Allehiany, Faiza Mohammad (2012) *Frequency and time domain analysis of networks of interacting processes: what can be achieved from records of short duration.*

PhD thesis

http://theses.gla.ac.uk/3741/

Frequency And Time Domain Analysis of Networks of Interacting processes: What can be achieved from records of short duration

by

Faiza Mohammad Allehiany

A thesis submitted to the

School of Mathematics and Statistics

College of Science and Engineering

at the University of Glasgow

for the degree of

Doctor of Philosophy

November 2012

Abstract

Recently, there has been increasing interest in investigating the interrelationships among the component stochastic processes of a dynamical system. The applications of these studies are to be found in various fields such as Economics, Neuroscience, Engineering and Neurobiology. Also the determination of the direction of the information flow is one of the important subjects studied widely. These investigations have usually been implemented in the time and frequency domains. Consequently, several mathematical and statistical procedures have been developed to facilitate these analyses.

The aim of this thesis is to discuss the relationships between stochastic processes of a relatively short time duration. Specifically, the research concerns the analysis of the electrical activity of the dysfunctional brain, where the available data belong to a right-handed focal epileptic patient. EEG signals are recorded directly from the scalp using numbered electrodes according to the International 10/20 system introduced by Jasper [1958]. The analysis is only performed for processes of the left hemisphere as they represent the dominant hemisphere. Moreover, since each region of the brain is responsible for a special function, we have chosen five processes to represent the five main lobes of the brain; the frontal lobe, the central region, the parietal lobe, the occipital lobe and the temporal lobe.

The analyses of these signals are carried out using four spectral density estimation procedures, namely the multivariate autoregressive model of order 2; the average of periodograms of adjacent segments of the single record; the smoothed periodogram approach for the entire record; and the multi-taper method. Thereafter comparisons among the results of these methods are made. The strength of the correlation between signals is measured by coherence and partial coherence functions. Also, the Granger causality concept is implemented for these data in the form of determining the direction of the information flow between these signals using the partial directed coherence (PDC) proposed by Baccalá

and Sameshima [2001] using the statistical level of significance suggested by Schelter *et al.* [2005]. The structure of the causal influences produced by the PDC shows that there are statistically significant reciprocal direct causal effects between processes representing the brain's region, the frontal lobe, the central area, the parietal lobe and the temporal lobe. However, there are two uni-directed causal influence relations, one is between the central area and the occipital lobe and the second one is between the occipital and temporal lobes. The indirect causal influences are detected between these processes throughout the process representing the temporal lobe. Generally, the values of the PDC in the anterior-posterior direction are larger than the values of the PDC in the opposite direction. Also, the causal influences of each process on the temporal lobe process is larger than the causal influences in the opposite direction.

The spectral analyses show that the estimated power spectra and coherences of these signals are approximately peak in the δ-wave band of frequency $[1, 4)$ Hz. The significant non-zero estimated coherences are captured between the brain's lobes except for the occipital lobe which is uncorrelated with any of the other lobes. The depth of non-zero significant estimated coherences is given by partial coherence, which measures the strength of the estimated coherence between any two processes after removing the linear influence of one or more other processes. For the current data, we found that the depth of correlations depends on the spectral estimation method adopted. For example, the depth of correlation is of order 2 for the method of averaging across periodograms of adjacent segments of the single record and the method of smoothed periodogram of the entire single record and is of order one for the multi-taper method. However, the depth of correlations is unknown for the multivariate autoregressive model of order 2. The comparisons made between the results of the four spectral estimation methods mentioned previously, indicated that MVAR is not sensitive to rapid changes occurring in the signal such as the effect of the notch filter at 60Hz and a calibration signal at 47Hz, while the other three methods exhibited good sensitivity to these changes with different strengths of responses. Furthermore, the smoothed periodogram and the multi-taper methods persistently detect the notch filter effect at 60Hz in the ordinary estimated coherence curves, while the method of averaging across periodograms of adjacent segments of the single record does not.

Acknowledgements

Statement

This thesis is submitted in accordance with the regulations for the degree of Doctor of Philosophy at the University of Glasgow, UK. Chapter 1 includes an introduction and literature review. Chapter 2, 3 and 4 involve preliminaries and theorems used throughout the thesis. Chapter 5, 6 and 7 are the author's original work conducted under the supervision of Prof. Kenneth Lindsay, unless explicitly mentioned.

Contents

C **222**

List of Figures

Chapter 1

Introduction

Interactions among large numbers of stochastic processes are commonplace and arise naturally when a number of different input or control processes combine together to produce output processes. Examples of this phenomenon are ubiquitous in Neurophysiology, Economics and in several biological systems.

One of the most important properties exhibited by such interacting processes is that their history, which is an intrinsic feature of the interaction. What happens at any instant in time is a consequence of the accumulated effect of the past history of the interacting processes. Therefore conventional differential equation models relating the rate of change of state variables to their current values are inappropriate. Suitable models are delay equations or integral equations where the integration is taken over the past history of the processes.

Investigating of the relationship between two interacting processes and fixing the other processes were common prototypes in experimental work. By performing this procedure for various pairs of interacting processes and various choices for the values of the fixed processes, the general structure of these interacting processes may be investigated. However, since the pairwise results of this strategy are based on fixing the simultaneous influences of the other processes, it does not reflect the actual behaviour of the interacting processes in many real world structures.

Recently, the multivariate model of a system has been developed to investigate the communication between various pairs of processes among large networks taking into account the influences of the others processes. Many researchers have used multivariate models of the system to perform the time-frequency analysis of multi-channel data sets in the different disciplines such as Economics, Biology, Physics, Neuroscience and Applied

1

Mathematics.

In Neuroscience, the simultaneous interfacing of large numbers of neurons generates electrical signals. In turn, these signals reflect the electrical activity of the brain. Furthermore, understanding of the concept of the functional connectivity of brain regions facilitates the recognition of brain behaviour in normal and abnormal cases.

One of the most widespread brain disorders is epilepsy. Epilepsy is a chronic disease characterised mainly by recurrent and unexpected interruptions of normal brain electrical activities, called epileptic seizures (Robert *et al.* [2005]). Also epilepsy is distinguished by a neurobiological, cognitive psychological and social consequences of this condition.

Electroencephalogram (EEG) is generally used to detect the brain electrical activity and to clinically diagnose epileptic seizures. EEG is characterised by its high temporal resolution, which measures the electrical signals over a short time in milliseconds. EEG can measure the electrical signals of the brain directly from the scalp by multiple electrodes. Each EEG signal represents a difference in voltage between two electrodes, and the presentation of the EEG signals is called a *montage*. Practically, there are different types of EEG montages, such as the **dipole montage**, where the voltage is taken between two contiguous electrodes, and the **referential montage**, where the voltage is measured with respect to a specific electrode appointed to be a reference for all recording electrodes. The common locations for the reference electrode are the midline, where the signals at these places are not amplified in one hemisphere against the other and the linked ear locations, where the signals usually encounter the same conditions in both hemispheres (Miller *et al.* [1991]). The **average reference montage**, defined as an average of all signals, then the averaged signal refer as a common reference for each channel. Finally, the **Laplacian montage** represents the voltage between a recording electrode and a weighted average of the surrounding electrodes (Nunez and Pilgreen [1991]).

There are various clinical and research applications for the EEG. Clinically, for example, the EEG is used to distinguish epilepsy from other types of brain disorders, to determine the location of an epileptic seizure onset, to specify the epilepsy syndromes and characterise the strength of the seizure in order to accustom anti-epileptic medications (Smith [2005], Aurlien *et al.* [2007], Ferris *et al.* [2006]). For research purposes, the EEG is used to investigate different sleep stages and to monitor the brain during physical tasks such as hand movements (Mima *et al.* [2000], Brovelli *et al.* [2004]). Although, all previous benefits of EEG, it apparently has the limitation of lower spatial resolution compared with

the other methods such as functional magnetic resonance imaging (fMRI). Since the EEG electrodes are mostly sensitive to the electrical activity of the brain in the cortical surface, which is beneath the scalp, while the sources of electrical activity are separated from the EEG electrodes by three layers; cerebrospinal fluid (CSF), the skull and the scalp (Srinivasan [1999]). However, the functional magnetic resonance imaging (fMRI) measures the brain electrical activities indirectly through the variations in the blood oxygen level that are coupled with these activities (Ogawa *et al.* [1990], Ogawa *et al.* [1992], Logothetis and Wandell [2004]). The most important feature of fMRI is the high spatial resolution which is as good as (1mm) (Engel *et al.* [1997]). Also it provides measurements of the brain's electrical activities from all its regions, not only the cortical surface. Thus, fMRI is a powerful tool with which to record small electrical signals from the deep brain areas. However, the disadvantages of the fMRI are its low temporal resolution and indirect measurements of the brain's electrical signals, where such signals are recorded during the changes of the blood oxygen level over time.

Recently, many authors have considered the incorporation of two techniques; EEG and fMRI in order to improve the temporal-spatial resolutions. That is, the brain is scanned simultaneously with EEG recording. For instance, Goldman *et al.* [2002] and Laufs *et al.* [2003] have investigated the behaviour of the alpha rhythm [8, 12] Hz in normal subjects during eyes-closed rest. They mapped the brain regions where the α-wave power changes as the MRI signals change.

Electrical activities of the brain produced by different techniques, are performed as sequences of observations over time called *time series*. The analysis of these time series has been carried out in many literatures with respect to time and frequency domains. For the spectral analysis, the times series are first transformed from their original time domain into frequency domain using the Fourier transform. Fourier transform is used to map functions in state space into functions in a frequency space. The fast Fourier algorithms are applied to compute the Fourier transform coefficients. The application of the fast Fourier algorithms conserves the time and effort required to carry out the massive computations (see, Cooley and Tuckey [1965], Cooly *et al.* [1967], Bergland ([1967], [1968]), Gentleman and Sande [1966]).

The Power spectrum is a fundamental part of the spectral analysis and represents the energy contained in the signal and how this energy is distributed with respect to frequency. Several methods have been developed to estimate the power spectra and cross-spectra

of multi-channel data. These methods can be classified into two types: a parametric approach performed in the form of a "multivariate autoregressive model (MVAR)", and a non-parametric approach, represented by "averaging periodograms across the adjacent sections of the record", a "smoothed periodogram of the entire record" and a "multi-taper method". The main concern of these methods is to estimate the spectra and cross-spectra of the data of interest taking into account the natural dependence between these data.

Using the multivariate autoregressive model means imposing the data to fit this model. Then the MVAR parameters are estimated to extract the spectral properties of these data. To estimate the multivariate autoregressive parameters we will use the method of maximum likelihood, which provides identical parameter estimates to those obtained by the Yule-Walker approach. However, owing to the ability of the maximum likelihood approach to determine the standard error of the parameter estimates, it is preferred. The standard error is used as a statistical test of the null hypothesis that, the coefficients of the multivariate autoregressive model of order p are significantly zero. Moreover, the accuracy of the multivariate autoregressive model to fit the data depends on the determination of the model order p which represents the number of the past observations used to predict the current observations. Anderson *et al.* [1998] applied the multivariate autoregressive model for EEG data to derive properties from the human EEG records when the subject was performing physical and mental tasks. The multivariate autoregressive model was proposed for the fMRI multi-subjects by Harrison *et al.* [2003] in order to investigate the connectivity between the fMRI signals. Many authors have applied the multivariate autoregressive model to obtain the spectral properties of the EEG signals (Bressler *et al.* [1999], Moller *et al.* [2001], Ding *et al.* [2000]). See also (Percival and Walden [1993], Cryer and Chan [2008]).

Since the periodogram is a natural estimate of the power spectrum, the first two non-parametric spectral estimates are developed to smooth the periodogram. The method of averaging peridograms across contiguous segments of the single record, also known as a disjoint sections method, was firstly proposed by Bartlett [1948]. In this method, the data is segmented into a number of sections, namely L, each of which contains the same number of data, say N, such that the whole length of the data T equals the product of the number of segments L and the sample size N. The periodogram for each segment is computed at every frequency in the spectral domain and hence, the estimate of the spectrum at a specific frequency is obtained by averaging the L periodograms together at

that frequency. Welch [1967] modified this method to involve overlapped segments in an attempt to reduce the bias error that stemmed from finite data (See also, Brillinger [1981], Amjad *et al.* [1997] and Barbé *et al.* [2010]). This spectral estimation algorithm is suitable when large numbers of data are available. However, increasing the size of the sub-samples reduces the bias error, improves the frequency resolution and attenuates the spectral side lobe leakage, but at the same time the risk of variability error is increased.

The smoothed periodogram procedure, which is also called a frequency average, is based on the fact that the variation of the power spectra in the relatively small band of frequency is small, so that the estimated power spectrum for each signal is achieved by taking the average of the periodograms over a small band of frequency around a specific frequency. This means every periodogram at a frequency lying within this small band represents an estimate of the power spectrum at the frequency at the centre of the band. In this method the variance error of the estimate is reduced while the number of the frequency ordinates is increased. However, the increment of the frequency ordinates has to be bounded by the sample size, that is the number of frequency ordinates must be less than the length of data (see, Brillinger [1981], Bloomfield [2000], Cryer and Chan [2008]).

The multi-taper method invented by Thomson [1982], plays an important role in estimating the power spectra by reducing the estimation variance and attenuating the spectral leakage due to the finite length of the data. These tapers are known as Slepian functions or Discrete Prolate Slepian Sequences (DPSS) (Slepian [1978]). In this method, the data is weighted by a number of orthogonal tapers *prior* to being transformed into the spectral domain by using the finite Fourier transform. Thereafter, the estimated power spectrum for each taper is computed and the multi-taper estimate of the power spectrum is captured by averaging the estimated power spectrum for all applied tapers (see, Percival and Walden [1993]).

The interrelation between a pair of processes within a dynamic neural network of stationary time series is captured by applying the ordinary coherence. Coherence is a real-valued function of frequency based on the coefficients of spectra and cross-spectra. Coherence measures the strength of correlation between a pair of processes and can be conditioned to take into account the influences of the other processes. The value of coherence ranges between one, which indicates perfect correlation, and zero, which indicates incoherent relation between these processes (see, Brillinger [1981], Bendat and Piersol [1986], Halliday *et al.* [1995]).

Furthermore, the determination of the information flow between stationary processes during the interactions is also one of the important concepts in time series analysis. Wiener [1956] was the first to realise the existence of causal relations between different time series and suggested that, for two simultaneously observed time series, one can be called causal to the other if the predictability of the second time series is improved by incorporating information from the first. This concept of causality was formalised by Granger [1969] in the context of the linear regression models of stochastic process theory. Granger argues that if the variance of the prediction of the second time series at the current time is reduced by involving past records from the first time series in the linear regression model, then the first time series can be called G-causal for the second time series. Geweke [1982] constructed the measures of linear dependence and feedback between multivariate autoregressive processes in frequency space, thereby making the interpretation of Granger causality more straightforward.

The concept of Granger causality has attracted researchers from various disciplines, and in particular the ideas have been used widely in Economics. Thereafter, there has been a growing interest in the use of G-causality to identify causal interactions within networks of stationary time series in various fields. In Engineering, Caines and Chan [1975] investigated the feedback between input and output variables. Their measures of causality were consistently presented in the spectral domain and have been subsequently adopted by researchers in other fields.

Causal measures have been applied in Neurology since the beginning of the 1980s in order to understand the mechanism of neuron connectivity and identify the direction of interaction between multivariate time series. Saito and Harashima [1981] and Kamitake *et al.* [1984] presented the method of directed coherence (DC) to demonstrate the relation between a pair of data channels described by a bivariate autoregressive model. Thereafter various methods have been developed to specify the direction of interaction between pairs of processes within the dynamic system in the frequency domain. Kaminski and Blinowska [1991], Baccalá and Sameshima [1998] and Baccalá *et al.* [1998] have developed a method, after referred to as the *directed transfer function* (DTF), for calculating directed coherence in multichannel systems modulated by the multivariate autoregressive model framework. Wang and Takigawa [1992] used directed coherence to describe the direction of information flow of the EEG signals between the two hemispheres of the brain in the frequency domain. They found that the directed coherences of the right-left direction

in the parietal and occipital lobes were greater than the directed coherences of the left-right direction. Bernasconi and König [1999] applied Geweke's spectral measures to detect causal influences among different areas in the cat visual cortex.

Brovelli *et al.* [2004] used Granger causality to investigate the functional relationship between neuronal activity in the precentral and postcentral areas of the brains of two monkeys as each pressed a hand lever during the wait period of a visual discrimination task. The partial directed coherence (PDC) was introduced by Baccalá and Sameshima [2001] as a frequency approach to investigate the direction of information transmission among multivariate autoregressive models. Schelter *et al.* [2005] proposed a statistical significance level for non-zero partial directed coherence at a specific frequency. Mima *et al.* [2001] applied the directed transfer function to the electroencephalogram-electromyogram (EEG-EMG), where EMG reflects the muscles activities, in order to determine the information flow between EEG and EMG signals. They found that the DTF of EEG$\rightarrow$EMG was greater than DTF of EMG$\rightarrow$EEG.

Takigawa *et al.* [1996] debated the two directional influences between the frontal and the occipital lobes in the EEGs of healthy and epileptic subjects using directed coherence. They found that the dominant direction of information flow in the healthy control subject's EEG was the occipital-frontal direction. Also, they found that the high values of directed coherence occurred in the α band. The values of the directed coherences of the EEG belonged to the epileptic patient were significantly small when compared with those of the healthy control, and spread over δ-, θ- and β-bands, where δ, θ, α, β, and γ are the waves that EEG is classified into, and they will be discussed in Chapter 2. Sato *et al.* [2009] introduced the application of the partial directed coherence for the fMRI data to identify the causal relations between the fMRI signals and determine the information flow direction.

The purpose of this work is to compare the spectral analysis methods of short duration time series of EEG signals belonging to a right-handed focal epileptic patient. The electrical activities of the brain due to the interaction between its various regions, including frontal, central, parietal, occipital and temporal, are described by EEG signals recorded from the scalp. This study investigates the correlations between the neural processes collected from the left hemisphere. Since each region of the brain is responsible for a special function, we will choose one process to act as a representative process for that area. We will estimate the spectra and the cross-spectra of the chosen processes using a multivariate

autoregressive model as a parametric approach, where the data are forced to fit this model, using a disjoint sections method, a frequency averaging estimate and finally a multi-taper method. Once the coherence coefficients are estimated, the associations between each pair of chosen processes will be estimated. The partial coherence, which measures a linear relation between two stationary processes after removing the linear influence or influences of one or more processes, is employed to determine the correlation depth. The structure of these processes will be built up by applying the partial directed coherence, which is calculated from the estimated parameters of the MVAR model.

In this work, we will also carry out a statistical test of an independence property of coherences of finite uncorrelated samples of two independent stationary stochastic processes, as well as the independence relation between coherences and frequencies using the maximum likelihood principle.

This work is arranged as follows:

- **Chapter two**: This chapter gives a description of the electroencephalogram and describes the procedure of collecting these data from the scalp. In addition, we explain the important statistical and mathematical properties of the data. Since the data we will use in this research belongs to an epileptic patient, we give a general view of this brain disorder (epilepsy) and describe the different types of epilepsy and their symptoms.

- **Chapter three**: This chapter involves the definition and properties of the Fourier transform of signals and stochastic processes. In addition, we provide three non-parametric procedures to estimate the spectral density: an averaging periodograms across adjacent segments of single records method; a smoothed periodogram of the entire single record method; and, a multi-taper method. The statistical properties of these methods are demonstrated.

- **Chapter four** : In this chapter we provide the theoretical definitions of coherence and partial coherence, and explain the conditional spectra and cross-spectra. The procedure of estimating ordinary and partial coherences directly from data is described. The multivariate autoregressive model and the methods of estimating its parameters such as the Yule-Walker and maximum likelihood methods are also described. The partial directed coherence for multivariate system and its statistical significance level are defined here.

- **Chapter five**: In this chapter we provide a statistical test of the significant independence feature of coherences, which are detected between finite samples of uncorrelated stationary stochastic processes, and the independence property of coherences and frequencies. The maximum likelihood principle is used to examine how likely it is that the observed data fit likelihood function under the null hypothesis that the coherences of finite samples of independent stationary processes are significantly independent and coherences and frequencies are significantly independent.

- **Chapter six**: In this chapter we include the spectral analysis and results based on the spectral estimators mentioned in Chapter 3 and draw comparisons between these different methods.

- **Chapter seven**: This chapter concludes the work and provides ideas for future research.

Results included in this work have been produced by C programming and graphs have been plotted using MATLAB software.

Chapter 2

EEG Data Description

2.1 Introduction

The EEG data plays a crucial role in detecting and diagnosing brain abnormality and it has been employed intensively in various applications either for clinical or research purposes. This work investigates the connectivity among processes of a neural network and will use data consisting of brain signals belong to an epileptic patient as input for mathematical and statistical analyses. Since the neurons are the generators of the EEG signals, we will give a brief description of the neuron structure and the mechanism of generating EEG signals. We will give a description of the technical method used to collect EEG data from the scalp and present a general view of the brain regions and their basic functions. Also, we will define the EEG signal and explain its properties, such as amplitude and frequency bands. The epileptic activity is described and the most common types of epilepsy will be explained along with their symptoms. The statistical properties of the signal, such as its amplitude, mean and variance, will also be explained. Finally, the assumption of the stationarity and the mixing condition will be demonstrated.

2.2 Neuron Structures

Central nerves system (CNS) essentially consists of nerve cells or neurones that are imbedded in glia cells. Neurones play an important role in transmitting electrical signals and information to the CNS, while glia cells provide support, protection and supply nutrients to neurones. (See, Niedermeyer and Lopes Da Silva [2005]).

Basically, the neuron consists of a body (soma) in which a nucleolus is located, axon

and dendrites, as shown in figure (2.1). Axon is a long fibre extends from the body, and covered by fat membrane, called myelin sheath. Dendrites are short structures that branch off from the cell body forming small neurones. Axons and dendrites are considered as two different types of processes, which provide contact with and pass signals to different targets, either organs or other neurones. The nerve cells are covered with thousand of small structures, known as synapses, which are other mean of contact that pass signals.

Functionally, dendrites serve as receptors that receive and carry signals to the body of the neuron, while axon serves as a transmitter that carry signals away from the body toward other targets. Synapses are essential to pass signals to targets. At a synapse, the plasma membrane of the target is called postsynaptic terminal, while the plasma membrane of the neuron is called presynaptic terminal. These two synaptic terminals are connected by ionic channels.
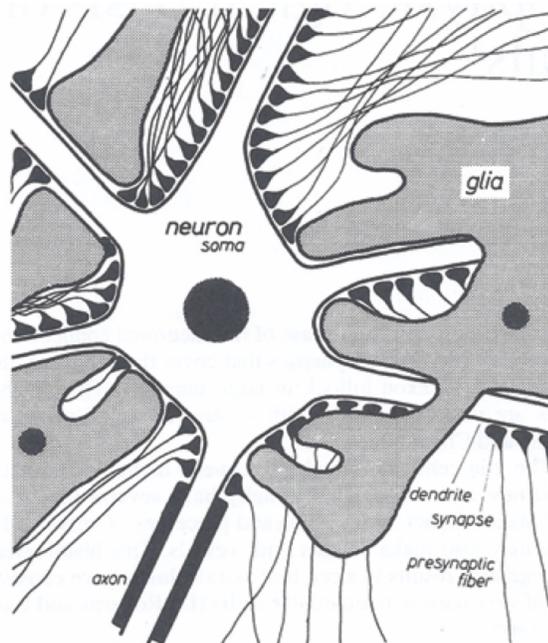
Figure 2.1: Neuron structure (Niedermeyer and Lopes Da Silva [2005]).

2.3   Generation and Propagation of Action Potentials

Neuron, likes all cells in our bodies, maintains a voltage across its membrane. The potential inside the neuron cell is about -70 μV, which means that the intracellular space is negatively charged relative to the extracellular space. This potential is subject to vary when the neuron is stimulated. Fluctuation of the neuron membrane potential gives rise to flow of

electrical current by synaptic activities.

At rest, the neuron cell membrane is selective to certain ions in order to maintain its potential. As soon as the cell body is stimulated, ion channels in synapses junctions begin to open, causing a flow of positive ions into the cell. Once the potential inside the cell reaches its threshold, around $+40$ μV, an action potential travels along the axon.

An excitatory postsynaptic potential (EPSP) occurs in the target neuron, if the action potential ends in an excitatory synapse. In this case, the current flows to the following neurones, hence propagates to the target. An inhibitory postsynaptic potential (IPSP) occurs, if the action potential ends in an inhibitory synapse. Therefore, hyperpolarization , a decrease of the intracellular potential relative to the extracellular potential, takes place due to an outflow of positive ions from the neuron cell.

2.4   Electroencephalogram (EEG) signal

The electroencephalogram (EEG) is a record of fluctuations of electrical activity produced by the firing of neurons within the brain. The brain itself consists of billions of neurons making up a large complex neural network. This record can be taken from the head using two different procedures either scalp or intracranial. In the former the electrical activity of the brain is recorded on the surface of the scalp by using small metal discs, called electrodes, with a conductive gel or paste to ensure good mechanical and electrical contact. The intracranial method of recording the EEG signals depends on special electrodes being implanted in the brain during surgery to record special kinds of EEG signal. In order to provide accurate detection of the potential electrical activity of the brain, the electrodes must be of impedance less than 5 kΩ (Adeli *et al.* [2003]). EEG has various applications either for research purposes, such as distinguishing the different sleep stages, or for medical purpose to diagnose brain disfunction such as epilepsy.

2.5   Description of Scalp EEG recording

In the scalp EEG recording, the brain's electrical activities are recorded from the scalp by placing electrodes, each of which is attached to an individual wire connected to an EEG amplifier that measures the voltage differences between two points on the scalp. Consequently, each channel is connected to two electrodes one of which is considered to be identical to all channels and is called the system reference electrode. An amplifier am-

plifies the voltage between the active electrode and the reference electrode typically by a factor from 10^3 to 10^5. The analog signal carried by each electrode then passes through three filters before being digitized for computer storage and processed via an analog-digital converter. These filters are a high-pass filter, a low-pass filter and a notch filter. The high-pass filter is set to (0.5 -1 Hz) in order to filter out all the low-frequency noise, whereas the low-pass filter allows passing for high frequency ranging between (35-70 Hz) and cuts off all the higher frequency. The notch filter is used to remove the noise caused by the electrical power lines at the main frequency of 50 Hz or 60 Hz.

In order to understand the mechanism of the EEG recordings and how the electrode locations and names have been chosen, it will be useful to have a close look at the brain. Anatomically, the brain consists of two similar hemispheres. Each hemisphere is responsible for sensory and motor processes on the contralateral side of the body. In other words, the sensory information coming from the left side of the body crosses over to the right side of the nervous system (and of course to the brain) and vice versa. Similarly, the motor cortex in one hemisphere controls the movements of the opposite side of the body. Each hemisphere of the brain is divided functionally into four significant lobes: frontal lobe, parietal lobe, temporal lobe and occipital lobe (Kandel *et al.* [1995]).

2.6   Anatomical Brain regions

1- The frontal lobe is largely concerned with planning and movement. This lobe involves the motor cortex, which lies in the back of the frontal lobe and controls the movement of the limbs and trunk. The Broca's area (or language area), which is located in the left frontal lobe, controls the muscles of the mouth involved in the production of speech.

2- The parietal lobe is the segment of the brain that lies behind the frontal lobe and contains the somatic sensory cortex, which is located in front of the parietal lobe just behind the motor cortex in the frontal lobe. This lobe is responsible for feeling touch sensation from the body. The top of the sensory cortex is concerned with the bottom of the body, while the bottom of the cortex is responsible for the top of the body.

3- The occipital lobe is the smallest portion of the brain and is located in the back of

the head just behind the parietal and temporal lobes. This lobe involves the vision
area, which helps in the visual recognition of shapes and colours. Damage to this
area of the brain can cause blindness.

4- The temporal lobe is responsible for audition, learning, memory and emotion. Also,
the left temporal lobe contains the Wernicke's area, which is concerned with inter-
preting the written and spoken language.

Moreover, there are three association areas concerned with integrating information for sig-
nificant action. They are all involved, to varying degrees, in the control of the three major
functions of the brain, namely perception, movement and motivation. The **prefrontal as-
sociation cortex** occupies most of the rostral part of the frontal lobe and is responsible for
planning voluntary movement. The **parietal-temporal-occipital cortex** is located in
the interface between the three lobes and controls the higher perceptual functions related
to somatic sensation, such as hearing and vision. Information from these different lobes
combine in the association cortex to form complex perceptions. The **limbic association
cortex** is found in the medial and inferior surfaces of the cerebral hemispheres, in por-
tions of the parietal, frontal and temporal lobes, and is mainly responsible for motivation,
emotion and memory. These lobes and their functions are displayed in figure (2.2)
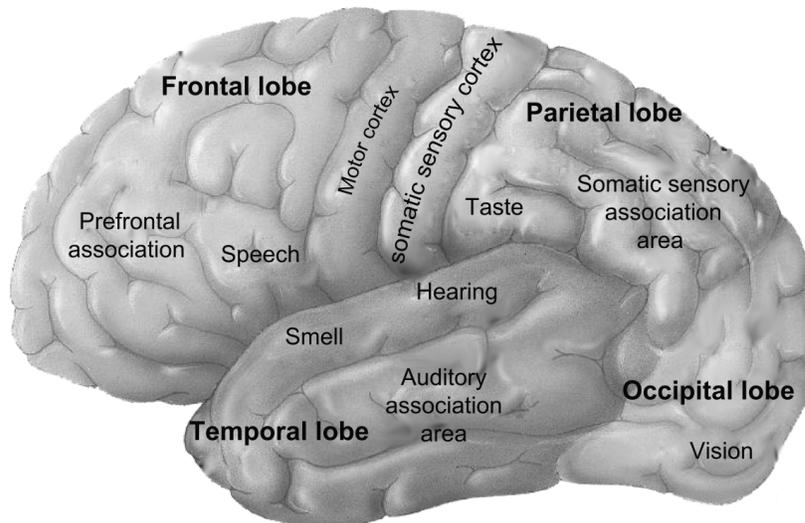
Figure 2.2: Brain lobes and their functions

2.7   Electrode position and nomenclature

Electrodes names and positions are specified by the International 10-20 system proposed by Jasper [1958], which depends on a proportional measurement strategy to be applicable for differences in head size and shape. This system consists of 19 recording electrodes as well as the ground and system reference.

The location of electrodes is based on the landmarks of the skull, namely nasian(Nz), which is a dent at the upper root of the nose; inion(Iz), which is an external occipital protuberance; the left pre-auricular (LPA); and, the right pre-auricular (RPA). The main coronal contour between LPA and RPA intersects with the main sagittal contour between Nz and Iz at 50% of their lengths; this intersection point indicates the vertex of the skull. The electrodes positioned along the main central coronal contour LPA-RPA are T3, C3, Cz, C4, T4, located at 10% above the LPA, 20%, 20%, 20%, 20% and lastly 10% above the RPA, respectively. The main sagittal contour Nz-Iz holds the electrode locations Pz, Cz and Fz. The electrodes positioned in the first coronal contour, which is placed at a 10% distance from the Nz between LPA and RLA, are FP1 and FP2. The second intermediate coronal contour is placed at a 20% distance from both the first coronal contour and from the central coronal contour, and holds the electrode positions F7, F3, Fz F4, F8. The third intermediate coronal contour holds the electrode positions T5, P3, Pz, P4, T4 and is located at a 20% distance from both the central coronal contour and from the fourth coronal contour of the electrode positions O1 and O2. This, in turn, is placed at a 20% distance from the last contour, which is located at a10% distance from the Iz (Vernon *et al.* [1993]). Each electrode site is characterised by a letter and a number. The letters indicate the corresponding brain lobe, where F stands for frontal lobe, T for temporal lobe, P for parietal lobe, O for occipital lobe, C indicates the central area of the skull as there is no central lobe but it is named for nomenclature purposes, and 'z' refers to the electrode located on the main sagittal contour (midline). The numbers indicate the hemisphere, where odd numbers applying to the left hemisphere and even numbers applying to the right hemisphere.

However, the rapid development of topographic methods to study spontaneous and evoked potentials and improvements of multi-channel EEG hardware systems, has necessitated the standardisation of a large number of channels. Hence, in 1985 the International 10-20 system was modified to involve 74 electrodes instead of 21 electrodes by Chatrian *et al.*

([1985] , [1985]). This extended 10-20 system is called the10-10 system or 10% system of electrode placements. In the extended 10-20 system the positions of the extra electrodes were designated to be placed on the coronal contours and sagittal contours lying halfway between the standard coronal and sagittal contours. This means that the coronal contour AF lies halfway between coronal contours FP and F; the coronal contour FC lies between contours F and C; the coronal contour CP between contours C and P; and the coronal contour PO lies between contours P and O. The sagittal contours hold numbers 1 and 5 lie halfway between z and 3, and between 3 and 7 respectively and the sagittal contour holds number 7 locates halfway between the contours 5 and 9. The right hemisphere sagittal contours have the same structures as the left contours. Consequently, the main sagittal contour Nz-Iz is divided in a 10% increments to generate the electrode positions FPz, AFz, Fz, FCz, Cz, CPz POz and Oz, and the main coronal contour LPA-RPA holds the equispaced electrode locations T9, T7, C5, C3, C1, Cz, C2, C4, C6, T8, T10.

For consistency the extended 10-20 system replaces the electrode names T3/ T4 by T7/T8 and T5/ T6 by P7/P8. All electrode positions along the same sagittal contour have the same postscript number, except FP1/ FP2 and O1/O2 and all electrode locations on the same coronal contour have the same letter(s). Electrode positions combining two letters indicate the position in the intermediate area, that is, FP means the fronto-polar position, AF means the anterior-frontal electrode position, FC indicates the fronto-central position, FT indicates the pronto-temporal position, CP is the centro-parietal location, TP is the temporal-posterior electrode position, and PO designates the parieto-occipital region. (See, American Electroencephalographic Society [1994], Nuwer [1987], Nuwer *et al.* [1998], Klem *et al.* [1999]).

However, modifications to the standard 10-20 international system have been continued by researchers in order to increase the resolution of EEG channels by increasing the number of electrode settings. Currently 128 EEG channel systems and 256 EEG channel systems are commercially available (see, Suarez *et al.* [2000]). Recently, Oostenveld and Praamstra [2001] presented their modification of the 10% system, known as the 5% system, which accommodates 345 channels and have proposed the nomenclature of the new electrode positions (see, Jurcak *et al.* [2007]). Figure (2.3) illustrates the extended 10-20 international system or the 10% system of the electrodes names and placements.
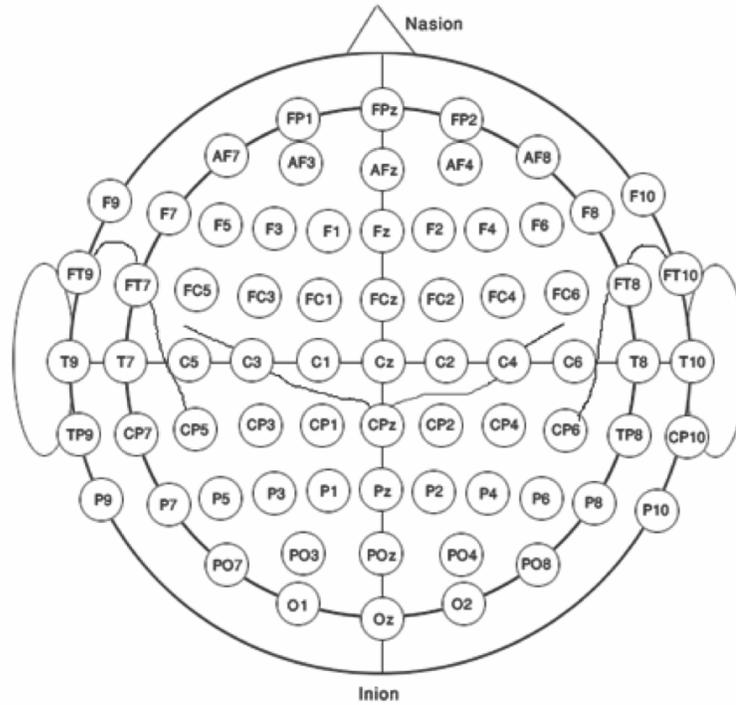
Figure 2.3: Modified 10-20 international system of EEG electrodes names and places

The approximate placements of the electrodes beneath the skull according to the modified 10-20 international system are displayed in figure (2.4).
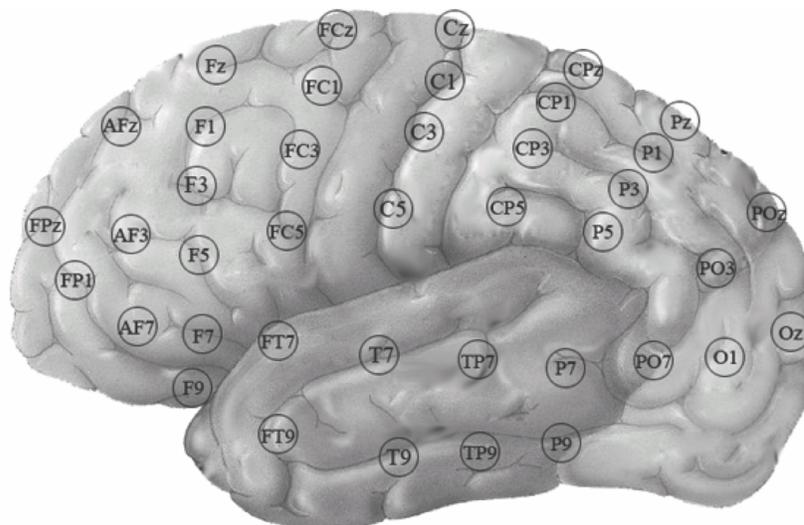
Figure 2.4: Electrodes positions in the left hemisphere of the brain

2.8   Frequency bands and amplitude of the EEG signal

2.8.1   Frequency bands

The EEG has a wide range of frequency components which may relate to different physiological processes. However, the range of clinical and physiological interests lies between 0.5Hz and 30Hz. This range is approximately classified into a number of frequency bands as follows (see, Pfurtscheller and Lopes da Silva [1999], Adeli *et al.* [2003], Hema *et al.* [2009]).

Delta: The δ- rhythm is known as the slowest wave ranging between 0.5Hz and 4Hz and possesses the highest amplitude compared with the other EEG's waves. It appears normally in adults and babies during deep sleep stages and indicates the slow brain activities.

Theta: θ- rhythms lie within the range (4-8) Hz and are found normally in children during sleep or in adults during drowsiness. Theta waves can also appear in small amounts in the normal waking adult. The existence of high theta activity in awake adults represents abnormal activity.

Alpha: α- rhythms lie within the range of frequencies (8-13) Hz and have been thought to indicate both a relaxed and mentally inactive awakeness. They appear strongly in the occipital area. These waves can be eliminated by opening eyes, hearing unfamiliar sounds, or by anxiety or mental concentration.

Beta: β-rhythms range between 13-30 Hz and appear mostly in the frontal area. A β-wave is usually associated with active thinking, active attention, focus on the outside world or solving concrete problems. It can reach frequencies near 50 Hz during intense mental activity.

Gamma: γ- rhythms lie within the high band of frequencies greater than 30Hz. Gamma rhythms are characterized by quite a wide range extending between 25Hz to 100Hz. The first detection of gamma activity was in 1956 by Sem-Jacobsen and Kuchera [1956] who pointed out that the gamma rhythms emerge in the depth of the cortex as a response to a light flash, especially when the patient looked at a light. According to Takano and Ogawa [1998], the gamma activity (35-45) Hz is correlated to informational processing and cognitive functions, such as attention. These authors demonstrated that the gamma band increases in the childhood between 3 and 4 years of age and then peaks at around

4 to 5 years of age, especially in the frontal areas . Tsang *et al.* [2004] showed that there is a significant increase in power within the gamma band (35-45) Hz, over the frontal and occipital areas in response to pulsed magnetic fields. The gamma wave around 40Hz occurs as a response to auditory stimuli (Jokeit and Makig [1994]), Funk and Epstein [2004] suggest that this may reflect the binding of various information and coordinating of sensory and motor cortices activities. Niedermeyer [2003] reported that the gamma activity within the range (40-80) Hz is apparently observed during arousal and may well be related to consciousness.

2.8.2   Amplitude

The EEG amplitudes typically vary between 20 to 100 μV and are attenuated by the skull and scalp. In general the EEG amplitude is inversely dependent on frequency, so that the amplitude of the wave increases as the frequency decreases (see, Pfurtscheller and Lopes da Silva [1999]). For instance, the δ-rhythm, associated with the lowest band of frequency (0.5-4) Hz in EEG, is classified as the highest amplitude wave. Also the θ-rhythm with frequency band (8-13) Hz, has a larger amplitude than the β-rhythm, which is associated with the frequency band (13-30) Hz, and so on.

2.9   Epilepsy

Epilepsy is a common chronic neurological disorder characterized by recurrent, unprovoked seizures. These seizures are caused when a large collection of neurons discharge in abnormal synchrony. There are two essential types of epileptic seizures, which are classified according to the degree of the seizures, that is whether the seizure appears in a restricted part of the brain in one hemisphere (focal epilepsy) or arises in both hemispheres simultaneously (generalized epilepsy), (see, Kandel *et al.* [1995], Shorvon *et al.* [2009]).

2.9.1   Partial or focal epilepsy

Partial epilepsy has been divided into three groups, based on whether the consciousness is impaired or if the seizure develops into generalized epilepsy as the abnormal activity spreads to the other hemisphere. These three groups are called simple partial, complex partial and partial secondary-generalized. The focal seizure is presented as a sharp spike of electrical activity, called an EEG spike.

Simple partial

A patient experiencing a simple partial seizure is still fully conscious. The seizure symptoms depend on the region of the brain where the epileptic seizure takes place. For example, if the partial seizure originates in the visual or auditory cortex, it will result in visual and auditory hallucinations respectively. However, there are some commonly observed sensory and physical symptoms that the patient may demonstrate during the seizure, such as a feeling of fear or extreme terror, where the patient may run to escape or try to find assistance; a feeling of extreme pleasure or fun; an enhanced sense of taste or smell; a numbness and tingling in a part of the body, involuntary twitching of the striated muscles of the body or jerky movement. If the abnormal activity attacks the speech center it will lead to vocalizing or repeating items of speech. This phenomenon of ictal repetition of the syllables or phrases is called "epileptic palilalai"(see, Shorvon *et al.* [2009]).

Complex partial seizures

Complex partial seizures are characterized by impairment or full loss of consciousness, as the epileptic activity affects a larger part of the brain and often lasts longer than other types of epileptic seizures. The common symptoms associated with this type of epilepsy include, a chewing or swallowing movement; a feeling of fear; a wandering around in a confused way; walking often in circles (cursive seizures), categorized as "ambulatory"; picking at or fumbling with garments (gestural automatisms); and, vocalizing or repeating items of speech.

Secondary-generalized seizures

These kind of seizures starts either a simple or complex partial seizure, then develops into a generalized seizure when the abnormal activity extends into the other hemisphere and consciousness is lost. The most common types of focal epilepsy are *Temporal lobe epilepsy*, where the seizures arise in one or both temporal lobes of the brain, and, *Frontal lobe epilepsy*, when epileptic activities occur in one or both frontal lobes of the brain, often while the patient is sleeping. The seizures of these two widespread types of epilepsy can be either simple or complex partial.

2.9.2   Generalized epilepsy

Generalized seizures happen when the epileptic discharges affect both hemispheres simultaneously and consciousness is lost. The generalized seizures involve six major types of seizure: absence seizures or "petit mal"; myoclonic seizures; clonic seizures; tonic seizures; tonic-clonic seizures or "grand mal"; and atonic seizures (see, Shorvon *et al.* [2009]). Generalized seizures present as EEG traces over all electrodes.

Absence seizures

Absence seizures, also known as "petit mal", begin in childhood or adolescence, but may persist into adulthood. They are brief episodes of lost consciousness and often last less than 20 seconds. Absence seizures are usually accompanied with facial automatisms, specifically repetitive blinking, but they do not involve involuntary jerking movements.

Myoclonic seizures

Myoclonic seizures are abrupt, brief arrhythmic, involuntary jerking movements that may last less than a second, but often cluster within a few minutes. If the myoclonic seizures evolve into rhythmic jerking movements, they are classified as clonic seizures.

Clonic seizures

Clonic seizures are characterized by rhythmic repetitive jerking movements. Each single clonic movement consists of a rapid contraction followed by a slower relaxation.

Tonic seizures

This type of seizure, which is characterized by a loss of consciousness and rigid, violent muscular spasms with posturing axial and limb muscular, typically last less than 30 seconds.

Tonic-Clonic seizures

Tonic-Clonic seizures were previously referred to as "grand mal"seizures and consist of two stages; a *Tonic* phase followed by a *Clonic* phase. Conventionally, these seizures are characterized by a loss of consciousness and, either a sudden fall or dramatic violent, involuntary jerking and muscular spasm of the limbs and the body.

Atonic seizures

Atonic seizures are characterized by a reduction or loss of postural tone, which causes a so-called drop attacks, whereby the patient falls or slumps to the ground. .

2.10   Statistical properties

In this section we will demonstrate some of the most important statistical terminologies that are needed to analyse EEG signals.

2.10.1   Time series

A *time series* is a series of observations made over a period of time. The observations are typically enumerated in the format $X_{t_1}, X_{t_2}, \cdots$, where X may be a scalar-valued function of time (**univariate** process), or a vector-valued function of time (**multivariate** process). Our interest here involves the latter in which X is taken to be a column vector of dimension T. In practice, observations of a time series are taken at uniformly spaced intervals, say Δt, and X_t conventionally denotes the observation taken at time t, X_{t-1} is the observation taken at time $t - \Delta t$ and so on with X_{t-k} denoting the observation taken at time $t - k\Delta t$. Time series models are classified in terms of the number of lags contained in their specification. A model containing $X_t, X_{t-1}, \cdots, X_{t-p}$ is said to contain p lags (see, Brillinger [1981], Rooss [1983]).

2.10.2   Stochastic Processes

The first step in the analysis of time series is the selection of a suitable mathematical model for the data. The time series $\{x_k(t), t \in T_o, k = 1, 2, \cdots, T\}$ is the realization of the family of random variables $\{X_t, t \in T_o\}$. These considerations suggest modelling the data as a realization (or part of a realization) of a stochastic process $\{X_t, t \in T\}$, where T is a set of time points, very often $0, \pm 1, \pm 2, \cdots$, and $T_o \subseteq T$. In the following section we shall define precisely the stochastic process and its realization (Brockwell and Davis [1991]).

Definition   (Stochastic Process) A stochastic process $X = \{X_t | t \in T\}$ is a family of random variables. That is, for each t in the index set T, X_t is a random variable, hence it has an associated cumulative distribution function (cdf) given by

$$F_t(a) = Prob(X_t \leq a)$$

Definition (Realization of a Stochastic Process).

The functions $\{X(t), t \in T_o\}$ on T are known as a **realization** or **sample-path** of the process or the time series X.

An EEG signal considered as a time series (stochastic processes) is characterised by the statistical properties that facilitate its analysis and physical interpreting. The most common statistical properties that are employed intensively in the time series analysis are: amplitude, sample mean, variance, autocorrelation and cross-correlation.

2.10.3 Amplitude

Let X, be a time series and let A denote the amplitude of X, then A can be defined as an absolute value of the X and be written as

$$A = |X|. \tag{2.1}$$

2.10.4 Sample mean

The sample mean at a specific time, say t, is usually denoted by $\mu(t)$ and represents the central tendency of the data in the stochastic process at that specific time t. It can be estimated simply by taking the instantaneous value of each sample, summing the values and then dividing by the number of samples(see, Brillinger [1981], Bendat and Pirsol [1986], Percival and Walden [1993]). Suppose we have a time series X, with samples $x_1(t), x_2(t), \cdots, x_T(t)$, then the mean value of the time series X at the time t is obtained from the formula

$$\mu_X(t) = \frac{1}{T} \sum_{k=1}^{T} x_k(t). \tag{2.2}$$

This definition considers the mean value as a function of time; hence the mean values at two different times may be different, *i.e.*

$$\mu_X(t_1) \neq \mu_X(t_2) \quad \text{if} \quad t_1 \neq t_2. \tag{2.3}$$

2.10.5 Variance estimate

The variance of the stochastic process X at time t, denoted by $\sigma_X^2(t)$, measures the dispersion of data around the mean value of the sample and can easily be estimated by computing the average of the squares of the instantaneous data values after subtracting the mean value estimate (see, Brillinger [1981], Bendat and Pirsol [1986], Percival and

Walden [1993]). Consequently, the variance can be written as

$$\sigma_X^2(t) = \frac{1}{T-1} \sum_{k=1}^{T} \left[x_k(t) - \mu_X(t) \right]^2 , \tag{2.4}$$

or equivalently, the variance can be expressed as

$$\sigma_X^2(t) = \mathbb{E}\left[(X_t - \mu_X(t))^2 \right] , \tag{2.5}$$

where $\mathbb{E}$ denotes the expectation operator. Thus the variance, in this case, depends on time, so that when the time changes the variance changes as well. Consequently, for $t_1 \neq t_2$, the variances $\sigma_X^2(t_1)$ and $\sigma_X^2(t_2)$ need not be equal, *i.e.*

$$\sigma_X^2(t_1) \neq \sigma_X^2(t_2) \quad \text{if} \quad t_1 \neq t_2 . \tag{2.6}$$

2.10.6   Autocorrelation function

The autocorrelation function (ACF) is a linear relationship which measures the interdependency between samples of time series at different times. The autocorrelation of the stochastic process X at the two times t_1 and t_2 is denoted by $r_{XX}(t_1, t_2)$ and defined by

$$r_{XX}(t_1, t_2) = \frac{C_{XX}(t_1, t_2)}{\left[\sigma_X^2(t_1) \sigma_X^2(t_2) \right]^{1/2}} , \tag{2.7}$$

where $C_{XX}(.)$ denotes the auto-covariance function of the stochastic process X at two different times t_1, t_2, which can be expressed as

$$C_{XX}(t_1, t_2) = \mathbb{E}\left[\left(X_{t_1} - \mu_X(t_1) \right) \left(X_{t_2} - \mu_X(t_2) \right) \right] , \tag{2.8}$$

where $\mathbb{E}$ denotes the expectation operator, and $\sigma_X^2(.)$ represents the variance of the process X at the required time. The values of the correlation coefficient $r_{XX}(.)$ range between -1 and +1, where -1 indicates anti-correlation in which the value of X_{t_2} decreases as the value of X_{t_1} increases, while +1 denotes positive correlation when the value of X_{t_2} increases rapidly as the value of X_{t_1} increases. The zero value of the auto-correlation represent the independence between the components of the process. In practice, $r_{XX}(.)$ is called the sample estimate of the autocorrelation coefficient of the population autocorrelation, which is denoted by ρ_{XX} (see, Brillinger [1981], Bendat and Pirsol [1986], Percival and Walden [1993]).

2.10.7   Cross-correlation function

The cross-correlation function between two stochastic processes X and Y, where $t = \pm 1, \pm 2, \cdots$, is denoted by r_{XY} at the time (t_1, t_2) and can be defined in terms of the variances $\sigma_X^2(t_1)$ and $\sigma_Y^2(t_2)$ of the time series X and Y respectively, and the covariance C_{XY} between the processes X and Y at the time (t_1, t_2) (see, Brillinger [1981], Bendat and Pirsol [1986]), Percival and Walden [1993]) . Thus $r_{XY}(t_1, t_2)$ takes the form

$$r_{XY}(t_1, t_2) = \frac{C_{XY}(t_1, t_2)}{\left[\sigma_X^2(t_1)\sigma_Y^2(t_2)\right]^{1/2}} \, , \tag{2.9}$$

where the covariance between the processes X and Y can be calculated from the formula

$$C_{XY}(t_1, t_2) = \mathbb{E}\left[\left(X_{t_1} - \mu_X(t_1)\right)\left(Y_{t_2} - \mu_Y(t_2)\right)\right], \tag{2.10}$$

where the values of the correlation $r_{XY}(t_1, t_2)$ lie within the interval [-1, 1]. Therefore, the two stochastic processes X and Y are said to be uncorrelated if the values of correlation are zero, that is $r_{XY}(t_1, t_2) = 0$.

2.10.8   Stationarity

The stochastic process X is called a stationary process or (**strictly stationary or strongly stationary process**) if the cumulative function of the joint distribution of X at times, $t_1 + s, t_2 + s, \cdots, t_k + s$ contained in the index set T, does not vary with respect to s also contained in the index set for all $t_1, t_2, \cdots, t_k$ and for all $k \geq 1$ (see, Brillinger [1981], Bendat and Pirsol [1986], Percival and Walden [1993]). In other words, the stochastic process X is said to be a stationary process if the cumulative function of the joint distribution of X is independent of a shift in time

$$F_{t_1+s, t_2+s+\cdots, t_k+s}\left(X_{t_1}, X_{t_2}, \cdots, X_{t_k}\right) = F_{t_1, t_2, \cdots, t_k}\left(X_{t_1}, X_{t_2} \cdots, X_{t_k}\right)$$

In addition, all the moments of the strictly stationary process X are independent of time. However, the **weak or wide-sense stationary Process** is the second most common type of stationarity, as it only requires its first and second orders of moments (mean and covariance) to be independent of time. That is, for all $k \geq 1$, for all $t_1, t_2, \cdots, t_k$ contained in the index set, for any s such that $t_1 + s, t_2 + s, \cdots, t_k + s$ are contained in the index set, all the joint first and second order moments of $X_{t_1}, X_{t_2}, \cdots, X_{t_k}$, exist and equal the corresponding joint moments associated with $X_{t_1+s}, X_{t_2+s}, \cdots, X_{t_k+s}$. We note that the strictly stationary process is a weakly stationary process. To simplify, we

will use the expression "stationary process" to indicate the wide-sense stationary process. Thus, the statistical quantities that are defined in the foregoing sections will be redefined for the stationary process. Consequently, the mean and the variance that are given in equations (2.2) and (2.5) respectively, will take the new formulae that are associated with the stationary process

$$\mathbb{E}\left[X_t\right] = \mu_X(t) = \mu_X(t+s) = \mu_X \,,$$
$$\mathbb{E}\left[(X_t - \mu_X)^2\right] = \sigma_X^2(t) = \sigma_X^2(t+s) = \sigma_X^2 \,. \tag{2.11}$$

The auto-covariance function $C_{XX}(.)$ of the stationary process X for arbitrary choices $t_1 = t$ and $t_2 = t + s$ can be expressed as

$$
\begin{aligned}
C_{XX}(t,t+s) &= \mathbb{E}\left[\left(X_t - \mu_X(t)\right)\left(X_{t+s} - \mu_X(t+s)\right)\right], \\
&= \mathbb{E}\left[\left(X_t - \mu_X\right)\left(X_{t+s} - \mu_X\right)\right], \\
&= \mathbb{E}\left[X_t X_{t+s}\right] - \mu_X^2 \,.
\end{aligned}
\tag{2.12}
$$

Hence, the auto-covariance function $C_{XX}(.)$ between members of the stationary process X separated by s units, where the variable s is an integer and called a *lag*, represents the ordinary variance of the stationary process X when $s = 0$. Similarly, the auto-covariance function of the stationary process Y, denoted by $C_{YY}(.)$ and given by

$$
\begin{aligned}
C_{YY}(t,t+s) &= \mathbb{E}\left[\left(Y_t - \mu_Y(t)\right)\left(Y_{t+s} - \mu_Y(t+s)\right)\right], \\
&= \mathbb{E}\left[\left(Y_t - \mu_Y\right)\left(Y_{t+s} - \mu_Y\right)\right], \\
&= \mathbb{E}\left[Y_t Y_{t+s}\right] - \mu_Y^2 \,,
\end{aligned}
\tag{2.13}
$$

also represents the ordinary variance of the stationary process Y, when the time difference s takes the value zero. The cross-covariance function $C_{XY}(.)$ between the two stationary processes X and Y for arbitrary choices $t_1 = t$ and $t_2 = t + s$, is given by

$$
\begin{aligned}
C_{XY}(t,t+s) &= \mathbb{E}\left[\left(X_t - \mu_X(t)\right)\left(Y_{t+s} - \mu_Y(t+s)\right)\right], \\
&= \mathbb{E}\left[\left(X_t - \mu_X\right)\left(Y_{t+s} - \mu_Y\right)\right].
\end{aligned}
\tag{2.14}
$$

To simplify, the argument $t + s$ of the auto- and cross-covariance functions, defined in equations (2.12), (2.13) and (2.14) for stationary processes X and Y, can be replaced by s, since these functions depend only on the time difference s and not on the time itself. The auto-covariance functions $C_{XX}(s)$ and $C_{YY}(s)$ are even real-valued functions, satisfying the property

$$
\begin{aligned}
C_{XX}^{\tau}(s) &= C_{XX}(-s) \\
C_{YY}^{\tau}(s) &= C_{YY}(-s) \,,
\end{aligned}
\tag{2.15}
$$

where A^τ denotes the transpose of A. The cross-covariance function $C_{XY}(s)$ is neither odd nor even, but satisfies the property

$$C_{XY}^\tau(s) = C_{XY}(-s).$$ (2.16)

Property (2.16) can be proved as follows, by the definition

$$C_{XY}(-s) = \mathbb{E}\left[(X_t - \mu_X)(Y_{t-s} - \mu_Y)\right].$$ (2.17)

As mentioned above, the cross-covariance function $C_{XY}(s)$ is independent of time which implies it will not vary with any time translation. We can therefore replace t wherever it appears in equation (2.17) by $t + s$ before taking the expectation

$$
\begin{aligned}
C_{XY}(-s) &= \mathbb{E}\left[(X_{t+s} - \mu_X)(Y_{(t+s)-s} - \mu_Y)\right], \\
&= \mathbb{E}\left[(Y_t - \mu_Y)(X_{t+s} - \mu_X)\right], \\
&= C_{XY}^\tau(s).
\end{aligned}
$$ (2.18)

The relations in equation (2.15) are special cases of the relation (2.16) when $X = Y$. Finally, the auto-correlation function $r_{XX}(.)$ of the stationary stochastic process X, can be obtained by substituting the values of the variance and covariance from equations (2.11) and (2.12) respectively, into the equation (2.7), thus

$$
\begin{aligned}
r_{XX}(s) &= \frac{C_{XX}(s)}{\left[\sigma_X^2 \sigma_X^2\right]^{1/2}}, \\
&= \frac{C_{XX}(s)}{\sigma_X^2}.
\end{aligned}
$$ (2.19)

Similarly, the cross-correlation function between the two stationary processes X and Y can be obtained by substituting the values of the variances and covariance from equations (2.11, 2) and (2.14) respectively into the equation (2.9). Thus

$$
\begin{aligned}
r_{XY}(s) &= \frac{C_{XY}(s)}{\left[\sigma_X^2 \sigma_Y^2\right]^{1/2}}, \\
&= \frac{C_{XY}(s)}{\left[\sigma_X^2 \sigma_Y^2\right]^{1/2}}.
\end{aligned}
$$ (2.20)

It is obvious from equations (2.19) and (2.20) that the auto- and cross-correlations for the stationary stochastic processes X and Y, $r_{XX}(s)$, $r_{YY}(s)$ and $r_{XY}(s)$, are invariant with respect to time.

2.10.9   Confidence interval of the ACF

In order to evaluate the accuracy of the estimated autocorrelation r_{XX} it is convenient to apply Fisher's transform, $\tanh^{-1}$ to the r_{XX}, since the variance of the transformed correlation coefficient is given by the constant value (see, Bendat and Piersol [1986]). Hence, Fisher's transform takes the form

$$z = \tanh^{-1}(r_{XX}) = \frac{1}{2}\log\left[\frac{1+r_{XX}}{1-r_{XX}}\right], \tag{2.21}$$

and the random variable z has an approximately normal distribution with mean μ_z and variance σ_z^2 defined by

$$\mu_z = \frac{1}{2}\log\left[\frac{1+\rho_{XX}}{1-\rho_{XX}}\right] \tag{2.22}$$

$$\sigma_z^2 = \frac{1}{T-3}. \tag{2.23}$$

The confidence interval of the statistically significant nonzero values of the estimated correlation coefficients, which indicate the actual existence of the correlation, can be achieved by testing the hypothesis that $\rho_{XX} = 0$, where a significant correlation is indicated if the hypothesis is rejected. From equation (2.22) and (2.23), the sampling distribution of the random variable z, given $\rho_{XX} = 0$, is a normal distribution with mean $\mu_z = 0$ and variance $\sigma_z^2 = 1/(T-3)$. Consequently, the acceptance region for the hypothesis of zero correlation at level α of significance is given by

$$\left[-z_{\alpha/2} \leq \frac{\sqrt{T-3}}{2}\log\left[\frac{1+\rho_{XX}}{1-\rho_{XX}}\right] \leq z_{\alpha/2}\right], \tag{2.24}$$

where z is a standardized normal variable. The values of the correlation that lie outside this interval provide a proof of the statistical correlation at the level α of significance.

2.10.10   Mixing Condition

The mixing condition of the stationary stochastic process $X = \{X_t; t \in T\}$ can be defined as a short span of dependence between its elements (see, Brillinger [1981]). That is, if X_t and X_{t+s} are random variables of the stochastic process X, then the process X is said to satisfy the mixing condition if the relation between X_t and X_{t+s} tends to zero as the gap in time s between "past " and "future " goes to infinity or at least becomes sufficiently large. The dependence between random variables can be measured by conditions expressed in terms of covariance or correlation coefficients or may be conditions placed on an order, a time or a distance between random variables.

Rosenblatt [1956] introduced a strong mixing condition in order to obtain a central limit theorem for strictly stationary processes. Loynes [1965] obtained some limit theorems for maxima of stationary processes when the strong mixing condition is satisfied. A big inconvenience of the strong mixing condition is the difficulty in verifying it. Many researches alternatively have tended to investigate the weakly stationary mixing condition or simply weak dependence. Withers [1981], Tran [1990], Birkel [1989], and others obtained central theorems under some notions of weak dependence. Newman [1984] investigated the independence structures of associated sequences of random variables and obtained some useful limit theorems under certain conditions placed on the covariance structure of a strong mixing condition. A new weak dependence condition for time series has been proposed by Doukhan and Louhichi [1999]. They stated their definition in terms of the decay of the covariance between random variables of the weakly stationary time series, providing that the gap of time between the two variables tends to infinity.

2.10.11 Weakly dependence

Let $X = \{X_t; t \in T\}$ be a stationary time series. This means that X has constant mean μ_X and variance σ_X^2, and that the auto-covariance coefficient $C_{XX}(s)$ depends only on the time difference s. Then X is said to be weakly dependent, if $C_{XX}(s) \to 0$ as $s \to \infty$. One of the most common weakly dependent time series is that generated by autoregressive models of order one, denoted by AR(1).

2.10.12 Autoregressive process of order 1; AR(1)

Suppose that X is a weakly stationary stochastic process derived from an AR(1) model, then

$$X_t = aX_{t-1} + E_t, \tag{2.25}$$

where $t = 1, 2, \cdots$; E_t is an identical independent random variable with zero mean and variance σ^2 and it is assumed to be uncorrelated with X_t and X_{t-1}; and a represents the autoregressive model parameter and satisfies the condition $|a| < 1$. Since we assumed that X is a stationary process, then $\mathbb{E}(X_t) = \mathbb{E}(X_{t-1})$, but since $a \neq 1$ the means become equal if and only if $\mathbb{E}(X_t) = 0$. Because of independence between X_{t-1} and E_t, the variance of

the process X can be written as

$$
\begin{aligned}
\text{Var}(X_t) &= a^2 \text{Var}(X_{t-1}) + \text{Var}(E_t) \\
&= a^2 \sigma_X^2 + \sigma_E^2 .
\end{aligned}
\tag{2.26}
$$

This equation yields

$$
\sigma_X^2 = \frac{\sigma_E^2}{1 - a^2}
\tag{2.27}
$$

In general, the stationary stochastic process X_{t+s} can be written as an AR(1)

$$
\begin{aligned}
X_{t+s} &= aX_{t+s-1} + E_{t+s} , \\
&= a^2 X_{t+s-2} + aE_{t+s-1} + E_{t+s} ,
\end{aligned}
\tag{2.28}
$$

then, after s substitutions, we get

$$
X_{t+s} = a^s X_t + \sum_{j=0}^{s-1} a^j E_{t+s-j} .
\tag{2.29}
$$

The covariance coefficient between X_{t+s} and X_t can be obtained as

$$
C_{XX}(t, t+s) = \mathbb{E}\Big[\big(X_t - \mathbb{E}[X_t]\big)\big(X_{t+s} - \mathbb{E}[X_{t+s}]\big) \Big] .
\tag{2.30}
$$

Note that from the previous assumption $\mathbb{E}[X_t] = \mathbb{E}[X_{t+s}] = 0$, and that E_t is not correlated to X_t across time, this means $\mathbb{E}[X_t E_t] = 0$. Substituting the values of X_{t+s} from equation (2.29) into equation (2.30) then gives

$$
\begin{aligned}
C_{XX}(t, t+s) &= \mathbb{E}\Big[X_t \big(a^s X_t + \sum_{j=0}^{s-1} a^j E_{t+s-j}\big) \Big] , \\
&= \mathbb{E}\Big[a^s X_t X_t + X_t \sum_{j=0}^{s-1} a^j E_{t+s-j} \Big] , \\
&= a^s \sigma_X^2 .
\end{aligned}
\tag{2.31}
$$

Therefore, the correlation between X_t and X_{t+s} is given by

$$
\begin{aligned}
r_{XX}(t, t+s) &= \frac{C_{XX}(t, t+s)}{\sigma_X^2} , \\
&= \frac{a^s \sigma_X^2}{\sigma_X^2} = a^s .
\end{aligned}
\tag{2.32}
$$

Hence, for $|a| < 1$, $\lim_{s \to \infty} a^s = 0$, and X is a weakly dependent process.

2.11   Conclusion

The importance of the electroencephalogram signal is its capability to detect the abnormal activities of the brain that are associated with diseases such as epilepsy. EEG has been used

widely for research purposes, including the study of various sleep stages. In this chapter, we have presented the definition of the EEG, explained the method that is used to collect the EEG data from the scalp and described the electrode placement and nomenclature according to the 10-20 international system and 10% system. Moreover, we have given a general view of the anatomical brain lobes and their functions in order to understand the EEG mechanism. We have shown that EEG signals are characterised by the specific range of an amplitude, usually between $10\mu V$ and 100 μV, and specific bands of frequencies associated to well known waves including δ-wave in [0.5, 4) Hz; θ-wave between 4Hz and 7Hz; α-wave in [8, 13) Hz; β-wave in [13, 30) Hz; and, finally γ-wave related to higher frequencies between 30Hz and 100Hz. Each of these waves indicates specific normal and abnormal activity of the brain. As we have shown, epilepsy is the most common abnormal brain activity and has various types according to the location of the seizures at onset and the degree of the seizures, so that the types of epilepsy are widespread. Finally, we have presented the statistical properties of the signals in the time domain. The spectral properties of the signals will be discussed in Chapter 3.

Chapter 3

Mathematical considerations

3.1 Introduction

Fourier analysis is an important tool that can play a crucial role in spectral analysis. Fourier analysis is classified into two types: the continuous Fourier transform and the discrete Fourier transform. Both are applied in various disciplines such as Physics, Engineering and Physiology. This analysis has been described in detail and applied extensively in several books, such as Brillinger [1981], Bloomfield [2000], Percival and Walden [1993], Edwards [1979], Chatfield [1996]. Specifically, in this chapter we will give a brief description of the Fourier transform for signals and stationary stochastic processes with their analytical and statistical properties. Also, we will present the concept of a *periodogram* as a natural estimate of the power spectrum and *cross-periodogram* as the estimate of the cross-spectrum. We will discuss three procedures commonly used to smooth the periodogram or the cross-periodogram namely; averaging the periodogram (or cross-periodogram) ordinates in the neighbourhood of a particular frequency, averaging the periodograms (or cross-periodograms) of contiguous sections of the stretch of data and a multi-taper method. The confidence interval for the spectra will be constructed for each estimate.

3.2 Fourier transform

The Fourier transform is a procedure which takes a function represented in state space, *i.e.* typically space or time, and transforms the function into an equivalent representation in the frequency domain (Edwards [1979]). In practice, the Fourier transform arises in the form of a continuous transform and a discrete transform.

Definition

Let $g(t)$ be a continuous function (signal), and let $g(t)$ be square-integrable, that is $g(t)$ has finite energy such that

$$\mathcal{E} = \int_{-\infty}^{\infty} |g(t)|^2 \, dt < \infty \,,$$

then the continuous Fourier transform of $g(t)$ at frequency f is defined by

$$G(f) = \int_{-\infty}^{\infty} g(t)e^{-2i\pi ft} \, dt, \tag{3.1}$$

and the discrete Fourier transform of the function $g(t)$ is denoted by d_g and given by

$$d_g(f) = \sum_{t=-\infty}^{\infty} g(t)e^{-2i\pi ft}. \tag{3.2}$$

Conventionally, almost all literatures use the partial sums of the discrete Fourier transform of the function $g(t)$, instead of the infinite sum defined in equation (3.2) (see, Edwards [1979], Brillinger [1981], Percival and Walden [1993]), that is

$$d_g(f) = \sum_{t=-n}^{n} g(t)e^{-2i\pi ft}. \tag{3.3}$$

For a specific value of n, this definition of the Fourier transform also known as the **finite Fourier transform** will be used intensively in the current work.

3.3 Finite Fourier transform

Consider a real-valued function (signal) $g(t)$ with T samples of the form $g(0), g(1), \cdots, g(T-1)$. Further, assume that the sequence outside the range $0, T-1$ is extended T-periodic, hence, $g(t) = g(t+T)$ for all t. The finite Fourier transform for the sequence $g(t)$ will be denoted by d_g^T, and it will also have T samples. Therefore, the finite Fourier transform may be defined as

$$d_g^T(f_k) = \sum_{t=0}^{T-1} g(t)e^{-2i\pi f_k t}; \qquad \frac{-T}{2} < k \le \frac{T}{2}, \tag{3.4}$$

where $f_k = \dfrac{k}{T}$ are the ordinary frequencies, also known as the Fourier frequencies. That is, the ordinary frequencies f_k are observed in the region (-1/2, 1/2], and associated with the angular frequencies ω_k by $\omega_k = 2\pi f_k$.

Of course, although the functions here are described as complex sequences, real-valued sequences can be represented by setting the imaginary part to 0. In general, the transform

into the frequency domain will be a complex valued function, that is,

$$d_g^T(f_k) = a + ib \,,$$

with respective magnitude and phase.

$$|d_g^T(f_k)| = \sqrt{a^2 + b^2} \,, \quad \tan^{-1}\left(\frac{b}{a}\right).$$

The quantity $|d_g^T(f_k)|^2$ is called the power spectrum of the signal $g(t)$. From the definition (3.4), we note that

$$d_g^T(f_k) = d_g^T(f_{k+T}) \,. \tag{3.5}$$

Proof:

The proof of the periodicity property of the discrete (finite) Fourier transform can be obtained directly from the definition of the transform given in equation (3.4). By definition we have

$$d_g^T(f_k) = \sum_{t=0}^{T-1} g(t) e^{-2i\pi f_k t} \,, \tag{3.6}$$

but since $e^{-2i\pi t} = \cos(2\pi t) - i\sin(2\pi t) = 1$, then equation (3.6) can be written as

$$d_g^T(f_k) = \sum_{t=0}^{T-1} g(t) e^{-2i\pi f_k t} e^{-2i\pi t} \,.$$

Once again, since $e^{-2i\pi t} = e^{-2i\pi f_T t} = 1$, then we get

$$
\begin{aligned}
d_g^T(f_k) &= \sum_{t=0}^{T-1} g(t) e^{-2i\pi f_k t} e^{-2i\pi f_T t} \\
&= \sum_{t=0}^{T-1} g(t) e^{-2i\pi \frac{k+T}{T} t} \\
&= d_g^T(f_{k+T}) \,.
\end{aligned}
$$

The complex conjugate of the Fourier transform of the real-valued function $g(t)$ at frequency f_{-k} is equal to its Fourier transform at frequency f_k, where $-T/2 < k \leq T/2$, that is

$$\overline{d_g^T(f_{-k})} = d_g^T(f_k) \,, \tag{3.7}$$

where $\overline{A}$ denotes the complex conjugate of the variable A.

Proof:

To prove the symmetry property for the discrete Fourier transform of the real-valued

signal, we first recall the definition of the finite Fourier transform given in equation(3.4), that is

$$d_g^T(f_{-k}) = \sum_{t=0}^{T-1} g(t)e^{-2i\pi f_{-k}t}. \tag{3.8}$$

Applying the complex conjugate operator on both sides of the equation (3.8), yields

$$\overline{d_g^T(f_{-k})} = \sum_{t=0}^{T-1} \overline{g(t)e^{-2i\pi f_{-k}t}}.$$

By employing the property of the complex conjugate operator, which states that the complex conjugate of the product of complex variables is equal to the product of their complex conjugate, $i.e.$, $\overline{Z_1 \times Z_2} = \overline{Z_1} \times \overline{Z_2}$, and since the signal $g(t)$ is a real-valued function of time, then the complex conjugate of the signal $g(t)$ is $g(t)$, that is $\overline{g(t)} = g(t)$. Hence we obtain

$$\begin{aligned}
\overline{d_g^T(f_{-k})} &= \sum_{t=0}^{T-1} g(t)\,\overline{e^{-2i\pi f_{-k}t}}, \\
&= \sum_{t=0}^{T-1} g(t)\,e^{2i\pi f_{-k}t},
\end{aligned}$$

since $f_{-k} = -\frac{k}{T}$, we get

$$\begin{aligned}
\overline{d_g^T(f_{-k})} &= \sum_{t=0}^{T-1} g(t)\,e^{-2i\pi f_k t}, \\
&= d_g^T(f_k).
\end{aligned}$$

The finite Fourier transform of a real-valued function $g(t)$ at frequencies f_0 and $f_{T/2}$, $i.e.$, $d_g^T(f_0)$ and $d_g^T(f_{T/2})$, are real-valued functions. Moreover, the periodicity and symmetry properties of the finite Fourier transform, expressed in equations (3.5) and (3.8) respectively, indicate that the main domain of the finite Fourier transform of the signal $g(t)$, $d_g^T(f)$, may be chosen to be $f_0 \leq f \leq f_{T/2}$ or $f \in [0, 1/2]$.

