

**Multivariate Time Series Analysis
of Heteroscedastic Data with
Application to Neuroscience**

Kin Foon Kevin WONG

DOCTOR OF PHILOSOPHY

Department of Statistical Science

School of Multidisciplinary Science

The Graduate University for Advanced Studies

2005

This work was carried out under the supervision of

Professor Tohru Ozaki

Acknowledgments

First, I am greatly indebted to my advisor, Prof. Tohru Ozaki, for the many years of guidance and teaching he dedicated to our research. I treasure those hours of discussion, brainstorming and programming together with him.

Special thanks go to Prof. Juan Carlos Jimenez-Sobrino who taught me the skill of numerical simulations, helped and gave his advice in many discussions. He also applied his considerable skills to make the presentation of the works as smooth and coherent as possible while maintaining scientific accuracy.

Additional thanks to Prof. Andreas Galka who coauthored the paper submitted to *Computer in Biology and Medicine*. He taught me important lessons to help me get on board. From time to time his genius scientific jokes did spirit me up from the tedious research work.

I would like to extend warmest thanks to my other examiners, Prof. Yoshiyasu Tamura, Prof. Tomoko Matsui and Prof. Hiroyuki Nakahara. They went through the thesis many times and gave their valuable comment and suggestion to improve the thesis.

I am particularly grateful to Mrs Miyuki Sawada, who was the secretary of the Institute of Statistical Mathematics during these years. Her support along the way were important.

I would like to thank, Prof. Jorge Bosch for teaching me various computer skills; Dr Roy John and Dr Leslie Prichep for providing the EEG data and for dedicating their precious time to giving me comments and guidance; Prof. Norihiro Sadato for providing the head movement data and bimanual rotation data, allowing us to join all these inspiring projects.

Further thanks go to Prof. Takashi Tsuchiya, Prof. Yosihiko Ogata, Prof. Tomoyuki Higuchi and Prof. Takashi Nakamura. They gave me a chance to learn various topics in statistics.

I would also like to thank the Iwatani Naoji Foundation and the Atsumi International Scholarship Foundation for their financial support to my study.

Last but not least, I would like to extend my gratitude to my family: my parents, Bill, Sharon, Stephen, Aunt Sammy and John. They have supported me through the years and continued to encourage me to pursue and achieve my goals. Especially I would like to thank John, my greatest cousin. He is always there to stand by me.

Contents

Acknowledgments	iii
List of Figures	x
List of Tables	xi
List of Abbreviations	xiii
List of Symbols	xv
Abstract	xix
1 Introduction	1
2 Compartment Model	7
2.1 Prerequisite	7
2.1.1 ARMA Model	7
2.1.2 State Space Model	8
2.1.3 Kalman Filter	9
2.1.4 Analogy of ARMA and State Space	9
2.2 Transformation of State Space Model	12
2.3 Compartment Model	14
2.3.1 Companion Form	14
2.3.2 Rotation Form	17
2.4 Log-likelihood Function and Akaike Information Criterion	19
2.5 Appendix	20
2.5.1 Transformation Matrices	20
2.5.2 AR Compartment Model	20
3 Spectrum and Causality in State Space	23
3.1 Spectrum	23
3.1.1 Multivariate ARMA Spectral Density	24
3.1.2 State Space Spectral Density	25
3.2 Noise Contribution Ratio Causality	28

3.2.1	Multivariate ARMA Model NCR Causality	31
3.2.2	State Space Model NCR Causality	31
3.2.3	Time varying NCR Causality	32
4	Heteroscedasticity in State Space	35
4.1	Chi-square Statistics based on Innovations	35
4.2	State Space GARCH Model	38
4.2.1	GARCH Type Models	38
4.2.2	Compartment GARCH model formulation	40
4.2.3	Substitution of Unobservable Noise	40
4.3	Smoothing Technique	42
4.3.1	Simple Moving Average	42
4.3.2	Exponential Moving Average	43
4.4	Exogenous Variables	43
4.5	Appendix	44
4.5.1	Derivation of Conditional Expectation of Square of System Noise	44
5	Application 1 - Detection of Consciousness in Surgery	47
5.1	Introduction	47
5.2	Data	48
5.3	Analysis	50
5.3.1	Segmented Data Analysis	50
5.3.2	Compartment Modelling	51
5.3.3	GARCH Modelling	54
5.3.4	Simulation Study	56
5.4	Discussions	58
6	Application 2 - Neural Activity of a Coma Patient	59
6.1	Introduction	59
6.2	Data	60
6.3	Analysis	61
6.4	Discussions	64
7	Application 3 - Causality in Human Head Movement	65
7.1	Introduction	65
7.2	Data	65
7.3	Analysis	66
7.4	Discussions	73
8	Application 4 - Study of Synchronous Rotation of Fingers	75
8.1	Introduction	75
8.2	Data	76
8.2.1	Data Preprocessing	76

8.3	Analysis	79
8.3.1	NCR Causality of Multivariate AR Model	80
8.3.2	NCR Causality of State Space Model	84
8.4	Discussions	89
9	Conclusion	91
	References	93

List of Figures

2.1	Top left: data; bottom left: decomposed components with data; top right: spectrum of individual decomposed components and their sum; bottom right: characteristic roots of decomposed components.	17
3.1	Model spectrum and NCR	30
3.2	Estimated states, state variance, original data and NCR of the bivariate EEG data at 4s	33
3.3	Decomposition, noise variance and NCR of compartment-GARCH model at 4 instants.	34
4.1	Chi-square statistics of compartment model with and without exogenous input of noise variance	37
5.1	Time series plot of a raw EEG data, patient 58012, condition c01.	49
5.2	Time series plot of a transformed EEG data, patient 58012, condition c01.	49
5.3	Best AR order periodogram of the raw EEG data	50
5.4	Segmentwise AR(9) spectrum, patient 58012, condition c01.	50
5.5	Segmentwise compartment(4,0) spectrum, patient 58012, condition c01, fixed frequencies.	50
5.6	Segmentwise AR(8) Spectra	52
5.7	Estimated Compartments	55
5.8	Variance of Compartments	56
5.9	Simulated states and observation	57
5.10	Histogram of parameter estimates	57
6.1	Time series plot of data, and time varying FFT spectrum of side view and top view.	60
6.2	Time series plot of data, and time varying best AR spectrum of side view and top view.	61
6.3	State decomposition and its corresponding driving noise variance.	62
6.4	Time series plot of data, and time varying state space GARCH spectrum of side view and top view.	63

6.5	Time series plot of data, and time varying de-trended state space GARCH spectrum of side view and top view.	63
6.6	Time series plot of innovations, and time varying best AR spec- trum of side view and top view.	63
6.7	Time series plot of normalized innovations, and time varying best AR spectrum of side view and top view.	64
7.1	A photo for illustrating the set up of the experiment.	66
7.2	Time series of the angular movement of 2 subject and the object.	67
7.3	Time series of yawing of the 2 subjects	67
7.4	Spectrum of the time series of yawing of the 2 subjects.	68
7.5	AR(5) model spectrum and NCR	68
7.6	AR(8) model spectrum and NCR	69
7.7	AR model spectrum and NCR of order 13, 21, 34 and 55	70
7.8	AR(8) model residual	71
7.9	Time varying NCR at 4 different time points.	72
8.1	Finger rotation data	77
8.2	First and second difference finger rotation data and their FFT spectrum	78
8.3	Matrix plot of time series and FFT spectra at 3 different sub- sampling rate and 2 different differencing order of finger rotation data	79
8.4	Preprocessed finger rotation data	80
8.5	AIC of AR models up to order 30.	81
8.6	Spectra and NCR of a fitted AR(11) model	83
8.7	Innovation from a fitted AR(11) model	83
8.8	Argand diagram of characteristic roots of (a) AR(11) model and (b) state space model	85
8.9	Trajectory of $H(j)$	87
8.10	NCR of the state space model with different H	88
8.11	Innovation from a state space model	89

List of Tables

2.1	A list of transformation matrices	20
5.1	Compartment (2,1) Model	51
5.2	Summary of the parameters of fitted compartment (3,0), (3,1), (4,0), (4,1).	53
5.3	Mean and standard error of parameter estimates	57
7.1	AIC of fitted multivariate AR models of different orders.	70

List of Abbreviations

Hz	Hertz
s	second of time
ACF	Autocorrelation Function
AIC	Akaike Information Criterion
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroscedasticity
ARMA	Autoregressive Moving Average
BOLD	Blood Oxygenation Level-Dependent
EEG	Electroencepharogram
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedasticity
FFT	Fast Fourier Transformation
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
MA	Moving Average
NCR	Noise Contribution Ratio
PACF	Partial Autocorrelation Function
VAR	Vector Autoregressive

List of Symbols

A	(capital letter) matrix of any size
$ A $	determinant of matrix A
A^{-1}	inverse of matrix A
A'	transpose of matrix A
A^H	adjoint (or conjugate transpose or Hermitian) of matrix A
A^-	pseudo-inverse of matrix A
\mathbf{a}	(bold-face) vector of any size
$a^{(j)}$	j -th element of vector \mathbf{a} or matrix A
$ a $	absolute value of real number a or modulus of complex number a
$\arg(a)$	argument of complex number a
\bar{a}	conjugate of complex number a
a	real part of eigenvalue λ
b	imaginary part of eigenvalue λ
C^2	chi square statistics
F	$m \times m$ state transition matrix
\hat{F}	transition matrix of compartment model in rotation form
\check{F}	transition matrix of compartment model in companion form
f	frequency
G	$m \times k$ system noise matrix
H	$\ell \times m$ observation matrix
I	identity matrix
i	unit imaginary number, $i^2 = -1$
K	Kalman gain
k	length of system noise vector
L	log likelihood
ℓ	length of observation vector in state space model or length of data vector in multivariate ARMA model
m	length of state vector

N	number of data
n	length of noise vector in multivariate ARMA model
p	AR order
Q	system noise variance matrix
q	MA order
R	observation noise variance matrix
r	number of pairs of complex conjugate eigenvalues
S	transformation matrix
s	number of real eigenvalues
t	time
u	ARCH order
v	GARCH order
$V_{t t}$	state variance filter estimate
$V_{t t-1}$	state variance predictor estimate
w	state noise
x or \mathbf{x}	state vector
y or \mathbf{y}	observation vector or data vector
α	ARCH coefficient in GARCH model
β	GARCH coefficient in GARCH model
γ_k	autocovariance at lag k
ϵ or $\boldsymbol{\epsilon}$	observation noise
η or $\boldsymbol{\eta}$	noise vector in multivariate ARMA model
Θ	MA coefficient matrix in multivariate ARMA model
θ	MA coefficient in univariate ARMA model
$\boldsymbol{\nu}$	innovation
λ	eigenvalue
ν	any standard normal white noise
ξ	exogenous variable
π	circumference-diameter ratio
ρ_k	ACF at lag k
Σ	Noise covariance matrix in multivariate ARMA model
σ^2	diagonal elements of observation noise variance
τ^2	diagonal elements of system noise variance
Φ	AR coefficient matrix in multivariate ARMA model
ϕ	AR coefficient in univariate ARMA model
χ^2	chi square distribution

Ω	inverse of covariance of innovation
$\mathcal{F}_f(\Phi)$	Fourier Transform of AR coefficients
$\mathcal{F}_f(\Theta)$	Fourier Transform of MA coefficients
$\text{diag}(\lambda_1, \dots, \lambda_m)$	diagonal matrix with elements $\lambda_1, \dots, \lambda_m$
$E(x)$	expectation of a random variable x
$\text{Var}(x)$	variance of a random variable x
$\text{Cov}(x, y)$	covariance between two random variables x and y
$\text{NCR}(\sigma^2, y)$	NCR from noise variance σ^2 to y

Abstract

This thesis summarizes statistical analysis of some multivariate heteroscedastic time series data, including 2 sets of data from physiological experiments and 2 sets of EEG data about anaesthesia and coma.

The aim of this thesis is to provide a statistical tool for analyzing multivariate data which contains non-stationary and heteroscedastic characteristics.

The main contribution of this thesis is that we combine the linear state space model and GARCH model to develop a state space-GARCH model. The state space-GARCH model can describe the non-stationary characteristics of the system noise variance. In particular we adopt a special structure of the linear state space model to decompose a data into components by their frequencies. Combining a heteroscedasticity model and a state space model is carried out by fully utilizing the information of innovations and expected values from the filtering process.

Another contribution of the thesis is that we extend Akaike's NCR from constant noise variance to heterogeneous noise variance in order to study time-varying causality. By applying heteroscedasticity models, the phenomenon of an evolving causality relationship can be depicted.

All these methods are illustrated by their application to EEG data including the study of consciousness under anaesthesia and coma, and also to a physical data of head and finger movement.

Keywords:

multivariate time series, state space model, Kalman filter, frequency decomposition, autoregressive conditional heteroscedasticity, heterogeneity of variance, compartment-GARCH model, noise contribution ratio, spectral causality, EEG, anaesthesia, coma, bimanual coherence, handedness score.

Chapter 1

Introduction

This thesis focuses on two problems in modelling multivariate time series of neuroscience. The first is how can the non-stationarity of the data be described by means of statistical time series model in order to explain the non-stationary activities inside the brain. The second problem is what the functional role of statistical time series model is in explaining causality and feedback. These two problems in neuroscience research are important but often neglected.

The frequency approach is widely adopted in EEG study. In the scope of time series analysis in neuroscience application, the earliest work was John et al. (1977), who advocated the Broad Band Spectral Model, smoothing the FFT spectrum and looking at the power intensity of the five major frequency bands, which is considered as a pioneering work. Pascual-Marqui et al. (1988) characterized EEG spectra by the $\xi\alpha$ model, which is a mixture function in the frequency domain containing a few parameters associated with spectral peaks. This model is based on the theoretical assumption that there are different unobserved generations for each spectral peaks. Later, Valdés-Sosa et al. (1990) proposed the Narrow Band Spectral Model which is more flexible than the Broad Band Spectral Model and the $\xi\alpha$ Model. All the three models describe the EEG data in frequency domain and especially characterize the activities in terms of frequency bands. Frequency, number of oscillation in a second of time, characterizes different neural activities induced by neuron transmission at different rate, which is caused by and resulted in respective phenomena.

On the other hand, multivariate autoregressive model have been used for EEG data analysis (see Rappelsberger & Petsche, 1975; Gersch & Yonemoto, 1977; Franaszczuk et al., 1985). A multivariate AR model can be easily estimated by an iterative matrix calculation of successive AR coefficients by the Whittle (1963) algorithm. Despite a fast computation it can be interpreted in the time domain in addition to the frequency domain. A smooth spectrum can be drawn based on the AR coefficients and the noise variance of the model. Power intensity of several frequency bands can be seen apparently. However, because of the assumption of stationarity of the signal, this approach

is useful in clinical application to characterize the EEG of only homogeneous pattern. In real application, an EEG data of homogeneous pattern is fulfilled only by a subjective selection of piece of data by medical doctors and neuroscientists. We are trying to propose models which automatically captures the non-homogeneous pattern.

In most real situations, brain activity gives complicated EEG data measurement which does not show a homogeneous pattern. Either the data changes between different levels, gives a non-stationary mean, or fluctuates in different amplitude size, shows inhomogeneity of variance.

The basic idea of heterogeneity of variance is that in a stationary process the variance of input noise to the model is not constant in time. In the case of frequency approach of non-stationary EEG analysis, it is possible to split a EEG data into many segments, and fit a model to each segments. Each model corresponding to a segment takes the characteristics of the segment, so that the change of power intensity can be seen by comparing the estimated power intensity of frequency bands. This method of splitting a data into segments requires an assumption of stationarity within each segment. If the assumption does not hold, these spectral model are not reliable. Although the length of the segments can be shortened so as to obtain locally stationary data, such approach would have the disadvantage of reduced resolution in frequency domain. Improved resolution in time domain, desirable in order to pick out distinctive temporal characteristics in the data, has to be paid by reduced resolution in frequency domain, and vice versa. Therefore, using the spectral model to analyze a non-stationary data is unsuitable.

On the other hand, multivariate AR model is also useful to solve the problem of heterogeneity of variance, when it is applied together with volatility models such as GARCH model, jump model or stochastic volatility model, multivariate AR model. In terms of model estimation, volatility models cannot reduce the magnitude of innovation, but adjust the innovation variance to adapt to the size of innovation, leading to an increase of likelihood for a better model. Consequently, volatility models cannot further improve point prediction beyond the AR model, but they are important to predicting interval estimates and volatility prediction. This would be the reason why they have been widely used in risk management in finance.

We use the class of state space model to formulate the multivariate AR model. In general a linear multivariate AR model and a linear state space model are equivalent, which can be transformed from one to another one. We choose state space model over the other because state space model can reconstruct the hidden unobservable state behind the data and it is capable to differentiate the observation noise out of the system noise. With a special structure of the transitional matrix of the state space model, we divide the EEG data into various components (like the $\xi\alpha$ model mentioned above, each of which are oscillating at a specified frequency and driven by an independent

noise. We name this model a *compartment model*.

Then we apply the GARCH model inside the state space model to build up the state-space GARCH model. The GARCH model adjusts variance of noise driving each decomposed component. Adaptively adjusting the system noise variance, we are modifying the innovation variance so as to improve the likelihood, and as a result we get a better model.

By a compartment model we can clearly see the decomposition of data over a time axis, by the same token of a model spectrum showing the power intensity over a frequency axis as the spectral models do. The time-varying variance provides us good information of the changes of power of driving noise. State space-GARCH model is useful for investigating the principles that underlie the non-stationary neuroscience time series data, which will be shown step by step in this thesis.

Another important element of this thesis concerns with causal relations in multivariate time series, in particular, in the context of noise contribution ratio introduced by Akaike (1968). Causal relations, or causality, is a vague term since it has a slightly different meaning in different literatures. Intuitively, the study of causality is to relate the “cause” and the “result”, and to conclude any “the cause precedes the result” phenomenon. For noise contribution ratio, based on a multivariate time series model, any random noise is the “cause” and the time series is the “result”. We are interested in how time series cause each other, or equivalently, how the driving noises cause the time series.

Noise contribution ratio, in contrast to a yes or a no from Granger’s causality (see Granger, 1969), gives the level of causality quantitatively in a 0 to 100% scale. The level of causality is shown by the contribution ratio of a noise variance to an observation variable. Looking at the ratio and the spectral density, we can conclude the significance of each contributing noise to each time series. This is an important tool when we want to compare the level of two way causality, ie feedback, among several time series, especially when the causal relations is biased. In this thesis, we have two development on the topic of NCR. First, we generalize the NCR from multivariate AR model to multivariate ARMA model and further to linear state space model. Second, we advocate time varying NCR, when the variance is varying or when the model is changing over the time.

There is a huge library of study on the causal relations of EEG data and fMRI data, for example, see Friston et al. (1995); McIntosh (2000); Yamashita et al. (2005). Knowing the causality between the time series, we can know better the mechanism of the brain activity on the causal relations issue. We will apply the NCR method to two physical data also in order to show the new development of NCR, and some conclusions which is related to neuroscience study.

Notwithstanding a broad contents of this thesis, the materials of state space-GARCH model, the new NCR causality theory and several real appli-

cations on neuroscience are integrated and allocated into three chapters of theoretical issues and four chapters of applications.

In chapter 2, we introduce the compartment model, which is capable to decompose a time series into several components. To do so we first review the linear ARMA model and its relation with the state space model. We will see that by multiplying a non-singular transformation matrix, we can preserve the same model structure even though the state vector and transition matrix are projected onto another space. This is useful to explain why we can have good results using the compartment model even though the parameters have so strong restrictions.

In chapter 3, we study the new development of NCR in multivariate time series analysis, especially the NCR of a multivariate ARMA model or a state space model, and the time-varying occasion. We start from deriving the matrix spectral density of a multivariate ARMA model and of a state space model, then move on to the NCR and causality.

In chapter 4, we discuss how the GARCH model is combined with a state space model. We use chi-square statistics to show the importance of implementing a heteroscedasticity model together to model non-stationary data. The chi-square statistics is also a new method of diagnostic checking of innovation, as well as a quantity for feeding the GARCH model in state space model. Then we demonstrate how we calculate a conditional expectation for an underdetermined system residual so as to put into the GARCH equation. At the end of the chapter we bring out the idea of smoothing by moving average so that we can better operate the GARCH model.

Chapters 5 to 8 serve to give the analysis of 4 data sets. All data are related to neuroscience. In particular they are multivariate time series and possessing a lot of interesting phenomena of non-stationarity.

In chapter 5, we will analyze an EEG data during the onset of anaesthesia. A common practice of analyzing an EEG data is cutting out segments of data with a subjective judgement of a medical doctor before a spectral analysis. Such analysis requires the assumption of stationarity of variance, which is not always conceivable. Therefore, during a surgery, we could expect the spectral density of the EEG data to change over time, thus we need a model of heteroscedastic property so as to explain the variation of variance across time. The changes of variance directly affect the power intensity of the spectrum shape, and consequently give information of the onset of anaesthesia to help monitor the level of consciousness.

In chapter 6, we follow the topic of consciousness in chapter 5 to analyze an EEG data of a coma patient, who became unawaken after a traffic accident. The data was measured when the patient was receiving verbal stimuli, that a medical doctor told him to think of a favorite song, to recall his programmer job and so on. The stimuli led to some variations in the EEG. With the compartment-GARCH model we can calculate the power intensity of each main

frequency domain at different time. This is useful to show medical doctors the real time response of the coma patient to stimulus.

In chapter 7, we analyze a head movement data and study the non-stationary causal relations by NCR based on a bi-variate AR-GARCH model. Two subjects are sitting face to face and their head movement was measured. One subject observed and followed another subject turning his head to look at an object placed at a meter away from them. The same motion was repeated for several times. The data possesses a strong causal relations as a result. It is a one-way causal relation because there is only one subject following another subject's movement but not the opposite way. Using the Granger's causality theory one may conclude from the non-zero AR coefficients that there is a two-way feedback between two time series. We, instead, show that there is only a one-way causality when we interpret with the NCR theory. In addition to the bi-variate AR model with a stationary variance assumption we apply the GARCH model to the residual and capture the causal relations through different size of variance at different time, which on one hand provides a better pattern of residuals and on the other hand assures our conclusion on the causality through NCR. The time varying NCR causality is important to the study of attention and attraction. The tool can be applied to data of head movement from experiments such as attention of infants and mothers, attention of students and teachers in a classroom and so on. The functioning of a human body in responding to any attraction is interesting to psychologists and physio-scientists. These psychological and physical theory may also be beneficial to neuroscience study to give any intuitive idea of the unnecessary head movement motion during data measurement.

In chapter 8, we further analyze a bimanual movement of index fingers with the NCR method. In contrast to the classical NCR on multivariate AR models proposed by Akaike (1968) and the generalization by Tanokura & Kitagawa (2003) we consider NCR of multivariate ARMA models and state space models. Taking the advantage of estimating state variables of different dynamics in state space model, we break the three-parameter covariance matrix of a bi-variate AR model into three variances of three independent noises in the state equation. We let one of the three noises drive a state variable which becomes a common driving force to the two time series. We will explain the NCR based on the state space model, that the causality of the motion of the fingers after removing a significant common driving force.

Chapter 2

Compartment Model

In this chapter, explanations will be given of some statistical concepts and methods that are essential for understanding the discussion in the following chapters. First, some basic concepts of time series model which play fundamental roles in building a statistical model will be discussed. Second, idea and details of compartment model will be elucidated. In the end of the chapter, some extensions of the compartment model will be mentioned.

Compartment model is a special case of a state space model which we will use throughout the whole thesis. The compartment model was introduced by Ozaki (2003) and was applied to an EEG data by Wong et al. (2006). It is designed for a given time series to be decomposed into a set of noise driven processes, each corresponding to a unique frequency oscillation. The compartment model makes use of the unimodal property of AR(2) spectrum to reconstruct the spectral density structure of a given time series.

2.1 Prerequisite

2.1.1 ARMA Model

ARMA model is a basic starting time series model in many time series textbooks: Box & Jenkins (1970); Brockwell & Davis (1991, 1996); Wei (1990, 2006). An ARMA model comprises 2 parts, namely the autoregressive part and the moving average part. The AR part contains a weighted sum of the lags of the previous data, automatically fitted into the model as regressors at every consecutive time point. The MA part contains a current noise plus a weighted sum of the previous noise and acts as an average of driving noise. We will use equation 2.1 to represent a univariate ARMA(p, q) process and equation 2.2 to represent a multivariate ARMA(p, q) process in the whole thesis.

Let y be a univariate time series. ϕ is AR coefficients and θ is MA coefficients. p is the AR order and q is the MA order of the model. η is Gaussian

noise with zero mean and a constant variance σ^2 .

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \theta_0 \eta_t + \theta_1 \eta_{t-1} + \cdots + \theta_q \eta_{t-q} \quad (2.1)$$

In most literatures, it is assumed that $\theta_0 = 1$. In this thesis, we will also assume $\theta_0 = 1$ for any univariate AR model unless it is specified.

Analogously, in the multivariate case, we let \mathbf{y} be an observation vector storing the ℓ -variate data. Φ is an $\ell \times \ell$ matrix which is the AR coefficient for the model. Again, p is the AR order and q is the MA order of the model. $\boldsymbol{\eta}$ is a noise vector of size $n \times 1$ and Θ is the MA coefficient of size $\ell \times n$. Σ is the variance of $\boldsymbol{\eta}$ of size $n \times n$, so that $\boldsymbol{\eta} \sim N(0, \Sigma)$. A multivariate AR model can be expressed in equation 2.2.

$$\mathbf{y}_t - \Phi_1 \mathbf{y}_{t-1} - \cdots - \Phi_p \mathbf{y}_{t-p} = \Theta_0 \boldsymbol{\eta}_t + \Theta_1 \boldsymbol{\eta}_{t-1} + \cdots + \Theta_q \boldsymbol{\eta}_{t-q} \quad (2.2)$$

Note that when Θ_0 is $\ell \times \ell$ identity matrix and the other Θ are zero, equation 2.2 becomes a multivariate AR model, which is used in most literature on multivariate time series analysis. In this thesis, we are aiming to have a general case for the multivariate time series modelling, by not only allowing the MA part involved, but also considering a more complicated situation when the noise vector is not as the same size as the observation vector. By using the compartment model which will be discussed later, we can obtain a sensible meaning for every element of the noise vector even though it is longer than the observation vector.

2.1.2 State Space Model

On the other hand, state space model is another useful time series model in many real application. Let \mathbf{x}_t be the state vector at time t of size $m \times 1$, \mathbf{y}_t be the observation vector at time t of size $\ell \times 1$. F is an $m \times m$ matrix representing the transition matrix of the state vector, G is an $m \times k$ matrix and H is an observation matrix of size $\ell \times m$. We formulate the state space model in equation 2.3.

$$\begin{cases} \mathbf{x}_t = F \mathbf{x}_{t-1} + G \mathbf{w}_t \\ \mathbf{y}_t = H \mathbf{x}_t + \boldsymbol{\epsilon}_t \end{cases} \quad (2.3)$$

The two equations above are commonly known as system equation and observation equation. In the system equation the state \mathbf{x}_t is filtered by F and driven by a system noise \mathbf{w}_t . \mathbf{x}_t is projected onto an observation space by the observation matrix H and further driven by an observation noise $\boldsymbol{\epsilon}_t$ to form \mathbf{y}_t , where \mathbf{w}_t is assumed to follow a multivariate Gaussian distribution $\mathbf{w}_t \sim N(0, Q)$ and $\boldsymbol{\epsilon}_t$ follows a univariate Gaussian distribution $\boldsymbol{\epsilon}_t \sim N(0, R)$.

Throughout the thesis we assume both Q and R are diagonal, ie $Q = \text{diag}(\tau^{(1)2}, \tau^{(2)2}, \dots, \tau^{(k)2})$ and $R = \text{diag}(\sigma^{(1)2}, \sigma^{(2)2}, \dots, \sigma^{(\ell)2})$.

2.1.3 Kalman Filter

Kalman (1960) introduced a filtering technique for state space models which can efficiently calculate the conditional prediction and the conditional filtered estimation of unobserved states. A comprehensive introduction to state space models and Kalman filtering has been provided by Kalman (1960); Harrison & Stevens (1976); Harvey (1989); Grewal & Andrews (2001); Durbin & Koopman (2001).

Kalman filter is an algorithm of computing the conditional expectation and variance in a state space model. The Kalman filter algorithm consists of a prediction step and a filtering step.

$$\text{Prediction} \quad \begin{cases} \mathbf{x}_{t|t-1} = F\mathbf{x}_{t-1|t-1} \\ V_{t|t-1} = FV_{t-1|t-1}F' + GQ_tG' \end{cases}, \quad (2.4)$$

$$\text{Filter} \quad \begin{cases} \boldsymbol{\nu}_t = \mathbf{y}_t - H\mathbf{x}_{t|t-1} \\ \Omega_t = (HV_{t|t-1}H' + R)^{-1} \\ K_t = V_{t|t-1}H'\Omega_t \\ \mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + K_t\boldsymbol{\nu}_t \\ V_{t|t} = (I - K_tH)V_{t|t-1} \end{cases}. \quad (2.5)$$

where $\mathbf{x}_{t|s} = E(\mathbf{x}_t|\mathbf{y}_1, \dots, \mathbf{y}_s)$ and $V_{t|s} = \text{Var}(\mathbf{x}_t|\mathbf{y}_1, \dots, \mathbf{y}_s)$ are the expectation and the variance of \mathbf{x}_t conditional on the observations $\mathbf{y}_1, \dots, \mathbf{y}_s$ and conditional on the state space model. $\boldsymbol{\nu}_t$ and Ω_t are respectively the innovation and the innovation variance inverse. By iterating between the prediction step and the filtering step at each time point $t = 1, \dots, N$, we could obtain the filter $\mathbf{x}_{t|t}$. The filter estimate is known to be the optimal estimate of \mathbf{x}_t based on the observations up to current under the assumption of Gaussian noise.

2.1.4 Analogy of ARMA and State Space

It is always possible to transform an ARMA model 2.2 to a state space model 2.3 and vice versa (see Akaike, 1968; Aoki, 1987; Ozaki, 2003; Gilbert, 1993). On one hand, there are many ways to convert an ARMA model to a state space model, which is comprehensively explained in Akaike's literatures (see Akaike, 1968). In example 2.1 we will show how a univariate ARMA(2,1) model is transformed into a state space model. This example is a necessary building block for the compartment model in the latter sections.

