
GAZİ ÜNİVERSİTESİ
Gazi
1926

MODE MIXING PROBLEM IN EMPIRICAL MODE DECOMPOSITION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
GAZİ UNIVERSITY

BY
Muhammed I. JAITEH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

DECEMBER 2018

The thesis study titled "MODE MIXING PROBLEM IN EMPIRICAL MODE DECOMPOSITION" is submitted by Muhammed I. JAITEH in partial fulfillment of the requirements for the degree of Master of Science in the Department of Statistics, Gazi University by the following committee.

Supervisor : Prof. Dr. M. Akif BAKIR
Department of Statistics, Gazi University
I certify that this thesis is a graduate thesis in terms of quality and content .....

Chairman : Prof. Dr. Yılmaz AKDİ
Department of Statistics, Ankara University
I certify that this thesis is a graduate thesis in terms of quality and content .....

Member : Assoc. Prof. Dr. Filiz KARDİYEN
Department of Statistics, Gazi University
I certify that this thesis is a graduate thesis in terms of quality and content .....

Thesis Defence Date : 21/12/2018

I certify that this thesis, accepted by the committee, meets the requirements for being a Master of Science Thesis.

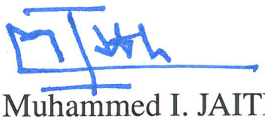
.....
Prof.Dr. Sena YAŞYERLİ
Dean of Graduate School of Natural and Applied Sciences

ETHICAL STATEMENT

I hereby declare that in this thesis study I prepared in accordance with thesis writing rules of Gazi University Graduate School of Natural and Applied Sciences;

- All data, information and documents presented in this thesis have been obtained within the scope of academic rules and ethical conduct,
- All information, documents, assessments and results have been presented in accordance with scientific ethical conduct and moral rules,
- All material used in this thesis that are not original to this work have been fully cited and referenced,
- No change has been made in the data used,
- The work presented in this thesis is original,

or else, I admit all loss of rights to be incurred against me.



Muhammed I. JAITEH

21/12/2018

MODE MIXING PROBLEM IN EMPIRICAL MODE DECOMPOSITION

(M.Sc. Thesis)

Muhammed I. JAITEH

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

December 2018

ABSTRACT

Empirical Mode Decomposition (EMD) has been broadly used to analyse nonlinear and nonstationary time series data in a wide range of fields in the recent years. Due to the adaptive nature of EMD and under the assumption that every data consist of oscillations of intrinsic mode functions; EMD decomposes nonlinear and nonstationary data into series of intrinsic mode functions (IMFs) via the sifting process in order to achieve reliable and credible analysis. Despite its sophistication, mode mixing problem occurs during the decomposition process. The issue of mode mixing is remedied by an adjusted noise assisted algorithm of the original EMD algorithm known as the Ensemble Empirical Mode Decomposition (EEMD). However, at the end of the process of EEMD, the added noise may not be entirely separated from the signal or data leading to spurious results of the analysis. Hence, the Complete Ensemble Empirical Mode Decomposition with Adaptive noise (CEEMDAN) is applied so as to completely segregate the noise from the signal. Furthermore, in the event of no prior knowledge of the signal, realisation of existence of mode mixing in an IMF without implementing any mode mixing detection test is a daunting challenge or near impossible. Therefore, in this thesis, the main focus will be on exploring the core concept of mode mixing in EMD as well as the solution to it and its methods of detection in an IMF.

Science Code	: 20514
Keywords	: Empirical Mode Decomposition, Mode Mixing, Time Series, Signal Processing
Number of pages	: 79
Supervisor	: Prof. Dr. M. Akif BAKIR

DENEYSEL MOD AYRIŞTIRMASINDA MOD KARIŞIKLIĞI PROBLEMİ

(Yüksek Lisans Tezi)

Muhammed I. JAITEH

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Aralık 2018

ÖZET

Deneysel Mod Ayrıştırması (EMD), bir çok alanda doğrusal olmayan ve durağan olmayan zaman serisi verilerini analiz etmek için özellikle son yıllarda yaygın olarak kullanılmıştır. EMD uyarlanabilir doğası gereği zaman serisi verisinin içsel (intrinsic) mod fonksiyonları salınımlarından oluştuğu varsayımı altında; EMD, doğrusal olmayan ve durağan olmayan verileri güvenilir bir şekilde analiz etmek için kaydırma işlemi (sifting process) yoluyla zaman serisini içsel mod fonksiyonlarına (IMF'ler) ayrıştırır. Oldukça gelişkin bir yöntem olmasına rağmen, EMD ile mod ayrıştırmasında mod karışımı problemi ortaya çıkabilmektedir. Bu problemin çözümüne ilişkin orijinal EMD algoritmasının uyarlanmış hali olan Ensemble Deneysel Mod Ayrıştırması (EEMD) yöntemi geliştirilmiştir. Ancak EEMD sürecinin sonunda da seriye eklenmiş olan gürültü, sinyalden tamamıyla ayrılmayarak yanlış sonuçlara yol açabilir. Böylece gürültüyü sinyalden tamamıyla ayrıştırmak için, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) yöntemi uygulanmaktadır. Ayrıca, sinyal hakkında önsel bilgi olmaması durumunda, bir IMF' de mod karışması probleminin var olup olmadığını herhangi bir formel yöntemde uygulamaksızın belirlemek zor ya da hemen hemen mümkün değildir. Bu nedenle, bu tezde temel hedef, EMD' nin çekirdek kavramı olan mod karışımını ele almak ve IMF' deki mod karışmasının varlığını belirleyen yöntemleri çalışmaktır.

Bilim Kodu : 20514
Anahtar Kelimeler : Deneysel Mod Ayrıştırması, Mod Karışıklığı
Zaman Serileri, Sinyal İşleme
Sayfa Adedi : 79
Danışman : Prof. Dr. M. Akif BAKIR

ACKNOWLEDGEMENTS

I will extend my sincere gratitude to my entire family more so to my wonderful mother Mrs Isatou Sanneh and my dearest amazingly gorgeous wife Dr Maimuna Danso Jaiteh for their understanding and support throughout my study period.Words cannot fully describe how thankful and appreciative of every help my supervisor Prof. Dr. Mehmet Akif BAKIR has rendered me from the very first day I met him. I had so many fruitful discussions with him and he has availed me the platform to freely speak my mind and come up with my own ideas for this thesis. Not only has he been my supervisor but a fatherly figure for me as well. I have learnt so many great ideas and values from him that will stay with me for the rest of my academic career in future and life in general. I have made lifelong friends from both international and local students, particularly Mr. Atacan ERDİŞ (Turkish), the level of support Mr. Atacan has given me is immeasurable. Despite his busy schedule, anytime I knock on his door he was ever ready to help me. I definitely cannot thank him enough. "Atacan bey, sana ne kadar teşekkür etsem azdır". My deepest thanks to Mr Ammar Homaida (Syrian) and Mr Muctarr Sonko (Gambian) for all the support. You guys really rock. My heartfelt appreciation goes to the entire staff of Statistics department of Gazi University. I will happily say definitely “Gazili olmak ayrıcalıktır”.

TABLE OF CONTENTS

Page

ABSTRACT iv

ÖZET v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS vii

LIST OF TABLES ix

LIST OF FIGURES x

SYMBOLS AND ABBREVIATIONS xii

1. INTRODUCTION 1

2. BASIC CONCEPTS AND DEFINITIONS 5

2.1. Hilbert – Huang Transform (HHT) 5

2.1.1. Hilbert spectral analysis (HSA) 6

2.1.2. Empirical mode decomposition method (The sifting process) 7

2.2. The Disagreement Between The Implementation Of EMD Algorithm And The Definition Of IMF 12

2.3. Statistical Significance of IMFs 16

3. MODE MIXING PROBLEM IN EMPIRICAL MODE DECOMPOSITION 21

3.1. Other Algorithms (Approaches) For Empirical Mode Decomposition 22

3.1.1. Ensemble empirical mode decomposition (EEMD) 22

3.1.2. Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) 23

4. APLICATION TO SIMULATED AND REAL DATA FOR MODE MIXING 25

4.1. Simulated Data Set in Matlab Without Mode Mixing Problem 25

4.2. Simulated Data Set in Matlab With Mode Mixing Problem 27

4.3. EMD Study in Matlab 28

Page

4.4. EMD Study in R 29

4.5. Detection of Mode Mixing in IMFs 31

4.5.1. Cross-correlation 32

4.5.2. Instantaneous frequency (IF) 35

4.5.3. Detecting the problem of mode mixing authomatically via the use of nonstationary detection method: application to selecting IMFs of interest and denoising 35

4.6. Illustrations of Detection of Mode Mixing Problem 38

4.7. Application of EMD to Real Datasets 42

4.7.1. Data set of total export of consumer goods from Turkey 42

4.7.2. Data set of US dollar – Turkish lira exchange rate 46

4.8. Determination of Mode Mixing Problem by Signal Masking Method 50

4.8.1. Applications to simulated data 54

4.8.2. Application to real data set 65

5. DISCUSSION AND CONCLUSION 73

REFERENCES 75

CURRICULUM VITAE 78

LIST OF TABLES

Table **Page**

Table 4.1. Values of parameters to estimate f_2/f_1 in Simulation 1 60

Table 4.2. Values of parameters to estimate f_2/f_1 in Simulation 2 65

Table 4.3. Values of parameters to estimate f_2/f_1 in Real Data Set 71

LIST OF FIGURES

Figure **Page**

Figure 2.1. Demonstration of an IMF's upper and lower envelopes with their extrema (Wang et al., 2010) 15

Figure 2.2. Flow chart of EMD sifting process (Crowley, 2009)..... 16

Figure 4.1. Flow chart of simulated datasets 25

Figure 4.2. No mode mixing problem..... 26

Figure 4.3. Empirical mode decomposition 27

Figure 4.4. Simulated data set and illustration of mode mixing problem..... 28

Figure 4.5. Empirical mode decomposition 29

Figure 4.6. Empirical mode decomposition 30

Figure 4.7. Ensemble empirical mode decomposition..... 31

Figure 4.8. IMFs extracted from the EMD of original data in Figure 4.5 39

Figure 4.9. Cross correlation of IMF1 and IMF2 in Figure 4.8 40

Figure 4.10. Cross correlation of IMF1 and IMF3 in Figure 4.8 40

Figure 4.11. Cross correlation of IMF2 and IMF3 in Figure 4.8 41

Figure 4.12. Instantaneous frequency of IMF1, IMF2 and IMF3 in Figure 4.8 42

Figure 4.13. Total export of consumer goods from Turkey..... 43

Figure 4.14. Empirical mode decomposition of data in Figure 4.13 44

Figure 4.15. Cross-correlations of IMFs in Figure 4.14 45

Figure 4.16. Instantaneous frequencies of IMFs in Figure 4.14 46

Figure 4.17. US dollar – Turkish lira exchange rate 47

Figure 4.18. Empirical mode decomposition of data in figure 4.17..... 48

Figure 4.19. Cross-correlations of IMFs in figure 4.18 49

Figure 4.20. Instantaneous frequencies of IMFs in figure 4.18 50

Figure 4.21. Vision Algorithm Flowchart..... 53

Figure 4.22. Boundary Map..... 54

Figure **Page**

Figure 4.23. Flow chart of algorithm and simulation procedure 55

Figure 4.24. Original signal, mixed signal, masking signal and masked signals 56

Figure 4.25. EMD for positive masked signal 57

Figure 4.26. EMD for negative masked signal 57

Figure 4.27. Instantaneous frequencies obtained from AIMFs 58

Figure 4.28. Hilbert envolops for AIMFs 58

Figure 4.29. Instantaneous amplitude obtained from AIMFs 59

Figure 4.30. Instantaneous amplitudes obtained from AIMFs 59

Figure 4.31. Instantaneous frequencies obtained from AIMFs 60

Figure 4.32. Original signal, mixed signal, masking signal and masked signals 61

Figure 4.33. EMD for positive masked signal 62

Figure 4.34. EMD for negative masked signal 62

Figure 4.35. Instantaneous frequencies obtained from AIMFs 63

Figure 4.36. Instantaneous amplitude obtained from AIMFs 63

Figure 4.37. Instantaneous amplitudes obtained from AIMFs 64

Figure 4.38. Instantaneous frequencies obtained from AIMFs 64

Figure 4.39. Original signal, mixed signal, masking signal and masked signals 66

Figure 4.40. EMD for positive masked signal 66

Figure 4.41. EMD for negative masked signal 67

Figure 4.42. Instantaneous frequencies obtained from AIMFs 67

Figure 4.43. Hilbert envolops for AIMFs 68

Figure 4.44. Instantaneous amplitude obtained from AIMFs 68

Figure 4.45. Instantaneous amplitudes obtained from AIMFs 69

Figure 4.46. Instantaneous frequencies obtained from AIMFs 69

Figure 4.47. Cross correlation between AIMFs 70

SYMBOLS AND ABBREVIATIONS

The symbols and abbreviations used in this study are given below with their explanations.

Symbols	Explanation
α	Alpha
β	Beta
ω	Omega
θ	Theta
ρ	Rho
τ	Tau
ψ	Phi

Abbreviations	Explanation
ACF	Autocorrelation Function
BCI	Brain Computer Interface
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
EWT	Empirical Wavelet Transform
FEMD	Fast Empirical Mode Decomposition
HHT	Hilbert-Huang transform
HSA	Hilbert Spectral Analysis
IA	Instantaneous Amplitute
IF	Instantaneous Frequency
IMF	Intrinsic Mode Function

1. INTRODUCTION

Traditionally, methods of time series data analysis are mostly based on the assumptions of linearity and stationarity. Thus, there is familiarity with methods which are designed to analyse linear but nonstationary data (wavelet analysis and Wagner–Ville distribution), as well as nonlinear time series analysis methods built for nonlinear but stationary data (Huang and Shen, 2005).

However, most of the practical processes in real life are complicated encompassing many components. These components interact with each other and alter the regularities needed to be identified. Most of these real life processes entail both nonlinearity and nonstationarity. Analysing data of such nature is a daunting problem. Recently, a method called the Hilbert – Huang transform (HHT) designed by Huang et al. (1996, 1998, 1999) is able to handle some of the challenges posed by the analysis of nonlinear and nonstationary data. HHT shows the real physical meaning of a lot of nonlinear and nonstationary data. It should be emphasised that this method has been validated only empirically.

Despite the fact that it has effective applications in diverse fields, EMD has its own limitations. One of its limitations or challenges is the issue of mode mixing.

Related works

Huang et al. (1998) transformed the method of analysing nonlinear and nonstationary data by introducing a new adaptive method, the most pivotal part of which was the Empirical Mode Decomposition (EMD). This proposed method was an efficient way of decomposing complicated data into series of functions known as the intrinsic mode functions (IMF). The IMFs produced instantaneous frequencies (IFs) with the help of Hilbert transform. The IFs clearly visualise hidden properties. They pointed out that the core concept of this method was the small and finite components called the IMFs based on the local characteristics of the data, which leads to the meaningfulness of the instantaneous frequency. They used data representing natural process and numerical results to provide evidence for the efficiency and power of this method.

Verma (2016) pointed out that although EMD is use in diverse fields it has its own limitation when the frequency and amplitude ratios are very close together. He reviewed the basis of mode mixing from the available resources at the time and came up with Empirical Wavelet Transform (EWT) as a solution to mode mixing problem of two harmonic decomposition. Verma (2016) used numerical examples to establish a comparison between EWT and EMD. He came up a conclusion that the results obtained from EWT provided evidence that EWT detected valid components where EMD experienced mode mixing.

Fosso and Molinas (2017) based their work on the research done by Rilling and Flandrin (2008) where Rilling and Flandrin introduced a solid mathematical analysis that clarifies how EMD turns out in the event of a composite two-tones signal, also frequency and amplitude proportion by which EMD accomplish a wonderful segregation of tones. Fosso and Molinas did not suggest a solution for isolating neighbouring tones that will normally stay blended after the EMD process. Based on the discoveries of Rilling and Flandrin, they introduced a technique that has the ability to isolate neighbouring spectral components that would normally stay in a one IMF. Fosso and Molinas's technique is based on turning around the conditions by which mode mixing takes place according to Rilling and Flandrin. Fosso and Molinas conducted numerical experiments with signal within which was closed spaced spectral components to prove the efficiency of their method.

Xu, Zhengxiang and Wang (2016) indicated that the desired result is not always achieved EMD due to mode mixing problem. They conducted research on the issue of mode mixing and came up with two factors responsible for causing mode mixing during the sifting process of EMD. These two factors are: (1) the jumping change of the decomposed signal result and (2) mode mixing caused by amplitude and frequency relationship of the signal.

Gupta, Kumar, Chakraborti and Sharma (2016) applied EMD and two more of its adjusted form, EEMD and CEEMDAN on electroencephalograph signal for Mental Task Classification (MTC) based Brain Computer Interface (BCI). The MTC is divided into single task and multi task classification with the help of support vector machine. They decomposed signal into phases. The first phase consisted of decomposition into disparate oscillatory function using all the three algorithms. In the second phase, eight parameters were evaluated for every function. Based on the results obtained from their experiment,

they concluded that EEMD was a better algorithm.

In this thesis upon exploring the fundamental concepts of Empirical Mode Decomposition (EMD) in Chapter 2, the issue of mode mixing will be delve into as well as the algorithms used to overcome the problem of mode mixing in Chapter 3. In Chapter 4 via simulation study in both R and Matlab the concept of mode mixing will be visualised. Furthermore, in real datasets it might be extremely challenging to detect mode mixing in an IMF without performing any detection test. Hence, methods of detecting mode mixing in an IMF will also be looked into and EMD will be implemented on a couple of real datasets in Chapter 4. Then discussion where recommendation is made for a future possible research study and conclusion will follow in Chapters 5.

2. BASIC CONCEPTS AND DEFINITIONS

2.1. Hilbert – Huang Transform (HHT)

HHT is a combination of both EMD and HSA. Thus, it is a more sophisticated tool for analysis of nonlinear and nonstationary data. Physical natural processes are most of the time nonlinear and nonstationary but yet most data analysis methods are not effective with those kinds of processes. A common property of nonlinear processes is the intra–wave frequency modulation, which shows the instantaneous frequency (IF) within a cycle of one oscillation (Huang and Shen, 2005). The physically meaningful way of explicitly describing the dynamics of a nonlinear and nonstationary process or data should be by the instantaneous frequency obtained from the data.

Hilbert–Huang transform (HHT) is the easiest way to determine instantaneous frequency, via which the conjugate complex $y(t)$ of any real valued function $x(t)$ of class L^p can be computed by:

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{2.1}$$

where P indicates the principal value of the singular integral. With the Hilbert transform, the analytic function is then:

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)} \tag{2.2}$$

where

$$a(t) = (x^2 + y^2)^{1/2}; \theta(t) = \arctan(y/x) \tag{2.3}$$

Here a is the instantaneous amplitude, and θ is the phase function; and the instantaneous frequency is simply:

$$\omega = -\frac{d\theta}{dt} \tag{2.4}$$

The instantaneous frequency is dependent on the derivative of the phase function which does not only describe the interwave frequency changes because of nonstationarity but also describes the intrawave frequency modulation because of nonlinearity (Huang and Shen, 2005).

2.1.1. Hilbert spectral analysis (HSA)

Once the IMF components have been obtained from the data it will be less daunting to apply HHT on each of the components so as to compute the instantaneous frequency. After HHT has been performed on each of the components, the data is expressible as real part $\Re$ as:

$$X(t) = \Re\left\{\sum_{j=1}^{n} a_j(t)\exp\left[i\int \omega_j(t)dt\right]\right\} \tag{2.5}$$

where a and ω are amplitude and frequency respectively and are both functions of time. In this case the residue r_n is eschewed because it is either a constant or a monotonic function.

When the same data is expanded in Fourier format:

$$X(t) = \Re\left[\sum_{j=1}^{n} a_j e^{i\omega_j(t)t}\right] \tag{2.6}$$

Where a_j and ω_j are constants. The variable amplitude and the instantaneous frequency do not only improve the efficiency expansion but also to a great extend enables the accommodation of nonlinearity and nonstationarity expansion. The modulations of the frequency and the amplitude are explicitly separated with the IMF. Hence, restrictions with regards to constant amplitude and constants frequencies in Fourier expansion are settled with amplitude and frequency which are variable. This frequency–time distribution of the amplitude is known as the Hilbert amplitude spectrum (Hilbert spectrum), $H(\omega,t)$

From the Hilbert spectrum, the marginal spectrum $h(\omega)$ is:

$$h(\omega) = \int_0^T H(\omega,t)dt \tag{2.7}$$

The marginal spectrum gives the magnitude of the entire amplitude contribution for each frequency value.

2.1.2. Empirical mode decomposition method (The sifting process)

Empirical Mode Decomposition (EMD) is the pivotal part of HHT. EMD is a necessary method to decompose nonlinear and nonstationary data for analysis and credible forecasting. EMD is adaptive, meaning it has no specific defined mathematical function but instead it has a posteriori-defined basis derived from the decomposition of the data. Succinctly, during the data decomposition process certain functions are kind of artificially formed which were not part of the data initially while preserving the original characteristics of the data. The basis of EMD is the assumption that every data comprises oscillations of intrinsic mode functions (IMF). Every intrinsic mode be it linear or nonlinear is a representation of simple oscillation, which have the same number of extrema and zero–crossing. The data may entail different superimposing modes of oscillation at a particular given time. Each of these modes of oscillation is represented by a function known as the intrinsic mode function (IMF).

Intrinsic mode function (IMF)

In order to have a clear grasp of the characteristics of nonlinear and nonstationary data, it is expanded into a complete set of intrinsic mode functions. These functions are separately analysed and collectively they give an in-depth representation of the characteristics or embedded structures of the dataset.

For a function to be qualified to be referred to as an intrinsic mode function, it must satisfy the following conditions:

1. In the entire dataset, the number of extrema and the number of zero–crossings must be either equal or differ by at most one.
2. At any point, the mean value of the envelop defined by the local maxima and the envelop defined by the local minima is zero.

These conditions are trying to restrict an IMF to only one oscillatory mode at any given time and making so that oscillations are symmetric with respect to the time axis.

The amplitude and frequency of an IMF is variable with respect to time.

The procedure of extracting intrinsic mode functions from a dataset is known as the sifting process. The process has two main objectives: (1) to get rid of riding waves so as to get meaningful instantaneous frequency for HHT and (2) ensuring symmetry is achieved in wave profiles in case there is significant disparity in wave amplitude with neighbouring wave.

The sifting process is carried out as follows:

- Locate all the local maxima in the entire dataset and connect them all by a cubic spline called the upper envelop.
- Locate all the local minima in the entire dataset and connect them all by a cubic spline called the lower envelop.

The mean of the envelops (upper and lower) is denoted by m_1. h_1 is the first component which is the difference between the data $X(t)$ and m_1:

$$X(t) - m_1 = h_1 \tag{2.8}$$

The sifting process is iterative, hence it can be repeated a number of times until a stoppage criterion is met.

$h_1$must satisfy the conditions of an IMF. In the following sifting process, $h_1$is treated as the data.

$$h_1 - m_{11} = h_{11} \tag{2.9}$$

After a k number of repeating the sifting process in this way, h_{1k}becomes an IMF, meaning:

$$h_{1(k-1)} - m_{1k} = h_{1k} \tag{2.10}$$

Then it is denoted as:

$$c_1 = h_{1k} \tag{2.11}$$

Stoppage criteria of the sifting process

Squared difference:

This criterion requires that the normalised squared difference of two consecutive sifting processes computed by

$$SD_k = \sum_{t=0}^{T} \frac{|h_{k-1}(t) - h_k(t)|^2}{h_{k-1}^2(t)} \tag{2.12}$$

to be small. The sifting process stops if the squared difference,SD_k, is smaller than the previously computed value. However, there are two problems associated with this criterion. First, what value or what interval of values of SD_k is small enough to determine a reliable stopping point. Second, this criterion is not dependent on the definition of intrinsic mode functions. Meaning the SD_k may be small but there is no guarantee that the number of zero–crossing and extrema will be the same. Due to these problems Huang et al. (1999, 2003) came up with a second criterion called the S–number.

S – number

The agreement between the number of zero–crossing and extrema is the basis of this criterion. The S–number is preselected. The sifting process stops only if after S consecutive number of the process the number of zero–crossing and extrema remains the same or at most differ by one. However, this process has its own challenge which is determining the S–number. The S–number selection is ad hoc and there should be justification of the selected S–number instead of a trial and error.

In order to provide a remedy for the challenge post by S –number selection, Huang et al. (2003) by using a lot of possible choices formulated a set of IMF ensemble. From this set the mean and confidence of the ensemble was established. Via thorough comparisons with individual means an empirical guide was formulated. The range of S–number should be from 4 to 8 for an optimal sifting process.

None of the stoppage criterion is completely adequate but they will each serve the purpose

if applied accordingly. The issue is that there is no thorough mathematic standard based on which a decision can be taken. Selection of the stoppage criterion so as to produce physically meaning IMF still remains a challenging objective to reach in the implementation of EMD. Currently, all the stoppage criteria are executed with some degree of fastidiousness and to satisfy the definition of IMF only roughly. Undoubtedly, the fastidiousness, in the stoppage criteria emerges from the conflict between the definition of IMF and the currently used EMD implementation algorithm.

Once the stoppage criterion has been established and the first component c_1 was obtained, c_1 can then be separated from the rest of the data to obtain the residue r_1, $X(t) - c_1 = r_1$. But still $r_1$entails longer period variations in the data and hence is regarded as the new data in the subsequent sifting process and is an iterative process.

This procedure can be repeated for all the subsequent $r_j's$ and the result is

$$r_{n-1} - c_n = r_n \tag{2.13}$$

The sifting is stopped once the component c_n or the residue r_n becomes less than the predetermined value or when the residue becomes a monotonic function meaning no IMF can be further extracted from it. Summing up the components and the residuals gives:

$$X(t) = \sum_{j=1}^{n} c_j + r_n \tag{2.14}$$

Therefore, the decomposition of the data into n IMFs and residue, r_n, is achieved. Nonetheless, for excessively large datasets EMD is most likely not a suitable technique for the decomposition process due to its slow speed in decomposing such datasets. Hence, Fast Empirical Mode Decomposition (FEMD) can be implemented in such a situation.

Fast empirical mode decomposition (FEMD)

Despite the enormous success registered by EMD method over the years it lacks the speed of traditional methods like Wavelet and Fourier transform which are standard in the engineering, mathematics and statistics sector partly due to their fast transform algorithm

implementation. Hence, a new computational algorithm has been introduced called FEMD. The main motivation for FEMD is its fast computational speed (Blakely, 2005).

EMD works on the assumption that the data consist of unique oscillation of intrinsic modes thereby creating the need to solve systems of equations which are well conditioned at each stage. On the contrary, FEMD operates on the concept of the theory of reproducing kernel constructed in a moving least square. This new computational algorithm, FEMD, is completely matrix-free; therefore no linear systems need to be solved.

The main reason for FEMD branches off the fact that for a large time series data say $\bigcirc(10^3)$, the process of cubic spline interpolation to construct the envelopes can be computationally extremely daunting because a solution is necessary for a linear system at every sifting iteration. For the FEMD algorithm–free moving least square (MLS) method is applied at every sifting iteration producing kernel to the extrema of the time series, which is a linear operation $\bigcirc(K)$, where K denotes number of evaluation points of the envelopes. Despite the speed and accuracy of the FEMD method the weight function $W(x_i, x, \beta)$ of the kernels and its shape parameter β need to be chosen with great care. When these two items are proper chosen the resulting interpolation of the resulting with the reproducing kernels will be very close to the interpolation with the cubic spline (Blakely, 2005).