In addition, if $g(t)$ and $h(t)$ are real-valued signals and are observed at $t = 0, 1, \cdots, T-1$, and a, b are scalars, then

$$d_{(ag(t)+bh(t))}^T(f_k) = ad_g^T(f_k) + bd_h^T(f_k). \tag{3.9}$$

This property can be proved straightforwardly from the definition of the finite Fourier

transform, so that

$$d^T_{(ag(t)+bh(t))}(f_k) \;=\; \sum_{t=0}^{T-1} \Big(ag(t) + bh(t)\Big) e^{-2i\pi f_k t},$$

$$=\; a \sum_{t=0}^{T-1} g(t) e^{-2i\pi f_k t} + b \sum_{t=0}^{T-1} h(t) e^{-2i\pi f_k t}$$

$$=\; a\, d^T_g(f_k) + b\, d^T_h(f_k).$$

The original function $g(t)$ can be reconstructed from its finite Fourier coefficients using the following relation which is referred to as an inverse form of the finite Fourier transform

$$g(t) = \frac{1}{T} \sum_{k=-T/2}^{T/2-1} d^T_g(f_k) e^{2i\pi f_k t}; \qquad t = 0, 1, \cdots, T-1. \tag{3.10}$$

The inverse of the finite Fourier transform of the function $g(t)$, can be deduced by multiplying equation(3.4) by $e^{2i\pi f_k \acute{t}}$ to obtain

$$d^T_g(f_k) e^{2i\pi f_k \acute{t}} = \sum_{t=0}^{T-1} g(t)\, e^{-2i\pi f_k t} e^{2i\pi f_k \acute{t}}.$$

Taking the summation over k from $k = -T/2$ to $k = T/2 - 1$ gives

$$\sum_{k=-T/2}^{T/2-1} d^T_g(f_k) e^{2i\pi f_k \acute{t}} = \sum_{k=-T/2}^{T/2-1} \sum_{t=0}^{T-1} g(t)\, e^{-2i\pi f_k t} e^{2i\pi f_k \acute{t}}.$$

Reversing the order of summation in this double sum leads to

$$\sum_{k=-T/2}^{T/2-1} d^T_g(f_k) e^{2i\pi f_k \acute{t}} = \sum_{t=0}^{T-1} g(t) \sum_{k=-T/2}^{T/2-1} e^{-2i\pi f_k (t-\acute{t})}. \tag{3.11}$$

To simplify equation (3.11), we take $u = e^{-2i\pi(t-\acute{t})/T}$. Thus equation (3.11) becomes

$$\sum_{k=-T/2}^{T/2-1} d^T_g(f_k) e^{2i\pi f_k \acute{t}} = \sum_{t=0}^{T-1} g(t) \sum_{k=-T/2}^{T/2-1} u^k. \tag{3.12}$$

However, the second sum has value

$$\sum_{k=-T/2}^{T/2-1} u^k = \left[\begin{array}{lll} T & ; & u = 1, \\[2mm] u^{-T/2}\dfrac{(1-u^T)}{1-u} & ; & u \neq 1. \end{array} \right.$$

The case $u = 1$ corresponds to the situation $t = \acute{t}$, but under all circumstances u is a T-th root of unity, that is, $u^T = 1$. Thus for t and $\acute{t}$ taking integer values in the intervals $0 \leq t < T$ and $0 \leq \acute{t} < T$, it follows that

$$\sum_{k=-T/2}^{T/2-1} u^k = \left[\begin{array}{lll} T & ; & t = \acute{t}, \\[2mm] 0 & ; & t \neq \acute{t}. \end{array} \right.$$

Consequently, equation (3.12) leads to the solution

$$g(t) = \frac{1}{T} \sum_{k=-T/2}^{T/2-1} d_g^T(f_k) e^{2i\pi f_k t} ; \qquad t = 0, 1, \cdots, T-1 . \tag{3.13}$$

Thus the proof is completed.

3.4   Finite Fourier transform applications

3.4.1   Parseval's theorem

This theorem states that the sum of the square of a function $g(t)$ is equal to the average of the sum of the square of its transform. In other words, the energy contained in a function $g(t)$ is equivalent to the average of the energy contained in its Fourier coefficients. That is,

$$\sum_{t=0}^{T-1} |g(t)|^2 = \frac{1}{T} \sum_{k=-T/2}^{T/2-1} |d_g^T(f_k)|^2 . \tag{3.14}$$

Proof:

By using the definition of the inverse finite Fourier Transform, equation (3.10), we get

$$
\begin{aligned}
\sum_{t=0}^{T-1} |g(t)|^2 &= \frac{1}{T^2} \sum_{t=0}^{T-1} \left(\sum_{k=-T/2}^{T/2-1} d_g^T(f_k) \, e^{-\frac{2i\pi k}{T} t} \right) \left(\sum_{j=-T/2}^{T/2-1} \overline{d_g^T(f_j)} \, e^{\frac{2i\pi j}{T} t} \right) \\
&= \frac{1}{T^2} \sum_{t=0}^{T-1} \left(\sum_{k=-T/2}^{T/2-1} \sum_{j=-T/2}^{T/2-1} d_g^T(f_k) \overline{d_g^T(f_j)} \, e^{-\frac{2i\pi(k-j)}{T} t} \right) ,
\end{aligned}
\tag{3.15}
$$

where "overline", as mentioned before, denotes the complex conjugate. By choosing $m = k - j$ and interchanging sums we get

$$\sum_{t=0}^{T-1} |g(t)|^2 = \frac{1}{T^2} \sum_{k=-T/2}^{T/2-1} \sum_{j=-T/2}^{T/2-1} d_g^T(f_k) \overline{d_g^T(f_j)} \left(\sum_{t=0}^{T-1} e^{\frac{-2i\pi m}{T} t} \right) . \tag{3.16}$$

The sum between parentheses has a solution T when $m = 0$ and zero otherwise. Hence equation (3.16) becomes

$$\sum_{t=0}^{T-1} |g(t)|^2 = \frac{1}{T} \sum_{k=-T/2}^{T/2-1} |d_g^T(f_k)|^2. \tag{3.17}$$

The quantity $|d_g^T(f_k)|^2$ is called the energy spectral density for the signal $g(t)$, while the average of the energy spectral density gives the power spectral density for $g(t)$.

3.4.2 Convolution theorem

Let $g(t)$ and $h(t)$ be two T-dimensional sequences and let $d_g^T(f)$ and $d_h^T(f)$ be their finite Fourier transforms respectively, then

$$d_{(g*h)}^T(f) = d_g^T(f)d_h^T(f), \tag{3.18}$$

where the asterisk denotes the convolution between the two functions $g(t)$ and $h(t)$. This formula can be derived straightforwardly from the finite Fourier transform and convolution definition, as follows

$$
\begin{aligned}
d_{(g*h)}^T(f_k) &= \sum_{t=0}^{T-1} (g*h)(t)\, e^{-\frac{2i\pi}{T}kt} \\
&= \sum_{t=0}^{T-1} \left(\sum_{t-\acute{t}=0}^{T-1} g(\acute{t})h(t-\acute{t}) \right) e^{-\frac{2i\pi}{T}kt} \\
&= \left(\sum_{\acute{t}=0}^{T-1} g(\acute{t})e^{-\frac{2i\pi}{T}k\acute{t}} \right) \left(\sum_{t-\acute{t}=0}^{T-1} h(t-\acute{t})e^{-\frac{2i\pi}{T}k(t-\acute{t})} \right).
\end{aligned}
$$

Let $t_1 = t - \acute{t}$ to get

$$
\begin{aligned}
d_{(g*h)}^T(f_k) &= d_g^T(f_k) \sum_{t_1=0}^{T-1} h(t_1)\, e^{-\frac{2i\pi}{T}t_1} \\
&= d_g^T(f_k)\, d_h^T(f_k).
\end{aligned}
\tag{3.19}
$$

Thus the Fourier transform of the convolution of the signals $g(t)$ and $h(t)$ is the product of their Fourier transforms $d_g^T(f)$ and $d_h^T(f)$.

3.5 Fourier transform of a stationary stochastic process

In the previous section, we introduced the Fourier transform for deterministic functions, and the finite Fourier transform for relatively short lengths of record and its analytical properties. The reconstruction of the original signal from its Fourier coefficients has also been demonstrated. In Chapter 2, we introduced the definition of a stochastic process as a series of random variables, namely $\{X_t\}$, and noted that these random variables are observed at times $t = 0, \pm 1, \pm 2, \cdots$. Also it is known that every realization of the random variable represents a function (signal) of time. In addition, we demonstrated the concept of the stationarity of stochastic process. The Fourier transform of the stationary stochastic

process will now be defined.

Definition

Let $X = \{X_t\}$, where $t = 0, \pm 1, \pm 2, \cdots$, be a stationary stochastic process, then the Fourier transform of X at frequency f is given by

$$d_X(f) = \sum_{t=-\infty}^{\infty} X_t \, e^{-2i\pi ft}; \qquad -\infty < f < \infty, \tag{3.20}$$

where the Fourier coefficient $d_X(f)$ is a complex-valued function of frequency f. The spectral quantity $|d_X(f)|^2$ is a real-valued function of frequency f and it is called the power spectrum of the stationary process X. However, when the values of the stationary process X are available at the time points $t = 0, 1, 2, \cdots, T - 1$, then the finite Fourier transform of X is given by

$$d_X^T(f_k) = \sum_{t=0}^{T-1} X_t \, e^{-2i\pi f_k t}; \qquad -T/2 < k \le T/2, \tag{3.21}$$

where $d_X^T(f_k)$ are called the **finite Fourier coefficients** of the stationary stochastic process X at frequency $f_k = \dfrac{k}{T}$, and the superscript T denotes the size of sample from which the Fourier transform is calculated. For a sufficiently large number of observations T, and according to the central limit theorem, the Fourier coefficients for the stationary process X are asymptotically independent complex normally distributed variates. The following theorem gives the statistical distribution of the Fourier coefficients and their components (Brillinger [1981]).

Theorem 3.5.1 Let X be a stationary stochastic process whose values are available at $t = 0, 1, \cdots, T - 1$, and let $d_X^T(f_k), -T/2 < k \le T/2$, defined in equation (3.21), be the finite Fourier coefficients of X. Then as $T \to \infty$, the finite Fourier coefficients $d_X^T(f_k), -T/2 < k \le T/2$, are approximately independent complex normally distributed variables with zero mean value and variance $2\pi T |d_X^T(f_k)|^2$, that is, $d_X^T(f_k) \sim N^C\big(0, 2\pi T |d_X^T(f)|^2\big)$, $-T/2 < k \le T/2$,. The real, Re $d_X^T(f_k)$, and imaginary, Im $d_X^T(f_k)$, parts of the Fourier coefficients are asymptotically independent normally distributed variables $N\big(0, \pi T |d_X^T(f)|^2\big)$, $-T/2 < k \le T/2$.

The finite Fourier coefficients of the stationary process are invertible and we can reconstruct the stationary process X by recovering its realizations X_t from its finite Fourier coefficients using the relation

$$X_t = \frac{1}{T} \sum_{k=-T/2}^{T/2-1} d_X^T(f_k) e^{2i\pi f_k t}; \qquad t = 0, 1, \cdots, T - 1.$$

Since the spectral density function is the core of the spectral analysis, we will present two approaches that are broadly used in calculating the spectral density function from Fourier coefficients and from covariance functions.

3.6   Spectral density via Fourier coefficients

Let X be a stationary stochastic process whose values are available at times $t = 0, 1, \cdots, T-1$, then the power spectrum of the stationary stochastic process X at frequency f, denoted by $S_{XX}(f), f \in (-1/2, 1/2]$, is defined in terms of the finite Fourier coefficients by

$$
\begin{aligned}
S_{XX}(f) &= d_X^T(f) \times \overline{d_X^T(f)}, \\
&= |d_X^T(f)|^2,
\end{aligned}
\tag{3.22}
$$

where $\overline{d_X^T(f)}$ represents the complex conjugate of the Fourier coefficient $d_X^T(f)$ at frequency f. From this definition, we note that the spectrum is a non-negative definite function of frequency and describes the distribution of the power (variance) of the stationary stochastic process with respect to frequency. Similarly, the spectrum of the stationary process Y at a specific frequency $f \in (-1/2, 1/2]$, is denoted by $S_{YY}(f)$ and is defined as

$$
S_{YY}(f) = |d_Y^T(f)|^2.
\tag{3.23}
$$

Note that the power spectra $S_{XX}(f)$ and $S_{YY}(f)$ are real-valued functions of frequency f, whereas the cross-spectrum between the stationary processes X and Y at frequency f, denoted by $S_{XY}(f)$, is a complex-valued function of frequency f and is defined in terms of the finite Fourier coefficients by

$$
S_{XY}(f) = d_X^T(f) \times \overline{d_Y^T(f)}.
\tag{3.24}
$$

3.7   Spectral density via covariance function

The auto- and cross-spectral density functions, $S_{XX}(f)$, $S_{YY}(f)$ and $S_{XY}(f)$ for the stationary processes X and Y, correspond to the auto-covariances $C_{XX}(s)$ and $C_{YY}(s)$ for the stationary processes X and Y respectively, and the cross-covariance $C_{XY}(s)$ between these processes in the time domain, where s represents the difference between two consecutive times and takes the values $s = 0, \pm 1, \cdots$. Recall the definitions of the auto- and

cross-covariance from equations (2.12), (2.13) and (2.14), namely

$$
\begin{aligned}
C_{XX}(s) &= \mathbb{E}\left[X_t\,X_{t+s}\right] - \mu_X^2\,, \\
C_{YY}(s) &= \mathbb{E}\left[Y_t\,Y_{t+s}\right] - \mu_Y^2\,, \\
C_{XY}(s) &= \mathbb{E}\left[(X_t - \mu_X)(Y_{t+s} - \mu_Y)\right],
\end{aligned}
\tag{3.25}
$$

where $\mathbb{E}$ denotes the expectation, and μ_X and μ_Y are time-independent mean values of the processes X and Y respectively. The spectral density functions $S_{XX}(f)$, $S_{YY}(f)$ and $S_{XY}(f)$ can be defined in terms of the auto- and cross-covariance functions $C_{XX}(s)$, $C_{YY}(s)$ and $C_{XY}(s)$, provided these functions satisfy

$$
\begin{aligned}
\sum_{s=-\infty}^{\infty} |C_{XX}(s)| &< \infty\,, \\
\sum_{s=-\infty}^{\infty} |C_{YY}(s)| &< \infty\,, \\
\sum_{s=-\infty}^{\infty} |C_{XY}(s)| &< \infty\,.
\end{aligned}
\tag{3.26}
$$

The spectra and cross-spectra at a specific frequency f, can be expressed in terms of the Fourier transform of the auto- and cross-covariance as follows (see, Brillinger [1981], Percival and Walden [1993], Bendat and Piersol [1986])

$$
\begin{aligned}
S_{XX}(f) &= \sum_{s=-\infty}^{\infty} C_{XX}(s)e^{-2i\pi fs}\,, \\
S_{YY}(f) &= \sum_{s=-\infty}^{\infty} C_{YY}(s)e^{-2i\pi fs}\,, \\
S_{XY}(f) &= \sum_{s=-\infty}^{\infty} C_{XY}(s)e^{-2i\pi fs}\,.
\end{aligned}
\tag{3.27}
$$

These summations always exist over finite length of records. Due to the symmetry properties of the auto- and cross-covariance functions for the stationary processes stated in equations (2.15) and (2.16), the following properties are satisfied for the auto- and cross-spectral density functions

$$
\begin{aligned}
\overline{S_{XX}(f)} &= S_{XX}(-f) = S_{XX}^{\tau}(f)\,, \\
\overline{S_{YY}(f)} &= S_{YY}(-f) = S_{YY}^{\tau}(f)\,, \\
\overline{S_{XY}(f)} &= S_{XY}(-f) = S_{XY}^{\tau}(f)\,.
\end{aligned}
\tag{3.28}
$$

The superscript τ denotes the transportation operator. The power spectra $S_{XX}(f)$ and $S_{YY}(f)$ are real-valued functions of frequency f, while the cross-spectral density is a

complex-valued function of frequency f. From these properties, we note that the basic frequency domain for the spectral density function is $[0, 1/2]$. Since the first two relations in equation (3.28) are special cases of the last relation when $X = Y$, it is sufficient to prove (3.28, 3) and then deduce results (3.28, 1) and (3.28, 2). The proof of the relation (3.28, 3) can be obtained as follows. By definition

$$
\begin{aligned}
\overline{S_{XY}(f)} &= \sum_{s=-\infty}^{\infty} C_{XY}(s)e^{2i\pi fs}, \\
S_{XY}(-f) &= \sum_{s=-\infty}^{\infty} C_{XY}(s)e^{2i\pi fs}, \\
S_{XY}^{\tau}(f) &= \sum_{s=-\infty}^{\infty} C_{XY}^{\tau}(s)e^{-2i\pi fs}.
\end{aligned}
\tag{3.29}
$$

From equations (3.29, 1) and (3.29, 2), it is obvious that $\overline{S_{XY}(f)} = S_{XY}(-f)$. Now to prove that $S_{XY}(-f) = S_{XY}^{\tau}(f)$, we will use the the property $C_{XY}^{\tau}(s) = C_{XY}(-s)$ and rearrange the exponential function argument to be $2i\pi f(-s)$ in the definition of $S_{XY}^{\tau}(f)$. Hence,

$$
S_{XY}^{\tau}(f) = \sum_{s=-\infty}^{\infty} C_{XY}(-s)e^{2i\pi f(-s)}.
\tag{3.30}
$$

By choosing $u = -s$, equation (3.30) become

$$
S_{XY}^{\tau}(f) = \sum_{u=-\infty}^{\infty} C_{XY}(u)e^{2i\pi fu}.
\tag{3.31}
$$

The right hand side of the equation (3.31) is the definition of the $S_{XY}(-f)$, therefore $S_{XY}(-f) = S_{XY}^{\tau}(f)$. The proof is completed.

The expressions in equation (3.27) are invertible. Thus the auto- and cross-covariance functions can be obtained from the definition of the spectral density in equation (3.27). Thus

$$
\begin{aligned}
C_{XX}(s) &= \int_{-1/2}^{1/2} S_{XX}(f)e^{2i\pi fs}\, df \\
C_{XY}(s) &= \int_{-1/2}^{1/2} S_{XY}(f)e^{2i\pi fs}\, df \quad ; \quad s = 0, \pm 1, \cdots
\end{aligned}
\tag{3.32}
$$

For the particular case when $s = 0$

$$
C_{XX}(0) = \int_{-1/2}^{1/2} S_{XX}(f)\, df.
\tag{3.33}
$$

But equation (2.12) indicates that the variance of the stationary process X is $\sigma_X^2 = C_{XX}(0)$. Consequently

$$
\sigma_X^2 = \int_{-1/2}^{1/2} S_{XX}(f)\, df.
\tag{3.34}
$$

3.8   Nyquist frequency and aliasing

The highest frequency that can be extracted from the data is called the *Nyquist* frequency. Suppose that the observations have been taken at equally spaced intervals of length Δt, which is called the sample rate, then the *Nyquist* frequency is $f_{Nq} = 1/2\Delta t$ and the corresponding angular *Nyquist* frequency takes the value $\omega_{Nq} = 2\pi f_{Nq} = \pi/\Delta t$. The *Nyquist* frequency is also called the *folding* frequency since all higher frequencies are effectively folded down into the interval $[0, f_{Nq}]$. On the other hand, the lowest frequency (resolution) can be determined in terms of the *Nyquist* frequency and the sample size T as

$$f_{resl} = \frac{2 \times \text{Nyquist frequency}}{\text{sample size}} = \frac{2f_{Nq}}{T}.$$

Aliasing is a simple phenomenon which occurs naturally when the signal of interest is sampled at equal spacing in time, in which two frequencies f and $-f$ become indistinguishable. To explain this problem, consider the observation X_t at time $\acute{t} = t\Delta t = t/2f_{Nq}$ from a pure cosine wave at frequency f, where $0 < f \le f_{Nq}$. Since every observation of the stochastic process can be expressed as a deterministic function, then the observation at $\acute{t}$ will be (see, Bloomfield [2000], Percival [1993])

$$X_t = \cos(2\pi f \acute{t}),$$

$$= \cos(\pi ft/f_{Nq}).$$

This function X_t fluctuates rapidly as f increases from zero to $f = f_{Nq}$,

$$X_t = \cos(\pi t) = (-1)^t.$$

Now suppose that the value of the frequency f exceeds the value of the *Nyquist* frequency f_{Nq}, such that $f_{Nq} < f < 2f_{Nq}$ and let $\acute{f} = 2f_{Nq} - f$, then the new value of the observation X_t at a frequency $\acute{f}$ is given by

$$X_t = \cos(2\pi \acute{f} \acute{t}) = \cos(2\pi(2f_{Nq} - f)\acute{t}).$$

By substituting the value of $\acute{t} = t/2f_{Nq}$ and rearranging the argument of the cosine function, we get

$$X_t = \cos(2\pi t - \pi ft/f_{Nq}).$$

But, by applying the trigonometric identity $\cos(A \pm B) = \cos(A)\cos(B) \mp \sin(A)\sin(B)$, we obtain

$$X_t = \cos(\pi ft/f_{Nq}) = \cos(2\pi f \acute{t}).$$

Thus all data at frequencies, $\acute{f} = 2f_{Nq} - f$, have the same cosine function as the data at frequency f, when sampled at point $\acute{t} = t/2f_{Nq}$. Hence the frequencies f and $\acute{f}$ are indistinguishable and they may be said to be aliasing of each other. Consequently, we can say that every frequency not in the interval $0 \leq f \leq f_{Nq}$ has an alias in that interval.

Shannon's [1949] sampling theorem places restrictions on the frequency contents of the time signal X_t and can be simply stated as "The signal X_t can be recovered exactly from its Fourier coefficients, if it is sampling at a rate greater than twice its highest frequency". However, the sampling of the signal X_t at rates below its highest frequency f_{Nq} results in aliasing.

3.9   Fast Fourier transform

The discrete Fourier transform (DFT), demonstrated in section 3.5 can be applied to any real or complex valued series. For example if we have the series $X = \{X_t, t = 0, 1, 2, \cdots, T - 1\}$ then the discrete Fourier transform of X is given by

$$d_X^T(f_k) = \sum_{t=0}^{T-1} X_t\, e^{-2i\pi f_k t}; \qquad -T/2 < k \leq T/2\,.$$

The calculation of the discrete Fourier transform of the series X directly from this definition requires T^2 multiplications. In practice, when the length of the series becomes large, the computation of the discrete Fourier transform will become more computationally intensive and will take a considerable amount of time, as the times taken are proportional to the square of the number of points in the series. Therefore it is important to look for a technique which reduces the time consumed in calculating the discrete Fourier transform of the series $\{X_t, t = 0, 1, 2, \cdots, T-1\}$ and is also more accurate. A much faster algorithm has been developed by Cooley and Tukey [1965] and is called the FFT "*Fast Fourier Transform*", while the basic ideas were discussed much earlier (see, Cooley *et al* [1967]). The concept of this technique depends on decomposing the series length T into small integer factors, hence, the simple version of the FFT occurs when T is a power of 2. On the other hand, Good ([1958], [1971]) presented an algorithm to decompose T when it either has a prime factor or cannot be written as a power of 2. For more information (see, Brillinger [1981], Chatfield [1996], and Bloomfield [2000]).

Let $T = T_1 T_2$, where T_1 and T_2 are integers, and suppose that time and frequency indices take the form $k = k_1 T_2 + k_2$ and $t = t_2 T_1 + t_1$, then the discrete Fourier transform

of the series $\{X_t, t = 0, 1, 2, \cdots, T - 1\}$ may be written as

$$d_X^T(f_k) = \sum_{t_1=0}^{T_1-1} \sum_{t_2=0}^{T_2-1} X_{(t_2 T_1 + t_1)}\, e^{-\frac{2i\pi}{T_1 T_2}(k_1 T_2 + k_2)(t_2 T_1 + t_1)}; \qquad -T/2 < k \leq T/2, \qquad (3.35)$$

where $f_k = k/T$, $-\frac{T_1}{2} < k_1 \leq \frac{T_1}{2}$ and $-\frac{T_2}{2} < k_2 \leq \frac{T_2}{2}$.

Equation (3.35) can be rearranged as

$$
\begin{aligned}
d_X^T(f_k) &= \sum_{t_1=0}^{T_1-1} \sum_{t_2=0}^{T_2-1} X_{(t_2 T_1 + t_1)}\, e^{-\frac{2i\pi}{T_1 T_2}[k_1 T_2 t_2 T_1 + k_1 T_2 t_1 + k_2 t_2 T_1 + k_2 t_1]} \\
&= \sum_{t_1=0}^{T_1-1} e^{-\frac{2i\pi}{T} k_2 t_1} \left(e^{-\frac{2i\pi}{T_1} k_1 t_1} \sum_{t_2=0}^{T_2} X_{(t_2 T_1 + t_1)} e^{-\frac{2i\pi}{T_2} k_2 t_2} \right).
\end{aligned}
\tag{3.36}
$$

The inner sum represents the discrete Fourier transform of size T_2, the outer sum represents the discrete Fourier transform of size T_1 and the exponential term $e^{-\frac{2i\pi}{T} k_2 t_1}$ is called the twiddle factor. It can be shown that the term $(k_1 T_2 t_2 T_1)$ vanishes because of primitivity $(e^{-\frac{2i\pi}{T_1 T_2} k_1 T_2 t_2 T_1} = e^{-2i\pi k_1 t_2} = 1)$. Thus the total number of complex multiplications required for this process is $(T_1 + T_2) T_1 T_2 = (T_1 + T_2) T$, and the computational cost is reduced, and is proportional to $T \log T$, instead of T^2 if we compute the finite Fourier coefficients directly from the DFT definition.

If $T_1, T_2, \cdots, T_m$ are integers, such that $T = T_1 T_2 \cdots T_m$, then the speed comparison is between T^2 and $T(T_1 + T_2 + \cdots + T_m)$. In practice, $T_1 + T_2 + \cdots + T_m$ is of the order of $\log T$ and, therefore, the FFT has a computational speed of order $T \log T$ for large T.

3.10 Spectra Estimation

The natural estimate of the spectra and cross-spectra is the *periodogram* or *sample periodogram* . The definition of the periodogram is given by Brillinger [1981], Bloomfield [2000], Percival [1993] and Cryer and Chan [2008] as follows.

Definition 3.10.1 Let $X = \{X_t, t = 0, \pm 1, \cdots\}$ be a stationary stochastic process with theoretical power spectrum $S_{XX}(f), -\infty < f < \infty$. Its finite Fourier coefficients are given by

$$d_X^T(f_k) = \sum_{t=0}^{T-1} X_t\, e^{-2i\pi f_k t}; \qquad -T/2 < k \leq T/2.$$

The estimate of the power spectrum (periodogram) denoted by $I_{XX}^T(f_k)$ is given by

$$I_{XX}^T(f_k) = \frac{1}{2\pi T} |d_X^T(f_k)|^2; \qquad -T/2 < k \leq T/2. \tag{3.37}$$

Due to its dependence on the finite Fourier coefficients, the symmetry, non-negativity and periodicity properties are held for $I_{XX}^T(f_k), f_k = \dfrac{k}{T}$. As a result of the symmetry property of the periodogram $I_{XX}^T(f_k)$, we will henceforth be concerned only with values at $f_k \in [0, 1/2]$. The statistical properties of the sample periodogram have been discussed intensively in various literatures including (Brillinger [1981], Percival [1993], Cryer and Chan [2008]) and others.

Theorem 3.10.1 Let $X = \{X_t, t = 0, \pm 1, \cdots\}$ be a stationary stochastic process with power spectrum $S_{XX}(f)$, and let $I_{XX}^T(f_k)$ be a sample periodogram of X at frequency f_k, where $0 < k < T/2$, then the periodogram ordinates $I_{XX}^T(f_k)$ are asymptotically independent $S_{XX}(f_k)\chi_2^2/2$ variates for $0 < k < T/2$, whereas $I_{XX}^T(f_k)$ are asymptotically independent $S_{XX}(f_k)\chi_1^2$ variates for $k = 0$ or $k = T/2$.

The χ_ν^2 denotes a chi-squared variable with ν degrees of freedom, while $\chi_2^2/2$ is an exponential variate with mean 1. Statistically, the chi-squared variate with ν degrees of freedom is a summation of the squares of ν independent identical standard normally distributed variates, where the expected value of the chi-squared variable is ν and its variance is 2ν.

Proof:

To prove this theorem, we rewrite equation (3.37) in terms of the components of $d_X^T(f_k)$ as

$$I_{XX}^T(f_k) = \frac{1}{2\pi T}\Big\{ \big[\operatorname{Re} d_X^T(f_k)\big]^2 + \big[\operatorname{Im} \, d_X^T(f_k)\big]^2 \Big\}; \qquad 0 \le k \le T/2 \,. \qquad (3.38)$$

But theorem (3.5.1) indicated that $\operatorname{Re} d_X^T(f_k)$ and $\operatorname{Im} d_X^T(f_k)$ are asymptotically independent normally distributed with zero mean values and variances $\pi T S_{XX}(f_k)$ for $-T/2 < k \le T/2$. That is

$$\Big(\frac{\operatorname{Re} d_X^T(f_k)}{\sqrt{\pi T S_{XX}(f_k)}}\Big) \sim N(0,1) \, \text{and} \, \Big(\frac{\operatorname{Im} d_X^T(f_k)}{\sqrt{\pi T S_{XX}(f_k)}}\Big) \sim N(0,1) \,.$$

Consequently, the sum of the squares of the standard normal variables $\dfrac{1}{\pi T S_{XX}(f_k)} \times \Big\{ \big[\operatorname{Re} d_X^T(f_k)\big]^2 + \big[\operatorname{Im} \, d_X^T(f_k)\big]^2 \Big\}$ is distributed as a chi-squared variable according to the statistical fact that, the sum of the squares of the standard normally distributed variables is a chi-squared variate with a degree of freedom equals to the number of the variates being summed. Thus from equation (3.38) we get

$$\frac{2I_{XX}^T(f_k)}{S_{XX}(f_k)} = \frac{1}{\pi T S_{XX}(f_k)}\Big\{ \big[\operatorname{Re} d_X^T(f_k)\big]^2 + \big[\operatorname{Im} \, d_X^T(f_k)\big]^2 \Big\} \sim \chi_2^2 \,.$$

Thus, $I_{XX}^T(f_k) \sim S_{XX}(f_k) \chi_2^2/2$ for $0 < k < T/2$. In cases $k = 0$ or $k = T/2$, the term $\operatorname{Im} d_X^T(f_k)$ is cancelled out. This in turn leads to a reduction in the degrees of freedom to 1

instead of 2. Hence, the periodogram will be distributed as chi-squared with 1 degree of freedom, that is $I_{XX}^T(f_k) \sim S_{XX}(f_k) \chi_1^2/2$.

The expected value and variance of the periodogram $I_{XX}^T(f_k)$ are given by

$$\begin{aligned}
\mathbb{E}\, I_{XX}^T(f_k) &= S_{XX}(f_k) + O(T^{-1})\,, \\[2mm]
\operatorname{Var} I_{XX}^T(f_k) &= S_{XX}^2(f_k) + O(T^{-1})\,; \qquad 0 \le k \le T/2\,.
\end{aligned} \tag{3.39}$$

It is obvious from these statistics that the periodogram estimator is an inconsistent estimator of the power spectrum due to its poor bias and, it has been proved to have poor variance. The error in the bias stems from the truncation of the record at length T, which causes the spectral leakage or loss of energy. This in turn leads to $\mathbb{E}\, I_{XX}^T(f_k) \ne S_{XX}(f_k)$, the expected value of the periodogram estimator is not equal to the power spectrum. Although increasing the sample size T used in calculating the periodogram improves the resolution of the frequency and reduces the bias, the variance of the periodogram remains at the level of $S_{XX}^2(f_k)$, and does not reduce because it is independent of the sample size. In order to resolve the issues of bias and variability of the priodogram estimator, several alternative methods have been proposed to provide reasonable spectral density estimators take into account the trade-off between reducing variability and introducing bias. For example, smoothed periodogram, and averaging periodograms across adjacent sections of the single record.

3.10.1   Smoothed periodogram

The idea of this method is based upon the fact that the changes occur in the spectral density over small frequency intervals are relatively small. The smoothed periodogram averages the values of the sample periodogram over small intervals of frequency, which consequently reduces the variability. This method has been investigated by Brillinger [1981], Percival [1993], Diggle [1990] and Cryer and Chan [2008].

The smoothed periodogram denoted by $\hat{S}_{XX}(f_k)$ at frequency $f_k = \dfrac{k}{T}$, can be obtained by averaging the values of the sample periodogram over the small band of frequency centred by f_k and then extending m Fourier frequencies on either side of f_k. This means that every periodogarm at frequency f_j lies in the interval $[f_{k-m}, f_{k+m}]$ represents an estimate of the power spectrum $S_{XX}(f_k)$. Thus we have a set of $2m+1$ independent estimators for one quantity $S_{XX}(f_k)$

$$I_{XX}^T(f_{k-m})\,, \cdots, I_{XX}^T(f_k)\,, \cdots, I_{XX}^T(f_{k+m})\,.$$

The average of these estimators provides a reasonable unbiased and less variable estimator for the spectrum $S_{XX}(f_k)$

$$\hat{S}_{XX}(f_k) = \frac{1}{2m+1} \sum_{j=-m}^{m} I_{XX}^T(f_{k+j}); \qquad 0 \le k \le T/2. \qquad (3.40)$$

Note that the power spectrum properties of symmetry, periodicity and non-negativity are also satisfied for this estimator of the power spectrum, since it depends on the finite Fourier coefficients $d_X^T(f_k)$. The statistical properties of the smoothed periodogram $\hat{S}_{XX}(f_k)$ are

$$\mathbb{E}\,\hat{S}_{XX}(f_k) \approx S_{XX}(f_k), \qquad 0 < k < T/2,$$

$$\operatorname{Var}\hat{S}_{XX}(f_k) \approx \frac{S_{XX}^2(f_k)}{2m+1}; \qquad 0 < k < T/2, \qquad (3.41)$$

$$\approx \frac{S_{XX}^2(f_k)}{m}; \qquad k = 0 \,\text{or}\, k = T/2.$$

Thus, in the limit, the estimator $\hat{S}_{XX}(f_k)$ is unbiased, since the expected value of $\hat{S}_{XX}(f_k)$ approaches asymptotically the value of the power spectrum $S_{XX}(f_k)$ at frequency f_k. However, the number of adjacent periodogram ordinates m should not be too large when compared with the length of data T, as this may increase the bias of the estimator $\hat{S}_{XX}(f_k)$. In terms of variability, the large number of adjacent periodogram ordinates m at frequencies close to f_k, produced good stability of the periodogram $\hat{S}_{XX}(f_k)$, as the variance of the average of $(2m+1)$ adjacent periodogram ordinates is approximately $1/(2m+1)$ times the variance of the periodogram at the frequency f_k, when $0 < k < T/2$, and $(1/m)$ times the variance of the periodogram at the frequency f_k when the frequency index is $k = 0$ or $k = T/2$. The following theorem gives the asymptotic distribution of the smoothed periodogram $\hat{S}_{XX}(f_k)$ (see, Brillinger [1981]).

Theorem 3.10.2 Let $X = \{X_t, t = 0, \pm 1, \cdots\}$ be a stationary stochastic process and suppose that $\hat{S}_{XX}(f_k), 0 < k < T/2$ is the estimation of the real power spectrum $S_{XX}(f_k)$ at frequency f_k, such that

$$\hat{S}_{XX}(f_k) = \frac{1}{2m+1} \sum_{j=-m}^{m} I_{XX}^T(f_{k+j}); \qquad 0 \le k \le T/2.$$

Then $\hat{S}_{XX}(f_k)$ at frequencies $f_k, 0 < k < T/2$ is approximately distributed as $S_{XX}(f_k)\chi^2_{(4m+2)}/(4m+2)$ and is asymptotically distributed as $S_{XX}(f_k)\chi^2_{2m}/(2m)$, if $k = 0$ or $k = T/2$.

Remark 3.10.3 The periodograms $I_{XX}^T(f_q)$, where $q = j + k; -m < q < m + N/2$, are approximately independent $S_{XX}(f_q)\chi_2^2/2$ variates for $-m < q < T/2 + m$, and asymp-

totically $S_{XX}(f_q)\,\chi_1^2$ if $q = -m$ or $q = m + T/2$, as indicated in theorem (3.10.1). Consequently, the spectral estimator $\hat{S}_{XX}(f_k)$ distributes approximately as chi-squared variate with $(4m + 2)$ degrees of freedom, *i.e.*

$$\hat{S}_{XX}(f_k) \sim S_{XX}(f_k)\chi_{2(2m+1)}^2/2(2m+1) \quad \text{for} \quad 0 < k < T/2,$$

and asymptotically as chi-squared variate with $2m$ degrees of freedom, *i.e.*

$$\hat{S}_{XX}(f_k) \sim S_{XX}(f_k)\chi_{2m}^2/(2m) \quad \text{if} \quad k = 0 \quad \text{or} \quad k = T/2.$$

3.10.2 Averaging across periodograms of adjacent sections of single records

The smoothed periodogram estimate of the theoretical power spectra, based on taking the average of periodograms of contiguous segments of the set of data, was first proposed by Bartlett [1948]. Welch [1967] proposed the use of the average of overlapped periodograms. Thereafter, this estimate of the power spectra has been applied widely in various disciplines because of its high precision and the relative low cost of calculations produced by the Fast Fourier algorithm compared with the other estimators (see, Brillinger [1981], Bloomfield [2000], Percival [1993], Rosenberg *et al.* [1989], Attivissimo *et al.* [1995], and Amjad *et al.* [1997]). The periodograms asymptotic distribution and statistical properties have been discussed intensively by Brillinger [1981] and Percival [1993].

For large number of observations, it is convenient to estimate the spectrum using the disjoint structure, which depends on partitioning the main sample of length T into L disjoint sections of equal length N. One must first perform the finite Fourier coefficients for each segment, then calculate its periodogram, thereafter take the average of these periodograms with respect to the number of subsamples.

Disjoint sections procedures

Let $X = \{X_t, t = 0, \pm 1, \cdots, T - 1\}$ be a stationary stochastic process, and suppose that there are L non-overlapping disjoint sections each of which has a length of N, such that $T = LN$. Then the finite Fourier transform for each segment will be implemented as

$$d_X^N(f_k, l) = \sum_{t=(l-1)N}^{lN-1} X_t\, e^{-2i\pi f_k t}; \qquad 0 \le k \le N/2, \tag{3.42}$$

where $d_X^N(f_k, l); l = 1, \cdots, L$ and denotes the finite Fourier coefficient for the lth section at frequency $f_k = \dfrac{k}{N}$. The periodogram of the lth section at frequency f_k is then denoted by

$$I_{XX}^N(f_k, l) = \frac{1}{2\pi N}|d_X^N(f_k, l)|^2; \qquad 0 \le k \le N/2, \tag{3.43}$$

where $l = 1, 2, \cdots, L$. The average of these periodograms at a specific frequency f_k gives the estimation $\hat{S}_{XX}(f_k)$ of the theoretical power spectrum $S_{XX}(f_k)$ at that specific frequency, that is

$$\hat{S}_{XX}(f_k) = \frac{1}{L} \sum_{l=1}^{L} I_{XX}^N(f_k, l); \qquad 0 \leq k \leq N/2. \tag{3.44}$$

The $\hat{S}_{XX}(f)$ is symmetric, non-negative, and a periodic function of frequency f identical to that of function $I_{XX}^N(f)$. The theorem (3.10.4) gives the approximate distribution of the averaged periodograms of the contiguous segments of the data.

Theorem 3.10.4 Let $X = \{X_t, t = 0, \pm 1, \cdots\}$ be a stationary stochastic process with power spectrum $S_{XX}(f)$. Let $I_{XX}^N(f_k, l)$ be the sample periodogram of the lth section of N observations at frequency f_k, that is

$$I_{XX}^N(f_k, l) = \frac{1}{2\pi N} |d_X^N(f_k, l)|^2, \tag{3.45}$$

for $0 \leq k \leq N/2$, $l = 1, 2, \cdots, L$. Let

$$\hat{S}_{XX}(f_k) = \frac{1}{L} \sum_{l=1}^{L} I_{XX}^N(f_k, l); \qquad 0 \leq k \leq N/2, \tag{3.46}$$

where $T = LN$, then the periodogram $\hat{S}_{XX}(f_k)$ is asymptotically distributed as $S_{XX}(f_k)\chi_{2L}^2/(2L)$ for $0 < k < T/2$, and $\hat{S}_{XX}(f_k)$ are asymptotically independent $S_{XX}(f_k)\chi_L^2/L$ variables for $k = 0$ or $k = T/2$.

Remark 3.10.5 The periodograms $I_{XX}^N(f_k, l), l = 1, \cdots, L$, where $0 < k < N/2$ are approximately independent $S_{XX}(f_k)\chi_2^2/2$ variates for $0 < k < T/2$ and asymptotically $S_{XX}(f_k)\chi_1^2$ if $k = 0$ or $k = T/2$ as indicated in theorem (3.10.1).

The average of periodograms over non-overlapped segments of the time series is unbiased and provides an accurate estimator of the power spectrum. The reduction in variance of the estimator is inversely proportional to the number of segments L, that is, the variance of the estimator decreases as the number of disjoint sections L increases. For a fixed size of subsample N, the decrease of the variance of the estimator implies an increase in its bias, so we must be careful in employing this method. The expected value of $\hat{S}_{XX}(f_k)$ and its variance are given by

$$\begin{aligned}
\mathbb{E}\, \hat{S}_{XX}(f_k) &\approx S_{XX}(f_k); \qquad 0 < k < N/2, \\
\operatorname{Var} \hat{S}_{XX}(f_k) &\approx \frac{S_{XX}^2(f_k)}{2L}; \qquad 0 < k < N/2, \\
&\approx \frac{S_{XX}^2(f_k)}{L}; \qquad k = 0 \operatorname{or} k = N/2.
\end{aligned} \tag{3.47}$$

3.11 Confidence intervals for the spectrum

In order to evaluate the accuracy of an estimate to the parameter, it is often useful to construct a confidence interval for the parameter based on the distributional properties of the estimator. Thus, we will use the asymptotic distributions for the power spectrum estimate $\hat{S}_{XX}(f)$ defined in equations (3.40) and (3.44) to obtain the required confidence intervals. In general, the estimate $\hat{S}_{XX}(f)$ is distributed as chi-squared variable $\frac{S_{XX}}{\nu}\chi_\nu^2$ with ν degrees of freedom. Let $\chi^2_{\nu,\alpha/2}$ be the $(\alpha/2)100\%$ percentage point of the chi-squared distribution χ_ν^2, such that

$$\text{prob}\left[\chi^2_{\nu,\alpha/2} < \chi_\nu^2 < \chi^2_{\nu,(1-\alpha/2)}\right] = 1 - \alpha.$$

Since the estimate $\hat{S}_{XX}(f)$ is distributed asymptotically as $\frac{S_{XX}(f)}{\nu}\chi_\nu^2$, it follows that the variable $\frac{\nu\hat{S}_{XX}(f)}{S_{XX}(f)}$ is distributed approximately as χ_ν^2. Thus,

$$\chi^2_{\nu,\alpha/2} < \frac{\nu\hat{S}_{XX}(f)}{S_{XX}(f)} < \chi^2_{\nu,(1-\alpha/2)}.$$

Consequently, the confidence interval for the power spectrum $S_{XX}(f)$ is

$$\frac{\nu\hat{S}_{XX}(f)}{\chi^2_{\nu,(1-\alpha/2)}} < S_{XX}(f) < \frac{\nu\hat{S}_{XX}(f)}{\chi^2_{\nu,\alpha/2}}.$$

This interval represents a $(1-\alpha)100\%$ confidence interval for the power spectrum at a particular frequency f. This means that for every frequency in the domain of the function $S_{XX}(f)$ there is a different confidence interval at the level of significance $\alpha\%$, as the width of the confidence interval changes according to the value of the estimate $\hat{S}_{XX}(f)$ which varies from frequency to frequency. However, taking the logarithm of the confidence interval, provides a confidence interval that is independent of frequency, that is

$$\log(\hat{S}_{XX}(f)) + \log\left(\frac{\nu}{\chi^2_{\nu,(1-\alpha/2)}}\right) < \log(S_{XX}(f)) < \log(\hat{S}_{XX}(f)) + \log\left(\frac{\nu}{\chi^2_{\nu,\alpha/2}}\right). \quad (3.48)$$

Thus, the inequality in (3.48) represents the $(1-\alpha)100\%$ confidence interval for the $\log(S_{XX}(f))$. It is obvious from (3.48) that the width of the confidence interval

$$\left[\log(\hat{S}_{XX}(f)) + \log\left(\frac{\nu}{\chi^2_{\nu,(1-\alpha/2)}}\right), \log(\hat{S}_{XX}(f)) + \log\left(\frac{\nu}{\chi^2_{\nu,\alpha/2}}\right)\right],$$

is given by

$$\log\left(\frac{\nu}{\chi^2_{\nu,\alpha/2}}\right) - \log\left(\frac{\nu}{\chi^2_{\nu,(1-\alpha/2)}}\right) = \log\left(\frac{\chi^2_{\nu,(1-\alpha/2)}}{\chi^2_{\nu,\alpha/2}}\right),$$

and is independent of $\log(\hat{S}_{XX}(f))$. Consequently, it is independent of frequency. There-fore, the confidence interval for the logarithm of the power spectrum relates to the esti-mate based on averaging periodogram ordinates near the frequency f_k, defined in equation (3.40), is

$$\left[\log(\hat{S}_{XX}(f_k)) + \log\left(\frac{(4m+2)}{\chi^2_{(4m+2),\,(1-\alpha/2)}} \right), \log(\hat{S}_{XX}(f_k)) + \log\left(\frac{(4m+2)}{\chi^2_{(4m+2),\,\alpha/2}} \right) \right],$$

for frequency indices, $0 < k < T/2$. The confidence interval at the level of significance $\alpha\%$ for frequency indices $k = 0$ or $k = T/2$ is given by

$$\left[\log(\hat{S}_{XX}(f_k)) + \log\left(\frac{2m}{\chi^2_{2m,\,(1-\alpha/2)}} \right), \log(\hat{S}_{XX}(f_k)) + \log\left(\frac{2m}{\chi^2_{2m,\,\alpha/2}} \right) \right].$$

The confidence interval for the logarithm of the power spectrum relates to the esti-mate based on averaging periodograms of contiguous sections of the single data, given in equation(3.40) is

$$\left[\log(\hat{S}_{XX}(f_k)) + \log\left(\frac{2L}{\chi^2_{2L,\,(1-\alpha/2)}} \right), \log(\hat{S}_{XX}(f_k)) + \log\left(\frac{2L}{\chi^2_{2L,\,\alpha/2}} \right) \right],$$

for $0 < k < N/2$. For the case of $k = 0$ or $k = N/2$, the confidence interval for the $\log(S_{XX}(f_k))$ takes the form

$$\left[\log(\hat{S}_{XX}(f_k)) + \log\left(\frac{L}{\chi^2_{L,\,(1-\alpha/2)}} \right), \log(\hat{S}_{XX}(f_k)) + \log\left(\frac{L}{\chi^2_{L,\,\alpha/2}} \right) \right].$$

3.12 Cross-spectra estimate

Until this point, we have presented the periodogram as an estimate of the power spec-trum for a stationary process with finite length. We have also discussed smoothing peri-odogram(s) through averaging the periodogram ordinates near one specific frequency or averaging periodograms of contiguous segments of the whole data. Moreover, we have dis-cussed the asymptotic distribution and the related statistical properties for each estimate. In this section, we will discuss the estimate of the cross-spectra of stationary processes. In general, the periodogram has a natural extension to multiple time series.

Let X and Y be two stationary stochastic processes with partial values $\{X_0, X_1, \cdots, X_{T-1}\}$ and $\{Y_0, Y_1, \cdots, Y_{T-1}\}$. Let $d_X^T(f_k)$ and $d_Y^T(f_k)$ represent finite Fourier coefficients for X

and Y respectively at frequency f_k, $-T/2 < k \le T/2$, that is

$$
\begin{aligned}
d_X^T(f_k) &= \sum_{t=0}^{T-1} X_t\, e^{-2i\pi f_k t}\,, \\
d_Y^T(f_k) &= \sum_{t=0}^{T-1} Y_t\, e^{-2i\pi f_k t}\,; \qquad -T/2 < k \le T/2\,.
\end{aligned}
\tag{3.49}
$$

Suppose that the cross-spectra between the stationary processes X and Y at a frequency f where $-\infty < f < \infty$ is $S_{XY}(f)$, then the estimate cross-periodogram of the cross-spectra $S_{XY}(f)$ is denoted by $I_{XY}^T(f_k)$ and given by

$$
I_{XY}^T(f_k) = \frac{1}{2\pi T} d_X^T(f_k)\, \overline{d_Y^T(f_k)}\,; \qquad -T/2 < k \le T/2\,.
\tag{3.50}
$$

The cross-periodogram satisfies the identities

$$
I_{XY}^T(-f_k) = I_{YX}^T(f_k) = \overline{I_{XY}(f_k)}\,.
\tag{3.51}
$$

The proofs of these identities are straightforward from the definition of the cross-peridogram and the properties of the Fourier coefficients. Thus, by definition,

$$
I_{XY}^T(-f_k) = \frac{1}{2\pi T} d_X^T(-f_k)\, \overline{d_Y^T(-f_k)}\,; \qquad -T/2 < k \le T/2\,.
$$

Using the property $d_X^T(-f_k) = \overline{d_X^T(f_k)}$, gives

$$
I_{XY}^T(-f_k) = \frac{1}{2\pi T} \overline{d_X^T(f_k)}\, \overline{\overline{d_Y^T(f_k)}}\,.
$$

Since $\overline{\overline{Z}} = Z$ and by using the definition of the cross-periodogram again, we get

$$
I_{XY}^T(-f_k) = \frac{1}{2\pi T} d_Y^T(f_k)\, \overline{d_X^T(f_k)} = I_{YX}^T(f_k)\,.
$$

Now we want to prove that $I_{XY}^T(-f_k) = \overline{I_{XY}(f_k)}$. By definition,

$$
I_{XY}^T(-f_k) = \frac{1}{2\pi T} d_X^T(-f_k)\, \overline{d_Y^T(-f_k)}\,.
$$

Using the property $\overline{d_Y^T(-f_k)} = d_Y^T(f_k)$ yields

$$
I_{XY}^T(-f_k) = \frac{1}{2\pi T} d_X^T(-f_k)\, d_Y^T(f_k)\,.
$$

Once again, since $Z = \overline{\overline{Z}}$, then we get

$$
I_{XY}^T(-f_k) = \frac{1}{2\pi T} \overline{d_X^T(f_k)}\, \overline{\overline{d_Y^T(f_k)}}\,.
$$

Since the complex conjugate of the product of two complex variables equals the product of their complex conjugates, that is $\overline{Z_1 Z_2} = \overline{Z_1}\,\overline{Z_2}$, we get

$$I_{XY}^T(-f_k) = \frac{1}{2\pi T}\overline{d_X^T(f_k)\,\overline{d_Y^T(f_k)}} = \overline{I_{XY}^T(f_k)}\,.$$

Therefore, the proof is completed. Note that, the cross-periodogram $I_{XY}^T(f_k)$ is a real-valued function of frequency at f_0 and $f_{T/2}$, since the imaginary parts of the Fourier coefficients are vanished at those frequencies. however, this value is not necessarily positive. Also, there is another important identity which relates the cross-periodogram $I_{XY}^T(f_k)$ with periodograms $I_{XX}^T(f_k)$ and $I_{YY}^T(f_k)$ where $-T/2 < k \leq T/2$, namely

$$|I_{XY}^T(f_k)|^2 = I_{XX}^T(f_k)\,I_{YY}^T(f_k)\,; \qquad -T/2 < k \leq T/2\,. \tag{3.52}$$

The proof of this identity can be deduced immediately from the definition of auto- and cross-periodograms, that is

$$
\begin{aligned}
|I_{XY}^T(f_k)|^2 &= I_{XY}^T(f_k)\,\overline{I_{XY}^T(f_k)} \\
&= \left(\frac{1}{2\pi T}d_X^T(f_k)\,\overline{d_Y^T(f_k)}\right)\left(\overline{\frac{1}{2\pi T}d_X^T(f_k)\,\overline{d_Y^T(f_k)}}\right).
\end{aligned}
$$

By using the properties of the Fourier coefficients, mentioned above and the complex variables multiplication properties, we obtain

$$
|I_{XY}^T(f_k)|^2 = \left(\frac{1}{2\pi T}d_X^T(f_k)\,\overline{d_X^T(f_k)}\right)\left(\frac{1}{2\pi T}d_Y^T(f_k)\,\overline{d_Y^T(f_k)}\right).
$$

Obviously, the two terms on the right hand side represent the definitions of the periodograms $I_{XX}^T(f_k)$ and $I_{YY}^T(f_k)$, thus $|I_{XY}^T(f_k)|^2 = I_{XX}^T(f_k)\,I_{YY}^T(f_k)$. An estimate of the cross-specrum $S_{XY}(f), -\infty < f < \infty$, can be constructed by smoothing the cross-periodogram in the same way that we smoothed the periodogram when estimating the power spectrum in the previous section.

3.12.1 Smoothing the cross-periodogram

To smooth the periodogram in order to estimate the power spectrum at a particular frequency, we average periodogram ordinates close to that frequency. Here we will average cross-periodogram ordinates in the neighbourhood of a specific frequency to produce a consistent estimate of the cross-spectrum.

Let $X = \{X_t; t = 0, \pm 1, \cdots\}$, and $Y = \{Y_t; t = 0, \pm 1, \cdots\}$ be two stationary stochastic processes with cross-spectrum $S_{XY}(f), -\infty < f < \infty$, and suppose that the two partial

sequences $\{X_0, X_1, \cdots, X_{T-1}\}$ and $\{Y_0, Y_1, \cdots, Y_{T-1}\}$ are known for the processes X and Y respectively, then the estimate of the cross-spectrum, denoted by $\hat{S}_{XY}(f_k)$ is given by

$$\hat{S}_{XY}(f_k) = \frac{1}{2m+1} \sum_{j=-m}^{m} I_{XY}^{T}(f_{k+j}); \qquad -T/2 < k \leq T/2, \qquad (3.53)$$

where $I_{XY}^{T}(f_k)$ represents the cross-periodogram defined in equation (3.50), and m represents the number of frequencies on each side of the main frequency f_k. The cross-periodogram ordinates are calculated at those frequencies *prior* to averaging them in order to produce the estimate of the cross-spectrum $\hat{S}_{XY}(f_k)$ at that specific frequency. The estimate $\hat{S}_{XY}(f)$ has the same symmetry and periodicity properties as the cross-spectrum $S_{XY}(f)$. Moreover, this estimate of the cross-spectrum has been proved to be an unbiased estimate when the length of the times series T tends to infinity (see, Brillinger [1981]).

3.12.2 Averaging cross-periodograms of the contiguous sections

This procedure is elaborated clearly in section (3.10.2) during the discussion of smoothing the periodogram as an estimate of the power spectrum in order to improve its statistical properties. Here we will apply this method to smooth the cross-spectrum.

Let $X = \{X_t; t = 0, \pm 1, \cdots\}$ and $Y = \{Y_t; t = 0, \pm 1, \cdots\}$ be two stationary stochastic processes with the theoretical cross-spectrum $S_{XY}(f), -\infty < f < \infty$, and suppose that the two partial sequences $\{X_0, X_1, \cdots, X_{T-1}\}$ and $\{Y_0, Y_1, \cdots, Y_{T-1}\}$ are known for the processes X and Y respectively. Suppose that the two stretches are divided into L disjoint blocks, in which each block involves N observations. The finite Fourier coefficients at every frequency f_k are computed for each segment as follows,

$$\begin{aligned} d_X^N(f_k, l) &= \sum_{t=(l-1)N}^{lN-1} X_t \, e^{-2i\pi f_k t}, \\ d_Y^N(f_k, l) &= \sum_{t=(l-1)N}^{lN-1} Y_t \, e^{-2i\pi f_k t}; \qquad -N/2 < k \leq N/2, \end{aligned} \qquad (3.54)$$

where $d_X^T(f_k, l)$ and $d_Y^T(f_k, l)$ represent the finite Fourier coefficients at frequency f_k for the lth section; $l = 0, 1, \cdots L - 1$, of the processes X and Y respectively. Then the cross-periodogram at frequency f_k for the lth section, denoted by $I_{XY}^N(f_k)$, is

$$I_{XY}^N(f_k, l) = \frac{1}{2\pi N} d_X^N(f_k, l) \, \overline{d_Y^N(f_k, l)}; \qquad -N/2 < k \leq N/2. \qquad (3.55)$$

The cross-spectrum estimate $\hat{S}_{XY}(f_k)$ where $-N/2 < k \leq N/2$ can be obtained by taking

the average of cross-periodograms, to get

$$\hat{S}_{XY}(f_k) = \frac{1}{L} \sum_{l=0}^{L-1} I_{XY}^N(f_k, l); \qquad -N/2 < k \le N/2 . \tag{3.56}$$

All previous estimates of the spectral density have performed good results for long sequences of data. However, when the length of the data is short, these methods exhibit undesirable results due to the paucity of the data. In the following section, we will discuss the "multi-taper estimate"of the spectral density function that suites the situation of short data.

3.13   Multi-taper method

In the two previous sections (3.10) and (3.12), we discussed the estimation of the auto- and cross-spectrum for finite time series. We began with a simple periodogram as an estimation for the power spectrum and a cross-periodogram for the cross-spectrum. We also described their limitations, namely bias due to the spectral leakage and high variance because of the data finiteness. Also, we presented the most common types of spectral density estimation methods, which have been developed to smooth the simple periodogram (or cross-periodogram) with the aim of reducing its bias and variability. Averaging across adjacent frequencies exhibits a good reduction of the bias for a compromised number of frequencies lying close to a particular frequency, while the improvement in the variance is small. The average across contiguous preriodograms (or cross-periodograms) represents a consistent, unbiased estimator of the power or cross-spectrum with acceptable variance for long stretches of data.

In this section, we will introduce a method that has been proved to overcome the problems of bias and variability that usually accompany the spectral density estimate, especially for the short data sets. This method was developed by Thomson [1982] as a new estimate for the signal power spectrum. The new technique is based on the multiplication of the whole data sequence by a set of weights or tapers, which makes the Thomson scheme different from the Blackman-Tukey method, which is based on multiplying the data by a single, constant taper. The main characteristic of Thomson tapers is that they form an orthogonal basis, which not only leads to a reduced variance of the spectral estimates, but also to a minimization of the spectral leakage due to the finite-length of the time series (see, Slepian [1983], Thomson [1982]) .

Each window is constructed so that it provides an independent sample of the data sequence, while maximising the resistance to spectral leakage (Park *et al.* [1987], Thomson [1990]). The statistical information discarded from the first taper is partially regained from the second one, while the third taper retrieves the statistical information which is discarded from the first two tapers and so on. Only a few low-order tapers are used as the higher-order tapers allow an unacceptable level of spectral leakage (Percival and Walden [1993]). These tapers belong to a family of functions known as Discrete Prolate Spheroidal Sequences (DPSS) or Slepian Functions. These functions simultaneously optimise energy concentration in time and frequency when either or both has a definite limit and define the dimensionality of the time-frequency region. However, the application of the Slepian functions is only suitable for continuous time and continuous frequency problems.

3.14 Slepian functions

The Discrete Prolate Spheroidal Sequences (DPSS) or Slepian functions were developed by Slepian [1978] and arise as the eigenvectors of the Toeplitz eigenvalue problem

$$\sum_{t=0}^{T-1} \frac{\sin 2\pi W(t' - t)}{\pi(t' - t)} \, v_{t,k}(W) = \lambda_k(W) \, v_{t',k}(W); \quad t' = 0, 1, \cdots, T - 1, \qquad (3.57)$$

where T is the length of the data and W is the half-bandwidth parameter, which describes a small local frequency band centred around $f\!:\!|f - f'| \leq W$. Thus, the Slepian sequences are orthogonal time-limited functions, highly concentrated in the frequency band $[-W, W]$. The eigenfunctions λ_k are usually sorted in descending order of associated eigenvalue, that is $(1 > \lambda_0 > \lambda_1 > ... > \lambda_{K-1})$, where λ_k gives the fraction of total energy retained within the band $[-W, W]$. Only the first $K = 2TW$ eigenvalues are used because these are the only tapers for which leakage is small.

3.14.1 The multi-taper procedure in estimating the power spectrum

The multi-taper method weighs the entire data sequence to get an estimated spectrum for each taper. The final power spectrum is then computed by averaging these individual spectra.

Let $X_0, \cdots, X_{T-1}$ be realizations of the stationary stochastic process X with the theoretical power spectrum $S_{XX}(f), -\infty < f < \infty$. Suppose that the sampling interval between observations is Δt, then the *Nyquist* frequency will be $f_{Nq} = \dfrac{1}{(2\Delta t)}$ and the

fundamental Fourier frequency (*frequency resolution*) is $f_{resl} = \dfrac{1}{(T\Delta t)}$. The multi-taper procedure is described in the following steps (Percival and Walden [1993]).

[*i*] Determine the value of the half-bandwidth parameter W. The usual strategy is to choose W to be a multiple $j > 1$ of the frequency resolution $(T\,\Delta t)^{-1}$, *i.e.* $W = \dfrac{j}{(T\Delta t)}$, and the common values of j are 2, 3 or 4. Note that the value of j may be taken to be larger than 4 in some applications, and some applications are interested in taking j to be non-integer multiples (see, Thomson [1982]).

[*ii*] According to the equation (3.57), the first K of the eigenvalues λ_k and the eigenvectors $v_{t,k}$ are required, that is the values $\lambda_0, \cdots, \lambda_{K-1}$ and $v_{t,0}, \cdots, v_{t,K-1}$, must be computed, where K is the greatest integer less than or equal to $2TW$.

[*iii*] Each eigenvector $v_{t,k}$, which is the kth Slepian function, is applied to the whole data sequence, and the finite Fourier coefficients are computed at every frequency f from the sum (using the fast Fourier transform, defined in section 3.9)

$$d_X^T(f_j, k) = \sum_{t=0}^{T-1} X_t\, v_{t,k}\, e^{-2i\pi ft}\,; \qquad k = 0, \cdots, K-1, \qquad 0 \le j \le T/2, \quad (3.58)$$

where $f_j = j/T$ is the Fourier frequency, and $d_X^T(f_j, k)$, is called the kth eigencoefficient, is the discrete Fourier transform of the product of the data and the kth Slepian function of the length T at frequency f_j. Consequently, the smoothed periodogram corresponding to the kth Slepian function, denoted by $I_{XX}^T(f_j, k)$, is the squared magnitude of the kth eigencoefficient, *i.e.*

$$I_{XX}^T(f_j, k) = \frac{1}{2\pi T}\left|d_X^T(f_j, k)\right|^2; \qquad k = 0, \cdots, K-1, \qquad 0 \le j \le T/2\,,$$

and is known as the kth eigenspectrum.

[*iv*] The multi-taper spectrum estimate, denoted by $\hat{S}_{XX}(f_j)$ of the theoretical power spectrum $S_{XX}(f)$, is the average of the K eigenspectra, that is,

$$\hat{S}_{XX}(f_j) = \frac{1}{K}\sum_{k=0}^{K-1} I_{XX}^T(f_j, k)\,; \qquad 0 \le j \le T/2\,. \qquad (3.59)$$

3.14.2   Cross-spectra estimate

To achieve the estimate of the cross-spectrum between two stationary processes, namely X and Y, we follow the same procedures of estimating the power spectra, which are elaborated in the previous section.

Suppose that $X = \{X_t, t = 0, 1, \cdots\}$ and $Y = \{Y_t, t = 0, 1, \cdots\}$ are two stationary stochastic processes with the theoretical cross-spectrum $S_{XY}(f)$, where $-\infty < f < \infty$, and suppose that the finite T realizations of these two processes are known, then the finite Fourier coefficients at frequency f_j for each process are obtained from the expressions

$$
\begin{aligned}
d_X^T(f_j, k) &= \sum_{t=0}^{T-1} X_t\, v_{t,k}\, e^{-2i\pi f_j t}\,, \\
d_Y^T(f_j, k) &= \sum_{t=0}^{T-1} Y_t\, v_{t,k}\, e^{-2i\pi f_j t}\,; \qquad k = 0, \cdots, K-1, \qquad 0 \le j \le T/2\,.
\end{aligned}
\tag{3.60}
$$

Thus, the k cross-eigenspectrum of the process X with the process Y at frequency f_j, is denoted by $I_{XY}^T(f_j, k)$, and defined by

$$
I_{XY}^T(f_j, k) = \frac{1}{2\pi T} d_X^T(f_j, k)\, \overline{d_Y^T(f_j, k)}\,; \qquad k = 0, \cdots, K-1, \qquad 0 \le j \le T/2\,. \tag{3.61}
$$

Consequently, the estimated cross-spectrum at frequency f_j, denoted by $\hat{S}_{XY}(f_j)$ of the theoretical cross-spectrum $S_{XY}(f)$, is produced by taking the average of the K cross-eigenspectra *i.e.*

$$
\hat{S}_{XY}(f_j) = \frac{1}{K} \sum_{k=0}^{K-1} I_{XY}^T(f_j, k)\,; \qquad 0 \le j \le T/2\,. \tag{3.62}
$$

3.14.3 Asymptotic distribution of the eigenspectrum

Although the periodogram is considered as a natural estimate of the spectral density, as we mentioned in sections 3.10 and 3.12, it has been proved that, it is a biased estimator because of its bias and variance. Indeed, the increase in the length of the data reduces the bias and improves the frequency resolution, but unfortunately, it cannot improve the variability of the estimator. In order to improve the accuracy of the periodogram, several procedures were developed to reduce the bias and variance of the periodogram. We have already provided two methods that are widely used in smoothing periodograms, namely averaging across the interior frequency for the whole data and averaging periodograms across adjacent sections from single records. As shown previously, all these estimators were distributed approximately as squared-chi variable with different degrees of freedom according to the nature of the estimator.

We will now investigate the asymptotic distribution of the multi-taper estimator. Each eigenspectrum $I_{XX}^T(f_j, k)$ represents an estimate of the power spectrum at frequency f_j, which is in fact a kind of smoothing periodogram that is produced by tapering the data

prior to calculating its finite Fourier coefficients (eigencoefficients) $d_X^T(f_j, k)$. Hence, since the kth eigenspectrum $I_{XX}^T(f_j, k)$ is the squared magnitude of a complex Gaussian random variable, as given in theorem (3.5.1), it is distributed as $S_{XX}(f_j) \chi_2^2/2$ for $k = 0, 1, \cdots, K-1$ and frequency $0 < j < T/2$, and as $S_{XX}(f_j) \chi_1^2$ for $(j = 0)$ or $(j = T/2)$. Consequently, the multi-taper estimator of the power spectrum $\hat{S}_{XX}(f_j)$ will asymptotically distribute as $S_{XX}(f_j) \chi_{2K}^2/2K$ for $j < 0 < T/2$. The variance of the multi-taper estimator of the power spectrum is approximately $S_{XX}^2(f_j)/K$ (see, Percival and Walden [1993], Walter and Soleski [2005]).

3.14.4 Estimating the Slepian functions

Slepian functions are the eigenvectors computed from equation (3.57). However, a direct approach based on this equation is rarely used in practice because of the infeasibility of dealing with $T \times T$ matrices when T is large, and, of course, the computational effort required to determine the eigenvalues and eigenvectors of such matrices. The standard method for computing Discrete Prolate Spheroidal functions, uses an expansion in Legendre polynomials or an expansion in Bessel functions. Here we will use the approach based on Legendre polynomials (Miranian [2004], Boyd [2004], Gosse [2010]).

The Slepian functions are defined in the Fourier domain and were originally considered as possible solutions for energy concentration problems (see, Slepian [1983], Percival and Walden [1993]).

Let $g(t)$ be a signal and suppose that the values of the signal $g(t)$ are available at $t = 0, 1, \cdots, T-1$, then its continuous Fourier transform is given by

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-2i\pi ft} dt, \tag{3.63}$$

and the recovery of the signal from its Fourier transform is captured by

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{2i\pi ft} df, \tag{3.64}$$

where f denotes the frequency in Hertz (Hz). The ratio of the energy of the signal $g(t)$ lying in a time interval $[-T/2, T/2]$ to that in the entire signal, is defined by

$$\lambda = \frac{\displaystyle\int_{-T/2}^{T/2} |g(t)|^2 dt}{\displaystyle\int_{-\infty}^{\infty} |g(t)|^2 dt}. \tag{3.65}$$

Evidently $\lambda \in (0, 1)$. The function $g(t)$ is now replaced by its Fourier transform in the numerator and denominator of this ratio, to get

$$\lambda = \frac{\int_{-T/2}^{T/2} \left(\int_{-\infty}^{\infty} G(f) \, e^{2i\pi ft} \, df \right) \left(\int_{-\infty}^{\infty} \overline{G}(f') \, e^{-2i\pi f't} \, df' \right) dt}{\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} G(f) \, e^{2i\pi ft} \, df \right) \left(\int_{-\infty}^{\infty} \overline{G}(f') \, e^{-2i\pi f't} \, df' \right) dt} \, . \tag{3.66}$$

Interchanging the time integral with the frequency integrals, and gathering the Fourier coefficients out of the time integral, gives

$$\lambda = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(f) \overline{G}(f') \left(\int_{-T/2}^{T/2} e^{2i\pi(f-f')t} \, dt \right) df \, df'}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(f) \overline{G}(f') \left(\int_{-\infty}^{\infty} e^{2i\pi(f-f')t} \, dt \right) df \, df'} \, . \tag{3.67}$$

The numerator time integral can be solved as follows

$$\begin{aligned} \int_{-T/2}^{T/2} e^{2i\pi(f-f')t} \, dt &= \frac{1}{2i\pi(f-f')} \left[e^{2i\pi(f-f')t} \right]_{-T/2}^{T/2} \\ &= \frac{1}{2i\pi(f-f')} \left[e^{i\pi(f-f')T} - e^{-i\pi(f-f')T} \right] . \end{aligned} \tag{3.68}$$

Due to the evenness of the cosine function, and oddness of the sine function, equation (3.68) becomes

$$\begin{aligned} \int_{-T/2}^{T/2} e^{2i\pi(f-f')t} \, dt &= \frac{1}{2i\pi(f-f')} \left[2i \sin(T\pi(f-f')) \right] , \\ &= \frac{1}{\pi(f-f')} \sin(T\pi(f-f')) . \end{aligned} \tag{3.69}$$

The denominator infinite integral $(\int_{-\infty}^{\infty} e^{2i\pi ft} \, df)$ is basically the inverse Fourier transform of the Dirac $\delta(t)$ function, since

$$\hat{\delta}(f) = \int_{-\infty}^{\infty} \delta(t) e^{-2i\pi ft} \, dt = 1 \, ,$$

where $\hat{\delta}(f)$ represents the Fourier transform of $\delta(t)$ at a frequency f. Thus

$$\delta(t) = \int_{\infty}^{\infty} e^{2i\pi ft} \, df \, . \tag{3.70}$$

Substituting equations (3.69) and (3.70) into the equation (3.67) yields

$$\lambda = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(f) \overline{G}(f') \frac{\sin T\pi(f-f')}{\pi(f-f')} \, df \, df'}{\int_{-\infty}^{\infty} G(f) \overline{G}(f') \left(\int_{-\infty}^{\infty} e^{-2i\pi f't} \delta(t) \, dt \right) df'} \, , \tag{3.71}$$

which simplifies further, to give

$$\lambda = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(f)\,\overline{G}(f')\,\dfrac{\sin T\pi(f - f')}{\pi(f - f')}\,df\,df'}{\int_{-\infty}^{\infty} |G(f)|^2\,df}. \tag{3.72}$$

Suppose that $g(t)$ has compact support in the frequency domain, by which is meant that $G(f)$ is zero-valued for all $|f| > \sigma$. In this case, equation (3.71) becomes

$$\lambda = \frac{\int_{-\sigma}^{\sigma} \int_{-\sigma}^{\sigma} G(f)\,\overline{G}(f')\,\dfrac{\sin T\pi(f - f')}{\pi(f - f')}\,df\,df'}{\int_{-\sigma}^{\sigma} |G(f)|^2\,df}. \tag{3.73}$$

The objective of the subsequent analysis is to identify which function $G(f)$ maximizes the value of λ. Suppose that $G(f) = F(f)$ maximizes the value of λ, and let $G(f) = F(f) + \varepsilon h(f)$ in expression (3.73), then λ becomes a function of the parameter ε and the function $h(f)$, as expressed in the representation

$$\lambda(\varepsilon; h) = \frac{P(\varepsilon; h)}{Q(\varepsilon; h)}, \tag{3.74}$$

where $P(f)$ and $Q(f)$ are defined by the respective formulae

$$\begin{aligned} P(\varepsilon) &= \int_{-\sigma}^{\sigma} \int_{-\sigma}^{\sigma} \big(F(f) + \varepsilon\,h(f)\big)\big(\overline{F}(f') + \varepsilon\,\overline{h}(f')\big)\frac{\sin T\pi(f - f')}{\pi(f - f')}\,df\,df', \\ Q(\varepsilon) &= \int_{-\sigma}^{\sigma} \int_{-\sigma}^{\sigma} \big(F(f) + \varepsilon\,h(f)\big)\big(\overline{F}(f') + \varepsilon\,\overline{h}(f')\big)\,df\,df'. \end{aligned} \tag{3.75}$$

The function $\lambda(\varepsilon; h)$ has a stationary value at $G(f) = F(f)$, provided

$$\frac{\partial \lambda(0; h)}{\partial \varepsilon} = 0 \tag{3.76}$$

for arbitrary choice of the function $h(f)$. It follows immediately by logarithmic differentiation of expression (3.75), that

$$\frac{1}{\lambda(0; h)}\frac{\partial \lambda(0; h)}{\partial \varepsilon} = \frac{1}{P(0; h)}\frac{\partial P(0; h)}{\partial \varepsilon} - \frac{1}{Q(0; h)}\frac{\partial Q(0; h)}{\partial \varepsilon}.$$

According to equation (3.76), the condition of the stationary point of the function $\lambda(\varepsilon; h)$ at the value $G(f) = F(f)$, is required to satisfy

$$\frac{\partial P(0; h)}{\partial \varepsilon} - \lambda(0; h)\frac{\partial Q(0; h)}{\partial \varepsilon} = 0 \tag{3.77}$$

for arbitrary choice of the function $h(f)$. By replacing the derivatives of the functions $P(\varepsilon; h(f))$ and $Q(\varepsilon; h(f))$, with respect to the variable ε in equation (3.77), by their values

from equation (3.75), we get

$$
\int_{-\sigma}^{\sigma} \int_{-\sigma}^{\sigma} \left\{ h(f) \left[\overline{F}(f') + \varepsilon \, \overline{h}(f') \right] + \left[F(f) + \varepsilon \, h(f) \right] \overline{h}(f') \right\} \frac{\sin T\pi(f - f')}{\pi(f - f')} \, df \, df'
$$

$$
-\lambda(\varepsilon; h) \int_{-\sigma}^{\sigma} \int_{-\sigma}^{\sigma} \left\{ h(f) \left[\overline{F}(f') + \varepsilon \, \overline{h}(f') \right] + \left[F(f) + \varepsilon \, h(f) \right] \overline{h}(f') \right\} df \, df' \, . \tag{3.78}
$$

Therefore, the stationary value of $\lambda(\varepsilon; h)$ at $\varepsilon = 0$ satisfies

$$
\int_{-\sigma}^{\sigma} h(f) \left(\int_{-\sigma}^{\sigma} \overline{F}(f') \frac{\sin T\pi(f - f')}{\pi(f - f')} \, df' - \lambda(0; h) \, \overline{F}(f) \right) df
$$

$$
+ \int_{-\sigma}^{\sigma} \overline{h}(f') \left(\int_{-\sigma}^{\sigma} F(f) \frac{\sin T\pi(f - f')}{\pi(f - f')} \, df - \lambda(0; h) \, F(f') \right) df' = 0 \, , \tag{3.79}
$$

for all values of $h(f)$ and therefore

$$
\int_{-\sigma}^{\sigma} F(f) \frac{\sin T\pi(f - f')}{\pi(f - f')} \, df = \lambda \, F(f') \, . \tag{3.80}
$$

Thus $F(f)$ is an eigenfunction of a *Fredholm equation of the second kind* with eigenvalue λ, the value of the ratio, *i.e.* the fraction of the signal's energy concentrated in the frequency interval $[-\sigma, \sigma]$. Recall that the definition of λ, given in (3.65), ensures that the eigenvalues of equation (3.80) lie in $(0, 1)$, thereby guaranteeing that these eigenvalues can be organised in decreasing order of magnitude, that is

$$
1 \geq \lambda_0 \geq \lambda_1 \geq \cdots \geq 0 \, .
$$

The function $\left[\dfrac{\sin T\pi(f - f')}{\pi(f - f')} \right]$ is known as a kernel of the integral equation, hence the solution of the integral equation depends on the properties of the kernel function.

