usually based on statistical approaches. Some basic questions like "Which kind of signals can be analyzed by EMD?" still cannot be answered. Like its name "Empirical Mode Decomposition" described, EMD is still an algorithm based on empirical research.

In 2006, Z. Youming et al.[16] did some research for the theoretical basis of EMD. They focused on the definition of IMF. They concluded that, to make the definition of instantaneous frequency

$$f(t) = \frac{1}{2\pi} \frac{d(\theta(t))}{dt} \quad (3.12)$$

has a physical meaning, IMF should be described as

$$c(t)=a(t)\cos(\theta(t)) \quad (3.13)$$

In EMD algorithm, the two requirements for being an IMF means IMF could be written in a form like Eq.(3.13), when $a(t)$ and $\cos(\theta(t))$ fulfill some preconditions. They also pointed out that the definition of IMF should be bound up with the Bedrosian theorem which was used in Hilbert transform. They analyzed the theoretical basis of the IMF definition, and then gave some proofs for their theory. However, those proofs lack of mathematical rigor. Their theory still needs more detailed demonstration.

3.2.2 Stopping Criteria for sifting

In EMD, the stopping criteria determine when the sifting process should stop. It directly influences the result of EMD. Insufficient sifting process cannot generate an IMF. Too much sifting process will turn the IMF into a pure frequency modulated signal, and its amplitude will be constant. To make sure the IMF retain enough physical sense of both amplitude and frequency modulations, the stopping criteria must be chosen carefully.

When Huang first time introduce the EMD algorithm[11], he used a stopping criterion by calculating the standard deviation of two consecutive sifting results:

$$S_d = \sum_{t=0}^T \frac{|(h_{k-1}(t) - h_k(t))|^2}{h_{k-1}^2(t)}$$

Usually the S_d value was set between 0.2~0.3. If calculated S_d is less than the setting value, the sifting process is stopped. In the original paper, Huang did not give the theoretical explanation of choosing this stopping criterion.

In 2003, Rilling et al.[17] proposed some improvement for the EMD. For the stopping criteria, they introduced two thresholds θ_1 and θ_2 . In addition to the original mean $m(t) = (e_{min}(t) + e_{max}(t))/2$, they introduce a mode amplitude $a(t) = (e_{max}(t) - e_{min}(t))/2$ and the evaluation function $\sigma(t)$:

$$\sigma(t) = \left| \frac{m(t)}{a(t)} \right| = \left| \frac{e_{max}(t) + e_{min}(t)}{e_{max}(t) - e_{min}(t)} \right| \quad (3.14)$$

When the $\sigma(t)$ in the setting fraction $(1 - \alpha)$ of the total duration fulfill $\sigma(t) < \theta_1$, and in the rest of the fraction fulfill $\sigma(t) > \theta_2$, the sifting process is stopped. The typical value of those thresholds could be $\theta_1 \approx 0.05$, $\theta_2 \approx 10\theta_1$ and $\alpha \approx 0.05$. The goal of setting θ_1 is to make sure the globally small fluctuations are in the mean, while θ_2 is to deal with locally large excursions.

Comparing with the stopping criterion used by Huang, $\sigma(t)$ focuses on the second condition of IMF's definition. It confines the large fluctuations only in some local areas, therefore, guarantees the calculated IMF's mean values are zero.

Experiments also proved that, in some situation, Rilling's method generated better sifting results.

3.2.3 End effect

When EMD is used to analyze finite length signals, the results usually will be influenced by end effect. During the sifting process, EMD interpolates extrema by cubic splines to get the upper and lower envelopes. The most serious problem of the spline fitting is at the ends. At each end of the data series, because the signal may not be at the extreme points, and there is not enough information outside the end point, the divergence will appear on both ends of data series. The calculated envelopes will deviate from the real signals' trend. Furthermore, because the EMD sifting process is actually a series of iterations, the errors generated at the early stage of iterations will be introduced in the further calculation, which will cause more data contamination. For long data series, the data contamination can be controlled by throwing away some data at each end of the signal. For short data series, it becomes complicated because there are not enough data to throw away.

When Huang introduced the EMD in 1998[11], he also proposed a method to reduce the end effect by adding characteristic waves to each end of the data. Characteristic waves are the last (first) waves at each end of the original signal. By introducing more information for the sifting process, this method has confined the large swings successfully.

In 2003, Rilling et al.[17] proposed another method to deal with the end effect. To add more data to each end of the original signal, Rilling added extrema by symmetrically mirroring with the extrema that are closed to the data edges. This method put mirror faces at each end of the data, as shown in Figure 3-2, and then finds the nearest extrema to the mirror face, i.e. Min(1) and Max (N). The points which mirror symmetrically with the found extrema are mirror extrema (Min (0) and Max (N+1)). Mirror method adds extra data for sifting process. At the same time it can guarantee that for any processed signals, the end of the data must be extrema, and the trend of the signal at the edge could be preserved. Therefore, the divergence during the sifting process could be reduced. Mirror method works well when the original signal does not have big asymmetry. Currently Rilling's mirror method is widely used in EMD algorithm for signal analysis.

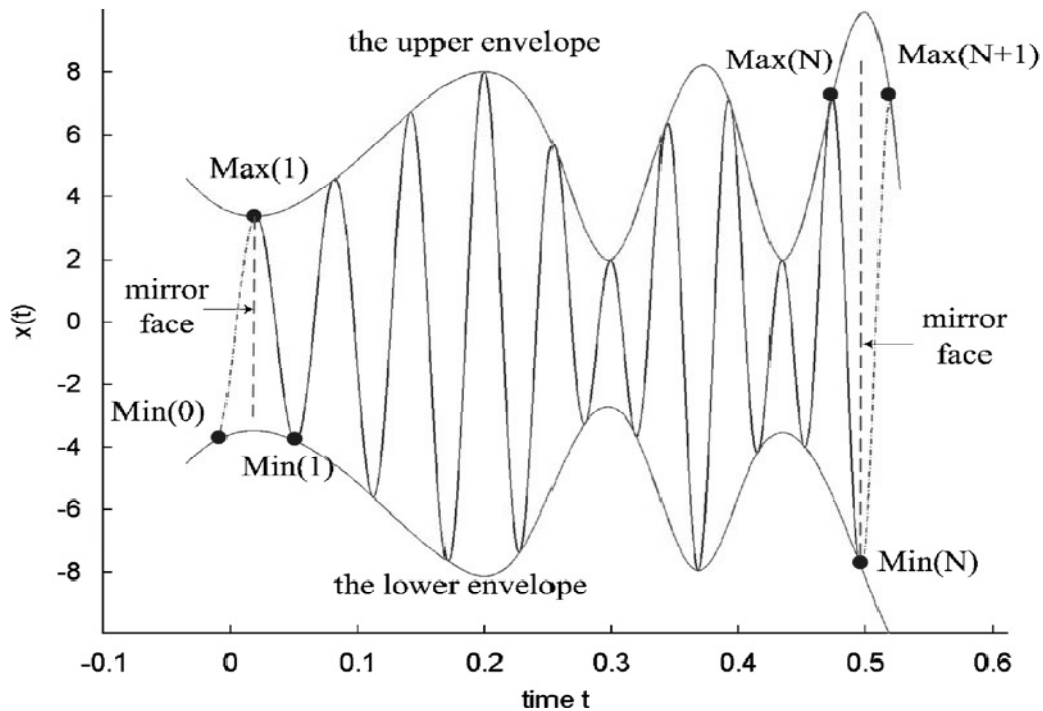


Figure 3-2 Rilling's mirror method

End effect is one of the most discussed topics in EMD research papers. Many people proposed different methods to overcome the end effect for different situations. For example, in 2001, Y. J. DENG et al.[18] proposed neural network method. In 2004, H. T. LIU et al.[19] proposed orthogonal polynomial fitting algorithm. In 2008, F. Wu et al.[20] proposed improved slope-based method. Those methods are proposed for different situations, and all have their limitations. For the end effect problem in EMD algorithm, there are still a lot of work to do.

3.2.4 Mode Mixing

Mode mixing is another problem in EMD algorithm. It frequently appeared in EMD processing. Mode mixing 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. Usually, mode mixing is caused by intermittency. Signal intermittency is a kind of disturbance signals without fixed form. Intermittency will make the physical meaning of decomposed IMF unclear. After studying the white noise decomposed by EMD, in 2009, Huang et al.[21] proposed a noise-assisted data analysis (NADA) method to deal with the mode mixing in EMD, which will be discussed in next chapter.