Example 2.1 (ARMA(2,1) to state space) *Let y_t be a univariate time series following ARMA(2,1) model.*

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t + \theta_1 w_{t-1}$$

The equivalent state space model is

$$\begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1}^{(1)} \\ x_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix} w_t$$

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix}.$$

It is easy to verify that the models are equivalent.

$$\begin{aligned} x_t^{(1)} &= \phi_1 x_{t-1}^{(1)} + x_{t-1}^{(2)} + w_t \\ x_t^{(2)} &= \phi_2 x_{t-1}^{(1)} + \theta_1 w_t \\ x_t^{(1)} &= \phi_1 x_{t-1}^{(1)} + \phi_2 x_{t-1}^{(2)} + w_t \\ &= \phi_1 x_{t-1}^{(1)} + \phi_2 x_{t-2}^{(1)} + w_t + \theta_1 w_{t-1} \\ y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t + \theta_1 w_{t-1} \end{aligned}$$

□

When we have any higher order ARMA model, we can use the same technique to formulate a state space model. Let $m = \max(p, q + 1)$, the state vector \mathbf{x} will be an $m \times 1$ column vector. F will be an $m \times m$ matrix, in which the first column contains the AR coefficients of the ARMA model, and upper diagonal entries are all 1 and other entries are all 0. G will be an $m \times 1$ column vector, in which the first element is 1 and the other entries are placed with the MA coefficients of the ARMA model. H will be an $m \times 1$ row vector in which the first element is 1 and the other elements are all 0. Both in the state space model and the ARMA model \mathbf{w} are equivalent, so both of them follow a normal distribution with zero mean and a common variance. In example 2.2, we will show the formulation from a state space model to a general ARMA model.

Example 2.2 (State Space to ARMA) Suppose we have a state space model 2.3. We assume the eigenvalues λ of the $m \times m$ transitional matrix F satisfy a characteristic equation 2.6.

$$\lambda^m - \phi_1 \lambda^{m-1} - \phi_2 \lambda^{m-2} - \dots - \phi_{m-1} \lambda - \phi_m = 0 \quad (2.6)$$

Then, by Cayley Hamilton Theorem, we have equation 2.7.

$$F^m - \phi_1 F^{m-1} - \phi_2 F^{m-2} - \dots - \phi_{m-1} F - \phi_m I = 0 \quad (2.7)$$

We can rewrite the state equation into ARMA($m, m - 1$) model using the

Cayley-Hamilton Theorem.

$$\begin{aligned}
& \mathbf{x}_t - \phi_1 \mathbf{x}_{t-1} - \phi_2 \mathbf{x}_{t-2} - \cdots - \phi_m \mathbf{x}_{t-m} \\
&= (F \mathbf{x}_{t-1} + G \mathbf{w}_t) - \phi_1 \mathbf{x}_{t-1} - \phi_2 \mathbf{x}_{t-2} - \cdots - \phi_m \mathbf{x}_{t-m} \\
&= (F - \phi_1 I) \mathbf{x}_{t-1} - \phi_2 \mathbf{x}_{t-2} - \cdots - \phi_m \mathbf{x}_{t-m} + G \mathbf{w}_t \\
&= (F - \phi_1 I) (F \mathbf{x}_{t-2} + G \mathbf{w}_{t-1}) - \phi_2 \mathbf{x}_{t-2} - \cdots - \phi_m \mathbf{x}_{t-m} + G \mathbf{w}_t \\
&= (F^2 - \phi_1 F - \phi_2 I) \mathbf{x}_{t-2} - \cdots - \phi_m \mathbf{x}_{t-m} + G \mathbf{w}_t + (F - \phi_1 I) G \mathbf{w}_{t-1} \\
&= \dots \\
&= (F^m - \phi_1 F^{m-1} - \cdots - \phi_{m-1} F - \phi_m I) \mathbf{x}_{t-m} \\
&\quad + G \mathbf{w}_t + (F - \phi_1 I) G \mathbf{w}_{t-1} + (F^2 - \phi_1 F - \phi_2 I) G \mathbf{w}_{t-2} + \dots \\
&\quad + (F^{m-1} - \phi_1 F^{m-2} - \cdots - \phi_{m-2} F - \phi_{m-1} I) G \mathbf{w}_{t-m+1} \\
&= G \mathbf{w}_t + (F - \phi_1 I) G \mathbf{w}_{t-1} + (F^2 - \phi_1 F - \phi_2 I) G \mathbf{w}_{t-2} + \dots \\
&\quad + (F^{m-1} - \phi_1 F^{m-2} - \cdots - \phi_{m-2} F - \phi_{m-1} I) G \mathbf{w}_{t-m+1}
\end{aligned}$$

And we can also rewrite the observation equation,

$$\begin{aligned}
& \mathbf{y}_t - \phi_1 \mathbf{y}_{t-1} - \phi_2 \mathbf{y}_{t-2} - \cdots - \phi_m \mathbf{y}_{t-m} \\
&= (H \mathbf{x}_t + \boldsymbol{\epsilon}_t) - \phi_1 (H \mathbf{x}_{t-1} + \boldsymbol{\epsilon}_{t-1}) - \cdots - \phi_m (H \mathbf{x}_{t-m} + \boldsymbol{\epsilon}_{t-m}) \\
&= H [G \mathbf{w}_t + (F - \phi_1 I) G \mathbf{w}_{t-1} + (F^2 - \phi_1 F - \phi_2 I) G \mathbf{w}_{t-2} + \dots \\
&\quad + (F^{m-1} - \phi_1 F^{m-2} - \cdots - \phi_{m-2} F - \phi_{m-1} I) G \mathbf{w}_{t-m+1}] \\
&\quad + [\boldsymbol{\epsilon}_t - \phi_1 \boldsymbol{\epsilon}_{t-1} - \phi_2 \boldsymbol{\epsilon}_{t-2} - \cdots - \phi_m \boldsymbol{\epsilon}_{t-m}] \\
&= [H G \mathbf{w}_t + \boldsymbol{\epsilon}_t] + [H (F - \phi_1 I) G \mathbf{w}_{t-1} - \phi_1 \boldsymbol{\epsilon}_{t-1}] \\
&\quad + [H (F^2 - \phi_1 F - \phi_2 I) G \mathbf{w}_{t-2} - \phi_2 \boldsymbol{\epsilon}_{t-2}] + \dots \\
&\quad + [H (F^{m-1} - \phi_1 F^{m-2} - \cdots - \phi_{m-2} F - \phi_{m-1} I) G \mathbf{w}_{t-m+1} - \phi_{m-1} \boldsymbol{\epsilon}_{t-m+1}] \\
&\quad - \phi_m \boldsymbol{\epsilon}_{t-m} \\
&\equiv \Theta_0 \boldsymbol{\eta}_t + \Theta_1 \boldsymbol{\eta}_{t-1} + \Theta_2 \boldsymbol{\eta}_{t-2} + \cdots + \Theta_{m-1} \boldsymbol{\eta}_{t-m+1} + \Theta_m \boldsymbol{\eta}_{t-m}
\end{aligned}$$

where Θ are block matrices of size $\ell \times (k + \ell)$,

$$\begin{aligned}
\Theta_0 &= \left(HG \mid I \right) \\
\Theta_1 &= \left(H (F - \phi_1 I) G \mid -\phi_1 I \right) \\
\Theta_2 &= \left(H (F^2 - \phi_1 F - \phi_2 I) G \mid -\phi_2 I \right) \\
&\vdots \\
\Theta_{m-1} &= \left(H (F^{m-1} - \phi_1 F^{m-2} - \cdots - \phi_{m-2} F - \phi_{m-1} I) G \mid -\phi_{m-1} I \right) \\
\Theta_m &= \left(0 \mid -\phi_m I \right) \\
\boldsymbol{\eta}_{t-j} &= \begin{pmatrix} \mathbf{w}_{t-j} \\ \boldsymbol{\epsilon}_{t-j} \end{pmatrix} \sim N \left(0, \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix} \right)
\end{aligned}$$

Each Θ contains two matrix blocks of sizes $\ell \times k$ and $\ell \times \ell$. $\boldsymbol{\eta}$ are $(k+l)$ column vectors of the state noise and the observation noise stacked vertically. We write $\Phi_j = \phi_j I$ then the state space model is of the same form as model 2.2. This completes the conversion. \square

In the state space model framework, the state equation is a multivariate AR(1) model, on the other hand, it can be expressed in a form of multivariate ARMA($m, m-1$) model. In the form of multivariate ARMA, the AR coefficients are all identity matrices multiplied by scalars. This form implies that the elements of the state vector are filtered by the same AR process to become different moving average of the system noise. We can see that the AR coefficients in this ARMA($m, m-1$) model is characterized solely by the transitional matrix F . In many real applications, AR coefficients play the main role of picking out spikes in a spectrum. In other words, F is the core of the model describing the dynamics of the state.

2.2 Transformation of State Space Model

Suppose \tilde{F} is similar to F , then there exists a non-singular matrix S , so that $\tilde{F} = SFS^{-1}$. The state equation and the observation equation can be transformed as follow.

$$\begin{aligned} S\mathbf{x}_t &= S(F\mathbf{x}_{t-1} + G\mathbf{w}_t) = SFS^{-1}S\mathbf{x}_{t-1} + SG\mathbf{w}_t \\ \tilde{\mathbf{x}}_t &= \tilde{F}\tilde{\mathbf{x}}_{t-1} + \tilde{G}\mathbf{w}_t \\ \\ \mathbf{y}_t &= H\mathbf{x}_t + \boldsymbol{\epsilon}_t = HS^{-1}S\mathbf{x}_t + \boldsymbol{\epsilon}_t \\ &= \tilde{H}\tilde{\mathbf{x}}_t + \boldsymbol{\epsilon}_t \end{aligned}$$

where

$$\tilde{F} = SFS^{-1}, \quad \tilde{G} = SG, \quad \tilde{H} = HS^{-1}, \quad \tilde{\mathbf{x}}_t = S\mathbf{x}_t.$$

The transformed model is also a state space model. The transformation does not affect the observation \mathbf{y} , the noises \mathbf{w} and $\boldsymbol{\epsilon}$. Therefore, the same state space model can be characterized by different combination of F , G and H . There is an invariant property of the state noise and observation noise once the state space model is identified. It is equivalent to say that the state space model can be rotated to allow a state vector and be viewed in any angle without affecting the transitional process.

Example 2.3 (Companion to Rotation) Suppose we have a state space model as in example 2.1, where

$$F = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Suppose F has a pair of complex conjugate roots. Then we can express the roots in $a + bi$ and $a - bi$, where a and b are positive real numbers. ϕ_1 and $-\phi_2$ represent the sum of roots and the product of roots, so $\phi_1 = 2a$ and $-\phi_2 = a^2 + b^2$. There exists a non-singular S and \tilde{F} such that $\tilde{F} = SFS^{-1}$, where

$$F = \begin{bmatrix} 2a & 1 \\ -(a^2 + b^2) & 0 \end{bmatrix}, \quad \tilde{F} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}, \quad S = \begin{bmatrix} \frac{a^2+b^2}{\sqrt{b(a^2+b^2)}} & \frac{a}{\sqrt{b(a^2+b^2)}} \\ 0 & \frac{b}{\sqrt{b(a^2+b^2)}} \end{bmatrix}.$$

\tilde{G} and \tilde{H} are given by

$$\begin{aligned} \tilde{G} &= SG \\ &= \begin{bmatrix} \frac{a^2+b^2}{\sqrt{b(a^2+b^2)}} & \frac{a}{\sqrt{b(a^2+b^2)}} \\ 0 & \frac{b}{\sqrt{b(a^2+b^2)}} \end{bmatrix} \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} \frac{a^2+b^2+\theta_1 a}{\sqrt{b(a^2+b^2)}} \\ \frac{\theta_1 b}{\sqrt{b(a^2+b^2)}} \end{bmatrix}, \\ \tilde{H} &= HS^{-1} \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{b}{\sqrt{b(a^2+b^2)}} & \frac{-a}{\sqrt{b(a^2+b^2)}} \\ 0 & \frac{a^2+b^2}{\sqrt{b(a^2+b^2)}} \end{bmatrix} = \begin{bmatrix} \frac{b}{\sqrt{b(a^2+b^2)}} & \frac{-a}{\sqrt{b(a^2+b^2)}} \end{bmatrix}. \end{aligned}$$

□

The transformation can be done only under the assumption that F and \tilde{F} are similar to each other, ie $\tilde{F} = SFS^{-1}$, for some non-singular S . If two matrices are similar, then they have the same set of eigenvalues, but the converse may not be true. It is possible that \tilde{F} has the same set of eigenvalues as F but it is not similar to F . Then there is no way to transform between the two models.

It is common that even F and \tilde{F} contain real entries, S can be complex. In such a case, the transformed counterparts $\tilde{\mathbf{x}}_t$, \tilde{G} and \tilde{H} become complex, which is not a desirable situation if we want to make use of the state estimates for any purposes.

However, we can take the real part of \tilde{G} and \tilde{H} as \tilde{G} and \tilde{H} to obtain a 'realistic' state space model. In real practice by doing this there is only a slight difference. The main part of a state space model is the transition matrix F or equivalently its eigenvalues λ .

If the characteristic equation has r conjugate pairs of complex roots, $\lambda_1, \bar{\lambda}_1, \lambda_2, \bar{\lambda}_2, \dots, \lambda_r, \bar{\lambda}_r$ and s real roots, $\lambda_{r+1}, \dots, \lambda_{r+s}$, a square matrix whose diagonal entries are λ is similar to F , ie $\text{diag}(\lambda_1, \bar{\lambda}_1, \dots, \lambda_{r+s}) = SFS^{-1}$. It is obvious that S is a composition of $r + s$ non-zero eigenvectors of F . This is known as spectral decomposition.

2.3 Compartment Model

We have discussed in chapter 1 about the literature of modelling EEG data, including the Broad Band Spectral Model by John et al. (1977), the $\xi\alpha$ model by Pascual-Marqui et al. (1988), and the Broad Band Spectral Model by Valdés-Sosa et al. (1990). The common point of these models is the analysis of spectral power in frequency domain. The major frequency bands are thought to characterize neural activity of a brain. For example, John (2002a) stated that during visual processing, gamma frequency range (25-50 Hz) evolves from visual cortex region; during multimodal semantic processing, lower beta frequency range (12-18 Hz) evolves from neighboring temporal region and parietal cortical region; and during tasks of mental imagery, alpha frequency range (8-12 Hz) and theta frequency range (4-8 Hz) evolves from fronto-parietal region.

Compartment model is proposed by Ozaki (2003) for characterizing dynamics of major frequency bands of EEG data and can be fitted in a state space framework. It is a modification of the parallel seasonal model Ozaki (1997a,b) which is based on the idea from Ameen & Harrison (1985) and Harvey (1985); at the same period different state space models with similar model structure were proposed, for example, cyclical component model by West (1995), individual AR component model by Kitagawa & Gersch (1996) and quasi-periodic oscillation model by Higuchi (1999). The compartment model assembles a weighted sum of state components to reconstruct the multivariate time series data, each observed variable of which shares common noise driven periodic dynamics of different absorption in terms of magnitude and phase lag.

2.3.1 Companion Form

For any state space model F is defined by its eigenvalues. The eigenvalues are either real numbers or pairs of complex conjugate numbers. Therefore, the characteristic polynomial of F can be factorized into first order polynomials for the real roots and second order polynomials for each conjugate pair of complex roots. Let λ be the eigenvalues of F .

$$\begin{aligned} \phi(\lambda) &= (\lambda - \lambda_1) (\lambda - \bar{\lambda}_1) (\lambda - \lambda_2) (\lambda - \bar{\lambda}_2) \dots (\lambda - \lambda_r) (\lambda - \bar{\lambda}_r) \\ &\quad \times (\lambda - \lambda_{r+1}) (\lambda - \lambda_{r+2}) \dots (\lambda - \lambda_{r+s}) \\ &= \left(\lambda^2 - \phi_1^{(1)} \lambda - \phi_2^{(1)} \right) \left(\lambda^2 - \phi_1^{(2)} \lambda - \phi_2^{(2)} \right) \dots \left(\lambda^2 - \phi_1^{(r)} \lambda - \phi_2^{(r)} \right) \\ &\quad \times \left(\lambda - \phi_1^{(r+1)} \right) \left(\lambda - \phi_1^{(r+2)} \right) \dots \left(\lambda - \phi_1^{(r+s)} \right) \end{aligned}$$

It is possible to fill in non-zero values into the off-diagonal blocks in \check{G} . This will allow the driving noises to be coupled before going into the states, which gives an interaction between state components. When \check{G} is a full matrix then the model is of no difference from a general state space model, which is out of scope of this thesis but will be saved as a future research. In this thesis we would concentrate on the class of independent compartment model.

When we have a univariate observation vector y then we choose $\check{H} = (\underbrace{1, 0, \dots, 1}_{2r \text{ elements}}, \underbrace{0, 1, \dots, 1}_{s \text{ elements}})$. This gives a meaning that we decompose the observation into ARMA(2) and AR(1) components described in the state equation, and we resemble the components to predict the data.

On the other hand, when we have multivariate data, the observation matrix \check{H} contains different weights of states for each variable. The first row of \check{H} is fixed to $(1, 0, \dots, 1, 0, 1, \dots, 1)$ so that the first element of the observation vector is a simple of states which is not necessary in general, but here we have this setting in order to avoid parameter redundancy.

$$\check{H} = \begin{bmatrix} 1 & 0 & \dots & 1 & 0 & 1 & \dots & 1 \\ h^{(21)} & h^{(22)} & \dots & h^{(2, 2r-1)} & h^{(2, 2r)} & h^{(2, 2r+1)} & \dots & h^{(2, 2r+s)} \\ h^{(31)} & h^{(32)} & \dots & h^{(3, 2r-1)} & h^{(3, 2r)} & h^{(3, 2r+1)} & \dots & h^{(3, 2r+s)} \\ \vdots & \vdots \\ h^{(\ell 1)} & h^{(\ell 2)} & \dots & h^{(\ell, 2r-1)} & h^{(\ell, 2r)} & h^{(\ell, 2r+1)} & \dots & h^{(\ell, 2r+s)} \end{bmatrix}$$

The compartment model allows every single time series to pick up a weighted combination of components of its own special. The pairs $(h^{(21)}, h^{(22)})$, for instance, acts an interpolation coefficients of first two state elements. This is an approximation to remedy the phase lag between time series. Therefore, it is good to have time series in phase or constant phase lag.

Moreover, those decomposed components must be mutual components to all variables. If two observed variables have same characteristics at a frequency but driven by two different noise, the model would be unsuitable unless we add another component.

Example 2.4 *EEG data contains some high frequency, such as delta, theta, alpha, beta and gamma frequency. They appear significantly in the sample spectrum and brain scientists believe they are source to the brain's dynamics. In this case, 4 compartments are used for theta, alpha, beta and gamma frequency, and the remaining root is for the delta frequency.*

Figure 2.1 shows how the decomposition was resulted. The data, on the top left hand corner, is decomposed into 4 components, illustrated with four colors in the bottom left hand corner, of which each is characterized by a pair of conjugate complex roots in the Argand plane in the bottom right hand corner. Each decomposed component contributes to a peak in the power spectrum. The four spectra constitutes a model spectrum for the data. \square

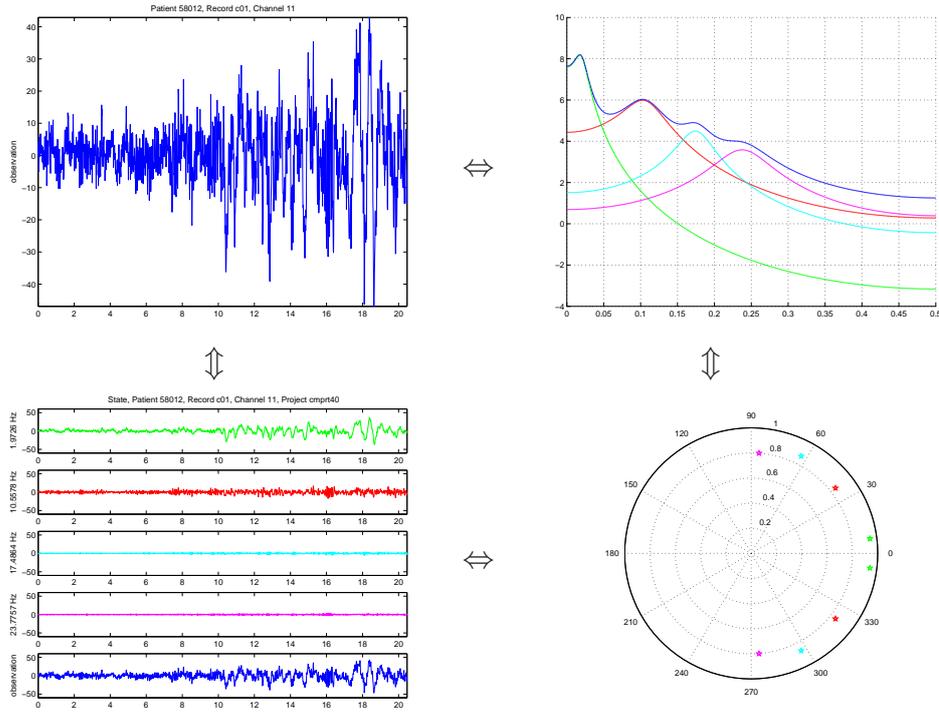


Figure 2.1: Top left: data; bottom left: decomposed components with data; top right: spectrum of individual decomposed components and their sum; bottom right: characteristic roots of decomposed components.

2.3.2 Rotation Form

In section 2.2 we studied the invariant property of linear state space models. Any linear state space model projecting onto another space does not affect its statistical properties. We can transform a state space model in order to view the state in another angle.

Following to the previous subsection let λ be the eigenvalues of a transition matrix F . Let the complex $\lambda_j = a_j + b_j i$ for $j = 1, \dots, r$ and the real $\lambda_j = a_j$ for $j = r + 1, \dots, r + s$. Without any loss of generality we can assume b_j are the non-zero positive real numbers for $j = 1, \dots, r$ and zero for $j = r + 1, \dots, r + s$, so as to ensure the real eigenvalues and the complex eigenvalues are well separated. We can have the following matrix \hat{F} .

\hat{F} is a diagonal block matrix of r 2×2 rotation matrices for r pairs of conjugate complex eigenvalues and s 1×1 matrices of the real eigenvalues.

In example 2.3, we showed that the transition matrix in companion form \tilde{F} and in rotation form \hat{F} are similar to each other. It is true in general when we have the following transformation matrix.

2.5 Appendix

2.5.1 Transformation Matrices

We list out the transformation matrices in table 2.1.

F	S	$\tilde{F} = SFS^{-1}$
$\begin{bmatrix} 2a & 1 \\ -(a^2 + b^2) & 0 \end{bmatrix}$	$\begin{bmatrix} \frac{a^2+b^2}{\sqrt{b(a^2+b^2)}} & \frac{a}{\sqrt{b(a^2+b^2)}} \\ 0 & \frac{b}{\sqrt{b(a^2+b^2)}} \end{bmatrix}$	$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$
$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$	$\begin{bmatrix} \frac{b}{\sqrt{b(a^2+b^2)}} & \frac{-a}{\sqrt{b(a^2+b^2)}} \\ 0 & \frac{a^2+b^2}{\sqrt{b(a^2+b^2)}} \end{bmatrix}$	$\begin{bmatrix} 2a & 1 \\ -(a^2 + b^2) & 0 \end{bmatrix}$
$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$	$\begin{bmatrix} \frac{a^2+b^2}{\sqrt{-b(a^2+b^2)}} & 0 \\ \frac{a}{\sqrt{-b(a^2+b^2)}} & \frac{-b}{\sqrt{-b(a^2+b^2)}} \end{bmatrix}$	$\begin{bmatrix} 2a & -(a^2 + b^2) \\ 1 & 0 \end{bmatrix}$
$\begin{bmatrix} 2a & -(a^2 + b^2) \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} \frac{-b}{\sqrt{-b(a^2+b^2)}} & 0 \\ \frac{-a}{\sqrt{-b(a^2+b^2)}} & \frac{a^2+b^2}{\sqrt{-b(a^2+b^2)}} \end{bmatrix}$	$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$
$\begin{bmatrix} a + bi & 0 \\ 0 & a - bi \end{bmatrix}$	$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$	$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$

Table 2.1: A list of transformation matrices

2.5.2 AR Compartment Model

In this subsection we give another form of compartment model used in Wong et al. (2006). This compartment model is a special case of the companion form of compartment model in section 2.3.1 when the ARMA(2,1) compartments become AR(2) compartments, equivalently when θ are all zero. Since this special case restricts the MA part of compartments that would reduce the flexibility of model fitting, we leave this model in this appendix.

$$\begin{aligned}
\phi(B) &= (1 - \lambda_1 B)(1 - \bar{\lambda}_1 B)(1 - \lambda_2 B)(1 - \bar{\lambda}_2 B) \dots (1 - \lambda_r B)(1 - \bar{\lambda}_r B) \\
&\quad \times (1 - \lambda_{r+1} B)(1 - \lambda_{r+2} B) \dots (1 - \lambda_{r+s} B) \\
&= \left(1 - \phi_1^{(1)} B - \phi_2^{(1)} B^2\right) \left(1 - \phi_1^{(2)} B - \phi_2^{(2)} B^2\right) \dots \left(1 - \phi_1^{(r)} B - \phi_2^{(r)} B^2\right) \\
&\quad \times \left(1 - \phi_1^{(r+1)} B\right) \left(1 - \phi_1^{(r+2)} B\right) \dots \left(1 - \phi_1^{(r+s)} B\right)
\end{aligned}$$

Chapter 3

Spectrum and Causality in State Space

In view of the wide use of state space model in this thesis, discussing the theory in chapter 2 and showing the result in the latter chapters, we shall discuss the spectral properties of the linear state space model. Through spectral density we can discuss causality by the theory of NCR proposed by Akaike (1968). NCR is commonly applied on multivariate AR model while it will be applied to a more general multivariate ARMA model or a linear state space model. At the end of the chapter we will briefly discuss the time varying causality which leads us to the topic of heteroscedasticity in chapter 4.

3.1 Spectrum

Spectrum is power intensity against frequency. It is a smooth function showing the characteristic of a time process. It is also a polygon curve, approximating the true spectrum of a time series data, which can be produced by Fourier transformation of auto-covariances or fast Fourier transformation of the data.

There are several ways to smooth the sample spectrum from a zigzag polygon curve. Examples include averaging spectrum windows (Brinllinger, 1981; Priestley, 1981; Harvey, 1989) and approximating spectral density by a mixture of quadratic functions (Pascual-Marqui et al., 1988). Instead of applying any smoothing technique directly to spectrum in the frequency domain, we can fit an AR model to the data and plot the AR model spectrum.

It is also fast and easy to estimate AR model first from zero order to a reasonably high order, choose one of them with minimum AIC over the others and then plot out the spectrum, which will be called best AR spectrum in the rest of this thesis. It is also wise to take windows of data and draw the spectral density of each window on a time axis. It helps to see how the spectrum shape changing in the progression of time. We will use the symbol P_f to denote the

spectral density at frequency f . We shall plot the log spectrum in decibel scale, ie $10 \times \log_{10} P_f$, for easy visualization.

3.1.1 Multivariate ARMA Spectral Density

Let f be the frequency ranged from 0 to 0.5, i be the unit imaginary number and π be the ratio of circumference to diameter of a circle. The spectral density of a univariate ARMA model 2.1 is formulated in equation 3.1.

$$P_f = \frac{\sigma^2 |\theta_0 + \theta_1 e^{-2i\pi f} + \dots + \theta_q e^{-2qi\pi f}|^2}{2\pi |1 - \phi_1 e^{-2i\pi f} - \dots - \phi_p e^{-2pi\pi f}|^2} \quad (3.1)$$

The spectrum P_f is an averaged variance multiplied by a coefficient depending on frequency f . This coefficient is obtained by Fourier transformation of the ARMA coefficients. The Fourier transformation expands the ARMA coefficients to a sum of sine functions and cosine functions. By putting different value between 0 to 0.5 into f , we get the amplitude of each ARMA coefficient of f . The amplitude constitutes the power contributing to the observed variable.