It is convenient at this instant to non-dimensionalize the eigenvalue problem posed in (3.80) by making the change of variable $x = f'/\sigma$, $y = f/\sigma$ and introducing the function $\psi(x) = F(x\sigma)$. The result of this non-dimensionalisation procedure is that the eigenvalue problem posed in (3.80) is modified into the eigenvalue problem

$$
\int_{-1}^{1} \frac{\sin c(y - x)}{\pi(y - x)} \, \psi(y) \, dy = \lambda \, \psi(x) \, ; \qquad c = T\pi\sigma \, , \qquad |x| \leq 1 \, . \tag{3.81}
$$

Thus, the solution $\psi(x)$ of the equation (3.81) depends on the values of T and σ through the parameter $c = T\pi\sigma$, due to the dependency of the kernel function on the value of the parameter c. However, there are countably infinite solutions of the equation (3.81), namely $\psi_0(x), \psi_1(x), \cdots$ corresponding to the fractions of the energy $\lambda_0, \lambda_1, \cdots$, where the eigenfunctions $\psi(x)$'s are known as prolate spheroidal wave functions.

Associated with the eigenvalue problem for $\psi(x)$, is the problem of reconstructing the

signal $g(t)$ from its Fourier transform $G(f)$. This reconstruction is achieved by using the inverse Fourier transform

$$g(t) = \int_{-\sigma}^{\sigma} e^{2\pi i f t} G(f) \, df$$

in which it has been noted that $G(f)$ has its support in $[-\sigma, \sigma]$. Under the change of variables $y = f/\sigma$, with time rescaled from the interval $[-T/2, T/2]$ to the interval $[-1, 1]$ by the mapping $x = 2t/T$, the previous equation becomes

$$g(Tx/2) = \sigma \int_{-1}^{1} e^{T\pi i\sigma yx} \psi(y) \, dy,$$

To simplify, we choose $f(x) = g(Tx/2)$ and $c = T\pi\sigma$, then the above expression will take the form

$$f(x) = \sigma \int_{-1}^{1} e^{icxy} \psi(y) \, dy; \qquad |x| \le 1.$$

To summarize, we want to recover the signal $g(t)$ from the eigenfunctions $\psi(y)$. To do so, we first need to compute the function $f(x)$, using the mathematical system which relates $f(x)$ with the eigenfunction $\psi(x)$,

$$
\begin{aligned}
\lambda\,\psi(x) &= \int_{-1}^{1} \frac{\sin c(y-x)}{\pi(y-x)} \psi(y) \, dy, \\
f(x) &= \sigma \int_{-1}^{1} e^{icxy} \psi(y) \, dy.
\end{aligned}
\tag{3.82}
$$

3.15 Characterization of the eigenfunctions $\psi(x)$

To achieve the required objective, the crucial idea is to construct an ordinary differential equation satisfied by $\psi(x)$, and use the solutions of this equation to deduce $f(x)$. The definition of the eigenfunction in (3.82) is first used to deduce that

$$\lambda \frac{d\psi}{dx} = \frac{1}{\pi} \int_{-1}^{1} \frac{d}{dx}\left(\frac{\sin c(y-x)}{(y-x)}\right) \psi(y) \, dy = -\frac{1}{\pi} \int_{-1}^{1} \frac{d}{dy}\left(\frac{\sin c(y-x)}{(y-x)}\right) \psi(y) \, dy.$$

Integration by parts applied to this identity, gives

$$
\begin{aligned}
\lambda \frac{d\psi}{dx} &= -\frac{1}{\pi}\left[\frac{\sin c(y-x)}{(y-x)} \psi(y)\right]_{-1}^{1} + \frac{1}{\pi} \int_{-1}^{1} \left(\frac{\sin c(y-x)}{(y-x)}\right) \frac{d\psi}{dy} \, dy \\
&= -\frac{1}{\pi}\left[\frac{\sin c(1-x)}{(1-x)} \psi(1) - \frac{\sin c(1+x)}{(1+x)}\psi(-1)\right] + \frac{1}{\pi} \int_{-1}^{1} \frac{\sin c(y-x)}{(y-x)} \frac{d\psi}{dy} \, dy.
\end{aligned}
$$

The next step in the calculation is to note that

$$
\begin{aligned}
\lambda \frac{d}{dx}\left[(1-x^2)\frac{d\psi}{dx}\right] &= \frac{1}{\pi} \frac{d}{dx} \int_{-1}^{1} \frac{(1-x^2)\sin c(y-x)}{(y-x)} \frac{d\psi}{dy} \, dy \\
&\quad + \frac{1}{\pi} \frac{d}{dx}\left[(1-x)\sin c(1+x)\psi(-1) - (1+x)\sin c(1-x)\,\psi(1)\right].
\end{aligned}
\tag{3.83}
$$

Now we will concentrate on the integral on the right hand side of this equation. Using the identity $1 - x^2 = (1 - y^2) + (y - x)(y + x)$, this integral may be re-expressed in the form

$$\frac{d}{dx} \int_{-1}^{1} \frac{\sin c(y - x)}{(y - x)} \left[(1 - y^2) \frac{d\psi}{dy} \right] dy + \frac{d}{dx} \int_{-1}^{1} (y + x) \sin c(y - x) \frac{d\psi}{dy} \, dy \,,$$

or may be equivalently represented by

$$-\int_{-1}^{1} \frac{d}{dy} \left(\frac{\sin c(y - x)}{(y - x)} \right) \left[(1 - y^2) \frac{d\psi}{dy} \right] dy$$

$$+ \int_{-1}^{1} \left(\sin c(y - x) - c(y + x) \cos c(y - x) \right) \frac{d\psi}{dy} \, dy \,. \tag{3.84}$$

Each integral of expression (3.84) is integrated by parts, to get

$$\int_{-1}^{1} \frac{\sin c(y - x)}{(y - x)} \frac{d}{dy} \left[(1 - y^2) \frac{d\psi}{dy} \right] dy - c^2 \int_{-1}^{1} (y + x) \sin c(y - x) \, \psi(y) \, dy$$

$$+ \left[\left(\sin c(y - x) - c(y + x) \cos c(y - x) \right) \psi(y) \right]_{-1}^{1} \,. \tag{3.85}$$

Replacing the integral on the right hand side of equation (3.83) by expression (3.85) yields

$$\lambda \frac{d}{dx} \left[(1 - x^2) \frac{d\psi}{dx} \right] = \frac{\psi(-1)}{\pi} \left[-\sin c(1 + x) + c(1 - x) \cos c(1 + x) \right]$$

$$- \frac{\psi(1)}{\pi} \left[\sin c(1 - x) - c(1 + x) \cos c(1 - x) \right]$$

$$+ \frac{1}{\pi} \int_{-1}^{1} \frac{\sin c(y - x)}{(y - x)} \frac{d}{dy} \left[(1 - y^2) \frac{d\psi}{dy} \right] dy - \frac{c^2}{\pi} \int_{-1}^{1} (y + x) \sin c(y - x) \, \psi(y) \, dy$$

$$+ \frac{\psi(1)}{\pi} \left[\sin c(1 - x) - c(1 + x) \cos c(1 - x) \right]$$

$$- \frac{\psi(-1)}{\pi} \left[-\sin c(1 + x) + c(1 - x) \cos c(1 + x) \right] \,,$$

which in turn simplifies, to give

$$\lambda \frac{d}{dx} \left[(1 - x^2) \frac{d\psi}{dx} \right] = \frac{1}{\pi} \int_{-1}^{1} \frac{\sin c(y - x)}{(y - x)} \frac{d}{dy} \left[(1 - y^2) \frac{d\psi}{dy} \right] dy$$

$$- \frac{c^2}{\pi} \int_{-1}^{1} (y + x) \sin c(y - x) \, \psi(y) \, dy \,. \tag{3.86}$$

The expression $(y + x)$, is now replaced by the ratio $(y^2 - x^2)/(y - x)$ into the equation (3.86), to give

$$\lambda \frac{d}{dx} \left[(1 - x^2) \frac{d\psi}{dx} \right] = \frac{1}{\pi} \int_{-1}^{1} \frac{\sin c(y - x)}{(y - x)} \frac{d}{dy} \left[(1 - y^2) \frac{d\psi}{dy} \right] dy$$

$$- \frac{c^2}{\pi} \int_{-1}^{1} (y^2 - x^2) \frac{\sin c(y - x)}{(y - x)} \, \psi(y) \, dy \,. \tag{3.87}$$

Rearranging this equation leads to

$$\lambda \frac{d}{dx}\left[(1-x^2)\frac{d\psi}{dx}\right] = \frac{1}{\pi}\int_{-1}^{1}\frac{\sin c(y-x)}{(y-x)}\frac{d}{dy}\left(\left[(1-y^2)\frac{d\psi}{dy}\right] - c^2y^2\,\psi(y)\right)dy$$

$$+c^2x^2\int_{-1}^{1}\frac{\sin c(y-x)}{\pi(y-x)}\,\psi(y)\,dy\,. \tag{3.88}$$

But the second integral on the right-hand side of equation (3.88) equals to $\lambda\,\psi(x)$, as defined in equation (3.82), thus

$$\lambda\left(\frac{d}{dx}\left[(1-x^2)\frac{d\psi}{dx}\right] - c^2x^2\psi(x)\right) = \frac{1}{\pi}\int_{-1}^{1}\frac{\sin c(y-x)}{y-x}\left(\frac{d}{dy}\left[(1-y^2)\frac{d\psi}{dy}\right] - c^2y^2\psi(y)\right)dy. \tag{3.89}$$

In effect, $-\frac{d}{dx}\left[(1-x^2)\frac{d\psi}{dx}\right] + c^2x^2\psi(x)$, satisfies the same equation as $\psi(x)$ and is therefore a multiple of $\psi(x)$. In conclusion, $\psi(x)$ satisfies the ordinary differential equation

$$-\frac{d}{dx}\left[(1-x^2)\frac{d\psi}{dx}\right] + c^2x^2\psi(x) = \gamma\,\psi(x)\,. \tag{3.90}$$

3.16 Computation of $f(x)$

The derivation of the properties of the function $f(x)$ begins by first multiplying equation (3.90) by e^{icxy} and integrating the resulting equation with respect to y over the interval $[-1,1]$ to get

$$-\int_{-1}^{1}\frac{d}{dy}\left[(1-y^2)\frac{d\psi}{dy}\right]e^{icxy}\,dy + c^2\int_{-1}^{1}y^2\,e^{icxy}\psi(y)\,dy = \gamma\int_{-1}^{1}e^{icxy}\psi(y)\,dy\,. \tag{3.91}$$

One integration by parts applied to the integral on the left hand side of equation (3.91) gives

$$\int_{-1}^{1}\frac{d}{dy}\left[(1-y^2)\frac{d\psi}{dy}\right]e^{icxy}\,dy = \left[(1-y^2)\frac{d\psi}{dy}e^{icxy}\right]_{-1}^{1} - icx\int_{-1}^{1}(1-y^2)\,e^{icxy}\frac{d\psi}{dy}\,dy$$

$$= -icx\int_{-1}^{1}(1-y^2)\,e^{icxy}\frac{d\psi}{dy}\,dy\,.$$

A further integration by parts gives

$$\int_{-1}^{1}(1-y^2)\,e^{icxy}\frac{d\psi}{dy}\,dy = \left[(1-y^2)\,e^{icxy}\psi(y)\right]_{-1}^{1} - \int_{-1}^{1}\frac{d}{dy}\left[(1-y^2)\,e^{icxy}\right]\psi(y)\,dy$$

$$= \int_{-1}^{1}\left[2y\,e^{icxy} - icx(1-y^2)\,e^{icxy}\right]\psi(y)\,dy\,.$$

Consequently

$$\int_{-1}^{1} \frac{d}{dy}\left[(1-y^2)\frac{d\psi}{dy}\right] e^{icxy}\, dy = \int_{-1}^{1}\left[-2icxy\, e^{icxy} - c^2x^2(1-y^2)\, e^{icxy}\right]\psi(y)\, dy$$

$$= -c^2x^2 f(x) - \frac{d}{dx}\left(x^2 ic \int_{-1}^{1} y\, e^{icxy}\psi(y)\, dy\right)$$

$$= -c^2x^2 f(x) - \frac{d}{dx}\left(x^2 \frac{d}{dx} \int_{-1}^{1} e^{icxy}\psi(y)\, dy\right)$$

$$= -c^2x^2 f(x) - \frac{d}{dx}\left(x^2 \frac{df}{dx}\right).$$

(3.92)

Furthermore,

$$c^2 \int_{-1}^{1} y^2\, e^{icxy}\psi(y)\, dy = -\frac{d^2}{dx^2}\int_{-1}^{1} e^{icxy}\psi(y)\, dy = -\frac{d^2 f}{dx^2}. \tag{3.93}$$

When results (3.92) and (3.93) are introduced into (3.91), after reorganisation the function $f(x)$, is seen to satisfy

$$-\frac{d}{dx}\left[(1-x^2)\frac{df}{dx}\right] + c^2x^2 f(x) = \gamma\, f(x), \tag{3.94}$$

which is precisely the same ordinary differential equation satisfied by $\psi(x)$. Therefore $f(x)$ is a multiple of $\psi(x)$. Let $\psi_1(x)$ and $\psi_2(x)$ be eigenfunctions corresponding to distinct eigenvalues γ_1 and γ_2 then

$$-\int_{-1}^{1} \frac{d}{dx}\left[(1-x^2)\frac{d\psi_1}{dx}\right]\psi_2\, dx + c^2 \int_{-1}^{1} x^2\psi_1(x)\psi_2(x)\, dx = \gamma_1 \int_{-1}^{1} \psi_1(x)\psi_2(x)\, dx, \quad (3.95)$$

and for $\psi_2(x)$

$$-\int_{-1}^{1} \frac{d}{dx}\left[(1-x^2)\frac{d\psi_2}{dx}\right]\psi_1\, dx + c^2 \int_{-1}^{1} x^2\psi_1(x)\psi_2(x)\, dx = \gamma_2 \int_{-1}^{1} \psi_1(x)\psi_2(x)\, dx. \quad (3.96)$$

The result of the integration by parts of the first expression in equation (3.95) is

$$\int_{-1}^{1} \frac{d}{dx}\left[(1-x^2)\frac{d\psi_1}{dx}\right]\psi_2\, dx = \left[\psi_2\,(1-x^2)\frac{d\psi_1}{dx}\right]_{-1}^{1} - \int_{-1}^{1} (1-x^2)\frac{d\psi_2}{dx}\frac{d\psi_1}{dx}\, dx.$$

Applying integration by parts, once again, for the first term in equation (3.95), gives

$$\int_{-1}^{1} \frac{d}{dx}\left[(1-x^2)\frac{d\psi_1}{dx}\right]\psi_2\, dx = -\left[(1-x^2)\frac{d\psi_2}{dx}\,\psi_1\right]_{-1}^{1} + \int_{-1}^{1} \frac{d}{dx}\left[(1-x^2)\frac{d\psi_2}{dx}\right]\psi_1\, dx.$$

Thus, by introducing the last result into equation (3.95), we will obtain

$$-\int_{-1}^{1} \frac{d}{dx}\left[(1-x^2)\frac{d\psi_2}{dx}\right]\psi_1\, dx + c^2 \int_{-1}^{1} x^2\psi_1(x)\psi_2(x)\, dx = \gamma_1 \int_{-1}^{1} \psi_1(x)\psi_2(x)\, dx.$$

The left hand-side of this equation is identical to the left hand-side of equation (3.96), and consequently, we get

$$(\gamma_1 - \gamma_2) \int_{-1}^{1} \psi_1(x)\psi_2(x)\, dx = 0.$$

Thus eigenfunctions corresponding to distinct eigenvalues are orthogonal.

3.17 Slepian function properties

From the foregoing discussion, we can summarize some properties of the eigenfunctions $\psi(x)$ and of the corresponding eigenvalues λ (see, Slepian [1983], Percival and Walden [1993], Boyd [2004]).

[*i*] The eigenfunctions are real-valued and orthogonal on the interval $[-1, 1]$, that is

$$\int_{-1}^{1} \psi_j(x)\psi_k(x)\, dx = 0; \quad \text{for} \quad j \neq k.$$

[*ii*] The eigenvalues are positive and satisfy the decreasing sequence

$$1 \geq \lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \cdots$$

and also satisfy the property $\lim_{k \to \infty} \lambda_k = 0$.

[*iii*] The definition of the eigenfunction $\psi(x)$, given in equation (3.81), was restricted to the values $x \in [-1, 1]$. The definition can be extended to the values $|x| > 1$, because the left hand-side of this equation is well-defined for all x, thus

$$\psi(x) = \frac{1}{\lambda} \int_{-1}^{1} \frac{\sin c(y - x)}{\pi(y - x)} \psi(y)\, dy; \qquad c = T\pi\sigma, \qquad |x| > 1.$$

Moreover, the orthogonality property is also satisfied by the eigenfunctions on the interval $(-\infty, \infty)$, as well as on the interval [-1, 1].

[*iv*] The Fourier transform for the eigenfunction $\psi_k(x)$ restricted to $|x| \leq 1$, has the same form as $\psi_k(x)$, except for scale change, *i.e.*

$$\psi_k(2\pi t/c) = \frac{1}{\alpha_k} \int_{-1}^{1} \psi(x)\, e^{-2i\pi xt}\, dx, \quad -\infty < t < \infty$$

where α_k is independent of x.

[*v*] The eigenfunction $\psi_k(x)$ is even or odd according to whether k is even or odd; $\psi_k(x)$ has exactly k zeros in the interval [-1, 1].

[*vi*] The normalization of the eigenfunctions contributes to the energy of unity over $(-\infty, \infty)$, that is

$$\int_{-\infty}^{\infty} \psi_j(x)\psi_k(x) = \delta_{jk}$$

where δ_{jk} is *Kronecker's delta* function,

$$\delta_{jk} = \begin{cases} 1 & \text{if} \quad j = k, \\ 0 & \text{if} \quad j \neq k. \end{cases}$$

This follows

$$\int_{-1}^{1} \psi_j(x)\psi_k(x) = \lambda_k \delta_{jk}\,.$$

The eigenfunctions are orthogonal on the interval $[-1, 1]$ and orthonormal on $(-\infty, \infty)$, where λ_k represents the fraction of the energy contained in $[-1, 1]$.

[*vii*] The eigenfunctions $\psi_k(x)$ and the eigenvalues λ_k are dependent variables of the parameter c. In addition, the eigenvalue λ_k steeply decreases from approximately unity to nearly zero value at $k = 2c/\pi = 2WT$, where k is called the *Shannon number*.

In the following section we will solve the eigenvalue problem (3.94), which is satisfied by the Slepian functions, using Legendre polynomials.

3.18 The Slepian eigenvalue problem

Prior to considering the Slepian eigenvalue problem, it will be useful to note some properties of Legendre polynomials. Let $P_k(x)$ be the Legendre polynomial of order h, then the family of such polynomials is defined by the generating formula (Xiao *et al.* [2001], Boyd [2004]).

$$\frac{1}{\sqrt{1 - 2xh + h^2}} = \sum_{k=0}^{\infty} P_k(x)\,h^k\,. \tag{3.97}$$

It is straightforward to show that $P_k(x)$ has the following properties.

[*i*] $P_k(x)$ is a polynomial of degree k satisfying the ordinary differential equation

$$\frac{d}{dy}\left[(1 - x^2)\frac{d\psi}{dx}\right] + k(k+1)\psi = 0\,. \tag{3.98}$$

[*ii*] The polynomials 1, x, x^2 and x^3 have unique representations in terms of Legendre polynomials, given by the formulae

$$1 = P_0(x)\,, \quad x = P_1(x)\,, \quad x^2 = \frac{2P_2(x) + P_0(x)}{3}\,, \quad x^3 = \frac{2P_3(x) + 3P_1(x)}{5}\,. \tag{3.99}$$

[*iii*] The family of Legendre polynomials satisfy the recurrence relation

$$x\,P_k(x) = \frac{(k+1)}{2k+1}\,P_{k+1}(x) + \frac{k}{2k+1}\,P_{k-1}(x)\,. \tag{3.100}$$

[*iv*] The family of Legendre polynomials are mutually orthogonal over the interval $[-1, 1]$, with respect to the weight function $w(x) = 1$, that is,

$$\int_{-1}^{1} P_j(x) P_k(x) \, dx = \frac{2\delta_{jk}}{j + k + 1}, \tag{3.101}$$

where δ_{jk} is *Kronecker function.*

The strategy is to approximate the eigenfunction $\psi(x)$ by the Legendre series

$$\psi(x) = \sum_{k=0}^{K} a_k \sqrt{\frac{2k + 1}{2}} P_k(x). \tag{3.102}$$

Of course, expression (3.102) is an approximation because it cannot satisfy equation (3.94) in the strict algebraic sense. Instead, the coefficients a_k are chosen to ensure that equation (3.94) is satisfied in the integrated sense, namely that

$$\int_{-1}^{1} \left[-\frac{d}{dx} \left[(1 - x^2) \frac{d\psi}{dx} \right] + c^2 x^2 \psi(x) \right] P_k(x) \, dx = \gamma \int_{-1}^{1} \psi(x) P_k(x) \, dx$$

for each integer $k = 0, \cdots, K$. This criterion guarantees that the remainder on substitution of approximation (3.102) into equation (3.94) is orthogonal to all polynomials of degree less than or equal to K. The self-adjoint nature of the differential operator in the previous equation, in combination with the properties of Legendre polynomials, allows this condition to be expressed in the simplified form

$$k(k + 1) \int_{-1}^{1} \psi(x) P_k(x) \, dx + c^2 \int_{-1}^{1} x^2 \psi(x) P_k(x) \, dx = \gamma \int_{-1}^{1} \psi(x) P_k(x) \, dx, \tag{3.103}$$

where k takes all integer values from $k = 0$ to $k = K$ inclusive. For notational convenience, it is beneficial to define

$$\psi_k = \int_{-1}^{1} \psi(x) P_k(x) \, dx. \tag{3.104}$$

When $k = 0$ and when $k = 1$ the equation (3.103) leads to the respective conditions

$$c^2 \int_{-1}^{1} x^2 \psi(x) P_0(x) \, dx = \gamma \int_{-1}^{1} \psi(x) P_0(x) \, dx,$$

$$2 \int_{-1}^{1} \psi(x) P_1(x) \, dx + c^2 \int_{-1}^{1} x^2 \psi(x) P_1(x) \, dx = \gamma \int_{-1}^{1} \psi(x) P_1(x) \, dx,$$

which, in view of properties (3.99) and definition (3.104,) become

$$\frac{c^2}{3} \psi_0(x) + \frac{2c^2}{3} \psi_2 = \gamma \psi_0,$$

$$\left(2 + \frac{3c^2}{5} \right) \psi_1 + \frac{2c^2}{5} \psi_3 = \gamma \psi_1. \tag{3.105}$$

See Appendix A. The conditions for $k \geq 2$ are determined using the identity

$$\int_{-1}^{1} x\psi(x)P_k(x)\,dx = \frac{k+1}{2k+1} \int_{-1}^{1} \psi(x)P_{k+1}(x)\,dx + \frac{k}{2k+1} \int_{-1}^{1} \psi(x)P_{k-1}(x)\,dx$$

which is a direct consequence of the recurrence relation (3.100). A second application of this identity gives, (see Appendix A),

$$
\begin{aligned}
\int_{-1}^{1} x^2\psi(x)P_k(x)\,dx \;=\;& \frac{(k+2)(k+1)}{(2k+3)(2k+1)}\,\psi_{k+2} + \frac{2k^2+2k-1}{(2k+3)(2k-1)}\,\psi_k \\
& + \frac{k(k-1)}{(2k+1)(2k-1)}\,\psi_{k-2} \,.
\end{aligned}
\tag{3.106}
$$

When results (3.103) and (3.106) are combined, the conditions to be satisfied for $k \geq 2$ are

$$
\begin{aligned}
c^2 \frac{(k+2)(k+1)}{(2k+3)(2k+1)}\,\psi_{k+2} &+ \left[k(k+1) + c^2 \frac{2k^2+2k-1}{(2k+3)(2k-1)} \right] \psi_k \\
&+ c^2 \frac{k(k-1)}{(2k+1)(2k-1)}\,\psi_{k-2} = \gamma\,\psi_k \,.
\end{aligned}
\tag{3.107}
$$

Finally, it is clear from the orthogonality of Legendre polynomials and the definition of $\psi(x)$ in (3.102), that the coefficients a_k in expression (3.104) are connected to ψ_k via the formula

$$
\psi_k = \int_{-1}^{1} \psi(x)P_k(x)\,dx =
\begin{cases}
a_k\sqrt{\dfrac{2}{2k+1}}, & 0 \leq k \leq K , \\[2mm]
0 & k > K .
\end{cases}
$$

3.18.1 Determination of $\psi(x)$

It is clear from (3.105) and (3.107) that these equations can be subdivided into separate determination of odd and even coefficients. Specifically the even coefficients satisfy

$$\frac{c^2}{3}\,\psi_0(x) + \frac{2c^2}{3}\,\psi_2 \;=\; \gamma\,\psi_0 \,,$$

$$
\begin{aligned}
\frac{2k(2k-1)\,c^2}{(4k+1)(4k-1)}\,\psi_{2(k-1)} &+ \left[2k(2k+1) + \frac{(8k^2+4k-1)c^2}{(4k+3)(4k-1)} \right] \psi_{2k} \\
&+ \frac{2(k+1)(2k+1)\,c^2}{(4k+3)(4k+1)}\,\psi_{2(k+1)} \;=\; \gamma\,\psi_{2k},
\end{aligned}
\tag{3.108}
$$

and the odd coefficients satisfy

$$\left(2 + \frac{3c^2}{5}\right)\psi_1 + \frac{2c^2}{5}\psi_3 = \gamma\,\psi_1$$

$$\frac{2k(2k+1)c^2}{(4k+3)(4k+1)}\,\psi_{2k-1} + \left[2(2k+1)(k+1) + \frac{(8k^2+12k+3)\,c^2}{(4k+5)(4k+1)}\right]\psi_{2k+1}$$

$$+ \frac{2(2k+3)(k+1)\,c^2}{(4k+5)(4k+3)}\,\psi_{2k+3} = \gamma\,\psi_{2k+1}\,.$$

(3.109)

In conclusion, each set of coefficients is obtained by solving an eigenvalue problem for a tri-diagonal matrix with positive off-diagonal entries. The eigenvalues and eigenvectors are determined by the QR algorithm (see, Appendix A). Both the eigenvalues and eigenvectors are guaranteed to be real-valued, with eigenvectors corresponding to distinct eigenvalues being mutually orthogonal.

3.19 Computation of the eigenvalues λ

It has been shown that $\psi(x)$ satisfies the respective eigenvalue problems

$$\lambda\,\psi(x) = \int_{-1}^{1} \frac{\sin c(y-x)}{\pi(y-x)}\,\psi(y)\,dy\,,$$

$$\mu\,\psi(x) = \int_{-1}^{1} e^{icxy}\,\psi(y)\,dy$$

(3.110)

in which μ may be complex-valued. It follows directly from the second of equations (3.110), that

$$|\mu|^2 \int_{-1}^{1} \psi(x)^2\,dx = \int_{-1}^{1}\left(\int_{-1}^{1}\int_{-1}^{1} e^{icx(y-z)}\,\psi(y)\psi(z)\,dy\,dz\right)dx$$

$$= \int_{-1}^{1}\int_{-1}^{1}\left(\int_{-1}^{1} e^{icx(y-z)}\,dx\right)\psi(y)\psi(z)\,dy\,dz.$$

(3.111)

However, it is a straightforward calculation to confirm that

$$\int_{-1}^{1} e^{icx(y-z)}\,dx = \frac{2\sin c(y-z)}{c(y-z)}\,,$$

and when this result is incorporated into (3.111), the outcome is that

$$|\mu|^2 \int_{-1}^{1} \left[\psi(x)\right]^2\,dx = \frac{2\pi}{c}\int_{-1}^{1}\left(\int_{-1}^{1} \frac{\sin c(y-z)}{\pi(y-z)}\,\psi(y)\,dy\right)\psi(z)\,dz$$

$$= \frac{2\pi}{c}\int_{-1}^{1} \lambda\left[\psi(z)\right]^2\,dz\,.$$

(3.112)

The conclusion of this analysis is that $\lambda = c|\mu|^2/2\pi = T\sigma|\mu|^2/2$. The quality of any particular eigenfunction $\psi(x)$, as measured by its ability to concentrate energy, and is determined by the value of λ, which in turn is computed indirectly from the value of μ. Specifically,

$$\mu = \frac{\int_{-1}^{1}\int_{-1}^{1} e^{icxy}\,\psi(y)\psi(x)\,dx\,dy}{\int_{-1}^{1}[\psi(x)]^2\,dx}. \tag{3.113}$$

It is straightforward to show from equation (3.110), that the eigenfunctions $\psi(x)$ are either even-valued functions of $x \in [-1,1]$, in which case μ is real-valued, or alternatively, are odd-valued functions of $x \in [-1,1]$ in which case the eigenvalue μ is purely complex. The calculation of the numerator of expression (3.113) is simplified by taking note of these two properties of the eigenfunctions to obtain

$$\mu_{\text{real}} = \frac{\int_{-1}^{1}\int_{-1}^{1} \cos(cxy)\,\psi(y)\psi(x)\,dx\,dy}{\int_{-1}^{1}[\psi(x)]^2\,dx}. \tag{3.114}$$

When $\psi(x)$ is an even-valued function of x, and when $\psi(x)$ is an odd-valued function of x, the corresponding result is

$$\mu_{\text{imag}} = i\,\frac{\int_{-1}^{1}\int_{-1}^{1} \sin(cxy)\,\psi(y)\psi(x)\,dx\,dy}{\int_{-1}^{1}[\psi(x)]^2\,dx}. \tag{3.115}$$

In either case, when $\psi(x)$ is specified by expression (3.102), it is easy to show that

$$\begin{aligned} \int_{-1}^{1}[\psi(x)]^2\,dx &= \sum_{j=0}^{K}\sum_{k=0}^{K} a_j a_k \frac{\sqrt{(2j+1)(2k+1)}}{2}\int_{-1}^{1} P_j(x)P_k(x)\,dx \\ &= \sum_{j=0}^{K}\sum_{k=0}^{K} a_j a_k \frac{\sqrt{(2j+1)(2k+1)}}{2}\frac{2\,\delta_{jk}}{j+k+1} = \sum_{k=0}^{K} a_k^2, \end{aligned} \tag{3.116}$$

where all coefficients with odd indices are zero when $\psi(x)$ is an even-valued function of x, i.e. $a_{2k+1} = 0$, and all coefficients with even indices are zero when $\psi(x)$ is an odd-valued function of x, i.e. $a_{2k} = 0$. The numerators of expressions (3.114) and (3.115) are computed by Gauss-Legendre quadrature.

The primary result of Gauss-Legendre quadrature is that

$$\int_{-1}^{1} f(t)\,dt = \sum_{k=1}^{M} w_k\,f(x_k), \tag{3.117}$$

has maximum precision $(2M - 1)$ when the nodes x_k are the zeros of $P_M(x) = 0$ where $P_M(x)$ is the Legendre polynomial of order M. Let the vector Ψ and the symmetric matrices $M \times M$ matrices C and S be respectively defined by

$$\Psi = \left[\begin{array}{cccc} \psi(x_1) & \psi(x_2) & \cdots & \psi(x_M) \end{array}\right]^{\mathrm{T}},$$

$$C = \begin{bmatrix} w_1^2 \cos(cx_1^2) & w_1 w_2 \cos(cx_1 x_2) & \cdots & w_1 w_M \cos(cx_1 x_M) \\ w_2 w_1 \cos(cx_2 x_1) & w_2^2 \cos(cx_2^2) & \cdots & w_2 w_M \cos(cx_2 x_M) \\ \vdots & \vdots & \ddots & \vdots \\ w_M w_1 \cos(cx_M x_1) & w_M w_2 \cos(cx_M x_2) & \cdots & w_M^2 \cos(cx_M^2) \end{bmatrix},$$

$$S = \begin{bmatrix} w_1^2 \sin(cx_1^2) & w_1 w_M \sin(cx_1 x_2) & \cdots & w_1 w_M \sin(cx_1 x_M) \\ w_2 w_1 \sin(cx_2 x_1) & w_2^2 \sin(cx_2^2) & \cdots & w_2 w_M \sin(cx_2 x_M) \\ \vdots & \vdots & \ddots & \vdots \\ w_M w_1 \sin(cx_M x_1) & w_M w_2 \sin(cx_M x_2) & \cdots & w_M^2 \sin(cx_M^2) \end{bmatrix},$$

then it follows immediately from equations (3.114) and (3.115) that the numerators of μ_{real} and μ_{imag} are respectively $\Psi^{\mathrm{T}} C \Psi$ and $i \Psi^{\mathrm{T}} S \Psi$. The computation of μ, and indirectly λ, therefore require two further steps, namely the determination of the nodes and weights of Gauss-Legendre quadratures, and second, the computation of $\psi(x)$ at these nodes. The latter is achieved with high accuracy using Clenshaw's algorithm which is now described.

3.19.1 Summation of Legendre sum by Clenshaw's algorithm

The elegance of Clenshaw's algorithm for the summation of a weighted sum of Legendre polynomials stems from two important properties. First, it is never necessary to evaluate high order Legendre polynomials, thereby avoiding the need to have explicit forms for these polynomials. Second, the subtle cancelations that take place between polynomials of different orders in the computation of the Legendre sum are recognised and incorporated into the algorithm.

Suppose it is required to compute

$$\psi(x) = \sum_{k=0}^{K} b_k P_k(x)$$

for a given value of x. The algorithm begins by introducing the sequence $y_0, \cdots, y_{K+2}$, which is initiated by the specifications $y_{K+1} = y_{K+2} = 0$ with the remaining members of the sequence generated by the backward recursion

$$y_k = \frac{(2k+1)x}{k+1} y_{k+1} - \frac{k+1}{k+2} y_{k+2} + b_k, \qquad k = 0, \cdots, K. \tag{3.118}$$

The expression for $\psi(x)$, when written in terms of the iterates $y_0, \cdots, y_{K+2}$, becomes

$$\psi(x) = \sum_{k=1}^{K} \left(y_k - \frac{(2k+1)x}{k+1} y_{k+1} + \frac{k+1}{k+2} y_{k+2} \right) P_k(x)$$

which may be conveniently re-indexed to obtain

$$\psi(x) = y_0 + \sum_{k=0}^{K-1} y_{k+1} P_{k+1}(x) - \sum_{k=0}^{K} \frac{(2k+1)x}{k+1} y_{k+1} P_k(x) + \sum_{k=1}^{K+1} \frac{k}{k+1} y_{k+1} P_{k-1}(x) \,.$$

The term $y_{K+2} P_{K+2}(x)$ is added to the first summation and the term $\frac{(K+1)}{K+2} y_{K+2} P_K(x)$ is eliminated from the third summation. Both terms are zero-valued but allow the previous expression for $\psi(x)$ to be recast into the format

$$\psi(x) = y_0 + y_1 \left(P_1(x) - x P_0(x) \right) + \sum_{k=1}^{K} y_{k+1} \left(P_{k+1}(x) - \frac{(2k+1)x}{k+1} P_k(x) + \frac{k}{k+1} P_{k-1}(x) \right) \,.$$

The recursive properties of Legendre polynomials ensure that each term in this summation is zero-valued. Moreover, $P_1(x) = x P_0(x)$, so that $\psi(x) = y_0$.

3.19.2 Nodes and weights of the Gauss-Legendre quadrature

The general result underlying numerical integration based on Gaussian quadrature is that if $f(x)$ is continuously differentiable $2N$ times in the interval $[a, b]$, that is $f \in C^{2N}[a, b]$, and $L_0(x), L_1(x), \cdots, L_N(x)$ is a family of orthogonal polynomials associated with the weight function $w(x)$, then

$$\int_a^b w(x) f(x) = \sum_{k=1}^{N} a_k f(x_k) + \frac{f^{2N}(\xi)}{(2N)!} \int_a^b w(x) L_N^2(x) \, dx \,. \tag{3.119}$$

Evidently if $f(x)$ is a polynomial of degree less than or equal to $2N - 1$, then the Gaussian quadrature is exact. The quadrature is commonly stated to have a degree of precision $2N - 1$, which is the maximum possible degree of precision for a quadrature sampling the function value at N nodes. For this reason, Gaussian quadratures are regarded as optimal quadratures. When $w(x) = 1$ and the interval is finite, the associated family of polynomials are the Legendre polynomials and in this case the optimal Gaussian quadrature is

$$\int_{-1}^{1} f(x) \, dx = \sum_{k=1}^{N} w_k f(x_k) \tag{3.120}$$

where $x_1, \cdots, x_N$ are the roots of $P_N(x)$ and $w_1, \cdots, w_N$ are the associated weights.

Suppose that $L_0(x), \cdots$ is a family of monic polynomials defined by the iterative scheme $L_{-1}(x) = 0$, $L_0(x) = 1$ and $L_{k+1}(x) + a_k\, L_k(x) + b_k\, L_{k-1}(x) = x\, L_k(x)$ then it is straightforward to show that $x_1, \cdots, x_N$ are the eigenvalues of the $N \times N$ matrix

$$
M_N = \begin{bmatrix}
a_0 & \sqrt{b_1} & 0 & 0 & \cdots & 0 \\
\sqrt{b_1} & a_1 & \sqrt{b_2} & 0 & \cdots & 0 \\
0 & \sqrt{b_2} & a_2 & \sqrt{b_3} & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \sqrt{b_N} \\
\cdots & \cdots & \cdots & \cdots & \sqrt{b_N} & a_N
\end{bmatrix}.
\tag{3.121}
$$

The argument proceeds as follows. Let $L_{k+1}(x) = (-1)^{k+1}\,|M_k - xI_k|$ and consider the evaluation of $L_{k+1}(x)$ by computing the determinant of the matrix $(M_k - xI_{k+1})$ about its last row. The result of this calculation is the identity

$$
\begin{aligned}
L_{k+1}(x) &= (-1)^{k+1}|M_k - xI_{k+1}| \\
&= (-1)^{k+1}(a_k - x)(-1)^k L_k(x) - (-1)^{k+1}b_k(-1)^{k-1}L_{k-1}(x)
\end{aligned}
$$

Consequently, the eigenvalues of M_k are the roots of the polynomial $L_{k+1}(x)$ which is now seen to satisfy the recursive relation

$$
L_{k+1}(x) + a_k L_k(x) + b_k L_{k-1}(x) = x\, L_k(x)\,.
\tag{3.122}
$$

In the case of Legendre polynomials the coefficients a_n and b_n can be estimated by the following procedure. Given that Legendre polynomials are defined by the iterative scheme $(k+1)P_{k+1}(x) = (2k+1)xP_k(x) - kP_{k-1}(x)$, define

$$
P_k(x) = \frac{(2k)!}{2^k (k!)^2}\, L_k(x),
$$

so that $L_k(x)$ satisfies the iterative scheme

$$
\frac{(k+1)(2k+2)!}{2^{k+1}[(k+1)!]^2}\, L_{k+1}(x) = \frac{(2k+1)(2k)!}{2^k (k!)^2}\, x\, L_k(x) - \frac{k(2k-2)!}{2^{k-1}[(k-1)!]^2}L_{k-1}(x)\,.
$$

This identity is multiplied by $2^k (k!)^2$ and divided by $(2k+1)!$ to obtain the final form

$$
L_{k+1}(x) + \frac{k^2}{4k^2 - 1}L_{k-1}(x) = x\, L_k(x),
\tag{3.123}
$$

which provides explicit expressions for a_k and b_k, thereby allowing the matrix M_N to be constructed and its eigenvalues calculated using the QR algorithm.

3.19.3   Conclusion

Regarding the highly important role of the Fourier transform in facilitating the analysis of the interacting processes in the spectral domain, rather than the original time domain, we have defined, in this chapter, the conventional Fourier transform, and the concept of the finite Fourier transform. Also, we have given the algorithm for the fast Fourier transform, which is used extensively to reduce the number of computations of the discrete Fourier coefficients. In addition, we have presented the definition of power and the cross-spectrum and how they relate to the time domain quantities; autocorrelation and cross-correlation. The three common used non-parametric spectral estimators have been debated such as, the method of averaging across periodograms of the adjacent segments of the single record, the method of frequency averaging and the multi-taper method. We have demonstrated that the frequency averaging method reduces the bias error, when the number of frequencies m lying in the neighbourhood of a specific frequency, is consistent. However, it has small variance improvement. Averaging across periodograms of the adjacent segments of the single record is regarded as an unbiased spectral estimator with acceptable variance when the number of available observations is sufficiently large. The multi-taper method, which based on multiplying the whole data with a number of orthogonal tapers, is characterized as an unbiased spectral estimator with highly reducing variability, especially for the relatively short data. Since each taper plays a crucial role in reducing the sidelobe spectral leakage and provides a good eigenspectrum, then the average of the K eigenspectra increase the resistance against spectral leakage to produce an approximately leakage-free estimator for the spectra. Moreover, the average of the K different eigenspectra reduces the variance of the spectral estimator.

Chapter 4

Measure of association among processes

4.1 Introduction

Interactions among large numbers of stochastic processes are commonplace and arise naturally when a number of different input or control processes combine together to produce various output processes. Examples of this phenomenon can be found in Neurophysiology, in Economics and in many biological systems. In the frequency domain, coherence function is the natural measure of the connectivity between various couplings of processes.

4.2 Coherency and coherence functions

Definition 4.2.1 (Coherency function) The coherency function between two stationary stochastic processes, namely $X = \{X_t; t = 0, \pm 1, \cdots\}$ and $Y = \{Y_t; t = 0, \pm 1, \cdots\}$, at a specific frequency f is denoted by $R_{XY}(f)$ and defined in terms of power and cross-spectral densities of these processes at that specific frequency by

$$R_{XY}(f) = \frac{S_{XY}(f)}{[S_{XX}(f)S_{YY}(f)]^{1/2}}, \qquad -\infty < f < \infty. \qquad (4.1)$$

In this definition $S_{XY}(f)$ is a complex valued-function of frequency f and therefore the coherency itself is a complex valued-function of frequency f, whereas the spectral quantities $S_{XX}(f)$ and $S_{YY}(f)$ represent the power spectrum of the processes X and Y respectively and are real-valued functions of frequency.

Definition 4.2.2 (Coherence function) The coherence between two stationary stochastic process $X = \{X_t\}$ and $Y = \{Y_t\}$, where $t = 0, \pm 1, \cdots$, at a specific frequency f is defined as the squared-magnitude of the coherency between these processes, that is,

$$|R_{XY}(f)|^2 = \frac{|S_{XY}(f)|^2}{S_{XX}(f)S_{YY}(f)}. \qquad (4.2)$$

The coherence is therefore a real valued-function of frequency f and gives a measure of the linear correlation between two processes, say X and Y, at frequency f. In particular, the Cauchy-Schwarz inequality $|\mathbf{a} \cdot \mathbf{b}|^2 \le |\mathbf{a}|^2|\mathbf{b}|^2$ indicates that coherence must lie in $[0, 1]$. If the coherence between two processes has value zero at frequency f then these processes are uncorrelated at that frequency, whereas coherence with value one means that the processes are perfectly correlated at that frequency.

4.2.1 Asymptotical Coherence

By definition, the measure of linear relationship between two stationary stochastic processes is given by the squared-magnitude of the coherency function in equation (4.2). Since coherence of two stationary processes is a normalisation of the cross-spectra to the power spectra and since the exact spectral densities, especially in the real world, are not known we must estimate the coherence through estimating the spectral densities. A good estimation of the spectral densities leads to an accurate estimate for the coherence. In Chapter 3, we discussed the most common methods that are usually used to estimate auto- and cross-spectral densities. Thus the estimated coherence, which is denoted by $|\hat{R}_{XY}(f)|^2$, will take the form,

$$|\hat{R}_{XY}(f)|^2 = \frac{|\hat{S}_{XY}(f)|^2}{\hat{S}_{XX}(f)\hat{S}_{YY}(f)}, \qquad (4.3)$$

where $\hat{S}_{XY}(f)$ represents the estimated cross-spectral densities at frequency f between the two stochastic processes X and Y with estimated auto-spectral densities $\hat{S}_{XX}(f)$ and $\hat{S}_{YY}(f)$. This approximation is asymptotically unbiased. In order to facilitate the investigation of the statistical properties of the estimated coherence, we apply Fisher's transform to the magnitude of the estimated coherency, $|\hat{R}_{XY}(f)|$, thus

$$Z = \tanh^{-1}(|\hat{R}_{XY}(f)|) = \frac{1}{2}\log\left(\frac{1 + |\hat{R}_{XY}(f)|}{1 - |\hat{R}_{XY}(f)|}\right). \qquad (4.4)$$

The random variable Z computed from the formula (4.4) may be shown to be asymptotically normally distributed.

Goodman (Goodman [1965], Brillinger [1981]) derived the probability density function (pdf) of $|\hat{R}_{XY}(f)|^2$ as

$$\frac{\Gamma(n)}{\Gamma(n-r)\Gamma(r)}(1-|R_{XY}(f)|^2)^n \, _2F_1(n,n;r;|R_{XY}(f)|^2|\hat{R}_{XY}(f)|^2)\times$$
$$|\hat{R}_{XY}(f)|^{2r-2}(1-|\hat{R}_{XY}(f)|^2)^{n-r-1} \tag{4.5}$$

where $_2F_1(n,n;r;|R_{XY}(f)|^2|\hat{R}_{XY}(f)|^2)$, is a *hypergeometric function*, n represents the number of independent samples that effectively contribute to the estimate of the coherence. This means, n may indicate the number of non-overlapping sections into which each record is divided when "averaging across periodograms of contiguous sections from single records " is applied to estimate spectral densities, as detailed in section 3.10.2. Also n can be considered as the number of independent frequency points in a relatively small band of frequency centred on a specific frequency, namely f, when the procedure of "smoothed periodogram for the entire single record ", discussed in section 3.10.1, is used to estimate spectra and cross-spectra. The parameter r is the dimension of the stochastic processes of interest, *i.e.*, X. In the case of independent processes, that is $|R_{XY}(f)|^2 = 0$, expression (4.5) becomes

$$\frac{\Gamma(n)}{\Gamma(n-r)\Gamma(r)}|\hat{R}_{XY}(f)|^{2r-2}(1-|\hat{R}_{XY}(f)|^2)^{n-r-1}, \tag{4.6}$$

which is asymptotically normally distributed with mean value (Enochson and Goodman [1965], Brillinger [1981])

$$\mathbb{E}\left[\tanh^{-1}|\hat{R}_{XY}(f)|\right] = \tanh^{-1}|\hat{R}_{XY}(f)| + \frac{r}{2(n-r)} \tag{4.7}$$

and variance

$$\mathrm{Var}\left[\tanh^{-1}|\hat{R}_{XY}(f)|\right] = \frac{1}{2(n-r-1)}. \tag{4.8}$$

4.2.2 Confidence interval for the asymptotic coherence

To decide whether the asymptotic coherence between two stochastic processes, namely X and Y, is significant or not, we will examine the null hypothesis that the two processes of interest are independent at level of significance $\alpha\%$, that is $|R_{XY}(f)|^2 = 0$. The probability distribution function of the asymptotic coherence whose density distribution function given by (4.6) when $r = 1$ is given by (Brillinger [1981])

$$\mathrm{Prob}\left[|\hat{R}_{XY}(f)|^2 < x\right] = 1 - (1-x)^{n-1} \quad 0 < x < 1. \tag{4.9}$$

and

$$\text{Prob}\left[|\hat{R}_{XY}(f)|^2 < x\right] = \alpha. \tag{4.10}$$

The combination of the two expressions (4.9) and (4.10) yields the solution

$$1 - (1-x)^{n-1} = \alpha,$$
$$\Rightarrow 1 - \alpha = (1-x)^{n-1},$$
$$\Rightarrow (1-\alpha)^{(1/n-1)} = 1 - x,$$

therefore

$$x = 1 - (1-\alpha)^{(1/n-1)}, \tag{4.11}$$

where n denotes the disjoint segments L that the data has been divided into, when the method of averaging periodograms across contiguous sections of single records, is used to estimate the spectra and cross-spectra. Also this confidence interval can be applied to the estimated coherence, when its components are estimated by the smoothed periodogram method, in this procedure, the average of periodogram ordinates is taken with respect to the number of frequencies lying in the neighbourhood of a specific frequency. That is n is the number of independent frequencies $2m + 1$ in a small band centered on that specific frequency.

The estimated coherence based on spectral densities, which are obtained from the multi-taper method, possesses the same features as the estimated coherence, which is produced by either the "disjoint sections method" and the "smoothed periodogram method". Consequently, it is distributed according to equation (4.6). Since the spectral densities are estimated by applying K independent tapers, then the confidence interval of the significant asymptotic coherence of X with Y under the null hypothesis of independence at level $\alpha\%$ is given by, (see, Thompson ([1982], [2007]))

$$x = 1 - (1-\alpha)^{(1/K-1)}. \tag{4.12}$$

4.3 Partial spectra and partial coherence

4.3.1 Derived spectra

An important feature of the subsequent analysis is the use of processes that are derived from other processes. Such processes are not measurable and are referred to here as *derived processes*. A particularly important class of derived process occurs when an observable process, say X, is conditioned on a group of other observable processes, say $\boldsymbol{Z}$, such that

the derived process of X is independent of each process of Z. For example, suppose $\mathbf{Z}$ denotes the family of n processes $Z_1, \ldots, Z_n$ with respective Fourier coefficients $\mathbf{d}_Z = (d_{Z_1}, \ldots, d_{Z_n})$ where it is assumed *a priori* that the family $\mathbf{Z}$ have no common processes, and let the process X have Fourier coefficients d_X. The derived process of X conditioned on $\mathbf{Z}$ is the process with Fourier coefficients

$$d_{X|Z} = d_X - A_Z \mathbf{d}_Z, \tag{4.13}$$

where the n dimensional row vector A_Z is chosen to ensure that

$$\mathbb{E}\left[d_{X|Z} \overline{\mathbf{d}_Z}\right] = 0. \tag{4.14}$$

In this equation superscript $\overline{\mathbf{d}_Z}$ denotes the complex conjugate of $\mathbf{d}_Z$. Thus A_Z is required to satisfy the equation

$$\mathbb{E}\left[d_X \overline{\mathbf{d}_Z}\right] - A_Z \mathbb{E}\left[\mathbf{d}_Z \overline{\mathbf{d}_Z}\right] = 0. \tag{4.15}$$

4.3.2   Partial spectra

Let X, Y and Z be three stationary stochastic processes, and let $d_{X|Z}(f)$ and $d_{Y|Z}(f)$ be the Fourier coefficients of the derived processes X and Y with respect to the process Z. The partial cross-spectral density between the processes X and Y given Z at a frequency f is denoted by $S_{XY|Z}(f)$ and defined as, (see, Brillinger [1981], Bendat and Piersol [1986], Rosenberg *et al.* [1989])

$$
\begin{aligned}
S_{XY|Z}(f) &= \mathbb{E}\left[d_{X|Z}(f)\overline{d_{Y|Z}(f)}\right], \\
&= \mathbb{E}\left[\left(d_X(f) - S_{XZ}\,S_{ZZ}^{-1}(f)d_Z(f)\right)\overline{\left(d_Y(f) - S_{YZ}\,S_{ZZ}^{-1}(f)d_Z(f)\right)}\right], \\
&= \mathbb{E}\left[\left(d_X(f) - S_{XZ}\,S_{ZZ}^{-1}(f)d_Z(f)\right)\left(\overline{d_Y(f)} - \overline{d_Z(f)}\,S_{ZZ}^{-1}(f)S_{ZY}\right)\right], \\
&= S_{XY}(f) - S_{XZ}(f)S_{ZZ}^{-1}(f)S_{ZY}(f); \qquad -\infty < f < \infty,
\end{aligned}
\tag{4.16}
$$

where the spectral functions $S_{XY}(f)$, $S_{XZ}(f)$ and $S_{ZY}(f)$ represent the cross-spectral densities at frequency f between the processes (X, Y), (X, Z) and (Z, Y) respectively. The spectral quantity $S_{ZZ}^{-1}(f)$ denotes the inverse auto-spectral density matrix of the process Z at frequency f. In the same way, the partial auto-spectral density of the process X conditioned on the process Z is defined by $S_{XX|Z}(f) = \mathbb{E}\left[d_{X|Z}(f)\overline{d_{X|Z}(f)}\right]$. Thus, by a

simple calculation, we can obtain

$$S_{XX|Z}(f) \;=\; \mathbb{E}\Big[d_{X|Z}(f)\,\overline{d_{X|Z}(f)}\Big] ,$$

$$=\; \mathbb{E}\Big[\big(d_X(f) - S_{XZ}\,S_{ZZ}^{-1}(f)d_Z(f)\big)\big(\overline{d_X(f)} - \overline{d_Z(f)}\,S_{ZZ}^{-1}(f)S_{ZX}\big)\Big], \quad (4.17)$$

$$=\; S_{XX}(f) - S_{XZ}(f)S_{ZZ}^{-1}(f)S_{ZX}(f) .$$

Similarly, the partial auto-spectrum of the process Y conditioned on the process Z will be defined as

$$S_{YY|Z}(f) = S_{YY}(f) - S_{YZ}(f)S_{ZZ}^{-1}(f)S_{ZY}(f) . \qquad (4.18)$$

The definition of the partial spectra follows by the definition of the partial coherency between two stationary stochastic processes given a third process as follows.

4.3.3 Partial coherency

The coherency function between two stationary stochastic processes, namely X and Y conditioned on the stationary stochastic process Z, is denoted by $R_{XY|Z}(f)$ and defined as (see, Brillinger [1981], Bendat and Piersol [1986])

$$R_{XY|Z}(f) \;=\; \frac{S_{XY|Z}(f)}{[S_{XX|Z}(f)S_{YY|Z}(f)]^{1/2}} ; \quad -\infty < f < \infty . \qquad (4.19)$$

This type of coherency is called the partial coherency of the process X with the process Y after removing the linear influence of the process Z at a specific frequency f. The components of the partial coherency are calculated from the partial cross-spectrum between the processes X and Y conditioned on the process Z, namely $S_{XY|Z}(f)$, and the partial auto-spectral densities of the processes X, and, Y conditioned on the process Z, namely $S_{XX|Z}(f)$ and $S_{YY|Z}(f)$ respectively. Consequently, the partial coherence of the process X with the process Y after eliminating the linear effect of the process Z, can be defined as a squared magnitude of the partial coherency.

Definition 4.3.1 (Partial coherence) The coherence function between two stationary stochastic processes, namely X and Y conditioned on the stationary stochastic process Z, is denoted by $|R_{XY|Z}(f)|^2$ and defined as (see, Brillinger [1981], Bendat and Piersol [1986])

$$|R_{XY|Z}(f)|^2 \;=\; \frac{|S_{XY|Z}(f)|^2}{S_{XX|Z}(f)S_{YY|Z}(f)} ; \quad -\infty < f < \infty . \qquad (4.20)$$

The partial coherence function, $|R_{XY|Z}(f)|^2$, measures the strength of the correlation between two processes X and Y at a specific frequency f after removing the linear influence of the process Z. The values of the partial coherence lie within the interval $[0, 1]$. The zero value of the partial coherence, *i.e.* $|R_{XY|Z}(f)|^2 = 0$, idicates that the correlation between two processes X and Y is vanished after removing the linear effect of Z. That is the correlation between X and Y can be predicted by Z.

4.3.4 Combining Partial Spectra

Let $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ denote respectively families of n processes $\phi_1, \ldots, \phi_n$ and m processes $\psi_1, \ldots, \psi_m$ with respective Fourier coefficients $\mathbf{d}_\phi = (d_{\phi_1}, \ldots, d_{\phi_n})$ and $\mathbf{d}_\psi = (d_{\psi_1}, \ldots, d_{\psi_m})$ where it is assumed *a priori* that the families $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ share no common processes. Let $\boldsymbol{\xi} = \boldsymbol{\phi} \cup \boldsymbol{\psi}$, that is, the family $\mathbf{Z}$ consists of $(n+m)$ processes. The spectral density matrix of $\mathbf{Z}$ may be expressed in the block form

$$S_{ZZ} = \begin{bmatrix} S_{\phi\phi} & S_{\phi\psi} \\ S_{\psi\phi} & S_{\psi\psi} \end{bmatrix} . \tag{4.21}$$

Moreover, the inverse of the spectral matrix S_{ZZ} can be obtained straightforwardly as

$$S_{ZZ}^{-1} = \frac{1}{\det S_{ZZ}} \begin{bmatrix} S_{\psi\psi} & -S_{\phi\psi} \\ -S_{\psi\phi} & S_{\phi\phi} \end{bmatrix} . \tag{4.22}$$

where the determinant of the spectral matrix S_{ZZ} is calculated by

$$\begin{aligned} \det S_{ZZ} &= S_{\phi\phi} S_{\psi\psi} - S_{\phi\psi} S_{\psi\phi} \\ &= S_{\phi\phi}\big(S_{\psi\psi} - S_{\psi\phi} S_{\phi\phi}^{-1} S_{\psi\phi}\big) \tag{4.23} \\ &= S_{\psi\psi}\big(S_{\phi\phi} - S_{\phi\psi} S_{\psi\psi}^{-1} S_{\psi\phi}\big) . \end{aligned}$$

The partial auto-spectrum of the process ψ with respect to the process ϕ is given by $S_{\psi\psi|\phi} = S_{\psi\psi} - S_{\phi\psi} S_{\phi\phi}^{-1} S_{\psi\phi}$, and the auto-spectrum of the process ϕ conditioned on the process ψ is $S_{\phi\phi} - S_{\phi\psi} S_{\psi\psi}^{-1} S_{\psi\phi}$, and so equation (4.23) will take the form

$$\det S_{ZZ} = S_{\phi\phi} S_{\psi\psi|\phi} = S_{\psi\psi} S_{\phi\phi|\psi} . \tag{4.24}$$

Therefore, the inverse of the spectral matrix S_{ZZ} is given by

$$S_{ZZ}^{-1} = \begin{bmatrix} S_{\phi\phi|\psi}^{-1} & -S_{\phi\phi}^{-1} S_{\phi\psi} S_{\psi\psi|\phi}^{-1} \\ -S_{\psi\psi}^{-1} S_{\psi\phi} S_{\phi\phi|\psi}^{-1} & S_{\psi\psi|\phi}^{-1} \end{bmatrix} . \tag{4.25}$$

Although it is not apparent from expression (4.25), the inverse of the spectral density matrix is symmetric. To appreciate how the symmetry comes about, note that the identity

$$\left(S_{\psi\psi} - S_{\psi\phi}S_{\phi\phi}^{-1}S_{\phi\psi}\right)S_{\psi\psi}^{-1}S_{\psi\phi} = S_{\psi\phi} - S_{\psi\phi}S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi}^{-1}S_{\psi\phi} = S_{\psi\phi}S_{\phi\phi}^{-1}\left(S_{\phi\phi} - S_{\phi\psi}S_{\psi\psi}^{-1}S_{\psi\phi}\right)$$

can be rewritten in the form $S_{\psi\psi|\phi}S_{\psi\psi}^{-1}S_{\psi\phi} = S_{\psi\phi}S_{\phi\phi}^{-1}S_{\phi\phi|\psi}$, or equivalently,

$$S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1} = S_{\psi\psi|\phi}^{-1}S_{\psi\phi}S_{\phi\phi}^{-1} = \left(\overline{S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}}\right).$$

There are two different representations of the partial spectrum $S_{XY|\phi\psi}$, depending on the choice of family by which to perform the initial conditioning. However, to demonstrate the independence of the initial choice of conditioning family of processes, we begin the analysis with the calculation of $S_{XY|Z} = S_{XY|\phi\psi}$ from expression (4.16), to get

$$S_{XY|Z} = S_{XY|\phi\psi} = S_{XY} - \begin{bmatrix} S_{X\phi} & S_{X\psi} \end{bmatrix} \begin{bmatrix} S_{\phi\phi|\psi}^{-1} & -S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1} \\ -S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1} & S_{\psi\psi|\phi}^{-1} \end{bmatrix} \begin{bmatrix} S_{\phi Y} \\ S_{\psi Y} \end{bmatrix}.$$

$$(4.26)$$

It is a matter of straightforward matrix multiplication to demonstrate that

$$\begin{aligned} S_{XY|\phi\psi} &= S_{XY} - S_{X\phi}S_{\phi\phi|\psi}^{-1}S_{\phi Y} + S_{X\phi}S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}S_{\psi Y} \\ &\quad - S_{X\psi}S_{\psi\psi|\phi}^{-1}S_{\psi Y} + S_{X\psi}S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1}S_{\phi Y}. \end{aligned}$$

$$(4.27)$$

Case I - Initial conditioning taken with respect to ϕ The analysis of this case begins by replacing S_{XY} in expression (4.27) by $S_{XY|\phi} + S_{X\phi}S_{\phi\phi}^{-1}S_{\phi Y}$ to obtain

$$\begin{aligned} S_{XY|\phi\psi} &= S_{XY|\phi} + S_{X\phi}S_{\phi\phi}^{-1}\left(S_{\phi\phi|\psi} - S_{\phi\phi}\right)S_{\phi\phi|\psi}^{-1}S_{\phi Y} + S_{X\phi}\left(S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}\right)S_{\psi Y} \\ &\quad - S_{X\psi}S_{\psi\psi|\phi}^{-1}S_{\psi Y} + S_{X\psi}\left(S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1}\right)S_{\phi Y}. \end{aligned}$$

$$(4.28)$$

The difference $\left(S_{\phi\phi|\psi} - S_{\phi\phi}\right)$ in equation (4.28) is now replaced by $-S_{\phi\psi}S_{\psi\psi}^{-1}S_{\psi\phi}$, to obtain

$$\begin{aligned} S_{XY|\phi\psi} &= S_{XY|\phi} - S_{X\phi}S_{\phi\phi}^{-1}S_{\phi\psi}\left(S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1}\right)S_{\phi Y} + S_{X\phi}\left(S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}\right)S_{\psi Y} \\ &\quad - S_{X\psi}S_{\psi\psi|\phi}^{-1}S_{\psi Y} + S_{X\psi}\left(S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1}\right)S_{\phi Y} \\ &= S_{XY|\phi} + \left(S_{X\psi} - S_{X\phi}S_{\phi\phi}^{-1}S_{\phi\psi}\right)\left(S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1}\right)S_{\phi Y} \\ &\quad - \left(S_{X\psi} - S_{X\phi}S_{\phi\phi}^{-1}S_{\phi\psi}\right)S_{\psi\psi|\phi}^{-1}S_{\psi Y}. \end{aligned}$$

$$(4.29)$$

The combination $\left(S_{X\psi} - S_{X\phi}S_{\phi\phi}^{-1}S_{\phi\psi}\right)$ is now replaced in equation (4.29) by its definition $S_{X\psi|\phi}$ and the identity $S_{\psi\psi|\phi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1} = S_{\psi\psi|\phi}^{-1}S_{\psi\phi}S_{\phi\phi}^{-1}$ is used, to obtain

$$
\begin{aligned}
S_{XY|\phi\psi} &= S_{XY|\phi} + S_{X\psi|\phi}\left(S_{\psi\psi|\phi}^{-1}S_{\psi\phi}S_{\phi\phi}^{-1}\right)S_{\phi Y} - S_{X\psi|\phi}S_{\psi\psi|\phi}^{-1}f_{\psi Y} \\
&= S_{XY|\phi} - S_{X\psi|\phi}S_{\psi\psi|\phi}^{-1}\left(S_{\psi Y} - S_{\psi\phi}S_{\phi\phi}^{-1}S_{\phi Y}\right) \qquad (4.30)\\
&= S_{XY|\phi} - S_{X\psi|\phi}S_{\psi\psi|\phi}^{-1}S_{\psi Y|\phi}\,.
\end{aligned}
$$

Thus $S_{XY|\phi\psi} = S_{XY|\phi} - S_{X\psi|\phi}S_{\psi\psi|\phi}^{-1}S_{\psi Y|\phi}$ in the situation in which the initial conditioning is taken with respect to the family of processes ϕ.

Case II - Initial conditioning taken with respect to ψ Suppose now that the initial conditioning is taken with respect to the family of processes ψ. In this case the analysis begins by replacing S_{XY} in expression (4.24) by $S_{XY|\psi} + S_{X\psi}S_{\psi\psi}^{-1}S_{\psi Y}$ to obtain

$$
\begin{aligned}
S_{XY|\phi\psi} &= S_{XY|\psi} + S_{X\psi}S_{\psi\psi}^{-1}\left(S_{\psi\psi|\phi} - S_{\psi\psi}\right)S_{\psi\psi|\phi}^{-1}S_{\psi Y} + S_{X\phi}\left(S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}\right)S_{\psi Y} \\
&\quad - S_{X\phi}S_{\phi\phi|\psi}^{-1}S_{\phi Y} + S_{X\psi}\left(S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1}\right)S_{\phi Y}\,.
\end{aligned}
$$
$$(4.31)$$

The difference $\left(S_{\psi\psi|\phi} - S_{\psi\psi}\right)$ in equation (4.31) is now replaced by $-S_{\psi\phi}S_{\phi\phi}^{-1}S_{\phi\psi}$, to obtain

$$
\begin{aligned}
S_{XY|\phi\psi} &= S_{XY|\psi} - S_{X\psi}S_{\psi\psi}^{-1}S_{\psi\phi}\left(S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}\right)S_{\psi Y} + S_{X\phi}\left(S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}\right)S_{\psi Y} \\
&\quad - S_{X\phi}S_{\phi\phi|\psi}^{-1}S_{\phi Y} + S_{X\psi}\left(S_{\psi\psi}^{-1}S_{\psi\phi}S_{\phi\phi|\psi}^{-1}\right)S_{\phi Y} \\
&= S_{XY|\psi} + \left(S_{X\phi} - S_{X\psi}S_{\psi\psi}^{-1}S_{\psi\phi}\right)\left(S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}\right)S_{\psi Y} \\
&\quad - \left(S_{X\phi} - S_{X\psi}S_{\psi\psi}^{-1}S_{\psi\phi}\right)S_{\phi\phi|\psi}^{-1}S_{\phi Y}\,.
\end{aligned}
$$
$$(4.32)$$

The combination $\left(S_{X\phi} - S_{X\psi}S_{\psi\psi}^{-1}S_{\psi\phi}\right)$ is now replaced in equation (4.32) by its definition $S_{X\phi|\psi}$ and the identity $S_{\phi\phi|\psi}^{-1}S_{\phi\psi}S_{\psi\psi}^{-1} = S_{\phi\phi}^{-1}S_{\phi\psi}S_{\psi\psi|\phi}^{-1}$ is used, to obtain

$$
\begin{aligned}
S_{XY|\phi\psi} &= S_{XY|\psi} + S_{X\phi|\psi}\left(S_{\phi\phi|\psi}^{-1}S_{\phi\psi}S_{\psi\psi}^{-1}\right)S_{\psi Y} - S_{X\phi|\psi}S_{\phi\phi|\psi}^{-1}S_{\phi Y} \\
&= S_{XY|\psi} + S_{X\phi|\psi}S_{\phi\phi|\psi}^{-1}\left(S_{\phi\psi}S_{\psi\psi}^{-1}S_{\psi Y} - S_{\phi Y}\right) \qquad (4.33)\\
&= S_{XY|\psi} - S_{X\phi|\psi}S_{\phi\phi|\psi}^{-1}S_{\phi Y|\psi}\,.
\end{aligned}
$$

Thus $S_{XY|\phi\psi} = S_{XY|\psi} - S_{X\phi|\psi}S_{\phi\phi|\psi}^{-1}S_{\phi Y|\psi}$ in the situation in which the initial conditioning is taken with respect to the family of processes ψ.

To summarise, the partial cross-spectrum $S_{XY|\phi\psi}$ may be calculated by either conditioning first on the processes of ϕ or on the processes of ψ. Both approaches, although superficially different, lead to equivalent expressions for $S_{XY|\phi\psi}$, namely

$$S_{XY|\phi\psi} = S_{XY|\psi} - S_{X\phi|\psi}S_{\phi\phi|\psi}^{-1}S_{\phi Y|\psi} = S_{XY|\phi} - S_{X\psi|\phi}S_{\psi\psi|\phi}^{-1}S_{\psi Y|\phi}. \tag{4.34}$$

4.3.5 Estimating the partial spectra and partial coherence directly from the data

Consider the case of $(n+2)$ processes divided into the family of n processes Z and the single processes X and Y in that order. These $(n+2)$ processes give rise to the $(n+2) \times (n+2)$ dimensional spectral matrix with block form

$$S = \begin{bmatrix} S_{ZZ} & S_{ZX} & S_{ZY} \\ S_{XZ} & S_{XX} & S_{XY} \\ S_{YZ} & S_{YX} & S_{YY} \end{bmatrix} \tag{4.35}$$

The recursive application of formula (4.16) provides a mechanism for the computation of $S_{XY|Z}$ starting with the ordinary spectral densities represented by the elements of the $(n+2) \times (n+2)$ dimensional spectral matrix set out in expression (4.35). The ordinary spectra can be estimated by using one of the methods elaborated in Chapter 3. Moreover, ordinary and partial coherences can be computed directly from these ordinary and partial spectral densities. However, partial coherences at any order can be constructed directly without the need to construct all partial coherences or partial spectra at lower orders.

It is a standard result from matrix algebra that S^{-1}, the inverse of the matrix S set out in block form in expression (4.35), is

$$S^{-1} = \frac{1}{\det S} \operatorname{adj} S, \tag{4.36}$$

where $\operatorname{adj} S$, commonly called the adjugate matrix of S, is the transposed matrix of cofactors of S. The cofactor of S, corresponding to the (j, k)-th element of S, has the value $(-1)^{j+k}M_{jk}$, where M_{jk} is the determinant of the matrix (or minor) formed from S by striking out the j-th row and k-th column of S. Suppose now that it is desired to compute the partial coherence

$$\left|R_{XY|Z}\right|^2 = \frac{\left|S_{XY|Z}\right|^2}{S_{XX|Z}S_{YY|Z}}. \tag{4.37}$$

The key idea is to calculate the entry of S^{-1} in the slot currently occupied by S_{XY}. Bearing in mind that the adjugate matrix is the transposed matrix of cofactors, then the required

entry will be a signed multiple of the determinant of the matrix formed by striking out the row and column of S containing the entry S_{YX}, that is, the determinant,

$$
\begin{vmatrix} S_{ZZ} & S_{ZY} \\ S_{XZ} & S_{XY} \end{vmatrix}.
\tag{4.38}
$$

Because S_{ZZ} is a nonsingular $n \times n$ matrix, then it is straightforward matrix algebra to establish the identity

$$
\begin{bmatrix} I & 0 \\ -S_{XZ}S_{ZZ}^{-1} & 1 \end{bmatrix} \begin{bmatrix} S_{ZZ} & S_{ZY} \\ S_{XZ} & S_{XY} \end{bmatrix} \begin{bmatrix} I & -S_{ZZ}^{-1}S_{ZY} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} S_{ZZ} & 0 \\ 0 & S_{XY} - S_{XZ}S_{ZZ}^{-1}S_{ZY} \end{bmatrix}
\tag{4.39}
$$

where I is the $n \times n$ identity matrix. Taking determinants of the matrix identity immediately gives

$$
\begin{vmatrix} S_{ZZ} & S_{ZY} \\ S_{XZ} & S_{XY} \end{vmatrix} = \left(S_{XY} - S_{XZ}S_{ZZ}^{-1}S_{ZY} \right) \left| S_{ZZ} \right| = S_{XY|Z} \left| S_{ZZ} \right|.
\tag{4.40}
$$

Consider now the determinant of the matrix formed from S by striking out the row and column of S containing the entry S_{XX}, that is, the determinant

$$
\begin{vmatrix} S_{ZZ} & S_{ZY} \\ S_{YZ} & S_{YY} \end{vmatrix}.
\tag{4.41}
$$

In this case it is straightforward matrix algebra to establish the identity

$$
\begin{bmatrix} I & 0 \\ -S_{YZ}S_{ZZ}^{-1} & 1 \end{bmatrix} \begin{bmatrix} S_{ZZ} & S_{ZY} \\ S_{YZ} & S_{YY} \end{bmatrix} \begin{bmatrix} I & -S_{ZZ}^{-1}S_{ZY} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} S_{ZZ} & 0 \\ 0 & S_{YY} - S_{YZ}S_{ZZ}^{-1}S_{ZY} \end{bmatrix}
\tag{4.42}
$$

from which it immediately follows that

$$
\begin{vmatrix} S_{ZZ} & S_{ZY} \\ S_{YZ} & S_{YY} \end{vmatrix} = \left(S_{YY} - S_{YZ}S_{ZZ}^{-1}S_{ZY} \right) \left| S_{ZZ} \right| = S_{YY|Z} \left| S_{ZZ} \right|.
\tag{4.43}
$$

An equivalent calculation, in which the row and column of S containing S_{YY} is eliminated, yields

$$
\begin{vmatrix} S_{ZZ} & S_{ZX} \\ S_{XZ} & S_{XX} \end{vmatrix} = \left(S_{XX} - S_{XZ}S_{ZZ}^{-1}S_{ZX} \right) \left| S_{ZZ} \right| = S_{XX|Z} \left| S_{ZZ} \right|.
\tag{4.44}
$$

The immediate conclusion of this analysis is that the partial coherence (4.37) can be computed directly from the entries of S^{-1}, which is the inverse of the spectral density matrix

in (4.35). Of course, in general, the procedure computes the partial coherence between any pair of processes conditioned on all the other processes. However, lower order coherences may be computed for the same set of processes by striking out appropriate rows and columns (see, Medkour *et al.* [2009]).