Chapter 4 Ensemble Empirical Mode Decomposition

4.1 Introduction

As an adaptive time-frequency data analysis method, EMD has been proved to be effective in nonlinear and non-stationary data analyzing. However, during the implementation of EMD, some drawbacks also emerged. One of the major problems of the original EMD is mode mixing. Mode mixing frequently appeared in EMD processing, it interfered with the decomposition result, sometimes even made the generated IMF lost its physical meaning. Mode mixing is caused by signal intermittency. How to overcome the influence brought by intermittency has been an important topic of EMD research. In 2004, after analyzing the statistical characteristics of white noise using EMD method, Flandrin et al.[22] and Wu and Huang[23] reported some important properties of EMD. They found that when EMD was used to decompose white noise, the EMD serves as an adaptive dyadic filter bank for various types of noise. This means that, for the IMFs decomposed from white noise, the mean period of any IMF component almost exactly doubles that of the previous one. Because the intermittency is the reason of mode mixing, and the intermittency means some scales are missed in a signal, Huang considered introducing white noise into the original signal to compensate missed scales. Therefore, in 2009, Huang et al. proposed a noise-assisted data analysis (NADA) method to deal with the mode mixing in EMD.

4.1.1 Noise-Assisted Data Analysis

From the beginning of signal processing, noise always appears with the true signal. In most case, noise plays an undesired role in signal analysis. Noise can obscure or reduce the clarity of a signal. Many data analysis methods were designed to reduce or remove the noise from the true signal. In general, all data are mixture of real signal $s(t)$ and noise $n(t)$:

$$x(t) = s(t) + n(t)$$

There are some issues that should be considered during separating the signal and the noise: (1) Some assumptions about the character of the data (For example, data should be linear, stationary.) need to be made. (2) There is no perfect noise removing method. No matter which method was used, the processed data still contain some noise. (3) During de-noise processing, some portion of the real signal also are removed or deformed as part of the noise.

In some situation, noise could be used to aid data analysis. The earliest work of noise assisted data analysis was done by Press and Tukey. They used a method called pre-whiting in spectral estimation, by adding white noise to flatten the narrow spectral peaks. Now the pre-whiting method has been widely used in signal analysis.

Noise also can be used in algorithm research, by adding noise to data to test the sensitivity and robustness of a given data analysis algorithm. Another example of

using noise to assist data analysis is introducing a physical phenomenon called stochastic resonance into data analysis. By adding an appropriate measure of noise to a specific system, the noise will help to increase the signal-to-noise ratio of the system output, like the resonance phenomenon in mechanics.

To resolve the mode mixing problem in EMD method, Huang et al. proposed to add white noise to the data, which will help to compensate some missed scales in the original signal. They also used the cancellation effects associated with the white noise, to remove most of the added noise to get a better result.

4.1.2 The mode mixing problem

Based on the definition of mode mixing in EMD, mode mixing includes two different situations:

1. A single IMF consists of signals of widely disparate scales.
2. A signal of a similar scale resides in different IMF components.

According to Huang's research [13, 24], mode mixing is caused by signal intermittency. When a signal with intermittent noise (also called abnormal signal) was decomposed by EMD, the generated IMFs will consist of both useful signal and abnormal signal. Figure 4-1 is an example of signal intermittency.

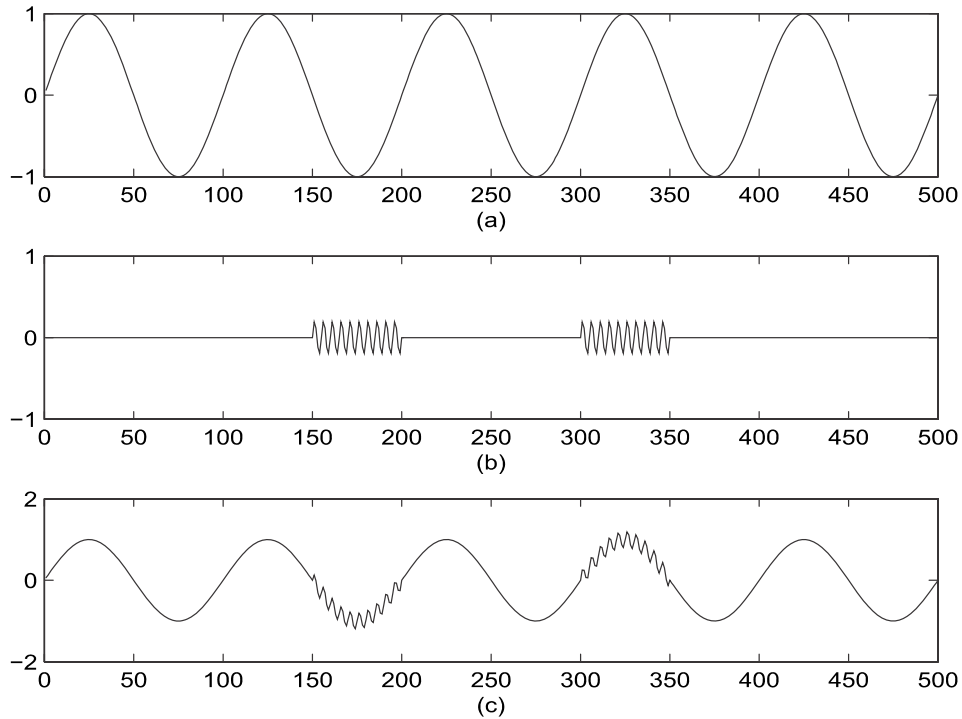


Figure 4-1 (a) is a low frequency signal. (b) is a small intermittent signal with high frequency. (c) = (a)+(b).

Signal a is a sine wave, $a=\sin(0.02\pi t)$. Signal b is a high frequency intermittent signal with amplitude of 0.2. Signal c is the sum of signal a and b. For period 150-200 and 300-350, the intermittent signal b rides on the low frequency signal a, form an intermittent signal c. Figure 4-2 is the decomposed result of signal c, using the original EMD algorithm.

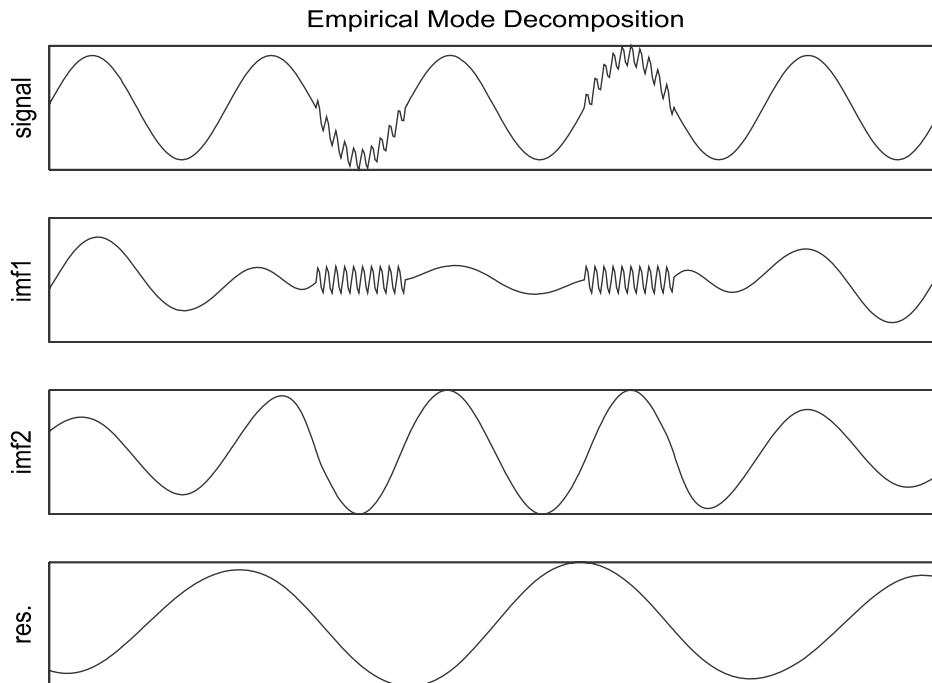


Figure 4-2 Decomposed IMFs of the signal c

The decomposed result shows that the EMD method is not successful for this situation. In Figure 4-2, obviously IMF1 contains both high and low frequency signals. Furthermore, because the mode mixing occurred during the sifting of first IMF, the error data was introduced into following sifting processes, which caused the distortions at each end of IMF2.

To overcome the mode mixing problem, Huang first propose the intermittency test method in one of his paper[13]. Intermittency test will set a criterion based on the period length. Only the point with a period length lower than the criterion can be included in any given IMF component. This method was reported to be effective in decomposing intermittent signal. However, this method also has its limitations: First, intermittency test needs a setting criterion, which means it

needs outside intervention. The intervention compromised EMD's self-adaptive property. Second, Many signals in real world are continuously distributed, and consisted of a variable range of frequency. It is very difficult to select a single criterion of intermittence test.

