FUNCTIONAL MRI DATA ANALYSIS: DETECTION, ESTIMATION AND MODELLING

LUO HUAIEN

(M.Eng., Huazhong University of Science and Technology)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2007

Acknowledgements

I would like to thank all those who provided invaluable advice and assistance to my research work during the past four years in the National University of Singapore.

First of all, I would like to express my deepest gratitude to my advisor Dr. Sadasivan Puthusserypady for his patient discussion, inspiring encouragement and prompt guidance.

My special thanks also go to my dear parents. It is their love that lead me through many difficulties.

I also would like to thank my friend Zheng Yue, who plays a pivotal role during the course of my Ph.D studies and especially helps me to recover from many setbacks.

Thanks to my friends Chen Huiting, Zhou Xiaofei, Zuo Ziqiang, Zhang Jing, Rajesh, Ajeesh, Wang Zhibing etc for all the good times we had together.

Luo Huaien

June 2007

Contents

Summary

Functional Magnetic Resonance Imaging (fMRI) is an important technique for neuroimaging. Through the analysis of the variation of blood oxygenation level-dependent (BOLD) signals, fMRI links the function of the brain and its underlying physical structures by using the MRI techniques. The low signal-to-noise ratio (SNR) and complexity of the experiment poses major difficulties and challenges to the analysis of fMRI data.

This thesis presents robust (less false positive rate) and efficient (easy estimation procedure) signal processing methods for fMRI data analysis. It aims to complement the current methods of fMRI data analysis in order to achieve accurate *detection* of the activated regions of the brain, better *estimation* of the hemodynamic response (HDR) of the brain functions and *modelling* of the dynamics of fMRI signal.

The fMRI data are first investigated under the Bayesian framework. Based on the conventional general linear model (GLM), a flexible design matrix determination method through sparse Bayesian learning is proposed. It integrates the advantages of both data-driven and model-driven analysis methods. This method is then extended to incorporate the nonstationary noise to the model. Two nonstationary noise (time-varying variance

noise and fractional noise) models are examined. The covariance matrices of these two noises share common properties and are successfully estimated using a Bayesian estimator. Considering that the fMRI signal also contains drift, a modified GLM model is proposed which could effectively model and remove the drift in the fMRI signal. Through mathematical manipulations, updating algorithms are derived for these proposed methods. The proposed Bayesian estimator could provide accurate probability of the activation and hence avoid the multiple comparison problems encountered in the traditional null hypothesis methods.

The second part of the thesis is focused on the estimation of the HDR of the human brain. Both linear and nonlinear properties of the event-related fMRI experiment are examined based on the inter-stimulus intervals (ISI). A linear spatiotemporal adaptive filter method is proposed to model the spatial activation patterns as well as the HDR. The equivalence of the proposed method to the canonical correlation analysis (CCA) method is also demonstrated. It is reported that when the ISI is small, the fMRI signal shows nonlinear properties. Thus, nonlinear methods of fMRI signal analysis are also examined. A method based on the radial basis function (RBF) neural network is proposed to regress the measured fMRI signal on the input stimulus functions. The relation between the parameters of the RBF neural network and Volterra series are demonstrated. The HDR is then obtained from the parameters of the RBF neural network which shows significant advantages.

The third part of the thesis examines the nonlinear autoregressive with exogenous inputs (NARX) neural network to model the fMRI signal. With the knowledge of experimental paradigm (input) and measured data (output), the NARX neural network could identify the complex human brain system and reconstruct the BOLD signal from noisy fMRI signal. This results in an enhanced SNR of the measured signal and a robust estimation of the activated regions of the brain.

Extensive simulation studies on synthetic as well as experimental fMRI data are carried out in this thesis. Results show that these methods could complement the traditional methods to cope with the difficulties and challenges in fMRI data analysis. This may contribute to the better understanding of the nature of the fMRI signal as well as the underlying mechanisms.

List of Tables

List of Figures

List of Abbreviations

AIC	Akaike Information Criterion
AR	Autoregressive
ARD	Automatic Relevance Determination
ARX	Autoregressive Model with Exogenous Inputs
ATP	Adenosine Triphosphate
BLUE	Best Linear Unbiased Estimator
BOLD	Blood Oxygenation Level Dependent
BSS	Blind Source Separation
CBF	Cerebral Blood Flow
CBV	Cerebral Blood Volume
CCA	Canonical Correlation Analysis
CIC	Confidence Interval Criterion
$CMRO_2$	Cerebral Metabolic Rate of Oxygen
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform

EEG	Electroencephalography
EPI	Echo Planar Imaging
ERP	Event-Related Potentials
FIR	Finite Impulse Response
fMRI	Functional Magnetic Resonance Imaging
FPR	False Positive Ratio
FWHM	Full Width at Half Maximum
GLM	General Linear Model
GLS	Generalized Least Squares
GRE	Gradient-Echo Imaging
HDR	Hemodynamic Response
HRF	Hemodynamic Response Function
ICA	Independent Component Analysis
ICA-R	ICA with Reference
ISI	Inter-Stimulus Intervals
KLT	Karhunen-Loéve Transform
LMS	Least Mean Square
LS	Least Squares
LTI	Linear Time-Invariant
MEG	Magnetoencephalography
MLP	Multi-Layer Perceptrons
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MSE	Mean Square Errors
NARX	Nonlinear Autoregressive with Exogenous Inputs
NMR	Nuclear Magnetic Resonance

NMSE	Normalized Mean Square Error
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PEB	Parametric Empirical Bayesian
PET	Positron Emission Tomography
PPM	Posterior Probability Map
RBF	Radial Basis Function
rCBF	regional CBF
RF	Radio Frequency
ROC	Receiver Operator Characteristic
SD	Standard Deviation
SE	Spin-Echo Imaging
SI	Spiral Imaging
SIC	Schwartz Information Criterion
SNR	Signal-to-Noise Ratio
SOM	Self-Organizing Maps
SPM	Statistical Parametric Mapping
TE	Echo Time
TMS	Transcranial Magnetic Stimulation
TPR	True Positive Ratio
TR	Repetition Time
WLS	Weighted Least Squares

List of Symbols

a_0	Zeroth-order Volterra Kernel
a_1	First-order Volterra Kernel
a_2	Second-order Volterra Kernel
B_0	Static Magnetic Field
B_1	Radio-frequency Pulse
$\mathbf{B}$	Noise Precision Matrix
$\mathbf{b}$	BOLD Response Vector
$\mathbf{c}$	Contrast Vector
$d(n)$	Desired Signal
f	Drift
$\mathbf{f}$	Drift Vector
$\mathbf{G}(t)$	Magnetic Field Gradient
$h(t)$	Hemodynamic Response Function (HRF)
M_0	Initial Magnetization
M_z	Magnetization along z direction

M_{xy}	Magnetization in the xy plane
$\mathbf{r}$	Residual Vector
$\mathbf{S}$	Scaling Matrix
$s(n)$	Stimulus Signal
$\mathbf{w}$	Weight Vector
$y_b(t)$	BOLD Response
$\mathbf{y}$	Data Vector
ϵ	Noise
$\boldsymbol{\epsilon}$	Noise Vector
ρ	Correlation Coefficient
σ^2	Variance
γ	Gyromagnetic Ratio
Ψ	Normal cumulative distribution function
$\boldsymbol{\phi}$	Regressors in the Design Matrix
$\mathbf{\Phi}$	Design Matrix in GLM
$\mathbf{\Sigma}$	Noise Covariance Matrix
$\mathbf{\Lambda}$	Weight Covariance Matrix

Chapter 1

Introduction

If you know, to recognize that you know, if you don't know, to realize that you don't know: That is knowledge. —— *Confucius*

The brain is the most amazing organ in the human body and the most mysterious as well as complex. With the development of cognitive neuroscience, many mysteries are gradually becoming clear to us. Cognitive neuroscience reveals the relation between cognitive processes (the immaterial mind) and the material brain [1] [2]. It shows what happens in the brain when human beings are thinking, talking, learning, memorizing, seeing, acting, etc. To study these cognitive processes in terms of brain-based mechanisms (i.e., which parts of the brain are involved, in what kind of ways, what is the neural basis underlying these processes), many measurement methods have been developed. These measurement methods can be grouped into four categories: the drug-based methods, lesion-based methods, electrophysiological methods and neuroimaging methods. Drug-based methods are used to study how the human brain functions under the control of drugs. Lesion-based methods analyze the influence of naturally occurring lesions or that of "virtual lesions" induced by transcranial magnetic stimulation (TMS) [3] on the brain's functioning. Electrophysiological methods measure the action potentials

or ensemble of the brain action potentials during the execution of a specific task [4]. It includes single-cell recordings, multiple-cells recordings, electroencephalography (EEG), event-related potentials (ERP) and magnetoencephalography (MEG). Although these methods show good temporal resolutions, they provide little spatial information about the activation regions of the human brain. With the help of neuroimaging methods, these functional images of the physiological processes can be visualized. The neuroimaging methods include positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) [5]. These four categories of measurement methods in cognitive neuroscience complement each other to give detailed structural (at which region the neural activities occur) and functional explanations (in which way the brain functions) of the cognitive processes. Table 1.1 shows the summary of the properties of these major methods used in the measurements of the cognitive neuroscience.

As shown in Table 1.1, fMRI possesses advantages of non-invasiveness as well as better spatial and temporal resolution (It has better spatial resolution compared to EEG and MEG and better temporal resolution compared to PET). It can adapt to many types of experimental paradigms. These advantages enable fMRI to provide important information about the brain beyond what is obtained from other techniques. Since its introduction in the early 1990s, fMRI has become the most influential modality for functional neuroimaging. It opens new possibilities to investigate how the human brain works. Many previously unthinkable experiments about cognition and the brain can now be carried out in the laboratories using fMRI.

The fMRI experiments scan the whole or part of the brain repeatedly and generate a sequence of 3-D images. Because of the size and complexity of the fMRI data, powerful analysis methods are essential to the successful interpretation of fMRI experiments. The main aims of the fMRI analysis are both detection and estimation. Detection means to localize the activated regions of the human brain. Estimation, on the other hand, tries to

Table 1.1 Major techniques used for the study of brain functioning

Methods		**Method type**	**Invasiveness**	**Brain property used**	**Temporal Resolution**	**Spatial Resolution**
Drug-based		–	Invasive	Biochemical	–	–
Lesion-based	TMS	Stimulation	Non-invasive	Electromagnetic	Millisecond	Millimetre
Electro-physiological	EEG/ERP	Recording	Non-invasive	Electrical	Millisecond	poor
	Single-cell (multi-units) recordings	Recording	Invasive	Electrical	Millisecond	Neuron level
	MEG	Recording	Non-invasive	Magnetic	Millisecond	Potentially good
Neuroimaging	PET	Recording	Invasive	Haemodynamic	Minutes	Millimetre
	fMRI	Recording	Non-invasive	Haemodynamic	Seconds	Millimetre

study the time course of an activated region related to a specific neural process. However, difficulties such as complexity of the data, low signal-to-noise ratio (SNR) and nonlinear properties, render the analysis of fMRI data a challenging problem. This thesis aims to deal with these difficulties through advanced signal processing and analysis methods. In our work, we only treat the analysis of single-subject (first-level) experiments. Generalization to group studies (second-level experiments) requires further investigation and analysis.

The rest of this chapter is organized as follows. Section 1.1 introduces the basic properties of fMRI, including its Magnetic Resonance Imaging (MRI) fundamentals, the physiological effects that underpin current neuroimaging techniques, its experimental design and so on. Section 1.2 deals with the main methods that have been proposed to analyze the fMRI data. Section 1.3 gives an overview of the thesis and discusses its main contributions.

1.1 Functional Magnetic Resonance Imaging

The basic idea in fMRI is to use MRI to measure the changes in blood oxygenation, which are closely related to the activities of the neurons. The development of fMRI could be traced back to the 1920s and spans almost the whole twentieth century, from Nuclear Magnetic Resonance (NMR) to MRI, and then to fMRI. Figure 1.1 shows some milestones in the development of fMRI. In this section, a short overview of the concepts related to fMRI is given. The experimental fMRI data sets used in this thesis are also introduced.

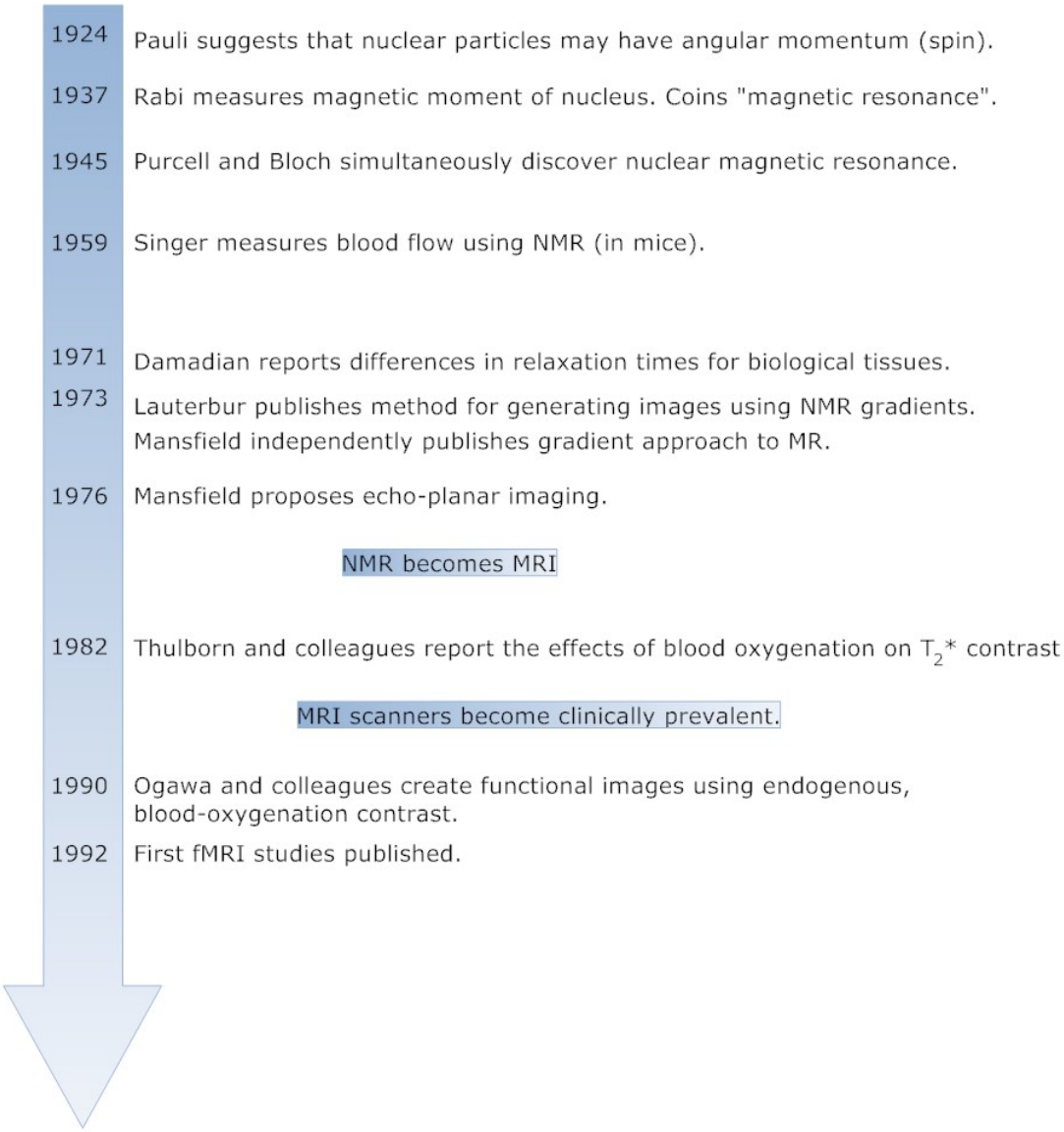


Figure 1.1 Some milestones in the development of fMRI.

1.1.1 Nuclear Magnetic Resonance – the Basis

MRI and fMRI are based on the NMR phenomenon. It concerns primarily the hydrogen nuclei present in the body (most of the tissues are water-based and different tissues contain different amount of water; this hydrogen density difference can be used to construct the 3-D images of the tissues). Each hydrogen nucleus behaves as a tiny magnet, with an angular momentum, called the spin [6]. In the absence of an external magnetic field, the sum of the moments of a sample of molecules is zero; but in the presence of a static magnetic field B_0, the spins align themselves either parallel (low energy state) or anti-parallel (high energy state) to the external static magnetic field, with a slight preference for the first. Thus, the resulting magnetic moment M_0 of a sample is oriented to the direction of B_0. Figure 1.2 shows the illustration of the spins' alignment at equilibrium before and after a static magnetic field B_0 is applied.

The hydrogen nuclei also experiences a torque from the externally applied magnetic field, which causes the spins to rotate, or precess around the direction of the external magnetic field. The frequency of precession is given by the Larmor equation [7]:

$$v_0 = \gamma |B_0|, \tag{1.1}$$

where γ is gyromagnetic ratio (the value of γ depends on the nature of the nuclei) and the precession frequency v_0 is called the Larmor frequency.

The energy state of the nuclei can be changed by transmitting energy to the nuclei using a second oscillating magnetic field pulse tuned to the precession frequency of the nuclei; it is called the resonance. This radio-frequency (RF) pulse B_1 is orthogonal to B_0 and rotates at the Larmor frequency v_0 of the nuclei. As a result, the moment M of a sample is flipped and the flip angle is $\theta = 2\pi\gamma B_1 t$, where t is the duration of the RF pulse. The magnetization vector M has two components: M_z — the longitudinal component aligned with B_0, and M_{xy} — the transverse component in the plane orthogonal

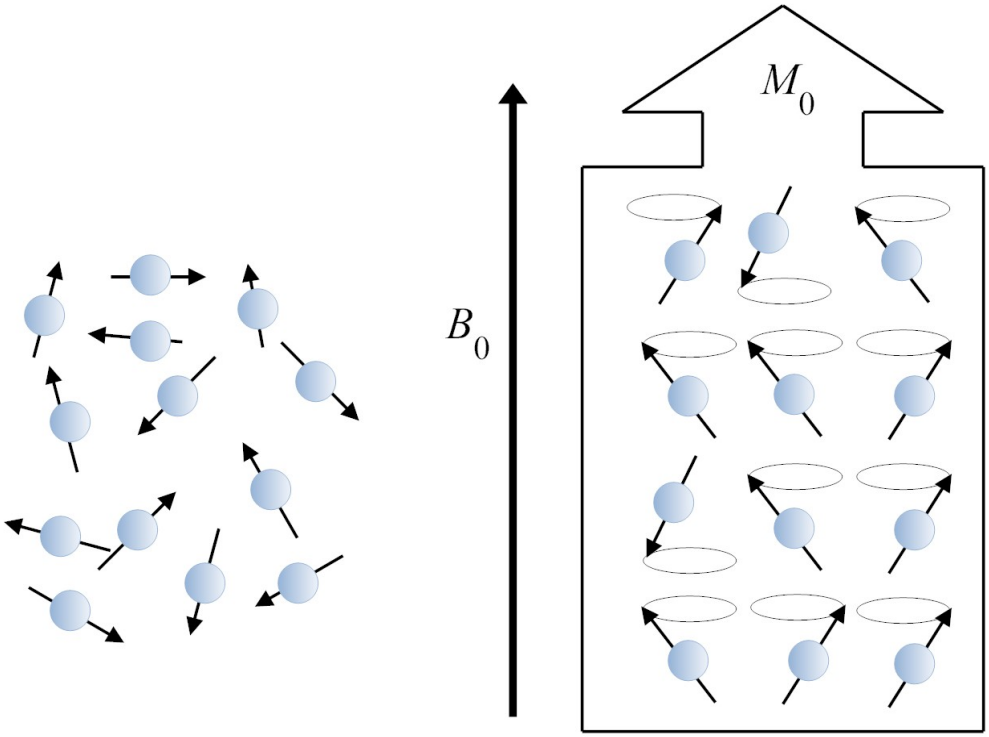


Figure 1.2 Illustration of the spins' alignment at equilibrium before (left side) and after (right side) the magnitude field B_0 is applied.

to B_0. Before applying the RF pulse B_1, M is in the equilibrium state where M_z is maximum and M_{xy} is zero. After applying B_1, M_z becomes small and M_{xy} becomes large. When B_1 is switched off, M will return to the equilibrium state. The recovery of M_z to the initial magnetization M_0 after the RF pulse (longitudinal relaxation) is characterized by the relaxation time constant T_1:

$$M_z(t) = M_0(1 - e^{-t/T_1}). \tag{1.2}$$

The decay of M_{xy} after the RF pulse (transverse relaxation) is characterized by the relaxation time constant T_2:

$$M_{xy}(t) = M_0 e^{-t/T_2}. \tag{1.3}$$

The T_1 relaxation is due to spin-lattice interactions and is the time it takes for the protons to come to equilibrium with each other; and T_2 relaxation is due to spin-spin interactions and is the time it takes for the protons to come to equilibrium with each other. However, since the nuclei in the studied ensemble are spatially distributed, they may experience slightly different magnetic field strength due to a number of reasons. These local magnetic field inhomogeneities greatly accelerate the decay of M_{xy}; the time constant T_2^*, which is shorter than T_2, characterizes the combined effect of random nuclei interactions and magnetic field inhomogeneities.

Two important factors govern the time at which the magnetic resonance (MR) images are collected. The first factor is the repetition time TR, which is the time interval between successive excitation pulses. Since the longitudinal magnetization is not fully recovered at time TR, the transverse magnetization, which determines the detected MR signal, is described as:

$$M_{xy}(t) = M_0(1 - e^{-TR/T_1})e^{-t/T_2}. \tag{1.4}$$

The second factor is the echo time, TE, which is the time interval between the measurement of the received signal and excitation of RF pulse. According to Eq. (1.4), the

Table 1.2 Comparison of different TR, TE and pulse sequence used in different MR contrast images.

Contrast	**TR**	**TE**	**Pulse Sequence**
Proton-density	long	short	GRE[a]/SE[b]
T_1	intermediate	short	GRE/SE
T_2	long	intermediate	SE
T_2^*	long	intermediate	GRE

[a]GRE: gradient-echo imaging

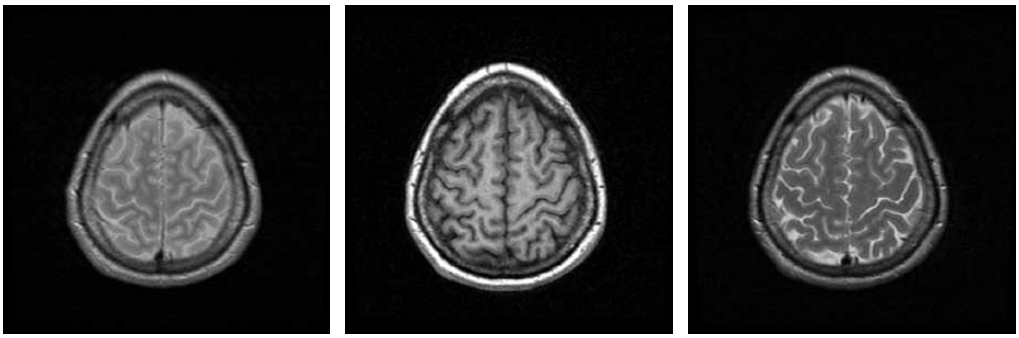
[b]SE: spin-echo imaging

acquired signal is determined by:

$$M_{xy}(t)|_{t=TE} = M_0(1 - e^{-TR/T_1})e^{-TE/T_2}. \tag{1.5}$$

From Eq. (1.5), it is shown that the MR signal depends on two quantities: M_0, determined by the original magnetization or proton density; and $(1 - e^{-TR/T_1})e^{-TE/T_2}$ determined by the properties of the tissue being imaged (different tissues have different time constants T_1 and T_2). It can also be seen from Eq. (1.5) that by manipulating TR and TE, MR images based on different contrast can be obtained. The most basic MRI maps the distribution of hydrogen nuclei (^1H), which is called proton density contrast. MR images based on T_1, T_2 or T_2^* relaxation times emphasize different features of the tissue and the resulting MR images are respectively called T_1-weighted, T_2-weighted or T_2^*-weighted images. The commonly used blood oxygenation level dependent (BOLD) fMRI relies on T_2^* contrast and the structural anatomical image of the brain is commonly T_1-weighted image. Table 1.2 summarizes the TR, TE and pulse sequence used in different contrast based MR images.

Figure 1.3 shows three types of MR images of the same slice in the brain. The



(a) Proton density -weighted (b) T_1-weighted (c) T_2-weighted

Figure 1.3 Three types of MR images of the same slice in the brain.

contrast and the imaging parameters used in these three images are respectively: a) Proton density-weighted (TR = 6000ms, TE = 30ms); b) T_1-weighted (TR = 600ms, TE = 53ms); and c) T_2-weighted (TR = 6000ms, TE = 105ms). All of these images are obtained under the magnetic field 1.5 Telsa.

1.1.2 Magnetic Resonance Imaging

To spatially encode the measurements of proton density, T_1, T_2 or T_2^*, magnetic field gradients $\mathbf{G}(t)$ defined below are used:

$$\mathbf{G}(t) = G_x(t)\mathbf{i} + G_y(t)\mathbf{j} + G_z(t)\mathbf{k}, \tag{1.6}$$

where $G_x(t)$, $G_y(t)$ and $G_z(t)$ are respectively the magnitudes of the gradient magnetic field along x, y and z directions. In the equation above, **i**, **j** and **k** are unit vectors along the x, y and z directions respectively.

This gradient field alters the precession frequency of spins depending on their spatial

location. The total signal measured in MRI combines the changes in the net magnetization generated at every excited voxel and it is represented as:

$$s(t) = \int_x \int_y \int_z M_{xy}(x, y, z, t) dx dy dz. \tag{1.7}$$

By ignoring the terms without spatial information, the acquired signal is:

$$s(t) = \int_x \int_y \int_z M_{xy0}(x, y, z) e^{-i\gamma \int_0^t (G_x(\tau)x + G_y(\tau)y + G_z(\tau)z) d\tau} dx dy dz, \tag{1.8}$$

where $M_{xy0}(x, y, z)$ is the original magnetization at spatial location (x, y, z); $G_x(t)$, $G_y(t)$ and $G_z(t)$ are respectively the gradient magnetic fields $\mathbf{G}(t)$ at this spatial location at time t. The term $-\gamma \int_0^t (G_x(\tau)x + G_y(\tau)y + G_z(\tau)z) d\tau$ is the accumulated phase at this location due to the gradient fields and it is the integral of its precession frequency from the time it is created to the time that is observed.

The MRI image formation comprises three steps. In the first step, the spins in a particular slice are excited (*slice selection*). Then, the spatial distribution of the spins in the selected slice is coded by the two-dimensional gradient impulse and this gives the MR signal in k-space (*spatial encoding*) which will be defined next. Finally, the MR images are reconstructed from the signals in the k-space (*image reconstruction*). Figure 1.4 shows these three steps.

Suppose we want to generate an image centered at the longitudinal location $z = z_0$ (as shown in Fig. 1.5), then the total magnetization $M(x, y)$ of the selected slice along the z-direction with thickness Δz is:

$$M(x, y) = \int_{z_0 - \frac{\Delta z}{2}}^{z_0 + \frac{\Delta z}{2}} M_{xy0}(x, y, z) dz. \tag{1.9}$$

Thus, after the slice selection step, Eq. (1.8) can be simplified as:

$$s(t) = \int_x \int_y M(x, y) e^{-i\gamma \int_0^t (G_x(\tau)x + G_y(\tau)y) d\tau)} dx dy. \tag{1.10}$$

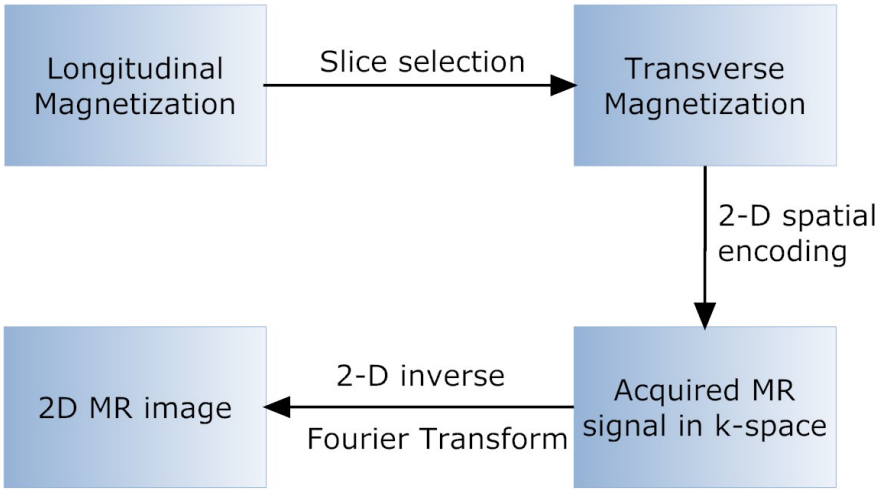


Figure 1.4 Three stages in the formation of MR images.

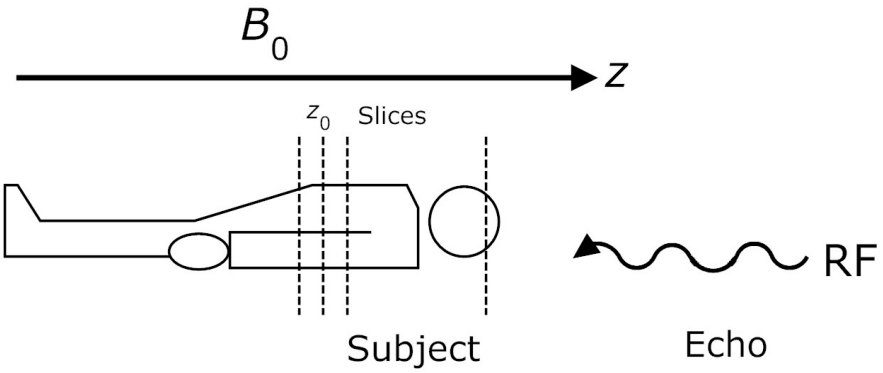


Figure 1.5 Illustration of the slice selection.

If we define:

$$k_x(t) = \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau, \tag{1.11}$$

and

$$k_y(t) = \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau, \tag{1.12}$$

Eq. (1.10) can be represented as:

$$s(t) = \int_x \int_y M(x,y) e^{-i2\pi(k_x(t)x + k_y(t)y)} dx dy. \tag{1.13}$$

This expression has a form similar to the 2-D Fourier transform. At time t, the signal we receive ($s(t)$) is simply the value of the Fourier transform of $M(x,y)$ sampled at the spatial frequency $(k_x(t), k_y(t))$, which is called the k-space. After sampling the spatial frequency content in the k-space, the MR images can be reconstructed through the inverse Fourier transform.

1.1.3 BOLD Functional MRI

The Human brain contains roughly 100 to 150 billion neurons, the activities of which support all the cognitive, sensory and motor processes of the body [8]. Basically, the neurons carry electrical information and interact with other neurons through their synapse. The information is exchanged through the release of neurotransmitters. The transmission of the information along the neuron requires the exchange of ions (e.g. K^+ and Na^+) and the consumption of Adenosine Triphosphate (ATP). The ATP consumption requires the supply of glucose and oxygen, which is provided by increase of cerebral blood flow (CBF). Thus, the brain activity could be assessed by PET which measures the regional CBF (rCBF). fMRI on the other hand measures another effect: the BOLD effect.

When there is neural activity in brain, the oxygen consumption (measured in terms of cerebral metabolic rate of oxygen ($CMRO_2$)) increases, but in less amount than the

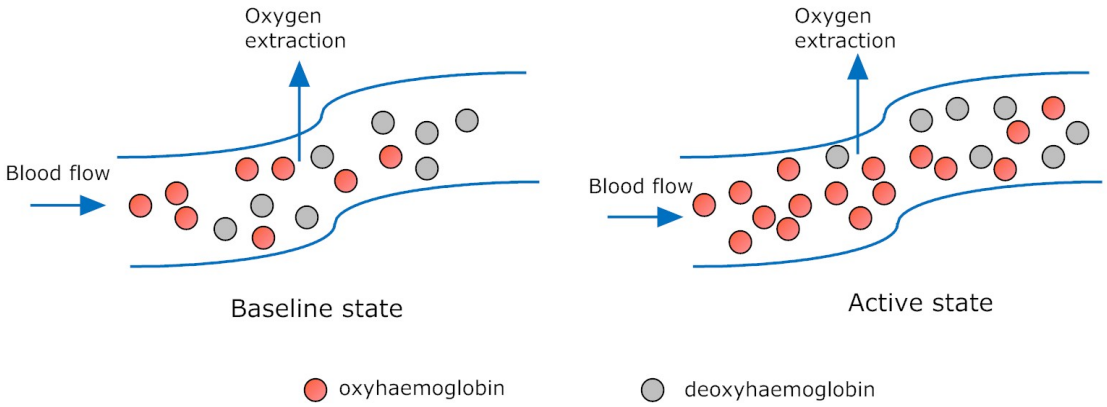


Figure 1.6 Illustration of the change of deoxyhemoglobin content in the venous blood when the neuron is in the baseline (left) and active (right) states. In active state, the oversupply of oxygen by CBF results in the decrease of the concentration of deoxyhemoglobin.

blood flow. This results in the drop in oxygen extraction and a corresponding dilution of deoxyhemoglobin content of the venous blood as shown in Figure 1.6.

The deoxyhemoglobin, without oxygen attached, is paramagnetic, which means that it interacts with and distorts an applied magnetic field. At low oxygen concentrations (baseline state), there are many paramagnetic hemoglobin molecules that locally modulate the main magnetic field B_0 and as a consequence make the hydrogen nuclei excited by an RF-pulse dephase faster. Hence, the T_2^* time constant becomes shorter in areas with low oxygen concentration and longer in areas with high oxygen concentration. MR images reflecting the T_2^* time constant are therefore brighter (longer T_2^*) when a brain is in an active state compared to the baseline state. This effect is referred to as the BOLD effect [9][10][11]. The effect is however very small, an intensity change of around 2-5 percent in the magnetic field 1.5T is expected, and it is therefore hard to be detected. Figure 1.7 summarizes the physiological changes during the brain activation. The actual mechanism underlying the BOLD effect, however, is much more complicated than what

we have demonstrated here. And more importantly, there are some other explanations about the cause of BOLD effect. What we present here is only one possible explanation.

Functional imaging also requires fast acquisition of the images in order to understand the fast physiological changes that are taking place in the brain. To have a high temporal resolution, fast imaging sequences like Echo Planar Imaging (EPI) and Spiral Imaging (SI) are commonly used in fMRI. These gradient-echo imaging techniques are sensitive to the T_2^* time constant, and are capable of capturing an image slice in less than 100 milliseconds and an entire brain volume in just a few seconds (2s or even less). Compared to structural anatomical images, those acquired under the fast imaging sequences are of relatively lower spatial resolution and after time TR, the set of brain images is acquired again. This process results in a 4-D (three spatial dimensions plus time) spatio-temporal dataset.

1.1.4 Hemodynamic Response

The change in MR signal following the firing of neurons is known as the hemodynamic response (HDR), which gives the information of how the BOLD signal evolves over time in response to the brief neuronal activity. The HDR normally has three phases [12]:

1. *Initial Dip*: This initial negative-going dip spans 1 to 2s and is attributed to a transient increase in the amount of deoxyhemoglobin as neurons consume oxygen. (This phenomenon may not be observed in the standard 1.5T magnets.)

2. *Overcompensation*: In this phase, more oxygen is supplied than is extracted, and this results in a decrease of the concentration of deoxyhemoglobin, and hence significant increase of BOLD signal.

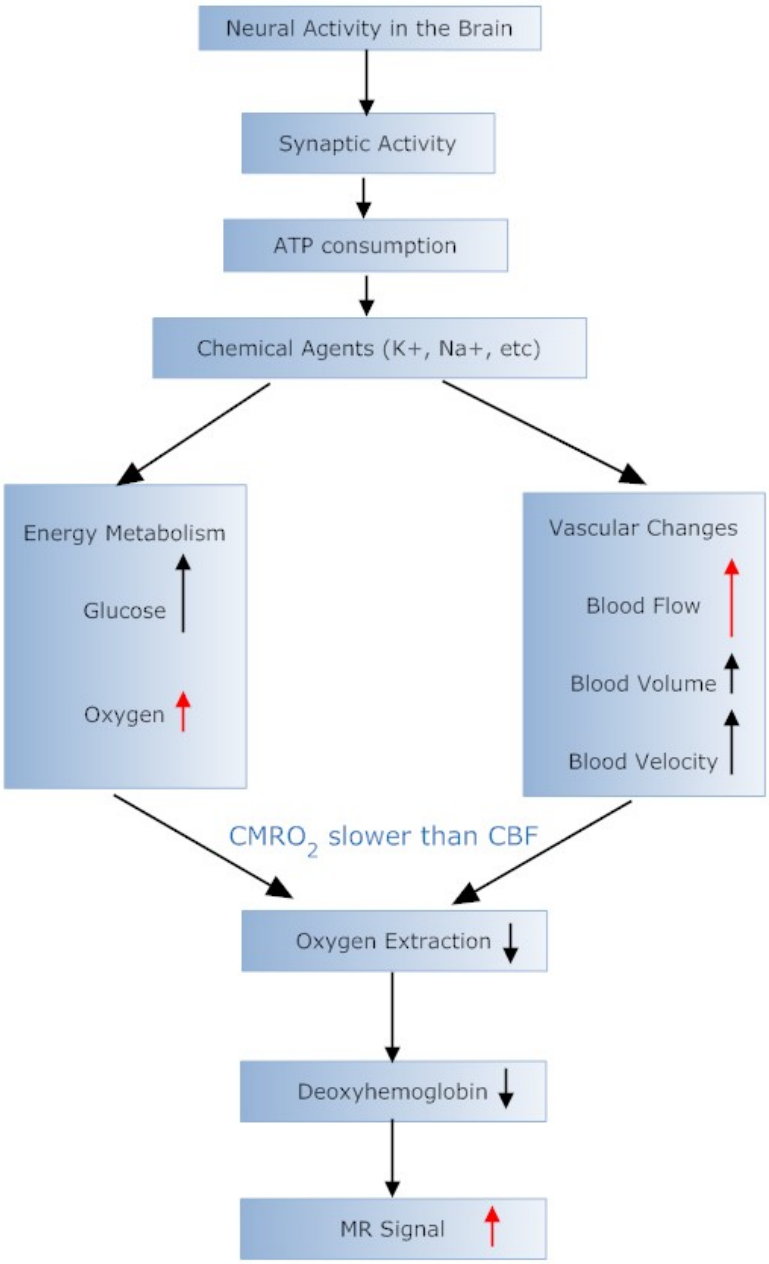
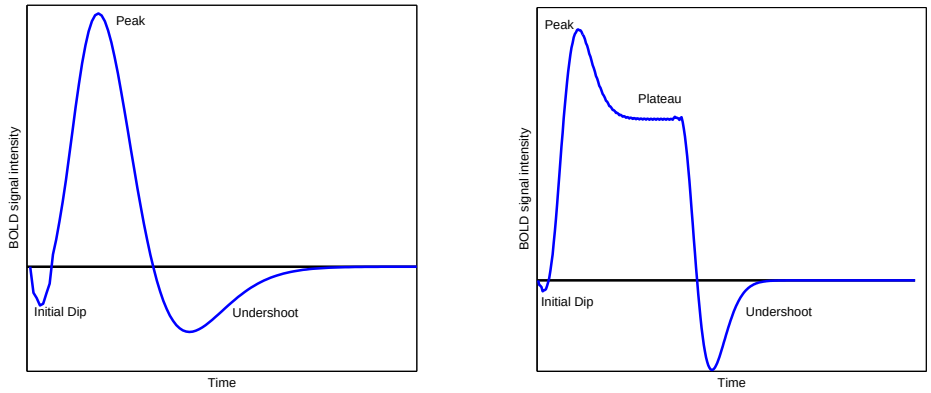


Figure 1.7 Physiological changes accompanying brain activation



(a) (b)

Figure 1.8 Schematic representations of the fMRI BOLD hemodynamic responses. (a) HDR to a single short duration event; (b) HDR to a block of multiple consecutive events.

3. *Undershoot*: Finally, the blood flow and oxygen consumption return to the baseline level. However, the blood flow decreases more rapidly than blood volume, causing temporary increase in deoxyhemoglobin again.

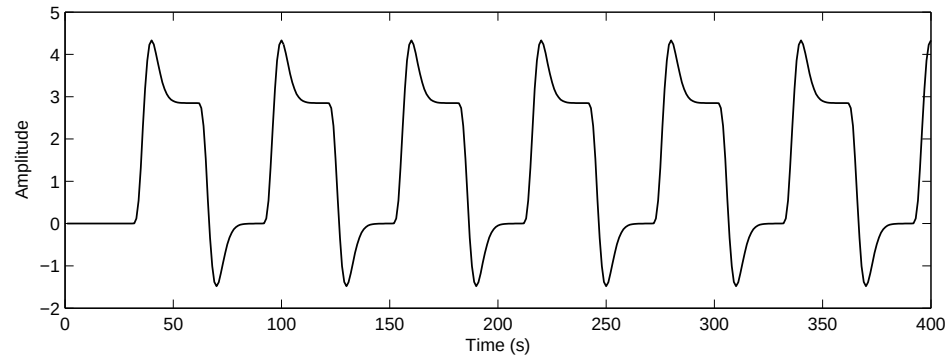
Figure 1.8 shows the representative waveforms for the hemodynamic responses to the single event and multiple events respectively. The hemodynamic response function (HRF) is relatively stable across sessions with the same participant in the same region, but for different regions within the same individual or between individuals, the HDR varies greatly [13].

1.1.5 Experimental Designs in fMRI

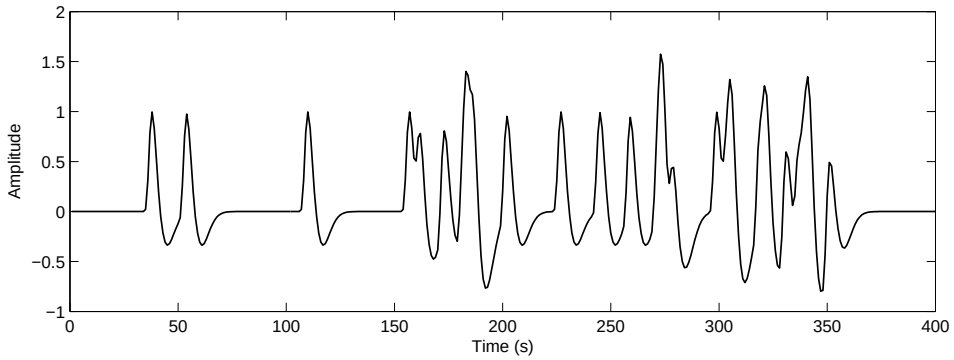
As shown in Figure 1.9, the fMRI experiment begins with an experimental design. The experimental design includes hypothesis formation, choice of experimental conditions

Figure 1.9 The basic steps in an fMRI experiment.

and presentation of stimuli to manipulate the experimental conditions. A good experimental design is the key to the success of the fMRI experiment. In fMRI experiments, two schemes of experimental designs are generally used: the *block design* and the *event-related design* [14]. In block design, each cognitive condition is presented repeatedly for an extended time interval and the different conditions alternate periodically. This design is shown to be an optimum design for brain activation detection because sufficient SNR can be obtained. However, measuring the temporal integration of the response signals instead of the response to individual stimuli limits the flexibility of the block design. In event-related design, on the other hand, the discrete stimuli are presented briefly one at a time separated by interstimulus intervals (ISI) rather than together in a block. Compared to the block design, the event-related fMRI experimental design is more versatile and is a suitable scheme to naturally event-related experimental tasks such as the 'oddball' experiments. Furthermore, the event-related design could capture the temporal properties of the response, thereby providing us with the ability to investigate the timing of the HDR. Figure 1.10 illustrates simulated BOLD signals of block and event-related designs.