In the following section, we discuss the parametric spectral estimation, represented by the multivariate autoregressive model of order p.

4.4   Autoregressive Model of order p AR(p)

Supposing that X is an M real vector valued stochastic process whose values are available at $t = 0, \pm 1, \pm 2, \cdots$, then the value of the process X at a specific time t can be expressed in terms of its past values using the notion of an autoregressive model of order p as

$$X_t = \sum_{j=1}^{p} A_j X_{t-j} + E_t , \tag{4.45}$$

where $X_t = [X_{1,t}, X_{2,t}, \cdots, X_{M,t}]^\tau$, the superscript (τ) denotes the vector transpose, $A_1, \cdots, A_p$ are $M \times M$ matrices of coefficients and $E_t = [E_{1,t}, E_{2,t}, \cdots, E_{M,t}]^\tau$ is a residual error vector of dimension M with $M \times M$ covariance matrix Σ. In particular, error vectors at different times are assumed to be independent random variables with mean value zero. In the following two sections we shall describe two different ways in which the values of the matrices $A_1, \cdots, A_p$ may be estimated from experimental observations.

4.4.1   Yule-Walker approach

The Yule-Walker approach estimates the coefficients matrices $A_1, \cdots, A_p$, by multiplying equation (4.45) by historical values of X_t, and then taking expectations of the resulting equations. This procedure is illustrated with reference to multiplication of equation (4.45) by X_{t-1}^τ. In this case,

$$X_t X_{t-1}^\tau = \sum_{j=1}^{p} A_j X_{t-j} X_{t-1}^\tau + E_t X_{t-1}^\tau ,$$

and the expectation of this equation gives

$$\langle X_t, X_{t-1} \rangle = \sum_{j=1}^{p} A_j \langle X_{t-j}, X_{t-1} \rangle + \langle E_t, X_{t-1} \rangle , \tag{4.46}$$

where $\langle\, X, Y\, \rangle$ denotes the expected value of the outer product of the vectors X and Y, that is, it is the expectation of the matrix XY^τ. The assumption that residual errors at different times are independent processes indicates that $\langle\, E_t, X_{t-1}\, \rangle = 0$, which in turn leads to the first Yule-Walker equation,

$$\langle\, X_t, X_{t-1}\, \rangle = \sum_{j=1}^{p} A_j \langle\, X_{t-j}, X_{t-1}\, \rangle \,. \tag{4.47}$$

Let C_j be the $M \times M$ matrix of auto-correlations of X at lag j then

$$C_j = \langle\, X_t, X_{t-j}\, \rangle \,. \tag{4.48}$$

Note that C_j are antisymmetric matrices. This means that

$$C_j^\tau = \langle\, X_{t-j}, X_t\, \rangle \,,$$

and in terms of this family of matrices, equation (4.47) may be expressed in the form

$$C_1 = \sum_{j=1}^{1} A_j C_{j-1} + \sum_{j=2}^{p} A_j C_{j-1}^\tau \,, \tag{4.49}$$

which further simplifies to give

$$C_1 = A_1 C_0 + \sum_{j=2}^{p} A_j C_{j-1}^\tau \,. \tag{4.50}$$

This strategy can be continued by multiplying both sides of equation (4.45) by X_{t-k}^τ where k takes all lags from $k = 2$ to $k = p$ to obtain

$$X_t X_{t-k}^\tau = \sum_{j=1}^{p} A_j X_{t-j} X_{t-k}^\tau + E_t X_{t-k}^\tau \,.$$

After taking the expectation of this equation and splitting the sum, the matrices $A_1, \cdots, A_p$ are seen to satisfy the $(p-1)$ equations

$$C_k = \sum_{j=1}^{k} A_j C_{|j-k|} + \sum_{j=k+1}^{p} A_j C_{|j-k|}^\tau \tag{4.51}$$

where k takes values from $k = 2$ to $k = p$. It is now convenient to scale the matrices $C_0, C_1, \cdots, C_p$ with respect to the covariance C_0 by introducing the $M \times M$ matrix

$$\rho_j = C_0^{-1} C_j \,, \qquad \rho_j^\tau = C_0^{-1} C_j^\tau \,.$$

In this construction $\rho_0 = I$, and the matrices, $C_1, \cdots, C_p$ are not symmetric. Consequently, equations (4.50) and (4.51) can be rewritten in the form

$$\rho_1 = A_1\rho_0 + A_2\rho_1^\tau + A_3\rho_2^\tau + A_4\rho_3^\tau + \cdots + A_p\rho_{p-1}^\tau,$$

$$\rho_2 = A_1\rho_1 + A_2\rho_0 + A_3\rho_1^\tau + A_4\rho_2^\tau + \cdots + A_p\rho_{p-2}^\tau,$$

$$\rho_3 = A_1\rho_2 + A_2\rho_1 + A_3\rho_0 + A_4\rho_1^\tau + \cdots + A_p\rho_{p-3}^\tau,$$

$$\rho_4 = A_1\rho_3 + A_2\rho_2 + A_3\rho_1 + A_4\rho_0 + \cdots + A_p\rho_{p-4}^\tau, \tag{4.52}$$

$$\vdots$$

$$\rho_N = A_1\rho_{p-1} + A_2\rho_{p-2} + A_3\rho_{p-3} + A_4\rho_{p-4} + \cdots + A_{p-1}\rho_1 + A_p\rho_0.$$

The transposed form of equations (4.52), when expressed in block matrix form, is

$$
\begin{bmatrix}
\rho_0 & \rho_1 & \rho_2 & \rho_3 & \cdots & \rho_{p-1} \\
\rho_1^\tau & \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{p-2} \\
\rho_2^\tau & \rho_1^\tau & \rho_0 & \rho_1 & \cdots & \rho_{p-3} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{p-1}^\tau & \rho_{p-2}^\tau & \rho_{p-3}^\tau & \rho_{p-4}^\tau & \cdots & \rho_0
\end{bmatrix}
\begin{bmatrix}
A_1^\tau \\
A_2^\tau \\
A_3^\tau \\
A_4^\tau \\
\vdots \\
A_p^\tau
\end{bmatrix}
=
\begin{bmatrix}
\rho_1^\tau \\
\rho_2^\tau \\
\rho_3^\tau \\
\rho_4^\tau \\
\vdots \\
\rho_p^\tau
\end{bmatrix}. \tag{4.53}
$$

Succinctly, $\mathbf{R}\mathbf{A}^\tau = \boldsymbol{\rho}^\tau$ where $\mathbf{R}$ is a symmetric matrix of full-rank, so that $\mathbf{R}^{-1}$ is guaranteed to exist. In particular,

$$\mathbf{A}^\tau = \mathbf{R}^{-1}\boldsymbol{\rho}^\tau. \tag{4.54}$$

4.4.2 Maximum Likelihood

The Maximum Likelihood (ML) procedure estimates the values of the parameters of a model to be those values which maximise the probability (likelihood) that a given sample of data is a realization of the model. From a practical point of view, ML estimation is robust and gives parameter estimates with good statistical properties in large samples. By this we mean that the procedure is asymptotically unbiased and consistent, that is, the standard error of parameter estimates decays as the square root of sample size increases. To estimate the matrices $A_1, \cdots, A_p$ using an ML procedure (Kalbfleisch [1985]), rewrite the model (4.45) as

$$E_t = X_t - \sum_{j=1}^{p} A_j X_{t-j}.$$

Supposing that E_t is distributed as a multivariate normal with mean value zero, then the associated probability density function of E is

$$f(E) = \frac{1}{(2\pi)^{M/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} E^\tau \Sigma^{-1} E\right), \quad (4.55)$$

where $|\Sigma|$ is the determinant of Σ, the covariance of E. Each observation of E_t is based on the observations of X_t and its past history up to and including X_{t-p}, and so the likelihood of observing E_t is

$$\mathcal{L}_t = \frac{1}{(2\pi)^{M/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\left(X_t - \sum_{j=1}^{p} A_j X_{t-j}\right)^\tau \Sigma^{-1}\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)\right]. \quad (4.56)$$

The negative log-likelihood function is therefore given by

$$-\log \mathcal{L}_t = \frac{M}{2}\log 2\pi + \frac{1}{2}\log|\Sigma| + \frac{1}{2}\left(X_t - \sum_{j=1}^{p} A_j X_{t-j}\right)^\tau \Sigma^{-1}\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right). \quad (4.57)$$

Since observations of E_t at different times are independent random variables then the likelihood of observing the sample is

$$\mathcal{L} = \prod_t \mathcal{L}_t,$$

and therefore the negative log-likelihood of observing the sample is

$$-\log \mathcal{L} = -\sum_t \log \mathcal{L}_t \propto -\mathbb{E}\left[\log \mathcal{L}_t\right],$$

where the constant of proportionality is the number of observation of E_t, that is, the number of transitions in the sample. Thus the computation of the negative log-likelihood of the sample is formally equivalent to the task of computing the expected value of expression (4.57). To facilitate this calculation we use equation (4.48), *i.e.*

$$C_j = \mathbb{E}\left[X(t)X^\tau(t-j)\right].$$

Taking the expectation of the negative log-likelihood gives

$$\mathbb{E}\left[-\log \mathcal{L}_t\right] = \mathbb{E}\left[\frac{M}{2}\log 2\pi + \frac{1}{2}\log|\Sigma| + \frac{1}{2}\left(X_t^\tau - \sum_{j=1}^{p} X_{t-j}^\tau A_j^\tau\right)\Sigma^{-1}\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)\right]. \quad (4.58)$$

Since the first and second terms on the right hand side of equation (4.58) are unaffected by the expectation (the expected value of a deterministic function is constant), our focus

will be on the third term of expression (4.58). This term is first expanded to obtain

$$\mathbb{E}\left[\left(X_t^{\tau} - \sum_{j=1}^{p} X_{t-j}^{\tau} A_j^{\tau}\right)\Sigma^{-1}\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)\right]$$

$$= \mathbb{E}\left[X_t^{\tau}\Sigma^{-1}X_t\right] - \sum_{j=1}^{p}\mathbb{E}\left[X_{t-j}^{\tau}A_j^{\tau}\Sigma^{-1}X_t + X_t^{\tau}\Sigma^{-1}A_jX_{t-j}\right]$$

$$+ \sum_{j=1}^{p}\sum_{k=1}^{p}\mathbb{E}\left[X_{t-j}^{\tau}A_j^{\tau}\Sigma^{-1}A_kX_{t-k}\right].$$

To simplify this expression each part is treated separately as shown in calculations (4.59).

(a) $\mathbb{E}\left[X_t^{\tau}\Sigma^{-1}X_t\right] = \mathbb{E}\left[X_{i,t}\,\Sigma_{is}^{-1}X_{s,t}\right] = C_{0:\,is}\Sigma_{is}^{-1}$,

(b) $\sum_{j=1}^{p}\mathbb{E}\left[X_{t-j}^{\tau}\,A_j^{\tau}\Sigma^{-1}X_t + X_t^{\tau}\Sigma^{-1}A_jX_{t-j}\right]$

$$= \sum_{j=1}^{p}\left[\mathbb{E}\left[X_{i,t-j}\,X_{s,t}\right]A_{j:\,mi}\Sigma_{ms}^{-1} + \mathbb{E}\left[X_{i,t}\,X_{s,t-j}\right]\Sigma_{im}^{-1}A_{j:\,ms}\right]$$

$$= \sum_{j=1}^{p}C_{j:\,is}\left[A_{j:\,mi}\Sigma_{ms}^{-1} + A_{j:\,ms}\Sigma_{im}^{-1}\right],$$

(4.59)

(c) $\sum_{j=1}^{p}\sum_{k=1}^{p}\mathbb{E}\left[X_{t-j}^{\tau}A_j^{\tau}\Sigma^{-1}A_kX_{t-k}\right] = \sum_{j=1}^{p}\sum_{k=1}^{p}\mathbb{E}\left[X_{i,t-j}\,X_{r,t-k}\right]A_{j:\,mi}\Sigma_{ms}^{-1}A_{k:\,sr}$

$$= \sum_{j=1}^{p}\sum_{k=1}^{p}C_{|j-k|:\,ir}A_{j:\,mi}\Sigma_{ms}^{-1}A_{k:\,sr}.$$

The individual calculations in (4.59) are now substituted into the equation (4.58) to give

$$-\mathbb{E}\left[\log\mathcal{L}_t\right] \quad = \quad \frac{M}{2}\log 2\pi + \frac{1}{2}\log|\Sigma| + \frac{1}{2}C_{0:\,is}\Sigma_{is}^{-1}$$

$$- \frac{1}{2}\sum_{j=1}^{p}C_{j:\,is}\left[A_{j:\,mi}\Sigma_{ms}^{-1} + A_{j:\,ms}\Sigma_{im}^{-1}\right]$$

(4.60)

$$+ \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p}C_{|j-k|:\,ir}A_{j:\,mi}\Sigma_{ms}^{-1}A_{k:\,sr}.$$

The matrices $A_1, \cdots, A_p$ are now chosen to minimize expression (4.60). The minimum value of $-\mathbb{E}\left[\log\mathcal{L}_t\right]$ will be attained at its stationary point with respect to variations in the values of the matrices $A_1, \cdots, A_p$, that is, where

$$\frac{\partial\left(\log\mathcal{L}_t\right)}{\partial A_{u:\,nq}} = 0.$$

Calculation gives

$$
\begin{aligned}
-\frac{\partial\big(\log\mathcal{L}_t\big)}{\partial A_{u:\,nq}} &= -\frac{1}{2}\sum_{j=1}^{p} C_{j:\,is}\big[\delta_{ju}\delta_{mn}\delta_{iq}\Sigma^{-1}_{ms} + \delta_{ju}\delta_{mn}\delta_{sq}\Sigma^{-1}_{im}\big] \\
&\quad + \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} C_{|j-k|:\,ir}\big[\delta_{ju}\delta_{mn}\delta_{iq}\Sigma^{-1}_{ms}A_{k:\,sr} + A_{j:\,mi}\Sigma^{-1}_{ms}\delta_{ku}\delta_{sn}\delta_{rq}\big] \\
&= -\frac{1}{2}\sum_{j=1}^{p}\big[C_{j:\,qs}\delta_{ju}\Sigma^{-1}_{ns} + C_{j:\,iq}\delta_{ju}\Sigma^{-1}_{in}\big] \\
&\quad + \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p}\big[C_{|j-k|:\,qr}\delta_{ju}\Sigma^{-1}_{ns}A_{k:\,sr} + C_{|j-k|:\,iq}A_{j:\,mi}\Sigma^{-1}_{mn}\delta_{ku}\big] \\
&= -\frac{1}{2}\big[C_{u:\,qs}\Sigma^{-1}_{ns} + C_{u:\,iq}\Sigma^{-1}_{in}\big] + \frac{1}{2}\sum_{k=1}^{p} C_{|u-k|:\,qr}\Sigma^{-1}_{ns}A_{k:\,sr} \\
&\quad + \frac{1}{2}\sum_{j=1}^{p} C_{|j-u|:\,iq}A_{j:\,mi}\Sigma^{-1}_{mn}\,.
\end{aligned}
$$

$$(4.61)$$

Taking account of the symmetry of Σ^{-1}, it follows immediately from calculation (4.61) that $A_1,\cdots,A_p$ should be chosen to satisfy

$$
-\frac{1}{2}\big[C_{u:\,qs}\Sigma^{-1}_{sn} + C_{u:\,iq}\Sigma^{-1}_{in}\big] + \frac{1}{2}\sum_{k=1}^{p} C_{|u-k|:\,qr}A_{k:\,sr}\Sigma^{-1}_{sn} + \frac{1}{2}\sum_{j=1}^{p} C_{|j-u|:\,iq}A_{j:\,mi}\Sigma^{-1}_{mn} = 0\,.
$$

$$(4.62)$$

Equation (4.62), when expressed in matrix notation, takes the simple form

$$
-C_u\Sigma^{-1} + \sum_{k=1}^{p} C_{|u-k|}A_k\Sigma^{-1} + \sum_{j=1}^{p} C_{|j-u|}A_j^{\tau}\Sigma^{-1} = 0\,,
$$

$$(4.63)$$

where u takes all integer values from $u=1$ to $u=p$ inclusive. Moreover, it is clear that all occurrences of Σ^{-1} can be removed by post-multiplying equation (4.63) by Σ, to get

$$
-C_u + \sum_{j=1}^{p} C_{|j-u|}A_j^{\tau} + \sum_{k=1}^{p} C_{|u-k|}A_k = 0\,, \qquad u = 1,2,\cdots,p\,.
$$

$$(4.64)$$

This equation can be rearranged to have the form

$$
C_k = \sum_{j=1}^{k} A_j C_{|j-k|} + \sum_{j=k+1}^{p} A_j C^{\tau}_{|j-k|}\,.
$$

$$(4.65)$$

Equation (4.65) is identical to that derived by the Yule-Walker procedure with, of course, the same solutions for the matrices $A_1,\cdots,A_p$.

Similarly, the optimal value of the residual covariance matrix Σ, which is chosen to minimise the value of expression(4.58), is obtained by setting to zero the derivative of $-\mathbb{E}\big[\log\mathcal{L}_t\big]$ with respect to the (i,j)-th component of Σ, that is

$$
\frac{\partial\big(\log\mathcal{L}_t\big)}{\partial\Sigma_{ij}} = 0\,.
$$

Differentiating equation (4.58) with respect to Σ_{ij} gives

$$-\frac{\partial\left(\log \mathcal{L}_t\right)}{\partial \Sigma_{ij}} = \frac{1}{|\Sigma|}\frac{\partial|\Sigma|}{\partial \Sigma_{ij}} + \left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)^{\tau}\frac{\partial \Sigma^{-1}}{\partial \Sigma_{ij}}\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right). \qquad (4.66)$$

To simplify equation (4.66), we use the result

$$|\Sigma| = \sum_{j=1}^{M} \Sigma_{ij}\alpha_{ij}, \qquad \alpha_{ij} = (-1)^{(i+j)}M_{ij}, \qquad (4.67)$$

where α_{ij} is the cofactor of Σ_{ij} in det Σ. Thus α_{ij} is a polynomial function of the elements of Σ but is independent of Σ_{ik} and Σ_{kj} for all k, and M_{ij} is the determinant of the matrix that results from Σ by removing the ith row and j^{th} column. Therefore

$$\frac{\partial|\Sigma|}{\partial \Sigma_{ij}} = \alpha_{ij}. \qquad (4.68)$$

Equation (B.14) can be expressed in the matrix form

$$\frac{\partial|\Sigma|}{\partial \Sigma} = \mathrm{Adj}\left(\Sigma\right), \qquad (4.69)$$

where $\mathrm{Adj}\left(\Sigma\right)$ is the cofactor matrix the entries of which are α_{ij}. The following calculations have been done as a preamble to the computation of $\partial \Sigma^{-1}/\partial \Sigma$. From the property $\Sigma\Sigma^{-1} = I$, it follows that

$$\frac{\partial\left(\Sigma\Sigma^{-1}\right)}{\partial \Sigma_{ij}} = 0,$$

and therefore

$$\frac{\partial \Sigma_{rn}}{\partial \Sigma_{ij}}\Sigma_{nq}^{-1} + \Sigma_{rn}\frac{\partial \Sigma_{nq}^{-1}}{\partial \Sigma_{ij}} = 0, \quad \Longrightarrow \quad \delta_{ri}\Sigma_{jq}^{-1} + \Sigma_{rn}\frac{\partial \Sigma_{nq}^{-1}}{\partial \Sigma_{ij}} = 0.$$

The pre-multiplication of this equation by Σ_{sr}^{-1} gives

$$\frac{\partial \Sigma_{sq}^{-1}}{\partial \Sigma_{ij}} = -\Sigma_{si}^{-1}\Sigma_{jq}^{-1}. \qquad (4.70)$$

By substituting equation (B.16) into (4.66), we obtain

$$\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)_n^{\tau}\Sigma_{ni}^{-1}\Sigma_{jq}^{-1}\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)_q = \frac{1}{|\Sigma|}\mathrm{Adj}(\Sigma)_{ji},$$

and when this equation is contracted with $\Sigma_{ir}\Sigma_{nj}$, the result is

$$\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)_n^{\tau}\Sigma_{ni}^{-1}\Sigma_{ir}\Sigma_{nj}\Sigma_{jq}^{-1}\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)_q = \frac{1}{|\Sigma|}\mathrm{Adj}(\Sigma)_{ji}\Sigma_{ir}\Sigma_{nj}. \qquad (4.71)$$

The identity $B\,\mathrm{Adj}(B) = |B|I$ is now used to deduce that

$$\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)_r^{\tau}\left(X_t - \sum_{k=1}^{p} A_k X_{t-k}\right)_n = \delta_{jr}\Sigma_{nj} = \Sigma_{nr}. \qquad (4.72)$$

Equation (4.72) can be expressed in block form as

$$\Sigma = \left(X_t - \sum_{j=1}^{p} A_j X_{t-j}\right)^{\tau} \left(X_t - \sum_{j=1}^{p} A_j X_{t-j}\right).$$ (4.73)

The standard error of the estimates of the matrices $A_1, \cdots, A_p$ are given by the square roots of the diagonal entries of the inverse Hessian of the negative log-likelihood function. In this case the Hessian of the negative log-likelihood function is

$$Hess = -\frac{\partial^2 \mathbb{E}\left[\log \mathcal{L}\right]}{\partial A_m \partial A_u} = (\text{Number of realisations of } E) \times \Sigma^{-1} C_{|m-u|},$$

with inverse Hessian

$$Hess^{-1} = \frac{C_{|m-u|}^{-1}\Sigma}{\text{Number of realisations of } E}.$$

4.4.3 Confidence interval for the estimated AR(p) coefficients

To validate the estimated autoregressive model coefficients A_j computed in the previous section, we will apply the t-statistical test, which is a measure of the deviation of the estimated parameter value from its natural value and the standard error. The null hypothesis to be examined is that the processes construct the autoregressive model are independent, that is $A_j = 0$, where j represents the lag order of the autoregressive model and $\{j = 1, \cdots, p\}$. Moreover, the matrix A_j is equal to the zero matrix, if and only if, all its entries $A_j(l, m) = 0$, where $\{l, m = 1, 2, \cdots, M\}$. That is, the interdependence between processes vanishes at lag number j.

Let $\hat{A}_j(l, m)$ denote the estimated values of the entries of the matrix A_j, and let $SE_j(l, m)$ denote the standard error of the estimated autoregressive coefficients $\hat{A}_j(l, m)$, then

$$\left|\frac{\hat{A}_j(l, m)}{SE_j(l, m)}\right| < 2; \quad j = 1, 2, \cdots, p.$$ (4.74)

This relation can be written as

$$-2SE_j(l, m) < \hat{A}_j(l, m) < 2SE_j(l, m); \quad j = 1, 2, \cdots, p.$$ (4.75)

Thus the values of the estimated AR coefficients , $\hat{A}_j(l, m)$, lying within the interval

$$\left(-2SE_j(l, m), 2SE_j(l, m)\right)$$

indicate that the estimated parameters $\hat{A}_j(l, m)$ are significantly zero, otherwise the null hypothesis is rejected.

4.4.4 Spectral analysis for AR(p)

The spectral components of the data are estimated from the autoregressive model of order p in equation (4.45). We began by applying the continuous Fourier transform defined in equation (3.1) to the model equation (4.45), to get

$$
\begin{aligned}
\mathcal{X}(f) = \int_{-\infty}^{\infty} X_t e^{-2i\pi ft}\, dt &= \int_{-\infty}^{\infty} \sum_{k=1}^{p} A_k X_{t-k} e^{-2i\pi ft}\, dt + \int_{-\infty}^{\infty} E_t e^{-2i\pi ft}\, dt \\
&= \sum_{k=1}^{p} A_k \int_{-\infty}^{\infty} X_{t-k} e^{-2i\pi f(t-k\Delta t)} e^{-2i\pi fk\Delta t}\, dt + \mathcal{E}(f),
\end{aligned}
$$

$$(4.76)$$

where $\mathcal{X}(f)$ denotes the Fourier transform of $X(t)$ and $\mathcal{E}(f)$ is the Fourier transform of E_t, and f represents the ordinary frequency, which is in turn linked to the angular frequency ω, by $\omega = 2\pi f$. A simple translation of integration variable in the integral on the right-hand side of the equation (4.76), gives the final equation

$$
\mathcal{X}(f) = \sum_{k=1}^{p} A_k e^{-2i\pi fk\Delta t}\, \mathcal{X}(f) + \mathcal{E}(f).
$$

$$(4.77)$$

Let $B(f)$ denote the transfer matrix, which is defined by

$$
B(f) = I - \sum_{k=1}^{p} A_k e^{-2i\pi fk\Delta t}.
$$

$$(4.78)$$

Then the expression (4.77) may be rearranged into the form $B(f)\mathcal{X}(f) = \mathcal{E}(f)$ with solution

$$
\mathcal{X}(f) = H(f)\mathcal{E}(f),
$$

$$(4.79)$$

where $H(f) = B(f)^{-1}$ represents the inverse of the transfer matrix. The power spectrum of X is therefore

$$
S^p(f) = \mathcal{X}(f)\mathcal{X}^*(f) = H(f)\mathcal{E}(f)\mathcal{E}^*(f)\, H^*(f),
$$

$$(4.80)$$

where the superscript asterisk (*) denotes the matrix Hermitian operator, namely the process of taking the complex conjugate of the matrix transpose.

The essential assumption made in the analysis of the error process E_t is that it has instantaneous covariance Σ but otherwise is uncorrelated at different times, that is,

$$
\int_{-\infty}^{\infty} E_t E_{t-s}^{\tau}\, dt = \Sigma\, \delta(s),
$$

$$(4.81)$$

where $\delta(s)$ is the Dirac distribution. This function takes the value of zero everywhere except at $s = 0$ and also satisfies

$$
\int_{-\infty}^{\infty} f(x)\delta(x - s)\, dx = f(s).
$$

Equation (4.81) is multiplied by $e^{-2i\pi fs}$ and then integrated over $\mathbb{R}$ with respect to s to obtain

$$\int_{-\infty}^{\infty} e^{-2i\pi fs}\left(\int_{-\infty}^{\infty} E_t E_{t-s}^{\tau}\, dt\right) ds \;=\; \int_{-\infty}^{\infty} E_t e^{-2i\pi ft}\, dt \int_{-\infty}^{\infty} E_{t-s}^{\tau} e^{2i\pi f(t-s)}\, ds$$

$$=\; \int_{-\infty}^{\infty} E_t e^{-2i\pi ft}\, dt \int_{-\infty}^{\infty} E_u^{\tau} e^{2i\pi fu}\, du = \mathcal{E}(f)\mathcal{E}^*(f)\,.$$

In conclusion,

$$\mathcal{E}(f)\mathcal{E}^*(f) = \int_{-\infty}^{\infty} e^{-2i\pi fs}\left(\int_{-\infty}^{\infty} E_t E_{t-s}^{\tau}\, dt\right) ds = \int_{-\infty}^{\infty} \Sigma\delta(s)e^{-2i\pi fs}\, ds = \Sigma\,.$$

Hence the model specification of the power spectrum of the error process is

$$S^p(f) = \mathcal{X}(f)\mathcal{X}^*(f) = H(f)\,\Sigma\,H^*(f)\,. \tag{4.82}$$

Here $S^p(f)$ represents an $M \times M$ parametric spectral density matrix of the process X . The superscript p has been included to distinguish the spectral density computed by the parametric approach, namely the multivariate autoregressive model system, from the spectral density, $S(f)$, that is computed directly from the data. The diagonal entries of $S^p(f)$ contain the auto-spectra or the power spectra of the vector X components at frequency f, say $S^p_{X_{i,t}X_{i,t}}(f)\,; i = 1, \cdots, M$, while the off-diagonal entries involve the cross-spectra between the elements of X. For example, $S^p_{X_{i,t}X_{j,t}}(f)$, where $i,j = 1, \cdots, M$, represents the cross-spectrum between the two process $X_{i,t}$ and $X_{j,t}$.

Consequently, the interrelation between components of the multivariate process X can be captured using the ordinary coherence and partial coherence measurements given in equations (4.2) and (4.20) respectively, by replacing the spectral density $S(f)$ by the parametric spectral density as follows

$$|R_{X_iX_j}(f)|^2 = \frac{|S^p_{X_iX_j}(f)|^2}{S^p_{X_iX_i}(f)S^p_{X_jX_j}(f)}\,, \quad i,j = 1, \cdots, M. \tag{4.83}$$

This equation gives the definition of the coherence between two components of the multivariate vector process X, say X_i and X_j. The partial coherence between two processes X_i and X_j given a process X_k, is defined as follows,

$$|R_{X_iX_j|X_k}(f)|^2 = \frac{|S^p_{X_iX_j|X_k}(f)|^2}{S^p_{X_iX_i|X_k}(f)S^p_{X_jX_j|X_k}(f)}\,, \quad i,j,k = 1, \cdots, M \text{ and } k \neq i \text{ or } j. \tag{4.84}$$

Thus, this equation defines the correlation between two processes X_i and X_j after removing the linear influence of the process X_k.

4.5 Granger causality

Granger causality is a statistical measure of causal or directional influence from one time series to another and is based on a regressive model which attempts to provide a linear prediction of the behaviour of network activity (a time series) from its history (Granger [1969], Dhamala *et al.* [2008]). The concept of Granger causality is based on the fact that the causes precede their influences in time. That is, given two stationary stochastic processes, the first process is said to be G-causal for the second process if knowledge involved in the history of the first process plays a crucial role in improving the prediction of the second process. In the following sections, we will demonstrate the concept of Granger causality in its simple performance for bi-variate autoregressive models of appropriate model order p, which is called bi-variate Granger causality (section (4.5.1)). While Granger causality in multivariate system, which is based on fitting the underlying data by autoregressive vector models, is called multivariate Granger causality, or simply MVGC, will be presented in section (4.5.2).

4.5.1 Bi-variate G-Causality

Suppose that the two stationary stochastic processes X_t and Y_t satisfy bi-variate autoregressive models (Kaminski *et al.* ([2001]))

$$
\begin{aligned}
X_t &= \sum_{j=1}^{p} A_{XX,j} X_{t-j} + \sum_{j=1}^{p} A_{XY,j} Y_{t-j} + E_{X,t} \\
Y_t &= \sum_{j=1}^{p} A_{YX,j} X_{t-j} + \sum_{j=1}^{p} A_{YY,j} Y_{t-j} + E_{Y,t},
\end{aligned}
\tag{4.85}
$$

where $A_{ik,j}, i, k = X, Y$, are matrices of coefficients at lags, $j = 1, 2, \cdots, p$, and $E_{i,t}, i = X, Y$, is a residual error vector of dimension 2 with 2×2 covariance matrix Σ. In particular, error vectors at different times are assumed to be independent random variables, as mentioned in the previous section, with mean value zero. If the variance of the prediction error $E_{X,t}$ (or $E_{Y,t}$) is reduced by the inclusion of the Y (or X) terms in the first (or second) equation, then, according to Granger causality Y (or X), is said to have a causal influence on X (or Y). To construct the spectral decomposition of Granger causality as proposed by Geweke [1982], we rewrite equation (4.85) in the matrix form

$$
\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{j=1}^{p} \begin{pmatrix} A_{XX,j} & A_{XY,j} \\ A_{YX,j} & A_{YY,j} \end{pmatrix} \begin{pmatrix} X_{t-j} \\ Y_{t-j} \end{pmatrix} + \begin{pmatrix} E_{X,t} \\ E_{Y,t} \end{pmatrix}.
\tag{4.86}
$$

The Fourier transform of equation (4.86) can be easily obtained as

$$
\begin{pmatrix} B_{XX}(f) & B_{XY}(f) \\ B_{YX}(f) & B_{YY}(f) \end{pmatrix} \begin{pmatrix} \mathcal{X}(f) \\ \mathcal{Y}(f) \end{pmatrix} = \begin{pmatrix} \mathcal{E}_X(f) \\ \mathcal{E}_Y(f) \end{pmatrix}, \tag{4.87}
$$

where $\mathcal{Z}(f) = \int_{-\infty}^{\infty} Z_t e^{-2i\pi ft}$ represents the Fourier transform of the process Z and the components of the coefficient matrix $B(f)$ take the form

$$
B(f) = I - \sum_{j=1}^{p} A_j e^{-2i\pi fj\Delta t} \tag{4.88}
$$

where $A_1, \cdots, A_p, B$, and I are 2×2 matrices. Equation (4.87) can be written as

$$
\begin{pmatrix} \mathcal{X}(f) \\ \mathcal{Y}(f) \end{pmatrix} = \begin{pmatrix} H_{XX}(f) & H_{XY}(f) \\ H_{YX}(f) & H_{YY}(f) \end{pmatrix} \begin{pmatrix} \mathcal{E}_X(f) \\ \mathcal{E}_Y(f) \end{pmatrix}, \tag{4.89}
$$

where $H(f) = B^{-1}(f)$ denotes the inverse of the transfer function matrix. Therefore, the spectral density matrix at frequency f is denoted by $S^p(f)$, and expressed as,

$$
\begin{aligned}
S^p(f) &= \mathcal{X}(f)\mathcal{X}^*(f) \\
&= H(f)\mathcal{E}(f)\mathcal{E}^*(f)\,H^*(f) \\
&= H(f)\Sigma H^*(f),
\end{aligned} \tag{4.90}
$$

and the covariance matrix of the residual errors has the form

$$
\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix},
$$

where the diagonal components of the covariance matrix have the definitions $\Sigma_{XX} = \mathrm{Var}(E_{X,t})$, $\Sigma_{YY} = \mathrm{Var}(E_{Y,t})$, and the off-diagonal components have the definitions $\Sigma_{XY} = \Sigma_{YX} = \mathrm{Cov}(E_{X,t}, E_{Y,t})$.

In order to illustrate the causal influence of Y on X, we investigate the spectral density component S^p_{XX}, so from equation(4.90) we have

$$
S^p_{XX} = H_{XX}\Sigma_{XX}H^*_{XX} + 2\Sigma_{XY}Re(H_{XX}H^*_{XY}) + H_{XY}\Sigma_{YY}H^*_{XY}.
$$

To evaluate the causal power contribution, it is beneficial to eliminate the cross terms (second term) of the spectral density S^p_{XX}, using the transformation that is introduced by

Geweke [1982]. For the process X, the transformation takes the form

$$\begin{pmatrix} 1 & 0 \\ -\Sigma_{XY}/\Sigma_{XX} & 1 \end{pmatrix}. \tag{4.91}$$

Multiplying equation (4.87) by the transformation (4.91) gives

$$\begin{pmatrix} B_{XX}(f) & B_{XY}(f) \\ \widehat{B}_{YX}(f) & \widehat{B}_{YY}(f) \end{pmatrix} \begin{pmatrix} \mathcal{X}(f) \\ \mathcal{Y}(f) \end{pmatrix} = \begin{pmatrix} \mathcal{E}_X(f) \\ \widehat{\mathcal{E}}_Y(f) \end{pmatrix},$$

where

$$\widehat{B}_{YX}(f) = B_{YX}(f) - \frac{\Sigma_{XY}}{\Sigma_{XX}} B_{XX}(f),$$

$$\widehat{B}_{YY}(f) = B_{YY}(f) - \frac{\Sigma_{XY}}{\Sigma_{XX}} B_{XY}(f),$$

$$\widehat{\mathcal{E}}_Y(f) = \mathcal{E}_Y(f) - \frac{\Sigma_{XY}}{\Sigma_{XX}} \mathcal{E}_X(f).$$

Now $\mathcal{E}_X$ and $\widehat{\mathcal{E}}_Y$ are uncorrelated with each other even at the same time,

$$\begin{aligned} \operatorname{cov}(\mathcal{E}_X, \widehat{\mathcal{E}}_Y) &= \mathrm{E}\left[\mathcal{E}_X \widehat{\mathcal{E}}_Y\right] \\ &= \mathrm{E}\left[\mathcal{E}_X\left(\mathcal{E}_Y - \frac{\Sigma_{XY}}{\Sigma_{XX}}\mathcal{E}_X\right)\right] \\ &= \mathrm{E}\left[\mathcal{E}_X\mathcal{E}_Y - \frac{\Sigma_{XY}}{\Sigma_{XX}}\mathcal{E}_X\mathcal{E}_X\right] \\ &= \Sigma_{XY} - \Sigma_{XY} \\ &= 0. \end{aligned}$$

Consequently, the transformed covariance matrix of the noise processes takes the form

$$\widehat{\Sigma} = \mathrm{E}\left[\begin{pmatrix} \mathcal{E}_X \\ \widehat{\mathcal{E}}_Y \end{pmatrix} \begin{pmatrix} \mathcal{E}_X & \widehat{\mathcal{E}}_Y \end{pmatrix} \right] = \begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \widehat{\Sigma}_{YY} \end{pmatrix}, \tag{4.92}$$

where $\widehat{\Sigma}_{YY} = \Sigma_{YY} - \Sigma_{XY}/\Sigma_{XX}$, and the new inverse transfer matrix $\widehat{H}(f)$ takes the form

$$\widehat{H} = \begin{pmatrix} \widehat{H}_{XX}(f) & H_{XY}(f) \\ \widehat{H}_{YX}(f) & H_{YY}(f) \end{pmatrix},$$

where the transformed components of the inverse transfer matrix, $\widehat{H}_{XX}$ and $\widehat{H}_{YX}$, take the values

$$\widehat{H}_{XX}(f) = H_{XX}(f) + \frac{\Sigma_{XY}}{\Sigma_{XX}} H_{XY}(f),$$

$$\widehat{H}_{YX}(f) = H_{YX}(f) + \frac{\Sigma_{XY}}{\Sigma_{XX}} H_{YY}(f).$$

Hence, the component of the causal density spectrum $\widehat{S}^p_{XX}$ can be written as a combination of the intrinsic power of X and the causal power due to the linear influence of Y on X.

$$\widehat{S}^p_{XX}(f) = \widehat{H}_{XX}(f)\Sigma_{XX}\widehat{H}^*_{XX}(f) + \widehat{H}_{XY}(f)\widehat{\Sigma}_{YY}\widehat{H}^*_{XY}(f). \tag{4.93}$$

Since Granger causality is the natural logarithm of the ratio of total power to the intrinsic power (Geweke [1982], Wang *et al.* [2007], Dhamala *et al.* [2008]), the causal influence of Y on X at frequency f is

$$F_{Y \to X}(f) = \log \left(\frac{\widehat{S}^p_{XX}(f)}{\widehat{S}^p_{XX}(f) - (\Sigma_{YY} - \Sigma^2_{XY}/\Sigma_{XX})|H_{XY}(f)|^2} \right). \tag{4.94}$$

Similarly, the causal influence of X on Y at frequency f, can be obtained by using the transformation matrix

$$\begin{pmatrix} 1 & -\Sigma_{XY}/\Sigma_{YY} \\ 0 & 1 \end{pmatrix},$$

and following the same steps that lead to equation (4.94). Hence the linear causality from X to Y can be interpreted as

$$F_{X \to Y}(f) = \log \left(\frac{\widehat{S}^p_{YY}(f)}{\widehat{S}^p_{YY}(f) - (\Sigma_{XX} - \Sigma^2_{YX}/\Sigma_{YY})|H_{YX}(f)|^2} \right). \tag{4.95}$$

The instantaneous linear causality as defined by (Geweke [1982], Wang *et al.* [2007]) is

$$F_{X.Y}(f) = \log \left(\frac{|\widehat{H}_{XX}(f)\Sigma_{XX}\widehat{H}^*_{XX}(f)||\widehat{H}_{YY}(f)\Sigma_{YY}\widehat{H}^*_{YY}(f)|}{\widehat{S}^p(f)} \right). \tag{4.96}$$

The combination of equations (4.94), (4.95) and (4.96) leads to the linear dependence between X and Y which takes the form

$$F_{X,Y}(f) = F_{Y \to X} + F_{X \to Y} + F_{X.Y}. \tag{4.97}$$

The measure of the linear dependence is the sum of the measures of the three types of linear causality.

To investigate the causal influences in a multivariate stochastic system with respect to Granger causality, the partial directed coherence will be elaborated in the following section.

4.5.2 Partial directed coherence for MVGC

Recall the mathematical model

$$X_t = \sum_{k=1}^{p} A_k X_{t-k} + E_t \,, \tag{4.98}$$

where X_t represents an M-dimensional stationary time series, A_k's, where $k = 1, \cdots, p$, are coefficient matrices of dimension $M \times M$ of the model and E_t is an M-dimensional residual error vector with mean value zero and covariance matrix Σ. As mentioned above, the essential assumption is that the error vectors are uncorrelated across time.

In the multivariate autoregressive model the coefficients $A_k(i,j)$ represent the linear dependence of the current values of the process X_i on the past values of the process X_j where $i,j = 1, \cdots, M$. This means that the process X_j has causal influence on the process X_i, if the information involved in the past values of the process X_j improves the predictability of the process X_i, taking into account the influences of the other processes.

The frequency domain of Granger causality in a multivariate system is based on the autoregressive model proposed by Baccalá and Sameshima [2001] through the concept of the *partial directed coherence* which is defined as

$$|\pi_{i \leftarrow j}(f)| = \frac{|B_{ij}(f)|}{\sqrt{\sum_m |B_{mj}(f)|^2}} \,, \tag{4.99}$$

where $B_{ij}(f)$ denotes the (i,j) component of the transfer matrix $B(f) = I - \sum_{k=1}^{p} A_k e^{-2i\pi f k \Delta t}$ at frequency f. This definition expresses the ratio of the influence of the process X_j on the process X_i to the joint influences of the process X_j on the other processes in the system. According to this definition, the partial directed coherence is equal to zero for all frequency f, if and only if, all coefficients $A_k(i,j)$ are zero, and hence the process X_j is not Granger causal to the process X_i conditioned on the other processes in the system. Furthermore, the values of the partial directed coherence are ranging between 0 and 1 due to the normalization in equation (4.99).

In order to assess whether the estimated values of the partial directed coherence from fitting the multivariate autoregressive model of order p to the data, are significant or not, it is essential to determine a significance level for $|\pi_{i \leftarrow j}(f)|$. Schelter *et al.* [2005] derived a significant level for the partial directed coherence using the statistical properties of the estimated autoregressive coefficients $A_k(i,j)$. Thus under the null hypothesis that $|\pi_{i \leftarrow j}(f)| = 0$, the proposed significance level is given by

$$\left(\frac{G_{ij(f)} \chi^2_{1,1-\alpha}}{T \sum_m |B_{mj}(f)|^2} \right)^{1/2}, \tag{4.100}$$

where T represents the number of observations, $\chi^2_{1,1-\alpha}$ denotes the $(1-\alpha)$quantile of the chi-squared distribution with one degree of freedom, and $G_{ij}(f)$ is defined by

$$G_{ij}(f) = \Sigma_{ii}\left[\sum_{k=1}^{p}\sum_{l=1}^{p}D_{jj}(k,l)(\cos(kf)\cos(lf) + \sin(kf)\sin(lf))\right], \qquad (4.101)$$

where $D_{jj}(k,l)$ denotes the components of the inverse of the covariance matrix R of the multivariate autoregressive model, that is $D = R^{-1}$. The entries of the covariance matrix are $M \times M$ sub-matrices as defined in equations (4.53) and (4.54). Therefore, values of the estimated partial directed coherence lying above the significance level line indicate the existence of a causal influence from the process X_j to the process X_i, taking into account the intermediate effects of the process X_j on the other processes in the considered system.

4.6 Conclusion

In this chapter, we have discussed the spectral association measurements, the ordinary and partial coherences. Coherence function is commonly used in detecting the spontaneous pairwise linear relations within a dynamic system and assessing statistically the strength of the correlation under a null hypothesis. The partial coherence is used to determine the depth of the coherence as it measures the linear relation between two stationary processes after removing the linear effects of the intermediate process (or processes). Moreover, we have demonstrated the mechanism of estimating spectral densities and partial spectral densities directly from the data, and hence estimate coherences and partial coherences at any order independently, without need to calculate the lower order. The significant confidence intervals of coherences under the null hypothesis that, the two considered stationary processes are uncorrelated at level of significance $\alpha\%$, are explained for the three non-parametric spectral estimators, the method of averaging across periodograms of adjacent segments from single records, the method of averaging frequency and the multi-taper method.

In addition, we have discussed the traditional multivariate autoregressive model as a parametric spectral estimate. We have also used the likelihood and the Yule-Walker methods to estimate the autoregressive model parameters A_j's, as well as the error covariance Σ. The MVAR estimated parameters obtained by the Yule-Walker are identical to those obtained by the maximum likelihood estimation. However, The maximum likelihood method facilitates the calculation of the estimated standard error, which is a statistically important to examine the accuracy of the estimation. The causal influences have been discussed

and the direction of the information flow has been defined for bi-variate systems using the direct Granger causality in the spectral domain and for multivariate autoregressive model systems using a measure called partial directed coherence.

Chapter 5

Test for bands of significant coherence in finite samples

5.1 Introduction

The likelihood principle is a statistical principle of inference which asserts that all the information in a sample is contained in its likelihood function (Berger [1988], Birnbaum [1962]). One of the most widely used applications of the likelihood principle is the estimation of model parameters by choosing them to maximize the likelihood of a sample. The procedure is called Maximum Likelihood Estimation (MLE) and is often employed in signal processing analysis to estimate parameters. Moreover, parameter estimates generated by MLE are asymptotically unbiased (consistent) and asymptotically Gaussian with minimum variance (efficient), with estimates achieving the Cramer-Rao lower bound, which is defined as the smallest variance that an unbiased or asymptotically unbiased estimator can achieve, (see, Kay [1993], Peleg and Porat [1991], Rice *et al.* [2001]).

Given two stationary time series, we aim to use the likelihood principle to compute the probability that two experimental data sets are uncorrelated in a particular band of frequencies. To achieve this objective, we shall use coherence to measure the strength of correlation between the two signals at each frequency of the band, and thereafter compute the value of the likelihood of observing such a sequence of coherences under the null hypothesis. A probability is then associated with this value of likelihood, and a decision is taken as to whether or not the signals are correlated or uncorrelated in this band of frequencies. Although the strategy is identical, in practice, we prefer to use the negative log-likelihood function instead of the traditional likelihood function.

To achieve this objective, we will use the known probability density function for the coherences between samples of the two independent time series X and Y, derived in Goodman [1965] and Brillinger [1981], and defined previously in Chapter 4 by equation (4.6). Asymptotically as the sample size increases, coherences at different frequencies are independent. Therefore the computation of negative log-likelihood comprises the sum of the negative log-likelihood of independent coherences, which in turn behaves asymptotically as the sum of independent random variables. Each random variable has a finite mean and a finite variance, so the conditions necessary for the validity of the central limit theorem (CLT) are satisfied. The theorem asserts that the value of the negative log-likelihood of the band of frequencies is increasingly well approximated by a Gaussian random variable as the number of frequencies in the band is increased. Indeed, the conventional view is that five or six independent frequencies are enough for the CLT to give a good approximation for their distribution. These asymptotic results provide a good way to validate our procedure: the distribution of our test statistic is expected to converge to that of a Gaussian distribution when a large amount of data is available and larger bands of frequencies are used.

Of course, the primary difficulty in the application of the likelihood procedure in this work, is that the available samples have a finite (and sometimes relatively small) size, and that the investigation may sometimes involve a band which has as few as four or five frequencies. In such applications, neither is the conclusion of the central limit theorem appropriate, nor are the calculated coherences at sequential frequencies independent. The statistical properties of small bands of frequencies calculated from finite samples must be determined by simulation under the null hypothesis that the signals are uncorrelated. Another important point to note is that the test statistic is one-sided, as large coherences push the test statistic into the right tail of its distribution. In the simulations, the null hypothesis will be set at the confidence levels $\alpha = 1\%, 2.5\%, 5\%$ and 10%, to follow its rejection region.

5.1.1 Asymptotic behaviour of the likelihood function for samples of coherences

Supposing we have two independent stationary stochastic processes X and Y, then the probability density function of the coherence $U = |R_{XY}(\omega)|^2$ between samples of these

processes, has been given by equation (4.6). Recall the pdf of the coherence U

$$f(u) = \frac{\Gamma(L)}{\Gamma(L-r)\Gamma(r)} u^{2r-2}(1-u)^{L-r-1} , \tag{5.1}$$

where r is the dimension of the stochastic processes of interest and L is the number of non-overlapping sub-samples of length N, into which a record of total length T has been sub-divided. In other words, if there are T observations of a stochastic process, namely X, then $L = T/N$ where N has to be chosen appropriately (see, Brillinger, [1981], Amjad *et al.* [1997]). Neurophyiological data is usually observed at millisecond intervals with a large sample typically consisting of 60 seconds of recording, some 60,000 data in total for each process. However, the samples at our disposal have as little as 5 seconds of data. When data is measured at millisecond intervals the choice $N = 1000$ leads naturally to physical frequencies in Hertz (Hz) whereas other popular choices of N (say $N = 512$ or $N = 1024$) necessarily require the natural Fourier frequency to be rescaled into Hertz.

When $r = 1$, then equation (5.1), becomes

$$f(u) = (L-1)(1-u)^{L-2} \tag{5.2}$$

and the likelihood function of observing a sequence $u_K, \ldots, u_{K+M}$ of independent coherences over the frequency band of K Hz to $K + M$ Hz inclusive is

$$\mathcal{L}(u_K, \ldots, u_{K+M}) = \prod_{i=0}^{M} f(u_{K+i}) . \tag{5.3}$$

Since the calculation of log-likelihood is more straightforward and is equivalent to calculating likelihood, then the log-likelihood of observing a sequence $u_K, \ldots, u_{K+M}$ of independent coherences is therefore,

$$\log \mathcal{L} = (M+1)\log(L-1) + (L-2)\sum_{i=0}^{M} \log(1-u_{K+i}) . \tag{5.4}$$

Although $\log(1-U)$ is not a Gaussian distributed random variable, it has a finite mean and variance, and therefore, for a sufficiently large value of M the central limit theorem indicates that $\log \mathcal{L}$ will behave as a Gaussian random variable, with the mean value and variance to be determined. In the specific case in which $U_K, \cdots, U_{K+M}$ are assumed to behave as independent random variables then

$$\mathbb{E}\left[\log \mathcal{L}\right] = (M+1) \times \mathbb{E}\left[\log f(u)\right], \qquad \mathrm{Var}\left[\log \mathcal{L}\right] = (M+1) \times \mathrm{Var}[\log f(u)]. \tag{5.5}$$

Therefore, in this instance, the task is to compute $\mathbb{E}\left[\log f(u)\right]$ and $\mathrm{Var}[\log f(u)]$ for the pdf given by equation (5.2). The calculation of $\mathbb{E}\left[\log f(u)\right]$ yields

$$
\begin{aligned}
\mathbb{E}\left[\log f(u)\right] &= \int_0^1 (\log(L-1) + (L-2)\log(1-u))f(u)\,du \\
&= \log(L-1) + (L-2)\int_0^1 \log(1-u)f(u)\,du \qquad (5.6) \\
&= \log(L-1) - \frac{L-2}{L-1}.
\end{aligned}
$$

Given the value of $\mathbb{E}\left[\log f(u)\right]$, in this case the computation of $\mathrm{Var}[\log f(u)]$, is most easily achieved from its definition, to get

$$
\begin{aligned}
\mathrm{Var}[\log f(u)] &= \int_0^1 \left(\log(L-1) + (L-2)\log(1-u) - \mathbb{E}\left[\log(f(u))\right]\right)^2 f(u)\,du \\
&= \int_0^1 \left((L-2)\log(1-u) + \frac{L-2}{L-1}\right)^2 f(u)\,du \qquad (5.7) \\
&= (L-2)^2 \int_0^1 \left(\log(1-u) + \frac{1}{L-1}\right)^2 f(u)\,du.
\end{aligned}
$$

Therefore $\mathrm{Var}[\log(f(u)] = (L-2)^2\mathrm{Var}[\log(1-u)]$. The identity $\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}\left[X\right])^2$ now yields

$$
\begin{aligned}
\mathrm{Var}[\log f(u)] &= (L-2)^2\left[\int_0^1 \left(\log(1-u)\right)^2 f(u)\,du - \frac{1}{(L-1)^2}\right] \\
&= (L-2)^2\left[\frac{2}{(L-1)^2} - \frac{1}{(L-1)^2}\right] \qquad (5.8) \\
&= \frac{(L-2)^2}{(L-1)^2}.
\end{aligned}
$$

The mean value and variance of the $\log \mathcal{L}$ are now determined by substituting results (5.6) and (5.8) into the equation (5.5), to recover

$$
\mathbb{E}\left[\log \mathcal{L}\right] = (M+1)\left[\log(L-1) - \frac{L-2}{L-1}\right], \qquad \mathrm{Var}[\log(\mathcal{L})] = (M+1)\frac{(L-2)^2}{(L-1)^2}. \quad (5.9)
$$

Therefore the asymptotic estimate of the standard deviation σ of $\log \mathcal{L}$ is

$$
\sigma = \sqrt{(M+1)}\frac{(L-2)}{(L-1)}. \qquad (5.10)
$$

In practice our simulation experiments achieve this result for samples of two independent processes when the sample of frequencies comprises 5 or 6, or more frequencies, which are chosen to be well separated, so that the correlation structure between successive frequencies has no role to play in shaping the properties of the sample. Figure (5.1) illustrates the Gaussian behaviour of the negative $\log \mathcal{L}$ for the coherence between samples of two

stationary uncorrelated processes of duration $T = 100$ seconds at one millisecond intervals, when the data has been subdivided into 100 sub-samples of size $N = 1000$ and for $M = 19$ (20 frequencies) well separated over a spectrum of $200\,\mathrm{Hz}$.
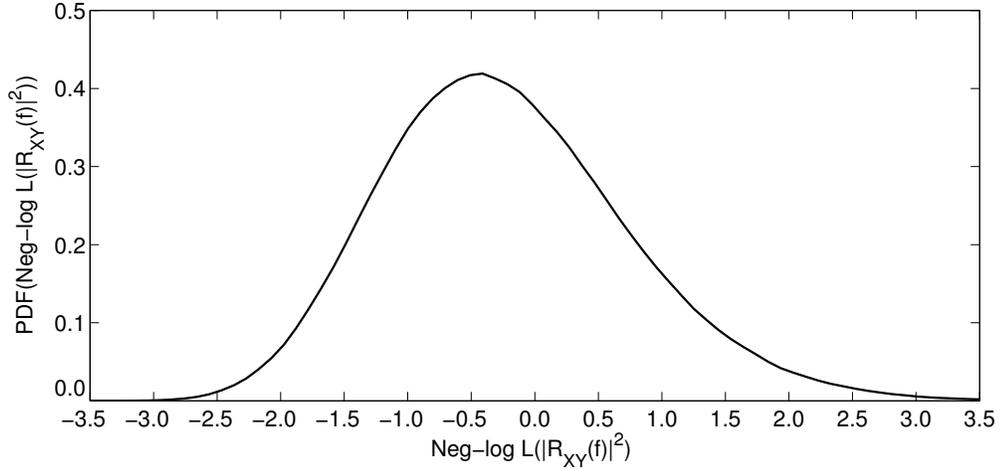

Figure 5.1: Distribution of negative log-likelihood of coherence at $M = 19$(20 frequencies) and $T = 100$ sec.

5.1.2 Behaviour of the negative log-likelihood function for bands of frequencies

The previous section has demonstrated the Gaussian behaviour of the negative log-likelihood function for the coherences between samples of two stationary stochastic processes known *a priori* to be independent for sub-samples of well-separated frequencies. This section will investigate the behaviour of the negative log-likelihood function of coherences between samples of independent stationary stochastic processes, say X and Y, for small bands of frequencies. The investigation will be carried out by simulating millisecond data for two independent time series using the autoregressive model

$$
\begin{aligned}
X_t &= \alpha X_{t-1} + E_{X,t}, \\
Y_t &= \beta Y_{t-1} + E_{Y,t},
\end{aligned}
\tag{5.11}
$$

where $|\alpha| < 1$ and $|\beta| < 1$ are the autoregressive coefficients of each model and $E_{X,t}$ and $E_{Y,t}$ are independent white noise processes with zero mean and unit variances for processes X and Y, respectively. The simulation exercise consists of one million trials of durations from 5 seconds to 60 seconds in 5 second intervals. Each simulated sample (X, Y) is first divided into L sub-samples of duration one second (*i.e.*, $N = 1000$), and the finite Fourier

transform is used to compute the associated sample of Fourier coefficients

$$
\begin{aligned}
d_X(f,l) &= \sum_{t=(l-1)N}^{lN-1} X(t)\, e^{-2i\pi ft} \\
d_Y(f,l) &= \sum_{t=(l-1)N}^{lN-1} Y(t)\, e^{-2i\pi ft}
\end{aligned}
\tag{5.12}
$$

for the lth sample at frequencies between $f = 3$Hz to $f = 100$Hz in intervals of 1Hz. The estimated cross-spectrum and power spectra for X and Y are computed from the formulae

$$
\hat{S}_{ij}(f) = \frac{1}{L}\sum_{l=1}^{L} I_{ij}^{N}(f,l),
\tag{5.13}
$$

where,

$$
I_{ij}^{N}(f,l) = \frac{1}{2\pi N} d_i(f,l)\overline{d_j(f,l)},
\tag{5.14}
$$

where $I_{ij}^{N}(f,l)$ represents the periodogram of the lth sample at frequency f, the overbar indicates a complex conjugate, and (i,j) denotes the pairs (X,X), (X,Y), (Y,X) and (Y,Y). The quantities $\hat{S}_{XX}(f)$ and $\hat{S}_{YY}(f)$ are the estimated power spectra of the processes X and Y, respectively, while $\hat{S}_{XY}(f)$ is the estimated cross-spectrum between the processes X and Y (see, Brillinger [1981], Amjad *et al.* [1997]). The estimated coherence between the processes X and Y is defined by

$$
U_{XY}(f) = |\hat{R}_{XY}(f)|^2 = \frac{|\hat{S}_{XY}(f)|^2}{\hat{S}_{XX}(f)\hat{S}_{YY}(f)},
\tag{5.15}
$$

where $\hat{R}_{XY}(f)$ is the estimated *coherency* between X and Y. The definition guarantees that coherence is a real valued-function of frequency, such that $U_{XY} \in [0,1]$. Coherence gives a measure of the linear correlation between two processes, say X and Y, at frequency f. If the coherence between two processes has value zero at frequency f, then these processes are said to be uncorrelated at that frequency, whereas a coherence with value one means that the processes are perfectly correlated at that frequency.

The negative log-likelihood function of independent coherences is computed for bands of frequencies ranging in size from $R = 3$ consecutive frequencies to $R = 13$ consecutive frequencies. The probability density function for the negative log-likelihood function $(-\log\mathcal{L})$ for each band of frequencies has been estimated using Kernel density estimation. Although not explicitly presented here, it has been verified by simulation that the distribution of negative log-likelihood depends only on the number of consecutive frequencies in the band. Specifically, the distribution is independent of the location of the band of frequencies.

The estimated distribution of the negative log-likelihood $(-\log \mathcal{L})$ of coherences be-tween finite independent samples, has been computed for different sets of data at distinct bands of frequencies, where the size of band ranges between $R = 3$ to $R = 13$. For a fixed number of sub-samples L, increasing the size of the frquency band from $R = 3$ to $R = 13$ results in a contraction in the length of the tail of the distribution, that is, the distribution moves gradually in the direction of becoming a Gaussian distribution, as illustrated in figure (5.2).
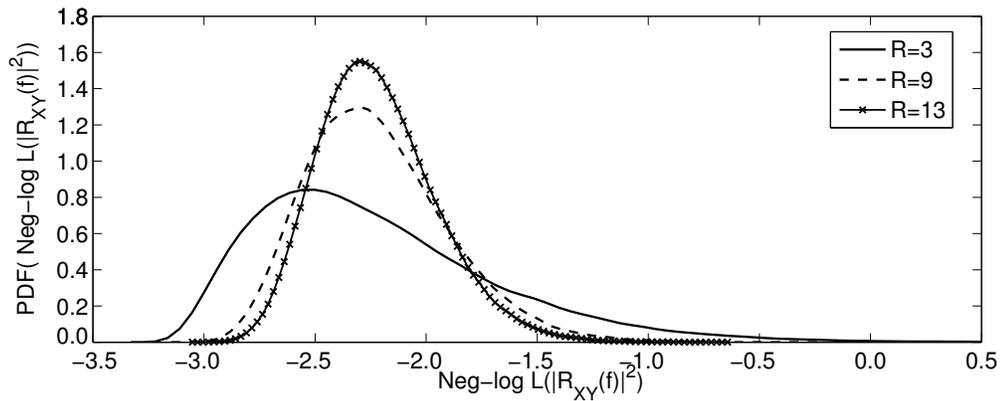
Figure 5.2: Distribution of negative log-likelihood of coherence, repre-sented by the solid-line at $R = 3$, dashed-line at $R = 9$ and crossed-line at $R = 13$.

The estimated probability density function of the negative log-likelihood $(-\log \mathcal{L}(u))$ of coherences is independent of frequencies, that is the distribution at the same size of two different group of frequencies is identical. Figure (5.3) illustrates the distribution of $(-\log \mathcal{L})$ for two bands of frequencies, both of size $R = 3$, where the first band, consisting of the frequencies $R_1 = [3, 4, 5]$, and the second band consisting of the frequencies $R_2 = [10, 11, 12]$, for $L = 25$, corresponding to 25 seconds of data.
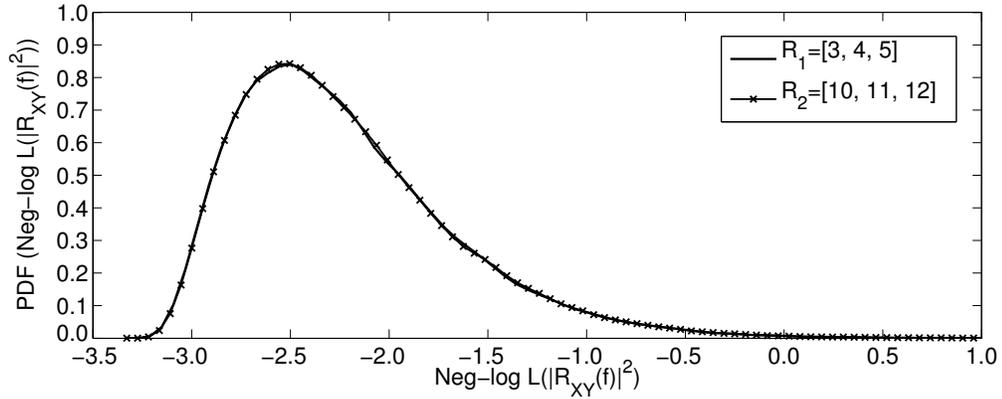
Figure 5.3: Distribution of negative log-likelihood of coherence for the data sample $T = 25$ seconds, represented by the solid-line for $R1 = [3, 4, 5]$ and crossed-line for $R2 = [10, 11, 12]$.

Note that the previous bands of frequencies were chosen to be well separated. The estimated probability density function of the negative log-likelihood $(-\log \mathcal{L}(u))$ of coherences is also identical for equal-sized contiguous bands of frequencies. Figure (5.4) illustrates the distribution of the negative log-likelihood $(-\log \mathcal{L}(u))$ of coherences for the two adjacent bands of frequencies, each of size $R = 3$, where the first band involves the frequencies $R_1 = [7, 8, 9]$, and, the second band involves the frequencies $R_2 = [10, 11, 12]$, for the data sample $T = 25$ seconds.
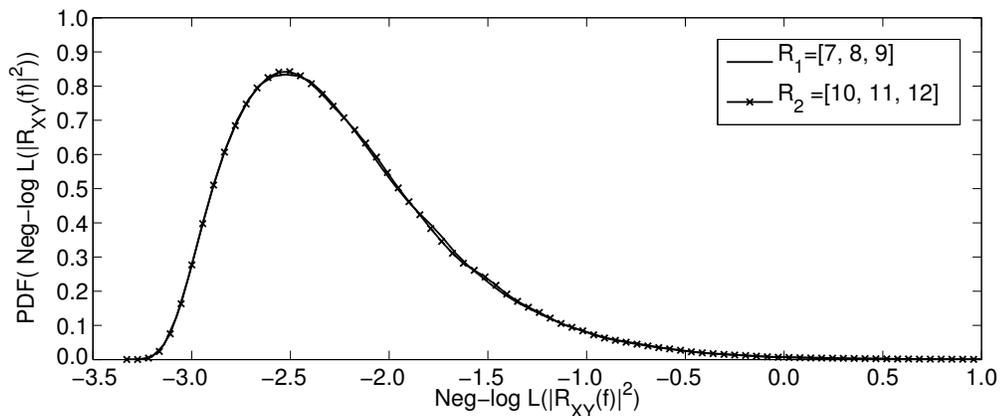

Figure 5.4: Distribution of negative log-likelihood of coherence for the data sample $T = 25$ seconds, represented by the solid-line for $R1 = [7, 8, 9]$ and crossed-line for $R2 = [10, 11, 12]$.

Furthermore, this result is also satisfied by the distribution of the negative log-likelihood $(-\log \mathcal{L}(u))$ of coherences for overlapped equal-sized bands of frequencies. Figure (5.5) displays the distributions of the negative log-likelihood $(-\log \mathcal{L}(u))$ of coherences for the

two overlapped bands of frequencies, each of size $R = 3$, where the first band consists of the frequencies $R_1 = [3, 4, 5]$, and the second band consists of the frequencies $R_2 = [5, 6, 7]$, when the number of sub-samples is $L = 40$, corresponding to 40 seconds of data.
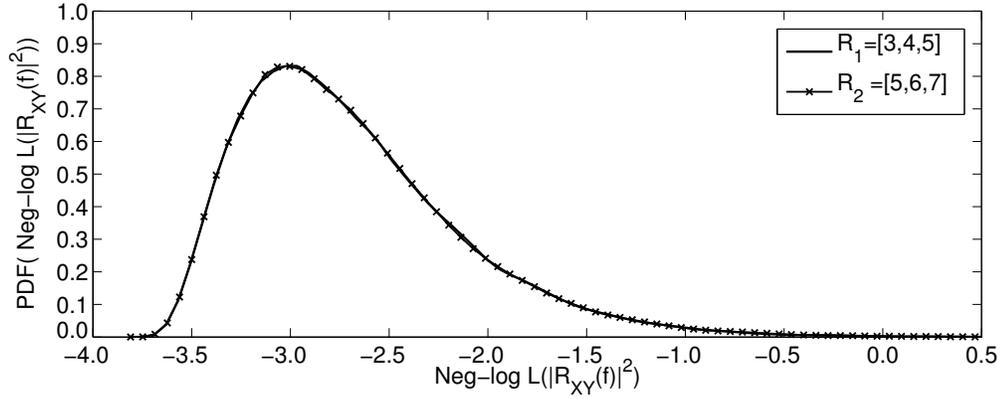

Figure 5.5: Distribution of negative log-likelihood of coherence for the data sample $T = 40$ seconds, represented by the solid-line for $R1 = [3, 4, 5]$ and crossed-line for $R2 = [5, 6, 7]$.

The distribution of the negative log-likelihood of coherences becomes less diffuse as more data becomes available. This effect is illustrated in figure (5.6)
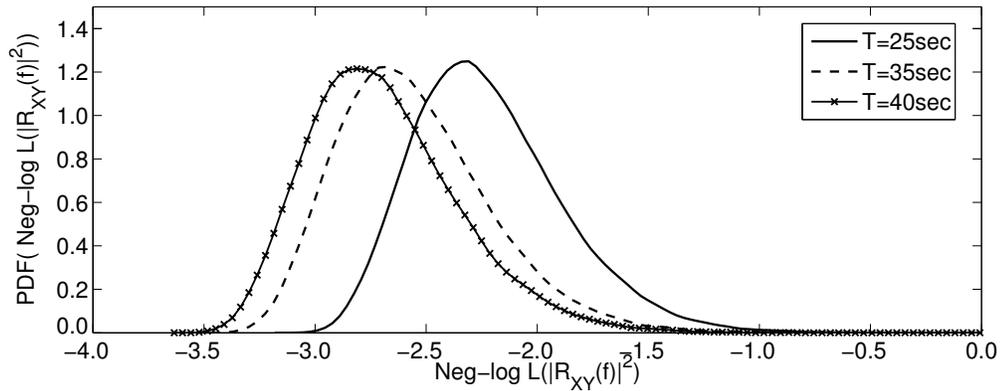

Figure 5.6: Distribution of negative log-likelihood depends on the volume of data for $N = 1000$. The solid-line corresponds to $L = 25$, the dashed-line to $L = 35$ and the crossed-line to $L = 40$.

In this analysis, we will use a one-sided statistical test of the null hypothesis. Given a negative log-likelihood function $(-\log \mathcal{L})$ estimated from L samples, a normalised random variable is constructed by subtracting the mean value for the given band of frequencies and dividing this difference by the standard deviation for bands of that size. Specifically,

the normalised random deviate is

$$Z = \frac{-\log \mathcal{L} - \left(-\log(L-1) + \frac{L-2}{L-1}\right)}{\frac{L-2}{L-1}} \times \sqrt{R} \qquad (5.16)$$

which is asymptotically distributed N(0,1) as $L \to \infty$ and as $R \to \infty$.

5.1.3 Confidence bounds

Confidence bounds for the 1%, 2.5%, 5% and 10% confidence intervals are constructed from simulated data by ordering the values of simulated negative log-likelihood in ascending order. The critical values for Z are displayed in tables (5.1) and (5.2), and plotted in figure (5.7) for 10%; in tables (5.3) and (5.4) and plotted in figure (5.8) for 5%; in tables (5.5) and (5.6) and plotted in figure (5.9) for 2.5%; and in tables (5.7) and (5.8) and plotted in figure (5.10) for 1%. From these tables, it is clear that critical values decrease as the volume of available data increases.

Sample size L

Band size R	5	10	15	20	25	30
3	1.342834	1.341998	1.337456	1.340030	1.342699	1.341123
4	1.341676	1.336733	1.340778	1.337048	1.337156	1.336827
5	1.337838	1.342567	1.335269	1.337006	1.338702	1.339840
6	1.335826	1.334253	1.338740	1.335731	1.333924	1.335803
7	1.339838	1.339042	1.338117	1.336204	1.333018	1.332716
8	1.331561	1.334763	1.333724	1.333022	1.333312	1.335280
9	1.327653	1.329546	1.329784	1.329891	1.330039	1.327031
10	1.337362	1.332822	1.327778	1.331017	1.330574	1.325646
11	1.330229	1.330303	1.328423	1.325019	1.328107	1.329091
12	1.328739	1.323474	1.325460	1.325809	1.328876	1.328557
13	1.325830	1.325288	1.326166	1.323608	1.322406	1.325656

Table 5.1: Critical values for the confidence bound level at 10% significance
for samples $L = 5, \cdots, 30$, and frequency bands $R = 3, 4, \cdots 13$.

Sample size L

Band size R	35	40	45	50	55	60
3	1.334961	1.336417	1.340235	1.341048	1.337544	1.340419
4	1.336244	1.335669	1.338019	1.335678	1.336491	1.333835
5	1.336829	1.339126	1.340525	1.341660	1.339817	1.336321
6	1.335620	1.335661	1.338113	1.337660	1.334820	1.337346
7	1.335077	1.336282	1.337177	1.336792	1.338362	1.334823
8	1.333487	1.336034	1.330677	1.328573	1.330327	1.333293
9	1.325812	1.327081	1.330947	1.327166	1.331452	1.326266
10	1.327333	1.330052	1.332908	1.332495	1.330539	1.330708
11	1.328769	1.325493	1.327090	1.330719	1.329691	1.328311
12	1.327554	1.331100	1.327862	1.327726	1.327440	1.328344
13	1.326207	1.323766	1.325723	1.324008	1.324640	1.326435

Table 5.2: Critical values for the confidence bound level at 10% significance for samples $L = 35, \cdots, 60$, and frequency bands $R = 3, 4, \cdots 13$.
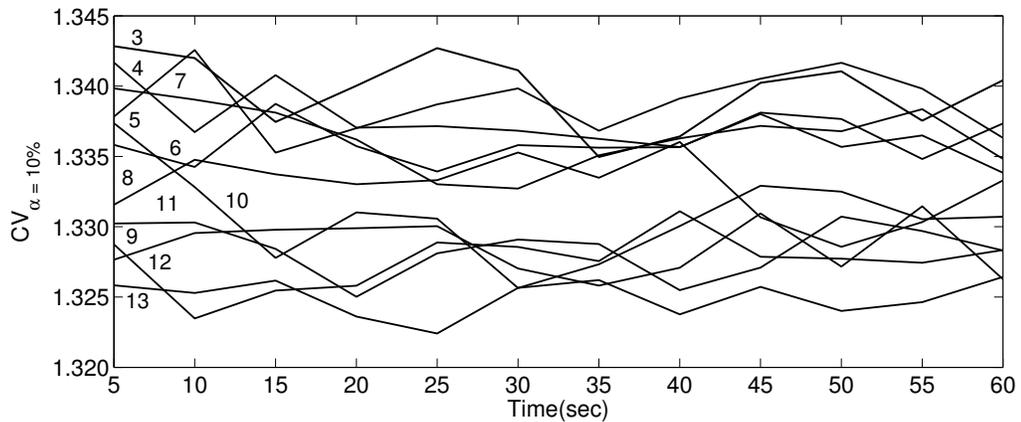
Figure 5.7: Confidence intervals of 10% level of significance for the negative log-likelihood of coherences indexed by R, the size of frequency bands.