In 2004, after studying the statistical characteristics of white noise using EMD method, Huang et al.[21] proposed an improved EMD method, named Ensemble Empirical Mode Decomposition (EEMD).

4.2 Ensemble Empirical Mode Decomposition

The idea of using noise to assist EMD came from studying the statistical characteristics of white noise using EMD. In 2003, Flandrin et al.[22] and Wu and Huang[23] used EMD to decompose white noise. After analyzing the result, they found that the frequency components in the white noise were separated regularly. EMD acts like a dyadic filter bank. In the decomposed result, except the first IMF, each IMF's mean period (mean period equals total number of data divided by total number of extrema) is almost exactly two times of its previous IMF. Because the reason of mode mixing is signal intermittency, which means some scales are missed in those signals, they considered using white noise as a uniformly distributed reference scale for the EMD method to overcome the intermittency problem.

In 2005, Flandrin et al.[25] tried to resolve one of the difficulties of EMD method. Because EMD decomposes signal base on the extrema in the data, if the data

does not have enough extrema, EMD will not work. When Flandrin et al.[25] decomposed a Dirac pulse, EMD did not work because the Dirac pulse only has one extrema in the data. Flandrin et al. tried to add noise with small amplitude to the Dirac pulse data, and then decomposed the data with EMD. Because the decomposition results are sensitive to the noise, they need to find a method to remove the noise from the results. They tested 5000 decompositions with different white noise each time, then they calculated the mean of those 5000 results as the final result because the white noise in the 5000 tests cancelled with each other when the 5000 results are add up.

When Huang et al.[21] wanted to use white noise as a reference scale for EMD, they noticed Flandrin's research. The white noise's cancellation effects make it the ideal candidate to improve EMD. White noise can be added before EMD processing to help sifting, then it can be removed from the final result. In 2009, they proposed the EEMD method.

The EEMD method is developed as follows:

1. White noise with small amplitude is added to the original data.
2. The data is decomposed by EMD, and the generated IMFs and residues contain added white noise.
3. The step 1 and step 2 will be repeated for many times (The times is set as the number of ensemble when EEMD starts.). Every time with different

white noise series added. Each time the decomposed IMFs and residues will be recorded. The same IMFs will be put in a group, for example, all the IMF1 will be put in one group, then IMF2 in another group... and so on.

4. Finally, calculate the mean of each group of IMFs and residues. Because of the cancellation effects associated with the white noise, most of the white noise is removed from the IMF. The ensemble mean of each IMF group is deemed as the final result of decomposition.

EEMD can reduce the chance of mode mixing significantly. For example, the signal in Figure 4-1 is decomposed by EEMD as follows:

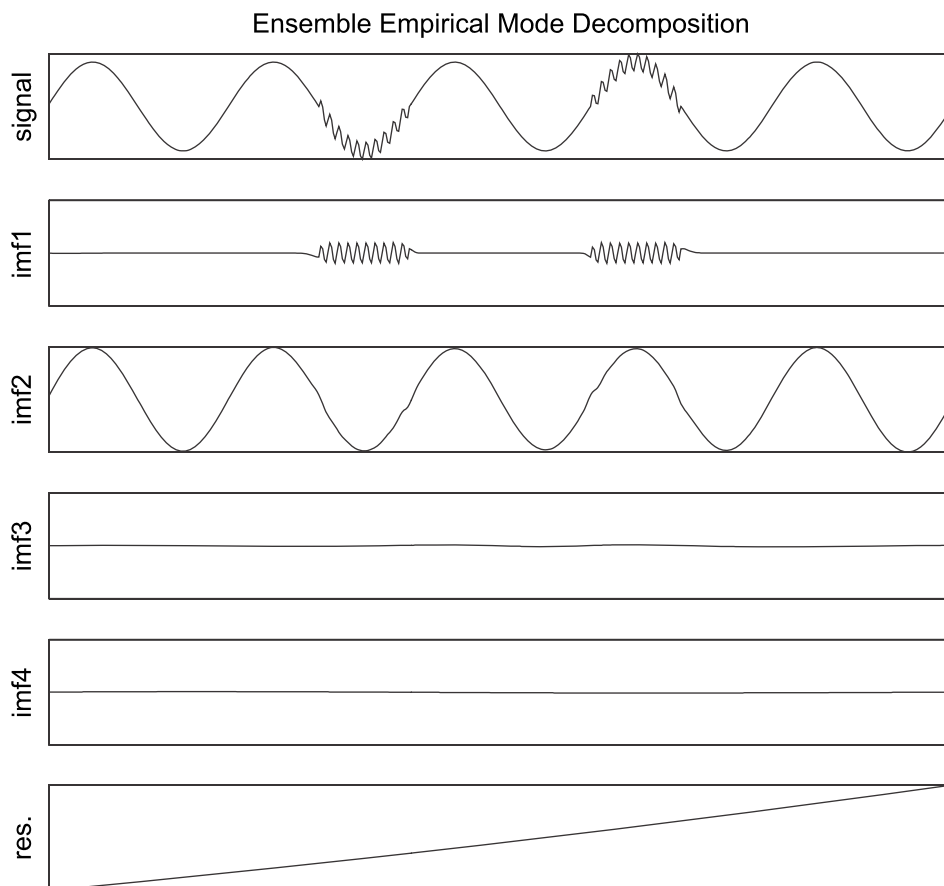


Figure 4-3 EEMD decomposition with ensemble number of 50

Figure 4-3 shows that the signal in Figure 4-1 is well decomposed with EEMD, the intermittency problem which cause EMD to generate false result was eliminated.

EEMD method was proposed after studying mode mixing problem in EMD. It was developed to deal with the intermittency phenomenon, which is the reason of mode mixing. Intermittency in signal will not only cause the mode mixing problem, in some case it will also cause the EMD to generate false signals at the position where intermittency exists. Therefore, by introducing white noise into the processing, EEMD not only significantly alleviates the mode mixing problem, it also reduces the influence of end effect and other phenomena caused by intermittency.

4.2.1 The number of ensemble for EEMD

The white noise added in EEMD follows the statistical rule:

$$\varepsilon_n = \frac{\varepsilon}{\sqrt{N}} \quad (4.1)$$

or

$$\ln \varepsilon_n + \frac{\varepsilon}{2} \ln N = 0 \quad (4.2)$$

N is the number of ensemble. ε is the amplitude of the white noise. ε_n is the final standard deviation of error, which is used to describe the difference between the input signal and the corresponding IMF(s). The statistical rule means, when the amplitude of the white noise is set, theoretically bigger ensemble number will

generate better result. However, it does not mean that, if the ensemble number is set, smaller amplitude of the white noise will generate better IMFs. Because if the amplitude of the added white noise is too small, it may not introduce enough influence on the sifting process.

4.2.2 The selection of the amplitude of added noise

Generally, EEMD result is not sensitive to the added white noise. The amplitude of white noise will influence the EEMD process in two ways: the amplitude of white noise is big enough to help the decomposition, or it is not enough to influence the sifting process. According to Huang's research[21], they suggested that the amplitude of the noise could be about 0.2~0.4 standard deviation of the amplitude of the raw data. However, since there is no equation to help choosing the right white noise amplitude, there are some basic rules could be followed.

When the raw data is dominated by high frequency signals, the noise amplitude could be smaller, and if the raw data mostly consists of low frequency signals, the noise amplitude could be increased. Because the white noise is introduced to deal with intermittency in the raw signal, if the amplitude of the intermittency can be estimated, it also can be a reference for setting noise amplitude. Higher intermittency amplitude needs stronger white noise. It should be noticed that, if the noise amplitude is increased for the decomposition, the ensemble number also should be increased to reduce the influence of the white noise in the final result.

Chapter 5 The implementation of the EMD and EEMD methods in particle detection

5.1 Introduction

For the experiment, at first, the EMD method will be investigated to determine the efficiency of the algorithm when it deals with the output signal of the oil sensor. Because of the limitations (for example, mode mixing) of EMD method, sometimes the decomposition result will contain errors. An experiment will demonstrate the decomposition problem when intermittency appears in the sensor output. To overcome the intermittency problem, EEMD will be introduced. Finally, to compensate the effect of introduced white noise in EEMD, a post processing technique will be implemented to achieve a better result.

