

MODELING AND ESTIMATION OF SIGNALS AND ARTIFACTS IN
FUNCTIONAL MAGNETIC RESONANCE IMAGING AND
COMPUTED TOMOGRAPHY

by

Negar Bazargani

APPROVED BY SUPERVISORY COMMITTEE:

Dr. Aria Nosratinia, Chair

Dr. John H. L. Hansen

Dr. Nasser Kehtarnavaz

Dr. Issa M. S. Panahi

Copyright © 2014

Negar Bazargani

All rights reserved

To My Family

MODELING AND ESTIMATION OF SIGNALS AND ARTIFACTS IN
FUNCTIONAL MAGNETIC RESONANCE IMAGING AND
COMPUTED TOMOGRAPHY

by

NEGAR BAZARGANI, BS, MS

DISSERTATION

Presented to the Faculty of
The University of Texas at Dallas
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY IN
ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

December 2014

ACKNOWLEDGMENTS

I would like to thank everyone who supported and helped me through the progress and completion of this dissertation.

I would like to express my appreciation and special thanks to my advisor, Dr. Aria Nosratinia, for his continuous advice, guidance and support throughout my graduate studies at UTD.

I would like to thank Dr. Nasser Kehtarnavaz, Dr. Issa Panahi and Dr. John Hansen for kindly serving on my supervisory committee, for giving their precious time to read my dissertation and for providing valuable feedback and comments on this work.

The work on baseline drift was supported jointly by the UTD Erik Jonsson School of Engineering and Computer Science and by the Department of Internal Medicine, University of Texas Southwestern Medical Center at Dallas. I would like to thank Dr. Richard Briggs and Dr. Kaundinya Gopinath.

I would like to thank my parents, my sister, my aunts, my brother in law and my cousins Peyman and Lesa for their love and support.

My special thanks to Dr. Thomas Morton, my colleague at Philips Healthcare, for his support and encouragement during this work. I appreciate his time, guidance and help.

I also acknowledge and thank my colleagues at the Multimedia Communications Lab for their support and encouragement. I would like to thank my friends Azadeh, Oldooz, Omid, Yasaman and Ali for their support.

September 2014

PREFACE

This dissertation was produced in accordance with guidelines which permit the inclusion as part of the dissertation the text of an original paper or papers submitted for publication. The dissertation must still conform to all other requirements explained in the “Guide for the Preparation of Master’s Theses and Doctoral Dissertations at The University of Texas at Dallas.” It must include a comprehensive abstract, a full introduction and literature review, and a final overall conclusion. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported.

It is acceptable for this dissertation to include as chapters authentic copies of papers already published, provided these meet type size, margin, and legibility requirements. In such cases, connecting texts which provide logical bridges between different manuscripts are mandatory. Where the student is not the sole author of a manuscript, the student is required to make an explicit statement in the introductory material to that manuscript describing the student’s contribution to the work and acknowledging the contribution of the other author(s). The signatures of the Supervising Committee which precede all other material in the dissertation attest to the accuracy of this statement.

MODELING AND ESTIMATION OF SIGNALS AND ARTIFACTS IN
FUNCTIONAL MAGNETIC RESONANCE IMAGING AND
COMPUTED TOMOGRAPHY

Publication No. _____

Negar Bazargani, PhD
The University of Texas at Dallas, 2014

Supervising Professor: Dr. Aria Nosratinia

The objective of this dissertation is to enhance and improve two medical imaging modalities: functional Magnetic Resonance Imaging (fMRI) and x-ray Computed Tomography (CT). For image analysis in both modalities accurate modeling of the signal is important as it affects the analysis and image quality. Several components of fMRI have resisted precise modeling, due to the complexity of the signals, their underlying causes and their variations across subjects and brain regions. We introduce a maximum likelihood method that jointly estimates the Hemodynamic Response Function (HRF) and activation level in fMRI, with a regularization that allows our method to work with smaller regions of interest and therefore to better capture the variations of the HRF across the areas of the brain. Improvements in modeling achieved via the proposed methods have direct impact on enhancing the activation detection, which is a primary goal in the fMRI analysis. We also introduce extracting the general structure of the signals, such as activation level, baseline drift and the hemodynamic response function, from each data set itself using the Minimum Description Length (MDL)

principle. The integration of the MDL principle with modeling the fMRI signal provides paradigm-shifting advances in the fMRI analysis. Results show that MDL-based methods offer benefits in comparison with other methods. In Computed Tomography the scatter radiation from the dose compensator causes artifacts in the image. This signal is hard to measure and it varies between each detector. This scattered radiation impacts high contrast edges such as bone-soft tissue and tissue-air interfaces, especially in brain imaging. We use Monte Carlo (MC) simulation for modeling this nuisance component of the CT data. We implement a variance reduction technique called forced detection (FD) to improve the computational efficiency. A correction algorithm for the scatter radiation is implemented which improves the image quality.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
PREFACE	vi
ABSTRACT	vii
LIST OF FIGURES	xii
LIST OF TABLES	xvi
LIST OF ABBREVIATIONS	xvii
CHAPTER 1 INTRODUCTION	1
1.1 Functional MRI	1
1.1.1 Overview	1
1.1.2 Background and Literature Review on Drift Model	3
1.1.3 Background and Literature Review on HRF Model	5
1.1.4 Motivation	8
1.1.5 Contributions	9
1.2 Computed Tomography	11
1.2.1 Overview and Motivation	11
1.2.2 Background on CT dose compensator scatter modeling and Forced Detection	13
1.2.3 Contribution	14
1.3 Organization	14
CHAPTER 2 FUNCTIONAL MRI AND MINIMUM DESCRIPTION LENGTH . .	16
2.1 Minimum Description Length principle	16
2.1.1 Why use MDL for fMRI?	19
2.2 MDL-based estimation of fMRI drift and activation level	20
2.2.1 Algorithm	20
2.2.2 Results	23

2.3	MDL-based estimation of HRF and activation level	32
2.3.1	Algorithm	32
2.3.2	Results	35
2.4	Conclusion	37
CHAPTER 3 FMRI: JOINT MAXIMUM LIKELIHOOD ESTIMATION OF HRF AND ACTIVATION LEVEL		40
3.1	Overview	40
3.2	Methods	43
3.2.1	Formulation	43
3.2.2	Rank-one Approximation in White Noise	45
3.2.3	Regularized Rank-one Approximation in White Noise	46
3.2.4	Rank-one Approximation in Colored Noise	50
3.2.5	Regularized Rank-one Approximation in Colored Noise	51
3.2.6	Iterative Estimation	51
3.2.7	Extension to Multiple Stimuli	53
3.2.8	Region of Interest Selection	53
3.3	Results	54
3.3.1	Simulated Data	54
3.3.2	Real fMRI Data: Event Related experiment	67
3.3.3	Real fMRI Data: Block Design experiment	74
3.4	Conclusion	77
CHAPTER 4 COMPUTED TOMOGRAPHY: MODELING AND CORRECTION FOR SCATTER RADIATION FROM DOSE COMPENSATOR		80
4.1	Background	80
4.2	Monte Carlo simulation	82
4.2.1	Conventional Monte Carlo	83
4.2.2	Variance reduction by Forced Detection technique	84
4.3	Dose compensator scatter kernel	88
4.4	Results	89
4.5	Conclusion	92

CHAPTER 5 CONCLUSION	94
5.1 Contributions	94
REFERENCES	96
VITA	

LIST OF FIGURES

1.1	Patient in MRI scanner with a block design stimulus, structural data and functional data across time, and activation map after the statistical analysis.	2
1.2	Summary of the fMRI time series model for each voxel.	3
1.3	fMRI time series of 9 neighboring voxels. The baseline drift varies between neighboring voxels.	4
1.4	Example of the hemodynamic response function with parameters defined in [41].	6
1.5	Schematic diagram of a dose compensator and scattered radiations from the patient (red long dashed line) and from the dose compensator (blue dashed line). The black dash-dot line shows a ray without scattering.	12
2.1	Comparing estimated drift from different methods with the benchmark drift for a block design simulated data with SNR=0.8. The MSE between estimated and benchmark drifts is shown for each method.	25
2.2	The mean square error between the benchmark and estimated drift for (a) a block design and (b) an event related experiment.	27
2.3	Histogram of autocorrelation of residual error at lag one for all voxels for a block design experiment.	28
2.4	Axial, Sagittal and Coronal views of a single slice activation maps. (a) benchmark activation, (b) MDL-based detrending, (c) No detrending, (d) spline detrending and (e) polynomial of order 3 detrending.	29
2.5	Axial and Coronal views of a single slice activation map of real fMRI data for (a) AFNI, (b) MDL-based detrending, (c) No detrending, (d) spline detrending and (e) polynomial of order 3 detrending.	31
2.6	Square error between drift estimates across iterations for one voxel of a real fMRI dataset.	32
2.7	The benchmark, Gamma, Poisson and estimated HRFs for an event related stimuli in a ROI with 27 active voxels.	36
2.8	Comparing t-statistics obtained by benchmark, Gamma, Poisson and estimated HRFs.	36

2.9	Axial and Coronal views of a single slice activation maps for MDL-based HRF estimation and AFNI. (a) HRF from MDL-based estimation for ROI 1 (on the left side of activation maps), (b) HRF from MDL-based estimation for ROI 2 (on the right side of activation maps), (c) MDL-based activation maps, (d) AFNI activation maps, (e) HRF from AFNI for one voxel in ROI 1, (f) HRF from AFNI for one voxel in ROI 2.	38
3.1	Average of the HRF estimate and the variance of the estimate at each point on a ROI with 50 voxels, with SNR = 0.2 for (a) a block design experiment without regularization, (b) a block design experiment with regularization.	56
3.2	Average of the HRF estimate and the variance of the estimate at each point on a ROI with 50 voxels, with SNR = 0.2 for (a) an event related experiment without regularization and (b) an event related experiment with regularization.	57
3.3	Mean square error between the benchmark HRF and the estimated HRF with different SNRs for a block design experiment with white noise assumption for (a) M=50 and (b) M=100 voxels.	59
3.4	Mean square error between the benchmark HRF and the estimated HRF with different SNRs for an event related experiment with white noise assumption for (a) M=50 and (b) M=100 voxels.	60
3.5	Effect of regularization constraint on MSE between the benchmark HRF and the estimated HRF with different SNRs for a block design experiment on (a) M=50 and (b) M=100 voxels.	61
3.6	Effect of regularization constraint on MSE between the benchmark HRF and the estimated HRF with different SNRs for an event related experiment on (a) M=50 and (b) M=100 voxels.	62
3.7	Comparing the estimated activation levels by imposing regularization for a block design experiment with 100 voxels in the ROI and SNR=0.5. (a) Absolute value of the difference between estimated and benchmark activation levels. (b) Variance of the activation levels estimates.	63
3.8	T-tests calculated with the benchmark correlation coefficient, compared with the t-tests calculated with and without whitening.	66
3.9	(a) HRF estimates and (b) t-tests for a ROI with 50 voxels including 40 active voxels. One realization of event related with SNR=0.2 and white noise assumption.	68
3.10	Effect of the t-test threshold on the estimated HRF, for the iterative algorithm in a ROI with 40 active and 10 inactive voxels for (a) a block design experiment with SNR = 0.8, (b) an event related experiment with SNR = 0.2.	69
3.11	Axial and coronal views of a single slice activation maps for ROI 1. (a) activation maps from the MLE method, (b) estimated HRF from MLE method, (c) ROI obtained from AFNI, (d) estimated HRF from AFNI.	71

3.12	Axial and coronal views of a single slice activation maps for ROI 2. (a) activation maps from the MLE method, (b) estimated HRF from MLE method, (c) ROI obtained from AFNI, (d) estimated HRF from AFNI.	72
3.13	Axial and coronal views of a single slice activation maps for ROI 3. (a) activation maps from the MLE method, (b) estimated HRF from MLE method, (c) ROI obtained from AFNI, (d) estimated HRF from AFNI.	73
3.14	Coronal and sagittal views of a single slice activation maps. (a) activation maps from the MLE method without prior knowledge on active ROI, (b) estimated HRF from MLE method, (c) activation map based on AFNI 3dDeconvolve algorithm, (d) estimated HRF from AFNI.	75
3.15	Fixation (Stimulus 1) and Attention (Stimulus 2) conditions over the duration of the scans. Amplitude of 1 represents the stimulus was on.	76
3.16	Axial and Sagittal views of a single slice activation maps for the block design attention to visual stimuli data set. Activation maps are for Stimulus 1. (a) activation maps from the MLE method without prior knowledge on active ROI, (b) estimated HRF from MLE method, (c) activation map based on AFNI 3dDeconvolve algorithm, (d) estimated HRF from AFNI, (e) activation maps from the MLE method with ROI from AFNI, (f) estimated HRF from MLE method with ROI from AFNI.	78
4.1	Example of an x-ray tube output energy spectrum at 120kVp.	82
4.2	Photon tracks for conventional Monte Carlo (a) and with application of forced detection (b). The path for the photon emitted from the source is shown in solid black line. The red and cyan dotted arrows show forced photons from two scattered events which are forced to travel in the detector directions.	85
4.3	The ratio of primary with the dose compensator to the primary without the dose compensator for the FD, conventional MC, and analytic model are plotted for all detectors.	88
4.4	Dose compensator scatter kernel for one of the detectors with respect to off-focal angles in X and Z direction.	89
4.5	Effect of collimation on the scatter to primary ratio across the detectors for 8×0.625 , 16×0.625 , 32×0.625 , 64×0.625 and 128×0.625 mm collimations.	90
4.6	Slice 1 of the (a) simulated data, (b) added dose compensator scatter radiation, (c) corrected for compensator scatter. (d) The horizontal profiles. The difference between original and added scatter is shown in (e), the difference between original and corrected image is in (f).	91

4.7	Simulated uncentered 20 cm diameter water phantom. (a) Horizontal profiles for original, added scatter with amplitude 1%, corrected for dose compensator, (b) difference between the original and added dose compensator scatter radiation, (c) difference between the original and corrected images and (d) and (e) Vertical profiles for original, added scatter with amplitude 1%, corrected for dose compensator zoomed in top and bottom section of image.	93
-----	---	----

LIST OF TABLES

3.1	Comparison of SPM, AFNI and proposed joint MLE	42
3.2	Monte Carlo comparison between SPM and joint MLE. Average of mean square error between benchmark and estimated values are shown for block and event related designs for a ROI with 100 voxels.	64

LIST OF ABBREVIATIONS

AFNI	Analysis of Functional NeuroImages
AR	Auto Regressive
ARMA	Auto Regressive Moving Average
BOLD	Blood Oxygenation Level Dependent
CBCT	Cone Beam Computed Tomography
CT	Computed Tomography
DCT	Discrete Cosine Transform
fMRI	functional Magnetic Resonance Imaging
FBP	Filtered Back Projection
FD	Forced Detection
FIR	Finite Impulse Response
GEANT	GEometry ANd Tracking
GLM	General Linear Model
HRF	Hemodynamic Response Function
ICA	Independent Component Analysis
ISI	Inter Stimulus Intervals
MC	Monte Carlo
MDL	Minimum Description Length
MLE	Maximum Likelihood Estimation
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
NML	Normalized Maximum Likelihood

PCA	Principle Component Analysis
ROI	Region Of Interest
SNR	Signal to Noise Ratio
SPECT	Single Photon Emission Computed Tomography
SPM	Statistical Parametric Mapping
SPR	Scatter to Primary Ratio
SVD	Singular Value Decomposition
TR	Repetition Time

CHAPTER 1

INTRODUCTION

This dissertation addresses some of the key challenges in modeling signal components of functional Magnetic Resonance Imaging (fMRI) and x-ray Computed Tomography (CT) imaging. Modeling of signal and artifact components affect image analysis and activation detection in fMRI and image quality in CT. The overview, motivations and challenges in each modality are covered in this chapter.

1.1 Functional MRI

1.1.1 Overview

Functional Magnetic Resonance Imaging is considered to be one of the most popular brain imaging modalities because it is a powerful tool to non-invasively study the relationship between cognitive stimulus and the hemodynamic response of the brain in real time. fMRI uses Magnetic Resonance Imaging (MRI) technology and works based on the increase in blood flow to the local vasculature that accompanies neural activity in the brain. This results in a corresponding local reduction in deoxyhemoglobin because the increase in oxygenated blood flow occurs without an increase of similar magnitude in oxygen extraction. One of the main types of fMRI employs the blood oxygenation level dependent (BOLD) contrast, which is based on the differing magnetic properties of oxygenated (diamagnetic) and deoxygenated (paramagnetic) blood. The focus in this dissertation is on BOLD fMRI signal. Since deoxyhemoglobin is paramagnetic, it alters the magnetic resonance signal [73]. These magnetic susceptibility differences lead to small, but detectable changes in susceptibility-weighted MR image intensity. Relatively low image signal to noise ratio (SNR) of the BOLD effect, head

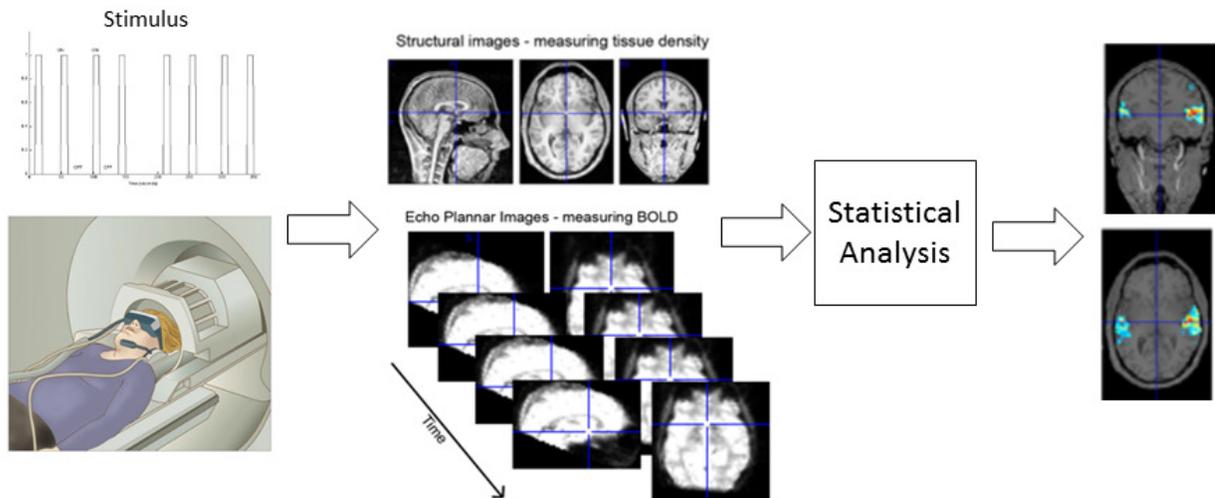


Figure 1.1. Patient in MRI scanner with a block design stimulus, structural data and functional data across time, and activation map after the statistical analysis.

movement, and undesired physiological sources of variability (cardiac, pulmonary) make detection of the activation-related signal changes difficult. The BOLD signal can be detected by appropriate design of experiments and imaging the brain during rest and activation. Rapid image acquisition techniques can be used to generate data sets with hundreds of images for each slice location which can be statistically analyzed to determine brain functional activation map.

Clinical applications of fMRI include neuropsychological studies of behavior, networks for cognitive processing, mapping of functional areas in neuropathology and brain surgery planning.

Figure 1.1 shows a patient in a MRI scanner, example of block design stimulus, the output BOLD and structural images and brain activation map. fMRI stimuli are either block paradigms, with alternating periods of action (activation) and rest (baseline) conditions each typically having duration of about 10-30 seconds, or event-related paradigms, with short periods of activation and longer periods of rest.

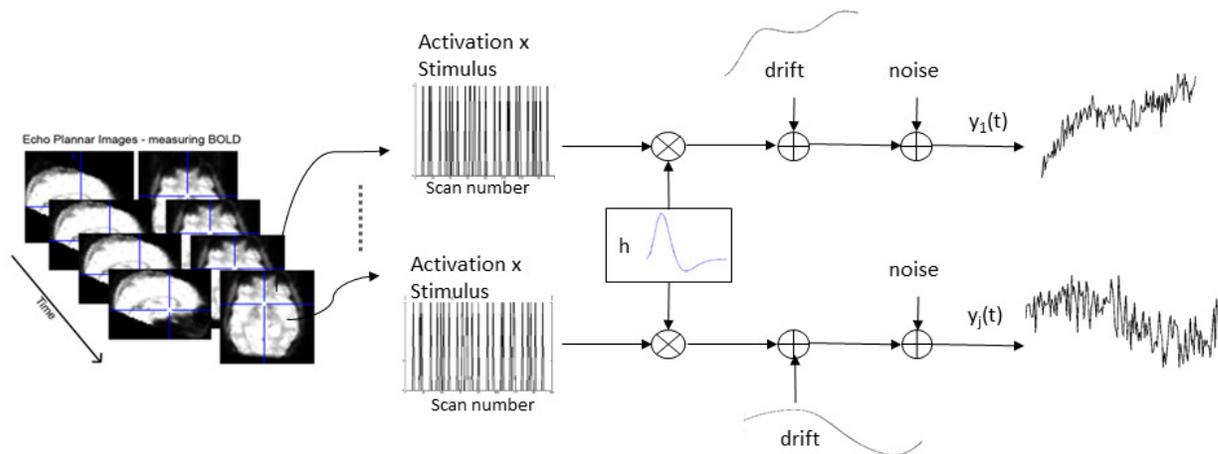


Figure 1.2. Summary of the fMRI time series model for each voxel.

The fMRI time series model at each voxel is shown in Figure 1.2. Stimulus is multiplied by activation level and is convolved with the hemodynamic response function (HRF), shown as h , then drift and noise are added to the signal.

The drift and hemodynamic response function and their background are defined in Sections 1.1.2 and 1.1.3. The proposed methods to model these components are covered in Chapter 2, Sections 2.2 and 2.3 and Chapter 3.

1.1.2 Background and Literature Review on Drift Model

Slow variations over time in the fMRI data are known as drift. Some causes of the drift are uncorrected head movement and instrumental instability. Drift has low frequencies of about 0-0.015 Hz and can have maximum amplitude of approximately 5% of the baseline signal intensity [85]. The amount and direction of the baseline drift is difficult to predict even between neighboring voxels [87]. Figure 1.3 shows the fMRI time series for 9 neighboring voxels, the pattern of the baseline drift is changing between neighboring voxels. The mean of each time series is subtracted to better show the variation in the same axis. Because of the low signal to noise ratio of the fMRI signal, estimating and removing the drift component can have a strong impact on the statistical analysis and accurate activation detection [87].

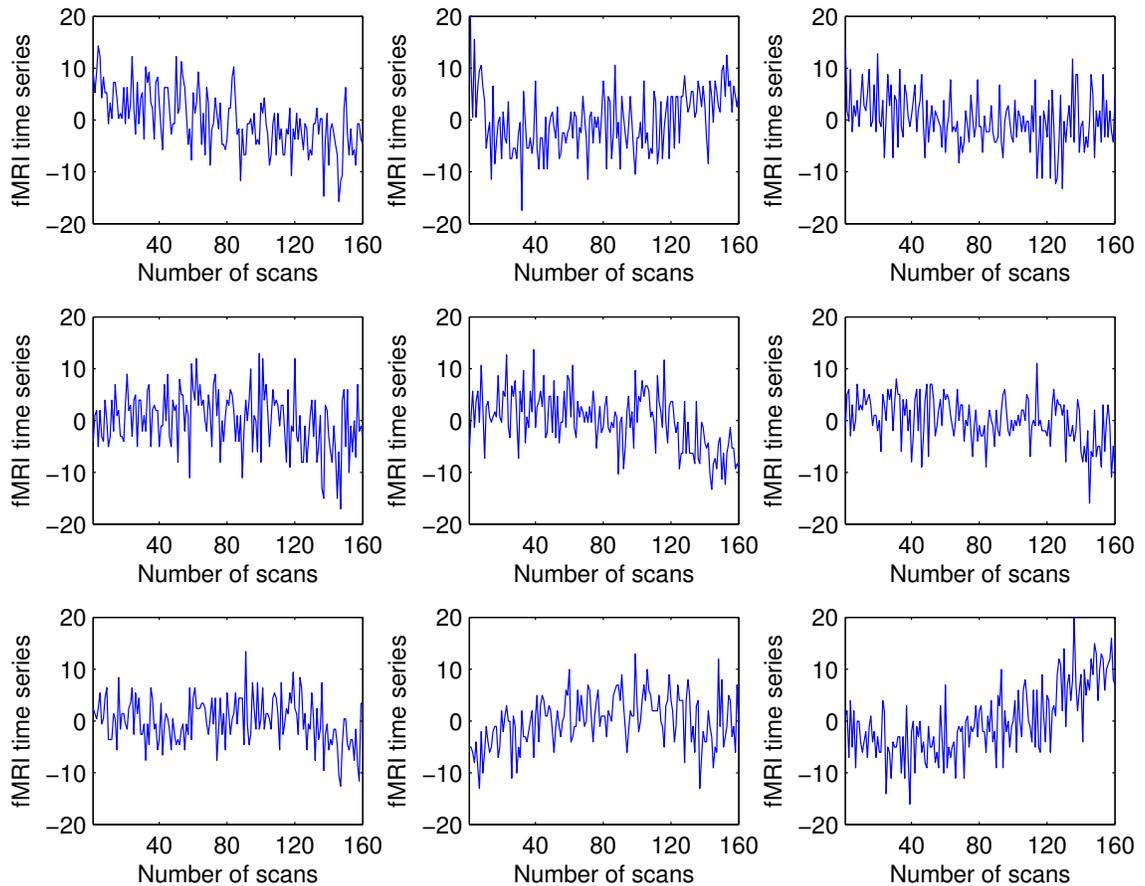


Figure 1.3. fMRI time series of 9 neighboring voxels. The baseline drift varies between neighboring voxels.

The two most popular methods for drift estimation and removal have been low pass filtering and subspace methods.