The FEMD algorithm is:

Initiate Set $n := 0,\ f_0 = f(t)$

Steps:

1. Set $h_0 := f_n$ and $k := 0$
2. Construct lower and upper envelopes for h_k
   - Identify all $N_{\max}$ local maxima $h_{k,i}^{\max}$ and $N_{\min}$ local minima $h_{k,i}^{\min}$ $t_i, t_j \in [0, T]$, of h_k.
   - Construct a sub-grid on $[0,1]$ of $N_{\max}$ uniformly spaced points and another sub-grid on $[0,1]$ of $N_{\min}$ uniformly spaced points while normalised values of $h_{\max}(t_i)$ and $h_{\min}(t_j)$ on their respective sub-grids.
   - If number of extrema for iteration K is different from $K-1$, construct new matrix-

free MLS reproducing kernels $\psi_i^{\max}(t)$, $\psi_j^{\min}(t)$ with respect to the two sub-grids of $N_{\max}$and $N_{\min}$ points. Evaluate the upper and lower envelopes $U(t)$ and $L(t)$ by applying the reproducing kernels as:

$$U(t) = \sum_{i=0}^{N_{\max}} h_{k,i}^{\max} \psi_i^{\max}(t) \ , \ \ t \in [0,1] \tag{2.15}$$

and

$$L(t) = \sum_{i=0}^{N_{\min}} h_{k,i}^{\min} \psi_i^{\min}(t) \ , \ \ t \in [0,1] \tag{2.16}$$

- Rescale uniform grid points back onto $[0,T]$ along with upper and lower envelopes $U(t)$ and $L(t)$.

3. Sifting process

   Since $L(t) \leq f_k(t) \leq U(t)$ the mean of the envelope can be defined as $m_k(t)$ The $K+1th$ component is then defined to be $h_{k+1} := h_k - m_k$. Next, h_{k+1} is checked for IMF criteria. If h_{k+1} does not meet IMF criteria, increment k and repeat from the second step (step 2). Else, set $IMF_n := h_{k+1}$ and set $f_{n+1} := f_n - IMF_n$ and reconstruct producing kernels $\psi_i^{\max}(t), \psi_j^{\min}(t)$. When the convergence criteria is satisfied for f_{n+1} stop. Or else, increment n and return to step 1.

2.2. The Disagreement Between The Implementation Of EMD Algorithm And The Definition Of IMF

The basis of the execution of EMD is the cubic natural spline due to its adaptability and simplicity. At the same time, it has been clearly understood that the definition of IMF could not be thoroughly realized: for a big number of sifting will create an IMF better adhering to IMF definition but such a function will not be physically meaningful (Wang, Chen, Qiao, Wu, and Huang, 2010). The real number of sifting process is chimeric. Hence, the stoppage criteria are to some degree fastidious. This predicament makes placing the EMD on a thoroughly mathematical establishment very daunting, if at all possible. We will demonstrate that this predicament arises from the conflict between the definition of IMF and the algorithm based on spline fitting. Therefore, the fastidiousness will have to be something that we will deal with until the spline implementation is adjusted. First, we start with the cubic natural algorithm as EMD is right now implemented. Now we will

demonstrate the following theorem that is the core of the conflict:

> *For any function with sparsely populated extrema, the upper envelop $u(t)$ and lower envelop $l(t)$, constructed according to the EMD sifting procedures, have to degenerate to a pair of symmetric straight lines.*

The proof will come in two folds: The first step is to set up that the envelop will have to be a unitary cubic spline and the second step is to set up that the unitary spline will degenerate to a straight line. $l_0(t)$

The first step is given as follows: Assuming that we have an IMF, $c(t)$, extracted from a signal $x(t)$ through the envelopes $u(t)$ and $l(t)$ and the mean $m(t)$. Suppose $c(t)$ strictly satisfies the definition of IMF, then

$$u(t) = -l(t) \tag{2.17}$$

Without generality loss, in the event that we make the assumption that the extrema of the IMF are sparsely distributed with regard to the data points, ie between the maximum and its neighbouring minimum there are at least two data points. Say we have a situation where t_1, t_3, t_5 are locations of the maxima of $c(t)$, and t_2, t_4 are locations of the minima; $u_1(t)$ and $u_2(t)$ are portions of $u(t)$ and $l_0(t)$ is also a portion of $l(t)$. Therefore,

$$u_1(t) = A_1 t^3 + A_2 t^2 + A_3 t + A_4 \ , \ \ t \in [t_1, t_3] \tag{2.18}$$

$$u_2(t) = B_1 t^3 + B_2 t^2 + B_3 t + B_4 \ , \ \ t \in [t_3, t_5] \tag{2.19}$$

$$l_0(t) = C_1 t^3 + C_2 t^2 + C_3 t + C_4 \ , \ \ t \in [t_2, t_4] \tag{2.20}$$

Since there are at slightest two data points in $[t_2, t_3]$, equation $u(t) + l(t) = 0$ has at the slightest four roots in $[t_2, t_3]$. Take note that $u_1(t)$ and $l_0(t)$ are cubic functions, the equations above indicates that they need to have the same coefficients in the expressions. Meaning, $A_i = -C_i \ , \ \ i = 1,2,3,4$. Likewise, over the interval $t \in [t_3, t_4]$, we ought to have $B_i = -C_i \ , \ i = 1,2,3,4$. Therefore, the two segmental cubic curves $u_1(t)$ and $u_2(t)$ are essentially two parts of the same cubic function. So we came to a conclusion that is the same as the

all the other segmental functions in other intervals. At long last, is found that the upper or lower envelope of an IMF is in reality a unitary cubic function, in spite of that it is composed of a number of segmental cubic functions (Wang et al., 2010). Meaning, all the maxima or minima of the $c(t)$ are on one single cubic curve,

$$u(t) = U_1 t^3 + U_2 t^2 + U_3 t + U_4 \tag{2.21}$$

Hence, the first step is completed. Next, let a and b be the initial and end time points. As $u(t)$ is the upper envelop of $c(t)$ and $u(t)$ is a cubic natural spline as well, as recommended by Huang et al (1998), we ought to have the zero curvature at the ends:

$$u''(a) = u''(b) = 0 \tag{2.22}$$

This condition forces U_1 and U_2 to identically become zero. Hence, $u(t)$ has to degenerate to a straight line. This ended the proof of the theorem.

A crucial corollary of the theorem above is this:

> *A part from the data with strictly monotonic sifting or varying amplitudes the slopes of the final straight line for the general data has to be zero.*

It is simple to see that the only cases where the slope is not zero would be for the cases for the strictly monotonic decreasing or increasing in amplitude of the oscillating data. Hence, the slope of the straight line has to be zero as most general data are not strictly monotonic in oscillating amplitude.

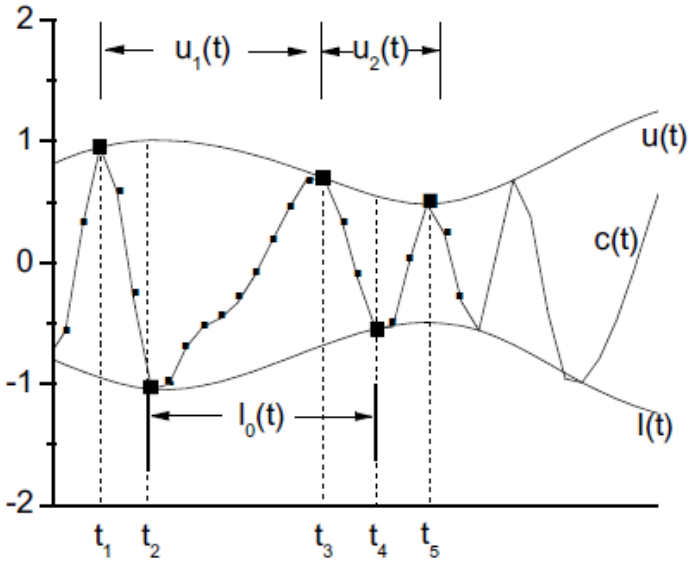


Figure 2.1. Demonstration of an IMF's upper and lower envelopes with their extrema (Wang et al., 2010)

This inevitable zero slope straight line envelope result concurs precisely with the asymptotic state in Wu and Huang [2010](Wang et al., 2010) empirically observed outcome that iteration of sifting ad infinitum will create IMF of steady amplitude. Moreover, in the event that the mirror extension is taken to treat the end effects by supply the extrema at the boundaries to expand the envelopes when executing EMD, the amplitude of the envelope will continuously be consistent. During the EMD decomposition process, components with highest frequency are extracted and proceed to the component with the second highest frequency in the residual data. The process continues until the trend function is gotten. With the execution of EEMD, distribution of the extrema of second or higher mode would satisfy the sparse condition set above. Appropriately, the amplitude of IMF2 or higher would have to be steady, in the event that the definition of IMF is strictly fulfilled. In the event that a spline function of higher order is utilized, the limitation of the straight line above will still be valid only the sparse condition will be fulfilled at a higher IMF component (Wang et al., 2010). Higher order spline, in any case, is in reality illogical and could indeed result in non-convergence and moreover deliver artificial higher order frequency variations. Envelope characterized with a function other than the natural spline zero curvature condition would need other hardness to characterize artificial conditions. Hence, the conclusion above could be expanded to any spline usage of EMD with any order.

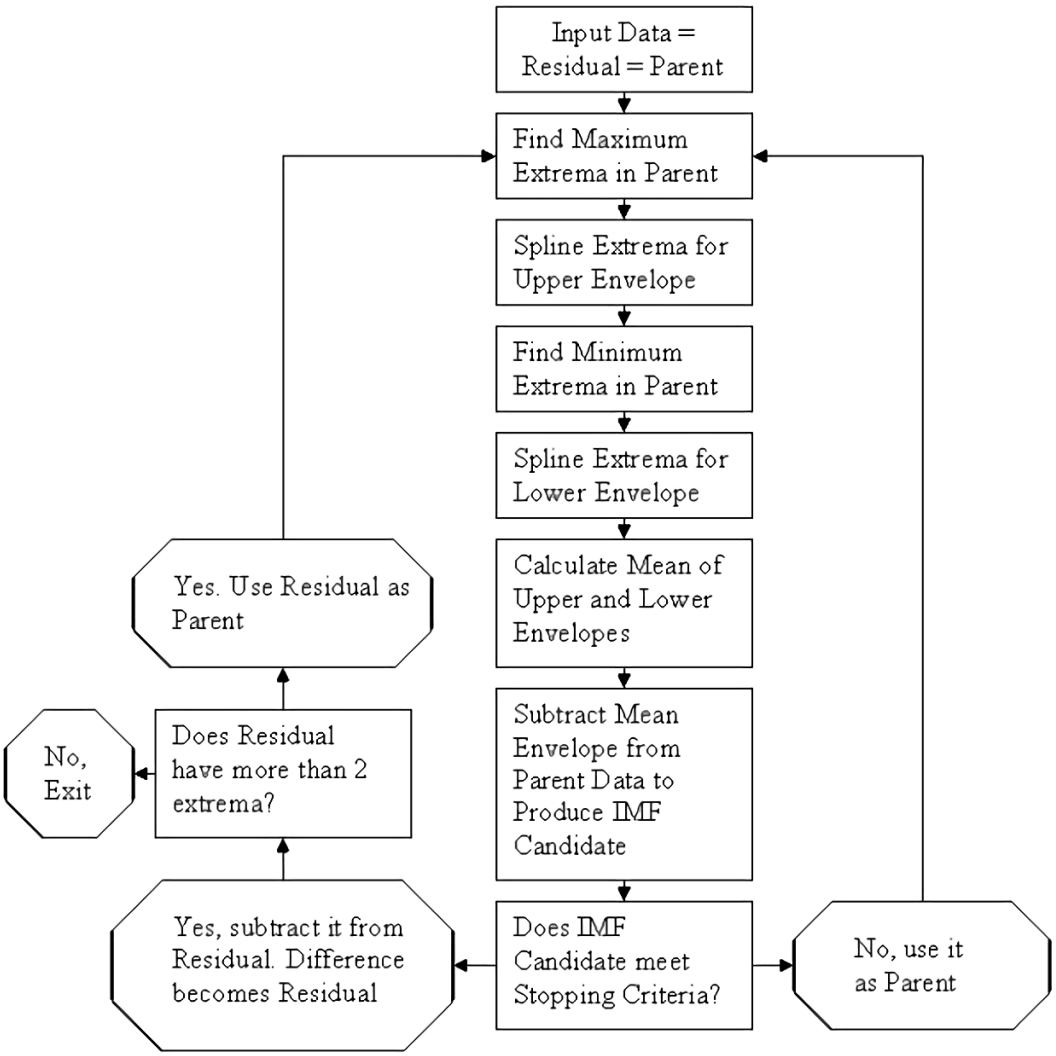


Figure 2.2. Flow chart of EMD sifting process (Crowley, 2009)

2.3. Statistical Significance of IMFs

EMD is a method that decomposes data into a series of components, IMFs, based on their scales. However, the statistical significance of IMF has always been a concern. How could noise be certainly separated from a dataset that has noise embedded within it? Via explicit study of noise, questions of this nature were addressed by Flandrin et al (2004), Flandrin and Gonçalves (2004) and Wu and Huang (2004).

Research was conducted on fractional Gaussian noise by Flandrin et al (2004) and Flandrin and Gonçalves (2004) and came out with the result that EMD is a dyadic filter. It was also founded by these researchers that a straight line is achieved when the root-mean-squared (RMS) values of IMFs as a function of the mean period obtained from fractional Gaussian

noise are plotted against log-log scales. The derivative (slope) of a white noise is -1. Nonetheless, with different Hurst indices there is regular change in values. Based on these research results, EMD can be used to identify any noise encountered (Huang and Shen, 2005).

Wu and Huang (2004) conducted research on only Gaussian white noise rather than fractional Gaussian noise. Furthermore, Wu and Huang (2004) examined the properties of the scattering of the data and analytically discovered the bound for the noise distribution. From the scattering, they came up with a 95% bound for the white noise. Hence, they came up with the final decision that when EMD is used for the analysis of data, if the RMS-mean period is found in the noise bound the component is highly likely a representation of the noise. On the contrary if the RMS-mean period goes beyond the noise bound, thus the IMF is a representation of statistically significant information. Therefore, with the research conducted on noise, Wu and Huang discovered a way of dissociating noise from the information or data.

One of the first key steps in data analysis is to make decision as to whether the data itself or its components contain vital information. The assignment is basically a binary hypothesis testing issue where the null hypothesis of pure noise is suggested. To test this null hypothesis, first the properties of the noise need to be known because these properties relates to the method of analysis to be used.

Scientifically, the term “noise” alludes disturbance, particularly a random and a determined unsettling disturbance that clouds and reduces the clarity of a signal. In nature, noise can be created by local and on-and-off instabilities, irresolvable grid phenomena, as well as by sensors and recording phenomena. Therefore, whenever we are dealing with data then we are unavoidably dealing with a combination of signal and noise,

$$x(t) = s(t) + n(t) \tag{2.23}$$

Where $x(t)$, $s(t)$ and $n(t)$ denotes data, signal and noise respectively. The discovery of the content of a noisy data as to signal and noise is crucial to decision making and extraction of information. Information extraction from dataset needs idea of the properties of both the

signal and the noise. If linear processes are used to generate the dataset, then in that case the noise in the dataset has properties from that of the true signal. Hence, effective filter can be developed depending on the properties of the noise and the signal in order to isolate the noise from the signal. However, cases of such nature are very few since the chance of having an idea about the signal in a noisy data is thoroughly slim before any analysis of the data. In the event that the signal is nonlinear and nonstationary, and the data is of restricted size then the issue becomes more complicated. In such events, even a short piece of pure noise can emulate a signal. Hence, all the conceivable behaviours of a short piece of pure noise ought to be considered. In such situations a less yearning objective is often set to determine whether there is any signal in a noisy dataset. Determining whether a noisy dataset contains any signal is often a binary hypothesis process where the null hypothesis assumes that the dataset contains no signal. For such hypothesis, the properties of the noise can be used a reference for segregating the signal from pure noise without having any previous idea about the signal. Moreover, knowing the properties of the noise is fundamental first step before any form of significance can be attached to the signal extracted from the data. The method of analysis use to investigate the noise is usually related to the properties of the noise.

There are numerous time series analysis methods available for use. The analysis method might possess some power of segregating the data from the noise if linear processes are used in generating the data and the data have distinct time and frequency scales different from those of the true signal. Nonetheless, quite a good number of these methods are affected by various setbacks even in linear and stationary phenomena. For instance if the real signal and the noise have particular essential frequencies, their harmonics can still blend with the noise during a Fourier spectrum analysis. This blending of the harmonics will discredit the Fourier spectrum analysis of being an effective method for segregating noise from the signal. The problem can be even more complicated if the time series data is both nonlinear and nonstationary. Therefore, more reliable effective methods and exclusive knowledge about the properties of noise pertaining to these methods are required in order to be able to estimate the content of the true signal.

In order to overcome the complications posed by data generated by nonlinear and nonstationary process, EMD will be applied, because due to its adaptive nature harmonics are not required. Therefore, the properties of Gaussian white noise will be investigated by

using EMD and a method of statistical test will be developed in order to recognise the IMFs of real data from those of the Gaussian white noise.

It was demonstrated that Gaussian white noise has nearly the same properties as those of the uniform white noise, which already stated (Wu and Huang, 2004): (1) the EMD is a compelling dyadic filter with the power of segregating white noise into IMFs having mean periods precisely twice the value of the predetermined value; (2) the IMFs are normally distributed; and (3) the Fourier spectra of the IMF components are the same in shape and cover the same area on the semi-logarithmic period scale. These outcomes are vital for deciding the connection between the product of the mean energy density of an IMF and its corresponding mean period and as well as the spread function of the energy density. The properties and the determined connections are confirmed by utilising the Monte-Carlo method, which analyses a large synthetically produced Gaussian white noise dataset. These quantities also give the essential information enabling the development of a statistical significance test method by utilising the bounds for the energy density spread function of the IMFs of Gaussian white noise.

Despite all the merits of the EMD there is a major problem it experiences called mode mixing. This problem arises from a single IMF contains signals of different scales or a signal of the same scale found in different IMFs. One of the ways to alleviate this problem is by the ensemble empirical mode decomposition method (EEMD).

3. MODE MIXING PROBLEM IN EMPIRICAL MODE DECOMPOSITION

Even though EMD frequently demonstrated remarkable effectiveness in terms of representing nonlinear and nonstationary datasets, it is confronted with the difficulty of being basically characterised by an algorithm, and therefore of not conceding an analytical formulation which would permit a theoretical basis of EMD in a simplified way via direct analysis of the separation of the harmonics (De Wu, Chiou, and Goldman, 2010). Mode mixing is defined as a single IMF either consisting of signals of greatly different scales (frequencies) or a signal containing the same scales (frequencies) found in different IMF components. When mode mixing takes effect an IMF can decline to have any physical meaning on its own proposing falsely that there might be disparate representation of the physical process in the mode (Wu and Huang, 2009). In another words, when mode mixing takes place due the sifting process, physical meaning of IMFs deteriorates significantly rendering the results of the analysis quite unreliable (Fosso and Molinas, 2017). For theoretical analysis the logical result is that the amplitude (A) and frequency (ω) ratio of the harmonics can be separated into three different categories (Verma, 2016):

(i.) Harmonics consisting of very close frequencies and small amplitude

$$\frac{(A_2)}{(A_1)} \leq \left(\frac{\omega_2}{\omega_1}\right)^2 \tag{3.1}$$

Not suitable for EMD.

(ii.) Close frequency harmonics

$$\left(\frac{\omega_2}{\omega_1}\right)^{2^2} \leq \frac{A_2}{A_1} \leq 2.4\left(\frac{\omega_2}{\omega_1}\right)^{1.75} \tag{3.2}$$

Needs a number of sifting iterations. Mode mixing normally occurs in this case.

(iii.) Distant frequencies and large amplitude harmonics

$$\frac{A_2}{A_1} > 2.4\left(\frac{\omega_2}{\omega_1}\right)^{1.75} \tag{3.3}$$

Well separated for a single iteration.

The cause of mode mixing in EMD is not limited to the relationship between amplitude and frequency but jumping changed of the decomposed signal result in mode mixing as well (Xu et al., 2016). For instance, discontinuity of the decomposed signal such as interference of unwanted noise in the original signal, discrete distribution of high frequency wave in the original signal which results to a local high frequency of the original signal, etc. The jumping change causes mutation of properties of the original signal resulting in the IMFs not agreeing with the time scale, making the mode mixing phenomenon occur in EMD during the decomposition process; thereby different frequency components cannot be effectively segregated based on the characteristics of the time scale.

The problem of mode mixing is solved by other algorithms for Empirical Mode Decomposition.

3.1. Other Algorithms (Approaches) For Empirical Mode Decomposition

3.1.1. Ensemble empirical mode decomposition (EEMD)

EEMD alleviates the problem of mode mixing by considering IMF components as the mean of the ensemble of trials within which is the original signal in addition to significant white noise with finite amplitude. In the end of the process the white noise cancels out and tends to stop the process from the problem of the combination of maxima and minima which might result to mode mixing (Gupta et al., 2016). Hence, upon addition of white noise the signal $X(t)$ in i^{th} trials can be expressed as:

$$X^i(t) = X(t) + a_0\omega^i(n), i = 1,2,...,l \tag{3.4}$$

where, $\omega^i(n)$ is the white noise in i^{th} trials with unit variance and a_0 amplitude. For every $i = 1,2,...,l$ the IMF_k^i is evaluated with different realisation of the white noise in the above equation. The average k^{th} $\overline{IMF_k}$ can be expressed as:

$$\overline{IMF_k} = {}^1/_l \sum_{i=1}^{l} IMF_k^i \tag{3.5}$$

The main concept of EEMD is:

1. With the help of the ensemble mean the added collection of the white noise cancels each other out at the end of the process. Therefore, the only ingredient left in the equation of the signal and white noise is just the signal.
2. In order to find all possible solutions, it is deemed fit for white noise with finite amplitude to be ensemble with the signal.
3. In order to get authentic physically meaningful IMF extracts from EMD, white noise needs to be added to the signal.

One problem posed by EEMD is that at the end of the process the added noise may not be completely separated from the signal. This problem is normally remedied by the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN).

3.1.2. Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)

The core problem of the signal and the noise not being completely segregated is that the resultant different IMFs might be a combination of the signal and the noise, hence leading to spurious results in the analysis of the IMFs (Gupta et al., 2016). The CEEMDAN algorithm provides a clear distinction of the modes spectrally to contain this problem (Torres, Colominas, Schlotthauer, and Flandrin, 2011). Moreover, it provides an extract reconstruction with lower computational ability of the original signal.

The first residue can be evaluated as

$$r_1 = X(t) - \overline{IMF_1} \tag{3.6}$$

where $\overline{IMF_1}$ denotes the average obtained by EEMD. The second average IMF can be evaluated by

$$\overline{IMF}_2 = {}^1\!/_l \sum_{i=1}^{l} E_1(r_1 + a_0 E_1(\omega^i(n))) \tag{3.7}$$

Upon finding k^{th} residue, for $k = 2, 3, ..., K$ the $k+1$ average IMF can be evaluated by:

$$\overline{IMF}_{k+1} = {}^{1}/_{l} \sum_{i=1}^{l} E_1(r_k + a_k E_k(\omega^i(n))) \tag{3.8}$$

where $E_k(.)$ is an operator to extract k^{th} IMF from given signal by EMD algorithm.

4. APLICATION TO SIMULATED AND REAL DATA FOR MODE MIXING

In the simulation process two datasets are generated with specified frequencies, one without any mode mixing problem (Kim et al., 2006), and the other with mode mixing problem (Xu et al., 2016) as illustrated in the flow chart below. The simulated data generated with mode mixing problem is demonstrated by a discontinuous signal. The discontinuous signal refers to the time domain frequency discontinuity where $x_1(t) = a_1 \sin(2\pi f_1 t)$, $0.05 \leq t \leq 0.1$; $x_2(t) = a_2 \sin(2\pi f_2 t)$, $0.15 \leq t \leq 0.25$; $x_3 = \sin(2\pi f_3 t)$, $0 \leq t \leq 0.3$. The frequencies $f_1 = 150Hz$, $f_2 = 50Hz$ and $f_3 = 10Hz$. The amplitudes are $a_1 = 0.2$, $a_2 = 0.4$ and $a_3 = 1$. Each of these simulated datasets will be decomposed accordingly in order to portray the dynamics and properties of extracted IMFs.

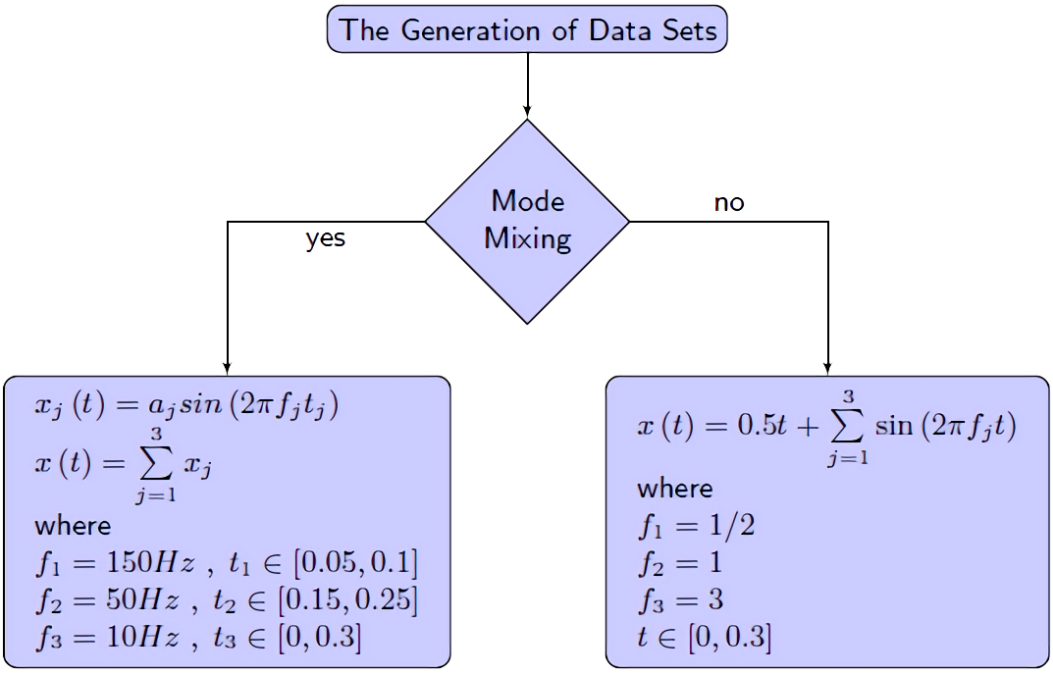


Figure 4.1. Flow chart of simulated datasets

4.1. Simulated Data Set in Matlab Without Mode Mixing Problem

Figure 4.2 is the plotted output of the dataset without any mode mixing problem from which seasonality and trend can clearly be depicted.