Sample size L

Band size R	5	10	15	20	25	30
3	1.905980	1.898001	1.899185	1.903682	1.901794	1.908924
4	1.877294	1.875737	1.878117	1.870973	1.872840	1.873265
5	1.853368	1.858847	1.854459	1.856493	1.854886	1.858248
6	1.844127	1.843433	1.841972	1.842051	1.836630	1.839643
7	1.836248	1.835600	1.832545	1.830744	1.833717	1.830251
8	1.818080	1.824916	1.818552	1.822887	1.818487	1.819648
9	1.810317	1.812138	1.809030	1.812371	1.811600	1.805118
10	1.809714	1.803434	1.805472	1.803989	1.808063	1.800398
11	1.798983	1.800827	1.798454	1.796276	1.793796	1.799518
12	1.790979	1.786701	1.790726	1.788147	1.793443	1.792843
13	1.787403	1.785392	1.787051	1.787481	1.787910	1.783824

Table 5.3: Critical values for the confidence bound of 5% level of significance for samples $L = 5, \cdots, 30$, and frequency bands $R = 3, 4, \cdots, 13$

Sample size L

Band size R	35	40	45	50	55	60
3	1.899715	1.900456	1.900873	1.902032	1.902768	1.903178
4	1.874695	1.869926	1.876958	1.870658	1.875474	1.868643
5	1.853686	1.856863	1.859390	1.859509	1.859554	1.857693
6	1.839876	1.841897	1.839951	1.841877	1.839719	1.844579
7	1.833153	1.833609	1.833819	1.835697	1.832430	1.830455
8	1.822732	1.821378	1.815450	1.813553	1.814434	1.817297
9	1.803982	1.807732	1.809689	1.807973	1.805669	1.805319
10	1.803029	1.805926	1.808030	1.804406	1.802083	1.803864
11	1.798549	1.794394	1.796393	1.797390	1.796998	1.796993
12	1.793475	1.794226	1.791518	1.793007	1.794719	1.791227
13	1.787103	1.785732	1.788485	1.786476	1.785866	1.787137

Table 5.4: Critical values for the confidence bound of 5% level of significance for samples $L = 35, \cdots, 60$, and frequency bands $R = 3, 4, \cdots, 13$.
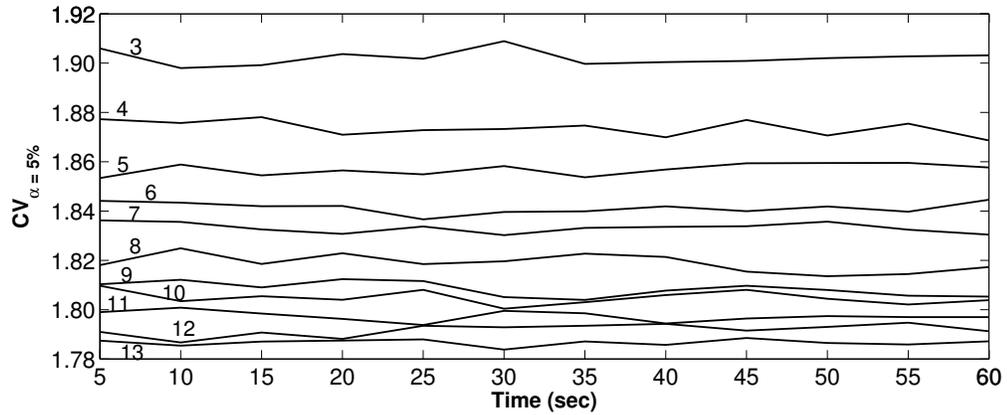
Figure 5.8: Confidence intervals of 5% level of significance for the negative log-likelihood of coherences indexed by R, the size of the frequency band.

<div align="center">Sample size L</div>

Band size R	5	10	15	20	25	30
3	2.441731	2.433056	2.435632	2.436052	2.434719	2.446202
4	2.384801	2.385301	2.387818	2.377896	2.378743	2.376821
5	2.344527	2.350677	2.338193	2.341252	2.341899	2.339987
6	2.309318	2.311871	2.312939	2.315495	2.307185	2.310784
7	2.297646	2.297382	2.291266	2.293322	2.294826	2.286953
8	2.260710	2.277594	2.270758	2.271683	2.263686	2.264133
9	2.256689	2.256112	2.246502	2.254222	2.252756	2.249213
10	2.245126	2.239773	2.242619	2.240027	2.243679	2.240495
11	2.230312	2.232448	2.225071	2.226759	2.222060	2.223289
12	2.217937	2.214207	2.215455	2.211415	2.214492	2.219919
13	2.208729	2.207789	2.209355	2.209205	2.207751	2.206530

Table 5.5: Critical values for the confidence bound of 2.5% level of significance for samples $L = 5, \cdots, 30$, and frequency bands $R = 3, 4, \cdots, 13$.

Sample size L

Band size R	35	40	45	50	55	60
3	2.439525	2.436577	2.438974	2.440095	2.442775	2.444274
4	2.379444	2.374200	2.380426	2.378700	2.379901	2.374239
5	2.340660	2.342443	2.342139	2.342213	2.344429	2.343432
6	2.307397	2.309294	2.306780	2.313340	2.309141	2.314306
7	2.290272	2.298239	2.292143	2.292333	2.295721	2.286555
8	2.274322	2.273548	2.269730	2.264676	2.268676	2.265840
9	2.248715	2.246421	2.251407	2.252495	2.244750	2.243727
10	2.237924	2.237184	2.245066	2.236115	2.238157	2.236057
11	2.221398	2.229195	2.227047	2.228148	2.227234	2.230487
12	2.217560	2.217978	2.216824	2.220419	2.217499	2.215059
13	2.206431	2.202549	2.210191	2.206447	2.209440	2.212456

Table 5.6: Critical values for the confidence bound of 2.5% level of significance for samples $L = 35, \cdots, 60$, and frequency bands $R = 3, 4, \cdots, 13$.
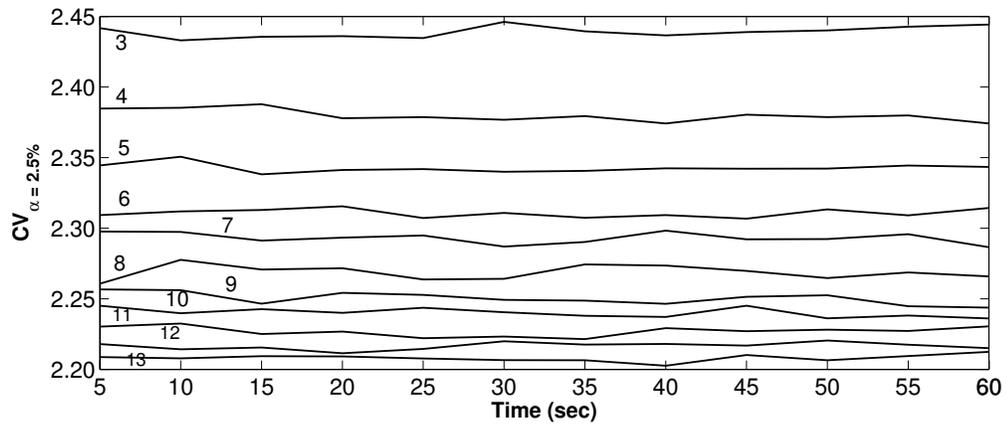

Figure 5.9: Confidence intervals of 2.5% level of significance for the negative log-likelihood of coherences indexed by R, the size of frequency bands.

Sample size L

Band size R	5	10	15	20	25	30
3	3.125212	3.107784	3.104866	3.123241	3.129997	3.124436
4	3.017014	3.026843	3.029781	3.018637	3.026708	3.020331
5	2.955838	2.955127	2.952360	2.940477	2.953851	2.942908
6	2.905423	2.894407	2.900126	2.910242	2.898123	2.900871
7	2.861484	2.872921	2.868190	2.863554	2.865099	2.862360
8	2.822964	2.833855	2.822102	2.829362	2.815902	2.818667
9	2.801921	2.807405	2.792819	2.810156	2.807732	2.798323
10	2.783738	2.784485	2.775517	2.781787	2.781244	2.769433
11	2.757280	2.765712	2.755530	2.753106	2.755636	2.753675
12	2.737436	2.732164	2.737552	2.731446	2.733644	2.732794
13	2.729947	2.724100	2.722570	2.723601	2.719365	2.719585

Table 5.7: Critical values for the confidence bound of 1% level of signifi-cance for samples $L = 5, \cdots, 30$, and frequency bands $R = 3, 4, \cdots, 13$.

Sample size L

Band size R	35	40	45	50	55	60
3	3.129315	3.115726	3.124341	3.119534	3.114501	3.121020
4	3.018749	3.012121	3.013897	3.015230	3.034061	3.009235
5	2.957804	2.954701	2.945864	2.950818	2.952831	2.948743
6	2.897422	2.894139	2.899168	2.901543	2.900022	2.895699
7	2.866115	2.865815	2.860510	2.859838	2.863527	2.859399
8	2.838740	2.822902	2.824445	2.817581	2.830935	2.814708
9	2.795475	2.790769	2.798602	2.795846	2.788758	2.795515
10	2.776942	2.772330	2.785609	2.775062	2.776470	2.773983
11	2.747478	2.758826	2.752427	2.753554	2.758709	2.757996
12	2.736303	2.744082	2.737092	2.735006	2.735639	2.731828
13	2.724202	2.721043	2.728733	2.726191	2.727522	2.720905

Table 5.8: Critical values for the confidence bound of 1% level of signifi-cance for samples $L = 35, \cdots, 60$, and frequency bands $R = 3, 4, \cdots, 13$.
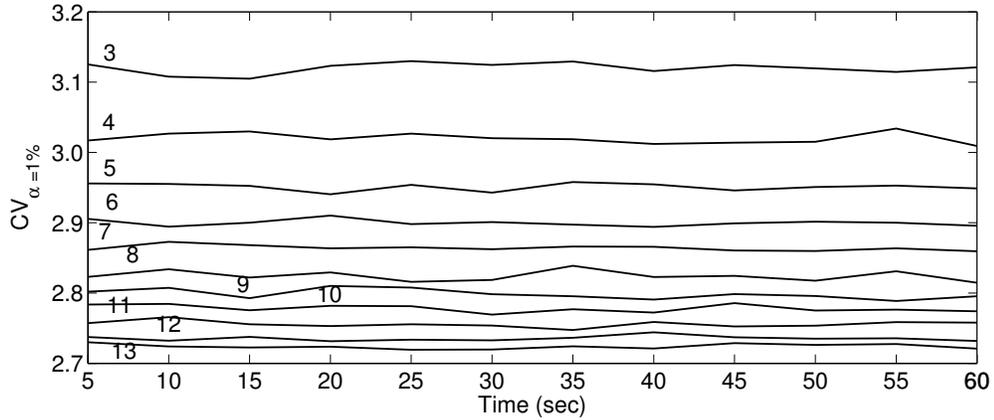
Figure 5.10: Confidence intervals of 1% level ofsignificance for the negative log-likelihood of coherences indexed by R, the size of frequency bands.

5.2 Conclusion

In this chapter, we proposed the use of the likelihood function as a measure of the independence between two samples in a specified band of frequencies. The basis of the algorithm is the analytical expression for the probability density function of coherence at any frequency, as provided by Goodman ([1965]) and Brillinger ([1981]), when it is known *a priori* that the processes are uncorrelated at that frequency. The difficulty in practice is that the analysis must be performed on samples of data; indeed for experimental reasons these samples may be collected over intervals of relatively short duration. Sequential measures of coherences calculated from such samples may be expected to be correlated and this effect may be compounded by the need to search for correlations across bands containing small numbers of consecutive frequencies. These effects conspire to complicate the identification of the information-carrying bands of frequencies. Under such circumstances, the distribution of any test statistic can only be established by extensive simulation, and the simulated data can be used to identify confidence bounds. This chapter has demonstrated how this procedure is achieved in the case of a test statistic based on the negative log-likelihood of a sample of coherences.

Chapter 6

Data Analysis

The EEG data set used in this analysis has been provided by King Faisal Hospital and Research Center KFHRC [Jeddah, Saudi Arabia. In KFHRC, all experiments are performed with ethical approval, and the patients (or their representative) give the informed consent for the procedures to be undertaken]. These data belong to an epileptic patient who suffers from focal epilepsy. The EEG recordings were collected from 64-electrodes placed on the scalp according to the International 10-10 system and the sampling rate was 200 Hz. The cut-off frequencies digital filters were set at 70 Hz and 1 Hz for the high-pass and low-pass respectively, and the power line noise was reduced using a 60 Hz notch filter.

In order to facilitate the investigation of the underlying data set, we have taken a subset of the channels (or signals), on the basis that signals collected from the same geographical region of the brain convey approximately the same information and, consequently, are equivalent functionally. As elaborated in Chapter 2, the brain is essentially divided into two distinct regions, namely the right and left hemispheres, each of which involves four main lobes: the frontal, the parietal, the occipital and the temporal lobes. Each hemisphere controls the contralateral side of the body by receiving information from different parts of the body, processing this information and then taking the convenient response among these lobes .

Since the subject whose EEG being analysed is right-handed, we will restrict the investigation to the left hemisphere only. Specifically, we will choose a representative signal from each lobe, for example channel F5 for the frontal lobe, C5 for the central area which lies between the frontal and parietal lobes, P5 for the parietal lobe, O1 for the occipital lobe and T9 for the temporal lobe. For mathematical simplicity these channels will be

renamed to hold $\{X_1, X_2, X_3, X_4, X_5\}$ for F5, C5, P5, O1 and T9 respectively. These signals will be investigated spectrally using both a parametric approach, whereby the data are fitted to the multivariate autoregressive model of order (2), since the MVAR parameters are significantly zero for the order 3, and a non-parametric approach, such as the method of averaging across periodograms of contiguous sections from single records, the smoothed periodogram method, and the multi-taper method or the Slepian sequence approach. Thereafter the comparison will be made between these methods.

We first prepare the data by subtracting the ensemble mean from the raw data to ensure the stationarity condition of the signals before being analysed. Then we investigate the autocorrelation coefficients (ACF) of the chosen signals, which provide a measure of the interdependence between the components of the signal itself in the time domain using equation (2.20). To decide whether or not the signal is significantly decaying across time, or in other words, whether the interdependence between its components is decreasing or not, we apply the statistical test under the null hypothesis that the signal's components are %95 uncorrelated, *i.e*, $\rho_{XX} = 0$. This statistical test is then examined with respect to Fisher's transformed variable of the estimated correlation, defined in equation (2.21), since it is distributed normally with a mean value and constant variance, calculated from the formulae (2.22) and (2.23), respectively. Thus, the confidence bounds of the independent statistical test is computed with respect to the Fisher variable z using the equation (2.24). Consequently, all values of the estimated correlation lie within the confidence bounds, construct the acceptance region for the null hypothesis. While the values of correlation lying outside the boundaries of the confidence interval contribute to reject the null hypothesis at level $\alpha = 5\%$ of significance. The maximum number of lags used in this calculation is $p = 2600$, where the whole length of the signal is $T = 13000$ milliseconds. Hence , the results of this statistical test exhibit that the signals are strongly autocorrelated across time. The ACF of the signals X_1 and X_2 are illustrated in figures (6.1:a) and (6.1:b), where the oscillating curve represents the ACF and the two horizontal lines denote the confidence interval limits at level of significance 5%,
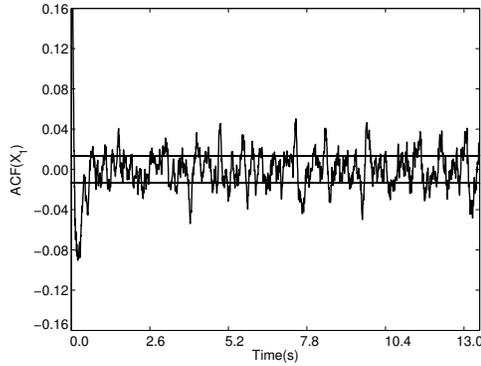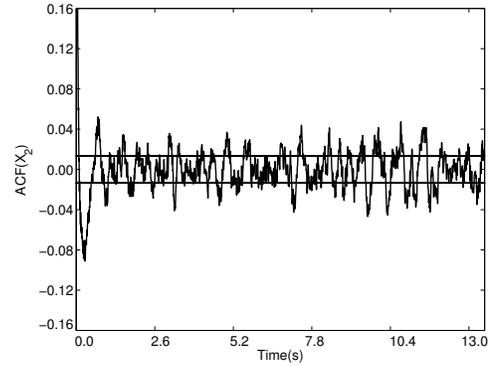
(a) The ACF of the signal X_1 (b) The ACF of the signal X_2

Figure 6.1: (a): The ACF of the signal X_1, (b): the ACF of the signal X_2. The two straight lines represent the confidence bounds of $\alpha = 5\%$ level of significance, where the low line is $x = -0.017$ and the upper line is $x = 0.017$.

The autocorrelation functions of the signals X_3 and X_4 are plotted in figures (6.2:a) and (6.2:b), respectively.
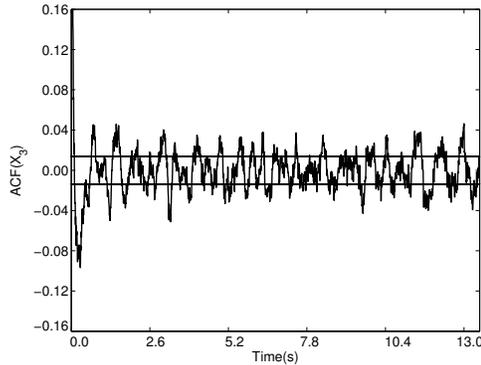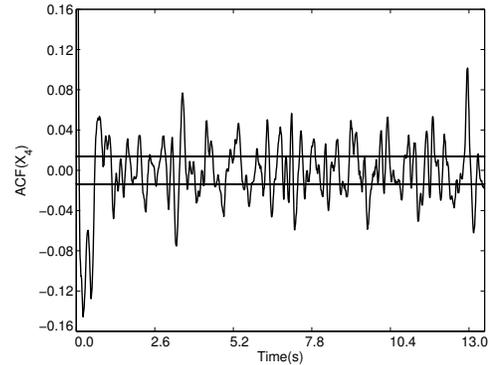

(a) The ACF of the signal X_3 (b) The ACF of the signal X_4

Figure 6.2: (a): The ACF of the signal X_3, (b): the ACF of the signal X_4. The two straight lines represent the confidence bounds of $\alpha = 5\%$ level of significance, where the low line is $x = -0.017$ and the upper line is $x = 0.017$.

The interdependence relationship between the components of the signal X_5 is illustrated in figure (6.3)
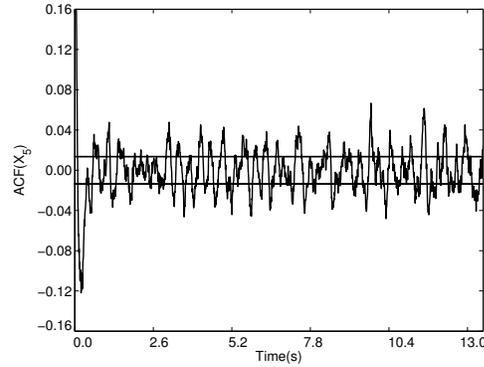
Figure 6.3: The ACF of the signal X_5 (oscillated curve) and the two straight lines represent the confidence bounds of $\alpha = 5\%$ level of significance, where the low line is $x = -0.017$ and the upper line is $x = 0.017$.

These figures show that the hypothesis of independence relationship between the signal's components is rejected for all signals, since there are statistically significant correlation values that lie outside the confidence interval of independence $[\,-0.017\,,\,0.017]$, which implies that the signals are not decaying. Although the autocorrelation function appears weak sometimes as in signal X_5, but still statistically significant.

6.1 Multivariate autoregressive model analysis (MVAR)

To perform the spectral analysis of the processes $\{X_1, X_2, X_3, X_4, X_5\}$, we model these signals by a multivariate autoregressive model of order $(p = 2)$. The coefficients A_j's of the multivariate autoregressive system are estimated from equation (4.54) using the maximum likelihood method, described in Chapter 4. Once the multivariate autoregressive coefficients are obtained, the Fast Fourier Transform is applied to the multivariate autoregressive model, and the transfer matrix $\left[H = B^{-1} = \left(I - \sum_{j=1}^{p} A_j \, e^{-2i\pi f j \Delta t} \right)^{-1} \right]$, where the identity matrix I and the coefficients matrices A_j are of dimension $M \times M$ with $(M = 5)$ in this case, is calculated.

The power spectrum is an important measure that is used widely in spectral analysis to capture the frequencies which correspond to the fluctuations of a signal in its time domain. The estimated power spectra of the signals X_1, X_2, X_3, X_4 and X_5 are calculated using the formula (4.80) with respect to the frequency band $f \in [1, 70]$ Hz. The logarithm to base 10 of the estimated power spectra of these signals are then plotted versus the frequency components in figure (6.4). Figure (6.4:a) depicts the estimated spectrum of the signal

X_1, figure (6.4:b) illustrates the estimated power spectrum of the signal X_2, the estimated power spectrum of the channel X_3 is plotted in figure (6.4:c), and finally, figure (6.4:d) shows the estimated power spectrum of the channel X_4.
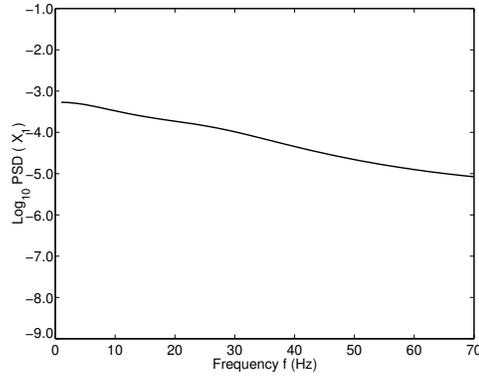
(a) The power spectrum of the signal X_1
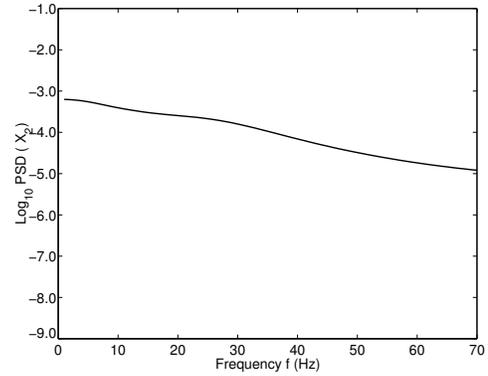
(b) The power spctrum of the signal X_2
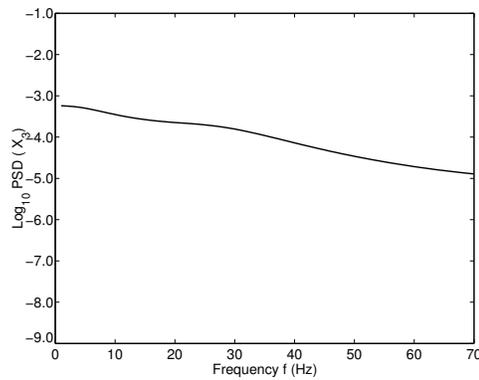
(c) The power spectrum of the signal X_3
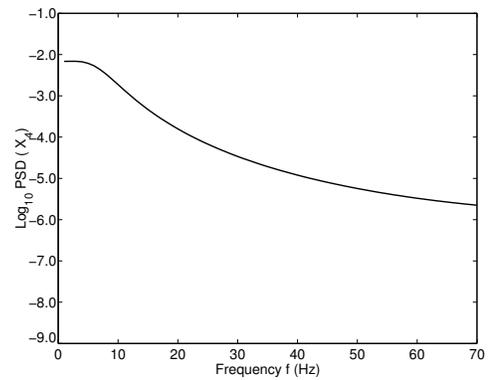
(d) The power spctrum of the signal X_4

Figure 6.4: The logarithm to base 10 of (a): the power spectrum of the signal X_1, (b): the power spectrum of the signal X_2, (c): the power spectrum of the signal X_3, and (d): the power spectrum of the signal X_4. The spectra plotted against the frequency which takes values between 1 Hz and 70 Hz.

The estimated power spectrum of the signal X_5, which describes the distribution of the power contained in the signal as a function of frequency, is displayed in terms of the logarithm to base 10 of the estimated power spectrum in figure (6.5).
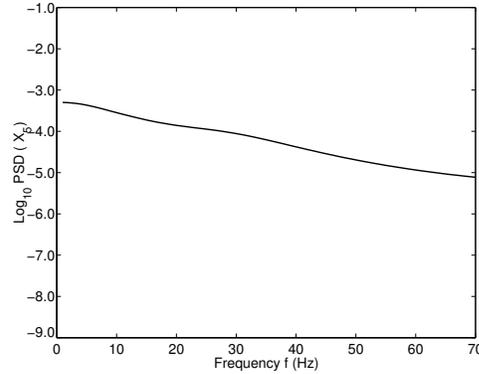
Figure 6.5: The logarithm to base 10 of the estimated power spectrum of the signal X_5 against the frequency components $f \in [1, 70]$ Hz.

From these figures of the estimated power spectra, produced by the multivariate autoregressive model of order 2, we note that this parametric spectral estimation procedure is not sensitive to the notch filter, because the effect of the notch filter is expected to appear as a spike at 60Hz. This means that the multivariate autoregressive model is only responsive to the observational properties that are consistent with its parameters features.

The direction of the information flow between signals among neuronal structures is determined by using the partial directed coherence, which is denoted by $|\pi_{i \leftarrow j}(f)|$. The partial directed coherence measurement $|\pi_{i \leftarrow j}(f)|$ describes the causal influence of the process j, on the process i at a specific frequency f, taking into account the influence of the process j on all the processes involved in the network being investigated. To decide whether the estimated values of the partial directed coherence between the components of the multivariate autoregressive model of order $(p = 2)$ are statistically significant or not, we examine the null hypothesis that $|\pi_{i \leftarrow j}(f)| = 0$, and use the significance level introduced by Schelter *et al.* [2005] and defined in equation (4.100) with $\alpha = 5\%$ level of significance. The graphical structure produced by the partial directed coherence is shown in figure (6.6).
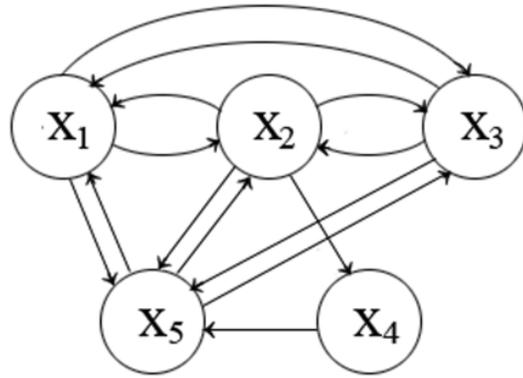
Figure 6.6: The directions of the causal influences between the signals in the neural structures

This diagram shows the direction of the interrelation between processes representing the main brain areas. It can be seen clearly that there are statistically significant causal influences of the frontal process X_1 on the central, parietal, and temporal processes X_2, X_3, X_5, respectively. That is, the values of the estimated partial directed coherence $|\pi_{2\leftarrow1}(f)|$, $|\pi_{3\leftarrow1}(f)|$ and $|\pi_{5\leftarrow1}(f)|$, are lying above the level of significance $\alpha = 5\%$, which represents the rejection region of the null hypothesis. Figure (6.7) illustrates the partial directed coherence from X_1 to X_2, denoted by the solid-line, and, the confidence bound of $\alpha = 5\%$ level of significance, denoted by the starred-line.
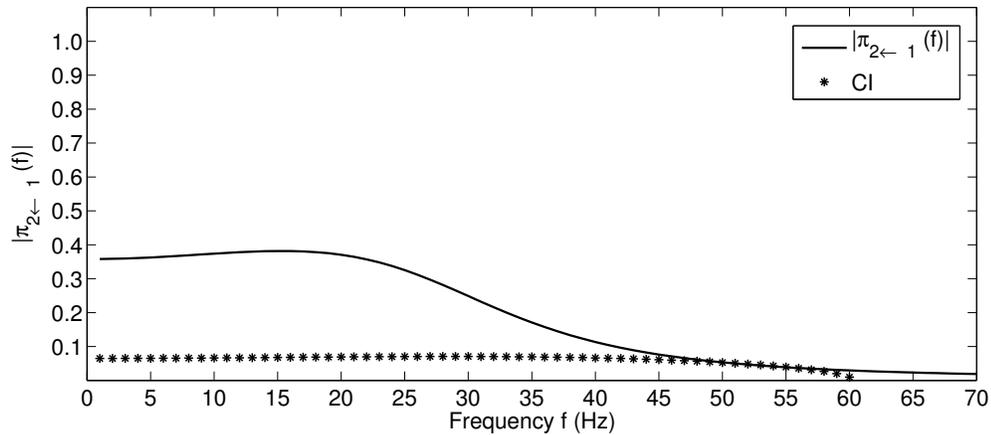

Figure 6.7: Partial directed coherence $|\pi_{2\leftarrow1}(f)|$ is denoted by the solid-line, and, the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

The causal influence of process X_1 on process X_3 at frequency f, is plotted in figure (6.8)
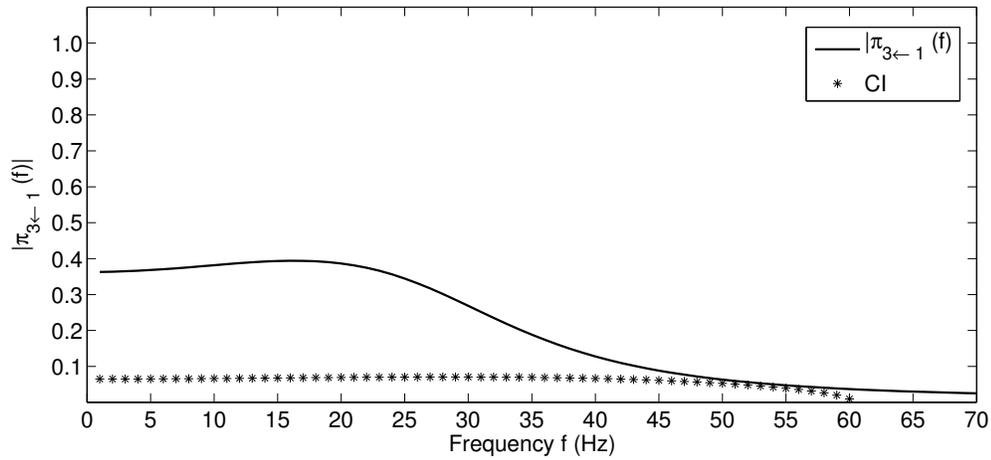
Figure 6.8: Partial directed coherence $|\pi_{3\leftarrow 1}(f)|$ is denoted by the solid-line, and, the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

The absolute value of the partial directed coherence is $|\pi_{5\leftarrow 1}(f)| \approx 0.5$ at frequency $f \in [1, 20)$ Hz, while this value gradually decreases and becomes smaller from 22Hz, onwards. Figure (6.9) illustrates the direct causal influence from X_1 to X_5, represented by the solid-line, and, the confidence bound of significance, plotted by the starred-line.
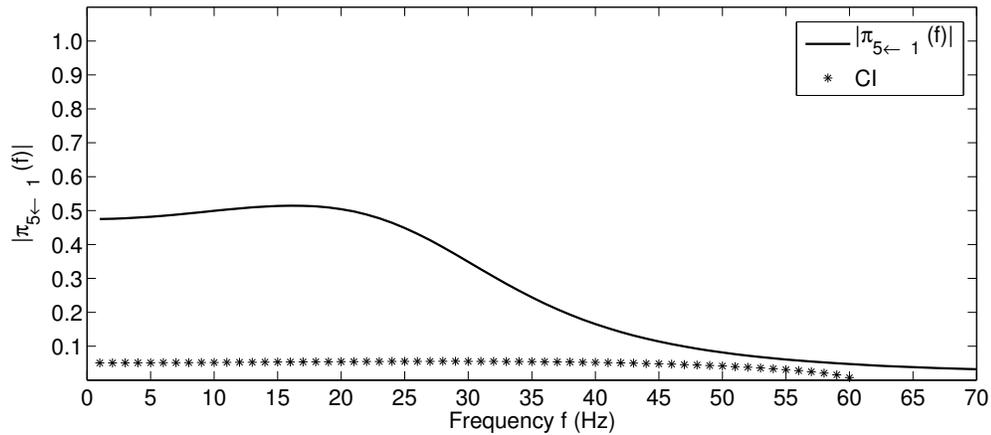

Figure 6.9: Partial directed coherence $|\pi_{5\leftarrow 1}(f)|$ is represented by the solid-line, and, the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

We can say generally that the historical information of process X_1 combined with the historical information of the processes X_2, X_3, X_5, collectively, contribute to improve the expectation of the present values of processes X_2, X_3 and X_5. Also, we can see clearly from figure (6.6) that the frontal process X_1 has no direct influence on the occipital process X_4.

This means that the estimated directed partial coherence $|\pi_{4\leftarrow1}(f)|$ is significantly zero. figure (6.10) illustrates that the partial directed coherence $|\pi_{4\leftarrow1}(f)|$, plotted in the solid-line and the confidence bound of $\alpha = \%5$ level of significance, marked by the starred-line are approximately identical.
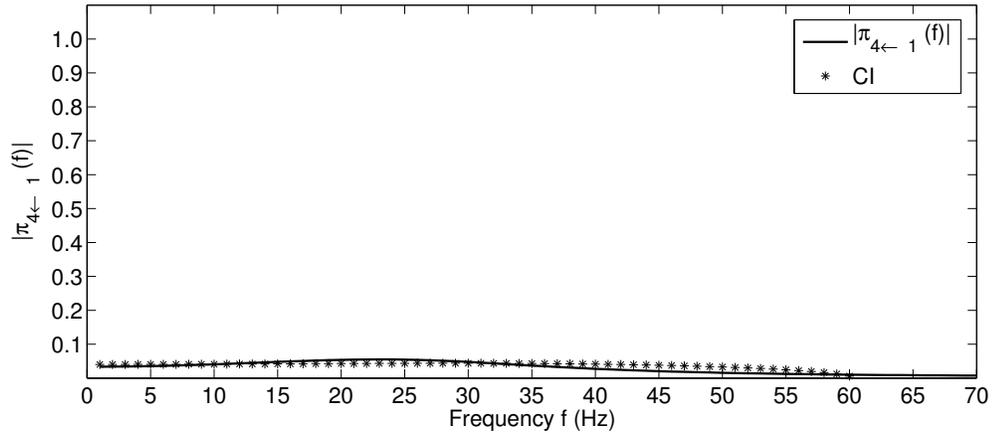
Figure 6.10: Partial directed coherence $|\pi_{4\leftarrow1}(f)|$ is represented by the solid line, and the confidence bound of the $\alpha = 5\%$ level of significance is marked by the starred line.

On the other hand, the direct casual influences from processes X_2, X_3 and X_5 to process X_1 are detected. That is, the values of the partial directed coherences $|\pi_{1\leftarrow2}(f)|$, $|\pi_{1\leftarrow3}(f)|$, and, $|\pi_{1\leftarrow5}(f)|$, are statistically significant. We note that, at the same points of frequency f, the values of the direct causal influence from process X_2 to process X_1, displayed in figure (6.11), is less than the values of the causal influence of process X_1 on process X_2, shown in figure (6.7).
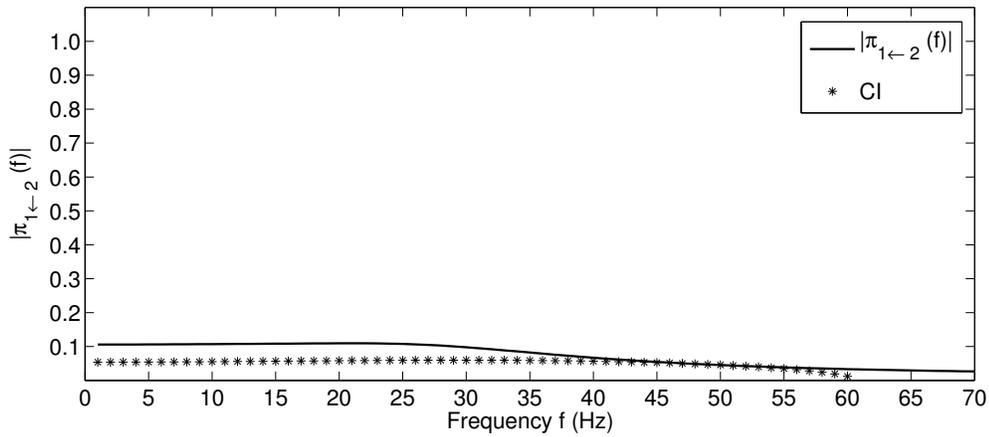
Figure 6.11: Partial directed coherence $|\pi_{1\leftarrow 2}(f)|$ is represented by the solid-line, and the confidence bound of $\alpha = 5\%$ level of significance is marked by the starred-line.

The direct causal influence of process X_3 on process X_1 is plotted by the solid-line, and the confidence bound of $\alpha = 5\%$ level of significance is marked by the starred-line in figure (6.12).
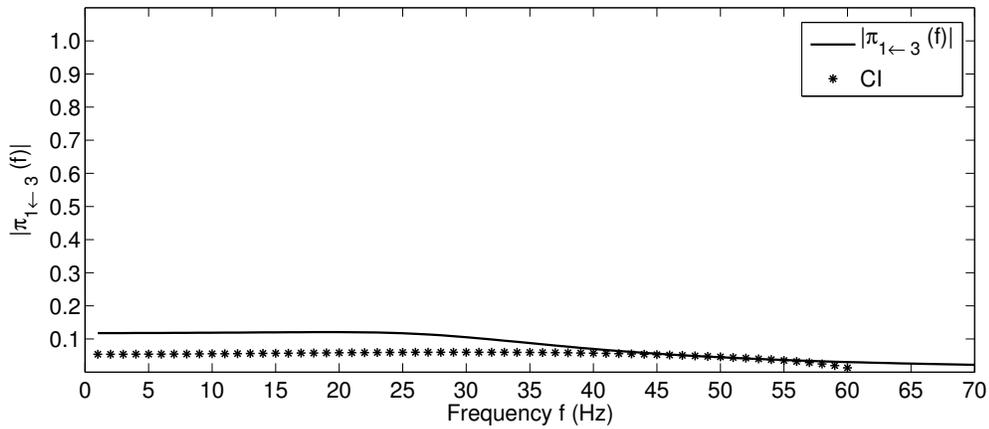
Figure 6.12: Partial directed coherence $|\pi_{1\leftarrow 3}(f)|$ is represented by the solid-line, and the confidence bound of $\alpha = 5\%$ level of significance is marked by the starred-line.

The values of the directed partial coherence $|\pi_{1\leftarrow 3}(f)|$ are statistically significant at frequencies $f \in [1, 38)$ Hz. Furthermore, the values of the partial directed coherence from process X_1 to process X_3 are larger than the values of the partial directed coherence from process X_3 to process X_1 at the same points of frequency. The direct causal influence of process X_5 on process X_1 is shown in figure (6.13), where the solid-line represents the

$|\pi_{1\leftarrow 5}(f)|$ and the starred-line denotes the confidence bound of $\alpha = 5\%$ level significance.
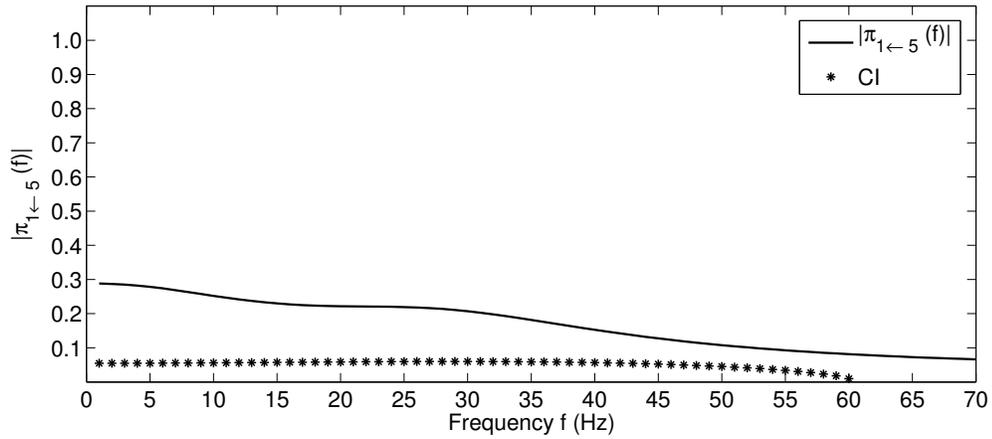
Figure 6.13: Partial directed coherence $|\pi_{1\leftarrow 5}(f)|$, represented by the solid-line, and the confidence bound of $\alpha = 5\%$ level of significance, marked by the starred-line.

Obviously, the partial directed coherences from process X_1 to processes X_2, X_3, and, X_5 are larger than the partial directed coherence of each process on process X_1 at the same points of frequency.

Now we will discuss the causal effects of process X_2 on processes X_3, X_4, and, X_5. Clearly, the graphical diagram (6.6) shows that there are reciprocal causal influences between processes X_2 and X_3, and between processes X_2 and X_5. This means that the partial directed coherences $|\pi_{3\leftarrow 2}(f)|$, $|\pi_{5\leftarrow 2}(f)|$, $|\pi_{2\leftarrow 3}(f)|$, and $|\pi_{2\leftarrow 5}(f)|$ are statistically significant. However, there is only one direct causal influence of process X_2 on process X_4. Figure (6.14) displays the causal influence of process X_2 on process X_3, denoted by the solid-line, the causal influence of process X_3 on the process X_2, which is denoted by the dashed-line and the significant confidence bound at level of significance $\alpha = 5\%$, indicated by the starred-line.
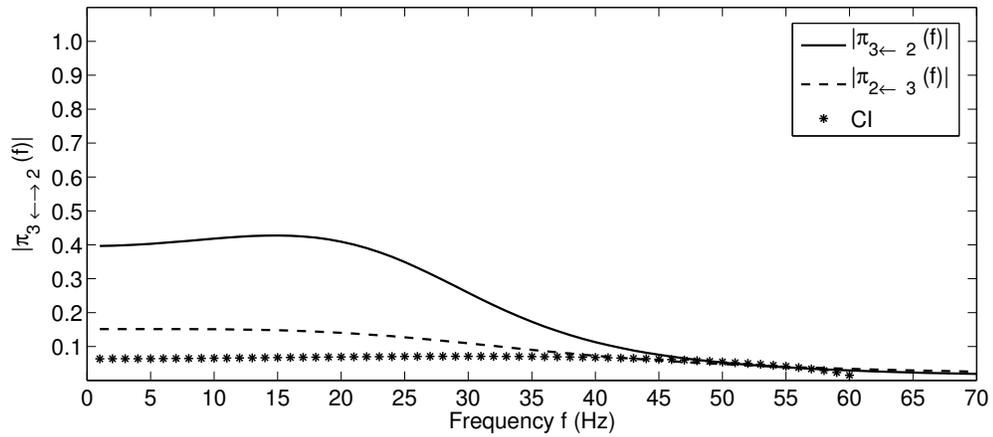
Figure 6.14: Partial directed coherence $|\pi_{3\leftarrow2}(f)|$ is plotted in the solid-line, the Partial directed coherence $|\pi_{2\leftarrow3}(f)|$ is represented by the dashed-line, and the confidence interval of $\alpha = 5\%$ level of significance is marked by the starred line.

From this figure, we can clearly see that process X_2 (or X_3) has a statistically significant causal influence on process X_3 (or X_2), taking into account the joint influences of process X_2 (or X_3) on the other processes in the current neural structure. Also, we note that the partial directed coherence from process X_2, which represents the central area of the brain, to process X_3, which indicates the parietal lobe of the brain, is larger than the partial directed coherence from process X_3 to process X_2 at the same points of frequency f. This means $|\pi_{3\leftarrow2}(f)| > |\pi_{2\leftarrow3}(f)|$, where $f \in [1, 45]$ Hz, and the partial directed coherence $|\pi_{2\leftarrow3}(f)|$ is entirely vanished at frequency $f \in [35, 70]$ Hz.

The partial directed coherences between processes X_2 and X_5 is displayed in figure (6.15), where the solid-line represents the $|\pi_{5\leftarrow2}(f)|$, the dashed-line denotes the $|\pi_{2\leftarrow5}(f)|$, and, the starred-line reprsents the confidence bound of $\alpha = 5\%$ level of significance. It is noticeable that there are two bands of frequency containing two different causal effects of process X_2 (or X_5) on process X_5 (or X_2). While the partial directed coherence $|\pi_{5\leftarrow2}(f)|$ is larger than the partial directed coherence of X_5 on X_2 within the band of frequency $f \in [1, 33)$ Hz, the causal effect of X_5 on X_2 becomes larger at frequency $f > 33$Hz.
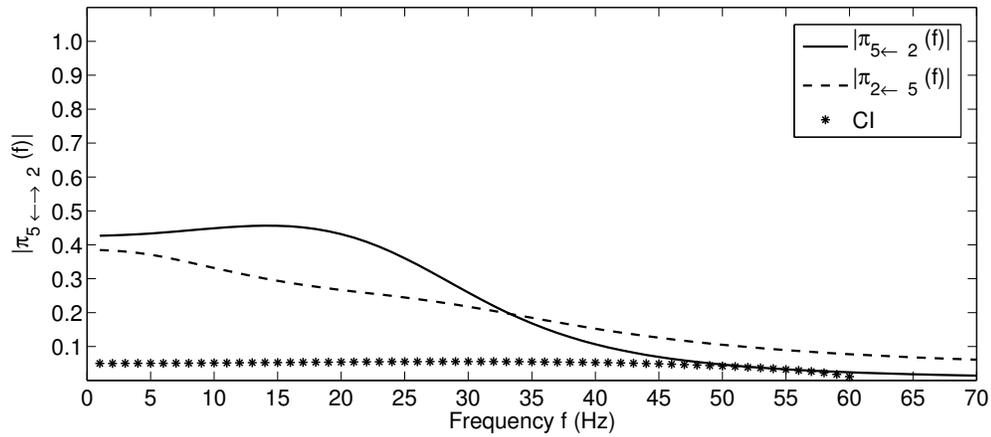
Figure 6.15: Partial directedcoherence $|5\pi_{\leftarrow 2}(f)|$ is plotted by the solid-line, the Partial directed coherence $|\pi_{2\leftarrow 5}(f)|$ is represented by the dashed-line, and the confidence interval of $\alpha = 5\%$ level of significance is marked by the starred line.

The direct causal influence of process X_2 on process X_4 is detected and illustrated in figure (6.16), where the solid-line indicates the $|\pi_{4\leftarrow 2}(f)|$ and the starred-line represents the confidence bound of $\alpha = 5\%$ level of significance.
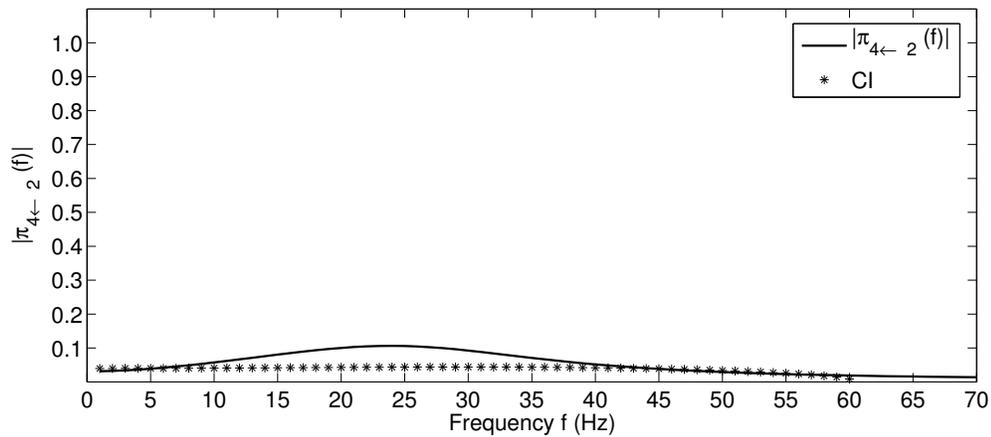
Figure 6.16: Partial directed coherence $|\pi_{4\leftarrow 2}(f)|$ is plotted by the solid line and the confidence interval of $\alpha = 5\%$ level of significance is represented by the starred-line.

Although the two direct causal influences are detected between the parietal lobe process X_3 and the temporal lobe process X_5 at frequency f, the strength of these causal influences are different. For instance, the causal effect of X_3 on X_5, taking into account the joint influences of process X_3 on the other processes in the neural structure at fre-

quency $f \in [1, 35)$, is stronger than the causal effect process X_5 on process X_3, considering the joint effects of process X_5 on the other processes corresponding to the same points of frequency. The values of the partial directed coherence $|\pi_{3\leftarrow5}(f)|$ are larger than the values of $|\pi_{5\leftarrow3}(f)|$ at frequency $f > 35$Hz and upwards. Figure (6.17) shows the direct causal influences between processes X_3 and X_5, where the partial directed coherence $|\pi_{5\leftarrow3}(f)|$ is represented by the solid-line, the partial directed coherence $|\pi_{3\leftarrow5}(f)|$ is denoted by the dashed-line, and, finally the confidence interval of $\alpha = 5\%$ level of significance is denoted by the starred-line.
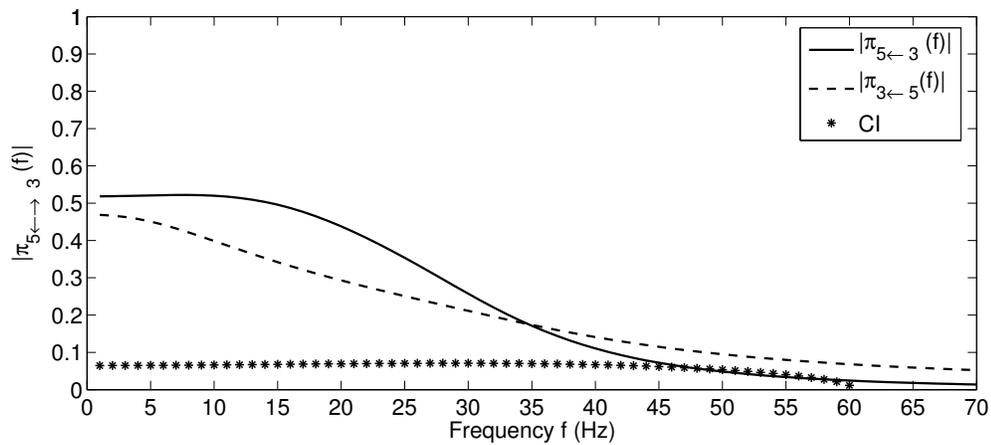
Figure 6.17: Partial directed coherence $|\pi_{5\leftarrow3}(f)|$ is represented by the solid-line, the partial directed coherence $|\pi_{3\leftarrow5}(f)|$ is dashed-line, and the confidence interval of $\alpha = 5\%$ level of significance is marked by the starred-line.

Finally, only one direct causal influence of the occipital lobe process X_4 on the temporal lobe process X_5 is detected. The partial directed coherence $|\pi_{5\leftarrow4}(f)|$ is statistically significant and is plotted in figure (6.18) by the solid-line, the dashed-line represents the values of $|\pi_{4\leftarrow5}(f)|$ and the starred-line denotes the confidence bound of $\alpha = 5\%$ level of significance.
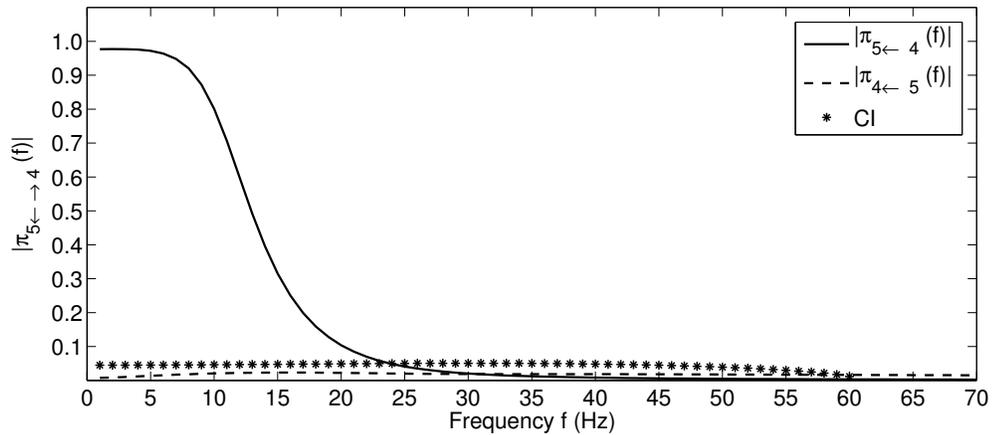
Figure 6.18: Partial directed coherence $|\pi_{5\leftarrow4}(f)|$ is represented by the solid-line, the dashed-line represents the $|\pi_{4\leftarrow5}(f)|$, and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

This figure shows that the causal influences between processes X_4 and X_5. The partial directed coherence $|\pi_{5\leftarrow4}(f)|$ takes its highest value (≈ 98) at frequency $f \in [1, 10)$, and then decreases dramatically towards zero at $f = 23$ Hz and onwards. The process X_5 has no direct causal effect on the process X_4. As can be seen from figure (6.18), almost all values of the partial directed coherence $|\pi_{4\leftarrow5}(f)|$, represented by the dashed-line, lie beneath the confidence bound, marked by the starred-line. However, the process X_4 possesses indirect causal influences on processes X_1, X_2, X_3 among process X_5. Similarly the processes X_1, X_3, X_5 have indirect causal effects on process X_4 via process X_2. That is, the processes X_1, X_3 and X_5 have direct causal effects on X_2, and, X_2 has a direct causal influence on X_4, consequently, these processes have indirect causal influences on the process X_4.

We will now investigate the correlation between the processes representing the various brain areas by using the measurement of coherence. As mentioned in Chapter 2, the coherence measurement is a function of frequency and its value varies between zero, which indicates no correlation between the processes, and the unit value, which means that the two processes are identical. The estimated coherence of the multivariate autoregressive model of order (p) is given in equation (4.83). The statistical significant estimated coherence of processes X_i and X_j is determined by testing the null hypothesis that these two processes are %95 not correlated at frequencies between 1 Hz and 70 Hz, that is $|R_{X_iX_j}(f)|^2 = 0$. The confidence bound with the $\alpha = 5\%$ level of significance proposed by Schelter et $al.$ [2005] and defined in equation (4.100) is applied.

In general, the significant estimated coherences are captured between the processes, which represent the different brain areas, except for the occipital lobe, which apparently is uncorrelated to any of those processes. The estimated coherence of process X_1 with process X_2 occurs mostly at frequency $f \in [1, 17]$ Hz, which involves the EEG's waves; δ-rhythm, which ranges between frequencies 1Hz and 3Hz and is thought to indicate the slow brain activities; and, θ-rhythm, which corresponds to the frequency band $[4, 8)$ Hz and reflects sleep and drowsiness in adults. The occurrence of a high theta wave indicates an abnormality, which is in this case is known to be epilepsy. Also, there is evidence of the existence of the α-wave within the frequency band $[8, 13)$ Hz, indicating a relaxed and mentally inactive awakeness. The β-rhythms, which usually occur at frequencies between 13Hz and 30Hz, and indicate the active brain activities, such as thinking, concentrating, and solving problems, are appeared with proportionally small magnitude almost reaching the value 0.2, at frequency ranging between 13Hz and 15Hz. Figure (6.19) illustrates the estimated coherence of process X_1 with process X_2, represented by the solid-line, the confidence bound of $\alpha = 5\%$ level of significance is marked by the starred-line.
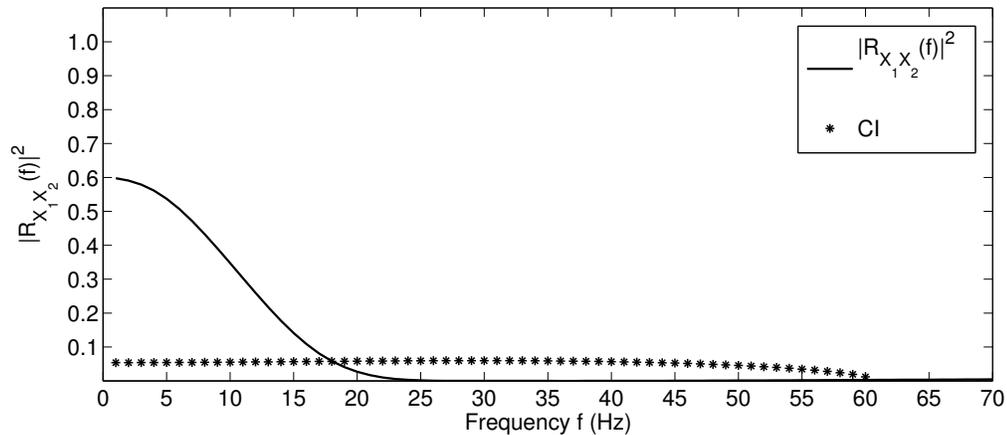
Figure 6.19: The estimated coherence function $|\hat{R}_{X_1 X_2}(f)|^2$ is plotted by the solid-line, and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

Also, the estimated coherence between the frontal representative process X_1 and the parietal representative process X_3, is statistically significant at frequencies between 1Hz and 17Hz, as can be seen clearly from figure (6.20), where the solid-line represents the estimated coherence $|\hat{R}_{X_1 X_3}(f)|^2$, and the starred-line denotes the confidence bound of $\alpha = 5\%$ level of significance.
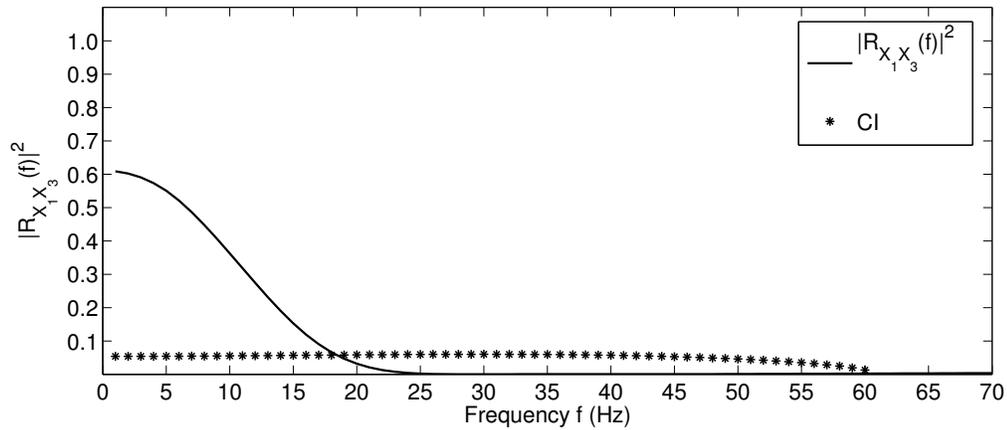
Figure 6.20: The estimated coherence $|\hat{R}_{X_1 X_3}(f)|^2$ is plotted by the solid-line, and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

Clearly, the two processes X_1 and X_3 are highly correlated at frequencies corresponding to the δ-wave, where the estimated coherence values almost hit the value 0.56 at frequencies $f \in (0.5, 4)$ Hz. Then, the estimated coherence decreases gradually within the band of frequency $[4, 8]$ Hz, which represents the θ-wave. Also, there is the existence of the EEG wave α in the band frequency $[8, 13)$ Hz and the beta wave at frequencies between 13Hz and 16Hz. The coherence between the two processes X_1 and X_5 is displayed in figure (6.21).
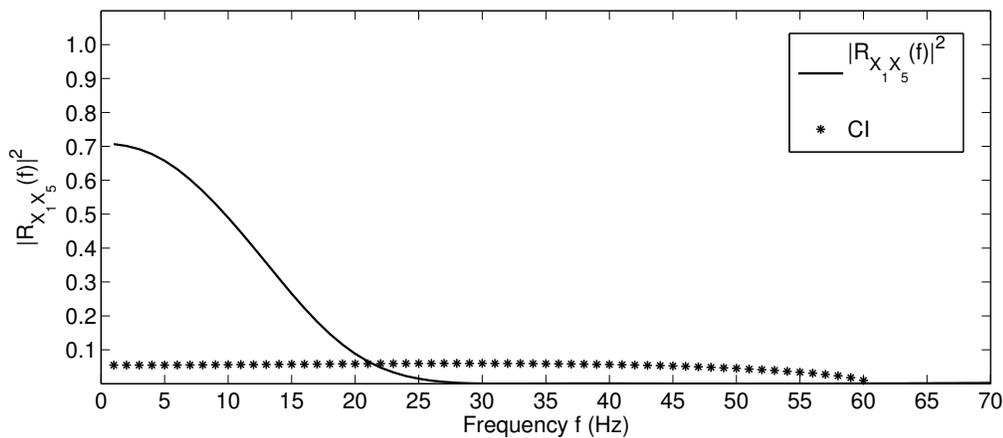
Figure 6.21: The estimated coherence $|\hat{R}_{X_1 X_5}(f)|^2$ is represented by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is marked by the starred-line.

where the solid-line denotes the estimated coherence of the process X_1 with the pro-

cess X_5 and the starred-line indicates the confidence bound of 5% level of significance. As can be seen from figure (6.21), the estimated coherence occurs mostly at points corresponding with frequencies in the interval $[1, 20)$ Hz. High correlation between these two processes happens within the δ-wave frequencies, $[1, 4)$ Hz, where the values of the coherence $|\hat{R}_{X_1 X_5}(f)|^2$ almost reach 0.7. Then, the coherence reduces to the value of 0.6 in the band of frequencies $[4, 8)$ Hz, which is the θ-wave range. Although the estimated coherence continuously decreases, the values of the estimated coherence remain strongly significant, which approximately range between 0.4 and 0.6 in the α-wave frequency band $[8, 13)$ Hz and between 0.1 and 0.4 in the β-wave frequency band $[13, 30)$ Hz. There are significant estimated coherences arising between the processes X_2 and X_3 in the frequency band $[1,16]$ Hz; and the processes X_2 and X_5 at frequency $f \in [1, 17]$ Hz; and between the two processes X_3 and X_5 in the frequency band $[1,18]$ Hz. According to the previous discussion, these intervals involve four types of EEG waves, each of which indicates a specific brain activity: δ-wave within the band $[1, 4)$ Hz; θ-wave in the frequency band $[4, 8)$ Hz; α-wave, which is ranging between 8Hz and 13Hz; and finally the β-wave, which is usually captured in the frequency band $[13, 30)$ Hz. Figure (6.22) illustrates the estimated coherence function of the process X_2 with the process X_3.
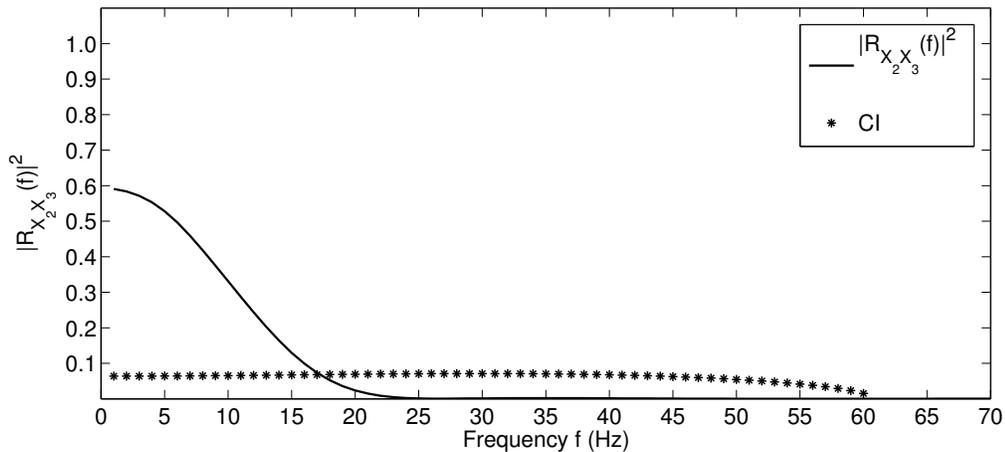

Figure 6.22: The estimated coherence $|\hat{R}_{X_2 X_3}(f)|^2$ is plotted by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is marked by the starred-line.

From this figure we can see clearly that the processes X_2 and X_3 appear strongly correlated within the δ-wave and θ-wave bands of frequency where the values of the estimated coherence range between 0.45 and 0.6. Thereafter, the estimated coherence reduces dra-

matically in the two following bands of frequency corresponding to alpha and beta waves until reaching the zero value.

The significant estimated coherence of the process X_2 with the process X_5 occurs within the frequency band $[1, 17)$ Hz. The estimated coherence function $|\hat{R}_{X_2 X_5}(f)|^2$ is plotted against frequencies f between 1Hz and 70Hz in figure (6.23), where the solid-line indicates the function $|\hat{R}_{X_2 X_5}(f)|^2$ and the confidence bound of 5% level of significance is marked by the starred-line.
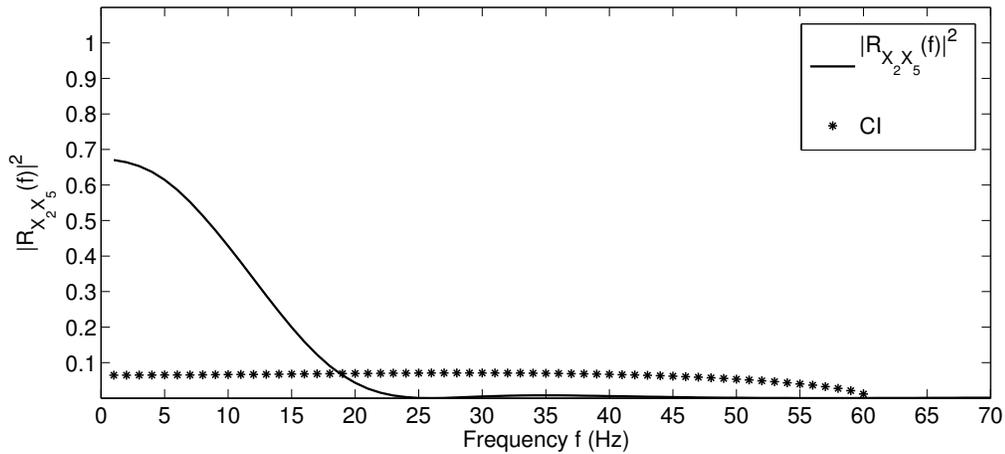
Figure 6.23: The estimated coherence function $|\hat{R}_{X_2 X_5}(f)|^2$ is represented by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is marked by the starred-line.

Again the strong correlation between the two processes occurs in the delta and theta bands of frequency $[1, 8)$ Hz, while the values of the estimated coherence decrease gradually in the alpha and beta bands of frequency. Similarly the estimated coherence $|\hat{R}_{X_3 X_5}(f)|^2$ presents a high correlation between processes X_3 and X_5 in the frequency band $[1, 4)$ Hz, where the coherence almost hits the value 0.7. This value of coherence reduces to 0.6 in the frequency interval $[4, 8)$ Hz, which corresponds to the theta wave. Thereafter, the strength of the correlation decreases in the α- and β- rhythms bands of frequency, as can be seen in figure (6.24), where the solid-line denotes the estimated coherence function $|\hat{R}_{X_3 X_5}(f)|^2$ versus frequency f, and the 5% statistical confidence bound of coherence, is marked by the starred-line.
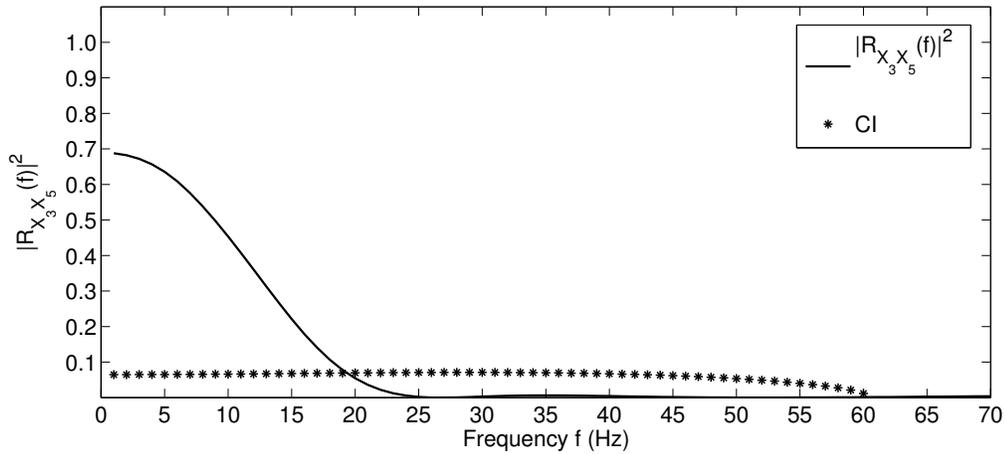
Figure 6.24: The estimated coherence $|\hat{R}_{X_3 X_5}(f)|^2$ is denoted by the solid-
line and the confidence bound of $\alpha = 5\%$ level of significance is marked by
the starred-line.

Note that although the occipital lobe process X_4 has a direct causal influence on the
temporal lobe process X_5, there is no significant estimated coherence between these two
processes. Figure (6.25) shows that all coherence values lie beneath the confidence bound
of $\alpha = 5\%$ level of significance.
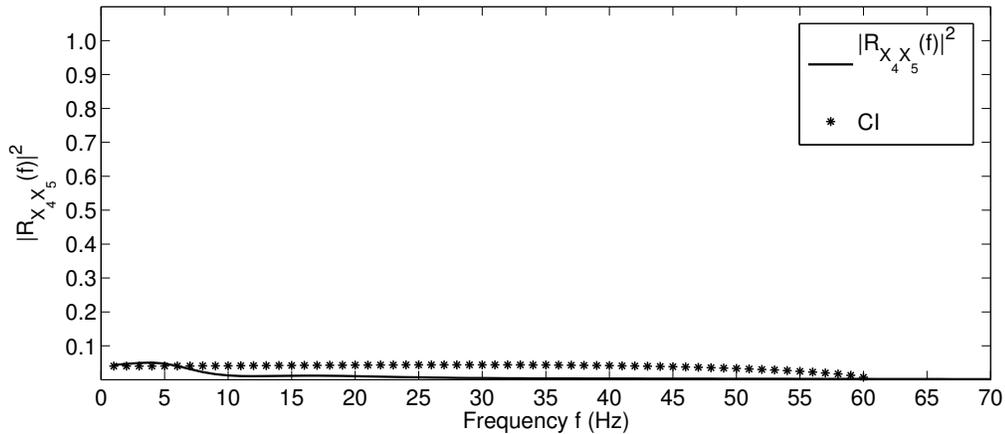
Figure 6.25: The estimated coherence $|\hat{R}_{X_4 X_5}(f)|^2$ is denoted by the solid-
line and the confidence bound of $\alpha = 5\%$ level of significance is marked by
the starred-line.

To determine the depth of the correlation between various pairs of processes among
the neural network, we investigate the partial coherence, defined in Chapter 4. Partial
coherence can be obtained from the relation (4.84). That is, the order of the depth
of correlation between two processes corresponds to the number of the processes that

have been used to condition the coherence relation. The estimated coherence between processes X_1 and X_2, after removing the linear influences of process X_3, reduces from 0.6 to approximately 0.2 and also results in the disappearance of the β-wave in the frequency range [13, 30) Hz, as shown in figure (6.26).
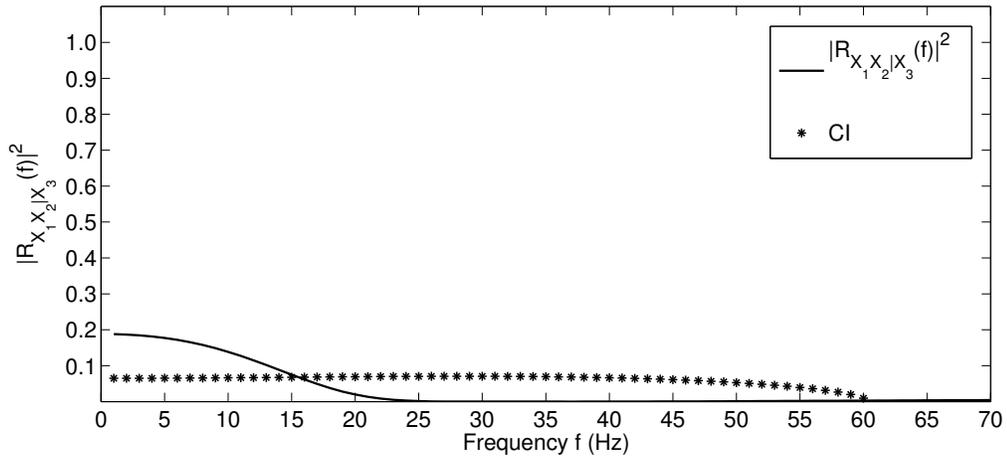

Figure 6.26: The estimated partial coherence $|\hat{R}_{X_1 X_2 | X_3}(f)|^2$ between processes X_1 and X_2, after removing the linear influences of process X_3 is marked by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

Also, eliminating the linear effects of process X_2 from the estimated coherence between processes X_1 and X_3 generally leads to weaken the correlation between these processes, and causes the β-wave in [13, 30) Hz to decay. The changes in the estimated coherence of processes X_1 and X_3, after removing the linear effects of process X_2, are presented in figure (6.27). Similarly the estimated partial coherence between processes X_2 and X_3 conditioned on X_1, reduces to less than 0.2, as revealed in figure (6.28).
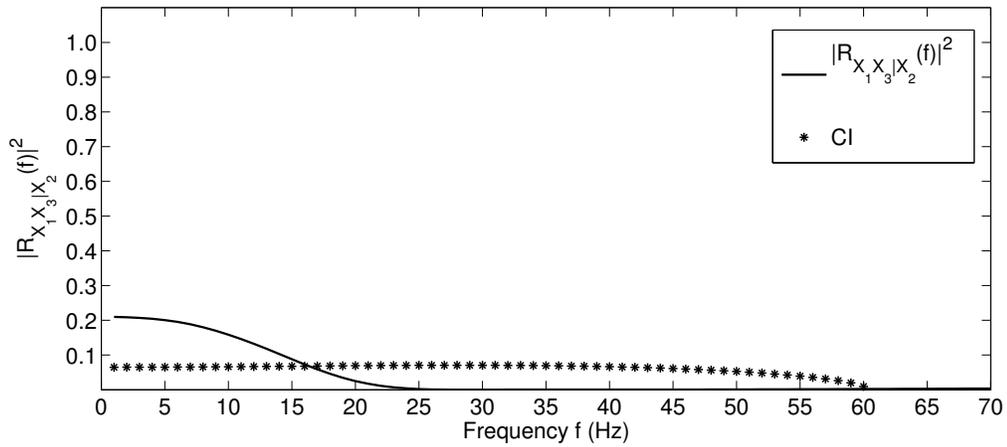
Figure 6.27: The estimated partial coherence $|\hat{R}_{X_1X_3|X_2}(f)|^2$ between processes X_1 and X_3, after removing the linear influences of process X_2, is marked by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.
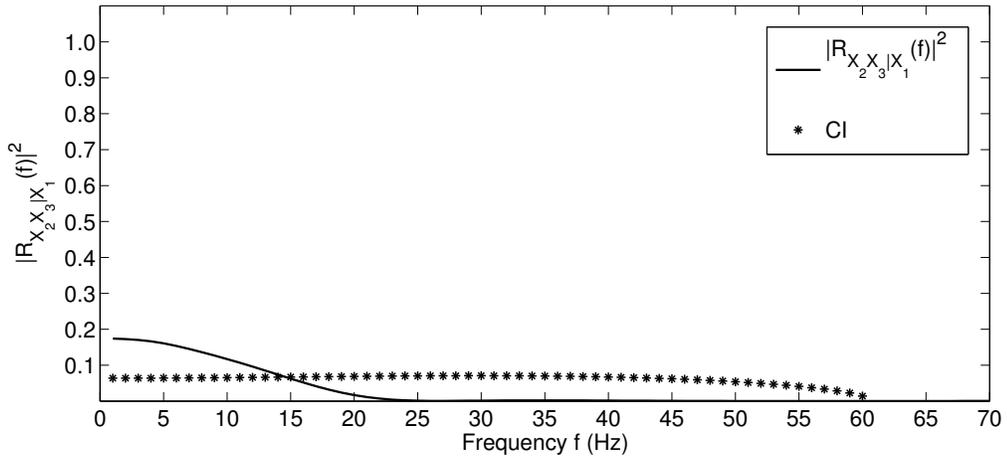
Figure 6.28: The estimated partial coherence $|\hat{R}_{X_2X_3|X_1}(f)|^2$ between processes X_2 and X_3, after removing the linear influence of process X_1, is marked by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred line.