The spectral density of a multivariate ARMA model can be calculated analogously by using a matrix manner. First we need to calculate the Fourier transformation of the AR and MA coefficients in model 2.2 on page 8 by equation 3.2.

$$\begin{aligned} A_f &= \mathcal{F}_f(\Phi)^{-1} \mathcal{F}_f(\Theta) \\ &= \left[I - \sum_{j=1}^p \Phi_j e^{-2ji\pi f} \right]^{-1} \left[\sum_{j=0}^q \Theta_j e^{-2ji\pi f} \right] \\ &= \begin{pmatrix} a_f^{(11)} & a_f^{(12)} & \dots & a_f^{(1n)} \\ a_f^{(21)} & a_f^{(22)} & \dots & a_f^{(2n)} \\ \vdots & \vdots & \ddots & \vdots \\ a_f^{(\ell 1)} & a_f^{(\ell 2)} & \dots & a_f^{(\ell n)} \end{pmatrix} \end{aligned} \quad (3.2)$$

The power spectral density matrix P_f is given by equation 3.3.

$$\begin{aligned} P_f &= \frac{1}{2\pi} A_f \Sigma A_f^H \\ &= \frac{1}{2\pi} [\mathcal{F}_f(\Phi)^{-1} \mathcal{F}_f(\Theta)] \Sigma [\mathcal{F}_f(\Phi)^{-1} \mathcal{F}_f(\Theta)]^H \end{aligned} \quad (3.3)$$

Example 3.1 Suppose we have a multivariate AR(p) model as equation 3.4. The data \mathbf{y} and the noise $\boldsymbol{\eta}$ are 3-variate vector.

$$\begin{aligned} \mathbf{y}_t - \Phi_1 \mathbf{y}_{t-1} - \Phi_2 \mathbf{y}_{t-2} - \dots - \Phi_p \mathbf{y}_{t-p} &= \boldsymbol{\eta}_t \\ \boldsymbol{\eta}_t &\sim N(0, \Sigma) \end{aligned} \quad (3.4)$$

$$\mathbf{y}_t = \begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \end{pmatrix} \quad \boldsymbol{\eta}_t = \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \\ \eta_t^{(3)} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma^{(1)^2} & \sigma^{(12)} & \sigma^{(13)} \\ \sigma^{(12)} & \sigma^{(2)^2} & \sigma^{(23)} \\ \sigma^{(13)} & \sigma^{(23)} & \sigma^{(3)^2} \end{pmatrix}$$

A_f can be obtained by equation 3.2 with $\Theta_0 = I$. Then we can calculate the Spectral density function P_f .

$$\begin{aligned} P_f &= \frac{1}{2\pi} A_f \Sigma A_f^H \\ &= \frac{1}{2\pi} \begin{pmatrix} a_f^{(11)} & a_f^{(12)} & a_f^{(13)} \\ a_f^{(21)} & a_f^{(22)} & a_f^{(23)} \\ a_f^{(31)} & a_f^{(32)} & a_f^{(33)} \end{pmatrix} \begin{pmatrix} \sigma^{(1)^2} & \sigma^{(12)} & \sigma^{(13)} \\ \sigma^{(12)} & \sigma^{(2)^2} & \sigma^{(23)} \\ \sigma^{(13)} & \sigma^{(23)} & \sigma^{(3)^2} \end{pmatrix} \begin{pmatrix} \bar{a}_f^{(11)} & \bar{a}_f^{(21)} & \bar{a}_f^{(31)} \\ \bar{a}_f^{(12)} & \bar{a}_f^{(22)} & \bar{a}_f^{(32)} \\ \bar{a}_f^{(13)} & \bar{a}_f^{(23)} & \bar{a}_f^{(33)} \end{pmatrix} \\ &= \begin{pmatrix} P_f^{(11)} & P_f^{(12)} & P_f^{(13)} \\ P_f^{(21)} & P_f^{(22)} & P_f^{(23)} \\ P_f^{(31)} & P_f^{(32)} & P_f^{(33)} \end{pmatrix} \end{aligned}$$

The first element is

$$\begin{aligned} P_f^{(11)} &= \frac{1}{2\pi} \left[\left(a_f^{(11)} \sigma^{(1)^2} + a_f^{(12)} \sigma^{(12)} + a_f^{(13)} \sigma^{(13)} \right) \bar{a}_f^{(11)} \right. \\ &\quad + \left(a_f^{(11)} \sigma^{(12)} + a_f^{(12)} \sigma^{(2)^2} + a_f^{(13)} \sigma^{(23)} \right) \bar{a}_f^{(12)} \\ &\quad \left. + \left(a_f^{(11)} \sigma^{(13)} + a_f^{(12)} \sigma^{(23)} + a_f^{(13)} \sigma^{(3)^2} \right) \bar{a}_f^{(13)} \right], \end{aligned}$$

and in general for all elements, $i = 1, 2, 3; j = 1, 2, 3$,

$$\begin{aligned} P_f^{(ij)} &= \frac{1}{2\pi} \left[\left(a_f^{(i1)} \sigma^{(1)^2} + a_f^{(i2)} \sigma^{(12)} + a_f^{(i3)} \sigma^{(13)} \right) \bar{a}_f^{(j1)} \right. \\ &\quad + \left(a_f^{(i1)} \sigma^{(12)} + a_f^{(i2)} \sigma^{(2)^2} + a_f^{(i3)} \sigma^{(23)} \right) \bar{a}_f^{(j2)} \\ &\quad \left. + \left(a_f^{(i1)} \sigma^{(13)} + a_f^{(i2)} \sigma^{(23)} + a_f^{(i3)} \sigma^{(3)^2} \right) \bar{a}_f^{(j3)} \right]. \end{aligned}$$

$P_f^{(11)}$ is the spectral density function of $y_t^{(1)}$ based on the AR(3) model 3.4. With this function we can draw the model spectrum, a plot of power intensity against frequency. Similarly other diagonal elements of P_f give the spectral density of other variables. \square

3.1.2 State Space Spectral Density

The state space model contains a system noise and an observation noise so its model structure is more complicated than a general ARMA model. It is possible to draw the spectrum of a state space model but not in a straightforward way. To calculate the spectrum, one should first transform the state space

model to an ARMA model representation to retrieve the model coefficients and the noise covariance from the state space model as shown in example 2.2. Then we can put these Φ , Θ and Σ into equation 3.3 to obtain the spectrum for the general state space model 2.3.

While we have a linear state space model, we can easily calculate the AR and MA coefficients and then substitute into these calculation to obtain the spectrum. In example 3.2 we illustrate such calculation with a simple case of compartment model.

Example 3.2 *Suppose we have a compartment model in companion form (section 2.3.1) of $r = 1$ and $s = 1$. Let \mathbf{y} be a bivariate data. We assume both the system noise variance Q and the observation noise variance R to be diagonal.*

$$\begin{cases} \mathbf{x}_t = F\mathbf{x}_{t-1} + G\mathbf{w}_t \\ \mathbf{y}_t = H\mathbf{x}_t + \boldsymbol{\epsilon}_t \end{cases}$$

$$\mathbf{w}_t \sim N(0, Q_t), \quad \boldsymbol{\epsilon}_t \sim N(0, R)$$

$$F = \begin{pmatrix} \phi_1^{(1)} & 1 & \\ \phi_2^{(1)} & 0 & \\ & & \phi_1^{(2)} \end{pmatrix}, \quad G = \begin{pmatrix} 1 & \\ \theta_1^{(1)} & \\ & & 1 \end{pmatrix},$$

$$H = \begin{pmatrix} 1 & 0 & 1 \\ h^{(21)} & h^{(22)} & h^{(23)} \end{pmatrix}, \quad Q = \begin{pmatrix} \tau^{(1)2} & 0 \\ 0 & \tau^{(2)2} \end{pmatrix}, \quad R = \begin{pmatrix} \sigma^{(1)2} & 0 \\ 0 & \sigma^{(2)2} \end{pmatrix}.$$

The characteristic polynomial of F is the product of an AR(2) factor and an AR(1) factor. We can expand the product of the factors to get a characteristic polynomial of F .

$$\begin{aligned} \phi(\lambda) &= (\lambda^2 - \phi_1^{(1)}\lambda - \phi_2^{(1)}) (\lambda - \phi_1^{(2)}) \\ &= \lambda^3 - (\phi_1^{(1)} + \phi_1^{(2)})\lambda - (-\phi_1^{(1)}\phi_1^{(2)} + \phi_2^{(1)})\lambda - (-\phi_1^{(2)}\phi_2^{(1)}) \\ &\equiv \lambda^3 - \phi_1\lambda^2 - \phi_2\lambda - \phi_3 \end{aligned}$$

Next we can use the formula in example 2.2 to obtain the ARMA coefficients.

$$\begin{aligned} \Phi_1 &= \phi_1 I = \begin{pmatrix} \phi_1^{(1)} + \phi_1^{(2)} & 0 \\ 0 & \phi_1^{(1)} + \phi_1^{(2)} \end{pmatrix} \\ \Phi_2 &= \phi_2 I = \begin{pmatrix} -\phi_1^{(1)}\phi_1^{(2)} + \phi_2^{(1)} & 0 \\ 0 & -\phi_1^{(1)}\phi_1^{(2)} + \phi_2^{(1)} \end{pmatrix} \\ \Phi_3 &= \phi_3 I = \begin{pmatrix} -\phi_1^{(2)}\phi_2^{(1)} & 0 \\ 0 & -\phi_1^{(2)}\phi_2^{(1)} \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
\Theta_0 &= \left(HG \mid I \right) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ h^{(21)} + h^{(22)}\theta_1^{(1)} & h^{(23)} & 0 & 1 \end{pmatrix} \\
\Theta_1 &= \left(H(F - \phi_1 I)G \mid -\phi_1 I \right) \\
&= \begin{pmatrix} -\phi_1^{(2)} + \theta_1^{(1)} & -\phi_1^{(1)} & -\phi_1^{(1)} - \phi_1^{(2)} & 0 \\ \left\{ \begin{array}{l} h^{(21)}(-\phi_1^{(2)} + \theta_1^{(1)}) + \\ h^{(22)}[\phi_2^{(1)} - (\phi_1^{(1)} + \phi_1^{(2)})\theta_1^{(1)}] \end{array} \right\} & -h^{(23)}\phi_1^{(1)} & 0 & -\phi_1^{(1)} - \phi_1^{(2)} \end{pmatrix} \\
\Theta_2 &= \left(H(F^2 - \phi_1 F - \phi_2 I)G \mid -\phi_2 I \right) \\
&= \begin{pmatrix} -\phi_1^{(2)}\theta_1^{(1)} & -\phi_2^{(1)} & \phi_1^{(1)}\phi_1^{(2)} - \phi_2^{(1)} & 0 \\ \left\{ \begin{array}{l} -h^{(21)}\phi_1^{(2)}\theta_1^{(1)} - \\ h^{(22)}(\phi_1^{(2)}\phi_2^{(1)} - \phi_1^{(1)}\phi_1^{(2)}\theta_1^{(1)}) \end{array} \right\} & -h^{(23)}\phi_2^{(1)} & 0 & \phi_1^{(1)}\phi_1^{(2)} - \phi_2^{(1)} \end{pmatrix} \\
\Theta_3 &= \left(0 \mid -\phi_3 I \right) = \begin{pmatrix} 0 & 0 & \phi_1^{(2)}\phi_2^{(1)} & 0 \\ 0 & 0 & 0 & \phi_1^{(2)}\phi_2^{(1)} \end{pmatrix} \\
\Sigma &= \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix} = \begin{pmatrix} \tau^{(1)2} & 0 & & \\ 0 & \tau^{(2)2} & & \\ & & \sigma^{(1)2} & 0 \\ & & 0 & \sigma^{(2)2} \end{pmatrix}
\end{aligned}$$

We can write the explicit expression only up to this step. Next step we put Φ , Θ and Σ into equation 3.2 and equation 3.3 to get P_f .

$$\begin{aligned}
A_f &= \mathcal{F}_f(\Phi)^{-1} \mathcal{F}_f(\Theta) = \begin{pmatrix} a_f^{(11)} & a_f^{(12)} & a_f^{(13)} & 0 \\ a_f^{(21)} & a_f^{(22)} & 0 & a_f^{(24)} \end{pmatrix} \\
P_f &= \frac{1}{2\pi} A_f \Sigma A_f^H = \begin{pmatrix} P_f^{(11)} & P_f^{(12)} \\ P_f^{(21)} & P_f^{(22)} \end{pmatrix}
\end{aligned}$$

Then the power spectral density of $y_t^{(1)}$ and $y_t^{(2)}$ are

$$\begin{aligned}
P_f^{(11)} &= \frac{1}{2\pi} \left(|a_f^{(11)}|^2 \tau^{(1)2} + |a_f^{(12)}|^2 \tau^{(2)2} + |a_f^{(13)}|^2 \sigma^{(1)2} \right) \\
P_f^{(22)} &= \frac{1}{2\pi} \left(|a_f^{(21)}|^2 \tau^{(1)2} + |a_f^{(22)}|^2 \tau^{(2)2} + |a_f^{(24)}|^2 \sigma^{(2)2} \right)
\end{aligned}$$

and the cross spectrum is

$$P_f^{(12)} = \bar{P}_f^{(21)} = \frac{1}{2\pi} \left(a_f^{(11)} \bar{a}_f^{(21)} \tau^{(1)2} + a_f^{(12)} \bar{a}_f^{(22)} \tau^{(2)2} \right)$$

In general for all elements, $i = 1, 2; j = 1, 2$,

$$P_f^{(ij)} = \frac{1}{2\pi} \left(a_f^{(i1)} \bar{a}_f^{(j1)} \tau^{(1)2} + a_f^{(i2)} \bar{a}_f^{(j2)} \tau^{(2)2} + a_f^{(i3)} \bar{a}_f^{(j3)} \sigma^{(1)2} + a_f^{(i4)} \bar{a}_f^{(j4)} \sigma^{(2)2} \right)$$

where $a_f^{(14)} = a_f^{(23)} = 0$. □

3.2 Noise Contribution Ratio Causality

Causality is a vague term since it has a slightly different meaning in different literatures. Intuitively, the study of causality is to relate the cause and the result, and is to conclude a “the cause precedes the result” phenomenon. A broad meaning of causality includes the cause of a random noise to an observation (Brockwell & Davis, 1996), in other words, a linear AR model is an example of causality. A narrow meaning of causality is the pairwise relationship between time series. One way to explain causality is to look at the causal relation from one time series to another (Granger, 1969). Another way, that exhibits the marrow of statistical modelling, ie transforming data to white noise, is to look at the causal relation from noise to observed data (Akaike, 1968).

The literature of causality can be traced back to Wiener (1956), in a part of his book chapter he discussed the causality based on prediction in the language of Lebesgue measure and integration theory. Granger (1963) and Granger (1969) proposed a definition of causality based on multivariate time series. This evolves the popularity of Granger’s causality which bases on the predictability of some series for the others, that the coefficients of a multivariate time series model is a signal of causality was interpreted. The weakness of Granger’s approach is the negligence of an important element in statistical time series model: the noise variance. Akaike (1968) proposed his own measure in view of causality, using the NCR theory. Spectrum based on multivariate AR model was considered. A spectrum, when the noise are uncorrelated, becomes a simple weighted sum of the noise variance. Other than the work of Akaike and Granger, different measures and tests of causality were proposed (Geweke, 1982, 1984; Hosoya, 1991; Kaminski et al., 2001).

Akaike (1968) began the concept of NCR in order to interpret the causality of multivariate AR models. Calculation of NCR is easy once we have the multivariate spectral density function. NCR is a measure of the proportion of power dedicating to a spectral density from independent noise variance. In explicit words, when the noise variance covariance matrix Σ is and only is diagonal, the spectral density function of each variable is a weighted sum of noise variance. The ratio of part of the weighted sum to the total weighted sum is NCR.

$$\text{NCR}(\sigma^2, y_t) = \frac{\text{spectral power intensity of } y_t \text{ contributed by } \sigma^2}{\text{total spectral power intensity of } y_t}$$

Example 3.3 *We will continue from example 3.1, except that we let Σ be diagonal, ie $\sigma^{(12)} = \sigma^{(13)} = \sigma^{(23)} = 0$. The power spectral density of $y_t^{(1)}$ is*

given by

$$\begin{aligned} P_f^{(11)} &= \frac{1}{2\pi} \left[\left(a_f^{(11)} \sigma^{(1)^2} \right) \bar{a}_f^{(11)} + \left(a_f^{(12)} \sigma^{(2)^2} \right) \bar{a}_f^{(12)} + \left(a_f^{(13)} \sigma^{(3)^2} \right) \bar{a}_f^{(13)} \right] \\ &= \frac{1}{2\pi} \left[\left| a_f^{(11)} \right|^2 \sigma^{(1)^2} + \left| a_f^{(12)} \right|^2 \sigma^{(2)^2} + \left| a_f^{(13)} \right|^2 \sigma^{(3)^2} \right], \end{aligned}$$

and therefore, for instance, the noise contribution from $\sigma^{(3)^2}$ to $y_t^{(1)}$ is

$$\text{NCR} \left(\sigma^{(3)^2}, y_t^{(1)} \right) = \frac{\frac{1}{2\pi} \left| a_f^{(13)} \right|^2 \sigma^{(3)^2}}{P_f^{(11)}}.$$

In general,

$$\text{NCR} \left(\sigma^{(j)^2}, y_t^{(i)} \right) = \frac{\frac{1}{2\pi} \left| a_f^{(ij)} \right|^2 \sigma^{(j)^2}}{P_f^{(ii)}}.$$

□

Only when the noise variance covariance matrix Σ being diagonal, each element of the noise vector $\boldsymbol{\eta}_t$ is the only residual of the corresponding element in \mathbf{y}_t , ie each observed variable can drain its unexplained residual into its corresponding noise but not any other one. It makes the spectral density of each variables be a weighted sum of the diagonal entries of the noise variance matrix.

For a graphical representation, the cumulative power spectrum and the cumulative NCR will be used.

Example 3.4 Suppose we have the following multivariate AR(2) model.

$$\begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \end{pmatrix} = \begin{pmatrix} 1.6 & 0.2 & 0.3 \\ 0.4 & 0.9 & 0.1 \\ 0.3 & 0.1 & 1.2 \end{pmatrix} \begin{pmatrix} y_{t-1}^{(1)} \\ y_{t-1}^{(2)} \\ y_{t-1}^{(3)} \end{pmatrix} + \begin{pmatrix} -0.81 & 0.10 & 0.42 \\ -0.10 & -0.81 & 0.24 \\ 0.22 & 0.38 & -0.96 \end{pmatrix} \begin{pmatrix} y_{t-2}^{(1)} \\ y_{t-2}^{(2)} \\ y_{t-2}^{(3)} \end{pmatrix} + \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \\ \eta_t^{(3)} \end{pmatrix}$$

$$\begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \\ \eta_t^{(3)} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$$

In figure 3.1 we show the model spectrum of three observed variables on the left hand side and their corresponding NCR on the right hand side. The horizontal axes are all in frequency, ranging from 0 to the Nyquist frequency, ie half of the sampling rate.

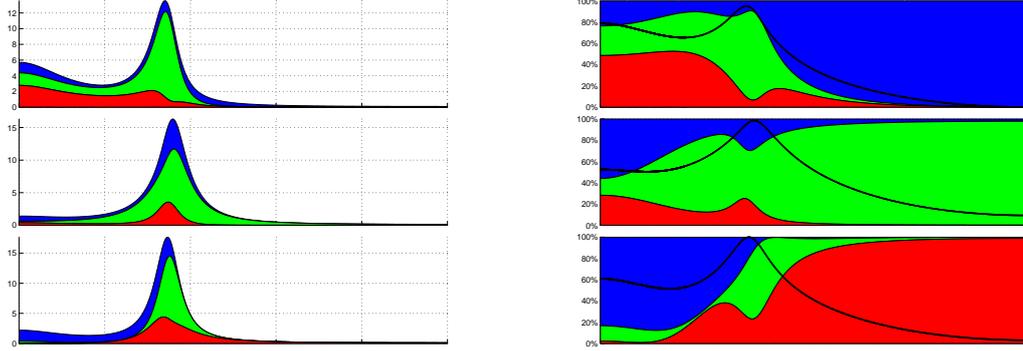


Figure 3.1: Model spectrum and NCR

The three plots on the left hand side show the power intensity of each of the three variables. The color under them is the proportion of contribution of the noise variances. In this example, the blue, green and red colors respectively stand for $\eta_t^{(1)}$, $\eta_t^{(2)}$ and $\eta_t^{(3)}$. Therefore, $\eta_t^{(2)}$, represented by green color, contributes the most to all three observed variables.

The three plots on the right hand side are the corresponding NCR. They are obtained as if each value on the spectral density curve is stretched or squeezed to fit the rectangular box of 100%. In each 100% box we still have 3 colors, and a black curve crossing on the colors of each box is the log spectrum. The NCR helps to see the causality among variables at various frequencies, especially around the low frequency range and the high frequency range as their power is less intense than the range in between.

One important note is that we should pay more attention to the main characteristics of the model. In other words, we should look at the colors around the region having relatively higher spectral values. \square

The causality on a frequency domain can be drawn by distinguishing the contribution ratio of the noises of which each is uniquely revealing the dynamics of its corresponding observed variable. In this situation the interpretation of causality is different from Granger's causality because we use both the AR coefficients and noise variance of the model for computing the spectral power density.

Moreover, non-zero AR coefficients do not directly imply an absolute causal effect between variables. It is better to look at the coefficients jointly instead of individually. An example on this issue will be given in chapter 7.

As we know that the NCR is a weighted sum of noise variances, we can tell how important role do the size of the variances play. However it is the same argument to say that causality should be described by not only the noise variance but the product with Fourier transformation of the coefficients.

A deficiency of NCR for causality analysis is the assumption of diagonal structure in the noise covariance matrix. If the off-diagonal entries in the noise

covariance is non-zero, ie the noise is correlated, then the power spectrum becomes a weighted sum of individual noise variances and also covariances.

Although a model with a diagonal noise variance matrix is often worse, in terms of likelihood, than a model without restriction, in practice the two models have slightly changes in the coefficients, and a same shape of spectral density. Therefore, it is not a problem when we use the former model to describe the causality.

3.2.1 Multivariate ARMA Model NCR Causality

Until presently, NCR has not been applied to any models except multivariate AR models. Difficulty in identifying a multivariate ARMA model would be the most probable reason. Our generalization would not be a big surprise to those who know the multivariate AR models very well. But in essence, we have an innovative assumption for the multivariate ARMA models that the dimension of the observed data and the dimension of driving noise are unequal. This assumption is essential when we carry on to the next section, in explaining the NCR causality of a linear state space model.

We preserve Akaike (1968)'s approach to keep the noise variance matrix diagonal. By equation 3.2 we get matrix A_f . In general, if we have a ℓ -variate AR(p) model,

$$\text{NCR} \left(\sigma^{(j)2}, y_t^{(i)} \right) = \frac{\frac{1}{2\pi} \left| a_f^{(ij)} \right|^2 \sigma^{(j)2}}{P_f^{(ii)}} \quad (3.5)$$

for $i = 1, 2, \dots, \ell$ and $j = 1, 2, \dots, \ell$.

3.2.2 State Space Model NCR Causality

We have studied in section 2.1.4, that any state space model can be written as a general multivariate ARMA model. So we can also interpret the contribution of every individual noise to each observed variable in a state space model. Calculation of NCR under state space model is easy, if we can assume all the noises are independent to each other.

The example below shows the formula of the NCR for a simple case of compartment model.

Example 3.5 *Continuing from example 3.2, the power spectral density of $y_t^{(1)}$,*

$P_f^{(11)}$, is a sum of $\frac{1}{2\pi} |a_f^{(11)}|^2 \tau^{(1)2}$, $\frac{1}{2\pi} |a_f^{(12)}|^2 \tau^{(2)2}$ and $\frac{1}{2\pi} |a_f^{(13)}|^2 \sigma^{(1)2}$. We have

$$\begin{aligned} \text{NCR} \left(\tau^{(1)2}, y_t^{(1)} \right) &= \frac{\frac{1}{2\pi} |a_f^{(11)}|^2 \tau^{(1)2}}{P_f^{(11)}}, \quad \text{NCR} \left(\tau^{(2)2}, y_t^{(1)} \right) = \frac{\frac{1}{2\pi} |a_f^{(12)}|^2 \tau^{(2)2}}{P_f^{(11)}}, \\ \text{NCR} \left(\sigma^{(1)2}, y_t^{(1)} \right) &= \frac{\frac{1}{2\pi} |a_f^{(13)}|^2 \sigma^{(1)2}}{P_f^{(11)}}, \quad \text{NCR} \left(\sigma^{(2)2}, y_t^{(1)} \right) = 0. \end{aligned}$$

□

The state space NCR causality will be applied to the finger rotation data in chapter 8. A hidden variable driven by an independent noise acts as a common component on two observed time series. The common

Tanokura & Kitagawa (2003) explained the contribution ratio from the covariances, and called this an extended power contribution. They claimed that this extended power contribution is useful for any multivariate AR model, if the row sum of absolute correlation coefficients is less than 2. This brings the queries of the generality of the method, especially when the multiple correlation among 3 or more variables is inclusive in some pairwise correlation. We shall leave the comparison of Tanokura & Kitagawa (2003) approach and our approach as future work.

3.2.3 Time varying NCR Causality

In a recent paper of Yamashita et al. (2005), NCR was applied to a multivariate AR model in which a delay parameter switches the AR coefficients. The delay parameter comes from a known stimulation in BOLD signals data, so that connectivity of two visual cortex region against time is depicted. Their threshold type AR model, as well as any other time (coefficient) varying AR models, modifies the A_f matrix in equation 3.2.

In the case of heteroscedasticity the noise variance changes in time, causes the model spectrum varying over time. At the same time, the NCR will follow the changes of the spectrum and therefore it varies over time also. The spectrum of a linear time series model is defined by the time varying spectrum can be easily extended by either making the ARMA coefficients changing over time, or making the driving noise variance changing over time.

Therefore, to obtain a time vary spectrum and a time varying NCR is equivalent to allow the two components, A_f and Σ , of the spectrum vary over time.

Example 3.6 *In figure 3.2 we show the result of the compartment-GARCH model on this bivariate data. Although it is more suitable to have a higher*

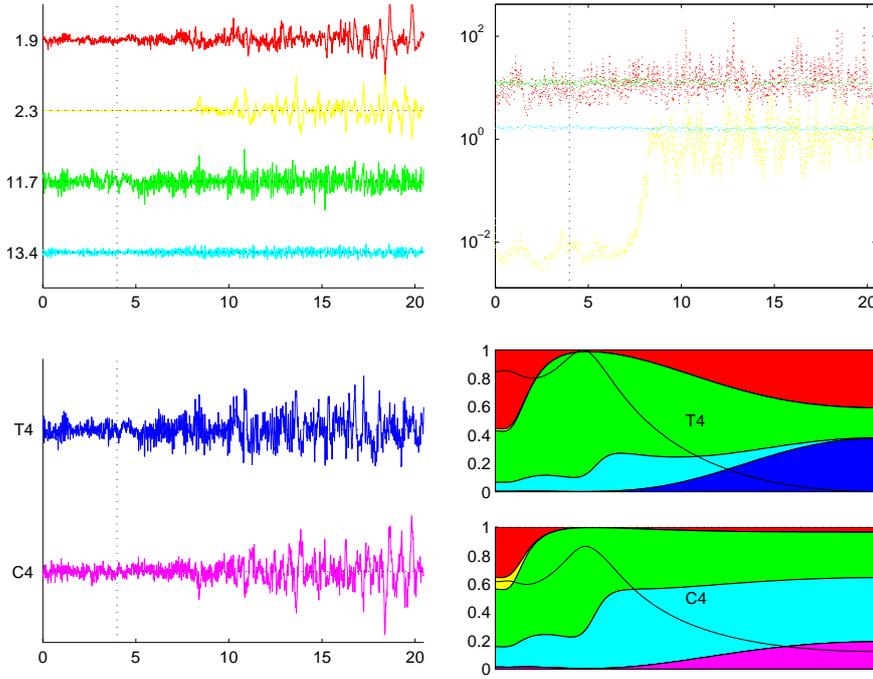


Figure 3.2: Estimated states, state variance, original data and NCR of the bivariate EEG data at 4 s

order of compartment model, we choose to keep a 4 compartment state space structure. We see that the compartment-GARCH model identifies 4 frequency at 1.9 Hz, 2.3 Hz, 11.7 Hz and 13.4 Hz and their estimated decomposition are on the top left corner of the figure. On the top right corner of the figure we see the variance corresponding to 2.3 Hz shows a significant rise but that of another low frequency 1.9 Hz shows a relative dimmer upward change.

To explain the NCR result, in addition to figure 3.2, we show the NCR of compartment-GARCH model at 4 different time points in figure 3.3. At 4 s, T4 electrode is mainly explained by the driving noise from 1.9 Hz and 11.7 Hz components because the corresponding colors in the rectangular box have a greater proportion of area. The C4 electrode is mainly explained by the driving noise from 11.7 Hz and 13.4 Hz components when the low frequency activity is not so active. At 8 s the 2.3 Hz starts showing activity in C4, and increases gradually until the end of the data. The blue and purple colors correspond to the observation noise of the 2 channels, when they were only contributing the less important high frequency activities in the data. \square

In this thesis we apply the same technique to the two physical data. In our approach we use the technique of inhomogeneity of variance, for instance the GARCH model in chapter 4, on the noise variance so that the NCR is changing over time.

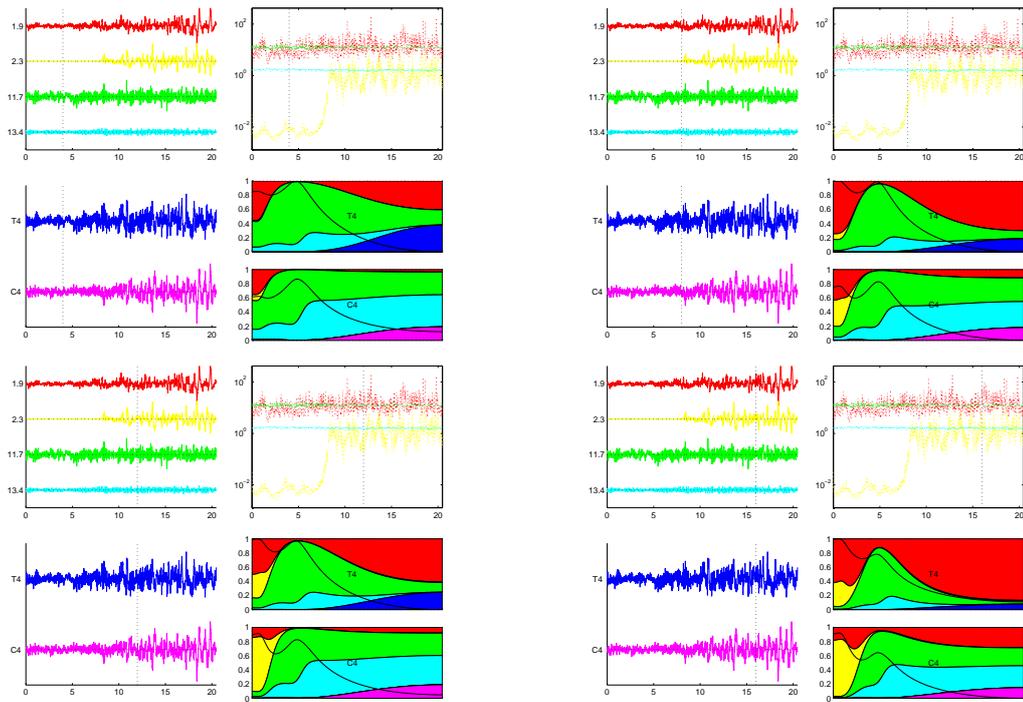


Figure 3.3: Decomposition, noise variance and NCR of compartment-GARCH model at 4 instants.

Chapter 4

Heteroscedasticity in State Space

When a model is established, we may want to investigate the normality assumption of the residuals. Although the residuals appear to be normal and uncorrelated, the square of residuals may be correlated. This suggests that the second moment, ie variance, is not constant. ‘Non-stationary in variance’, ‘heterogeneity of variance’ or ‘heteroscedasticity’ are always used to describe this situation.

In this chapter, the technique of modelling non-stationarity and heteroscedasticity will be first discussed. Then, the way how to implement non-stationarity and heteroscedasticity to state space model will come after. An important trick for making the GARCH model working in the space space model framework will be introduced in details. Last but not least, several alternative methods including jump detection, exogenous input variables and chi-square statistics of innovations will be presented along with examples.

4.1 Chi-square Statistics based on Innovations

The innovations, or prediction error, always provide us a good information about the goodness of the model, since we choose to use a maximum likelihood to decide parameter estimate, and likelihood function basically depends on the innovation. In the state space model, innovation is a collection of prediction error coming from system noise and observation noise. If there is a moment when innovation goes up to a large value, we may suspect that there is a sudden event happened in the data, and we may need to modify our model accordingly.

During the process of Kalman filter, in the prediction step and the filtering step at every time point, an innovation and its conditional variance are generated in order to produce state filter estimate and state variance filtered estimate. A large value in the innovation can push the filter estimate away from the predictor estimate. At this point, the innovation variance plays a

key role to adjust the conditional expectation and the Kalman gain to help innovation to do its job well.