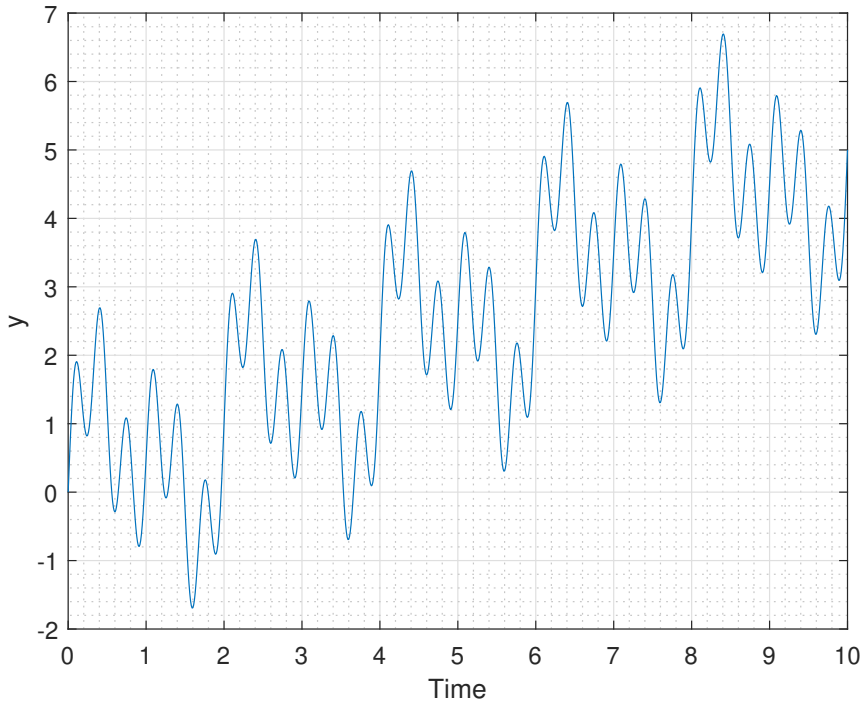


Figure 4.2. No mode mixing problem

The dataset in Figure 4.2 is decomposed using the EMD technique and three different IMFs are extracted: IMF_1, IMF_2, and IMF_3 as in Figure 4.3 all of which are in conformity with the definition of an IMF . Each of the IMFs is a representation of a single intrinsic mode of oscillation within the simulated dataset. Hence, uniqueness in each of IMFs is achieved, providing a clear evidence of no mode mixing problem within the dataset. The residue shows the trend of the dataset.

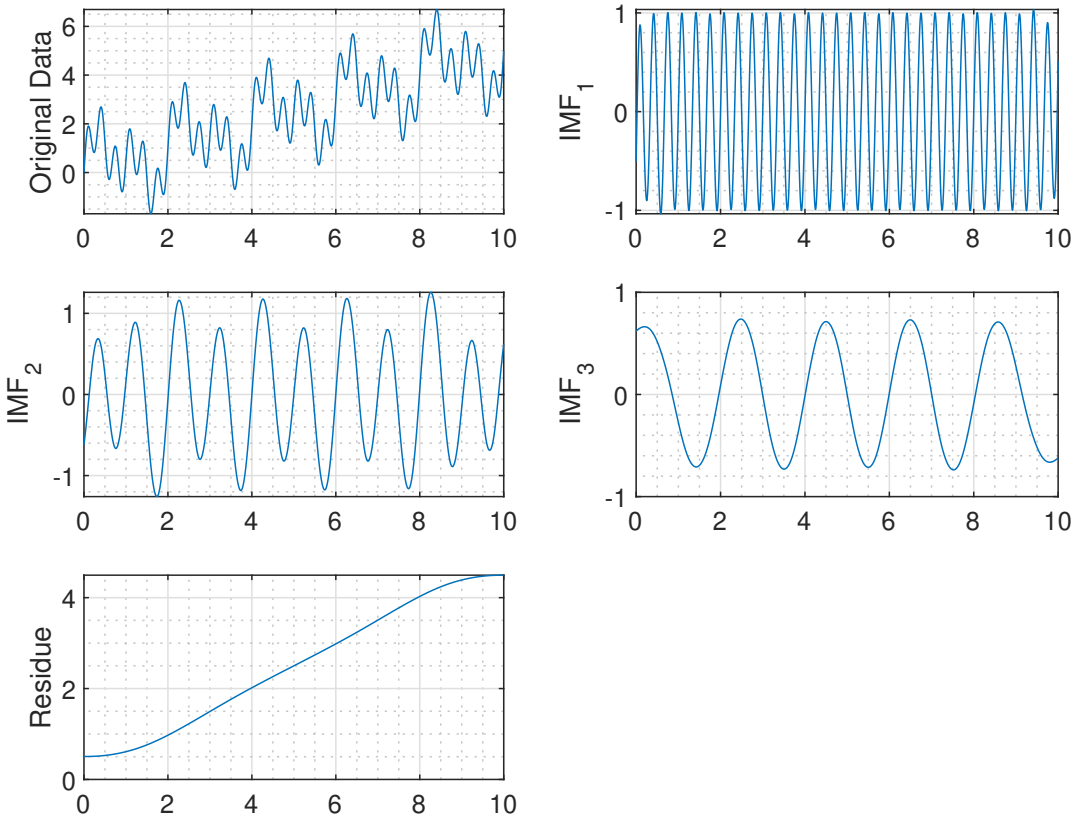


Figure 4.3. Empirical mode decomposition

4.2. Simulated Data Set in Matlab With Mode Mixing Problem

Three different IMFs are generated, x_1, x_2, and x_3, each with its own unique characteristics and dynamics as indicated in Figure 4.4. The combination of these IMFs produced the plotted dataset in red colour and is named original data in Figure 4.4. This data will be decomposed by the EMD technique and the expected result is to obtain the IMFs combined to form the original data each standing out on its own similar to what is shown in Figure 4.3.

Due to the fact that we have prior knowledge about the signal, we can explicitly interpret the decomposed output of the signal without applying any detection of mode mixing in IMF technique.

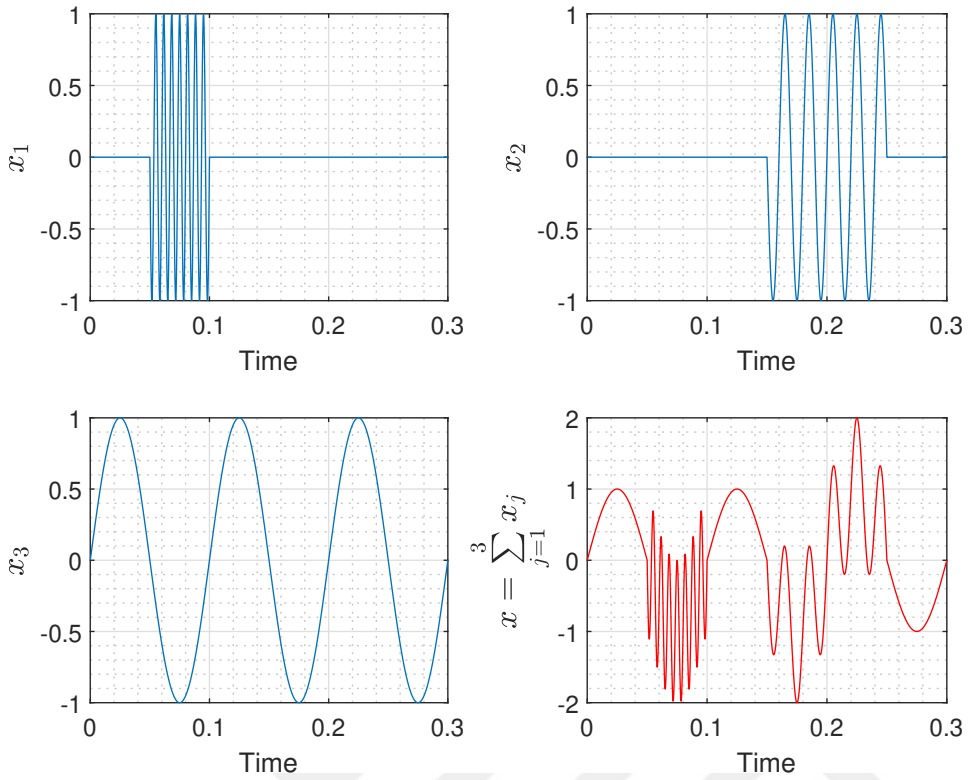


Figure 4.4. Simulated data set and illustration of mode mixing problem

4.3. EMD Study in Matlab

EMD technique is implemented in the decomposition of original data in Figure 4.5 which yields two IMFs. It can be identified that IMF_1 in Figure 4.5 consist of properties of x_1 and x_2 in Figure 4.4, also IMF_2 has characteristics of both x_2 andx_3. Therefore, there is the problem of mode mixing in both IMFs in Figure 4.5. In order to segregate the IMFs into their unique single intrinsic modes of oscillation, the EEMD technique which is noise assisted will be implemented.

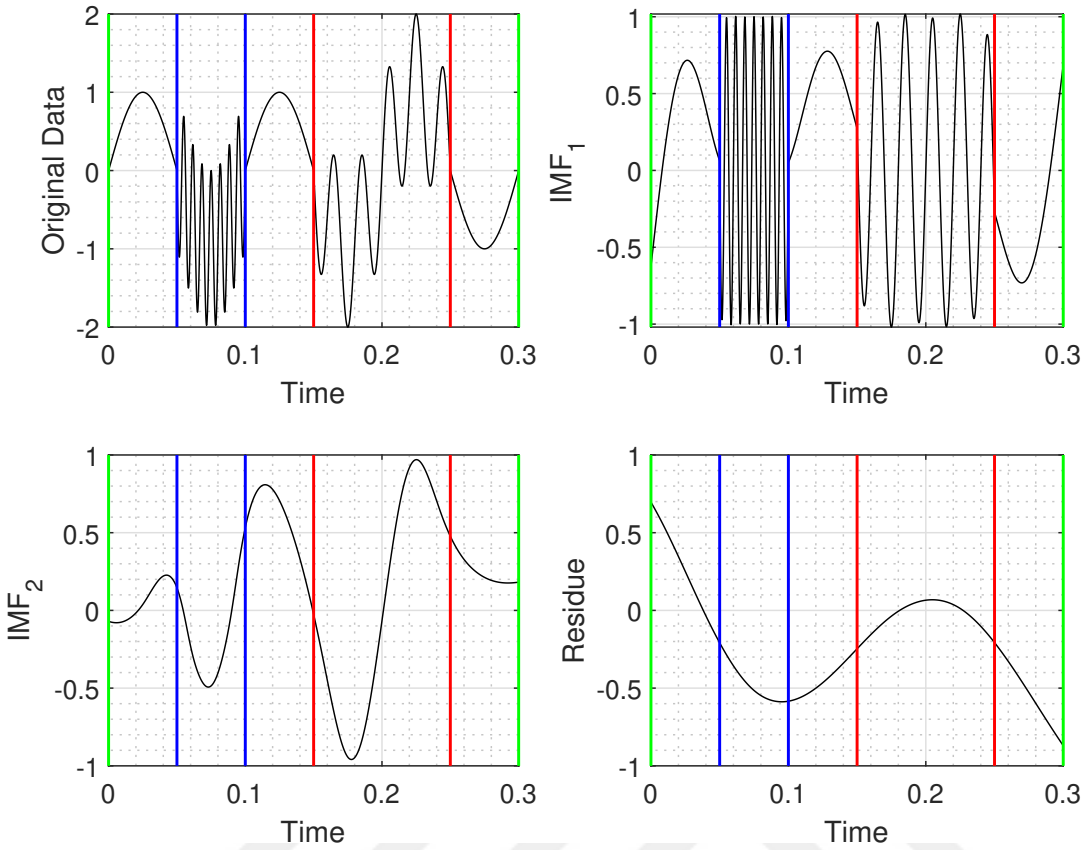


Figure 4.5. Empirical mode decomposition

4.4. EMD Study in R

The same input data in Figure 4.5 named original data was decomposed using the EMD technique in R from which the output displayed more IMFs than usual as in Figure 4.6. The reason responsible for this is that in R, the decomposition process did not make use of the stoppage criteria mentioned above. Nevertheless, IMF_1 is containing frequencies of 150Hz and 50Hz which are characteristics of x_1 and x_2 respectively in Figure 4.4. Hence, there is evidence of mode mixing problem. In order to solve the mode mixing problem, EEMD technique will be implemented.

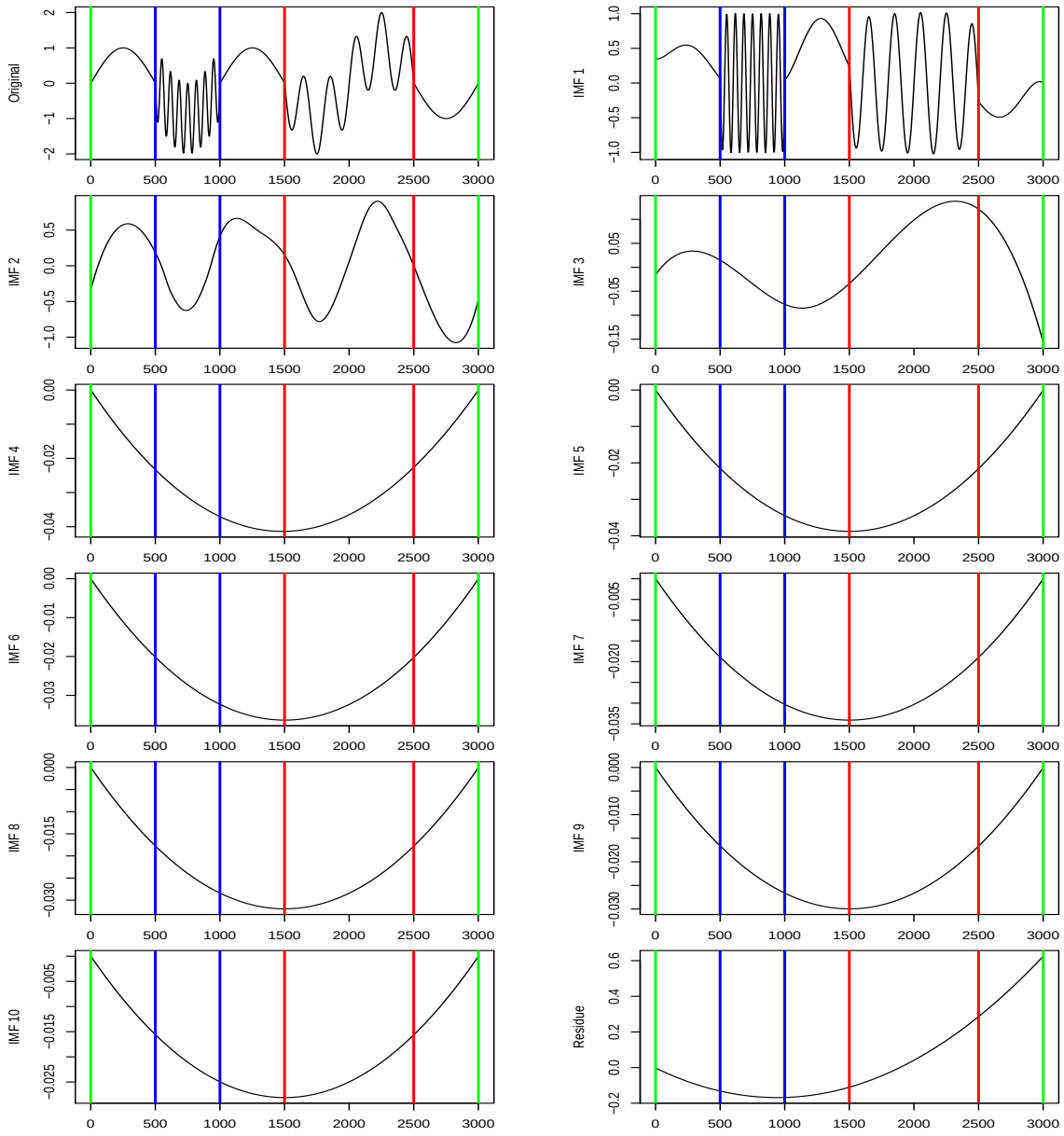


Figure 4.6. Empirical mode decomposition

The components in Figure 4.7 are obtained from the EEMD process which is manifested in the noise showcased in IMF_1, IMF_2 and IMF_3 as a result of the added white noise to the input data. IMF_4 is a representation of a single intrinsic mode of oscillation with a frequency of 150Hz. IMF_5 and IMF_6 both have the same frequency of 50Hz meaning they both represent a single intrinsic mode of oscillation with a frequency of 50Hz; suggesting two different periodicities in the IMF x_2 in Figure 4.4. Moreover, there is no remnant of the added noise in any of the IMFs, thus, CEEMDAN will not be implemented.

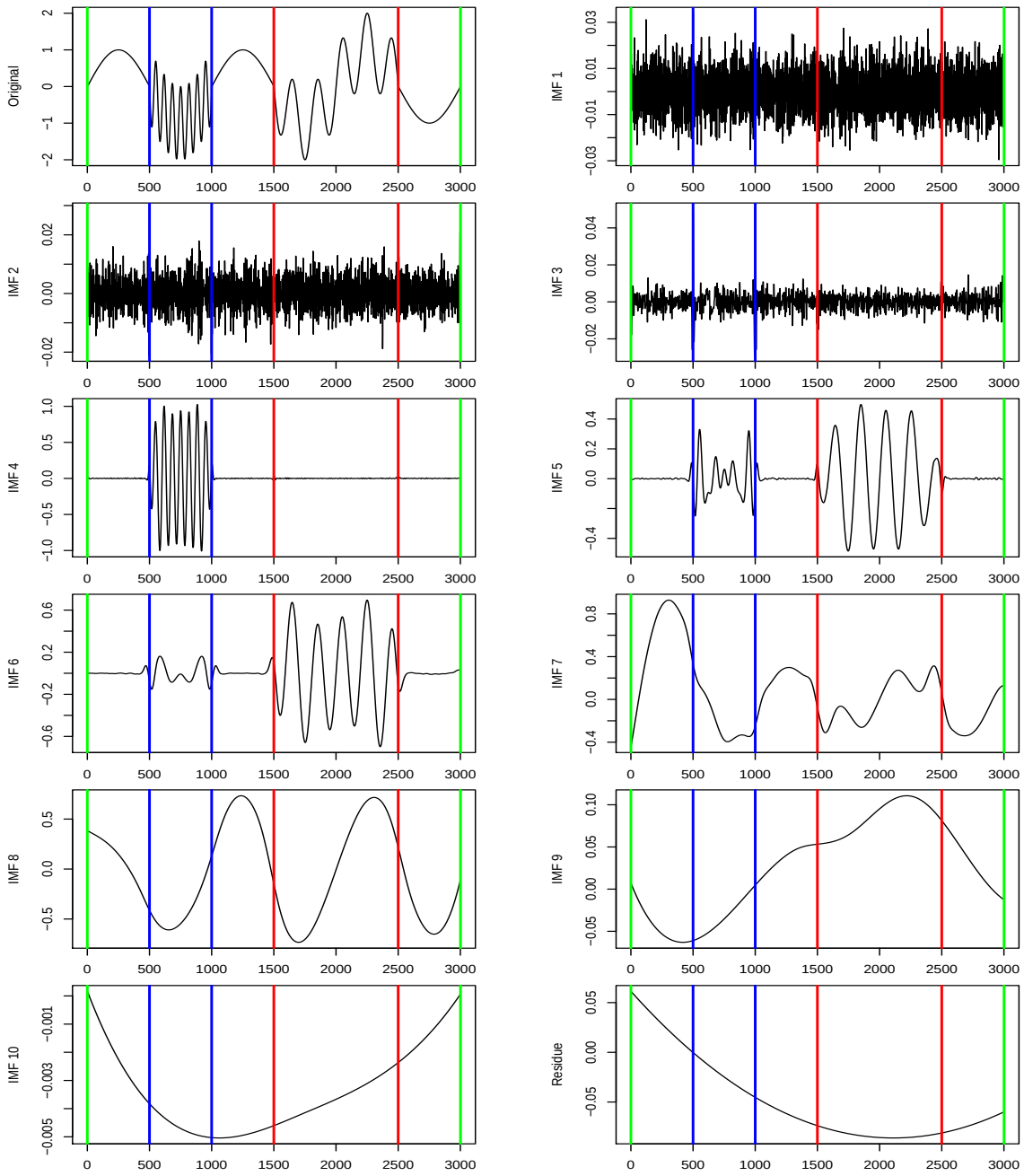


Figure 4.7. Ensemble empirical mode decomposition

However, in instances where there is no prior knowledge of the signal, detection of mode mixing in IMF technique(s) will be implemented so as to confirm whether mode mixing exist within the IMF(s) or otherwise.

4.5. Detection of Mode Mixing in IMFs

In cases where there is no pre-existing knowledge of the signal or dataset, upon implementation of EMD it is quite challenging to be able to distinguish scales within

extracted IMFs making the detection of mode mixing extremely hard. Therefore, in order to provide a clear evidence of the existence of mode mixing or otherwise within extracted IMFs, cross-correlation and instantaneous frequency will be implemented on the extracted IMFS.

4.5.1. Cross-correlation

Cross correlation considers circumstances where two functions, signals or datasets are connected by a consistent shift in time or space. Correlation gives measures of such shifts even whereas there is presence of significant noise corruption (Belmont, 1997). In other words, cross correlation measures the similarities between two signals or waveforms as a function of time-lag applied to one of them. It could be referred to as the sliding dot product or the sliding inner product.

Consider two signals $u(t)$ and $v(t)$, the cross correlation between them is (Brooks, 2015):

$$w(t) = u(t) \otimes v(t) = \int_{-\infty}^{+\infty} u^*(\tau)v(\tau+t)d\tau = \int_{-\infty}^{+\infty} u^*(\tau-t)v(\tau)d\tau\ ,\ \ [sub : \tau \to \tau - t] \quad (4.1)$$

The complex conjugate $u^*(\tau)$makes no distinction in case $u(t)$ is real–valued but makes the definition work even if $u(t)$ is complexed–valued (Brooks, 2015).

Here $w(t)$ is the lag (delay in u versus v).

Correlation versus convolution

$$u(t) \otimes v(t) = \int_{-\infty}^{+\infty} u^*(\tau)v(\tau+t)d\tau \qquad \text{[Correlation]} \quad (4.2)$$

$$u(t) * v(t) = \int_{-\infty}^{+\infty} u^*(\tau-t)v(\tau)d\tau \qquad \text{[Convolution]} \quad (4.3)$$

Different from convolution, the sign of the variable of integration τ, is the same for both arguments of $u(...)$ and $v(...)$. Hence, there is constant difference in the arguments instead of a constant sum (meaning $v(t)$ is not time flipped) (Brooks, 2015).

Normalised cross correlation

$$\text{Correlation: } w(t) = u(t) \otimes v(t) = \int_{-\infty}^{+\infty} u^*(\tau - t)v(\tau)d\tau \tag{4.4}$$

Provided that $y(t) = u(t - t_0)$ for some fixed t_0, hence $E_y = E_u$:

$$E_y = \int_{-\infty}^{+\infty} |y(t)|^2 dt = \int_{-\infty}^{+\infty} |u(t - t_0)|^2 dt = \int_{-\infty}^{+\infty} |u(\tau)|^2 d\tau = E_u \ , \ \ [t \to \tau + t_0] \tag{4.5}$$

$$\text{Cauchy-Schwarz inequality: } \left| \int_{-\infty}^{+\infty} y^*(\tau)u(\tau)d\tau \right|^2 \leq E_y E_u \tag{4.6}$$

$$\Rightarrow |w(t_0)|^2 = \left| \int_{-\infty}^{+\infty} u^*(\tau - t_0)v(\tau)d\tau \right|^2 \leq E_y E_v = E_u E_v \tag{4.7}$$

Since t_0 was arbitrary, then

$$|w(t)| \leq \sqrt{E_u E_v} \ , \ \ \forall t \tag{4.8}$$

Normalised cross correlation can be defined as:

$$z(t) = \frac{u(t) \otimes v(t)}{\sqrt{E_u E_v}} \tag{4.9}$$

Having properties:

1. $|z(t)| \leq 1 \ , \ \ \forall t$
2. $|z(t_0)| = 1 \Leftrightarrow v(\tau) = \alpha u(\tau - t_0)$ with α constant.

Autocorrelation

Autocorrelation is the correlation of a signal with itself (Brooks, 2015). ie

$$w(t) = u(t) \otimes u(t) = \int_{-\infty}^{+\infty} u^*(\tau - t)u(\tau)d\tau \tag{4.10}$$

Autocorrelation at lag zero:

$$w(0) = \int_{-\infty}^{+\infty} u^*(\tau)u(\tau)d\tau = \int_{-\infty}^{+\infty} |u(\tau)|^2 d\tau = E_u \tag{4.11}$$

$w(0)$ which is autocorrelation at lag 0, is the energy of the signal.

Normalised autocorrelation:

$$z(t) = \frac{u(t) \otimes u(t)}{E_u} \tag{4.12}$$

Normalized autocorrelation satisfies to followings:

1. $z(0) = 1$ and
2. $|z(t)| \leq 1$ for any t.

Dangers of cross-correlation

The vital point in this instance is that meaningless relationship exists between autocorrelated independent pair of time series which are independent. Say there are two independent time series x_t and y_t (meaning the values of one at any time does not influence or give information about the values of other), there is still a possibility that the cross correlation between the series can come off essentially nonzero when judged against standard criteria for significance for independent pair of measurements (Dean and Dunsmuir, 2016). Consider a circumstance in which both series are strongly positively autocorrelated. Those types of series have moderately smooth sample path in which the low frequency behaviour stands out. Now say there are two such independent series and shift the sample pathway of one relative to the other and correlate the adjusted pairs. In this case, it is highly possible that in each series the wave type movement or behaviour can roughly align creating the impression that the series are correlated both numerically and visually while in actual sense there is no actual correlation between the series. Due to this danger another method will be looked into (Dean and Dunsmuir, 2016).

A more reliable method of detecting mode mixing in IMFs is by the use of instantaneous frequency.

4.5.2. Instantaneous frequency (IF)

The frequency of a sinusoidal signal is normally a well characterised quantity. Nonetheless, usually in practice signals are really not sinusoidal in nature. IF is fundamental in getting a clear picture of the dynamics of nonlinear and nonstationary process (Huang et al., 2009), nonstationary stationary signals especially, which do not mostly follow the pattern of sinusoidal component upon decomposition. For signals of such nature the concept of frequency loses its viability. Hence, a parameter should be utilized which accounts for the time-varying nature of the process. Due to this need the idea of instantaneous frequency has been developed (Boashash, 1992). Therefore, instantaneous frequency portrays the difference of local phase and permits detection of variations of frequency. These variations show the processed data and are therefore cryptanalytically vital (Korkikian, Naccache, and de Almeida, 2013). By the use of Hilbert Huang transform, IMFs creates instantaneous frequencies as a function of time which clearly displays hidden structures (Huang et al., 1998).

Another method for the detection of mode mixing is automatic detection technique using nonstationary detection.