The process X_5, which represents the temporal lobe, plays a crucial role in supporting estimated coherence functions $|\hat{R}_{X_1X_2}(f)|^2$, $|\hat{R}_{X_1X_3}(f)|^2$ and $|\hat{R}_{X_2X_3}(f)|^2$. The removal of the linear effects of process X_5 from these coherence functions causes the complete destruction of these correlations. Figure (6.29) shows the decay of the estimated coherence between processes X_1 and X_2 as a result of eliminating the linear effects of process X_5.
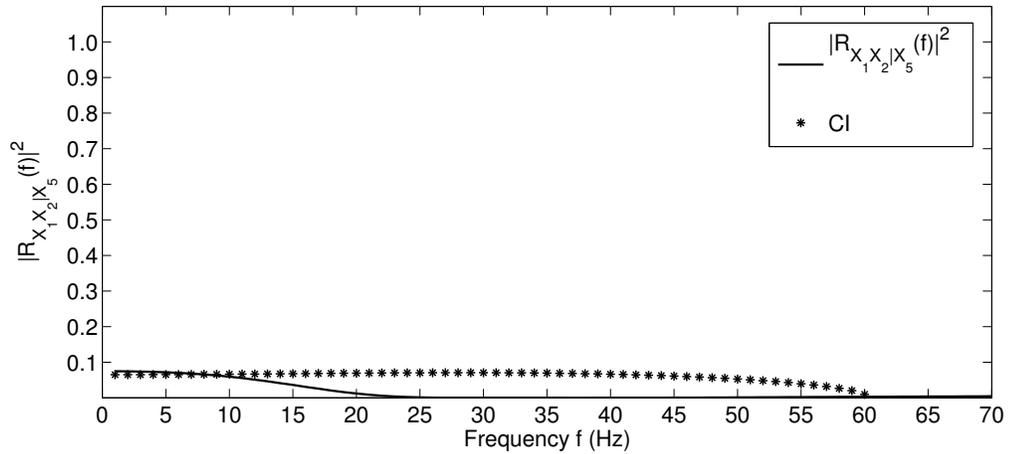
Figure 6.29: The estimated partial coherence $|\hat{R}_{X_1 X_2 | X_5}(f)|^2$ between processes X_1 and X_2, after removing the linear influence of process X_5, is marked by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

Figure (6.30) illustrates the estimated coherence of processes X_1 and X_3 conditioned on process X_5.
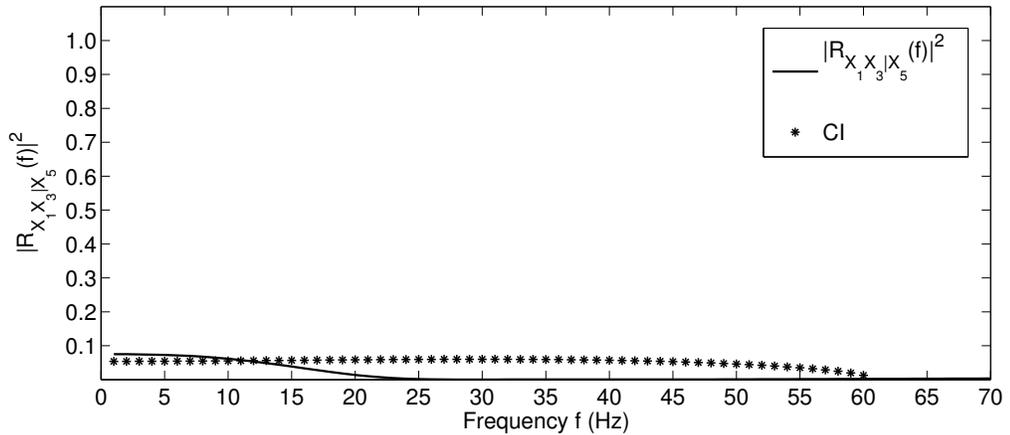
Figure 6.30: The estimated partial coherence $|R_{X_1 X_3 | X_5}(f)|^2$ between processes X_1 and X_3, after removing the linear influences of process X_5, is marked by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

Figure (6.31) displays the decay of the estimated coherence between the processes X_2 and X_3 after removing the linear influences of process X_5 .
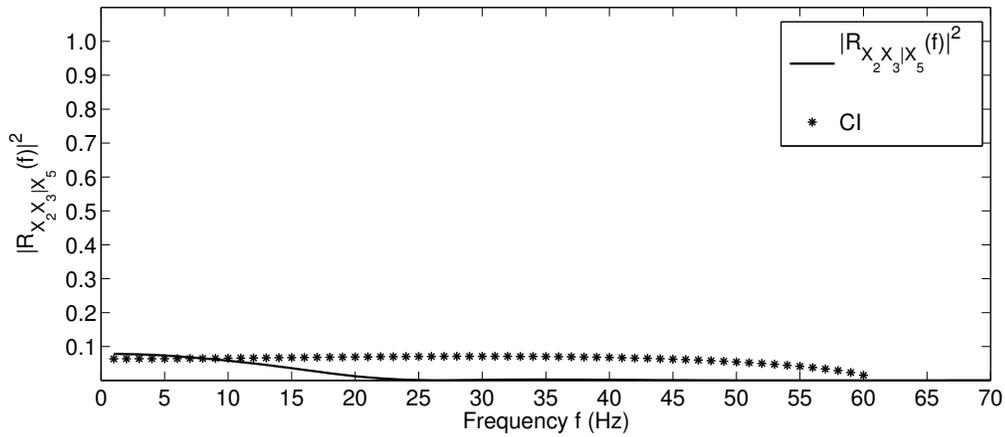
Figure 6.31: The estimated partial coherence $|\hat{R}_{X_2 X_3 | X_5}(f)|^2$ between processes X_2 and X_3, after removing the linear influences of process X_5, is marked by the solid-line and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

Consequently, we can say that the depth of the correlation between the frontal process X_1 and the central process X_2 is of order one, as the estimated coherence vanishes completely after removing the linear influences of only one process, in this case X_5. Similarly the depth of the estimated coherence of the frontal process X_1 with the parietal process X_3; and the estimated coherence of the central process X_2 with the parietal process X_3 are also of order one.

However, the investigation of the correlations between process X_5 and the other processes involved in the current neural network, with the exception of process X_4, reveals that the coherence remains strong and significant even for the second order. Note that we do not consider the third order of coherence, since process X_4 is entirely independent from all other processes. Figure (6.32) illustrates the estimated coherence of X_1 with X_5, after removing the linear influences of process X_2, plotted by the dashed-line and the effect of eliminating the linear influences of process X_3 on the estimated coherence between X_1 and X_5, marked by the dotted-line. That is, the first order estimated partial coherences $|\hat{R}_{X_1 X_5 | X_2}(f)|^2$ and $|\hat{R}_{X_1 X_5 | X_3}(f)|^2$ are approximately similar, whereas the removal of simultaneous linear influences of processes X_2 and X_3 from the estimated coherence between processes X_1 and X_5, contributes to reduce the estimated coherence function $|\hat{R}_{X_1 X_5 | X_2 X_3}(f)|^2$ to the smaller values.
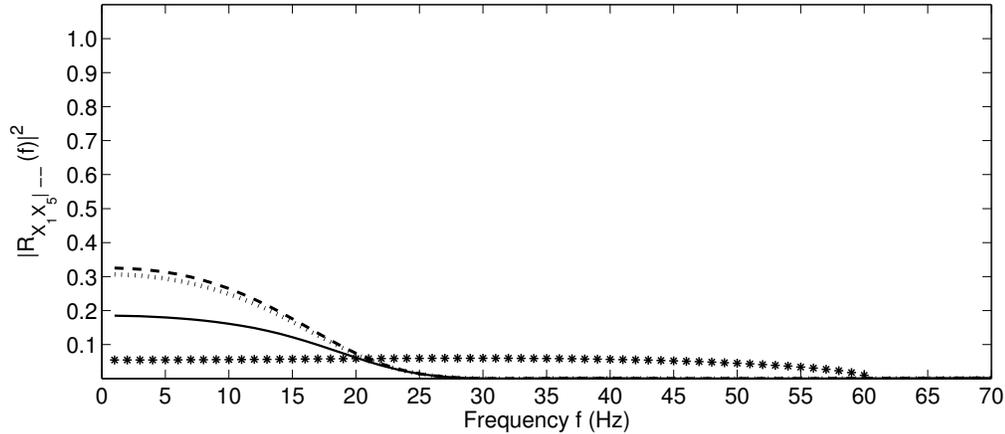
Figure 6.32: The estimated partial coherence $|\hat{R}_{X_1X_5|X_2X_3}(f)|^2$, after removing the linear influences of the two processes X_2 and X_3 simultaneously, is marked by the solid-line, the dashed-line represents the estimated partial coherence of order one $|\hat{R}_{X_1X_5|X_2}(f)|^2$, the dotted-line denotes the estimated partial coherence of order one $|R_{X_1X_5|X_3}(f)|^2$ and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred line.

Similarly, the removal of the linear influences of process X_1 from the estimated coherence of processes X_2 with X_5, contributes to reduce the value of coherences to less than 0.3. Approximately the same result is obtained from removing the linear effects of process X_3 from the estimated coherence of processes X_2 with X_5. These results are illustrated in figure (6.33) where the $|\hat{R}_{X_2X_5|X_1}(f)|^2$ is represented by the dashed-line, the $|\hat{R}_{X_2X_5|X_3}(f)|^2$ is denoted by the dotted-line and the starred-line represents the confidence bound of $\alpha = 5\%$ level of significance. The estimated second order partial coherence $|\hat{R}_{X_2X_5|X_1X_3}(f)|^2$, which is produced by removing the simultaneous linear influences of processes X_1 and X_3, is still statistically significant with relatively small values just above 0.1, and is plotted by the solid-line in figure (6.33).
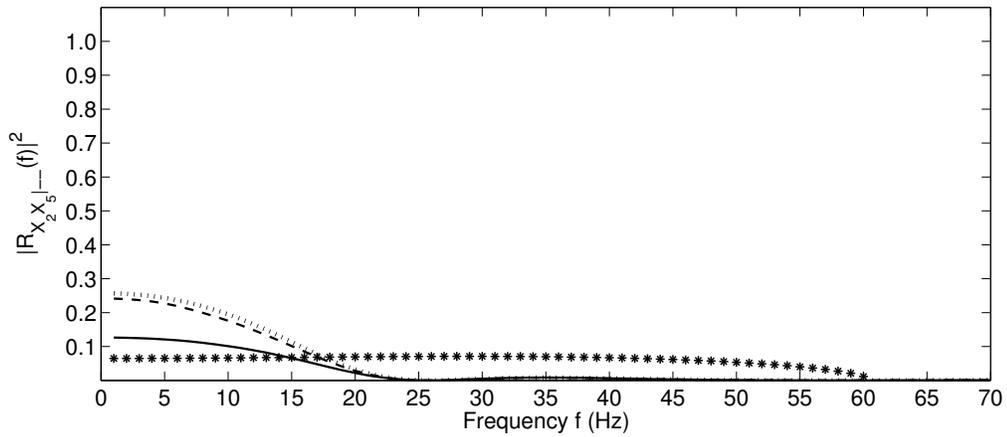
Figure 6.33: The estimated partial coherence $|\hat{R}_{X_2X_5|X_1X_3}(f)|^2$ between processes X_2 and X_5, after removing the linear influences of processes X_1 and X_3 simultaneously, is marked by solid-line, the dashed-line represents the partial coherence of order one $|\hat{R}_{X_2X_5|X_1}(f)|^2$, the dotted-line denotes the partial coherence of order one $|R_{X_2X_5|X_3}(f)|^2$ and the confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.

Figure (6.34) illustrates the estimated first order partial coherence $|\hat{R}_{X_3X_5|X_1}(f)|^2$, denoted by the dashed-line, the estimated first order partial coherence $|\hat{R}_{X_3X_5|X_2}(f)|^2$, marked by the dotted-line and the estimated second order partial coherence $|\hat{R}_{X_3X_5|X_1X_2}(f)|^2$, represented by the solid-line. The confidence bound of $\alpha = 5\%$ level of significance is represented by the starred-line.
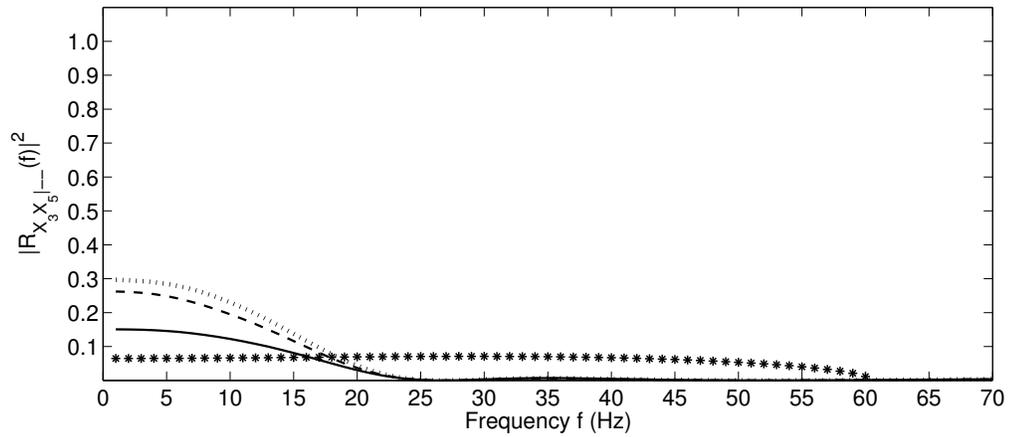
Figure 6.34: The estimated second order partial coherence $|\hat{R}_{X_3X_5|X_1X_2}(f)|^2$ is denoted by the solid-line, the dashed-line represents the estimated first order partial coherence $|\hat{R}_{X_3X_5|X_1}(f)|^2$, the dotted-line denotes the estimated first order partial coherence $|\hat{R}_{X_3X_5|X_2}(f)|^2$. The confidence bound of $\alpha = 5\%$ level of significance is marked by the starred-line.

A quick look at the ordinary coherence of processes X_3 and X_5, displayed in figure (6.24) reveals how much support is provided from the processes X_1 and X_2, either separately or collectively, to reinforce the correlation between the processes X_3 and X_5. Figure (6.34) shows that the absence of the linear influences of the process X_1 from the estimated coherence of X_3 with X_5 , denoted by the dashed-line, plays a crucial role in reducing the values of the estimated coherence from almost 0.7 to less than 0.3 in the δ- wave band [1, 4) Hz, and less than that value for the rest of the EEG's waves over their bands of frequencies. The estimated first order partial coherence, occurred as a result of removing the linear effects of process X_2, sketched by the dotted-line, appears approximately as equal as the estimated first order partial coherence of X_3 and X_5, produced by removing the linear effects of process X_1, denoted by the dashed-line. However, the removal of simultaneous linear influences of these processes from the estimated coherence of processes X_3 andX_5, leads to a huge reduction in the estimated coherence values and almost eliminates the β-wave in [13, 30) Hz.

Generally speaking, the temporal area is strongly associated with the frontal, central and parietal areas, in spite of removing the linear effects of one or two processes. Moreover, the EEG's waves; δ-wave, θ-wave, α-wave still remain in the estimated second order partial coherence of X_3 with X_5, while the β- wave almost vanishes.

6.1.1   Results summary of MVAR

In light of the previous analyses of the chosen neural network using a multivariate autoregressive model, we encapsulate the following results:

- The multivariate autoregressive model of order$(p = 2)$ is not sensitive to the rapid changes that occur in the data, such as the calibration signal at $(f = 47\text{Hz})$, and the loss of information caused by the notch filter at $(f = 60\text{Hz})$.

- The occipital lobe, represented by the process X_4, is uncorrelated with the other brain lobes.

- The four brain regions, the frontal lobe, central area, parietal lobe and temporal lobe exhibit different non-zero levels of correlation.

- The temporal lobe process X_5 plays a crucial role in supporting the correlation between the other lobe processes. Consequently, the removal of the linear influences of process X_5 leads to the destruction of these correlations, as shown in figure (6.29), figure (6.30) and figure (6.31).

- The estimated coherences of the process X_5 with the other processes, except X_4, decrease, but do not vanish after removing the linear effects of one process such as X_2, or two processes such as X_2 and X_3 . See figure (6.32), figure (6.33) and figure (6.34).

- The four types of the EEG's waves that appear in the MVAR analysis are δ-wave, θ-wave, α-wave and β-wave.

6.2   Non-parametric spectral estimation

In this section, in order to estimate the spectral density, we provide the spectral analysis of the current data using non-parametric approaches, namely the averaging periodograms across contiguous sections of single records; smoothing the periodogram of the entire record (also known as averaging across frequency method); and the multi-taper method. These methods are demonstrated extensively in Chapter 3.

Firstly, we give particular descriptions for these methods, which are consistent with the underlying data, and discuss the power spectra of the considered data. Secondly, we debate the coherence and partial coherence analyses of these methods in parallel.

6.2.1 Averaging periodograms across contiguous sections of single records

It is well known that the power spectrum is an important measure widely used in spectral analysis to capture frequencies corresponding to the fluctuations of the signal in its time domain. We begin by estimating the power spectrum of each channel, $\{X_i; i = 1, \cdots, 5\}$, by portioning the whole record, for each signal, of length T milliseconds into L non-overlapping segments, each of size N, in this case $N = 200$, and $L = 65$ subsamples. The choice of the sample size $N = 200$ implies naturally to physical frequencies in Hertz. Thereafter, the finite Fourier transform is applied to every single section $\{l = 1, 2, \cdots, L\}$, and the estimated power spectrum (periodogram) for each segment $\{l = 1, 2, \cdots, L\}$ at frequency f_k, where the frequency index, $-N/2 < k \le N/2$, $I_{XX}^N(f_k, l)$ is computed directly from equation (3.45), by taking the average of the squared-magnitude of the finite Fourier coefficients $|d_X^N(f_k, l)|^2$ with respect to the size of subsample N. Once the periodograms are computed, the estimated power spectrum for the whole record of the signal is obtained from relation (3.46), by taking the average of these periodograms over number of subsamples L. Similarly, the cross-spectra can be estimated directly from the data using equation (3.56), which is based on averaging the cross-periodograms of disjoint sections as elaborated in Chapter 3.

In order to validate the calculation of the estimated power spectra of the underlying data set we add to each observed signal a calibration signal consisting of a cosine signal at a specific frequency, namely ($f = 47$Hz), multiplied by the convenient root mean square RMS to the signal of interest. This leads to the appearance of a spike corresponding to that specific frequency ($f = 47$Hz) in the power spectra curves. The logarithm to base 10 of the estimated power spectra are plotted against frequencies between 1Hz and 70Hz for the processes X_1 and X_2, in figures (6.35:a) and (6.35:b), respectively.

(a)   (b)
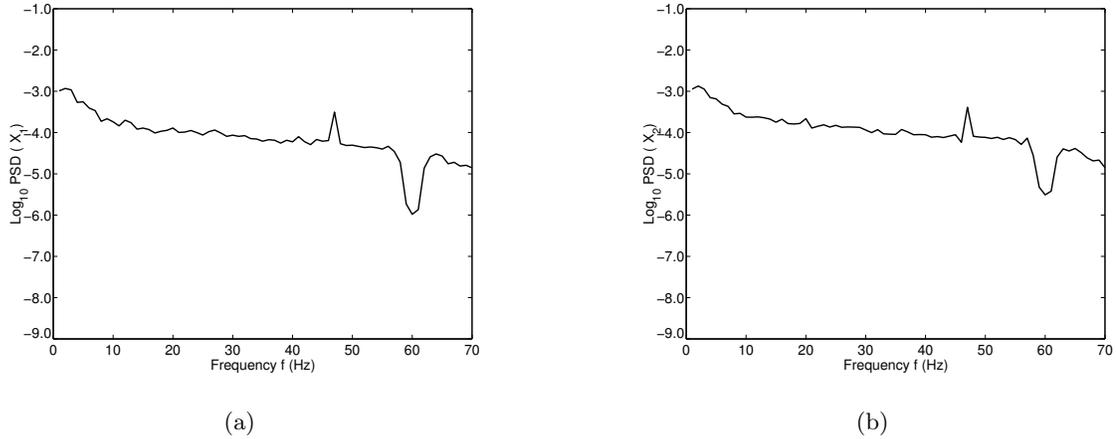
Figure 6.35: The logarithm to base 10 of: (a) the power spectrum of the signal X_1; (b) the power spectrum of the signal X_2. The estimated spectra plotted against the frequency ranging between 1Hz and 70Hz.

From figures (6.35:a, 6.35:b) we can see clearly that the method of disjoint sections, successfully captures the artificial calibration signal at frequency ($f = 47$Hz), while the multivariate autoregressive model failed to do so in the previous section. Also there is an appearance of a trough at a frequency 60Hz due to the reduction of the main frequency, in Saudi Arabia is of 60 Hz, by the notch filter. The logarithm to base 10 of the estimated power spectra of the processes X_3 and X_4 are displayed in the figures (6.36:a, 6.36:b) respectively,


(a)   (b)
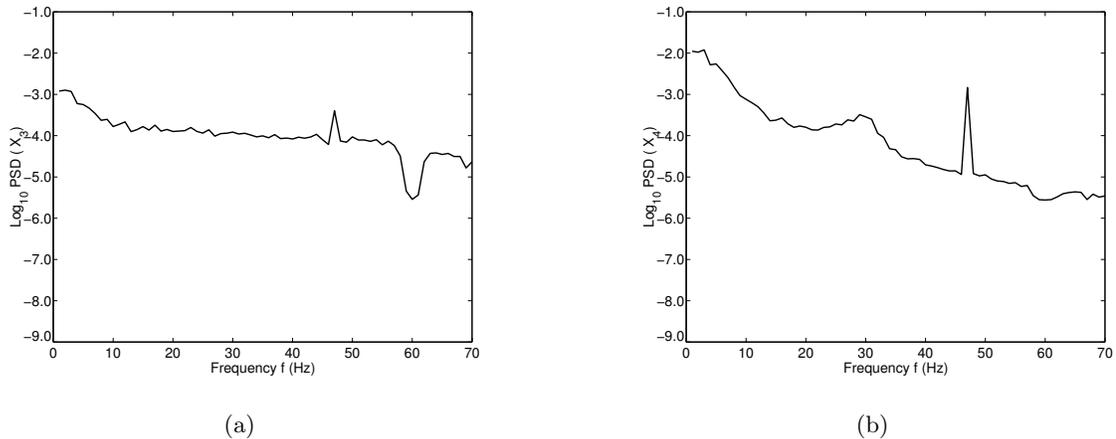
Figure 6.36: The logarithm to base 10 of: (a) the power spectrum of the signal X_3; (b) the power spectrum of the signal X_4. The spectra plotted against frequencies between 1Hz and 70Hz.

and the logarithm to base 10 of the power spectrum of the process X_5 is illustrated in

figure (6.37).
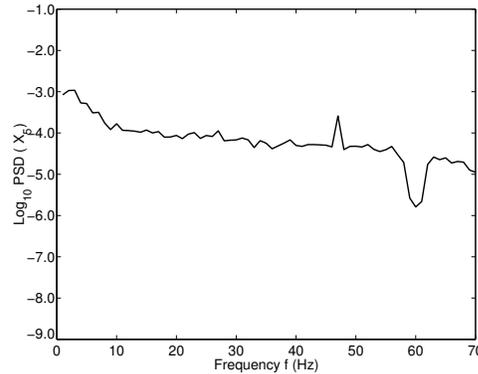
Figure 6.37: The logarithm to base 10 of the power spectrum of the signal X_5.
The spectra plotted against frequencies between 1 Hz and 70 Hz.

The calibration signal peak at frequency ($f = 47$Hz), and the reduced power line noise at ($f = 60$Hz) have emerged. Since the power spectrum describes the distribution of the energy contained in the signal at each frequency, and from figures (6.35:a, 6.35:b, 6.36:a, 6.37), we note that these signals almost behave similarly, but with different levels of energy. The process X_4, which represents the occipital lobe, contains energy which differs in strength and behaviour from those involved in the other signals as illustrated in figure (6.36:b).

6.2.2 Smoothed periodogram

In this section we perform the spectral analysis for the same neural network by employing the smoothed periodogram method to estimate the spectral densities. As elaborated in Chapter 3, the Fourier coefficients $d_X^T(f)$ are obtained for the entire record of every single process, then the periodogram $I_{XX}^T(f_k)$ is computed at frequency ($f_k = k/T$) where $k \in (-T/2, T/2]$. Thereafter, the estimated spectra and cross-spectra are captured by taking the average of periodograms across ($2m + 1$), where $m = 7$ in this case, independent frequencies in the neighbourhood of the frequency f_k using expressions (3.40) and (3.53) respectively. Also, we add the typical calibration signal at frequency ($f = 47$Hz) to the observed signals in order to validate the calculations of the estimated spectra and cross-spectra. Thus the logarithm to base 10 of the spectra of processes X_1 and X_2 are plotted against frequency in figures (6.38:a, 6.38:b).

(a)   (b)
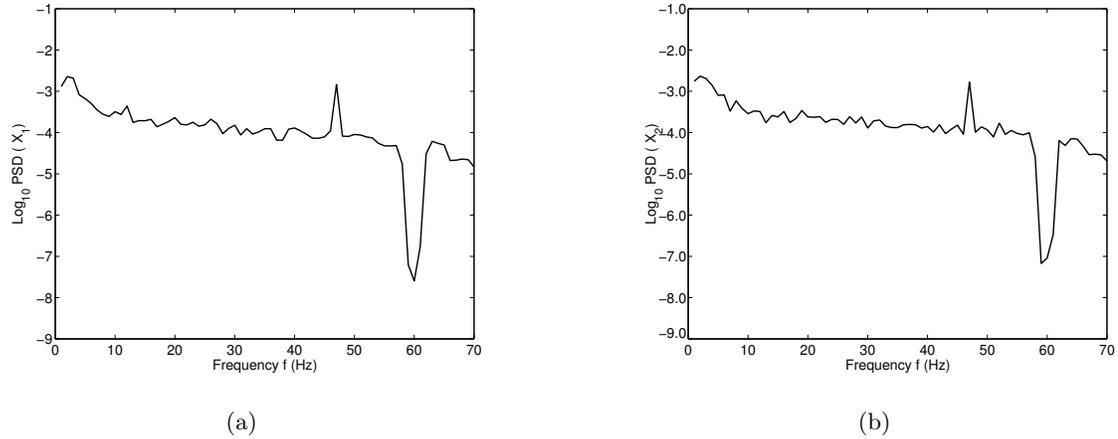
Figure 6.38: The logarithm to base 10 of:(a) the power spectrum of the signal X_1, (b) the power spectrum of the signal X_2. The spectra plotted against the frequencies between 1Hz and 70Hz.

From these figures, we can clearly see a spike at frequency ($f = 47$Hz), due to the addition of the calibration signal, and a deep trough appearing at frequency ($f = 60$Hz), as a result of the notch filter, which reduces the line power noise. Similarly the logarithm to base 10 of the power spectra of processes X_3, X_4 are displayed in figures (6.39:a, 6.39:b) respectively. The estimated power spectrum of the process X_5 is illustrated in figure (6.40).


(a)   (b)
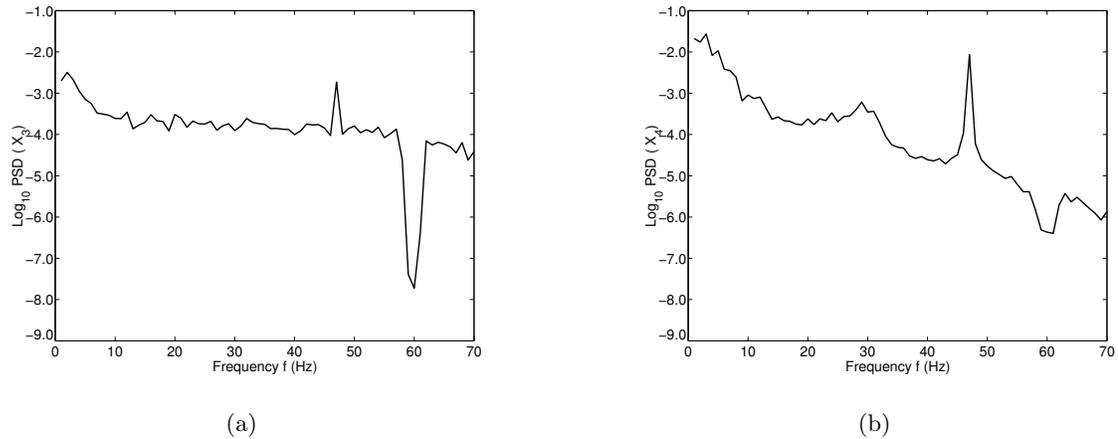
Figure 6.39: The logarithm to base 10 of: (a) the power spectrum of the signal X_3; (b) the power spectrum of the signal X_4. The spectra plotted against the frequencies between 1Hz and 70Hz.
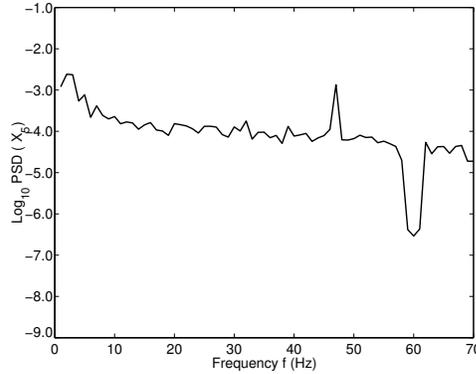
Figure 6.40: The logarithm to base 10 of the power spectrum of the signal X_5.
The spectra plotted against frequencies between 1 Hz and 70 Hz.

Again, the calibration signal peak emerges at the correct frequency and ensures that the calculations of the power spectra are executed properly. The trough stems from decreasing the noise of the power line by the notch filter at frequency ($f = 60$Hz), appears more clearly in the power spectra estimated by the smoothed periodogram procedure, than with those estimated by the disjoint sections method, especially for the process X_4.

6.2.3   Multi-taper method

Multi-taper methods have been developed to estimate spectral densities for records of relatively short duration. In this approach, observations are weighted by independent tapers. The finite Fourier transform is then applied to the product of the data with tapers using the logarithm of the Fast Fourier Transform. Next the power spectrum is obtained for every single taper and the final power spectrum is constructed, by taking the average of these spectra. This method was described in Chapter 3. We shall now perform the spectral analysis for the same neural network using the multi-taper method, where the number of independent tapers we use is $K = 10$, and the estimated spectra and cross-spectra will be obtained from the expressions (3.59) and (3.62) respectively. Typically, the artificial pure signal is added to the observed signals in order to ensure the authenticity of the estimated spectra and cross-spectra. Then, the logarithm to base 10 is applied to the estimated spectra of the processes in the study. The estimated spectra of the processes X_1, representing the frontal lobe and X_2, representing the central area of the brain, are illustrated in figures (6.41:a), (6.41:b), respectively.

(a)   (b)
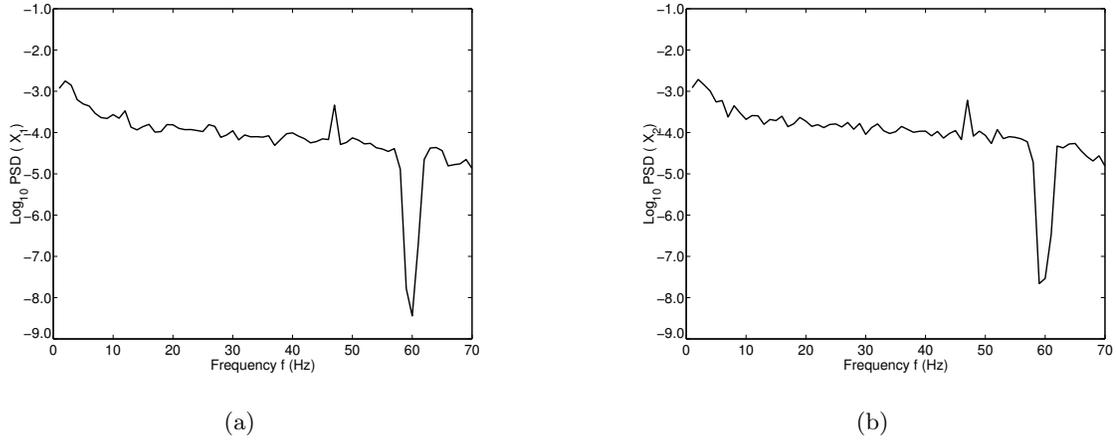
Figure 6.41: The logarithm to base 10 of : (a) the power spectrum of the signal X_1; (b) the power spectrum of the signal X_2. The spectra plotted against the frequencies between 1Hz and 70Hz.

The estimated spectra of the parietal lobe, represented by process X_3 , and the occipital lobe process X_4, are displayed in figures (6.42:a), (6.42:b), respectively.


(a)   (b)
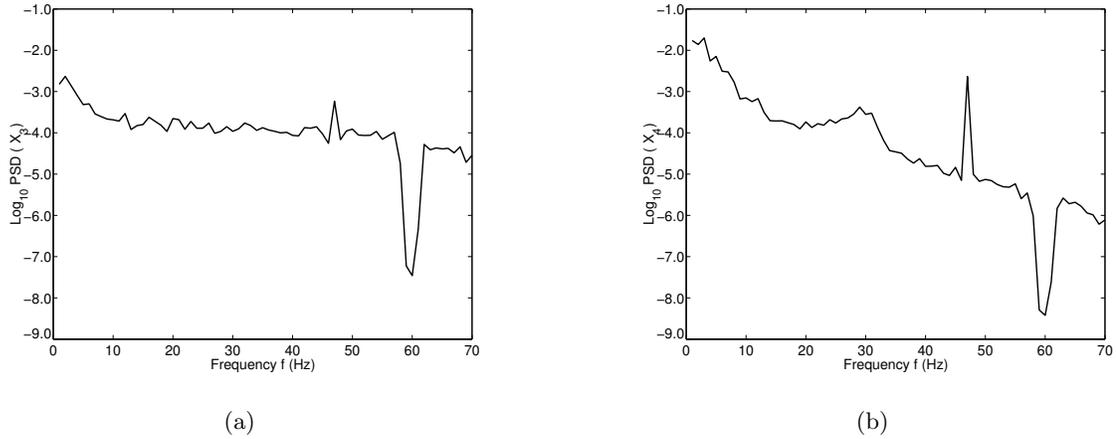
Figure 6.42: The logarithm to base 10 of : (a) the power spectrum of signal X_3; (b) the power spectrum of the signal X_4. The spectra plotted against the frequencies between 1Hz and 70Hz.

The estimated spectrum of the temporal lobe of the brain, is presented in figure (6.43).
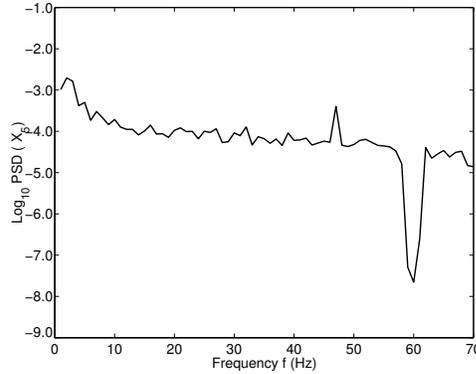
Figure 6.43: The logarithm to base 10 of the power spectrum of the signal X_5, plotted against the frequencies between 1Hz and 70Hz.

The detection of the loss of information by the notch filter at frequency 60Hz, is presented as a deep narrow trough is revealed more clearly in this method, compared with previous estimation methods, especially for the process X_4. Also, the emergence of the calibration signal at frequency ($f = 47$Hz), indicates the correctness of the calculations.

In general, the estimated spectra of signals X_1, X_2, X_3 and X_5 decrease slightly at frequencies between 1Hz and 10Hz, then the steady manner of the energy dominates. However, the power spectrum of the signal X_4, displayed in figure (6.41:b), shows that the energy of X_4 decreases dramatically at frequencies lying in the interval $[1, 20)$ Hz followed by an increase of the signal power peaking at almost 30Hz, therefore descending behaviour appears again with appearance of the calibration signal peak at 47Hz and the notch filter effect at 60Hz.

We shall now investigate the correlation between these processes using the measurements of the coherence and partial coherence, defined in equations (4.3) and (4.20) respectively. The components of the partial coherence, partial spectra, are computed from the mechanism described in Chapter 4. In order to decide whether the estimated coherence of two processes is significant or not, we examine the null hypothesis that two processes are %95 uncorrelated, that is $|R_{X_i X_j}(f)|^2 = 0$. The confidence interval for the estimated coherence is defined in expression (4.11) in Chapter 4. Thus we have three level of confidence bound at %5 level of significance according to the adopted spectral estimation method. That is, for the the disjoint sections method, the confidence bound is given by $\left(x = 1 - 0.05^{(1/64)} \right)$; for the smoothed periodogram method, is given by $\left(x = 1 - 0.05^{(1/14)} \right)$; and finally for the multi-taper method, the confidence bound is given by $\left(x = 1 - 0.05^{(1/9)} \right)$. Consequently, all values of the estimated coherence or partial

coherence locate below the confidence bound construct the acceptance region for the null hypothesis.

The estimated coherences between the processes X_1 and X_2 at frequency f in $[1, 70]$ Hz, is displayed in figure (6.44), where the narrow spike appears at ($f = 47$Hz) represents the calibration signal. Figure (6.44:a) displays the estimated coherence function $|\hat{R}_{X_1 X_2}(f)|^2$, calculated by the disjoint sections method, figure (6.44:b) represents the estimated coherence between those processes, calculated by the frequency averaging method, and finally, figure (6.44:c) represents the estimated coherence of X_1 with X_2, produced by the multi-taper method. The estimated coherences are plotted by solid-line, where the confidence bound of the significant coherence at level 5% is marked by the starred-line. Thus the estimated coherence values, $|\hat{R}_{X_1 X_2}(f)|^2$, lying above the starred-line indicate the significant correlation between processes X_1 and X_2, while the coherence values locate below the starred-line represent frequencies at which processes are not significantly correlated.

Figure (6.44:a) shows that the significant coherence between X_1 and X_2 occurs within the frequency interval $[1, 11]$ Hz, which is the frequency domain of three types of EEG's waves. For example, the δ-wave appears clearly in the frequency band $[1, 4)$ Hz, where the high values of coherence are raised. The θ-wave frequency band $[4, 8)$ Hz includes different events, as the estimated coherence reduces steeply from the value 0.497 at frequency 5Hz to the value 0.2271 at frequency 6Hz, followed immediately by a sharp rise in the estimated coherence value to 0.3365 at 7Hz. Thereafter, the value decreases slightly in the α-rhythm frequency band $[8,11]$ Hz.

Figure (6.44:b) shows that the estimated coherence between X_1 and X_2 occurs mostly in the band frequency $[1, 4)$Hz, which is related to δ-rhythm and partially in the θ-rhythm band frequency $[4,6]$ Hz. The spike at 47Hz denotes the calibration signal that has been added to the real signals to authenticate the calculations of the spectral densities, and the second spike at frequency 60Hz due to the notch filter, which is reducing the power line noise in both processes X_1 and X_2. Continuing, the estimated coherence between processes X_1 and X_3 is also captured in the δ-rhythm and θ-rhythm frequency bands, that is, the significant estimated coherence occurred in $[1, 8]$ Hz. It also can be seen that the spike at frequency ($f = 60$Hz), which represents the notch filter effect.
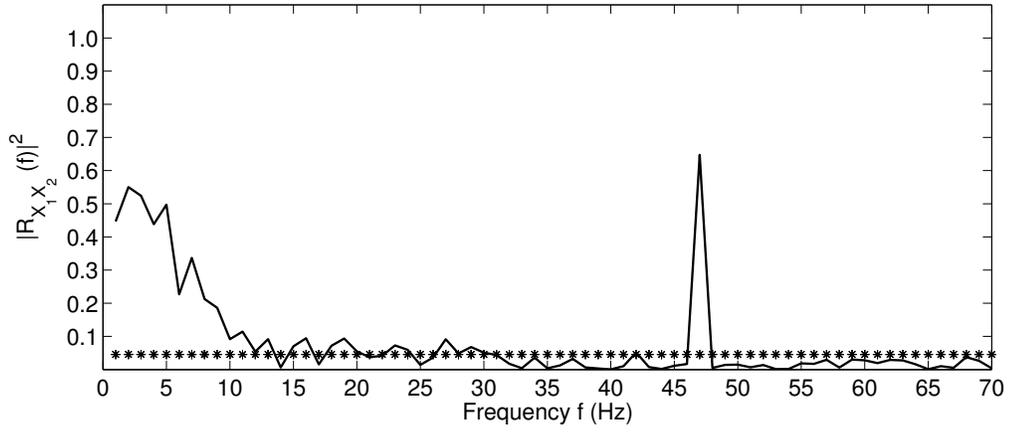
(a)
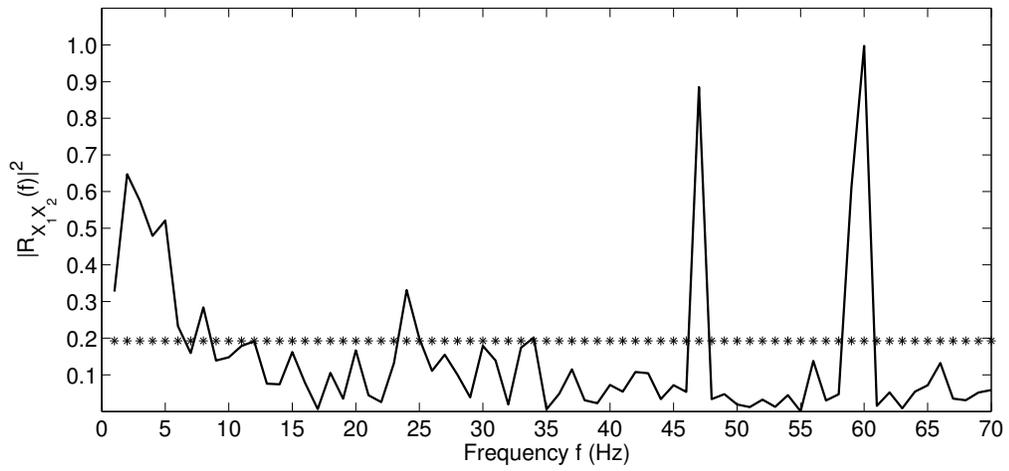
(b)
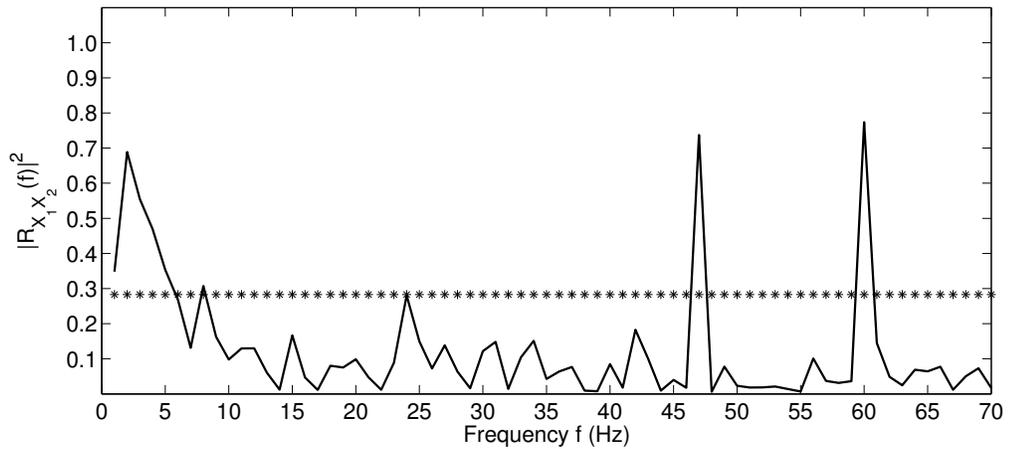
(c)

Figure 6.44: The estimated coherence function, $|\hat{R}_{X_1 X_2}(f)|^2$: (a) the disjoint sections method, (b) the frequency averaging method, (c) the multi-taper method. The estimated coherences are plotted against the frequencies between 1 Hz and 70 Hz.

From figure (6.44:c), we can see that the high value of the estimated coherence, $|\hat{R}_{X_1 X_2}(f)|^2$, occurs within the δ-wave band of frequency $[1, 4)$ Hz, which denotes the slow brain activity. Also, there are two significant values of the estimated coherence corresponding to frequencies 4Hz and 5Hz within the θ-wave frequency band $[4, 8)$ Hz.

Continuing, the significant estimated coherences between the processes X_1 and X_3, computed by the three non-parametric methods, are illustrated in figure (6.45). Figure (6.45:a) displays the estimated coherence, calculated by the disjoint sections method; and figure (6.45:b) represents the estimated coherence between processes X_1 and X_3, calculated by the smoothed periodogram method; and finally the estimated coherence of X_1 with X_3, produced by the multi-taper method, is shown in

figure(6.45:c). Where the estimated coherence $|\hat{R}_{X_1 X3}(f)|^2$ is plotted by the solid-line and the starred-line represents the confidence interval of the significant coherence at level 5%.

For the disjoint sections method, the estimated coherence of X_1 and X_3 occurs entirely within the frequency interval $[1, 10]$ Hz, as shown in figure (6.45:a). It can be seen that the estimated coherence value rises between frequencies 1Hz and 3Hz, then falls sharply from the value 0.6078 at frequency 3Hz to the value 0.3642 at frequency 4Hz, constructing a spike of the δ-wave in $[1, 4)$ Hz. Also, another peak appears within the band frequency $[4, 8)$ Hz, which is associated with the θ-wave. Thereafter, the value of the estimated coherence decreases gradually within the α-wave frequency band $[8, 10]$ Hz.

For the smoothed periodogram method, the estimated coherence between processes X_1 and X_3 is captured within the frequency band $[1, 8]$ Hz. This band, in turn, involves two types of the EEG's waves: the δ-rhythm in $[1, 4)$ Hz; and θ-rhythm in the frequency band $[4, 8)$ Hz. Figure (6.45:b) shows three consecutive spikes correspond to the frequencies 3Hz, 5Hz and 8Hz.

Furthermore, the significant estimated coherence between processes X_1 and X_3, calculated by the multi-taper method, occurs in the δ-wave band frequency $[1, 4)$ Hz as illustrated in figure (6.45:c). Although the calibration signal at frequency $(f = 47)$ Hz appears in the three figures, the notch filter at $(f = 60)$ Hz emerges in the last two figures.
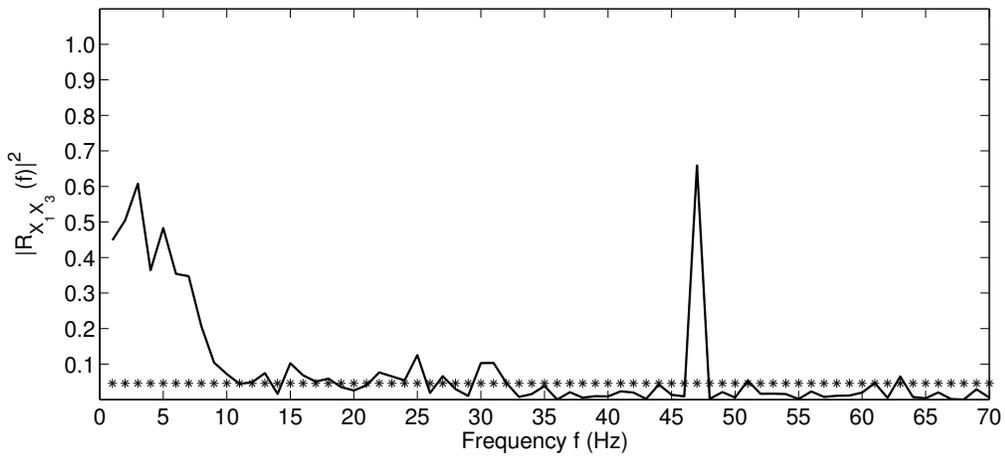
(a)
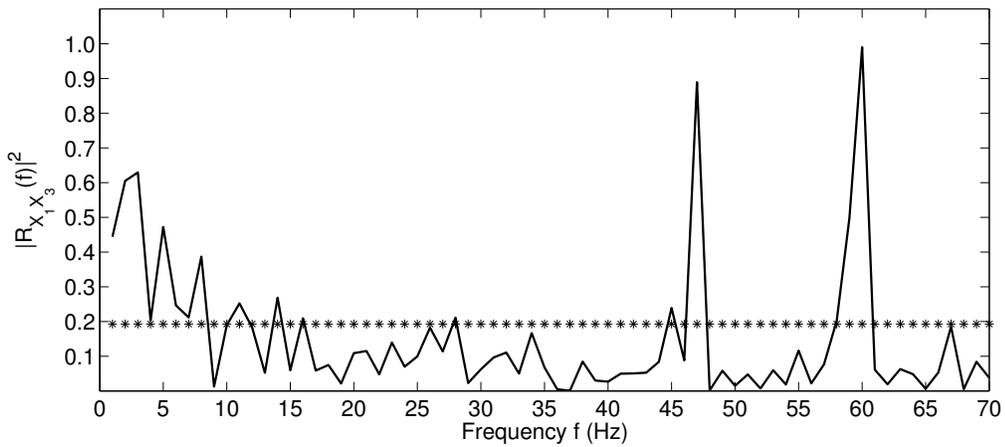
(b)
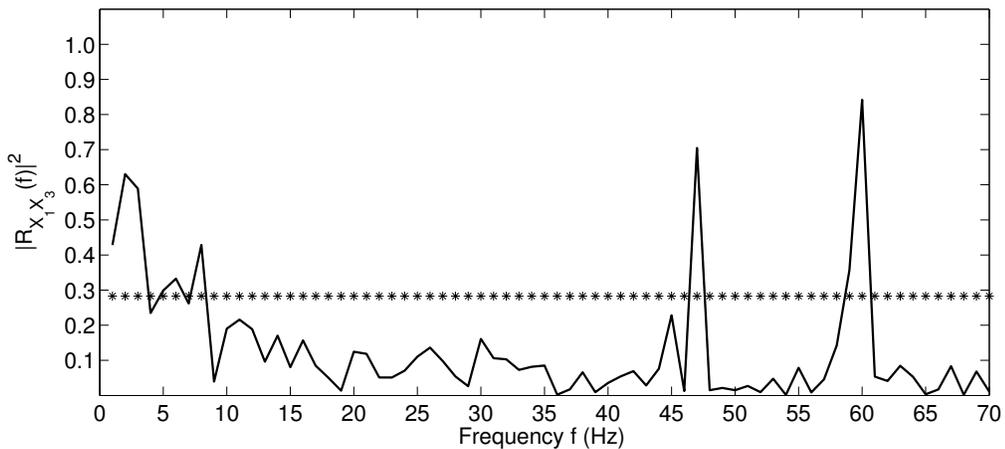
(c)

Figure 6.45: The estimated coherence function, $|\hat{R}_{X_1 X_3}(f)|^2$: (a) the disjoint sections method, (b) the frequency averaging method, (c) the multi-taper method. The estimated coherences are plotted against the frequencies between 1 Hz and 70 Hz.

The significant estimated coherence is captured between processes X_1 and X_5. Figure (6.46:a) shows that the estimated coherence, calculated by the disjoint sections method, occurs in the wide frequency band [1, 33] Hz, which consists of five frequency bands linked to the EEG's waves. The largest value of the estimated coherence 0.6898 occurs at frequency 3Hz within the δ-wave frequency band [1, 4) Hz. This value then decreases steeply to 0.4572 corresponding to the frequency 4Hz in the θ-wave frequency band [4, 8) Hz. In addition, there is evidence of the α-wave in [8, 13) Hz, while the β-wave emerges in two adjacent bands of frequency [15, 22) Hz and [22, 28) Hz. Moreover, the first appearance of the γ-wave can be seen within the frequency band [29, 33) Hz.

Figure (6.46:b) displays the estimated coherence of X_1 with X_5, produced by the smoothed periodogram method. The significant estimated coherence $|\hat{R}_{X_1 X_5}(f)|^2$ occurs mostly in the δ-wave frequency band [1, 4) Hz and the θ-wave in frequencies between 4Hz and 8Hz. Also, there is a small spike appears at ($f = 10$Hz) in the α-wave frequency band [8, 11) Hz.

Moreover, the significant estimated coherence between X_1 and X_5, calculated by the multi-taper method and plotted in figure (6.46:c), happens in the frequency band [1, 8] Hz. High values of coherence appear in the δ-wave band [1, 4) Hz, while the low values of coherence occur within the θ-wave band [4,8) Hz. Figure (6.46) shows that, as the confidence bound increases, from the top of the panel to the bottom, the significant band of frequency contracts and coherence curve becomes more smooth.

Nevertheless, the process X_4, which represents the occipital lobe of the brain, is uncorrelated with the other processes. That is the estimated coherences between this process and any of the processes $\{X_1, X_2, X_3, X_5\}$ are significantly zero. For instance, the independent relation between processes X_1 and X_4 is illustrated in figure (6.47). The small values of the estimated coherence appear at 28Hz and 29Hz are irrelevant, because the statistical test of the significant coherence requires at least three consecutive values of the significant estimated coherence.
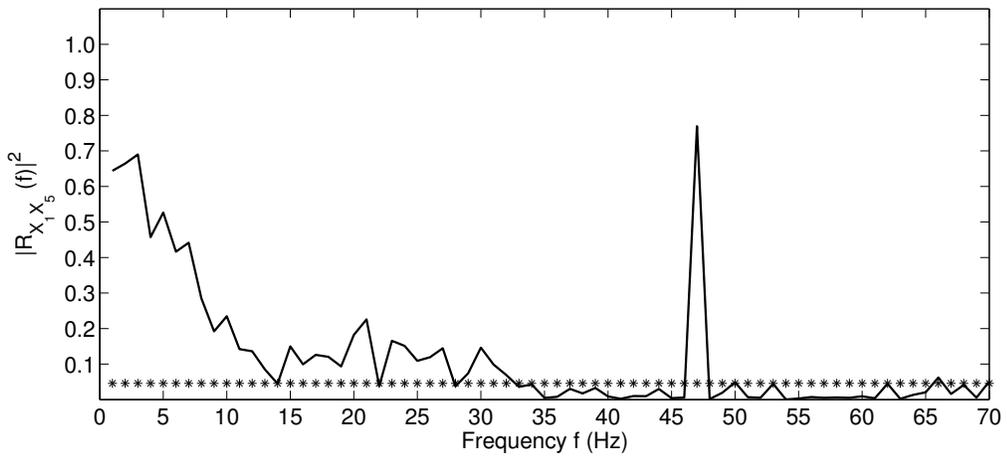
(a)
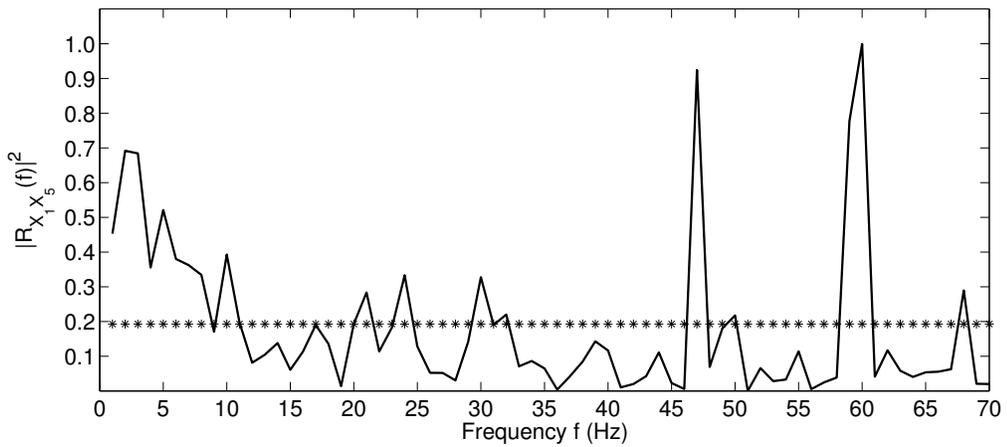
(b)
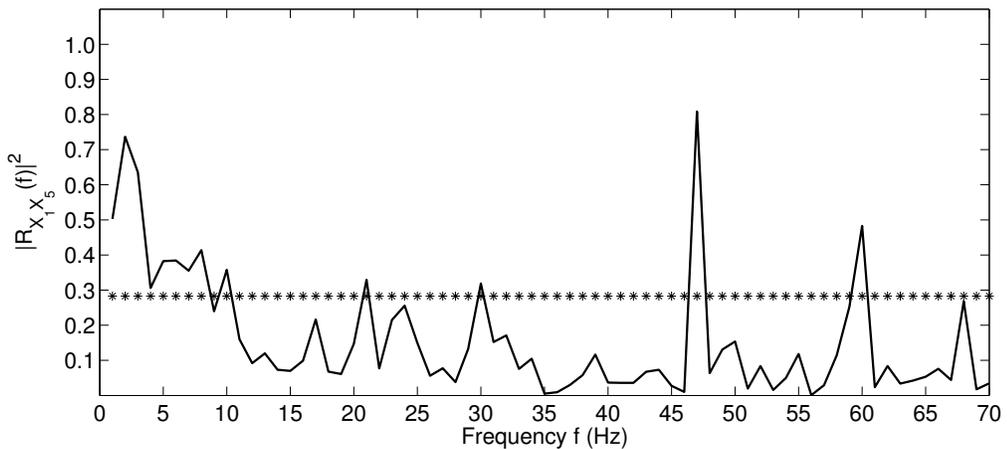
(c)

Figure 6.46: The estimated coherence function, $|\hat{R}_{X_1 X_4}(f)|^2$: (a) the disjoint sections method, (b) the frequency averaging method, (c) the multi-taper method. The estimated coherences are plotted against the frequencies between 1Hz and 70Hz.
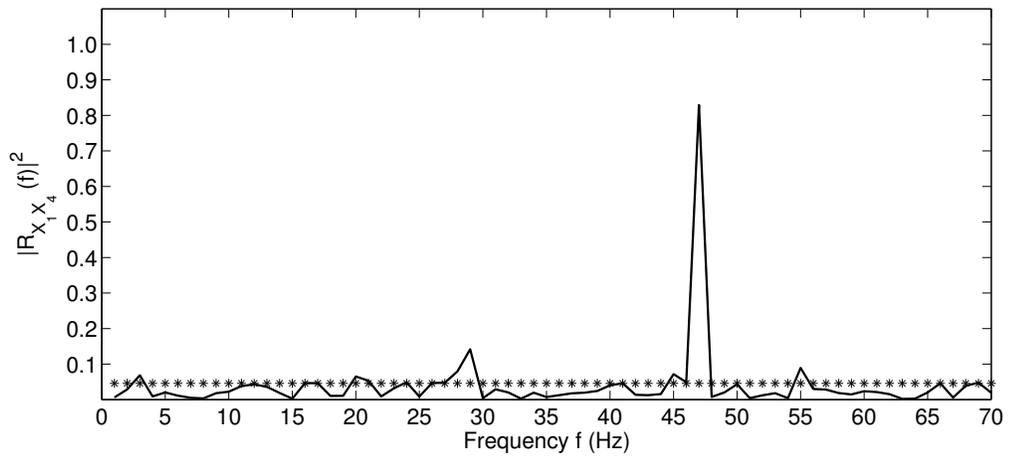
(a)
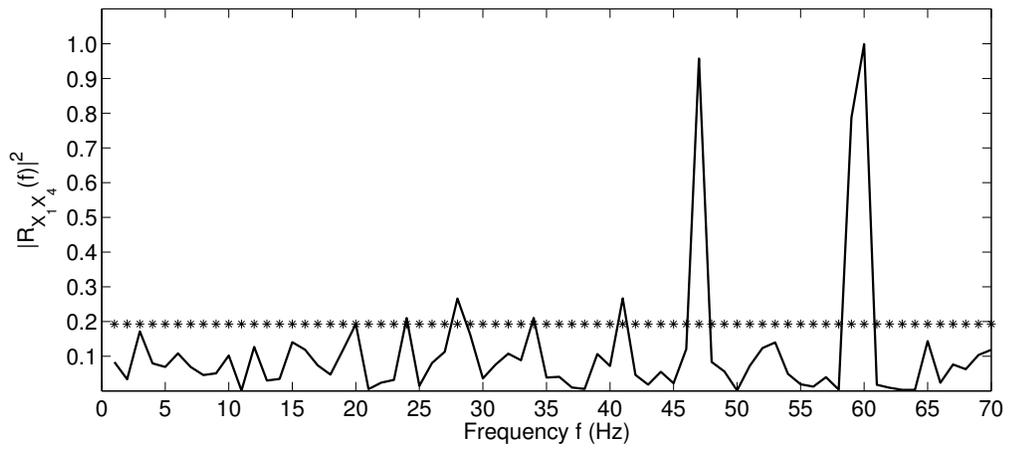
(b)
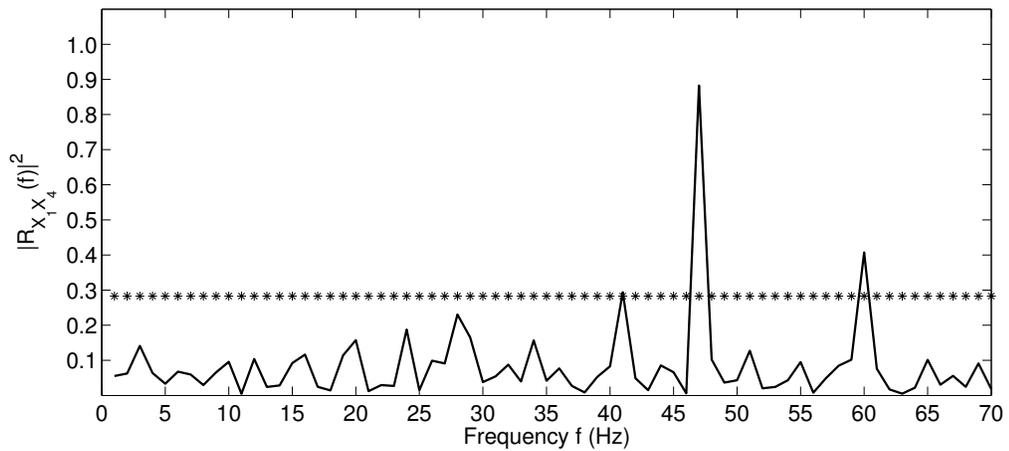
(c)

Figure 6.47: The estimated coherence function, $|\hat{R}_{X_1 X_4}(f)|^2$: (a) the disjoint sections method, (b) the frequency averaging method, (c) the multi-taper method. The estimated coherences are plotted against the frequencies between 1Hz and 70Hz.

Following the coherence discussion, there is a significant estimated coherence between processes X_2 and X_3 within the frequency interval $[1, 12]$ Hz, as displayed in figure (6.48:a) with respect to the disjoint sections method. This figure shows three adjacent peaks; the first one represents the δ-wave in $[1,4)$ Hz; while the second and third peaks lie within the θ-wave frequency band $[4, 8)$ Hz. The estimated coherence decreases sharply between frequencies 7Hz and 8Hz. The α-wave also emerges between 8Hz and 12Hz with relatively low values of the estimated coherence.

For the smoothed periodogram method, the significant estimated coherence of X_2 with X_3 occurs in the frequency band $[1, 8)$ Hz, this band involves two types of the EEG's waves; δ-wave in $[1, 4)$ Hz, and θ-wave in $[4, 8)$ Hz. The highest value of the estimated coherence is 0.6451 at frequency ($f = 4$Hz) in the θ-wave frequency band, this value then sharply decreases to hit the value 0.4006 at frequency ($f = 5$Hz). Figure (6.48:b) shows that the estimated coherence curve become more smooth compared with the curve produced by the disjoint sections method. That is, for small amount of data the smoothed periodogram method becomes more suitable, while the disjoint sections method becomes more consistent estimation when the number of available observations is sufficiently large. These results are broadly demonstrated in Chapter 3.

However, the multi-taper method exhibits high capability in reducing the estimation variance. This result can be clearly seen as a smoothness of the estimated coherence curve in figure (6.48:c). In addition, the process X_2 strongly correlates with the process X_5 at frequencies lying between 1Hz and 13Hz. That is the significant estimated coherence occurs in δ-rhythm band $[1, 4)$ Hz, θ-rhythm band $[4, 8)$ Hz, and α-rhythm band $[8,13]$ Hz. Figure (6.49) shows that the highest value of $|\hat{R}_{X_2 X_5}(f)|^2$ is 0.6494 at frequency ($f = 2$Hz), and this value decreases gradually to reach the value 0.3796 at frequency ($f = 7$Hz). Then the estimated coherence value steeply declines towards the value of 0.2104 at frequency ($f = 8$Hz), following that this value decays gradually towards the significant zero value.

Similarly, the significant estimated coherence is found between processes X_3 and X_5 over the frequency bands $[1,12]$ Hz, $[14,18]$ Hz and $[21, 24)$ Hz, as shown in figure (6.50). In the first frequency band $[1, 12]$ Hz, the estimated coherence appears in the form of δ-waves in $[1, 4)$ Hz, θ-waves in $[4, 8)$ Hz and α-waves in $[8,12]$ Hz. Also, the estimated coherence appears in frequency intervals $[14, 18]$ Hz and $[21, 24)$ Hz, gives rise to β-waves.
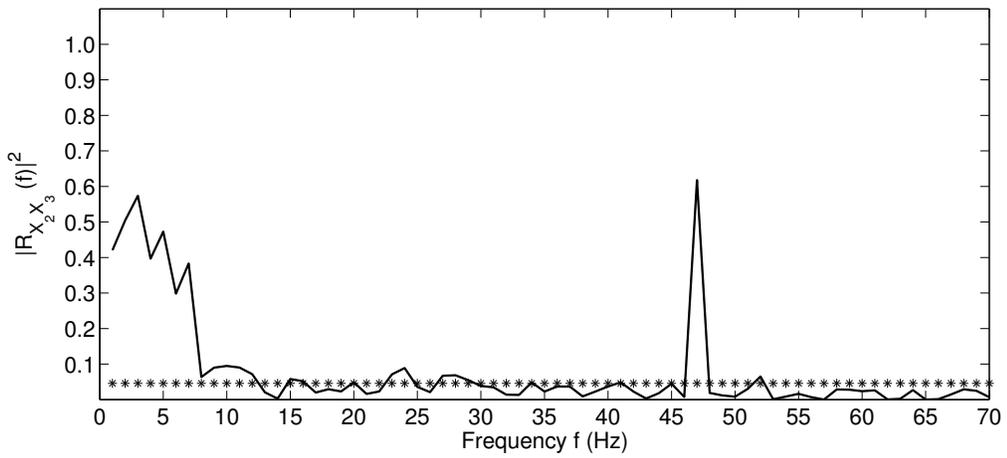
(a)
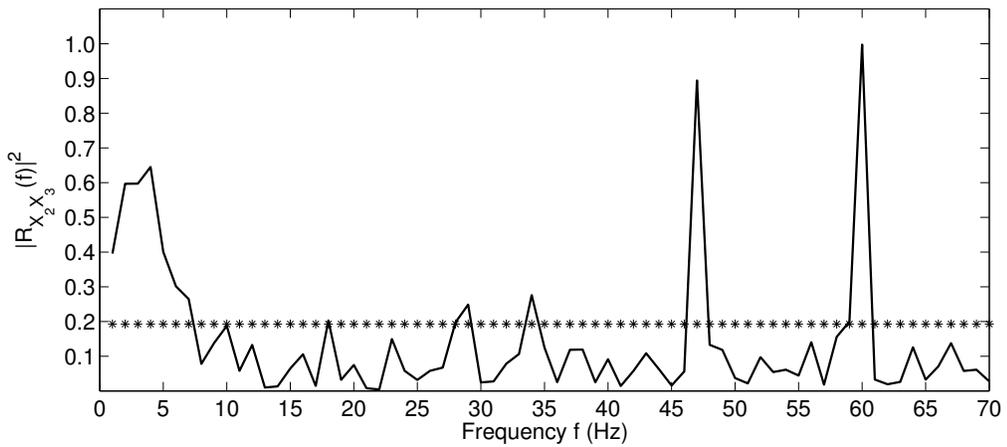
(b)
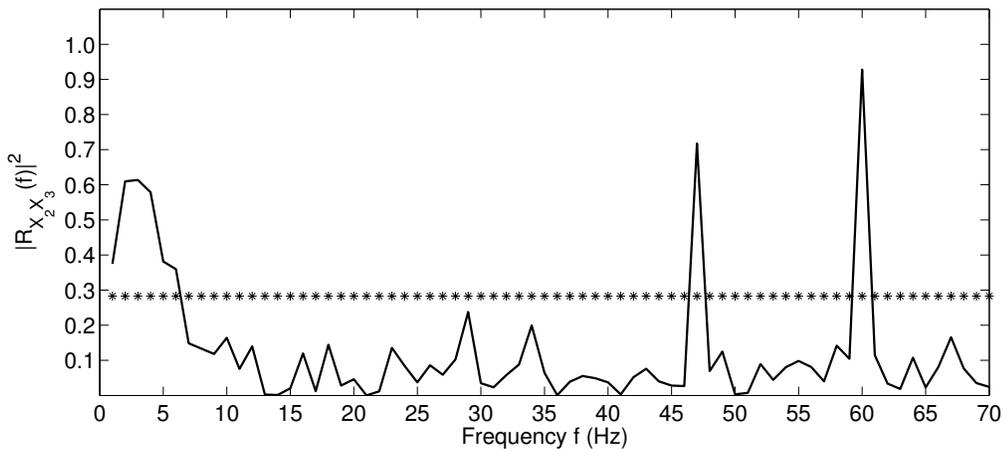
(c)

Figure 6.48: The estimated coherence function, $|\hat{R}_{X_2 X_3}(f)|^2$: (a) the disjoint sections method, (b) the frequency averaging method, (c) the multi-taper method. The estimated coherences are plotted against the frequencies between 1Hz and 70Hz.
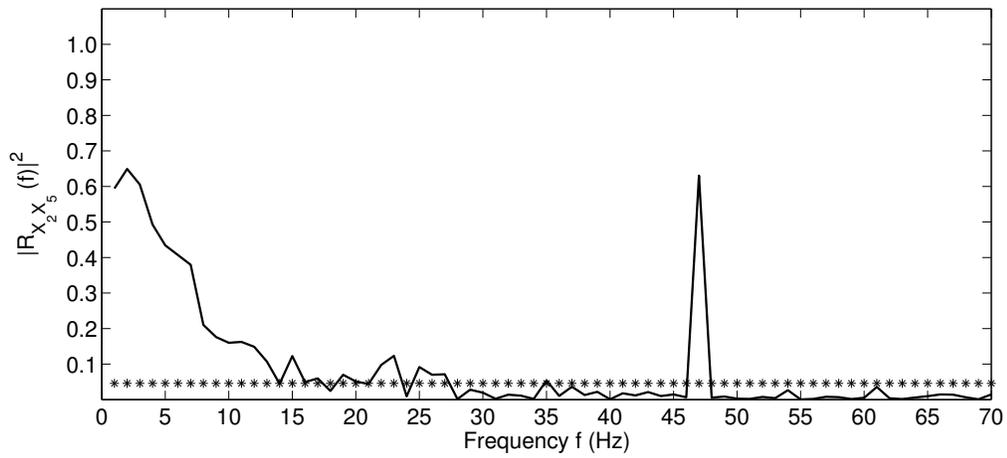
(a)
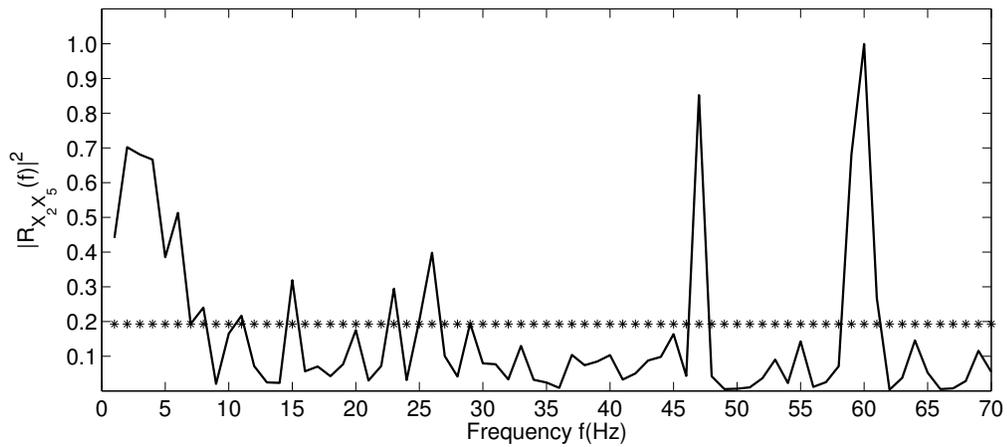
(b)
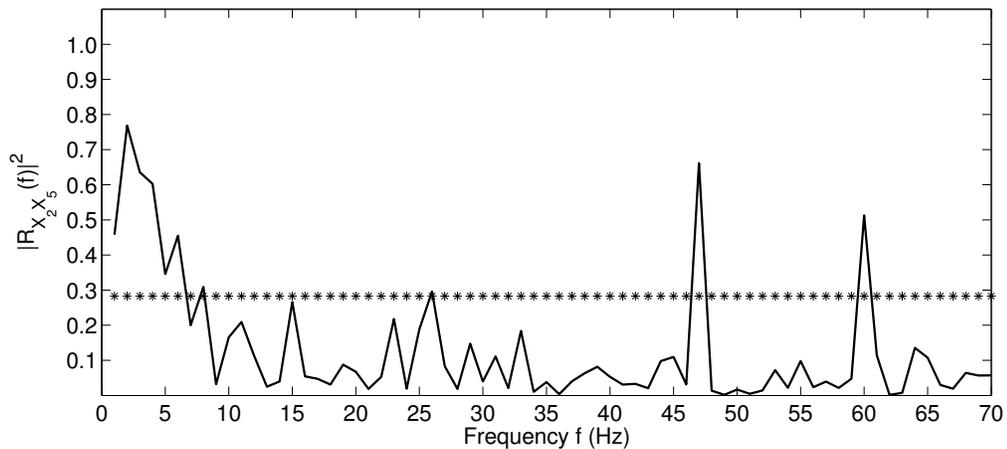
(c)

Figure 6.49: The estimated coherence function, $|\hat{R}_{X_2X_5}(f)|^2$: (a) the disjoint sections method, (b) the frequency averaging method, (c) the multi-taper method. The estimated coherences are plotted against the frequencies between 1Hz and 70Hz.
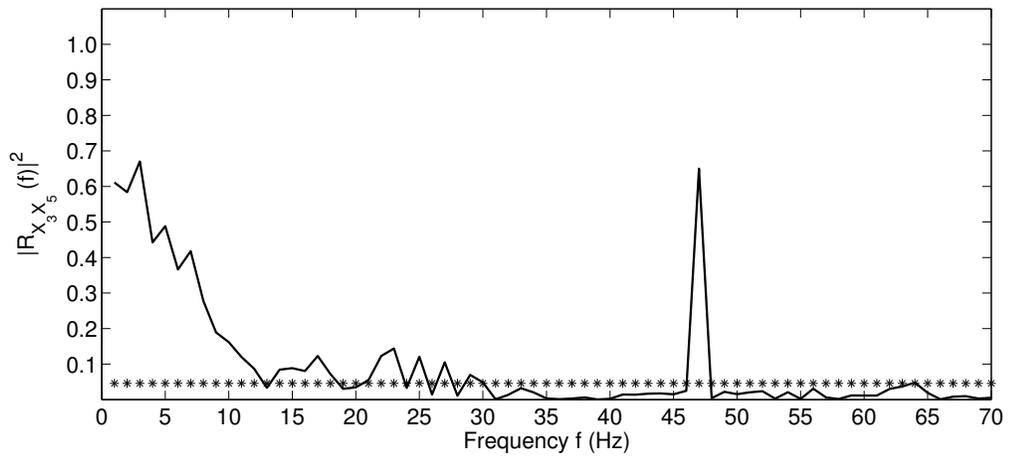
(a)
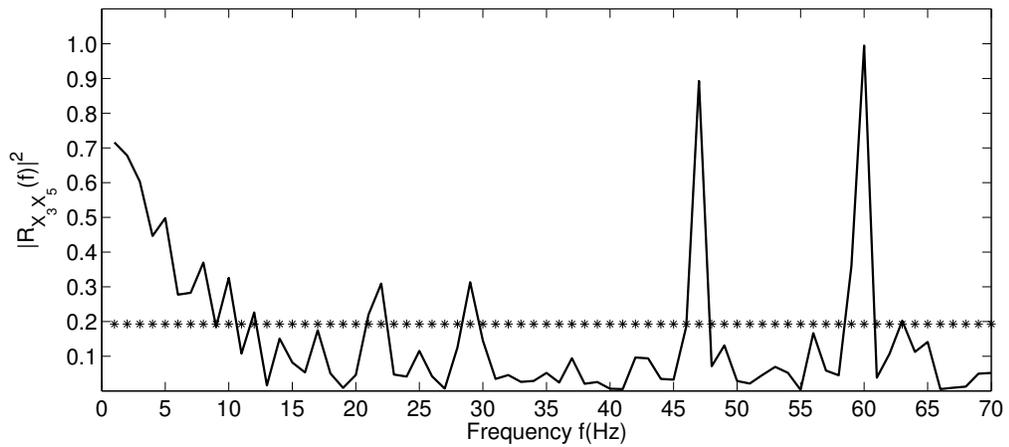
(b)
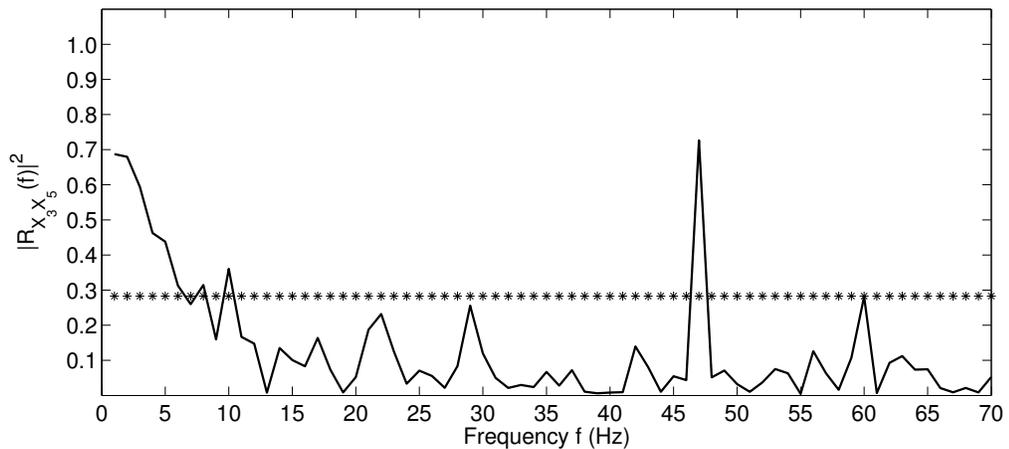
(c)

Figure 6.50: The estimated coherence function, $|\hat{R}_{X_3X_5}(f)|^2$: (a) the disjoint sections method, (b) the frequency averaging method, (c) the multi-taper method. The estimated coherences are plotted against the frequencies between 1Hz and 70Hz.