5.2 The experiments

5.2.1 EMD and the noise problem

The data which was used for EMD method were collected through the experiments with the oil sensor. During the experiments, some metal particles were passing through the oil sensor with a speed. The output signal of the sensor was received and converted by a data acquisition device made by National Instruments. The sampling frequency of the data acquisition device is 8000Hz. The recorded data for each experiment usually contains two or three metal particles. By overlapping the different data at different position, a signal can be

constructed to simulate the situation that a group of metal particles pass through the oil sensor. Experiments had proved that this kind of constructed signals are consistent with the output signals when multiple particles pass through the oil sensor.

The test signal will contain 16 metal particles, including ferromagnetic and non-ferromagnetic particles superimposed with each other. The noise come with the output signal in the experiment is also kept in the constructed signal. During the construction of the signal, some special cases were avoided. For example, two ferromagnetic particles with similar mass, their output signal will be two sinusoidal waves with similar amplitude when they pass through the oil sensor at different time. When those two particles pass through oil sensor at the same time, their signals are superimposed with each other at the same phase. The result signal will be one bigger sinusoidal wave. For this kind of signal, there will be infinite solutions for the signal decomposition. No algorithm can precisely decompose this kind of signal. The particle signals from experiments and the constructed signal are shown as follows:

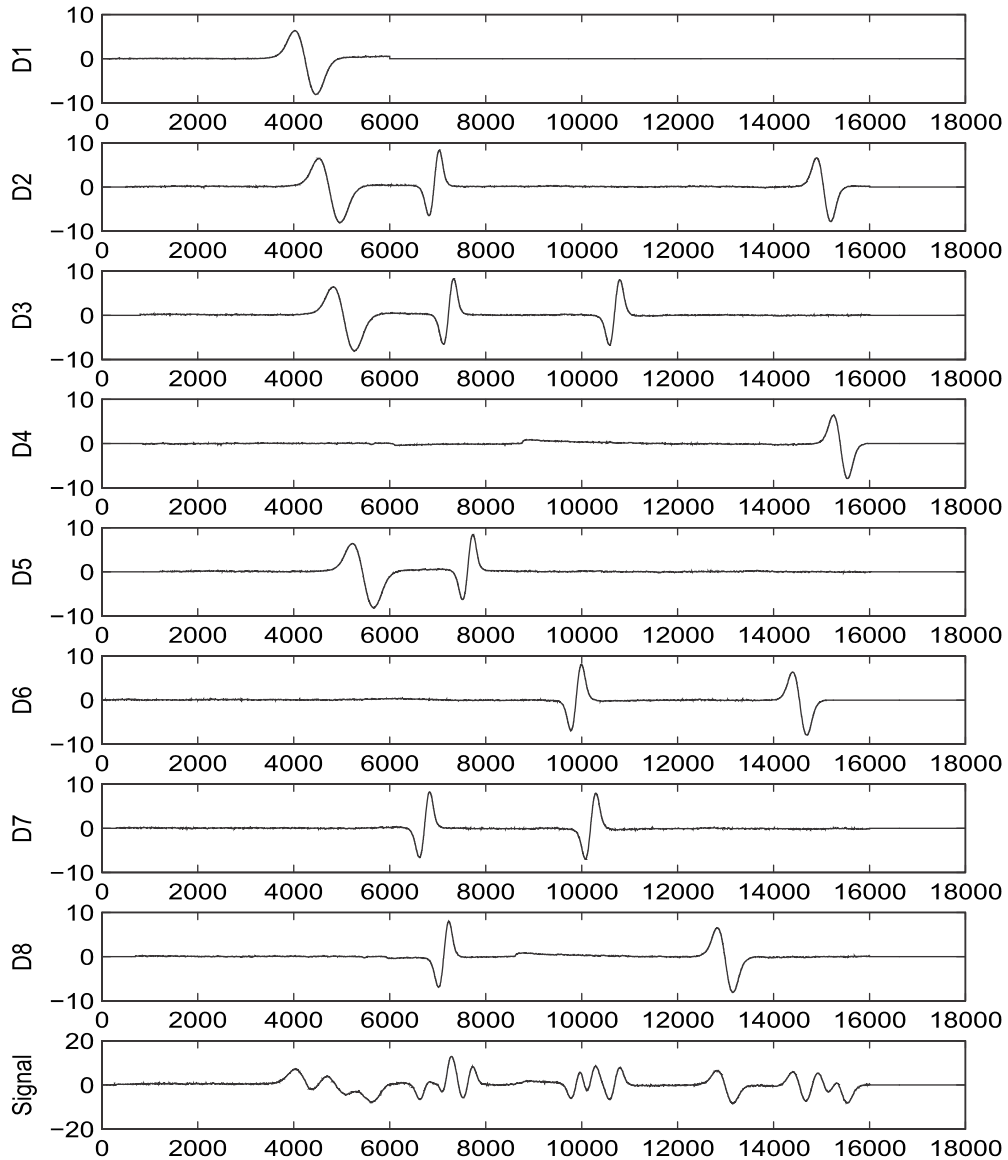


Figure 5-1 Constructed signal and its components

The constructed signal is composed by eight different data collected from experiments. D1~D8 are the data files. The last signal is the sum of those data files. The Y axis represents voltage. The X axis represents time. For the last constructed signal, the data from 3000th point to 17000th point are chosen for the EMD decomposition. Because the sampling frequency of the data acquisition

device is 8000Hz, 14000 data points is equivalent to 1.75 seconds. Figure 5-3 is the decomposition result of the constructed signal.

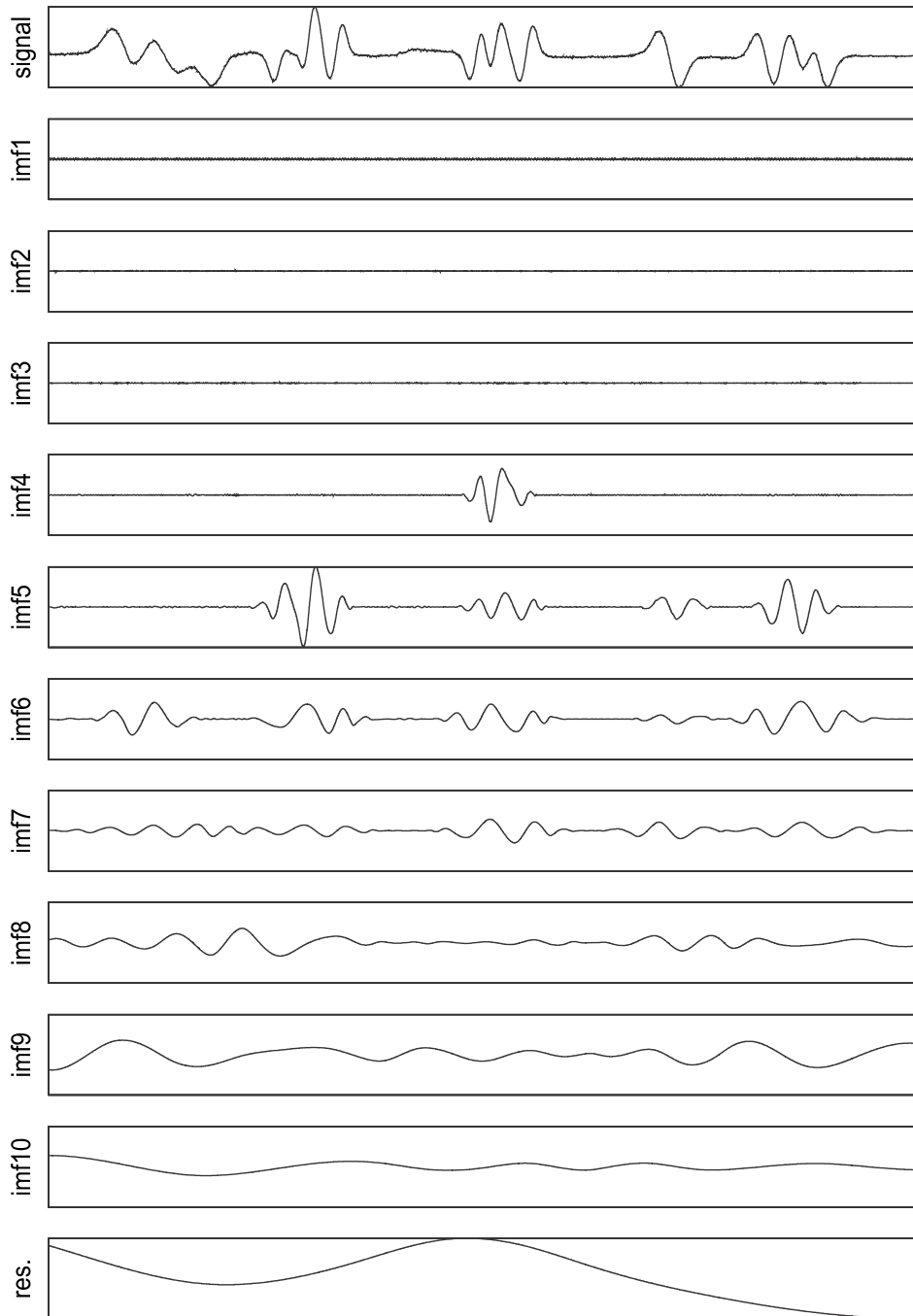


Figure 5-2 EMD Decomposition result of the signal in Figure 5-1

In Figure 5-2, IMF1 to IMF3 are noise come from the experiment data. Comparing with the particle signals, they have higher frequency and smaller

amplitude. Because the plotted IMFs are using the same scale, IMF1 to IMF3 look like a straight line. During the sifting process, EMD acts like a filter bank. The IMFs contain higher frequency components will be sifted out first.