4.5.3. Detecting the problem of mode mixing authomatically via the use of nonstationary detection method: application to selecting IMFs of interest and denoising

This technique is of a more robust approach and was recently proposed by Terrien, Marque, and Karlsson (2011). The basis for the algorithm of this method is the statistical analysis depicting noise repartition between IMFs. This algorithm employs measurement of the signal noise content using the energy of the first IMF, which is supposed to entail a specific portion of the entire noise and contain noise only. This method proposes the utilisation of detection of mode mixing based on the stationary test implemented on the first IMF. In the event of mode mixing, noise estimation will be rectified by extracting from the first IMF the portion compatible with the signal of interest.

Noise within the signal of interest will cause defilement of each of the modes by a portion of the noise. The Study conducted on the spectral content and that of the statistical

properties of the modes of EMD, evaluated from Gaussian white noise, drove to definition of a model that can measure the information details of each IMF (Wu and Huang, 2004). Generalisation of this mode in the case of fractional Gaussian white noise was later done by Flandrin, Goncalves, and Rilling (2005). These models can be utilised for either denoising or suppression of unneeded baseline meander (detrending) of the signal of interest. In the utilisation of these models the fundamental hypothesis developed is that, the first IMF accommodates a particular portion of the signal as determined from the models above. The energy of the first IMF is estimated and then utilised to adapt the denoising algorithm to the signal and to its undermining noise (Terrien et al., 2011). The algorithm performance can drastically be decreased in case there is error in estimating the degree of noise from the first IMF. One cause of such error is the mode mixing problem.

In this application, mode mixing consists of the high frequency portion of the noise and the signal's first high frequency component. This sort of mixing will nullify the basic assumption of the method that the first IMF accommodates only noise, and causes an exaggeration of the level of noise in the signal (Terrien et al., 2011).

Firstly, the classical algorithm will be introduced, is for selecting germane modes of the signal.

The classical algorithm

A statistical model of the energy distribution of noise between IMFs has been derived by studying the energy in the modes of a fractional Gaussian noise after EMD (Wu and Huang, 2004). This model makes it possible to come up with a statistical cut-off point below which the IMF is not considered as the dominant portion of the signal. With the use of this model, noise contaminated first IMF of the signal is always regarded as pure noise.

Essentially, the energy of the first signal is:

$$W_H[1] = \sum_{n=1}^{N} d_1^2[1] \tag{4.13}$$

For the rest of the IMFs, the energy of the noise is evaluated for a given Hurst exponent as:

$$W_H[k] = C_{H\rho H}^{-2(1-H)k}, k \geq 2 \tag{4.14}$$

With $C_H = W_H[1]/\beta H$, $\beta_H = 0.719$, for $H = 0.5$, $\rho_H = 2.01 + 0.2 \times (H - 0.5) + 0.12 \times (H - 0.5)^2$

Furthermore, confidence level, $T_H[k]$, can be drawn with this model. For various noise realisations of an IMF, the fractional Gaussian noise decomposition leads to an energy distribution. There exists a linear relationship between a number of IMF, k, and the logarithm of the confidence interval such that (Terrien et al., 2011):

$$\log_2(\log_2(T_H[k])/W_H[k]\ )) = a_H k + b_H \tag{4.15}$$

Where, if $H = 0.5$and for 99% confidence interval: $a_H = 0.45$ and $b_H = -1.95$. B_H, a_H, b_H can be evaluated from the equation above.

Flandrin et al. (2005) proposed the algorithm selection as follows:

1. Asses the energy of the noise $W_H[1]$ on $d_1[n]$
2. Assess the content of pure noise $W_H[2,...,k]$, using the assessment in 1.
3. Asses the confidence interval $T_H[1,...,k]$
4. Conduct comparison of energy between IMF 2 to k and the confidence interval
5. All IMFs with energy higher than that of the confidence interval are regarded as components of the signal but not as pure noise.

In case the required result is to expel the noise, the signal is somewhat (partially) reconstructed by adding together the selected IMFs and the residual $r[n]$.

Terrien, Marque and Karlsson's Proposed approach

In the event of nonexistence of mode mixing, the first IMF $(d_1[n])$ consists of noise alone and is hence stationary. In the presence of mode mixing, the first IMF $(d_1[n])$ is nonstationary. Henceforth, the test of stationarity for the detection of mode mixing was proposed by Terrien et al. (2011). In order to verify the stationarity of the first IMF, an algorithm created by Xiao,

Borgnat, and Flandrin (2007) is applied. It is generally a robust method that only requires a selection of the significance level.

In the event of existence of mode mixing in the first IMF ($d_1[n]$), there is the necessity to differentiate between the portion due to energy and the portion due to noise within the IMF. In order to execute this, the first IMF will be regarded as a noisy signal. Hence, a wavelet filter, $f_{wavelet}$, will be applied so as to distinguish between the original signal, $s[n]$, and the portion due to noise.

The adjusted algorithm is as follows:

1. Test whether the first IMF ($d_1[n]$) is stationary, if it is stationary, move to step 4
2. Extract from the first IMF ($d_1[n]$) the noisy portion $b[n]$ and the signal portion $s[n]$ by wavelet filtering ($f_{wavelet}(x)$).
   a. $s[n] = f_{wavelet}(d_1[n])$
   b. $b[n] = d_1[n] - s[n]$
3. Mode de-mixing
   a. $d_1[n] = b[n]$
   b. $d_2[n] = d_2[n] + s[n]$
4. Implement the classical algorithm for the selection of IMFs

All types of wavelet filter may be used depending on the properties of the signal of interest.

4.6. Illustrations of Detection of Mode Mixing Problem

It would be illustrated how to detect the similarities between the IMFs using cross-correlation and IF to the data from Figure 4.5.

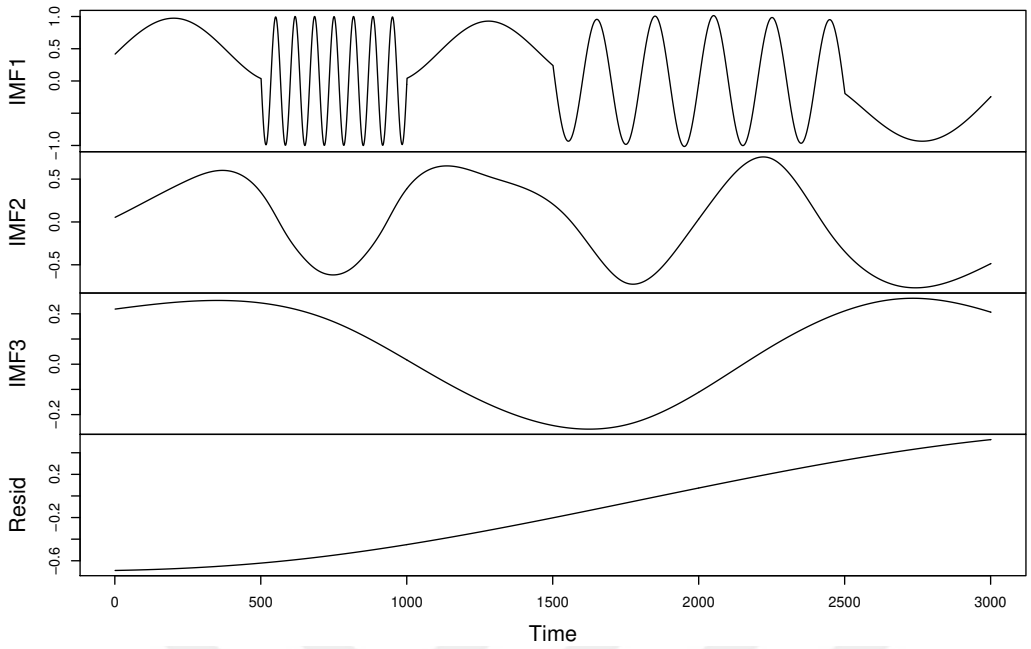


Figure 4.8. IMFs extracted from the EMD of original data in Figure 4.5

Cross-correlation

Figure 4.9 is the plot of the cross correlation which indicates as to whether there is similarity between IMF1 and IMF2 from Figure 4.8. The dashed blue lines represent the confidence interval. If all the vertical lines on the time lags are within the confidence interval would mean there is no significant similarities between IMF1 and IMF2. However, in Figure 4.9 all the vertical lines crossed the dashed blue lines, providing evidence of significant similarities between IMF1 and IMF2. Hence, there is confirmation of the existence of the problem of mode mixing.

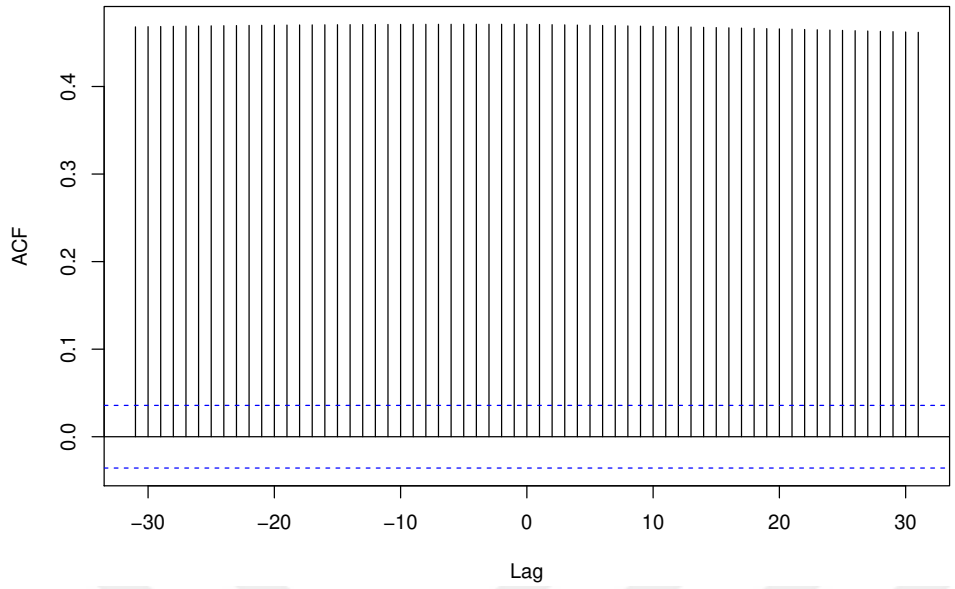


Figure 4.9. Cross correlation of IMF1 and IMF2 in Figure 4.8

Figure 4.10 is the cross correlation plot IMF1 and IMF3 in Figure 4.8 measuring the similarities between the two IMFs. In this plot the vertical lines on the lags are in the downward direction. However, for the IMFs which of the two leads or which lags which is trivial. The pivotal part is whether there is similarity or otherwise. In this plot the entire vertical lines passed the confidence interval, thereby indicating the presence of mode mixing.

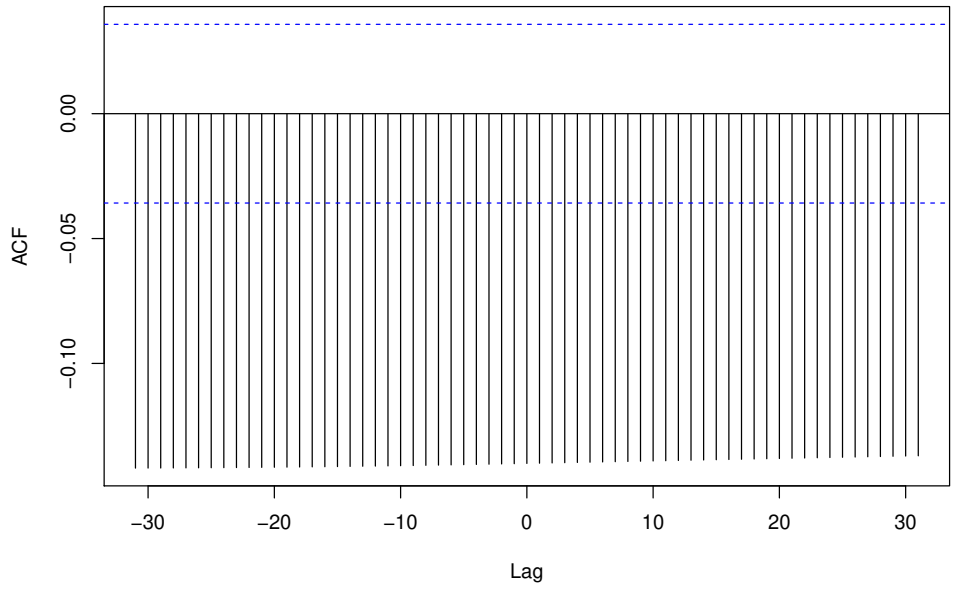


Figure 4.10. Cross correlation of IMF1 and IMF3 in Figure 4.8

Figure 4.11 is another plot for the cross correlation of IMF2 and IMF3. Similar to Figure

4.10, there is indication of strong similarity of IMF2 and IMF3 in Figure 4.8. Thus, there is evidence of the existence of mode mixing.

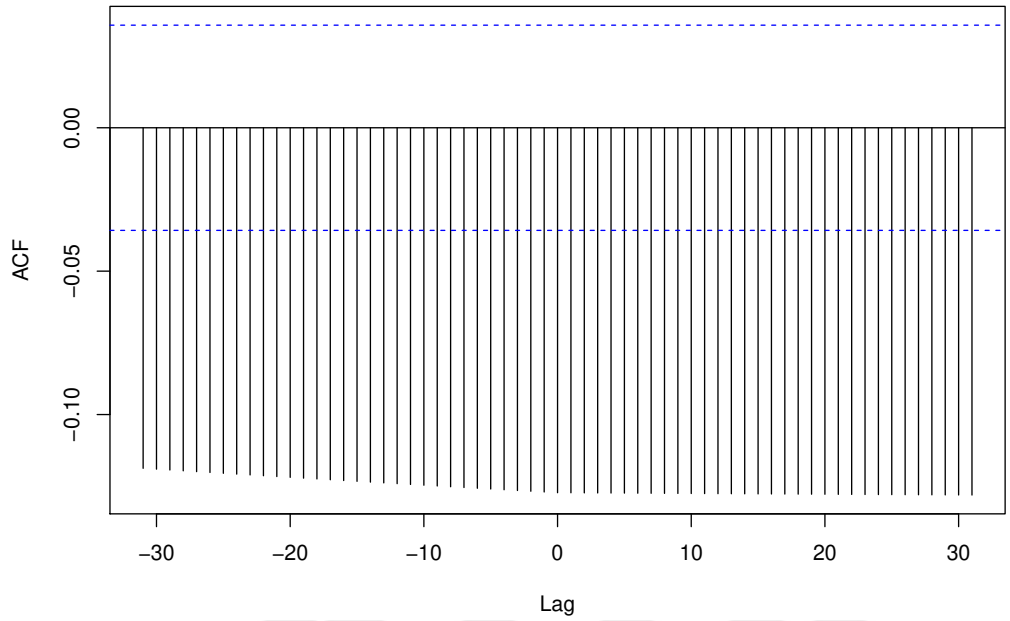


Figure 4.11. Cross correlation of IMF2 and IMF3 in Figure 4.8

Instantaneous frequency

In Figure 4.12 IF1, IF2 and IF3 are the plotted instantaneous frequencies derived from IMF1, IMF2 and IMF3 respectively in Figure 4.8. Under normal circumstances, the IF plots should all show smooth transition from one cycle to another or in certain instances near straight for every IMF because each IMF has the same scale. IF is very efficient in deciphering every scale within an IMF. In the presence of mode mixing, the IF plot for an IMF does not come off smooth transition between cycles or near straight. Instead the IF plot clearly visualises the different scales within an IMF due to mode mixing. Thus, in Figure 4.12 IFs portrays the existence of mode mixing in the extracted IMFs in Figure 4.8 more so in IMF1.

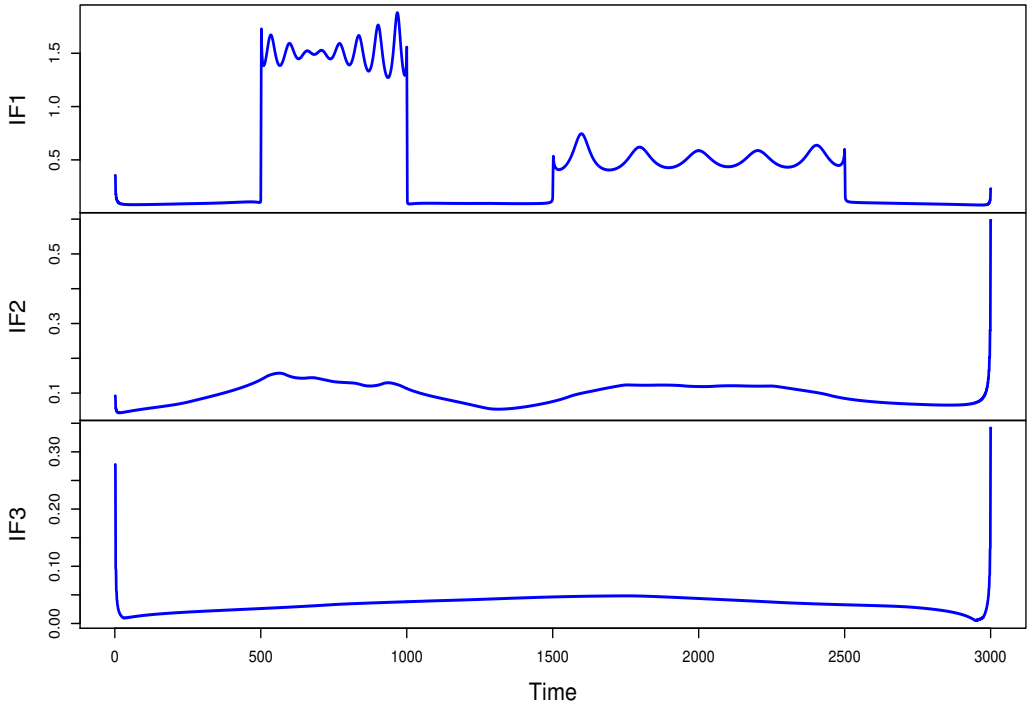


Figure 4.12. Instantaneous frequency of IMF1, IMF2 and IMF3 in Figure 4.8

4.7. Application of EMD to Real Datasets

4.7.1. Data set of total export of consumer goods from Turkey

This first data is the monthly total export of consumer goods from Turkey from 1989 to 2018 which is plotted in Figure 4.13. This data clearly exhibits nonstationarity and nonlinearity properties, hence fit for EMD.

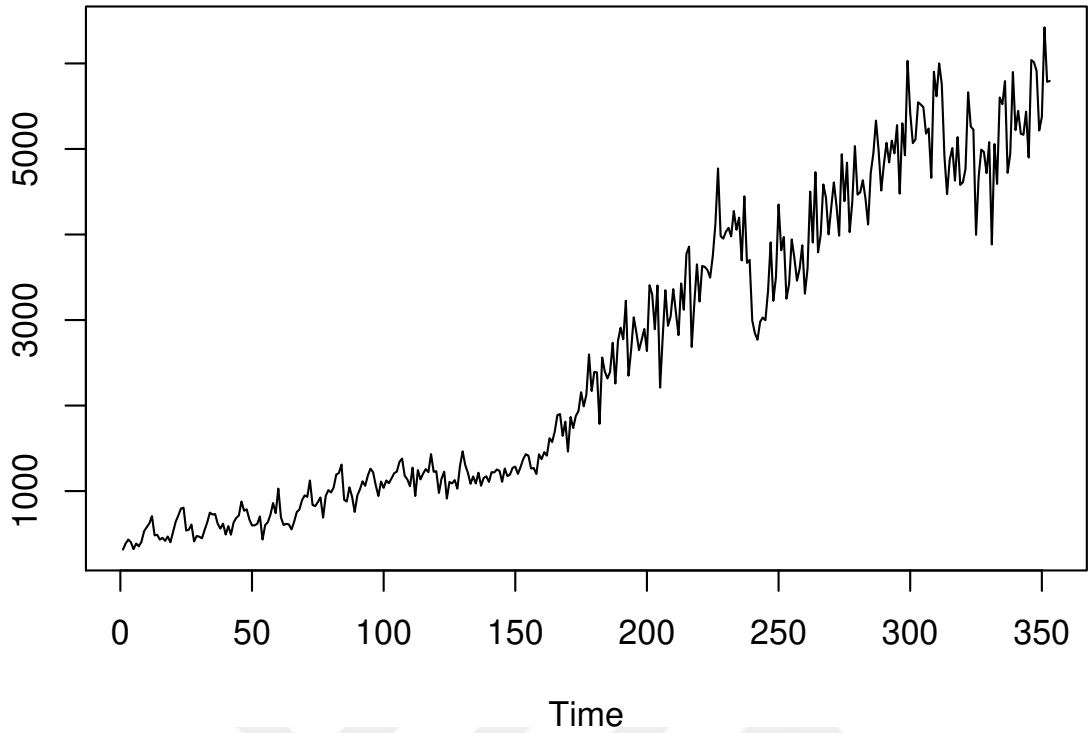


Figure 4.13. Total export of consumer goods from Turkey

After implementing EMD on the data above in Figure 4.13 five IMFs are extracted. The residue shows the trend of the series indicating increase in the series. IMF1 and IMF2 appear to consist of mainly noise while IMFs 3 and 4 showcases some non-regular cycles as shown in Figure 4.14. IMF5 although has different amplitudes indicates seasonality in the export of consumer goods from Turkey. Nonetheless, in order for the structure of data and analysis derived from the extracted IMFs to be credible there should not be any mode mixing within any of the IMFs. Therefore, detection of mode mixing techniques will be implemented to establish evidence that there is or otherwise mode mixing within the extracted IMF(s) in Figure 4.14.

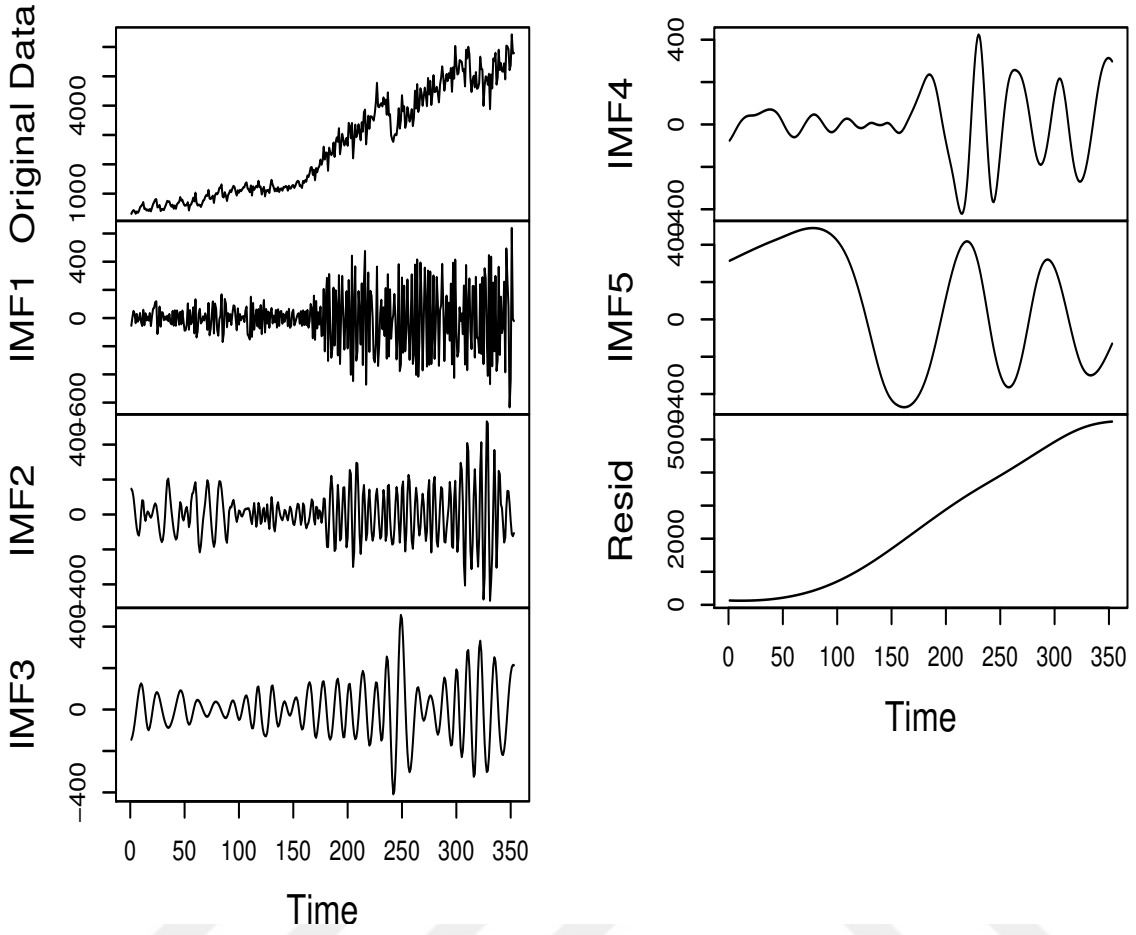


Figure 4.14. Empirical mode decomposition of data in Figure 4.13

Illustrations of detection of mode mixing problem

In order to verify the existence or nonexistence of mode mixing in the IMFs extracted from the EMD process in Figure 4.14 mode mixing detection techniques mentioned above like cross-correlation, instantaneous frequency will be applied.

Cross-correlation

The point of cross-correlation is to confirm whether there is any similarity between any two IMFs at every time shift from Figure 4.14 as indicated in Figure 4.15.The dashed horizontal lines represents the confidence interval and if the vertical lines on the time lags are within the confidence interval, it implies that there is no similarity between the IMFs. Hence, in Figure 4.15 all the cross-correlations are indicating no similarity between the IMFs with exception of the cross-correlation between IMF4 and IMF5. However, recall that one the

dangers of cross-correlation is that in independent series it is highly possible that in each series the wave type movement or behaviour can roughly align creating the impression that the series are correlated both numerically and visually while in actual sense there is no actual correlation between the series. This is the case for the cross-correlation of IMF4 and IMF5 where the wave movements align creating the impression similarity but in actual sense there is no similarity in terms of the scales of the two IMFs 4 and 5. Therefore, we can conclude that there is no mode mixing in any of the extracted IMFs in Figure 4.14 based on the cross-correlations in Figure 4.15.