(a) Block Design.



(b) Event-related Design.

Figure 1.10 Illustration of BOLD signals of (a) block design and (b) event-related design.

1.1.6 Description of the Experimental Data Used in This Thesis

In this thesis, two data sets from fMRI experiments are analyzed. One is a block design and the other is an event-related design; we will refer to these two data sets respectively as *DATA-BLOCK* and *DATA-EVENT*, hereafter. These data sets are obtained from the National fMRI data center (http://www.fmridc.org) with the accession number 2-2000-111JJ (*DATA-BLOCK*) and 2-2000-11127 (*DATA-EVENT*). In this section, a brief description of these two fMRI experiments is presented.

DATA-BLOCK

This fMRI experiment was designed for visuospatial processing task – judgement of line orientation [15]. In this data, 100 T_2^*-weighted images were obtained with an in-plane resolution of 3mm (Magnetic Field Strength = 1.5T, TR = 3000ms, TE = 40ms, flip angle = 90^o). There are five cycles of alternating 30-sec epochs of baseline and activation conditions in this experiment which gives a total experimental duration of 5 minutes and sample length of 10 in each epoch. During the baseline condition, the subjects were asked to ascertain if the two stimulus lines on the screen were on the same level. During the activation condition, the subjects were shown two stimulus lines in the top half of the screen, and an exemplar consisting of nine radial lines arranged in a semi-circle in the bottom of the screen. These two stimulus lines can be any two of the nine radial lines in the exemplar. The subjects had to decide if these two stimulus lines matched the two highlighted lines on the exemplar. This experiment activates the striate cortex and extrastriatal cortex of the brain.

DATA-EVENT

This real fMRI experiment was designed for detection of visual-mental imagery and perception — transient activity in the human calcarine cortex. The functional data were acquired using EPI with the imaging parameters: Magnetic Field Strength = 3T, TE = 40 ms, TR = 2000 ms, flip angle = 90^o and 64×80 matrix. The details of the experiment can be found in [16]. During the experiment, after hearing the name of an animal, the subjects were told to form an image mentally; then, the subjects evaluated the characteristics of the named animal after hearing an auditory instruction of a specific property of the named animal. The total number of data points for this experiment is 308, with several visual-mental processes. From the description of the experiment, in addition to the visual-mental process, the subjects were also given auditory stimuli. Each of these auditory stimuli lasting for 2 seconds are presented with fixed interval and separated 14 seconds apart (7 scans). As a result, the auditory cortex should also be activated. In this thesis, the activation of the auditory cortex is examined.

1.2 fMRI Data Analysis

This section describes the various techniques of fMRI data analysis, including the preprocessing, signal models and analysis methods.

1.2.1 Preprocessing

Realignment

Even though the heads of the subjects are restrained from movements physically in the scanner and the subjects are instructed to keep as still as possible, it is inevitable that the head may change its position during the relatively long period of scanning. This may

pose the problem that a given voxel does not uniquely represent a specified area in the brain. Realignment tries to reduce the effect of the head motion. This involves basically two steps: the motion estimation and correction.

Co-registration

This step is to co-register the functional data and the structural (anatomical) data, that is, to map the individual functional image onto anatomical image of the same subject or onto a standard template (stereotactic normalization) and then to interpret the functional results on the template. The commonly used templates are brain atlas provided by Talairach and Tournoux [17] or averaged brain images provided by Montreal Neurological Institute [18].

Smoothing

Smoothing spreads the activation of a voxel to its neighboring voxels. It has two advantages [19][20]: i) to increase the SNR and ii) to make the data satisfy the later processing assumptions, for example, the interpretation of the images as Gaussian random field [21].

1.2.2 Modelling the fMRI Data

From an engineering perspective, fMRI data analysis can be viewed as the analysis of the response of a system given the information about the input and output. This is shown in Figure 1.11. The system is a black box which contains the subject undergoing the experiment (including experimental task, other cognitive/sensory/motor tasks, motion, etc.) and the measurement settings (including environment, BOLD effect, data acquisition, spatial reconstruction, artifacts, preprocessing etc.). Since fMRI data is a 4-D data

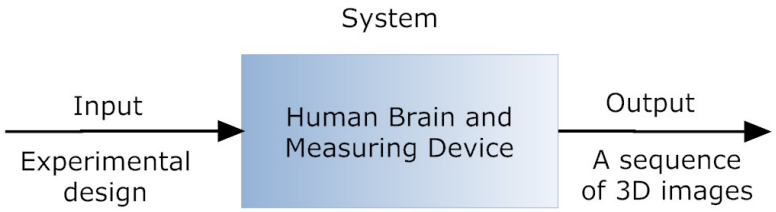


Figure 1.11 fMRI data acquisition as a system with input and output.

set, it is necessary to investigate the temporal and spatial properties of the fMRI data in order to have a full understanding of this complex data.

Temporal Modelling of fMRI Data

The measured fMRI data are often interpreted in terms of voxel time series and the dynamic behavior of the intensity in a voxel indicates whether this voxel is active or not in a specified experimental task. Roughly speaking, an observed voxel time series could be decomposed as:

Observed voxel time series = BOLD response + Drift + Noise.

Each of these parts are briefly explained below. In the subsequent chapters, the related signal models are described in greater detail.

Models for the BOLD Response

1. *Boxcar model*

   This is the simplest BOLD response model which ignores how the system (brain and MR-scanner) modulates the applied experimental paradigm and just use a square-wave function to represent the observed BOLD response (only applicable

to block design). The resulting BOLD signal $y_b(n)$ for a block design with the block duration p is:

$$y_b(n) = \begin{cases} 0, & \text{if } n = 1, \cdots, p; 2p+1, \cdots, 3p; \cdots; T-2p+1, \cdots, T-p \\ 1, & \text{if } n = p+1, \cdots, 2p; 3p+1, \cdots, 4p; \cdots; T-p+1, \cdots, T \end{cases} \tag{1.14}$$

where T is the total number of the time points of the BOLD response.

2. *Convolutive model*

This model assumes the system to be linear and hence the BOLD signal $y_b(t)$ is the convolution between the system impulse response and the experimental design:

$$y_b(t) = h(t) \otimes s(t) = \int_0^\infty h(\tau) s(t-\tau) d\tau \tag{1.15}$$

where $\otimes$ denotes the convolution operation, $s(t)$ is the signal representing experimental paradigm, $h(t)$ is the impulse response called HRF. The models proposed to describe the HRF waveform include the Poisson model [22], Gamma model [23], Gaussian model [24] and difference of gamma functions model [25]. Among these models, the difference of gamma functions model could model the undershoot of the HDR and hence it is widely used. This model is represented by:

$$h(t) = \Big(\frac{t}{d_1}\Big)^{a_1} \exp\Big(\frac{-(t-d_1)}{b_1}\Big) - c\Big(\frac{t}{d_2}\Big)^{a_2} \exp\Big(\frac{-(t-d_2)}{b_2}\Big) \tag{1.16}$$

where $d_j = a_j b_j$ is the time (in seconds) to the peak amplitude of $h(t)$. The common choice of these parameters are $a_1 = 6$, $a_2 = 12$, $b_1 = b_2 = 0.9s$ and $c = 0.35$, which can generate the canonical HRF commonly used in fMRI data analysis.

3. *Balloon model*

The Balloon model is a physiologically inspired model introduced by Buxton et

al [26]. It is a state-space model which describes the dynamics among blood flow, blood volume and blood oxygen concentration changes. The model is inherently nonlinear and can be used to explain the nonlinearities appearing in the BOLD signals. Compared to the convolutive model which is used for brain activity detection only, the Balloon model has been mainly used for understanding the mechanisms underlying the BOLD effect. The detailed description of the Balloon model is given in Section 6.4 where the nonlinear models of fMRI data analysis are discussed.

4. *Subspace model*

   This model considers the variations in the BOLD response between different subjects and different brain areas by defining a linear signal subspace in which realistic BOLD responses reside [37]. A BOLD response shape $y_b(t)$ can then be realized as a linear combination of several temporal basis functions:

$$y_b(t) = w_1 y_1(t) + \cdots + w_n y_n(t) \tag{1.17}$$

   where $w_1, \cdots, w_n$ are the weights and $y_1(t), \cdots, y_n(t)$ are the basis functions. These basis functions could be Fourier series consisting of sine and cosine functions with different frequencies, or Taylor series consisting of different derivatives and so on. It may be noted that this subspace model should be constrained to avoid over-fitting. A general principle is that the model subspace should be as small as possible while it still should capture the important variations in actual BOLD response shapes.

Drift and Trends

Drift and trends are the slowly varying interferences which are generally present in the fMRI time series. The sources of these drifts and trends are not well understood, and

may come from the instability of the magnetic field [27], physiological processes such as respiration and cardiac processes and so on [28]. It is necessary to model and remove the drifts in order to make correct inferences about the data. The most common drift model consists of linear subspaces spanned either by a set of polynomials or a set of cosine functions. In the fMRI analysis software, *Statistical Parametric Mapping* (SPM) [29], a highpass filter using discrete cosine transform (DCT) basis functions is implemented to remove the lower frequency components (drift). In Chapter 4, the drift in the fMRI data is modelled to reside in the subspace spanned by a few large scale wavelets and a criterion for choosing the model order is proposed.

Noise

The random noise is basically what is left when the BOLD signal as well as the drift or trends have been specified. This includes not only the "genuine" random physiological noise and scanner noise, but also the residual variance due to an imperfect model as well. The most popular model is the white Gaussian noise. Even if this is not always specified explicitly, it is the implicit model underlying the simple *t*-test methods and correlation methods, which are widely used. However, it is found that the noise is not independent. For this reason, we must take some care to model the correlation structure. The simplest one is the first order autoregressive model (AR(1)):

$$\varepsilon_i = a\varepsilon_{i-1} + \eta_i, \tag{1.18}$$

where ε_i and ε_{i-1} are the error component at time point i and $i-1$, a is the AR(1) coefficient and $\eta_i \sim \mathcal{N}(0, \sigma^2)$. A more complex model is AR(p) model or an AR(1)+white noise model [30].

Besides these traditional noise models, the $1/f-$like noise or nonstationary fractional noises in the fMRI data have been reported and investigated by many researchers recently. In Chapter 3 and 4, we discuss this noise model and show how the wavelet

transform which whitens the noise in fMRI could be used in the proposed Bayesian estimator.

Spatial Modelling of fMRI Data

Even though it is the temporal behavior of the voxel time series that determines whether the voxels are active or not, the voxel time series in a spatial context can be exploited to improve the detection of the active voxels. The active areas, in general, have a spatial extent of several millimeters, i.e., a few voxels. This causes that the *t* statistics or *F* statistics occur above a specified threshold in a cluster manner in the SPM. Gaussian random field theory is applied in SPM to account for the spatial correlation of the fMRI data.

Single Voxel Model

It is the simplest model that each voxel time series is analyzed in isolation from its neighbors. The underlying assumption is that the active brain areas are smaller than the area covered by a single voxel.

Gaussian Model

This model is currently the most commonly used model in fMRI analysis due to its equivalence with a pre-smoothing of the fMRI images with a Gaussian filter. The size of the filter is commonly measured by the Full Width at Half Maximum (FWHM)– the distance between two opposite points on the Gaussian at which the function reaches half its maximum value.

Flexible Model

Due to the variation of the shape, the activation profile can be described as the linear combination of several spatial basis functions $f_i(\mathbf{x})$:

$$f(\mathbf{x}) = w_1 f_1(\mathbf{x}) + \cdots + w_m f_m(\mathbf{x}), \tag{1.19}$$

where $\mathbf{x}$ represents spatial locations. When examining a voxel for activity, the voxel time series in the neighborhood can be first weighted and averaged according to the spatial basis functions to improve the detection accuracy. However, the flexibility of this model may cause an over-fitting problem.

1.2.3 Data Analysis and Inference

The analysis methods of fMRI data could be divided into two categories:

1. *Model-driven (hypothesis testing) methods*: These methods assume a certain form of the response to the experimental stimulation and then statistically test the presence or absence of the response in the analyzed data. The methods in this category differ either by the signal estimation procedure or by the statistical method employed to assess the activation. The potential weakness of model-driven methods is that the assumed response may be inadequate to describe the real response, thus yielding biased conclusions. The commonly used general linear model (GLM) method belongs to this category.

2. *Data-driven (exploratory) methods*: These methods consider all the voxels simultaneously. They search for kinds of patterns appearing in the data set, and how these patterns are temporally/spatially structured. This information is extracted from the data set without any prior knowledge of the experimental paradigm. Roughly speaking, these methods can be divided into two groups: blind source separation (BSS) and clustering. BSS tries to find a generative model of the data which assumes that the measured signals are mixtures of different kind of sources. Principal component analysis (PCA) [31][32], independent component analysis (ICA) [33][34][35][36] and canonical correlation analysis (CCA) [37] methods

are used to unmix these mixtures to obtain the source signals. By contrast, clustering [38][39] and self-organizing map (SOM) techniques are based on the assumption that the set of voxels can be split into different sets on which one effect is predominant.

General Linear Model (GLM)

The GLM has been popularized by the SPM software. It is the most fundamental and basic method to analyze the fMRI data sets. Compared to the common *t* test and correlation analysis, GLM could eliminate the effects that may confound the analysis by using a suitable model.

Let $y(t)$ denote the measured response in one voxel of the fMRI data set at time t. We assume that the subject undergoes C different conditions of a given experimental paradigm. The conditions are defined by the time course of the stimulation function $s_c(t)$, where $c = 1, \ldots, C$.

The GLM assumes that the response to each experimental task is proportional to the convolution of the stimulation $s_c(t)$ with a filter $h(t)$ known as the canonical HRF discussed earlier (Eq. (1.16)). Using the assumption that the system is linear time invariant (LTI), the most basic signal model is:

$$y(t) = w_0 + \sum_{c=1}^{C} w_c h(t) \otimes s_c(t) + \epsilon(t) \tag{1.20}$$

where w_0 is a constant, w_c is the amplitude of the responses to the stimulus c and $\epsilon(t)$ is the noise term.

Let us denote $\phi_c(t) = h(t) \otimes s_c(t)$ and specify other explicative variables such as head motion estimation during the session, or low frequency sinusoids, or the convolutions of the first and second temporal derivatives of the HRF $h(t)$ with $s_c(t)$ to account for the variations. We obtain a set of temporal regressors $\phi_r(t)$, with $r = 1, \ldots, R$;

where R ($\geq C$) is the total number of regressors. Putting these regressors into a linear model, the following equation is obtained:

$$y(t) = \sum_{r=1}^{R} w_r \phi_r(t) + \epsilon(t). \tag{1.21}$$

The discrete version of the GLM can be succinctly denoted in a matrix form,

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \boldsymbol{\epsilon}. \tag{1.22}$$

Here, $\mathbf{y} = [y_1, \cdots, y_T]^T$ is the vector of observed/measured voxel values with dimensions $T \times 1$, where T is the length of the time series or indices of observations (e.g. scan). $\mathbf{w} = [w_1, \cdots, w_R]^T$ is the vector of parameters and $\boldsymbol{\epsilon} = [\epsilon_1, \cdots, \epsilon_T]^T$ is the vector of error terms which is assumed to be normal $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{\Sigma})$. The design matrix $\mathbf{\Phi}$ (dimensions $T \times R$) is defined as:

$$\mathbf{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1R} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{T1} & \phi_{T2} & \cdots & \phi_{TR} \end{pmatrix}, \tag{1.23}$$

where ϕ_{ij} is the j^{th} regressor at the i^{th} time points of the time series. Each row of $\mathbf{\Phi}$ corresponds to one time point (scan) of the regressors, and the columns represent different explanatory variables or regressors in the model.

With the assumption that $\boldsymbol{\epsilon}$ is white noise with $\mathbf{\Sigma} = \mathbf{I}$, $\mathbf{\Phi}$ is of full column rank and consequently $\mathbf{\Phi}^T\mathbf{\Phi}$ invertible, the least squares (LS) estimation of the parameter $\mathbf{w}$ is:

$$\hat{\mathbf{w}} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{y} \tag{1.24}$$

with the covariance matrix

$$\mathbf{\Lambda}_{\hat{\mathbf{w}}} = \sigma^2(\mathbf{\Phi}^T\mathbf{\Phi})^{-1}. \tag{1.25}$$

This technique yields the best linear unbiased estimator (BLUE) [40] of $\mathbf{w}$ ($E(\hat{\mathbf{w}}) = \mathbf{w}$). The estimated parameters are also Gaussian distributed, i.e., $\hat{\mathbf{w}} \sim \mathcal{N}(\mathbf{w}, \sigma^2(\mathbf{\Phi}^T\mathbf{\Phi})^{-1})$, and since the variance can be consistently estimated, the statistical inference can be made based on this.

Considering the temporal autocorrelation of the noise ($\mathbf{\Sigma} \neq \mathbf{I}$), the following procedures can be performed to make the statistical inferences easier to derive and interpret:

1. Whiten the signal $\mathbf{y}$ together with the regression model $\mathbf{\Phi}$; this procedure requires a careful estimation of the autocorrelation [41].

2. Estimate the noise covariance $\mathbf{\Sigma}$, and modify Eq. (1.24) and Eq. (1.25) by:

$$\hat{\mathbf{w}} = (\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{y} \tag{1.26}$$

$$\mathbf{\Lambda}_{\hat{\mathbf{w}}} = \sigma^2(\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi})^{-1}. \tag{1.27}$$

   This method is at the risk of biasing the result if the estimate of $\mathbf{\Sigma}$ is poor [42].

3. Add more correlation than what is actually in the data (replace the unknown correlation with known correlation), and derive a new noise covariance matrix [20]. This model has the form:

$$\mathbf{Ky} = \mathbf{K\Phi w} + \mathbf{K}\boldsymbol{\epsilon} \tag{1.28}$$

   where $\mathbf{K}$ is the filter written in matrix form. Eq. (1.24) and Eq. (1.25) now become:

$$\hat{\mathbf{w}} = (\mathbf{\Phi}^T\mathbf{K}^T\mathbf{K\Phi})^{-1}\mathbf{\Phi}^T\mathbf{K}^T\mathbf{Ky} \tag{1.29}$$

$$\mathbf{\Lambda}_{\hat{\mathbf{w}}} = \sigma^2(\mathbf{\Phi}^T\mathbf{K}^T\mathbf{K\Phi})^{-1}\mathbf{\Phi}^T\mathbf{K}^T\mathbf{K\Sigma K}^T\mathbf{K\Phi}(\mathbf{\Phi}^T\mathbf{K}^T\mathbf{K\Phi})^{-1}. \tag{1.30}$$

   This estimate is biased, however it is argued that this bias is inferior to the bias induced by an improper whitening [20].

After having estimated the parameter $\mathbf{w}$, it is possible to assess whether the elements in $\mathbf{w}$ corresponding to the response are large enough to ascertain that the response truly exists in the data. The null hypothesis here is that there is no activation anywhere in the brain, that is, $H_0 : \mathbf{c}^T\mathbf{w} = 0$, where $\mathbf{c}$ is a contrast which forms a linear combination of the estimates $\hat{\mathbf{w}}$. To justify the activation of the voxel, the null hypothesis needs to be rejected. Under the null hypothesis, the contrast of the parameter estimates $\mathbf{c}^T\hat{\mathbf{w}}$ is normally (Gaussian) distributed with mean $\mathbf{c}^T\mathbf{w}$ and variance $\mathbf{c}^T\mathbf{\Lambda}_{\hat{\mathbf{w}}}\mathbf{c}$ as shown below:

$$\mathbf{c}^T\hat{\mathbf{w}} \sim \mathcal{N}(\mathbf{c}^T\mathbf{w}, \mathbf{c}^T\mathbf{\Lambda}_{\hat{\mathbf{w}}}\mathbf{c}). \tag{1.31}$$

A *t*-statistic can then be formed by dividing a contrast of the estimated parameters $\mathbf{c}^T\hat{\mathbf{w}}$ by its estimated standard deviation [43]:

$$\frac{\mathbf{c}^T\hat{\mathbf{w}}}{\sqrt{\mathbf{c}^T\mathbf{\Lambda}_{\hat{\mathbf{w}}}\mathbf{c}}} \sim t_{df} \tag{1.32}$$

where t_{df} is the Student distribution with df degrees of freedom, which is derived and adjusted using the *Satterthwaite* approximation [44]:

$$df = \frac{trace(\mathbf{R\Sigma})^2}{trace(\mathbf{R\Sigma R\Sigma})} \tag{1.33}$$

where $\mathbf{R} = \mathbf{I} - \mathbf{P}_{\mathbf{\Phi}} = \mathbf{I} - \mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T$ is the orthogonal projection matrix (projector onto the residual space of $\mathbf{\Phi}$). For practical applications, the *t* scores are then converted to a normal variable *z* through standard procedures. Each voxel has a *t* score or *z* score, thus a statistical parametric map *SPM*$\{t\}$ or *SPM*$\{z\}$ is formed. This map can be thresholded for a certain significance value or *P* value, which is the probability that *z* or *t* is above a given threshold under the null hypothesis. Then the activated regions in the brain can be inferred by rejecting the null hypothesis given the statistical score.

The *P*-value defined above is voxel-based which may create a problem when trying to control the overall or family wise error (a multiple comparison problem). In

order to control the number of false positive voxels for a given map, we should consider map-wise threshold. If we assume that the voxels are independent, then the Bonferroni correction procedure [45] could be applied. However, the Bonferroni correction is too conservative since the voxels are spatially correlated. Therefore, the Gaussian random field theory [21] is applied. The idea is that, under the null hypothesis, the z-map is a Gaussian random field, with a given smoothness. The probability of family wise error can be approximated by the expected Euler characteristic [1]. This gives a new way of setting a threshold on a smooth Gaussian random field and hence controlling the family wise error.

For fMRI, inference can be drawn at different levels using different tests. The following tests pertain to different levels of inference for an activation profile, namely a voxel, a cluster and a set level with decreasing localizing power but potentially increasing sensitivity.

1. Test for the intensity of an activation (voxel level).

2. Test for both peak height and spatial extent of an activation (cluster level).

3. Test for significance of a set of regions (set level).

After the inference on single subject (first-level analysis) has been obtained, the inference can be generalized to the population (second-level analysis or group study). This is achieved by the success of the following two procedures. i) Spatial normalization: The commonly used methods include whole-brain normalization, sulcal matching and functional matching [53]. ii) Group inference: This is achieved by random-effect analysis [21] and Bayesian analysis [66]. The basic idea is to treat the effect size or parameters obtained in the single subject analysis as samples from a large population

[1]The Euler characteristic of an image is a property of the image after it has been thresholded.

and by using random effect analysis or Bayesian analysis, the inference of the effect on the group level can be obtained.

1.3 Thesis Contribution and Organization

This thesis presents robust and efficient algorithms for fMRI signal processing and data analysis which is validated by both simulated and experimental fMRI data. Advanced schemes are developed in order to achieve better detection of the activated regions of the brain and estimation of the hemodynamic response of the brain functions. We only consider the single-subject (first-level) analysis in this thesis. Analysis of group studies (second-level experiments) requires further analysis.

The remainder of the thesis is divided into three parts.

The first part (Chapters 2 – 4) deals with the Bayesian analysis of fMRI data. Different from the traditional analysis methods which are based on the null hypothesis, the Bayesian methods could give the accurate probability of an activation and avoid the multiple comparison problem encountered in the traditional null hypothesis methods.

In Chapter 2, a sparse Bayesian learning method to flexibly determine the design matrix in the GLM is proposed. The construction of the design matrix is critical to the accurate detection of activation regions of the brain in fMRI. It should be flexible to capture the unknown slowly varying drifts as well as robust enough to avoid overfitting. The method proposed in this chapter lets the data itself determine the form of the regressors in the design matrix. It automatically finds those regressors that are relevant to the generation of the fMRI data and discards the others that are irrelevant. The approach integrates the advantages of model-driven and data-driven methods for fMRI data analysis.

The assumption of noise stationarity in the fMRI data analysis may lead to the loss

of crucial dynamic features of the data and thus result in inaccurate activation detection. In Chapter 3, a Bayesian approach is proposed to analyze the fMRI data with two non-stationary noise models (the time-varying variance noise model and the fractional noise model). The covariance matrices of the time-varying variance noise and the fractional noise after wavelet transform are diagonal matrices. This covariance matrix is estimated using a Bayesian estimator which also gives an accurate estimate of the weights in the GLM. Therefore, this method could enhance the accuracy of the detection of activated regions in the brain.

In Chapter 4, we extend the method introduced in Chapter 3, and incorporat the drift in the GLM. Based on the observation that the slowly varying drift resides in a subspace spanned only by large scale wavelets, we propose a modified GLM in the wavelet domain under a Bayesian framework. This modified GLM estimates the activation parameters at each scale of wavelet decomposition. Then, a model selection criterion – Confidence Interval Criterion (CIC) – based on the results from the modified GLM is proposed to model the drift. The Bayesian estimator accurately captures the noise structure and hence results in a robust estimation of the parameters in the GLM. Besides, the proposed model selection criterion works well and efficiently models and removes the drift.

The second part (Chapters 5 and 6) of this thesis investigates the estimation of the HDR under different experimental scenarios (linear and nonlinear). In Chapter 5, a linear spatiotemporal adaptive filter is proposed to model the spatial activation patterns as well as the HDR to the event-related stimulus. The well-known least mean square (LMS) adaptive algorithm is used for estimating the coefficients of the spatiotemporal filter. The method proposed is proved to be equivalent to the CCA method. It is then extended to multiple event type scenarios to estimate the HDRs of each event type.

In Chapter 6, we investigate the nonlinear method to estimate the form of the HDR.

The measured BOLD signals to a particular processing task (for example, rapid event-related fMRI design) show nonlinear properties and vary with different brain regions and subjects. In this chapter, radial basis function (RBF) neural network (a powerful technique for modelling nonlinearities) is proposed to model the dynamics underlying the fMRI data. The equivalence of the proposed method to the existing Volterra series method has been demonstrated and validated in this chapter. It is shown that the first order Volterra kernel which is equivalent to the HDR can be deduced from the parameters of the RBF neural network.

The third part (Chapter 7) of this thesis is mainly focused on the modelling methods of fMRI data. Being complex and noisy, 4-D fMRI data sets show low SNR and nonlinear properties. To enhance the SNR, signal modelling methods are proposed to reconstruct the BOLD signal from the noisy fMRI data. To capture the nonlinear dynamics of the fMRI signal, nonlinear autoregressive with exogenous inputs (NARX) neural networks are investigated to model the fMRI system from the measured fMRI signal and available experimental designs. The NARX neural network is able to capture the dynamics of the fMRI signal and could provide a promising method to model and reconstruct the dynamics existing in the fMRI data.

Chapter 8 provides a summary and general conclusions of the thesis along with recommendations for future research.

This thesis could provide neuroscientists and psychologists with an advanced analysis tool to cope with various difficulties faced in fMRI data analysis. This may contribute to the better understanding of the nature of the fMRI signal and be of importance in explaining the underlying mechanisms.

Chapter 2

Sparse Bayesian Method for Determination of Flexible Design Matrix in fMRI Data Analysis

2.1 Introduction

In Section 1.2.3, we briefly introduced the model-driven and data-driven methods used in fMRI data analysis. Both methods have their own advantages and disadvantages. The data-driven methods are flexible especially when an appropriate data generation model is not available. However, the necessity to explore the whole data set leads to high computational demands and difficulties in interpreting the results. The model-driven methods on the other hand assumes an underlying model taking into account the information provided by the experimental paradigm; they often require less computation and lead to an easier interpretation of the results. However, the model-driven methods may impose an improper model on the data and lead to misinterpretation. In particular, the BOLD response and the interferences may vary for different subjects and for different

regions of the brain. The complexity of the fMRI data requires a method flexible enough to accommodate the variations of the signal and that can adapt to different noises and interferences.

Recent work has suggested that the combination of data-driven and model-driven methods could give better results for fMRI data analysis. In [36], the author proposed a hybrid approach which uses the results of ICA to form the regressors in the GLM. In [48], the ICA with reference (ICA-R) which uses the information of the input stimuli as reference was proposed. In [49], a semi-blind ICA of fMRI incorporating prior information about the paradigm time course was introduced. These methods are based on the powerful data-driven techniques and use some information of the experimental paradigm to guide the analysis. These methods help to understand the fMRI data analysis problem from data-driven methods towards model-driven methods.

In this chapter, a mixed model which starts with model-driven methods and utilizes the data-driven methods to guide the selection of the regressors is proposed. This method assumes a generalized linear model. However, instead of specifying the whole design matrix before analysis (as in the GLM method), only the BOLD response regressor in the design matrix is specified and the data automatically determines the remaining regressors through sparse Bayesian learning. Furthermore, some evidence about the existence of the regressors in the design matrix is obtained through the learning procedure and the irrelevant regressors are discarded to avoid overfitting. This new method integrates the advantages of the data-driven and model-driven methods and gives full flexibility to let the data itself determine its regressors in the design matrix.

2.2 General Linear Model

The aim of fMRI data analysis is to determine the activated regions of the brain (i.e. to determine which voxels embody the BOLD responses to the experimental task). Normally, it is the temporal properties that determine whether a voxel is activated or not. Hence, the fMRI data is processed voxel by voxel and the time series at each voxel is investigated [50]. The GLM imposes a general linear model to the time series of each voxel. The parameters of the model are then tested to draw a conclusion on the activation of the voxel.

The GLM has been introduced in detail in Section 1.2.3. Here, we reiterate its major properties for the convenience of reference. Let $\mathbf{y}$ denote the vector of observed/measured intensity changes in one voxel of the fMRI dataset, with dimensions $T \times 1$, where T is the total length of the time series or indices of observations (e.g. scans). GLM is then given by:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \boldsymbol{\epsilon} \tag{2.1}$$

in which $\mathbf{w}$ is a vector of parameters with dimensions $R \times 1$ which could be found by the least square methods. $\boldsymbol{\epsilon} = [\epsilon_1, \cdots, \epsilon_T]^T$ is the vector of error terms which is assumed to be normal $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ (after prewhitening). The design matrix $\mathbf{\Phi}$ has dimensions $T \times R$ with each row corresponding to one time point (scan) of the regressors, and the columns corresponding to the different explanatory variables or regressors in the model.

In GLM, the design matrix is specified prior to the analysis and will not change during the analysis procedure. The regressors (or the column vectors of the design matrix) often consist of a canonical BOLD response (and its derivatives), a vector of constant value 1 representing mean value and several discrete cosine transform (DCT) basis functions representing the high-pass filter to remove the unwanted low-frequency components from the data [53].

The design matrix in GLM is not flexible and may cause problems. The number of the DCT waveforms to be included in the design matrix should be determined carefully before analysis. In *SPM*, this is implemented by specifying the highpass cutoff. Generally, too many basis functions would lead to an overfitting problem, while too few basis functions may not filter out the slowly varying interference efficiently. So, the selection of the number of basis functions is a very tricky issue. Furthermore, the inclusion of the canonical BOLD response regressors into the design matrix may result in a deviated model if the voxel is not activated. For accurate detection, an efficient method is required, which could switch the BOLD response regressor ON and OFF according to the measured data during the data learning procedure.

In the method proposed, the initial design matrix includes the BOLD response regressor, the vector of constant value 1 and a set of general nonlinear functions (instead of DCT basis fucntions) that account for the slowly varying drifts and trends. Compared to the design matrix used in the GLM, the proposed model is much more flexible due to the use of general nonlinear functions which in fact construct a flexible subspace and could capture signal variations more efficiently. Besides, different from the GLM in which the design matrix will not change during the whole analysis, the proposed method could adapt/adjust the design matrix according to the data. This is implemented under the Bayesian framework. By using the sparse Bayesian learning, the regressors could be learned from the data automatically to best account for the observed signal. The unwanted regressors are then removed from the design matrix to avoid the overfitting problem.

2.3 Sparse Bayesian Learning

Suppose the initial number of regressors in the design matrix is $(T+2)$ and the design matrix is $T \times (T+2)$, which is defined as

$$\mathbf{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_T, \mathbf{b}, \mathbf{1}], \tag{2.2}$$

where the first T regressors in the design matrix are flexible radial basis functions which are further defined as Gaussian:

$$\boldsymbol{\phi}_i(j) = \exp\Big(-\frac{1}{2v^2}(i-j)^2\Big), \tag{2.3}$$

where $i, j = 1, \cdots, T$ are the time points (or scan indices), and v is the width of the basis function. The $(T+1)$th regressor $\mathbf{b}$ is the vector of the canonical BOLD response which is the convolution of the experimental paradigm and the HRF [53]; the $(T+2)$th regressor is a vector of constant value 1.

The aim of sparse Bayesian learning is to determine a suitable set of M (where $M \leq T+2$) regressors and their corresponding weighting coefficients $\mathbf{w}$. Since the noise $\epsilon_n; n = 1, \cdots, T$ in Eq. (2.1) are assumed as independent samples of Gaussian noise with zero mean and variance σ^2 (for temporally correlated errors, the prewhitening procedure is required), the output signal $y_n; n = 1, \cdots, T$ is also Gaussian distributed with variance σ^2. With the assumption of the independence of the output y_n, the likelihood of the whole data set can be derived as:

$$p(\mathbf{y}|\mathbf{w}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\{-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{\Phi}\mathbf{w}\|^2\} \tag{2.4}$$

where $\mathbf{y} = [y_1, \cdots, y_T]^T$ is the vector formed by the measured fMRI signal, which is defined by the GLM; $\mathbf{w} = [w_1, \cdots, w_M]^T$ and $\mathbf{\Phi}$ is the $T \times M$ ($M \leq T+2$) design matrix. For the initial settings, M is equal to $T+2$ and $\mathbf{\Phi}$ has the form defined in Eq. (2.2).

To avoid over-fitting, some of the parameters are constrained by defining prior probability distributions over them. Especially, the weight vector $\mathbf{w}$ is treated as a random variable with Gaussian prior probability (with zero-mean and variance α^{-1}) [54]:

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} \mathcal{N}(0, \alpha_i^{-1}) \tag{2.5}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_M]^T$ is a vector of M hyperparameters. This prior probability in Eq. (2.5) is known as an automatic relevance determination (ARD) prior [54, 55]. This prior could also be set up using the general parametric empirical Bayesian (PEB) framework [56] by assigning a covariance basis function Q_i to each regression coefficients w_i. This prior means, at this time, that the best guess about the value of w_i is 0, and that α_i^{-1} represents the uncertainty about this guess. Furthermore, the α_i's and the noise variance σ^2 are defined with uniform distributions.

The basic idea of Bayesian learning is to maximize the posterior probability over the weights $\mathbf{w}$ and the hyperparameters $\boldsymbol{\alpha}$ and σ^2 given the data $\mathbf{y}$, i.e. maximizing $p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{y})$. This posterior is further decomposed as:

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{y}) = p(\mathbf{w}|\mathbf{y}, \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2|\mathbf{y}). \tag{2.6}$$

Maximizing the left hand side of Eq. (2.6) is equivalent to maximizing the two probabilities on the right hand side. The first probability on the right hand side is also a Gaussian distribution given by

$$p(\mathbf{w}|\mathbf{y}, \boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-M/2} |\boldsymbol{\Lambda}|^{-1/2} \exp\Big(-\frac{1}{2}(\mathbf{w} - \mathbf{u})^T \boldsymbol{\Lambda}^{-1} (\mathbf{w} - \mathbf{u}) \Big) \tag{2.7}$$

with the posterior covariance $\boldsymbol{\Lambda}$ and mean $\mathbf{u}$ being

$$\boldsymbol{\Lambda} = (\sigma^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{A})^{-1} \tag{2.8}$$

$$\mathbf{u} = \sigma^{-2} \boldsymbol{\Lambda} \boldsymbol{\Phi}^T \mathbf{y} \tag{2.9}$$

where $\mathbf{A} = diag(\alpha_1, \alpha_2, \cdots, \alpha_M)$.

The maximum is clearly at the mean, i.e. the best estimate of the weights $\mathbf{w}$ is

$$\hat{\mathbf{w}} = \mathbf{u}. \tag{2.10}$$

The second probability on the right hand side of Eq. (2.6) is further decomposed and its maximization is equivalent to the maximization of $p(\boldsymbol{\alpha}, \sigma^2|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha})p(\sigma^2)$ with respect to $\boldsymbol{\alpha}$ and σ^2. For uniform prior distributions, the following update equations are derived [54, 55]:

$$\alpha_i^{new} = \frac{\gamma_i}{\hat{w}_i^2} \tag{2.11}$$

$$\gamma_i = 1 - \alpha_i \boldsymbol{\Lambda}_{ii} \tag{2.12}$$

$$(\sigma^2)^{new} = \frac{\|\mathbf{y} - \boldsymbol{\Phi}\hat{\mathbf{w}}\|^2}{T - \sum_{i=1}^{M} \gamma_i} \tag{2.13}$$

where $\hat{w}_i$ is the i-th element of the posterior mean vector $\hat{\mathbf{w}}$ in Eq. (2.10) and $\boldsymbol{\Lambda}_{ii}$ is the i-th diagonal element of the posterior covariance matrix $\boldsymbol{\Lambda}$ in Eq. (2.8), T is the total number of data points (in this case, the length of the fMRI time series), γ_i can be interpreted as a measure of how "well-determined" its corresponding parameter $\hat{w}_i$ is by the data [55]. If $\gamma_i \approx 0$, from Eq. (2.12) it can be seen that $\boldsymbol{\Lambda}_{ii} \approx \alpha_i^{-1}$, which means that $\hat{w}_i$ is highly constrained by the prior. Conversely, if $\gamma_i \approx 1$, α_i is small and $\hat{w}_i$ fits the data.

The learning algorithm is to iteratively update Eq. (2.11) to Eq. (2.13), together with $\boldsymbol{\Lambda}$ and $\hat{\mathbf{w}}$ from Eq. (2.8) to Eq. (2.10) until some suitable convergence criteria are satisfied. In practice, during the learning process, many of the α_i's approach infinity, which means the probability density $p(w_i|\mathbf{y}, \boldsymbol{\alpha}, \sigma^2)$ becomes infinitely large at zero. This shows that it is certain to some extent that the particular w_i should be zero given the data at hand. Thus, the corresponding regressor functions could be 'pruned' and fewer regressors are kept to construct a suitable design matrix. These regression vectors

are called Relevance Vectors in [54]. Specifically, the drifts and trends will be modelled by automatically selecting a minimum number of flexible basis functions which could capture the slow variations. The canonical BOLD response regressor may be 'pruned' as well given the fMRI data. Thus a method is achieved to determine a flexible design matrix in fMRI data analysis. This method could capture the underlying slowly varying drift in the fMRI data and avoid overfitting.

Furthermore, besides the estimation of the weights $\mathbf{w}$, Bayesian learning also provides the estimate of the additive noise level σ^2. This term is important in the statistical evaluation of the weights $\mathbf{w}$, which are evaluated by defining a contrast vector $\mathbf{c}$ and calculating t as [57]:

$$t = \frac{\mathbf{c}^T \hat{\mathbf{w}}}{\sqrt{\mathbf{c}^T \mathbf{\Lambda} \mathbf{c}}}. \tag{2.14}$$

These t obtained at each voxel are then used to form a Statistical Parametric Map and a threshold is determined to find the activation regions of the brain.

2.4 Results and Discussion

The proposed approach was first tested on simulated data with the total number of time points (or scans) equal to 100. These signals simulate block design fMRI signals. Figure 2.1 shows a sketch of a block design experiment and its square waveform representation. When the stimulus is applied (ON), the representing waveform has the value 1; while when the stimulus is absent (OFF), the representing waveform has the value 0. A comparison between the proposed approach and the conventional t-test method is also given. Then, this new approach is applied on the experimental fMRI data. The results are discussed in detail in the following sections.

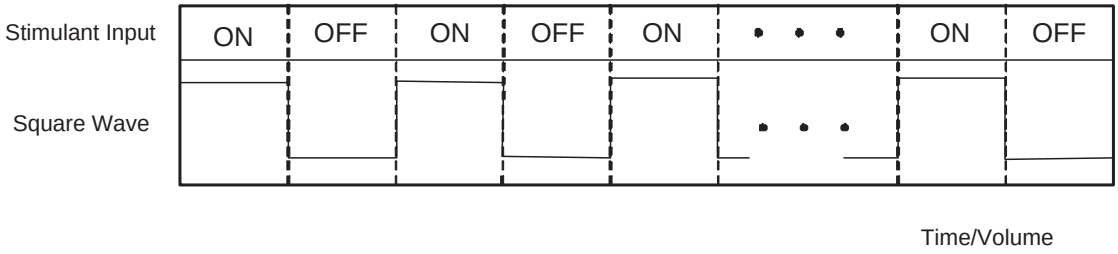


Figure 2.1 Illustration of a block design and its square waveform representation.

2.4.1 Simulated Data

The aim of this study is to investigate whether the proposed approach could capture the underlying slowly varying drift and can give some evidence to include the BOLD response regressor in the design matrix or not. Particularly, four types of signals are simulated and tested:

Type 1: BOLD response + Constant mean value + Noise;

Type 2: Constant mean value + Noise;

Type 3: BOLD response + Constant mean value + Drift + Noise;

Type 4: Constant mean value + Drift + Noise.

The BOLD response is simulated by assuming the brain and MR acquisition system to be a linear system. Thus, BOLD responses are generated as the convolution of the experimental paradigm and the HRF. The HRF is chosen as the difference between two gamma functions as introduced in Chapter 1 (Section 1.2.2). The mean value is a randomly generated constant representing the grey level of the specified voxel in the brain image. The drift is simulated by a slowly varying sine wave. The Gaussian noise with zero mean and variance 1 is added to the simulated signal, which results a signal-to-noise ratio (SNR) of about -7dB.