From the result in Figure 5-2, EMD method can decompose the constructed signal, although if we try to count the number of particles from decomposed IMFs, the result is not satisfactory. Except the IMF1~IMF3, which obviously only contain noise, IMF7 contains the largest number of maxima or minima in those IMFs. Therefore, IMF7 is chosen to count the number of particles. Because the decomposed result in each IMF also contains some influence from noise, some small waveforms were not deemed as particle signals when their amplitudes are lower than 10% of the average amplitude of the decomposed signals. In IMF7, 12 particles are counted.

When Huang et al. studied the mode mixing problem in EMD method, they noticed that the mode mixing was caused by the intermittency appeared in the original signal. Then they introduced white noise to compensate the influence of the intermittency. In fact, mode mixing is not the only phenomenon that intermittency will cause in EMD process. In some situation, it will cause EMD to generate false signals at the position where intermittency exists.

Usually, EMD method is stable and not sensitive to the noise. However, in some situation when the frequencies of useful signals are close, EMD process also can

be disturbed by the change of noise. Sometimes when the intermittency occurs in the data segment that only contains noise, the EMD decomposition result will be different. For example, for the signal in Figure 5-1, we fill the data after 16000 points with zero, where there is only noise in that data segment. No particle signal is changed. Only the intermittency is created because the noise is cut off after the 16000th point. Then the data from 3000th point to 17000th point are chosen for the EMD decomposition, same data length as the signal used in Figure 5-2. The decomposition result is shown in Figure 5-3.

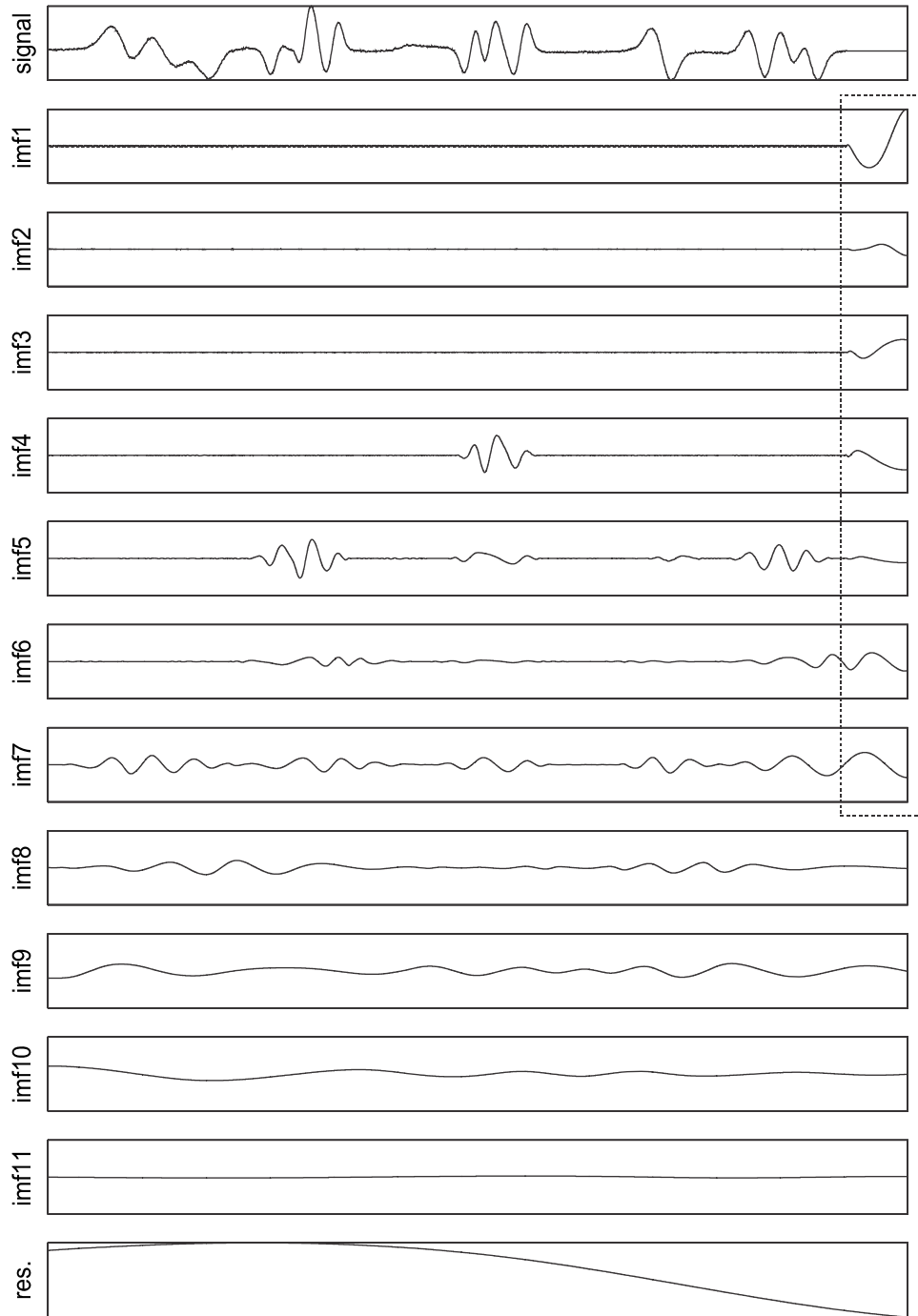
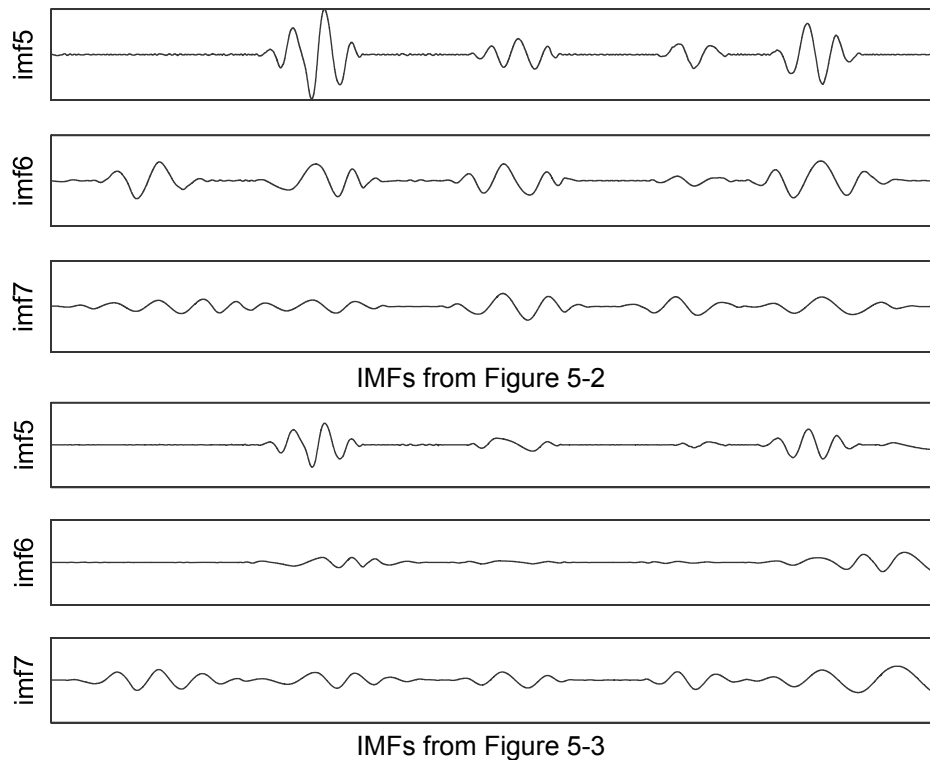


Figure 5-3 EMD decomposition result for the signal with intermittency

The first noticeable difference is the distorted wave at the end of each IMF. The distortions start from the point where the noise was removed. Theoretically, in that data segment, the amplitude of each IMF should be zero.