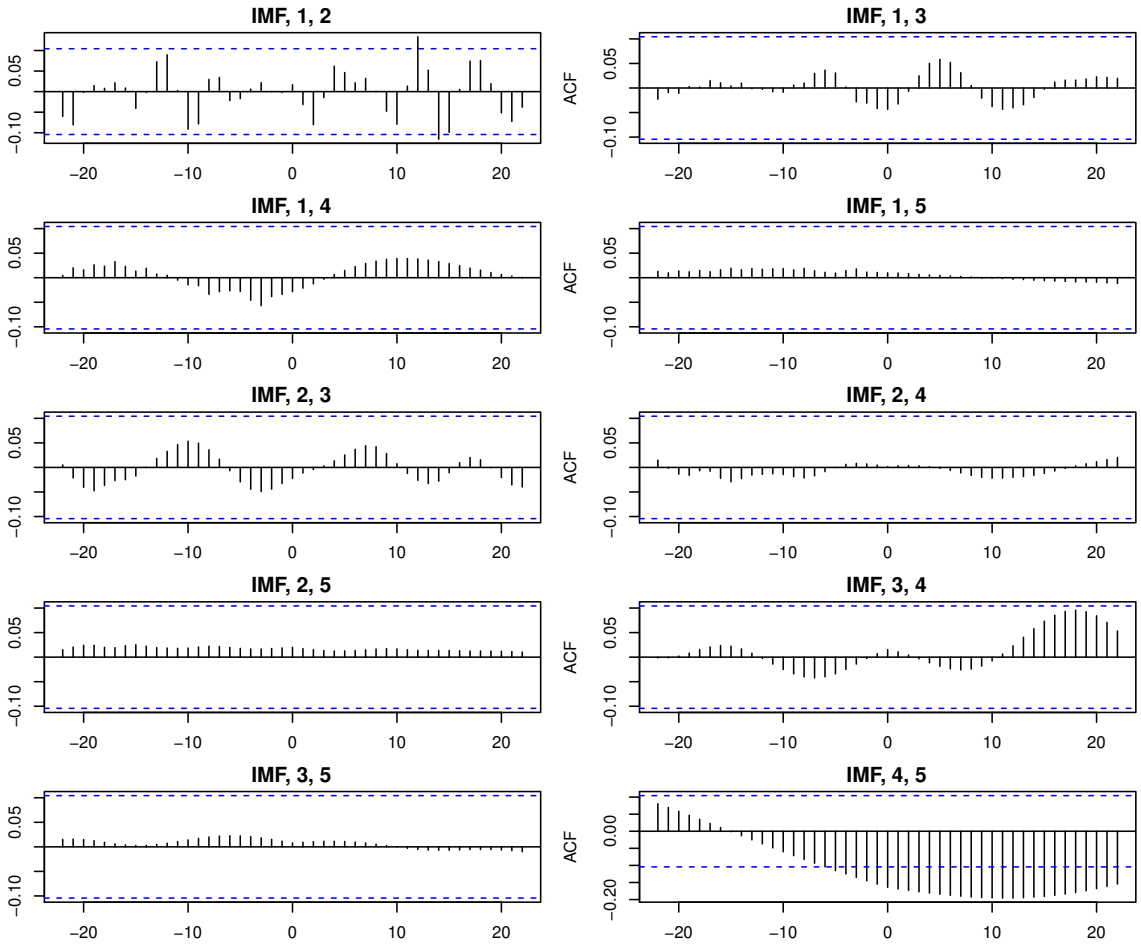


Figure 4.15. Cross-correlations of IMFs in Figure 4.14

Instantaneous frequency

In Figure 4.16 IF1, IF2, IF3, IF4 and IF5 are the plotted instantaneous frequencies of the IMFs in figure 4.14. Each of the plotted IFs displays the dynamics of the corresponding IMF without showcasing any frequency variation in any of the IMFs. Hence, a smooth transition throughout each IF is achieved. Therefore, since IFs have not shown any frequency variation

with smooth transition, then it implies that there is no mode mixing existing in any of the IMFs. Hence, all the extracted IMFs are in conformity with the definition of an IMF and authentic in describing the structure of the dataset.

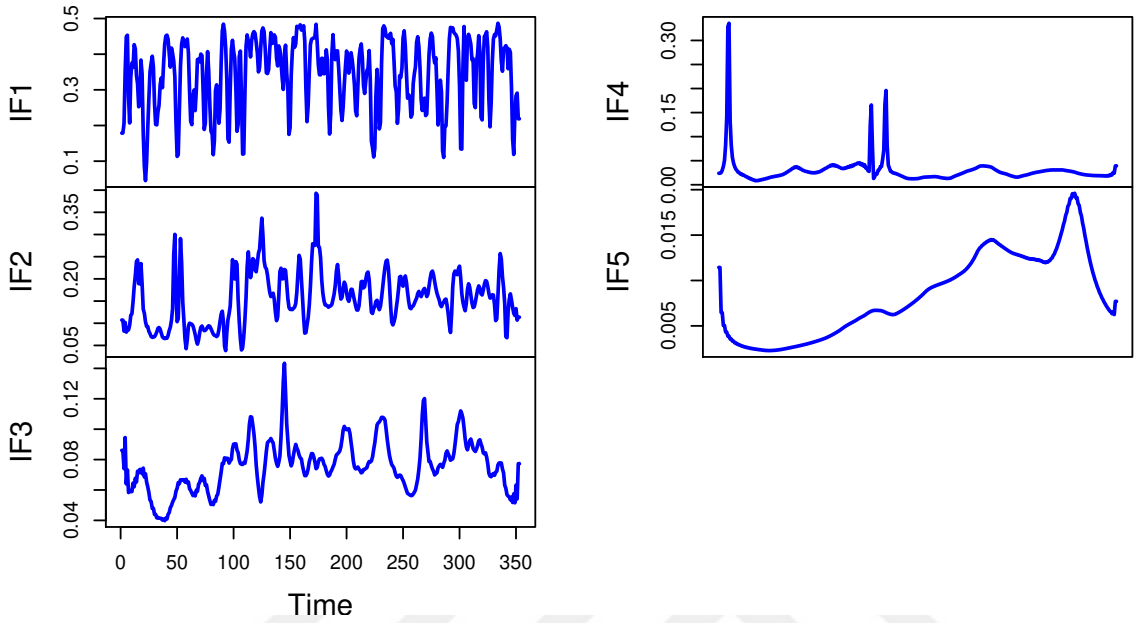


Figure 4.16. Instantaneous frequencies of IMFs in Figure 4.14

4.7.2. Data set of US dollar – Turkish lira exchange rate

This second data is the daily US dollar – Turkish lira exchange rate from 9th of February 2018 to 9th of July 2018 which is plotted in Figure 4.17 below. This data also possesses the properties of nonlinearity and nonstationarity, therefore meets the requirements for the implementation the EMD technique.

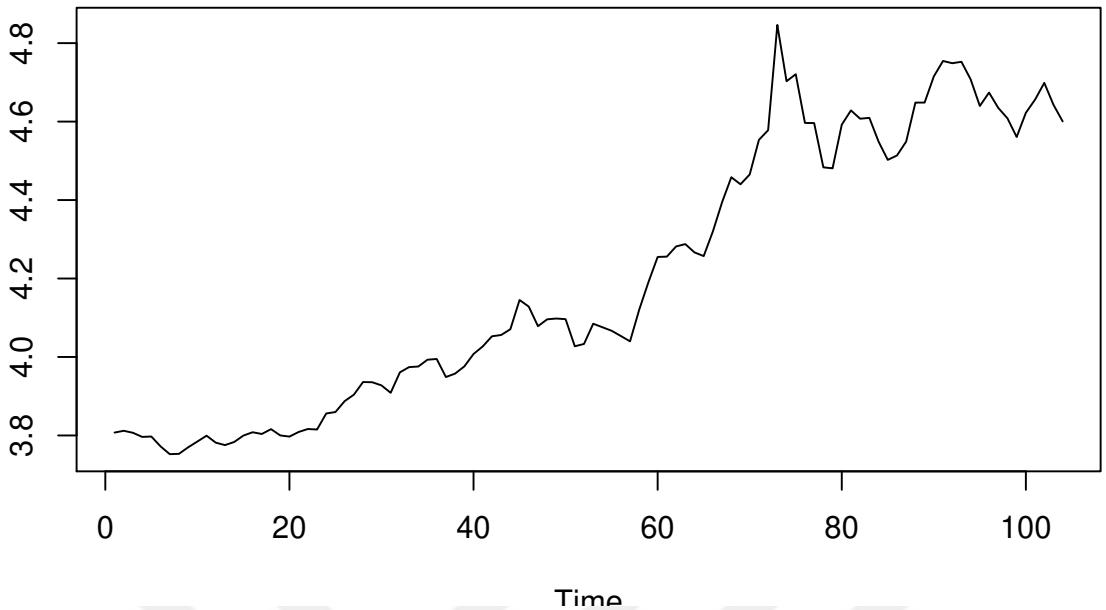


Figure 4.17. US dollar – Turkish lira exchange rate

Upon the implementation of the EMD technique on the US dollar – Turkish lira exchange rate data in Figure 4.17 above four IMFs were extracted as indicated in Figure 4.18. IMF1, IMF2 and IMF3 exhibits a synchronised steadily increasing volatility while IMF4 showcases cycles implying the existence of cycles within the US dollar – Turkish lira dataset. The residue shows the trend of the series indicating steadiness between the two currencies then a sharp increase portraying US dollar gaining value against Turkish lira and then a relative stability at the end between the two currencies. As done with previous dataset detection of mode mixing techniques will be implemented so as to establish whether or not there is any mode mixing in any of the extracted IMFs.

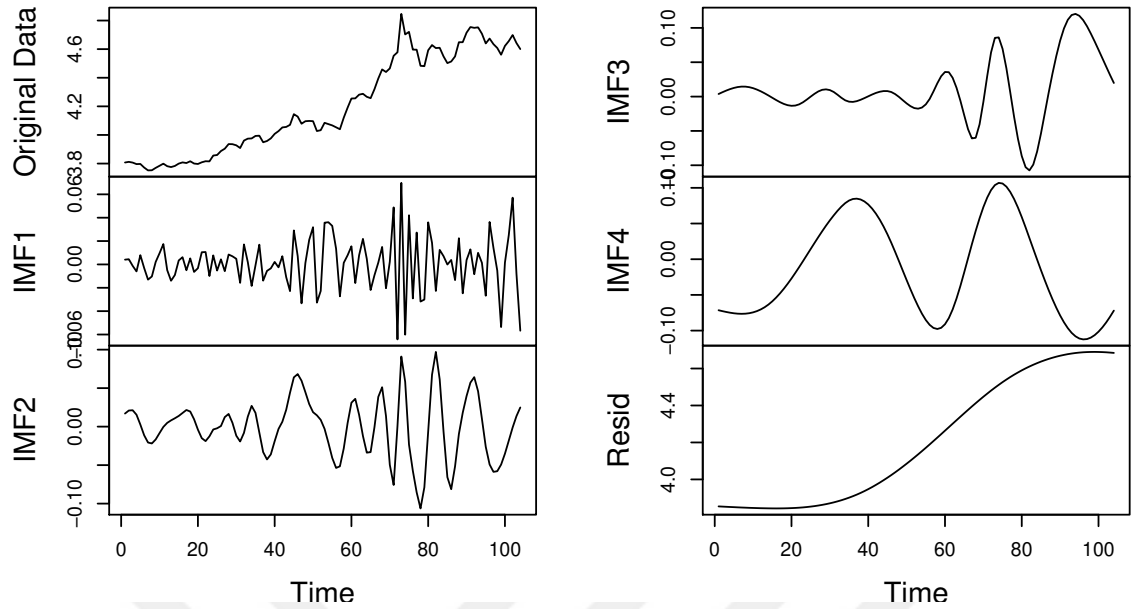


Figure 4.18. Empirical mode decomposition of data in figure 4.17

Illustration of detection of mode mixing problem

In an attempt to verify the existence or nonexistence of mode mixing in the IMFs extracted from the EMD process in Figure 4.18 mode mixing detection techniques like cross-correlation, instantaneous frequency will be applied.

Cross-correlation

Cross-correlation will be used to confirm whether there is any similarity between any two IMFs at every time shift from Figure 4.18 as indicated in Figure 4.19.The dashed horizontal lines represents the confidence interval and if the vertical lines on the time lags are within the confidence interval, it implies that there is no similarity between the IMFs. Hence, in Figure 4.19 all the cross-correlations are indicating no similarity between the IMFs with exception of the cross-correlation between IMF3 and IMF4. The wave movements of IMFs 3 and 4 can roughly align creating spurious similarity between IMF3 and IMF4 in terms of their cross-correlation. However, in reality there no concrete similarity between IMFs 3 and 4. Henceforth, we can conclude that there is no similarity between any two IMFs in Figure 4.18 based on the cross-correlation technique.

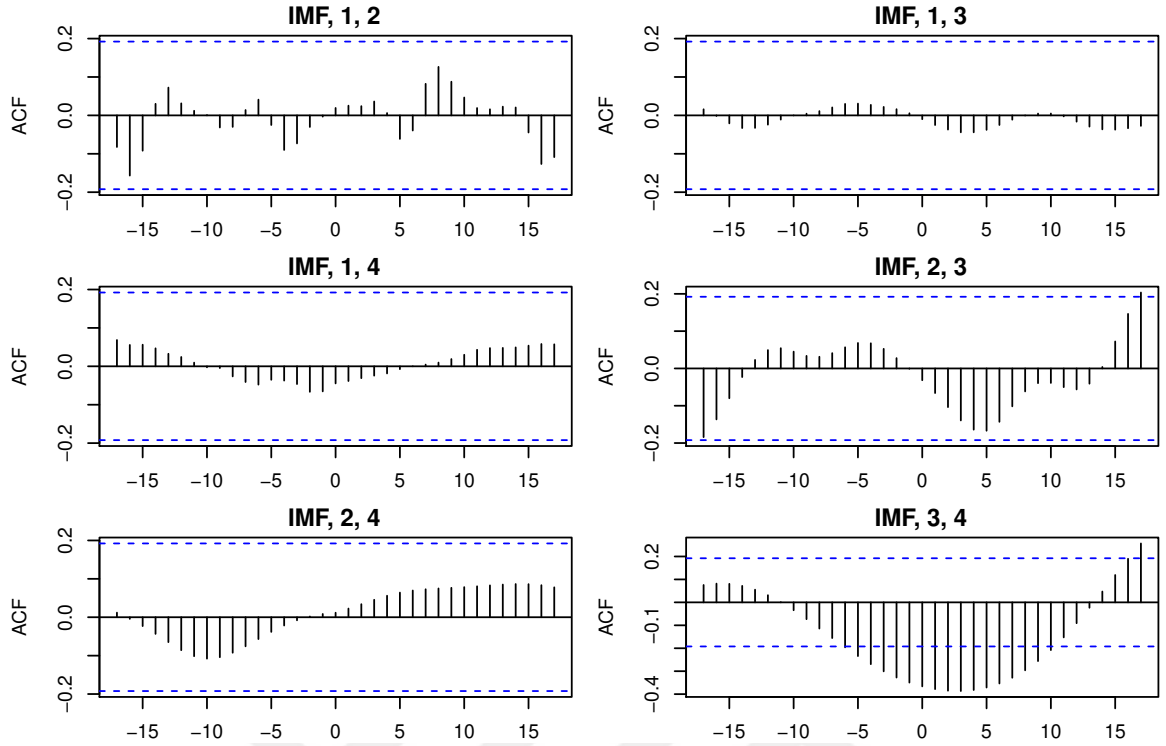


Figure 4.19. Cross-correlations of IMFs in figure 4.18

Instantaneous frequency

In Figure 4.20 IF1, IF2, IF3, and IF4 are the plotted instantaneous frequencies of the IMFs in Figure 4.18. Each of the plotted IFs displays the dynamics of the corresponding IMF without indication of any frequency disparity in any of the IMFs. Hence, a smooth transition throughout each IF is achieved. Therefore, since IFs have not shown any frequency variation with smooth transition, then it implies that there is no mode mixing existing in any of the IMFs. Hence, all the extracted IMFs are in conformity with the definition of an IMF and authentic in describing the structure of the dataset.

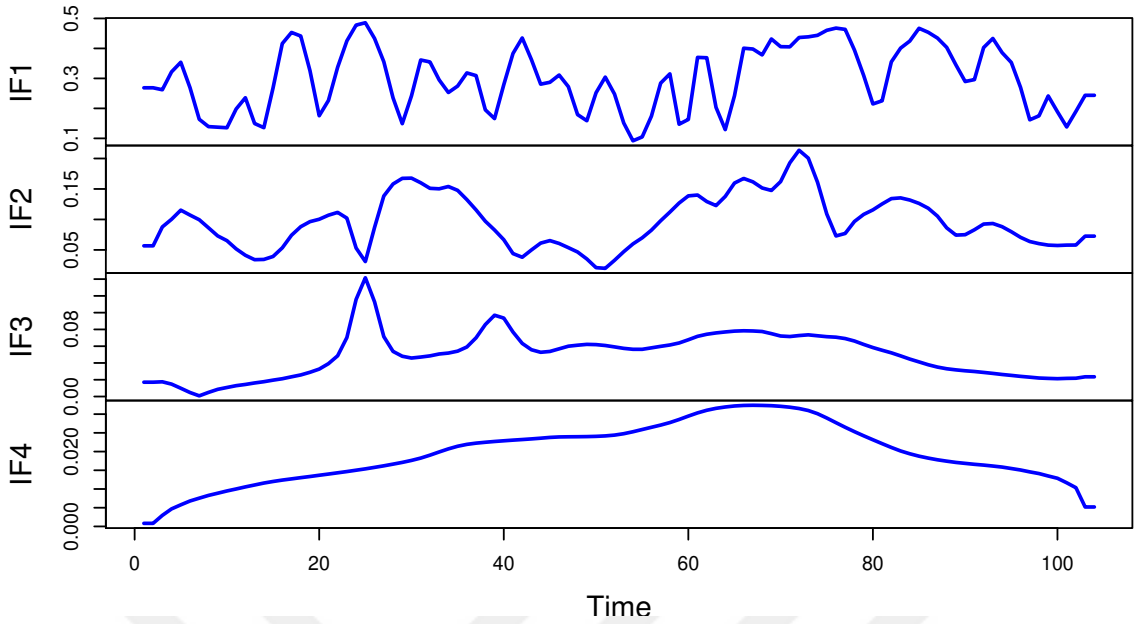


Figure 4.20. Instantaneous frequencies of IMFs in figure 4.18

4.8. Determination of Mode Mixing Problem by Signal Masking Method

Spectral tones mainly consist of frequency[1] and amplitude[2]. Hence, frequency and amplitude collectively determine the characteristic structure of the spectral tone. Spectral tones with equal amplitudes can only be distinquished by their frequencies. In such a situation, it is not feasible to distinguish between the tones as the difference between the frequencies of the spectral tones decreases. Therefore, as a solution to this problem signal masking methods have been developed in studies by Deering and Kaiser (2005), Fosso and Molinas (2017).

Signal masking can be applied to spectral tones with signals of similar structure and based on the structure of the IMFs this situation is one of the most suitable examples. As mentioned before, the problem of mode mixing exist as a result of the IMFs obtained from EMD consisting of the same scale. The similarity of IMFs with spectral tones leads to usage of signal masking techniques, thus gives an aternative method for detection of the problem of mode mixing.

[1]The frequency is defined as the frequency of repetition of the signal in 1 second.

[2]The amplitude is defined as half of the distance between the top of a wave and its peak. As its amplitude increases, its energy and intensity increase.

Method

Assume that an signal X is defined the sum of two sinusoidal functions as follows

$$X = A_1 sin(2\pi f_1 t) + A_2 sin(2\pi f_2 t) \ , \ \ t = [0, 2s] \tag{4.16}$$

Here the hypothetically generated signal X has mode mixing problem created as a mixture of close frequency spectral tones with the parameters given in Figure 4.23.

The mixed signal generated with close frequency spectral tones as shown in Figure 4.24 cannot be easily distinguished from the original signal. Such a situation will be observed in IMFs obtained from the EMD result. Signal masking method proposed by Fosso and Molinas (2017) is used in order to determine the existence of this problem and eliminate the problem even if the effect is low.

In order to implement the signal masking method, first the maximum (or minimum) signal which constitute the mixing signal is transformed and thus the frequency (f_m) is updated. A new spectral tone is generated with the new frequency obtained and two spectral tones are obtained by adding $\left(IMF_{j+}\right)$ and subtracting $\left(IMF_{j-}\right)$ this signal to the mixed signal X. Then EMD is applied to the masked signal, IMFs obtained are $\left(IMF_{j+} \ and \ IMF_{j-}\right)$ and the average of the IMFs is computed $\left(AIMF_j = \frac{IMF_{j+}+IMF_{j-}}{2}\right)$. As a result, the new IMF will be seperated futher according to the initial condition.

Instantaneous frequencies and instantaneous amplitudes obtained by Hilbert – Huang Transform of the new IMF estimates the frequencies of the spectral tones of the mixed signal. This process is introduced below.

The following expressions are derived from the instantaneous amplitudes of the transformed signal using Hibert Transform

$$K_{min} = \sqrt{A_1^2 + A_2^2 - 2A_1A_2} = (A_1 - A_2) \tag{4.17}$$

$$K_{max} = \sqrt{A_1^2 + A_2^2 + 2A_1A_2} = (A_1 + A_2) \tag{4.18}$$

Similarly, the expressions below are obtaiined from the extreme frequency values of the instantaneous frequency graph.

$$F_{min} = \frac{A_1 \Delta f}{(A_1 + A_2)} + f_2 \tag{4.19}$$

$$F_{max} = \frac{A_1 \Delta f}{(A_1 - A_2)} + f_2 \tag{4.20}$$

Here,

K_{min} : Minimum value of the amplitude plot

K_{max} : Maximum value of the amplitude plot

F_{min} : Minimum value of the instantaneous frequency plot

F_{max} : Maximum value of the instantaneous frequency plot

Δf : Difference between the two frequencies $(f_1 - f_2)$

With the help of Hilbert – Huang transform of the new IMF, inserting (K_{min}) and (K_{max}) of instantaneous amplitude in the Equations (4.17) and (4.18) an equation systems of first and second order with two unknown variables are established. The solution of the eqution system results in the values of A_1 and A_2. It is possible to obtain the value of Δf using similar processes with the help of (F_{min}) and (F_{max}) values and previously computed values of A_1 and A_2. At this point, using Δf, A_1 and A_2, then f_2 value is estimated from either Equation (4.19) or Equation (4.20), after obtaining the estimated f_2 value, then the estimated f_1 value can be computed.

Using these estimates, with the help of $\hat{f}_2/\hat{f}_1$ ratio and boundary map existence of mode mixing can be detected.

Boundary Map

The procedure of boundary map proposed by Deering and Kaiser (2005) and given in Figure 4.22 can be summarized as follows

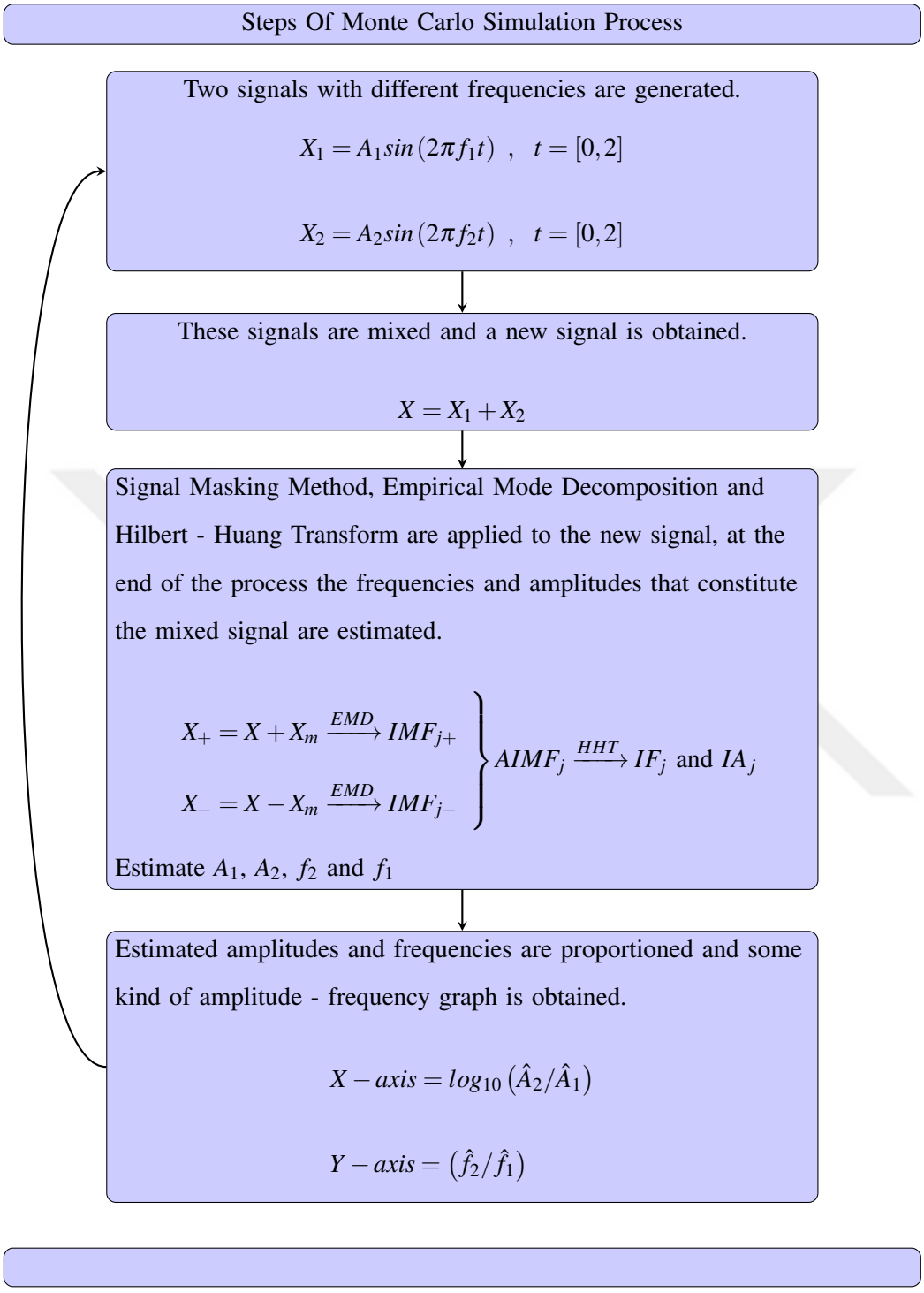


Figure 4.21. Vision Algorithm Flowchart

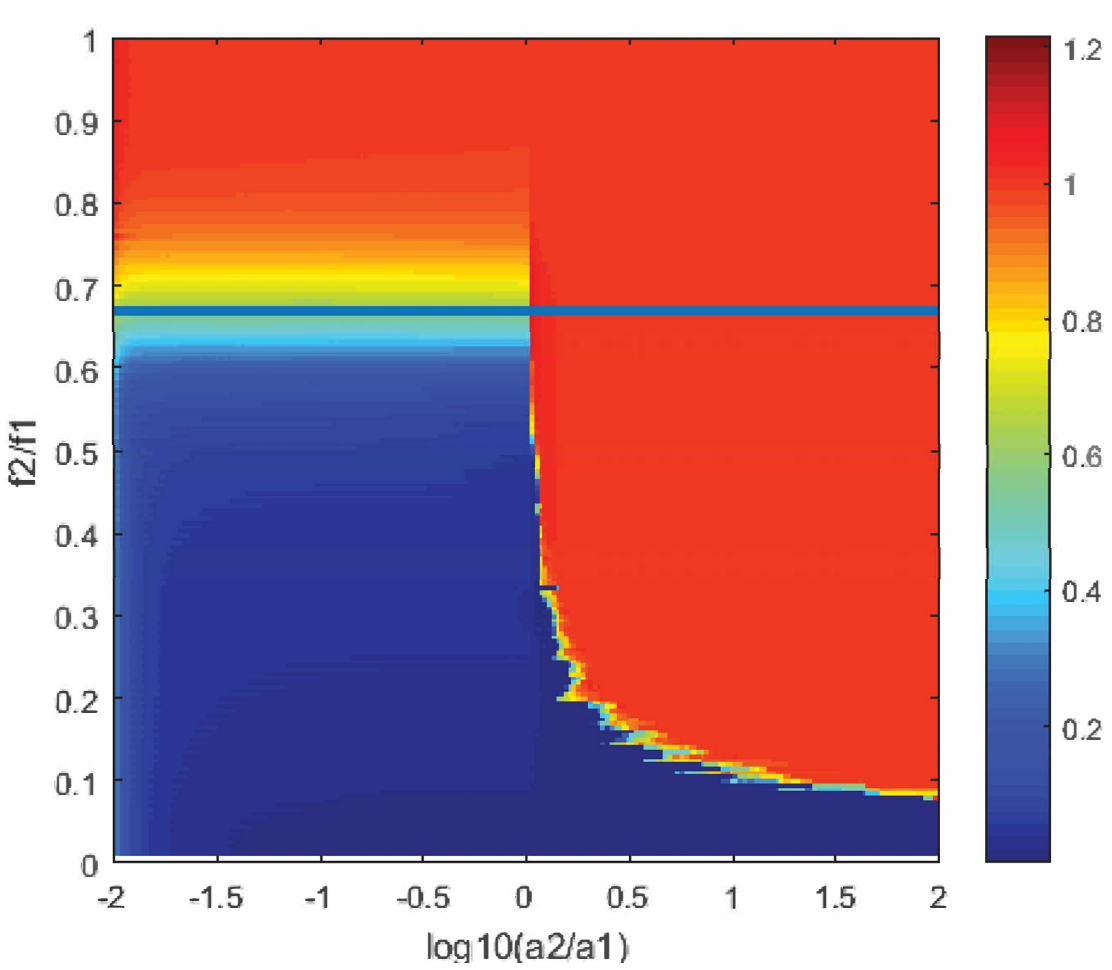


Figure 4.22. Boundary Map

4.8.1. Applications to simulated data

Simulation Study

Here the data with mode mixing problem and data without mode mixing problem generated through Figure 4.23, hypothetically how it is generated and how the above mentioned method is applied is given in an algorithmic format.