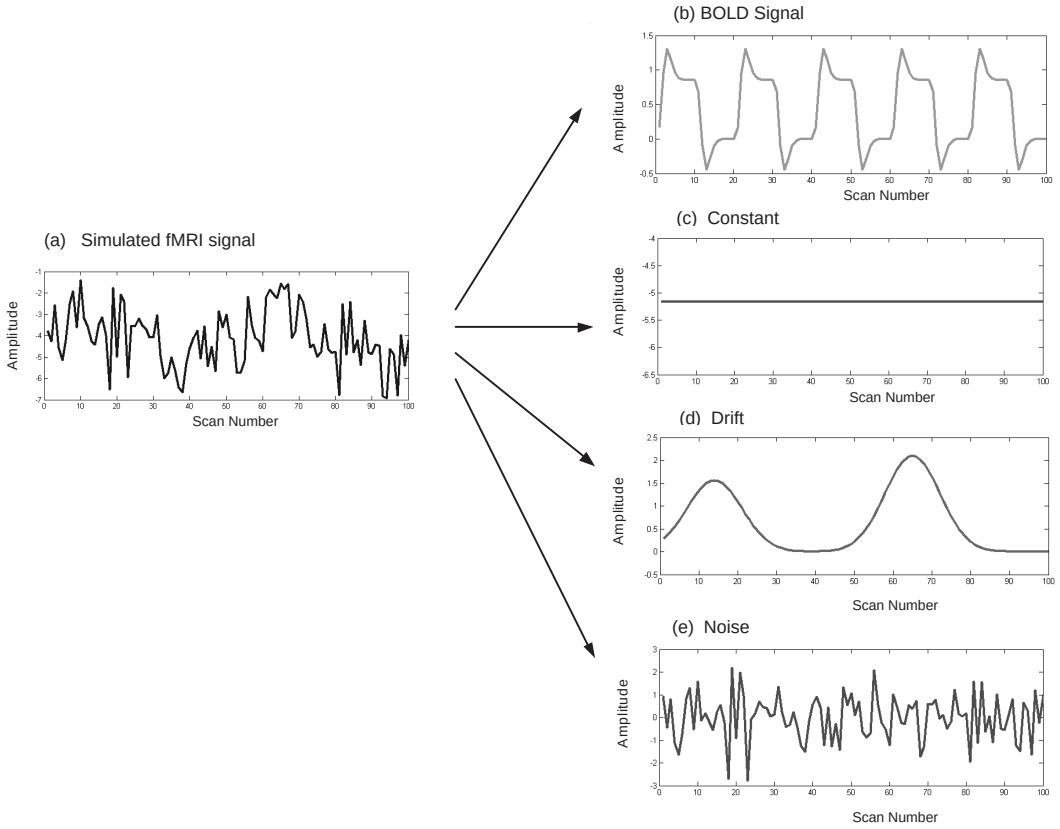
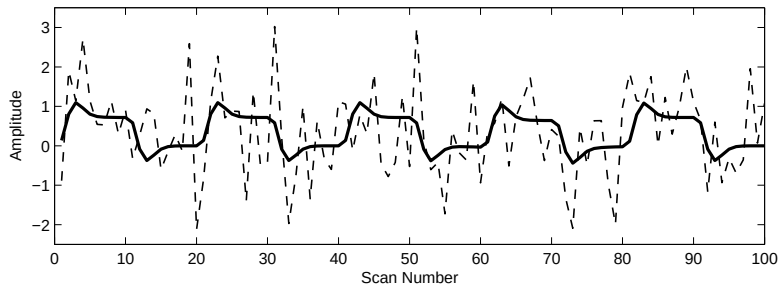


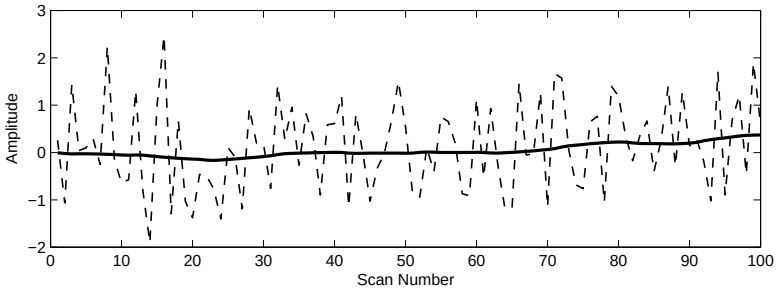
Figure 2.2 A simulated BOLD signal corrupted by drift and noise (Type 3) is decomposed by the proposed approach into different sources. (a) Simulated noisy fMRI signal; (b) BOLD response; (c) Constant mean value; (d) Slowly varying drift; (e) Noise.

Figure 2.2 shows the result of analyzing the Type 3 signal. Figure 2.2(a) is the simulated BOLD signal corrupted by the drift and noise. This noisy signal is decomposed into the pure BOLD response (Fig. 2.2(b)), the constant mean value (Fig. 2.2(c)), the slowly varying drift (Fig. 2.2(d)) and the noise (Fig. 2.2(e)) after fitting the simulated data to the model learned by the proposed method. Initially, the design matrix has the dimensions 100×102 ($T = 100$). Through sparse Bayesian learning, the proposed approach discards the irrelevant columns in the design matrix and retains only the relevant columns. The number of final regressors in the design matrix for this simulated signal is reduced to 4 (i.e. the final design matrix has the dimensions 100×4), with one BOLD response regressor (with $w_4 = 1.2040; \alpha_4 = 0.6670$), one constant mean value regressor (with $w_3 = -5.1650; \alpha_3 = 0.0374$) and two basis functions to account for the slowly varying drift (with $w_2 = 1.5631; \alpha_2 = 0.3913$ and $w_1 = 2.0976; \alpha_1 = 0.2215$). The values of w_i's (different from zero) and α_i's (not too large) give us some evidence that these four regressors are relevant regressors and the design matrix formed by these four regressors is suitable for the simulated signal. The value of t for this signal is 4.145.

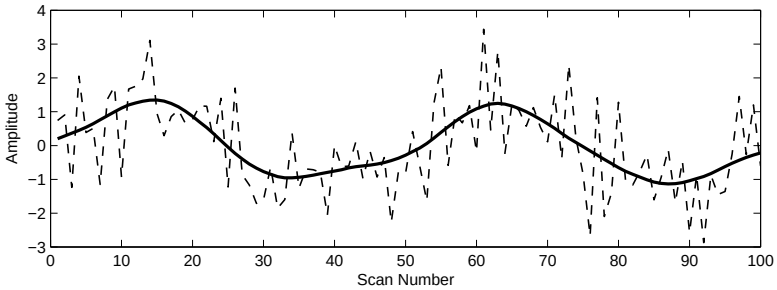
Figure 2.3 shows the result of using the proposed model determination method to analyze the other three types of simulated signals. The simulated noisy signal is denoted in dashed line, while the reconstructed signal is denoted in solid line. The initial design matrix is formed as in Eq. (2.2), however, during the learning process, the BOLD response regressor as well as some of the other regressors in the design matrix may be discarded since they are irrelevant to the generation of the data. In Fig. 2.3(a), when there is BOLD response in the simulated signals, the corresponding regressor column is kept in the design matrix after the learning. While in Fig. 2.3(b) and Fig. 2.3(c), when the signal contains no BOLD response and only consists of either drift and noise or only noise, the proposed method will automatically switch OFF the BOLD response regressor column in the design matrix through learning. In this simulation, for Type 2 and Type 4



(a)



(b)



(c)

Figure 2.3 The simulated signals and their reconstruction. (a) Type 1: BOLD response corrupted by noise; (b) Type 2: No BOLD response, only noise; (c) Type 4: No BOLD response, only noise and drift.

Table 2.1 The error rate of different *t*-value thresholds for different types of signals

***t*-threshold**	**Type 1** (%)	**Type 3** (%)	**Type 2** (%)	**Type 4** (%)
1.0	0.25	0.85	10.45	5.70
1.5	0.65	1.85	5.25	3.25
2.0	1.95	3.80	2.50	1.60
2.5	4.05	8.05	0.75	0.60
3.0	9.60	13.5	0.20	0.15

signals, in most cases (around 90%), the BOLD response regressor is discarded through learning and the *t*-test values are thus 0; but in a few cases (around 10%), the BOLD regressor is not discarded in the design matrix, resulting in the *t*-test value greater than zero. However, these *t*-test values are normally small.

Table 2.1 summarizes the error rates under different *t*-value thresholds for the four types of simulated signals. For this simulation, 10^4 realizations of each type of signals were generated and processed using the proposed method. The *t*-test was carried out under the null hypothesis H_0 that there is no BOLD response in the simulated signal. Both Type I error (rejecting the null hypothesis H_0 when it is true) and Type II error (not rejecting the null hypothesis H_0 when it is false) [43] are computed. In this table, the errors are displayed under different signal types and different *t*-test values. For Type 1 and Type 3 signals, the error rates are of Type II error, while for Type 2 and Type 4 signals, the error rates are of Type I error. From this table, it could be seen that the proposed method can make correct decisions with small error rates by choosing the suitable t threshold.

A comparison of the detection ability of the proposed method and the conventional

t-test method (it tests whether the means of the BOLD signal in the experimental condition and control condition are the same) is also investigated by using receiver operator characteristic (ROC) analysis [58]. The ROC method reflects the ability of different processing methods to detect most of the real activations while minimizing the detection of false activations. For ROC analysis, two values were computed: the true positive ratio (proportion of correctly detected voxels to all voxels with added activations) and the false positive ratio (proportion of voxels that were incorrectly recognized as active to all voxels without added activation). The ROC curve is a plot of true positive ratio versus false positive ratio under different threshold values. In this simulation, one slice from the fMRI data set was used to form the background and Gaussian noise was added to construct the 2-D time series. Simulated BOLD responses were added to specific areas to simulate the active brain areas. For the simulated activated voxels, the signal is Type 1, while for inactivated voxels, the signal is Type 2. In this simulation, the drift is not added. The conventional *t*-test and the proposed method were applied to this simulated data. Figure 2.4 shows the ROC curves under both the conventional *t*-test method and the proposed Bayesian learning method.

Observing the ROC curves of Fig. 2.4, it can be seen that at the same false positive rate, the proposed method could actually detect more real activations. This clearly proves the better performance of the proposed method compared to the conventional *t*-test method.

2.4.2 Experimental fMRI Data

The proposed method is also validated on the experimental fMRI data (*DATA-BLOCK*) introduced in Chapter 1. Figure 2.5 shows the results of both the conventional *t*-test and the proposed method. It is clearly seen from the figure that the visual cortical area has

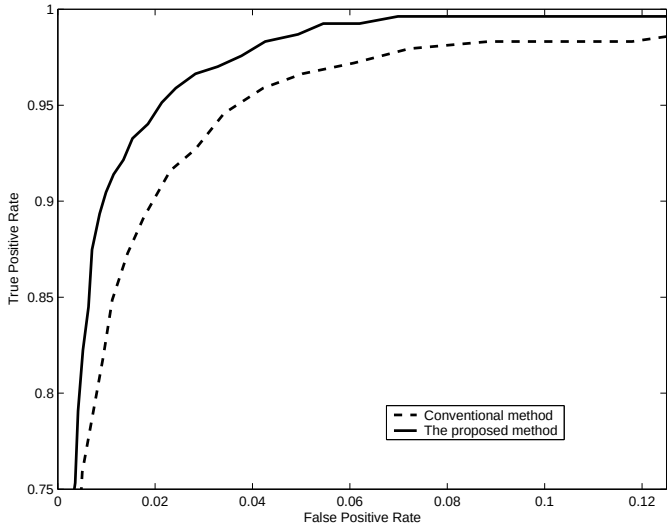
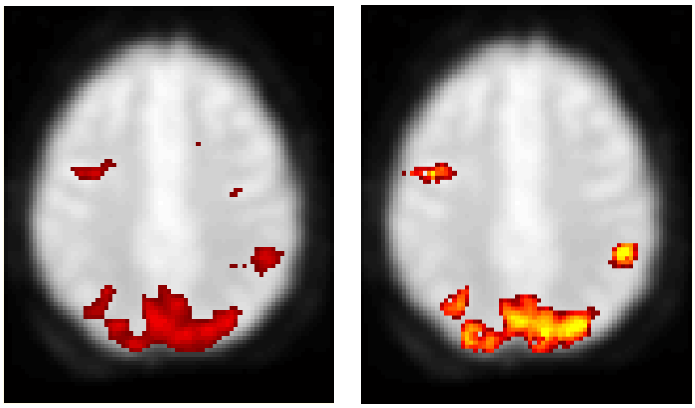


Figure 2.4 ROC curves for simulated noisy data (2D plus time).

been identified.

For the experimental fMRI data, a quantitative comparison of the performance between the proposed method and conventional method is difficult. This is because that we lack a reference which could serve as the true activation of the brain. Therefore, we refer to the *t*-test scores for a quantitative comparison. If the degrees of freedom are the same, the higher *t*-test score means that the false positive would be smaller under the null hypothesis. And a method which could detect the activated regions in the brain with higher *t*-test score is better compared to the other methods. As shown in Fig. 2.5, the proposed method could detect the activated regions of the brain with *t*-test threshold 6.3. However, under this threshold, the conventional *t*-test method cannot detect the activated regions. If the threshold is decreased to 3.8, the activated regions of the brain could be detected by the conventional *t*-test method. The results that the proposed method could detect the activation regions with less erratic points and higher *t*-threshold values clearly confirms that the proposed method is more robust than the conventional *t*-test methods.



(a) (b)

Figure 2.5 Results of fMRI data analysis to a visuospatial processing task. (a) Conventional t-test ($t > 3.8, p < 0.05$); (b) The proposed method with Sparse Bayesian Learning ($t > 6.3, p < 0.05$).

2.5 Conclusion

In fMRI data analysis, especially in GLM, the construction of the design matrix is very important to the data analysis. A flexible design matrix that can account for the measured data while not being too flexible to induce interference is desired. In this chapter, the sparse Bayesian learning is applied to determine the regressors in the design matrix. The initial design matrix is a very flexible one which may induce the overfitting problem. Through sparse Bayesian learning, some confidence about the existence of the regressors is obtained and those regressors that are unlikely present and irrelevant to the measured fMRI data are discarded. Thus, a method to determine a flexible design matrix is achieved. This method could capture any unknown underlying slowly varying drift and avoid overfitting problem. It imposes a flexible model to the data, and lets the data itself determine what the model should be like. This new method integrates the advantages of the data-driven and model-driven methods and gives full flexibility to let

the data itself determine its regressors in the design matrix. Validation results from both simulated and experimental fMRI data show that the proposed method provides a much better performance than the conventional *t*-test method and enhances the ability of brain activity detection.

The method proposed in this chapter assumes that the noise is stationary, that is, the variance does not change with time. However, the noises in fMRI data also show nonstationary properties. In Chapter 3, this method is generalized to cope with two types of nonstationary noises in fMRI.

Chapter 3

fMRI Data Analysis with Nonstationary Noise Models: A Bayesian Approach

3.1 Introduction

In the previous chapter, a flexible design matrix determination method based on sparse Bayesian learning is proposed. This method assumes that the noise is stationary, that is, the underlying noise model is time-invariant. However, since fMRI noise is inherently time-varying, the stationarity assumption may not be appropriate considering the complexity of the data.

In [59], a spatially nonstationary but temporally stationary spatio-temporal noise model was developed to fit the fMRI time series. However, the nonstationarities can also exist temporally. Several factors may induce the nonstationarities in the fMRI time series. Neurophysiological processes, such as the number of neurons involved in a specific activity at different time points and the extraneous auditory and visual stimuli or background memory processes, may cause the variance of the noise to change [60]. Besides, the abrupt movements of the subjects are another source of the nonstationarity of

the noise in fMRI [61]. For example, the abrupt movements of the lower jaw of subjects may affect only a few consecutive fMRI images, causing the variance of these images to be high and violating the stationarity assumption of the noise. Moreover, the $1/f-$like noise or nonstationary fractional noises in the fMRI data have been reported and investigated by many researchers recently [62][63]. These factors show that the assumption of stationary noise may not be realistic to cope with the complex fMRI data. The general nonstationary noise model to incorporate the time-varying stochastic properties of fMRI noise (such as transients and sudden changes in the fMRI data) is needed for a more accurate fMRI data analysis.

In this chapter, two noise models (time-varying variance model [61] and fractional noise model [63]) are investigated to capture the nonstationarities in the fMRI data. Different from the assumption (spatially nonstationary but temporally stationary) used in [59], these two noise models are temporally nonstationary. The covariance matrices of the time-varying variance noise and the fractional noise after wavelet transform are diagonal matrices. This property is investigated under the Bayesian framework. The Bayesian methods are utilized both in single-subject (first-level) analysis [64] [65] [66] and group studies (multi-level experiments) [67] [68]. In [64], a variational Bayesian method was proposed to infer on the GLM for fMRI data. The haemodynamic response basis functions in the GLM using variational Bayesian are then constrained to sensible HRF shape in [65]. In [69], a Bayesian method with sparse spatial basis function priors was proposed to incorporate the spatial properties of the fMRI data. In [67] and [68], the Bayesian approach was utilized to deal with the inference problem on the hierarchical linear models for fMRI group analysis. These works show the advantages of the Bayesian analysis for the fMRI data. The classical approaches, such as SPM, working under the mechanism of rejecting or accepting the null hypothesis, have some limitations. For example, the p values in SPM are the probabilities of the effects under the

null hypothesis which states there is no activation in each voxel. These p values do not reflect the probabilities of the actual effects when the voxel is truly activated. Bayesian approaches, on the contrary, can give the probability that a voxel is activated or the probability that the effect is greater than some threshold value. Hence, these limitations of the classical approaches could be overcome using the Bayesian methods.

3.2 Nonstationary Noise Models

Let $\mathbf{y}_n$ and $\boldsymbol{\epsilon}_n$ represent the the measured fMRI signal and noise at the n^{th} voxel of fMRI data respectively. The ordinary least squares (OLS) estimator of the weight vector $\mathbf{w}_n$ in the GLM is:

$$\hat{\mathbf{w}}_{n(\text{OLS})} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{y}_n. \tag{3.1}$$

This estimator assumes that the noise is independent and identically distributed (i.i.d.) Gaussian white noise. Clearly, this assumption is inappropriate considering the temporal auto-correlations and the nonstationary nature of the fMRI signals. To deal with the temporal auto-correlations in the fMRI data, coloring and pre-whitening methods are proposed [42][52]. Many nonstationary noise models [59][60][63] are also proposed to model the noises in the fMRI data. Among these proposed noise models, two nonstationary noise models are considered in this chapter.

3.2.1 Time-varying Variance Model

The abrupt physical movements of the subjects in the fMRI experiment, such as the lower jaw movements, may affect a few fMRI images, causing the variances of these fMRI images to rise and breaking the stationarity assumption. In [61], the time-varying variance noise model is introduced and it is reported that such increase in noise variance

is multiplicative. That is, the noise interferences affect the variances in the fMRI images multiplicatively, but not additively. Hence, the variance at each time point is modelled as a scaled version of the overall variance in that voxel. This time-dependent noise is modelled as a Gaussian process $\boldsymbol{\epsilon}_n \sim \mathcal{N}(0, \mathbf{B}_n^{-1})$ with the precision matrix $\mathbf{B}_n$ (i.e, the inverse of the covariance matrix) defined as:

$$\mathbf{B}_n = \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_T \end{pmatrix} \beta_n = \mathbf{S}\beta_n \tag{3.2}$$

where $s_1, s_2, \dots, s_T$ are the scaling parameters, $\mathbf{S}$ is a $T \times T$ scaling diagonal matrix and β_n is a scalar representing the overall noise precision in the n^{th} voxel. This noise precision matrix shows that the precisions (or inverse of variance) of the n^{th} voxel at different time points are scaled versions of the overall precision β_n observed in that voxel. In [61], it is testified that the scaling change of the precision has a spatial uniformity. This shows that if the noise of the voxels in some portion of the image changes, the noise in all other voxels in the same image will also change to a similar degree. Thus, although the overall variance is different in different voxels (i.e. different β_n in different voxels), the scaling parameters are assumed to be the same across the whole image (i.e. the scaling matrix $\mathbf{S}$ is assumed to be the same for all the voxels).

Under the assumption that the precision matrix of noise $\boldsymbol{\epsilon}_n$ is $\mathbf{B}_n = \mathbf{S}\beta_n$, the maximum likelihood estimate of $\mathbf{w}_n$ in the GLM is a weighted least squares (WLS) estimate:

$$\hat{\mathbf{w}}_{n(\text{WLS})} = (\mathbf{\Phi}^T\mathbf{S}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{S}\mathbf{y}_n. \tag{3.3}$$

However, the WLS estimate requires us to have an accurate estimate of the scaling matrix $\mathbf{S}$.

A traditional methods to estimate the scaling matrix $\mathbf{S}$ make use of the residuals $\mathbf{r}_n$ of the ordinary least squares (OLS) estimates, where the residuals are defined as:

$$\mathbf{r}_n = \mathbf{y}_n - \mathbf{\Phi}\hat{\mathbf{w}}_{n(\text{OLS})} = \mathbf{y}_n - \mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{y}_n. \tag{3.4}$$

The overall precision β_n is estimated by [44]:

$$\hat{\beta}_n = \frac{T - \text{rank}(\mathbf{\Phi})}{\mathbf{r}_n^T\mathbf{r}_n}, \tag{3.5}$$

where $\mathbf{r}_n^T\mathbf{r}_n$ is the sum of the squares of the residuals and $(T-\text{rank}(\mathbf{\Phi}))$ is the appropriate degrees of freedom.

Considering that $\mathbf{r}_n\mathbf{r}_n^T$ is the estimate of the covariance matrix of the noise (i.e. $\mathbf{B}_n^{-1}$ as shown in Eq. (3.2)), the inverse of the scaling parameters $\mathbf{s}_{inv} = \text{diag}(\mathbf{S}^{-1}) = [s_1^{-1}, s_2^{-1}, \cdots, s_T^{-1}]^T$ is estimated by averaging these variance estimates weighted by β_n over the N voxels:

$$\hat{\mathbf{s}}_{inv} = \frac{\sum_{n=1}^{N} \text{diag}(\beta_n\mathbf{r}_n\mathbf{r}_n^T)}{N}, \tag{3.6}$$

where the operator $\text{diag}(\cdot)$ transforms the diagonals of a square matrix into a column vector. N is the total number of voxels considered due to the spatial uniformity of the scaling parameters.

Although this classical method is simple to implement, it is a biased estimate of the scaling parameters $\mathbf{s}$ as shown in [61] that the precision matrix of $\mathbf{r}_n$ is not equal to $\mathbf{S}\beta_n$. This bias comes from the estimation of $\mathbf{s}$ using the residuals of the OLS estimates which assume the covariance matrix of the residuals to be an identity matrix. This bias may cause the t value in this voxel to be invalid and hence to draw incorrect inference about the activation status of this voxel if the noise covariance matrix is not an identity matrix. In Section 3.3, a Bayesian estimator to accurately estimate the scaling parameters $\mathbf{s}$ is proposed.

3.2.2 Fractional Noise Model

It is reported that the noise in the fMRI time series obtained under the resting or null conditions exhibits long-range auto-correlation in time and $1/f-$like spectral properties [70]. This means that the spectral density $S(|f|) \propto |f|^{\lambda}$ is with the spectral exponent $\lambda < 0$. One of the $1/f-$like processes is the fractional Brownian motion (fBm). It is a zero mean, nonstationary, and nondifferentiable process with the auto-covariance (r) between time t_1 and t_2 defined by the Hurst exponent (H) [71]:

$$r(t_1, t_2) = \frac{1}{2}\delta^2(|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}) \tag{3.7}$$

where

$$\delta^2 = \Gamma(1 - 2H)\frac{\cos(\pi H)}{\pi H}. \tag{3.8}$$

Let the length of the fMRI time series be $T = 2^J$, where J is an integer. Applying the discrete wavelet transform (DWT) to both sides of the GLM, we get a GLM in the wavelet domain [72] as follows:

$$\mathbf{y}_n^W = \mathbf{\Phi}^W \mathbf{w}_n + \boldsymbol{\epsilon}_n^W \tag{3.9}$$

where $\mathbf{y}_n^W$ and $\boldsymbol{\epsilon}_n^W$ are the results of applying the DWT respectively to the data $\mathbf{y}_n$ and noise $\boldsymbol{\epsilon}_n$ up to the maximum scale J, and $\mathbf{\Phi}^W$ is the wavelet transform applying to each columns of the design matrix $\mathbf{\Phi}$.

The wavelet transform of the noise $\boldsymbol{\epsilon}_n$ up to the maximum scale J is denoted as:

$$\boldsymbol{\epsilon}_n^W = [a_{J,1}, d_{J,1}, d_{J-1,1}, d_{J-1,2}, \cdots, d_{1,1}, \cdots, d_{1,2^{J-1}}]^T \tag{3.10}$$

where $a_{J,1}$ is the scaling coefficient (or approximation coefficient) at level J, and $d_{m,k}$ ($k = 1, \cdots, 2^{J-m}$) are the wavelet coefficients (or detail coefficients) at level m ($m = 1, \cdots, J$).

The orthonormal wavelet decomposition is an approximate whitening filter for fBm [73]. The correlation between the wavelet coefficients within any of the scales and the cross-correlation coefficients between different scales is very small [74] and hence can be ignored for any wavelet provided that the number of vanishing moments is sufficiently large (greater than $2H+1$). It is also shown in [74],[75] and [76] that for $1/f-$like noises, the wavelet coefficients $d_{m,k}$ at level m and the scale coefficient $a_{J,1}$ at level J are normally distributed with zero mean and denoted by:

$$\begin{aligned} d_{m,k} &\sim \mathcal{N}(0, V_{d_m}), \\ &\quad m = 1, \cdots, J; k = 1, \cdots, 2^{J-m} \qquad (3.11) \\ a_{J,1} &\sim \mathcal{N}(0, V_{a_J}), \qquad (3.12) \end{aligned}$$

where $V_{d_m} = \text{var}\{d_{m,k}\}$ ($k = 1, \cdots, 2^{J-m}$) is the variance of the wavelet coefficients $d_{m,k}$ at level $m (m = 1, \cdots, J)$ and $V_{a_J} = \text{var}\{a_{J,1}\}$ is the variance of the scaling coefficients $a_{J,1}$ at the level J.

Thus, for $1/f-$like noises, the orthonormal wavelet decomposition behaves like a KLT and the transformed fractional noise $\boldsymbol{\epsilon}_n^W$ could be modelled by Gaussian process $\boldsymbol{\epsilon}_n^W \sim \mathcal{N}(\mathbf{0}, \mathbf{V}^W)$ with the $T \times T$ diagonal covariance matrix defined as:

$$\begin{aligned} \mathbf{V}^W &= Cov(\boldsymbol{\epsilon}_n^W) \\ &= \begin{pmatrix} V_{a_J} & & & & & & \\ & V_{d_J} & & & & & \\ & & V_{d_{J-1}} & & & & \\ & & & V_{d_{J-1}} & & & \\ & & & & \ddots & & \\ & & & & & \begin{pmatrix} V_{d_1} & & \\ & \ddots & \\ & & V_{d_1} \end{pmatrix} \end{pmatrix}. \end{aligned} \qquad (3.13)$$

The diagonal matrix $\mathbf{V}^W$ is composed of V_{a_J} (the variance of scaling coefficients at level J) and J sub-matrices representing the covariance matrix of the wavelet coefficients $d_{m,k}$ at level m ($m = J, J-1, \cdots, 1$). From Eq.(3.11), it is seen that the wavelet

coefficients $d_{m,k}$ at level m are with the same variance V_{d_m}. Hence, the sub-matrix at level m is of identical value V_{d_m} at the diagonal and of dimension $2^{J-m} \times 2^{J-m}$.

Since the wavelet function at different scales are bandpass filters and the scaling function is a low-pass filter, it is verified in [76] that the variance of the wavelet coefficients at level m and the scaling coefficient at level J can be denoted as scaled version of σ^2:

$$\begin{aligned} V_{d_m} &= \sigma^2 S_{d_m}; \quad m = 1, 2, \cdots, J \\ V_{a_J} &= \sigma^2 S_{a_J} \end{aligned} \tag{3.14}$$

where S_{d_m} ($m = 1, 2, \cdots, J$) and S_{a_J} are scaling constants determined by H [72]. Hence, the diagonal elements of the covariance matrix $\mathbf{V}^W$ are scaled versions of σ^2 and can be represented by:

$$\mathbf{V}^W = \begin{pmatrix} S_{a_J} & & & & & & \\ & S_{d_J} & & & & & \\ & & S_{d_{J-1}} & & & & \\ & & & S_{d_{J-1}} & & & \\ & & & & \ddots & & \\ & & & & & S_{d_1} & \\ & & & & & & \ddots \\ & & & & & & & S_{d_1} \end{pmatrix} \sigma^2 \tag{3.15}$$

$$= \text{diag}^{-1}\{(S_{a_J}, S_{d_J}, S_{d_{J-1}}, S_{d_{J-1}}, \cdots, S_{d_1}, \cdots, S_{d_1})\sigma^2\}, \tag{3.16}$$

where $\text{diag}^{-1}\{\cdot\}$ transforms a vector into diagonals of a square matrix.

From the above descriptions, it is seen that the covariance matrix of the fractional noise in the wavelet domain (Eq. (3.16)) has almost the same structure as the covariance matrix of the time-varying variance noise in the time domain (see Eq. (3.2)). The only difference is that for wavelet-transformed fractional noise, the scaling parameters at the same level are of the same values. Since the covariance matrices have similar structure in both cases, these two cases can be put under the same framework and in the next section, we do not distinguish between these two cases unless it is necessary.

3.3 Bayesian Estimator

The fMRI data analysis aims at a sensitive and robust detection of activation regions of the brain under different sensory, motor and cognitive experiments. This requires an accurate estimate of the weights $\mathbf{w}$ in the GLM. For the nonstationary noises presented in the previous section, the WLS and OLS estimators may fail because: (i) the WLS estimator requires an accurate estimate of the covariance matrix which is difficult and (ii) the OLS estimator does not consider the nonstationary properties of the noise. In Chapter 2, a Bayesian method to estimate the parameter $\mathbf{w}$ under the white Gaussian noise with fixed variance is provided. Here, we extend the Bayesian method to the situation where the variance of the noise is changing with the time.

The noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{B}^{-1})$ is assumed as a Gaussian noise with zero mean and diagonal precision matrix $\mathbf{B} = \text{diag}^{-1}\{(s_1, s_2, \cdots, s_T)\beta\}$ as introduced in Section 3.2 (for both time-varying variance noise and fractional noise after wavelet transform). At time k ($k = 1, 2, \cdots, T$), the noise ϵ_k are independent samples from zero-mean Gaussian with precision $s_k\beta$ which may be different at each time point k. Thus, the output signal y_k at time point k is also a Gaussian distributed process with varying precision at each time point. With the assumption of independence of the output at different time points, the likelihood of the whole time series $\mathbf{y}$ can be derived as a multivariate Gaussian density function:

$$\begin{aligned} p(\mathbf{y}|\mathbf{w}, \mathbf{B}) &= \frac{|\mathbf{B}|^{T/2}}{(2\pi)^{T/2}} \exp\{-\frac{1}{2}(\mathbf{y} - \mathbf{\Phi w})^T \mathbf{B}(\mathbf{y} - \mathbf{\Phi w})\} \\ &\sim \exp\{-\frac{1}{2}(\mathbf{y} - \mathbf{\Phi w})^T \mathbf{B}(\mathbf{y} - \mathbf{\Phi w})\}. \end{aligned} \tag{3.17}$$

We also treat the weight vector $\mathbf{w}$ as random and assume its prior distribution is Gaussian with zero mean and precision matrix $\mathbf{A}$:

$$p(\mathbf{w}|\mathbf{A}) = \mathcal{N}(0, \mathbf{A}^{-1}) \sim \exp\{-\frac{1}{2}\mathbf{w}^T \mathbf{A} \mathbf{w}\}, \tag{3.18}$$

where $\mathbf{A}$ is a diagonal matrix formed by M hyperparameters as follows:

$$\mathbf{A} = \text{diag}^{-1}(\alpha_1, \alpha_2, \cdots, \alpha_M). \tag{3.19}$$

This prior represents, at this time, the best guess about the value of the i^{th} weight w_i is 0, and α_i^{-1} represents the uncertainty about this guess. Moreover, we specify the hyperpriors over $\mathbf{A}$ and $\mathbf{B}$ to be uniform.

The basic idea of a Bayesian estimator is to maximize the posterior probability over the weights $\mathbf{w}$ and the hyperparameters $\mathbf{A}$ and $\mathbf{B}$ given the data $\mathbf{y}$, i.e. maximizing $p(\mathbf{w}, \mathbf{A}, \mathbf{B}|\mathbf{y})$. This posterior is further decomposed as:

$$p(\mathbf{w}, \mathbf{A}, \mathbf{B}|\mathbf{y}) = p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})p(\mathbf{A}, \mathbf{B}|\mathbf{y}). \tag{3.20}$$

Maximizing the left hand side of Eq.(3.20) is equivalent to maximizing the two probabilities on the right hand side, that is, maximizing $p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})$ and $p(\mathbf{A}, \mathbf{B}|\mathbf{y})$.

To maximize the first probability of the right hand side of Eq.(3.20), we rewrite $p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})$ according to Bayes rule,

$$\begin{aligned} p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B}) &= \frac{p(\mathbf{w}, \mathbf{y}|\mathbf{A}, \mathbf{B})}{p(\mathbf{y}|\mathbf{A}, \mathbf{B})} \\ &= \frac{p(\mathbf{y}|\mathbf{w}, \mathbf{B})p(\mathbf{w}|\mathbf{A})}{p(\mathbf{y}|\mathbf{A}, \mathbf{B})}. \end{aligned} \tag{3.21}$$

Rather than evaluating Eq. (3.21) directly (this needs to evaluate the term $p(\mathbf{y}|\mathbf{A}, \mathbf{B})$ which is not easy), we can rearrange Eq. (3.21) and obtain the probabilities $p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})$ and $p(\mathbf{y}|\mathbf{A}, \mathbf{B})$ simultaneously as shown below (Appendix A):

$$p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B}) = (2\pi)^{-\frac{M}{2}}|\mathbf{\Lambda}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{w} - \mathbf{u})^T\mathbf{\Lambda}^{-1}(\mathbf{w} - \mathbf{u})\}, \tag{3.22}$$

and

$$\begin{aligned} p(\mathbf{y}|\mathbf{A}, \mathbf{B}) &= (2\pi)^{-\frac{T}{2}}|\mathbf{B}^{-1} + \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}\mathbf{y}^T(\mathbf{B}^{-1} + \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{y}\}, \end{aligned} \tag{3.23}$$

where

$$\mathbf{\Lambda}^{-1} = (\mathbf{A} + \mathbf{\Phi}^T \mathbf{B} \mathbf{\Phi}) \tag{3.24}$$

and

$$\mathbf{u} = \mathbf{\Lambda} \mathbf{\Phi}^T \mathbf{B} \mathbf{y}. \tag{3.25}$$

The maximum of the first probability $p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})$ on the right hand side of Eq.(3.20) is clearly at the mean, i.e. the best estimate of the weights $\mathbf{w}$ is

$$\hat{\mathbf{w}} = \mathbf{u}. \tag{3.26}$$

The second probability on the right hand side of Eq.(3.20) is further decomposed and its maximization is equivalent to the maximization of $p(\mathbf{A}, \mathbf{B}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{A}, \mathbf{B}) p(\mathbf{A}) p(\mathbf{B})$ with respect to $\mathbf{A}$ and $\mathbf{B}$. For uniform hyperprior distributions, it is equivalent to maximizing $p(\mathbf{y}|\mathbf{A}, \mathbf{B})$ with respect to hyperparameters $\alpha_i; i = 1, \cdots, M$ in the matrix $\mathbf{A}$, the scaling parameters $s_i; i = 1, 2, \cdots, T$ and overall precision β in matrix $\mathbf{B}$.

It is convenient if we maximize the logarithm of this quantity $p(\mathbf{y}|\mathbf{A}, \mathbf{B})$ and accordingly the objective function becomes $\mathcal{L} = \log(p(\mathbf{y}|\mathbf{A}, \mathbf{B}))$. Without considering the constant term, the objective function is:

$$\mathcal{L} \;=\; -\frac{1}{2} \log |\mathbf{B}^{-1} + \mathbf{\Phi} \mathbf{A}^{-1} \mathbf{\Phi}^T| - \frac{1}{2} \mathbf{y}^T (\mathbf{B}^{-1} + \mathbf{\Phi} \mathbf{A}^{-1} \mathbf{\Phi}^T)^{-1} \mathbf{y}.$$

Maximizing the objective function $\mathcal{L}$ with respect to α_i, s_i and β, the following update equations are derived (The detailed derivation is given in Appendix B):

$$\alpha_i = \frac{1}{\mathbf{\Lambda}_{ii} + \hat{w}_i^2} \tag{3.27}$$

$$s_i = \frac{1}{\text{trace}(\mathbf{\Lambda} \beta \boldsymbol{\phi}_i^T \boldsymbol{\phi}_i) + \beta (\mathbf{y} - \mathbf{\Phi} \hat{\mathbf{w}})_i^2} \tag{3.28}$$

$$\beta = \frac{T}{\text{trace}(\mathbf{\Lambda} \mathbf{\Phi}^T \mathbf{S} \mathbf{\Phi}) + (\mathbf{y} - \mathbf{\Phi} \hat{\mathbf{w}})^T \mathbf{S} (\mathbf{y} - \mathbf{\Phi} \hat{\mathbf{w}})} \tag{3.29}$$

where $\mathbf{\Lambda}_{ii}$ is the i-th diagonal element of the posterior covariance matrix $\mathbf{\Lambda}$ in Eq.(3.24), $\hat{w}_i$ is the i-th element of the posterior mean vector $\hat{\mathbf{w}}$ in Eq.(3.26), $\boldsymbol{\phi}_i$ is the i-th row

vector of the design matrix $\mathbf{\Phi}$ and $(\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{w}})_i$ is the i-th element of the estimated error $\mathbf{r}_b = \mathbf{y} - \mathbf{\Phi}\hat{\mathbf{w}}$.

By iterative updating of Eq.(3.27) to Eq.(3.29), together with $\mathbf{\Lambda}$ and $\hat{\mathbf{w}}$ from Eq.(3.24) and Eq.(3.26), this update algorithm converges to the optimum solution. For the time-varying variance noise model, different voxels are assumed to have the same variance scaling parameters s_i and hence at the end of each updating cycle, s_i could be averaged over all the relevant voxels. For the fractional noise model, after wavelet transform, the variance scaling parameters of the transformed noise in the wavelet domain are testified to have the same value at the same decomposition level as shown in Eq.(3.13) and Eq.(3.14). During the updating process, the scaling parameters s_i at the same level are averaged to obtain an accurate estimation.

In practice, some of the α_i's in Eq.(3.19) will approach infinity, which means the w_i should be zero given the data at hand. Thus, the corresponding regressor functions could be 'pruned' and the remaining regressors are kept to construct a suitable design matrix. This is the flexible design matrix determination method in the GLM for fMRI data analysis as introduced in Chapter 2.

Having estimated the posterior probability density function of the weights $\mathbf{w}$ at each voxel, a map of the activation regions in the brain could be obtained by computing the posterior probability that a voxel is activated or the probability that an effect is greater than some threshold value. Given the effect size γ, the posterior probability is [56]:

$$1 - \Psi\Big(\frac{\gamma - \mathbf{c}^T\mathbf{u}}{\sqrt{\mathbf{c}^T\mathbf{\Lambda}\mathbf{c}}}\Big) \tag{3.30}$$

where $\mathbf{c}$ is the contrast vector as introduced earlier, $\Psi(\cdot)$ is the normal cumulative distribution function (cdf), $\mathbf{u}$ and $\mathbf{\Lambda}$ are defined in Eq.(3.25) and Eq.(3.24), respectively. These posterior probabilities will form posterior probability maps (PPMs) of the fMRI activation detection.

In the classical approach, the point estimate of the weights $\mathbf{w}$ by OLS and WLS methods is used to calculate a t at each voxel:

$$t = \frac{\mathbf{c}^T \hat{\mathbf{w}}}{\sqrt{\mathbf{c}^T \mathbf{\Lambda}_{\hat{\mathbf{w}}} \mathbf{c}}} \tag{3.31}$$

where $\hat{\mathbf{w}}$ is the least-square estimate of the parameter $\mathbf{w}$ and $\mathbf{\Lambda}_{\hat{\mathbf{w}}}$ is the covariance matrix of the estimate $\hat{\mathbf{w}}$. The statistic t is of Student's t distribution with the effective degree of freedom (df) as derived in [50].

These t values obtained at each voxel are then mapped to form statistical parametric maps (SPMs) for further determination of the activated regions in the brain.

3.4 Results and Discussion

In this section, simulation studies are carried out to compare the performance of the proposed approach to the OLS and the WLS estimators introduced in the previous sections. Both simulated as well as experimental data are examined. We first compare the accuracy of the estimated weights in GLM. The activation detection ability of these methods are then investigated. Lastly, results from the experimental fMRI data are given and discussed.

3.4.1 Simulated Data

Time-varying variance noise

We first show that the proposed Bayesian approach could give more efficient (lower variance) estimate of the weights $\mathbf{w}$ than the OLS and the WLS estimators. A simulated block experimental fMRI design is investigated in this study. The BOLD signal is represented by a square waveform as shown in Fig. 2.1.

Table 3.1 Standard deviation (SD) of estimated $\hat{w}$ on simulated data with different weight and noise.

		OLS	**WLS**	**Bayesian**
i.i.d. noise	$w = 0$	0.3564	0.3579	0.1673
	$w = 0.5$	0.3558	0.3577	0.3318
	$w = 1$	0.3527	0.3545	0.3472
	$w = 2$	0.3527	0.3546	0.3065
time-varying variance noise	$w = 0$	0.4086	0.1153	0.0321
	$w = 0.5$	0.4100	0.1162	0.0583
	$w = 1$	0.4101	0.1170	0.0578
	$w = 2$	0.4084	0.1159	0.0570

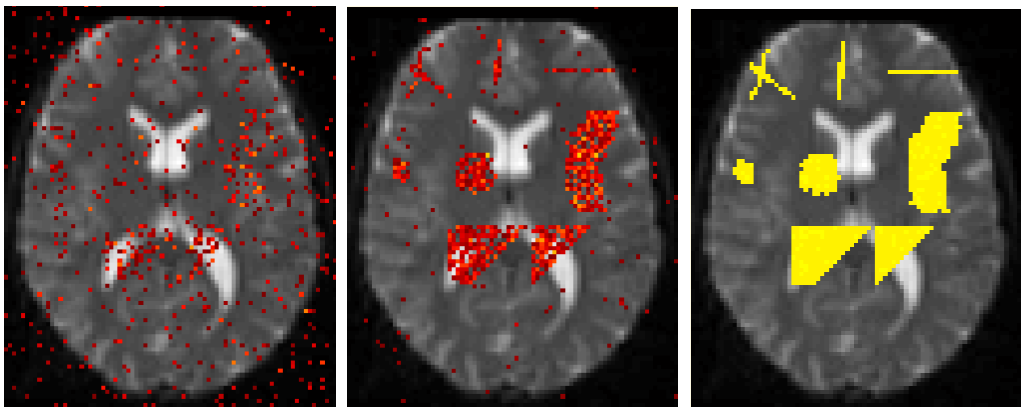
Two hundred realizations of 128-point time series (i.e. 128 (T) images with 200 (N) voxels in fMRI data sets) under different weighting of the square waveform are generated. Both the i.i.d. noise and time-varying variance noise are then added to the data set to simulate the noisy fMRI data. These data are fitted with a GLM model with a design matrix $\mathbf{\Phi}$ (dimension 128×2) composed of two regressors : the constant value 1 to model the mean grey level in fMRI voxels and a square waveform to model the experimental design. Since we are only concerned about the weight parameter related to the square waveform regressor, only the standard deviation (SD) of this weight w is reported. Table 3.1 shows the SD of the estimate $\hat{w}$ under different values of w and different noise properties.