Low pass filters, autoregressive filters, Kalman filters [56] and nonlinear low pass filters such as those suggested in [52] can be used for drift estimation. The filtering methods have the advantage of simplicity but they face the problem of determining the appropriate cutoff frequency. This is very challenging, because one needs to be very careful not to remove the stimulus frequencies. Furthermore, it is not even clear that one may be able

to sufficiently estimate the drift via low pass filters without harming activation analysis, because the separation of stimulus and drift signals may not be well described by a single cut-off frequency.

Another technique for estimating drift is to assume that a certain subspace can represent the drift. Several works model drift by assuming a predefined basis, such as linear functions of time [6, 14, 62] and low order polynomials [41, 84], set of vectors like polynomials and harmonics of sine and cosine [25, 24], discrete cosine transform (DCT) [48, 38] and large scale wavelets [71, 32]. There are other techniques that use exploratory analysis like principal component analysis, independent component analysis (ICA), clustering method or canonical correlation analysis to find the basis for the drift component [5, 35]. In [40, 66] smoothing splines and running lines smoothers are used for modeling the drift.

Subspace methods are, in general, more powerful than filtering methods, but they also have certain shortcomings. The main difficulty is that our understanding of the origin of drift does not provide us with a specific function or a parametric model of drift. Moreover, most of the existing drift estimation methods have user defined parameters. In general, subspace models assume that the drift component belongs completely to a given subspace, and this might not be true. There is growing evidence that detrending (drift removal) techniques should be adapted to the subject or even the voxel being analyzed, but at the moment there is no consensus as to which detrending approach should be used [86].

1.1.3 Background and Literature Review on HRF Model

The most precise approaches to modeling the link between stimulus and the induced BOLD activity, are physiological nonlinear models [15, 39, 77] but because of their computational cost and their identifiability issues their use is restricted to a number of specific regions and to a few experimental conditions [18]. In contrast, the common approach, being the focus of this dissertation, relies on linear systems which appear more robust and tractable.

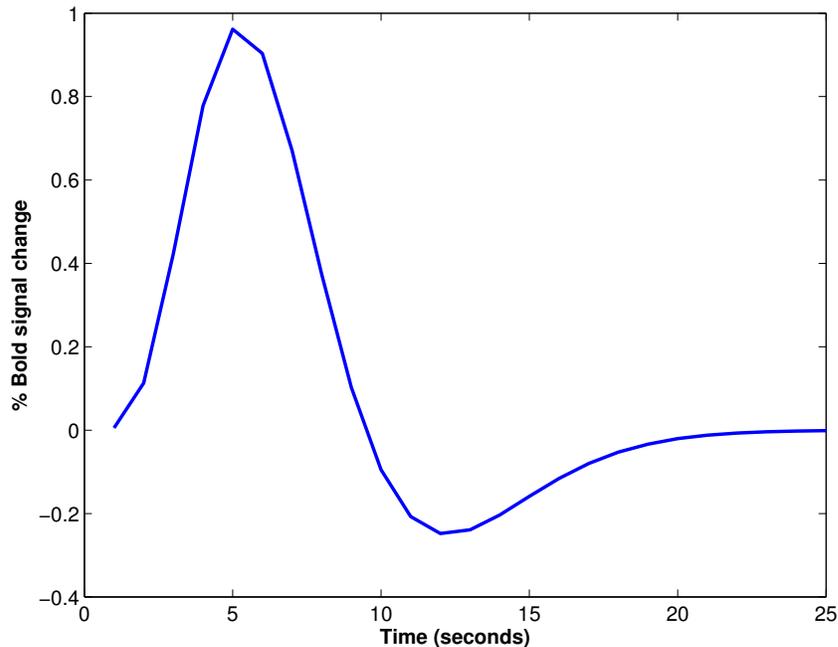


Figure 1.4. Example of the hemodynamic response function with parameters defined in [41].

It has been observed that the neurovascular system is well approximated by a linear and stationary model if the inter-stimulus interval (ISI) is more than a few seconds [12, 26]. Using a linear time invariant model the hemodynamic changes in BOLD fMRI can be characterized as the convolution of the stimulus with the impulse response function of the system, which is called the hemodynamic response function. Figure 1.4 shows the hemodynamic response function as described by [41]. The hemodynamic response starts immediately after the stimulus is applied, and experiences a peak after 3-10 seconds delay. There is a post stimulus undershoot and then it returns to its baseline [52].

HRF estimation is important in two aspects, estimation of temporal dynamics and detection of where activity occurs. It is well known that the HRF varies across different regions of the brain, different subjects and tasks [4]. Knowledge of the hemodynamic response function is necessary for the estimation of activation levels, which are the responses at each brain locality to a given stimulus. In the absence of exact prior knowledge, the estimation of HRF

and activation levels are inter-related: a precise HRF model results in better activation detection, but at the same time active voxels need to be known in order to estimate the HRF, since inactive voxels do not carry any information that is useful for HRF estimation, and their inclusion will introduce noise into the process.

Methods for estimating HRF can be categorized by the amount of assumptions on the shape of the HRF. At one extreme, one may assume a pre-determined shape for the HRF with varying magnitude. Parametric models fit in this category. Among the parametric models one may name the Poisson density [36], the Gaussian function [76], the Gamma density [21, 57], shifted Gamma density [12], difference of two Gamma functions [41] and truncated Gaussian functions [44]. Parametric models may introduce bias in the HRF estimate, since they have limited ability to capture HRF shape variability across voxels.

On the other extreme, one may assume no prior knowledge about the shape, thus the model would have as many free parameters as the data points [41, 45, 74], and is able to estimate HRF with any arbitrary shape. The nonparametric models fall in this category. This may seem like a superior method, however, its shortcoming is sensitivity to measurement noise. For that reason, the latter methods need a much larger data set to train on. Bayesian nonparametric estimations of the HRF were proposed in [20] and [67]. Afonso et al. [2] also used a Bayesian framework to jointly estimate the HRF and activation levels, the HRF was estimated voxel-wise in their work.

In between the two extremes, there are basis-set methods, whose degrees of freedom are fewer than the time samples. Past work that models the HRF as a linear combination of basis functions includes combining canonical shape for the HRF and its temporal derivative [37], spline basis sets, finding basis sets using principal components [4, 90], cosine functions [94], radial basis functions [77], linear combination of Fourier basis functions [53] and spectral basis functions [58]. Lindquist et al. [59] performed a comparative study using various basis functions. There is a trade off between the flexibility of the model and its ability to capture

variabilities. Too little flexibility will not allow adaptation of HRF shape to the variability of brain regions or subjects, while too much flexibility may produce unlikely shapes.

The HRF varies across different regions of the brain. Therefore, assuming any fixed shape for HRF is inaccurate in most parts of the brain [46]. Because of the low signal-to-noise ratio in the fMRI data, voxel-wise estimation methods may be too noisy. It has been observed [20, 43] that for a specific task, the shape of the HRF is constant across neighboring voxels in a homogeneous region [20], thus a more robust HRF estimation may be obtained from the data corresponding to a set of voxels within a neighborhood, instead of a single voxel.

A region-based formulation for the HRF estimation was proposed in [18, 28, 64, 65], where a single HRF shape was used with varying amplitude within a region of interest (ROI). Donnet et al. [28] proposed expectation maximization for estimating HRF and activation levels, they used pre-defined ROI on event-related experiment. Makni et al. and Chaari et al. [64, 65, 18] used a Bayesian framework that requires knowledge of the prior distributions for activation levels, HRF and other parameters. They used a parcellation technique to select active ROIs. Pedregosa et.al [75] used a rank constraint and quasi-Newton method for joint estimation of HRF and activation levels at each voxel.

1.1.4 Motivation

A significant task in fMRI data analysis is to develop activation maps showing a subject's localized brain responses to pre-specified stimuli. fMRI data analysis methods can be categorized as model-driven and data-driven techniques. In model-driven methods each component of the fMRI signal is often modeled separately and these *a priori* models are fitted to the fMRI time series. Some of the fMRI components are the neural response and nuisance terms such as baseline drift and noise. The quality of model-based fMRI data analysis, which includes pre-processing, fitting a model to the data, calculation of statistics and activation

detection, depends on accurate modeling of the data. But due to the complexity of biological systems, and because the models cannot be measured directly, fMRI modeling has been *ad hoc*.

For data-driven methods, no underlying model is assumed and, rather, the method attempts to extract meaningful spatial or temporal trends in the measurement. The data-driven approach is useful when an appropriate model for the data is not available. However, these methods are computationally expensive and their results do not easily lend themselves to interpretation. The independent component analysis and principle component analysis (PCA) [16, 69] are among data-driven approaches.

More specifically, despite their convenience, simple and rigid models do not provide reliable results when the expected neural response to the stimuli is not known *a priori* or when it is varying across the brain, subjects and/or experiments. It is well-known that large qualitative variations are often observed in fMRI signals, which explains the need for adaptive methods. Adaptability to signal variations can be achieved with data-driven methods. But these methods need large data sets to train on and are computationally more expensive than model-based methods.

The large variations of fMRI signal components motivates the need for adaptive methods that can capture variabilities of the HRF and the baseline drift across regions and tasks while not being as computationally demanding as data-driven methods.

1.1.5 Contributions

The contributions of this dissertation on fMRI include adaptive methods to estimate and model the baseline drift and the HRF.

1. In model-based fMRI data analysis, which includes pre-processing, calculation of statistics and activation detection, accurate modeling of the data plays an important role

in final activation map. The model for each component has to be as flexible as possible and be able to capture variabilities across voxels. The minimum description length principle is introduced to model the baseline drift and the hemodynamic response function. In the MDL principle any regularity in the data can be used to compress the data and fMRI models try to match to regularities. So the MDL principle is proposed for modeling and capturing structures and regularities in the fMRI data with as few assumptions as possible. Unlike previous works which applied MDL for model order selection [9, 19, 50], in this dissertation MDL is used as a tool to guide us throughout estimating the fMRI baseline drift and HRF components.

- (a) Estimation and removing the baseline drift artifact in fMRI using the MDL principle: The proposed method is based on the iterative estimation of the activation level and the drift component, using least squares estimation and the concept of MDL denoising. Results show that the proposed algorithm can estimate the activation level and the drift term without having any prior knowledge or assuming an overly restrictive model for the drift. The proposed method captures variability of drift across voxels and improves the statistical analysis results. This is covered in Section 2.2.
- (b) Estimation of the HRF and activation level in fMRI using the MDL principle: Estimating the HRF is important for statistical analysis of the fMRI data and for obtaining information about the temporal dynamics of the brain. HRF models commonly assume varying amounts of prior knowledge about the shape of the HRF. The proposed algorithm is based on the iterative estimation of the HRF and activation level. This provides a nonparametric approach to HRF estimation without any explicit assumptions. The proposed method is capable of capturing variabilities of the HRF across regions and improving statistical analysis. This is discussed in detail in Section 2.3

2. Joint maximum likelihood estimation of the fMRI hemodynamic response function and activation level: Modeling the HRF and estimating the activation level are two important aspects in the statistical analysis of the fMRI signal. These two are intertwined. Knowledge of the HRF is necessary for the reliable estimation of the activation levels in fMRI. Furthermore, estimation of the HRF requires knowledge of activation levels. The cross dependency motivates the joint approach for estimation of these quantities. Our work takes a maximum likelihood approach to this problem and, since it does not require the prior probability distributions of the HRF or the activation levels, it is useful in situations where obtaining reliable and verifiable prior probability distributions on the HRF or the activation levels are difficult. It is known that the shape of the HRF varies across brain regions and subjects. This suggests that a data-driven estimation of this function is statistically more powerful when modeling BOLD fMRI data. A different HRF is estimated depending on the stimulus presented for each ROI. The proposed method has the flexibility that our own algorithm can be used to define the active ROIs without the need to use another algorithm to provide pre-defined ROIs as is the practice in most of the region based analysis methods. This is explained in Chapter 3.

1.2 Computed Tomography

1.2.1 Overview and Motivation

Computed tomography is a medical imaging technique for imaging cross-sections using a series of x-ray projection measurements taken from different angles around the patient. Using mathematical reconstruction algorithms, the volumetric information inside the patient can be reconstructed slice by slice. CT is a valuable medical tool that can help a physician diagnose disease, trauma or abnormality; plan and guide interventional or therapeutic procedures and monitor the effectiveness of therapy (e.g., cancer treatment).

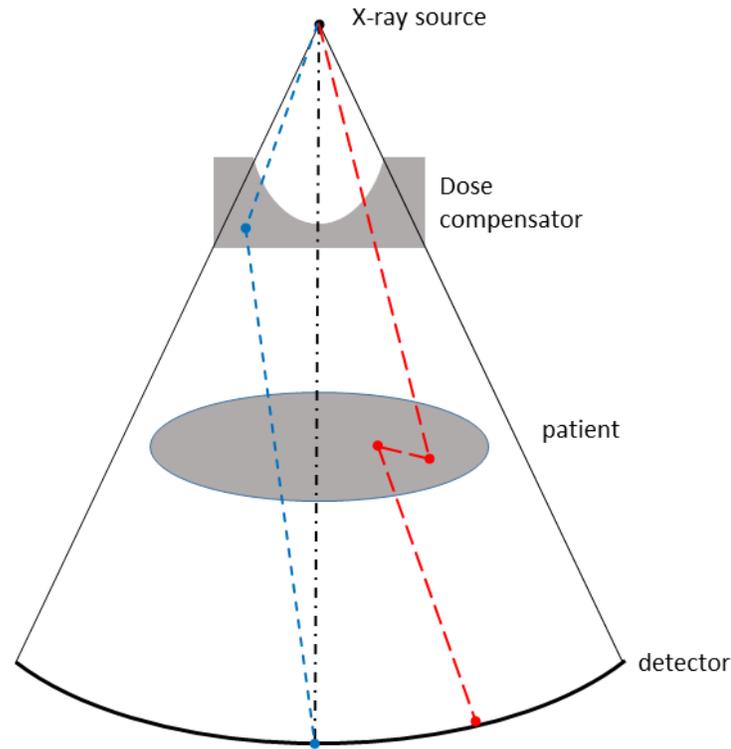


Figure 1.5. Schematic diagram of a dose compensator and scattered radiations from the patient (red long dashed line) and from the dose compensator (blue dashed line). The black dash-dot line shows a ray without scattering.

Figure 1.5 shows a schematic diagram of the x-ray source, the dose compensator or bowtie filter (which is a pre-patient filter), the patient and the detector. One of the challenging problems that affects the accuracy of the projection measurement is scattered radiation. Sources of scatter radiation are the patient and the dose compensator. This figure shows examples of two scattered and one primary rays. The blue dashed lines represent a ray that scattered once in the dose compensator before reaching the detector. The red long dashed lines represent a ray that scatters two times in the patient. The black dash-dot line represents a primary ray without scattering.

Dose compensators serve to optimize the skin radiation dose received by the patient by preferentially attenuating x-rays that are incident at the edge of the patient where the

thickness is lower [49] and a lower flux is needed to achieve a given level of image quality. It is an important part of the overall strategy of maximizing the image quality to patient dose ratio. Such dose compensation is usually achieved with a pre-patient filter of variable thickness across the fan angle that is thinner at the center of the fan and thicker towards the outer edges (shaped like a bowtie). However, as with all x-ray filtration, the beam is not entirely absorbed and there is some degree of scattered radiation and some transmission without scatter. Dose compensators are typically constructed of materials that are relatively similar to water in their attenuation properties, which implies the use of materials with low atomic number elements. Scatter is then the primary mechanism of attenuation due to dominance of Compton scatter in low atomic number elements [54]. Some portion of the radiation scattered from the dose compensator can reach the detector system and lead to a degradation in image quality. In traditional Filtered Back Projection (FBP) reconstruction, these scattered x-rays confound the image reconstruction as they do not follow a straight ray path from detector to focal spot and are a source of off-focal radiation [10] that can blur edges and may require correction algorithms.

There is a need to model and estimate the scatter from the dose compensator and correct for it as part of the pre-processing steps. Since detectors are far from the scatter source and have small solid angle, traditional MC is not computationally efficient. We have implemented a variance reduction technique called forced detection (FD) to improve the computational efficiency.

1.2.2 Background on CT dose compensator scatter modeling and Forced Detection

Part of the work represented in this dissertation on the scatter from the dose compensator appeared in SPIE [7], and was independently repeated by others in [60]. In [60] Monte Carlo simulation was done without using forced detection and with monochromatic x-ray

source assumption. However X-ray sources are polychromatic in nature. In our work we used polychromatic x-ray source, analyzed the effect of scatter from dose compensator and implemented a software correction algorithm.

The forced detection technique has been used in Single Photon Emission Computed Tomography (SPECT) scanner simulations [8, 27, 61]. FD was used to simulate all projections in SPECT. In Cone Beam Computed Tomography (CBCT) the scatter contribution from the patient was studied in [63]. The scatter radiations from the patient and the dose compensator were investigated in [11, 70] for CBCT. The compensator shape and scanner geometry are different between CT and CBCT.

1.2.3 Contribution

The contribution in the CT modality is:

1. Modeling the scatter radiation from the CT dose compensator using Monte Carlo simulation: In order to increase the computational efficiency of the calculations, a forced detection method which is a variance reduction technique is implemented. We find that FD increases the efficiency of our computations by a factor of more than 40,000. The impact of the scatter radiation on image quality is studied on simulated data and a correction algorithm is implemented and applied to a set of simulated data. This is covered in Chapter 4.

1.3 Organization

The remainder of this dissertation is organized as follows. Chapter 2 explains the MDL principle, the proposed MDL-based algorithm for fMRI baseline drift and activation level estimation and the MDL-based algorithm for the HRF and activation level estimation. Chapter 3 describes the joint maximum likelihood estimation of HRF and activation levels. Chapter 4

explains the scatter contribution from CT dose compensator, the Monte Carlo simulation for modeling the scatter radiation, the variance reduction method and the proposed correction algorithm. The concluding remarks are in Chapter 5.

CHAPTER 2

FUNCTIONAL MRI AND MINIMUM DESCRIPTION LENGTH

We introduce methods based on the minimum description length (MDL) principle to estimate two important components of the fMRI signal, the baseline drift and the hemodynamic response function. In the following subsections, a detailed explanation of the MDL principle and its use in estimating the baseline drift and the hemodynamic response function are provided.

Throughout this chapter vectors and matrices are denoted by bold lowercase and uppercase symbols, respectively. Vectors are defined as column vectors.

2.1 Minimum Description Length principle

The minimum description length principle is an important tool in statistics, information theory, and learning theory, which was introduced by Rissanen [78] in 1978 and is related to *Occam's razor*. Any data can be represented by a string of numbers, also known as *codes*. In MDL, we only work with a subset of all possible codes, the prefix codes. The length of prefix codes can be shown to follow a pattern according to: $\ell(x) = -\log P(x)$, where x is the event, $P(x)$ is its probability, and $\ell(x)$ is the length of its codeword [22]. The length of description is thus related to the probability distribution of the samples. The more likely an event, the shorter its representation can be. The average length of the description will be: $L = E[\ell(x)] = -E[\log P(x)]$. If the data are distributed according to $P(x)$, then the code corresponding to $P(x)$ will give the shortest average codelength. If the probability distribution $P(x)$ is unknown and an arbitrary distribution $Q(x)$ is assumed, it can be shown that the mismatch will result in inefficiencies and the average length of the calculated

descriptions will increase.

$$L' = E_p[-\log Q(x)] > E_p[-\log P(x)] = L \quad (2.1)$$

This is known as the *information inequality*. Now imagine a number of samples with unknown distribution. One can try to build representations based on all candidate distributions or models until come across one that gives the smallest average length. The inequality (2.1) shows that the minimum description length can lead to a good approximation of the probability $P(x)$ that samples are drawn from. The class of probability distributions, out of which our observed data has been drawn is denoted by \mathcal{M} and called a family. It is easier to imagine that the class is parameterized by one parameter θ from the parameter space Θ , so finding a desired distribution is equivalent to finding a good θ . This group is then called a parametric family, and denoted as

$$\mathcal{M} = \{P(\cdot|\theta); \theta \in \Theta\} \quad (2.2)$$

MDL methods calculate the description length for the model and the data and minimize the total description length. The description length has two components: description of θ , the member of the family of distributions to which data belongs, and a description of $P(D|\theta)$. This is a consequence of conditional probabilities, because the distribution of the data is $P(D) = P(D|\theta)P(\theta)$, therefore

$$-\log P(D) = -\log P(D|\theta) - \log P(\theta). \quad (2.3)$$

Having a set of candidate models, the best model to explain the data D is the one which minimizes the combined length of the description of the model, plus the description of the data under that model.

Often parametric descriptions for each component are used so that they can be more easily optimized. A useful example is MDL denoising via wavelets [79], where each subset

of the wavelet coefficients are considered as a model class. The description of model, θ , is constant, thus simplifying the optimization, and the length of $P(\cdot|\theta)$ is described by the number of non-zero wavelet coefficients.

Denoising is related to data compression; many denoising techniques are designed based on sparseness and compressibility in the wavelet domain. Methods suggested by Donoho and Johnstone [29, 30] are the pioneering denoising techniques among the wavelet-based denoising methods. By taking advantage of the property of the orthonormal wavelet basis, the noise will have coefficients with small absolute value. So denoising can be done by finding a threshold such that if the square of the coefficient exceeds the threshold it remains while others are set to zero. The problem with wavelet denoising is that the noise is defined by the estimated signal, which depends on the estimated variance of the noise. The MDL principle has another approach to denoising. MDL takes advantage of the fact that noise is the incompressible part in the data while the compressible part defines the meaningful information-bearing signal. Based on this concept, one may decompose the signal by minimizing the ideal code length describing the data. Because the noise part is incompressible, one can recover the data itself.

Denoising is a special case of linear regression with regressor selection. MDL denoising uses wavelets. Orthonormal regressor matrices such as wavelet transforms are used as regressors. The wavelet transform takes the data and converts it into coefficients representing the data at various frequencies and scales (high-detail vs. low-detail). The solution to the denoising problem is to consider each subset of the regressors separately as a model class. The model class in the case of wavelets is going to be defined as: “the non-zero coefficients,” which includes only the coefficients which are above the threshold. The coefficients below that threshold will be set to zero. In other words, we will partition all signals according to which of their wavelet coefficients are non-zero. The minimization of description length is equivalent to the minimization of the number of non-zero coefficients. Models are compared by their ability to compress a data set. The goal is to select a model that gives the most

compact description of the data and the model. This mathematical model is motivated by the practical fact that many real-world signals have a relatively small number of significant coefficients, with a large number of zero or near-zero coefficients.

2.1.1 Why use MDL for fMRI?

The fundamental idea behind the MDL principle in modeling is that any regularity in a given set of data can be used to compress the data. Since we want to select the hypothesis/model that captures the most regularity in the data, we look for the hypothesis/model with which the best compression can be achieved.

fMRI models only try to match to the regularities in the observed data from which scientists aim to learn, and then to propose and test hypotheses about the biological phenomena. Modeling is therefore closely related to answering the following question: *how much structure is there in the data?* Unstructured measurements are categorized as noise, while appropriately structured measurements are inferred as active voxels. Thus, if detecting and learning from data regularities is our aim, it stands to reason to develop methods and algorithms that specifically pursue and uncover these regularities with the fewest assumptions about the data. Thus, the MDL principle is a very good match for the fMRI modeling problem.

MDL has been successfully used in some image analysis problems, i.e., image restoration and edge location, segmentation, denoising, simultaneous noise suppression and compression [33, 72, 55, 81]. We note that, in the past, MDL has been used as a tool to solve small subproblems in fMRI. For example, certain models have been assumed and then their order was calculated with MDL [9, 19, 50]. In contrast, we propose that MDL can be used as a comprehensive principle that can guide us throughout the solution and algorithm development.

2.2 MDL-based estimation of fMRI drift and activation level¹

2.2.1 Algorithm

The contribution of this chapter for drift estimation is to have fewer prior assumptions about the drift component and let the data itself find the best model for it. No subspace is assumed for the drift component; moreover, drift can have different structures across voxels. In this approach, data are considered as sequences to be compressed by the model. We implement a compression-based method for drift estimation based on MDL denoising as introduced in [79]. Wavelets are used as regressors for MDL denoising.

The fMRI time series model at each voxel j at a time point t_i is modeled as

$$y_j(t_i) = \alpha_j \underbrace{(S(t_i) * h(t_i))}_{X(t_i)} + d_j(t_i) + e_j(t_i), \quad j = 1, \dots, J \quad \text{and} \quad i = 1, \dots, N \quad (2.4)$$

where the first term represents the BOLD response model which is the brain response to a given stimulus. Index j represents the spatial position of a voxel, J is the total number of voxels, α_j is the activation level, $S(t_i)$ is the stimulus time series which is convolved with the HRF represented by $h(t_i)$. The convolution of stimulus and HRF is denoted as $X(t_i)$. The slowly varying baseline drift is represented as \mathbf{d}_j and \mathbf{e}_j is random noise. The time for each scan is $t_i = iTR$, with TR as the time of repetition and i is between 1 and N , the number of scans. J is the total number of voxels.

By rewriting this equation as vectors and matrices, the fMRI time series at voxel j is

$$\mathbf{y}_j = \mathbf{X}\boldsymbol{\alpha}_j + \mathbf{d}_j + \mathbf{e}_j \quad j = 1, \dots, J \quad (2.5)$$

where the $N \times 1$ vector \mathbf{y}_j is the fMRI time series for voxel j , \mathbf{X} is the $N \times m$ design matrix which includes the desired activation signal for all the stimuli and a vector of all-ones

¹© 2007 IEEE. Adapted/Reprinted with permission from N. Bazargani, A. Nosratinia, K. Gopinath, and R. Briggs (2007, April). fMRI baseline drift estimation method by MDL principle. In IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), Washington, D.C., 472-475.

representing the average of the signal. In this case the number of stimuli is $m - 1$. The desired activation signal is the convolution of the HRF with each of the stimulus time series. $\boldsymbol{\alpha}_j$ is $m \times 1$ vector represents the activation level and mean of the time series, \mathbf{d}_j represents the drift component and \mathbf{e}_j denotes Gaussian noise. For cases with one stimulus, \mathbf{X} is $N \times 2$ and $\boldsymbol{\alpha}_j$ is 2×1 . Since the goal of this study is to focus on the drift estimation, we do not concern ourselves with the power spectral density of the noise, and assume it to be white. The baseline drift is estimated voxel-wise.