Furthermore, the influence of the intermittency is not just confined at the position where the intermittency exists. We can choose IMF5 to IMF7 from Figure 5-2 and Figure 5-3 to make a close comparison.



The comparison shows that the IMFs from two EMD decompositions are different. Not only the position where the intermittency occurred, but also the whole IMFs were changed. This is because the EMD sifting process is actually a series of iterations. The errors appeared at the end of IMF1 would participate in the following calculations. Following with the iterations, the data contamination expanded to all the IMFs and made the decomposition result lost its physical meaning.

In most case, the noise in a system is unpredictable and uncontrollable. For the implementation of EMD into particle detection, if the changes in the noise will totally distort the output of EMD, that will make the EMD method useless for this project. To overcome this problem, an improved EMD method has to be used to eliminate the influence of intermittency from the beginning. This is the reason why the EEMD method is introduced.

5.2.2 EEMD method

Comparing with EMD, EEMD method is not totally self adapting. Because EEMD introduces white noise into the decomposition process, there are two extra parameters need to be set before the data processing. One parameter is the amplitude of the white noise, another is the ensemble number.

Huang[21] suggested that the noise amplitude could be about 0.2~0.4 standard deviation of the amplitude of the raw data. For this project, many experiments were done and the conclusion is, 0.2 standard deviation of the amplitude of the raw data will be enough to generate a good result. Noise amplitude should be set to an appropriate value. Increase the noise amplitude will require more ensemble operations to reduce the influence of the white noise.

Ensemble number determines how many ensemble operations will be done in the decomposition process. In each ensemble operation, EEMD will add a series of white noise to the original data and perform the EMD decomposition. To

reduce the effect of the added white noise to a negligibly small level, a higher ensemble number will be better. However, increasing ensemble number will require more calculation time, reduce the efficiency of EEMD.

Some papers[21, 26] had discussed the value setting for white noise amplitude and ensemble number. According to these papers, usually an ensemble number of a few hundred could be enough to remove most of the added white noise. To reduce the effect of the added noise, one should first consider reducing the amplitude of the white noise, than increasing the ensemble number, because the amplitude of white noise has more influence on the decomposed result. For metal particle signals, after the comparison of decomposition results with different preset parameters, 0.2 standard deviation of the amplitude of raw data and 100 for the ensemble number are chosen for the EEMD decomposition. The EEMD result of the same signal in Figure 5-3 is shown as follows:

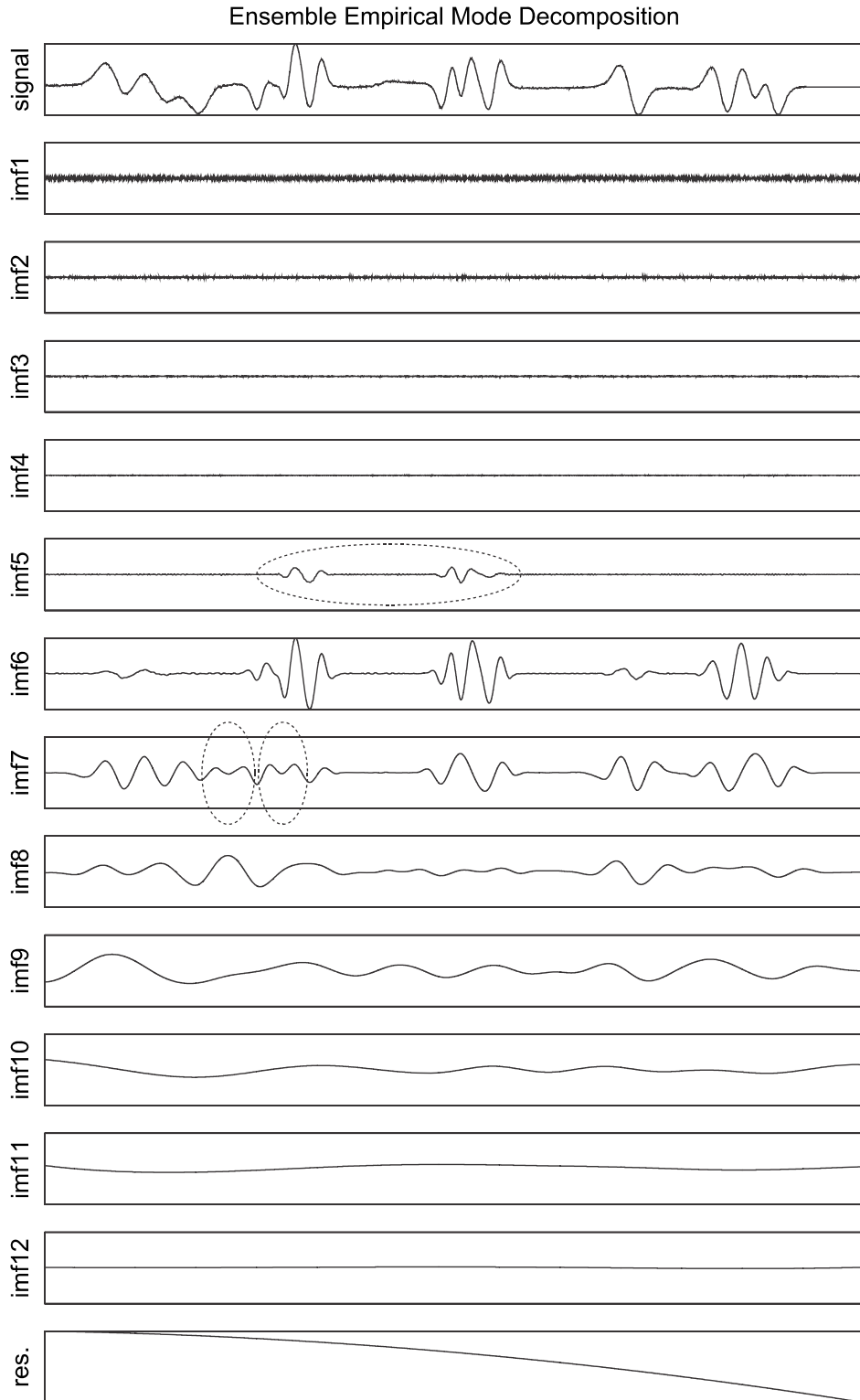


Figure 5-4 EEMD result of the signal in Figure 5-3

Comparing with Figure 5-3, the distortions at the end of first 7 IMFs are eliminated. In IMF7, If we count the maxima number as the number of particles, it is 15. However, if we count the minima number, it will be different. This problem will be discussed in the next section.

Basically, EEMD result is the mean of many EMD decompositions for the same signal plus different white noise series. Because EEMD need to calculate the ensemble mean of the IMFs to reduce the influence of the white noise, some changes need to make for the EMD in each ensemble operation.

Firstly, it is the IMF number. In original EMD method, sifting process is totally data adaptive. How many IMFs will be generated depends on the character of the original signal. In EEMD, the following process of calculating the ensemble mean of the IMFs requires each EMD operation must generate same number of IMFs. Wu et al.[21, 23] had studied the character of the white noise. After many experiments, they conclude that when EMD is used to decompose the white noise, the total number of IMFs is close to $\log_2 N$, in which N is the number of the total data points. When the data is not pure noise, it will have fewer scales than white noise. The number of generated IMFs might be fewer than $\log_2 N$.

Therefore, in EEMD the IMF number is set to $\lceil \log_2 N \rceil - 1$. For the signal in Figure 5-3, original signal contains 14000 data points. The EEMD IMF number is $\lceil \log_2 14000 \rceil - 1 = 12$.

The second change is the number of iterations for calculating each IMF. In EMD the stopping criteria for sifting process determine when the iteration should stop. Different stopping criteria were proposed and implemented in different researches. Those stopping criteria have a common problem; the decompositions using those criteria are sensitive to the local disturbance. Based on those stopping criteria, a local data change of the component during the sifting process may require more iterations. According to Huang's research[21], for some data which contain local high-frequency oscillations, more iterations will generate new local extrema, which will distort the final result. For this reason, Wu and Huang[23] proposed to fix the sifting number for the decomposition. They suggested that setting the number of iterations to 10 will be enough for most decompositions to keep the upper and lower envelopes of IMFs almost symmetric with respect to the zero line.