By assumption, when the state space model is specified correctly, innovation ought to be identically independently normally distributed, and would not show any pattern. Otherwise, we may see some pattern in the innovations. The pattern can be seen in many ways; one way is summing all the squared normalized innovations to form a chi-square statistic. This chi-square statistic, C^2 , follows a chi-square distribution under the normal assumption.

$$C^2_{t-1} = \sum_{j=1}^u \boldsymbol{\nu}_{t-j} \boldsymbol{\Omega}_{t-j} \boldsymbol{\nu}'_{t-j} \sim \chi^2(u)$$

When any changes happens, the innovation become large and the innovation variance may not be able to react immediately. In this situation the normalized innovation becomes very large and pull the chi-square statistic up to a large value. Also, when the system noise variance or observation noise variance are over-estimated, innovation variance is retained at an excessive value, giving a small value of chi-square statistic. Therefore, comparing the chi-square statistic with a chi-square distribution can help to find out insufficiency of the model.

The choice of u depends on data characteristics. If u is chosen to be small, chi-square statistic is sensible to an instantaneous outlier in the innovation; if u is chosen to be large, chi-square would be more conservative and showing changes when there is consecutive improper values of normalized innovations.

For example, if we observe the chi-square statistic of the sum of 50 squared normalized innovations, ie $q = 50$, then the 95% confidence region for $\chi^2(50)$ distribution is $[\chi^2(.025, 50), \chi^2(.975, 50)] = [32.36, 71.42]$, and has a probability of 0.975 to cover the chi-square statistics.

Example 4.1 *We apply a compartment model to an EEG data of a coma patient. We use 4 AR(2) compartments and 2 AR(1) compartments for our compartment modelling. In figure 4.1(a), we show the state decomposition in the first 6 panels and the data in the bottom panel. We can see that there is a sudden pulse between 4s to 4.5s, which is captured by a 0 Hz AR(1) compartment. In figure 4.1(c) we show the compartment driving noise variance in the first 6 panels and the normalized innovation in the bottom panel. We can clearly see that the normalized innovation is not white around that moment.*

In this case we apply the same compartment model to the same data again with an exogenous input on the noise variance as in equation 4.10. In figure 4.1(b) we show the decomposition of 6 compartments and the data although we can only see a tiny difference between this model and the previous model. In figure 4.1(d) we show the state noise variance in the first 6 panels and the normalized innovation in the bottom panel. We can see that the exogenous

input raises the variance up in the first panel, and the spike around the 4.5 s region disappears in the normalized innovation in the bottom panel.

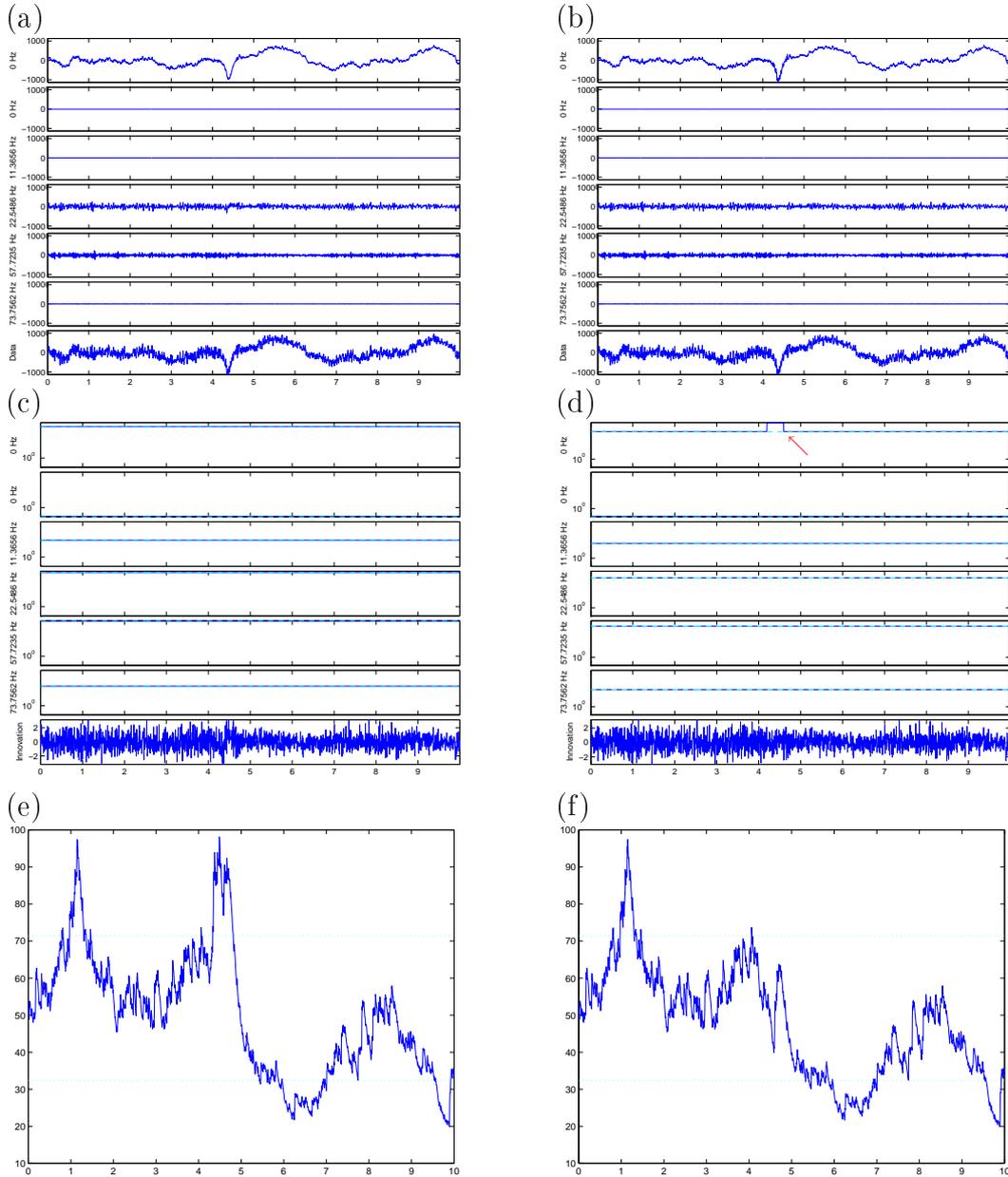


Figure 4.1: Chi-square statistics of compartment model with and without exogenous input of noise variance

In figures 4.1(e) and 4.1(f) we respectively show the exponential smoother Chi-square statistics as defined in equation 4.9, $q = 50$. The cyan dotted lines are 95% confidence region for $\chi^2(50)$ distribution, ie $[\chi^2(.025, 50), \chi^2(.975, 50)] = [32.36, 71.42]$. We can clearly see that after applying an exogenous input to the noise variance the chi-square statistics does not rise up beyond the upper

threshold. We change the intensity of the state driving noise variance and it consequently changes the innovation variance and improves likelihood. Except the region between 4s and 4.5s, both the chi-square statistics show similar shape, ie a large value around 1s and small values around 6s to 7s and 9.5s to 10s. This implies that the chi-square statistics is adaptive to the change of state noise variance, such that a sudden change in the dynamics in the state can be reflected immediately to the innovation. \square

The chi-square statistics is not only a tool for diagnostic checking, but also a clue for predicting the variance size next step. A large chi-square statistic implies that the variance is not large enough and a small chi-square statistic implies that the variance is too large. Using the chi-square statistics, or equivalently using the square of innovations to predict variance is the core idea of the GARCH type model which will be explained in the following section.

4.2 State Space GARCH Model

There were studies of combining the use of GARCH model and the state space model, by making the variance of the noise be the state of the state space model. Although this approach takes the benefit of state space model that the state, ie conditional variance, is unobserved, it also consumes the space of system equation from modelling the dynamics of the data.

The concept of employing the GARCH model within a state space has been introduced by Galka et al. (2004) in a study on the estimation of inverse solutions from EEG time series; for simulated data they obtained improved the reconstruction of true states by this technique.

In this thesis we will have the GARCH model applied on the driving noise variance of the system equation. This driving noise variance will become time dependent and also predictable. The time varying driving noise variance gives a conclusion of how the intensity of the driving noise changes in time. We will study how we can implement the GARCH model to the system noise of the state space model, with (1) an additional step in the Kalman filter algorithm and (2) a simple substitution of the conditional expected driving noise square.

4.2.1 GARCH Type Models

GARCH model stands for the generalized autoregressive conditional heteroscedasticity model. The idea of conditional heteroscedasticity is first introduced by Engle (1982) as the ARCH model that a noise variance which is non-homogeneous from time to time can be modelled by regressing it on the past

heteroscedastic noises.

$$\begin{aligned}\eta_t &= \sigma_t \nu_t, \quad \nu_t \sim N(0, 1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^p \alpha_i \eta_{t-i}^2\end{aligned}\quad (4.1)$$

Bollerslev (1986) made a trivial additional step to give a generalized ARCH model, which allows the variance regressing not only past noises but also past variances. In this way we may be able to reduce the number of parameters in order to replace a high order of ARCH model. A typical GARCH model can be written as equation 4.2.

$$\begin{aligned}\eta_t &= \sigma_t \nu_t, \quad \nu_t \sim N(0, 1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^u \alpha_i \eta_{t-i}^2 + \sum_{j=1}^v \beta_j \sigma_{t-j}^2\end{aligned}\quad (4.2)$$

σ_t^2 is regressed on its past values like an AR part of the GARCH model and the past squared noises η_t^2 like an AR part of another component of a multivariate model. α is usually called ARCH coefficient and β is usually called GARCH coefficient. The ARCH model is a special case of GARCH model when the GARCH coefficients are all zeros.

Nelson (1991) observed that in financial time series conditional heteroscedasticity has unequal effect from positive innovation and negative innovation, so he put an asymmetric weighted innovation into the GARCH model in addition to taking logarithm of the conditional variance for his EGARCH model.

$$\begin{aligned}\nu_t &= \sigma_t \eta_t \\ \eta_t &\sim N(0, 1) \\ \log \sigma_t^2 &= \alpha_0 + \sum_{i=1}^u \alpha_i g(\nu_{t-i}^2) + \sum_{j=1}^v \beta_j \log \sigma_{t-j}^2\end{aligned}\quad (4.3)$$

where $g(\nu_{t-i}^2)$ is a function in ν , such a function can be defined in many different ways, see Nelson (1991); Tsay (2005). We choose a simple case where $g(\nu_{t-i}^2) = \log \nu_{t-i}^2$. In essence, only taking the logarithm of σ and ν is so useful to avoid a negative value of σ next time steps.

Note that if $\alpha_1 = 1$ and all other α and all β are zeros, the variance at each time is taking the same value as the previous time, which means that there is no GARCH effect on the noise variance. This is true for ARCH, GARCH and EGARCH models.

GARCH model is not a stochastic model because it has no randomness. All the input variables and parameters are specified at each time step. And the conditional variance follows the residuals to change and has no impact on the residuals to give better prediction next step.

4.2.2 Compartment GARCH model formulation

We apply a GARCH type model to the system noise variance of the state space model. The general state space GARCH model can be written as

$$\begin{cases} \mathbf{x}_t = F\mathbf{x}_{t-1} + G\mathbf{w}_t \\ \mathbf{y}_t = H\mathbf{x}_t + \boldsymbol{\epsilon}_t \end{cases}$$

$$\mathbf{w}_t \sim N(0, Q_t)$$

Q_t denotes the variance matrix of the dynamical noise \mathbf{w}_t .

$$Q_t = \begin{bmatrix} \tau_t^{(1)2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \tau_t^{(2)2} & \ddots & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \tau_t^{(r)2} & \ddots & & \vdots \\ \vdots & & & \ddots & \tau_t^{(r+1)2} & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \tau_t^{(r+s)2} \end{bmatrix}.$$

The variance of the dynamical noise are obtained from the diagonal elements of Q . We write $\tau_t^{(k)2}$ for the variance of the dynamical noise of k -th compartment. The nonzero elements $\tau_t^{(k)2}$ are time-dependent; according to the GARCH approach in its general form, they are modelled by

$$\log \tau_t^{(k)2} = \alpha_0^{(k)} + \sum_{i=1}^u \alpha_i^{(k)} \log w_{t-i}^{(k)2} + \sum_{j=1}^v \beta_j^{(k)} \log \tau_{t-j}^{(k)2}. \quad (4.4)$$

The first term on the right-hand side of this equation, α_0 represents a constant term. The first sum represents an ARCH part of order u for the logarithm of the noise square, while the second sum corresponds to a GARCH part of order v of the noise variance. Using EGARCH model has the beneficial effect of preventing the variance from becoming negative.

In the original EGARCH model 4.3 the ARCH part $g(\nu_{t-j}^2)$ would be given by previous values of the prediction error of the data, but in the state space approach the state prediction errors are not directly accessible, therefore we estimate it by applying a trick in next section.

4.2.3 Substitution of Unobservable Noise

As we started the idea of combining a GARCH model and a state space model, the residual of the state equation, \mathbf{w} , is not calculated in the Kalman filter

algorithm. There is no direct access to this noise, and at this point we have no way to make use of the GARCH model.

However, in Kalman filtering the conditional expected value of the state and the conditional expected covariance matrix of the state are calculated at each time point. We adopt the idea of conditional expectation, a measure which is based on the observation up to the current time point. The ARCH part of equation 4.4 requires a square of the noise, therefore we replace it by the expected square of state dynamic noise conditional to the observations up to the current time point.

$$\begin{aligned}
& E(G\mathbf{w}_{t-j}(G\mathbf{w}_{t-j})' | \mathbf{y}_{t-j}) \\
&= E(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j}) E(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j})' + \text{Var}(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j}) \\
&= E(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j}) E(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j})' + \text{Var}(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j-1}, \boldsymbol{\nu}_{t-j}) \\
&= E(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j}) E(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j})' + [\text{Var}(G\mathbf{w}_{t-j} | \mathbf{y}_{t-j-1}) \\
&\quad - \text{Cov}(G\mathbf{w}_{t-j}, \boldsymbol{\nu}_{t-j} | \mathbf{y}_{t-j-1}) \text{Var}(\boldsymbol{\nu}_{t-j})^{-1} \text{Cov}(G\mathbf{w}_{t-j}, \boldsymbol{\nu}_{t-j} | \mathbf{y}_{t-j-1})'] \\
&= K_{t-j} \boldsymbol{\nu}_{t-j} \boldsymbol{\nu}_{t-j}' K_{t-j}' + GQ_{t-j}G' - GQ_{t-j}G'H'\Omega_{t-j}HGQ_{t-j}G'
\end{aligned}$$

Therefore, we have an estimate for w^2 in equation 4.4 by forming the matrix $E(G\mathbf{w}_{t-j}(G\mathbf{w}_{t-j})' | \mathbf{y}_{t-j})$ and extract from its diagonal the first element for $\log \tau_t^{(1)^2}$, the third element for $\log \tau_t^{(2)^2}$ up to the $(2r-1)$ th for the r -th AR(2) compartment, and the $(2r+1)$ th element for $\log \tau_t^{(r+1)^2}$ and so on for the other AR(1) compartment.

It is obvious to see that the GARCH model is applied to system noise variance, so it is a business in the system noise space. To obtain a conditional expectation of system noise we require the information of innovation, resulting we can only get the conditional expectation of system noise on the state space. As a result we have only taken out values from $E(G\mathbf{w}_{t-j}(G\mathbf{w}_{t-j})' | \mathbf{y}_{t-j})$ and sacrificed some information in it. The above estimate is of the size $m \times m$, but only k out of $m \times m$ ($k < m = k + r$) elements are used.

If we have the system error estimate then it is perfect. Since we do not have it so projecting the innovation onto the system noise space is the ideal way. Mathematically if we can construct a $k \times m$ pseudo-inverse G^- for G , we obtain a simplified estimate $E(\mathbf{w}_{t-j}\mathbf{w}_{t-j}' | \mathbf{y}_{t-j})$, which is of size $k \times k$. Then each diagonal value is an estimate for w^2 .

$$E(\mathbf{w}_{t-j}\mathbf{w}_{t-j}' | \mathbf{y}_{t-j}) = G^- K_{t-j} \boldsymbol{\nu}_{t-j} \boldsymbol{\nu}_{t-j}' K_{t-j}' G^{-'} + Q_{t-j} - Q_{t-j}G'H'\Omega_{t-j}HGQ_{t-j}$$

The properties and the way of calculation of pseudo-inverse can be found in many textbooks, for example Goldberg (1991); Golub & Loan (1996).

For example, in the companion form of compartment model, when the MA coefficients of every compartment are all zero, ie G is defined as equation 2.11

4.3.2 Exponential Moving Average

When we use the simple moving average method, we have to store u ARCH variable \mathbf{w} for the GARCH model next time step. In order to reduce storage and to compute GARCH variance faster, we use the idea of exponential smoother in Engle (2002).

$$\alpha_1^{(k)} = \mu^{-1}\alpha_2^{(k)} = \dots = \mu^{-u}\alpha_{u+1}^{(k)} = \dots$$

Equation 4.6 will become

$$\log \tau_t^{(k)^2} = \alpha_0^{(k)} + \alpha_1^{(k)} \sum_{i=1}^{t-1} \mu^{i-1} \log w_{t-i}^{(k)^2} + \sum_{j=1}^v \beta_j^{(k)} \log \tau_{t-j}^{(k)^2}. \quad (4.7)$$

Or equivalently it can be written as

$$\begin{aligned} \log \tau_t^{(k)^2} &= \alpha_0^{(k)} + \alpha_1^{(k)} \log \bar{w}_{t-1}^{(k)^2} + \sum_{j=1}^v \beta_j^{(k)} \log \tau_{t-j}^{(k)^2} \\ \log \bar{w}_{t-1}^{(k)^2} &= \log w_{t-1}^{(k)^2} + \mu \log \bar{w}_{t-2}^{(k)^2} \end{aligned} \quad (4.8)$$

μ is a factor between 0 and 1. If we let $\mu = \frac{u-1}{u}$ for some positive integers u , then an exponential moving average $\log \bar{w}_{t-1}^{(k)^2}$ has a value of approximately u times $\log w_{t-1}^{(k)^2}$. Every step the exponential moving average is discounted by the factor μ and remedied by the current error estimate. The method gives a low consumption of storing error estimates and allows a faster computation.

We can also compute an exponential moving average of a chi-square statistics introduced previously. This exponential smoother chi-square statistic is approximately following a $\chi^2(u)$ distribution.

$$\begin{cases} C^2_0 &= q \\ C^2_t &= \frac{q-1}{q} C^2_{t-1} + \boldsymbol{\nu}_{t-1} \Omega_{t-1} \boldsymbol{\nu}'_{t-1} \\ C^2_{t-1} &= \sum_{j=1}^q \left(\frac{q-1}{q} \right)^{j-1} \boldsymbol{\nu}_{t-j} \Omega_{t-j} \boldsymbol{\nu}'_{t-j} \end{cases} \quad (4.9)$$

In this case we need not store q innovations but only one weighted sum. It provides a smoothing function to the basic version of chi-square statistics if we choose q to be small to detect quick but short changes.

4.4 Exogenous Variables

A straightforward and easy way to deal with a heteroscedastic data is to add an exogenous variable to the noise variance. When we have a prior knowledge of

any external stimulation, or when we observe that there is any obvious changes in the driving noise variance, we can apply this exogenous variable method. To apply the method we need a indicating variable ξ , storing 0 and 1 values. Let σ_c^2 be the homogeneous noise variance and σ_d be an adjustment term.

$$\sigma_t^2 = (\sigma_c + \xi_t \sigma_d)^2 \quad (4.10)$$

In example 4.1 we had applied the method of exogenous variables by switching the variance between a large and a small value. We assumed there is a change in the strength of a driving noise from 4 s to 4.5 s. We added a constant to the driving noise variance of one of the compartments.

This method is easy to apply because we just need to add an adjustment term to a homogeneous value. The meaning of the model is natural while we can see the variance is switching between values clearly.

However, this method depends on the indicating variable, therefore in the real time situation, we can apply it only when we have a prior knowledge.

4.5 Appendix

4.5.1 Derivation of Conditional Expectation of Square of System Noise

We will give the derivation of $E(G\mathbf{w}_{t-j}(G\mathbf{w}_{t-j})'|\mathbf{y}_{t-j})$, i.e. the estimator of the noise term $w_{t-j}^{(k)2}$ in equation 4.4.

Since the state prediction error w_{t-j} is not directly accessible, we derive a estimator with similar meaning. This estimator is chosen as the expectation of the product $\mathbf{w}_{t-j}\mathbf{w}'_{t-j}$, conditional on the data up to time $t-j$.

Let K_{t-j} , $\boldsymbol{\nu}_{t-j}$ and Ω_{t-j} denote the Kalman gain, innovations and inverse covariance matrix of innovations, respectively, at time $t-j$. These quantities are obtained naturally through the application of the Kalman filter (for details see Harvey (1989)). Then we have equation 4.11.

$$\begin{aligned} E(G\mathbf{w}_{t-j}(G\mathbf{w}_{t-j})'|\mathbf{y}_{t-j}) = \\ \text{Var}(G\mathbf{w}_{t-j}|\mathbf{y}_{t-j}) + E(G\mathbf{w}_{t-j}|\mathbf{y}_{t-j})E(G\mathbf{w}_{t-j}|\mathbf{y}_{t-j})' \end{aligned} \quad (4.11)$$

The expectation $E(G\mathbf{w}_{t-j}|\mathbf{y}_{t-j})$ is equal to $K_{t-j}\boldsymbol{\nu}_{t-j}$. The term $\text{Var}(G\mathbf{w}_{t-j}|\mathbf{y}_{t-j})$ represents the conditional variance of the system noise, which can be expressed as equation 4.12

$$\begin{aligned}
& \text{Var} (G\mathbf{w}_{t-j}|\mathbf{y}_{t-j}) \\
&= \text{Var} (G\mathbf{w}_{t-j}|\mathbf{y}_{t-j-1}, \boldsymbol{\nu}_{t-j}) \\
&= \text{Var} (G\mathbf{w}_{t-j}|\mathbf{y}_{t-j-1}) \\
&\quad - \text{Cov} (G\mathbf{w}_{t-j}, \boldsymbol{\nu}_{t-j}|\mathbf{y}_{t-j-1}) \text{Var} (\boldsymbol{\nu}_{t-j})^{-1} \text{Cov} (G\mathbf{w}_{t-j}, \boldsymbol{\nu}_{t-j}|\mathbf{y}_{t-j-1})'
\end{aligned} \tag{4.12}$$

$$\begin{aligned}
& \text{Cov} (G\mathbf{w}_{t-j}, \boldsymbol{\nu}_{t-j}|\mathbf{y}_{t-j-1}) \\
&= \text{E} (G\mathbf{w}_{t-j}\boldsymbol{\nu}_{t-j}|\mathbf{y}_{t-j-1}) - \text{E} (G\mathbf{w}_{t-j}|\mathbf{y}_{t-j-1}) \text{E} (\boldsymbol{\nu}_{t-j}|\mathbf{y}_{t-j-1}) \\
&= \text{E} \left(G\mathbf{w}_{t-j} (\mathbf{y}_{t-j} - H\mathbf{x}_{t-j|t-j-1})' |\mathbf{y}_{t-j} \right) \\
&= \text{E} \left(G\mathbf{w}_{t-j} \{ [H(F\mathbf{x}_{t-j-1} + G\mathbf{w}_{t-j}) + \boldsymbol{\epsilon}_{t-j}] - H(F\mathbf{x}_{t-j-1|t-j-1}) \}' |\mathbf{y}_{t-j} \right) \\
&= \text{E} \left(G\mathbf{w}_{t-j} \{ HF(\mathbf{x}_{t-j-1} - \mathbf{x}_{t-j-1|t-j-1}) \}' |\mathbf{y}_{t-j} \right) \\
&\quad + \text{E} (G\mathbf{w}_{t-j}\mathbf{w}'_{t-j}G'H'|\mathbf{y}_{t-j}) + \text{E} (G\mathbf{w}_{t-j}\boldsymbol{\epsilon}'_{t-j}|\mathbf{y}_{t-j}) \\
&= 0 + GQ_{t-1}G'H' + 0 \\
&= GQ_{t-1}G'H'
\end{aligned} \tag{4.13}$$

In equation 4.13, we substitute $\text{E}(\boldsymbol{\nu}_{t-j}|\mathbf{y}_{t-j-1}) = 0$ in the second line; in the fifth line $\text{E}(G\mathbf{w}_{t-j}\{HF(\mathbf{x}_{t-j-1} - \mathbf{x}_{t-j-1|t-j-1})\}'|\mathbf{y}_{t-j}) = 0$; in the sixth line $\text{E}(G\mathbf{w}_{t-j}\mathbf{w}'_{t-j}G'H'|\mathbf{y}_{t-j}) = GQ_{t-1}G'H'$ and $\text{E}(G\mathbf{w}_{t-j}\boldsymbol{\epsilon}'_{t-j}|\mathbf{y}_{t-j}) = 0$. By substituting equations 4.12 and 4.13 into equation 4.11, we get equation 4.14.

$$\begin{aligned}
& \text{E} (G\mathbf{w}_{t-j} (G\mathbf{w}_{t-j})' |\mathbf{y}_{t-j}) = \\
& \quad K_{t-j}\boldsymbol{\nu}_{t-j}\boldsymbol{\nu}'_{t-j}K'_{t-j} + GQ_{t-j}G' - GQ_{t-j}G'H'\Omega_{t-j}HGQ_{t-j}G' \tag{4.14}
\end{aligned}$$

Chapter 5

Application 1 - Detection of Consciousness in Surgery

5.1 Introduction

When an open surgery is carried out, different kinds of anaesthesia may be needed to buffer the pain of invasive operations. Anaesthesia has many benefits beside its side effect. Modern surgery could not move forward if anaesthesia were not applied. Although anaesthetic mishap happens rarely nowadays, they still occur sometimes. The damages suffered by medical doctors and hospitals can be very costly. Also, on the patients' side, death can result from malpractice of anesthesia. Therefore anaesthesia should be applied with absolutely no error.

In the operating room, anesthesiologists have to be professional to make correct decision, but they can easily make a mistake in many ways in that complicated environment. By quantifying the level of consciousness during surgical anaesthesia with an FDA-approved monitor, John (2002b) found that an increase in absolute power of the low frequency band occurs invariably when patients lose consciousness. Through observing the change at the low frequency band it is helpful to ensure adequate surgical anaesthesia.

The method underlying the monitors of depth of anaesthesia is based on frequency spectrum. We reviewed in the introduction chapter that various spectral models were used because of an eagerness of explaining the EEG data in the frequency domain. A frequency spectrum can be done by parametric or nonparametric methods. Applying FFT, a well-known nonparametric method, to a time series data we can immediately obtain the intensity of the power at different frequency bands (Priestley, 1981), while fitting of autoregressive (AR) models is a prominent example of a parametric method (Box & Jenkins, 1970; Gersch, 1970).

However, in the case of the presence of pronounced non-stationarity in the EEG, such as time-dependent changes of the power in different frequency

bands, direct application of the FFT to the data would be inappropriate. Although in this case it is still possible to apply the FFT to a window moving over the data, such approach would have the disadvantage of reduced resolution either in time or in frequency domain; improved resolution in time domain, desirable in order to pick out distinctive temporal characteristics in the data, has to be paid by reduced resolution in frequency domain, and vice versa.

In contrast, parametric spectral estimation by AR models offers various advantages over the FFT, since it represents a more general and flexible framework for parsimonious dynamical modelling of time series data, which can be readily employed for purposes such as prediction, classification or causality analysis of time series (Shumway, 2000); in the case of non-stationarity, parametric spectral estimation may also be applied to a moving window (Ozaki & Tong, 1975), but there is an alternative approach for this situation which avoids the introduction of a moving window. This alternative approach is to introduce a volatility model to the noise, in order to allow a time varying noise variance. The volatility model, or specifically, the GARCH model, has been introduced in chapter 4. In this chapter we will study the topic of loss of consciousness by anaesthesia, and to confirm the previous result by a parametric approach.

5.2 Data

The EEG time series which we will study in this chapter was retained from a recent study of John et al. (2001) and John (2002a) who have studied the change of spectral content of clinical EEG accompanying the loss and subsequent recovery of consciousness due to initiation and termination of anaesthesia during surgery.

We select from their data a segment of 2048 samples, sampled at 100 Hz, such that the segment extends over about 20 seconds. It was measured at 19 electrodes fixed to the scalp according to the international 10/20 System. This data set covers the transition from awake conscious state to anaesthesia. The detailed experimental procedures have been described in John et al. (2001) or Prichep et al. (2004).

In figure 5.1 we show the raw data, while in figure 5.2 we show the transformed data (versus average reference), on which we are mainly working. In the figure we can see the data is mainly split into two halves. Especially in the second half, there is a surge of a low frequency activity, at about 4 oscillations in 2 seconds. Channels in the frontal-parietal (Fp), the temporal (T) and the parietal (P) regions show high frequency activities before the transition. Alpha oscillation can be seen in the whole period of the data.

In figure 5.3 we show the best AR model periodogram of the raw EEG data. It is an overlay plot of periodograms of the 20 channels. Each periodogram is

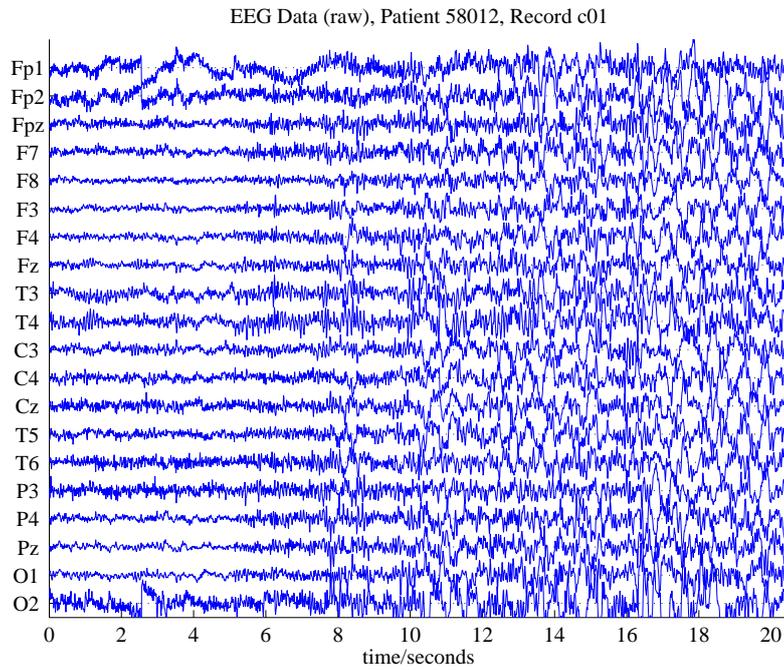


Figure 5.1: Time series plot of a raw EEG data, patient 58012, condition c01.