The following iterative algorithm is proposed to estimate the unknown parameters \mathbf{d}_j and $\boldsymbol{\alpha}_j$ for each voxel:

1. Define design matrix \mathbf{X} by convolving a known HRF with the stimulus time series.
2. Initialize drift to zero, $\hat{\mathbf{d}}_j^{(0)} = 0$.
3. For iteration i , find an initial estimate of $\boldsymbol{\alpha}_j$ via least squares estimation

$$\hat{\boldsymbol{\alpha}}_j^{(i)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{y}_j - \hat{\mathbf{d}}_j^{(i-1)}). \quad (2.6)$$

4. Use the activation level estimate and the MDL principle to find a new estimate of the drift

$$\hat{\mathbf{d}}_j^{(i)} = f_{MDL}(\mathbf{y}_j - \mathbf{X} \hat{\boldsymbol{\alpha}}_j^{(i)}). \quad (2.7)$$

5. Set $i = i + 1$.
6. Repeat steps 3-5 until the convergence condition $\|\hat{\mathbf{d}}_j^{(i)} - \hat{\mathbf{d}}_j^{(i-1)}\|_2 < \epsilon$ holds. $\|\cdot\|_2$ represents ℓ^2 norm.

The function f_{MDL} in step 4 represents the use of MDL principle for estimating the drift component. We use MDL denoising, which operates in the wavelet domain. The MDL principle is described in Section 2.1. Specifically, we use a variation of the method

suggested in [80], which itself is a refined and extended version of the earlier MDL denoising criterion [79].

For each model class in MDL denoising a single representative of the whole class is constructed which is universal. This means that it can represent any densities in the whole model class. One of the universal models is the normalized maximum likelihood (NML) model. The NML code length is used to determine the length of the description and gives the minimax optimal universal model for each model class. NML was originally proposed by Shtarkov [83] for data compression. After finding NML code length for each model class, the model class with the shortest description length for the data was selected.

In the MDL-based denoising, model selection is performed by considering each subset of coefficients as a model class and minimizing the summation of the code length of the model and the code length of the data given the model. The MDL criterion proposed in [80] for finding the best subset of coefficients is as follows

$$\min_k \frac{(n-k)}{2} \ln \frac{C - C_\gamma}{(n-k)^3} + \frac{k}{2} \ln \frac{C_\gamma}{k^3} \quad (2.8)$$

where n is the total number of coefficients, γ denotes a set of k nonzero coefficient indices, $k \in (0, n)$, C is the sum of the squares of all the wavelet coefficients and C_γ is the sum of the squares of the coefficients that are in the set γ . As described in [80], for orthonormal bases, the subset of coefficients that minimizes the MDL criterion corresponds to the k largest coefficients in absolute value for some k . The rest of the coefficients will be set to zero.

The Student's t-test is calculated for testing the significance of the activation level estimate. The contrast vector is defined as $\mathbf{c} = [1 \ 0]^T$ for a study with one stimulus time series to check the significance of each of the elements of vector $\hat{\boldsymbol{\alpha}}_j$. The null and alternative hypotheses at each voxel are

$$H_0 : \mathbf{c}^T \boldsymbol{\alpha}_j = 0$$

$$H_1 : \mathbf{c}^T \boldsymbol{\alpha}_j \neq 0.$$

Under the null hypothesis, $\frac{\mathbf{c}^T \hat{\boldsymbol{\alpha}}_j}{\sqrt{\mathbf{c}^T \text{cov}(\hat{\boldsymbol{\alpha}}_j) \mathbf{c}}}$ has a Student's t-distribution with ν degrees of freedom. For MDL-based drift estimation algorithm the degree of freedom is $\nu = N - \text{rank}(\mathbf{X}) - p$, where p is the number of remaining wavelet coefficients.

2.2.2 Results

The performance of the algorithm was tested using simulated and real fMRI data. For MDL denoising, we used Daubechies symlet 8 wavelet basis with 6 decomposition levels. In order to implement wavelet detrending, the wavelet set should be compact. Daubechies wavelets and symlets with small orders are therefore reasonable candidates.

We evaluated our method against existing detrending approaches like linear, quadratic and cubic polynomial detrending methods [41, 84] and spline detrending [40].

For polynomial detrending, a polynomial with the pre-defined degree was fitted to each voxel time series. Then, the detrended time series was obtained by subtracting the fitted polynomial model from the voxel time series. The degrees of freedom for calculating t-test of polynomial detrending is $\nu = N - \text{rank}(\mathbf{X}) - \text{polynomial order}$.

Smoothing splines have been used in the literature [40] to estimate the fMRI baseline drift. Using MATLAB spline toolbox, a cubic smoothing spline was fitted to the fMRI time series at each voxel. The smoothing parameter is a number between 0 and 1, which is selected based on giving a good estimate of the benchmark drift in the simulated data. This parameter has been used for experiments on real data. The spline $\theta(t)$ is the solution to the optimization problem

$$\min_{\theta} \rho \sum_{k=1}^N |y_j(t_k) - \theta(t_k)|^2 + (1 - \rho) \int |\theta''|^2 d\theta \quad (2.9)$$

where $y_j(t_k)$ is the fMRI time series at voxel j at time sample t_k , the parameter ρ controls the degrees of smoothness. The fitted spline is subtracted from the original time series to obtain the detrended time series.

Simulated Data

The simulated time series at each voxel j was constructed using model in Equation (2.5). We defined one stimulus time series $S(t)$, a known benchmark HRF $h(t)$, a known activation level α_{1j} which was multiplied by $(h(t) * S(t))$, a known drift term \mathbf{d}_j , a known DC level α_{2j} and white Gaussian noise \mathbf{e}_j with the signal to noise ratio (SNR) observed in the fMRI data. The HRF was defined as the difference of two Gamma functions with known parameters as in [41].

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} e^{-\frac{(t-d_1)}{b_1}} - c\left(\frac{t}{d_2}\right)^{a_2} e^{-\frac{(t-d_2)}{b_2}} \quad (2.10)$$

where $d_j = a_j b_j$, $a_1 = 6$, $a_2 = 12$, $b_1 = b_2 = 0.9$, $c = 0.35$. This HRF was convolved with the stimulus time series $S(t)$ to yield the first column of matrix \mathbf{X} . The benchmark drift used in the simulated time series was extracted by kernel smoothing the resting state fMRI data. The added drift at each voxel was tuned to be significant so that the ratio of the norm of the drift over the norm of the drift free signal was more than 0.5.

Figure 2.1 shows one time series realization of a simulated data with $SNR = 0.8$ and block design stimuli with period 40 seconds and known activation level. The length of the time series was $N = 160$. The benchmark drift and estimated drift via each of detrending methods have been shown. The mean square error (MSE) between the benchmark and estimated drifts was calculated and was shown in each plot. MDL detrending gives the lowest MSE compare to the other approaches.

We performed Monte Carlo (MC) studies on the simulated fMRI time series which was generated with the benchmark activation and the benchmark drift, the length of the time series was $N = 160$. Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to obtain numerical results. We used Monte Carlo for error analysis. The signal to noise ratio was within the typical range of the SNR in the real fMRI data, which is between 0.2 and 1 [52]. For each SNR, $Q = 300$ Monte Carlo realizations

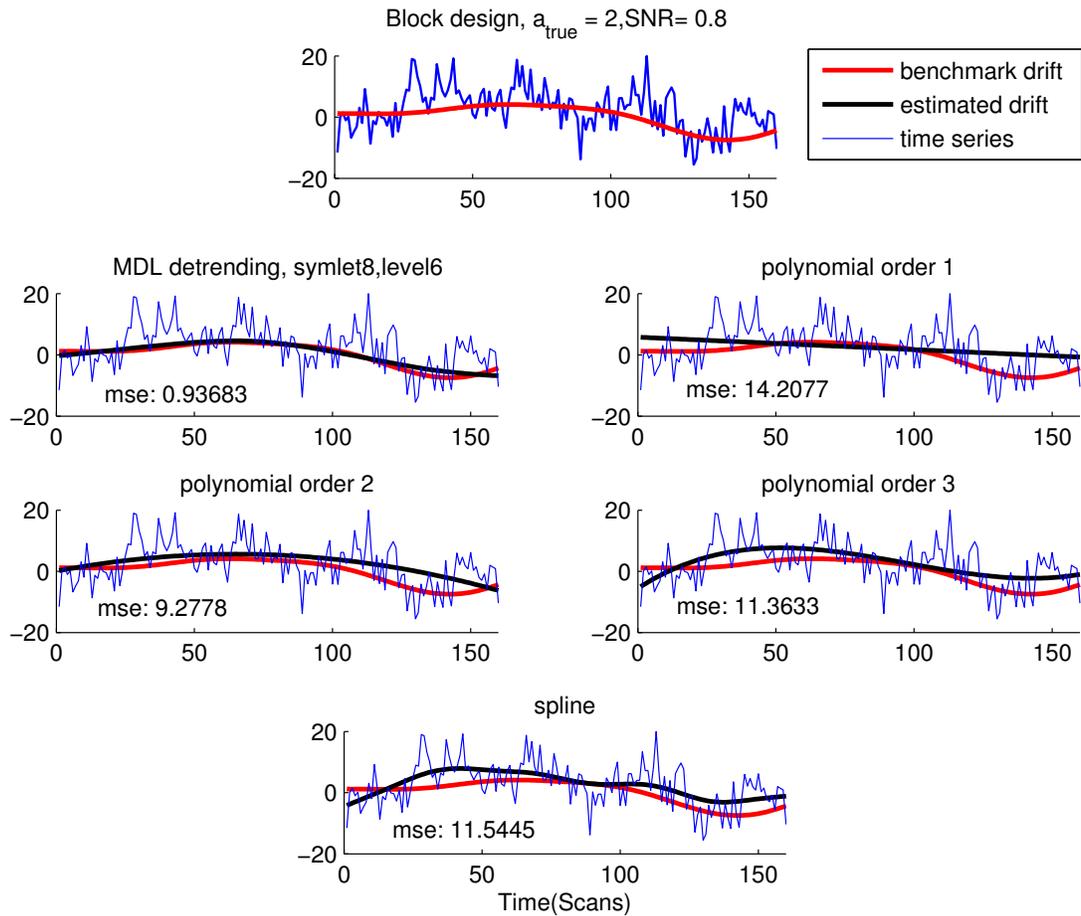


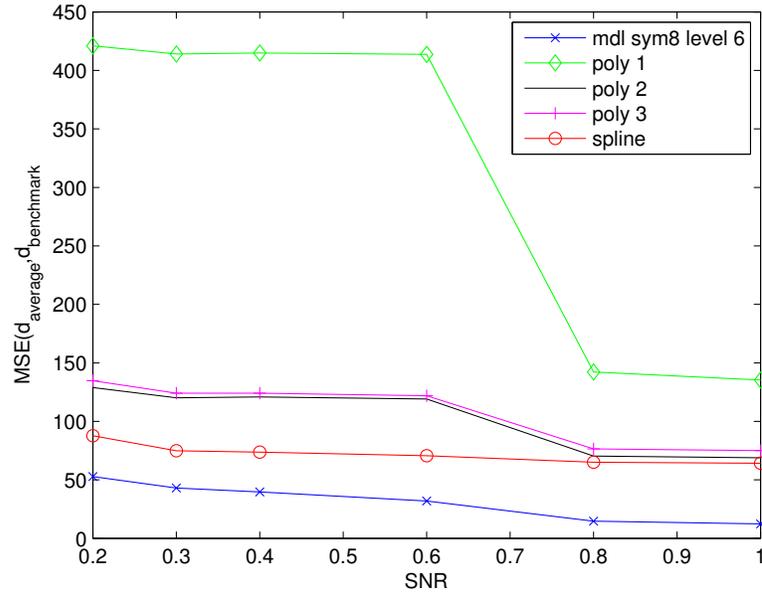
Figure 2.1. Comparing estimated drift from different methods with the benchmark drift for a block design simulated data with $\text{SNR}=0.8$. The MSE between estimated and benchmark drifts is shown for each method.

have been performed. For each noise realization the square error between the estimated and the benchmark drift was calculated, and then the average was calculated over all the realizations. Figure 2.2 shows the mean square error (MSE) between the estimated and the benchmark drift for different SNR values, for a block design experiment with period of 40 seconds and for an event-related experiment, respectively. The true activation level was set to 5. The performances of various detrending methods were compared with each other. The MDL detrending results in smaller MSE compare to other detrending methods for all SNR

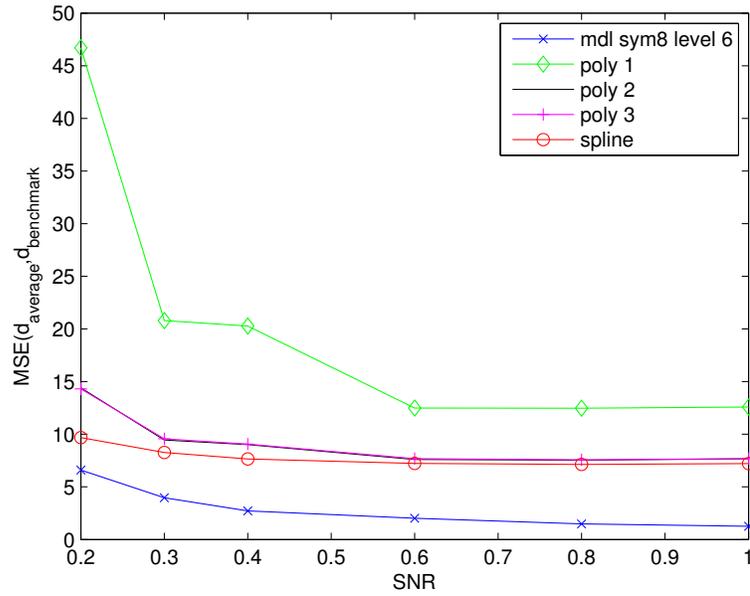
levels. As the SNR increases, the MSE decreases for all of the detrending methods. Event related experiment shows lower MSE in comparison with the block design experiment.

The drift has slow variations, so it induces correlation to the fMRI time series. The autocorrelation will be decreased by detrending. Thus, we checked the performance of the various detrending methods in removing autocorrelation from the voxel time series. The histogram of the residual autocorrelation at lag one was plotted for all the voxels before and after removing drift and is shown in Figure 2.3. The data set we used for this experiment was a resting data set with added activation to it. The activation level was 1% of the mean of the signal. MDL-based compression-based detrending, spline detrending and polynomial order 3 result in the lowest 95th percentile of the residual error at lag one, respectively. The 95th percentile is the autocorrelation in the most autocorrelated residual time series. The result was consistent both for event related and block design stimuli added to the resting state data. So, in terms of removing autocorrelation of the residual error, MDL-based detrending gives better performance compare to polynomials and spline detrending. To compare detection performance, activation was added to some voxels of the resting state data at known locations. This generated benchmark activation level.

The stimulus was a block design stimulus with the period of 40. The known activation with $\text{SNR} = 0.2$ was added to 459 voxels in known locations. The benchmark activation for one slice is shown in Figure 2.4. Axial, sagittal and coronal views are shown for MDL-based, without detrending, spline detrending and polynomial detrending with order 3. The MDL-based activation map shows less false positives compare to no detrending and polynomial detrendings, and at the same time detects more true positives compare to spline detrending. The t-test threshold is the same for all of these images.



(a) Block design



(b) Event related

Figure 2.2. The mean square error between the benchmark and estimated drift for (a) a block design and (b) an event related experiment.

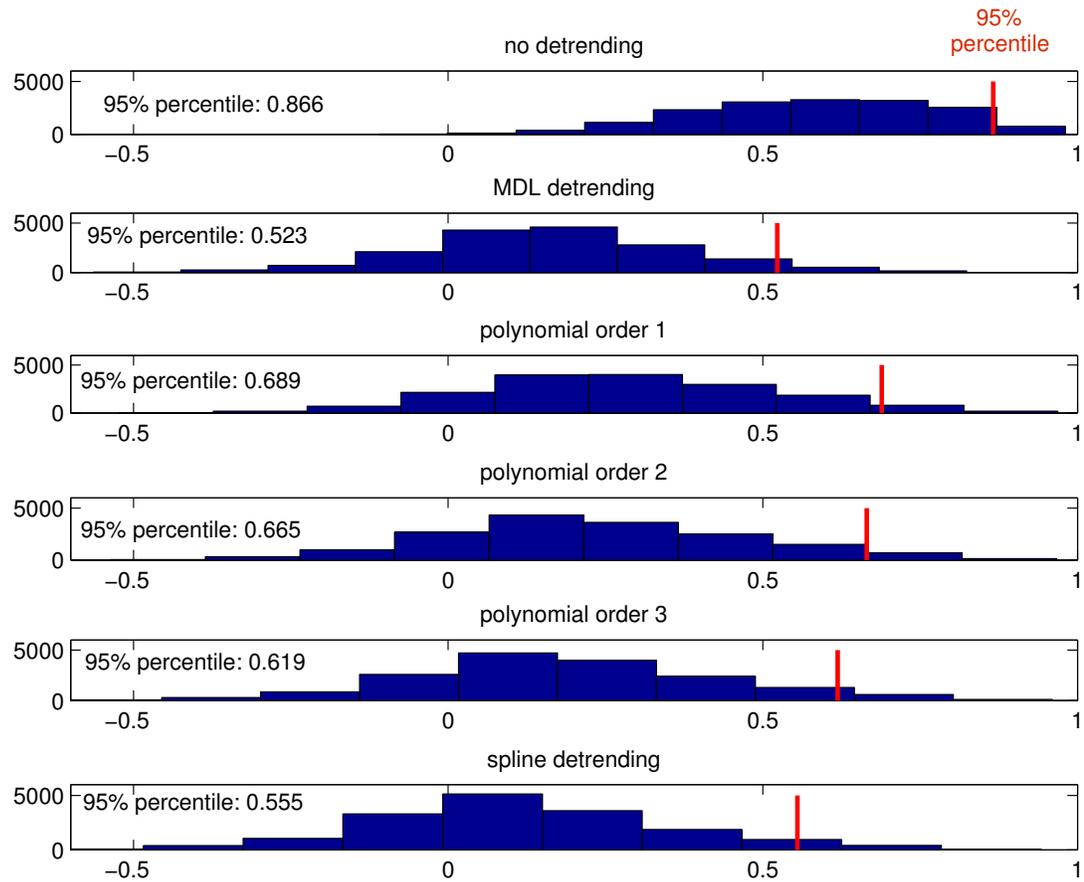


Figure 2.3. Histogram of autocorrelation of residual error at lag one for all voxels for a block design experiment.

Real fMRI data

Figure 2.5 shows results of MDL-based detrending on a real fMRI “attention to visual motion” data set available on the SPM web site.² The data set was used with permission from Methods group at the Wellcome Trust Centre for Neuroimaging. This data set was obtained by Christian Büchel and is described in [13]. The experiment was a block design consisting of 4 conditions: fixation, static, no attention and attention. Each of these conditions were

²<http://www.fil.ion.ucl.ac.uk/spm>

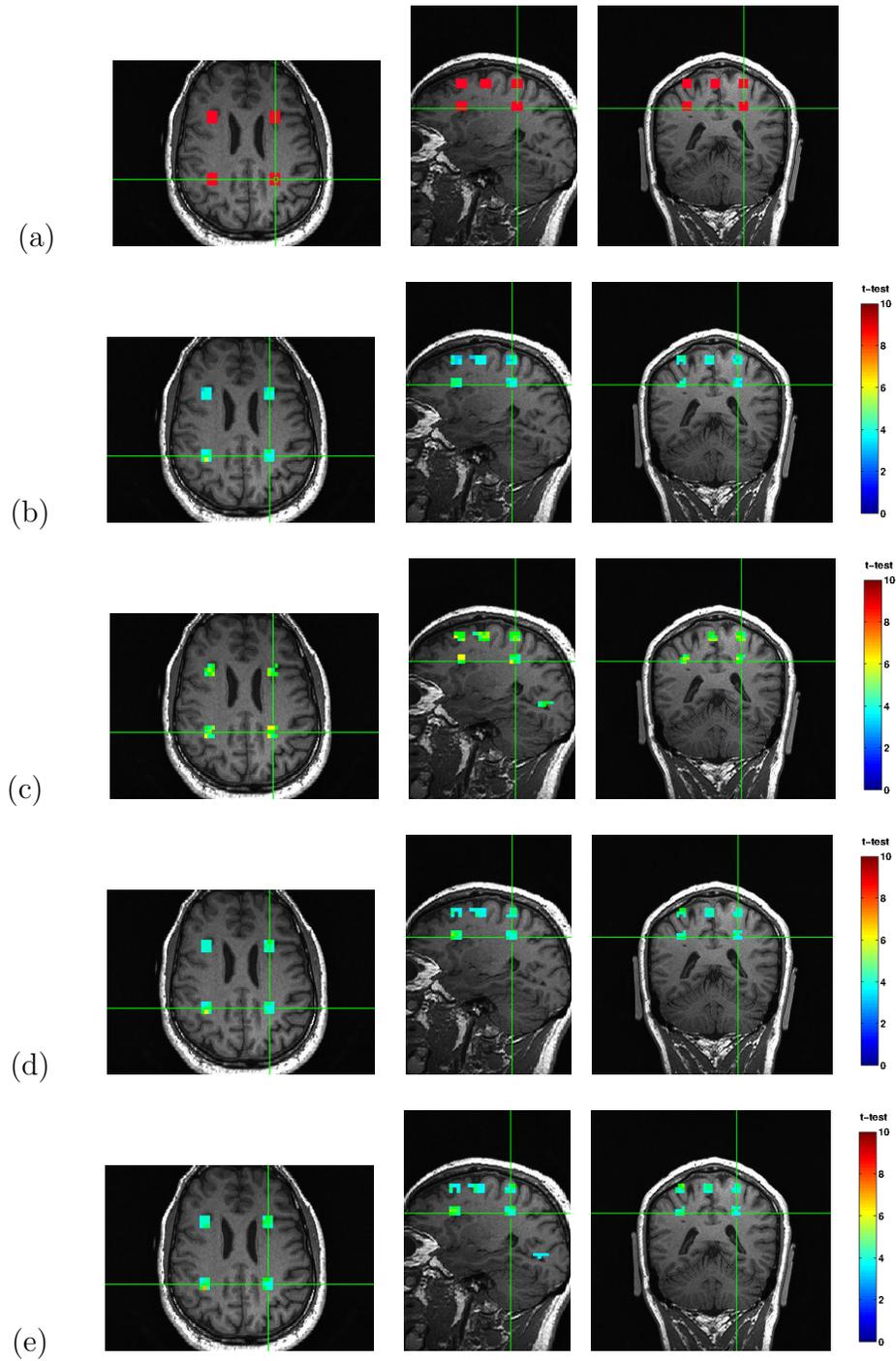


Figure 2.4. Axial, Sagittal and Coronal views of a single slice activation maps. (a) benchmark activation, (b) MDL-based detrending, (c) No detrending, (d) spline detrending and (e) polynomial of order 3 detrending.

blocks of 32.2 seconds. The repetition time (TR) was 3.22 seconds. There were 360 scans in this study. The experiment was performed on a 2 Tesla Magnetom VISION (Siemens, Erlangen) whole-body MRI system equipped with a head volume coil. Contiguous multislice T2*-weighted fMRI images, with 64×64 pixels, 32 slices and slice thickness of 3 mm. Image processing and statistical analysis were carried out using SPM8 and MATLAB. All volumes were realigned to the first volume and a mean image was created using the realigned volumes. A structural MRI image, acquired using a standard three-dimensional T1 weighted sequence ($1 \times 1 \times 3$ mm voxel size) was co-registered to this mean (T2) image.

The activated brain regions identified by our approach were compared with the activation maps from spline detrending, polynomial of order 3 detrending and AFNI (Analysis of Functional NeuroImages) imaging analysis tool.³ AFNI is a widely used set of programs that implement well-known algorithms for processing, analyzing and displaying fMRI data [23], and is often used for benchmarking new developments. Figure 2.5 shows axial and coronal views of activation maps from AFNI, compared with MDL-based detrending, no detrending, spline detrending and polynomial of order 3 detrending. The activation maps are superimposed on the anatomical data.

For AFNI analysis the data was detrended (using Legendre polynomials of degree 3) and then analyzed using `3dDeconvolve` to locate the active areas of the brain. The HRF modeled as a finite impulse response model in AFNI, while for other methods in Figure 2.5 the HRF was based on Equation (2.10). The overall activation location of the MDL-based method is not different from AFNI. We observe slight differences with AFNI, which are explained by the differences between F-test and t-test and the differences between the algorithms used in AFNI and our method. The number of false positives are higher in spline detrending and polynomial detrending compare to AFNI and MDL-based detrending.