For the EEMD method used for this project, the number of iterations is set to 10. However, from the decomposition result in Figure 5-4, one can notice that some part of the IMF is not symmetric with respect to the zero line, for example, the wave part in the two dotted ellipses in IMF7. If the number of particles is counted manually, it is not difficult to pick the right extrema. If the counting work is performed by a program that based on some preset rules, this kind of asymmetry might be a problem. According to Huang, after the ensemble mean calculation of EEMD, the final results do not satisfy the strict definition of IMF, because there is no guarantee that the sum of IMFs will still be an IMF. To make the IMF more

symmetric, some experiments were done, different number of iterations from 10 to 15 was tested. However, the improvement was limited. Furthermore, in Figure 5-4 there still some mode mixing problem left, like the wave in the dotted ellipse in IMF5. For this drawback of EEMD, Wu and Huang[21] suggested that a post processing could be used to get a better result.

5.2.3 Post processing

The basic idea for post processing is simple. Since the results of EEMD do not satisfy the strict definition of IMF, they can be treated like signals, and be processed by EMD method. After the process of the original EMD method, the new IMFs will fit the strict definition of IMF. For the detail of the processing, considering the signal integrity, each IMF will need an EMD process. The post processing is developed as follows:

1. For the first IMF of EEMD decomposition result, IMF_1 will be processed by EMD, then get the new IMF, named C_1 , set $r_0=0$.
2. Calculate the residue $r_n = (IMF_n + r_{n-1}) - C_n$, $n=1, 2, 3, \dots$
3. Add r_n to $IMF_{(n+1)}$, then processed by EMD, get $C_{(n+1)}$.
4. Repeat step 2 to step 3, until all the IMFs are processed by EMD, the last calculated residue will be the trend.

The original signal $x(t)$ can be expressed as the sum of C_n and the trend:

$$x(t) = \sum_{i=1}^n C_n + trend$$

The post processing result of the EEMD output in Figure 5-4 is shown as follows:

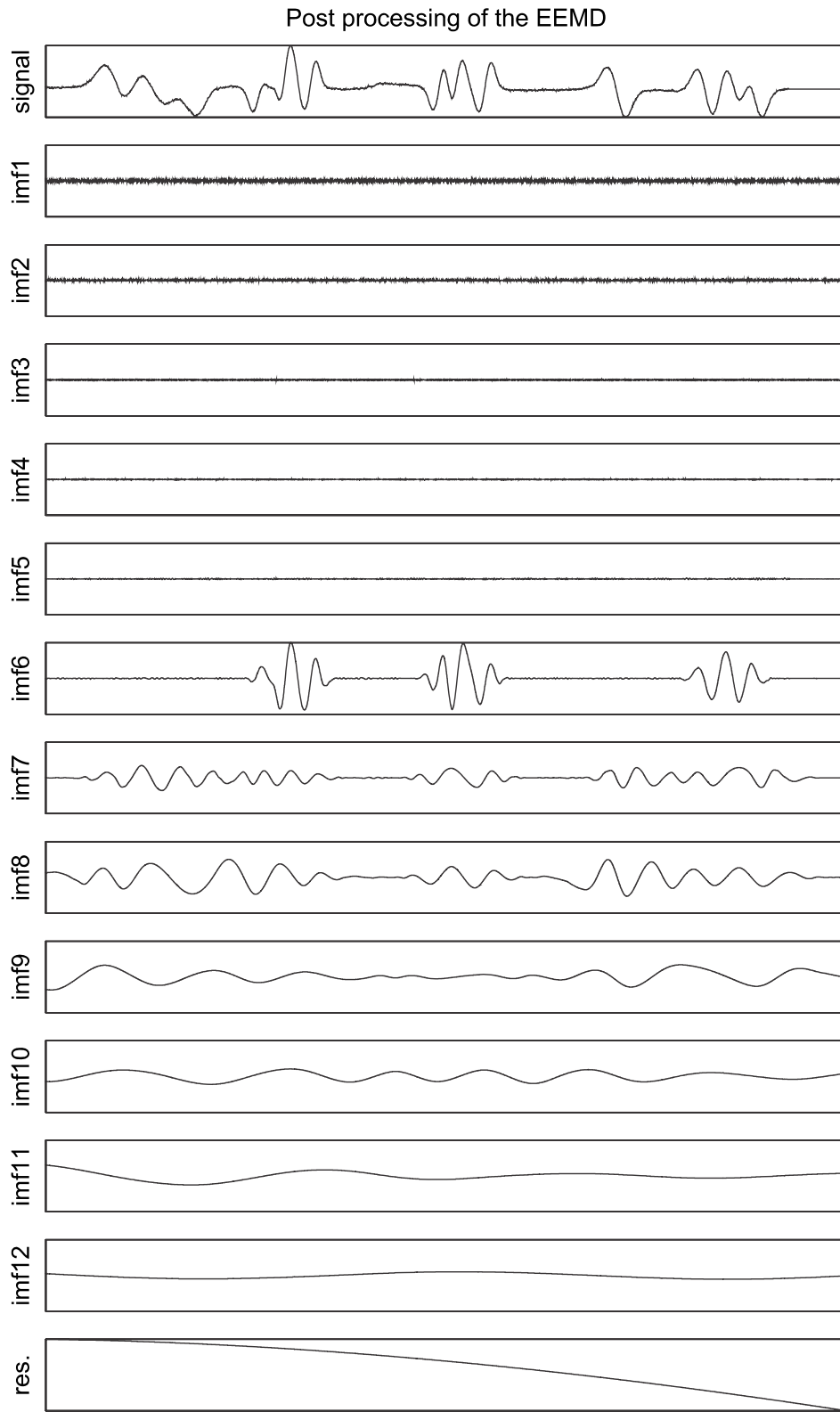


Figure 5-5 Post processing of the EEMD output in Figure 5-4

From Figure 5-5 we can see the IMFs are more symmetric with respect to the zero line, and the mode mixing in IMF5 is eliminated. For the IMF7, if we count the maxima as the number of particles, it is 15.

One of the reasons why EEMD does not include the post processing in the algorithm is the efficiency. For example, if the ensemble number is set to 100, the EEMD will need to perform 100 times EMD decomposition during the process. It will be much slower than the original EMD operation. If the post processing is implemented, for the data previously used in this chapter, there will be 12 more times of EMD operation. For the specific data, after analysis, if the post processing is necessary, it is still possible to reduce the calculation time. In this situation, EEMD will become the preconditioning for the following process. It is not necessary for EEMD to decompose the data into perfect IMFs. Then the number of iterations could be reduced to save time. To test the result of reducing the number of iterations, an experiment was done by using the particle data in Figure 5-4 which contains 14000 data points. The ensemble number is set to 100. The EEMD process will generate 12 IMFs. If the number of iterations for each IMF is changed from 10 to 8, the EEMD process will save $(10-8) \times 12 \times 100 = 2400$ iterations. After post processing, the final result is similar to the decomposition when the number of iterations is set to 10.

5.2.4 Other experiments

To test the stability of the EEMD method, many experiments were done with different number of particles in the constructed test signal. The decomposition results were shown in Table 1. Because the number of particles is manually counted, some small waveforms were not deemed as particle signals when their amplitudes are lower than 10% of the average amplitude of the decomposed signals. In Table 1, the column of exact number of particles lists the number of different particle signals used to construct the test signal in each experiment. The column of detected number of particles lists the manually counted number of particles in each decomposition result. The error percentage is used to measure the deviation between the detected result and the exact number of particles.

$$\text{Error percentage} = \left| \frac{\text{Detected number of particles} - \text{Exact number of particles}}{\text{Exact number of particles}} \right|$$

Table 1 EEMD Decomposition results of different number of particle signals

Experiment No.	Exact number of particles	Detected number of particles	Error percentage
1	12	12	0
2	14	15	7.1%
3	16	15	6.3%
4	18	19	5.6%
5	20	18	10%
6	22	20	9.1%
7	23	18	21.7%
8	24	19	20.8%
9	26	19	26.9%

From Table 1, one can notice that when the exact number of particles changed from 22 to 23, the error percentage changed from 9.1% to 21.7%, which means the decomposition result in experiment No.7 is not as reliable as experiment No.6. Because it will take 1.75 seconds for the data acquisition device to collect 14000 data points, 22 particle signals in 1.75 seconds means about 12 particles per second.

Chapter 6 Conclusions and future work

6.1 Conclusions

The goal of this thesis is to find a working algorithm to decompose the particle signal from the oil sensor. Then the number of particles could be found. By comparing the number of particles with the setting threshold, one can conclude that if the maintenance is necessary for the mechanical system. The difficulty for this project is, when multiple particles pass through the sensor in a short time interval, the particle signals will superimpose with each other. The characteristic waveform of a single particle signal cannot be found in the composed output data directly. Because the particle detection device will be used for different mechanical system, there is no more specific information for the particle detection device except the particle signal data itself.