To compare the performance of different detection methods, the standard deviation of these estimates are calculated. The method that results in lower SD estimate is more

accurate since a stringent confidence interval of the estimated parameter could be obtained by using lower SD. From Table 3.1, it is clear that the proposed Bayesian method always results in lower SD estimate of the weight w for all the values of w tested no matter whether the noise is i.i.d. or time-varying variance. This implies that the Bayesian estimator is more accurate than the OLS and WLS estimators. This improvement of our approach comes from the fact that the Bayesian estimator not only captures the true variance structure of the noise better than OLS and WLS, but also utilizes a flexible design matrix which is more suitable for the actual nonstationary data. Accurate regression weights are desired since these regression weights are fitted with a hierachical linear model for second-level analysis or random effect analysis [21]. Lower SD estimates will give more sensitive results for these higher level analysis.

Then, the detection ability of these estimators is compared. A slice from a real fMRI data set is used as the background image. First, the i.i.d noises are added to the simulated fMRI data set. The detection results of these three estimators are comparable as shown in Table 3.1 . Next, time-varying noises with same scaling parameters but different overall variances are then added to each voxel of the data to form a 3-D fMRI time series. At some specific regions of the image, simulated BOLD signals are added. The BOLD signal is simulated by the convolution of the experimental paradigm and conventional HRF, which is the difference between two gamma functions. These simulated fMRI data are fitted by the GLM with the design matrix composed of two regressors: the simulated BOLD response and a constant vector of value 1. For the Bayesian method, the PPM is obtained as introduced in Section 3.3. For the OLS and WLS methods, SPMs can be obtained. Figure 3.1 shows the detection results of simulated fMRI data using these methods.

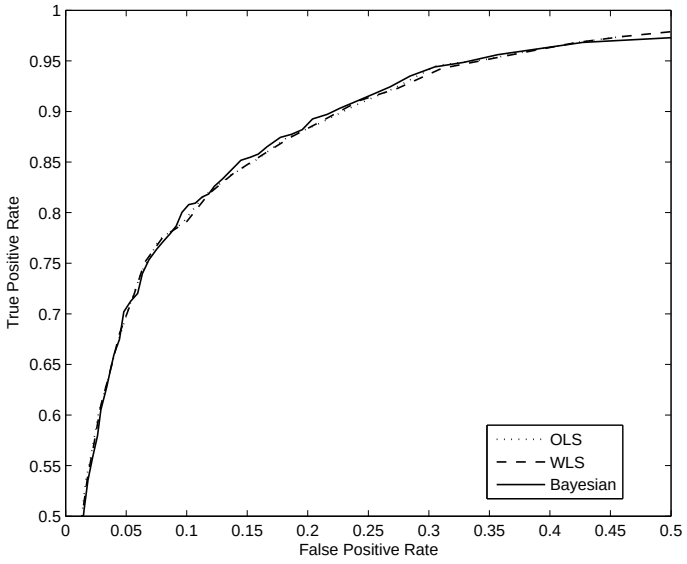
From Fig. 3.1, it is clear that the proposed Bayesian method is more robust and sensitive compared to the OLS and WLS methods with more true activations and less



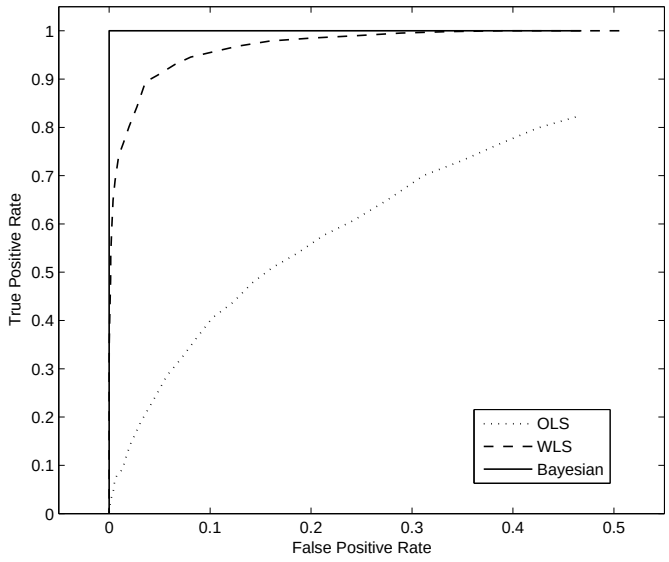
(a) (b) (c)

Figure 3.1 Detection results of simulated fMRI data using different methods: (a) OLS method with thresholded statistical parametric map (SPM) ($t > 1.7, p < 0.05$); (b) WLS method with thresholded SPM ($t > 1.7, p < 0.05$); (c) Bayesian method with posterior probability map (PPM) ($P(\text{effect} > 0.4) > 0.9$).

false activations detected (the simulated pattern is shown in Fig. 5.8 (a) as a comparison reference). This is because the noise structure estimated by the proposed method is closer to the true noise. To have a clearer comparison of the detection ability, the receiver operator characteristic (ROC) analysis [58] is used to investigate the activation maps of Bayesian, OLS and WLS methods. The ROC curve is a plot of true positive ratio (TPR) versus false positive ratio (FPR) under different threshold values. The method that can detect most of the real activations while minimizing the detection of false activations is more desirable. Figure 3.2 (a) and Fig. 3.2 (b) show the ROC curves under the i.i.d. noise and time-varying variance noise, respectively. For i.i.d. noise, the three methods have comparable performances (Fig. 3.2 (a)). While for the time-varying variance noise, the ROC curves (Fig. 3.2 (b)) indicate that under the same FPR, the proposed Bayesian method could actually detect more real activations. This clearly shows the superior performance of the proposed method compared to the OLS and WLS methods.



(a)



(b)

Figure 3.2 ROC curves for simulated noisy data: (a) for i.i.d. noise; (b) for time-varying variance noise.

Fractional Noise

To investigate the effect of fractional noise or $1/f-$like noise, we first synthesize realizations of fBm noise based on the statistical model of fBm using the method introduced in [62]. The total length of the simulated fBm noise is $T = 128$ with the decomposition level $J = 7$ (since $T = 2^J$). These noises are added to the simulated BOLD signals to form the simulated fMRI data. The design matrix is constructed as the same as the one introduced previously. The OLS method is first applied to the original simulated data in the time domain (denoted as "OLS-time"). After transforming the original simulated data and design matrix in the GLM into the wavelet domain as shown in Eq.(3.9), the Bayesian method and OLS method (denoted as "OLS-wavelet") are applied to the transformed data in the wavelet domain. Daubechies wavelets with vanishing moments 4 ('db4') are used for the DWT since they are the most compactly supported wavelets with sufficient vanishing moments to whiten the fBm noise. To compare the detection results of different methods, the generalized least squares (GLS) estimator is also investigated. In the GLS estimator, the noise covariance matrix is known according to the specified value of H and σ in Eq.(3.7) and Eq.(3.8). This GLS method is used as a comparison reference for the efficiency of the Bayesian method in the wavelet domain. Table 3.2 shows the SD of estimate $\hat{w}$ under different weight value w and different H using GLS, OLS-time, OLS-wavelet and Bayesian methods in the wavelet domain.

From Table 3.2, it is seen that the Bayesian method in wavelet domain is a robust and efficient estimator when the noise considered is fBm noise. Both the OLS method in time domain (OLS-time) and in wavelet domain (OLS-wavelet) perform worse when the Hurst exponent H is high since the noise assumption does not match the simulated data. The SD of the Bayesian estimator is close to that of the GLS method, which shows that the Bayesian estimator is an accurate estimate of the weights $\mathbf{w}$ and could lead to

Table 3.2 Standard deviation (SD) of estimated $\hat{w}$ on simulated data with different weight and Hurst exponent.

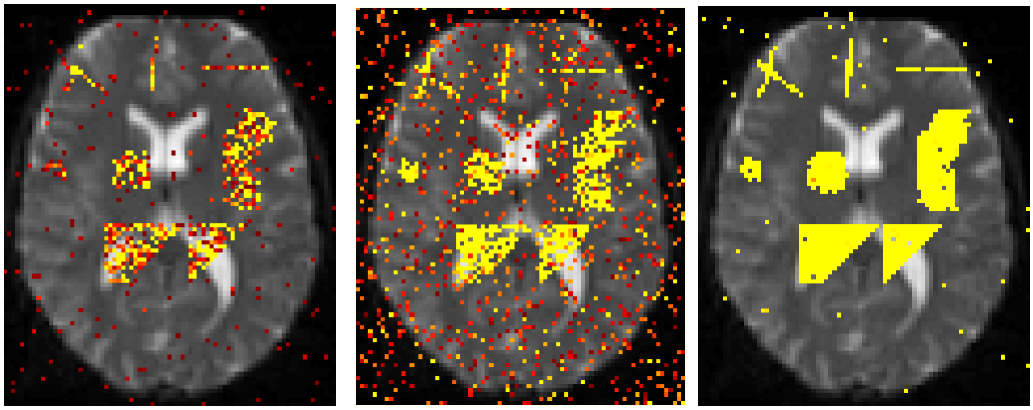
		GLS	**OLS-time**	**OLS-wavelet**	**Bayes**
	$H = 0.1$	0.0503	0.0518	0.0537	0.0384
	$H = 0.3$	0.0664	0.0881	0.0937	0.0485
$w = 0$	$H = 0.5$	0.0667	0.1305	0.1397	0.0484
	$H = 0.7$	0.0608	0.2389	0.2433	0.0438
	$H = 0.9$	0.0408	0.4499	0.3304	0.0332
	$H = 0.1$	0.0501	0.0523	0.0569	0.0529
	$H = 0.3$	0.0678	0.0868	0.1002	0.0725
$w = 0.5$	$H = 0.5$	0.0711	0.1326	0.1389	0.0797
	$H = 0.7$	0.0642	0.2262	0.1980	0.0700
	$H = 0.9$	0.0414	0.4601	0.3307	0.0491
	$H = 0.1$	0.0531	0.0556	0.0628	0.0551
	$H = 0.3$	0.0685	0.0874	0.0924	0.0701
$w = 1$	$H = 0.5$	0.0674	0.1277	0.1550	0.0730
	$H = 0.7$	0.0637	0.2179	0.2155	0.0686
	$H = 0.9$	0.0390	0.4542	0.3658	0.0442

almost the same performance as the GLS method. In the case of $w = 0$, the SD of the Bayesian method is even lower than that of the GLS method due to the pruning property of the proposed Bayesian method. These results show that the wavelet transform actually whitens the noise and the diagonal error covariance matrix is accurately estimated by the Bayesian method.

The simulated 3-D fMRI time series are also investigated. The simulated data are generated as explained before. The difference is that the noises added here are synthesized fBm noises with different H at different voxels. Figure 3.3 shows the activation detection results of these methods. It is clear from Fig. 3.3 that the proposed Bayesian estimator in the wavelet domain performs better than the OLS method (in time domain and wavelet domain) with more true activations and fewer false activations detected (the simulated pattern is shown in Fig. 5.8 (a) as a comparison reference). The ROC curves of the OLS method (OLS-time and OLS-wavelet) and the Bayesian method are also given in Fig. 3.4 with a clear illustration of the better detection ability of the Bayesian method.

3.4.2 Experimental fMRI Data

The experimental data used in this study are the *DATA-BLOCK* introduced in Chapter 1. Figure 3.5 shows the results of applying the OLS, WLS and Bayesian methods assuming the time-varying variance noise model and the Bayesian method assuming fractional noise model. For the fractional noise model, the time series are up-sampled in order to make the length of the time series to be a power of 2 and at the same time keep the scaling properties of the fractional noise. From these figures, we can see that the activation of the visual cortical areas are detected. Figure 3.5 (a) and (b) are respectively



(a) (b) (c)

Figure 3.3 Detection results of simulated data using fBm noise model: (a) OLS in time domain with thresholded SPM ($t > 3.4, p < 0.001$); (b) OLS in wavelet domain with thresholded SPM ($t > 3.4, p < 0.001$); (c) Bayesian method in wavelet domain with PPM ($P(\text{effect} > 1) > 0.99$).

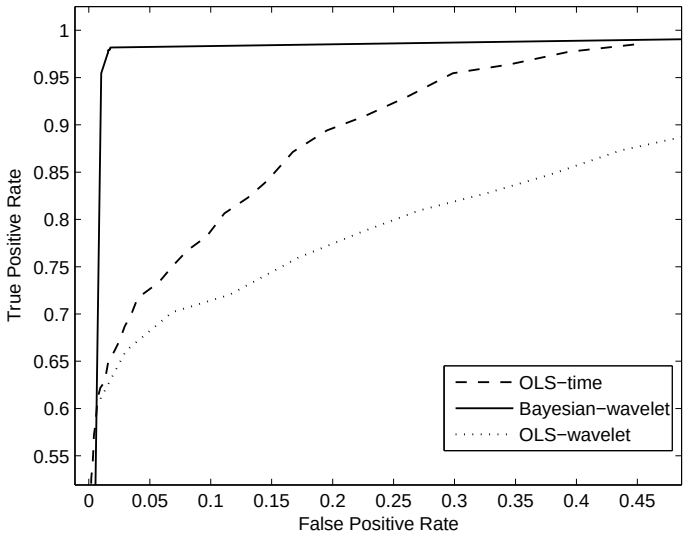


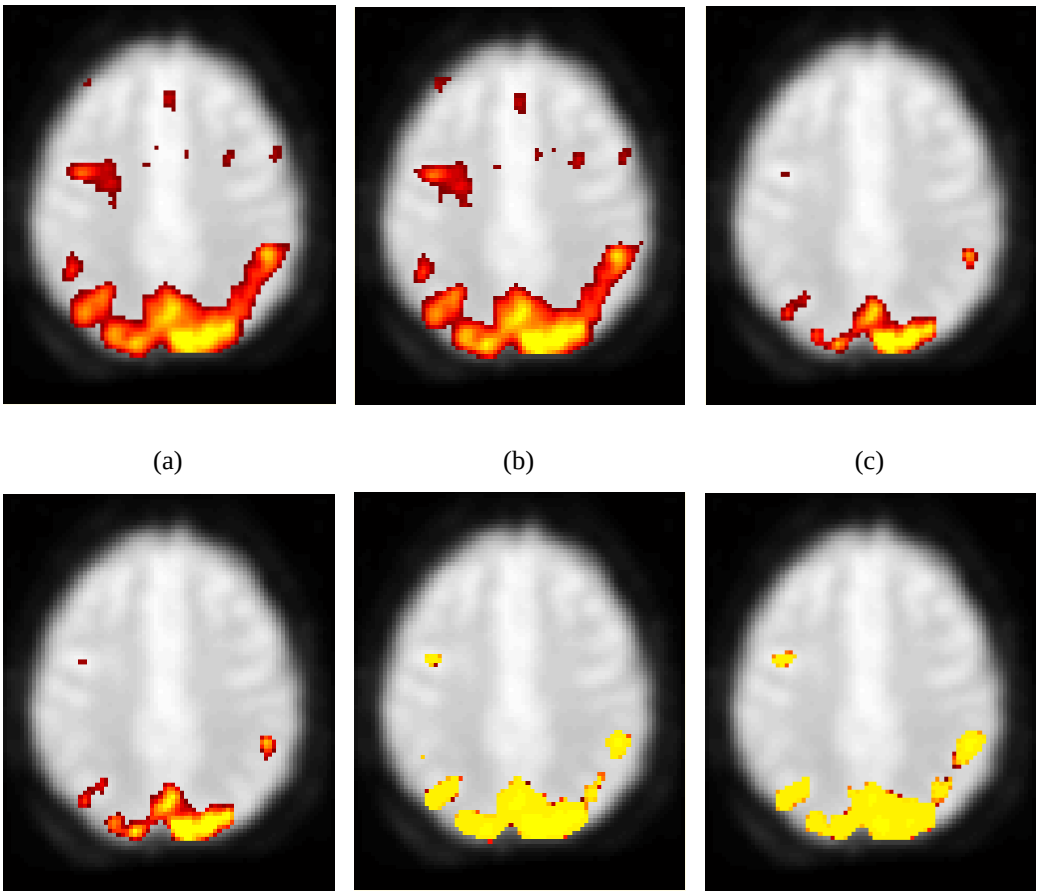
Figure 3.4 ROC curves of OLS (in both time domain and wavelet domain) and Bayesian (after DWT) methods for simulated fMRI data corrupted with fBm noises.

the detection results of the OLS and WLS method with the significance level p uncorrected. It is seen that at this p level, the detection results may include some erratically scattered noise. Considering the multiple comparisons problem, a stringent correction using the Bonferroni correction [21] is used here as shown in Fig. 3.5 (c) and (d) which illustrate that some potentially activated voxels may not be detected. Compared to the OLS and WLS methods, the Bayesian methods perform better (Fig. 3.5 (e) and (f)). Besides, the Bayesian methods provide the probabilities of the activation effect, avoiding the problems encountered in SPM to correct or adjust the p levels due to the multiple comparisons problem. These results show that the Bayesian estimators under these two noise assumptions can provide alternative estimators and better detection results.

3.5 Conclusion

In fMRI data analysis, the noise is generally assumed to be stationary. However, this is an inappropriate assumption given the complex nature of the fMRI data. The fMRI data is sensitive to the subjects' movements, resulting in the time-varying variance noise. In addition, it is reported that noise in the fMRI data can be of fractional and/or $1/f$-like. In this chapter, a Bayesian method is proposed to detect the activated voxels under two nonstationary noise structures (time-varying variance noise and fractional noise). For time-varying variance noise, the variance of the noise at each time point is a scaled version of the overall variance in that voxel. For fractional noise, the wavelet transform is first applied as a whitening filter. The coefficients of the transformed noise are i.i.d. and the variance at each decomposition level is a scaled version of the overall variance. Thus, the covariance matrices of these two assumed noise models have similar structures. This similarity led to similar estimators or updating algorithms.

The weight estimator is investigated under the Bayesian framework. This estimator



(a) (b) (c)

(d) (e) (f)

Figure 3.5 Results of a block visuospatial processing task fMRI data: (a) thresholded SPM of OLS method ($t > 3.4$, $p < 0.001$, uncorrected); (b) thresholded SPM of WLS method ($t > 3.4$, $p < 0.001$, uncorrected); (c) thresholded SPM of OLS method with Bonferroni correction ($t > 7$, $p < 0.05$, corrected); (d) thresholded SPM of WLS method with Bonferroni correction ($t > 7$, $p < 0.05$, corrected); (e) PPM of Bayesian method using time-varying variance noise model ($P(\text{effect} > 0.8) > 0.99$); (f) PPM of Bayesian method using fractional noise model ($P(\text{effect} > 0.8) > 0.99$).

could provide a probability that a voxel is activated or the probability that the activation of this voxel is greater than some threshold. This advantage enhances the ability to detect activated regions of the brain and avoids the limitations of the classical methods. The proposed methods are compared to the ordinary least squares (OLS) and the weighted least squares (WLS) methods both on simulated as well as experimental fMRI data. The ROC analysis validates that the proposed Bayesian methods are more accurate than the OLS and WLS methods. These results show that the proposed Bayesian methods under the time-varying variance noise model and the fractional noise model are efficient and robust methods for brain activity detection in fMRI data analysis.

The method proposed in this chapter does not consider the drift in fMRI data. In the following chapter, this method is extended to incorporate the drift into the GLM.

Chapter 4

Analysis of fMRI Data with Drift: Modified General Linear Model and Bayesian Estimator

4.1 Introduction

As introduced in Section 1.2.2, the fMRI signal consists basically of three components: the BOLD response related to the stimuli, drift and noise. Let the fMRI signal be y_i at a specific voxel measured at scan number i (with total T scans). Then, GLM of fMRI data can be written as:

$$y_i = wb_i + f_i + \epsilon_i; \qquad i = 1, \cdots, T \tag{4.1}$$

where b_i is the explanatory BOLD response. The parameter w in Eq. (4.1) is a scalar which represents the contribution of the BOLD signal to y_i, f_i is the drift and ϵ_i is the noise.

Written in a matrix form, Eq. (4.1) becomes:

$$\mathbf{y} = w\mathbf{b} + \mathbf{f} + \boldsymbol{\epsilon}, \tag{4.2}$$

or in a traditional way (GLM in Section 1.2.3):

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{w} + \boldsymbol{\epsilon}, \tag{4.3}$$

where $\mathbf{y}$, $\mathbf{b}$, $\mathbf{f}$ and $\boldsymbol{\epsilon}$ are of dimensions $T \times 1$. $\boldsymbol{\Phi} = [\mathbf{b}, \mathbf{f}]$ is the design matrix with the dimensions $T \times 2$ and $\mathbf{w} = [w \;\; 1]^T$ is the parameter vector. The detection of the activated voxels/regions of the human brain is based on the inference drawn from the estimation of the parameter w.

To remove the drift before proceeding to the statistical analysis, preprocessing procedures or drift models have been proposed. Some of the well-known preprocessing techniques are the high-pass filtering [21] and the median filtering method [77]. The drift can also be removed by introducing drift models in the GLM. These drift models assume that the drift resides in a linear subspace spanned by a set of polynomials or a set of cosine functions [21], then the drift can be projected into this subspace and removed. In [78], a wavelet-based semiparametric GLM was proposed to estimate the parameters in the GLM as well as to estimate the drift. However, the accuracy of the estimated parameters in this method is related to the accuracy of the estimation of the noise covariance matrix. The inaccuracy of the estimation of the covariance matrix may bias the results and lead to wrong judgements about the activation of the voxels. In this chapter, based on the observation that the noise in fMRI is long memory fractional noise and the slowly varying drift resides in a subspace spanned by large scale wavelets, we examine a modified GLM in wavelet domain under Bayesian framework. By estimating the noise covariance efficiently, the Bayesian method can give accurate estimation of the parameters in GLM. Besides, a model selection scheme based on the results from the

modified GLM is proposed to accurately estimate the parameter w in GLM as well as to remove the drift efficiently.

4.2 Models

Using the discrete wavelet transform (DWT), the parameters of activation w and the drift can be estimated efficiently. The wavelet transform is used because of the following reasons.

1. Drift can be efficiently removed from the fMRI signal since it resides only in the subspace spanned by large scale wavelets [78].

2. The noise in the fMRI data exhibits long-range auto-correlation in time and $1/f-$like spectral properties [70]. The DWT applied to this kind of $1/f-$like processes works as a Karhunen-Loéve transform (KLT) to decorrelate the fractional noise process [79][63]. The resulting wavelet coefficients are independently Gaussian distributed, which simplifies the estimation of the parameters and makes statistical analysis easy.

Assume the length of the fMRI signal is $T = 2^J$, where J is an integer. Applying the DWT to both sides of Eq.(4.2), the GLM in the wavelet domain is of the form:

$$\mathbf{y}^W = w\mathbf{b}^W + \mathbf{f}^W + \mathbf{n}^W \tag{4.4}$$

where $\mathbf{y}^W$, $\mathbf{b}^W$, $\mathbf{f}^W$ and $\boldsymbol{\epsilon}^W$ are the results of applying the DWT respectively to the data $\mathbf{y}$, BOLD response $\mathbf{b}$, drift $\mathbf{f}$ and noise $\boldsymbol{\epsilon}$ up to the maximum scale J.

4.2.1 Noise Model

The wavelet transform of the noise $\boldsymbol{\epsilon}^W$ is:

$$\boldsymbol{\epsilon}^W = [a\epsilon_{J,1}, d\epsilon_{J,1}, d\epsilon_{J-1,1}, d\epsilon_{J-1,2}, \cdots, d\epsilon_{1,1}, \cdots, d\epsilon_{1,2^{J-1}}]^T \tag{4.5}$$

where $a\epsilon_{J,1}$ is the scaling coefficient at level J, and $d\epsilon_{m,k}$ ($k = 1, \cdots, 2^{J-m}$) are the wavelet coefficients at level m ($m = 1, \cdots, J$). The noise in the fMRI time series obtained under the resting or null conditions exhibit long-range auto-correlation in time and $1/f-$like spectral properties [70]. The properties of the noise is detailed in Section 3.2.2. After the wavelet transform, the transformed fractional noise $\boldsymbol{\epsilon}^W$ could be modelled by Gaussian process $\boldsymbol{\epsilon}^W \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ with the $T \times T$ diagonal covariance matrix defined in Eq. (3.13) and rewritten here:

$$\begin{aligned} \mathbf{V} &= Cov(\boldsymbol{\epsilon}^W) \\ &= \begin{pmatrix} V_{a_J} & & & & & & \\ & V_{d_J} & & & & & \\ & & V_{d_{J-1}} & & & & \\ & & & V_{d_{J-1}} & & & \\ & & & & \ddots & & \\ & & & & & \begin{pmatrix} V_{d_1} & & \\ & \ddots & \\ & & V_{d_1} \end{pmatrix} \end{pmatrix}. \end{aligned} \tag{4.6}$$

4.2.2 Drift Model

The drift is assumed to vary slowly with large scales, and hence the wavelet coefficients at fine scales are almost zero [78] as shown here:

$$\mathbf{f}^W = [af_{J,1}, df_{J,1}, \cdots, df_{J_0,1}, \cdots, df_{J_0,2^{J-J_0}}, 0, \cdots, 0]^T \tag{4.7}$$

where $af_{J,1}$ is the scaling coefficient (or approximation coefficient) at level J, and $df_{m,k}$ ($k = 1, \cdots, 2^{J-m}$) are the wavelet coefficients (or detail coefficients) at level m ($m = J_0, \cdots, J$). The wavelet coefficients at levels lower than J_0 (i.e. fine scales) are zero since at these levels, the drift does not vary greatly over a short interval of time.

4.3 Modified GLM

As seen from the properties of the drift, at finer scales, the effect of drift can be ignored and hence the GLM at finer scales becomes:

$$\mathbf{y}_i^W = w\mathbf{b}_i^W + \boldsymbol{\epsilon}_i^W \tag{4.8}$$

where $\mathbf{y}_i^W$, $\mathbf{b}_i^W$ and $\boldsymbol{\epsilon}_i^W$ are respectively wavelet decomposition coefficients at the finer scales (scales less than J_0).

However, since we do not know exactly at which level the effect of drift can be ignored, an approach is to estimate the parameter w at different scales and use a scheme to find a suitable level for modelling the drift as well as accurately estimating the parameter. Therefore, a modified GLM is obtained (for convenience, the superscript 'W' is dropped):

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{w} + \boldsymbol{\epsilon}, \tag{4.9}$$

where the matrix $\boldsymbol{\Phi}$ is defined as:

$$\boldsymbol{\Phi} = \begin{pmatrix} ab_{J,1} & & & & & & & \\ & db_{J,1} & & & & & & \\ & & \ddots & & & & & \\ & & & db_{J_0,1} & & & & \\ & & & \vdots & & & & \\ & & & db_{J_0,2^{J-J_0}} & & & & \\ & & & & \ddots & & & \\ & & & & & db_{1,1} & & \\ & & & & & \vdots & & \\ & & & & & db_{1,2^{J-1}} & & \end{pmatrix}, \tag{4.10}$$

which is a block diagonal matrix of dimensions $T \times (J+1)$. The block in the diagonal of matrix $\boldsymbol{\Phi}$ is a vector composed of the wavelet coefficients at the same scale.

The parameter $\mathbf{w}$ is defined by:

$$\mathbf{w} = [w_{J+1}, w_J, \ldots, w_{J_0}, \ldots, w_1]^T, \tag{4.11}$$

which consists of estimates of w at different levels. The parameter $\mathbf{w}$ can be estimated using Bayesian methods introduced in Section 3.3. Among these estimates, the estimates of w at levels lower than J_0 can be thought of as good estimates because the effect of drift can be ignored at these levels; however, other estimates of w at level J_0 or levels higher than J_0 are bad due to the effect of drift. In the next section, the schemes for model selection are discussed.

If the effect of drift can be ignored at the levels lower than J_0, the estimates $\hat{w}_i$ $(i = 1, \cdots, J_0 - 1)$ could be treated as $(J_0 - 1)$ observations of the unknown parameter w. Thus, we could obtain the following observation equation:

$$\begin{pmatrix} \hat{w}_{J_0-1} \\ \hat{w}_{J_0-2} \\ \vdots \\ \hat{w}_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} w + \mathbf{n} \tag{4.12}$$

where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_{J_0})$ and $\mathbf{\Lambda}_{J_0}$ is defined as:

$$\mathbf{\Lambda}_{J_0} = \begin{pmatrix} \Lambda_{J_0-1} & & & \\ & \Lambda_{J_0-2} & & \\ & & \ddots & \\ & & & \Lambda_1 \end{pmatrix}, \tag{4.13}$$

where Λ_i $(i = 1, \cdots, J_0 - 1)$ are the variances of the estimated parameters $\hat{w}_i$ $(i = 1, \cdots, J_0 - 1)$ respectively, which are the lower $J_0 - 1$ diagonal elements of $\mathbf{\Lambda}$ in Eq. (3.24) obtained through Bayesian estimator.

The maximum likelihood estimate of w is then given by (the detailed derivation is given in Appendix C):

$$\hat{w}_{ML} = \sum_{i=1}^{J_0-1} \frac{\Lambda_i^{-1}}{\sum_{k=1}^{J_0-1} \Lambda_k^{-1}} \hat{w}_i, \tag{4.14}$$

which is also Gaussian distributed with the variance:

$$\sigma^2_{\hat{w}_{ML}} = \sum_{i=1}^{J_0-1} \frac{\Lambda_i^{-1}}{(\sum_{k=1}^{J_0-1} \Lambda_k^{-1})^2}. \tag{4.15}$$

Thus, the significance of $\hat{w}_{ML}$ can be tested by calculating the t scores with degrees of freedom $T - 2^{-J_0}T$, which is determined by the choice of level J_0. In the next section,

schemes for choosing the optimum J_0 are described. The estimation of drift is then calculated by setting the coefficients of $\mathbf{y}^W - \hat{w}_{ML}\mathbf{b}^W$ at finer scales (scales lower than J_0) to be zeros and performing the inverse wavelet transform.

4.4 Model Selection

Model selection aims to find a suitable value J_0, where the effect of drift at levels lower than J_0 can be ignored. Some model selection criteria have been proposed in fMRI analysis using semiparametric GLM [78]. The first criterion is an improved version of the Akaike information criterion (AIC) [78][80], which is represented by:

$$\text{AIC}_c = \log \hat{\sigma}^2 + \frac{T + n_0}{T - n_0 - 2} \tag{4.16}$$

where n_0 is the number of wavelet coefficients used to estimate the drift and is determined by the selection of drift level J_0 as $n_0 = 2^{-J_0}T$, $\hat{\sigma}^2$ is the measure of the goodness of fit and is given by the mean squared error:

$$\hat{\sigma}^2 = \frac{1}{T}\|\mathbf{y} - \hat{w}_{ML}\mathbf{b} - \hat{\mathbf{f}}\| \tag{4.17}$$

where $\hat{\mathbf{f}}$ is the estimated drift.

The second criterion is the Schwartz information criterion (SIC) [78][81], given by:

$$\text{SIC} = \log \hat{\sigma}^2 + \frac{n_0 \log n_0}{T}. \tag{4.18}$$

For every scale, AIC_c and SIC are calculated and the optimum scale J_0 is set to the one which results in the minimum value of these two criteria.

Both the AIC_c and SIC criteria need to calculate the estimate of w for each assumed model, which is not efficient. In this thesis, we propose a more efficient criterion, namely, the Confidence Interval Criterion (CIC) to estimate the order of the drift (J_0) as well as the parameter w as shown below:

1. Start from the finest scale (i.e. $i = 1$) and set lower and upper limits $[l, u] = [-\infty, \infty]$;

2. Calculate the $100(1-\alpha)\%$ confidence interval of w_i given by:

$$[l_i, u_i] = [\hat{w}_i - z_{\alpha/2}\sqrt{\Lambda_i}, \hat{w}_i + z_{\alpha/2}\sqrt{\Lambda_i}]$$

   where $z_{\alpha/2}$ is the z score with the probability greater than z being $\alpha/2$;

3. Derive the overlap interval of $[l, u]$ and $[l_i, u_i]$ and update $[l, u]$;

4. Judge whether $\hat{w}_{i+1} \in [l, u]$. If yes, increase i by 1 and go to step 2. If not, $J_0 = i + 1$.

After finding the optimum scale of the drift J_0, the parameter w is then estimated according to Eq. (4.14).

4.5 Results and Discussion

4.5.1 Simulated Data

We first test the efficiency of the proposed approach on simulated data. A simulated block experimental fMRI design is investigated in this study. The BOLD signal is represented by a square waveform alternating with 6 scans per block. One thousand realizations of fBm noise are synthesized based on the statistical model of fBm using the method introduced in [62]. The total length of the simulated fBm noise is $T = 128$ with the decomposition level $J = 7$ (since $T = 2^J$). These noises are added to the simulated BOLD signals with different weights to form the simulated fMRI signal. Then, slowly varying drifts which are simulated using the summation of sinusoidal waves are added

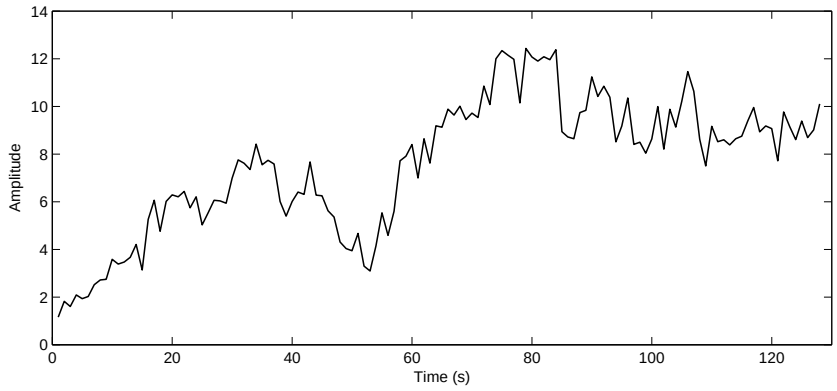


Figure 4.1 Simulated fMRI signal.

to the simulated fMRI data. Figure 4.1 shows a realization of the synthetic fMRI signal with the fBm noise generated using the parameter $H = 0.3$.

The simulated fMRI signals are then decomposed using Daubechies wavelets with vanishing moments 4 (‘db4’) since they are the most compactly supported wavelets with sufficient vanishing moments to whiten this fBm noise. The decomposed signals in the wavelet domain are then processed using the proposed Bayesian method introduced in Section 3.3. The ability of the Bayesian method to estimate the noise covariance matrix $\mathbf{V}$ in wavelet domain is first tested. For a comparison, the wavelet decomposition is also applied to the simulated fBm noise and the variance at different scales is calculated as the true value of the variance. Table 4.1 compares the estimated variance at different scales with the true values. It is clearly seen from the table that the Bayesian estimator accurately estimates the variances of wavelet coefficients of noise at different scales. These accurate estimates can be used to estimate the Hurst exponent H of fBm noise according to the method introduced in [74]. For this example, the estimated H is equal to 0.2666 which is very close to the true value of 0.3.

Then the three model selection criteria are used to find the optimum scale of drift

Table 4.1 Comparison of estimated variance of wavelet coefficients of noise at different scale with the true value.

Scale i	**Estimated Value**	**True Value**
1	0.8193	0.8471
2	1.4885	1.5783
3	4.1338	4.1340
4	20.1372	21.5596

Table 4.2 Model selection of CIC, AIC$_c$ and SIC criteria.

Scale i	$\hat{w}_i$	**Updated** $[l, u]$	**AIC**$_c$	**SIC**
1	0.6465	$[-\infty, \infty]$	1.9382	0.8737
2	0.6281	[0.0876, 1.2055]	0.7622	-0.0768
3	0.4617	[0.0876, 1.2055]	0.7312	-0.2393
4	0.6682	[0.0876, 1.0984]	0.7844	-0.2318
5	5.6985	[0.0876, 1.0984]	0.8434	-0.1966

and hence an accurate estimate of the parameter w. For the CIC criterion, the α value is set to be 0.01 and hence $z_{\alpha/2} = 2.576$. Table 4.2 compares the results of these three criteria. From the table, we can see that the CIC criterion chooses the scale J_0 to be 4 since $\hat{w}_5 \notin [l, u]$ at scale 5. Both SIC and AIC$_c$ criteria reach the minimum value at the scale of 3 and hence they give the optimum value of J_0 to be 3. Figure 4.2 shows the estimated drift when the optimum scale of drift is set to be $J_0 = 4$. It is seen from the figure that the estimated drift captures the properties of slowly varying drift without tracking the variation of BOLD signal.

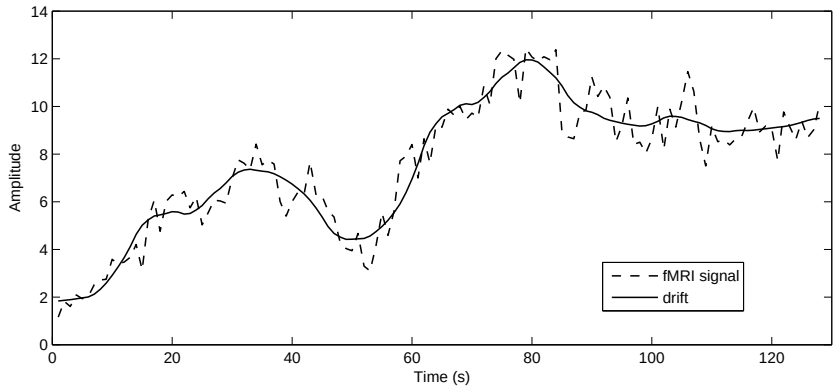


Figure 4.2 Simulated fMRI signal and the estimated drift.

Table 4.3 MSE comparison of three model selection criteria with drift added.

Noise	**CIC**	**AIC**$_c$	**SIC**
White	4.0127e-005	9.7032e-005	9.7032e-005
$H = 0.1$	3.5144e-004	3.5144e-004	3.5144e-004
$H = 0.3$	1.8624e-005	1.8820e-005	1.8820e-005
$H = 0.7$	7.2508e-007	6.3683e-006	6.3683e-006
$H = 0.9$	3.5122e-005	3.5122e-005	3.5122e-005

In order to see the effect of the choice of scale J_0 on the accuracy of estimated parameter $\hat{w}_{ML}$, we also compute the deviation of the estimated parameter $\hat{w}_{ML}$ from the true values. Table 4.3 compares the mean square errors (MSE) of the estimates obtained from these three criteria when drift is added to the data. For completeness, the situation without drift is also considered and the comparison results are presented in Table 4.4.

From Table 4.3 and 4.4, it is clearly seen that the proposed CIC criterion performs as

Table 4.4 MSE comparison of three model selection criteria without drift.

Noise	**CIC**	**AIC$_c$**	**SIC**
White	6.5180e-005	2.9001e-004	2.9001e-004
$H = 0.1$	0.0015	0.0015	0.0015
$H = 0.3$	1.1690e-004	1.1690e-004	1.1690e-004
$H = 0.7$	3.9381e-004	3.9381e-004	3.9381e-004
$H = 0.9$	1.9526e-004	2.8953e-004	2.8953e-004

well as the AIC$_c$ and SIC criterion. In some cases, it performs even better than the other two. The difference between these criteria results from the selection of the drift model. Regardless of the small differences among these three model selection criterion, the MSE is small for all of them. This shows that the Bayesian method in the wavelet domain is a robust and efficient estimator when the noise considered is fBm noise. Accurate estimations are desired since these weights are fitted with a hierachical linear model for second-level analysis or random effect analysis [67][56][68]. The estimate with lower MSE will give more sensitive results for these higher level analysis.

4.5.2 Experimental fMRI Data

The experimental data used in this set of studies are the *DATA-BLOCK* introduced in Chapter 1. The first ten volumes (i.e. the first block) of the data are discarded due to T_1 effects. The time series are up-sampled in order to make the length (T) of the time series to be a power of 2 and at the same time keep the scaling properties of the fractional noise. These data are not detrended in the preprocessing step before applying the proposed method. Figure 4.3 shows the results of activation detection. It is clearly seen that the

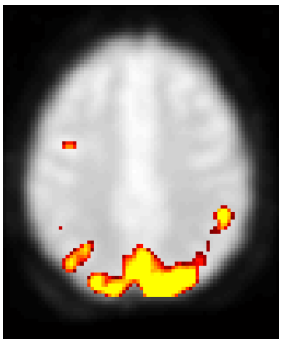


Figure 4.3 Results of the proposed method to a visuospatial processing task ($t > 5.3$ $P < 0.05$).

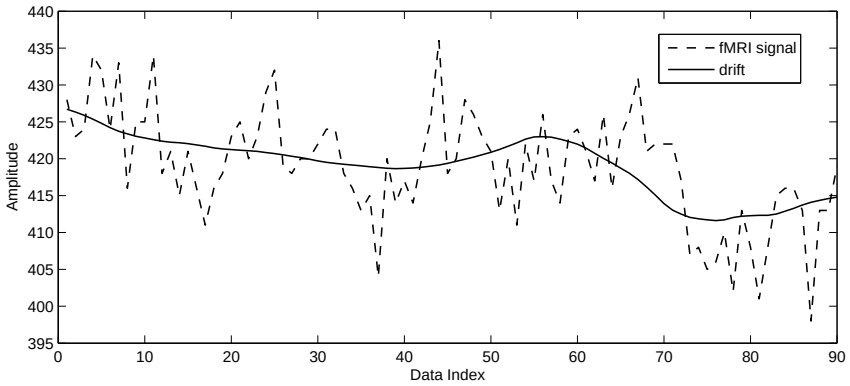


Figure 4.4 Time series in one voxel and the estimated drift.

activation of visual cortical areas of the brain are successfully detected. Figure 4.4 shows the time series of one voxel in the activated regions shown in Figure 4.3. The estimated drift is also shown here. It is seen from the figure that the proposed method successfully captures the slowly varying drift.

4.6 Conclusion

In the fMRI data analysis, the slowly varying drift poses a major problem for accurate detection of the activation. The commonly used subspace models for drift may cause overfitting problems. In this chapter, the drift is modelled to reside in the subspace spanned by the large scale wavelets. In addition, research shows that the noise in the fMRI data has a property of fractional noise or $1/f-$like noise. By first applying a wavelet transform to the fMRI data as a whitening filter, we propose a modified GLM which can estimate the parameter of activation at each scale of wavelet decomposition by using a Bayesian estimator. Then, a model selection criterion based on the parameters estimated from the modified GLM is proposed. Experiments on simulated as well as experimental fMRI data are carried out. These results show that the proposed Bayesian methods estimate accurately the covariance matrix of the noise and hence can give an efficient and robust estimator for the activation parameter. Besides, by applying the proposed model selection criterion, the drift in the fMRI data can be effectively estimated and removed. These advantages obtained by modelling the drift with only a few large wavelets can improve the brain activity detection in fMRI data analysis.

The Bayesian methods discussed till now mainly focused on the detection of the activated regions in the brain. It is also important to investigate the time course of the HDR to neural activities. In the following chapters (5 and 6), methods to estimate the HDR is discussed in detail.