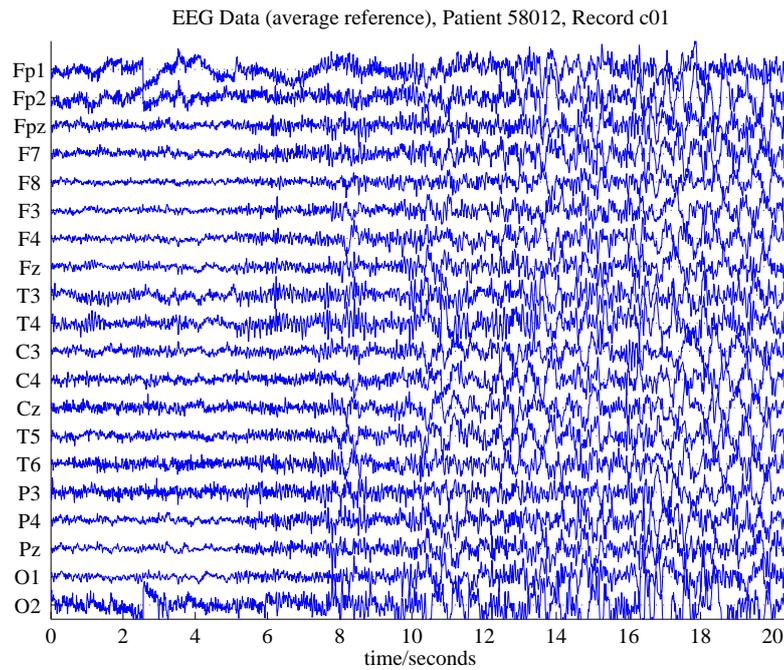


Figure 5.2: Time series plot of a transformed EEG data, patient 58012, condition c01.

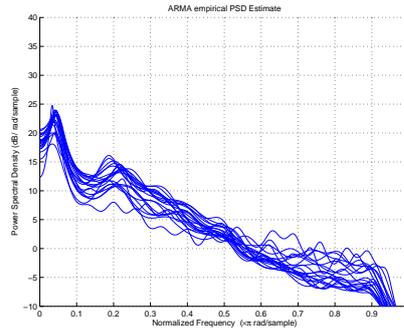


Figure 5.3: Best AR order periodogram of the raw EEG data

obtained from an AR model of data of a single channel. The best AR model for each channel is one of the AR model of 1 to order 30 which has the minimum AIC over the others. From the plot, we can see that all the 20 channels provide an information about the strong alpha frequency (10 Hz, 0.2π) and a strong low frequency (~ 2 Hz, 0.04π), which are those what we have seen from the data plot.

5.3 Analysis

5.3.1 Segmented Data Analysis

We select the T4 electrode for a segment-wise data analysis. We will look at segments of this univariate time series. We split into 15 segments which are 1 s to 256 s, 129 s to 384 s, 257 s to 512 s, ..., 1793 s to 2048 s.

(a)

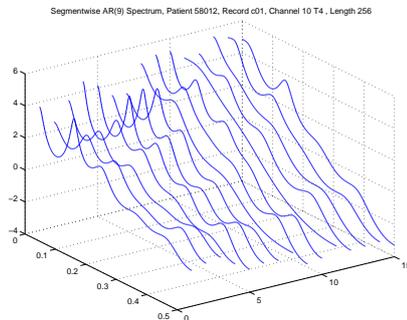


Figure 5.4: Segmentwise AR(9) spectrum, patient 58012, condition c01.

(b)

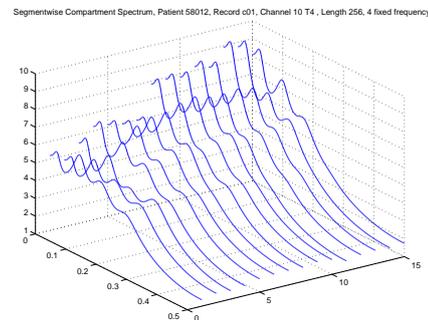


Figure 5.5: Segmentwise compartment(4,0) spectrum, patient 58012, condition c01, fixed frequencies.

In figure 5.4 we show a 3 dimensional segmentwise AR(9) spectrum. The

vertical axis represents the power intensity of the spectrum. Each contour line is the spectrum of a fitted AR(9) model for a moving window of data of 256 points. Putting the lines in the sequence of time we see the changes of the spectrum with respect to time.

The low frequency shows a rise from the middle of the data. The 10 Hz(0.1) frequency is persistently at a constant level of power, and so as 20 Hz(0.2) and high frequency range (0.3-0.5). All these peaks are blurred out with the upwards shifting of the whole spectrum in the second half, when the low frequency appears. The AR(9) spectrum is trying to capture the significance of the low frequency but could not show the 2 Hz peak in the spectrum, due to a limited number of data of moving window that consequently reduces frequency resolution.

Therefore we show a segmentwise compartment spectrum in figure 5.5. The transition matrix of compartment model is fixed to be the same in each segment, therefore the peaks stay at the same frequencies in the figure. Figure 5.5 clearly shows that the low frequency is going up in the second half. The alpha frequency rises in the first half but not showing the dropping in the second half.

5.3.2 Compartment Modelling

We choose the C3 electrode in this modelling analysis. We aim at decomposing the data into a set of source components, we choose a special structure for the state space model, such that pairs of elements within the state vector x_t represents AR(2) models, as we have discuss in chapter 2. Each AR(2) model is capable of describing one main frequency found in the data.

This channel 11 (C3) is similar to channel 10 (T4), that the amplitude of data rises up in the second half. The surge of low frequency oscillation can be seen clearly in both data and segmentwise spectra in figure 5.6.

In table 5.1 we show the summary of compartment model (2,1), ie, 2 AR(2) compartments and 1 AR(1) compartment.

Compartment(2,1) Model					
	phi 1	phi 2	magnitude	frequency	s.d
Compartment 1	1.8959	-0.91178	0.95487	1.91741	0.47007
Compartment 2	1.0559	-0.54686	0.7395	12.3453	4.0997
Compartment 3	0.0171	0	0.0171	0	0.0

$Q = I_3, R = 0.01, -2 \log\text{-likelihood} = 12035.3793,$
number of parameter = 8, AIC = 12051.3793

Table 5.1: Compartment (2,1) Model

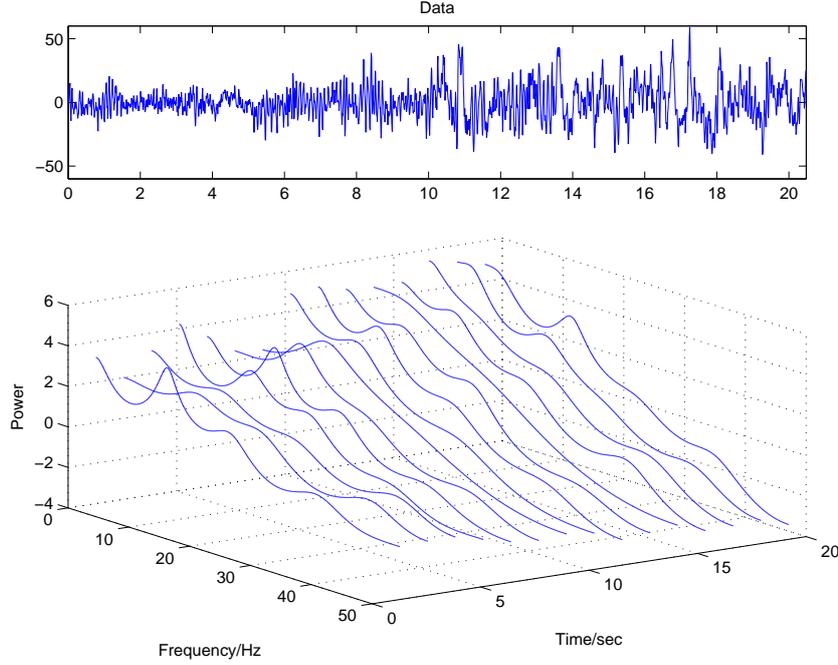


Figure 5.6: (top) An EEG data from the T4 electrode (versus average reference) of about 20 seconds, covers the transition from awake conscious state to anaesthesia. (bottom) A moving window spectral estimation of AR(8) models fitted to 15 segments of data, each of length 256.

It means

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + G\mathbf{w}_t$$

$$\mathbf{y}_t = H\mathbf{x}_t + \boldsymbol{\epsilon}_t$$

$$\mathbf{x}_t = \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \\ x_t^{(4)} \\ x_t^{(5)} \end{bmatrix} \quad F = \begin{bmatrix} 1.895 & 1 & & & \\ -0.911 & 0 & & & \\ & & 1.055 & 1 & \\ & & -0.546 & 0 & \\ & & & & 0.017 \end{bmatrix} \quad G = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{w}_t = \begin{bmatrix} w_t^{(1)} \\ w_t^{(2)} \\ w_t^{(3)} \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{w}_t \sim N(0, Q), \quad \boldsymbol{\epsilon}_t \sim N(0, R), \quad Q = \text{diag}(0.47^2, 4.10^2, 0.00^2), \quad R = 0.01.$$

In table 5.2 we show the fitted compartment models of different model order. The AR(1) compartment does not seem to be necessary as we see that all the compartment (2,1), (3,1) and (4,1) models choose a zero weight for the AR(1) compartment. Among these models AR(4,0) is the suitable model.

Compartment(3,0) Model					
	phi 1	phi 2	magnitude	frequency	s.d
Compartment 1	1.8796	-0.89687	0.94703	1.96594	0.53795
Compartment 2	1.3087	-0.70181	0.83774	10.7336	2.6018
Compartment 3	0.49447	-0.60964	0.7808	19.8722	2.3118
$Q = I_3, R = 0.01, -2 \log\text{-likelihood} = 11973.1552,$ number of parameter = 9, AIC = 11991.1552					
Compartment(3,1) Model					
	phi 1	phi 2	magnitude	frequency	s.d
Compartment 1	1.8796	-0.89687	0.94703	1.96594	0.53795
Compartment 2	1.3087	-0.70181	0.83774	10.7336	2.6018
Compartment 3	0.49447	-0.60964	0.7808	19.8722	2.3118
Compartment 4	0.85923	0	0.85923	0	0.0
$Q = I_4, R = 0.01, -2 \log\text{-likelihood} = 11973.1552,$ number of parameter = 11, AIC = 11995.1552					
Compartment(4,0) Model					
	phi 1	phi 2	magnitude	frequency	s.d
Compartment 1	1.8769	-0.89435	0.9457	1.97258	0.54913
Compartment 2	1.3318	-0.71426	0.84514	10.5578	2.4823
Compartment 3	0.79103	-0.75646	0.86974	17.4864	1.453
Compartment 4	0.12307	-0.64117	0.80073	23.7757	1.5183
$Q = I_5, R = 0.01, -2 \log\text{-likelihood} = 11953.9385,$ number of parameter = 12, AIC = 11977.9385					
Compartment(4,1) Model					
	phi 1	phi 2	magnitude	frequency	s.d
Compartment 1	1.8769	-0.89435	0.9457	1.97258	0.54913
Compartment 2	1.3318	-0.71426	0.84514	10.5578	2.4823
Compartment 3	0.79103	-0.75646	0.86974	17.4864	1.453
Compartment 4	0.12307	-0.64117	0.80073	23.7757	1.5183
Compartment 5	0	0	0	0	0
$Q = I_4, R = 0.01, -2 \log\text{-likelihood} = 11953.9385,$ number of parameter = 14, AIC = 12981.9385					

Table 5.2: Summary of the parameters of fitted compartment (3,0), (3,1), (4,0), (4,1).

5.3.3 GARCH Modelling

We choose to employ a state space model consisting of $r = 4$ AR(2) models, such that 4 major frequency bands can be described. By fitting the model to the data shown in Figure 5.6 these delta (2.4 Hz), alpha (10.3 Hz), mid-range beta (17.6 Hz) and low-range gamma (24.5 Hz) frequencies are found.

We find that the GARCH parameters $\beta_i^{(k)}$ in Equation 4.4 do not differ significantly from zero, therefore we set the GARCH order to zero, ie $v = 0$, which helps to reduce the number of parameters to be fitted. We also impose the simple moving average to the ARCH parameters, ie $\alpha_1^{(k)} = \dots = \alpha_u^{(k)} =: \alpha^{(k)}$, such that the variances can change smoothly, and we have a further reduction of the number of parameters. Since the likelihood does not improve significantly for ARCH order u larger than 2, we set $u = 2$. The resulting model estimate is

$$\begin{aligned}\mathbf{x}_t &= F\mathbf{x}_{t-1} + G\mathbf{w}_t \\ \mathbf{y}_t &= H\mathbf{x}_t + \boldsymbol{\epsilon}_t\end{aligned}$$

$$\mathbf{x}_t = \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \\ x_t^{(4)} \\ x_t^{(5)} \\ x_t^{(6)} \\ x_t^{(7)} \\ x_t^{(8)} \end{bmatrix} \quad F = \begin{bmatrix} 1.777 & 1 & & & & & & \\ -0.808 & 0 & & & & & & \\ & & 1.359 & 1 & & & & \\ & & -0.723 & 0 & & & & \\ & & & & 0.774 & 1 & & \\ & & & & -0.746 & 0 & & \\ & & & & & & 0.054 & 1 \\ & & & & & & -0.672 & 0 \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & & & & & & & \\ 0 & & & & & & & \\ & 1 & & & & & & \\ & 0 & & & & & & \\ & & 1 & & & & & \\ & & 0 & & & & & \\ & & & 1 & & & & \\ & & & 0 & & & & \\ & & & & 1 & & & \\ & & & & 0 & & & \end{bmatrix} \quad \mathbf{w}_t = \begin{bmatrix} w_t^{(1)} \\ w_t^{(2)} \\ w_t^{(3)} \\ w_t^{(4)} \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$\mathbf{w}_t \sim N(0, Q_t)$, $\boldsymbol{\epsilon}_t \sim N(0, R)$, $Q_t = \text{diag}(\tau_t^{(1)2}, \tau_t^{(2)2}, \tau_t^{(3)2}, \tau_t^{(4)2})$, $R = 0.01$, and

$$\begin{bmatrix} \tau_t^{(1)^2} \\ \tau_t^{(2)^2} \\ \tau_t^{(3)^2} \\ \tau_t^{(4)^2} \end{bmatrix} = \begin{bmatrix} -1.097 \\ 2.186 \\ 0.182 \\ 0.546 \end{bmatrix} + \begin{bmatrix} 0.829 \left(\log w_{t-1}^{(1)^2} + \log w_{t-2}^{(1)^2} \right) \\ 0.153 \left(\log w_{t-1}^{(2)^2} + \log w_{t-2}^{(2)^2} \right) \\ 0.939 \left(\log w_{t-1}^{(3)^2} + \log w_{t-2}^{(3)^2} \right) \\ -0.227 \left(\log w_{t-1}^{(4)^2} + \log w_{t-2}^{(4)^2} \right) \end{bmatrix}.$$

We use the method in section 4.2.3 to obtain a conditional expectation of $w_{t-i}^{(k)^2}$.

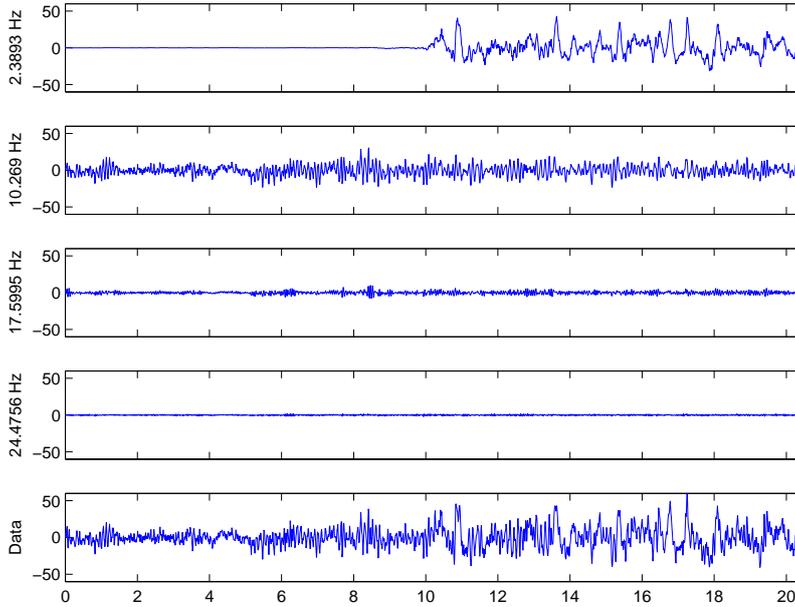


Figure 5.7: (top 4) Estimated source components of 4 different frequencies, scaled to the data space. (bottom) Data.

In Figure 5.7 we show the estimated components $x_t^{(k)}$, $k = 1, 3, 5, 7$; these components, each corresponding to one of the four frequencies, represent a decomposition of the original data (shown again in the bottom panel of the figure), such that by summing up these components the original data is reproduced. Note that a pronounced increase of amplitude occurs for the 2.4 Hz-component at about 10 seconds; this effect is solely obtained as a result of the maximum-likelihood model fit, without any input of prior knowledge concerning the change of the spectral composition of the data.

In Figure 5.8 we show explicitly the variances $\log \left(\tau_t^{(k)} \right)^2$ of the components as functions of time, according to the fitted GARCH models. It can be

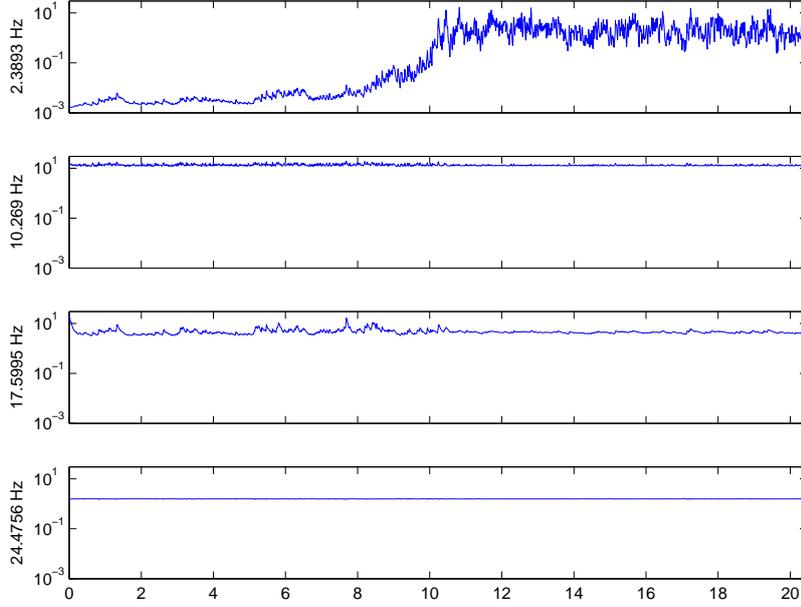


Figure 5.8: (top 4) Variance of prediction error of the 4 source components.

seen that for the 2.4 Hz-component the variance increases at about 8 seconds to a considerably larger value than before, and maintains that larger value within the second half of the data set. The variances of the other components do not display significant changes. This result can be readily interpreted by stating that the loss of consciousness at onset of anaesthesia is reflected almost exclusively by an increase of power in the delta band.

5.3.4 Simulation Study

In this subsection we do a simulation study. We simulate data with the compartment model from the result of the anaesthesia data in the previous subsection. The state space model is defined by the characteristic roots, and the variance of 2.4 Hz is one-step up function, looks like the variance shown in figure 5.8. We show one set of simulated states and observation in figure 5.9.

By fitting a compartment GARCH model to a simulated data, we obtain one set of parameter estimates. The simulation and the model fitting were done 200 times, and the summary of the true parameters and their mean estimates and standard error are shown in table 5.3. We also show the histogram plot of the estimates in figure 5.10. The result shows that the compartment-GARCH model can capture the parameters of the true model well. It guarantees the suitability of our model in this kind of data.

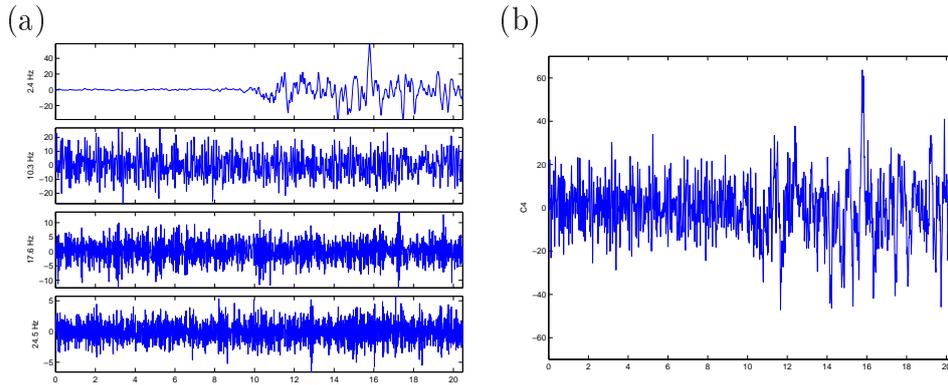


Figure 5.9: Simulated states and observation

parameter	true value	mean estimation	std errors
$ \lambda_1 $	0.8990	0.8991	0.0003
$ \lambda_2 $	0.8638	0.8638	0.0001
$ \lambda_3 $	0.8505	0.8505	0.0003
$ \lambda_4 $	0.8202	0.8202	0.0003
$\arg(\lambda_1)$	0.1501	0.1501	0.0001
$\arg(\lambda_2)$	1.1058	1.1055	0.0138
$\arg(\lambda_3)$	0.6452	0.6442	0.0060
$\arg(\lambda_4)$	1.5378	1.5388	0.0139

Table 5.3: Mean and standard error of parameter estimates

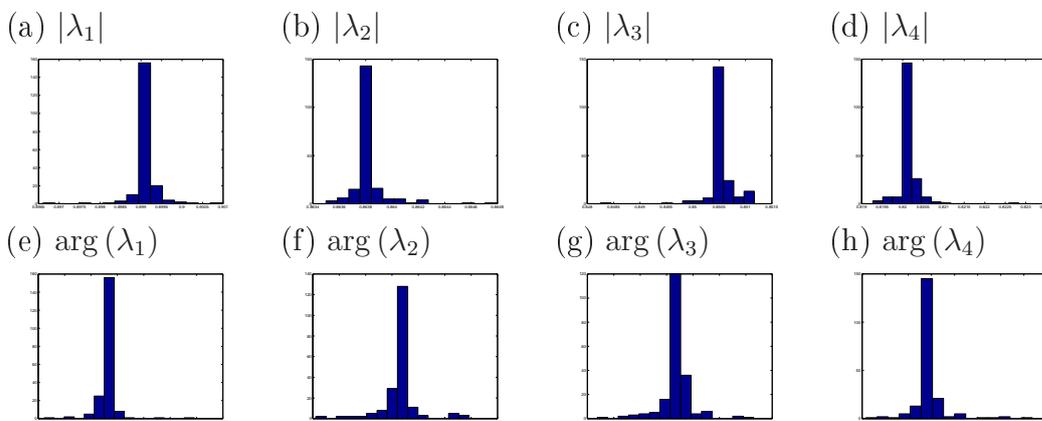


Figure 5.10: Histogram of parameter estimates

5.4 Discussions

We have introduced a model for decomposition of a given single-channel EEG time series into components defined by their main frequency, and we have shown how the variances of the dynamical noises driving these components can be made time-dependent by generalizing the concept of GARCH modelling to the situation of state-space modelling. As a result, changes of the distribution of power over the main spectral bands of the EEG can be traced over time.

Note that by choosing the GARCH approach for describing non-stationarity of variance we obtain a method which remains suitable for real-time monitoring, in contrast to approaches which describe non-stationarity retrospectively by fitting explicitly time-dependent functions to the non-stationary parameters; the GARCH model is fully compatible with predictive modelling, since it requires only information from the past.

Once a suitable model structure has been identified (with respect to the number of components in state space and the GARCH model orders u and v) and a corresponding set of appropriate parameters has been identified through maximum-likelihood, the Kalman filter can be applied very efficiently to new data without the need of any further time-demanding computations. This enables the application of this technique to real-time monitoring of patients during surgery, e.g. it would be possible to monitor the depth of anaesthesia quantitatively by the time-varying set of variances of the relevant frequency bands. Also applications to other kinds of data arising in the neuroscience are conceivable.

Chapter 6

Application 2 - Neural Activity of a Coma Patient

6.1 Introduction

Coma is an extended period of unconsciousness from which a person cannot be aroused with painful stimuli. It can be caused by severe head injury, seizure or metabolic problem. The outcome of a coma ranges from full recovery to death. Coma patients may remain in a persistent vegetative state, in which breathing, normal blood pressure, digesting and eliminating foods continues without the patient's awareness. The vegetative state can last for years or decades. Whether a patient dies, or recovers to moderate disability or full recovery, depends upon the cause of the coma and the type and extent of the brain damage.

A person in a coma does not experience reactivity or perceptivity (Young et al., 1997), where reactivity refers to the inborn functions of the brain and perceptivity refers to responses of the nervous system to learned stimuli. There is completely no clue to know a coma patient's condition from his physical appearance. Heartbeat is one clue to know the metabolic condition of the patient, but neuroscience data, such as EEG data, can give some hints to know more about the coma patient.

We apply the same technique used in chapter 5, in which we have a linear state space model, the compartment model, with a GARCH driving noise to decompose the EEG data into several components according to their frequency bands, allow the driving noise having a non-homogeneous variance. We use this model to depict how is the EEG changing in responding to stimuli. When we apply the model on-site we can be able to conclude whether the stimuli are effective and how strong the response it makes. This is a useful information to the medical doctors so that they would know on which kind of stimuli they should focus.

6.2 Data

The EEG data of the coma patient was obtained from Dr Roy John and Dr Leslie Prichep of New York University. The size of the data set is 135,168 time points by 24 channels. The sampling rate is 200 Hz and therefore the data covers 675.84s, about 11 minutes continuously.

The data is measured when the patient was receiving verbal stimulation. The stimulation, for example, telling him about his own history, asking him to think of a favorite song, or asking him to try moving his finger, is applied at a random sequence. Under these stimuli, the EEG data shows non-stationary pattern. Although we are not ready to identify to which non-stationary pattern what the stimulus was, we are able to show that the data is a useful medium for the coma patient to communicate with the medical doctors.

We choose from their data the first channel (Fp1) and the first 20,200 points for analysis. The data covers about 100 s. In figure 6.1 (a) we show the time series data and the segmentwise FFT spectrum at the side view. The time series data is plotted on the time in seconds. On the bottom right panel, it is a top view of the spectrum, of which the vertical axis is frequency axis in Hz and the horizontal axis is time axis in second. On the bottom left panel it is a side view of the spectrum which provide an image of the varying in time.

We see there is a strong low frequency in the data, especially we find there is an oscillation in every 4 s. Also we see a trough at around 40 Hz which shows some strong phenomenon in our statistical analysis.

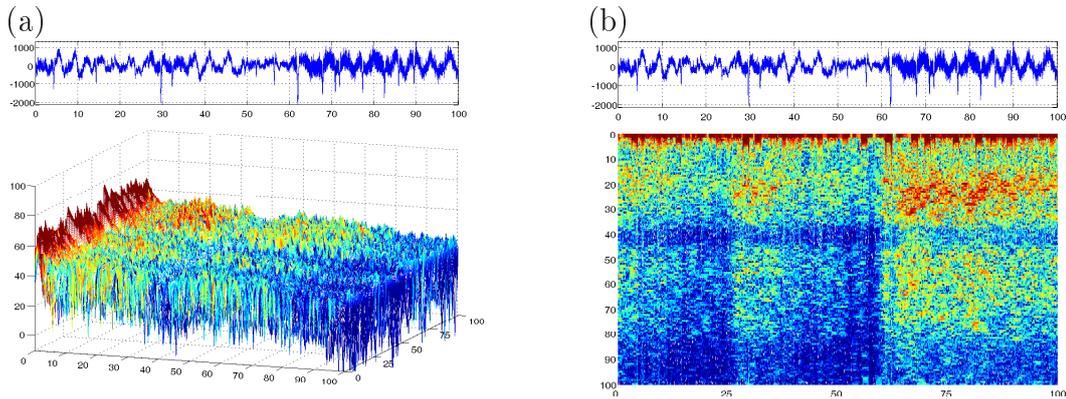


Figure 6.1: Time series plot of data, and time varying FFT spectrum of side view and top view.

The data is showing a significant low oscillation at a period of about 3 s or 4 s. This oscillation cause a sharp peak at the low frequency region in the time varying spectrum. Other than the low frequency oscillation, there is a strong activity between 15 Hz to 35 Hz, the so-called beta frequency, and another activity between 50 Hz to 80 Hz, the gamma frequency.

Within the 100 seconds of data selected, there are 4 significant changes

occurred at around 15 s, 26 s, 36 s and 58 s. We see a similar spectrum shape in 0-15 s and 26-36 s, and we see that both the beta frequency and the gamma frequency drops in intervals 15-26 s and 36-58 s. Also, after 58 s the spectrum tends to rise up at the 20-30 Hz, and a high gamma frequency range (50-70 Hz) evolves.

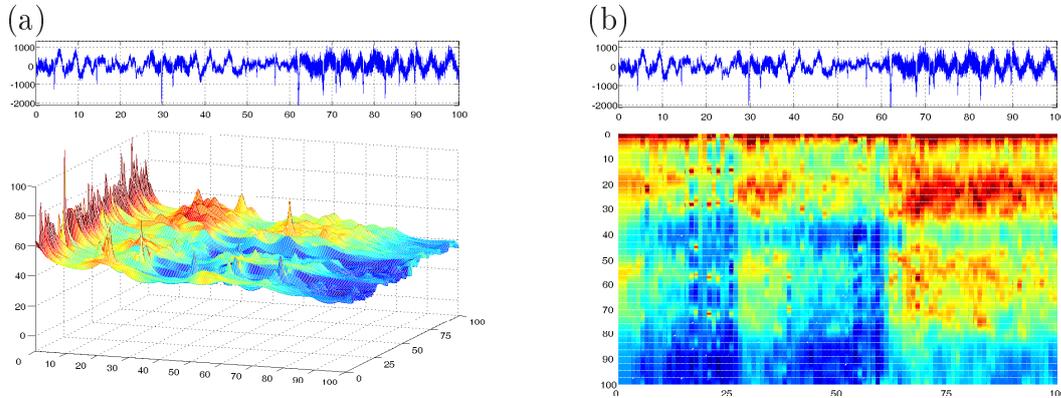


Figure 6.2: Time series plot of data, and time varying best AR spectrum of side view and top view.