³<http://afni.nimh.nih.gov>

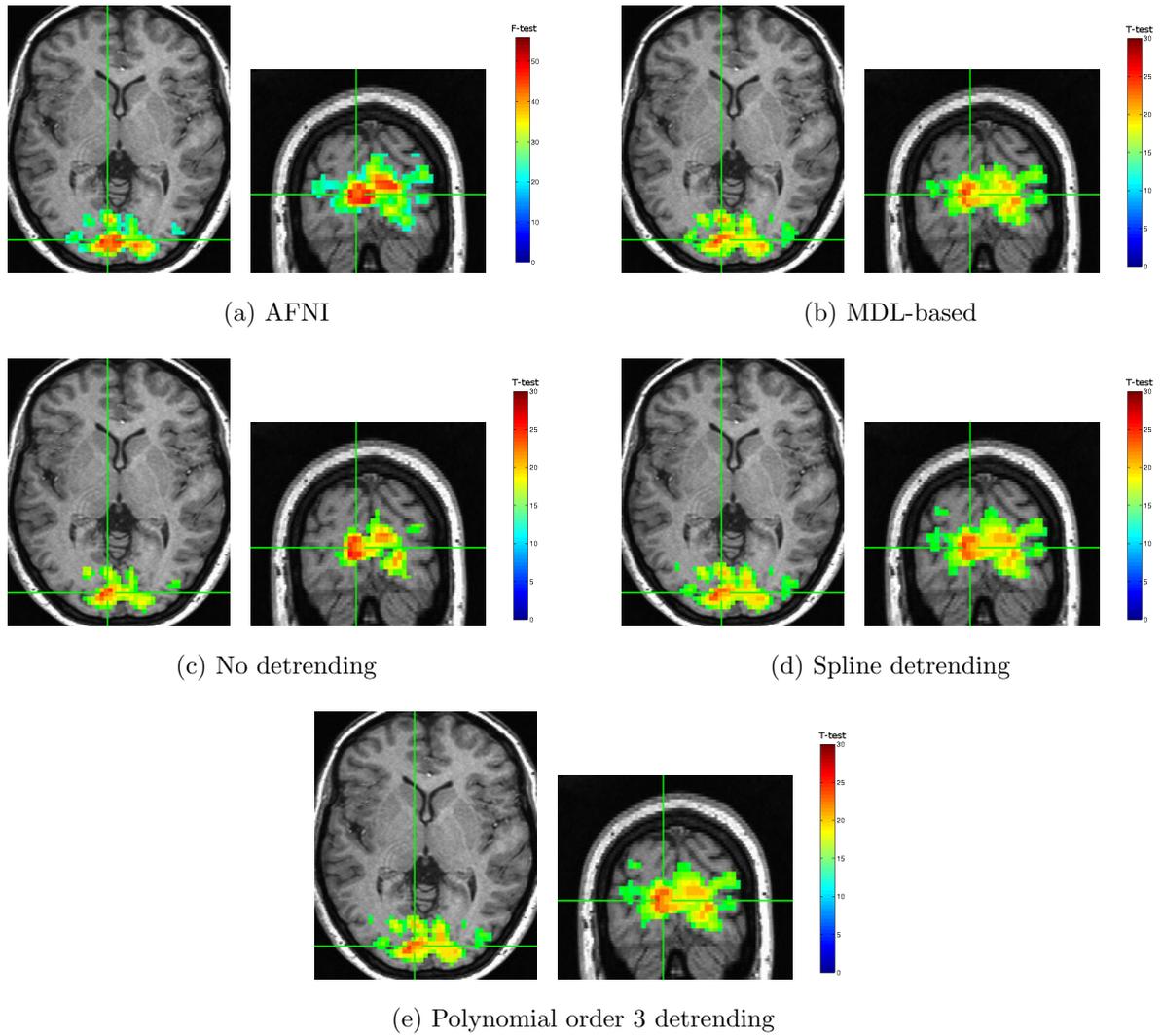


Figure 2.5. Axial and Coronal views of a single slice activation map of real fMRI data for (a) AFNI, (b) MDL-based detrending, (c) No detrending, (d) spline detrending and (e) polynomial of order 3 detrending.

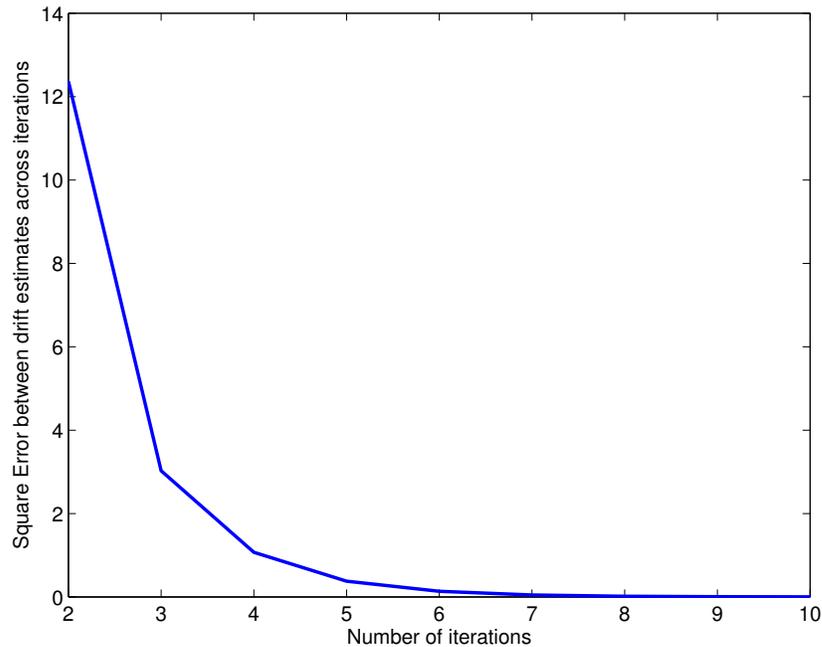


Figure 2.6. Square error between drift estimates across iterations for one voxel of a real fMRI dataset.

The MDL algorithm is an iterative algorithm, the condition for convergence is the square error between the drift estimates between the two iterations, $\|\hat{\mathbf{d}}^{(i)} - \hat{\mathbf{d}}^{(i-1)}\|_2 < \epsilon$, where $\|\cdot\|_2$ represents ℓ^2 norm and i is the iteration number. The ϵ is chosen to be 10^{-5} . For real fMRI data the maximum number of iteration needed was 20. Figure 2.6 shows the square error for drift estimates across iterations. As can be seen the square error decreases as the number of iterations increases and the algorithm converges.

2.3 MDL-based estimation of HRF and activation level

2.3.1 Algorithm

The contribution of this section is to use the MDL principle to estimate the HRF from the data itself, without any assumptions on its shape. MDL-based methods, which can extract models from data itself, are a natural choice for attacking this important problem,

because it allows the HRF to vary across different brain regions, experiments and subjects. The proposed algorithm iteratively estimates the HRF shape for a region of interest and activation levels for all voxels in that ROI, using the MDL principle and the least squares estimation.

We start with the model in Equation (2.5) without the drift component for each voxel time series as shown below

$$\mathbf{y}_j = \mathbf{X}\alpha_j + \mathbf{e}_j \quad j = 1, \dots, M \quad (2.11)$$

where index j indicates the spatial position of a voxel in the brain and M is the number of voxels in the region of interest. The $N \times 1$ vector \mathbf{y}_j shows the mean subtracted and detrended time series of voxel j . \mathbf{X} is a $N \times 1$ matrix that includes the convolution of the HRF, $h(t)$, with one stimulus time series $S(t)$. The HRF is unknown and will be estimated. Activation level for voxel j is denoted by the scalar α_j which is unknown and needs to be estimated and \mathbf{e}_j is the white Gaussian noise. We assume that HRF has the same shape over a homogeneous region but the magnitude can change. We can divide the brain into several sub-volumes, such that the HRF over each sub-volume is constant. For real fMRI data, the ROIs are obtained from analyzing the data with AFNI imaging analysis tool. In order to capture HRF variations, the estimation process is applied on each of the ROIs.

The general linear model for the set of voxels in each sub-volume is as follows

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{E} \quad (2.12)$$

where the matrix \mathbf{Y} includes the time series for M voxels in one cube, $\mathbf{Y} = [\mathbf{y}_{j+1} \ \dots \ \mathbf{y}_{j+M}]$. $\boldsymbol{\alpha}$ is a vector including α_{j+1} to α_{j+M} and \mathbf{E} is a matrix containing Gaussian noise for all the voxels in the ROI.

The following iterative algorithm is proposed to estimate the HRF and activation levels:

1. Set iteration number $i = 1$. Find an initial estimate of the BOLD response using one of the voxel's time series.

$$\hat{\mathbf{X}}^{(1)} = \frac{f_{MDL}(\mathbf{y}_j)}{\|f_{MDL}(\mathbf{y}_j)\|} \quad (2.13)$$

2. In first iteration, find the estimate of activation level via least squares from an initial estimate of the BOLD response and the time series of the M voxels in a cube, namely $\hat{\mathbf{X}}^{(1)}$

$$\hat{\boldsymbol{\alpha}}^{(1)} = (\hat{\mathbf{X}}^{(1)T} \hat{\mathbf{X}}^{(1)})^{-1} \hat{\mathbf{X}}^{(1)T} \mathbf{Y} \quad (2.14)$$

3. For iteration $i + 1$, use the activation estimate and the MDL principle to find an estimate for the BOLD response. $\hat{\boldsymbol{\alpha}}^{(i)\dagger}$ denotes pseudoinverse of $\hat{\boldsymbol{\alpha}}^{(i)}$.

$$\begin{aligned} \mathbf{Y} \hat{\boldsymbol{\alpha}}^{(i)\dagger} &= \hat{\mathbf{X}}^{(i+1)} + \mathbf{E} \hat{\boldsymbol{\alpha}}^{(i)\dagger} \\ \hat{\mathbf{X}}^{(i+1)} &= f_{MDL}(\mathbf{Y} \hat{\boldsymbol{\alpha}}^{(i)\dagger}) \end{aligned} \quad (2.15)$$

4. Normalize the estimate of \mathbf{X}

$$\hat{\mathbf{X}}^{(i+1)} = \frac{\hat{\mathbf{X}}^{(i+1)}}{\|\hat{\mathbf{X}}^{(i+1)}\|} \quad (2.16)$$

5. Use the latest $\hat{\mathbf{X}}$ to update the activation level estimate

$$\hat{\boldsymbol{\alpha}}^{(i+1)} = (\hat{\mathbf{X}}^{(i)T} \hat{\mathbf{X}}^{(i)})^{-1} \hat{\mathbf{X}}^{(i)T} \mathbf{Y} \quad (2.17)$$

6. Set $i = i + 1$.
7. Repeat steps 3-6 until the convergence condition $\|\hat{\mathbf{X}}^{(i+1)} - \hat{\mathbf{X}}^{(i)}\|_2 < \epsilon$ holds.
8. Calculate HRF, $h(t)$, by deconvolution of $\hat{\mathbf{X}}$ and the stimulus time series, using the constraints that $h(t) = 0$ for $t < 0$ and $t > t_n$. Divide $h(t)$ by its maximum value to get the peak of one for comparison purposes.

The function f_{MDL} in steps 1 and 3 represents the use of MDL principle for estimating the HRF. We use MDL denoising, which operates in the wavelet domain. Description of f_{MDL} is the same as Section 2.2.

2.3.2 Results

Simulated Data

The simulated time series were generated for a ROI containing 27 active voxels. The benchmark HRF was the difference of two Gamma functions as in Equation (2.10). The HRF was convolved with an event-related paradigm to generate the BOLD response. The SNR for each time series was 0.2. The benchmark and estimated HRF are shown in Figure 2.7, together with the Poisson and Gamma HRF plots. The MDL-based algorithm can estimate the benchmark with good precision. The t-statistics were calculated for all the voxels in the ROI with each of the HRFs. These t-statistics are shown in Figure 2.8. The t-values with the proposed MDL-based method are better estimates of the t-values compared with others, since they are closer to the t-values obtained with the benchmark HRF.

Real fMRI data

The real data were acquired from a healthy subject at the University of Texas Southwestern at Dallas on a 3-T scanner. The experiment was an event related paradigm with 27 auditory stimuli, the events were jittered with random inter stimulus intervals. The voxel size was $3.125 \text{ mm} \times 3.125 \text{ mm} \times 4 \text{ mm}$ and the number of scans was 160, repetition time (TR) was 2 seconds and the only preprocessing step done on the data was the registration. The size of the whole brain data set was $64 \times 64 \times 40$ voxels. The ROIs were obtained as follow. The data was detrended (using Legendre polynomials of degree 3) and then analyzed with AFNI toolbox using `3dDeconvolve` to locate the active areas of the brain. The criteria for defining the voxel clusters were the threshold for the activation map and the size of the cluster. The active contiguous voxels which were above these specific thresholds were defined as regions of interest. The program `3dclust` in AFNI toolbox finds clusters that fit the criteria. We set the threshold for the F-test to be greater than 3.045 and the number of voxels in each cluster to be greater than 20 voxels. We obtained 2 ROIs with these thresholds.

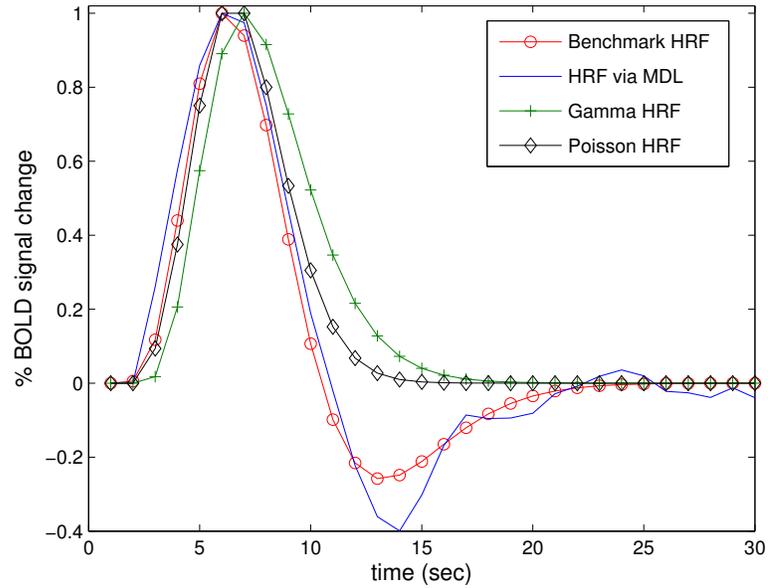


Figure 2.7. The benchmark, Gamma, Poisson and estimated HRFs for an event related stimuli in a ROI with 27 active voxels.

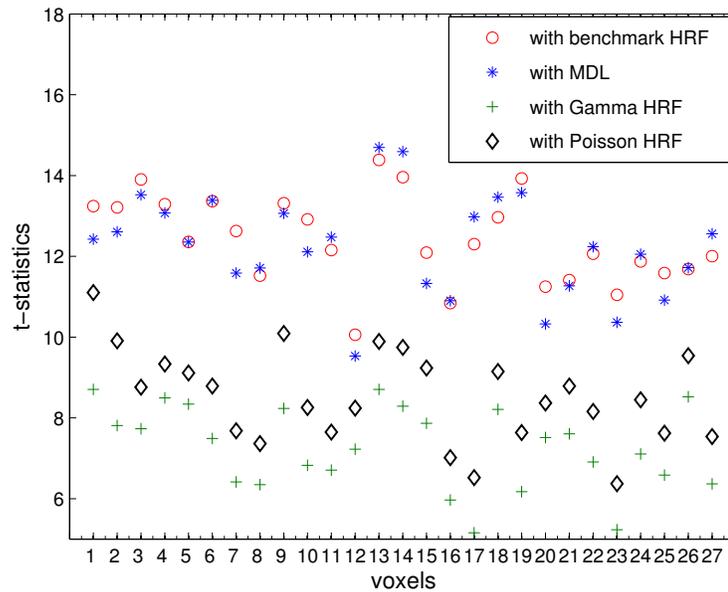


Figure 2.8. Comparing t-statistics obtained by benchmark, Gamma, Poisson and estimated HRFs.

The MDL-based HRF estimation algorithm was applied to each of the ROIs defined by AFNI analysis. For each voxel in the ROI the activation levels and one HRF for that ROI were estimated. The Student's t-test was calculated for each of the voxels in the ROI and the t-test activation map was superimposed on the anatomical data. Figure 2.9 shows the HRF estimates and activation maps for MDL-based and AFNI. The HRF estimates from the two methods for the two ROIs are shown in subfigures (a) and (b), and (e) and (f). First ROI is the left one and the second ROI is the right one shown in the activation maps. The activation map from the MDL-based algorithm is consistent with the AFNI activation maps. Since AFNI uses voxel-based analysis and MDL uses region-based methods, in low SNR regions there will be differences. MDL-based performs better in low SNR since it is region-based. The slight differences between the activation maps are because of differences between F-test used in AFNI and t-test used in our method, and differences between the two algorithms. The HRF estimates from MDL-based algorithm are smoother since the algorithm is ROI-based.

2.4 Conclusion

The use of MDL was introduced for estimating the baseline drift and the HRF in fMRI signal. For each of the algorithms activation levels for each voxel were estimated as well. The proposed algorithm based on MDL for the drift estimation needed no prior assumptions about the drift component and let the data itself find the best model for it. No subspace was assumed for the drift component; moreover, drift could have different structures across voxels. This algorithm showed improvement in activation detection and estimating the drift component compare with other detrending methods such as spline and polynomial methods. The MDL-based drift and activation level estimation algorithm was applied on a real fMRI dataset and the results were compared with AFNI, polynomial of order 3 detrending and spline detrending. The proposed MDL-based algorithm showed less false positives and

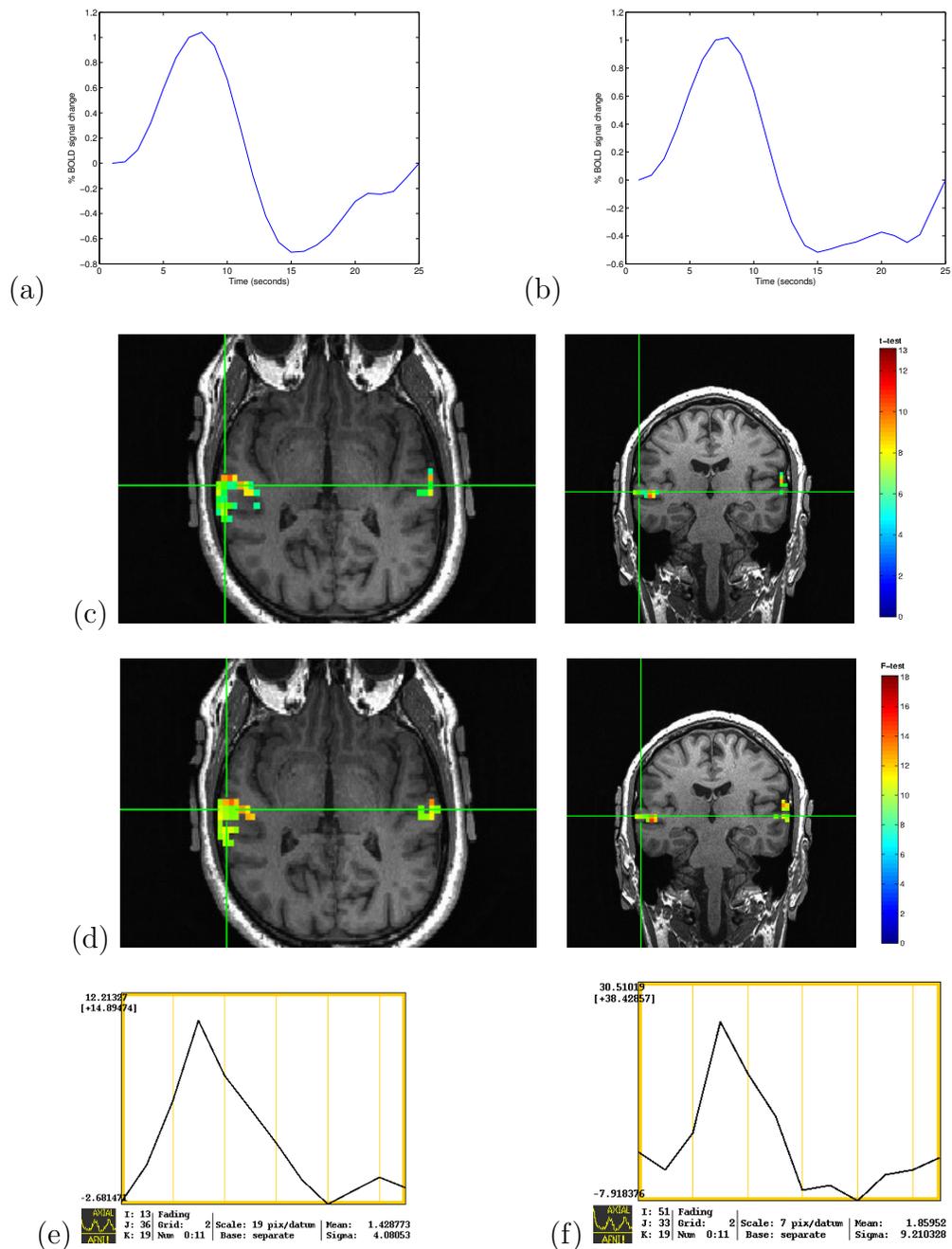


Figure 2.9. Axial and Coronal views of a single slice activation maps for MDL-based HRF estimation and AFNI. (a) HRF from MDL-based estimation for ROI 1 (on the left side of activation maps), (b) HRF from MDL-based estimation for ROI 2 (on the right side of activation maps), (c) MDL-based activation maps, (d) AFNI activation maps, (e) HRF from AFNI for one voxel in ROI 1, (f) HRF from AFNI for one voxel in ROI 2.

more true positives compared with other detrending methods. The benefits of the proposed algorithm were seen more in the low SNR regions as was shown in the simulations.

The MDL-based algorithm for the HRF and activation level estimation let the data itself find the best model for the HRF. No subspace was assumed for the HRF, and it could capture HRF variabilities across regions. The proposed algorithm was applied on a simulated data with a benchmark HRF and it showed that the proposed algorithm was capable of estimating the HRF and resulted in better t-test values compared with algorithms that assumed a fixed shaped HRF. The MDL-based HRF estimation was applied on a real fMRI dataset and the HRF and activation maps were compared with AFNI. Since the proposed algorithm was region based the HRF estimates were smoother compared to HRF estimates from AFNI. The activation maps were comparable with the ones from AFNI.

CHAPTER 3

fMRI: JOINT MAXIMUM LIKELIHOOD ESTIMATION OF HRF AND ACTIVATION LEVEL¹

3.1 Overview

In this Chapter, the joint estimation of the hemodynamic response function and activation levels is presented. The proposed algorithm is validated on simulated data and two real data sets. The result of real data sets are compared with AFNI. Throughout this chapter vectors and matrices are denoted by bold lowercase and uppercase symbols, respectively. Vectors are defined as column vectors.

As mentioned in the introduction, knowledge of the hemodynamic response function is necessary for the reliable estimation of the activation levels in fMRI. Furthermore, estimation of the HRF requires knowledge of activation levels. The cross dependency motivates a joint approach for estimation of these quantities. The main contribution of this chapter is a maximum likelihood estimation (MLE) method that jointly estimates the HRF and activation. We apply regularization on the HRF estimation to allow our method to work with smaller amounts of data, thus making it possible to focus on smaller regions of interest and therefore better capture the variations of the HRF across the areas of the brain. The method is developed in the presence of either white noise or colored noise. Our method does not require prior knowledge for the probability distributions of the HRF or the activation levels. This is useful in situations where obtaining reliable and verifiable prior distributions on the HRF or the activation levels is difficult.

¹© 2014 Elsevier. Adapted/Reprinted with permission from N. Bazargani, and A. Nosratinia (2014). Joint maximum likelihood estimation of activation and hemodynamic response function for fMRI. *Medical Image Analysis* 18(5), 711-724.

The shape of the HRF is observed to be constant across neighboring voxels in a homogeneous region [20]. Thus a more robust HRF estimation may be obtained from the data corresponding to a set of voxels within a neighborhood, instead of a single voxel. In order to capture HRF variations, our joint estimation process is applied individually on each of the regions of interest.

It has been common for ROI-based analyses to depend on other methods to determine the ROI. In [65] and in [18] regions are first identified as functionally homogeneous parcels in the mask of the grey matter using a specific procedure. In [64] the input ROIs are SPM clusters obtained from maps based on standard Statistical Parameter Mapping (SPM) software using MARSBAR toolbox.² Makni et al. [65] also required parcellation and a knowledge of active regions.

In this chapter, two approaches are explored for establishing the ROI. The first approach is to deduce the ROI via analyzing the data with AFNI (Analysis of Functional NeuroImages) imaging analysis tool. The second approach does not involve using another algorithm as the input to ours, but rather uses the proposed joint estimation algorithm in a two-step “bootstrap” approach, as follows. In the first step, the brain is segmented into equal-size cubes and the joint MLE method is applied to each cube, without *a priori* assumptions about any of the regions. The resulting activation levels are then used to determine the ROI. The activation levels are discarded after determining the ROI, and then the ROI obtained from the first round are used in a second round of joint MLE, which determines the final values of both activation level and HRF. Thus, the joint MLE introduced in this dissertation can also be used as a stand-alone method.

Table 3.1 shows a comparison between the main features of this work with SPM software and AFNI toolbox.

²<http://www.sourceforge.net/projects/marsbar>

Table 3.1. Comparison of SPM, AFNI and proposed joint MLE

	SPM	AFNI	Joint MLE
Activation estimation	Yes. For each voxel.	No. It is not part of the model.	Yes. For each voxel.
HRF estimation	Fixed basis functions, estimates parameters. Fixed across voxels.	1. FIR - estimate for each voxel 2. Fixed <i>a priori</i> (no estimation)	Estimates HRF for a ROI.
Variable HRF	Time and dispersion derivation, or using FIR model. Yes(voxel-by-voxel).	Yes (voxel-by-voxel)	Yes (across ROIs)
Per Voxel or Region?	voxel-based	voxel-based	region-based
Multiple stimuli?	Yes	Yes (different HRF for each stimulus can be estimated)	Yes (different HRF for each ROI)
Noise correlation	Yes, Auto regressive (AR(1) for classical, for Bayesian AR with order defined in the process)	Can model ARMA(1,1) with 3dREMLfit	AR(1)
Summary	Low flexibility in estimation of HRF via basis functions.	More flexibility in HRF via FIR, but noisy estimate of HRF.	Joint estimation of HRF and activation, ROI-based, regularized. More reliable HRF estimate for low SNR regions.