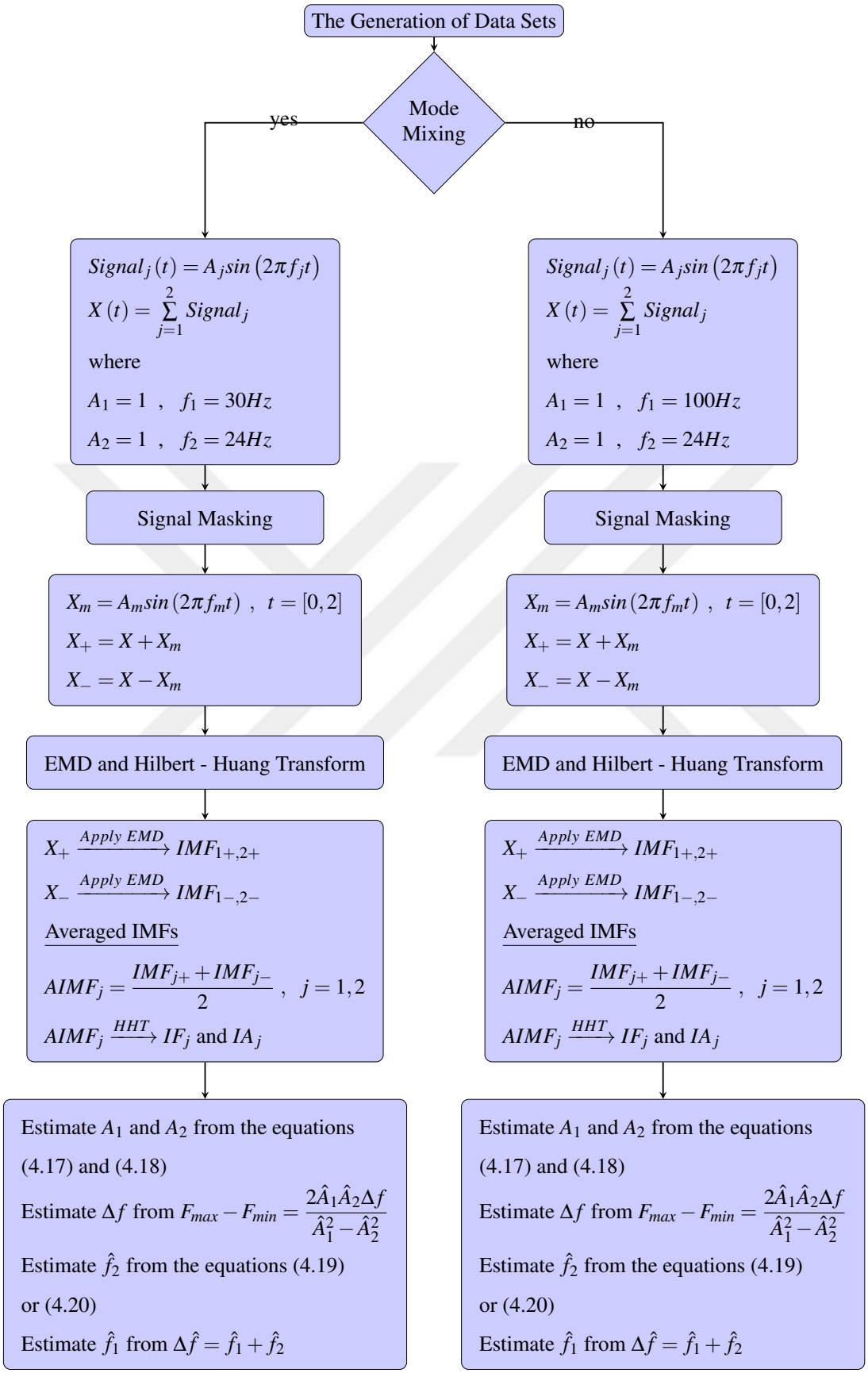


Figure 4.23. Flow chart of algorithm and simulation procedure

Simulation 1

In this part, the problem of mode mixing in time series is simulated by suggested signal masking based method.

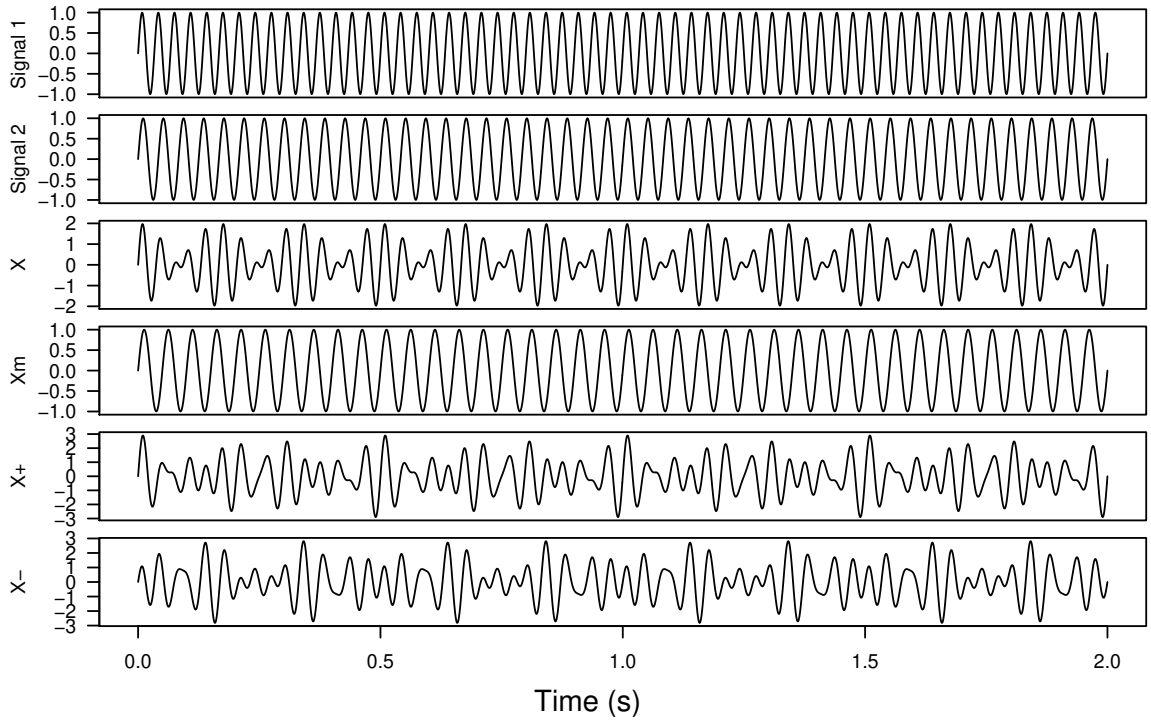


Figure 4.24. Original signal, mixed signal, masking signal and masked signals

Figure 4.24 displays the generated two spectral tones $(Signal_1, Signal_2)$ with close frequencies and signal (X) respectively, which is defined by the sum of these tones, masking signal (X_m) that will be used for signal masking process, positive masked signal (X_+) and negative masked signal (X_-) were illustrated.

The positive and negative masked signals given above are decomposed into IMFs by applying EMD procedure. In Figure 4.25 and Figure 4.26, empirical mode decomposition of positive and negative masked signal are illustrated, respectively.

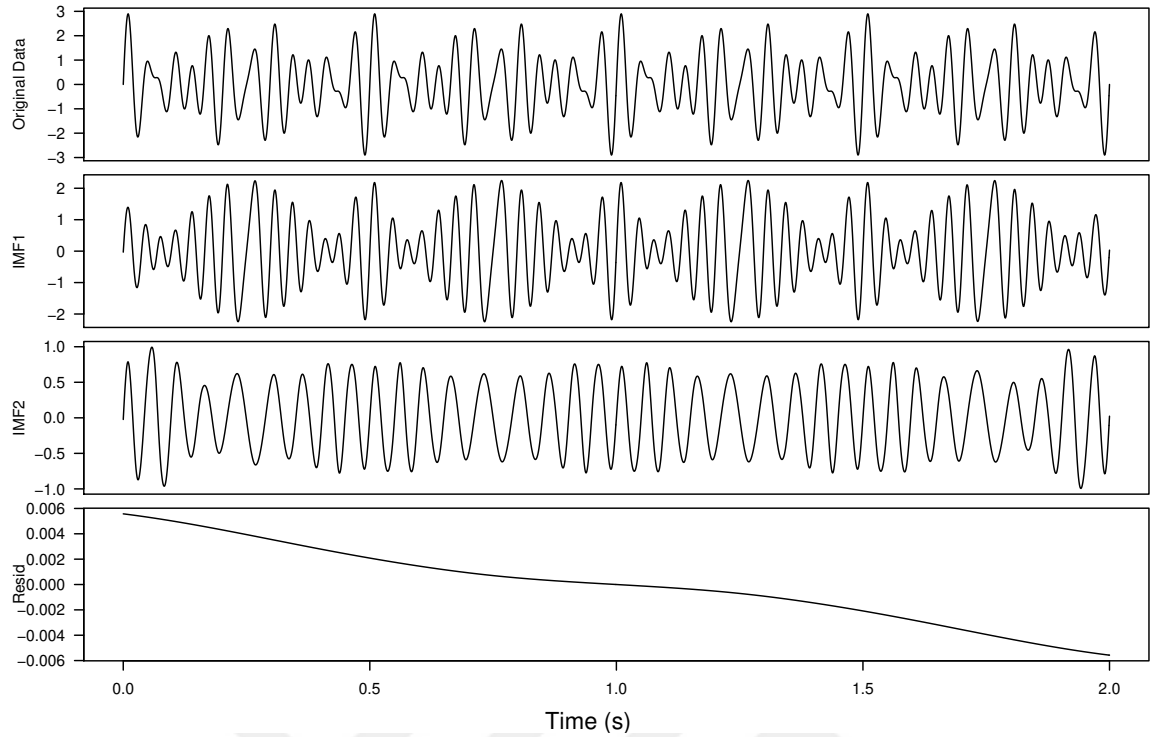


Figure 4.25. EMD for positive masked signal

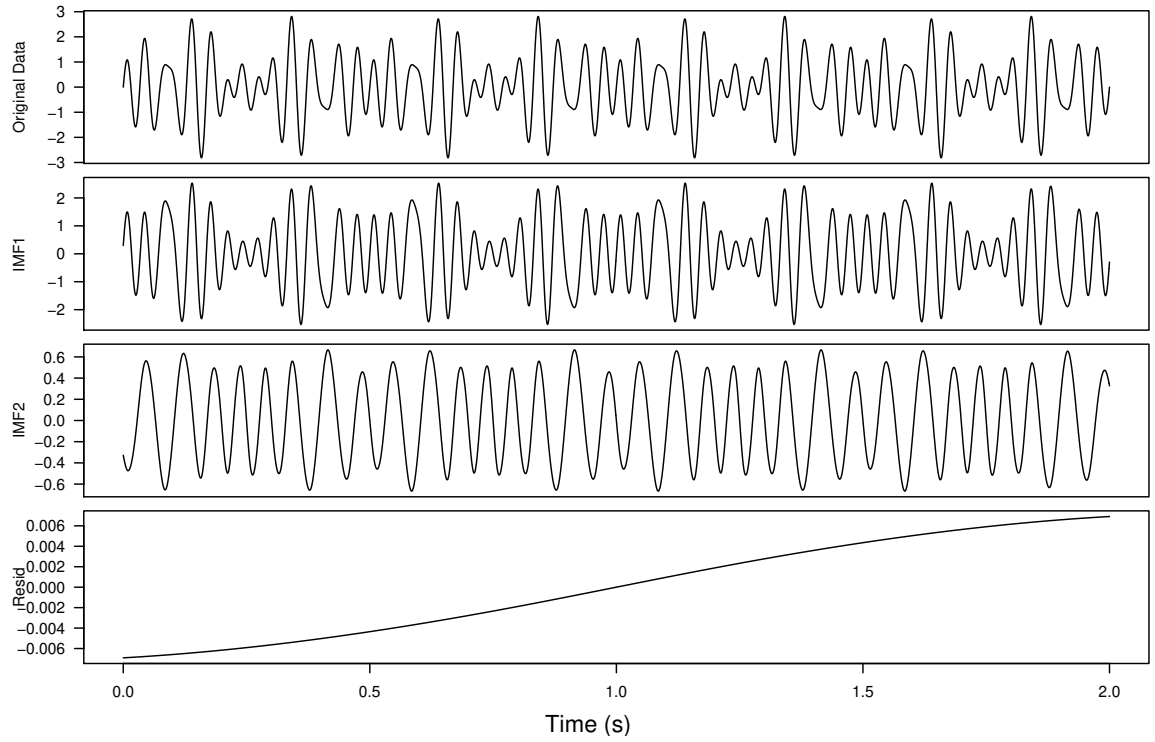


Figure 4.26. EMD for negative masked signal

Following the decomposition of IMFs, they are averaged and then by applying HHT the instantaneous frequencies are obtained as in Figure 4.27.

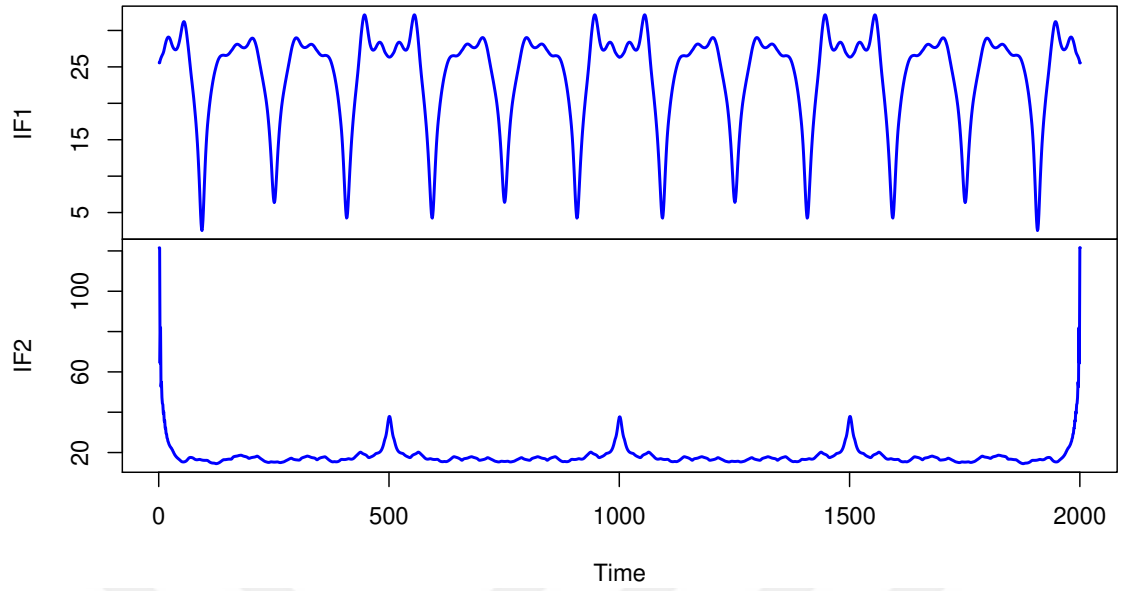


Figure 4.27. Instantaneous frequencies obtained from AIMFs

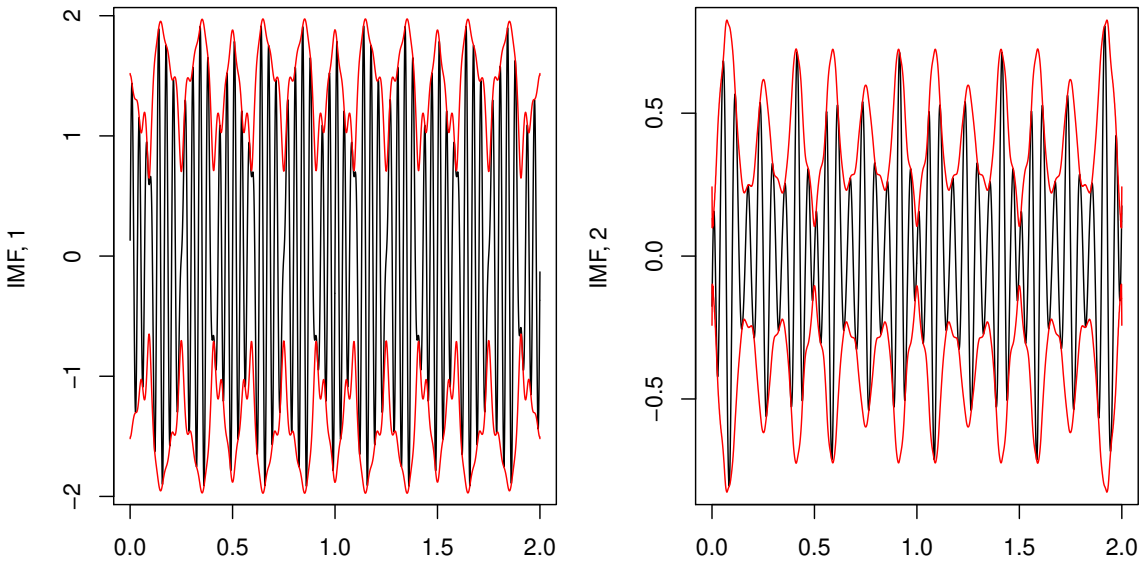


Figure 4.28. Hilbert envelops for AIMFs

In Figure 4.28, Hilbert envelops of AIMF were illustrated.

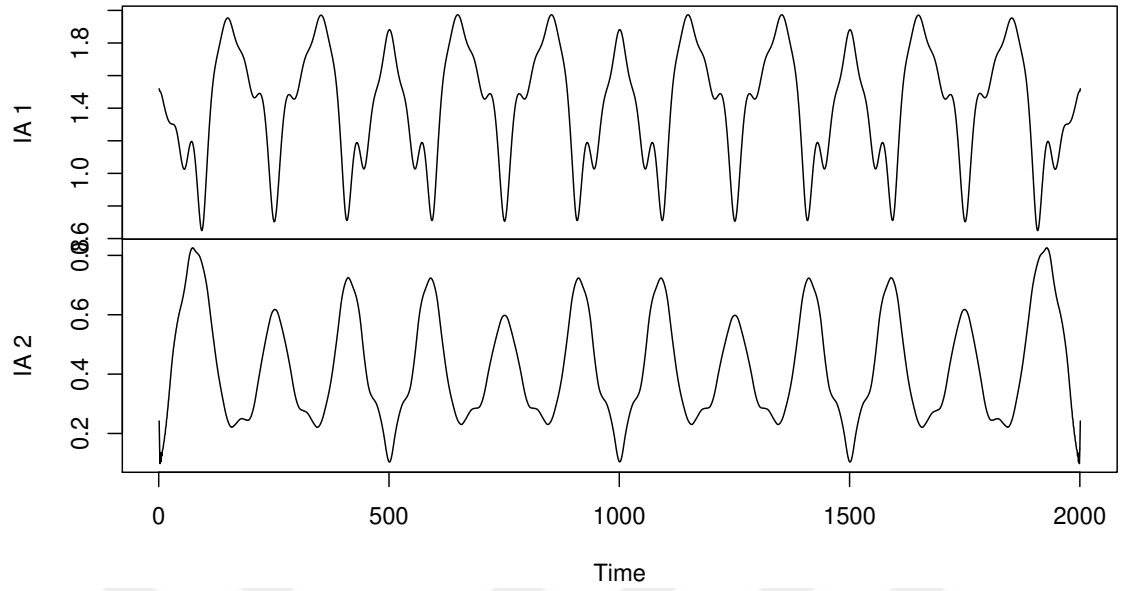


Figure 4.29. Instantaneous amplitude obtained from AIMFs

In Figure 4.29, instantaneous amplitudes obtained by HHT of AIMF were illustrated.

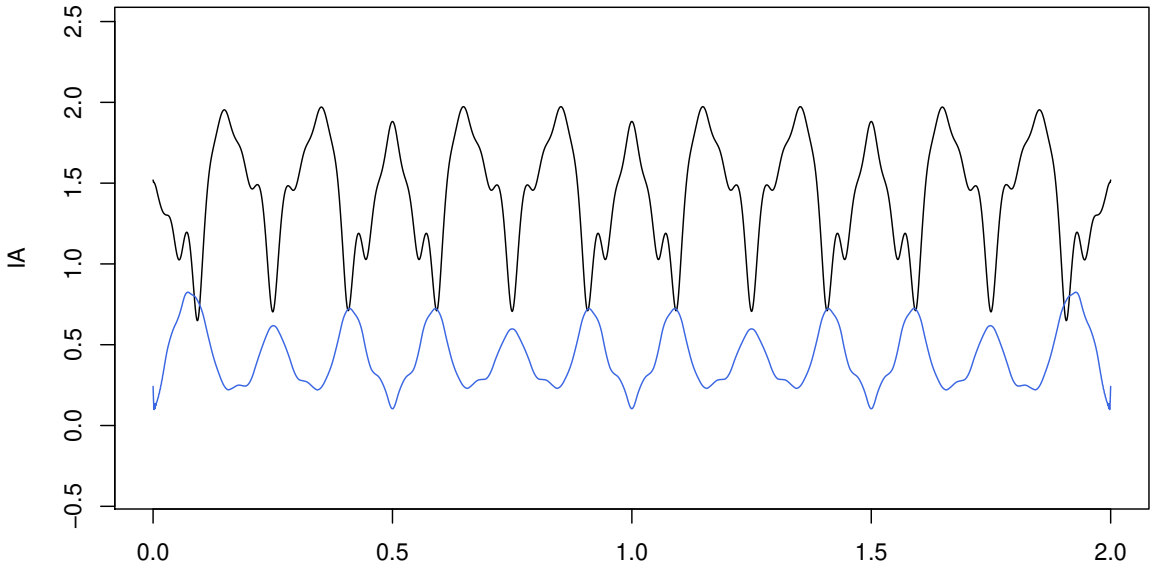


Figure 4.30. Instantaneous amplitudes obtained from AIMFs

In order to determine if two instantaneous amplitudes produced in the previous stage given in Figure 4.29 are superimposed, the resulting combined figures are displayed in Figure 4.30.

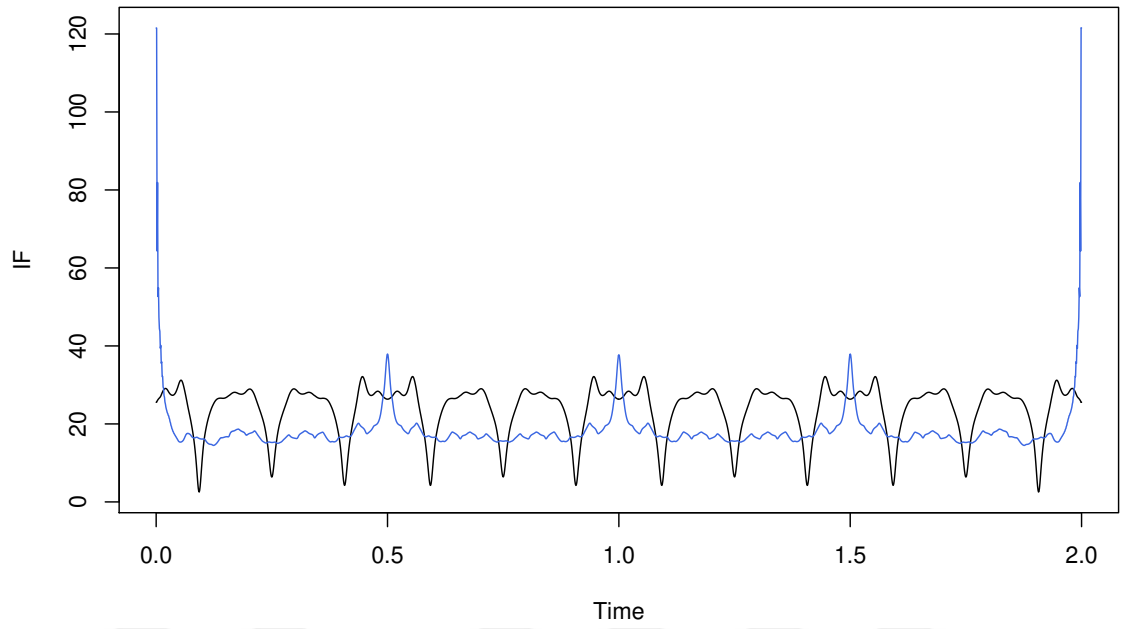


Figure 4.31. Instantaneous frequencies obtained from AIMFs

Similarly the IFs are combined into one graph displayed in Figure 4.31.

Examining Figure 4.30 and Figure 4.31, it is clearly observed that while the instantaneous frequencies are similar each other with respect to their structures, instantaneous amplitudes overlap with each other and thus are interpreted as similar with respect to their structures. This situation demonstrates the presence of mode mixing problem intuitively. To be concrete, it is also possible to calculate numerically through the procedure given Figure 4.23. The outputs of this alternative approach for determination of the mode mixing problem are given in Table 4.2.

Table 4.1. Values of parameters to estimate f_2/f_1 in Simulation 1

	K_{min}	K_{max}	F_{min}	F_{max}	f_1	f_2
Estimate Value	0.649	1.973	2.528	32.155	33.851	25.889
Real Value	0	2	0	30	30	24
$\hat{f}_2 \ / \ \hat{f}_1$	0.765					
$log10\left(\hat{A}_2 \ / \ \hat{A}_1\right)$	0					

The calculated values of $\hat{f}_2 \ / \ \hat{f}_1$ and $log10\left(\hat{A}_2 \ / \ \hat{A}_1\right)$ given in Table 4.2 shows that mode

mixing problem exists in the time series data since $\hat{f}_2 \ / \ \hat{f}_1$ falls into the red area in the boundary map given in Figure 4.22.