Chapter 5

Adaptive Spatiotemporal Modelling and Estimation of the Event-related fMRI Responses

5.1 Introduction

In Chapters 2 to 4, methods to detect the activated regions of the brain are discussed. The detection of the activated regions can give spatial information of how different locations in the brain relate to sensory, cognitive and motor functions. In order to obtain temporal information of the brain activity, the HDR is normally investigated. In Chapters 5 and 6, methods (both linear and nonlinear) to estimate the HDR is provided and discussed.

Estimating the HDR is of great importance in the study of human brain functions. It reflects the temporal properties of the human brain activities. The HDR is usually estimated by the signal averaging procedure [82]. It assumes that the stimuli are separated far apart (large ISI) so that the HDRs to consecutive stimuli do not overlap. However, considering the fact that the ISI are often a few seconds and the HDR has temporal extent

for about 10-15 seconds [23], the HDRs to consecutive stimuli overlap and bias the results of signal averaging. Hence, procedures for removing this overlap while performing the signal averaging are desired.

In [82], it is reported that the HDRs to consecutively presented stimuli add roughly linearly to form the BOLD signal. This validates the means of estimating the HDR under the linear assumption, which assumes the linear summation of responses to each stimulus. This linear assumption is true for ISI more than 4-6 seconds; while for ISI shorter than that, nonlinear phenomena were observed. In Chapter 6 and 7, a detailed discussion on nonlinear estimation and modelling of fMRI data is given. In the GLM, the HDR is estimated through a linear combination of basis functions [50][25]. The commonly used basis functions are the Fourier series functions [83] and two gamma functions together with their temporal and dispersion derivatives [44]. In special cases, if the basis functions are chosen to be the non-overlapping sampling time bins during the period of peri-stimulus time, finding the amplitudes of these basis functions boils down to the problem of estimating a finite impulse response (FIR) [77][42]. Hence, the estimation of the HDR is actually a deconvolution problem under the assumption that fMRI signals are the output of a linear time-invariant (LTI) system.

Spatial information can be included into the analysis of the fMRI data to improve the detection of activated regions of the brain since these regions span a few voxels and the neighborhood to an activated voxel are more likely to be activated than the others. Many methods have been developed to incorporate the time series of surrounding voxels into the analysis to improve the SNR as well as the detection accuracy [84][85]. In this chapter, a combined spatial and temporal adaptive filter method is proposed. This method makes use of a spatial adaptive filter to improve the SNR. The spatially filtered signal is then used as the desired signal for the temporal adaptive filter to estimate the HDR. This proposed spatiotemporal method is proved to be equivalent to the CCA method and

through simulations, it is shown that the method works well for HDR estimation.

5.2 HDR Function

With the assumption that the human brain and MR acquisition system could be treated as a LTI system, the measured fMRI signal $y(n)$ at an activated voxel could be obtained from Eq. (4.1) as shown below:

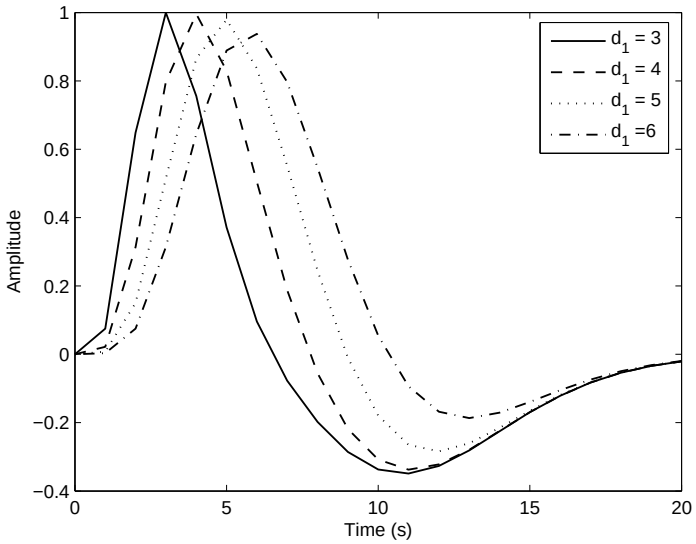
$$y(n) = h(n) \otimes s(n) + f(n) + \epsilon(n). \tag{5.1}$$

Here, the convolution of the stimulus function $s(n)$ and HDR function $h(n)$ represents the BOLD response. The fMRI signal $y(n)$ is then the summation of the BOLD response, drift $f(n)$ and noise $\epsilon(n)$. The polynomial method or median filtering method can be used to remove the drift $f(n)$ as a preprocessing step in the fMRI analysis [77].

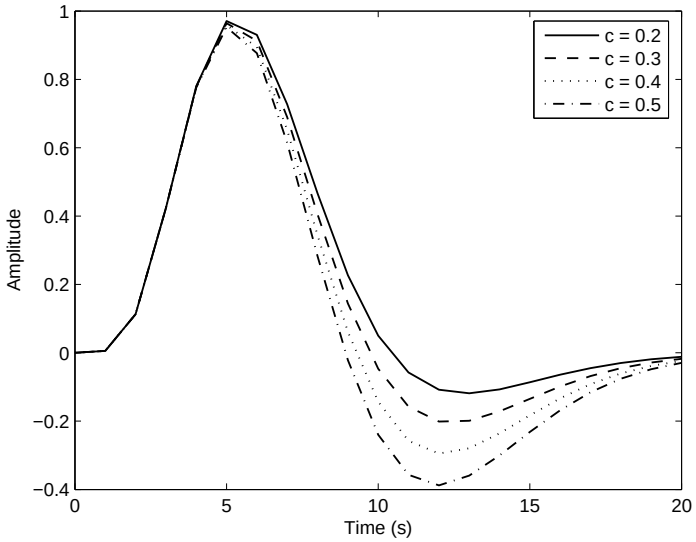
The HDR function is commonly described as the difference between two gamma functions [25] as shown in Eq. (1.16) and reiterated below:

$$h(t) = \Big(\frac{t}{d_1}\Big)^{a_1} \exp\Big(\frac{-(t-d_1)}{b_1}\Big) - c\Big(\frac{t}{d_2}\Big)^{a_2} \exp\Big(\frac{-(t-d_2)}{b_2}\Big) \tag{5.2}$$

where $d_j = a_j b_j$ is the time (in seconds) to the peak amplitude of $h(t)$. The common choice of these parameters are $a_1 = 6$, $a_2 = 12$, $b_1 = b_2 = 0.9s$ and $c = 0.35$ [44], which can generate the canonical HDR (represented as $\mathbf{h}_c$) commonly used in fMRI data analysis. By changing the values of these parameters, different shapes of HDR can be obtained. Figure 5.1 shows different shapes of the HDR function. For example, in Fig. 5.1 (a), $c = 0.35$ and d_1 varies between 3 and 6. In Fig. 5.1 (b), $d_1 = 5.4$ and c varies between 0.2 and 0.5. The parameter d_1 determines the delay of the response (larger the value of d_1, larger the delay) and c determines the depth of the undershoot (larger the value of c, larger the depth).



(a)



(b)

Figure 5.1 Simulated HDR functions for different parameter settings. (a) different values of d_1 while keeping $c = 0.35$ constant; (b) different values of c while keeping $d_1 = 5.4$ constant.

The observations from the fMRI experiments reveal that the shapes of the HDR functions are different for different brain regions and subjects [86]. Thus, it is desirable to incorporate the variation of the HDR into the analysis in order to have a good detection of the activated regions of the brain. In the GLM analysis method, the variation of the HDR is modelled through several basis functions (the canonical haemodynamic response function, its temporal and dispersion derivatives) and it is assumed to reside in the space spanned by these basis functions. However, in this method, these basis functions can only capture small variations of the HDR; for large variations, it may underestimate or even fail to capture the diversities of the HDR. In the next section, the HDRs at each voxel are estimated using adaptive filters and it is found that these estimated HDRs can reflect better the variation of brain responses in different regions of the brain.

5.3 Spatial and Temporal Adaptive Estimation

5.3.1 Model derivation

Considering both the spatial and temporal properties of the fMRI data, the adaptive estimation include two processes: *spatial smoothing* and *temporal modelling*. Suppose $y_0(n)$ is the n^{th} sample of the time series at a given voxel, and $y_1(n), \cdots, y_L(n)$ are the n^{th} samples of the time series of the L surrounding (adjacent) voxels. A spatially smoothed signal $d(n)$ is given by

$$d(n) = \sum_{i=0}^{L} w_i y_i(n) = \mathbf{w}^T \mathbf{y}(n) \tag{5.3}$$

where w_i is the i^{th} coefficient of the spatial filter, $\mathbf{w} = [w_0, w_1, \cdots, w_L]^T$ is the weight vector, and $\mathbf{y}(n) = [y_0(n), y_1(n), \cdots, y_L(n)]^T$. If the voxels are activated, the smoothed signal should approximate the ideal BOLD response $r(n)$ to the experimental stimuli,

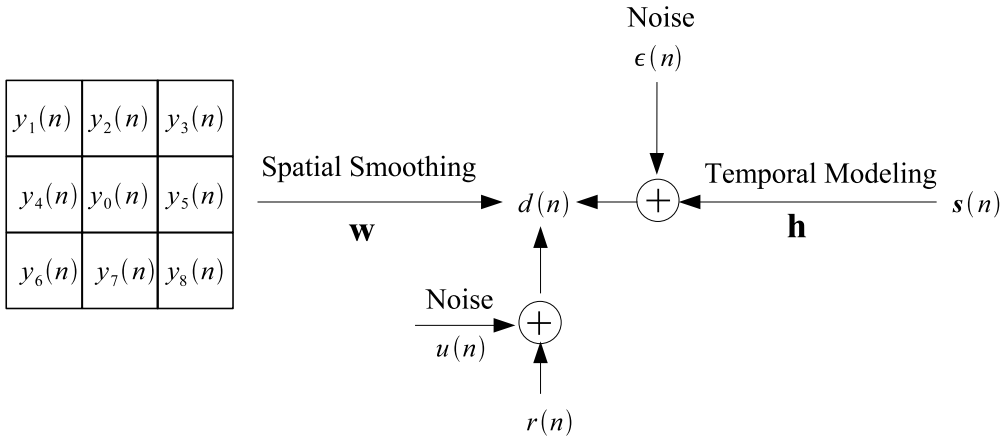


Figure 5.2 Illustration of the spatial smoothing filter and temporal modelling filter.

that is

$$d(n) = \mathbf{w}^T \mathbf{y}(n) = r(n) + u(n) \tag{5.4}$$

where $r(n)$ is the ideal BOLD response which is the convolution of the stimulus function and the canonical HDR, $u(n)$ is the white noise.

The spatially smoothed signal $d(n)$ is used as the desired signal for the temporal modelling process in Eq. (5.1). With the drift term $f(n)$ removed (by preprocessing), we get:

$$d(n) = \sum_{m=0}^{P} s(n-m)h_m + \epsilon(n) = \mathbf{h}^T \mathbf{s}(n) + \epsilon(n) \tag{5.5}$$

where $n = P, P+1, \ldots, N-1$, $\mathbf{h} = [h_0, h_1, \cdots, h_P]^T$, and $\mathbf{s}(n) = [s(n), s(n-1), \cdots, s(n-P)]^T$ is the vector formed by the delays of stimulus function. Here, P is the maximum temporal span of the HDR in time bins, which is related to the repetition time (TR) since the fMRI signal is sampled every TR. This is actually a transversal filter model, where $\mathbf{h}$ is the coefficient vector of the filter, $\mathbf{s}(n)$ is the input vector or regressor and $\epsilon(n)$ is the unmeasurable noise, which is assumed to be white.

Figure 5.2 shows an illustration of the spatial smoothing filter ($\mathbf{w}$) and temporal

modelling filter ($\mathbf{h}$). The goal of the analysis is to obtain an optimum temporal filter $\mathbf{h}_o$ so that the resultant BOLD signal (i.e. $\mathbf{h}^T\mathbf{s}(n)$) approximates the spatially smoothed signal $d(n)$ and meanwhile, $d(n)$ approximates the ideal BOLD signal $r(n)$ well. That is, both the mean-square errors (MSE) $E\{e_1^2(n)\} = E\{(r(n)-d(n))^2\}$ and $E\{e_2^2(n)\} = E\{(d(n) - \mathbf{h}^T\mathbf{s}(n))^2\}$ are minimum. Thus, the cost function to be minimized is given by

$$\begin{aligned} J &= E\{(r(n) - d(n))^2\} + E\{(d(n) - \mathbf{h}^T\mathbf{s}(n))^2\} \\ &= E\{(r(n) - \mathbf{w}^T\mathbf{y}(n))^2\} + E\{(\mathbf{w}^T\mathbf{y}(n) - \mathbf{h}^T\mathbf{s}(n))^2\} \end{aligned} \tag{5.6}$$

and at the optimum weights $\mathbf{w}_o$ and $\mathbf{h}_o$, J is minimum. Here, $E[\cdot]$ represents the expectation operation. The optimum weights can be found using widely used adaptive algorithms such as the least mean square (LMS) algorithm [87, 88].

The LMS algorithm minimizes the cost function $\hat{J} = e_1^2(n) + e_2^2(n)$ which is the instantaneous coarse estimate of J. The updating rule of the algorithm is shown below:

$$e_1(n) = r(n) - \hat{\mathbf{w}}^T(n)\mathbf{y}(n) \tag{5.7}$$

$$e_2(n) = \hat{\mathbf{w}}^T(n)\mathbf{y}(n) - \hat{\mathbf{h}}^T(n)\mathbf{s}(n) \tag{5.8}$$

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + 2\mu_1 e_1(n)\mathbf{y}(n) - 2\mu_1 e_2(n)\mathbf{y}(n) \tag{5.9}$$

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + 2\mu_2 e_2(n)\mathbf{s}(n) \tag{5.10}$$

where $\hat{\mathbf{h}}(n) = [\hat{h}_0(n), \hat{h}_1(n), \cdots, \hat{h}_P(n)]^T$ is the weight vector of the temporal modelling filter, $\hat{\mathbf{w}}(n) = [\hat{w}_0(n), \hat{w}_1(n), \cdots, \hat{w}_L(n)]^T$ is the weight vector of the spatial smoothing filter, μ_1 and μ_2 are the step sizes which control the speed of the convergence and the stability of the updating recursion.

Figure 5.3 shows the structure of the spatio-temporal adaptive filter. The estimation errors, $e_1(n)$ and $e_2(n)$ are fed back to the spatial adaptive filter to adjust the estimation of the coefficients $\hat{\mathbf{w}}$. The error $e_2(n)$ is fed back to the temporal modelling filter to

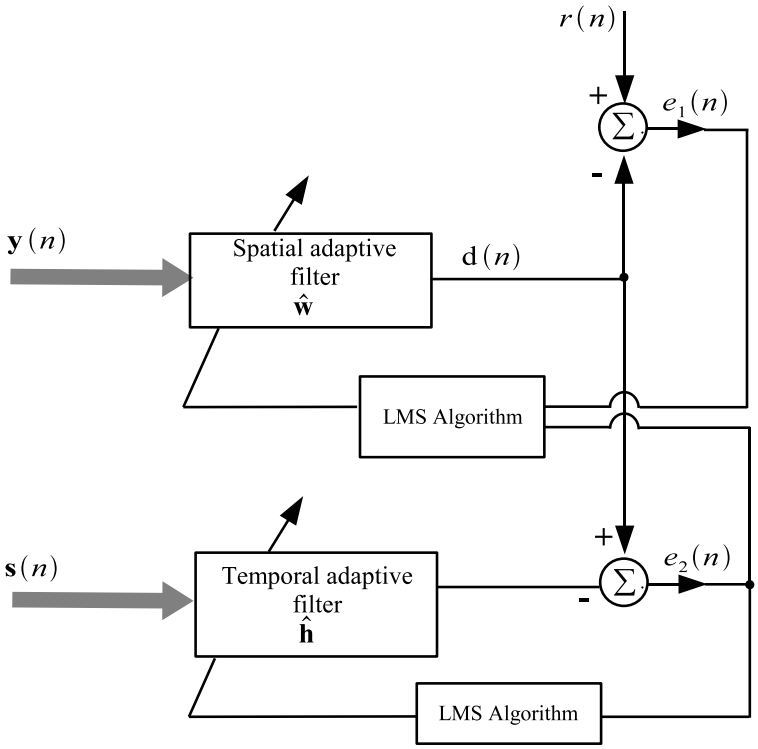


Figure 5.3 Spatio-temporal adaptive modelling of the fMRI system

adjust the estimation of the coefficients, $\hat{\mathbf{h}}$. As the algorithm converges, the estimated weights of the adaptive filter $\hat{\mathbf{w}}$ approximates the optimum spatial smoothing filter ($\mathbf{w}_o$), and the weights of the temporal adaptive filter $\hat{\mathbf{h}}$ approximates the HDR function.

5.3.2 Extension to Multiple Events

The HDR in the above algorithm is assumed to be fixed, that is, the optimum $\mathbf{h}_o$ does not change with time and/or with event types. Recently, some studies have shown that the hemodynamic function may vary from trial to trial [89][90]. To account for the variability of HDR in event-related fMRI data, several methods were proposed such as ICA to detect the variation of single-trial HDR [89] and single-trial variable model [90] where

the trial-to-trial variability is modelled as meaningful signal varying in the neighborhood of the initial HDR values. The extension of the proposed spatiotemporal adaptive filter method to model multiple event types or the time-varying effect of HDR for each trial is detailed in the following paragraphs. In these cases, the stimulus functions are multiple time series, $\mathbf{s}_1(n), \mathbf{s}_2(n), \cdots, \mathbf{s}_M(n)$, and the estimation of HDRs for each event types (multiple event types case) or different trials (time-varying cases) are $\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \cdots, \hat{\mathbf{h}}_M$, respectively. Here, M represents the number of events in the multiple event type fMRI experiment or the number of trials. The estimation error is

$$e_2(n) = d(n) - \hat{\mathbf{h}}_1^T(n)\mathbf{s}_1(n) - \hat{\mathbf{h}}_2^T(n)\mathbf{s}_2(n) - \cdots - \hat{\mathbf{h}}_M^T(n)\mathbf{s}_M(n). \tag{5.11}$$

Correspondingly, the update rule in Eq. (5.10) becomes

$$\hat{\mathbf{h}}_i(n+1) = \hat{\mathbf{h}}_i(n) + 2\mu e_2(n)\mathbf{s}_i(n); \qquad i = 1, 2, \cdots, M \tag{5.12}$$

with μ the step size. The desired signal $d(n)$ is the output of the spatial smoothing filter ($\hat{\mathbf{w}}^T(n)\mathbf{y}(n)$) as defined before. The point here is that there are M modelling filters to capture the properties of HDRs to different event types and different trials. The same error, $e_2(n)$, is used to update the filter coefficients of all the M transversal adaptive filters.

Due to the limited number of data points in one voxel time series (normally around a few hundred), the algorithm does not converge to the optimum value after one presentation of the entire training set (that is, one epoch in the learning process). Therefore, the training data set is presented repeatedly (many epochs) until the algorithm converges to the optimum solution.

5.3.3 Relation to the Canonical Correlation Analysis

In this section, we prove that minimizing the cost function in Eq. (5.6) is equivalent to the Canonical Correlation Analysis (CCA) method. CCA is a multivariate analysis

method [91] to detect the subcomponents of two multivariate data sets that are maximally correlated. Consider two zero mean random vectors $\mathbf{a}$ and $\mathbf{b}$, CCA is to find the weight coefficients $\mathbf{w}_a$ and $\mathbf{w}_b$ which maximizes the correlation ρ between the linear combinations, $a = \mathbf{w}_a^T\mathbf{a}$ and $b = \mathbf{w}_b^T\mathbf{b}$. The CCA problem can be expressed as the following maximization problem:

$$\max_{\mathbf{w}_a,\mathbf{w}_b} \rho(\mathbf{w}_a, \mathbf{w}_b) = \frac{\mathbf{w}_a^T\mathbf{C}_{ab}\mathbf{w}_b}{\sqrt{\mathbf{w}_a^T\mathbf{C}_{aa}\mathbf{w}_a\mathbf{w}_b^T\mathbf{C}_{bb}\mathbf{w}_b}}. \tag{5.13}$$

For zero mean vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{C}_{aa} = E[\mathbf{a}\mathbf{a}^T]$, $\mathbf{C}_{bb} = E[\mathbf{b}\mathbf{b}^T]$ and $\mathbf{C}_{ab} = E[\mathbf{a}\mathbf{b}^T]$ are the covariance matrices. The solutions $\hat{\mathbf{w}}_a$ and $\hat{\mathbf{w}}_b$ are found by solving the following eigenvalue problems:

$$\mathbf{C}_{aa}^{-1}\mathbf{C}_{ab}\mathbf{C}_{bb}^{-1}\mathbf{C}_{ba}\hat{\mathbf{w}}_a = \rho^2\hat{\mathbf{w}}_a \tag{5.14}$$

$$\mathbf{C}_{bb}^{-1}\mathbf{C}_{ba}\mathbf{C}_{aa}^{-1}\mathbf{C}_{ab}\hat{\mathbf{w}}_b = \rho^2\hat{\mathbf{w}}_b. \tag{5.15}$$

Suppose $\mathbf{y}(n)$ and $\mathbf{s}(n)$ are zero mean vectors (i.e. $E\{\mathbf{y}(n)\} = \mathbf{0}$ and $E\{\mathbf{s}(n)\} = \mathbf{0}$) in Eq. (5.6). By taking the partial derivatives of Eq. (5.6) with respect to $\mathbf{w}$, the following can be derived:

$$\begin{aligned}\frac{\partial J}{\partial \mathbf{w}} &= -2E\{\mathbf{y}(n)(r(n) - \mathbf{y}^T(n)\mathbf{w})\} + 2E\{\mathbf{y}(n)(\mathbf{y}^T(n)\mathbf{w} - \mathbf{s}^T(n)\mathbf{h})\} \\ &= -2E\{\mathbf{y}(n)r(n)\} + 2E\{\mathbf{y}(n)\mathbf{y}^T(n)\}\mathbf{w} \\ &\quad +2E\{\mathbf{y}(n)\mathbf{y}^T(n)\}\mathbf{w} - 2E\{\mathbf{y}(n)\mathbf{s}^T(n)\}\mathbf{h}\end{aligned} \tag{5.16}$$

The reference signal $r(n)$ is the result of convolution of $\mathbf{s}(n)$ and the canonical HDR functions $\mathbf{h}_c$, that is, $r(n) = \mathbf{h}_c^T\mathbf{s}(n)$. The temporal modelling filter, $\mathbf{h}$, in our model may be represented as a scaled and noisy version of the canonical HDR $\mathbf{h}_c$, i.e. $\mathbf{h} = \lambda\mathbf{h}_c + \boldsymbol{v}$, where $\boldsymbol{v}$ is a zero mean white noise. This means that the temporal modelling filter $\mathbf{h}$ is close to the scaled version of canonical HDR $\mathbf{h}_c$ and $\boldsymbol{v}$ capture the departure from the canonical HDR. The parameter, λ ($0 \leq \lambda \leq 1$) indicates the neural activity, when λ is

close to zero, the temporal modelling filter $\mathbf{h}$ is also close to zero showing that the voxel under consideration is not activated. On the other hand, when λ is close to one, the temporal modelling filter, $\mathbf{h}$, has properties in similar to those of the canonical HDR, $\mathbf{h}_c$, showing the voxel under consideration is activated. Hence, the reference signal, $r(n)$, can be represented as

$$r(n) = \mathbf{h}_c^T \mathbf{s}(n) = \mathbf{s}^T(n)\mathbf{h}_c = \mathbf{s}^T(n)\left(\frac{\mathbf{h} - \boldsymbol{v}}{\lambda}\right) \tag{5.17}$$

Substituting Eq. (5.17) into Eq. (5.16) and noting that $\boldsymbol{v}$ is zero mean and uncorrelated with $\mathbf{s}(n)$ and $\mathbf{y}(n)$, Eq. (5.16) becomes

$$\frac{\partial J}{\partial \mathbf{w}} = -\frac{2}{\lambda} E\{\mathbf{y}(n)\mathbf{s}^T(n)\}\mathbf{h} + 4E\{\mathbf{y}(n)\mathbf{y}^T(n)\}\mathbf{w} - 2E\{\mathbf{y}(n)\mathbf{s}^T(n)\}\mathbf{h}. \tag{5.18}$$

By taking the partial derivatives of Eq. (5.6) with respect to $\mathbf{h}$, we get

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{h}} &= 2E\{\mathbf{s}(n)(\mathbf{y}^T(n)\mathbf{w} - \mathbf{s}^T(n)\mathbf{h})\} \\ &= 2E\{\mathbf{s}(n)\mathbf{y}^T(n)\}\mathbf{w} - 2E\{\mathbf{s}(n)\mathbf{s}^T(n)\}\mathbf{h} \end{aligned} \tag{5.19}$$

Equating Eq. (5.18) and Eq. (5.19) to zero, and representing $\mathbf{C}_{yy} = E\{\mathbf{y}(n)\mathbf{y}^T(n)\}$, $\mathbf{C}_{ss} = E\{\mathbf{s}(n)\mathbf{s}^T(n)\}$, $\mathbf{C}_{sy} = E\{\mathbf{s}(n)\mathbf{y}^T(n)\}$ and $\mathbf{C}_{ys} = E\{\mathbf{y}(n)\mathbf{s}^T(n)\}$, we get

$$2\mathbf{C}_{yy}\mathbf{w} = (1 + \frac{1}{\lambda})\mathbf{C}_{ys}\mathbf{h} \tag{5.20}$$

and

$$\mathbf{C}_{sy}\mathbf{w} = \mathbf{C}_{ss}\mathbf{h}. \tag{5.21}$$

The solutions to the above equations are obtained as the eigenvalue problem:

$$\mathbf{C}_{yy}^{-1}\mathbf{C}_{ys}\mathbf{C}_{ss}^{-1}\mathbf{C}_{sy}\hat{\mathbf{w}} = \frac{2\lambda}{\lambda + 1}\hat{\mathbf{w}} \tag{5.22}$$

$$\mathbf{C}_{ss}^{-1}\mathbf{C}_{sy}\mathbf{C}_{yy}^{-1}\mathbf{C}_{ys}\hat{\mathbf{h}} = \frac{2\lambda}{\lambda + 1}\hat{\mathbf{h}} \tag{5.23}$$

Comparing Eq. (5.22) and Eq. (5.14) as well as Eq. (5.23) and Eq. (5.15), it is clear that the canonical correlation coefficient ρ is related to the neural activity parameter λ in our model:

$$\rho^2 = \frac{2\lambda}{\lambda + 1}, \qquad (0 \leq \lambda \leq 1) \tag{5.24}$$

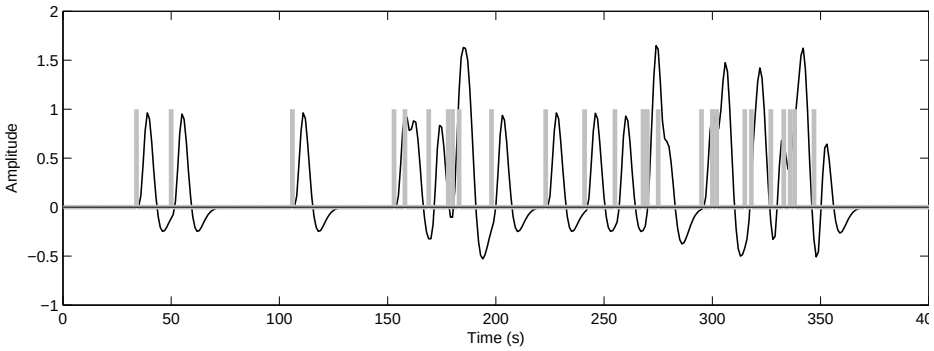
and $\mathbf{w}_a = \mathbf{w}$, $\mathbf{w}_b = \mathbf{h}$, $\mathbf{a} = \mathbf{y}(n)$ and $\mathbf{b} = \mathbf{s}(n)$. When the neural activity occurs, that is, $\lambda \approx 1$, the correlation coefficient ρ also approaches 1. When there is no neural activity, that is, $\lambda \approx 0$, the correlation coefficient ρ also approaches 0. This proves the equivalence of these two approaches.

5.4 Results and Discussion

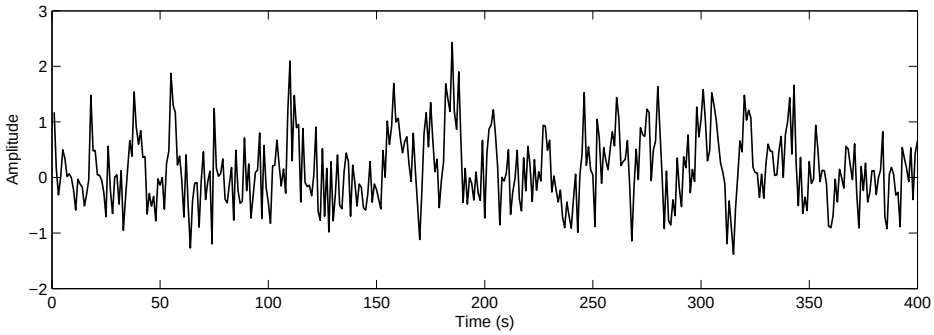
The proposed spatio-temporal adaptive modelling method is first tested on synthetic data and compared with the GLM method with different Gaussian smoothing width [92]. The scenarios for single event type and multiple event types are both examined. After this, the application of the proposed algorithm to experimental fMRI data is investigated. The results are discussed in detail in the following sections.

5.4.1 Simulated data

The simulated BOLD signal is generated through convolution of the input stimulus signal $s(n)$ and the HDR function described as the difference between two Gamma functions which is introduced in Eq. (5.2). The input stimulus signal $s(n)$ is generated as random impulses which represent an event-related fMRI experiment. When the stimulation is ON, $s(n) = 1$ and when the stimulation is OFF, $s(n) = 0$. Figure 5.4 (a) illustrates the simulated BOLD signal with the HDR function parameters $d_1 = 5.4, c = 0.35$. The thick vertical lines indicating the timing of the discrete random stimuli.



(a) Pure BOLD signal



(b) Noisy BOLD signal

Figure 5.4 Simulated BOLD signal. (a) pure BOLD signal and the timing of the stimuli; (b) noisy BOLD signal corrupted with Gaussian white noise.

Estimation of the Single Event Type HDR

In the first simulation, the relation between the spatial weighting coefficients and the levels of the noise variance is examined. The BOLD signal generated as shown in Fig. 5.4 (a) is first added to each voxel of the 3×3 window. Then, Gaussian noises with different variances are added to each voxel of the window to generate the simulated voxel time series. One of the simulated noisy voxel time series is shown in Figure 5.4 (b) with SNR = -0.8dB.

These simulated noisy signals are then investigated by the spatio-temporal adaptive filter. The step sizes μ_1 and μ_2 are chosen to be 0.0001. After one presentation of the entire training set, the average change of the coefficients are calculated; if this change is very small, the algorithm converges and the updating process stops, otherwise, the whole training data is used again until the algorithm converges. Figure 5.5 illustrates the learning curve of LMS algorithm for spatiotemporal adaptive filter. The mean square error is calculated at the end of each epoch (one presentation of the whole training data). It is seen that after 100 epochs, the algorithm almost converges.

Table 5.1 shows the relation of levels of the noise variances and the coefficients of the spatial adaptive filter after convergence. It is seen from this table that the coefficient values $\hat{w}_i$ of the spatial smoothing filter are approximately inversely proportional to the noise level at each voxel. When the voxel time series are noisier compared to other voxels, the weights are relatively smaller; while larger weights are found at the voxels which are less noisy. This validates the adaptivity of the spatial smoothing filter to different noise levels.

The estimated HDR is then investigated. In order to test the effect of $r(n)$ (formed by the conventional HDR function with $d_1 = 5.4, c = 0.35$ as introduced in Eq. (5.2)) on the estimation results, the BOLD signal in this simulation is generated using the difference

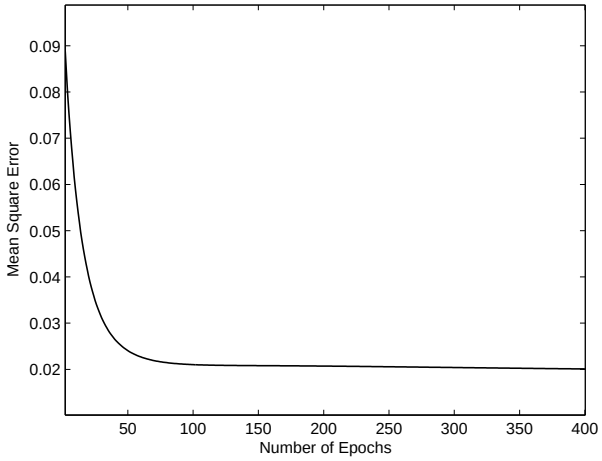


Figure 5.5 Learning curve of LMS algorithm for spatiotemporal adaptive filter.

Table 5.1 Relation of the levels of the noise variances and the coefficients of the spatial adaptive filter for a 3×3 window.

position	**noise variance**	**coefficients** $(\hat{w}_i)$
center	0.25	0.0326
top	0.01	0.5182
bottom	0.09	0.1151
left	0.49	0.0260
right	0.81	0.0121
top-left	1.00	0.0074
top-right	1.44	0.0068
bottom-left	1.96	0.0059
bottom-right	2.56	0.0052

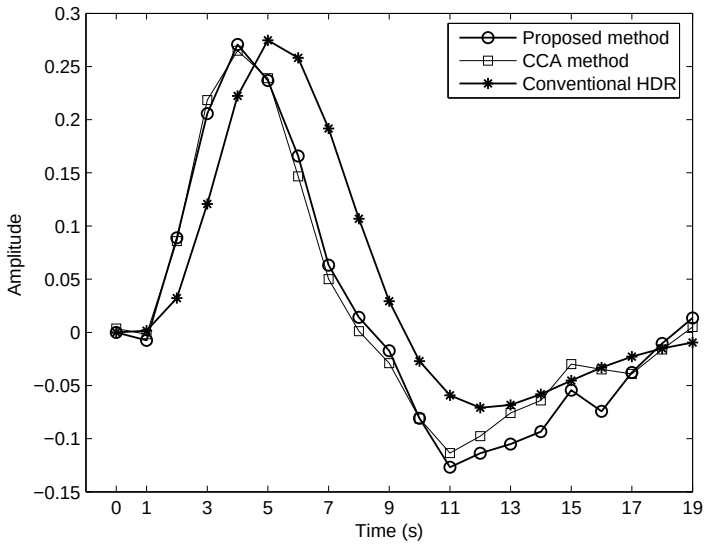


Figure 5.6 The HDRs estimated by the spatio-temporal adaptive filter and CCA methods.

gamma function with the following parameters $d_1 = 4, c = 0.35$, which has a different delay with the conventional HDR function. As before, noises with different variances are then added to the BOLD signal to generate noisy fMRI signals. After applying the proposed adaptive filter method, the estimated HDR is the impulse response of the temporal modelling filter, $\mathbf{h}$, once the algorithm is converged. For the CCA method, the estimated HDR is obtained from the eigenvector corresponding to the largest eigenvalue found in Eq.(5.15). Figure 5.6 shows the results of the estimated HDR and the difference with the conventional HDR function.

From Fig. 5.6, we can see that the similarity of the HDR estimated by the proposed method to that estimated by the CCA method. This result validates the conclusion that the proposed spatio-temporal adaptive filter is in fact equivalent to the CCA method. It also can be seen from the figure that $r(n)$ has no effect on the estimated HDRs. When the conventional HDR which forms $r(n)$ in the proposed algorithm is different from the

Table 5.2 Comparison of the proposed adaptive filter method, CCA and GLM for the estimation of HDR.

	proposed method	**CCA**	**GLM**
NMSE	1.4150	1.2063	2.3610

actual HDR, the proposed algorithm could still work well to estimate the actual HDR function. This is because that $r(n)$ is used as a part of the desired signal to update the coefficients of the spatial filter and it is the spatially smoothed signal which is used as the desired signal for the HDR estimation of the temporal filter as shown in Fig. 5.3.

The goodness-of-fit of the proposed method, CCA and GLM (with conventional HDR function and its temporal and dispersion derivatives) to the actual HDR of the above simulation, is evaluated using the normalized mean square error (NMSE) defined as:

$$NMSE = \frac{\sum_{i=0}^{L-1}(\hat{h}_i - h_i)^2}{\sum_{i=0}^{L-1} h_i^2} \tag{5.25}$$

where $\hat{h}_i$ and h_i are respectively the estimated and the true values of HDR function. Table 5.2 shows the NMSE values for the proposed method, CCA and GLM methods.

It is seen from Table 5.2 that the NMSE of the proposed adaptive filter method is close to the CCA method and both of them are lower than the GLM method. This means that the performance of the GLM method is inferior to the other two methods in estimating the HDR. This is because the GLM with the conventional HDR can only model small variations of the HDR; for large variations, the GLM lacks the flexibility to capture the large difference between the actual HDR and the conventional HDR.

Estimation of Multiple Event Type HDRs

Though the proposed method and the CCA method are equivalent, the advantage of the proposed spatio-temporal method is that it is easier to extend to multiple event types when subjects undergo multiple types of stimulation (for example, visual and auditory stimulations). To substantiate this point, in this simulation, the scenario of two different event types is studied by the proposed method. The two events stimulus onset functions are randomly generated and the HDRs (HDR_1 and HDR_2) corresponding to these two event types are assumed to have different delays ($d_1 = 4$ and 6.5 respectively for HDR_1 and HDR_2 in Eq. (5.2)). These two HDRs are convolved with different stimulus functions, resulting in the simulated BOLD signal. Gaussian white noise is then added to this simulated BOLD signal to generate the noisy fMRI signal with SNR = -2dB. The generated noisy signal and the respective stimulus functions are then fed into the filter to adaptively estimate the weights of the transversal filter as shown in Eq. (5.11) with two temporal adaptive filters. Figure 5.7 shows the estimated HDRs for these two event types. From this figure, it is seen that the delays of these two estimated HDRs are correctly estimated (respectively about 4 and 6.5). This clearly shows that the proposed method can capture the difference between these two HDRs.

Comparison of Activation Detection Ability

The detection ability of the proposed spatio-temporal adaptive filter method and the GLM method are compared. A slice from an experimental fMRI data set is used as the background image. At some specific regions of the image, the simulated BOLD signals generated by difference of gamma functions with different delays ($d_1 = 4, 5, 6$) are added. Gaussian white noises with different variances are then added to each voxel of the data to form a 3-D fMRI time series at different SNRs (ranging from -7dB to 0dB).

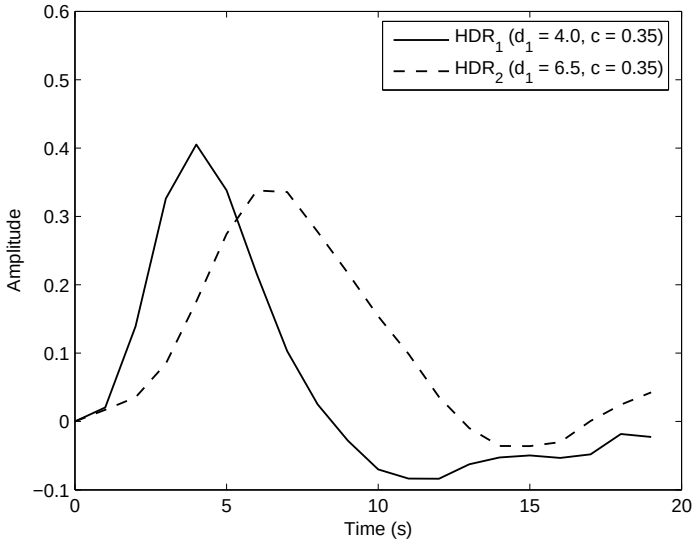


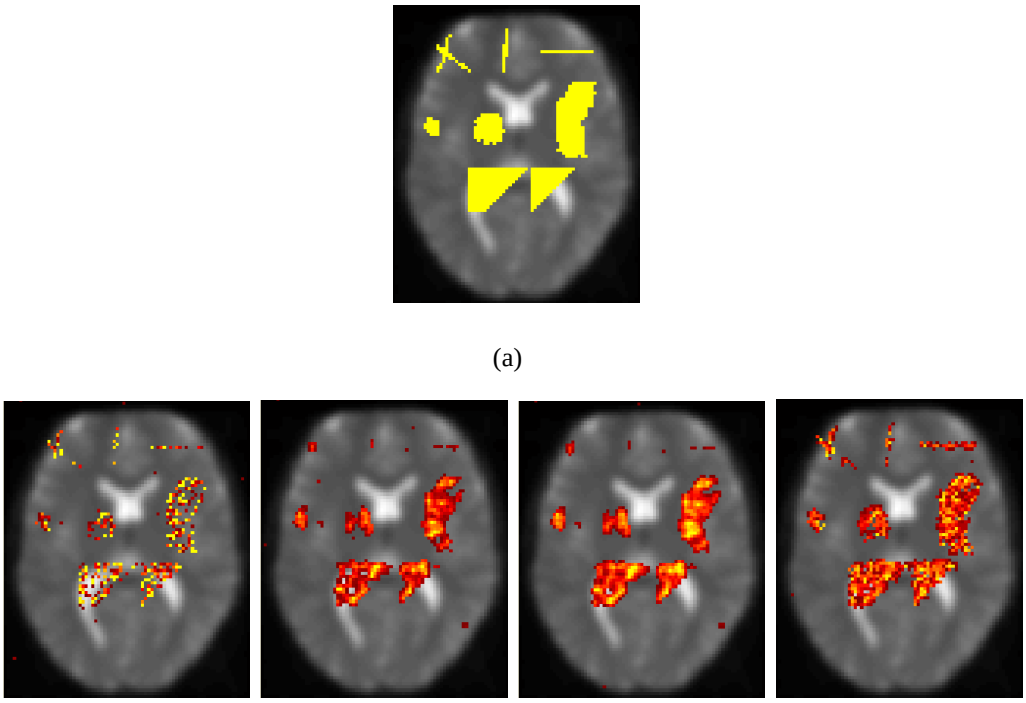
Figure 5.7 Estimated HDRs to two event types using the proposed method.

These simulated fMRI data are fitted by the GLM model with the design matrix composed of four regressors: the canonical HDR function, its temporal derivative and dispersion derivative (these three regressors model the variation of HDR) and a constant vector of value 1. The statistical parameter map (SPM) is then obtained based on the *t* statistic calculated at each voxel and a threshold is determined to find the activation regions of the brain. For the proposed adaptive filter method learned through the LMS algorithm, a 3×3 spatial window is considered. The initial values of the coefficients of the spatial filter are assumed to be $1/9$ and the initial weights of the temporal adaptive transversal filter are chosen to be 1. The values of parameter μ_1 and μ_2 in the LMS algorithm are chosen to be 0.0001, and the training data are used repeatedly until the algorithm converges. The reconstructed BOLD signal is then the convolution of the estimated HDR (after convergence) and the stimulant function. The correlation coefficients ρ between this reconstructed BOLD signal and the spatially smoothed signal are

calculated at each voxel. The statistical parametric map is then formed and thresholded to illustrate the activation regions.