Due to the level of noise and limitation on the data points arising from a region-dependent model on the HRF, we also consider smoothing constraints. This is, in part, motivated by the existing work in the literature, which we briefly outline below. It has been shown by Buxton [15] that the HRF is smooth. Earlier work on the use of smoothness constraint in the HRF estimation include smoothing a finite impulse response [45], as well as constraints on the second order discrete derivative operator [67, 20]. In addition, Tikhonov regularization [88] has been proposed for imposing smoothness constraints on the HRF estimate. Two approaches based on using Tikhonov regularization and smoothing splines are introduced in [95] and [89]. In both papers, the HRF estimation was done voxel-wise on a block design experimental paradigm. The impact of regularization constraint on the HRF estimates was studied in [17] by comparing regularization-based techniques to the least squares estimation methods.

3.2 Methods

3.2.1 Formulation

The time series at voxel j at time t is represented as Equation (2.4) without the drift component

$$y_j(t) = \alpha_j (S(t) * h(t)) + e_j(t), \quad e_j(t) \sim N(0, \mathbf{C}) \quad (3.1)$$

where \mathbf{y}_j is a $N \times 1$ vector, N represents the number of scans, t is the time which is between 1 and N , activation level at that voxel is represented by α_i , $s(t)$ is the stimulus time series which is convolved with the HRF represented by $h(t)$, and $e_j(t)$ is a zero-mean Gaussian random variable that represents the observation noise with the covariance matrix \mathbf{C} . In this formulation we only model activation signal and Gaussian noise, so other artifacts such as drift are removed from the measured signal. In this formulation only one stimulus time series is applied to the subject. The solution for the multiple stimuli experiment is discussed in Section 3.2.7.

We assume that the HRF is approximately invariant across small regions, which we denote as regions of interest. The activation levels within each ROI may vary. The size of ROI is denoted with M . The combined time series equation for a ROI is:

$$\mathbf{Y} = \mathbf{S} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix} [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_M] + \mathbf{E} \quad (3.2)$$

where \mathbf{Y} is a $N \times M$ matrix of observations, whose column \mathbf{y}_j is the time series observed at voxel j . The HRF is represented by $[h_1 \cdots h_p]^t$. In agreement with the existing results as well as common sense, we assume the HRF starts from zero and decays to zero eventually as shown in Figure 1.4. The time-sampled version of the HRF is represented as a finite impulse response (FIR) filter that has p nonzero values. The column vector $\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_M]^t$ represents the activation levels, \mathbf{E} contains the observation noise. For the purposes of exposition this noise is at first assumed to be white, and then the results are extended to the case of temporally correlated noise. In fMRI the noise is spatially correlated [91], as well. However, since the goal of this study is to look at activation levels and HRF estimations, we simplify the noise model by just accounting for the temporal correlation. The spatially uncorrelated noise model is prevalent in works where HRF is estimated, including [2, 17, 45, 59, 64, 65].

The matrix \mathbf{S} represents the stimulus data and has a Toeplitz structure in order to implement convolution by matrix multiplication. Specifically:

$$\mathbf{S} = \begin{bmatrix} s(1) & 0 & \cdots & 0 \\ s(2) & s(1) & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ s(N) & s(N-1) & \cdots & s(N-p) \end{bmatrix} \quad (3.3)$$

3.2.2 Rank-one Approximation in White Noise

By assuming that the noise is independent, identically distributed Gaussian $e_j(t) \sim \mathcal{N}(0, \sigma^2)$, the likelihood function for the voxel time series in a region with M voxels becomes

$$p(\mathbf{Y}; \mathbf{h}, \boldsymbol{\alpha}, \sigma) = \prod_{j=1}^M p(\mathbf{y}_j; \mathbf{h}, \alpha_j, \sigma) \quad (3.4)$$

$$\propto \prod_{j=1}^M \sigma^{-N} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y}_j - \mathbf{S} \mathbf{h} \alpha_j\|_2^2\right) \quad (3.5)$$

where σ is the standard deviation of the noise at each of the voxels in the ROI. It is assumed that the amount of noise is the same in each ROI. Maximizing the likelihood leads to the following optimization:

$$\min_{\mathbf{h}, \boldsymbol{\alpha}} \sum_{j=1}^M \|\mathbf{y}_j - \mathbf{S} \mathbf{h} \alpha_j\|_2^2 \quad (3.6)$$

where \mathbf{y}_j is the time series vector of voxel j , with activation level α_j . Recalling that $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_M]$, Equation (3.6) can be converted from a sum of vector norms to its equivalent in Frobenius norm, transforming the optimization problem as follows:

$$\min_{\mathbf{h}, \boldsymbol{\alpha}} \|\mathbf{Y} - \mathbf{S} \mathbf{h} \boldsymbol{\alpha}^T\|_F^2 \quad (3.7)$$

Since the product of two matrices is limited in its rank to the lower of the constituent matrix ranks, the product $\mathbf{S}(\mathbf{h} \boldsymbol{\alpha}^T)$ is a rank-one matrix. Thus, the problem represented by expression (3.7) is essentially to find a rank-one approximation to the observations \mathbf{Y} . From the Eckart and Young theorem [31] it follows that a low-rank approximation either in the Frobenius norm or the spectral norm is given by a truncated sum of the singular value decomposition of the original matrix. Specifically, the answer to:

$$\min_{\mathbf{B}} \|\mathbf{Y} - \mathbf{B}\|_F \quad (3.8)$$

$$\text{s.t. } \text{rank}(\mathbf{B}) \leq p \quad (3.9)$$

is given by:

$$\mathbf{B}_{opt} = \sum_{i=1}^p \eta_i \mathbf{u}_i \mathbf{v}_i^T$$

where η_i are the ordered singular values and \mathbf{u}_i and \mathbf{v}_i are the respective left and right singular vectors of matrix \mathbf{Y} . For a rank-one approximation, only one of the terms will be retained, therefore:

$$\mathbf{B}_{opt} = \mathbf{S} \hat{\mathbf{h}} \hat{\boldsymbol{\alpha}}^T = \eta_1 \mathbf{u}_1 \mathbf{v}_1^T$$

from which we set $\hat{\boldsymbol{\alpha}}^T = \eta_1 \mathbf{v}_1^T$ and $\mathbf{S} \hat{\mathbf{h}} = \mathbf{u}_1$, the estimates are as follows:

$$\hat{\boldsymbol{\alpha}} = \eta_1 \mathbf{v}_1 \quad \hat{\mathbf{h}} = \mathbf{S}^\dagger \mathbf{u}_1 \quad (3.10)$$

where \mathbf{S}^\dagger is the pseudo inverse of \mathbf{S} . To resolve the scaling ambiguity between \mathbf{h} and $\boldsymbol{\alpha}$, a normalization constraint is imposed on $\hat{\mathbf{h}}$, i.e., $\|\hat{\mathbf{h}}\|_2 = 1$.

3.2.3 Regularized Rank-one Approximation in White Noise

We now turn to the practical question of the required data for the reliable computation of the maximum likelihood solution. In our problem, this requires special care because we wish to obtain a space-varying HRF in homogeneous regions, which implies that the voxels whose observations are used to obtain the maximum likelihood solution must be concentrated around a certain locality. This imposes a practical upper limit on the number of voxels that can be used in our estimation.

To solve this problem, we turn to earlier investigations which have shown that the HRF is a smooth function of time [15]. We then try to incorporate this knowledge into our algorithm by imposing a smoothing constraint on the estimated HRF using the Tikhonov regularization technique [88]. The cost function to be minimized in this case is

$$\min_{\mathbf{h}, \boldsymbol{\alpha}} \{ \|\mathbf{Y} - \mathbf{S} \mathbf{h} \boldsymbol{\alpha}^T\|_F^2 + \lambda \|\mathbf{D} \mathbf{h}\|_2^2 \} \quad (3.11)$$

where λ is the regularization parameter. The smoothness constraint is imposed by a discrete second derivative matrix \mathbf{D} which is defined as

$$\mathbf{D} = \begin{bmatrix} -2 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & -2 & 1 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & 1 & -2 & 1 \\ 0 & \cdots & \cdots & 0 & 1 & -2 \end{bmatrix} \quad (3.12)$$

We use a regularized rank-one matrix approximation to minimize this penalty cost function. We first find the rank-one matrix approximation of \mathbf{Y} , which is $\mathbf{Y}^{(1)} = \eta_1 \mathbf{u}_1 \mathbf{v}_1^T$. We define $\mathbf{u} = \mathbf{u}_1$ and $\mathbf{v} = \eta_1 \mathbf{v}_1$. By substituting \mathbf{h} with $\mathbf{S}^\dagger \mathbf{u}_1$ and $\boldsymbol{\alpha}$ with \mathbf{v} , the iterative estimation procedure will be simplified. The minimization problem is transformed to:

$$\min_{\mathbf{u}, \mathbf{v}^T} \{ \|\mathbf{Y} - \mathbf{u}\mathbf{v}^T\|_F^2 + \lambda \|\mathbf{L}\mathbf{u}\|_2^2 \} \quad (3.13)$$

where $\mathbf{L} = \mathbf{D}\mathbf{S}^\dagger$. Now the vectors \mathbf{u} and \mathbf{v} are updated iteratively until convergence.³ A relaxation method is used to solve the minimization problem, resulting in the following update equations.

$$\begin{aligned} \mathbf{u}^{(k+1)} &= (\mathbf{v}^{(k)T} \mathbf{v}^{(k)} \mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{Y} \mathbf{v}^{(k)} \\ \mathbf{v}^{(k+1)} &= \frac{\mathbf{Y}^T \mathbf{u}^{(k+1)}}{\mathbf{u}^{(k+1)T} \mathbf{u}^{(k+1)}} \end{aligned} \quad (3.14)$$

After convergence and finding the optimum \mathbf{v} and \mathbf{u} , the estimates of \mathbf{h} and $\boldsymbol{\alpha}$ are

$$\hat{\boldsymbol{\alpha}} = \mathbf{v}_{\text{opt}} \quad \hat{\mathbf{h}} = \mathbf{S}^\dagger \mathbf{u}_{\text{opt}} \quad (3.15)$$

The computation of the update Equations (3.14) for a regularized rank-one minimization problem in Equation (3.13) is as follows. Our approach to find vectors \mathbf{u} and \mathbf{v} is found via a

³This is similar to the technique used in [82].

relaxation method, specifically, by the successive optimization of \mathbf{u} while \mathbf{v} is held constant, and vice versa.

First, for a fixed \mathbf{v} we optimize \mathbf{u} . The minimization criterion (3.13) can be rewritten as

$$\sum_j \sum_i (y_{ji} - u_j v_i)^2 + \lambda \mathbf{u}^T \mathbf{L}^T \mathbf{L} \mathbf{u} \quad (3.16)$$

By expanding the squares, we have

$$\min_{\mathbf{u}} \left\{ \sum_j \sum_i y_{ji}^2 - 2 \sum_j \sum_i y_{ji} u_j v_i + \sum_j \sum_i u_j^2 v_i^2 + \lambda \mathbf{u}^T \mathbf{L}^T \mathbf{L} \mathbf{u} \right\} \quad (3.17)$$

The first term does not have any effect on the minimization. Therefore, we need to minimize

$$\min_{\mathbf{u}} \left\{ -2 \mathbf{u}^T \mathbf{Y} \mathbf{v} + \mathbf{u}^T \mathbf{u} \mathbf{v}^T \mathbf{v} + \lambda \mathbf{u}^T \mathbf{L}^T \mathbf{L} \mathbf{u} \right\} \quad (3.18)$$

By taking the derivative with respect to \mathbf{u} and setting it to zero, we get

$$-2 \mathbf{Y} \mathbf{v} + 2 \mathbf{v}^T \mathbf{v} \mathbf{u} + 2 \lambda \mathbf{L}^T \mathbf{L} \mathbf{u} = 0 \quad (3.19)$$

In each iteration k , we can use the equation above to calculate \mathbf{u} in terms of observed values as well as the latest $\mathbf{v}^{(k)}$, the value of \mathbf{v} in iteration k .

$$\mathbf{u}^{(k+1)} = (\mathbf{v}^{(k)T} \mathbf{v}^{(k)} \mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{Y} \mathbf{v}^{(k)} \quad (3.20)$$

We now perform the remaining task, namely to find an update for \mathbf{v} given the latest calculated \mathbf{u} . By expanding the squares in a manner similar to Equation (3.17) and ignoring the terms that do not involve \mathbf{v} ,

$$\min_{\mathbf{v}} \left\{ -2 \sum_j \sum_i y_{ji} u_j v_i + \mathbf{u}^T \mathbf{u} \mathbf{v}^T \mathbf{v} \right\} \quad (3.21)$$

which is equivalent to

$$\min_{\mathbf{v}} \left\{ -2 \mathbf{Y}^T \mathbf{u} \mathbf{v} + \mathbf{u}^T \mathbf{u} \mathbf{v}^T \mathbf{v} \right\} \quad (3.22)$$

By taking derivative with respect to \mathbf{v} and setting it to zero

$$-2\mathbf{Y}^T\mathbf{u} + 2\mathbf{u}^T\mathbf{u}\mathbf{v} = 0 \quad (3.23)$$

The solution is

$$\mathbf{v}^{(k+1)} = \frac{\mathbf{Y}^T\mathbf{u}^{(k+1)}}{\mathbf{u}^{(k+1)T}\mathbf{u}^{(k+1)}}. \quad (3.24)$$

Selecting a suitable regularization parameter, λ , is an important issue. One of the prominent techniques that is used to find near-optimum regularization parameter is the generalized cross-validation criterion [42]. However, due to the nature of the cost function in Equation (3.13) and optimizing over two vectors, the derivation of cross validation is mathematically intractable. Therefore we introduce another approach for choosing the regularization parameter. We calculate the mean square error between the benchmark and the estimated HRF for each value of λ for the simulated data set and test 200 different values for λ , in the range from zero to 10^7 with linear spacing. The implementation is done in MATLAB, on Intel core2, 2.0 GHz processor. The computation time for selecting the optimum λ depends on the number of voxels in a ROI and ranges from 60 to 72 seconds for $M = 50$ and $M = 100$ voxels, respectively. The regularization parameter that corresponds to the least mean squares error is selected as the optimum value. The process of choosing the optimum λ was done on the block design and event related simulated data sets. The chosen optimum regularization parameter for the event related design was used for the real event related fMRI data experiments.

It is noteworthy that the basic matrix analysis and optimization techniques used in this section have a long and storied history, but their application to the specific problem at hand and the adaptation to the requirements of our problem are novel. In passing, we note the distinction of our approach with respect to earlier works that also used smoothing, e.g., [82] which finds the sparse principal component in their analysis. The work in [82] solved the problem for a sparseness constraint while we modify the regularization term to be the smoothing constraint and solve for the regularized rank-one matrix approximation.

3.2.4 Rank-one Approximation in Colored Noise

Various results, e.g., [92], indicate that the additive noise in fMRI is not purely white, but rather is better modeled by a random sequence that is correlated in time. Therefore we extend our results to capture the more accurate model for the noise in fMRI.

Using a formulation for a voxel time series as Equation (3.1), the likelihood function for all the voxels time series in the ROI becomes

$$p(\mathbf{Y}; \mathbf{h}, \boldsymbol{\alpha}, \mathbf{C}) = \prod_{j=1}^M p(\mathbf{y}_j; \mathbf{h}, \alpha_j, \mathbf{C}) \quad (3.25)$$

$$\propto \prod_{j=1}^M |\mathbf{C}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}_j - \mathbf{S} \mathbf{h} \alpha_j)^T \mathbf{C}^{-1} (\mathbf{y}_j - \mathbf{S} \mathbf{h} \alpha_j)\right) \quad (3.26)$$

where $\mathbf{h}, \boldsymbol{\alpha}$ and \mathbf{C} , the noise covariance matrix, are the unknown parameters. We model the fMRI noise as a first order autoregressive (AR) process. The use of AR(1) noise model has precedents in the literature, e.g., [14]. We assume that noise statistics are stationary within a homogeneous region of neighboring voxels, similar to a global AR(1) which was proposed in [34]. However, no assumptions are made in our work about global variations of the noise statistics. It follows that the noise covariance matrix of the voxels in a given region of interest is fixed and is represented by \mathbf{C} . For a AR(1) temporal noise model, $\mathbf{C} = \sigma_e^2 \boldsymbol{\Gamma}$, where σ_e^2 is the variance of the input white Gaussian noise in the AR(1) model and $\boldsymbol{\Gamma}$ is a matrix that its ℓ th element is $\gamma_\ell = \rho^{|\ell|}$, where ρ is the delay-one autocorrelation of noise. By substituting \mathbf{C} in Equation (3.26), the maximum likelihood solution takes the following form

$$\min_{\mathbf{h}, \boldsymbol{\alpha}, \boldsymbol{\Gamma}} \left\{ \sum_{j=1}^M (\mathbf{y}_j - \mathbf{S} \mathbf{h} \alpha_j)^T \boldsymbol{\Gamma}^{-1} (\mathbf{y}_j - \mathbf{S} \mathbf{h} \alpha_j) + \ln \boldsymbol{\Gamma} \right\} \quad (3.27)$$

Because $\boldsymbol{\Gamma}^{-1}$ is a positive definite matrix, it can be decomposed to square roots, i.e., $\boldsymbol{\Gamma}^{-1} = \mathbf{W}^T \mathbf{W}$. By substituting in Equation (3.27),

$$\min_{\mathbf{h}, \boldsymbol{\alpha}, \mathbf{W}} \left\{ \sum_{j=1}^M \|\mathbf{W}(\mathbf{y}_j - \mathbf{S} \mathbf{h} \alpha_j)\|_2^2 + \ln \boldsymbol{\Gamma} \right\} \quad (3.28)$$

Or,

$$\min_{\mathbf{W}} \min_{\mathbf{h}, \boldsymbol{\alpha}} \{ \|\mathbf{W}\mathbf{Y} - \mathbf{W}\mathbf{S}\mathbf{h}\boldsymbol{\alpha}^T\|_F^2 + \ln\Gamma \} \quad (3.29)$$

It should be noted that \mathbf{W} itself is completely characterized by one scalar parameter ρ . Several values of ρ in the range of 0 to 1, as mentioned in [93], are tested. The minimization is done in two steps. For each ρ , the matrix \mathbf{W} is generated. First we optimize the inner minimization with respect to \mathbf{h} and $\boldsymbol{\alpha}$, which gives a result which is a function of \mathbf{W} (or ρ). In the second step, the minimum over \mathbf{W} (effectively a scalar optimization over ρ) is found.

3.2.5 Regularized Rank-one Approximation in Colored Noise

The optimization for the colored noise case is as follows

$$\min_{\mathbf{W}} \min_{\mathbf{h}, \boldsymbol{\alpha}} \{ \|\mathbf{W}\mathbf{Y} - \mathbf{W}\mathbf{S}\mathbf{h}\boldsymbol{\alpha}^T\|_F^2 + \lambda\|\mathbf{D}\mathbf{h}\|_2^2 + \ln\Gamma \} \quad (3.30)$$

The noise model is AR(1) as the previous section and for each correlation coefficient ρ , the matrix \mathbf{W} is generated. For each \mathbf{W} , we optimize over \mathbf{h} and $\boldsymbol{\alpha}$ using the regularization algorithm explained in Section 3.2.3. Then we minimize over \mathbf{W} (effectively a scalar optimization over ρ) and the optimum values are obtained for the activation levels and the HRF estimates.

3.2.6 Iterative Estimation

In previous sections all the voxels in the region were used in the estimation of the HRF and were treated as active voxels. We extend the estimation algorithm to the case where inactive voxels or insignificantly activated voxels also exist in the ROI. Inactive voxels harm the reliability of the HRF and activation levels estimate. We address this problem by iterating between the estimation of the HRF and the activation levels.

The iterative process consists of two halves. In the first half-iteration, the HRF and activation levels are estimated via a rank-one matrix approximation in the same way discussed in the previous sections considering all the voxels in the ROI. This gives an initial

estimate for activation levels and the HRF. In the second half-iteration, the most recently estimated HRF is convolved with the stimulus time series and used as a design matrix in a generalized linear model (GLM). The new estimates for activation levels is obtained by the least squares estimation that is well-known in the literature. We then calculate Student's t-test for each voxel in a ROI to perform statistical inference on activation levels. A threshold for t-values is selected according to the desired significance level. The t-values for each voxel are converted to P-values based on the degrees of freedom of the dataset. The significance level of $P\text{-value} = 0.001$ which is corrected for multiple comparisons is chosen for selecting the t-test threshold. The Bonferroni correction method [1] is used for multiple comparison correction, where for a single voxel the significance level is calculated by dividing the significance level 0.001 by the number of the tests (voxels). If P-value of each voxel is less than $0.001/\text{Number of tests}$, then the voxel is considered active and the null hypothesis is rejected.

Those voxels with t-tests less than a threshold will be excluded from the next iteration. The iterations will stop when no further improvement on the HRF estimate is observed. After convergence, the last HRF estimate is used for calculating the t-tests for all of the voxels in the region.

It is important to select a good threshold for t-test values in order to exclude inactive voxels. If the threshold is set too high, the number of the participating voxels decreases, which makes the estimates less reliable, while setting the threshold too low will include too many low-quality voxels which will also harm the estimate. The iterative process outlined in this section can be used with any of the algorithms proposed previously. In particular, it can be used in the context of white or colored noise, and with or without regularization constraint. The effect of the selection of the t-test threshold on the estimations is studied and the results are shown in Figure 3.10, which appears later in this Chapter. The convergence of the iterative algorithm is studied by examining the changes in the estimated HRF from one iteration to another. At most 10 iterations are sufficient to achieve convergence.

3.2.7 Extension to Multiple Stimuli

All of the above proposed algorithms can be extended to multiple stimuli. If the responses to different stimuli are in different brain locations, the estimations are straightforward. For each stimulus we find the joint MLE solution in every ROI. Then for each ROI and each stimulus, we evaluate the objective function with its corresponding estimates. For each stimuli the estimates which give the lowest error at a given ROI, are selected as the response to that stimuli.

In the case where two stimuli give responses in the same region, the same approach as mentioned above can still work for the event related design as long as the stimulus time series are non-overlapping in time. There are several examples in the literature for the non-overlapping, orthogonal designs [47].

3.2.8 Region of Interest Selection

For real fMRI data two approaches are used to obtain ROIs.

- The first method is to use AFNI toolbox and analyze the data to locate active areas of the brain. Then use `3dclust` program in AFNI to find the active clusters. These ROIs from AFNI are used as inputs for the joint MLE algorithm.
- The second approach is to segment the brain into equal size cubes. The optimal cube size for this segmentation reflects a trade-off. The cubes (ROIs) should be small enough to guarantee the invariance of the HRF within each cube but large enough to contain reliable information for the estimation. Cubes of size 27 to 125 voxels represent a reasonable trade-off that produce good results. This cube segmentation uniformly divides the space.

For the second approach which is the analysis with the brain segmentation method, the joint MLE method is applied to each cube separately, so the cubes act as a rough

approximation to desired regions. The resulting activation levels are then used to determine the active regions. The ROIs from the first round are used in a second round of joint MLE to estimate the final activation levels and HRF inside each ROI. In other words, we can “bootstrap” our method by running it in the first round with cube-shaped regions.

3.3 Results

3.3.1 Simulated Data

In the first experiment, the proposed algorithm was applied to a simulated fMRI time series that was constructed by convolving a stimulus time series with a benchmark HRF. Our benchmark HRF was constructed by forming the difference of two Gamma functions with known parameters as in [41] which is mentioned in Equation (2.10). The simulated time series also reflects the magnitude of activation levels and a Gaussian noise. The noise statistics were fixed for the voxels in the ROI and were chosen such that the SNR was within the typical range of the SNR in the real fMRI data, which is between 0.2 and 1 [52]. The SNR was defined as $\|\mathbf{S}\mathbf{h}\alpha_i\|^2/(N \times \sigma^2)$. We applied our algorithm on both block design and event related experiments. For the block design experiment, the stimulus time series had a period of 60 seconds, with 30 seconds on and 30 seconds off intervals. For the event related experiment, there were 51 stimuli with random inter stimulus intervals. The time series had $N = 300$ samples. M different time series for the voxels in a ROI were generated in this way. We performed experiments with white Gaussian noise as well as with colored Gaussian noise having first order autoregressive model.

Simulated data with white Gaussian noise

The MLE algorithm with rank-one approximation was applied to the simulated data with additive white Gaussian noise. The results are presented for both regularized and non-regularized cases in block design and event related experiments.

To evaluate the estimate of the HRF, Monte Carlo simulation was performed on a block design paradigm and also on an event related paradigm. In these simulations $\text{SNR} = 0.2$, the number of voxels was $M = 50$, and Monte Carlo simulation was performed $Q = 500$ times. The mean of the HRF estimate and the variance of the estimation at each point are shown in Figures 3.1 and 3.2 for block design and event related paradigms. The HRF estimates are presented with and without imposing the smoothing constraint. It is observed that the smoothness constraint improves the accuracy of the mean of the HRF estimate (closer to the benchmark HRF). Inclusion of the smoothing constraint also improves the estimation variance. In the interest of brevity the results are reported only at a single SNR, but results at a wide range of SNRs have been obtained that show a similar trend.