Simulation 2

In this part, time series which do not have mode mixing problem are simulated by suggested signal masking based method.

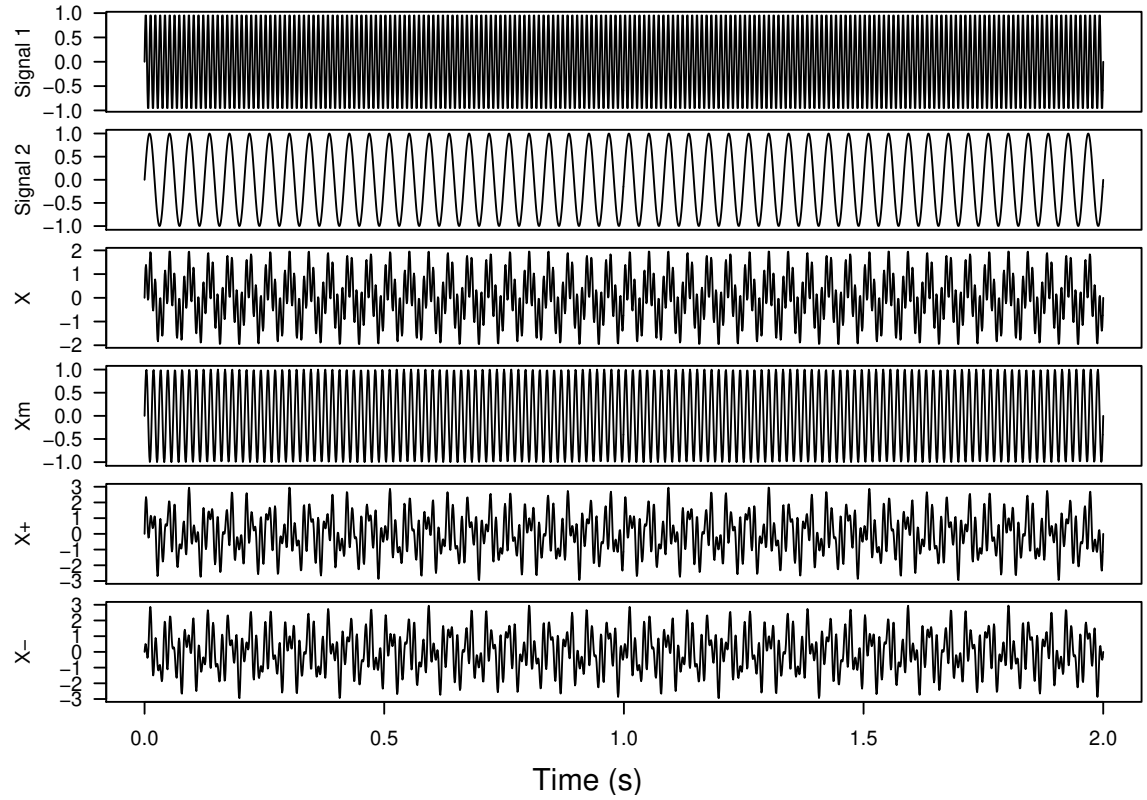


Figure 4.32. Original signal, mixed signal, masking signal and masked signals

The generated two spectral tones in Figure 4.32 displays signals with close frequencies as $(Signal_1, Signal_2)$ and (X) respectively, which is defined by the sum of these tones, masking signal (X_m) which will be utilized for signal masking process, positive masked signal (X_+) and negative masked signal (X_-) are demonstrated.

The IMFs extracted via the implementation of the EMD technique to the positive and negative masked signals are shown in Figure 4.33 and Figure 4.34, respectively.

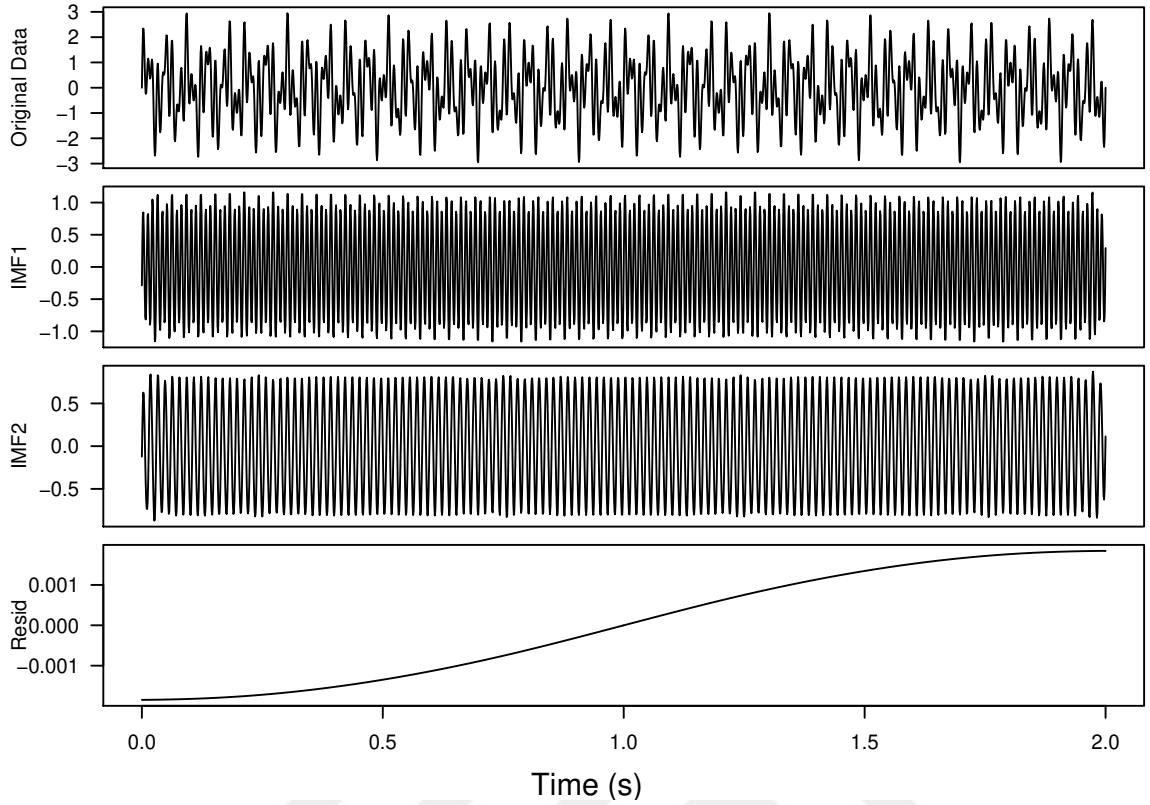


Figure 4.33. EMD for positive masked signal

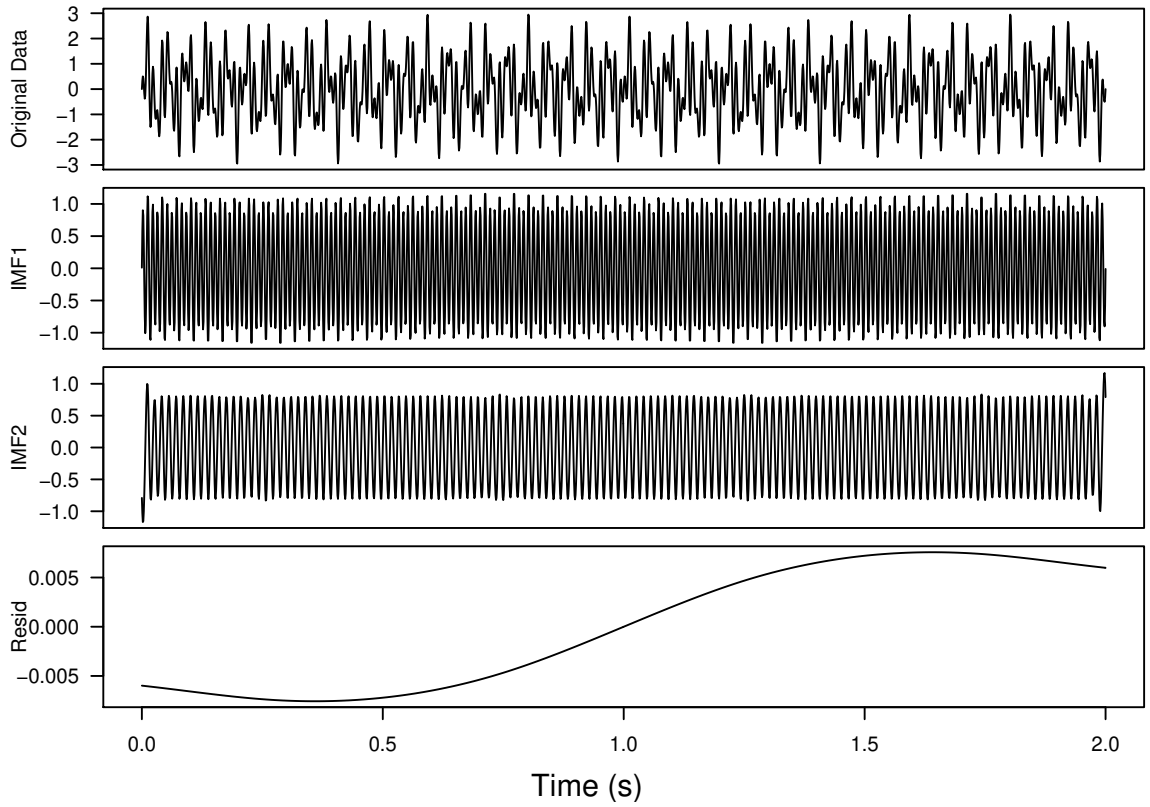


Figure 4.34. EMD for negative masked signal

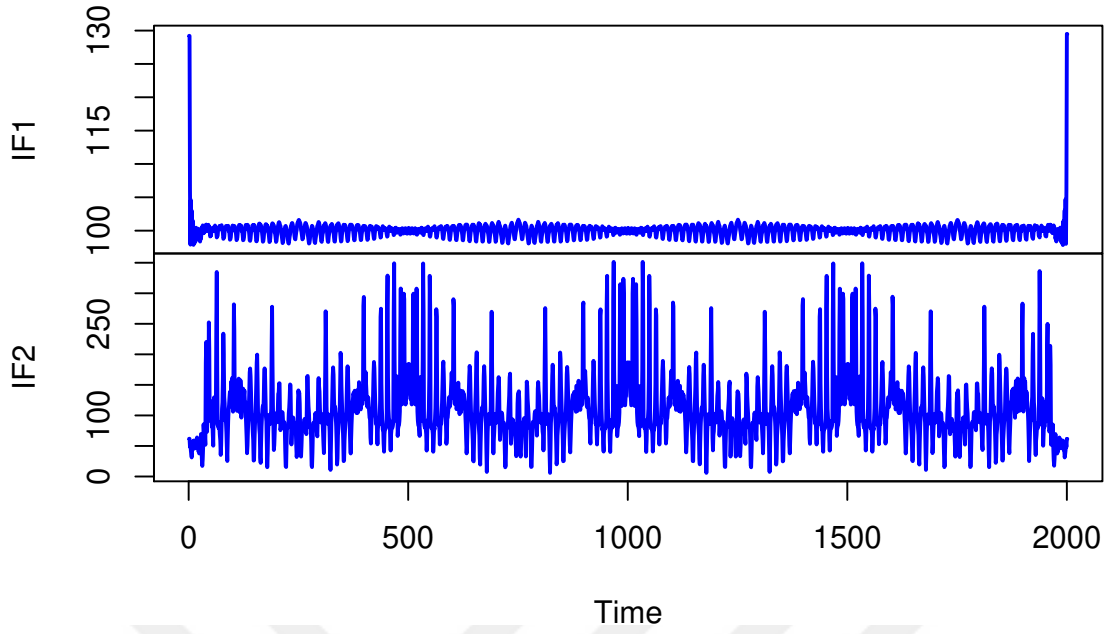


Figure 4.35. Instantaneous frequencies obtained from AIMFs

In Figure 4.35, instantaneous frequencies obtained by HHT of AIMF are illustrated.

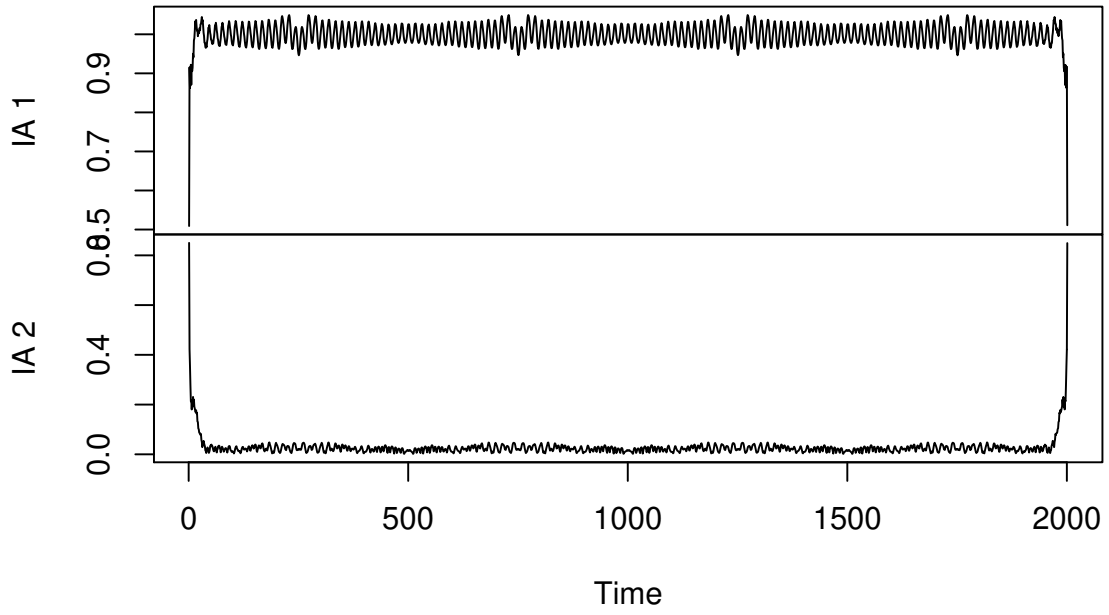


Figure 4.36. Instantaneous amplitude obtained from AIMFs

In Figure 4.36, instantaneous amplitudes obtained by HHT of AIMF are illustrated.

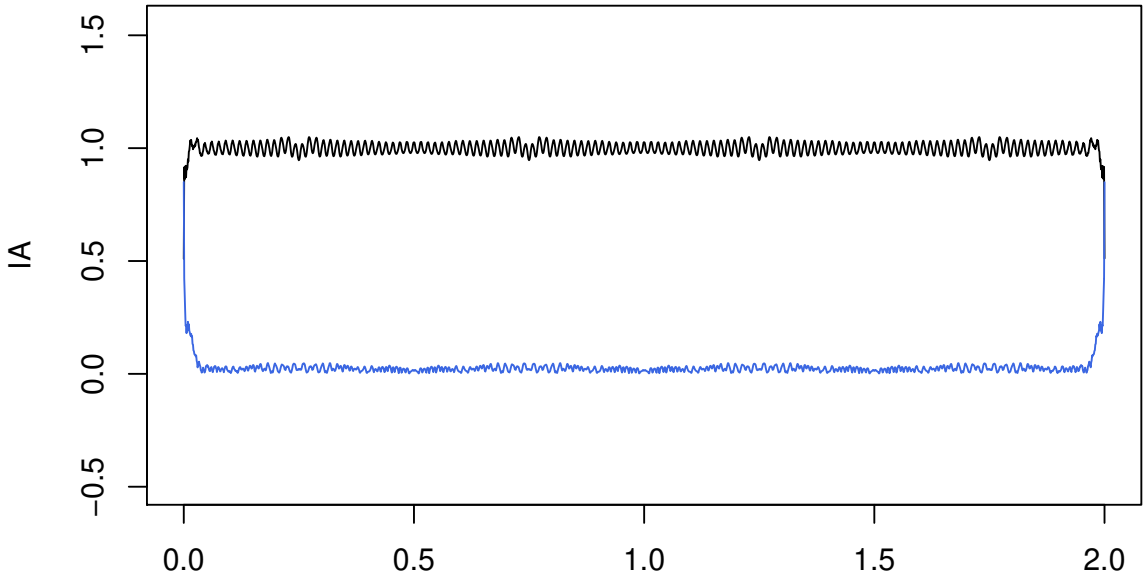


Figure 4.37. Instantaneous amplitudes obtained from AIMFs

In Figure 4.37, the combination of outputs in Figure 4.36 were illustrated in one graph.

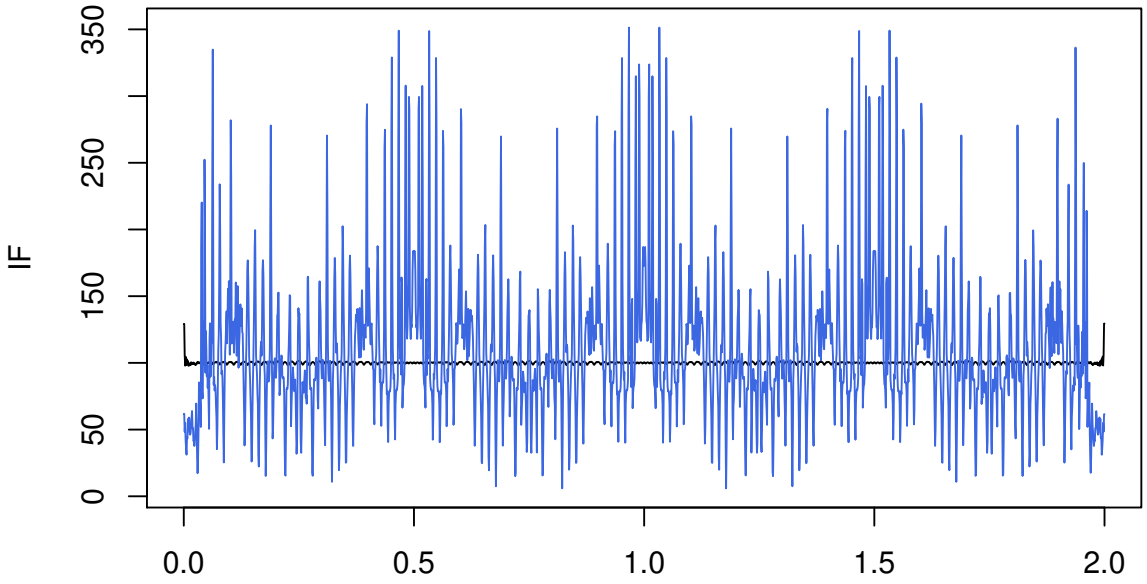


Figure 4.38. Instantaneous frequencies obtained from AIMFs

In Figure 4.38, the combination of outputs in Figure 4.35 were illustrated in one graph.

If Figure 4.37 and Figure 4.38 are examined, it is observed that while the instantaneous frequencies are not similar with aspect of their structures, instantaneous amplitudes do not overlap with each other. This situation demonstrates the absence of mode mixing problem

instuitively. The outputs of the alternative method for determination of the problem are given in Table 4.3.

Table 4.2. Values of parameters to estimate f_2/f_1 in Simulation 2

	K_{min}	K_{max}	F_{min}	F_{max}	f_1	f_2
Estimate Value	0.509	2.099	4.921	129.477	108.175	16.705
Real Value	0	2	0	100	100	24
f_2 / f_1	0.062					
$log10\left(\hat{A}_2 \,/\, \hat{A}_1\right)$	0					

The calculated values of $\hat{f}_2$ / $\hat{f}_1$ and $log10\left(\hat{A}_2 \,/\, \hat{A}_1\right)$ given in Table 4.3 shows that there is no apparent mode mixing problem in the data since $\hat{f}_2$ / $\hat{f}_1$ falls into the blue area in the boundary map given in Figure 4.22.

4.8.2. Application to real data set

In this part, the real data set of monthly total export in US Dolar of Turkey between 1989-2018 is studied to analyze whether the mode mixing problem exist in data by signal masking based method.

The method based on signal masking defined in Figure 4.23 can not easily be implemented to the total export data set. The reason for that sinusodial characteristic is not appearantly observed for this data set. However, we can mimic the similar behaiviour of seasonal data sets. It is well known that in the seasonal time series data, a repeated form of behaviour with a fixed period length is observed. The length of period in a seasonal data can be used as a required frequency of signal masking to implement the masking signal medhod defined here. However, only availability of frequency information is not sufficient to implement signal masking based method. In order to implement the method, the required information about amplitude is needed. This information can be obtained via instantenous amplitude resulted from IMF and HHT obtained by decomposing original data with EMD. The practice of procedure for the real data set is given in the following step by step.

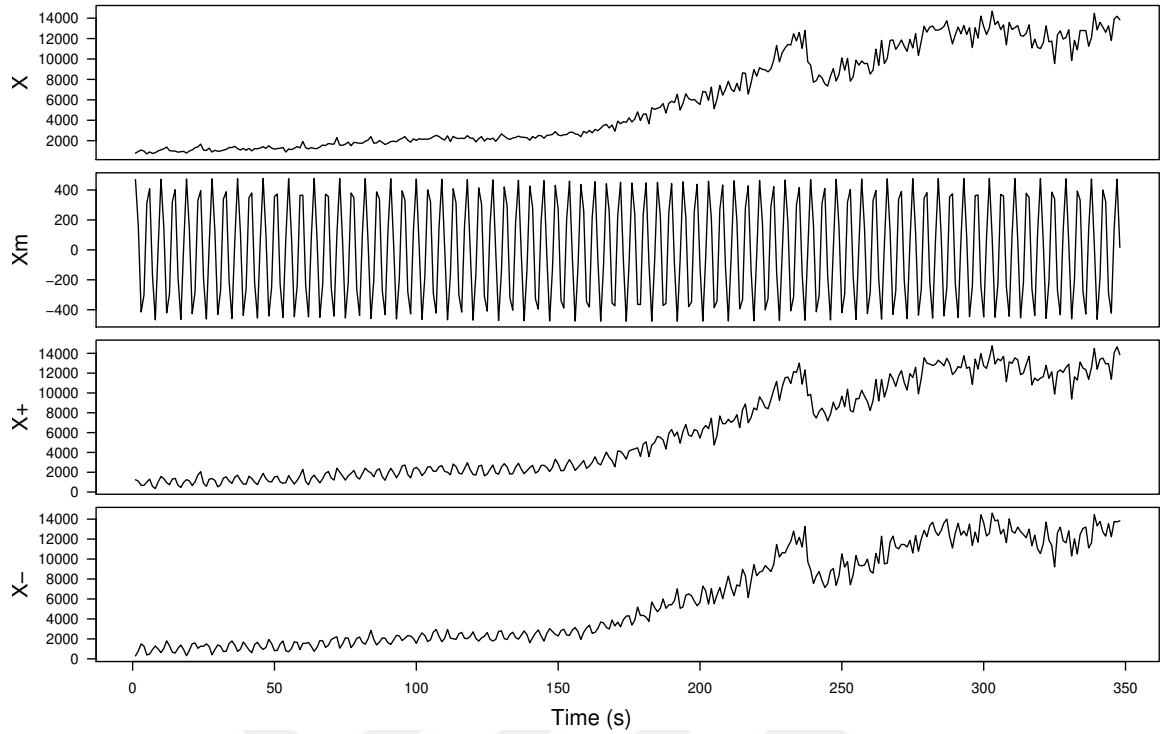


Figure 4.39. Original signal, mixed signal, masking signal and masked signals

In Figure 4.39, the original Total Export data (X), masking signal (X_m), positive masked signal (X_+) and negative masked signal (X_-) are given.

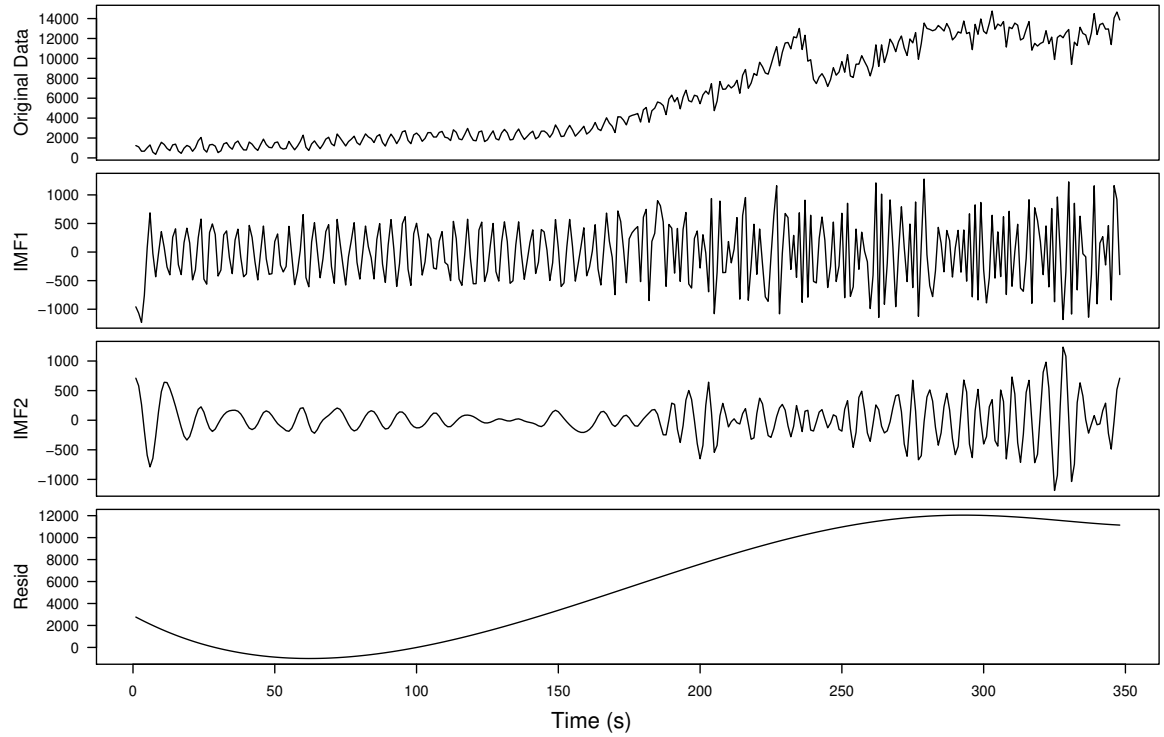


Figure 4.40. EMD for positive masked signal

In Figure 4.40, the empirical mode decomposition of positive masked signal is displayed.