Figure 5.8 shows the detection results of simulated fMRI data using these methods. Figure 5.8 (a) illustrates the simulated activation patterns. Figure 5.8 (b) is the detection results of the GLM method without pre-smoothing. As it is clear from the figure, this method almost fails to detect the simulated activation patterns with only a few sporadic voxels detected as active. Figure 5.8 (c) and (d) are the detection results of the GLM method with the image spatially smoothed by Gaussian spatial filters with different width (Full-Width-Half-Maximum (FWHM) = 3 voxels for (c) and FWHM = 5 voxels for (d)). Although Gaussian pre-smoothing gives better results compared to the one without pre-smoothing, it loses details of the activation patterns due to the isotropic property of Gaussian spatial smoothing, which spreads the energy of the activated voxels to the surrounding voxels. Figure 5.8 (e) shows the results of the proposed spatio-temporal adaptive filter method. Compared to the other results, it is clear that this method is more robust and sensitive than the GLM method with and without spatial pre-smoothing. The proposed method can give a good estimation of the activation patterns due to the ability to spatially adapt to the true activation patterns.

To have a clearer comparison of the detection ability, receiver operator characteristic (ROC) analysis is used to investigate the activation maps generated by the GLM and the proposed method. Figure 5.9 shows the ROC curves of the GLM method with and without pre-smoothing and the proposed spatio-temporal adaptive method, respectively. From these ROC curves, it is clear that under the same false positive ratio, the proposed spatio-temporal adaptive filter method could actually detect more real activations. This clearly shows that the proposed spatio-temporal adaptive filter method is a better method for activation detection and activation patterns estimation.



(a) (b) (c) (d) (e)

Figure 5.8 Detection results of simulated fMRI data: (a) Simulated activation pattern; (b) GLM without spatial smoothing ($t > 3$); (c) GLM with spatial smoothing the FWHM is 3 voxel ($t > 3$); (d) GLM with spatial smoothing the FWHM is 5 voxel ($t > 3$); (e) Spatio-temporal adaptive filter method ($\rho > 0.3$).

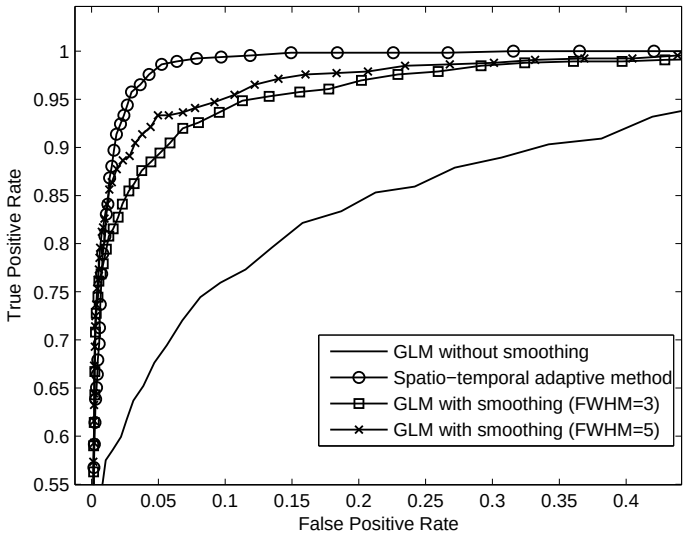


Figure 5.9 ROC curves for simulated noisy data.

5.4.2 Experimental fMRI Data

The proposed spatio-temporal adaptive filtering method is also validated on the experimental event-related fMRI data (*DATA-EVENT* introduced in Chapter 1). Before applying the proposed method to this experimental event-related fMRI data, the raw data are preprocessed by the SPM software for registration and normalization. The drift in each voxel time series is also removed. The resultant pre-processed data are analyzed using the proposed spatio-temporal adaptive filter method. Since the TR in this experiment is 2 seconds, the number of taps P in the adaptive transversal filter is chosen to be 10 so that the estimated HDR has a temporal extent of 18 seconds. The HDR estimated using the proposed method is then convolved with the stimulus function resulting in the reconstructed BOLD signal. The correlation coefficients ρ between this reconstructed BOLD signal and the spatially smoothed signal are calculated at each voxel. Figure 5.10 shows the results of the activation detection. The color bar shows the color corresponding to

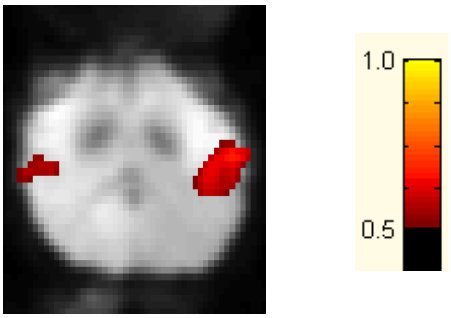


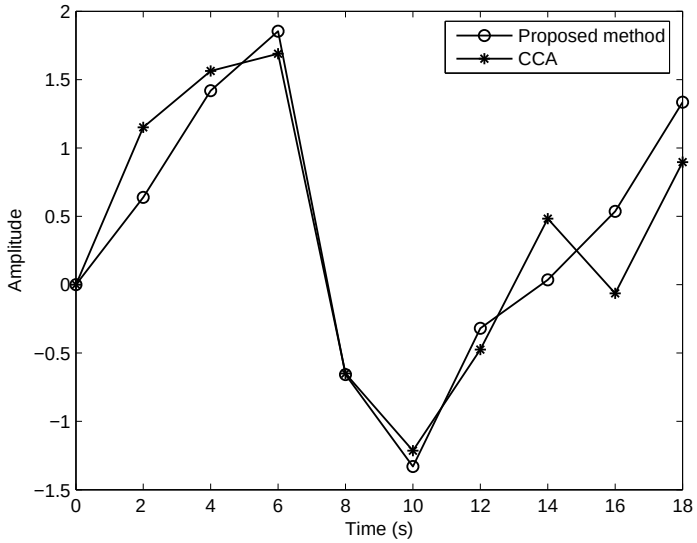
Figure 5.10 One slice showing the activation of the auditory cortex ($\rho > 0.5$).

the ρ from 0.5 to 1. When ρ is less than 0.5, it is represented by the original grey level at that voxel. And the From this figure, it is clear that the activation is detected in the auditory cortex.

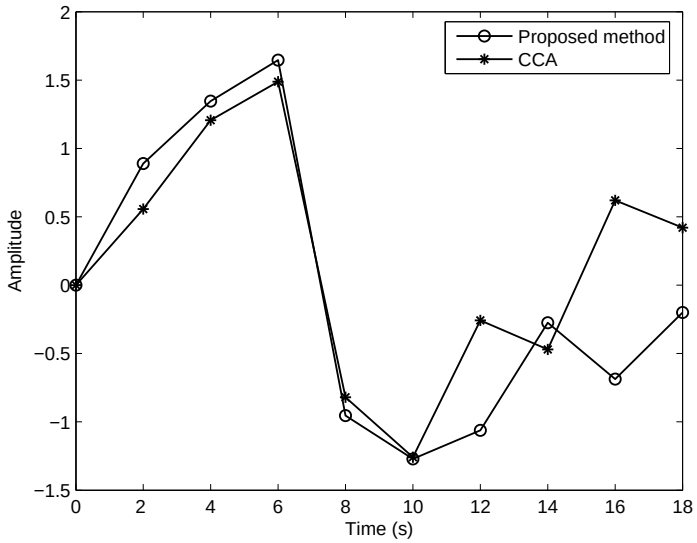
Figure 5.11 shows the estimated HDRs using the proposed spatio-temporal adaptive filter method and the CCA method to the time courses of activated voxels located respectively at the left and right auditory cortex shown in Figure 5.10. From this figure, it is seen that the estimated HDRs of the activated voxels show similar patterns of variations, achieving the peaks at the delay of about 4-6s and then decreasing below the baseline with the post stimulus undershoot. This validates our claim that the proposed method is equivalent to the CCA method.

5.5 Conclusion

From the signal processing perspective, fMRI data analysis can be considered as a system modelling problem. With the knowledge of experimental paradigm (input) and measured data (output), this complex system (the brain and scanner) could be identified. In this chapter, linear modelling methods are investigated for the event-related fMRI



(a)



(b)

Figure 5.11 The estimation of HDRs for the activated voxels using the proposed adaptive filter method and CCA method. (a) One voxel in the left auditory cortex; (b) One voxel in the right auditory cortex.

data.

Considered as a linear time-invariant (LTI) system, the human brain and fMRI data acquistion can be modelled using adaptive transversal filters. It is also known that the activation regions span several millimeters or a few voxels. Incorporating the spatial information into our analysis can increase the detection power of the methods. The Gaussian spatial smoothing filter is commonly used in fMRI data analysis as a preprocessing step to increase the SNR. However, due to the isotropic property of the Gaussian spatial filter, this method would blur the details of the activation patterns. In this chapter, an adaptive spatio-temporal modelling and estimation method is proposed. Two adaptive filters are used: (*i*) the spatial filter to adaptively estimate the activation patterns and (*ii*) the temporal filter to adaptively model the HDR. The weights of these filters are updated using the well-known LMS adaptive algorithm. The relation of the proposed method to the CCA method is also proven. Both the synthetic as well as experimental event-related fMRI data are examined. Results from extensive simulation studies show that the proposed spatio-temporal adaptive method is superior to GLM with Gaussian smoothing. The proposed method has the ability to detect the details of the activation pattern without blurring it. Besides, the HDR estimated by the temporal filter would capture the variation of brain responses between different brain regions and different persons. These simulation studies suggest that the spatio-temporal adaptive filter method is a superior approach to model and investigate the fMRI data.

With the advancement of the fMRI techniques as well as the experimental protocol, rapid event-related fMRI experiment becomes popular these days. The ISI in this type of fMRI experiments is less than 4-6 seconds, which violates the linear assumption underlying the spatio-temporal adaptive filter method proposed in this chapter. For the rapid event-related fMRI experiments, nonlinear phenomena are often observed. Thus, in the following chapter, nonlinear methods to estimate the HDR is investigated.

Chapter 6

Estimation of the Hemodynamic Response of fMRI Data using RBF Neural Network

6.1 Introduction

In Chapter 5, a spatiotemporal adaptive filter is proposed to estimate the HDR under the linear assumptions (when the ISI is larger than 4-6 seconds). However, the fMRI signals also exhibit nonlinear properties especially in the rapid event-related fMRI experiments when the stimulation duration or ISI is less than 4-6 seconds [95][96][97]. These non-linearities arise from both vascular and neuronal levels [98]. Several advanced nonlinear models have been proposed to analyze such fMRI data. The nonlinear dynamics of the BOLD signal is first described by the Balloon model, which is a physiologically derived model introduced by Buxton et al [26]. The couplings between blood flow and blood oxygen concentration changes on a biological level are incorporated into this nonlinear state-space model. The interpretability of the parameters in the Balloon model makes it

a suitable model to understand the nonlinear mechanisms underlying the BOLD effect. The Balloon model has undergone several extensions since its introduction. Friston et al. extended the original model to include a linear interaction between synaptic activity (or electrophysiology) and the microvascular control system [20]. Recently, Buxton et al. incorporated $CMRO_2$ and neural activity as new variables in the Balloon model and described the steps linking an external stimulus to the measured BOLD and CBF responses [99]. These physiologically derived nonlinear models have many advantages such as meaningful interpretations of the parameters. However, it is not an easy task to estimate the parameters in such nonlinear models. The non-physiological models described in the following sections, on the other hand, are easier to implement and more flexible for a better mapping of the input to the output if the signal is highly nonlinear.

The Volterra series method, which can model any dynamical input-output system, was also proposed to analyze and estimate the HDR using the 1st and 2nd order Volterra kernels [25]. Though this method represents the nonlinear properties of the BOLD signal, it has a number of limitations. To obtain a better representation of the dynamical system, higher order Volterra kernels are required. However, the inclusion of higher order Volterra kernels requires fitting a large number of parameters. Especially, the number of terms in the kernels of the series increases exponentially with the order of the series. Due to these difficulties and the complexity in identifying the higher order kernels, an alternative, yet efficient method to estimate the kernels in the Volterra series is desired.

In this chapter, a novel method to the estimation of the kernels is proposed. This method describes the non-linear dynamics of the fMRI hemodynamic response using the radial basis function (RBF) neural network. It is known that neural network could approximate any continuous functions to any degree of accuracy due to its universal approximation property. Modelling the fMRI system as nonlinear and dynamical, the

RBF neural network is capable of regressing nonlinearly the BOLD signal to the external stimuli. Additionally, the relationship between the parameters of the RBF neural network and Volterra kernels is provided and the HDR could easily be estimated from the parameters of the neural network. This provides an efficient method to calculate the large number of parameters in the Volterra series kernels.

In the following sections, we first introduce the Volterra series model and Balloon model for the analysis of fMRI data. Then, the RBF neural network method is utilized to model the fMRI data and the relation of proposed neural network method to the Volterra series method is demonstrated.

6.2 Volterra Series Model

The measured fMRI signal, $y(n)(n = 0, 1, \cdots, N-1)$, can be described as a nonlinear convolution of the stimulus function, $s(n)$, with the Volterra kernels where the nonlinear properties are captured by the higher order kernels. The Volterra series method to model the fMRI signal does not include the variables such as blood volume, blood flow, oxygenation etc introduced in the Balloon model which makes the parameters in the model easier to estimate. When the BOLD signal is represented in the form of a finite (2nd order) Volterra series, the output BOLD signal, $y_b(n)$, can be written as a function of the input stimulus, $s(n)$, as:

$$y_b(n) = a_0 + \sum_{i=0}^{P} a_1(i)s(n-i) + \sum_{i=0}^{P}\sum_{j=0}^{P} a_2(i,j)s(n-i)s(n-j) \tag{6.1}$$

where the length of the kernels is $(P+1)$. The constant a_0 is the zeroth-order kernel. The coefficients $a_1(\cdot)$ are the first-order kernels which relate the output as the weighted sum of the present and the recent past inputs; these coefficients represent the HDR in

fMRI. The coefficients $a_2(\cdot,\cdot)$ in the above equation are the second-order Volterra kernels which represent the output as the interactions between the present and/or the recent past inputs at different time points. It can be observed the existence of the kernels $a_2(i,j) = a_2(j,i), \forall i,j$, due to the symmetry property. The input stimulus $s(n)$ represents the timing of the external stimuli with the value equal to '1' when the stimulus is present (ON) and '0' when the stimulus is absent (OFF).

The measured fMRI signal, $y(n)$, is a noise corrupted version of the BOLD signal, $y_B(n)$, and can be represented as:

$$y(n) = y_b(n) + e(n) \tag{6.2}$$

where $e(n)$ is the additive noise. The solution to the Volterra series model could be obtained using a least-square method (assuming the noise to be white Gaussian). Substituting Eq. (6.1) into Eq. (6.2), the latter can be represented in the matrix form as:

$$\mathbf{y} = \mathbf{\Psi}\mathbf{x} + \mathbf{e} \tag{6.3}$$

where, $\mathbf{y} = [y(P), \cdots, y(N-1)]^T$ is the measured signal of dimension $(N-P)\times 1$; $\mathbf{e} = [e(P), \cdots, e(N-1)]^T$ is the noise vector with the same dimension. The coefficient vector $\mathbf{x}$ (of dimension $(L\times 1)$) and the matrix $\mathbf{\Psi}$ (of dimension $(N-P)\times L$) are respectively:

$$\mathbf{x} = [a_0, a_1(0), \cdots a_1(P), a_2(0,0), a_2(0,1), \cdots a_2(P,P)]^T, \tag{6.4}$$

$$\mathbf{\Psi} = \begin{pmatrix} 1 & s(P) & \dots & s(0) & s^2(P) & s(P)s(P-1) & \dots & s^2(0) \\ 1 & s(P+1) & \dots & s(1) & s^2(P+1) & s(P+1)s(P) & \dots & s^2(1) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s(N-1) & \dots & s(N-P-1) & s^2(N-1) & s(N-1)s(N-2) & \dots & s^2(N-P) \end{pmatrix}. \tag{6.5}$$

Here, N is the total number of measured output samples, and $L = (P+3)(P+2)/2$ considering the symmetry of the coefficients of $a_2(\cdot,\cdot)$. The least-squares (LS) estimate

of the coefficients ($\hat{\mathbf{x}}$) is given by:

$$\hat{\mathbf{x}} = (\mathbf{\Psi}^T\mathbf{\Psi})^{-1}\mathbf{\Psi}^T\mathbf{y}. \tag{6.6}$$

It needs to be noted that for the event-related fMRI design, the matrix $\mathbf{\Psi}$ may become singular because of the fact that the stimuli function $s(n)$ contains only zeros and ones, which may cause some columns in $\mathbf{\Psi}$ to be zero. Additionally, the Volterra series model needs to specify the order to which the representation is carried out. Inclusion of higher orders in Volterra series allows a better representation of the dynamical system under study. However, since all the related kernel parameters need to be estimated together if using the LS method, inclusion of higher order Volterra series causes the number of parameters to be estimated to increase exponentially with the order. For example, if the BOLD signal with TR = 1 second is modelled using Volterra series of order 3, the number of parameters to be estimated is more than 4000 if the kernels are to capture the temporal properties within 20 seconds after stimuli. Therefore, an efficient and accurate method for the calculation of the Volterra kernels is desired. In the following section, such a method (the RBF neural network method) is provided. Especially, the calculation of Volterra series kernels is derived from the parameters of the RBF neural network. Compared to the Volterra series method estimated by LS, the proposed method could model the systems nonlinearity even when higher order kernels are needed. Also since the kernel parameters are estimated independently, the lower order Volterra kernels could still be accurately calculated even when the higher order Volterra kernels are not estimated.

6.3 Neural Networks Model

Neural networks, a powerful method for modelling nonlinear systems [100], is used in this work to describe the nonlinear dynamics of HDR [101]. The present and the recent past inputs are applied to the neural network and the BOLD signal $y(n)$ from a specific voxel is used as the desired output signal. Thus, the measured BOLD signal can be expressed as the nonlinear function of the input stimulus as follows:

$$\hat{y}(n) = F^{NN}(\mathbf{s}(n)) \tag{6.7}$$

where, $\mathbf{s}(n) = [s(n), s(n-1), \cdots, s(n-P)]^T$ is the input vector of dimension $(P+1)\times 1$ formed by the present and recent past inputs with the maximum delay P. The output $\hat{y}(n)$ is a nonlinear function (denoted by F^{NN}) of the input vector $\mathbf{s}(n)$. This functional mapping is realized by the Radial Basis Function (RBF) network as shown in Fig. 6.1. The universal approximation property and straightforward computation using linearly weighted combination of single layer neurons have made RBF network a good choice in the dynamic reconstruction applications such as the one dealt in this thesis. Moreover, the RBF network has a simpler structure (one hidden layer) and easier to implement compared to the multi-layer perceptrons (MLP) neural network which may require more than one hidden layers and more time for learning the underlying nonlinearity.

Suppose there are M hidden units in the RBF network, then the output $\hat{y}(n)$ of the mapping is taken to be a linear combination of the basis functions, i.e.

$$\hat{y}(n) = \sum_{i=1}^{M} h_i G(\mathbf{s}(n), \mathbf{c}_i) \tag{6.8}$$

where h_i $(i = 1, \cdots, M)$ are the weighting coefficients (linear output layer) and $\mathbf{c}_i$ $(i = 1, \cdots, M)$ are the centers of the radial basis functions. Commonly used basis functions are Gaussian functions which are defined as:

$$G(\mathbf{s}(n), \mathbf{c}_i) = \exp\Big(-\frac{1}{\sigma_i^2} \|\mathbf{s}(n) - \mathbf{c}_i\|^2 \Big) \tag{6.9}$$

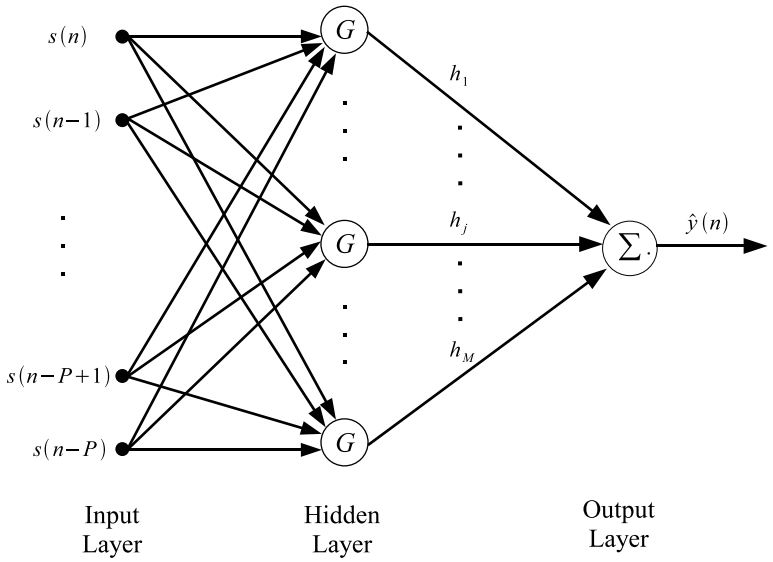


Figure 6.1 The structure of the RBF neural network.

where σ_i^2 is the variance of the ith Gaussian basis function.

The following subsection presents the equivalence of the RBF network and Volterra series methods. This equivalence could also be generalized to the MLP with simple structures (one hidden layer), whereas this equivalence is not so straightforward for structures with more than one hidden layers.

6.3.1 Relation between RBF neural network and Volterra series

According to Eqs. (6.8) and (6.9), $\hat{y}(n)$ can be written as,

$$\begin{aligned}
\hat{y}(n) &= \sum_{i=1}^{M} h_i e^{-\frac{\|\mathbf{s}(n)-\mathbf{c}_i\|^2}{\sigma_i^2}} \\
&= \sum_{i=1}^{M} h_i e^{-\frac{1}{\sigma_i^2}(\mathbf{s}(n)-\mathbf{c}_i)^T(\mathbf{s}(n)-\mathbf{c}_i)} \\
&= \sum_{i=1}^{M} h_i e^{-\frac{1}{\sigma_i^2}(\mathbf{s}^T(n)\mathbf{s}(n)-2\mathbf{s}^T(n)\mathbf{c}_i+\mathbf{c}_i^T\mathbf{c}_i)} \\
&= \sum_{i=1}^{M} \left(h_i e^{-\frac{1}{\sigma_i^2}\mathbf{c}_i^T\mathbf{c}_i} \right) e^{x_i}
\end{aligned} \tag{6.10}$$

Here, x_i is defined as:

$$\begin{aligned}
x_i &= -\frac{1}{\sigma_i^2}(\mathbf{s}^T(n)\mathbf{s}(n) - 2\mathbf{s}^T(n)\mathbf{c}_i) \\
&= -\frac{1}{\sigma_i^2}\left(\sum_{j=0}^{P} s^2(n-j) - 2\sum_{j=0}^{P} c_{ij}s(n-j) \right),
\end{aligned} \tag{6.11}$$

where, $\mathbf{c}_i$ is the i-th center and c_{ij} is the j-th element of $\mathbf{c}_i$.

According to the Taylor series expansion of e^{x_i} at origin:

$$e^{x_i} = \sum_{n=0}^{\infty} \frac{x_i^n}{n!} = 1 + x_i + \frac{1}{2!}x_i^2 + \frac{1}{3!}x_i^3 + \cdots \tag{6.12}$$

and substituting Eq. (6.12) into Eq. (6.10), we get

$$\begin{aligned}
\hat{y}(n) &= (h_1 e^{-\frac{\mathbf{c}_1^T\mathbf{c}_1}{\sigma_1^2}} + h_2 e^{-\frac{\mathbf{c}_2^T\mathbf{c}_2}{\sigma_2^2}} + \cdots + h_M e^{-\frac{\mathbf{c}_M^T\mathbf{c}_M}{\sigma_M^2}}) \\
&\quad +2\sum_{i=0}^{P}(h_1 e^{-\frac{1}{\sigma_1^2}c_1^T c_1}\frac{1}{\sigma_1^2}c_{1i} + h_2 e^{-\frac{1}{\sigma_2^2}c_2^T c_2}\frac{1}{\sigma_2^2}c_{2i} \\
&\quad +\cdots + h_M e^{-\frac{1}{\sigma_M^2}c_M^T c_M}\frac{1}{\sigma_M^2}c_{Mi})s(n-i) \\
&\quad +2\sum_{i=0}^{P}\sum_{j=0}^{P}(h_1 e^{-\frac{1}{\sigma_1^2}c_1^T c_1}\frac{1}{\sigma_1^4}c_{1i}c_{1j} + h_2 e^{-\frac{1}{\sigma_2^2}c_2^T c_2}\frac{1}{\sigma_2^4}c_{2i}c_{2j} + \cdots \\
&\quad +h_M e^{-\frac{1}{\sigma_M^2}c_M^T c_M}\frac{1}{\sigma_M^4}c_{Mi}c_{Mj})s(n-i)s(n-j) \\
&\quad +\sum_{j=0}^{P}(-\frac{1}{\sigma_1^2}h_1 e^{-\frac{1}{\sigma_1^2}c_1^T c_1} - \frac{1}{\sigma_2^2}h_2 e^{-\frac{1}{\sigma_2^2}c_2^T c_2} - \cdots \\
&\quad -\frac{1}{\sigma_M^2}h_M e^{-\frac{1}{\sigma_M^2}c_M^T c_M})s(n-j)^2 \\
&\quad +\cdots\cdots
\end{aligned} \tag{6.13}$$

Comparing the above equations with Eq. (6.1), it can be seen that the zeroth, first, and second-order Volterra kernels can be deduced as:

$$a_0 = h_1 e^{-\frac{\mathbf{c}_1^T\mathbf{c}_1}{\sigma_1^2}} + h_2 e^{-\frac{\mathbf{c}_2^T\mathbf{c}_2}{\sigma_2^2}} + \cdots + h_M e^{-\frac{\mathbf{c}_M^T\mathbf{c}_M}{\sigma_M^2}} \tag{6.14}$$

$$\begin{aligned}
a_1(i) &= 2(h_1 e^{-\frac{1}{\sigma_1^2}c_1^T c_1}\frac{1}{\sigma_1^2}c_{1i} + h_2 e^{-\frac{1}{\sigma_2^2}c_2^T c_2}\frac{1}{\sigma_2^2}c_{2i} \\
&\quad +\cdots + h_M e^{-\frac{1}{\sigma_M^2}c_M^T c_M}\frac{1}{\sigma_M^2}c_{Mi})
\end{aligned} \tag{6.15}$$

$$\begin{aligned}
a_2(i,j) &= 2(h_1 e^{-\frac{1}{\sigma_1^2}c_1^T c_1}\frac{1}{\sigma_1^4}c_{1i}c_{1j} + h_2 e^{-\frac{1}{\sigma_2^2}c_2^T c_2}\frac{1}{\sigma_2^4}c_{2i}c_{2j} \\
&\quad +\cdots + h_M e^{-\frac{1}{\sigma_M^2}c_M^T c_M}\frac{1}{\sigma_M^4}c_{Mi}c_{Mj}) \\
&\quad -(\frac{1}{\sigma_1^2}h_1 e^{-\frac{1}{\sigma_1^2}c_1^T c_1} + \frac{1}{\sigma_2^2}h_2 e^{-\frac{1}{\sigma_2^2}c_2^T c_2} + \cdots \\
&\quad +\frac{1}{\sigma_M^2}h_M e^{-\frac{1}{\sigma_M^2}c_M^T c_M})\delta(i-j)
\end{aligned} \tag{6.16}$$

$$a_2(j,i) = a_2(i,j), \forall i,j. \tag{6.17}$$

The above equations build the link between the Volterra series and RBF neural network. This implies that the Volterra kernels can be easily deduced from the parameters of the RBF neural network. It is also easy to extend the neural network method to deduce the coefficients of the third or higher order Volterra kernels if more terms from the Taylor series are incorporated into the expansion series.

As described in the previous section, the performance of the Volterra series method is determined by the choice of the order of Volterra series. Considering only the lower order Volterra kernels may not capture the dynamical properties of the system well. Although inclusion of higher order Volterra kernels could model the dynamic system better, this poses the problem of higher computational complexity in identifying the higher order kernels. Compared to the Volterra series method, the neural network method could model the nonlinear dynamical system well even when the system is highly nonlinear. Besides, to estimate the Volterra kernels using the LS method, all the relevant Volterra kernels need to be estimated together and this is difficult when the order of the Volterra kernels included is high. However, using the RBF neural network method, the lower order Volterra kernels can be estimated from the RBF parameters (independently) without the estimation of the higher order kernels. This is useful because for most cases, we are only interested in the lower order Volterra kernels. The proposed neural network method provides an efficient approach to estimate the lower order Volterra kernels and at the same time capture the nonlinearities of the system. In addition, the RBF neural network method avoids the possible singularity problem ($\mathbf{\Psi}$ in Eq. (6.6)) of the LS method.

6.3.2 Learning procedure

The centers $\mathbf{c}_i$ of the radial basis functions are chosen randomly from the training data set and the variances of the basis functions are fixed according to the spread of the centers. For fixed centers $\mathbf{c}_i$ and variance σ_i^2, the aim of the RBF network is to find the weights h_i such that the sum-squared-error is minimized. Considering the noise presented in the data, regularization is required to stabilize the solution [100]. The regularized RBF network gives the following estimates of the weight vector, $\mathbf{h} = [h_1, \cdots, h_M]^T$:

$$\mathbf{h} = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^T\mathbf{y} \tag{6.18}$$

where λ is the regularization parameter, $\mathbf{y}$ is the vector representing measured fMRI signal and $\mathbf{G}$ is an $(N-P)\times M$ interpolation matrix, which is defined below:

$$\mathbf{G} = \begin{pmatrix} G(\mathbf{s}(P), \mathbf{c}_1) & \dots & G(\mathbf{s}(P), \mathbf{c}_M) \\ \vdots & \ddots & \vdots \\ G(\mathbf{s}(N-1), \mathbf{c}_1) & \dots & G(\mathbf{s}(N-1), \mathbf{c}_M) \end{pmatrix} \tag{6.19}$$

The regularization parameter $\lambda > 0$ controls the balance between fitting the data and regularization. A small value of λ means the data can be fit tightly without causing a large regularization; a large value of λ means a tight fit has to be sacrificed to a smoothing output function.

Here, $\mathbf{h}$ and λ can be estimated through Bayesian learning by iteratively updating the following equations [102]:

$$\mathbf{\Sigma} = \beta^2(\mathbf{G}^T\mathbf{G} + \lambda\mathbf{I})^{-1} \tag{6.20}$$

$$\hat{\mathbf{h}} = \frac{1}{\beta^2}\mathbf{\Sigma}\mathbf{G}^T\mathbf{y} \tag{6.21}$$

$$\gamma = M - \frac{\lambda}{\beta^2}\text{trace}(\mathbf{\Sigma}) \tag{6.22}$$

$$\beta^2 = \frac{\|\mathbf{y} - \mathbf{G}\hat{\mathbf{w}}\|^2}{N - \gamma} \tag{6.23}$$

$$\lambda = \frac{\gamma\beta^2}{\hat{\mathbf{w}}^T\hat{\mathbf{w}}} \tag{6.24}$$

where, β^2 is the estimated variance of the noise.

The reconstructed BOLD signal $\mathbf{y}_B$ is then calculated as:

$$\hat{\mathbf{y}}_b = \mathbf{G}\hat{\mathbf{h}}. \tag{6.25}$$

One advantage of Bayesian learning of RBF neural network is that the cross-validation to find the suitable regularization parameter is not needed, which means that all the training data can be used. This is especially useful in the fMRI time series analysis since the number of the available data points is limited.

6.4 Balloon Model

In Chapter 1, Section 1.2.2, we have briefly introduced the Balloon model. It is a physiologically inspired model introduced by Buxton et al. to describe the dynamics of the BOLD signal [26]. In this section, a more detailed description of the Balloon model is provided. It is used to generate the simulated BOLD signal in the simulation studies of this chapter.

The Balloon model is a state-space model which describes the dynamics among blood flow, blood volume and blood oxygen concentration changes. The model is inherently nonlinear and can be used to explain the nonlinearities appearing in the BOLD

signals. The Balloon model is described by the following equations:

$$\begin{cases} \dot{f} &= u \\ \dot{u} &= \epsilon s - k_s u - k_f(f-1) \\ \dot{v} &= \frac{1}{\tau}(f - v^{1/\alpha}) \\ \dot{q} &= \frac{1}{\tau}(f\frac{1-(1-E_0)^{1/f}}{E_0} - v^{1/\alpha-1}q) \\ y_b &= V_0(7E_0(1-q) + 2(1-q/v) \\ & \quad +(2E_0 - 0.2)(1-v)) \end{cases} \tag{6.26}$$

where f is the CBF, v is the CBV, q is the deoxyhemoglobin content (dHb) of vein, u is the flow inducing signal (these values are normalized to their values at rest), s is the stimulus function and y_b is the BOLD signal. The time-varying intrinsic variables f, v, q and u summarize the hemodynamics of the system: how the changes of the CBF, CBV and dHb are coupled to each other and synaptic activity encoded in the stimulus function s. The other parameters in the above equations are the neural efficiency ϵ, the flow decay k_s, the flow time constant k_f, Grubb's parameter α, the venous transit time τ, the resting net oxygen extraction fraction E_0 and the resting blood volume fraction V_0.

Putting Eq. (6.26) in a succinct way, it can be written as:

$$\begin{cases} \dot{\mathbf{x}}(t) &= A(\mathbf{x}(t), s(t)) \\ y_b(t) &= B(\mathbf{x}(t)) \end{cases} \tag{6.27}$$

where A and B are nonlinear functions, $s(t)$ is the input signal representing the stimulus function, $\mathbf{x}(t) = [f(t), u(t), v(t), q(t)]^T$ is the state vector which is non-measurable and $y_b(t)$ is the output BOLD signal. The measured fMRI signal $y(t)$ is represented as:

$$y(t) = y_b(t) + e(t) \tag{6.28}$$

where $e(t)$ is the drift and measurement noise. The schematic diagram of the Balloon model is shown in Fig. 6.2.

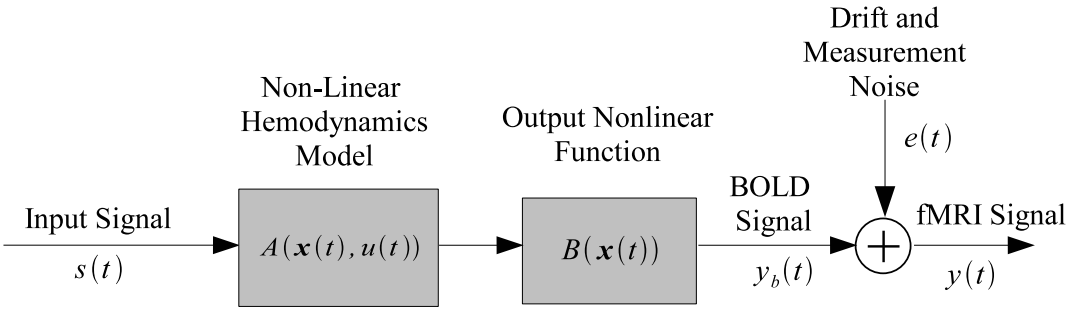


Figure 6.2 Schematic diagram for Balloon model.

6.5 Results and Discussion

6.5.1 Simulated Data

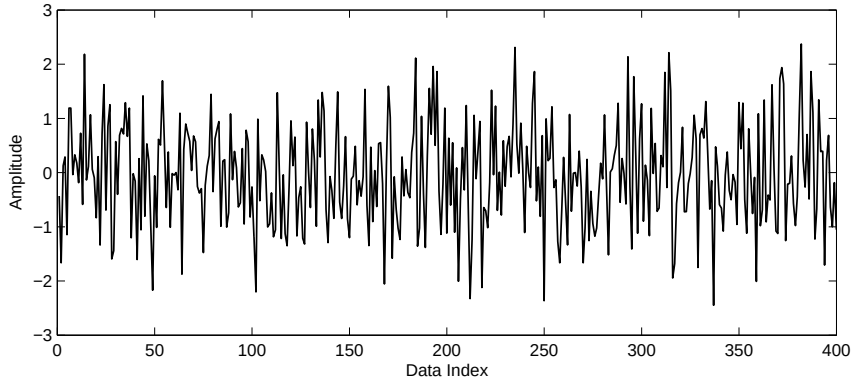
In this section, the proposed RBF neural network method was first tested on a simple example to validate the estimation accuracy of the Volterra kernels. Then, the simulated BOLD signal (using the Balloon model) with different noise levels and real event-related fMRI data were investigated to reconstruct the dynamics underlying the fMRI signal.

Example

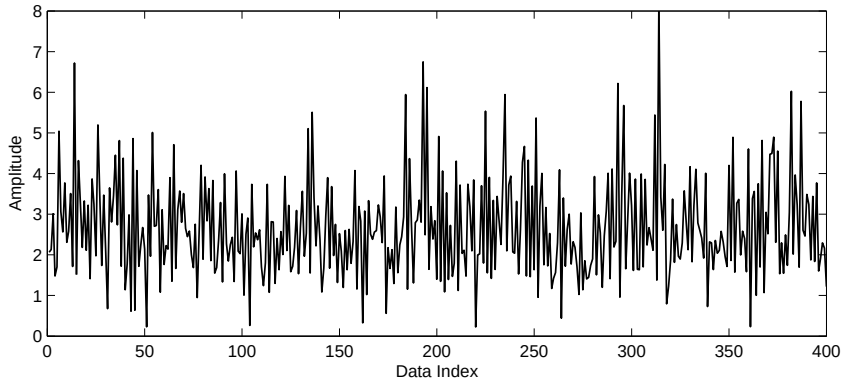
In this general example, the RBF neural network was applied to the simulated data to identify the coefficients of the Volterra kernels. A total of 400 data points were generated using the following Volterra series model ($P = 2$):

$$\begin{aligned} y(n) &= 2.4 + 0.9s(n) - 0.4s(n-1) + 0.74s(n-2) \\ &\quad -0.18s(n-1)s(n-2) + 0.36s(n)^2. \end{aligned} \tag{6.29}$$

The input signal $s(n)$ in this example is simulated by Gaussian white noise with unit variance. Figure 6.3 shows one realization of the input signal $s(n)$ (Fig. 6.3 (a)) and the simulated output signal $y(n)$ (Fig. 6.3 (b)) using Eq. (6.29).



(a)



(b)

Figure 6.3 One realization of the input signal and the simulated output signal using Eq. (6.29). (a) Input signal; (b) Output signal.

The RBF neural network with different number of hidden units ($M = 50, 100, 200, 300$) (respectively represented as RBF50, RBF100, RBF200, RBF300 in Table 6.1) are investigated. The Bayesian learning procedure described in Section 6.3.2 is applied to regress the output signal $y(n)$ on the input vector $\mathbf{s}(n)$ (From Eq. (6.29), the maximum delay P in the input vector $\mathbf{s}(n)$ is accordingly set to be 2). The M centers of the RBF network basis functions are randomly chosen from the training data set. The estimated Volterra kernels (using Eq. (6.14) – Eq. (6.17)) are shown in Table 6.1. Due to the symmetry of the second-order coefficients $a_2(i, j) = a_2(j, i)$, we only show the estimated coefficients where $i \leq j$. The goodness-of-fit of the RBF neural network with different number of hidden units is evaluated using the NMSE defined in Eq. (5.25) and rewritten here:

$$NMSE = \frac{\sum_{i=0}^{L-1}(\hat{x}_i - x_i)^2}{\sum_{i=0}^{L-1} x_i^2} \tag{6.30}$$

where $\hat{x}_i$ and x_i are respectively the estimated value and the true value of the Volterra kernel parameters.

From Table 6.1, it is clear that although the RBF with small number of hidden units (RBF50) shows some estimation errors (with NMSE = 0.004 which is relatively large), the estimation results are accurate when the number of hidden units is large (RBF100, RBF200 and RBF300 with NMSE equal to 1.593×10^{-4}, 8.728×10^{-6}, and 1.067×10^{-5} respectively which are relatively small). This study shows that the zeroth, first, and second-order Volterra kernels could be accurately estimated by the proposed method.

Example

To compare the proposed neural network method and Volterra series method using LS estimation, a nonlinear signal with third-order Volterra kernels is generated according to

Table 6.1 Estimation of Volterra kernel parameters ($P = 2$)

Parameters	**Value**	**RBF50**	**RBF100**	**RBF200**	**RBF300**
a_0	2.40	2.4475	2.4062	2.4030	2.4014
$a_1(0)$	0.90	0.8842	0.9207	0.9038	0.9055
$a_1(1)$	-0.40	-0.3511	-0.4169	-0.4020	-0.4017
$a_1(2)$	0.74	0.7157	0.7520	0.7422	0.7423
$a_2(0,0)$	0.36	0.2341	0.3628	0.3592	0.3569
$a_2(0,1)$	0.00	0.0191	-0.0069	-0.0004	-0.0003
$a_2(0,2)$	0.00	-0.0076	0.0046	-0.0033	-0.0029
$a_2(1,1)$	0.00	-0.0619	-0.0092	-0.0015	-0.0010
$a_2(1,2)$	-0.09	-0.0890	-0.0901	-0.0901	-0.0881
$a_2(2,2)$	0.00	-0.0669	0.0111	-0.0043	-0.0040
NMSE		0.0040	1.593e-4	8.728e-6	1.067e-5

the following equation:

$$\begin{aligned} y(n) &= 2.4 + 0.9s(n) - 0.4s(n-1) + 0.74s(n-2) - 0.18s(n-1)s(n-2) \\ &\quad +0.36s(n)^2 + 0.76s(n)s(n-1)s(n-2) \\ &\quad +0.85s(n-1)s(n-2)^2 + e(n). \end{aligned} \tag{6.31}$$

As before, the input signal $s(n)$ is Gaussian white noise with variance one and the total number of data points is 400. In this example, a small Gaussian noise $e(n)$ with variance 0.01 is also added to generate the noisy simulated signal $y(n)$. The maximum delay P in the input vector $\mathbf{s}(n)$ is set to be 2. For the RBF neural network, 200 hidden units ($M = 200$) are used to regress the simulated output signal $y(n)$ to the input signal $s(n)$ in order to have good results. For the LS estimation method, the choice of the highest Volterra kernel order need to be specified. The estimations with the highest assumed order being set to 2 (LS-2) and 3 (LS-3) are examined. Table 6.2 shows the estimation results of RBF, LS-2 and LS-3.

From Table 6.2, it is clear that the least-squares estimation is sensitive to the assumption of the highest order of Volterra kernels. If the chosen order mismatches with the actual order, the estimation results may deviate from the true values as shown in the LS-2 with large NMSE (0.1309). Including higher order Volterra kernels into the LS method may result in better estimation (as shown in LS-3 with small NMSE (9.8310e-007)), however, this is difficult for fMRI data analysis since the number of parameters to be estimated from LS method may become enormously high with the increasing order of the kernels included. The RBF neural network method, on the other hand, does not rely on the choice of the order and hence could accurately estimate the Volterra kernels even if the order of the Volterra series is unknown (with relatively small NMSE = 1.0573e-004). Another advantage of the RBF neural network method is that we can estimate Volterra

Table 6.2 Estimation of Volterra kernel parameters using RBF neural network method and least-squares (LS) method when the highest order of Volterra series is 3.