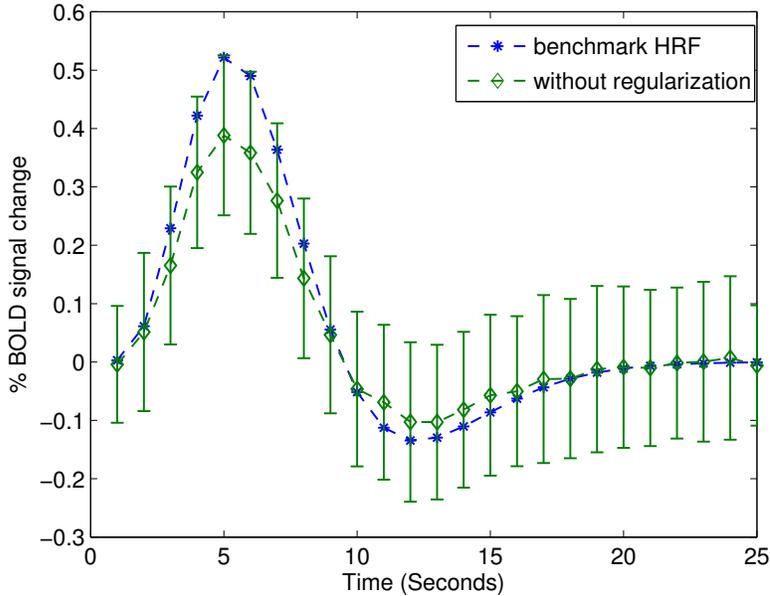
For the error analysis, we generated fMRI time series for one voxel with block design paradigm and performed $Q = 500$ Monte Carlo realizations over the entire ROI. The activation levels were drawn from a Normal distribution, $\alpha_i \sim \mathcal{N}(3, 0.01)$. The mean square error between the benchmark and the estimated HRF, and the MSE between the benchmark and the estimated activation levels were calculated. First, an average solution for the estimated HRF, $\bar{\mathbf{h}}$, over $Q = 500$ realizations was computed. Then the estimated variance was approximated by $\mathcal{V}(\bar{\mathbf{h}})$, and the mean square error was calculated as follows

$$\bar{\mathbf{h}} = \frac{1}{Q} \sum_{q=1}^Q \hat{\mathbf{h}}_q \quad (3.31)$$

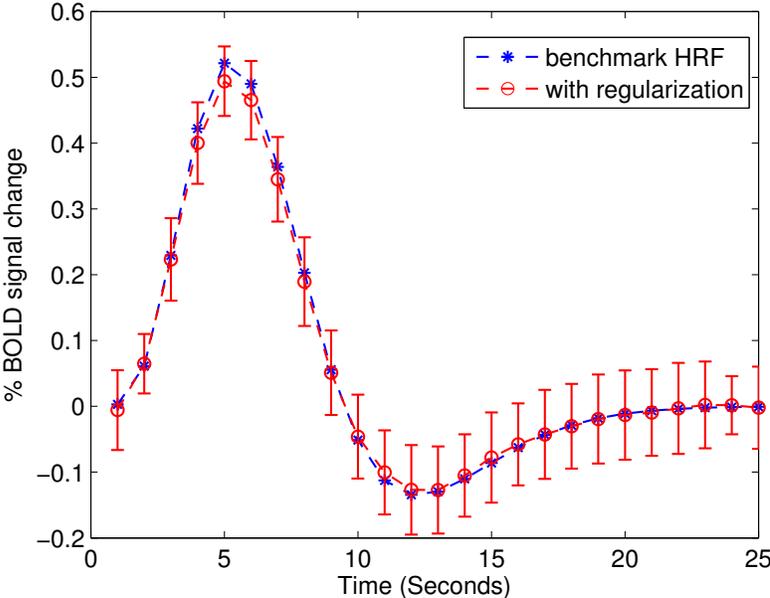
$$\mathcal{V}(\bar{\mathbf{h}}) = \frac{1}{Q} \sum_{q=1}^Q (\hat{\mathbf{h}}_q - \bar{\mathbf{h}})^2 \quad (3.32)$$

$$\text{MSE}(\hat{\mathbf{h}}; \mathbf{h}) \approx \mathcal{V}(\bar{\mathbf{h}}) + (\mathbf{h} - \bar{\mathbf{h}})^2 \quad (3.33)$$

Figures 3.3 and 3.4 show the MSE between the benchmark and the estimated HRF with no regularization for a block design experiment and for an event related experiment. Experimental data are produced for several SNRs and for observation sets containing $M = 50$ and 100 voxels. The results show that the performance of estimation improves with both SNR

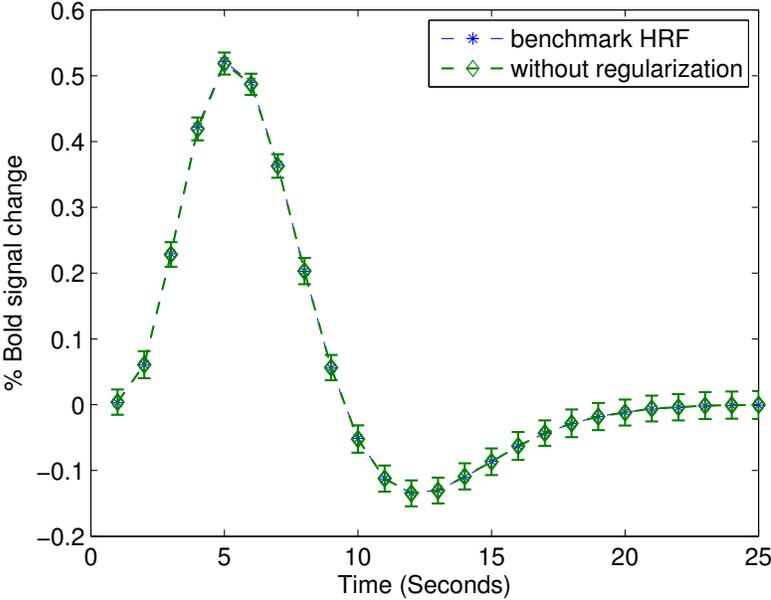


(a)

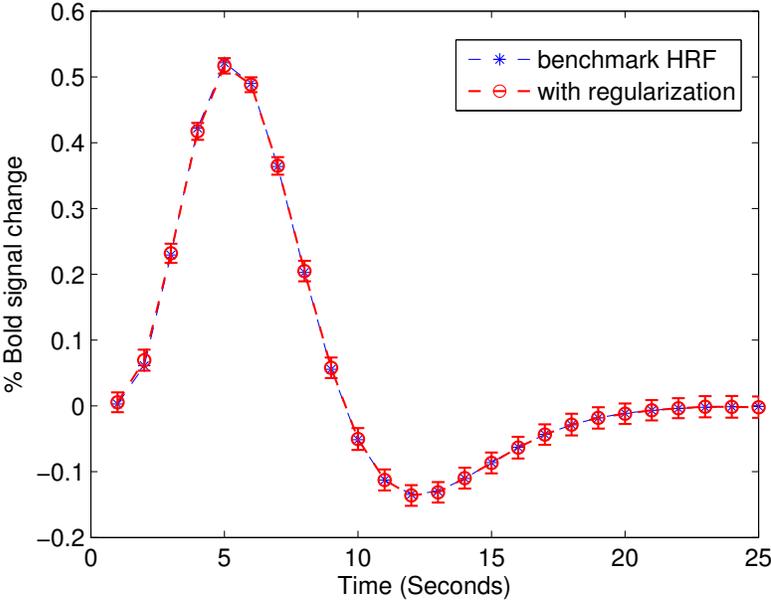


(b)

Figure 3.1. Average of the HRF estimate and the variance of the estimate at each point on a ROI with 50 voxels, with SNR = 0.2 for (a) a block design experiment without regularization, (b) a block design experiment with regularization.



(a)



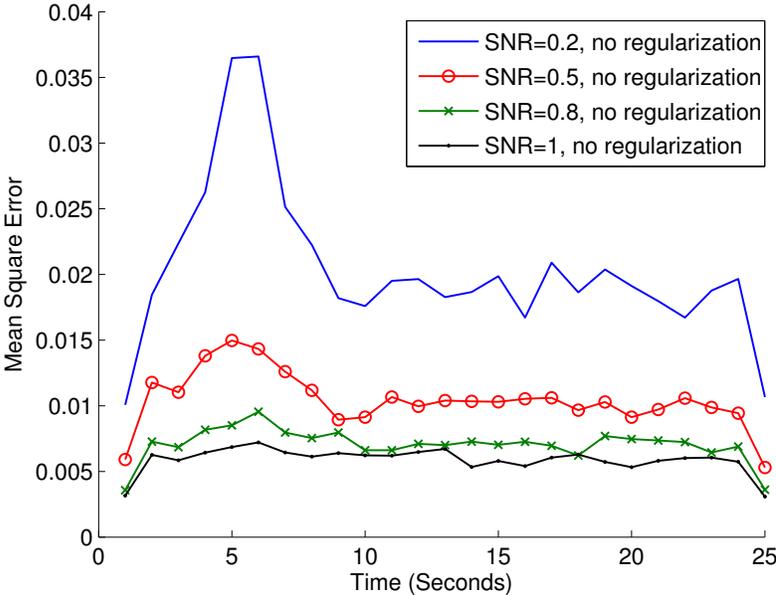
(b)

Figure 3.2. Average of the HRF estimate and the variance of the estimate at each point on a ROI with 50 voxels, with $SNR = 0.2$ for (a) an event related experiment without regularization and (b) an event related experiment with regularization.

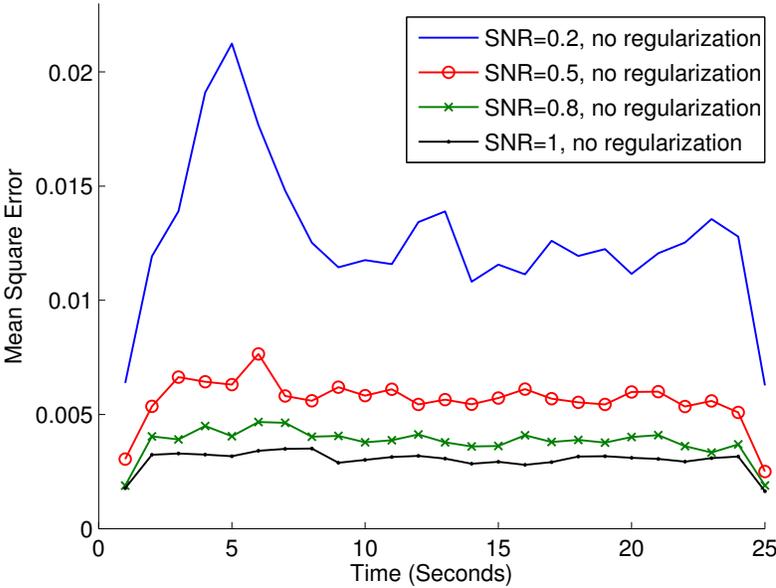
and M . The reported SNR at each case is the average of SNR among voxels in the ROI. The performance of the method including regularization constraint are shown in Figures 3.5 and 3.6 for observation sets containing $M = 50$ and 100 voxels for various SNR with white noise assumption. By imposing the smoothing constraint on the HRF shape, improvements are achieved in the form of smaller MSE between the benchmark and average of the HRF estimates compared to non-regularized case.

Similar error analysis experiments were performed with the block design paradigm and this time we evaluated the activation level estimates. The region of interest had $M = 100$ voxels and $\text{SNR} = 0.5$. The benchmark coefficients were drawn from a Gaussian distribution with mean 3 and variance of 0.01. For each voxel, the mean and variance of the activation level were calculated via Monte Carlo. For each voxel in the ROI, the average of the estimated activation level and the estimation variance were computed. In Figure 3.7 (a) the absolute value of the difference between the benchmark and estimated activation levels at each voxel are plotted. It is observed that the regularization constraint reduces the estimation error. The variances of the estimated activation levels are shown in Figure 3.7 (b). The regularization constraint results in a smaller estimation variance. Experiments at other SNR levels were also performed which verify the expected trends with respect to SNR and number of voxel observations, but were not included in this dissertation in the interest of brevity. The performance of the estimation algorithm on estimating activation levels was evaluated on the event related paradigm using Monte Carlo simulation, as well. It is observed that the regularization constraint improves the HRF estimate as well as the activation levels estimate. It also improves (reduces) the estimation variance. These experiments show that while event related and block design experiments both benefit from regularization, the improvements are more pronounced in the latter case.

In the case that voxels in a ROI have larger differences between their activation levels, the same results for error analysis with Monte Carlo have been obtained. The variances of

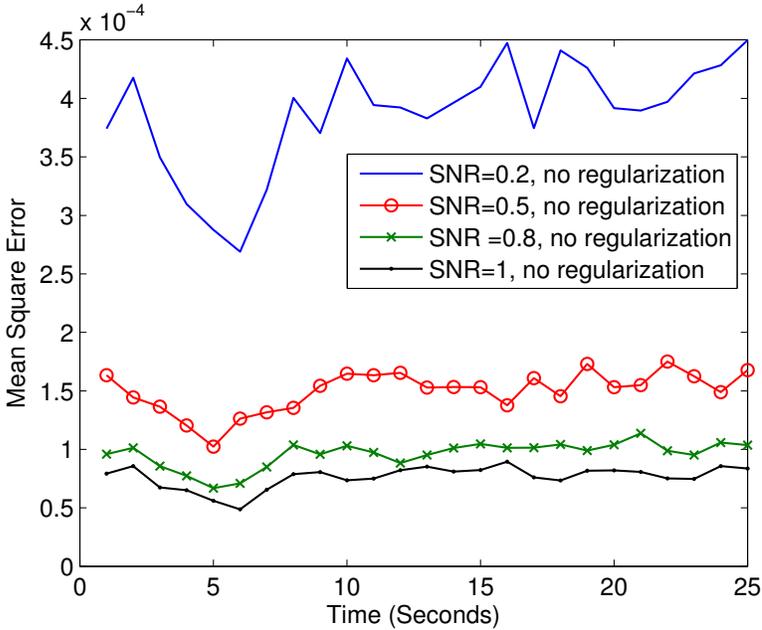


(a)

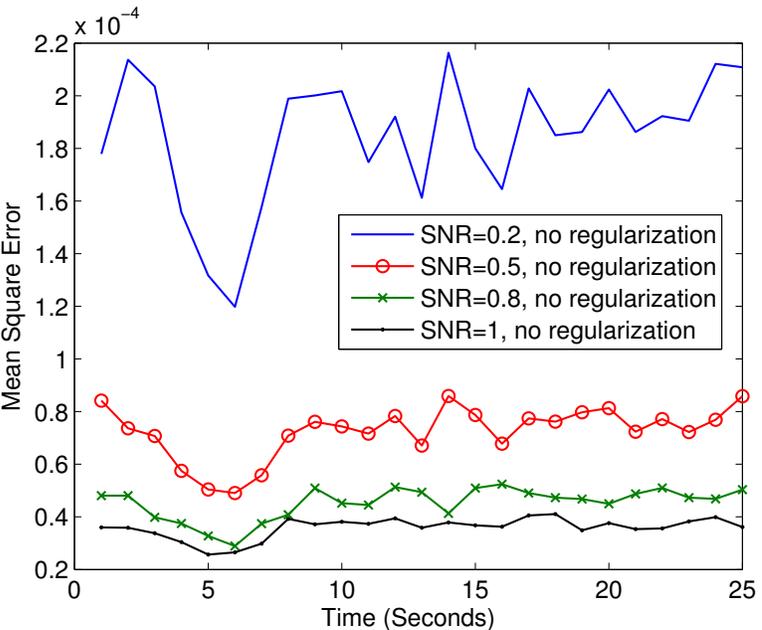


(b)

Figure 3.3. Mean square error between the benchmark HRF and the estimated HRF with different SNRs for a block design experiment with white noise assumption for (a) $M=50$ and (b) $M=100$ voxels.

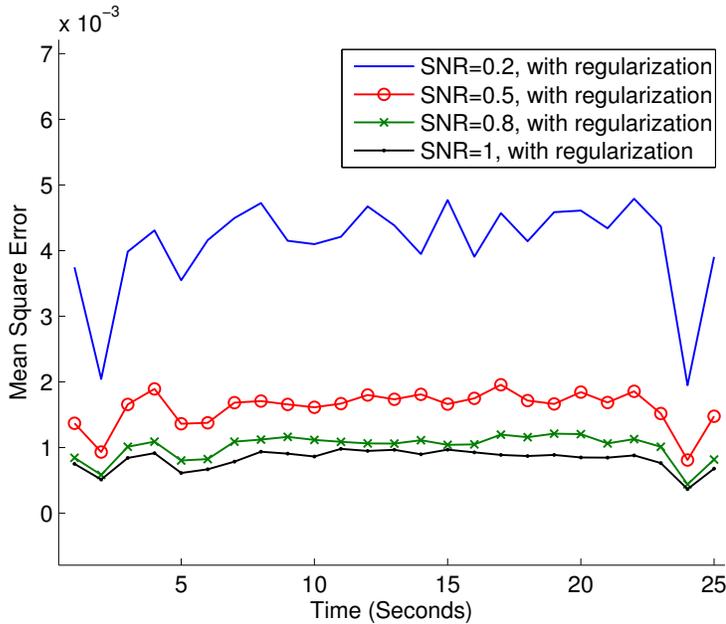


(a)

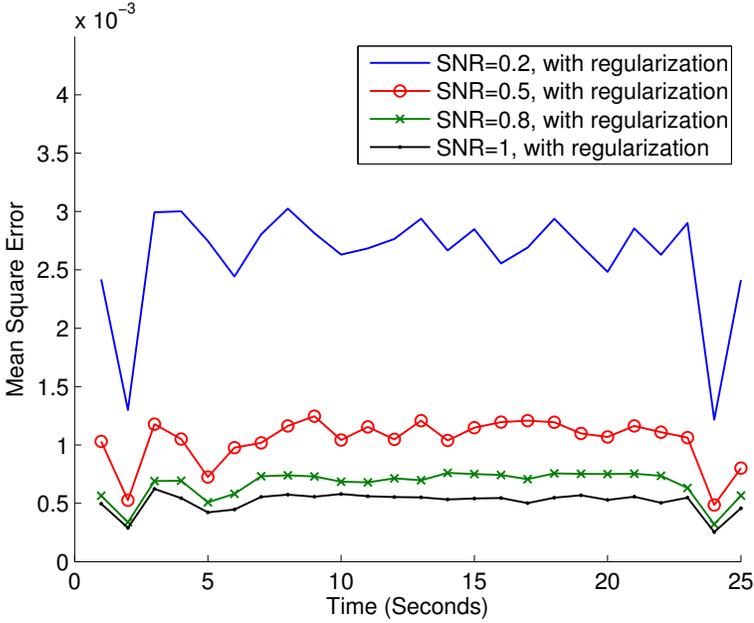


(b)

Figure 3.4. Mean square error between the benchmark HRF and the estimated HRF with different SNRs for an event related experiment with white noise assumption for (a) M=50 and (b) M=100 voxels.

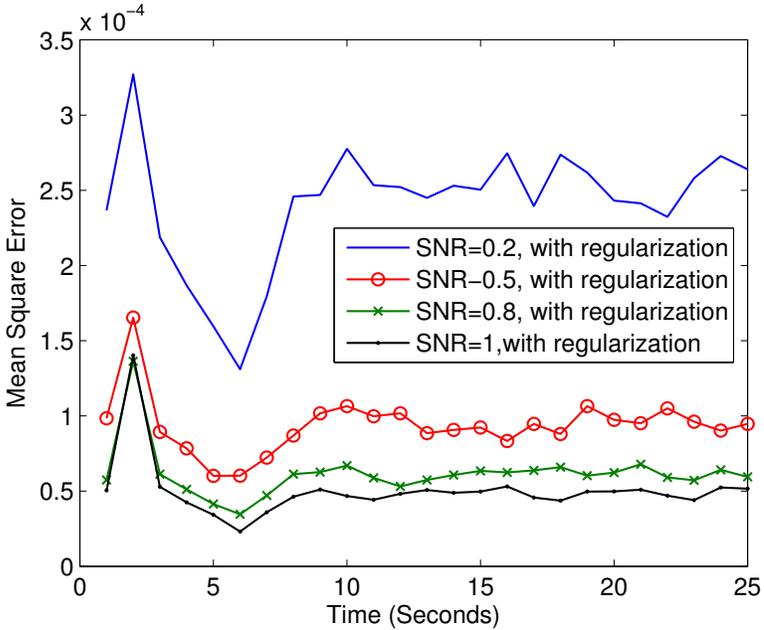


(a)

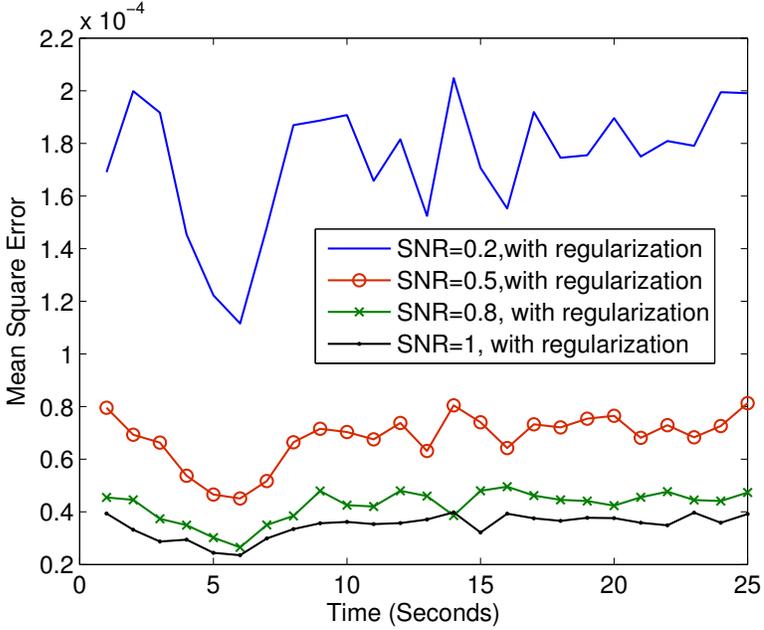


(b)

Figure 3.5. Effect of regularization constraint on MSE between the benchmark HRF and the estimated HRF with different SNRs for a block design experiment on (a) M=50 and (b) M=100 voxels.

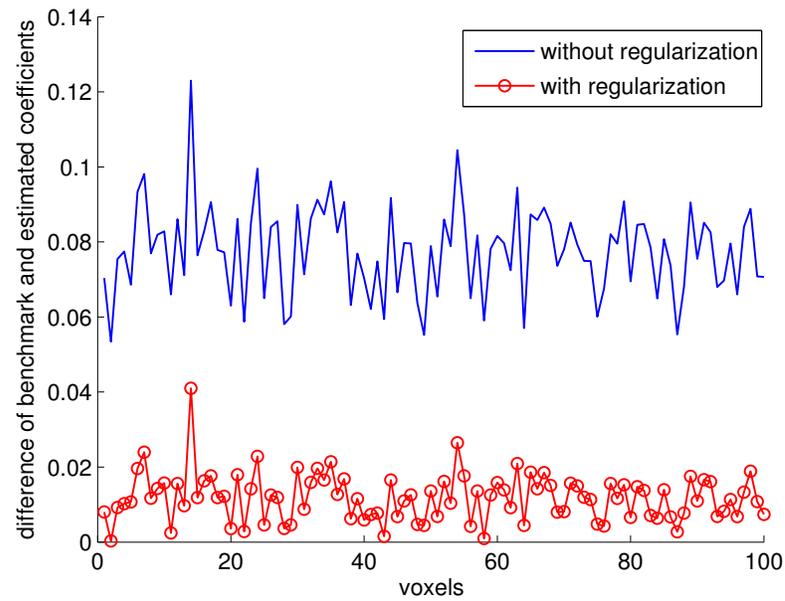


(a)

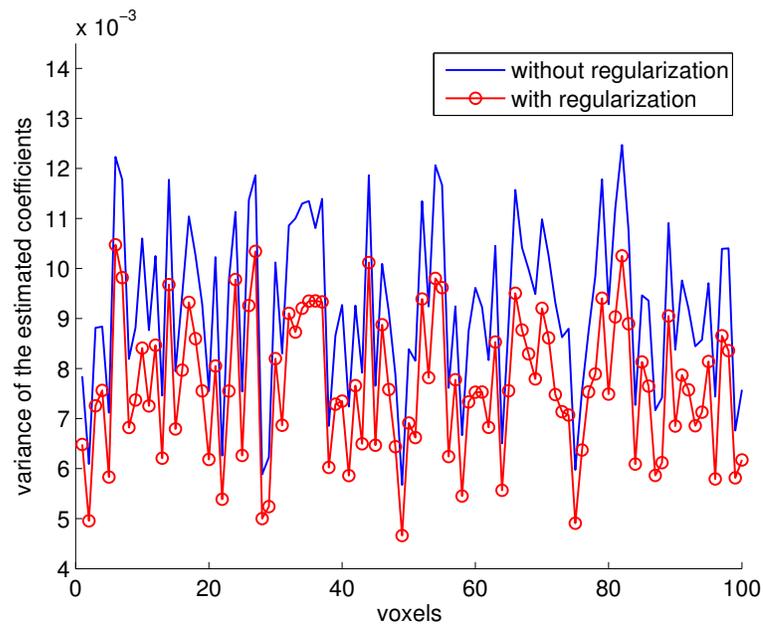


(b)

Figure 3.6. Effect of regularization constraint on MSE between the benchmark HRF and the estimated HRF with different SNRs for an event related experiment on (a) M=50 and (b) M=100 voxels.



(a)



(b)

Figure 3.7. Comparing the estimated activation levels by imposing regularization for a block design experiment with 100 voxels in the ROI and SNR=0.5. (a) Absolute value of the difference between estimated and benchmark activation levels. (b) Variance of the activation levels estimates.

Table 3.2. Monte Carlo comparison between SPM and joint MLE. Average of mean square error between benchmark and estimated values are shown for block and event related designs for a ROI with 100 voxels.

average of MSE	SNR = 0.5	SNR = 0.8	SNR = 1
$\text{HRF}_{\text{NoReg}}$, Block design	0.0078	0.0049	0.0041
HRF_{Reg} , Block design	0.0071	0.0044	0.0037
HRF_{SPM} , Block design	0.0162	0.0162	0.0162
$\text{HRF}_{\text{NoReg}}$, Event related	1.04×10^{-4}	6.43×10^{-5}	5.27×10^{-5}
HRF_{Reg} , Event related	9.95×10^{-5}	6.16×10^{-5}	5.02×10^{-5}
HRF_{SPM} , Event related	0.0162	0.0162	0.0162
$\hat{\alpha}_{\text{NoReg}}$, Block design	0.2093	0.0907	0.0684
$\hat{\alpha}_{\text{Reg}}$, Block design	0.1832	0.0812	0.0617
$\hat{\alpha}_{\text{SPM}}$, Block design	0.3169	0.2974	0.2963
$\hat{\alpha}_{\text{NoReg}}$, Event related	0.0657	0.0387	0.0297
$\hat{\alpha}_{\text{Reg}}$, Event related	0.0657	0.0387	0.0297
$\hat{\alpha}_{\text{SPM}}$, Event related	0.3077	0.2716	0.2565

the estimates are higher compared to the case with smaller difference between activation levels.