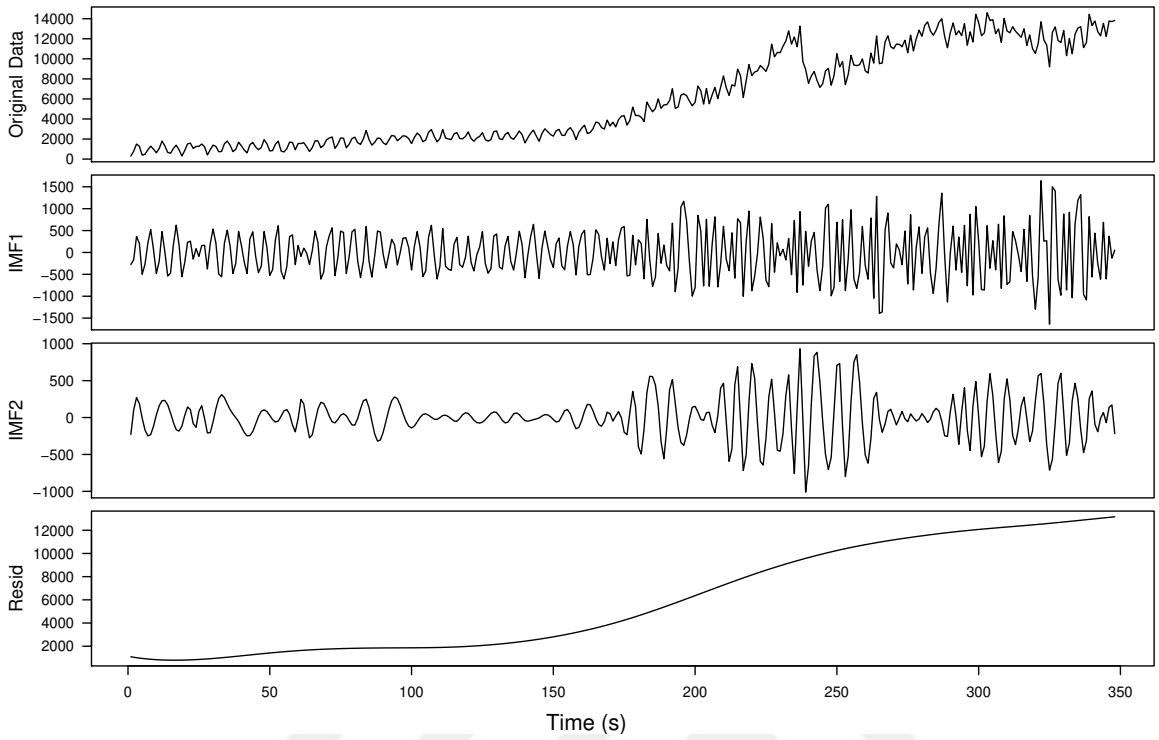


Figure 4.41. EMD for negative masked signal

Figure 4.41 shows the empirical mode decomposition of negative masked signal.

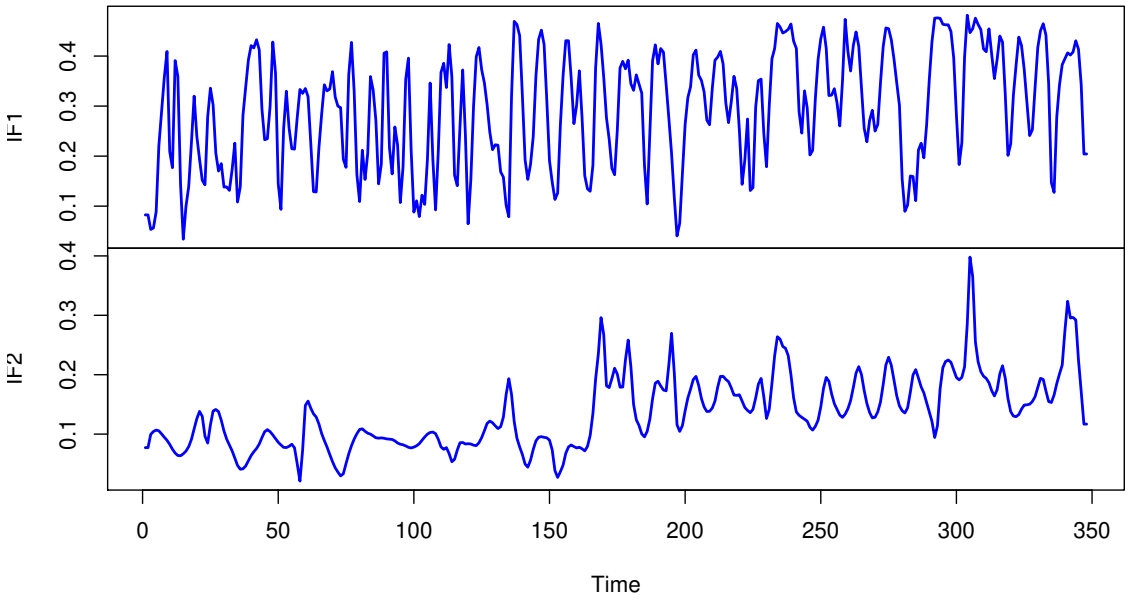


Figure 4.42. Instantaneous frequencies obtained from AIMFs

In Figure 4.42, the instantaneous frequencies obtained by HHT of AIMF are illustrated.

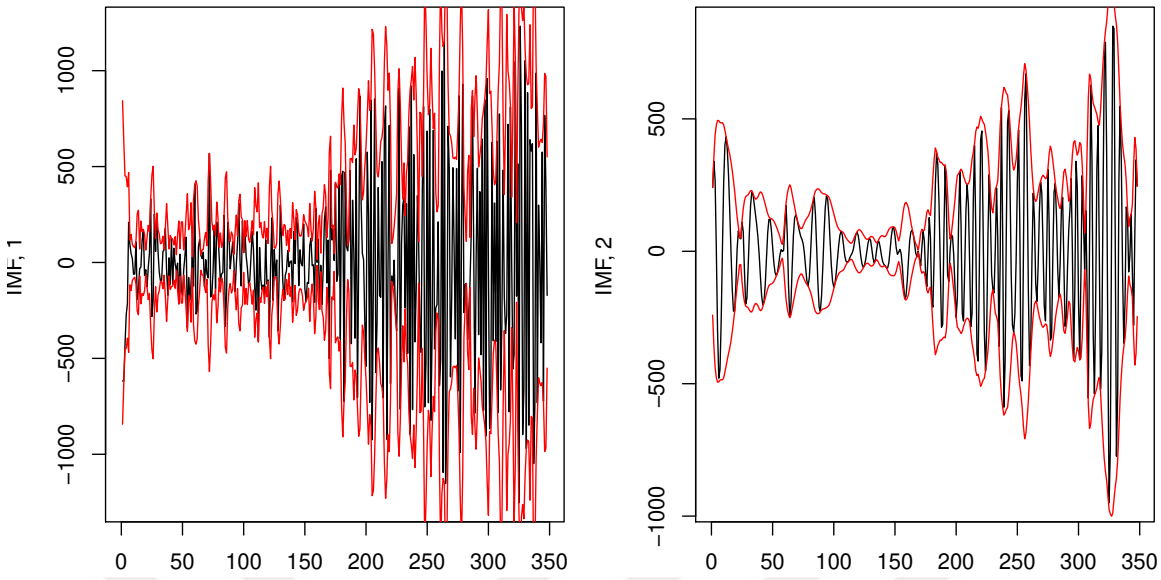


Figure 4.43. Hilbert envolops for AIMFs

Hilbert envelops of AIMF are given In Figure 4.43.

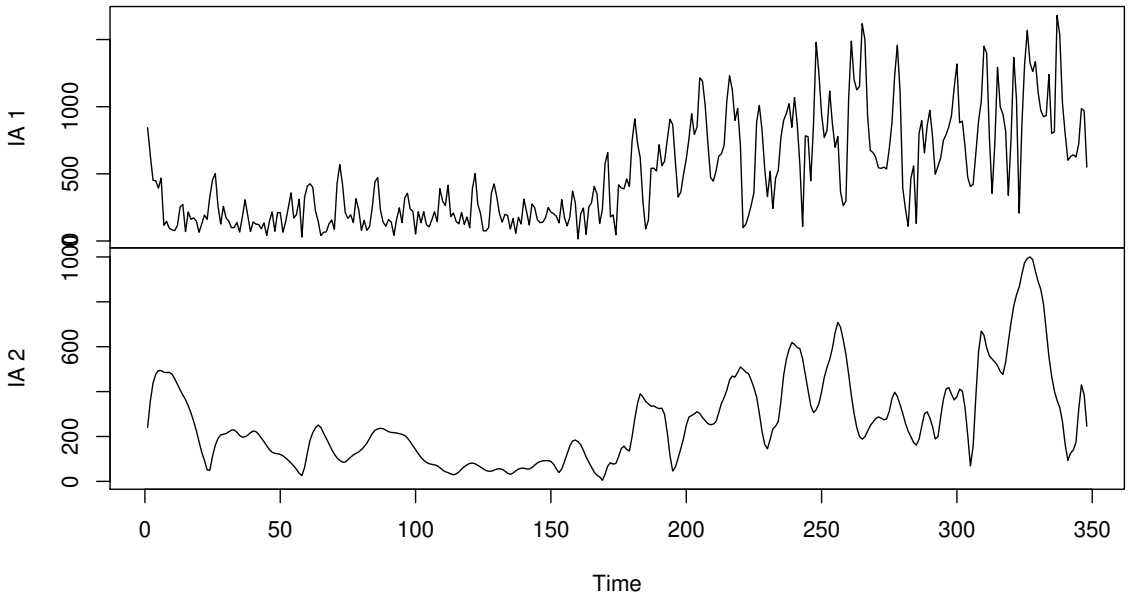


Figure 4.44. Instantaneous amplitude obtained from AIMFs

The following plot in Figure 4.44 is the instantaneous amplitudes obtained by HHT of AIMF.

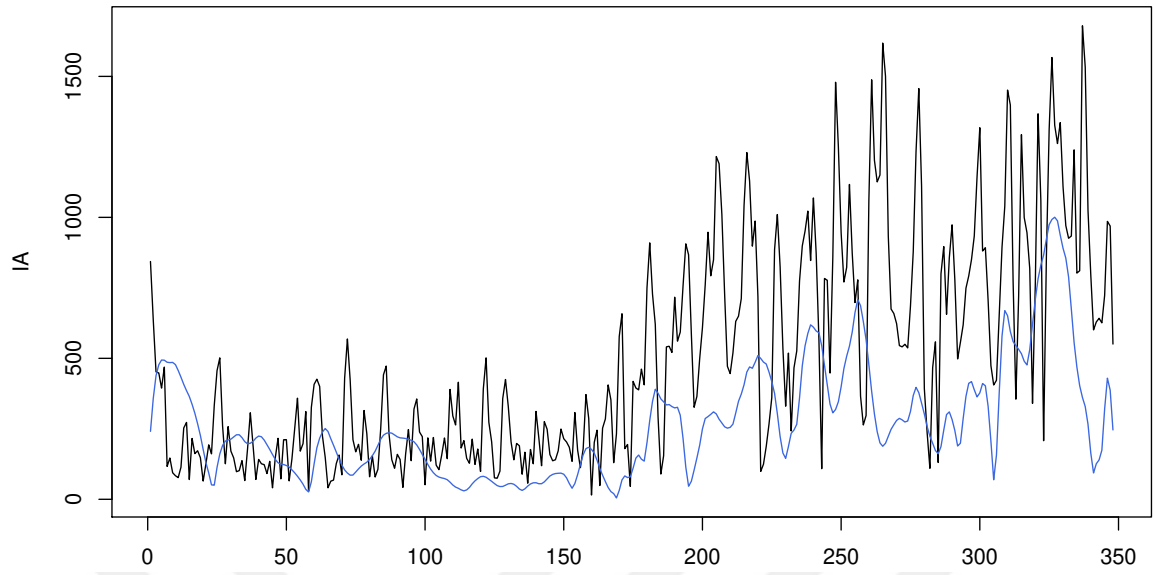


Figure 4.45. Instantaneous amplitudes obtained from AIMFs

In Figure 4.45, the combination of outputs in Figure 4.44 are illustrated in one graph.

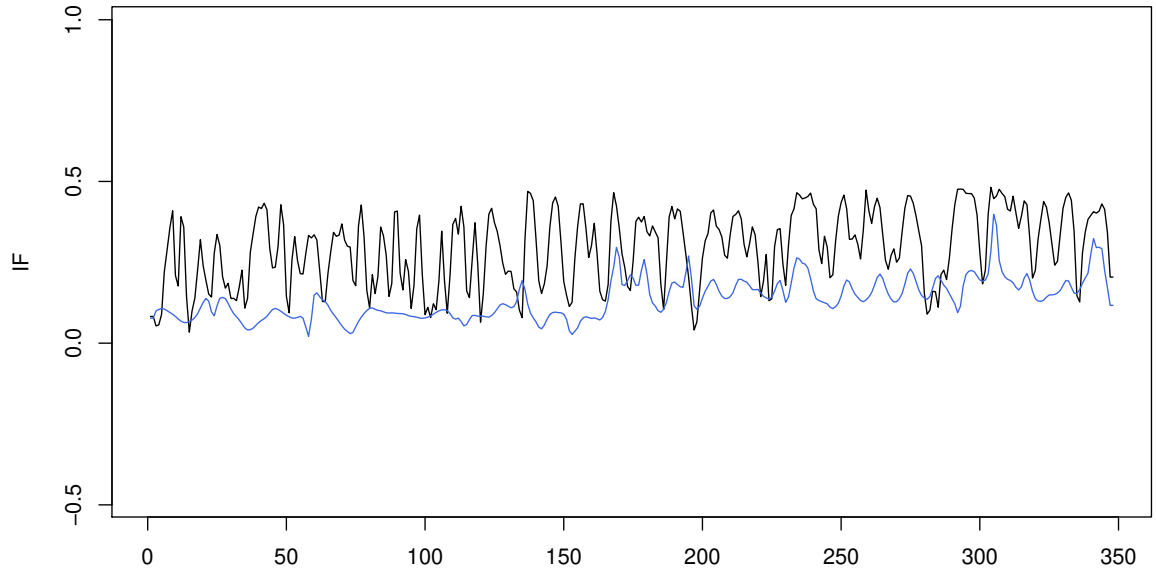


Figure 4.46. Instantaneous frequencies obtained from AIMFs

In Figure 4.46, the combination of instantenous frequencies in Figure 4.42 are depicted together.

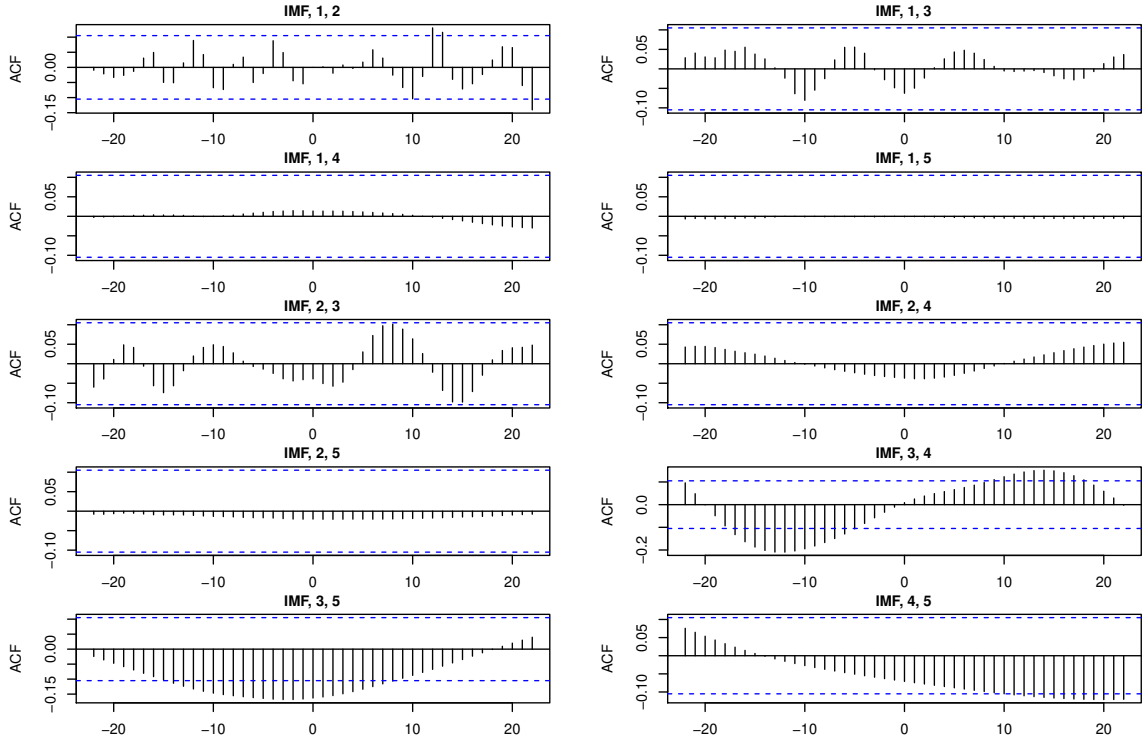


Figure 4.47. Cross correlation between AIMFs

In Figure 4.47, cross-correlations between AIMFs which are obtained by the EMD of the Total Export data are illustrated. As explained in subsection 4.5.1, cross-correlation between IMFs can be employed to detect mode mixing problem. The plotted cross-correlations in Figure 4.47 imply at first sight that there does not exist mode mixing problem in the data, since the first several cross-correlations are in the range of confidence limits.

However, looking at Figure 4.45 and Figure 4.46, it is observed that while the instantaneous frequencies are similar in terms of their structures, the instantaneous amplitudes overlap with each other. This situation demonstrates the presence of mode mixing problem intuitively.

Nonetheless, the numerical measures to interpret existency of mode mixing given in Table 4.4, it is concluded that there does not seem to be mode mixing in total export data set. As a conclusion, it can be said that the intutive decision based on plots would leads to misinterpretation of availabilty of mode mixing.

Table 4.3. Values of parameters to estimate f_2/f_1 in Real Data Set

	K_{min}	K_{max}	F_{min}	F_{max}	f_1	f_2
Estimate Value	4.516	1864.345	0.0941	0.4896	0.09501	0.02519
$\hat{f}_2 \ / \ \hat{f}_1$	0.2651959					
$log_{10}\left(\hat{A}_2 \ / \ \hat{A}_1\right)$	-0.1677212					

Looking at the coordinate of point $\left(\hat{f}_2 \ / \ \hat{f}_1, log_{10}\left(\hat{A}_2 \ / \ \hat{A}_1\right)\right) = (0.2651959, -0.1677212)$ given in Table 4.4 in the boundary map in Figure 4.22, it falls into blue area, no mode mixing area.

5. DISCUSSION AND CONCLUSION

Discussion

EMD which is an adaptive decomposition with which any complicated data (nonlinear and nonstationary) can be decomposed into its IMFs, corresponding to the physical reality that generated the data (Linderhed, 2004), is prone to the problem of mode mixing.

The results obtained in this thesis work elucidate the core concept and dynamics of the problem of mode mixing in EMD and effective techniques of detecting mode mixing in an IMF.

Previous research works on the problem of mode mixing neither presented the core concept of mode mixing nor delve into the techniques of detecting mode mixing in an IMF. For instance Fosso and Molinas (2017) did a brilliant work by categorising mode mixing into two categories: mode mixing caused by (1) presence of intermittence and (2) presence of closely spaced spectral components. However, Fosso and Molinas did not mention how the detection of mode mixing in an IMF can be executed. In a separate research, Verma (2016) without discussing neither the basis of mode mixing nor the detection method of mode mixing went straight ahead with the suggestion of a solution to mode mixing.

In this thesis, Figure 4.4 visualises the fundamental concept of how the problem of mode mixing prevails before jumping to suggestion of solution for it. Also Figure 4.5 displays explicitly how or what components derived from EMD of a dataset with the problem of mode mixing turns out. However, in real datasets to know whether mode mixing exist in a particular IMF might be a daunting task. In the light of this situation, in this thesis visualisation of the detection techniques of mode mixing were performed.

Future studies of mode mixing can focus on classifying mode mixing in terms of severity. For example high mode mixing, moderate mode mixing and low mode mixing or whatever the right nomenclature might be with regards to the degree of severity.

Conclusion

EMD decomposes nonlinear and nonstationary dataset into IMFs so as to depict all the individual characteristics embedded within the dataset in order to achieve credible prediction. However, despite its sophistication and merits, it is faced with the problem of mode mixing which leads to distortion of individual properties of the IMFs. The mode mixing problem is solved by the EEMD with the addition of significant white noise to the dataset. At the end of the EEMD process all the added white noise is expected to cancel each other. However, this is not always achieved. Therefore, the CEEMDAN is applied to finally get rid of all the added white noise. Moreover, detection of the problem of mode mixing in an IMF is a difficult task, hence some mode mixing detection techniques are applied to visualise or confirm its existence in an IMF.

The simulation studies performed indicate that mode mixing poses problem to EMD. Nonetheless, EEMD and CEEMDAN contain the issue of mode mixing and improve prediction reliability.

REFERENCES

1. Al-Naymat, G., Chawla, S. and Taheri, J. (2009). SparseDTW: A Novel Approach to Speed Up Dynamic Time Warping. *8th Australasian Data Mining Conference*, 101(1), 117–127.

2. Batista, G. E. A. P. A., Wang, X. and Keogh, E. J. (2011). A Complexity-Invariant Distance Measure For Time Series. *SIAM International Conference on Data Mining*, 1(1) 699–710.

3. Belmont, M. R. and Hotchkiss, A. J. (1997).Generalized Cross-Correlation Function For Enginnering Applications, Part I: Basic Theory. *Journal of Applied Mechanics*, 64(1), 321–326.

4. Blakely, C. D. (2005). A Fast Empirical Mode Decomposition Technique for Nonstationary Nonlinear Time Series. *Center For Scientific Computation and Modelling, University of Maryland*, 1(1) 1–14.

5. Boashash, B. (1992). Estimating and Interpreting the Instantaneous Frequency of a Signal. Part 1: Fundamentals. *Proceedings of the IEEE*, 80(4), 520–538.

6. Brooks, M. (2015). Correlation and Cross-Correlation. Fourier Series and Transforms - *Lecture Slides*, 1–11.

7. Crowley, P. M. (2009). How Do You Make a Time Series Sing Like a Choir ? Using the Hilbert-Huang Transform to Extract Embedded Frequencies From Economic of Financial Time Series. *Bank of Finland Research Discussion Papers*, 32(1), 12–40.

8. Dean, R. T. and Dunsmuir, W. T. M. (2016). Dangers and Uses of Cross-Correlation in Analyzing Time Series in Perception, Performance, Movement, and Neuroscience: The Importance of Constructing Transfer Function Autoregressive Models. *Behavior Research Methods*, 48(2), 783–802.

9. Deering, R., and Kaiser, J. F. (2005). The Use of a Masking Signal to Improve Empirical Mode Decomposition. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 4(1), 485–488.

10. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X. and Keogh, E. (2008). Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. *Proceedings of The VLDB Endowment*, 1(2), 1542–1552.

11. Flandrin, P., Goncalves, P. and Rilling, G. (2005). EMD Equivalent Filter Banks, From Interpretation to Applications. Hilbert-Huang Transform and Its Applications. *World Scientific*, 5(1), 67–87

12. Fosso, O. B. and Molinas, M. (2017). Method for Mode Mixing Separation in Empirical Mode Decomposition, *eprints in Physics, Mathematics, Statistics 1709.05547*,1(1), 1–7.

13. Gupta, A., Kumar, D., Chakraborti, A. and Sharma, K. (2016). Performance Evaluation of Empirical Eode Decomposition Algorithms For Mental Task Classification. *BioRxiv*, 1(1), 1–25.

14. Huang, N. E. and Shen, S. S. P. (2005). Hilbert-Huang Transform and its Applications. Transform, *World Scientific Publishing*, 5(1), 2–110.

15. Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Chen, X. and Blank, K. (2009). on Instantaneous Frequency. *Advances in Adaptive Data Analysis*, 1(2), 177–229.

16. Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Liu, H. (1998). The Empirical Mode Decomposition and the Hilbert Spectrum For Nonlinear and Non-stationary Time Series analysis. *Proceedings of The Royal Society*, 454(1971), 903-995.

17. Keogh, E. and Kasetty, S. (2002). On the Need For Time Series Data Mining Benchmarks. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2(1), 1–23.

18. Kim, D., Li, X., Li, D., Liang, Z., Voss, L. J. and Sleigh, J. W. (2006). Introduction to Empirical Mode Decomposition with Application to a Scientific Data. *Clinical Neurophysiology*, 119(11), 2465–2475.

19. Korkikian, R., Naccache, D. and de Almeida, G. O. (2013). Instantaneous Frequency Analysis. *Cryptology eprint Achirve, Report 2013/320*. 1(1), 1–23.

20. Linderhed, A. (2004). Adaptive Image Compression With Wavelet Packets and Empirical Mode Decomposition, Published Doctoral Thesis, Linköping University Department of Electrical Engineering, Sweden.

21. Rilling, G. and Flandrin, P. (2008). One or Two Frequencies?The Empirical Mode Decomposition Answers. *IEEE Transactions on Signal Processing*, 56(1), 85–95.

22. Spiegel, S. (2015). Time Series Distance Measures: Segmentation, Classification, and Clustering of Temporal Data (Doctoral dessertation, Berlin University, 2015). Dissertation Abstract International, 211.

23. Terrien, J., Marque, C. and Karlsson, B. (2011). Automatic Detection of Mode Mixing in Empirical Mode Decomposition Using Non-stationarity Detection: Application to Selecting IMFs of Interest and Denoising. *EURASIP Journal on Advances in Signal Processing*, 1(1), 1–8.

24. Torres, M. E., Colominas, M. A., Schlotthauer, G. and Flandrin, P. (2011). A Complete Ensemble Empirical Mode Decomposition With Adaptive Noise. In ICASSP, *IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 1(1), 4144–4147.

25. Verma, G. (2016). Solution to The Mode Mixing Problem in Two Harmonic Signals: Empirical Wavelet Transform. *International Journal of Engineering Research and Technology*, 5(6),154-157.

26. Wang, G., Chen, X. Y., Qiao, F. L., Wu, Z. and Huang, N. E. (2010). on Intrinsic Mode Function. *Advances in Adaptive Data Analysis*, 2(3), 277–293.

27. Wu, S. D., Chiou, J. C. and Goldman, E. (2010). Solution for Mode Mixing Phenomenon of the Empirical Mode Decomposition. *3rd International Conference on Advanced Computer Theory and Engineering, Proceedings*, 2(1), 1–6.

28. Wu, Z. and Huang, N. E. (2004). A Study of the Characteristics of White Noise Using the Empirical Mode Decomposition Method. *Proceedings of The Royal Society*, 460(2046), 1597–1611.

29. Wu, Z. and Huang, N. E. (2009). Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, 1(1), 1–41.

30. Xiao, J., Borgnat, P. and Flandrin, P. (2007). Testing Stationarity With Time-Frequency Surrogates. *European Signal Processing Conference*, 5(1), 2020–2024.

31. Xu, G., Zhengxiang, Y. and Wang, S. (2016). Study on Mode Mixing Problem of Empirical Mode Decomposition. *Joint International Information Technology, Mechanical and Electronic Engineering Conference*, 1(1), 389–394.

CURRICULUM VITAE

Personal Information



Surname, Name	: JAITEH, Muhammed I
Nationality	: Gambian
Date and Place of Birth	: 14.04.1991, Sankuya
Marital Status	: Married
Phone Number	: +905531565319
e-mail	: jaitehmuhammedi@yahoo.com

Education

Degree	School/Program	Graduation Date
Master's degree	Gazi University/Statistics	Ongoing
Graduate diploma	Met. Office College/Initial Forecasting Course	2013
Bachelor's degree	University of The Gambia/Mathematics	2012

Professional Experience

Year	Place of work	Position
2012 – 2015	Department of Water Resources, The Gambia	Weather Forecaster
2011 – 2012	Nusrat Senior Secondary School, The Gambia	Mathematics Teacher

Foreign Language

Turkish, English

Publications

-

Hobbies

Watching movie, Listening to music, Riding bicycle



GAZİ GELECEKTİR...