In figure 6.2 we show again the time series plot of the data. Instead of an FFT spectrum we show a segment-wise best AR spectrum, of which the side view on the bottom left panel and the top view on the bottom right panel. As we have defined in chapter 3, the segment-wise best AR spectrum is constructed by fitting an AR model for each segment of 200 data points, and plot the theoretical spectral density of each AR model and put them altogether. We can see the spectrum is smoother than the former counterpart. A beta frequency range (15-30 Hz) after 58 s is more clearly seen. The best AR spectrum even shows sharper peaks at the low frequency around 0 Hz.

6.3 Analysis

We try modelling this data by compartment-GARCH model. From figure 6.2 we find 5 major peaks in the spectrum, they are 0 Hz, 15 Hz, 30 Hz, 55 Hz and 70 Hz. One reasonable guess would be a compartment (4,1) model, with 2 AR(2) components going into the alpha-and-beta region and another 2 AR(2) components going into the gamma region, and an AR(1) component for capturing the low frequency oscillation.

Using the optimization procedure, we obtain the compartment-GARCH model consisted of 5 components, specified by 0 Hz, 11 Hz, 30 Hz, 58 Hz and 74 Hz.

In figure 6.3, on the left hand side we show the state decomposition of the data; on the right hand side we show the corresponding driving noise variance.

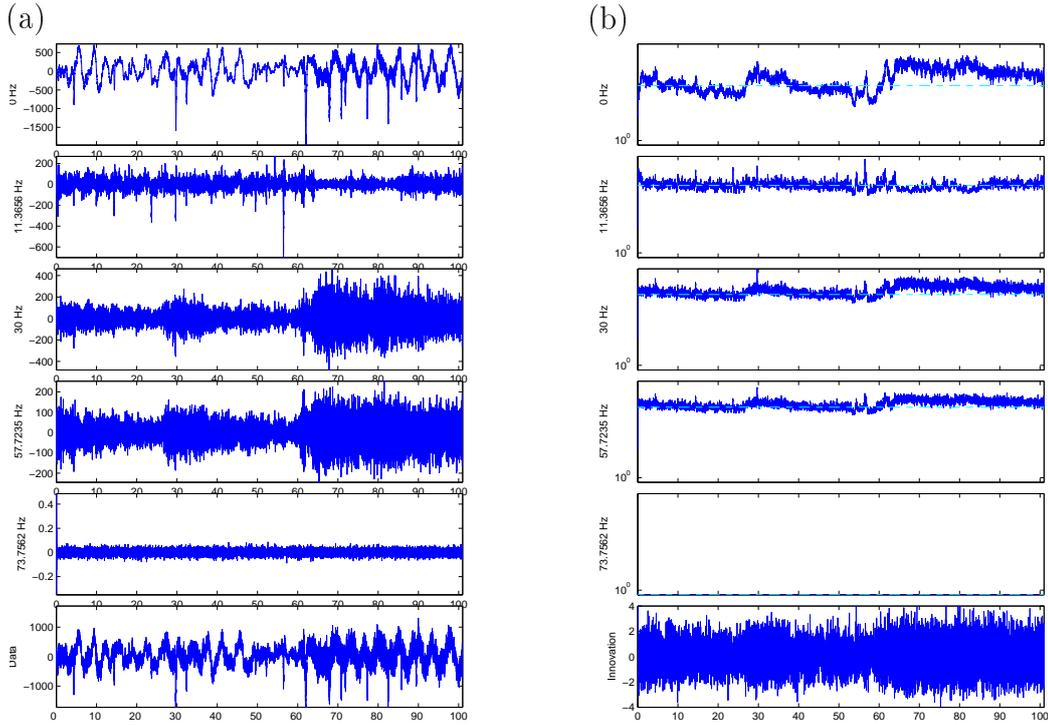


Figure 6.3: State decomposition and its corresponding driving noise variance.

We can see that a low frequency oscillation, about 3 s of period, goes into the 0 Hz component which has the widest amplitude. Other than this 3 s-period oscillation there are several spikes in the data and they go into the 0 Hz component and the 11 Hz component. and the remaining oscillation are decomposed into the other components. We can see that the 0 Hz, 30 Hz and 57.7 Hz components have the up and down transitional pattern as shown in the FFT time vary spectrum and the best AR spectrum, while the 11.4 Hz and the 73.6 Hz have relatively homogeneous movement.

In figure 6.4 we show the time series plot of the data and the time varying state space GARCH spectrum of top view on the bottom right panel. And in figure 6.5 we make the same plot by taking the 0 Hz component away. We can see the spectrum of the fitted model imitates the spectrum of the data with sharp peak at the 0 Hz frequency, some sharp peaks around the beta region and around the gamma region, and a pass around 40 Hz from 0 s to 62 s.

In figure 6.6 we show the best AR spectrum of innovations (or residuals) and in figure 6.7 we show the best AR spectrum of normalized innovations. Since we assume the variance is time varying so the innovations is good to compare with the original data while the normalized innovations is better for diagnostic checking.

The innovations is inhomogeneous and showing especially a low power around 40 Hz in the first 60 s and a high power around 20 Hz and 80 Hz af-

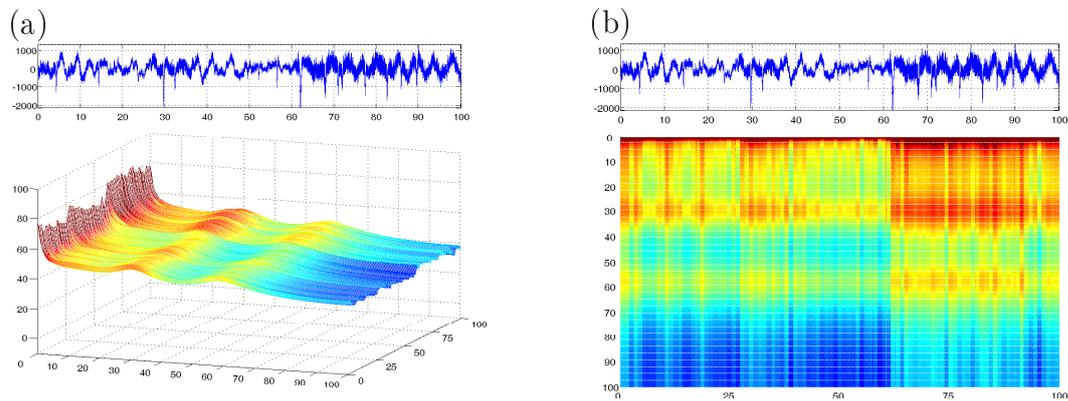


Figure 6.4: Time series plot of data, and time varying state space GARCH spectrum of side view and top view.

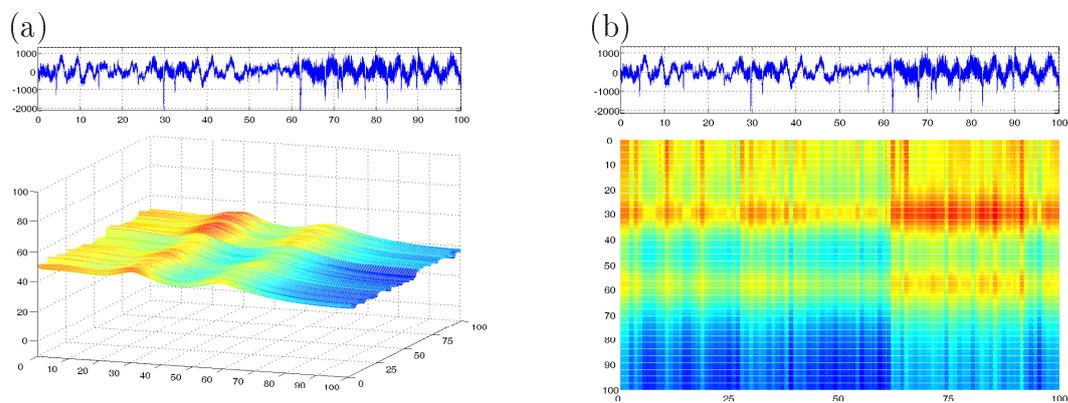


Figure 6.5: Time series plot of data, and time varying de-trended state space GARCH spectrum of side view and top view.

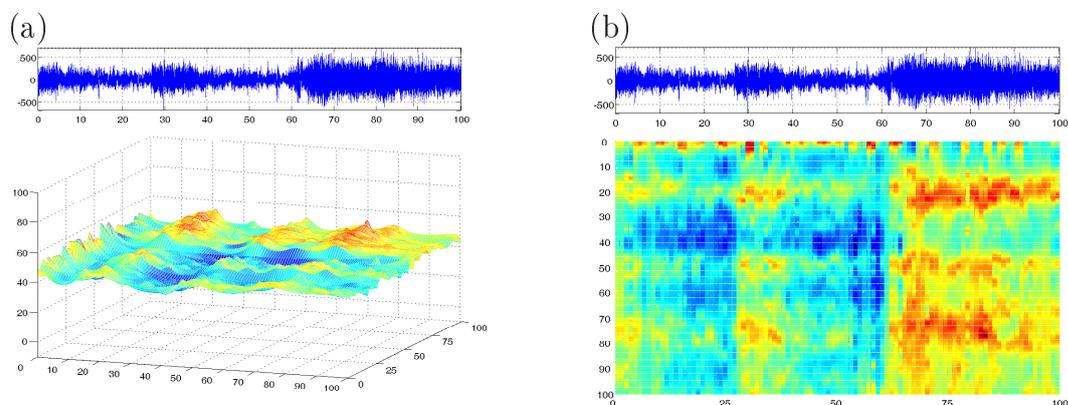


Figure 6.6: Time series plot of innovations, and time varying best AR spectrum of side view and top view.

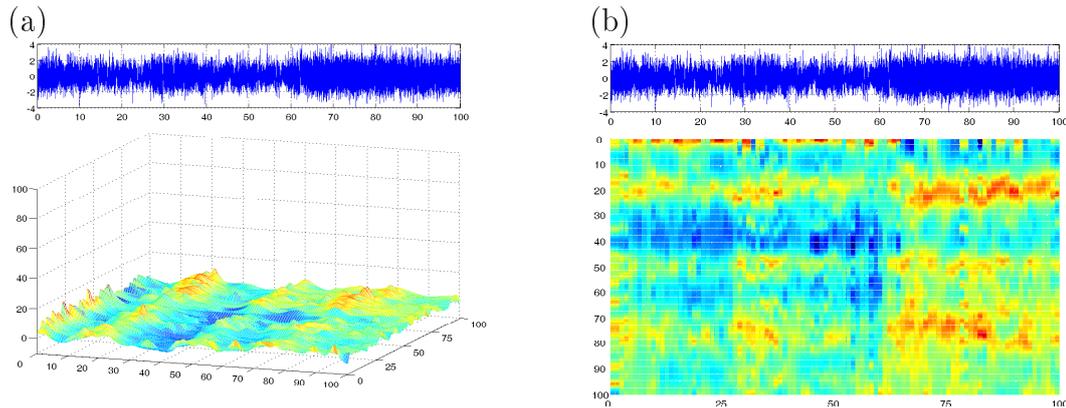


Figure 6.7: Time series plot of normalized innovations, and time varying best AR spectrum of side view and top view.

ter 70 s. These phenomenon are something which could not be managed by a compartment model solely. By applying the compartment-GARCH model and normalized the innovation, the spectrum of normalized innovations is obtained. The spectrum is flat which means that the normalized innovation is white. However, there is still a trough remaining around 40 Hz in the beginning 60 s which could not be handled by the compartment model with only AR(1) and AR(2) compartment. In this issue we need to have a generalized compartment model with Akaike's transposed transitional matrix to allow an MA(2) or even a higher MA component in the model.

6.4 Discussions

We have proposed a new tool for quantitative description of non-stationarity in EEG time series. We have analyzed the single channel of an EEG data of a coma patient. We applied the compartment model again for decomposing the selected time series into components. We also applied the compartment-GARCH model to the system noise variance, and able to see how the variance changes in the same way as the spectrum of data in each segment.

EEG time series are usually recorded not from just one electrode, but from a set of electrodes covering the whole scalp; in principle, the method could be applied to one or two channels of the data, but it would be desirable to have a modelling approach capable of building a single common model from all available channels simultaneously; thereby also the spatial information contained in the positions of electrodes could be incorporated. The generalization of the method to this case will be the subject of future work.

Chapter 7

Application 3 - Causality in Human Head Movement

7.1 Introduction

In this chapter, we are interested in: “Who is the leader in a discussion?” Two persons were sitting face to face. Each person was wearing a cap on which sensors are placed. Their head movement is immediately recorded by the detectors surrounding them. Our target is to study the causal effect of the two persons by applying statistical analysis to the head movement data. A multivariate autoregressive model is used to model the data, and a GARCH model is used to model the heteroscedastic residuals. We construct an NCR plot using the formula in chapter 3 and see the time varying causality.

7.2 Data

In figure 7.1 we show photos of the set up of the experiment. Two subjects are sitting face to face. Sensors were put on their caps in order to locate the position of their heads. Eight tripods around them containing detectors record the position of the caps and inscribed into a computer. The head movement is immediately visualized on a computer screen as shown in the bottom right panel of the figure.

In the experiment, the two subjects were looking at each other in the beginning. A commander standing outside is giving command. The first subject is listening to the commander to turn his head to look at the blue object. The second subject was watching and noticing the first subject, and turned his head to look at the object also to acted as he was following the first subject on a sudden distraction. After a short while the commander gave another command and then the two subjects turned their heads to the original position to look at each other again. This procedure was performed several times. The

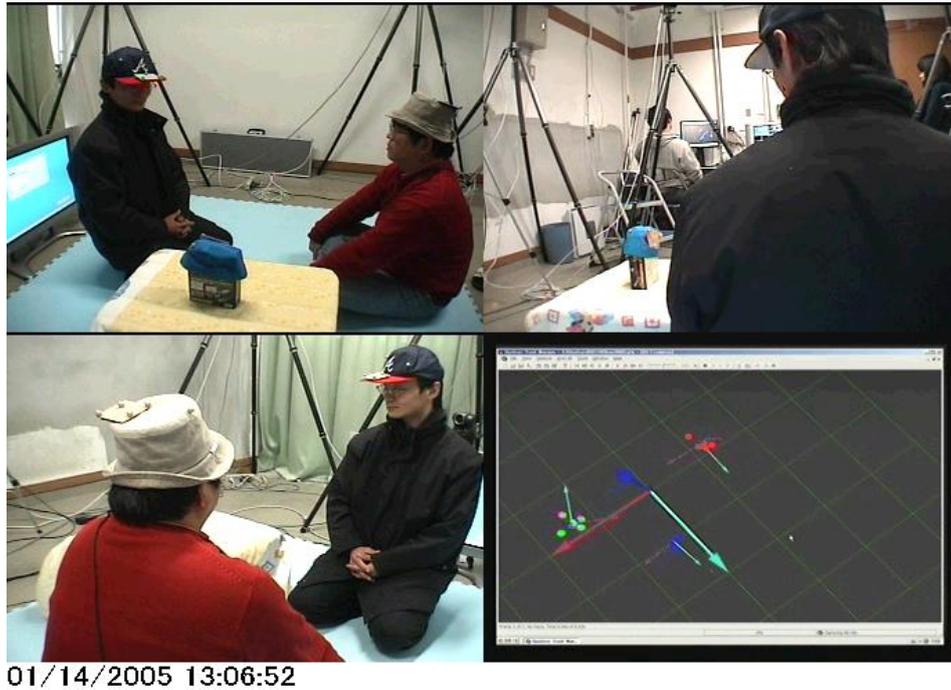


Figure 7.1: A photo for illustrating the set up of the experiment.

time interval between commands are different.

On each cap, a set of 5 sensors locates head position of one subject in a three-dimensional space, as well as estimates angular movement (pitch, roll and yaw) in the three-dimensional space. So, there are six variables per cap, including three variables measuring the position, and three variables measuring the angular movement.

The data was sampled at a rate of 60 Hz. In our analysis, we choose 2000 points (751 to 2750) from the data, covering about 33 s for our analysis. In figure 7.2 we show the data of the two subjects and the object. In each plot there are 3 time series which are the angular movements, pitch(blue), roll(green) and yaw(red) records of each of the two subjects and the object. We put concentration on the causality of yawing movement of the subjects. We plot the yawing of the two subjects in figure 7.3. The blue line belongs to the subject who listens to command and the green line belongs to the subject who follows the first subject. Note that the blue line starts to change before the green line.

7.3 Analysis

We start our analysis from spectral analysis. In figure 7.4, we show power spectral density of the 2 time series of figure 7.3. As we can see the time series

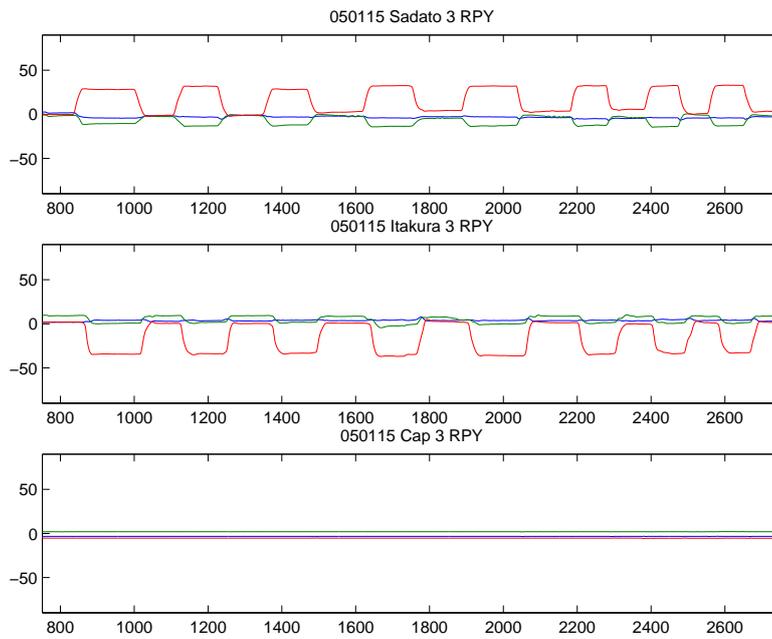


Figure 7.2: Time series of the angular movement of 2 subject and the object.

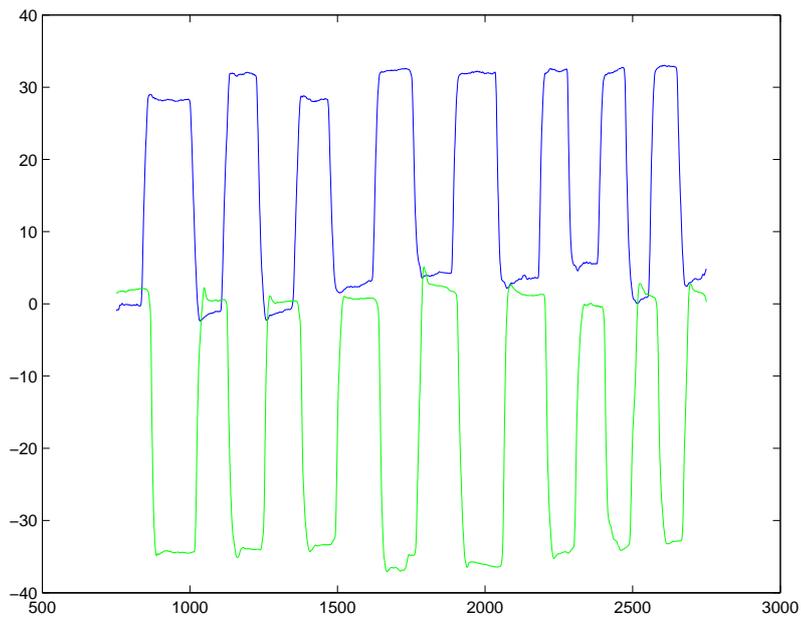


Figure 7.3: Time series of yawing of the 2 subjects

are non-stationary, there is a peak at the 0 Hz frequency.

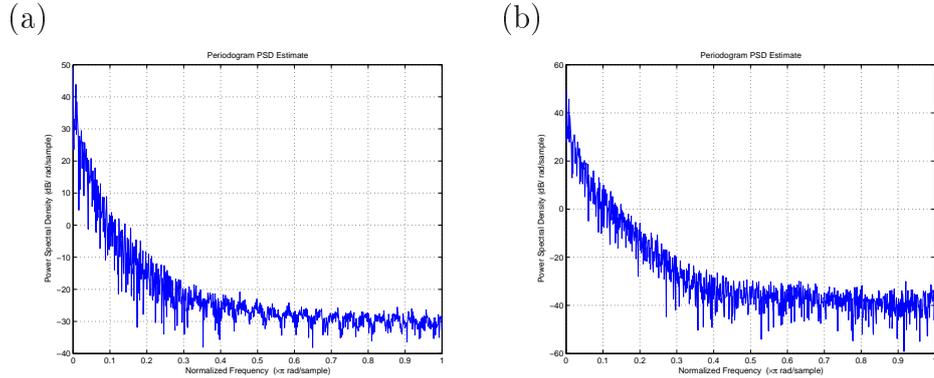


Figure 7.4: Spectrum of the time series of yawing of the 2 subjects.

We fit a bivariate AR model, equation 3.4 of different order to this data. We let $x_t^{(1)}$ be the time series of subject 1 and $x_t^{(2)}$ be the time series of subject 2. By applying Akaike's NCR introduced in chapter 3, we can see the causality of these 2 time series.

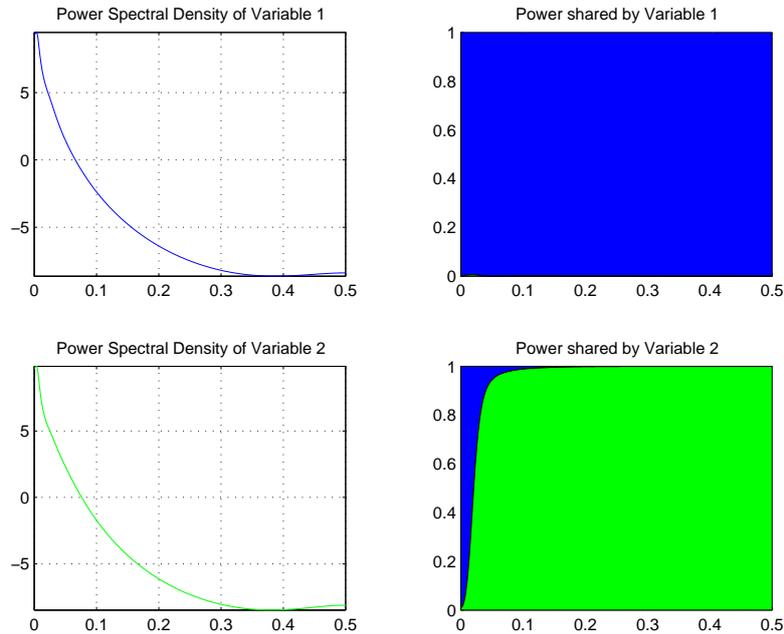


Figure 7.5: AR(5) model spectrum and NCR

In figure 7.5 we show the spectrum and NCR of the 2 time series of AR(5) model. The 2 plots on the left column are respectively the power spectrum of $x_t^{(1)}$ (blue) and $x_t^{(2)}$ (green). In these 2 plots the horizontal axis is the frequency

axis and the vertical axis is power. Since the sampling rate is 60 Hz, the 0.5 on the frequency axis corresponds to 30 Hz. The 2 plots on the right column are NCR. Again the horizontal axis is the frequency axis and the vertical axis is the ratio in percentage. The full blue-color on the top right panel means that the noise contributes from subject 1 to himself is almost 100%, while on the bottom right panel the blue color is occupying the region around 0 Hz meaning that the noise from subject 1 contributes some effect to subject 2, which is equivalent to say there is a causality from subject 1 to subject 2.

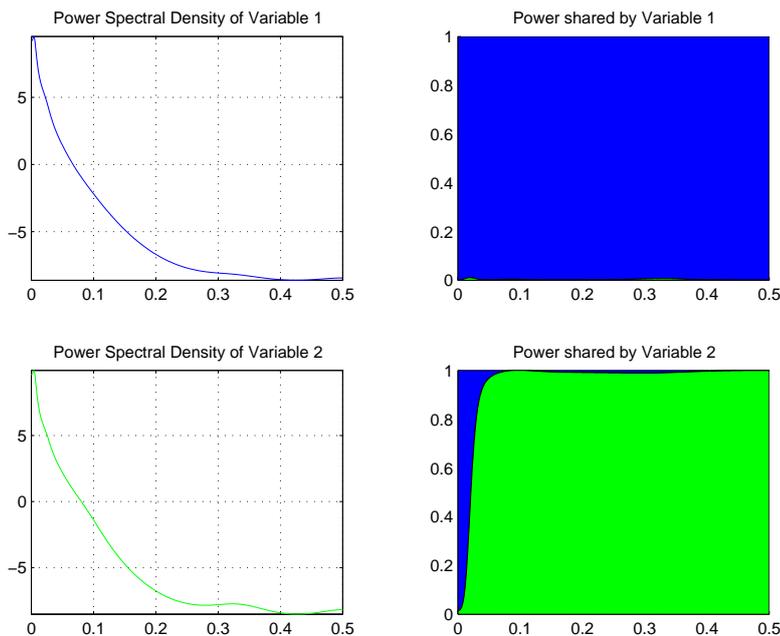


Figure 7.6: AR(8) model spectrum and NCR

In figure 7.6 we show the result when we use an AR(8) model. We can see that the result is almost the same as figure 7.5. It again shows that the subject 2 receives effects from subject 1 but not gives out any effect.

In figure 7.7 we show the result for another 4 higher order AR models. The result are coherent. We summarize the log likelihood and AIC of different model order in table 7.1. From the table we find that AR(36) gives the minimum AIC value.

It is ambiguous whether causality is confirmed upon looking at the coefficients in the AR model only. We see that by using Akaike's NCR approach, when the coefficients which correspond from one variable to another variable are zero, then there is no causal effect, however, the converse is not true, and is usually misunderstood. For example, we give here the numerical result of AR(8) model. We see that the coefficient from $x_{t-1}^{(2)}$ to $x_t^{(1)}$ is -0.0071, which

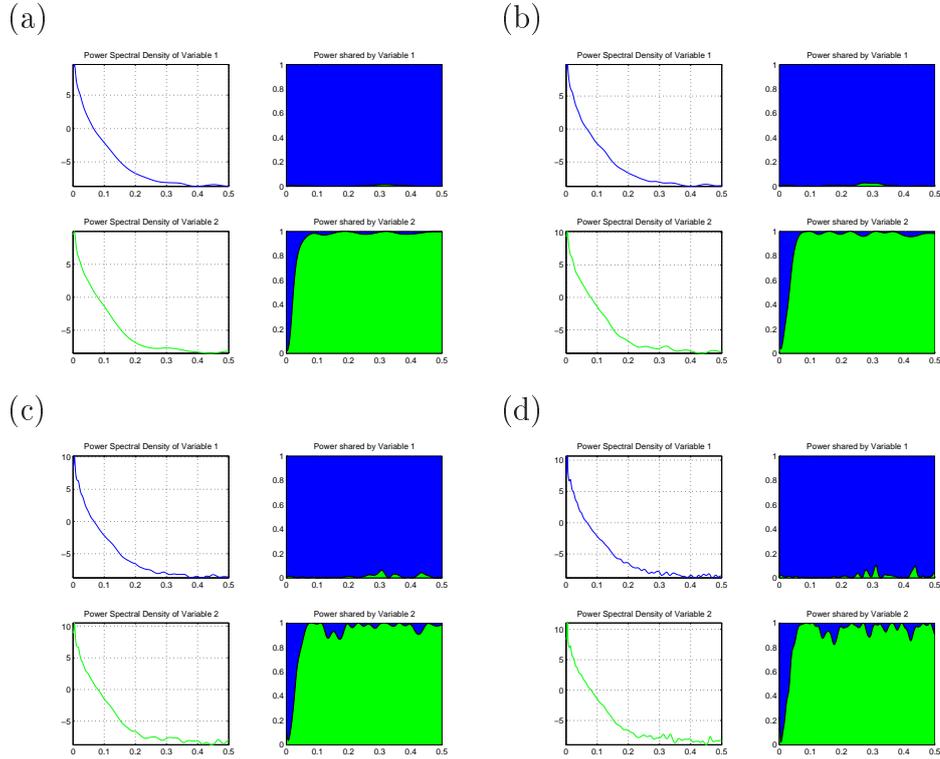


Figure 7.7: AR model spectrum and NCR of order 13, 21, 34 and 55

order	log-lik	AIC	order	log-lik	AIC	order	log-lik	AIC
1	703.6622	-1403.3243	16	8453.9826	-16843.9651	31	8496.1998	-16868.3996
2	6934.1117	-13860.2233	17	8460.4907	-16852.9814	32	8496.9381	-16865.8762
3	8207.4137	-16402.8274	18	8462.0264	-16852.0527	33	8500.6662	-16869.3325
4	8281.3433	-16546.6865	19	8466.0744	-16856.1488	34	8507.1949	-16878.3898
5	8341.9771	-16663.9541	20	8472.7998	-16865.5996	35	8508.6964	-16877.3928
6	8390.8306	-16757.6612	21	8475.3344	-16866.6687	36	8511.6727	-16879.3455
7	8398.3175	-16768.6349	22	8479.1146	-16870.2292	37	8513.4012	-16878.8025
8	8422.9423	-16813.8845	23	8479.6546	-16867.3092	38	8514.9970	-16877.9940
9	8424.7884	-16813.5767	24	8482.3779	-16868.7558	39	8515.3474	-16874.6948
10	8425.6215	-16811.2430	25	8483.0701	-16866.1403	40	8517.1184	-16874.2369
11	8429.9512	-16815.9024	26	8483.3192	-16862.6384	41	8518.0974	-16872.1947
12	8431.8095	-16815.6191	27	8486.3217	-16864.6434	42	8520.3459	-16872.6918
13	8439.8924	-16827.7849	28	8488.5101	-16865.0201	43	8522.7393	-16873.4786
14	8441.4036	-16826.8072	29	8489.9786	-16863.9572	44	8523.9461	-16871.8921
15	8450.9957	-16841.9915	30	8492.9160	-16865.8319	45	8524.3740	-16868.7480

Table 7.1: AIC of fitted multivariate AR models of different orders.

may be insignificantly different from zero, but still it does not surely imply that there is a causal effect. However when we make use of the NCR technique, we can tell there is no causal effect for sure.

$$\begin{aligned}
 \begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \end{pmatrix} &= \begin{pmatrix} 2.5230 & -0.0071 \\ -0.0221 & 2.4887 \end{pmatrix} \begin{pmatrix} y_{t-1}^{(1)} \\ y_{t-1}^{(2)} \end{pmatrix} + \begin{pmatrix} -1.7616 & 0.0139 \\ 0.0083 & -1.6884 \end{pmatrix} \begin{pmatrix} y_{t-2}^{(1)} \\ y_{t-2}^{(2)} \end{pmatrix} \\
 &+ \begin{pmatrix} -0.1598 & 0.0132 \\ 0.1095 & -0.2209 \end{pmatrix} \begin{pmatrix} y_{t-3}^{(1)} \\ y_{t-3}^{(2)} \end{pmatrix} + \begin{pmatrix} 0.3741 & -0.0676 \\ -0.1207 & 0.2654 \end{pmatrix} \begin{pmatrix} y_{t-4}^{(1)} \\ y_{t-4}^{(2)} \end{pmatrix} \\
 &+ \begin{pmatrix} 0.1009 & 0.0533 \\ -0.0829 & 0.2604 \end{pmatrix} \begin{pmatrix} y_{t-5}^{(1)} \\ y_{t-5}^{(2)} \end{pmatrix} + \begin{pmatrix} 0.0205 & 0.0564 \\ 0.1860 & 0.1245 \end{pmatrix} \begin{pmatrix} y_{t-6}^{(1)} \\ y_{t-6}^{(2)} \end{pmatrix} \\
 &+ \begin{pmatrix} -0.1608 & -0.1069 \\ -0.0744 & -0.3677 \end{pmatrix} \begin{pmatrix} y_{t-7}^{(1)} \\ y_{t-7}^{(2)} \end{pmatrix} + \begin{pmatrix} 0.0632 & 0.0449 \\ -0.0093 & 0.1334 \end{pmatrix} \begin{pmatrix} y_{t-8}^{(1)} \\ y_{t-8}^{(2)} \end{pmatrix} \\
 &+ \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix}
 \end{aligned}$$

$$\begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.0045 & 0 \\ 0 & 0.0053 \end{pmatrix} \right)$$

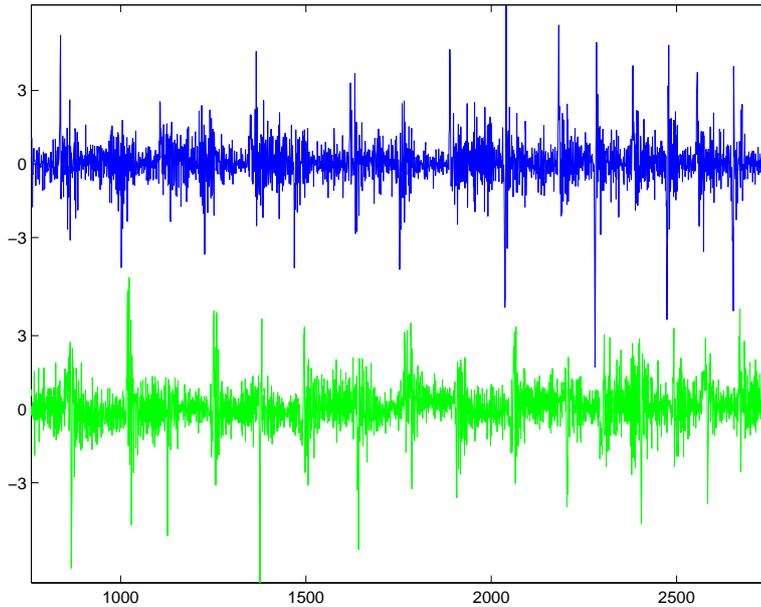


Figure 7.8: AR(8) model residual

In figure 7.8 we show the residual from a fitted AR(8) model. We can see the noise is driven by non-homogeneous variance under this AR model assumption. In the experiment, the 2 subjects are turning their heads to look at the object. At the moments when their heads starting to turn or stopping, the noises are especially larger.