We compared the performance of the rank-one MLE method with the Statistical parameter mapping (SPM) software which uses the generalized linear model (GLM) analysis. In SPM-GLM the HRF is pre-calculated using a mixture of Gamma functions and is fixed during the analysis, while in our approach the HRF is estimated jointly with activation levels. The comparison was on simulated data corrupted by white Gaussian noise. In these simulations the number of voxels was $M = 100$, activation levels were drawn from a Normal distribution $\mathcal{N}(3, 0.1)$ and Monte Carlo simulation was performed $Q = 500$ times. Several SNR values have been tested for block design and event related experiments. The average of the mean square error which was calculated in Equation (3.33), was computed for the HRF estimate and activation level estimates and reported in Table 3.2. Comparing the results show smaller average MSE for the joint MLE approach compared to SPM-GLM. The smallest average MSE is obtained with the regularized MLE approach.

Experiments with Colored Noise

As mentioned earlier, the noise correlations are modeled by a first-order autoregressive AR(1) model. No regularization constraint was imposed on the HRF in this experiment.

For the error analysis, $M = 50$ simulated fMRI time series were generated with the block design paradigm. The benchmark HRF was convolved with the stimuli, the result was multiplied by a benchmark activation level and then colored Gaussian noise with AR(1) model was added to it. The SNR used in this simulated experiment was selected to be in line with the SNR observed in actual fMRI data using present-day technologies. We performed $Q = 500$ Monte Carlo realizations over the entire ROI with M voxels. The estimation results were compared for various values of the correlation coefficient for noise. The goal of this experiment was to study the effect of whitening the colored noise on the HRF and activation level estimation.

The correlation coefficient $\rho = 0.4$ is similar to the noise correlation observed in the real fMRI data sets. For a block design experiment with $\text{SNR} = 0.3$, $\rho = 0.4$ and the activation levels drawn from a $\mathcal{N}(3, 0.01)$, no significant difference is observed in the HRF estimate by accounting for the colored noise. This result is in agreement with the observations made by previous investigators, e.g., [65, 68], which report that under the above-mentioned conditions the noise model has little impact on the estimated HRF shape.

We also investigated the effect of colored noise on detection performance. Figure 3.8 shows the t-tests calculated in three different scenarios: using the benchmark (ideal) noise correlation coefficient, estimating the noise correlation coefficient, and ignoring the correlation of noise (assuming it is white even though it is not). It is observed that taking into account the noise correlation result in t-values closer to the t-value calculated with the benchmark covariance.

We repeated this experiment for various other values of correlation coefficient and SNR, and the results demonstrated the same trend as seen in Figure 3.8. The improvements in the

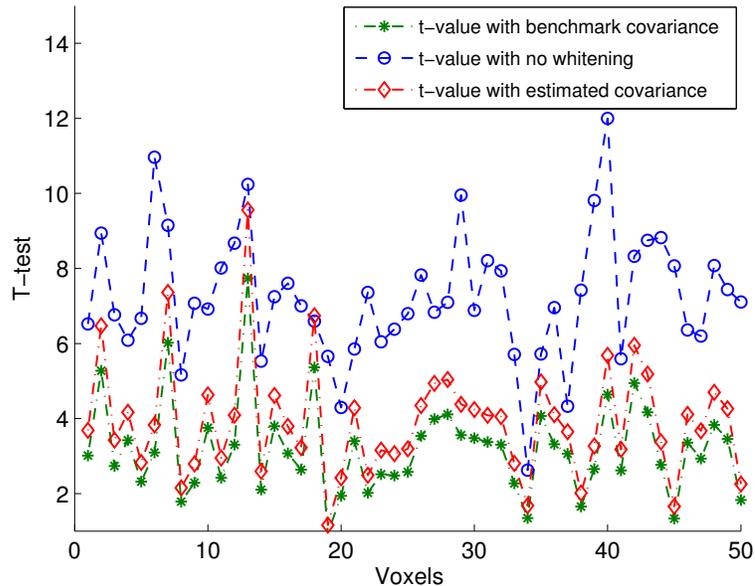


Figure 3.8. T-tests calculated with the benchmark correlation coefficient, compared with the t-tests calculated with and without whitening.

t-values diminish as the SNR increases. This is also in agreement with the result reported in [65]: that using an AR(1) noise model at low SNR slightly improves the detection.

Although colored noise models yield a small improvement at low SNR for activation detection, our simulations show that the corresponding improvements for the HRF estimation are not at all significant.

Iterative Estimation

We applied our iterative algorithm on a simulated data generated as follows: an ROI was constructed with 40 active and 10 inactive voxels. The observation length was $N = 300$ and observations were contaminated with additive white Gaussian noise. The iterative estimation algorithm was implemented with a regularization constraint.

For an event related paradigm with $SNR = 0.2$ and activation levels drawn from a normal distribution $\alpha_i \sim \mathcal{N}(3, 0.01)$, Figure 3.9 (a) shows the benchmark HRF as well as

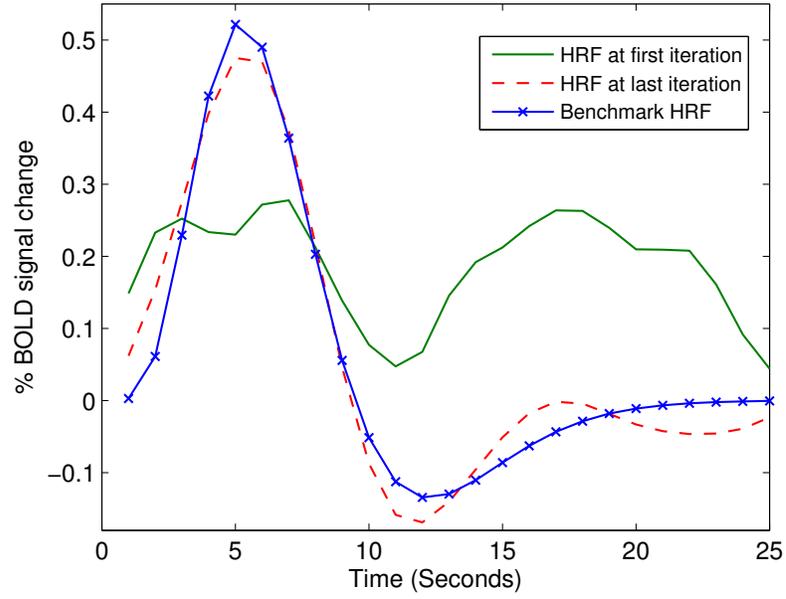
the HRF estimates at the first and last iterations. The t-tests are shown in Figure 3.9 (b). The HRF estimate and activation detection performance improve through iterations. The t-test statistics improve through iterations and approach the t-tests calculated with the benchmark HRF. We also performed simulations for SNR from 0.2 to 1. The results show that the advantage of the iterative algorithm is more pronounced at lower SNRs. We note that this was not unexpected, since the method was devised in the first place to address the small data sets resulting from a region-based estimation of the HRF. We observed the same improvement in the HRF and activation level estimations for the block design paradigm.

We investigated the effect of choosing the t-test threshold on the estimations for a ROI with 40 active and 10 inactive voxels. Figure 3.10 (a), shows the estimated HRF for a block design experiment with the SNR of 0.8 for different t-test thresholds. The activation levels were drawn from a Gaussian distribution with mean 2 and variance 0.5. The results for an event related design are shown in Figure 3.10 (b), the SNR is 0.2 for this experiment. By setting a high t-test threshold, the number of voxels whose t-test exceeds the threshold decreases and the estimates become less reliable. The significance level of $P\text{-value} = 0.001$ corrected for multiple comparisons, corresponds to a t-test threshold which will result in reliable estimates.

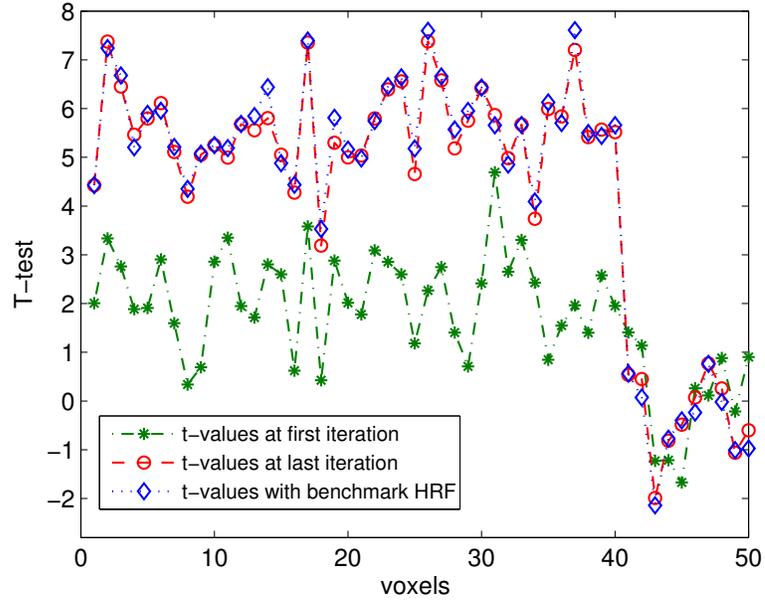
3.3.2 Real fMRI Data: Event Related experiment

Analysis with active ROIs from AFNI

The first real data is the same dataset used in Section 2.3.2. The data were acquired from a healthy subject at the University of Texas Southwestern at Dallas on a 3-T scanner. The experiment was an event related paradigm with 27 auditory stimuli, the events were jittered with random inter stimulus intervals. The voxel size was $3.125\text{ mm} \times 3.125\text{ mm} \times 4\text{ mm}$ and the number of scans was 160, repetition time (TR) was 2 seconds and the only

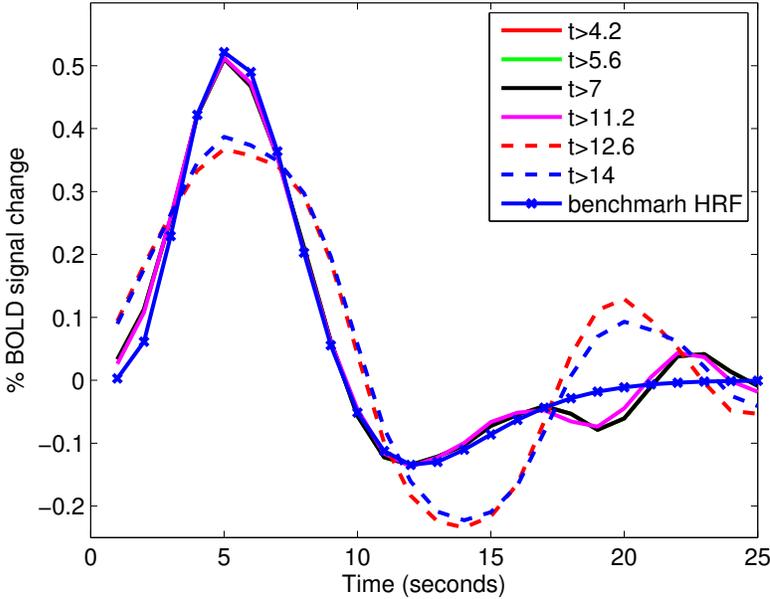


(a) HRF estimates

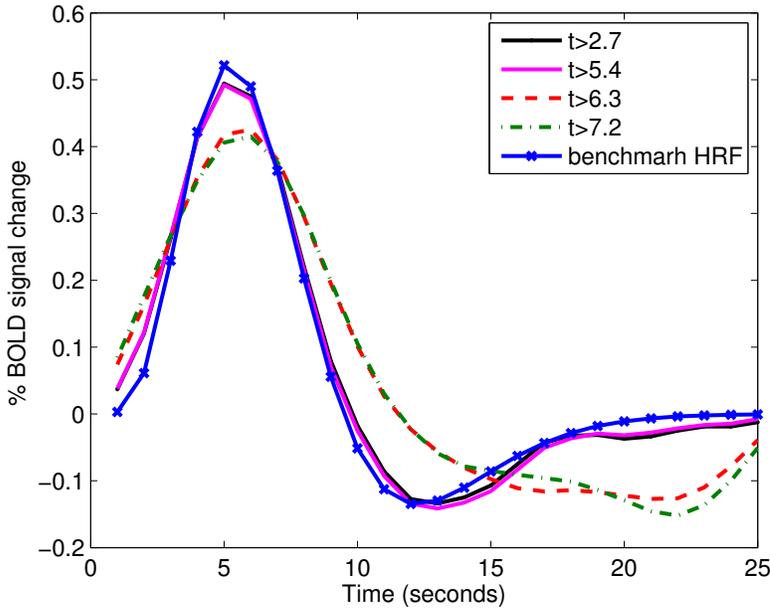


(b) T-tests

Figure 3.9. (a) HRF estimates and (b) t-tests for a ROI with 50 voxels including 40 active voxels. One realization of event related with $\text{SNR}=0.2$ and white noise assumption.



(a)



(b)

Figure 3.10. Effect of the t-test threshold on the estimated HRF, for the iterative algorithm in a ROI with 40 active and 10 inactive voxels for (a) a block design experiment with SNR = 0.8, (b) an event related experiment with SNR = 0.2.

preprocessing step done on the data was the registration. The size of the whole brain data set was $64 \times 64 \times 40$ voxels.

The regions of interest were obtained as follow. The data was detrended (using Legendre polynomials of degree 3) and then analyzed with AFNI toolbox using `3dDeconvolve` to locate the active areas of the brain. The criteria for defining the voxel clusters were the threshold for the activation map and the size of the cluster. The active contiguous voxels which were above these specific thresholds were defined as regions of interest. The program `3dclust` in AFNI toolbox finds clusters that fit the criteria. We set the threshold for the F-test to be greater than 3.045 and the number of voxels in each cluster to be greater than 10 voxels. We obtained 4 ROIs with these thresholds.

The iterative algorithm with regularization constraint and with the color noise assumption was applied to each ROI. The HRF and activation levels for each of the voxels in each region were estimated. The Student's t-test was calculated for each voxel.

The activated brain regions identified by our approach were compared with the activation maps from AFNI. The axial and coronal views of the activation maps for one of the ROIs (ROI 1) are shown in Figure 3.11. The cross-hair in Figure 3.11 shows the voxel of interest. The subfigure (a) shows activation maps from our MLE approach and Figure 3.11 (b) shows the HRF estimate with our iterative approach for that voxel and its corresponding ROI. The bottom row shows the region of interest and the estimated HRF obtained from AFNI. The activation maps are superimposed on the anatomical data. The HRF estimate with AFNI for the same voxel is shown in Figure 3.11 (d). The results on the two other ROIs are shown in Figure 3.12 and Figure 3.13, respectively.

The subfigure (c) in each of these figures shows the ROI from AFNI, which is the input to the MLE. Thus, the activate regions from iterative joint MLE algorithm which are shown in subfigure (a) should be consistent with AFNI activation maps in subfigure (c). Our method is not too far from AFNI when signal to noise ratio (SNR) is good, because in that case

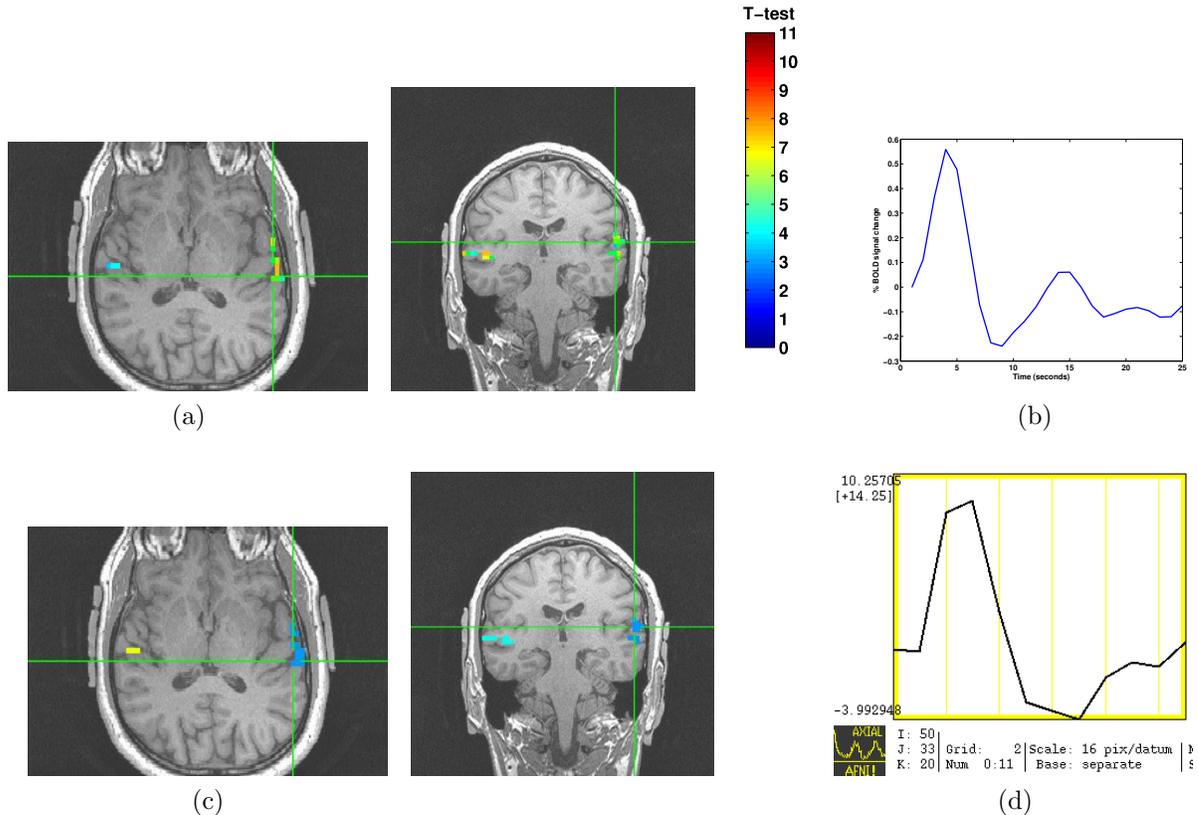


Figure 3.11. Axial and coronal views of a single slice activation maps for ROI 1. (a) activation maps from the MLE method, (b) estimated HRF from MLE method, (c) ROI obtained from AFNI, (d) estimated HRF from AFNI.

the precise determination of the HRF is not crucial since at high SNR the activation is somewhat robust to the precise form of the HRF. At very high SNR even a single voxel may have enough information to produce a sufficiently close approximation of the HRF estimate. The differences between our algorithm and AFNI, one would expect, are not in entirely different activation, but rather on the voxels in the activation region where the SNR is lowest. Thus it is not surprising that the overall location of the new activation is not different from AFNI, but the shape of the new region is slightly different and improved. In AFNI the activation analysis as well as the HRF estimation are performed voxel-wise. In our approach, the HRF estimation is performed over a region. The HRF is estimated in AFNI via a finite impulse response filter.

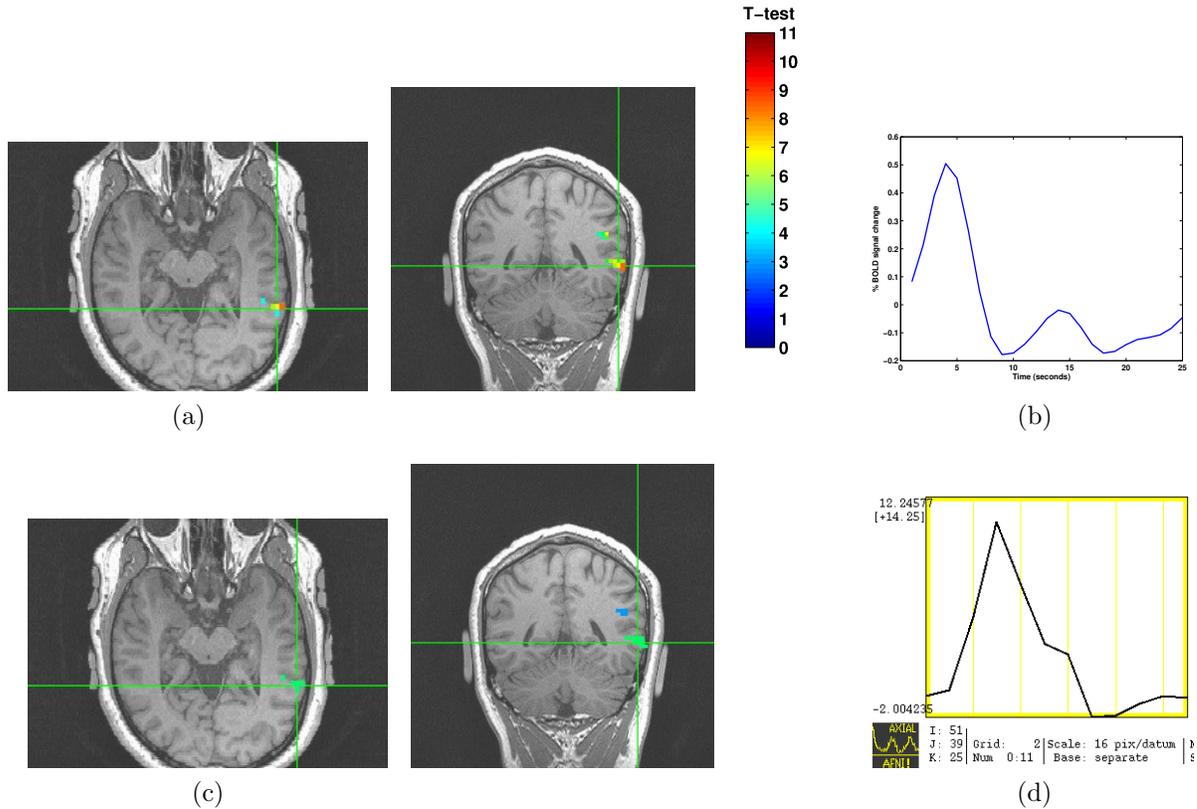


Figure 3.13. Axial and coronal views of a single slice activation maps for ROI 3. (a) activation maps from the MLE method, (b) estimated HRF from MLE method, (c) ROI obtained from AFNI, (d) estimated HRF from AFNI.

regions from first round were used in second round of joint MLE to estimate final activation levels and HRF.

The activation maps and HRF estimates from joint MLE were compared with the results from AFNI. Figure 3.14 subfigure (a) shows coronal and sagittal views of the t-test activation map for ROIs defined as active with joint MLE algorithm. The HRF estimate for the ROI which is shown with crosshairs is in subfigure (b). The activation map from AFNI which is F-test are shown in subfigure (c), the HRF estimate from AFNI is shown in subfigure (d). Figure 3.14 shows that even without prior knowledge of active ROI, our joint MLE approach is capable of detecting active regions. As mentioned previously, in AFNI the activation

analysis as well as the HRF estimation are performed voxel-wise. In our approach, the HRF estimation is performed over a region.

We observe slight differences with AFNI, which are explained by the differences between F-test and t-test and the differences between the algorithms used in AFNI and our method. However, we note that varying amounts of difference may be observed between AFNI and our method, for example when signal to noise ratio is low on a per-voxel basis, but the active ROI is not too small. In such cases, a more reliable HRF estimate and hence more reliable estimate of activation level, can be obtained by our proposed joint MLE method since it is ROI-based.