Up to this point, we have discussed the estimated coherence between two processes among the current neural network, taking into account the linear influences of the rest processes on this specific coherence. We now investigate the estimated coherence between the pair of processes, after removing linear effects of one or two processes, using the partial coherence measurement, that is derived directly from the data, as defined in Chapter 4. The estimated coherence between processes X_1 and X_2, after removing linear influences of process X_3 is illustrated in figure (6.51), where the solid-line represents the estimated ordinary coherence, $|\hat{R}_{X_1 X_2}(f)|^2$, the dashed-line denotes the estimated partial coherence, $|\hat{R}_{X_1 X_2 | X_3}(f)|^2$, and the starred-line denotes the confidence bound of %5 level of significance at frequencies ranging between 1Hz and 70Hz.

For the disjoint sections method, the partial coherence between processes X_1 and X_2, after removing linear influences of process X_3, $|\hat{R}_{X_1 X_2 | X_3}(f)|^2$ at frequency f in [1, 70] Hz, reveals a general decrease in the values of the estimated coherence. EEG's waves, that are found in the ordinary estimated coherence of X_1 with X_2, still emerge in the estimated first order partial coherence, $|\hat{R}_{X_1 X_2 | X_3}(f)|^2$, with less value, as can be seen from figure (6.51:a).

For the smoothed periodogram method, the removal of the linear influences of process X_3 from the estimated coherence of X_1 with X_2 reduces the value of the estimated coherence and creates two separated spikes, with one representing the δ-wave in [1, 4) Hz, and the other representing the θ-wave in [4, 6]. This results shown in figure (6.51:b).

However, the estimated coherence of X_1 with X_2, entirely decays, after removing the linear influences of the process X_3, using the multi-taper method as depicted in figure (6.51:c).
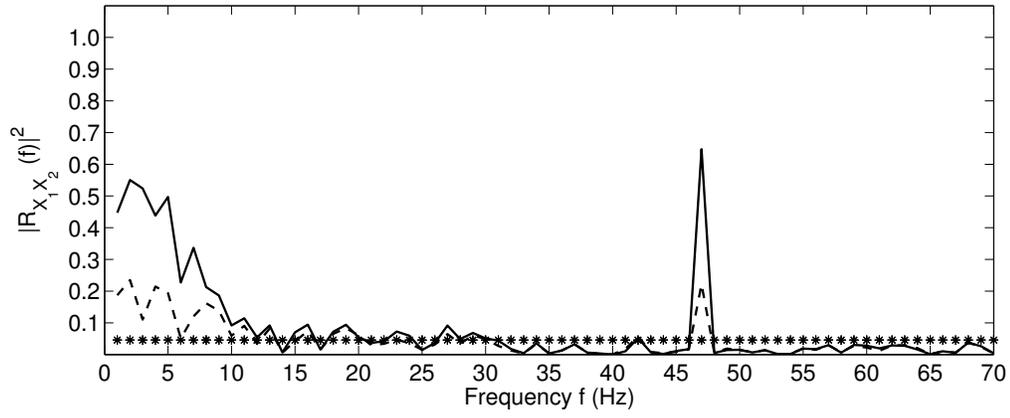
(a)
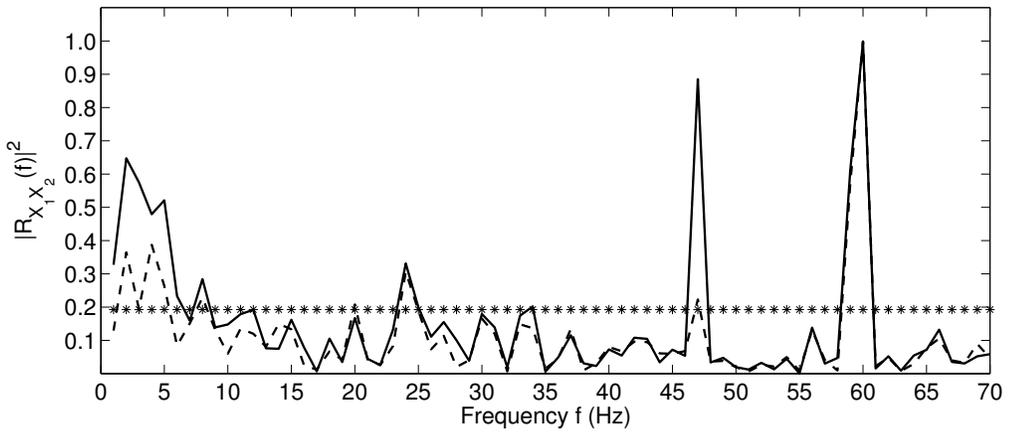
(b)
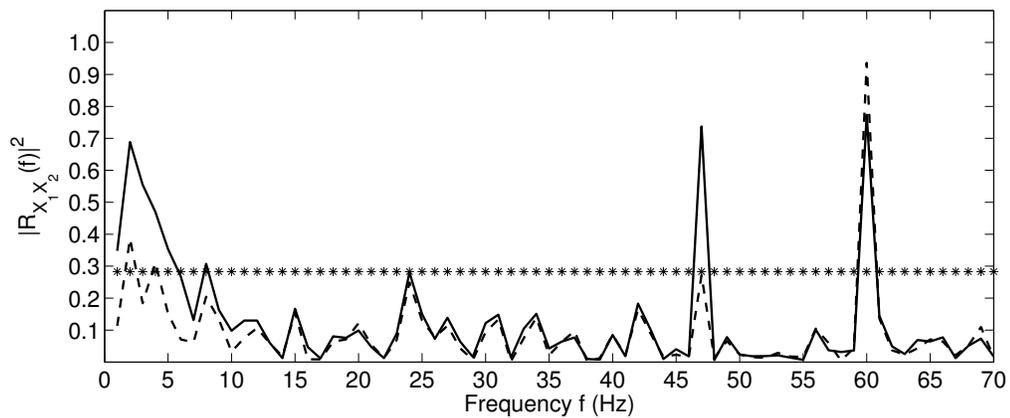
(c)

Figure 6.51: The estimated ordinary coherence, $|\hat{R}_{X_1X_2}(f)|^2$, represented by the solid-line, and the estimated first order partial coherence, $|\hat{R}_{X_1X_2|X_3}(f)|^2$, represented by the dashed-line. he starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

Similarly, figure (6.52) shows the estimated partial coherence of X_1 with X_2, after removing linear influences of the process X_5, depicted by the dashed-line and the estimated ordinary coherence between processes X_1 and X_2, denoted by the solid-line.

Figure (6.52:a) displays the decay of the estimated partial coherence between processes X_1 and X_2, given X_5, calculated by the disjoint sections method. Also, it can be clearly seen that the complete disappearance of the δ- and α- waves in frequency bands $[1, 4)$ Hz and $[8, 13)$ Hz, respectively. Meanwhile, the θ-wave within the frequency band $[4, 8)$ Hz, persists but with relatively small values for the estimated partial coherence.

In addition, the removal of linear influences of X_5 from the estimated coherence between X_1 and X_2 leads to entirely decrease of the estimated coherence using the smoothed periodogram method, shown in figure (6.52:b). the small spike appears at frequency ($f = 5$Hz) is irrelative, as the significant estimated coherence at %5 level of significance requires at least three consecutive frequencies.

The multi-taper method provides a clear view of the entirely decrease of the estimated coherence of X_1 and X_2, after eliminating linear effects of X_5. All values of the estimated partial coherence, $|\hat{R}_{X_1 X_2 | X_5}(f)|^2$, locate beneath the confidence bounds of %5 level of significance. Consequently, since the estimated coherence of X_1 with X_2 is entirely decayed, after removing linear influences of X_3 or X_5, one can say that the depth of the correlation between those process is of order one, using the multi-taper method.
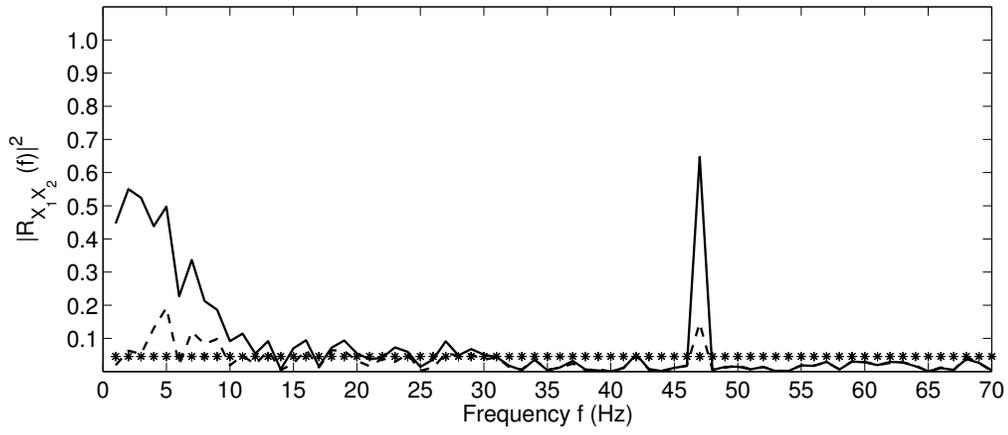
(a)
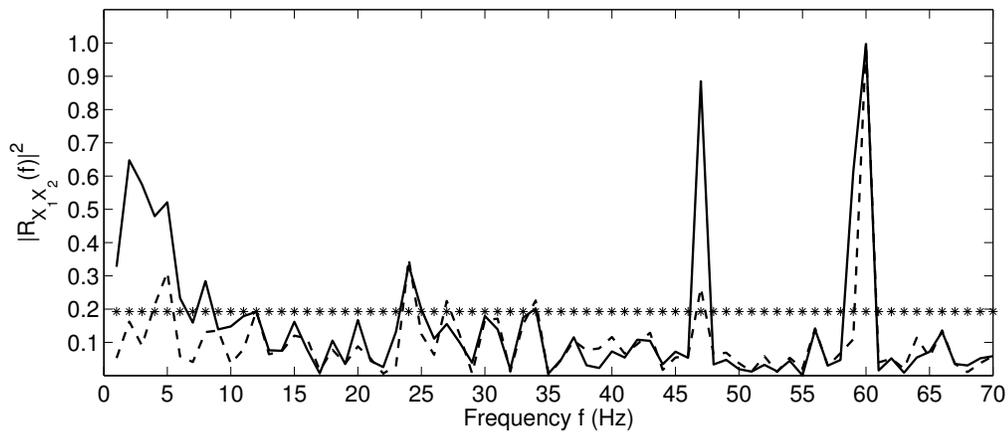
(b)
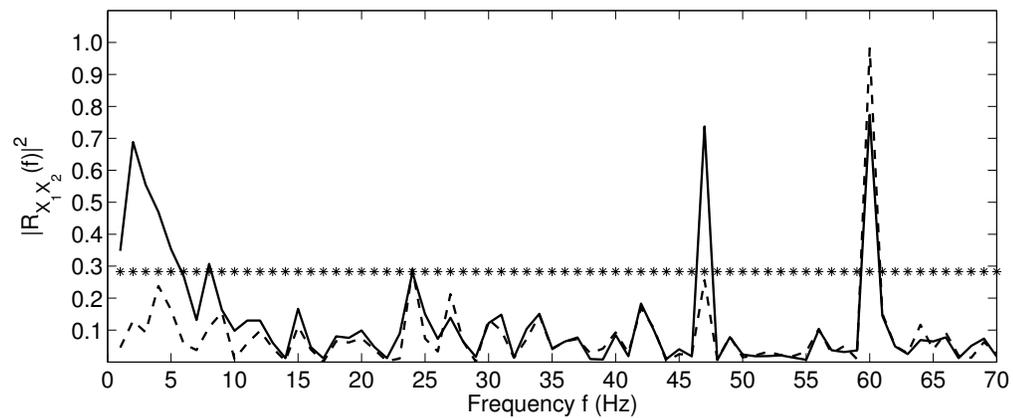
(c)

Figure 6.52: The estimated ordinary coherence, $|\hat{R}_{X_1 X_2}(f)|^2$, represented by the solid-line, and the estimated first order partial coherence, $|\hat{R}_{X_1 X_2 | X_5}(f)|^2$, represented by the dashed-line. he starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

Furthermore, eliminating the linear influences of processes X_3 and X_5 simultaneously, contributes to diminish the values of the estimated coherence between X_1 and X_2, as can be clearly seen in figure (6.53), where the solid-line indicates the estimated ordinary coherence, $|\hat{R}_{X_1 X_2}(f)|^2$, the dashed-line represents the estimated partial coherence, $|\hat{R}_{X_1 X_2 | X_3 X_5}(f)|^2$, and the starred-line denotes the confidence bound of 5% level of significance.

Figure (6.53:a) illustrates that the estimated coherence of X_1 with X_2, after removing the contemporaneous linear effects of X_3 and X_5, is decreased but not completely vanished. That is, the estimated coherence is still statistically significant at the consecutive frequencies 7Hz, 8Hz, and 9Hz within theta and alpha waves bands.

For the smoothed periodogram method, the estimated coherence between processes X_1 and X_2, after eliminating the concurrent linear influences of X_3 and X_5 is completely vanished. Consequently, the depth of correlation between these processes is of order 2. Similarly, the estimated second order partial coherence of X_1 and X_2 is entirely decayed using the multi-taper method.
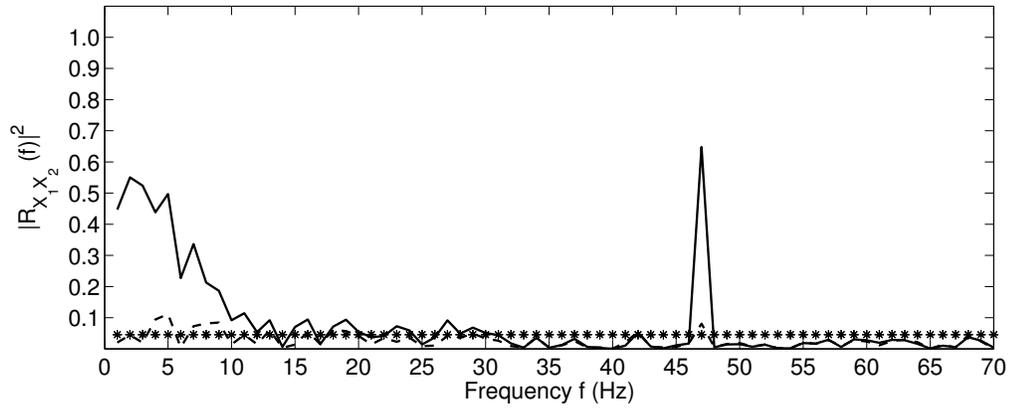
(a)
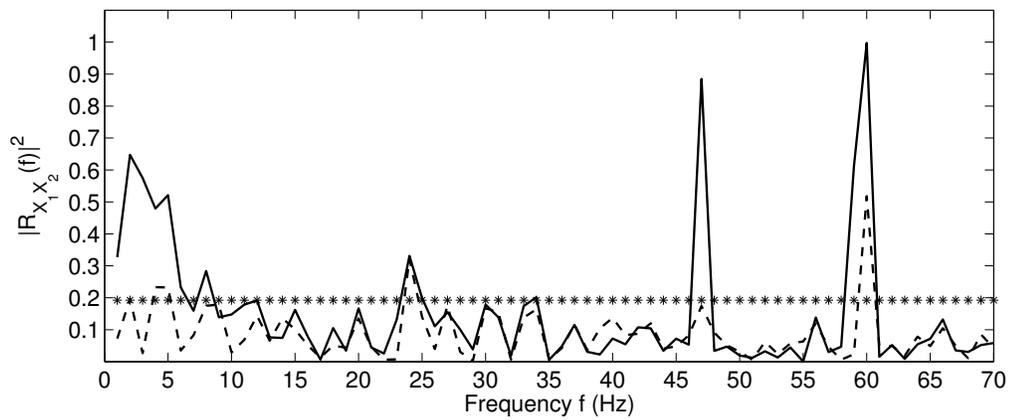
(b)
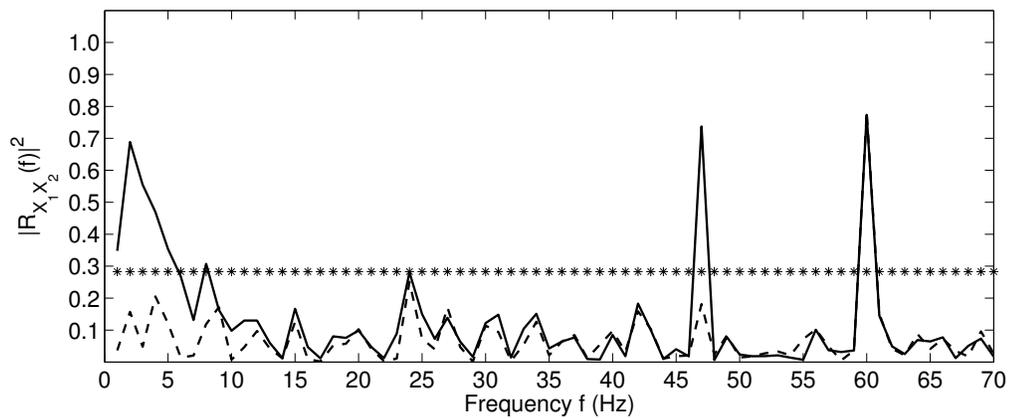
(c)

Figure 6.53: The estimated ordinary coherence, $|\hat{R}_{X_1 X_2}(f)|^2$, represented by the solid-line, and the estimated second order partial coherence, $|\hat{R}_{X_1 X_2 | X_3 X_5}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

The estimated first order partial coherence of X_1 and X_3, given X_2, depicted in figure (6.54), where the solid-line represents the estimated ordinary coherence, $|\hat{R}_{X_1 X_3}(f)|^2$, the dashed-line denotes the estimated partial coherence $|\hat{R}_{X_1 X_3|X_2}(f)|^2$, and the starred-line indicated the confidence bound of %5 level of significance. Figure (6.54:a) shows that the reduction in the value of the estimated coherence as a result of eliminating the linear influences of X_2, with respect to the disjoint sections method.

The estimated partial coherence, $|\hat{R}_{X_1 X_3|X2}(f)|^2$, calculated by the smoothed periodogram method and presented in figure (6.54:b) illustrates that the decrease of the estimated coherence values and completely disappearance of θ-waves in [4, 8) Hz. However, there is an entirely decay of the estimated coherence of X_1 and X_3, after removing linear influences of X_2, using the multi-taper method.

Similarly, although the removal of linear influences of X_5 from the estimated coherence of X_1 with X_3, reduces the values of the estimated coherence produced by the disjoint sections method, it contributes to vanish the estimated coherence between processes X_1 and X_3, with respect to the smoothed periodogram and multi-taper methods. That is $|\hat{R}_{X_1 X_3|X_5}|^2$ is significantly zero. Thus, the depth of the correlation between these two processes is of order one for the two latter methods. Figure (6.55) depicts the estimated ordinary and partial coherences of X_1 and X_3, after eliminating linear influences of X_5. Where the solid-line represents the estimated ordinary coherence, the dashed-line denotes the estimated partial coherence, $|\hat{R}_{X_1 X_3|X_5}|(f)^2$, and the starred-line indicates the confidence bound of %5 level of significance.

In addition, the removal of the simultaneous linear effects of X_2 and X_5 from the estimated coherence of X_1 with X_3, displayed in figure (6.56), leads to the complete destruction of the estimated coherence between these two processes, using the disjoint sections method. Thus the depth of the correlation between these two processes is of order two, with respect to this method. Moreover, the estimated second partial coherence, $|\hat{R}_{X_1 X_3|X_2 X_5}(f)|^2$, remains significantly zero, where all values of the estimated coherence lie beneath the confidence bound of %5 level of significance for the smoothed periodogram and multi-taper methods, as illustrated in figures (6.56:b) and (6.56:c), respectively.
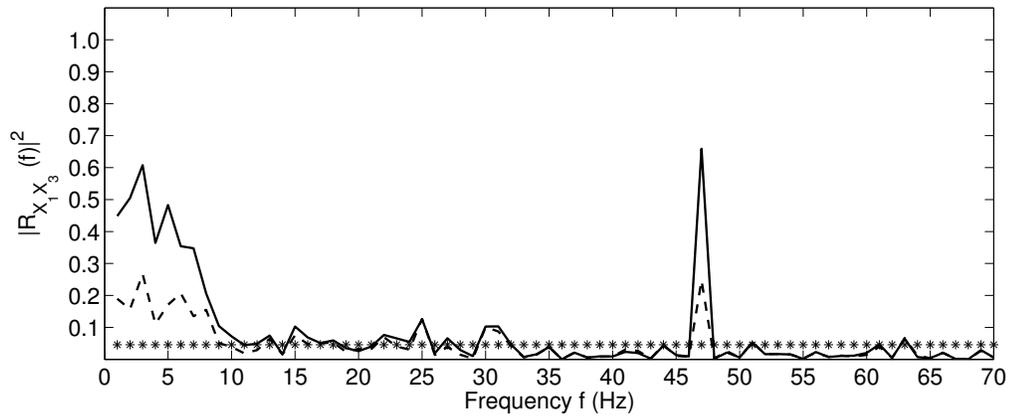
(a)
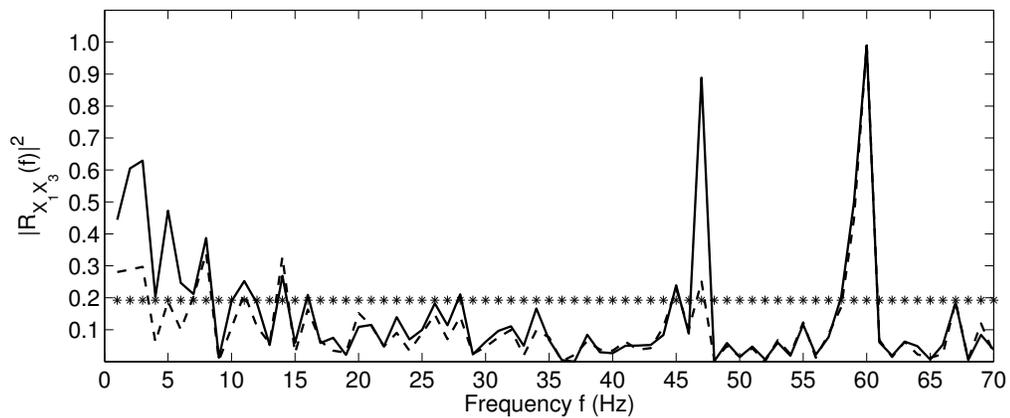
(b)
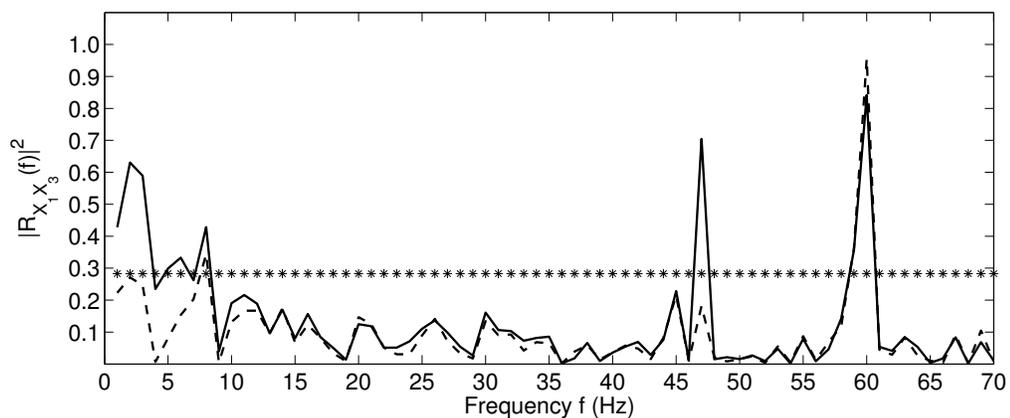
(c)

Figure 6.54: The estimated ordinary coherence, $|\hat{R}_{X_1 X_3}(f)|^2$, represented by the solid-line, and the estimated first order partial coherence, $|\hat{R}_{X_1 X_3 | X_2}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.
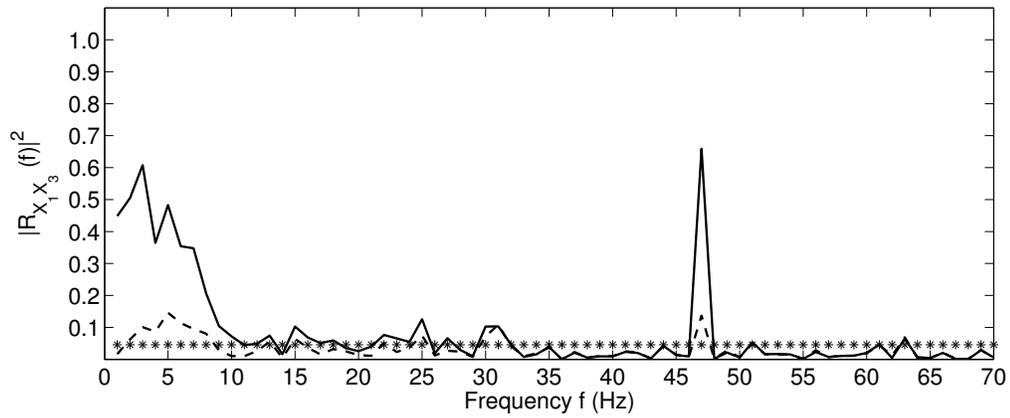
(a)
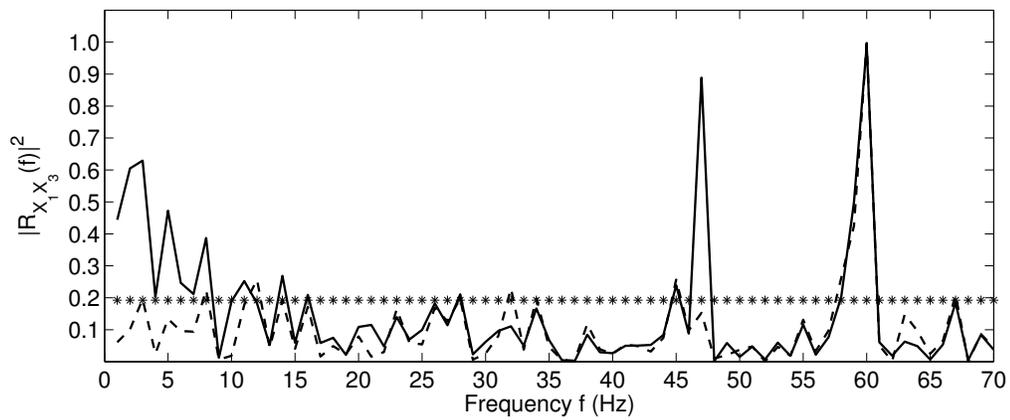
(b)
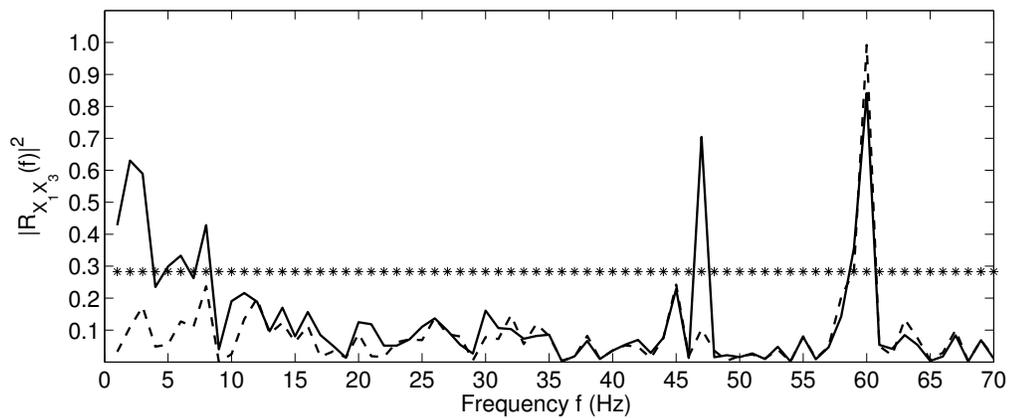
(c)

Figure 6.55: The estimated ordinary coherence, $|\hat{R}_{X_1 X_3}(f)|^2$, represented by the solid-line, and the estimated first order partial coherence, $|\hat{R}_{X_1 X_3 | X_5}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.
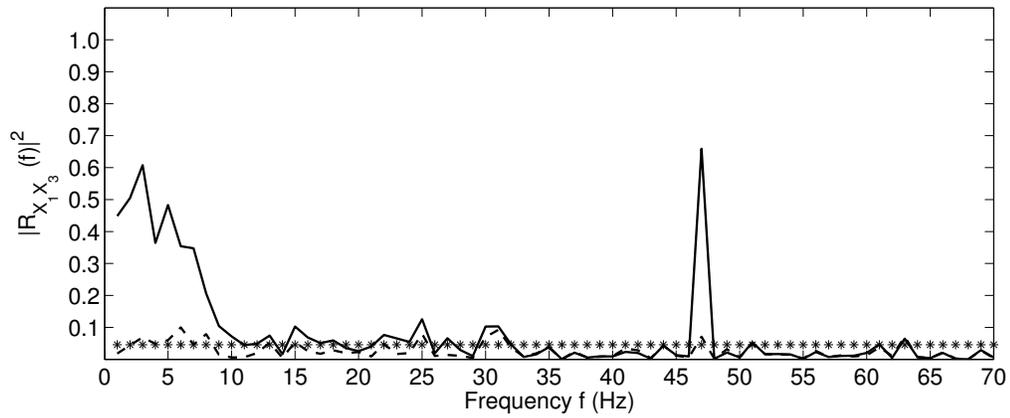
(a)
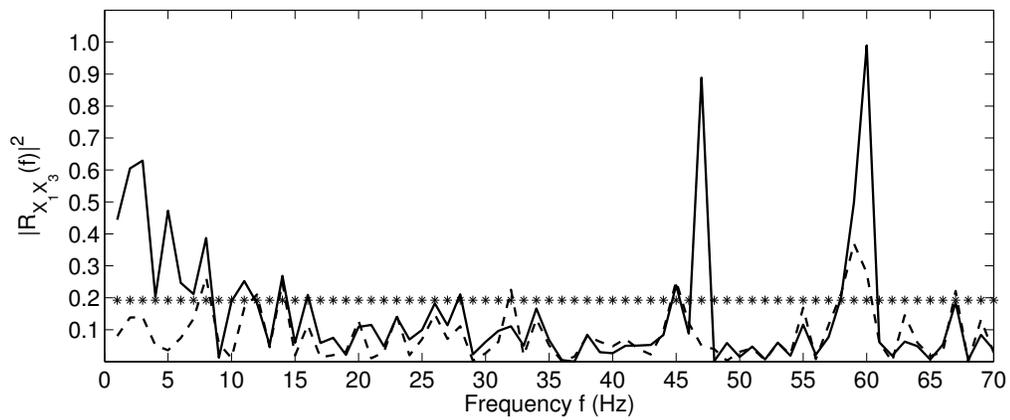
(b)
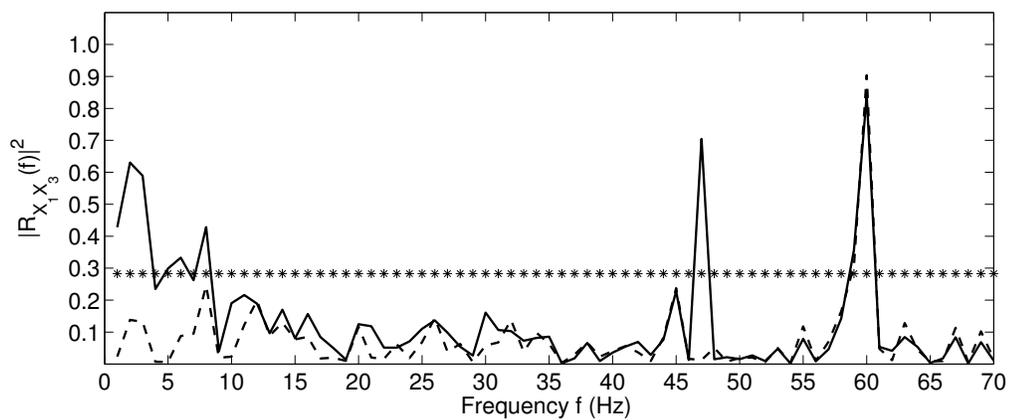
(c)

Figure 6.56: The estimated ordinary coherence, $|\hat{R}_{X_1 X_3}(f)|^2$, represented by the solid-line, and the estimated second order partial coherence, $|\hat{R}_{X_1 X_3 | X_2 X_5}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

Regarding the estimated coherence between processes X_2 and X_3, calculated by the disjoints sections method, the elimination of linear influences of the process X_1 (or X_5), contributes to reduced values of the estimated coherence, and to the elimination of the α-wave, which ranges between frequencies 8Hz and 13Hz.

Continuing, the removal of linear influences of process X_1 (or X_5) from the estimated coherence of processes X_2 and X_3, using the smoothed periodogram method, leads to a reduction in the values of the estimated coherence with a disappearance of the θ-wave in [4, 8) Hz. That is, the significant estimated first order partial coherence between processes X_2 and X_3, conditioned on X_1 exists only in the frequency band [1, 4) Hz.

However, eliminating linear influences of X_1 (or X_5) causes a completely destruction of the estimated coherence of X_2 with X_3. That is, $|\hat{R}_{X_2 X_3 | X_1}(f)|^2$ and $|\hat{R}_{X_2 X_3 | X_5}(f)|^2$ are significantly zero, using the multi-taper method. Figure (6.57) illustrates the estimated coherence between X_2 and X_3, after removing the linear effects of X_1. Where the ordinary coherence, represented by the solid-line, the estimated partial coherence $|\hat{R}_{X_2 X_3 | X_1}(f)|^2$, denoted by the dashed-line, and the starred-line indicates the confidence bound of %5 level of significance. The small spike appears at frequency 4Hz, in figure (6.57:c) is irrelative as we need at least three consecutive values of significant estimated coherence.

Furthermore, the removal of the simultaneous linear effects of X_1 and X_5 from the estimated coherence of X_2 with X_3, leads to a further reduction in the values of estimated coherence. Figure (6.58:a) displays the massive reduction of the estimated coherence of X_2 and X_3, where the significant estimated coherence only appears in θ-waves band with relatively small values.

The estimated second partial coherence, $|\hat{R}_{X_2 X_3 | X_1 X_5}(f)|^2$, calculated by the smoothed periodogram and multi-taper methods are significantly zero, where the small spike emerges at 4Hz in figure (6.58:b), is negligible.
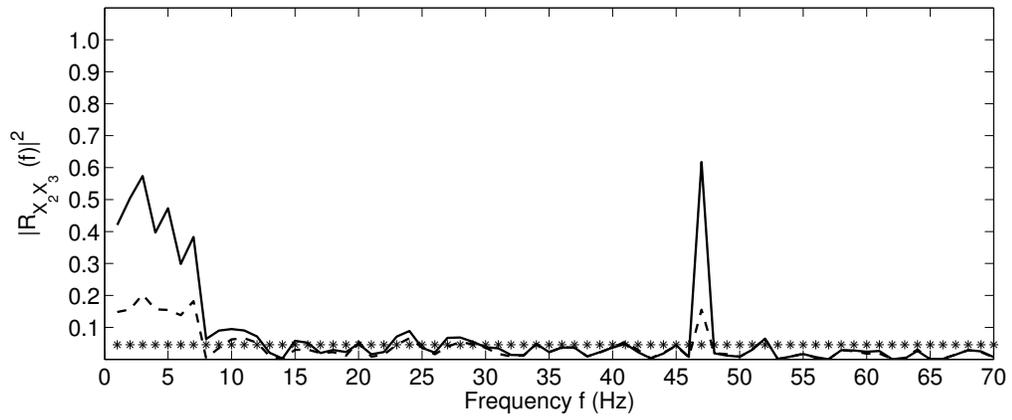
(a)
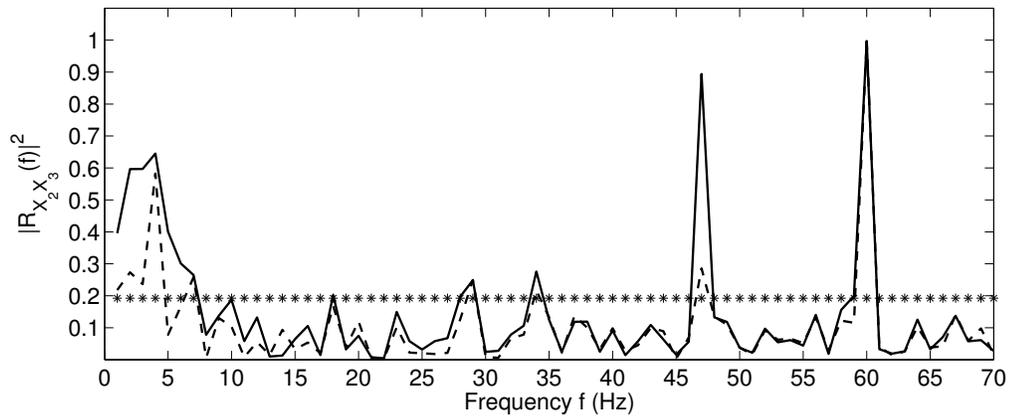
(b)
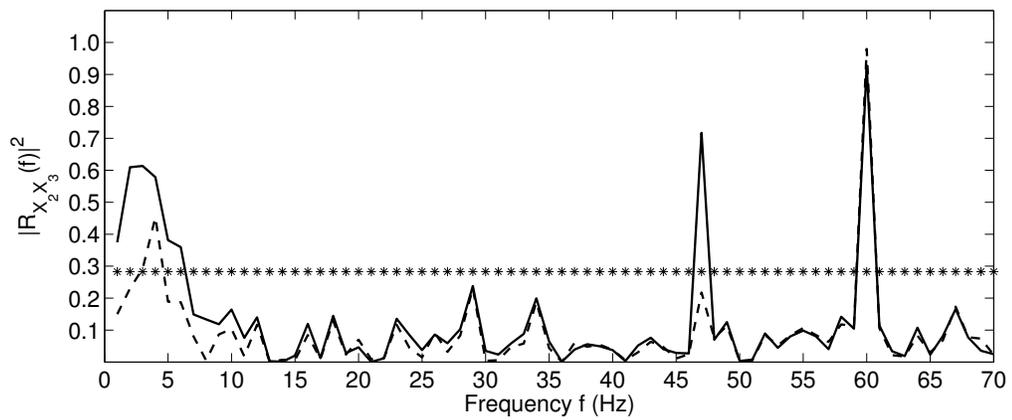
(c)

Figure 6.57: The estimated ordinary coherence, $|\hat{R}_{X_2X_3}(f)|^2$, represented by the solid-line, and the estimated first order partial coherence, $|\hat{R}_{X_2X_3|X_1}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.
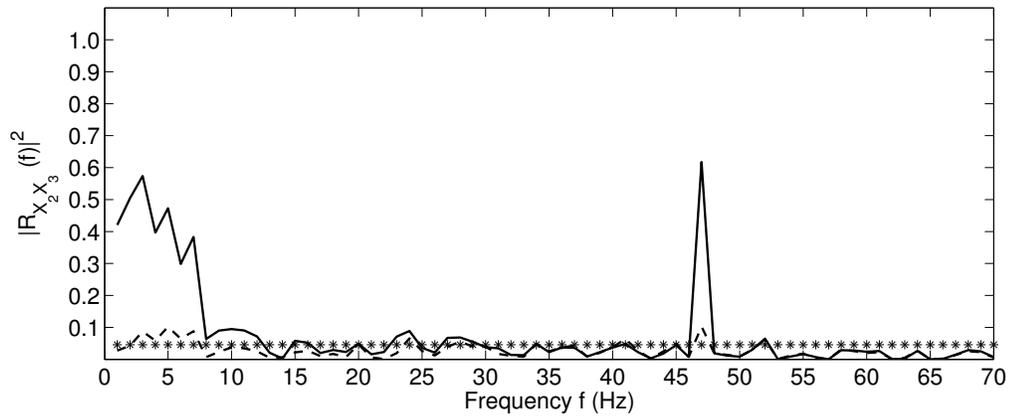
(a)
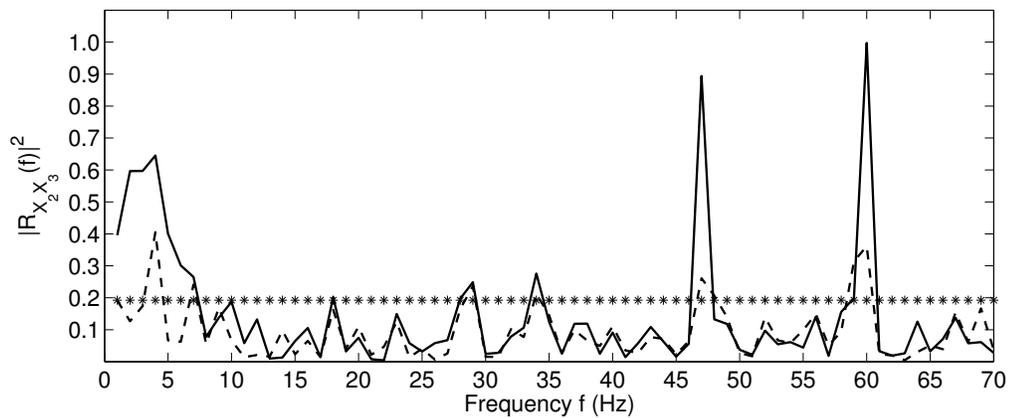
(b)
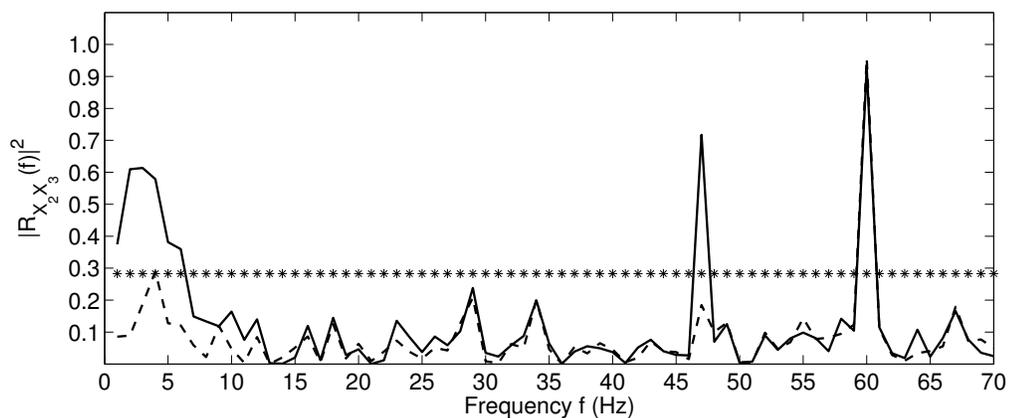
(c)

Figure 6.58: The estimated ordinary coherence, $|\hat{R}_{X_2X_3}(f)|^2$, represented by the solid-line, and the estimated second order partial coherence, $|\hat{R}_{X_2X_3|X_1X_5}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of $\%5$ level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

Continuing the discussion of estimated partial coherences, the process X_5 possesses strong and non-vanishing estimated coherences with processes $\{X_1, X_2, X_3\}$, despite of the removal of linear influences of one or two processes from the ordinary estimated coherences using the disjoint sections method.

For the smoothed periodogram method, the elimination of linear influences of one process from the estimated ordinary coherence of X_5 with any of the processes $\{X_1, X_2, X_3\}$ leads to reduce the values of the significant estimated coherence. However, the removal of the synchronized linear effects of two processes from the estimated coherence between X_5 and any of the processes $\{X_1, X_2, X_3\}$, causes the entire reduction of the estimated coherence. That is, the estimated second order partial coherences with respect to X_5 are significantly zero. Consequently, we can say that the depth of correlations between pairs (X_1, X_5), (X_2, X_5), and (X_3, X_5), is of order two.

The estimated first order partial coherences of X_5 with any of processes $\{X_1, X_2, X_3\}$, using the multi-taper method, are significantly zero. Thus the depth of correlations of pairs (X_1, X_5), (X_2, X_5), and (X_3, X_5), is of order one for this method.

Figure (6.59) illustrates the estimated first order partial coherence $|\hat{R}_{X_1 X_5 | X_2}(f)|^2$, plotted by the dashed-line, the estimated ordinary coherence, represented by the solid-line, and the confidence bound of $\%5$ level of significance, marked by the starred-line.

The removal of linear effects of X_2 from the estimated coherence of X_1 with X_5, calculated by the disjoint section method, reduces the values of the significant estimated coherence, as shown in figure (6.59:a). The estimated first order partial coherence of X_1 with X_5 conditioned on X_2, $|\hat{R}_{X_1 X_5 | X_2}(f)|^2$, calculated by the smoothed periodogram method, occurs in two separated bands of frequency, one associated with δ-waves, and the second, related to θ-waves, as shown in figure (6.59:b). Specifically, the first part of the estimated partial coherence curve locates in frequency band $[1, 3]$ Hz, while the second part of the estimated partial coherence curve lies between 5Hz and 8Hz.

In addition, the estimated first order partial coherence of X_1 with X_5, after removing linear influences of X_2, produced by the multi-taper method, is significantly zero as shown in figure (6.59:c). Similarly, although the removal of linear effects of process X_3 from the estimated coherence of X_1 with X_5 decreases the values of the significant estimated coherence, produced by the disjoint sections method, it causes an entire decay of the estimated partial coherence, $|\hat{R}_{X_1 X_5 | X_3}(f)|^2$, calculated by the smoothed periodogram or multi-taper methods.
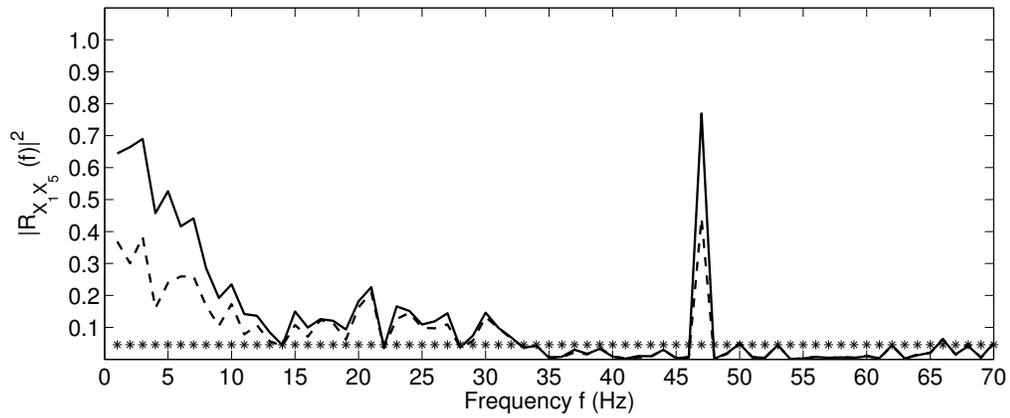
(a)
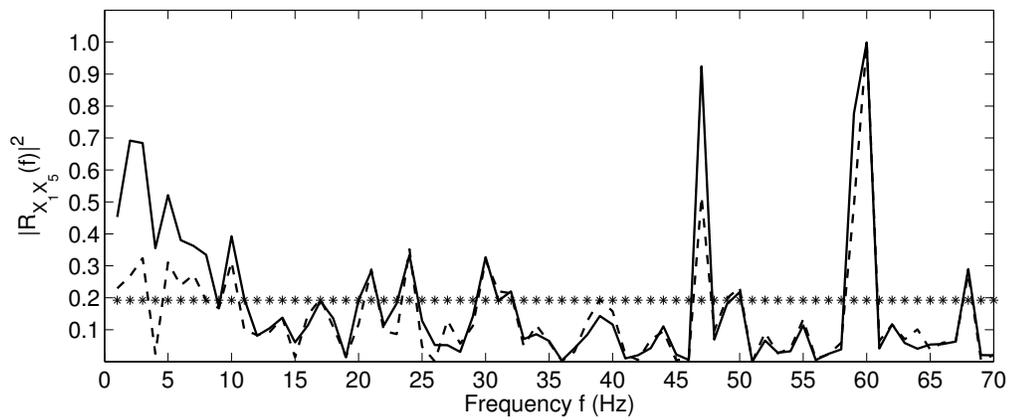
(b)
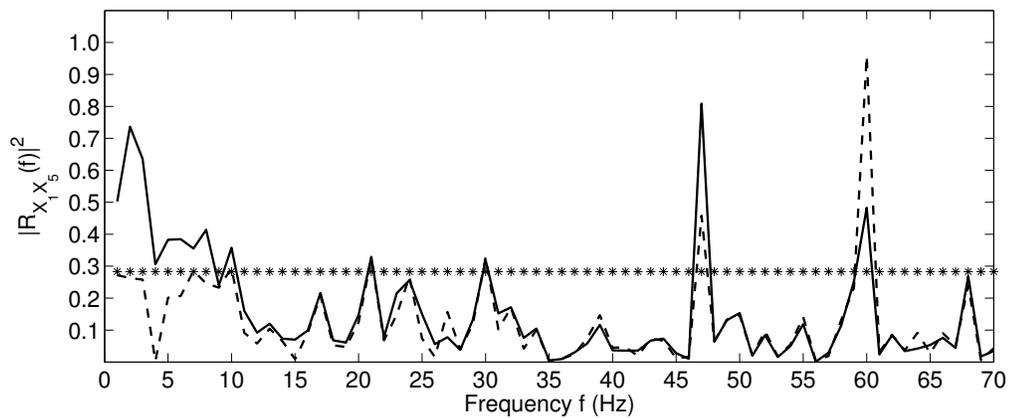
(c)

Figure 6.59: The estimated ordinary coherence, $|\hat{R}_{X_1 X_5}(f)|^2$, represented by the solid-line, and the estimated first order partial coherence, $|\hat{R}_{X_1 X_5 | X_2}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

The estimated second order partial coherences of X_1 with X_5, calculated by the three non-parametric estimations; the disjoint sections method; the smoothed periodogram method, and the multi-taper method, are displayed in figures (6.60:a), (6.60:b), and (6.60:c) respectively. Figure (6.60:a) displays the further decrease of the estimated second order partial coherence of X_1 with X_5 due to the elimination of the synchronized linear influences of X_2 and X_3. However, there is a complete decay of the estimated second order coherence, $|\hat{R}_{X_1X_5|X_2X_3}(f)|^2$, calculated by the smoothed periodogram method, as shown in figure (6.60:b).

For the multi-taper method, as discussed previously, the elimination of linear influences of X_2 (or X_3) from the estimated coherence of X_1 with X_5, contributes to destroy the estimated coherence. That is, $|\hat{R}_{X_1X_5|X_2}(f)|^2$, and $|\hat{R}_{X_1X_5|X_3}(f)|^2$ are significantly zero. Therefore, the removal of the concurrent linear effects of X_2 and X_3 from the estimated coherence of X_1 with X_5 is expected to decay this correlation, as displayed in figure (6.60:c).

Typical results are detected for the estimated coherence of X_2 with X_5, and X_3 with X_5, after removing either linear influences of one process or the simultaneous linear influences of two processes. However, the reduction caused by the removal of the concurrent linear influences of two processes is larger.

The estimated first order partial coherence, $|\hat{R}_{X_2X_5|X_1}(f)|^2$, illustrated in figure (6.61), where the solid-line represents the estimated ordinary coherence, the dashed-line denotes the first order partial coherence of X_2 with X_5 conditioned on X_1, and the starred-line indicates the confidence bound of %5 level of significance.

It can be clearly seen that the removal of linear influences of X_1 from the estimated coherence between X_2 and X_5 reduces the values of the estimated coherence, calculated by the disjoint sections method, as displayed in figure (6.61:a). In addition, the estimated first order partial coherence, $|\hat{R}_{X_2X_5|X_1}(f)|^2$, produced by the smoothed periodogram, appears only in the δ-wave range [1, 4) Hz with a complete disappearance of the θ- wave that is found in the estimated ordinary coherence. Again the small spike rises at 6Hz is negligible, as it is required at least three consecutive significant coherence values to say that there is a significant correlation between those processes.
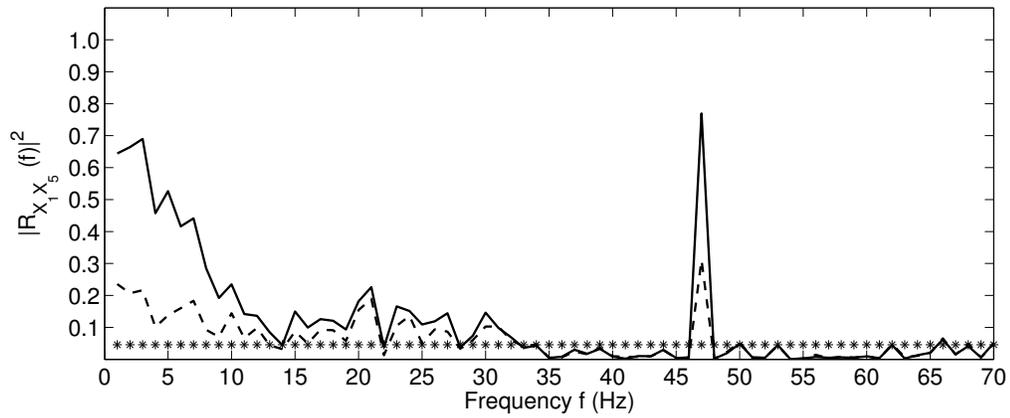
(a)
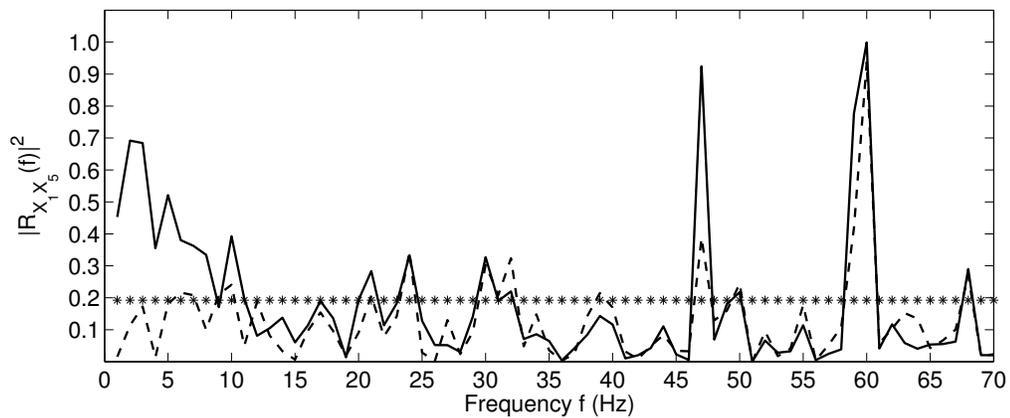
(b)
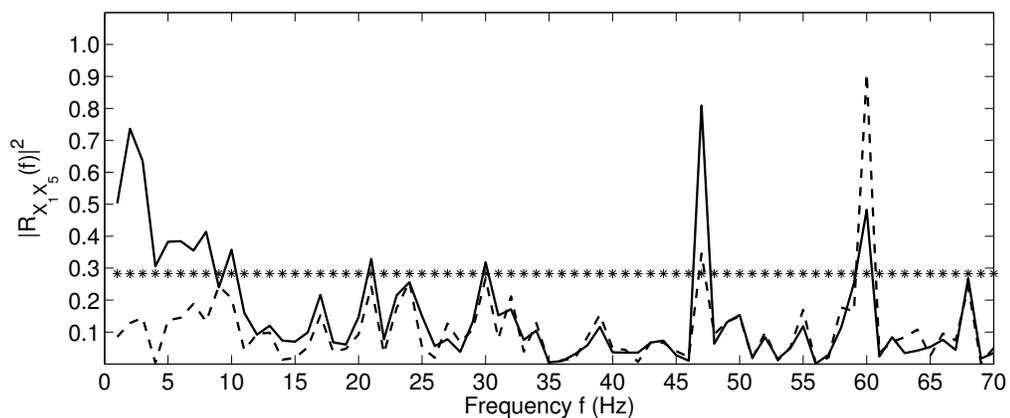
(c)

Figure 6.60: The estimated ordinary coherence, $|\hat{R}_{X_1X_5}(f)|^2$, represented by the solid-line, and the estimated second order partial coherence, $|\hat{R}_{X_1X_5|X_2X_3}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

However, the first order partial coherence $|\hat{R}_{X_2X_5|X_1}(f)|^2$, calculated by the multi-taper method is significantly zero, as illustrated in figure (6.60:c). Therefore, we can say that the depth of the correlation between processes X_2 and X_5 is of order one with respect to the latter two methods. The elimination of linear effects of X_3 from the estimated coherence of X_2 with X_5, gives the similar results of removing linear influences of X_1 for the three estimation methods.

The removal of the concurrent linear influences of X_1 and X_3 from the estimated coherence of X_2 with X_5, displayed in figure (6.61), where the solid-line represents the estimated ordinary coherence, the dashed-line denotes the estimated second order partial coherence, $|\hat{R}_{X_2X_5|X_1X_3}(f)|^2$, and the starred-line indicates the confidence bound of %5 level of significance.

The estimated second partial coherence, $|\hat{R}_{X_2X_5|X_1X_3}(f)|^2$, calculated by the disjoint sections method, occurs in two separated regions; with one appears in the δ-wave band $[1, 4)$ Hz, and the other part extends between the θ-wave in $[4, 8)$ Hz and the α-wave in $[8, 13)$ Hz. For the smoothed periodogram method, the elimination of linear effects of X_1 and X_3 from the estimated coherence causes a complete decay of the estimated coherence, where the small values emerge at 2Hz, 3Hz, and 6Hz are negligible, as shown in (6.62:b). That is, $|\hat{R}_{X_2X_5|X_1X_3}(f)|^2$ is significantly zero, which is identical to the result obtained by the multi-taper method and depicted in figure (6.62:c).
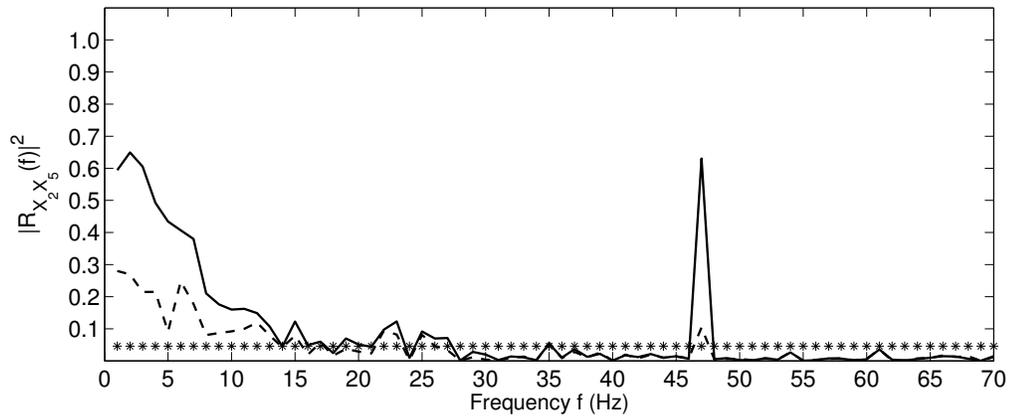
(a)
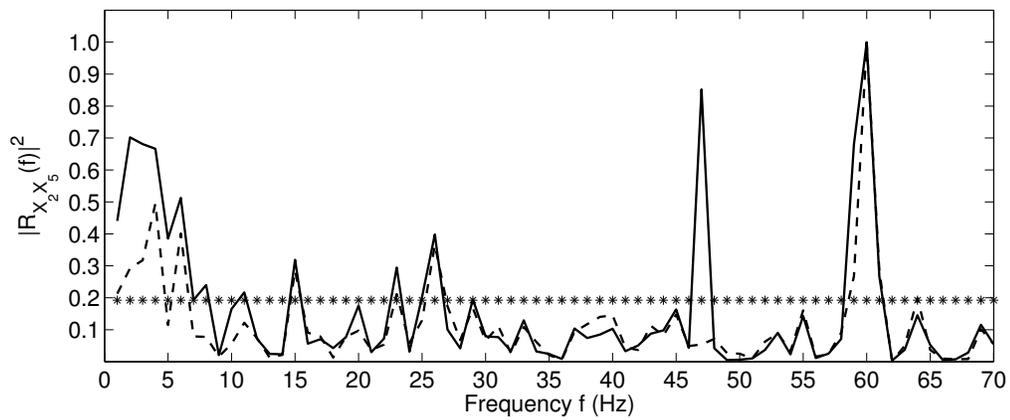
(b)
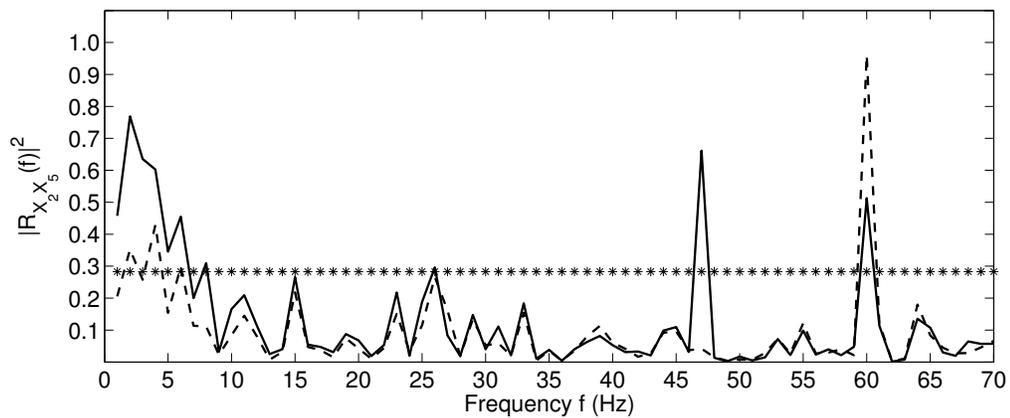
(c)

Figure 6.61: The estimated ordinary coherence, $|\hat{R}_{X_2X_5}(f)|^2$, represented by the solid-line, and the estimated first order partial coherence, $|\hat{R}_{X_2X_5|X_1}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.
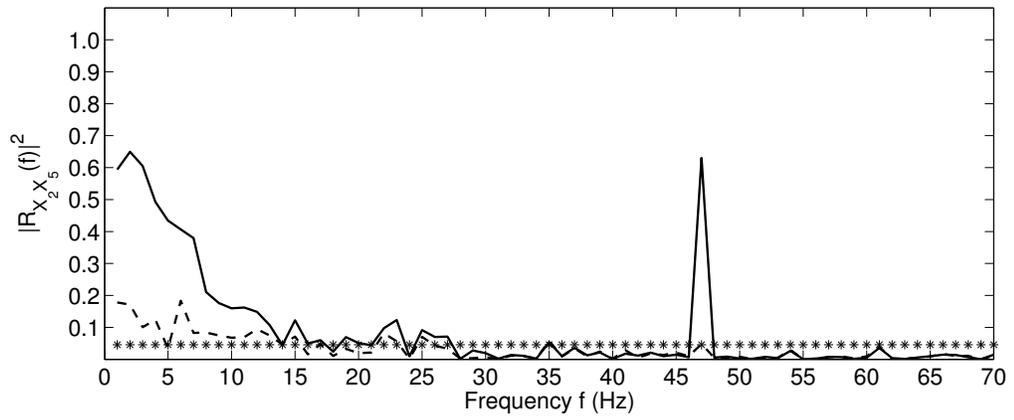
(a)
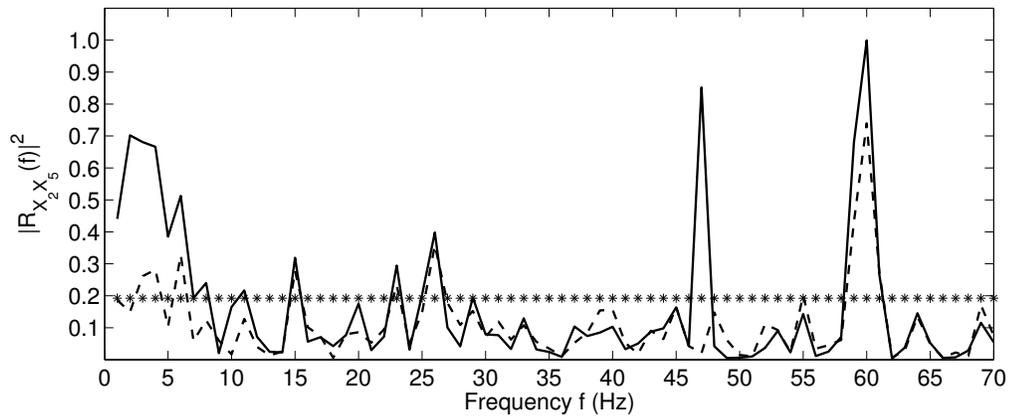
(b)
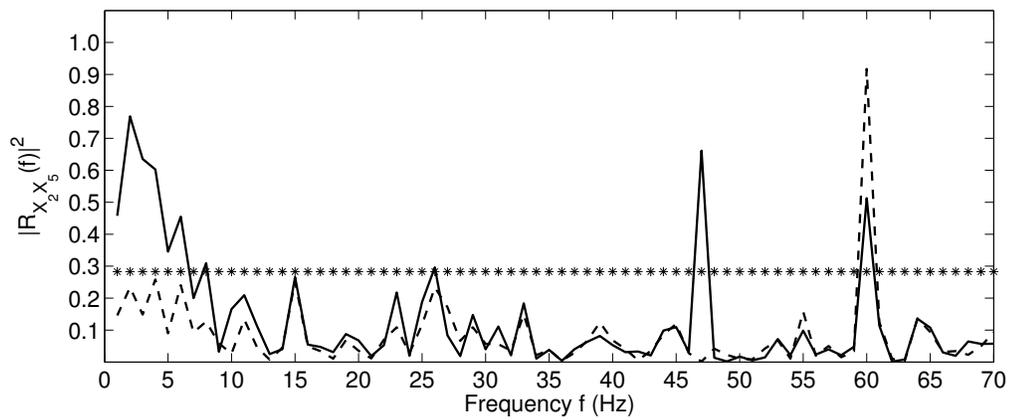
(c)

Figure 6.62: The estimated ordinary coherence, $|\hat{R}_{X_2 X_5}(f)|^2$, represented by the solid-line, and the estimated second order partial coherence, $|\hat{R}_{X_2 X_5 | X_1 X_3}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

Continuing, The estimated first order partial coherences of X_3 with X_5, after removing linear influence of X_1, calculated by the disjoint sections, the smoothed periodogram, and the multi-taper methods, are displayed in figures (6.63:a), (6.63:b), (6.63:c), respectively. The removal linear influences of X_1 from the estimated coherence of X_3 with X_5, decreases the values of the significant estimated coherence, and causes the entire disappearance of the β-wave at frequencies between 13Hz and 17Hz that emerges in the estimated ordinary coherence, $|\hat{R}_{X_3 X_5}(f)|^2$, as shown in figure (6.63:a). For the latter two spectral estimations, the elimination of linear influences of X_1 from the estimated coherence of X_3 with X_5 leads to the complete decay of the estimated coherences. That is, $|\hat{R}_{X_3 X_5 | X_1}(f)|^2$ is significantly zero. Consequently, the depth of the correlation between X_3 and X_5, is of order one with respect to these two methods. Similar results are obtained when the linear effects of X_2 are eliminated from the estimated coherence of X_3 with X_5, for the three spectral estimations.