We then apply a GARCH to the residuals. For each series of residual we fit a GARCH model. The normalized residuals after applying GARCH models is shown in the subplot on the top left corner of figure 7.9, showing the noise is further whiten.

$$\begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_t^{(1)^2} & 0 \\ 0 & \sigma_t^{(2)^2} \end{pmatrix} \right)$$

$$\sigma_t^{(1)^2} = 0.0038 + 0.4705\eta_t^{(1)^2} + 0.5295\sigma_{t-1}^{(1)^2}$$

$$\sigma_t^{(2)^2} = 0.0042 + 0.2653\eta_t^{(2)^2} + 0.7087\sigma_{t-1}^{(2)^2}$$

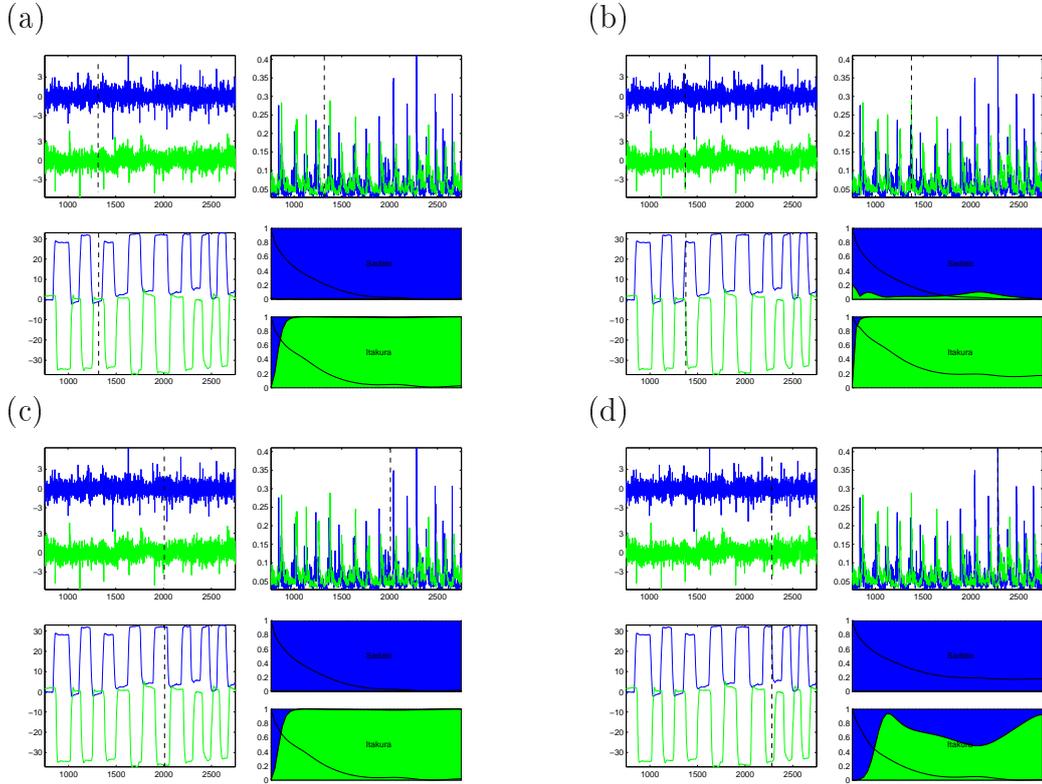


Figure 7.9: Time varying NCR at 4 different time points.

In figure 7.9 we show 16 subplots in 4 blocks. The 4 subplots in each block are GARCH residuals (top left), GARCH variances (top right), data (bottom left) and NCR (bottom right). When we apply the GARCH model to the residual, the noise variance becomes time varying. The time varying variance

changes the NCR at different time. Therefore, each block shows the same GARCH residual, the same GARCH variances and same data but different NCR.

We found that base on the AR(8)+GARCH(1,1) model there is a persistent causal effect coming from the first subject to the second subject at the 0 Hz oscillation. In the second block of figure 7.9 when the noise variance of the second subject is large, the NCR from the second subject to the first subject is negligibly small, but that from the first subject to the second subject is still large at the 0 Hz frequency. In the last block of the figure we can see even a larger proportion of NCR going from the first subject to the second subject. It is when the noise variance of the first subject is of the largest value.

7.4 Discussions

In this chapter we make use a very simple physical simulated data to look into the causality. We first applied multivariate AR model, a handy and fast modelling method. With different order of multivariate AR models we had the same causality conclusion. Based on the NCR causality theory, the first subject, who turned his head without noticing the second subject, has a strong causal effect on the second subject, who was distracted by the first subject. The NCR causality shows that causality happened at 0 Hz frequency but not any other frequency. Since this data does not contain any characteristics of frequency motion, the NCR causality is mainly explained in the low frequency region.

Moreover, this example strongly proved that non-zero coefficients do not imply causality when AR model is higher than order one. Nonetheless we choose NCR causality over Granger's causality because NCR causality takes the noise variance into consideration.

Due to the non-stationary variance of residuals we enforce the AR model with a GARCH model. The GARCH model allows the noise variance time-varying freedom to fix the heteroscedasticity. Time varying variance consequently brings NCR changing over time. A time varying NCR causality plot would be the only suitable way to explain any heteroscedastic time series data.

Chapter 8

Application 4 - Study of Synchronous Rotation of Fingers

8.1 Introduction

We will continue to study causality in this chapter with a data set concerning synchronous rotation of fingers. In this experiment, we investigate the causal relations of the two index fingers when they are both in a motion of drawing circles in the horizontal plane.

Continuous circle drawing task has been used extensively to study bimanual coordination. Finding the mechanisms of coordination of bimanual motion, we can learn how the hands interact while coordinate to move independently. In this paper, we perform statistical modelling and apply the Akaike NCR theory to explain causality of the hands interaction.

In the continuous circle drawing experiment, drawing circle in a symmetrical mode is easy no matter how high the speed is, but parallel mode is easy only when the speed is under a threshold (Kelso, 1984). When the speed increases the parallel mode of circle drawing may change into the symmetrical mode involuntarily (Kelso, 1984; Semjen et al., 1995). Moving the hands in mirror symmetry is a spontaneous tendency due to perceptual and spatial symmetry but not homologous muscles co-activation (Mechsner et al., 1971).

Various models were established in literature. Haken et al. (1985) use a second order differential equation model to capture the varying phase delay, which was modified by Bingham (2004) and turned to his phase driven model. Cattaert et al. (1999) advocated their neural cross-talk model assuming that bimanual circular motion are generated from two orthogonal oscillations coupled with a phase delay, with a proportion of force command sent to one hand dispatched to the other hand as a mirror image, and this theory was advocated by Viviani et al. (1998). A common characteristics of the models is their focus on the phase delay. The phase delay can be interpreted as the directed coherence or causal effect between the variables. In other words, a variable causes

other variables at a scale different to it was caused by the other variables.

A multivariate AR model can explain the temporal and the spatial characteristics of the data, however, it may not be suitable to explain the phase delay changing over time unless we make the AR coefficients as functions in time. Instead of multivariate AR model, we apply a linear state space model. We shall compare a result using a state space model with switching observation matrix to the AR model, and discuss how the state space model in explaining NCR causality.

8.2 Data

This data contains measurement of the bimanual movement of index fingers of a healthy subject in listening to an audio signal of 1 Hz, at the same time his index fingers are performing circle drawing at a frequency of 1 Hz in the horizontal plane. The subject cannot see his fingers during the experiment. We choose from the data a segment of 20 s, when the right-handed subject was performing a symmetrical rotation: the left index finger is rotating anticlockwise and the right index finger is rotating clockwise. In figure 8.1 we show the raw data plotting on an axis of angular displacement in radian against the time in second. In 20 s each of the fingers drew 20 circles. The green line represents the left index finger while the red line represents the other one. In this experiment the sampling rate is 1000 Hz and the measurement resolution is 1 degree.

8.2.1 Data Preprocessing

There are two problems in the data. The first problem is monotonicity. When we do time series analysis, especially when we use linear time series models, the data should fulfill the first order stationarity assumption, ie data need to fluctuate around a mean value. The characteristics of non-stationarity in mean is easily reflected in a heavy tail-off in ACF and a sharp spike at the 0 Hz in spectrum. In this situation, the overwhelming tail-off or spike may blind us from seeing other characteristics of the data, at the same time, no time series model requiring the assumption of stationarity would be suitable. To solve the problem we need a suitable pre-processing before we apply our statistical modelling. There are several possibility, including differencing, detrending and so on. Differencing can be easily done by subtracting pairs of consecutive data. For this data, we try the first order and the second order differencing method to look for the solution to monotonicity. However, it leads us to the second problem.

In figure 8.2 we show four subplots including two differenced time series and two FFT spectrum of another segments from the same data set. Subplot

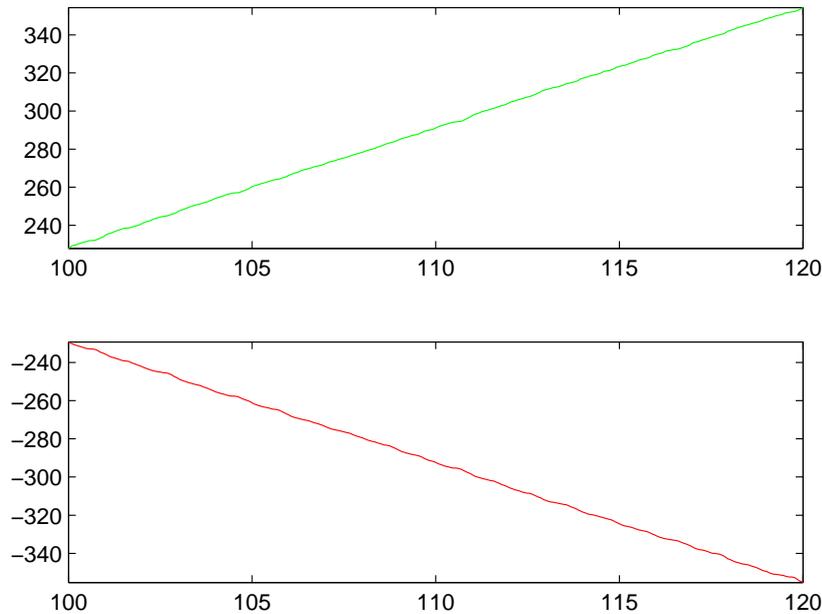


Figure 8.1: Finger rotation data

(a) is the first order differenced time series and subplot (c) is the second order differenced data. The time series plot show very strange pattern as we can see that the first order differenced time series in subplot (a) takes exactly eight values only, and the second order differenced time series gives completely no information about the dynamics of the data.

So, the second problem is poor measurement resolution in the data. The problem comes from an unsuitable setting of experiment, for instance, when a meter is set or is just capable to measure at a limited significant digits, then a large rounding off error in the measurement. To solve the problem, one have to increase the level of accuracy of measurement.

However, we will solve the problem using another method, say, by reducing the time resolution of the data. In figure 8.2(b) and (d) we show respectively an FFT spectrum of the first and the second differenced time series, in which the vertical axis is the log of absolute power and the horizontal axis is frequency ranged from 0 Hz to 500 Hz. Both spectra have a concave shape and a higher power at the high frequency. Unfortunately, these spectra are not useful, without the need of investigate the implication from the shape of the spectra. Since the data is recording bimanual rotation activity at an angular speed of 1 Hz, we are more interested in the dynamics related in the neighborhood of 1 Hz. In other words, we should study this time series on a 0.1 Hz to 10 Hz domain but not 100 Hz or 500 Hz, such a speed is almost meaningless in the

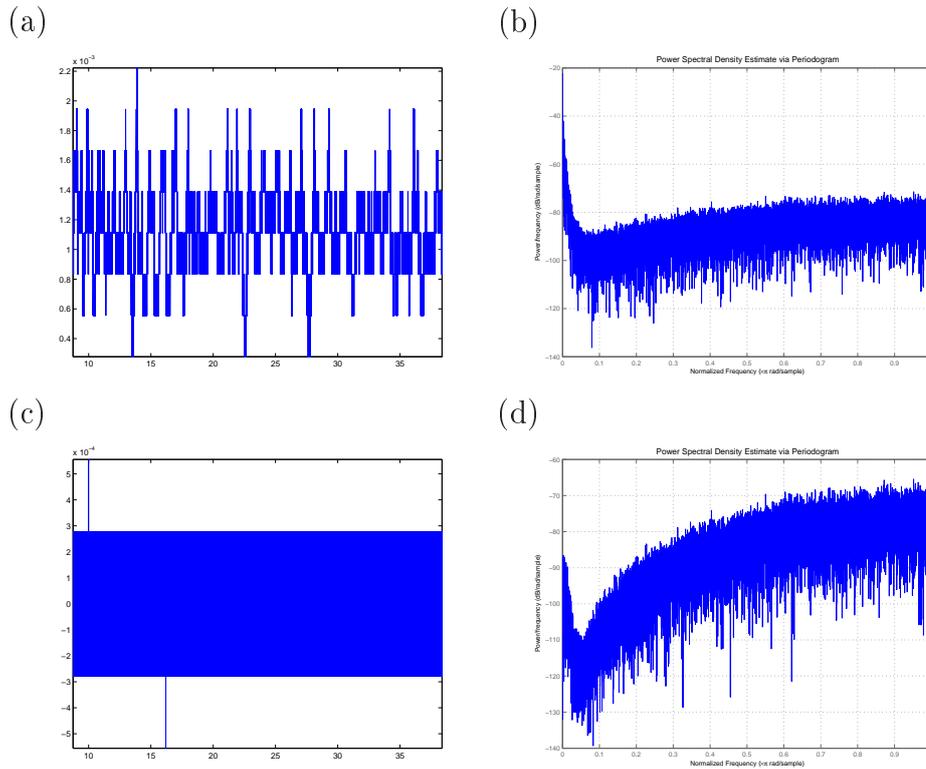


Figure 8.2: First and second difference finger rotation data and their FFT spectrum

physical motion of a human limb.

In our analysis since the fingers are bimanual rotation in a frequency of 1 Hz, we choose to reduce the time resolution of the data in order to analyze it. We perform a sub-sampling procedure on the data at 3 different sub-sampling rates.

In figure 8.3 the 12 plots are the first and the second order difference data and their FFT spectra of a sub-sampled data. In subplots (a)-(d), we have first and second order differenced data and their FFT spectra of 1/10 sub-sampling. In subplots (e)-(h) and in subplots (i)-(l) we have those of 1/20 and 1/50 respectively. The plots diminish into smaller size in order to ease comparison. In subplots (a), (e) and (i), we show the differenced data at three different sub-sampling rate. We can see how they are all so much different from figure 8.2(a). In subplots (c), (g) and (k) we can see the progressive improvement of resolution.

We can see the FFT spectrum of the processed data magnifies the low frequency part of the spectra in the subplot (b) from 8.2(b). Note that in subplot (j), we show the spectrum over the frequency domain of 0 Hz to 50 Hz, which is equivalently showing 1-fifth of subplot (b) and 2-fifth of subplot(f). When we further reduce the time resolution the difference time series are get-

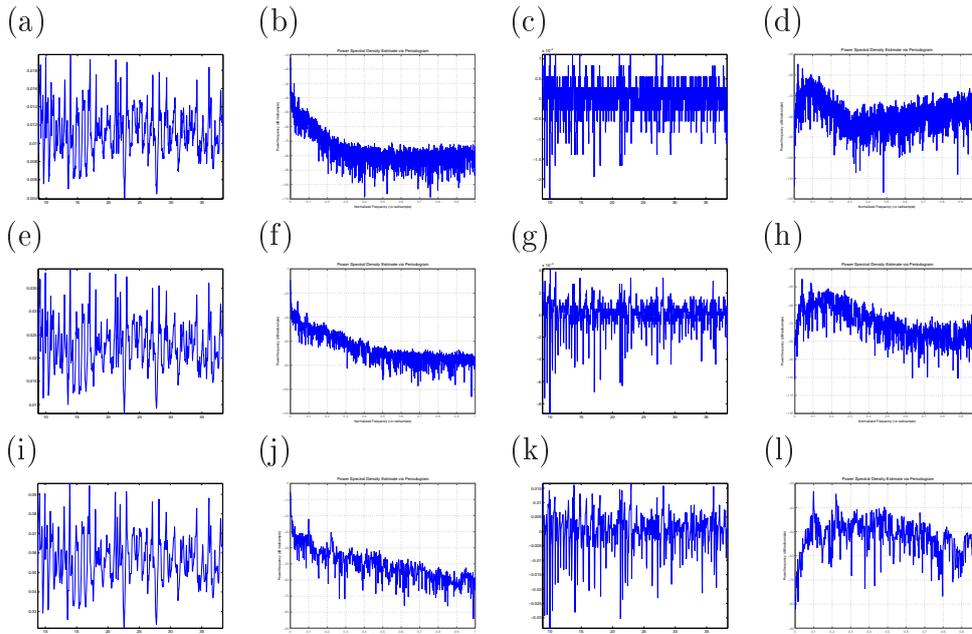


Figure 8.3: Matrix plot of time series and FFT spectra at 3 different sub-sampling rate and 2 different differencing order of finger rotation data

ting smoother and 2 significant peaks come out in the spectrum. In subplot (j) and (l), we can see clearly 2 peaks at 1 Hz and 2 Hz, which are the main characteristics of the time series.

Doing the sub-sampling, we do not lose the information in the data. Instead, we can enlarge the information at the low frequency region to look close into it. As shown in the spectrum of the first order difference and second order difference of the original data, Nyquist frequency is 500 Hz and we can hardly see the 1 Hz angular rotation. When the sub-sampling increase from 1/10 to 1/20 and further 1/50, the 1 Hz frequency can be clearly seen.

The differencing technique is not restricted to be applied one time only. It can be applied to data twice, which depends on the necessity of the data. It is so rare to be applied for 3 times or more, unless we know there are several unit roots existing in the data.

8.3 Analysis

After a differencing and a 1/50 sub-sampling on the raw data, we subtract the average angular speed (0.314) and re-scale then we obtain a working data for our data analysis as shown in figure 8.4. There are 2 variables and 400 samples in the selected data. Again, the green line represents the left index finger while the red line represents the right index finger. The sub-sampling reduces the sampling rate from 1000 Hz to 20 Hz. The horizontal is the time

axis in second. In the first 5 seconds we can see clearly 5 oscillations in both time series, which means there is a significant 1 Hz oscillation. The green line, ie the left finger, shows two to three oscillations with every 1 Hz oscillation, and the red line does show 2 Hz oscillation but rather weak.

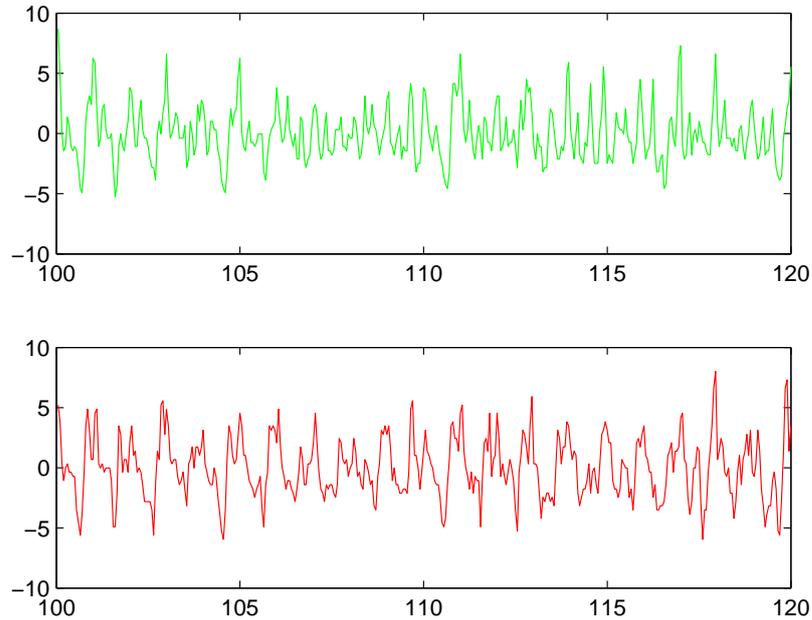


Figure 8.4: Preprocessed finger rotation data

8.3.1 NCR Causality of Multivariate AR Model

We fit multivariate AR model to the data as preliminary analysis. An AR model can be estimated by a computationally efficient algorithm by Whittle (1963). For each AR model from we have a likelihood. In figure 8.5 we show the AIC of the estimated model of order 1 to order 30 and we find that the AR(11) model gives the minimum AIC. So we will concentrate on the AR(11) model and compare it with other models.

By fitting the AR(11) model we obtain an AIC of 2883.1, when we assume the covariance matrix contains non-zero off-diagonal values. The correlation coefficient of the noise variance is $\frac{0.9341}{\sqrt{2.2906 \times 2.3463}} = 0.4029$.

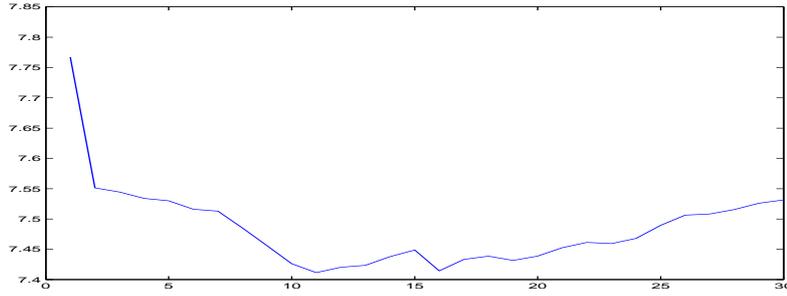


Figure 8.5: AIC of AR models up to order 30.

$$\begin{aligned}
\begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \end{pmatrix} &= \begin{pmatrix} 0.6337 & 0.2332 \\ 0.1420 & 0.6856 \end{pmatrix} \begin{pmatrix} y_{t-1}^{(1)} \\ y_{t-1}^{(2)} \end{pmatrix} + \begin{pmatrix} -0.4004 & -0.0569 \\ -0.1393 & -0.2093 \end{pmatrix} \begin{pmatrix} y_{t-2}^{(1)} \\ y_{t-2}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} -0.0790 & 0.1047 \\ 0.0569 & 0.0044 \end{pmatrix} \begin{pmatrix} y_{t-3}^{(1)} \\ y_{t-3}^{(2)} \end{pmatrix} + \begin{pmatrix} -0.1007 & 0.0289 \\ -0.0252 & -0.0234 \end{pmatrix} \begin{pmatrix} y_{t-4}^{(1)} \\ y_{t-4}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} 0.0182 & 0.1449 \\ -0.0187 & -0.1120 \end{pmatrix} \begin{pmatrix} y_{t-5}^{(1)} \\ y_{t-5}^{(2)} \end{pmatrix} + \begin{pmatrix} -0.1565 & -0.0416 \\ -0.0523 & 0.0798 \end{pmatrix} \begin{pmatrix} y_{t-6}^{(1)} \\ y_{t-6}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} 0.0097 & -0.0230 \\ 0.0381 & -0.0428 \end{pmatrix} \begin{pmatrix} y_{t-7}^{(1)} \\ y_{t-7}^{(2)} \end{pmatrix} + \begin{pmatrix} -0.0722 & 0.0758 \\ -0.0386 & -0.0555 \end{pmatrix} \begin{pmatrix} y_{t-8}^{(1)} \\ y_{t-8}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} -0.0607 & 0.0277 \\ 0.1067 & -0.0987 \end{pmatrix} \begin{pmatrix} y_{t-9}^{(1)} \\ y_{t-9}^{(2)} \end{pmatrix} + \begin{pmatrix} 0.0329 & -0.0285 \\ -0.0458 & -0.1038 \end{pmatrix} \begin{pmatrix} y_{t-10}^{(1)} \\ y_{t-10}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} -0.1470 & 0.0444 \\ -0.0146 & -0.1103 \end{pmatrix} \begin{pmatrix} y_{t-11}^{(1)} \\ y_{t-11}^{(2)} \end{pmatrix} + \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix}
\end{aligned}$$

$$\begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2.2906 & 0.9341 \\ 0.9341 & 2.3463 \end{pmatrix} \right)$$

However, in order to interpret the causal relation by NCR, we fit the model again with a restriction of diagonal covariance matrix. The estimated model is as follow.

$$\begin{aligned}
\begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \end{pmatrix} &= \begin{pmatrix} 0.6776 & 0.2517 \\ 0.1688 & 0.6882 \end{pmatrix} \begin{pmatrix} y_{t-1}^{(1)} \\ y_{t-1}^{(2)} \end{pmatrix} + \begin{pmatrix} -0.4323 & -0.1108 \\ -0.1555 & -0.2421 \end{pmatrix} \begin{pmatrix} y_{t-2}^{(1)} \\ y_{t-2}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} -0.0565 & 0.1523 \\ 0.0725 & 0.0345 \end{pmatrix} \begin{pmatrix} y_{t-3}^{(1)} \\ y_{t-3}^{(2)} \end{pmatrix} + \begin{pmatrix} -0.1265 & 0.0155 \\ -0.0387 & -0.0346 \end{pmatrix} \begin{pmatrix} y_{t-4}^{(1)} \\ y_{t-4}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} 0.0428 & 0.1518 \\ -0.0066 & -0.1166 \end{pmatrix} \begin{pmatrix} y_{t-5}^{(1)} \\ y_{t-5}^{(2)} \end{pmatrix} + \begin{pmatrix} -0.1838 & -0.0468 \\ -0.0636 & 0.0918 \end{pmatrix} \begin{pmatrix} y_{t-6}^{(1)} \\ y_{t-6}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} 0.0307 & -0.0305 \\ 0.0465 & -0.0526 \end{pmatrix} \begin{pmatrix} y_{t-7}^{(1)} \\ y_{t-7}^{(2)} \end{pmatrix} + \begin{pmatrix} -0.0876 & 0.0760 \\ -0.0511 & -0.0594 \end{pmatrix} \begin{pmatrix} y_{t-8}^{(1)} \\ y_{t-8}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} -0.0559 & 0.0376 \\ 0.1088 & -0.0908 \end{pmatrix} \begin{pmatrix} y_{t-9}^{(1)} \\ y_{t-9}^{(2)} \end{pmatrix} + \begin{pmatrix} 0.0332 & -0.0245 \\ -0.0336 & -0.1156 \end{pmatrix} \begin{pmatrix} y_{t-10}^{(1)} \\ y_{t-10}^{(2)} \end{pmatrix} \\
&+ \begin{pmatrix} -0.1374 & 0.0319 \\ -0.0099 & -0.1246 \end{pmatrix} \begin{pmatrix} y_{t-11}^{(1)} \\ y_{t-11}^{(2)} \end{pmatrix} + \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} \\
\begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2.1355 & 0 \\ 0 & 2.3333 \end{pmatrix} \right)
\end{aligned}$$

Then we use equation 3.3 to calculate the model spectra and NCR causality. In figure 8.6(a) we show the FFT spectra (blue) and the theoretical AR spectrum (black) which are plotted together on one plot. The top one is the spectrum for the rotation time series of the left finger and the bottom one is that of right finger. The vertical axis is the log of power intensity and the horizontal axis is the frequency ranged from 0 Hz to 10 Hz. In figure 8.6(b) we show the noise contribution ratio. The green color is the contribution from $\eta_t^{(1)}$ and the red color is the contribution from $\eta_t^{(2)}$. The vertical axis is the ratio from 0 to 1, and the horizontal axis is the frequency axis, again from 0 Hz to 10 Hz.