Kernels	**Parameters**	**Value**	**RBF**	**LS-2**	**LS-3**
zeroth	a_0	2.40	2.4064	2.1953	2.4006
first	$a_1(0)$	0.90	0.9018	0.8785	0.9013
	$a_1(1)$	-0.40	-0.3888	0.4764	-0.4013
	$a_1(2)$	0.74	0.7480	0.9776	0.7388
second	$a_2(0,0)$	0.36	0.3681	0.4146	0.3596
	$a_2(0,1)$	0.00	0.0023	0.0252	0.0012
	$a_2(0,2)$	0.00	0.0158	0.2670	0.0005
	$a_2(1,1)$	0.00	-0.0008	0.0177	0.0000
	$a_2(1,2)$	-0.09	-0.0840	-0.1943	-0.0895
	$a_2(2,2)$	0.00	0.0139	0.1340	-0.0001
	NMSE		1.0573e-4	0.1309	9.8310e-7

kernels at a specified order while not estimating Volterra kernels at other orders. Compared to the LS method which needs to estimate all the Volterra kernels together, this property of RBF neural network could estimate the lower order Volterra kernels independently without estimating the higher order kernels. This is efficient because the lower order Volterra kernels are of our interest for most cases.

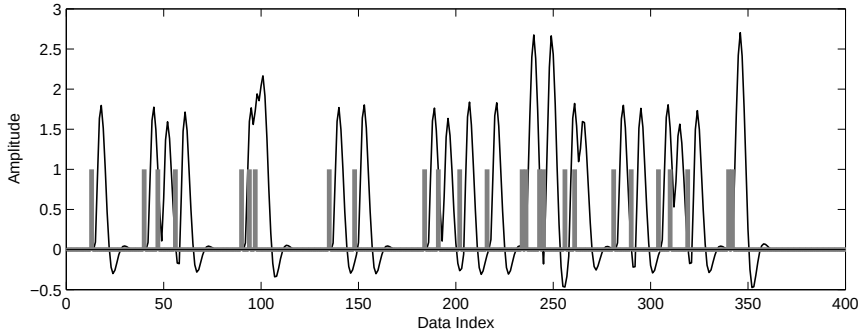
From the simulation studies on Example1 and Example2, it is shown that the proposed RBF neural network method works well in the general Volterra series models. The estimated Volterra kernels are accurate as long as enough hidden units are used in the network structure. In the following section, a case closer to the fMRI data – the simulated BOLD signal is tested.

Simulated BOLD Signal

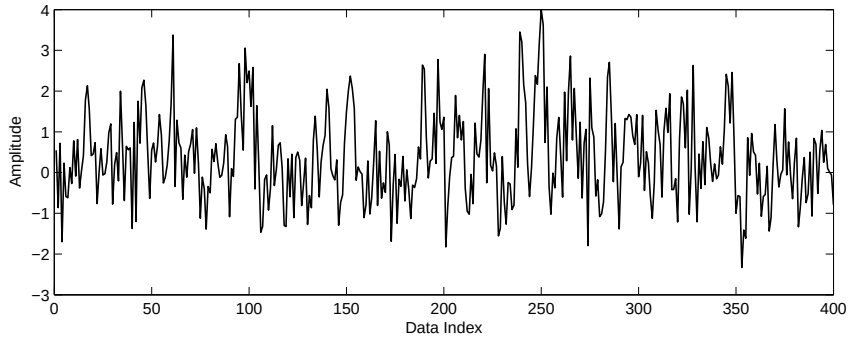
In this set of simulations, the simulated BOLD signal generated using the Balloon model is investigated using the proposed RBF neural network method. When the RBF neural network is applied to the fMRI data, the noise involved in the fMRI signal may pose a problem. In this section, both the Gaussian white noise and autocorrelation noise in the fMRI data are examined.

The simulated BOLD signal y_b is generated using Eq. (6.26) with the following set of parameters: $\epsilon = 0.5, k_s = 0.65, k_f = 0.4, \tau = 1, \alpha = 0.4, E_0 = 0.4$ and $V_0 = 0.02$ [26][20] and the total duration of the simulated BOLD signal is 400s. To simulate the event-related fMRI experiment, the input stimuli s are randomly generated with each lasting for 1s. The generated BOLD signal is then sampled with uniform sampling rate of 1s, which gives the simulated BOLD signal $y_b(n)$. Figure 6.4(a) illustrates the simulated BOLD signal $y_b(n)$ with the thick vertical line indicating the timing of the discrete random stimuli $s(n)$.

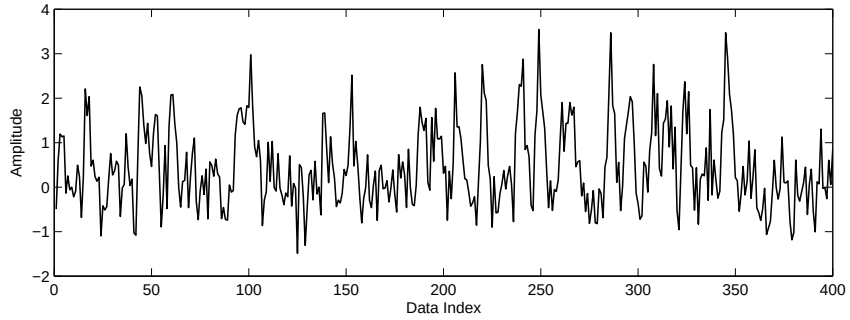
To test the effectiveness of the proposed method in the case of additive noise, the



(a)



(b)



(c)

Figure 6.4 Simulated BOLD signal generated by the Balloon model and noisy BOLD signals with different additive noise. (a) Simulated pure BOLD signal and the timing of the stimuli; (b) Simulated noisy BOLD signal corrupted with additive Gaussian white noise; (c) Simulated noisy BOLD signal corrupted with additive autocorrelation noise.

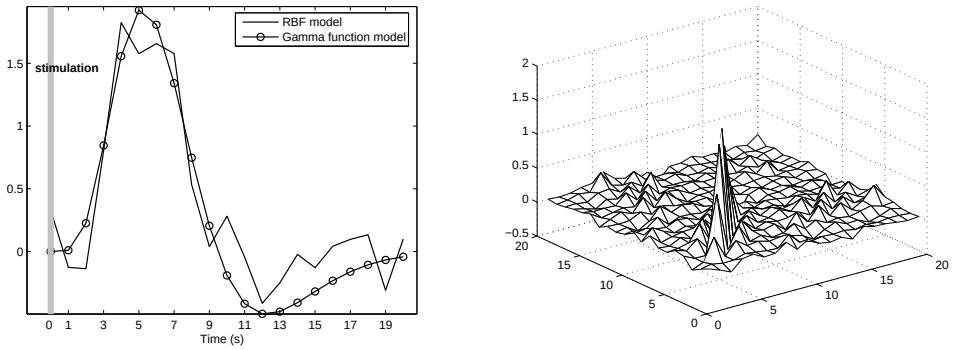
noise $e(n)$ was added to the BOLD signal $y_b(n)$ generated by the Balloon model as follows:

$$y(n) = y_b(n) + e(n) \tag{6.32}$$

where $e(n)$ is the additive noise (Gaussian white noise or noise with temporal autocorrelation) and $y(n)$ is the noisy BOLD signal. Then, the simulated input stimuli $s(n)$ were fed to the RBF network with the simulated noisy BOLD signal $y(n)$ as the target signal. The number of hidden units M in the RBF neural network is set to be 200 (enough to model the nonlinearity in the system) and the centers of the RBF are selected randomly from the input stimuli vectors. The parameters of the RBF neural network are estimated through the Bayesian learning. To cover the time span of the HDR, the maximum input delay P is chosen to be 20 since the HDR lasts for about 20s and the sampling rate is 1s in this simulation.

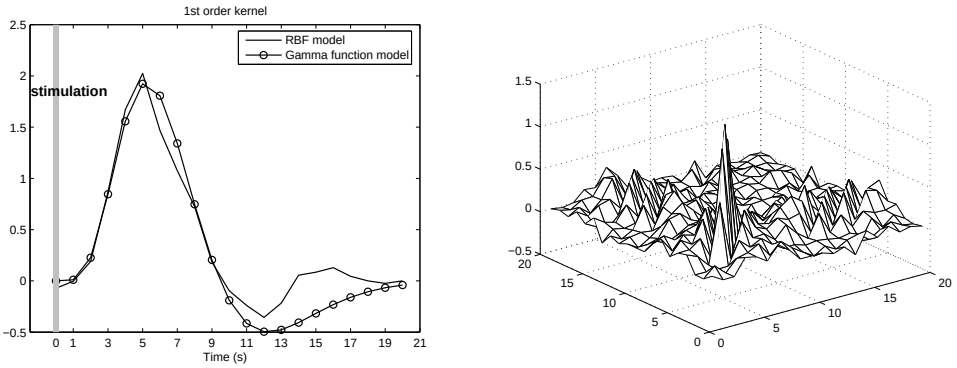
Figure 6.4(b) shows the simulated noisy BOLD signal $y(n)$ with additive Gaussian white noise. Different noise levels (with the signal-to-noise ratio (SNR) varying from -7dB to 5dB) are added to the simulated BOLD signal $y_b(n)$ to generate the noisy BOLD signal $y(n)$. Figures 6.5 and 6.6 show the the estimated 1st (a vector of dimensions 21×1) and 2nd order (a symmetric matrix of dimensions 21×21) Volterra kernels of the simulated noisy BOLD signal with the SNR of -7dB and 0dB respectively. It is clear from the figures that the estimated 1st order kernel of Volterra series shows the properties of the HDR and is analogous to the conventional HRF formed by the difference of two Gamma functions. When the noise level is high, the estimated HDR exhibits more variations as displayed in Fig. 6.5 (a). The 2nd order kernels in these figures indicate the effect of the interaction between adjacent inputs on output signal.

Next, the noise with temporal autocorrelation is investigated. The serial correlations



(a) (b)

Figure 6.5 Estimated 1^{st} (a) and 2^{nd} order (b) Volterra kernels using the proposed neural network method with SNR = -7dB.



(a) (b)

Figure 6.6 Estimated 1^{st} (a) and 2^{nd} order (b) Volterra kernels using the proposed neural network method with SNR = 0dB.

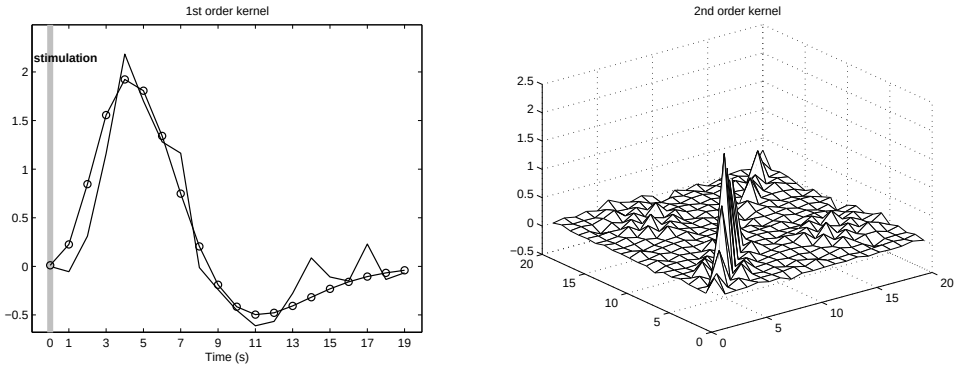
in fMRI data is simulated as autoregressive with order 1 (AR(1)) plus white noise [103]:

$$e(n) = z(n) + \eta(n) \tag{6.33}$$

$$z(n) = \rho z(n-1) + \xi(n) \tag{6.34}$$

where ρ is the AR(1) coefficient, $\eta(n)$ and $\xi(n)$ are the white noise terms ($\eta(n) \sim N(0, \sigma_\eta^2)$, $\xi(n) \sim N(0, \sigma_\xi^2)$). This noise model is to capture the short-range autocorrelations in the fMRI data. For long-range autocorrelations in the fMRI data, a detrend procedure is often included as a preprocessing step to relax the autocorrelation noise [104]. In the following part, the short-range autocorrelation noise is investigated.

The short-range autocorrelation noise synthesized using parameters $\rho = 0.4, \sigma_\eta^2 = 0.5, \sigma_\xi^2 = 0.6$ is added to the pure BOLD signal $y_b(n)$ to generate the noisy BOLD signal $y(n)$ as shown in Fig. 6.4(c). This simulated noisy BOLD signal is then modelled using the RBF neural network. The parameters of the RBF neural network are chosen according to the same scheme as introduced in the case of the Gaussian white noise. Figure 6.7 shows the estimated 1st and 2nd order Volterra kernels using the proposed RBF neural network method. These results indicate that the RBF neural network method performs well to estimate the Volterra kernels when the noise is of short-range temporal autocorrelation. This shows that the short-range autocorrelation noise is not so problematic for the proposed RBF neural network method. When the noise is of long-range autocorrelation, the performance of the RBF neural network would be affected by the long-range autocorrelation and may not work well. Hence, to ensure that the RBF neural network method works well, the fMRI data need to be detrended before applying the RBF neural network for further processing.



(a) (b)

Figure 6.7 Estimated 1^{st} (a) and 2^{nd} order (b) Volterra kernels using the proposed neural network method when the additive noise is autocorrelational.

6.5.2 Experimental Data

The proposed RBF neural network method for the estimation of the hemodynamic response is also tested on an experimental event-related fMRI data (*DATA-EVENT* in Chapter 1). The raw data are preprocessed by the SPM software for registration, normalization and smoothing. The drift (long-range autocorrelation noise) in each voxel time series is removed using the detrending procedure to ensure the proposed neural network method works well.

The preprocessed fMRI signal is then applied to the RBF neural network with the number of hidden units $M = 200$. The measured fMRI signal at each voxel is used as the desired signal $y(n)$ of the RBF neural network and the input signal $s(n)$ is constructed according to the description of the experimental design. When the stimuli are presented, $s(n) = 1$; while when the stimuli are absent, $s(n) = 0$. Since the sampling rate (TR) is 2 seconds, the maximum delay P of the input vector is chosen to be 10 in order to cover the time span (18s) of the HDR.

Estimation of HDR

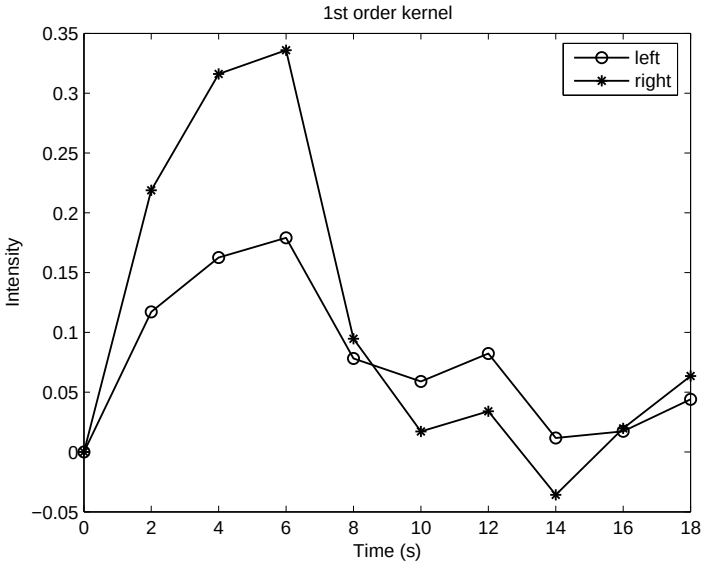
As mentioned in the Introduction section in this chapter, the HDR varies with different brain regions and different subjects. In this section we examine these variations. Figure 6.8 shows the estimated first order Volterra kernels of the left and right auditory cortex for two subjects using the proposed neural network method. Clearly, the variations of the HDR between different brain regions and different subjects can be captured using the proposed method. The estimated first order Volterra kernels show the properties of the HDR, such as a characteristic peak at around 6s. The estimated 2nd order kernels showed the same patterns as that is shown in Fig. 6.5(b) and hence not shown here. These figures illustrate that the proposed neural network method is able to provide a good estimation of the HDR. In addition, it is more flexible than the conventional HDR model formulated by the difference between two gamma functions. The gamma function model may miss the difference in the dynamics of different regions and subjects since it uses the same HDR model for all the voxels investigated. However, the proposed neural network method is applied on each voxel and is able to capture the differences in BOLD signals of different regions and subjects.

Detection of the Activated Regions

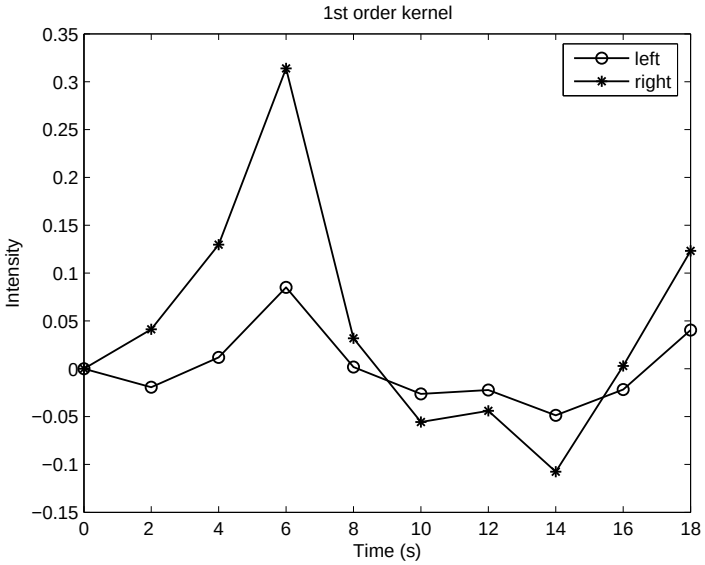
To detect the activated regions of the brain, the reconstructed BOLD signal $\hat{\mathbf{y}}$ at the output of the RBF neural network is investigated. The following test at each voxel is used as an index (R) for activation detection:

$$R = \frac{\|\hat{\mathbf{y}}\|}{\|\mathbf{y} - \hat{\mathbf{y}}\|} \tag{6.35}$$

where $\mathbf{y}$ is the measured fMRI signal in a voxel and $\hat{\mathbf{y}}$ is the reconstructed (or regressed) BOLD signal in this voxel. For the inactivated voxels, the reconstructed BOLD signal



(a)



(b)

Figure 6.8 Estimated 1st order Volterra kernels of the left and right auditory cortex for two subjects. (a) Subject 1; (b) Subject 2.

would be almost zero since the inactivated voxels do not involve the BOLD effect corresponding to the external stimuli. In this case, the value of R would be small. On the other hand, when the voxel is activated, the reconstructed BOLD signal should capture the dynamics corresponding to the input stimulations, that is, $\hat{\mathbf{y}} \neq 0$ and the value R would be large. The R value is analogous to the SNR. Figure 6.9 shows the results of the activated regions in the brain by thresholding the R values at each voxel ($R > 0.3$ is chosen to give best results). It is clear from this figure that the auditory cortex in the brain is activated in this event-related fMRI experiment.

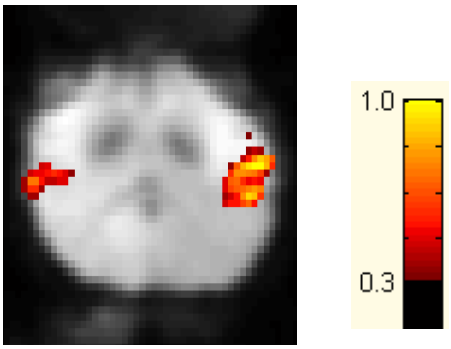


Figure 6.9 One slice showing the activation of the auditory cortex ($R > 0.3$).

6.6 Conclusion

The BOLD signal, as the foundation of the fMRI experiment, reflects the hemodynamic response of the human brain. To investigate how the brain responds to the stimulus, i.e., to study the dynamics of the human brain, system identification methods have been proposed to identify the complex functional relation between the input stimuli and the measured BOLD signal. Conventional methods are based on the linear system analysis which models the BOLD signal as the convolution of the HDR and the input stimuli.

However, this linear assumption holds only for ISI larger than 4-6 seconds. For rapid event-related fMRI experiment, nonlinear properties of the BOLD signal are often observed. This requires nonlinear methods to identify this complex (brain) system. One of these nonlinear models is the Balloon model which is a physiologically derived one. In this chapter, the Volterra series, which can represent any dynamical input-state-output system, was first investigated to provide a nonparametric framework for the system identification. Then, the RBF neural network is proposed as a general method to regress the measured BOLD signal on the input stimuli and to capture the system dynamics. The equivalence of the proposed RBF neural network method to the Volterra kernels has been derived. It is demonstrated that the 1st and 2nd order Volterra kernels can be easily deduced from the coefficients of the RBF neural network. Compared to the Volterra series model estimated through least-squares method, the RBF neural network method does not need to presume the highest order of the nonlinear system and hence is robust and efficient to estimate the Volterra kernels even when the nonlinear system is of higher order. Results from simulated as well as experimental event-related fMRI signals show that the proposed method can successfully estimate the HDR as well as capture the nonlinear dynamics of the BOLD signal. In addition, the proposed method could estimate the individual hemodynamic response (or the 1st order Volterra kernel) at each voxel. This helps us to investigate the variations of the HDR with different brain regions and different subjects.

In the next chapter, we continue our discussion of the neural network based method to analyze the fMRI data. Specifically, a recurrent neural network is proposed to reconstruct the BOLD signal from the measured noisy fMRI data.

Chapter 7

NARX Neural Networks for Dynamical Modelling of fMRI Data

7.1 Introduction

In Chapter 6, a detailed discussion of the methods to model and analyze the nonlinear dynamics of the BOLD responses have been provided. These methods can be categorized into state-space models and regression models. The state-space models account for the data by using intermediate state variables whose dynamics generate the BOLD signals. The commonly used state-space model is the Balloon model [26]. Regression methods model the observed BOLD signals to be some functions of the stimulation patterns. These include multivariate autoregressive (AR) models [105], autoregressive model with exogenous inputs (ARX) [106], specific basis functions implemented in the General Linear Model (GLM), Volterra kernels [20] and the RBF neural network method introduced in Chapter 6.

In this chapter, a recurrent neural network with one or more feedback loops is investigated. This neural network has been successfully used for system modelling and

identification because of their ability in approximating the nonlinearities. The nonlinear autoregressive with exogenous inputs (NARX) neural network [107] is applied to model and analyze event-related fMRI data. This neural network can capture the underlying nonlinearities and hence reconstruct the dynamics of the fMRI data. The reconstructed BOLD signal is less noisy and could facilitate the analysis.

7.2 NARX Model

The dynamical modelling of the fMRI data aims to capture the input-output mapping of the measured noisy fMRI data. Recurrent neural networks are particularly suitable for such applications due to their ability to identify arbitrary nonlinear dynamical systems. In this chapter, the NARX neural networks [100] are applied to model the input-output dynamics of the BOLD signal and hence identify this complex system. The NARX neural networks are powerful methods to model nonlinear systems with the advantage of faster convergence and better generalization ability [108].

Figure 7.1 shows the schematic diagram of the NARX model for reconstructing the fMRI data. In this model, the input signal $s(n)$ is applied to the network through $(q-1)$ delays, $\mathbf{s}(n) = [s(n), \cdots, s(n-q+1)]^T$. The measured fMRI signal $y(n)$ at each voxel is also fed back to the input via p delays, $\mathbf{y}(n) = [y(n-1), \cdots, y(n-p)]^T$. The output of the NARX model can be expressed as:

$$\hat{y}_b(n) = F^{NN}(\mathbf{y}(n), \mathbf{s}(n)) = F^{NN}(\mathbf{x}(n)) \tag{7.1}$$

where $\mathbf{x}(n) = [\mathbf{y}(n), \mathbf{s}(n)]^T$ is the full input vector of size $(p+q) \times 1$. The estimated output $\hat{y}_b(n)$ is a nonlinear transformation (denoted by F^{NN}) of the input vector $\mathbf{x}(n)$. This output $\hat{y}_b(n)$ is the estimated BOLD signal reconstructed by the NARX model from the measured noisy fMRI signal $y(n)$.

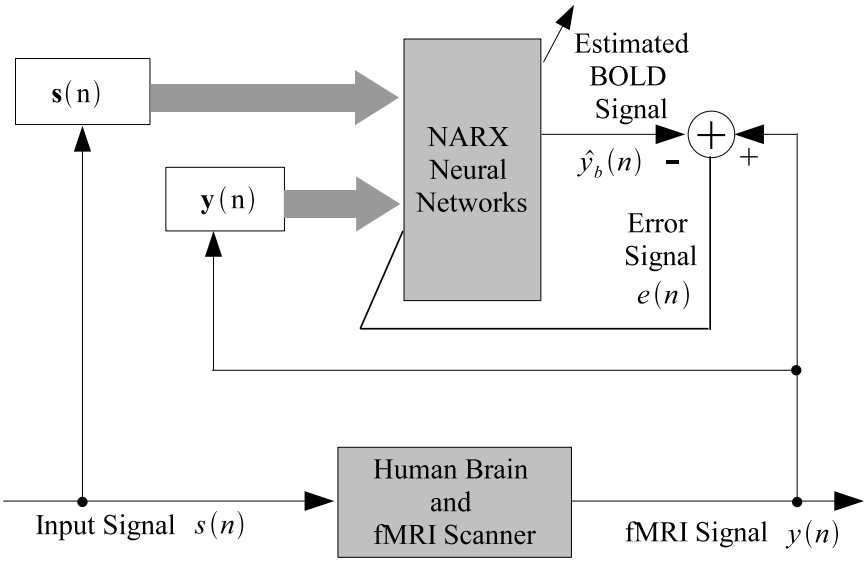


Figure 7.1 Schematic diagram for NARX model.

The proposed NARX model is realized through the RBF neural networks [100]. The learning of the RBF neural networks is detailed in Section 6.3.2. When the weights of the RBF neural networks have been found, a simulated single impulse could be fed into the trained neural networks to estimate the HDR which is usually used in the convolution model as the system impulse function. This estimated HDR could be compared with the conventional HDR formed by the difference of two gamma functions to validate the estimation ability of the NARX model.

7.3 Results and Discussion

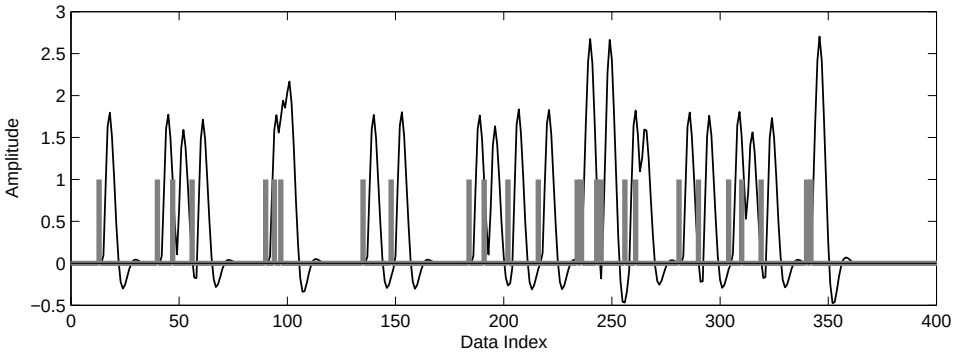
7.3.1 Simulated Data

The simulated BOLD signal is generated using the Balloon model (Eq. (6.26)) with the following set of parameters: $\epsilon = 0.5, k_s = 0.65, k_f = 0.4, \tau = 1, \alpha = 0.4, E_0 = 0.4$

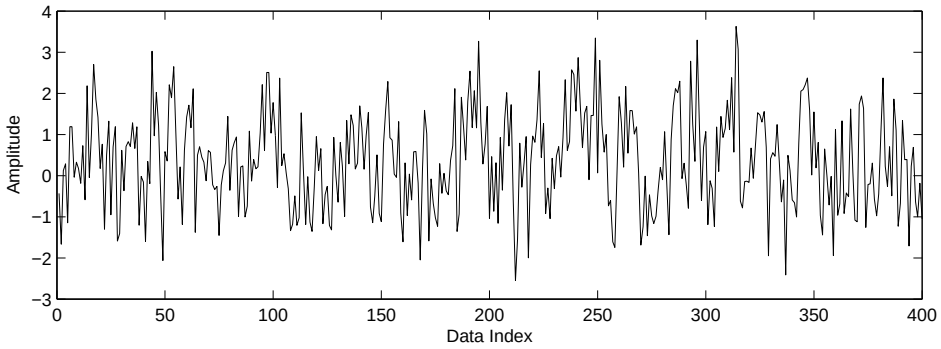
and $V_0 = 0.02$ [26][20]. The input stimulus signal $s(t)$ is generated as random impulses lasting for 1 second which represents an event-related fMRI experiment. When the stimulation is ON, $s(t) = 1$ and when it is OFF, $s(t) = 0$. The total duration of the simulated BOLD signal is 400s with uniform sampling period of 1s. Figure 7.2(a) illustrates the simulated BOLD signal with the thick vertical line indicating the timing of the discrete random stimuli. The white Gaussian noise is added to this clean BOLD signal resulting in a noisy BOLD signal with SNR = $-2.8dB$ which is shown in Fig. 7.2(b). Figure 7.2(c) shows the reconstructed BOLD signal from the NARX neural networks. The delay values p and q in the NARX model (exact interpolation with $N = M = 400$ for the interpolation matrix) are chosen to be 10 and 20, respectively. The value of the input delay q was chosen to cover the time span of the HDR. In this example, the sampling rate is 1s and the HDR lasts for about 20s, hence, the value of q is chosen to be 20.

It is clear from Fig. 7.2 that the reconstructed BOLD signal (Fig. 7.2(c)) accurately captures the properties of the pure BOLD response (Fig. 7.2(a)). This implies that the network has successfully identified and modelled the dynamical behavior of BOLD signal. Then a single simulated impulse is fed into the network, producing the impulse response at the output of the network. The impulse response is the estimated HDR to a single stimulation. Figure 7.3 shows the results of the estimated HDRs using the trained NARX neural networks and averaged over 20 runs. It resembles very well to the HDR of the Balloon model (mean square error, MSE = 0.004). We also observed that the choice of the value of q is related to the accuracy of the HDR estimation. If $q = 10$ is chosen, the HDR estimated by the NARX model does not fit very well (after 10s) to the HDR simulated using the Balloon model.

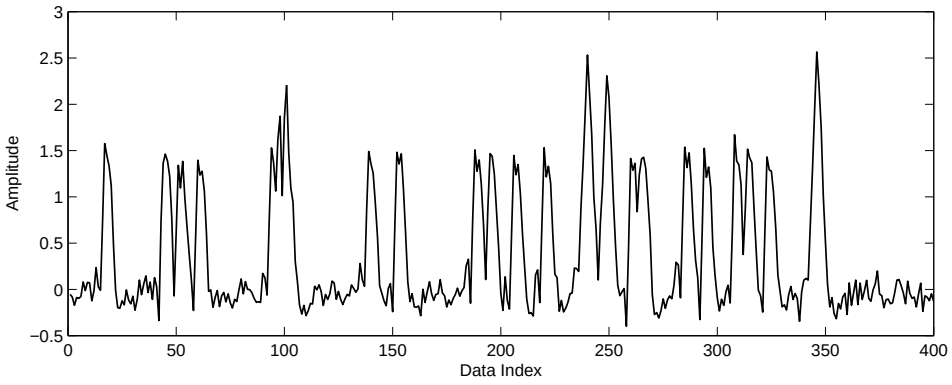
To validate the modelling ability of the NARX model, the reconstructed BOLD signal from the NARX model output is compared with the noisy BOLD signal using the



(a) Simulated pure BOLD signal and the timing of the stimuli.



(b) Simulated noisy BOLD signal.



(c) Reconstructed BOLD signal from the noisy BOLD signal.

Figure 7.2 Simulated BOLD signal and its reconstruction from the NARX neural network.

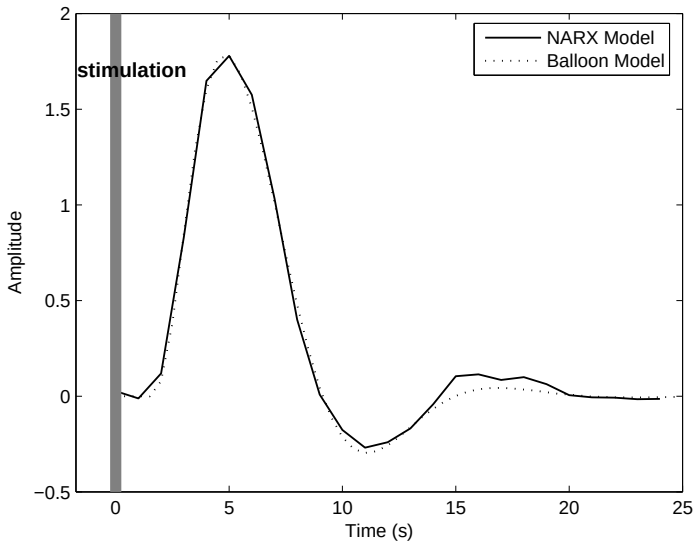


Figure 7.3 The estimated HDR of the simulated data.

Ljung and Box randomness test [109]:

$$Q_{LB} = N(N+2)\sum_{j=1}^{h}\frac{\rho^2(j)}{n-j} \tag{7.2}$$

where N is the sample size, $\rho(j)$ is the autocorrelations of the residuals at lag j, h is the maximum lag used to calculate the test statistics. For the level of statistical significance α, the null hypothesis of random residuals will be rejected if $Q_{LB} > \chi^2_{1-\alpha}(h)$.

Two thousand realizations of the noisy BOLD signals generated by the Balloon model with different random stimulus functions and corrupted by the Gaussian noise with different variances are simulated. The Ljung and Box statistic test shows that 99.65% realizations of the simulation survived the statistical test ($P = 0.05$). This clearly shows that the proposed NARX model is capable of capturing the dynamics of the simulated BOLD signals.

7.3.2 Experimental fMRI Data

The proposed dynamical modelling method based on the NARX neural network is first tested on an experimental event-related fMRI data set(*DATA-EVENT* in Chapter 1). Before applying the proposed NARX modelling method to this experimental event-related fMRI data, the raw data are preprocessed by the SPM software [29] for registration, normalization and smoothing. The drift in each voxel time series is also removed.

To detect the activated regions of the brain, the reconstructed BOLD signal $\hat{\mathbf{y}}_b$ at the output of the NARX model is investigated. For the inactivated voxels, the reconstructed BOLD signal would be almost zero since the inactivated voxels do not involve the BOLD effect corresponding to the external stimuli. For the activated voxels, on the other hand, the reconstructed BOLD signal should capture the dynamics corresponding to the input stimulations, that is, $\hat{\mathbf{y}}_b \neq 0$. Hence, it is desirable to use some parameters that would indicate how probable the reconstructed BOLD signal $\hat{\mathbf{y}}_b \neq 0$. We introduce the following test at each voxel as an index (R) for activation detection:

$$R = \frac{\|\hat{\mathbf{y}}_b\|}{\|\mathbf{y} - \hat{\mathbf{y}}_b\|} \tag{7.3}$$

where $\mathbf{y}$ is the measured fMRI signal in a voxel and $\hat{\mathbf{y}}_b$ is the reconstructed BOLD signal in this voxel. The R value is analogous to the SNR. When the voxel is activated, the test value R would be large; while when it is not activated, the value would be small. Since noise exists, $\mathbf{y} \neq \hat{\mathbf{y}}_b$, this prevents the value of R from approaching infinity. Figure 7.4 shows the results of the activated regions in the brain by thresholding the R values at each voxel. It is clear from this figure that the auditory cortex in the brain is activated in this event-related fMRI experiment.

Figure 7.5 compares the estimated HDR of an activated voxel from the NARX model with the conventional model formulated by the difference of two gamma functions. The solid line is the estimated HDR using the NARX model (averaged over the activated

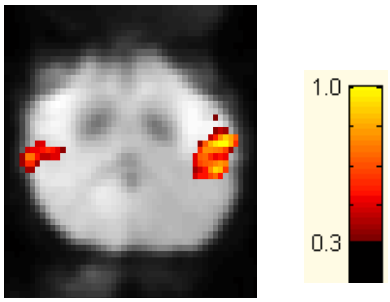


Figure 7.4 One slice showing the activation of the auditory cortex ($R > 0.3$).

regions) and the dotted line is the HDR formulated by the difference of gamma functions. These two waveforms are scaled to similar magnitudes in order to have better comparison. From this figure, it is clear that the two HDRs are analogous to each other. Though there are some differences between these two responses, the trends of the signal variation are the same, for example, both of them reach the peak at around 5s after the stimulus. The difference between these two responses reflect the ability of the proposed method to capture the variation of the HDR. This shows that the NARX model is able to reconstruct the dynamics of the fMRI data and provides a good estimation of the HDR. In addition, the NARX model is more flexible than the conventional HDR model formulated by means of gamma function model. The gamma function model may miss the different dynamics between regions and subjects since it uses the same HDR model for all the voxels investigated. However, the NARX model is trained on each voxel and is able to capture the differences in BOLD signals between regions and subjects.

The proposed NARX reconstruction method is also tested on the experimental block design dataset (*DATA-BLOCK* in Chapter 1). Figure 7.6 shows the original and the estimated time courses reconstructed by NARX model of both activated and inactivated voxels. From Figure 7.6, we can see that for the activated voxel, the reconstructed signals by using the proposed NARX method can approximate the real fMRI signal

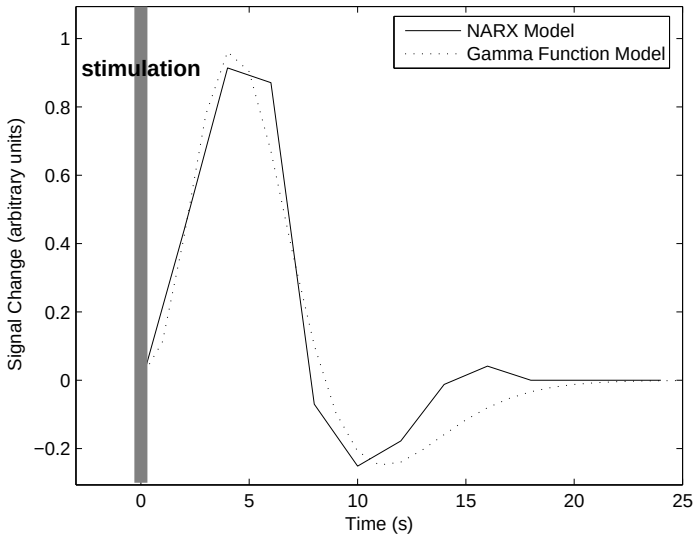
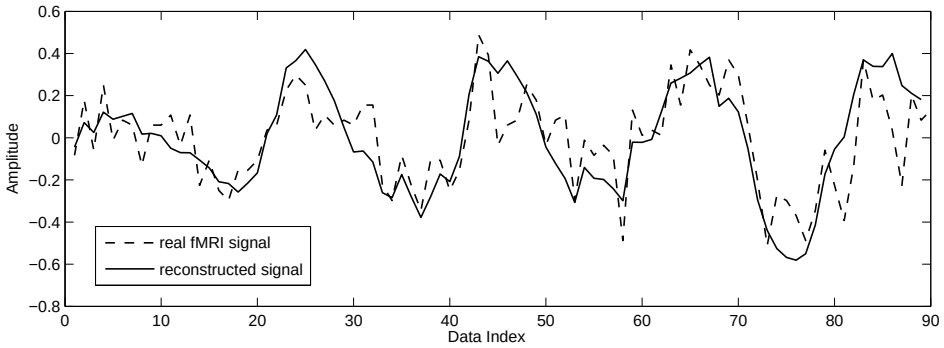


Figure 7.5 Comparison between the estimated HDR from the NARX model and the HDR formulated by difference of two Gamma functions.

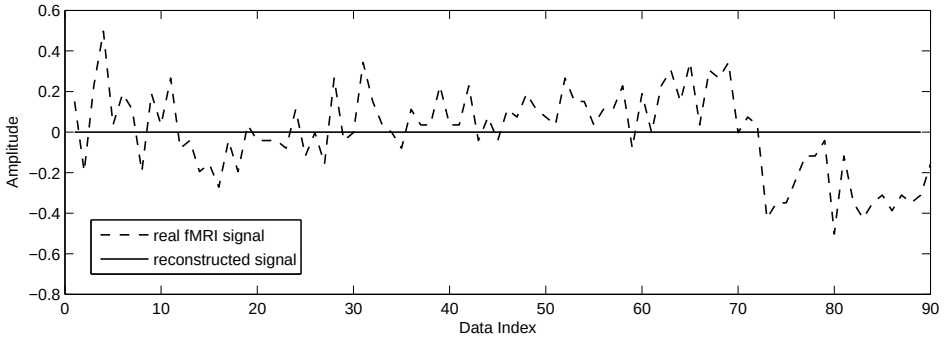
but less noisy than the original fMRI data. For the inactivated voxel, the reconstructed signal is almost zero since there is no BOLD signal related to the input stimulus in the measured data. This clearly confirms that the proposed NARX method can model and capture the complex and nonlinear properties underlying the fMRI data.

7.4 Conclusion

The objective of the fMRI data analysis is to accurately detect the activated regions of the human brain corresponding to a specific task and also to investigate how the brain responds to the stimulus, i.e., to study the dynamics of the human brain. The BOLD signal, as the foundation of the fMRI experiment, reflects the hemodynamics of human brain. How to reconstruct the BOLD signal from the noisy measured fMRI signals is of great importance to the fMRI data analysis. From an engineering perspective, this is a



(a) Activated voxel



(b) Inactivated voxel

Figure 7.6 Time courses of an activated and an inactivated voxel for the real block experimental fMRI data.

system modelling problem, that is, to identify the complex human brain system as well as to reconstruct the dynamics of the human brain.

In this chapter, the NARX model is proposed for the fMRI data analysis to capture the system dynamics. With the knowledge of experimental paradigm (input) and measured data (output), the NARX neural networks are investigated to identify the complex human brain system. The proposed scheme is realized through the RBF neural networks. Results from simulated as well as experimental event-related and block fMRI data show that the proposed model can successfully capture the underlying dynamics of the human brain. Comparison between the HDR estimated by the NARX model and the conventional difference of gamma functions model shows the effectiveness of the NARX model. This indicates that the NARX neural networks are good methods to model and capture the dynamics underlying the fMRI data.

Chapter 8

Conclusion and Future Directions

8.1 Summary and Conclusions

Functional MR Imaging is an important technique for neuroimaging. It can relate the structure of the brain to its function by using the MRI techniques to measure the human brain function. The processing and analysis of fMRI data are challenging due to its underlying physiological complexity. The research work presented in this thesis provide advanced signal processing and data analysis methods to the complex fMRI data.

The major objectives of this thesis are: (i) detection of the activated regions of the brain; (ii) estimation of the HDR and (iii) modeling the dynamics of the fMRI signal. The fMRI data were first examined under a Bayesian framework, which could efficiently detect the activated regions of the human brain. Then, the HDR which reflects the temporal properties of human brain function was estimated through both linear and nonlinear methods. Finally, the NARX neural network was proposed to model the dynamics of fMRI signal. Through the results of both simulated as well as experimental fMRI data, it is shown that these methods are robust, efficient and flexible. They can complement the traditional analysis methods to cope with diverse challenges of fMRI data analysis. The

findings from this research work could help the neurologist and psychologist interpret better the fMRI data.