However, the elimination of the synchronized linear influences of X_1 and X_2 from the estimated coherence, contributes to a further decrease of the estimated coherence values between X_3 and X_5. Figure (6.64:a) shows that the estimated second order partial coherence appears in two separated regions; with one in the frequency band [1, 5] Hz, which contains the δ-wave in [1, 4) Hz and a small amount of the θ-wave in [4, 6) Hz, and the other part of the estimated coherence extends between the θ-wave within the frequency band [6, 8) Hz, and the α-wave in the frequency band [8, 11) Hz. The estimated second partial coherence, $|\hat{R}_{X_3 X_5 | X_1 X_2},(f)|^2$ is significantly zero with respect to the smoothed periodogram and the multi-taper methods, as displayed in figures (6.64:b) and (6.64:c), respectively.
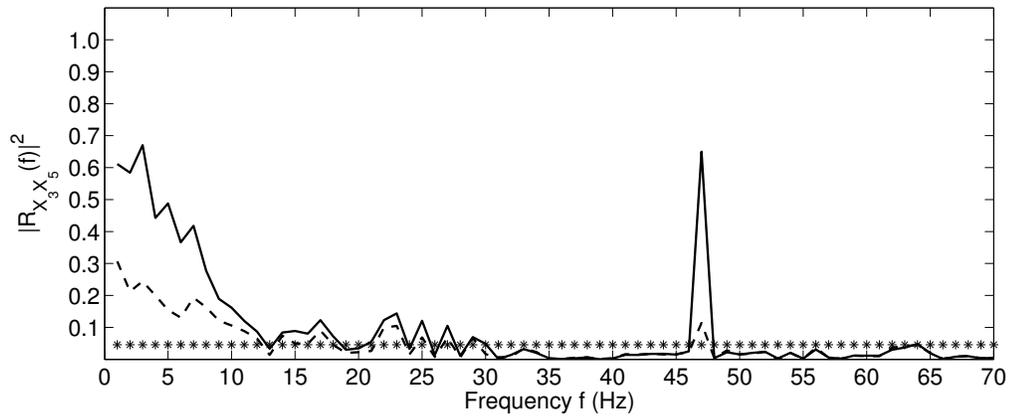
(a)
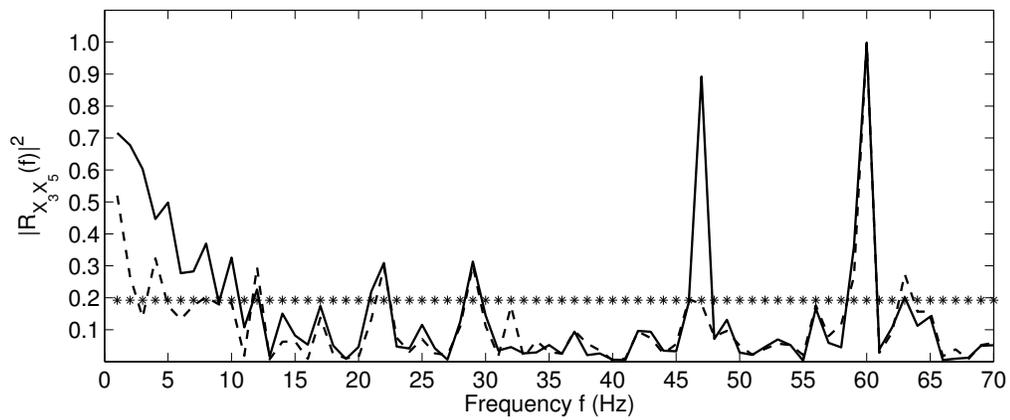
(b)
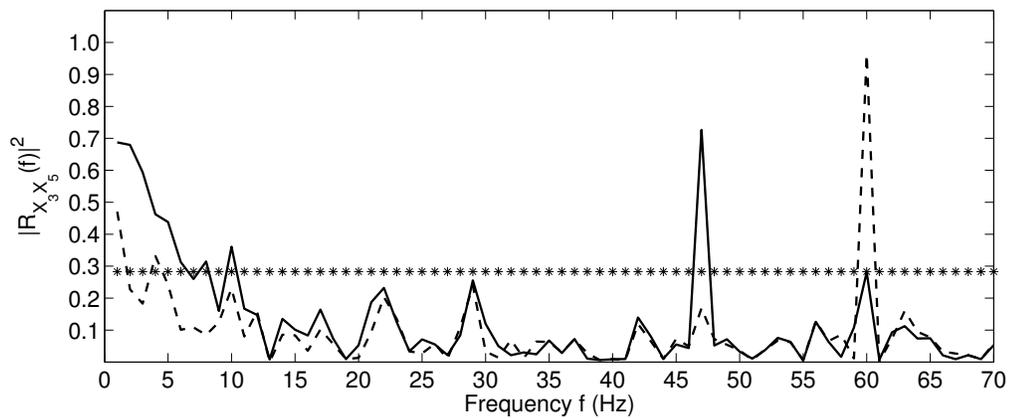
(c)

Figure 6.63: The estimated ordinary coherence, $|\hat{R}_{X_3 X_5}(f)|^2$, represented by the solid-line, and the estimated first order partial coherence, $|\hat{R}_{X_3 X_5 | X_1}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.
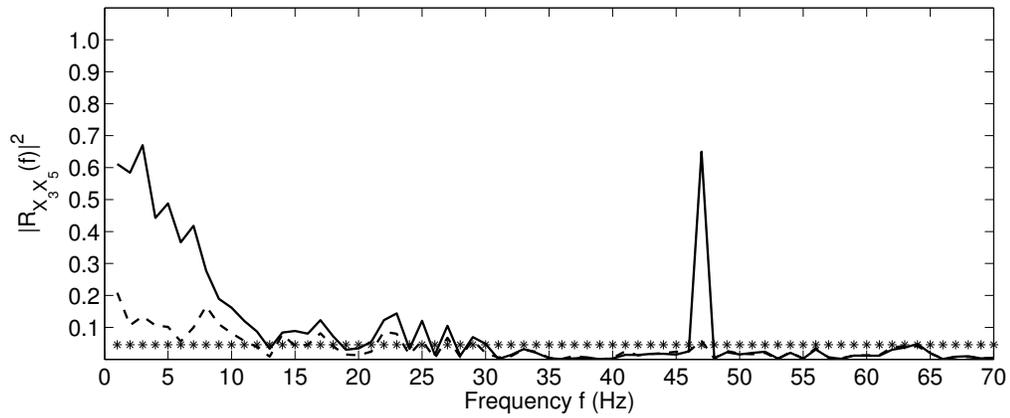
(a)
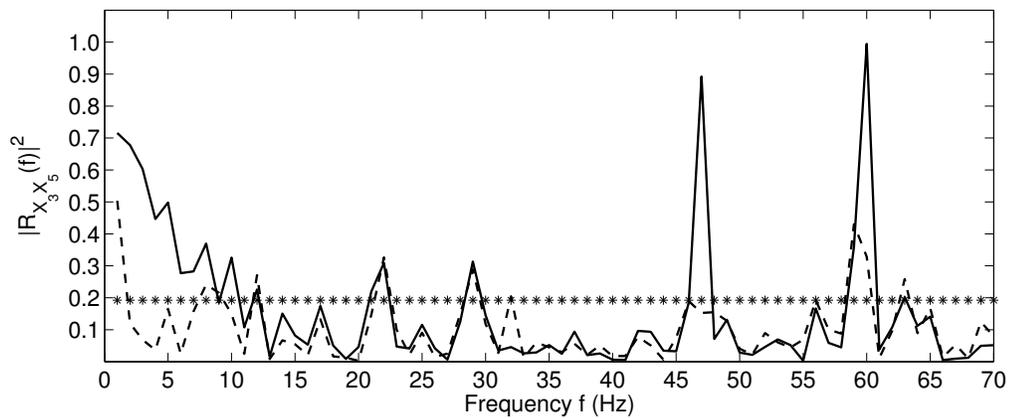
(b)
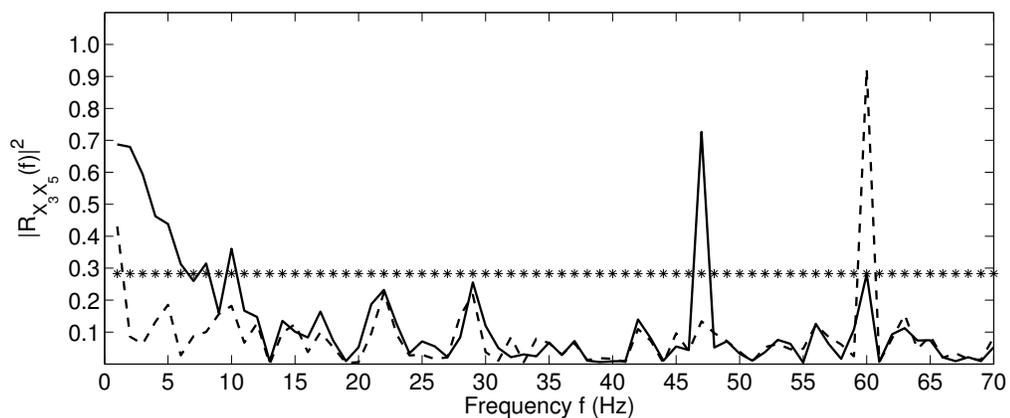
(c)

Figure 6.64: The estimated ordinary coherence, $|\hat{R}_{X_3X_5}(f)|^2$, represented by the solid-line, and the estimated second order partial coherence, $|\hat{R}_{X_3X_5|X_1X_2}(f)|^2$, represented by the dashed-line. The starred-line denotes the confidence bound of %5 level of significance at frequencies between 1 Hz and 70 Hz. Where (a) the disjoint sections method, (b) the smoothed periodogram method, (c) the multi-taper method.

6.2.4 Results summary of the non-parametric spectral estimations

We shall summarise the results that are discussed previously

- The occipital lobe of the brain, represented by the process X_4, is uncorrelated with the other brain areas, namely the frontal lobe represented by the process X_1; the central area represented by the process X_2; the parietal lobe denoted by the process X_3; and the temporal lobe represented by the process X_5.

- The temporal lobe process X_5 plays a crucial role in supporting the estimated coherences between the other processes. The elimination of linear influences of this process causes a massive decrease in values of the estimated coherence, such as $|\hat{R}_{X_1 X_2 | X_5}(f)^2|$, shown in figure (6.52:a) and $|\hat{R}_{X_1 X_3 | X_5}(f)|^2$, illustrated in figure (6.55:a). Moreover, the removal of linear influences of X_5 sometimes leads to the entire destruction of the estimated coherence, calculated by the smoothed periodogram or the multi-taper methods, such as the $|\hat{R}_{X_1 X_2 | X_5}(f)^2|$, shown in figures (6.52:b, 6.52:c), and $|\hat{R}_{X_1 X_3 | X_5}(f)|^2$, illustrated in figure (6.55:b, 6.55:c).

- The coherences $|\hat{R}_{X_1 X_2}(f)|^2$, $|\hat{R}_{X_1 X_3}(f)|^2$ and $|\hat{R}_{X_2 X_3}(f)|^2$ almost entirely vanish when the linear influences of any two other processes have been removed. On other words, the estimated second order partial coherences, associated with these estimated coherences, are significantly zero.

- The depth of correlations between processes (X_1, X_2), (X_1, X_3) and (X_2, X_3) depends on the adopted spectral estimation. For example, the depth of correlations is of order one when the multi-taper method is applied, and is of order two when the disjoint sections is used. For the smoothed periodogram method, the depth of correlations ranging between order one and order two according to the given process.

- The depth of correlations between processes (X_1, X_5) and (X_2, X_5), is of order one with respect to the multi-taper method, and is of order two when the smoothed periodogram method is used. However the depth of correlations between these processes is unknown for the disjoint section method.

- Although the depth of correlation between processes X_3 and X_5 is of order one for the smoothed periodogram and the multi- taper methods it is unknown for the disjoint sections method.

- The EEG's waves that appear frequently in this investigation are the δ-wave in the frequency band [1, 4) Hz, the θ-wave within the frequency band [4, 8) Hz, the α-wave in the frequency band [8, 13) Hz and the β-wave usually ranging between 13Hz and 30Hz.

6.2.5   Comparison of the spectral estimate methods

The spectral analyses of the current neural network $\{X_1, X_2, X_3, X_4, X_5\}$ using various spectral estimates either a parametric approach, represented by the "multivariate autoregressive model of order p ", where $p = 2$ in this case, or a non-parametric approach such as, the "disjoint sections method", the "smoothed periodogram method", and the "multi-taper method", exhibit independency of the occipital lobe process X_4, with processes representing other areas of the brain. Also, all these methods show strong correlations between the brain regions; the frontal lobe, the central area, the parietal lobe and the temporal lobe.

The multivariate autoregressive model is unable to detect rapid changes that occur in the signal, such as the effect of the notch filter which reduces the power line noise. This means that the MVAR only reflects the signal's properties which are consistent with properties of the MVAR parameters. Although, the three other methods respond to the reduction of the power line noise by the notch filter at frequency 60Hz and also to the calibration signal at frequency 47Hz, the strength of the responses are different. The multi-taper method displays high sensitivity in detecting the loss of information caused by the notch filter even for relatively small values of the power spectra. Furthermore, the smoothed periodogram and the multi-taper methods still present the notch filter effect at 60Hz in the ordinary estimated coherence curves while the disjoint sections method does not. This proves the accuracy of these two methods in estimating spectral densities, and, in turn, the coherences of the short length record. However, the orthogonality property of the tapers enables the multi-taper method to minimize the spectral leakage due to the finiteness of the data, and reduces the estimation variance. This gives the multi-taper method an advantage on the smoothed periodogram method.

The estimated coherences, computed from four spectral estimate methods, peak in the δ-wave range [1, 4) Hz, which corresponds to slow brain activities. In addition, the depth of coherence ranges between order one for the multi-taper method and order two for the smoothed periodogram and the disjoint section methods, but is unknown for the

multivariate autoregressive model since the correlations between the occipital lobe process and the other lobe representatives, X_1, X_2 and X_3 remain significant and do not completely vanish.

6.2.6 Application of the statistical test for bands of significant coherence in finite samples

In Chapter 5, we developed a statistical test for the significant coherences between finite samples of uncorrelated processes in particular band. Since the hypothesis of this test is that, the two processes are not correlated, we apply this test to the coherence between processes X_1 and X_4, which are proven to be uncorrelated at level of significance %5 in the frequency band [1, 70] Hz. The procedures of the test are

1- Calculate the coherence of the uncorrelated processes at each frequency in a particular band.

2- Choose the size of the band to be tested, where the size should be chosen between 3 to 13 frequencies.

3- Compute the negative log likelihood of the probability distribution values of observing such coherences, using equation (5.4).

4- compare the calculated values of the negative log-likelihood with the critical values listed in table (5.3), where $\alpha = 5\%$.

5- if the value of the negative log-likelihood of significant coherences in the chosen band of frequency is larger than the critical value then, the hypothesis is rejected in that band. If this value is less than the critical value, then the hypothesis is accepted.

According to the previous analysis of spectral densities and coherences, the frequency band of interest is [1, 30) Hz. In particular we choose the bands of size 3, 4, 5, where the first band consists of the frequency [1, 2, 3] Hz, the second band contains the frequencies [4, 5, 6, 7] Hz and the third band is [8, 9, 10, 11, 12] Hz. Since the length of the used data is 13 seconds, the corresponding critical value for the confidence bound of $\alpha = \%5$ level of significance is approximately $CV \approx 1.9$. The calculated values of the negative log-likelihood for the first band is 13.290721, for the second band is 18.821432, and , for the third band is 23.55084. Clearly these values are larger than the critical value CV, which lead to reject the hypothesis that the two processes are not correlated in

these particular bands at the level of significance of %5. Since the number of the available observations is small, the consecutive measures of coherences of finite samples are expected to be correlated. On other words, if the number of the underlying data is large enough, the hypothesis will be accepted.

Chapter 7

Conclusion and future work

7.1 Conclusion

The investigation of the interrelationships between interacting processes among dynamic systems has received a lot of interest from researchers. Furthermore, there has been an increase in the number of studies concerned with the direction of information flow within large networks in disciplines, such as Economics, Neuroscience, Engineering and Neurobiology. Several mathematical methods have been proposed in order to facilitate these studies in time and frequency domains.

Specifically, the current work has studied the electrical activity of the brain, which is produced by the interaction of large numbers of neurons. This electrical activity of the brain is recorded in different forms such as EEG , fMRI and MRI. The EEG's presentation is used as the application for this research, particularly the EEG signals which reflect the electrical activity of the brain under epileptic conditions. The time-frequency analyses are carried out for the EEG signals representing the left hemisphere of the brain. Since the EEG signals collected from the same geographical area are approximately functionally equivalent, only five signals are chosen to represent the five areas of the brain.

To conclude, we have described the mechanism of collecting the EEG records from the scalp and have described the electrode placements and nomenclature according to the 10-20 international system and 10% system in Chapter 2. Also, this chapter gave a succinct anatomical description of the brain lobes and their functions. As the data used in this work belongs to an epileptic subject, we have explained this brain disorder and provided information about its common types and their symptoms. The properties of the EEG signals, such as amplitude and frequency range, have been summarised in this chapter.

This chapter also presented the definition of a stochastic process, auto-correlation, cross-correlation, and mixing condition.

Chapter 3 dealt with the Fourier transform and its properties. The description of the Fast Fourier Transform algorithm was provided in this chapter. The spectral estimation methods such as averaging periodograms across contiguous sections of the single records, the averaging frequency method and the multi-taper method with their statistical properties have been described broadly in this chapter.

Chapter 4 involved definitions of coherence and partial coherence. The calculation of these measurements directly from the data using the concept of the derived spectra was explained. The asymptotic coherence and partial coherence were displayed in this chapter and the construction of confidence intervals of significant coherences and partial coherences was explained. The multivariate autoregressive model, and the methods that were used to estimate its parameters the Yule walker and likelihood estimations were presented. The concept of Granger causality was discussed in this chapter with respect to bivariate and multivariate autoregressive models. The concept of partial directed coherence (PDC) introduced by Baccalá and Sameshima [2001] was defined, while the significance level of the PDC proposed by Schelter *et al.* [2005] was also provided.

Chapter 5 provided a statistical test for bands of significant coherence in finite samples produced from uncorrelated stationary stochastic processes, using the likelihood principle. In this chapter, the likelihood function was used to measure the independence between two samples in a specified band of frequencies. The basis of this investigation was the analytical expression for the probability density function of coherence at any frequency, as provided by Goodman ([1965]) and Brillinger [1981] when the processes are known *a priori* to be uncorrelated. This test was implemented for extensive simulated data to determine the confidence interval bounds for significant coherences within a relatively small band of consecutive frequencies.

Chapter 6 involved the time-frequency analyses of the data provided by King Faisal Hospital and Research Center KFHRC [Jeddah; Saudi Arabia]. The analyses have been carried out for five EEG signals collected from the scalp of a focal right-handed epileptic patient. These signals were chosen to represent the main lobes of the brain in the left hemisphere, as explained broadly in Chapter 2. Coherence and partial coherence were used to measure the strength of the correlations of these processes, while the partial directed coherence introduced by Baccalá and Sameshima [2001] was applied to determine

the direct and indirect causal influence among these processes. The significance level of the PDC values proposed by Schelter *et al.* [2005], given in equation (4.100), was also employed.

The application of the partial directed coherence for the processes $\{X_1, X_2, X_3, X_5\}$ produced the causal influence structure displayed in figure (6.6). Clearly, there were reciprocal direct causal influences among the processes, representing the frontal lobe, the central area, the parietal lobe and the temporal lobe. That is, information flows between the processes in both directions at different levels. However, two uni-directed causal influences were captured between the processes X_2 and X_4, and the processes X_4 and X_5 where X_4 represents the occipital lobe of the brain. Indirect causal influences of the occipital lobe process, X_4, on the rest of the processes throughout the process X_5 were detected. Generally speaking, the values of the partial directed coherences in the anterior-posterior direction were larger and greater than the values of the PDC in the posterior-anterior direction. Similarly, the values of the partial directed coherences of all processes pairwise with the process X_5 were clearly larger than the values of the PDC in the opposite directions. The causal influences of the process X_5 on the processes X_1, X_2 and X_3, which were displayed in figures (6.13, 6.15, 6.17) respectively, are almost all spread out over the entire frequency range and gently decreasing towards higher frequencies.

The spectral analyses of the processes X_1, X_2, X_3 and X_5, using the parametric and non-parametric approaches, detailed in Chapter 3, revealed that the estimated power spectra of these processes peaked approximately at 2Hz in the δ-wave band, while the process X_4 had two peaks, one corresponding to the frequency 3Hz in the δ-wave band, and the second one at $f \approx 29$Hz, which is the last frequency in the β band. The power spectra of these processes, produced by applying the smoothed periodogram procedure in figures (6.38, 6.39, 6.40) and the multi-taper in figures (6.41, 6.42, 6.43), showed the efficiency of these two estimators to detect the noise caused by the power line even for relatively small values of the power spectra.

Significant coherences have been captured between the processes, representing the main regions of the brain in the left hemisphere, except the occipital process X_4, which exhibited independence from the other processes. Also, the significant estimated coherences were mostly detected in the δ and θ bands, and peaked in the δ-band. The partial coherence analysis, which specifies the depth of the correlations between these processes, gave slightly different results according to the spectral approach used in the analyses. Generally

speaking, significant estimated coherence still remains after removing linear influences of two processes in the analyses, produced by MVAR as in figures (6.32, 6.33, 6.34). For the method of averaging periodograms across non-overlapped segments of the single records, the estimated coherences are significantly non-zero between processes (X_1, X_5), (X_2, X_5) and (X_3, X_5) despite the removal of the simultaneous linear influence of two processes as in figures (6.49, 6.50, 6.52). However, the estimated second order partial coherences $|\hat{R}_{X_1 X_2 | X_3 X_5}(f)|^2$, $|\hat{R}_{X_1 X_3 | X_2 X_5}(f)|^2$ and $|\hat{R}_{X_2 X_3 | X_1 X_5}(f)|^2$ plotted in Figures (6.46, 6.47, 6.47) are significantly zero.

In the smoothed periodogram of the entire record, the elimination of the simultaneous linear influences of two processes decreases the estimated coherences. For example, the destruction of the estimated coherence of process X_1 with process X_5 occurred as a result of removing the concurrent linear effects of processes X_2 and X_3, see figure (6.60:b). However, in the multi-taper method, the pairwise significant estimated coherences are statistically significantly zero when linear influences of only one process has been removed as in figures (6.51:c, 6.52:c, 6.54:c).

To conclude, in the light of the results of spectral densities and coherences analyses, performed in Chapter 6 by using spectral estimations such as; the multivariate autoregressive method of order 2; averaging periodograms across adjacent sections of single records; smoothed periodogram of the entire record; and, the multi-taper method, we found that the multi-taper is a powerful tool estimation, that is not only able to estimate the spectral densities with reduced variance for the short duration data , but is also able to minimize the spectral leakage due to the finiteness of the data.

7.2   Future work

The analysis of the EEG signals of an epileptic subject has attracted interest for either clinical or research purposes. The research into EEG analysis will continue as it represents a broad area for researchers to know more about the underlying brain activities under normal and abnormal conditions. In the future I will continue to investigate the following points:

- According to the present results, as the power spectra and coherences peaked in the δ- band, the investigation of the behaviour of the δ-wave under normal and epilepsy conditions for waking adults is challenging, since the existence of the δ- wave in

adults during deep sleep stages is normal.

- Investigate the difference between the activities of the normal θ-wave and the epileptiform θ- wave in childhood. Since the θ wave manifests itself normally in children and in adults during the sleep or drowsiness, the presence of the high θ- wave in an awake adult indicates abnormal conditions.

- Investigate the effects of long-term anti-epilepsy medication on the vision levels in children with epilepsy.

References

[2003] Adeli, H., Zhou, Z., Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. J Neurosci Methods, Vol.(123), pp: 69-87.

[1994] American Electroencephalographic Society. (1994). Guideline thirteen: guidelines for standard electrode position nomenclature. Amer. Electroencephalogr. Soc. J Clin Neurophysiol. Vol.(11), pp: 111-113.

[1997] Amjad, A. M., Halliday, D. M. , Rosenberg , J. R., Conway, B. A. (1997). An extended difference of coherence test for comparing and combining several independent coherence estimates: Theory and application to the study of motor units and physiological tremor. J Neurosci Methods. Vol.(73), pp: 69-79.

[1998] Anderson, C. W., Stolz, E. A., Shamsunder, S. (1998). Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during Mental Tasks. IEEE Trans Biomed Eng., Vol.(45), No.(3), pp: 277-286.

[1995] Attivissimo, F., Savino, M., Trotta, A. (1995). Flat-top smoothing of periodograms and frequency averagings to increase the accuracy in estimating power spectral density. Measurements, Vol.(15), pp: 15-24.

[2007] Aurlien, H., Aarseth, J. H., Gjerde, I.O., Karlsen, B., Skeidsvoll, H., Gilhus, N., E. (2007). Focal epileptiform activity described by a large computerised EEG database. Clin Neurophysiol, Vol.(118), pp: 1369-1376.

[2001] Baccalá, L. A., Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. Biol. Cybern. Vol.(84), pp: 463-474.

[1998] Baccalá, L. A., Sameshima, K. (1998). Directed coherence: a tool for exploring functional interactions among brain structure. In: Nicolelis MAL (ed) Methods for neural ensemble recordings. CRC Boca Raton. Fla., pp: 179-192.

[1998] Baccalá, L. A., Sameshima, K., Ballester, G., Do Valle, A. C., Timo-Iaria, C. (1998). Studying the interaction between brain structures via directed coherence and granger causality. Applied Sig. Process., Vol.(5), pp: 40-48.

[2010] Barbé, K., Pintelon, R., Schoukens, J. (2010). Welch method revisited: Nonparametric power spectrum estimation via circular overlap. IEEE Trans signal proc., Vol.(58), No.(2), pp: 553-565.

[1948] Bartlett, M. S. (1948). Smoothing periodograms from time series with contagious spectra. Nature. Vol. (161), pp: 686-687.

[1986] Bendat, J. S., Piersol, A. G. (1986). Random Data: Analysis and measurement procedures. John Wiley and Sons, Inc.

[1988] Berger, J.O., Wolpert, R.L. (1988). The Likelihood Principle (Second Eddition). Haywood, CA: The Institute of Mathematical Statistics.

[1967] Bergland, G. D. (1967). The fast Fourier transform recursive equations for arbitrary length records. Math Comp., Vol.(21), pp: 236-238.

[1968] Bergland, G. D. (1968). A fast Fourier transform for real valued-series. Comm ACM., Vol.(11), No.(10), pp: 703-710.

[1999] Bernasconi, C., König, P., K. (1999). On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. Biol Cybern., Vol.(81), pp: 199-210.

[1999] Binnie C. D., Stefan H. (1999). Modern electroencephalography: its role in epilepsy management. Clin Neurophysiol., Vol.(110), pp:1671-1697.

[1989] Birkel, T. (1989). A note on the strong law of large numbers of positively dependent random variables. Statist Probab Lett., Vol.(7), pp: 17-20.

[1962] Birnbaum, A. (1962). On the foundations of statistical inference. J. Amer. Statist. Assoc., Vol.(57), No.(298), pp: 269-326.

[1958] Blackman, R.B., Tukey, J.,W. (1958). The measurement of power spectra from the point of view of communication engineering. Bell Syst Tech J. Vol.(37), pp: 183-282, 485-569.

[2000] Bloomfield, P. (2000). Fourier Analysis of Time Series: An Introduction. John Wiley and Sons, Inc.

[2004] Boyd, J. P. (2004). Prolate spheroidal wave functions as an alternative to Chebyshev and Legendre polynomials for spectral element and pseudospectral algorithms. J Comp Phys., Vol.(199), pp: 688-716.

[1999] Bressler, S. L., Ding, M., Yang, W. (1999). Investigation of cooperative cortical dynamics by multivariate autoregressive modeling of event-related local field potentials. Neuro Computing, Vol.(26-27), pp: 625-631.

[1981] Brillinger, D.R. (1981). Time Series: data analysis and theory. Holden-Day, Inc. Ross, S. M. (1983). Stochastic processes. John Wiley and Sons, Inc.

[1991] Brockwell, P. J., Davis, R. A. (1991). Time Series Theory and Methods. Second Eddition. Springer-Verlag New York,Inc.

[2004] Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R. Bressler, S.L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. PNAS, Vol.(101), No.(26), pp: 9849-9854.

[1975] Caines, P., E., Chan, C., W.(1975). Feedback between stationary stochastic processes. IEEE Trans Autom Control. Vol. (AC-20), No. (4), pp. 498-508.

[1996] Chatfield, C. (1996). The Analysis of Time series: An Introduction. Chapman and Hall.

[1985] Chatrian, G. E., Lettich, E., Nelson, P. L. (1985). Ten percent electrode system for topo- graphic studies of spontaneous and evoked EEG activity. Am J EEG Technol., Vol. 25), pp: 83-92.

[1985] Chatrian, G. E., Lettich, E., Nelson, P. L. (1988). Modified nomenclature for the 10% electrode system. J Clin Neurophysiol., Vol.(5), pp: 183-186.

[1967] Cooley, J. W. , Lewis, P. A. W., Welch, P. D. (1967). Historical notes on the fast Fourier transform. IEEE Trans. Audio Electroacoust., Vol.(AU-15), pp: 76-79.

[1965] Cooley, J. W., Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. Math Comput., Vol.(19), pp: 297-301.

[2008] Cryer, J. D., Chan, K. (2008). Time series analysis with applications in R. Springer Science+Businness Media, LLC. Second Eddition.

[1999] Denison, D. G. T., Walden, A. T., Balogh, A. Forsyth, R. J. (1999). Multitaper testing of spectral lines and the detection of the solar rotation frequency and its harmonics. Appl Statist., Vol.(48), pp: 427-439.

[2008] Dhamala, M. , Rangarajan, G. , Ding, M. (2008). Estimating Granger Causality from Fourier and Wavelet Transforms of Time Series Data. Phys Rev Lett., Vol.(100), No.(1), 018701, pp:1-4.

[1990] Diggle, P. J. (1990). Time Series: A Biostatistical Introduction. Oxford University Press, New York.

[2000] Ding, M., Bressler, S. L., Yang, W., Liang, H. (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. Biol Cybern, Vol.(83), pp: 35-45.

[1999] Doukhan, P., Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. Stoch Proc Appl., Vol.(84), pp: 313-342.

[1979] Edwards, R. E. (1979). Fourier Series: A modern introduction, Springer-Verlag. Vol. (1-2).

[1997] Engle S. A., Glover G. H., Wandell B. A. (1997). Retinotopic Organization in human visual cortex and the Spatial precision of functional MRI. Cerebral Cortex,Vol.(7), pp:181-192.

[1965] Enochson L. D, Goodman N. R. (1965). Gaussian approximation for the distribution of sample coherence. Air Force Flight Dynamics Laboratory. Tech. Rep. TR 6557, Wright- Patterson AFB, Ohio. ERP recording. J Psychophysiol., Vol.(5), pp: 273-276.

[2006] Ferris C. F., Febo M., Luo F., Schmidt K., Brevard M., Harder J. A, Kulkarni P., Messenger T., King J. A. (2006). Functional magnetic resonance imaging in Conscious Animals: A New tool in behavioural neuroscience Research. J Neuroendocrinol., Vol.(18), pp: 307-318.

[2005] Fisher, R. S., Boas, W. V. E., Blume, W., Elger, C., Genton, P., Lee, P., Engel Jr, J. (2005). Epilepsy seizures and Epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Blackwell publishing, Inc. ILAE. Epilepsia, Vol.(46), No.(4), pp: 470-472.

[1995] Florian, G., Pfurtscheller, G. (1995). Dynamic spectral analysis of event-related EEG data. Electroencephalogr Clin. Neurophysiol., Vol.(95), pp: 393-396.

[2004] Funk, A. P., Epstein, C. M. (2004). Natural rhythm: evidence for occult 40 Hz gamma oscillation in resting motor cortex. Neurosci Lett., Vol.(371), pp: 181-184.

[1966] Gentleman W. M., Sande, G. (1966). Fast Fourier transforms - for fun and profit. Fall Joint Computer Conf. AFIPS, Washington D.C., Spartan. Proc., Vol.(29), pp: 563-578.

[1982] Geweke, J. (1982). Measurements of linear dependence and feedback between multiple time series. J Am Stat Assoc, Vol.(77), No.(378), pp. 304-324.

[2002] Goldman R. I, John C. A., Stern M., Engel Jr J., Cohen M. S. (2002). Simultaneous EEG and fMRI of the alpha rhythm. LWW J., Vol.(13), No.(18), pp: 2487-2492.

[1958] Good, I. J. (1958). The interaction algorithm and practical Fourier analysis. J Roy Statist Soc Ser. B, Vol.(20), No.(2), pp: 361-372.

[1971] Good, I. J., (1971). The relationship between two fast Fourier transform. IEEE Trans Comput., Vol.(C-20), No.(3), pp: 310-317.

[1965] Goodman, N. R. (1965). Measurement of matrix frequency response functions and multiple coherence functions. Air Force Flight Dynamics Laboratory. Tech. Rep. Wright Patterson AFB, Ohio. AFFDL-TR 65-56.

[2010] Gosse, L. (2010). Practical band-limited extrapolation relying on Slepian series and compressive sampling. Comp Math Appl., Vol.(60), pp: 1259-1279.

[1969] Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica. Vol.(37), No.(3), pp: 424-438.

[1995] Halliday, D. M., Rosenberg, J. R., Amjad, A. M., Breeze, P., Conway, B. A., Farmer, S. F. (1995). A framework for the analysis of mixed time series/point process data-

theory and application to the study of physiological tremor, single unit discharges and electromyogram. Prog Biophys Mol Biol., Vol.(64), pp: 237-278.

[2003] Harrison, L., Penny, W. D., Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. Neuroimage, Vol.(19), pp: 1477-1491.

[2009] Hema, C. R. , Paulraj, M. P., Yaacob, S. , Adom, A. H., Nagarajan, R. (2009). An Analysis of the effect of EEG frequency bands on the classification of motor imagery signals. IJBSCHS., Vol.(16), No.(1), pp: 121-126.

[1958] Jasper, H. H. (1958). The 10-20 electrode system of the International Federation. Electroencephalogr Clin Neurophysiol., Vol.(10), pp: 370-375.

[1994] Jokeit, H., Makeig, S. (1994). Different event related patterns of gamma band power in brain waves of fast and slow-reacting subjects, Proc Natl Acad Sci., USA, Vol.(91), No.(14), pp: 6339-6343.

[2007] Jurcak, V., Okamoto, M., Singh, A. , Dan, I. (2007). 10/20, 10/10, 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. Neuroimage. Vol.(34), pp:1600-1611.

[1985] Kalbfleisch, J.G. (1985). Probability and Statistical Inference, Vol1: Probability, Vol2: Statistical Inference. Spring-Verlag New York, Inc.

[1991] Kaminski, M. J., Blinowska, K., J. (1991). A new method of the description of the information flow in the brain structures. Biol Cybern., Vol.(65), No.(3), pp: 203-210.

[1984] Kamitake. T., Harashima, H., Miyakawa, H. and Saito, Y. (1984). A time-series analysis method based on the directed trans information. Electron. Comm Jpn1., Vol.(67), No.(6), pp: 1-9.

[2001] Kamniski, M. , Ding, M., Truccolo, W. A. , Bressler, S. L. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biol Cybern., Vol.(85), pp: 145-157.

[1995] Kandel, E. R., Schwartz, J. H., Jessel,T. M., (1995). Essential of neural science and behavior. Appleton & Lange. A Simon & Schuster Company.

[1993] Kay, S. M. (1993). Fundamentals of Statistical Signal Processing: Estimation Theory. Volume I. Prentice-Hall.

[1999] Klem, G. H. , Luders, H. O., Jasper, H. H. , Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol., Suppl., Vol.(52), pp: 3-6.

[2003] Laufs, H., Kleinschmid, A., Beyerle, A., Eger, E., Salek-Haddadi, A., Preibisch, C., Krakowa, K. (2003). EEG-correlated fMRI of human alpha activity. Neuroimage. Vol.(19), pp: 1463-1476.

[2004] Logothetis, N. K., Wandell, B. A. (2004). Interpreting the BOLD signal. Annu Rev Physiol., Vol.(66), pp: 735-769

[1965] Loynes, R. M. (1965). Extreme values in uniformly mixing stationary stochastic processes. Ann Math Statist., Vol.(36), pp: 993-999.

[2009] Medkour, T., Walden, A. T., Burgess, A. (2009). Graphical modelling for brain connectivity via partial coherence. J Neurosci Methods, Vol.(180), pp: 374-383.

[1991] Miller, G. A., Lutzenberger, W., Elbert T. (1991). The linked-reference issue in EEG and ERP recording. J Psychophysiol., Vol.(5), pp: 273-276.

[2001] Mima, T., Matsuoka, T., Hallett, M. (2001). Information flow from the sensorimotor cortex to muscle in humans. Clin. Neurophysiol. Vol.(112), No.(1), pp: 122-126.

[2000] Mima, T., Stegera, J., Schulmanb, A. E., Gerloffa, C., Hallett, M. (2000). Electroencephalographic measurement of motor cortex control of muscle. Clin Neurophysiol., Vol.(111), No.(2), pp: 326-337.

[2004] Miranian, L. (2004). Slepian functions on the sphere, generalized Gaussian quadrature rule. Inst Phys. Pub., Vol.(20), pp: 877-892.

[2001] Moller, E., Schack, B., Arnold, M., Witte, H. (2001). Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models. J Neurosci Methods, Vol.(105), pp: 143-158.

[1984] Newman, C. M. (1984). Asymptotic independent and limit theorems for positively and negatively dependent random variables. In: Tong, Y. L. (ED.), Inequalities in statistics and probability, IMS Lect Notes-Monograph Ser. Vol.(5), pp: 127-140.

[2003] Niedermeyer, E. (2003). Electrophysiology of the frontal lobe. Clin Electroencephalogr. Vol.(34), pp: 5-12.

[2005] Niedermeyer, E., Lopes Da Silva, F. H. (2005).Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincott Williams & Wilkins.

[1991] Nunez, P.L., Pilgreen, K.L., (1991). The spline-Laplacian in clinical neurophysiology: a method to improve EEG spatial resolution. J Clin Neurophysiol., Vol.(8), No.(4), pp: 397-413.

[1998] Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J. M., Hinrichs, H., Ikeda, A., Luccas, F.J., Rappelsburger, P. (1998). IFCN standards for digital recording of clinical EEG. The International Federation of clinical neurophysiology. Electroencephalogr clin Neurophysiol Suppl., Vol.(106), No.(3), pp: 259-26.

[1987] Nuwer, M. R. (1987). Recording electrode site nomenclature. J. Clin. Neurophysiol. Vol.(4), No.(2), pp: 121-134.

[1990] Ogawa, S., Lee T. M., Kay, A. R., Tank, D.W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. USA, Vol. (87), pp: 9868-9872.

[1992] Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S., Merkle, H., Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. Proc. Nati. Acad. Sci. USA., Vol.(89), pp: 5951-5955.

[2001] Oostenveld, R., Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. Clin Neurophysiol. Vol.(112), No.(4), pp: 713-719.

[1987] Park, J., Lindberg, G. R., Thomson, D. J. (1987). Multiple-taper spectral analysis of terrestrial free oscillations. part I, Geophys J Royal Astr Soc., Vol.(91), pp: 755-794.

[1991] Peleg, S., Porat, B. (1991). The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase. IEEE Trans. Signal Proc., Vol.(39), No.(3), pp: 749-752

[1993] Percival, D. B., Walden, A. T. (1993). Spectral Analysis For Physical Applications: Multitaper and conventional univariate techniques. Cambridge University press.

[1999] Pfurtscheller, C., Lopes da Silva, F.H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol., Vol.(110), pp: 1842-1857.

[2001] Rice, F., Cowley, B., Moran, B., Rice, M. (2001). CramrRao Lower Bounds for QAM Phase and Frequency Estimation. IEEE Trans. Commun., Vol.(49), No.(9), pp: 1582- 1591.

[1989] Rosenberg, J. R., Amjad, A. M., Breeze, P., Brillinger, D. R. and Halliday, D. M. (1989). The Fourier approach to the identification of functional coupling between neural spike trains, Prog. Biophys Mol Biol., Vol.(53), pp: 1-31.

[1956] Rosenblatt, M. (1956). A central limit theorem and strong mixing condition. Proc Nat Acad Sci., USA, Vol.(42), pp: 43-47.

[1983] Ross, S. M. (1983). Stochastic processes. John Wiley and Sons, Inc.

[1981] Saito, Y., Harashima, H. (1981). Tracking of information within multichannel EEG record. In: Yamaguchi N., Fujisawa K. (Eds), Recent advances in EEG and EMG data processing. Elsevier, Amsterdam. pp. 133-146.

[2009] Sato, J. R., Takahashi, D. Y., Arcuri, S. M., Sameshima, K., Morettin, P. A., Baccala, L. A. (2009). Frequency domain connectivity identification: An application of partial directed coherence in fMRI. Hum Brain Mapp., Vol.(30), pp: 452-461.

[2005] Schelter, B. Winterhalder, M., Eichler, M. Peifer, M., Hellwig, B., Guschlbauer, B., Lucking, C. H., Dahlhaus, R., Timmer, J. (2005). Testing for directed influences among neural signals using partial directed coherence. J Neurosci Methods. Vol.(152), pp: 210-219.

[1956] Sem-Jacobsen, C. W. , Kuchera, W. E. (1956). An extra-short time constant for study of beta and gamma waves in depth recordings. Electroencephalogr Clin Neurophysiol Suppl., Vol.(8), No.(1), pp: 144-145.

[1949] Shannon, C. E. (1949). Communication in the presence of noise, Proc Instit Radio Engineers, Vol.(37), No.(1), pp: 10-21.

[2009] Shorvon, S., Perucca, E., Engel Jr, J. (2009). The treatment of epilepsy. Wiley-Blackwell. SIAM Rev., Vol.(25), No.(3), pp: 379-393.

[1978] Slepian, D., (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V-The discrete case. AT & T Tech J., Vol.(57), pp:1371-1430.

[1983] Slepian, D. (1983). Some comments on Fourier analysis , uncertainity and modeling. SIAM Rev., Vol.(25), No.(3), pp: 379-393.

[2005] Smith, S. J. M., (2005). EEG in the diagnosis, classification, and management of patients with epilepsy. J Neurol Neurosurg Psychiatry, Vol.(76), pp: ii2-ii7.

[1999] Srinivasan, R. (1999). Methods to Improve the Spatial Resolution of EEG. IJBEM. Vol. (1), No. (1), pp: 102-111.

[2000] Suarez, E., Viegas, M. D., Adjouadi, M., Barreto, A. (2000). Relating induced changes in EEG signals to orientation of visual stimuli using the ESI-256 machine. Biomed Sci Instrum., Vol(36), pp: 33-38.

[1998] Takano, T., Ogawa, T.(1998). Characterization of developmental changes in EEG-Gamma band activity during chilhdood using the autoregressive model. Acta Paediatr Jpn. Vol.(40), No.(5), pp: 446-452.

[1996] Takigawa, M., Wang, G., Kawasaki, H., Fukuzako, H. (1996). EEG analysis of epilepsy by directed coherence method: A data processing approach. Int J Psychophsiol., Vol.(21), pp: 65-73.

[1982] Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. Vol.(70). No.(9), pp: 1055-1095.

[2007] Thomson, D. J. (2007). Jackknifing multitaper spectrum estimates. IEEE Signal Proc Let., Vol.(24), No.(4), pp: 20-30.

[1990] Thomson, D.J. (1990). Quadratic-inverse spectrum estimates: Applications to palaeoclimatology. Philos T Roy Soc A., Vol. (332), No.(1627), pp: 539-597.

[1990] Tran, L. (1990). Recursive kernel density estimators under a weak dependence condition. Ann Inst Stat Math., Vol.(42), No.(2), pp: 305-329.

[2004] Tsang, E. W., Koren, S. A., Persinger, M. A. (2004). Power increases within the gamma range over the frontal and occipital regions during acute exposures to cerebrally counterclockwise rotating magnetic fields with specific derivatives of change. Int J Neurosci., Vol.(114), No.(9), pp: 1183-1193.

[1993] Vernon, L. T., José, B., Diane, S., Kim, T., Robert, G., David, N. L., Raif, C., Samuel, A. F., Jean-Paul, S. (1993). The spatial location of EEG electrodes: locating the best-fitting sphere relative to cortical anatomy. Electroencephalogr clin Neurophysiol., Vol.(86), No.(1), pp: 1-6.

[2005] Walter, G., Soleski, T. (2005). A new friendly method of computing prolate spheroidal wave functions and wavelets. Appl Comput Harmon Anal., Vol.(19), No.(3), pp: 432-443.

[1992] Wang, G., Takigawa, M., (1992). Directed coherence as a measure of interhemispheric correlation of EEG. Int J Psychophysiol., Vol.(13), No.(2), pp: 119-128.

[2007] Wang, X., Chen, Y., Bresseler, S. L., Ding, M. (2007). Granger Causality between multiple interdependent neurobiological time series: Blockwise versus pairwise methods. Int J Neural Syst., Vol.(17), No.(2), pp: 71-78.

[1967] Welch, P. D. (1967). The use of the fast Fourier transform for estimation of spectra: A method based on time averaging short, modified periodograms. IEEE Trans Electro Acoust., Vol.(AU 15), No.(2), pp: 70-73.

[1956] Wiener, N. (1956). The theory of prediction. In: Beckenbach EF, editors. Modern Math- ematics for Engineers. New York: McGraw-Hill; 1956(Chapter 8).

[1981] Withers, C. S. (1981). Central limit Theorems for dependent variables. Z Wahrsch Verw Gebiete., Vol.(57), pp: 509-534.

[2001] Xiao, H., V Rokhlin, V., Yarvin, N. (2001). Prolate spheroidal wavefunctions, quadrature and interpolation. Inst Phys Pub., Vol.(17), pp: 805-838.

Appendix A

In this appendix we provide some definitions for the mathematical terminologies mentioned and used in Chapter 3.

A.1   Chi-squared distribution

Let $\{X_1, X_2, \cdots, X_n\}$ be independent standard normal random variables, then the sum of the squares of these variables, say Z,

$$Z = \sum_{i=1}^{n} X_i^2\,,$$

is distributed as Chi-squared deviate with n degree of freedom and denoted by

$$Z \sim \chi_n^2\,.$$

With probability density function

$$f(x) = \begin{cases} \dfrac{1}{2^{n/2}\,\Gamma(n/2)}x^{n/2-1}\,e^{-x/2} & ; \quad x \geq 0 \\[2ex] 0 & ; \quad x < 0\,, \end{cases}$$

where the mean and variance values are n and $2n$ respectively.

A.2   QR algorithm

QR algorithm is a commonly used procedure to compute the eigenvalues and eigenvectors of a matrix. This procedure is based on decomposing a matrix, say A, into a product of an orthogonal matrix, Q, and an upper triangular matrix, R, such that

$$A = QR\,,$$

where $\overline{Q}Q = Q\overline{Q} = I$, and I is the identity matrix. The 'overline' represents the matrix transpose and satisfies the property $\overline{Q} = Q^{-1}$.

A.3 Derivation of equations (3.105) and (3.106)

Here we will explain how we obtained equations (3.105) and (3.106) delivered in Chapter 3 within the section on the calculation of the Slepian functions $\psi(x)$ using Legendre polynomials. For equation (3.105) we start with the equations

$$c^2 \int_{-1}^{1} x^2 \psi(x) \, P_0(x) \, dx = \gamma \int_{-1}^{1} \psi(x) P_0(x) \, dx \quad ; \quad \text{for } k = 0,$$

$$2 \int_{-1}^{1} \psi(x) \, P_1(x) \, dx + c^2 \int_{-1}^{1} x^2 \psi(x) \, P_1(x) \, dx = \gamma \int_{-1}^{1} \psi(x) P_1(x) \, dx \quad ; \quad \text{for } k = 1.$$
$$\text{(A.1)}$$

Substituting the Legendre polynomials,

$$1 = P_0, \quad x = P_1, \quad x^2 = \frac{2P_2(x) + P_0(x)}{3}, \quad x^3 = \frac{2P_3(x) + 3P_1(x)}{5},$$

into equations (A.1) yields

$$c^2 \int_{-1}^{1} \left[\frac{2P_2(x) + P_0(x)}{3} \right] \psi(x) \, dx = \gamma \int_{-1}^{1} \psi(x) P_0(x) \, dx \quad ; \quad \text{for } k = 0,$$

$$2 \int_{-1}^{1} \psi(x) \, P_1(x) \, dx + c^2 \int_{-1}^{1} \left[\frac{2P_3(x) + 3P_1(x)}{5} \right] \psi \, dx = \gamma \int_{-1}^{1} \psi(x) P_1(x) \, dx \quad ; \quad \text{for } k = 1.$$
$$\text{(A.2)}$$

Rearranging equations (A.2) and using the definition, $\psi_k = \int_{-1}^{1} \psi(x) P_k(x) \, dx$, give

$$\frac{2c^2}{3}\psi_2 + \frac{c^2}{3}\psi_0 = \gamma\psi(x)_0(x) \quad ; \quad \text{for } k = 0,$$

$$(2 + \frac{3c^2}{5})\psi_1 + \frac{2c^2}{5}\psi_3 = \gamma \int_{-1}^{1} \psi(x)_1(x) \quad ; \quad \text{for } k = 1.$$
$$\text{(A.3)}$$

Now for $k \geq 2$ we use the identity

$$\int_{-1}^{1} x\psi(x) P_k(x) \, dx = \frac{k+1}{2k+1} \int_{-1}^{1} \psi(x) P_{k+1}(x) \, dx + \frac{k}{2k+1} \int_{-1}^{1} \psi(x) P_{k-1}(x) \, dx$$

which is a direct consequence of the recurrence relation (3.100). A second application of this identity gives

$$\int_{-1}^{1} x^2 \psi(x) P_k(x) \, dx = \frac{k+1}{2k+1} \int_{-1}^{1} x\,\psi(x) P_{k+1}(x) \, dx + \frac{k}{2k+1} \int_{-1}^{1} x\,\psi(x) P_{k-1}(x) \, dx.$$
$$\text{(A.4)}$$

We apply the recurrence relation (3.100) for each term on the right hand side of equation (A.4), thus the first integral on the R.H.S. is simplified to

$$
\begin{aligned}
\int_{-1}^{1} x\,\psi(x)P_{k+1}(x)\,dx &= \frac{(k+1)+1}{2(k+1)+1}\int_{-1}^{1}\psi(x)P_{k+2}(x)\,dx + \frac{k+1}{2(k+1)+1}\int_{-1}^{1}\psi(x)P_{k}(x)\,dx\,, \\
&= \frac{k+2}{2k+3}\int_{-1}^{1}\psi(x)P_{k+2}(x)\,dx + \frac{k+1}{2k+3}\int_{-1}^{1}\psi(x)P_{k}(x)\,dx\,.
\end{aligned}
$$
(A.5)

And the second integral is solved as

$$
\begin{aligned}
\int_{-1}^{1} x\,\psi(x)P_{k-1}(x)\,dx &= \frac{(k-1)+1}{2(k-1)+1}\int_{-1}^{1}\psi(x)P_{k}(x)\,dx + \frac{k-1}{2(k-1)+1}\int_{-1}^{1}\psi(x)P_{k-2}(x)\,dx\,, \\
&= \frac{k}{2k-1}\int_{-1}^{1}\psi(x)P_{k}(x)\,dx + \frac{k-1}{2k-1}\int_{-1}^{1}\psi(x)P_{k-2}(x)\,dx\,.
\end{aligned}
$$
(A.6)

Substituting the solutions of the integrals obtained in equations (A.5) and (A.6) into equation (A.4) we get

$$
\begin{aligned}
\int_{-1}^{1} x^2\psi(x)P_k(x)\,dx = \frac{k+1}{2k+1}&\left(\frac{k+2}{2k+3}\int_{-1}^{1}\psi(x)P_{k+2}(x)\,dx + \frac{k+1}{2k+3}\int_{-1}^{1}\psi(x)P_k(x)\right) \\
+\frac{k}{2k+1}&\left(\frac{k}{2k-1}\int_{-1}^{1}\psi(x)P_k(x)\,dx + \frac{k-1}{2k-1}\int_{-1}^{1}\psi(x)P_{k-2}(x)\,dx\right).
\end{aligned}
$$
(A.7)

Rearranging this expression and using the definition of the ψ_k we obtain

$$
\begin{aligned}
\int_{-1}^{1} x^2\psi(x)P_k(x)\,dx = \frac{(k+1)(k+2)}{(2k+1)(2k+3)}\,\psi_{k+2} &+ \left[\frac{(k+1)^2}{(2k+1)(2k+3)} + \frac{k^2}{(2k+1)(2k-1)}\right]\psi_k \\
&+ \frac{k(k-1)}{(2k+1)(2k-1)}\,\psi_{k-2}\,.
\end{aligned}
$$
(A.8)

Simplifying the ψ_k coefficients yields

$$
\begin{aligned}
\int_{-1}^{1} x^2\psi(x)P_k(x)\,dx = \frac{(k+1)(k+2)}{(2k+1)(2k+3)}\,\psi_{k+2} &+ \frac{(2k^2+2k-1)}{(2k+3)(2k-1)}\,\psi_k \\
&+ \frac{k(k-1)}{(2k+1)(2k-1)}\,\psi_{k-2}\,.
\end{aligned}
$$
(A.9)

which is the form of the equation (3.106).

Appendix B

B.1 Introduction

In this appendix, we estimate the power spectra of two stationary correlated processes, namely X_t and Y_t using two different spectra estimations; the averaging across periodograms of adjacent sections (disjoint sections method) and the autoregressive model of order 1. The power spectra will be estimated for different length of data ranging between 5 and 65 seconds, then the comparison between the results of two approaches will be made using L_1 and L_2 error estimates.

B.2 Autoregressive Model AR(1)

Let X_t and Y_t be two correlated stationary processes of length T and they are connected by the autoregressive model

$$
\begin{aligned}
X_t &= \alpha\, X_{t-1} + \beta\, Y_{t-1} + \sigma\, E_{X,t}\,, \\[6pt]
Y_t &= \alpha\, Y_{t-1} + \beta\, X_{t-1} + \sigma\, E_{Y,t}\,,
\end{aligned}
\tag{B.1}
$$

where $E_{X,t}$ and $E_{Y,t}$ denote uncorrelated white noise processes of unit variance across time and within themselves, with the covariance matrix Σ. The observations X_t, Y_t are taken at time t, while X_{t-1} and Y_{t-1} are the observations taken at time $t - \Delta t$ and Δt is the time interval between two observations. Finally α, β and σ represent the AR(1) parameters of the trial model (B.1). To facilitate this calculation it is convenient to write equation (B.1) in the matrix form

$$
\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \sigma \begin{bmatrix} E_{X,t} \\ E_{Y,t} \end{bmatrix}
\tag{B.2}
$$

The Fourier transform

$$\mathcal{F}(f) = \int_{-\infty}^{\infty} f(t) e^{2i\pi ft} \, dt \tag{B.3}$$

is applied to equation (B.2) to get the frequency representation of this equation, that is,

$$\begin{bmatrix} \mathcal{X}(f) \\ \mathcal{Y}(f) \end{bmatrix} = e^{2i\pi f \Delta t} \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} \mathcal{X}(f) \\ \mathcal{Y}(f) \end{bmatrix} + \sigma \begin{bmatrix} \mathcal{E}_X(f) \\ \mathcal{E}_Y(f) \end{bmatrix}. \tag{B.4}$$

This equation can be rearranged into the form

$$\begin{bmatrix} 1 - \alpha e^{i\theta} & -\beta e^{i\theta} \\ -\beta e^{i\theta} & 1 - \alpha e^{i\theta} \end{bmatrix} \begin{bmatrix} \mathcal{X}(f) \\ \mathcal{Y}(f) \end{bmatrix} = \sigma \begin{bmatrix} \mathcal{E}_X(f) \\ \mathcal{E}_Y(f) \end{bmatrix}, \tag{B.5}$$

where $\theta = 2\pi f \Delta t$, and solved for the Fourier coefficients $\mathcal{X}(f)$ and $\mathcal{Y}(f)$ to obtain

$$Z(f) = \sigma A^{-1} \xi \tag{B.6}$$

where

$$Z(f) = \begin{bmatrix} \mathcal{X}(f) \\ \mathcal{Y}(f) \end{bmatrix}, \qquad A = \begin{bmatrix} 1 - \alpha e^{i\theta} & -\beta e^{i\theta} \\ -\beta e^{i\theta} & 1 - \alpha e^{i\theta} \end{bmatrix}, \qquad \xi = \begin{bmatrix} \mathcal{E}_X(f) \\ \mathcal{E}_Y(f) \end{bmatrix}. \tag{B.7}$$

As the spectrum of the innovations is the identity matrix then the spectral density matrix of Z is

$$S = \mathbb{E}\left[Z(f) Z^H(f) \right] = \sigma^2 A^{-1} I A^{-H} = \sigma^2 \Delta t \, A^{-1} A^{-H}, \tag{B.8}$$

where H denotes the complex-conjugate (Hermitian) transpose. The inverse of A is

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} 1 - \alpha e^{i\theta} & \beta e^{i\theta} \\ \beta e^{i\theta} & 1 - \alpha e^{i\theta} \end{bmatrix}, \tag{B.9}$$

where $\det A = (1 - \alpha e^{i\theta})^2 - (\beta e^{i\theta})^2 = 1 + (\alpha^2 - \beta^2) e^{2i\theta} - 2\alpha e^{i\theta}$. Substituting into equation (B.8) and performing the matrix multiplication gives

$$S = \frac{\sigma^2 \Delta t}{|\det A|^2} \begin{bmatrix} 1 + \alpha^2 + \beta^2 - 2\alpha \cos\theta & 2\beta(\cos\theta - \alpha) \\ 2\beta(\cos\theta - \alpha) & 1 + \alpha^2 + \beta^2 - 2\alpha \cos\theta \end{bmatrix}. \tag{B.10}$$

The calculation of the numerator of equation (B.9) yields

$$\begin{aligned} |\det A|^2 &= |1 + (\alpha^2 - \beta^2) e^{2i\theta} - 2\alpha e^{i\theta}|^2 \\ &= 1 + 2(\alpha^2 - \beta^2) \cos 2\theta - 4\alpha(1 + \alpha^2 - \beta^2) \cos\theta + (\alpha^2 - \beta^2)^2 + 4\alpha^2 \tag{B.11} \\ &= (1 - \alpha^2 + \beta^2)^2 + 4\left[(\alpha^2 - \beta^2) \cos^2\theta - \alpha(1 + \alpha^2 - \beta^2) \cos\theta + \alpha^2 \right]. \end{aligned}$$

Hence, the power spectra and cross-spectra of the two processes X_t and Y_t can be obtained directly from equations (B.10) and (B.11).

In the following section we will simulate two correlated processes using AR(1) for different length of data and apply the finite Fourier transform to these data.

B.3 Numerical spectra analysis

In this section we will generate two correlated processes X_t and Y_t using the autoregressive model AR(1) for different length of data. Specifically, $T = 10, 20, 30, \cdots 65$ seconds, and the parameters α, β and σ are given the values 0.5, 0.1, 1 respectively, that is

$$X_t = 0.5\,X_{(t-1}+0.1\,Y_{t-1} +\ N(0,1)\,,$$

$$\text{(B.12)}$$

$$Y_t = 0.5\,Y_{t-1} + 0.1\,X_{t-1} +\ N(0,1)\,.$$

The analytical power spectra derived in the previous section and the estimated power spectra using the disjoint section method will be computed. In fact we will consider the power spectrum of the process X_t but the power spectrum of the process Y_t can be obtained in a same way. For example, the analytical power spectrum of the process X_t when $T = 10,000$ milliseconds is calculated straightforwardly from equations (B.10) and (B.11) and plotted in figure (B.1),
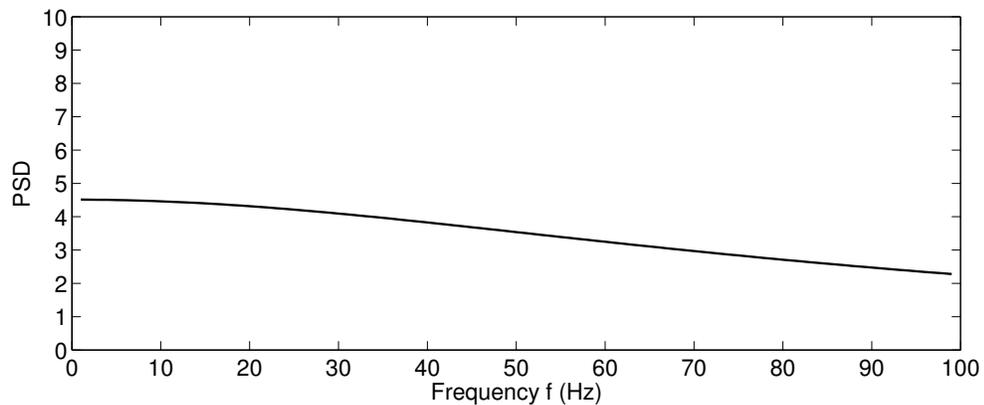

Figure B.1: This figure shows the AR(1) power spectrum of the both processes X_t when the length of records is $T = 10$ seconds.

To estimate the spectral density using the "Disjoint sections method" depends on partitioning the whole data T into L non-overlapping samples each of length N such that $L = T/N$. In this particular calculation we choose $N = 1000$ to give physical frequencies

in Hertz (Hz), consequently, the number of samples that will contribute to the estimate of the spectra and cross-spectra of the process X_t of the specific length T milliseconds is (L). The Finite Fourier transform is used to compute the Fourier coefficients for the associated sub-sample, namely for the l-th sample at frequencies between $f = 1\,\text{Hz}$ to $f = 100\,$ Hz in intervals of $1\,\text{Hz}$ by the formulae

$$
\begin{aligned}
d_X^T(f,l) &= \sum_{t=(l-1)N}^{lN-1} X(t)\, e^{-2i\pi ft} \\
d_Y^T(f,l) &= \sum_{t=(l-1)N}^{lN-1} Y(t)\, e^{-2i\pi ft} .
\end{aligned}
\tag{B.13}
$$

The estimated cross-spectra and spectra of the processes X_t and Y_t are computed from the formula

$$
I_{ij}^T(f) = \frac{1}{2\pi LT} \sum_{l=1}^{L} d_i^T(f,l)\overline{d_j^T}(f,l)\,; \qquad i,j = X,Y\,.
\tag{B.14}
$$

The periodogram of the processes X_t when the length of the data is $T = 10,000$ milliseconds is plotted in Figure (B.2)
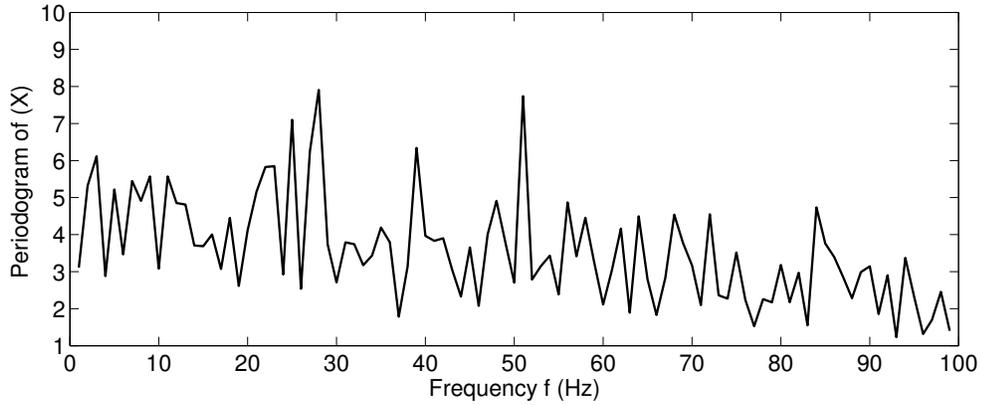

Figure B.2: This figure shows the periodogram of the processes X_t when the length of records is $T = 10$ seconds.

Figure (B.3) illustrates the comparison between the analytical power spectrum calculated directly from the autoregressive model AR(1), denoted by the solid line, and the estimated power spectrum of the processes X_t, denoted by the dotted-line, when the duration of the sample is ten seconds
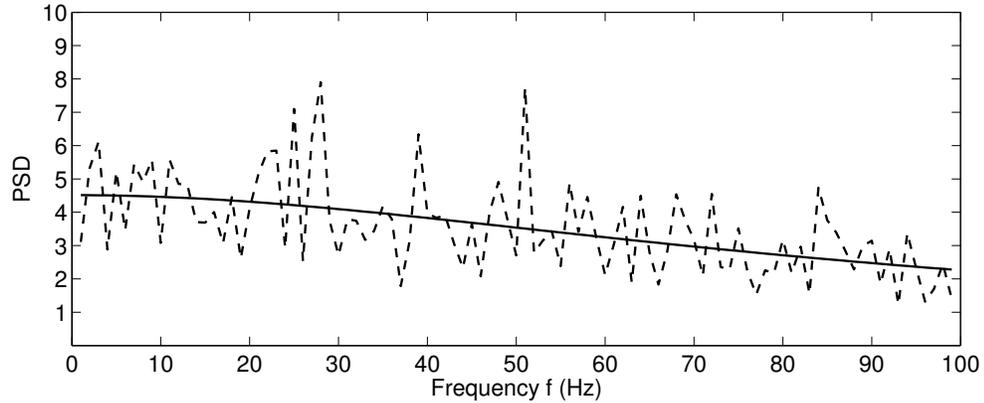
Figure B.3: This figure shows the comparison between the power spectrum calculated directly from the AR(1), solid-line, and the estimated power spectrum of the process X_t, dotted-line, when $T = 10$ seconds.

Similarly, the same procedures have been used to estimate the power spectra and cross-spectra of the processes X_t when the length of the data is $T = 15, 20, \cdots 65$ seconds by means of both approaches. We note that the AR(1) power spectrum remains having the same values since the AR(1) parameters are fixed. We note that as the size of sample increases the estimated power spectrum converges to the analytical power spectrum. Figures (B.4) and (B.5) display the analytical and the estimated power spectra (periodograms) of the processes X_t when the duration of time is 20 seconds and 30 seconds respectively.
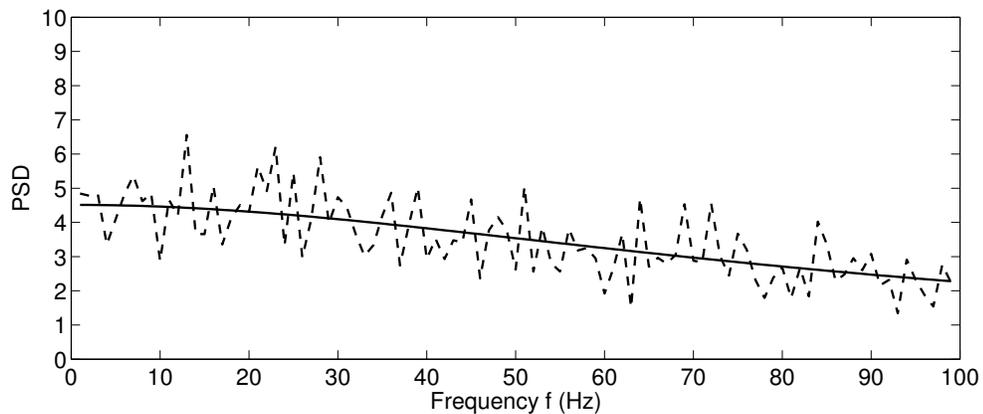

Figure B.4: This figure shows the comparison between the power spectrum calculated directly from the AR(1), solid-line, and the estimated power spectrum of the process X_t, dotted-line, when $T = 20$ seconds.
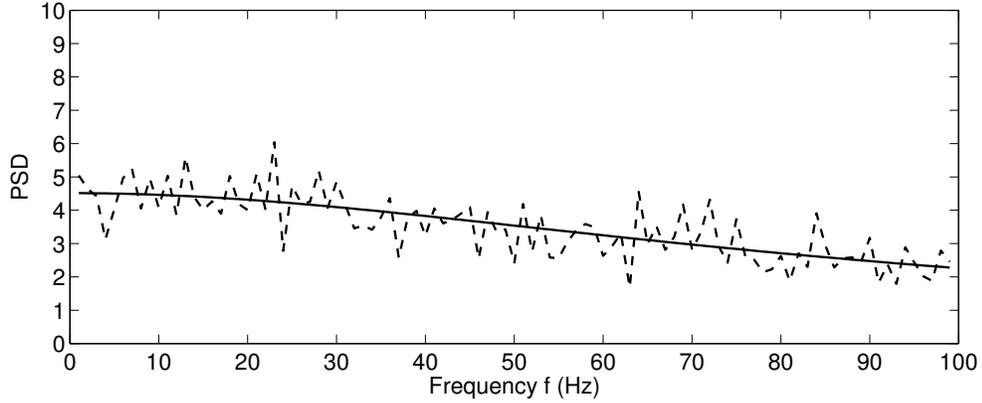
Figure B.5: This figure shows the comparison between the power spectrum calculated directly from the AR(1), solid-line, and the estimated power spectrum of the process X_t, dashed-line, when $T = 30$ seconds.

B.4   Error estimates

The convergence of the estimated power spectrum can be measured by using one of the two error norms L_1 and L_2 which defined as

$$L_1: \qquad Err \;\; = \;\; \frac{1}{n}\sum_{k=1}^{n}|ARPS - ESPS|\,; \qquad n = 1, 2, \cdots, 100\,, \tag{B.15}$$

$$L_2: \qquad Err \;\; = \;\; \sqrt{\frac{1}{n}\sum_{k=1}^{n}|ARPS - ESPS|^2}\,; \qquad n = 1, 2, \cdots, 100\,, \tag{B.16}$$

where **ARPS** represents the power spectrum derived from the autoregressive model and **ESPS** represents the estimated power spectrum obtained by applying the disjoint sections method, $|\,|$ denotes the absolute value of the difference at frequency $f \in [1, 100]$ Hz, and n is the number of the data. Specifically, the error estimates have been calculated from the natural logarithm of the power spectra instead of the power spectra themselves for the different size of samples, particularly $T = 5, 10, 15, 25, \cdots, 65$ seconds, five seconds apart, and arranged in the table (B.1) (upper panel) when the length of the used data ranging between 5 and 35 seconds and (lower panel) when the length of the data takes the higher values between 40 and 65 seconds of the process X_t. A similar strategy can be performed for the process Y_t. Thus, it can be seen clearly that the values of the two error norms decrease as the size of the data increases. This in turn means that the two approaches to the calculation of the spectra are convergent when the size of the sample is sufficiently large. Also, the estimated power spectra converges faster for the L_1 norm than for the L_2.

Sample Size	5	10	15	20	25	30	35
$\log_{10} L_1$	0.167530	0.113257	0.093697	0.080062	0.068631	0.061420	0.054126
$\log_{10} L_2$	0.202671	0.139274	0.119233	0.103447	0.088020	0.081384	0.070510

Sample Size	40	45	50	55	60	65
$\log_{10} L_1$	0.050303	0.046334	0.042947	0.040320	0.039753	0.036857
$\log_{10} L_2$	0.067820	0.062417	0.056559	0.050721	0.048580	0.046239

Table B.1: This table shows the L_1 and L_2 errors estimates between the logarithm to base 10 of the power spectra of the process X_t at different length of data , where $T = 5, 10, 15, \cdots, 35$ seconds, 5 seconds apart (upper panel) and for $T = 40, 45, \cdots, 65$ seconds, 5 seconds apart (lower panel).

The logarithm to base 10 of the L_1 and L_2 errors values are written in the table (B.1) for the sample sizes $T = 5, 10, 15, \cdots, 35$ seconds (upper panel), and $T = 40, 45, \cdots, 65$ seconds (lower panel) for the process X_t are explained in the figure
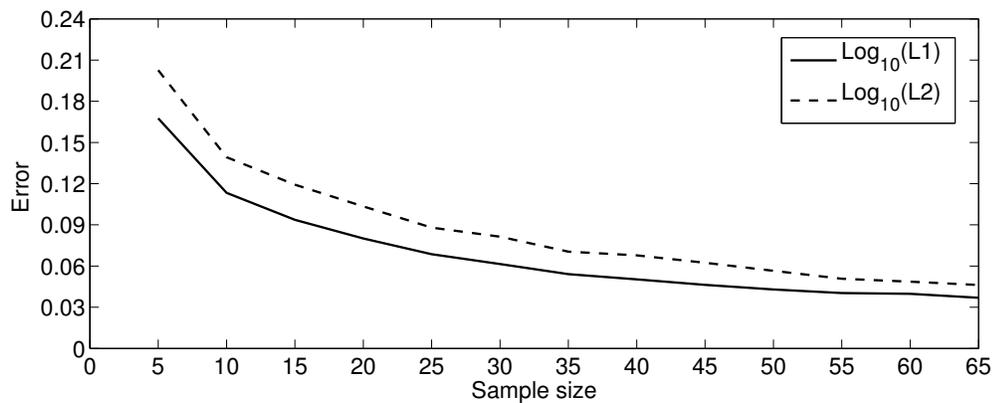

Figure B.6: This figure shows the comparison between the L_1 error, solid-line and L_2 error, dotted-line for the logarithm of estimated power spectrum of the process X_t, when the size of sample increases from 5 seconds to 65 seconds.

B.5   Conclusion

This work focussed on estimating the power spectra of two correlated stationary stochastic processes using two different spectral density estimations; the autoregressive model of order 1 and the disjoint sections method. The investigation has been made for different sizes of sample. Specifically, for the duration $T = 10, 15, \cdots 65$ seconds, five seconds apart, and

the two types of the error norms have been employed to examine the convergence between the analytical and the estimated power spectra. The results of this investigation showed that the difference between these values become less as the size of the sample increases.

Appendix C

In this appendix we provide a definition of the kernel density estimation

C.1   Kernel density estimation

Kernel density estimation is a non-parametric method for estimating the probability distribution of a random variable from a finite sample of independent realisations of that random variable.

Let $x_1, x_2, \cdots, x_m$ be a sample of m independent identically-distributed realisations of the random variable X, then the kernel density estimate of probability density is

$$f(x) = \frac{1}{mh} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right), \tag{C.1}$$

where K is some kernel function and h is a smoothing parameter called the bandwidth. The standard Gaussian function with zero mean value and variance 1 has been used frequently as a kernel function. In this case formula (C.1) becomes

$$f(x) = \frac{1}{mh} \sum_{i=1}^{m} \frac{1}{\sqrt{2\pi}} e^{\frac{(x - x_i)^2}{2h^2}}. \tag{C.2}$$

The kernel density function used in Chapter 5 is that implemented in MATLAB and called *ksdensity* in MATLAB.