In figure 8.6(a) we can see from the blue spectra that there are a lot of sharp spikes, especially at 1 Hz for both fingers and 2 Hz for left finger. When we look at the black curves we see that 1 Hz is well captured by the AR model, and a peak at around 3 Hz is also dominant, which may be a harmonic of the 1 Hz rotation, which means the bimanual rotation is strongly characterized by the 1 Hz rotating motion.

By figure 8.6(b) a strong causal relation from the right finger to the left finger can be seen at 1 Hz, and a trace amount of feedback at 3 Hz can be seen but not necessary to be an important causal relations.

In figure 8.7 we show the residuals of the AR(11) model. When we compare it to figure 8.4 we can see that most of the low frequency were taken away, but still some spikes are still remaining in the residuals, especially they are about

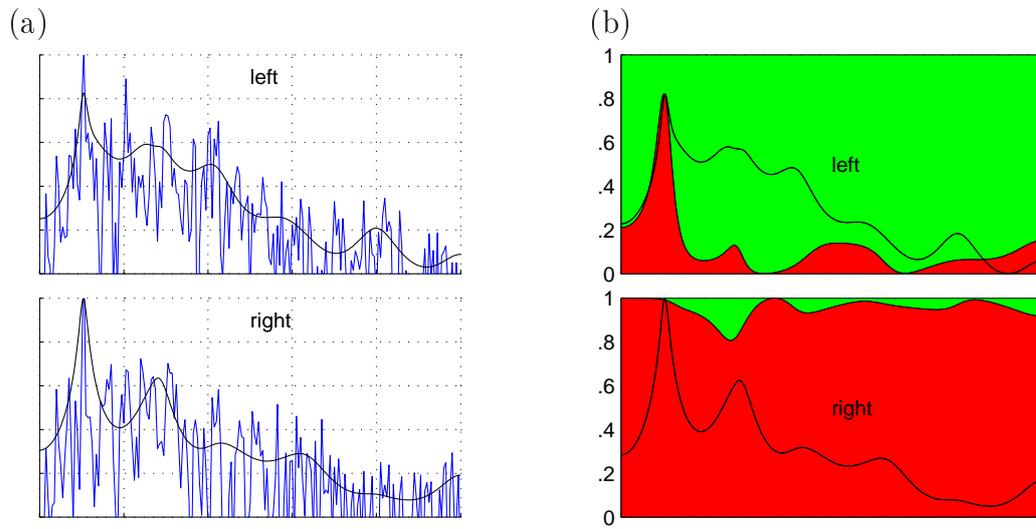


Figure 8.6: Spectra and NCR of a fitted AR(11) model

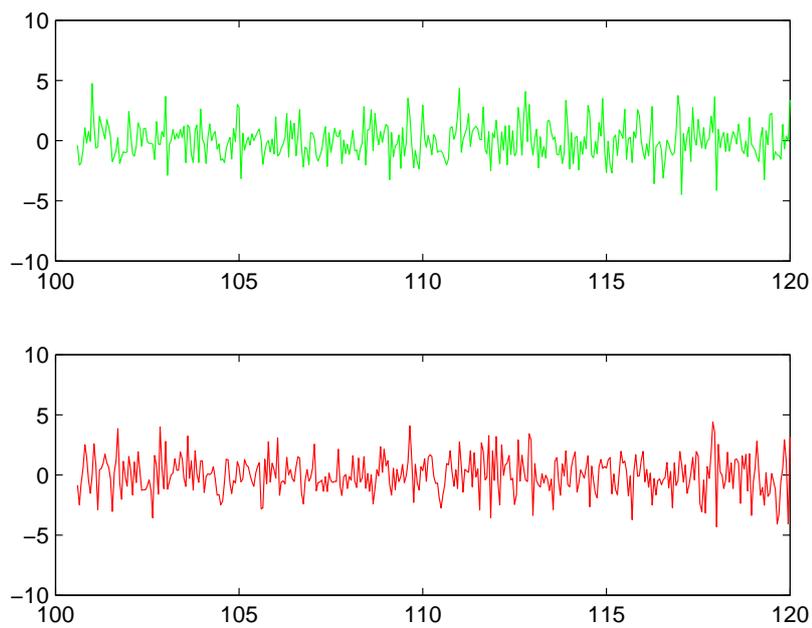


Figure 8.7: Innovation from a fitted AR(11) model

1 s in separation. However, as we have shown in the AIC plot in figure 8.5 that increasing the model order may not be suitable to improve the model fitting significantly in terms of a balance of the number of parameters.

8.3.2 NCR Causality of State Space Model

As an alternative to the multivariate AR model, we perform the analysis through the linear state space model. It has been stated in chapter 2 that any multivariate AR model can be written in the form of state space model. It is not our purpose to transform into a state space model, but we would like to interpret the NCR causality in another way.

We introduce a hidden variable which is neither observed nor be estimated perfectly. The hidden variable is taking the common dynamics of both the left finger and the right finger. The remaining uncommon dynamics can be explained by the individual dynamic of the fingers, which can produce their own causal relations. We use this approach because of two reasons. The first reason is due to the physical meaning, and the second reason is due to the characteristics of the time series based on the fitted AR(11) model.

As we have mentioned in the beginning of this chapter, Cattaert et al. (1999) used their neural cross-talk model to explain the bimanual motion. Similar to their two orthogonal oscillations theory, we assume that there would be three oscillations, of which they are all orthogonal to each other; one oscillation takes the common feature of the two time series and the other two oscillations explaining the individual characteristics of the two time series. The two time series are reconstructed by the three oscillations with a phase delay. Therefore, a common force command, which is sent to both limbs as a mirror image, is explained by this common hidden variable.

In figure 8.8 (a) we show an Argand diagram of the characteristic roots of the fitted AR(11) model. The horizontal and the vertical axes are respectively the real and the complex line. The diagram is symmetric about the horizontal axis, meaning that the roots can be real roots or complex roots of conjugate pair. In addition to the conjugate pairing, there is also a virtual pairing of the roots, as we can see that the 22 roots are closely located two by two. This suggests that the two time series have very similar but slightly different characteristics. This brings us idea of how to formulate a suitable model in the state space framework.

In figure 8.8 (b) we show another Argand diagram of the characteristic roots of the fitted state space model. The red asterisks are the characteristic roots of the common hidden oscillation, and the remaining blue asterisks are the characteristic roots of the two individual oscillations. The red asterisks are located very closed to the characteristic roots in figure (a), except the real characteristic roots.

And in the following we show the estimated model. The state space model

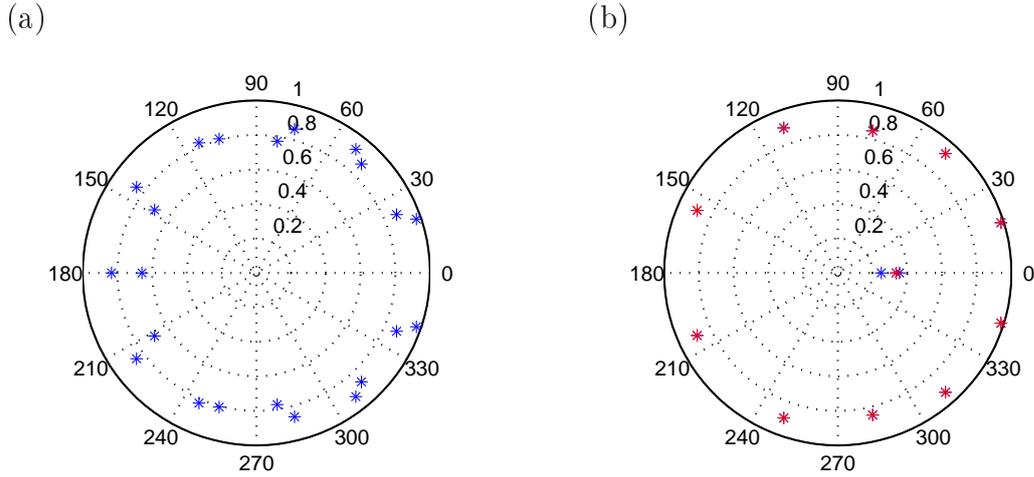


Figure 8.8: Argand diagram of characteristic roots of (a) AR(11) model and (b) state space model

is linear as defined in chapter 2. y_t is the observation which is a 2×1 vector. x_t is the state vector which is 13×1 in our example. F is a 13×13 square matrix, G is a 13 matrix and H is a 2×13 matrix. w_t is the system noise, and it is a 3×1 vector. An 11×11 block in F is used to explain the dynamics of the common hidden variable, and the 2×2 block is a coupling of the remaining two oscillations. G distributes the system noise to only three out of all states. H sums up the second state and the last state to reconstruct the second time series. It also sums up the second last state and one of the first four states to reconstruct the first time series. When we fit this model we allow at each time step H chosen from H_1, H_2, H_3 or H_4 , whichever maximizes the likelihood function at that time step. In this case we are maximizing the likelihood to choose the preferable H at every time point, which makes H as a prior information, and the Kalman filter is using this prior information, and so the fitted model is able to explain the data, but not predict any future observations.

$$\begin{aligned} x_t &= Fx_{t-1} + Gw_t \\ y_t &= Hx_t + \epsilon_t \end{aligned}$$

and H_4 . It is not suitable to draw conclusion about causal relations by looking at the delay of the variables, but we should plot the NCR to see the level of contribution of the driving noises in each of the four cases.

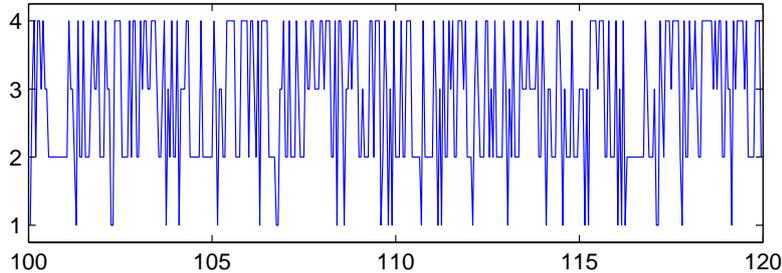


Figure 8.9: Trajectory of $H(j)$

In figure 8.10 we show the NCR for the state space model. From (a) to (d), each plot is showing the NCR of the state space model with a corresponding H . In each plot there is a black line and three colors partitioning the rectangle. The black line is the model spectrum; the yellow color is NCR of common driving oscillation, the green color and the red color are respectively the two independent system noises of the state equation. The vertical axis is ranged from 0 to 1, and the horizontal axis is the frequency axis from 0 Hz to 10 Hz.

Because the coefficients of the second row of H is fixed, the calculation makes no difference to the spectral density of the second time series, but the spectral density of the first time series switches between four different shapes. In figure 8.10 (a) to (d) the model spectra and the NCR plots of the right finger are the same, but those of the left finger vary. The noise contribution from the hidden common driving variable, the yellow portion, is significant particularly at 1 Hz and 3 Hz, at where the spectra, the black lines, contain high peaks. By this result we conclude that both time series is greatly caused by the common hidden driving force at the 1 Hz oscillation and its harmonics.

After eliminating the effect of the common driving force, we can see the causal relations of the two time series by the analysis of the remaining covariance. In figure 8.10 (a) and (b) the red color in the upper plots and the green color in the lower plots are of trace amount, meaning that the causal relations from the left finger to the right finger, and from the right finger to the left finger, conditioned on the hidden common driving force, is negligible. In figure 8.10 (c) and (d) there is only a thin green line in the lower plots, but there is a small proportion of red color in the upper plots. Especially there is about 20 % of NCR from the driving noise of the right finger to the time series of the left finger at 0 Hz. This implies that after taking away the effect of the common driving force of the 1 Hz oscillatory motion, there is evidence of showing the causal relations from the right finger to the left finger at the

low frequency motion. In other words, The right finger was leading the left finger in this experiment.

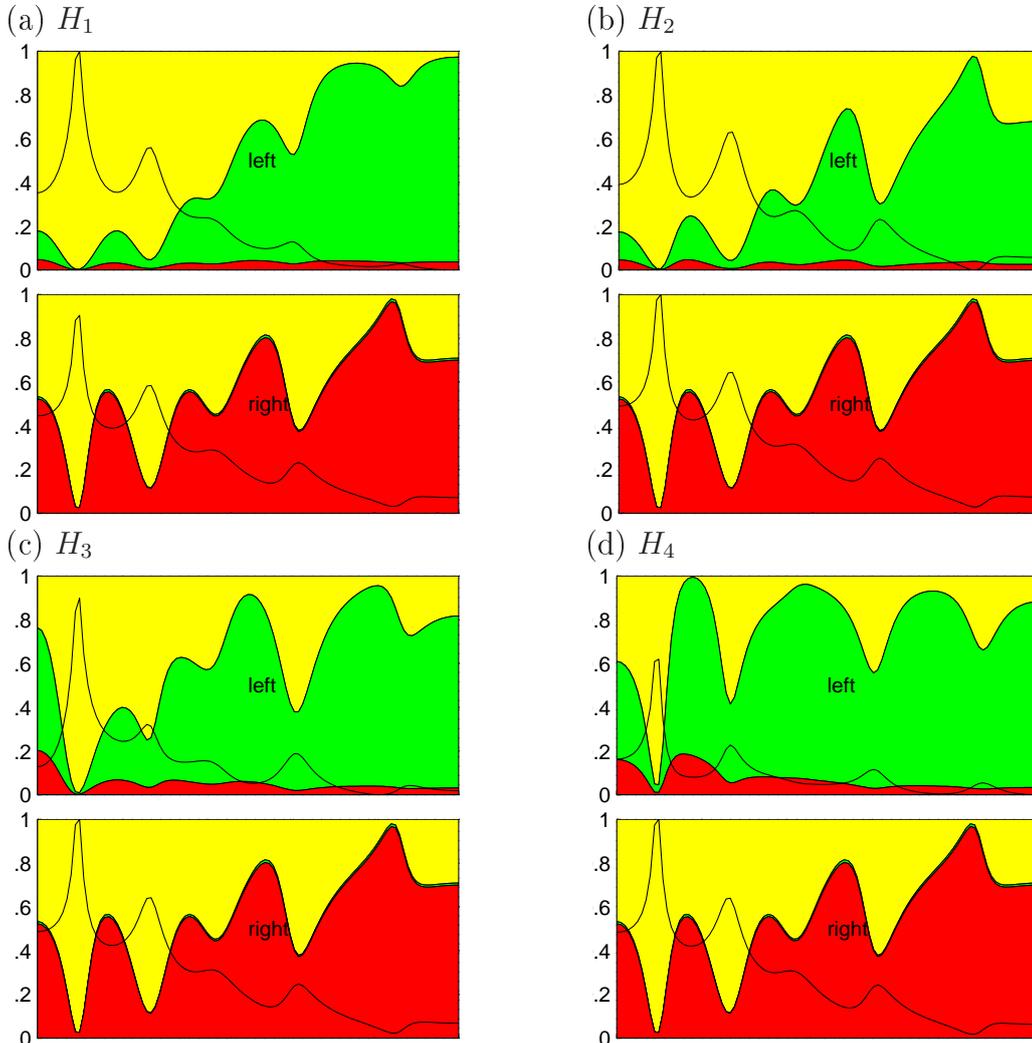


Figure 8.10: NCR of the state space model with different H

In figure 8.11 we show the innovations of the state space model. When we compare to the innovations of the AR(11) model in figure 8.7 as well as the data in figure 8.4, that most of the 1 Hz and 3 Hz oscillation in both time series, some of the low frequency in both time series and part of the high frequency in the green series have been explained by the state space model. However, there is still a number of spikes in both series which may require an exogenous to explain, which is believed not to affect the conclusion of the NCR causality.

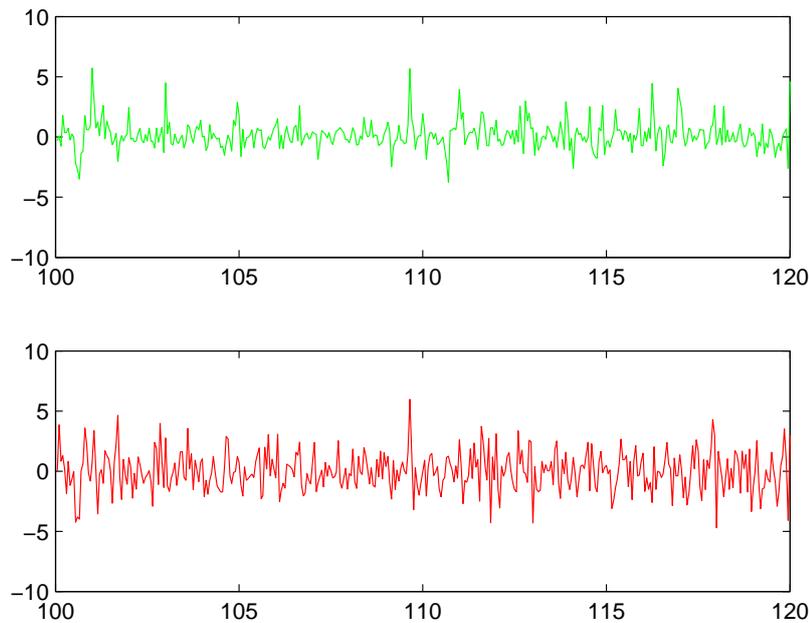


Figure 8.11: Innovation from a state space model

8.4 Discussions

In this chapter we analyzed a physical data of the bimanual finger movement. We apply the time varying NCR to a data of bimanual movement of fingers and see the causal relationship between the motion of the fingers.

We used a multivariate AR model and a linear state space model to fit the data in our analysis. The multivariate AR model contains constant coefficients which explain the homogeneous dynamics of the data and a constant phase between the time series solely. The linear state space model plays an additional role of modelling the varying delay of the two time series. By comparing the AIC we found that the state space model fits better than the AR model.

The essence of NCR causality is that it should be based on a model, so that the causal relations are focused in the major characteristics of the data. But it is also a disadvantage of NCR causality that if we do not fit the data well then the conclusion on causal relations is unreliable. Therefore, while we draw any conclusion from the NCR, we should pay more attention to the innovations.

It is important to identify the most important information we want to obtain from the data. As an example in this chapter, although we have a data of high frequency but the information from high frequency is not contributing to model fitting. Also, the monotonic trend in the data should be removed by either a differencing technique or any suitable trend model.

Although the data shows non-stationary variance, when we tried applying a GARCH model we cannot get a satisfactory result. The reason could be that the noise coming from the subject did not exist for long. GARCH model is a naive prediction of variance from the noise estimate that a sharp and immediately vanished noise impulse may not work well. More investigation is needed on this issue.

Chapter 9

Conclusion

We designed a special case of the linear state space model and named it compartment model. The compartment model is able to decompose a given time series into components defined by their main frequency. We combine the compartment model and GARCH model to develop compartment-GARCH model, suitable to model non-stationary variance of noise in the system.

We make use of different estimators of the square of system noise for the compartment-GARCH model. We try different orders of compartment-GARCH model and compare their performance in real application. We also apply the jump detection technique as an alternative approach to model the non-stationarity of variance.

We fit the compartment-GARCH model to an EEG data during the onset of anaesthesia and conclude that the loss of consciousness at onset of anaesthesia is reflected almost exclusively by an increase of power in the delta band. We fit the compartment-GARCH model to an EEG data of a coma patient and obtain several clear picture of the effectiveness in the detection of heteroscedasticity.

We released the constraint of constant noise variance of a multivariate AR model, to heterogeneous noise variance so that the non-stationary variance can be explained as well as the time-varying NCR causality can be observed. A data quantifying position and angular movement of head gazing between two subjects was analyzed. A unilateral noise contribution between two subjects was clearly portrayed. GARCH model brought a better residuals and improved the likelihood on one hand, and set the NCR free to be time-varying.

We also extend Akaike's NCR from multivariate AR model to multivariate ARMA model as well as state space model, so that more number of driving noise than number of time series variables becomes possible. Increasing number of driving noise allows covariance of driving noises being explained in Akaike's NCR theory, when the covariance is significantly different from zero. We applied the method to a bi-variate time series data of bimanual coordination of rotating index fingers. We obtained a result of causality from a latent variable to the two observed variable. The result was based on a state space model

which gives a better AIC than any ordinary multivariate AR model.

Bibliography

- AKAIKE, H. (1968). On the use of a linear model for the identification of feedback systems. *Annals of the Institute of Statistical Mathematics* 20 425–439.
- AKAIKE, H. (1973). Information theory and an extension of maximum likelihood principle. In *2nd International symposium on information theory*. 267–281.
- AKAIKE, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control* AC-19 716–723.
- AMEEN, J. R. M. & HARRISON, P. J. (1985). Normal discount Bayesian models. In J. M. Bernard, M. H. De Groot, D. V. Lindley & A. F. M. Smith, eds., *Bayesian Statistics 2*. Amsterdam: North-Holland, 271–298.
- AOKI, M. (1987). *State Space Modeling of Time Series*. Berlin: Springer.
- ÅSTRÖM, K. J. & KALLSTROM, C. G. (1973). Application of system identification techniques to the determination of ship dynamics. In P. Eykhoff, ed., *Identification and system parameter estimation*. Amsterdam: North-Holland.
- BINGHAM, G. P. (2004). A perceptually driven dynamical model of bimanual rhythmic movement (and phase perception). *Ecological Psychology* 16 45–53.
- BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 307–327.
- BOX, G. E. P. & JENKINS, G. (1970). *Time Series analysis: Forecasting and control*. Prentice Hall.
- BRINLLINGER, D. R. (1981). *Time series: Data analysis and theory*. McGraw-Hill, expanded ed.
- BROCKWELL, P. J. & DAVIS, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer, 2nd ed.

- BROCKWELL, P. J. & DAVIS, R. A. (1996). *Introduction to Time Series and Forecasting*. New York: Springer.
- CATTAERT, D., SEMJEN, A. & SUMMERS, J. J. (1999). Simulating a neural cross-talk model for between-hand interference during bimanual circle drawing. *Biological Cybernetics* 81 348–358.
- DURBIN, J. & KOOPMAN, S. J. (2001). *Time Series Analysis by State Space Models*. Oxford Univ. Press.
- ENGLE, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50.
- ENGLE, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20 339–350.
- FRANASZCZUK, P. J., BLINOWSKA, K. J. & KOWALCZYK, M. (1985). The application of parametric multichannel spectral estimates in the study of electrical brain activity. *Biological Cybernetics* 51 239–247.
- FRISTON, K. J., HOLMES, A. P., WORSLEY, K. P., POLINE, J. B., FRITH, C. D. & FRACKOWIAK, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* 2 189–210.
- GALKA, A., YAMASHITA, O. & OZAKI, T. (2004). Garch modelling of covariance in dynamical estimation of inverse solutions. *Physics Letter A* 333 261–268.
- GERSCH, W. (1970). Spectral analysis of EEG's by autoregressive decomposition of time series. *Mathematical Biosciences* 7 205–222.
- GERSCH, W. & YONEMOTO, J. (1977). Parametric time series model for multivariate EEG analysis. *Computers and Biomedical Research* 10 113–125.
- GEWEKE, J. F. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association* 77 304–324.
- GEWEKE, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* 79 907–915.
- GILBERT, P. D. (1993). State space and arma models: An overview of the equivalence. *Bank of Canada Working Paper* 4.

- GOLDBERG, J. L. (1991). *Matrix Theory with Applications*. McGraw-Hill.
- GOLUB, G. H. & LOAN, C. F. V. (1996). *Matrix Computations*. Baltimore: Johns Hopkins, 3rd ed.
- GRANGER, C. W. J. (1963). Economic processes involving feedback. *Information and control* 6 28–48.
- GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37 424–438.
- GREWAL, M. S. & ANDREWS, A. P. (2001). *Kalman filtering: Theory and Practice Using MATLAB 2nd edition*. New York: Wiley.
- HAKEN, H., KELSO, J. A. S. & BUNZ, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 51 347–356.
- HARRISON, J. & STEVENS, C. F. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Series B* 38 205–247.
- HARVEY, A. C. (1985). Trend and cycles in macroeconomic time series. *Journal of Business and Economics Statistics* 3 216–227.
- HARVEY, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- HIGUCHI, T. (1999). Applications of quasi-periodic oscillation models to seasonal small count time series. *Computational Statistics and Data Analysis* 30 281–301.
- HOSOYA, Y. (1991). The decomposition and measurement of the interdependency between second-order stationary process. *Probability Theory and Related Fields* 88 429–444.
- JOHN, E. R. (2002a). The neurophysics of consciousness. *Brain Research Reviews* 39 1–28.
- JOHN, E. R. (2002b). A neurophysiological model of consciousness. *International Journal of Psychophysiology* 45 67.
- JOHN, E. R., KARMEL, B. Z., CORNING, W. C., EASTON, P., BROWN, D., AHN, H., JOHN, M., HARMONY, T., PRICHEP, L. S., TORO, A., GERSON, I., BARTTLET, F., THATCHER, R., KAYE, H. & VALDÉS-SOSA, P. (1977). Neurometrics: numerical taxonomy identifies different profiles of brain functions within groups of behaviourally similar people. *Science* 196 1393–1410.

- JOHN, E. R., PRICHEP, L. S., KOX, W., VALDÉS-SOSA, P., BOSCH-BAYARD, J., AUBERT, E., TOM, M., DIMICHELE, F. & GUGINO, L. D. (2001). Invariant reversible QEEG effects of anesthetics. *Consciousness and Cognition* 10 165–183.
- KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82 35–45.
- KAMINSKI, M., DING, M., TRUCCOLO, W. A. & BRESSLER, S. T. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics* 85 145–157.
- KELSO, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology* 15 R1000–R1004.
- KITAGAWA, G. & GERSCH, W. (1996). *Smoothness Priors Analysis of Time Series*. New York: Springer.
- MCINTOSH, A. R. (2000). Towards a network theory of cognition. *Neural Networks* 13 861–870.
- MECHSNER, F., KERZEL, D., KNOBLICH, G. & PRINZ, W. (1971). Identification of stochastic linear dynamic systems. *American Institute of Aeronautics and Astronautics Journal* 9 28–31.
- MEHRA, R. K. (1971). Identification of stochastic linear dynamic systems. *American Institute of Aeronautics and Astronautics Journal* 9 28–31.
- NELSON, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59 347–370.
- OZAKI, T. (1997a). Dynamic X11 model and nonlinear seasonal adjustment I: Models and computational methods. In *Proceedings of the institute of Statistical Mathematics*, vol. 45. Tokyo, 265–285.
- OZAKI, T. (1997b). Dynamic X11 model and nonlinear seasonal adjustment II: Numerical examples and discussion. In *Proceedings of the institute of Statistical Mathematics*, vol. 45. Tokyo, 287–300.
- OZAKI, T. (2003). Compartment model, powerpoint note.
- OZAKI, T. & TONG, H. (1975). On the fitting of non-stationary autoregressive models in time series analysis. In *Proceedings of the 8th Hawaii International Conference on System Sciences*.

- PASCUAL-MARQUI, R. D., VALDÉS-SOSA, P. A. & ALVAREZ, A. (1988). A parametric model for multichannel EEG spectra. *International Journal of Neuroscience* 40 89–99.
- PRICHEP, L. S., GUGINO, L. D., JOHN, E. R., CHABOT, R. J., HOWARD, B., MERKIN, H., TOM, M. L., WOLTER, S., RAUSCH, L. & KOX, W. (2004). The patient state index as an indicator of the level of hypnosis under general anaesthesia. *British Journal of Anaesthesia* 92 393–399.
- PRIESTLEY, M. (1981). *Spectral Analysis and Time Series*. London: Academic Press.
- RAPPELSBERGER, P. & PETSCHKE, H. (1975). Spectral analysis of the EEG by means of autoregression .
- SEMJEN, A., SUMMERS, J. J. & CATTART, D. (1995). Hand coordination in bimanual circle drawing. *Journal of Experimental Psychology* 21 1139–1157.
- SHUMWAY, R. H. (2000). *Time series analysis and its applications*. New York: Springer.
- TANOKURA, Y. & KITAGAWA, G. (2003). Extended power contribution that can be applied without independence assumption. Tech. Rep. 886, The Institute of Statistical Mathematics.
- TSAY, R. S. (2005). *Analysis of Financial Time Series*. Wiley-Interscience.
- VALDÉS-SOSA, P., BISCAY, R., GALAN, L., BOSCH, J., SZAVA, S. & VIRUES, T. (1990). High resolution spectral EEG norms for topography. *Brain Topography* 1 281–282.
- VALDÉS-SOSA, P., JIMENEZ, J. C., RIERA, J., BISCAY, R. & OZAKI, T. (1999). Nonlinear EEG analysis based on a neural mass model. *Biological Cybernetics* 81 348–358.
- VIVIANI, P., PERANI, D., GRASSI, F., BETTINARDI, V. & FAZIO, F. (1998). Hemispheric asymmetries and bimanual asynchrony in left- and right-handers. *Experimental Brain Research* 120 531–536.
- WEI, W. W. S. (1990). *Time series analysis univariate and multivariate methods*. Addison-Wesley.
- WEI, W. W. S. (2006). *Time series analysis univariate and multivariate methods 2nd ed.* Addison-Wesley.
- WEST, M. (1995). Bayesian inference in cyclical component dynamic linear models. *Journal of the American Statistical Association* 90 1301–1312.

- WHITTLE, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika* 50 129–134.
- WIENER, N. (1956). The theory of prediction. In E. F. Beckenback, ed., *Modern mathematics for engineers*. NY: McGraw-Hill, 165–190.
- WONG, K. F. K., GALKA, A., YAMASHITA, O. & OZAKI, T. (2006). Modelling non-stationary variance in EEG time series by state space GARCH model. *Computers in Biology and Medicine* in press.
- YAMASHITA, O., SADATO, N., OKADA, N. & OZAKI, T. (2005). Evaluating frequency-wise directed connectivity of bold signals applying relative power contribution with the linear multivariate time series models. *Neuroimage* 25 478–490.
- YOUNG, G. B., ROPPER, A. H. & BOLTON, C. F. (1997). *Coma and Impaired Consciousness: A Clinical Perspective*. McGraw-Hill.