In the first part of the research, a Bayesian framework for fMRI data analysis was provided. Compared to the hypothesis test methods, Bayesian methods have many advantages such as easy extension to group analysis and the avoidance of the multiple comparisons problem. A sparse Bayesian learning for flexible design matrix determination in the GLM model was first proposed (Chapter 2). Through the comparison of ROC curves, it has been shown that the sparse Bayesian method is much more robust than the traditional GLM method. This improvement comes from the integration of the advantages of the data-driven and model-driven methodologies. Then, this method was extended to the scenarios where the noise variance is varying (either a time-varying noise or a fractional noise) (Chapter 3). Comparison with OLS and WLS methods showed that the proposed Bayesian method performs best when dealing with the nonstationary nature of the fMRI data. This improvement is mainly because the Bayesian method could accurately estimate and model the nonstationary noise structure in the fMRI data. In Chapter 4, the drift, which was modelled by a few large scale wavelets, was incorporated into the GLM under the Bayesian framework. A modified GLM and CIC for model selection were proposed. This method could successfully model and remove the drift in the fMRI data in order to obtain better activation results. The proposed Bayesian methods show significant improvement for the brain activation detection in fMRI data analysis.

The second part of the research was mainly focused on the estimation of the HDR. A spatio-temporal adaptive linear filter method for fMRI data analysis was first investigated (Chapter 5). It was proved in our research that this method is equivalent to the CCA method. Beyond the linear methods which are mainly suitable for fMRI experiment with large ISI, nonlinear methods are of great interest especially for rapid event-related fMRI data (or ISI less than 4-6 seconds). In Chapter 6, the RBF neural network

was proposed to analyze the nonlinearities in the fMRI data. It was shown that the RBF neural network is equivalent to the conventional Volterra-series method for nonlinear system modeling, but that it is more flexible and computationally efficient. This advantage is due to the universal approximation property of the RBF neural network and the disassociation of the nonlinear terms/kernels of different orders (by using Taylor expansion). This method is of considerable importance since it provides a completely new framework for nonlinear fMRI data analysis which is required by the advent of the more advanced and complex rapid event-related fMRI experimental design.

The third part of the research meant to model the dynamics of the fMRI data. In Chapter 7, the NARX neural network is proposed to capture the dynamics of the fMRI signal. With the knowledge of the experimental paradigm (input) and measured data (output), the NARX neural network could provide us with a promising method to model and reconstruct the dynamics existing in the fMRI data and hence identify this complex system.

In this research, novel and powerful advanced algorithms and analysis methods for fMRI data analysis have been proposed to cope with various difficulties that the neurologist and psychologist are facing while analyzing the measured fMRI data. The methods and techniques introduced in this research would allow the investigators to study different properties of the fMRI data and help to further elucidate the underlying physiological mechanisms.

8.2 Future Directions

The Bayesian methods proposed in the first part of this thesis were based on the assumption that the noise is Gaussian (in time domain or wavelet domain). However, this assumption may have some reservations considering the complex acquisition procedure

of fMRI data. Other noises either from the subject's non-experimental process or from the fMRI measurement process were not taken into account. Though the preprocessing step such as Gaussian smoothing may make the data satisfy Gaussian assumption, the detection ability can be improved by incorporating other factors in our model. Future work of the Bayesian method would be to refine the noise structures in the analysis to relax the assumption requirement. These could be done both with parametric and non-parametric methods. Specifically, Rician noises which are proven to exist in the MR images can be investigated.

In the second part of this thesis, methods to estimate the HDR are presented. However, since it is difficult to obtain the precise probability distribution of the reconstructed fMRI data, a traditional significance level of the hypothesis test cannot be derived from these methods. Future work would be needed to identify the distribution of the reconstructed data. A possible way is to obtain extra null data (the data is recorded when the subjects lie in the scanner and do not perform any tasks) for each fMRI experiment and derive the probability distribution from this null data. Another promising method would be to use parameter-free tests, such as permutation test and so on.

In chapter 7, the NARX neural network was presented to model the dynamics of the fMRI data. Further research of this method includes: i) incorporating the spatial data into this model; ii) analyzing the stability of the algorithm and iii) decreasing the computational complexity. It would be interesting to see the extension of this method to multidimensional signals as well.

Besides these future improvement of the methods, a quantitative comparison of the performance between different methods on experimental fMRI data is desired. This can be implemented by two ways. On the one side, a suitable criterion need to be proposed to assess the performance of different methods on experimental fMRI data. On the other side, these methods could be tested on some special experimental data,

such as the retinotopic stimulation experiment. For these special experimental data, the true activation of the brain regions could be obtained by using other methods. And these truly activated regions can serve as the reference to compare the performance of different detection and estimation methods.

Bibliography

[1] M. S. Gazzaniga, R. B. Ivry, and G. R. Mangun, *Cognitive neuroscience : the biology of the mind*, 2nd ed. W.W.Norton & Company, 2002, ch. 1.

[2] J. Ward, *The student's guide to cognitive neuroscience*. Psychology Press, 2006, p. 3.

[3] V. Walsh and A. Pascual-Leone, *Transcranial magnetic stimulation: a neurochronometrics of mind*. Cambridge: MIT Press, 2003, ch. 1.

[4] M. D. Rugg and M. G. H. Coles, *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford University Press, 1996, ch. 1.

[5] R. L. Savoy, "History and future directions of human brain mapping and functional neuroimaging," *Acta Psychologica*, vol. 107, pp. 9–42, 2001.

[6] M. A. Brown and R. C. Semelka, *MRI : Basic Principles and Applications*. Wiley-Liss, 2003, ch. 1.

[7] P. Jezzard, P. M. Matthews, and S. M. Smith, Eds., *Functional MRI: an introduction to methods*. Oxford University Press, 2001, ch. 3.

[8] J. W. Santrock, *Psychology*, updated 7th ed. McGraw-Hill, 2003, ch. 3.

[9] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with constrast dependent on blood oxygenation," *Proceedings of the National Academy of Sciences*, vol. 87, pp. 9868–9872, 1990.

[10] P. A. Bandettini, E. C. Wong, R. S. Hinks, R. S. Tikofsky, and J. S. Hyde, "Time course epi of human brain function during task activation," *Magnetic Resonance in Medicine*, vol. 25, pp. 390–397, 1992.

[11] J. W. Belliveau, D. N. Kennedy, R. C. McKinstry, B. R. Buchbinder, R. M. Weisskoff, M. S. Cohen, J. M. Vevea, T. J. Brady, and B. R. Rosen, "Functional mapping of the human visual cortex by magnetic resonance imaging," *Science*, vol. 254, pp. 716–719, 1991.

[12] P. Jezzard, P. M. Matthews, and S. M. Smith, Eds., *Functional MRI: an introduction to methods*. Oxford University Press, 2001, ch. 8.

[13] G. K. Aguirre, E. Zarahn, and M. D'Esposito, "The variability of human bold hemodynamic response," *NeuroImage*, vol. 8, pp. 360–369, 1998.

[14] S. A. Huettel, A. W. Song, and G.McCarthy, *Functional Magnetic Resonance Imaging*. Sinauer Associates, Inc, 2004, ch. 11.

[15] V. Ng, E. Bullmore, G. de Zubicaray, A. Cooper, J. Suckling, and S. Williams, "Identifying rate-limiting nodes in large-scale cortical networks for visuospatial processing: An illustration using fMRI," *Journal of Cognitive Neuroscience*, vol. 13, 2001.

[16] I. Klein, A. Paradis, J. Poline, S. Kosslyn, and D. Bihan, "Transient activity in the human calcarine cortex during visual-mental imagery: an event-related fMRI study," *Journal of Cognitive Neuroscience*, vol. 12, s2, pp. 15–23, 2000.

[17] J. Talairach and P. Tournoux, *A Co-planer Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers, 1988.

[18] D. Collins, P. Neelin, T. Peters, and A. Evans, "Automatic 3d intersubject registration of mr volumetric data in standardized talairach space," *Journal of Computer Assisted Tomography*, vol. 18, pp. 192–205, 1994.

[19] S. Kiebel, J. Poline, K. Friston, A. Holmes, and K. Worsley, "Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model," *Neuroimage*, vol. 10, pp. 755–756, 1999.

[20] K. Friston, O. Josephs, E. Zarahn, A. Holmes, S. Rouquette, and J. Poline, "To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis," *Neuroimage*, vol. 12, pp. 196–208, 2000.

[21] R. Frackowiak, K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penny, Eds., *Human Brain Function*, 2nd ed. Elsevier Academic Press, 2003, ch. 44.

[22] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time-series," *Human Brain Mapping*, vol. 1, pp. 153–171, 1994.

[23] G. Boynton, S. Engel, G. Glover, and D. Heeger, "Linear systems analysis of functional magnetic resonance imaging in human V1," *Journal of Neuroscience*, vol. 16, pp. 4207–4221, 1996.

[24] J. Rajapakse, F. Kruggel, J. Maisog, and D. Cramon, "Modeling hemodynamic response for analysis of functional MRI time series," *Human Brain Mapping*, vol. 6, pp. 283–300, 1998.

[25] K. Friston, P. Fletcher, O. Josephs, A. Holmes, M. Rugg, and R. Turner, "Event-related fMRI: Characterising differential responses," *Neuroimage*, vol. 7, pp. 30–40, 1998.

[26] R. Buxton, E. Wong, and L. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: the balloon model," *Magnetic Resonance in Medicne*, vol. 39, no. 6, pp. 855–864, 1998.

[27] A. Smith, B. Lewis, U. Ruttinmann, F. Ye, T. Sinnwell, Y. Yang, J. Duyn, and J. Frank, "Investigation of low frequency drift in fMRI signal," *Neuroimage*, vol. 9, pp. 526–533, 1999.

[28] J. Brosch, T. Talavage, J. Ulmer, and J. Nyenhuis, "Simulation of human respiration in fMRI with a mechanical model," *IEEE Transactions on Biomedical Engineering*, vol. 49, pp. 700–707, 2002.

[29] K. Friston, J. Ashburner, and et al, *SPM 97 course notes*, Wellcome Department of Cognitive Neurology, University College London.

[30] K. Worsley, C. Liao, J. Aston, V. Petre, G. Duncan, F. Morales, and A. Evans, "A general statistical analysis for fMRI data," *Neuroimage*, vol. 15, no. 1, pp. 1–15, 2002.

[31] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. Frackowiak, "Functional connectivity: the principal component analysis of large PET data sets," *Journal of Cerebral Blood Flow Metabolism*, vol. 13, no. 1, pp. 5–14, 1993.

[32] K. J. Friston, J. B. Poline, S. Strother, A. P. Holmes, C. D. Frith, and R. S. Frackowiak, "A multivariate analysis of PET activation studies," *Human Brain Mapping*, vol. 4, pp. 140–151, 1996.

[33] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.

[34] M. McKeown, S.Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.

[35] V. Calhoun, T. Adali, G. Pearlson, and J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Human Brain Mapping*, vol. 13, no. 43-53, 2001.

[36] M. J. Mckeown, "Detection of consistently task-related activations in fMRI data with hybrid independent component analysis," *Neuroimage*, vol. 11, pp. 24–35, 2000.

[37] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Exploratory fMRI analysis by autocorrelation maximization," *Neuroimage*, vol. 16, pp. 454–464, 2002.

[38] H. Chen, H. Yuan, D. Yao, L. Chen, and W. Chen, "An integrated neighborhood correlation and hierarchical clustering approach of functional MRI," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 452–458, 2006.

[39] M. L. Seghier, K. J. Friston, and C. J. Price, "Detecting subject-specific activations using fuzzy clustering," *NeuroImage*, vol. 36, pp. 594–605, 2007.

[40] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, 2002.

[41] P. L. Purdon and R. M. Weisskoff, "Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI," *Human Brain Mapping*, vol. 6, pp. 239–249, 1998.

[42] M. Burock and A.M.Dale, "Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach," *Human Brain Mapping*, vol. 11, pp. 249–260, 2000.

[43] J. L. Devore, *Probability and statistics for engineering and the sciences*, 5th ed. Duxbury, 2000.

[44] R. Frackowiak, K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penny, Eds., *Human Brain Function*, 2nd ed. Elsevier Academic Press, 2003, ch. 37.

[45] J. F. Hair, B. Black, B. Babin, R. E. Anderson, and R. L. Tatham, *Multivariate data analysis*, 6th ed. Pearson Prentice Hall, 2006, ch. 7.

[46] A. Andersen, D. Gash, and M. Avison, "Principal component analysis of the dynamic response measured by fMRI: a generalized linear systems framework," *Magnetic Resonance Imaging*, vol. 17, pp. 795–815, 1999.

[47] A. Baune, F. T. Sommer, M. Erb, D. Wildgruber, B. Kardatzki, G. Palm, and W. Grodd, "Dynamical cluster analysis of cortical fMRI activation," *NeuroImage*, vol. 9, pp. 477–489, 1999.

[48] J. C. Rajapakse and W. Lu, "Extracting task-related components in functional MRI," in *the 3rd International Conference on Independent Component Analysis and Blind Signal Separation, ICA 2001*, San Diego, California, 2001.

[49] V. Calhoun, T. Adali, M. Stevens, K. Kiehl, and J. Pekar, "Semi-blind ICA of fMRI: a method for utilizing hypothesis-derived time courses in a spatial ICA analysis," *NeuroImage*, vol. 25, pp. 527–538, 2005.

[50] K. Worsley and K. Friston, "Analysis of fMRI time-series revisited-again," *NeuroImage*, vol. 2, pp. 173–181, 1995.

[51] V. Solo, E. N. Brown, and C. J. Long, "Spatial wavelets for temporally correlated fMRI," in *International Conference on Image Processing, ICIP 2003*, Barcelona, Spain, 2003.

[52] M. W. Woolrich, B. D. Ripley, M. Brady, and S. M. Smith, "Temporal autocorrelation in univariate linear modeling of fMRI data," *NeuroImage*, vol. 14, pp. 1370–1386, 2001.

[53] R. Frackowiak, K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penny, Eds., *Human Brain Function*, 2nd ed. Elsevier Academic Press, 2003, ch. 40.

[54] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[55] D. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–417, 1992.

[56] K. Friston, D. Glaser, R. Henson, S. Kiebel, C. Phillips, and J. Ashburner, "Classical and bayesian inference in neuroimaging: Applications," *Neuroimage*, vol. 16, no. 2, pp. 484–512, 2002.

[57] P. Jezzard, P. M. Matthews, and S. M. Smith, Eds., *Functional MRI: an introduction to methods*. Oxford University Press, 2001, ch. 14.

[58] R. Constable, P. Skudlarski, and J. Gore, "An ROC approach for evaluating functional brain MR imaging and postprocessing protocols," *Magnetic Resonance in Medicine*, vol. 34, pp. 57–64, 1995.

[59] M. Woolrich, M. Jenkinson, J. Brady, and S. M. Smith, "Fully Bayesian spatio-temporal modeling of fMRI data," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 213–231, 2004.

[60] C. Long, E. Brown, C. Triantafyllou, I. Aharon, L. L. Wald, and V. Solo, "Nonstationary noise estimation in functional MRI," *NeuroImage*, vol. 28, pp. 890–903, 2005.

[61] J. Diedrichsen and R. Shadmehr, "Detecting and adjusting for artifacts in fMRI time series data," *NeuroImage*, vol. 27, pp. 624–634, 2005.

[62] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter, and M. Brammer, "Colored noise and computational inference in neurophysilogical (fMRI) time series analysis: resampling methods in time and wavelet domain," *Human Brain Mapping*, vol. 12, pp. 61–78, 2001.

[63] E. Bullmore, J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear, "Wavelets and functional magnetic resonance imaging of the human brain," *NeuroImage*, vol. 23, pp. S234–S249, 2004.

[64] W. Penny, S. Kiebel, and K. Friston, "Variational Bayesian inference for fMRI time series," *NeuroImage*, vol. 19, pp. 1477–1491, 2003.

[65] M. Woolrich, T. Behrens, and S. Smith, "Constrained linear basis sets for HRF modelling using variational bayes," *NeuroImage*, vol. 21, pp. 1748–1761, 2004.

[66] W. Penny, J. Kilner, and F. Blankenburg, "Robust bayesian general linear models," *NeuroImage*, vol. 36, pp. 661–671, 2007.

[67] K. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: theory," *NeuroImage*, vol. 16, no. 2, 2002.

[68] M. Woolrich, T. Behrens, C. Beckman, M.Jenkinson, and S. Smith, "Multilevel linear modelling for fMRI group analysis using Bayesian inference," *NeuroImage*, vol. 21, pp. 1732–1747, 2004.

[69] G. Flandin and W. D. Penny, "Bayesian fmri data analysis with sparse spatial basis function priors," *NeuroImage*, vol. 34, pp. 1108–1125, 2007.

[70] E. Bullmore, J. Fadili, M. Breakspear, R. Salvador, J. Suckling, and M. Brammer, "Wavelets and statistical analysis of functional magnetic resonance images of the human brain," *Statistical Methods in Medical Research*, vol. 12, pp. 375–399, 2003.

[71] S. Mallat, *A Wavlet Tour of Signal Processing*, 2nd ed. Academic Press, 1999, ch. 7.

[72] J. Fadili and E. Bullmore, "Wavelet-generalized least squares: a new BLU estimator of linear regression models with 1/f errors," *NeuroImage*, vol. 15, pp. 217–232, 2002.

[73] G. Wornell, "A karhunen-loève-like expansion for 1/f processes via wavelets," *IEEE Transactions on Information Theory*, vol. 36, pp. 859–861, 1990.

[74] P. Flandrin, "Wavelet analysis and synthesis of fractional brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, 1992.

[75] A. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 904–909, 1992.

[76] E. McCoy and A. Walden, "Wavelet analysis and synthesis of stationary long-memory processes," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 26–56, 1996.

[77] H. Hinrichs, M. Scholz, C. Tempelmann, M. G. Woldorff, A. M. Dale, and H.-J. Heinze, "Deconvolution of event-related fmri responses in fast-rate experimental designs: tracking amplitude variations," *Journal of Cognitive Neuroscience*, vol. 12, Supplement 2, pp. 76–89, 2000.

[78] F. Meyer, "Wavelet-based estimation of a semiparametric generalized linear model of fmri time-series," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 315–322, 2003.

[79] G. Wornell, "A Karhunen-loève-like expansion for 1/f processes via wavelets," *IEEE Transactions on Information Theory*, vol. 36, pp. 859–861, 1998.

[80] K. L. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, pp. 297–307, 1989.

[81] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[82] A. Dale and R. Buckner, "Selective averaging of rapidly presented individual trials using fMRI," *Human Brain Mapping*, vol. 5, pp. 329–340, 1997.

[83] O. Josephs, R. Turner, and K. Friston, "Event-related fMRI," *Human Brain Mapping*, vol. 5, pp. 243–248, 1997.

[84] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Adaptive analysis of fMRI data," *NeuroImage*, vol. 19, pp. 837–845, 2003.

[85] G. Hossein-Zadeh, B. Ardekani, and H. Soltanian-Zadeh, "Activation detection in fMRI using a maximum engergy ratio statistic obtained by adaptive spatial filtering," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 795–805, 2003.

[86] B. Bonakdarpour, T. Parrish, and C. Thompson, "Hemodynamic response function in patients with stroke-induced aphasia: Implications for fmri data analysis," *NeuroImage*, vol. 36, no. 2, 2007.

[87] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Application*. John Wiley & Sons, 1999.

[88] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 2002, p. 440.

[89] J. R. Duann, T. P. Jung, W. J. Kuo, T. C. Yeh, S. Makeig, J. C.Hsieh, and J. T. Sejnowski, "Single-trial variability in event-related BOLD signals," *NeuroImage*, vol. 15, no. 4, pp. 823–835, 2002.

[90] Y. Lu, T. Jiang, and Y. Zang, "Single-trial variable model for event-related fMRI data analysis," *IEEE Transanctions on Medical Imaging*, vol. 24, no. 2, pp. 236–245, 2005.

[91] O. Friman, J. Carlsson, P. Lundberg, M. Borga, and H. Knutsson, "Detection of neural activity in functional MRI using Canonical Correlation Analysis," *Magnetic Resonance in Medicine*, vol. 45, pp. 323–330, 2001.

[92] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R.S.J.Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, pp. 189–210, 1995.

[93] R. B. Buxton, *Introduction to functional magnetic resonance imaging : principles and techniques*. Cambridge University Press, 2002, ch. 3.

[94] D. A. Handwerker, J. M. Ollinger, and M. D'Esposito, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *NeuroImage*, vol. 21, pp. 1639–1651, 2004.

[95] R. M. Birn, Z. S. Saad, and P. A. Bandettini, "Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response," *NeuroImage*, vol. 14, pp. 817–826, 2001.

[96] K. L. Miller, W. M. Luh, T. T. Liu, A. Martinez, T. Obata, E. C. Wong, L. R. Frank, and R. B. Buxton, "Nonlinear temporal dynamics of the cerebral blood flow response," *Humain Brain Mapping*, vol. 13, pp. 1–12, 2001.

[97] A. Laird, B. Rogers, and M. Meyerand, "Investigating the nonlinearity of fMRI activation data," in *Proceedings of Second Joint EMBS/BMES Conference*, 2002, pp. 23–26.

[98] T. D. Wager, A. Vazquez, L. Hernandez, and D. C. Noll, "Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies," *NeuroImage*, vol. 25, pp. 206–218, 2005.

[99] R. Buxton, K. Uludag, D. J. Dubowitz, and T. T. Liu, "Modeling the hemodynamic response to brain activation," *NeuroImage*, vol. 23, pp. S220–S233, 2004.

[100] S. Haykin, *Neural Networks, A comprehensive foundation*. Prentice Hall, 1999.

[101] M. Misaki and S. Miyauchi, "Application of artificial neural network to fMRI regression analysis," *NeuroImage*, vol. 29, pp. 396–408, 2006.

[102] E. Rank, "Application of Bayesian trained RBF networks to nonlinear time-series modeling," *Signal Processing*, vol. 83, pp. 1393–1410, 2003.

[103] P. L.Purdon and R. Weisskoff, "Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI," *Human Brain Mapping*, vol. 6, pp. 239–249, 1998.

[104] J. Tanabe, D. Miller, J. Tregellas, R. Freedman, and F. G. Meyer, "Comparison of detrending methods for optimal fMRI preprocessing," *NeuroImage*, vol. 15, pp. 902–907, 2002.

[105] L. Harrison, W. D. Penny, and K. Friston, "Multivariate autoregressive modeling of fMRI time series," *NeuroImage*, vol. 19, pp. 1477–1491, 2003.

[106] M. Kamba, Y. Sung, and S. Ogawa, "A dynamic system model-based technique for functional MRI data analysis," *NeuroImage*, vol. 22, pp. 179–187, 2004.

[107] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transaction on Neural Networks*, vol. 1, pp. 4–27, 1990.

[108] T. Lin, B. G. Horne, P. Tino, and C. L. Giles, "Learning long-term dependencies in NARX recurrent neural networks," *IEEE Transaction on Neural Networks*, vol. 7, pp. 1329–1338, 1996.

[109] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, pp. 297–303, 1978.

[110] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, 2002, p. 536.

Appendix A

Derivation of Eq. (3.22) and Eq. (3.23)

From Eq. (3.17), Eq. (3.18) and Eq. (3.21), we get:

$$\begin{aligned} p(\mathbf{w}|\mathbf{y},\mathbf{A},\mathbf{B})p(\mathbf{y}|\mathbf{A},\mathbf{B}) &= p(\mathbf{y}|\mathbf{w},\mathbf{B})p(\mathbf{w}|\mathbf{A}) \\ &\sim \exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{\Phi w})^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi w}) - \frac{1}{2}\mathbf{w}^T\mathbf{A}\mathbf{w}\} \\ &\sim \exp\{-\frac{1}{2}\big[(\mathbf{y}-\mathbf{\Phi w})^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi w}) + \mathbf{w}^T\mathbf{A}\mathbf{w}\big]\}. \end{aligned} \tag{A.1}$$

Expanding the quantities in the square brackets in the exponential and grouping together all the terms containing $\mathbf{w}$, we get:

$$\begin{aligned} & (\mathbf{y}-\mathbf{\Phi w})^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi w}) + \mathbf{w}^T\mathbf{A}\mathbf{w} \\ = \;& \mathbf{y}^T\mathbf{B}\mathbf{y} - \mathbf{y}^T\mathbf{B}\mathbf{\Phi w} - \mathbf{w}^T\mathbf{\Phi}^T\mathbf{B}\mathbf{y} + \mathbf{w}^T\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi w} + \mathbf{w}^T\mathbf{A}\mathbf{w} \\ = \;& \mathbf{w}^T(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi})\mathbf{w} - \mathbf{y}^T\mathbf{B}\mathbf{\Phi w} - \mathbf{w}^T\mathbf{\Phi}^T\mathbf{B}\mathbf{y} + \mathbf{y}^T\mathbf{B}\mathbf{y} \\ = \;& \big[\mathbf{w}-(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{B}\mathbf{y}\big]^T(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi})\big[\mathbf{w}-(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{B}\mathbf{y}\big] \\ & + \mathbf{y}^T\big[\mathbf{B}-\mathbf{B}\mathbf{\Phi}(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{B}\big]\mathbf{y} \end{aligned} \tag{A.2}$$

According to the matrix inversion lemma [88], we get:

$$(\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1} = \mathbf{B}-\mathbf{B}\mathbf{\Phi}(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{B}, \tag{A.3}$$

Eq. (A.2) can be written as:

$$\begin{aligned} & (\mathbf{y}-\mathbf{\Phi w})^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi w})+\mathbf{w}^T\mathbf{A}\mathbf{w} \\ = \; & (\mathbf{w}-\mathbf{u})^T\mathbf{\Lambda}^{-1}(\mathbf{w}-\mathbf{u})+\mathbf{y}^T(\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{y} \end{aligned} \tag{A.4}$$

where

$$\mathbf{\Lambda}^{-1}=(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi}) \tag{A.5}$$

and

$$\mathbf{u}=\mathbf{\Lambda}\mathbf{\Phi}^T\mathbf{B}\mathbf{y}. \tag{A.6}$$

Thus,

$$\begin{aligned} & p(\mathbf{w}|\mathbf{y},\mathbf{A},\mathbf{B})p(\mathbf{y}|\mathbf{A},\mathbf{B}) \\ \sim \; & \exp\{-\frac{1}{2}(\mathbf{w}-\mathbf{u})^T\mathbf{\Lambda}^{-1}(\mathbf{w}-\mathbf{u})-\frac{1}{2}\mathbf{y}^T(\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{y}\} \end{aligned} \tag{A.7}$$

By inspection, we will get two Gaussian distributions:

$$p(\mathbf{w}|\mathbf{y},\mathbf{A},\mathbf{B})=(2\pi)^{-\frac{M}{2}}|\mathbf{\Lambda}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{w}-\mathbf{u})^T\mathbf{\Lambda}^{-1}(\mathbf{w}-\mathbf{u})\}, \tag{A.8}$$

and

$$\begin{aligned} p(\mathbf{y}|\mathbf{A},\mathbf{B}) \; = \; & (2\pi)^{-\frac{T}{2}}|\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T|^{-\frac{1}{2}} \\ & \exp\{-\frac{1}{2}\mathbf{y}^T(\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{y}\}. \end{aligned} \tag{A.9}$$

These are Eq. (3.22) and Eq. (3.23), respectively.

Appendix B

Derivation of Eq. (3.27) to Eq. (3.29)

In this part, we consider how to effectively compute the objective function $\mathcal{L}$ and the derivation of the hyperparameter updates.

B.1 Compute the the objective function $\mathcal{L}$

The objective function is:

$$\mathcal{L} \;=\; -\frac{1}{2}\log|\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^{T}| - \frac{1}{2}\mathbf{y}^{T}(\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^{T})^{-1}\mathbf{y}. \tag{B.1}$$

We first compute the first term in $\mathcal{L}$. According to the properties of matrix determinants [110]:

$$|\mathbf{C}||\mathbf{A}+\mathbf{B}\mathbf{C}^{-1}\mathbf{D}| = |\mathbf{A}||\mathbf{C}+\mathbf{D}\mathbf{A}^{-1}\mathbf{B}|, \tag{B.2}$$

where $\mathbf{A}\in\mathbb{R}^{s\times s}$, $\mathbf{C}\in\mathbb{R}^{r\times r}$ are nonsingular, $\mathbf{B}\in\mathbb{R}^{s\times r}$, $\mathbf{D}\in\mathbb{R}^{r\times s}$, we get:

$$|\mathbf{A}||\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^{T}| = |\mathbf{B}^{-1}||\mathbf{A}+\mathbf{\Phi}^{T}\mathbf{B}\mathbf{\Phi}|. \tag{B.3}$$

And this gives:

$$\log|\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^{T}| = -\log|\mathbf{\Lambda}| - \log|\mathbf{B}| - \log|\mathbf{A}|, \tag{B.4}$$

where $\mathbf{\Lambda}$ has been defined in Eq. (A.5).

The second term in objective function $\mathcal{L}$ is data dependent and we can further represent the term as:

$$\begin{aligned}
\mathbf{y}^T(\mathbf{B}^{-1}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{y} &= \mathbf{y}^T(\mathbf{B}-\mathbf{B}\mathbf{\Phi}(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{B})\mathbf{y} \\
&= \mathbf{y}^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi}(\mathbf{A}+\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{B}\mathbf{y}) \\
&= \mathbf{y}^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi}\mathbf{u}) \\
&= (\mathbf{y}-\mathbf{\Phi}\mathbf{u})^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi}\mathbf{u})+\mathbf{u}^T\mathbf{\Phi}^T\mathbf{B}\mathbf{y}-\mathbf{u}^T\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi}\mathbf{u} \\
&= (\mathbf{y}-\mathbf{\Phi}\mathbf{u})^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi}\mathbf{u})+\mathbf{u}^T\mathbf{\Lambda}^{-1}\mathbf{u}-\mathbf{u}^T\mathbf{\Phi}^T\mathbf{B}\mathbf{\Phi}\mathbf{u} \\
&= (\mathbf{y}-\mathbf{\Phi}\mathbf{u})^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi}\mathbf{u})+\mathbf{u}^T\mathbf{A}\mathbf{u}. \qquad \text{(B.5)}
\end{aligned}$$

And thus, the objective function $\mathcal{L}$ becomes:

$$\mathcal{L}=-\frac{1}{2}[-\log|\mathbf{\Lambda}|-\log|\mathbf{B}|-\log|\mathbf{A}|+(\mathbf{y}-\mathbf{\Phi}\mathbf{u})^T\mathbf{B}(\mathbf{y}-\mathbf{\Phi}\mathbf{u})+\mathbf{u}^T\mathbf{A}\mathbf{u}]. \qquad \text{(B.6)}$$

B.2 Derivatives and updates

The derivative of Eq. (B.6) with respect to α_i is:

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\alpha_i} &= -\frac{1}{2}[\text{trace}(\mathbf{\Lambda}\frac{\partial\mathbf{\Lambda}^{-1}}{\partial\alpha_i})-\frac{1}{\alpha_i}+u_i^2] \qquad \text{(B.7)} \\
&= -\frac{1}{2}[\mathbf{\Lambda}_{ii}-\frac{1}{\alpha_i}+u_i^2]. \qquad \text{(B.8)}
\end{aligned}$$

Setting the above equation to zero gives the estimate of α_i:

$$\alpha_i=\frac{1}{\mathbf{\Lambda}_{ii}+u_i^2}. \qquad \text{(B.9)}$$

The derivative of Eq. (B.6) with respect to s_i is:

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial s_i} &= -\frac{1}{2}[\text{trace}(\mathbf{\Lambda}\frac{\partial\mathbf{\Lambda}^{-1}}{\partial s_i})-\frac{1}{s_i}+\beta(\mathbf{y}-\mathbf{\Phi}\mathbf{u})_i^2] \qquad \text{(B.10)} \\
&= -\frac{1}{2}[\text{trace}(\mathbf{\Lambda}\beta\boldsymbol{\phi}_i^T\boldsymbol{\phi_i})-\frac{1}{s_i}+\beta(\mathbf{y}-\mathbf{\Phi}\mathbf{u})_i^2], \qquad \text{(B.11)}
\end{aligned}$$

where $\boldsymbol{\phi_i}$ is the i th row vector of $\mathbf{\Phi}$, $(\mathbf{y} - \mathbf{\Phi u})_i$ is the i-th element of the estimated error $\mathbf{r}_b = \mathbf{y} - \mathbf{\Phi u}$.

Setting the above equation to zero, we get the estimate of s_i:

$$s_i = \frac{1}{\text{trace}(\mathbf{\Lambda}\beta\boldsymbol{\phi}_i^T\boldsymbol{\phi_i}) + \beta(\mathbf{y} - \mathbf{\Phi u})_i^2} \tag{B.12}$$

The derivative of Eq. (B.6) with respect to β is:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta} &= -\frac{1}{2}[\text{trace}(\mathbf{\Lambda}\frac{\partial \mathbf{\Lambda}^{-1}}{\partial \beta}) - \frac{T}{\beta} + (\mathbf{y} - \mathbf{\Phi u})^T\mathbf{S}(\mathbf{y} - \mathbf{\Phi u})] & \text{(B.13)} \\ &= -\frac{1}{2}[\text{trace}(\mathbf{\Lambda\Phi}^T\mathbf{S\Phi}) - \frac{T}{\beta} + (\mathbf{y} - \mathbf{\Phi u})^T\mathbf{S}(\mathbf{y} - \mathbf{\Phi u})]. & \text{(B.14)} \end{aligned}$$

Setting this to zero gives the estimate of β:

$$\beta = \frac{T}{\text{trace}(\mathbf{\Lambda\Phi}^T\mathbf{S\Phi}) + (\mathbf{y} - \mathbf{\Phi u})^T\mathbf{S}(\mathbf{y} - \mathbf{\Phi u})}, \tag{B.15}$$

where

$$\mathbf{S} = \text{diag}^{-1}(s_1, s_2, \cdots, s_T). \tag{B.16}$$

B.3 A special case

The above derivation considers the situation where noise $\boldsymbol{\epsilon}$ in

$$\mathbf{y} = \mathbf{\Phi w} + \boldsymbol{\epsilon} \tag{B.17}$$

is assumed as $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{B}^{-1})$ which is a Gaussian noise with zero mean and diagonal precision matrix $\mathbf{B} = \text{diag}^{-1}\{(s_1, s_2, \cdots, s_T)\beta\} = \mathbf{S}\beta$ (see also Eq. (3.2)).

A special case is when $\mathbf{S} = \mathbf{I}$ (identity matrix), and thus $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I})$. The Equation (B.9) still holds and represented again here:

$$\alpha_i = \frac{1}{\mathbf{\Lambda}_{ii} + u_i^2}. \tag{B.18}$$

If we define the quantities

$$\gamma_i = 1 - \alpha_i \mathbf{\Lambda}_{ii}, \tag{B.19}$$

and cancel the the parameter $\mathbf{\Lambda}_{ii}$, the following updating equation is obtained:

$$\alpha_i^{new} = \frac{\gamma_i}{u_i^2}. \tag{B.20}$$

By substituting $\mathbf{S}$ with identity matrix $\mathbf{I}$ in Equation (B.14), the derivative of $\mathcal{L}$ respective to β is

$$\frac{\partial \mathcal{L}}{\partial \beta} = -\frac{1}{2}[\text{trace}(\mathbf{\Lambda}\mathbf{\Phi}^T\mathbf{\Phi}) - \frac{T}{\beta} + (\mathbf{y} - \mathbf{\Phi}\mathbf{u})^T(\mathbf{y} - \mathbf{\Phi}\mathbf{u})]. \tag{B.21}$$

From Eq. (A.5), it can be obtained that $\mathbf{\Phi}^T\mathbf{\Phi} = \beta^{-1}(\mathbf{\Lambda}^{-1} - \mathbf{A})$, and thus:

$$\begin{aligned} \text{trace}(\mathbf{\Lambda}\mathbf{\Phi}^T\mathbf{\Phi}) &= \text{trace}(\mathbf{\Lambda}\beta^{-1}(\mathbf{\Lambda}^{-1} - \mathbf{A})) \\ &= \text{trace}(\beta^{-1}(\mathbf{I} - \mathbf{\Lambda}\mathbf{A})) \\ &= \beta^{-1}\sum(1 - \alpha_i\mathbf{\Lambda}_{ii}) \\ &= \beta^{-1}\sum_i \gamma_i \end{aligned} \tag{B.22}$$

where γ_i is defined in Eq. (B.19).

Substituting Eq. (B.22) into Eq. (B.21) and noting that $(\mathbf{y} - \mathbf{\Phi}\mathbf{u})^T(\mathbf{y} - \mathbf{\Phi}\mathbf{u}) = \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|^2$, Eq. (B.21) is simplified as:

$$\frac{\partial \mathcal{L}}{\partial \beta} = -\frac{1}{2}[\beta^{-1}\sum_i \gamma_i - \frac{T}{\beta} + \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|^2]. \tag{B.23}$$

Setting the above derivative to zero gives the update equation:

$$\beta = \frac{T - \sum_i \gamma_i}{\|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|^2}. \tag{B.24}$$

Thus, we obtain the update equations for a special case where $\mathbf{B} = \beta\mathbf{I}$. Eq. (B.19), Eq. (B.20) and Eq. (B.24) are respectively Eq. (2.12), Eq. (2.11) and Eq. (2.13) introduced in Chapter 2.

Appendix C

Derivation of Eq. (4.14)

We represent the observation equation of Eq. (4.12)

$$\begin{pmatrix} \hat{w}_{J_0-1} \\ \hat{w}_{J_0-2} \\ \vdots \\ \hat{w}_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} w + \mathbf{n} \tag{C.1}$$

as

$$\hat{\mathbf{w}} = \mathbf{h}w + \mathbf{n} \tag{C.2}$$

where $\mathbf{h} = [1, 1, \cdots, 1]^T$ is the observation vector, w is the unknown parameter, $\hat{\mathbf{w}} = [\hat{w}_{J_0-1}, \hat{w}_{J_0-2}, \cdots, \hat{w}_1]^T$ is the measured data. $\mathbf{n}$ is the additive measurement noise and is Gaussian distributed $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_{J_0})$ and $\mathbf{\Lambda}_{J_0}$ is defined as:

$$\mathbf{\Lambda}_{J_0} = \begin{pmatrix} \Lambda_{J_0-1} & & & \\ & \Lambda_{J_0-2} & & \\ & & \ddots & \\ & & & \Lambda_1 \end{pmatrix}. \tag{C.3}$$

The likelihood function of $\hat{\mathbf{w}}$ given the unknown parameter w is:

$$\begin{aligned} f(\hat{\mathbf{w}}|w) &= f(\hat{\mathbf{w}} - \mathbf{h}w) \\ &= f(\mathbf{n}) \\ &= (2\pi)^{-\frac{J_0}{2}} |\mathbf{\Lambda}_{J_0}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\hat{\mathbf{w}} - \mathbf{h}w)^T \mathbf{\Lambda}_{J_0}^{-1} (\hat{\mathbf{w}} - \mathbf{h}w)\}. \end{aligned} \tag{C.4}$$

It is clear that maximizing the above likelihood function is equivalent to minimize the following cost function:

$$J = \frac{1}{2}(\hat{\mathbf{w}} - \mathbf{h}w)^T \mathbf{\Lambda}_{J_0}^{-1}(\hat{\mathbf{w}} - \mathbf{h}w). \tag{C.5}$$

By differentiating the cost function with respect to w and the let the result to be zero, we get:

$$\frac{\partial J}{\partial w} = \mathbf{h}^T \mathbf{\Lambda}_{J_0}^{-1}(\hat{\mathbf{w}} - \mathbf{h}w) = 0. \tag{C.6}$$

Then, we get the maximum likelihood estimation of the unknown parameter w as:

$$\hat{w}_{ML} = (\mathbf{h}^T \mathbf{\Lambda}_{J_0}^{-1}\mathbf{h})^{-1}\mathbf{h}^T \mathbf{\Lambda}_{J_0}^{-1}\hat{\mathbf{w}} \tag{C.7}$$

By substituting $\mathbf{h} = [1, 1, \cdots, 1]^T$, $\hat{\mathbf{w}} = [\hat{w}_{J_0-1}, \hat{w}_{J_0-2}, \cdots, \hat{w}_1]^T$ and $\mathbf{\Lambda}_{J_0}$ shown in Eq. (C.1) into Eq. (C.7), we will get:

$$\hat{w}_{ML} = \sum_{i=1}^{J_0-1} \frac{\Lambda_i^{-1}}{\sum_{k=1}^{J_0-1} \Lambda_k^{-1}} \hat{w}_i, \tag{C.8}$$

which is Eq. (4.14).

Appendix D

Papers Originated from this Work

Published/Accepted

1. Huaien Luo and S. Puthusserypady, “Analysis of fMRI Data with Drift: Modified General Linear Model and Bayesian Estimator,” *IEEE Transactions on Biomedical Engineering*, (to appear), 2008. [Chapter 4]

2. Huaien Luo and S. Puthusserypady, “Estimation of the hemodynamic response of the fMRI data using RBF neural networks,” *IEEE Transactions on Biomedical Engineering*, Vol. 54, pp. 1371-1381, 2007. [Chapter 6]

3. Huaien Luo and S. Puthusserypady, “fMRI data analysis with nonstationary noise models: A Bayesian approach,” *IEEE Transactions on Biomedical Engineering*, Vol. 54, pp. 1621-1630, 2007. [Chapter 3]

4. Huaien Luo and S. Puthusserypady, “Adaptive Spatiotemporal Modelling and Estimation of the Event-related fMRI Responses,” *Signal Processing*, Vol. 87, pp. 2810-2822, 2007. [Chapter 5]

5. Huaien Luo and S. Puthusserypady, “Spatio-temporal modeling and analysis of

fMRI data using NARX neural network," (Invited Paper) *International Journal of Neural Systems*, Vol.16, No.2, pp.139-149, 2006. [Chapter 7]

6. Huaien Luo and S. Puthusserypady, "A sparse Bayesian method for determination of flexible matrix for fMRI data analysis," *IEEE Transactions on Circuits and Systems, Part I Fundamental Theory and Applications*, Vol.52, No.12, pp.2699-2706, December 2005. [Chapter 2]

7. Huaien Luo and S. Puthusserypady, "Neural Networks for fMRI Spatio-Temporal Analysis," In *Proceeding of 11th International Conference on Neural Information Processing (ICONIP 2004)*, November 22-25 2004, Calcutta, India.
   Also published in *Lecture Notes in Computer Science*, Vol. 3316, pp. 1292-1297, Springer, 2004. [Chapter 7]

8. Huaien Luo and S. Puthusserypady, "Bayesian RBF network for modeling fMRI data," In *Proceeding of the 26th International Conference IEEE Engineering in Medicine and Biology Society (EMBC2004)*, September 1-5, 2004, San Francisco, USA. [Chapter 2]

9. Huaien Luo and S. Puthusserypady, "NARX neural networks for dynamical modeling of fMRI data," In *Proceedings of the International Joint Conference on Neural Networks, IEEE World Congress on Computational Intelligence (WCCI-2006)*, Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada, July 16-21, 2006. [Chapter 7]