Traditional signal analysis methods can provide the frequency distribution of a signal. They cannot directly separate a non-stationary signal into some components in time domain. Comparing with those methods, EMD is more fit for this project. It is intuitive and data adaptive, and more important, it can decompose a non-stationary signal into IMFs in time domain, by which the number of particles can be counted directly. As a new data analysis method, EMD still has some limitations, like lacking of theoretical support, end effect and

mode mixing etc. For this project, the problem is, EMD is sensitive to the intermittency in the input signal, even if the intermittency occurs in the noise part.

To deal with this problem, an improved EMD method, EEMD is used for signal decomposition. EEMD introduced white noise into the decomposition. White noise will compensate the intermittency in the signal, and it also provides a uniformly distributed reference scale for the decomposition. White noise's cancellation effect is used to remove the added white noise from the final result. EEMD performs many groups of EMD decompositions, and then calculates their ensemble mean. By this way most of the added white noise can be removed if the ensemble number is set to an appropriate value. However, this method is also the reason that EEMD is much slower than EMD. For example, when the particle data in this thesis was calculated on a Pentium 4 computer, to decompose the data in chapter 5, EMD needs 25 seconds, EEMD with ensemble number set to 100 needs more than 790 seconds. EEMD is more than 30 times slower than EMD. This speed makes implementing EEMD into on-line monitoring system impossible. Unfortunately, there is still no practical fast EMD or fast EEMD algorithm for the industry.

Also because of the calculation of ensemble mean in EEMD, the final results do not satisfy the strict definition of IMF. To get a better result, a post processing can be used on the EEMD output. Inevitably, EEMD with post processing also

has the computing speed problem. To calculate the previously mentioned same data, EEMD with post processing will need more than 850 seconds.

The Table 1 in the last part of Chapter 5 shows that EEMD also has its limitation. When the test data contains more than 22 particle signals, the error percentage doubled, which means the accuracy of the decomposition result dropped and the detected number of particles may be not reliable. This is mainly because the particle signals are so closed in the limited time interval that EEMD cannot distinguish them from each other. For the particles used in this thesis, in each second no more than 12 particles flow through the oil sensor will be the limitation of the EEMD algorithm.

6.2 Future work

As mentioned in Chapter 5, the number of particles is manually counted in each decomposition experiment. 10% of the average amplitude of the decomposed signals is used as a rough threshold to exclude small waveforms in the IMF. In the real engineering design, it is better to gather statistic data from different decomposition experiments on real particle signals, and find a more accurate threshold for different situations. Then a program can be used to count the number of particles based on the accurate threshold, which will increase the speed and accuracy of the system.

Furthermore, from the point of engineering, except a good algorithm, some other methods also can be use to improve the particle detection result. For example, because the mainly problem of this kind of particle detection is that particles stay too close with each other, some equipment can be used to shake the sample oil container to make particles evenly spread in the lubricating oil before the oil particle detection. For the even worse case, when there are too many particles in the sample oil that the particle detection algorithm does not work, some known volume clean lubricating oil can be used to dilute the sample oil and improve the detection condition.

Many people were first attracted by the result of EMD method. When EMD works on the right data, the result can be compared with the best outputs of the known algorithms. However, after people spending more time in the research of EMD, things become complicated. Just like its name described, EMD is an empirical method. More than ten years after it was first introduced, EMD still lack of theoretical background. The research about its characters and limitations are highly based on different experiments. There are still many questions without answers. Some people compare the current situation of EMD with the wavelet in 1980s, when wavelet still did not have a theoretical support. The principle of EMD is simple and intuitive. However, it seems that just the intuitive character of the EMD make it difficult to be described by mathematical theory. It is possible that there are some more effective theories behind the EMD method, and EMD is just some light of those theories projected out of a window.

References

- [1] Y. Yin, and et al., "An integrated on-line oil analysis method for condition monitoring," *Measurement Science and Technology*, vol. 14, no. 11, pp. 1973, 2003.
- [2] J. Miller, and D. Kitaljevich, "In-line oil debris monitor for aircraft engine condition assessment," in *Proc. 2000 IEEE Aerospace Conf.*, vol. 6, 2000.
- [3] X. Fan, M. Liang, and T. Yeap, "A joint time-invariant wavelet transform and kurtosis approach to the improvement of in-line oil debris sensor capability," *Smart Materials and Structures*, vol. 18 085010, no. 8, 2009.
- [4] J. P. Dron, F. Bolaers, and I. Rasolofondraibe, "Improvement of the sensitivity of the scalar indicators (crest factor, kurtosis) using a de-noising method by spectral subtraction: application to the detection of defects in ball bearings," *Journal of Sound and Vibration*, vol. 270, no. 1-2, pp. 61-73, 2004.
- [5] D. Gabor, "Theory of communication. Part 1: The analysis of information," *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, vol. 93, no. 26, pp. 429-441, 1946.
- [6] Y. Meyer, *Wavelets: Algorithms & applications*: Society for Industrial and Applied Mathematics (Philadelphia) 1993.
- [7] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674-693, 1989.
- [8] I. Daubechies, *Ten lectures on wavelets*, Philadelphia, Pa.: Society for Industrial and Applied Mathematics, 1992.
- [9] S. G. Mallat, *A wavelet tour of signal processing*, 2nd ed., San Diego: Academic Press, 1999.
- [10] N. Yen, "Wave packet decomposition," *The Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 889-896, 1994.
- [11] N. E. Huang, Z. Shen, S. R. Long *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London*, vol. 454, no. 1971, pp. 903-995, Mar 8, 1998.
- [12] N. E. Huang, and S. S. Shen, *Hilbert-Huang transform and its applications*, Singapore ; Hackensack, NJ ; London: World Scientific, 2005.
- [13] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: The Hilbert spectrum," *Annual Review of Fluid Mechanics*, vol. 31, pp. 417-457, 1999.
- [14] N. E. Huang, Z. H. Wu, S. R. Long *et al.*, "On instantaneous frequency," *Advances in Adaptive Data Analysis*, vol. 1, no. 2, pp. 177-229, 2009.
- [15] N. E. Huang, *An Introduction to Hilbert-Huang Transform: A Plea for Adaptive Data Analysis*, PowerPoint, RCADA, National Central University, Taiwan, 2006.

- [16] Z. Youming, and Q. Shuren, "Research on the uniform theoretical basis for Hilbert-Huang transform(HHT)," *Journal of vibration and shock*, vol. 25(3), pp. 40-43, 2006.
- [17] G. Rilling, P. Flandrin, and P. Goncalves, "On empirical mode decomposition and its algorithms," *In Proc. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, no. NSIL-03, Grado (I), 2003.
- [18] Y. J. DENG, W. WANG, and C. C. QIAN, "Boundary processing technique in EMD method and Hilbert transform," *Chinese Science Bulletin*, vol. 46(1), pp. 1-8, 2001.
- [19] H. T. LIU, M. ZHANG, and J. X. CHENG, "Dealing with the end issue of EMD based on orthogonal polynomial fitting algorithm," *Computer engineering and application*, vol. 40(16), pp. 84-86, 2004.
- [20] F. Wu, and L. Qu, "An improved method for restraining the end effect in empirical mode decomposition and its applications to the fault diagnosis of large rotating machinery," *Journal of Sound and Vibration*, vol. 314, no. 3-5, pp. 586-602, 2008.
- [21] Z. H. Wu, and N. E. Huang, "Ensemble Empirical Mode Decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, pp. 1-41, 2009.
- [22] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *Signal Processing Letters, IEEE*, vol. 11, no. 2, pp. 112-114, 2004.
- [23] Z. H. Wu, and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. Roy. Soc. London*, vol. 460, no. 2046, pp. 1597-1611, Jun 8, 2004.
- [24] N. E. Huang, M. L. C. Wu, S. R. Long *et al.*, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, vol. 459, no. 2037, pp. 2317-2345, Sep 8, 2003.
- [25] P. FLANDRIN, P. GONCALVES, and G. RILLING, "EMD Equivalent Filter Banks, from Interpretation to Applications," *Hilbert-Huang Transform and Its Applications*, pp. 57 -74, 2005.
- [26] G. Wei, and P. W. Tse, "Enhancing the ability of Ensemble Empirical Mode Decomposition in machine fault diagnosis," in *Prognostics and Health Management Conference, 2010. PHM '10.*, 2010, pp. 1-7.