3.3.3 Real fMRI Data: Block Design experiment

The second real fMRI data is the “attention to visual motion” data set available on the SPM web site. The data set was used with permission from Methods group at the Wellcome Trust Centre for Neuroimaging. We used this dataset in Section 2.2.2. This data set was obtained by Christian Büchel and is described in [13]. The experiment was a block design consisted of 4 conditions: fixation, static, no attention and attention. Each of these conditions were blocks of 32.2 seconds. The TR was 3.22 seconds. There were 360 scans in this study. The experiment was performed on a 2 Tesla Magnetom VISION (Siemens, Erlangen) whole-body MRI system equipped with a head volume coil. Contiguous multislice T2*-weighted fMRI images, with 64×64 pixels, 32 slices and slice thickness of 3 mm. Image processing and statistical analysis were carried out using SPM8 and MATLAB. All volumes were realigned to the first volume and a mean image was created using the realigned volumes. A structural MRI image, acquired using a standard three-dimensional T1 weighted sequence ($1 \times 1 \times 3$ mm voxel size) was co-registered to this mean (T2) image. Finally all the images were spatially normalized to a standard Tailarach template and smoothed using a 6 mm full width at half maximum (FWHM) isotropic Gaussian kernel. Figure 3.15 shows the stimuli during

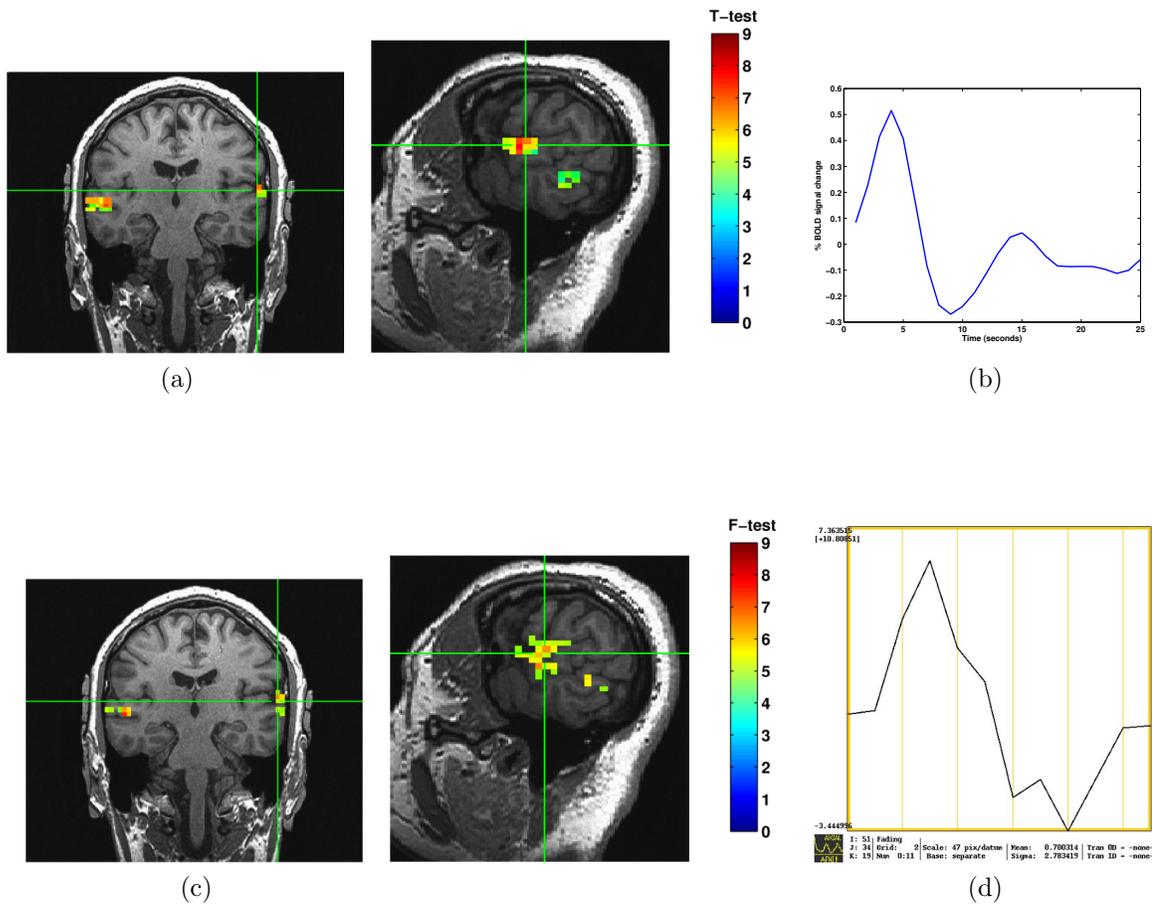


Figure 3.14. Coronal and sagittal views of a single slice activation maps. (a) activation maps from the MLE method without prior knowledge on active ROI, (b) estimated HRF from MLE method, (c) activation map based on AFNI 3dDeconvolve algorithm, (d) estimated HRF from AFNI.

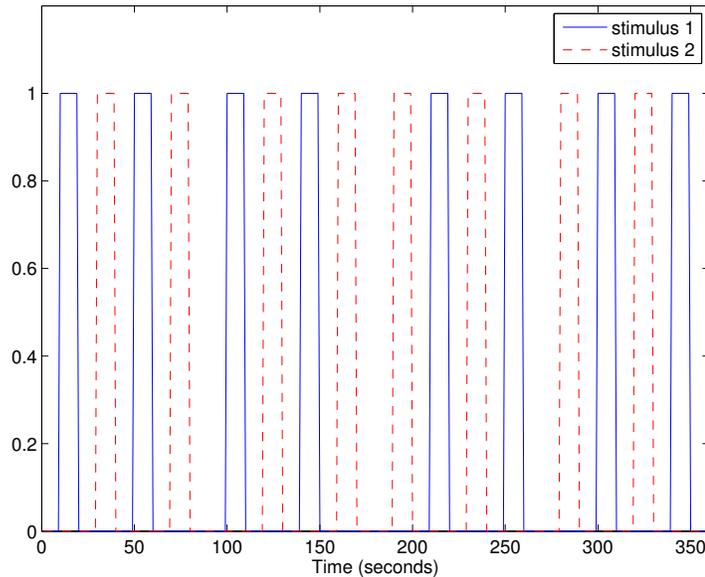


Figure 3.15. Fixation (Stimulus 1) and Attention (Stimulus 2) conditions over the duration of the scans. Amplitude of 1 represents the stimulus was on.

fixation and attention conditions. The fMRI time series was detrended with the third degree polynomial before applying the joint-MLE algorithm.

The data were analyzed using joint MLE with two methods for ROI selection as mentioned in Section 3.2.8. The first method was using AFNI to analyze the data and get active ROI. The data was analyzed with AFNI toolbox using `3dDeconvolve` to locate the active areas of the brain. The criteria for defining the clusters were the threshold for the activation map and the size of the cluster. The active contiguous voxels which were above these specific thresholds were defined as regions of interest. The program `3dclust` in AFNI toolbox finds clusters that fit the criteria. We set the F-test threshold for this data set to be greater than 25 and the number of voxels in each cluster to be greater than 25. We obtained one ROI with these thresholds. The analysis from joint MLE with the two different input methods for ROI selection was compared with AFNI results.

The second method was using the two-step “bootstrap” approach. The brain was segmented into equal-size cubes of $5 \times 5 \times 5$ voxels. The iterative algorithm with colored noise assumption was applied to each cube separately. For this multiple stimuli case the stimuli were orthogonal and $S_i^\dagger S_j = 0$ for each stimulus i and j . We applied the iterative MLE algorithm for each stimuli and estimate the HRF and activation levels for each stimuli as mentioned in Section 3.2.7. The active regions were determined based on the threshold on the t-test values. The threshold was chosen based on desired significance level of 0.001. These active regions were used as ROI input for the second round. Figure 3.16 shows activation maps and HRF estimates from joint MLE with ROIs from bootstrap method (subfigures a and b), AFNI analysis (subfigures c and d) and joint MLE with ROIs from AFNI (subfigures e and f). The HRF estimate for AFNI is for the voxel which is shown by crosshair. The HRF estimates for the joint MLE are for the whole ROI. The activation maps in subfigure (a) show that even without prior knowledge of active ROI, the joint MLE approach is capable of detecting active ROIs. The shape and location of the active regions are similar between AFNI analysis and joint MLE approach. The slight differences between our joint MLE results and AFNI results are because of differences between F-test and t-test, differences between algorithms in AFNI analysis and our method, and using voxels versus regions of interest for the analysis. Subfigures (b) and (f) show the HRF estimates for the two stimuli shown in Figure 3.15. As can be seen if there are multiple stimuli joint MLE approach can be used to estimate the HRF for all the individual stimuli as long as the stimuli are orthogonal.

3.4 Conclusion

This chapter introduced a maximum likelihood method for jointly estimating the activation levels and the hemodynamic response function for a region of interest in fMRI without using prior probability distributions on the HRF of activation levels. This enabled capturing variabilities in the HRF across the brain regions. The proposed algorithm can use ROIs

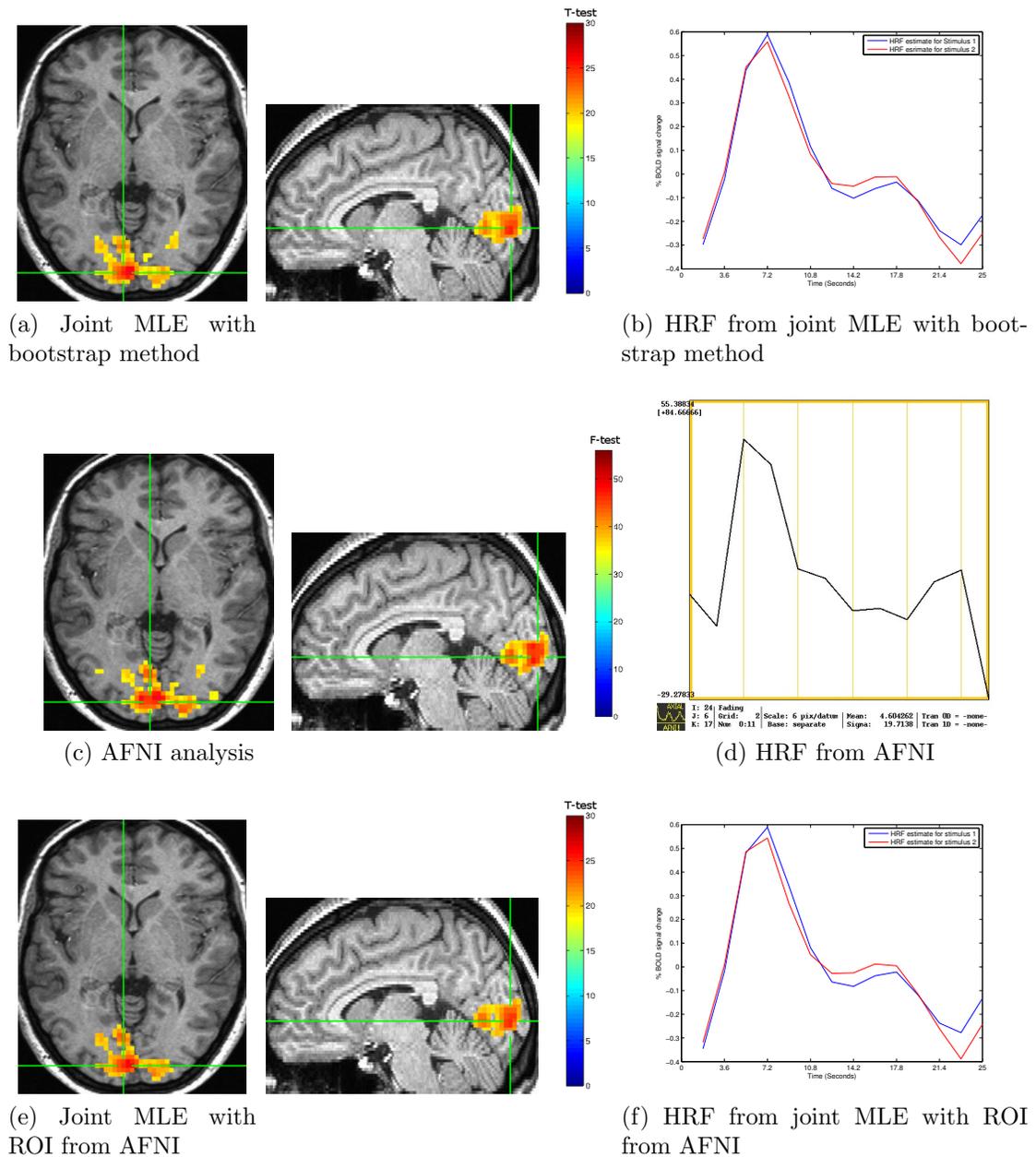


Figure 3.16. Axial and Sagittal views of a single slice activation maps for the block design attention to visual stimuli data set. Activation maps are for Stimulus 1. (a) activation maps from the MLE method without prior knowledge on active ROI, (b) estimated HRF from MLE method, (c) activation map based on AFNI 3dDeconvolve algorithm, (d) estimated HRF from AFNI, (e) activation maps from the MLE method with ROI from AFNI, (f) estimated HRF from MLE method with ROI from AFNI.

from AFNI or any other analysis tool as input. It also had the ability to work without pre-defined ROIs by applying our algorithm on segmented cubes in the first round. A rank-one approximation was used for estimating both the HRF and activation levels in a computationally tractable manner. A smoothness constraint was imposed on the HRF estimate to improve the results under limited observation data, which often happens in a region-based estimation of the HRF. Both block design and event related paradigms can be used with this method.

We investigated the effect of colored noise. Our results showed that the accurate modeling of noise autocorrelation did not significantly affect the HRF estimation, which was consistent with the results reported earlier in the literature.

To have accurate HRF estimates, inactive voxels must be excluded from HRF estimation, but the activation estimation itself is dependent on HRF values. To address this problem, an iterative method for the estimation of the HRF and activation values was introduced. The algorithms were tested using simulated and real data over a wide range of parameters and on block design and event related experiments.

CHAPTER 4

COMPUTED TOMOGRAPHY: MODELING AND CORRECTION FOR SCATTER RADIATION FROM DOSE COMPENSATOR¹

The focus of this chapter is to model the scatter radiation from the dose compensator using Monte Carlo (MC) simulations and apply the correction to simulated data. Scatter correction is part of the pre-processing steps applied to CT data.

4.1 Background

For better understanding of the scattered radiation artifact, an introduction of x-ray interaction with matter and projection data is given.

The interaction of x-ray with matter results in attenuation of the x-ray beam traveling through a medium. The attenuation of x-ray photons can be expressed by an exponential relationship for a monochromatic (monoenergetic) incident x-ray beam by Beer-Lambert law

$$I = I_0 e^{-\int_L \mu(l) dl} \quad (4.1)$$

where I and I_0 are the incident and transmitted x-ray intensities, L is the thickness of the material, $\mu(l)$ is the linear attenuation coefficient of the material at location l . The linear attenuation coefficient represents the sum of the possible interaction processes: photoelectric effect, Rayleigh scattering and Compton scattering and is a function of the incident x-ray photon energy [49]. The line integral of the linear attenuation coefficient along the x-ray

¹© 2012 SPIE, Adapted/Reprinted with permission from N. Bazargani, R. A. Thompson, and Y. Yoad (2012, February). Modeling scattered radiation from dose compensator in CT by forced detection Monte Carlo simulation. In Proc. SPIE 8313, San Diego. 83132L

trajectory through the object is described as

$$p = -\ln \frac{I}{I_0} = \int_L \mu(l) dl \quad (4.2)$$

where p is the projection measurement in CT. This describes an ideal situation. The x-ray beam produced by an x-ray tube covers a broad spectrum. Figure 4.1 shows an example of the output x-ray spectrum of an x-ray tube operating at 120 kilovolt peak (kVp). The projection line integrals for the polychromatic spectrum and a nonuniform material is described by the following

$$p = -\ln \frac{I}{I_0} = -\ln \left(\int \Omega(E) e^{-\int_L \mu(E,l) dl} dE \right) \quad (4.3)$$

where $\Omega(E)$ represents the incident x-ray spectrum, E is the energy and $\mu(E, l)$ shows the energy dependence of the linear attenuation coefficient. The reconstruction problem for CT is to estimate or calculate the attenuation distribution from the measured line integrals of an object. The mathematical derivation of the filtered back projection (FBP) reconstruction [49] assumes Equation (4.2) to be satisfied. The line integrals are distorted due to a number of effects, e.g., beam hardening, scattered radiation and detector artifacts. The beam hardening artifact happens because of the polyenergetic nature of the x-ray beam and the energy dependency of μ . The scattered radiation is discussed in this chapter. Remaining distortions of the line-integral values, possibly after application of corrections schemes, will lead to inconsistent data and cause artifacts in the reconstructed volume.

The scattered radiation affects the accuracy of the projection measurement. Not all of the x-ray photons that reach the detector are primary photons. Sources of scatter radiation are the patient and the dose compensator (bowtie filter). In this chapter we focus on the analysis of scatter radiation from dose compensator and its effect on image quality.

We introduce using a MC simulation to characterize the intensity of scattered x-rays from the dose compensator incident on a given detector in the absence of an object and developing an effective method for the correction of scatter induced artifacts.

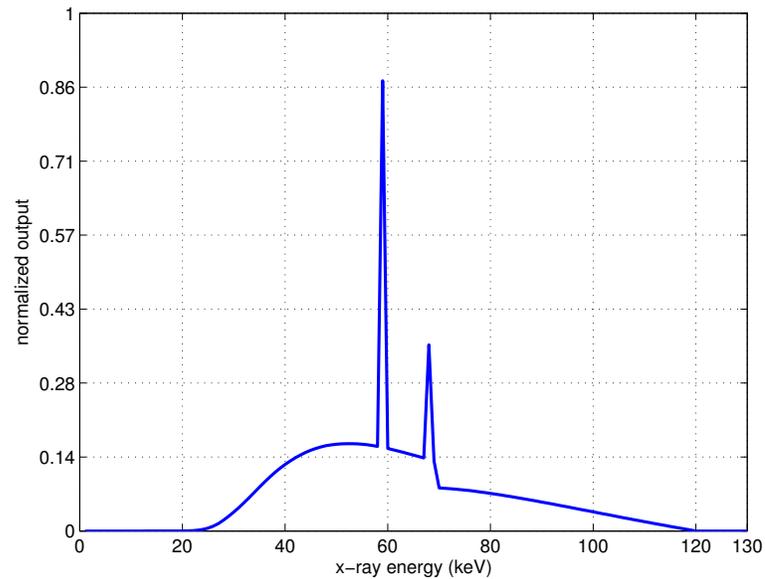


Figure 4.1. Example of an x-ray tube output energy spectrum at 120kVp.

4.2 Monte Carlo simulation

Many phenomena of physical nature are too complex to be modeled or to be calculated in an analytical way. The Monte Carlo method is a common statistical approach to overcome these limitations. The basic idea of the Monte Carlo method is to model the process of interest, e.g., the process of x-ray photon propagation through a scattering medium, by its known physical properties and to use these physical properties to calculate the tracks of individual photons through objects.

The physical properties of the different objects and their materials are specified by means of probability distributions, often referred to as physical cross sections that accurately describe the photon's interaction, transport and energy deposition characteristics within the respective material. Using this physical input data, a photon is tracked from its point of creation, e.g., the x-ray tube, through the dose compensator and to the detector by repeatedly generating random numbers and choosing associated actions defined by the probability distributions. Photons reaching the virtual detector surface are recorded and estimates of the

quantities of interest are derived from their distribution. For the simulations of the scattered radiation from the dose compensator, the geometry of the detectors, x-ray source and the dose compensator are modeled in a GEANT4 (GEometry ANd Tracking) [3] application.

4.2.1 Conventional Monte Carlo

The x-ray source is modeled as a polychromatic point source with 120 kVp and the low energy electromagnetic model of GEANT4 is used for physics processes. X-rays are tracked until they reach the surface of detector elements where the energy and angle of incidence on the face of the detector are recorded. The angle of incidence is characterized by two off-focal angles that are the angles that a given x-ray is incident on the face of the detector relative to the ray that traces directly to the focal spot. Off-focal angles of zero then indicate unscattered radiation or radiation that has undergone multiple scattering and is coincident with the unscattered direction. The intensity of scattered x-rays incident on a given detector in the absence of an object in the scan field of view is then a two dimensional function that we call the scatter kernel $K_i(\theta_x, \theta_z)$ where i denotes a particular detector element and θ_x and θ_z are the angles of incidence in a coordinate system local to detector i . A post-patient collimator (anti-scatter grid) is incorporated as an aperture function that is applied to the accumulated data. The objective of the MC simulations is to calculate the scatter kernel for each detector in the system. We are interested in the scatter radiation from the dose compensator for each detector in given off-focal directions. Since the solid angle of the detector relative to the point of scatter in the dose compensator is small, a large number of events is required to achieve reasonable accuracy. The most likely outcome of the simulation of a particular x-ray that scatters in the dose compensator is that it will be directed away from the detector. A high computational cost is required for the conventional MC simulation technique.

4.2.2 Variance reduction by Forced Detection technique

To increase the computational efficiency, a variance reduction technique called forced detection is implemented in a GEANT4 application. The different concepts of the conventional MC method and the forced detection technique are illustrated in Figure 4.2. The detectors are not shown but they are below the bowtie filters as shown in Figure 1.5. By using the conventional Monte-Carlo method, a large number of the simulated photons never reaches the detector because many photons are photo-absorbed or are scattered away from the detector area, the aim of the forced detection technique is to make use of as many contributions of the simulated photons as possible.

The basic strategy in forced detection is to force particles to interact with the detectors by creating multiple copies that are directed towards the detector elements and to adjust the weight of the copies to preserve the overall normalization. The forced detection technique is applied to the photons which scattered in the dose compensator so that they scatter into the direction of the detectors. At each Compton or Rayleigh scatter interaction, the Monte Carlo algorithm will determine a particular final state for the scattered photon. In forced detection variance reduction, a multitude of additional scattered photon is created. We call these copies forced photons. Each of the forced photons is forced to travel, without additional interactions, in the direction towards one of the detectors. Based on this final state, the scattering angle and the scattered photon energy are calculated and the weight of each forced photon is adjusted based on the detector that it is to hit and the interaction process. Figure 4.2 illustrates an incident photon emitted from the source and scattered in the dose compensator. Figure 4.2 (a) shows the conventional Monte Carlo without forced detection, the forced photons from two scatter events are shown in Figure 4.2 (b).

For Compton scattered photons the energy after the interaction is equal to [54]

$$E' = \frac{E}{1 + \frac{E}{m_0c^2}(1 - \cos \theta)} \quad (4.4)$$

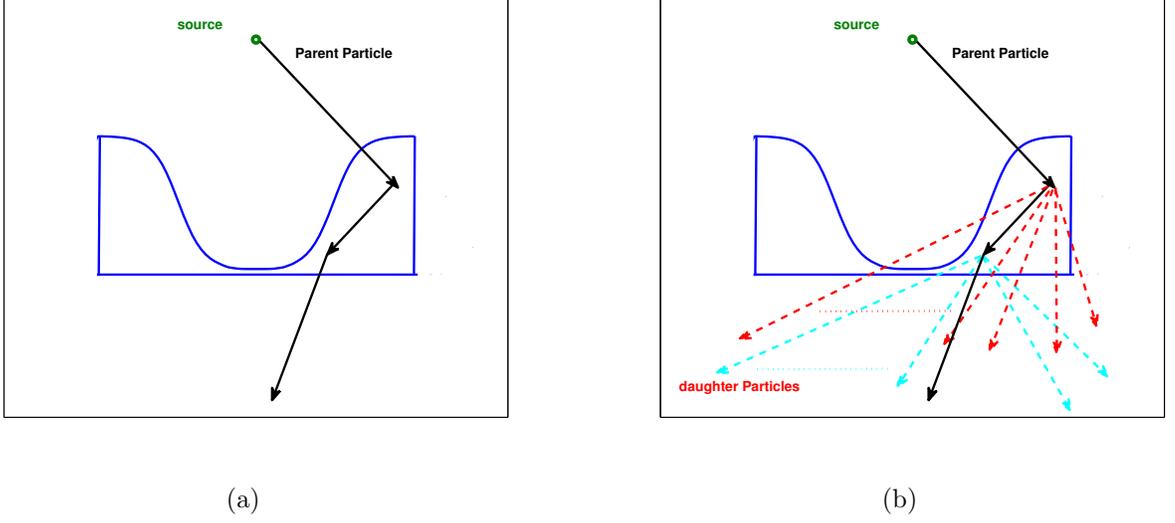


Figure 4.2. Photon tracks for conventional Monte Carlo (a) and with application of forced detection (b). The path for the photon emitted from the source is shown in solid black line. The red and cyan dotted arrows show forced photons from two scattered events which are forced to travel in the detector directions.

where E is the energy of the incident photon, m_0c^2 is the electron rest mass energy (0.511 MeV) and θ is the scatter angle. The probability of Compton scattering per atom of the absorber depends on the number of electrons available as scattering targets and therefore increases linearly with the atomic number represented as Z [54]. The angular distribution of scattered x-ray photons is predicted by Klein-Nishina formula for the differential scattering cross section [54].

$$\frac{d\sigma_{\text{KN}}}{d\Omega} = Zr_0^2 \left(\frac{1}{1 + \alpha(1 - \cos \theta)} \right)^2 \left(\frac{1 + \cos^2 \theta}{2} \right) \left(1 + \frac{\alpha^2(1 - \cos \theta)^2}{(1 + \cos^2 \theta)[1 + \alpha(1 - \cos \theta)]} \right) \quad (4.5)$$

where $\alpha \equiv E/m_0c^2$ and r_0 is the classical radius.

The differential Compton cross section is defined as the Klein-Nishina formula in Equation (4.5) multiplied by the incoherent scattering function $S(\chi, Z)$ and is angularly dependent. χ is the momentum transfer variable defined as $\chi = \frac{E}{hc} \sin(\frac{\theta}{2})$ and Z is the atomic number. If the electron is bound in an atom, Compton scattering in the forward direction

is suppressed due to electron binding effects. This is taken into account by the incoherent scattering function. The differential Compton cross section is defined as

$$\frac{d\sigma_{\text{Compton}}}{d\Omega} = \frac{d\sigma_{\text{KN}}}{d\Omega} \cdot S(\chi, Z) \quad (4.6)$$

The weight of the forced photons for the Compton process is

$$W_{\text{forced photon, Compton}} = W_{\text{incident}} \cdot e^{-\mu_{\text{total}}(E') \cdot L} \cdot \frac{1}{\sigma_{\text{Compton}}} \int \frac{d\sigma_{\text{Compton}}}{d\Omega} d\Omega \quad (4.7)$$

where W_{incident} is the weight of the incident photon at each scatter interaction, the second term, $e^{-\mu_{\text{total}}(E') \cdot L}$, is the probability of no further interaction of the forced photon in the dose compensator, μ_{total} is the total linear attenuation coefficient at energy E' . The total linear attenuation coefficient, μ_{total} , includes Compton, Rayleigh and photoelectric effects. L is the path length from the scattered point to the exit point from the dose compensator, σ_{Compton} is the total Compton cross section. The third term is the probability of scattering into the desired solid angle. Since the differential cross section is nearly constant over a small solid angle, the integral term can be approximated as $\frac{d\sigma_{\text{Compton}}}{d\Omega} \int d\Omega$. The solid angle from the scatter point to the detector, $\int d\Omega$, is calculated by another geometrical Monte Carlo simulation and a formula is derived for the solid angle based on the inverse of the square of the distance from the scatter point in the dose compensator to the detector.

During Rayleigh scattering, no energy is converted to kinetic energy or transferred to the medium. The scattered photon has the same energy as the incident photon ($E' = E$). Angular dependence for the process of Rayleigh scattering is based on the angular scattering probability of a free electron given by differential Thomson cross section [51]

$$\frac{d\sigma_0}{d\Omega} = \frac{r_0^2}{2} (1 + \cos^2 \theta) \quad (4.8)$$

where θ is the scattering angle. While interacting with a bound electron, the Thomson cross section must be weighted with an atomic form factor $F(\chi, Z)$ [51]. The differential Rayleigh cross section is

$$\frac{d\sigma_{\text{Rayleigh}}}{d\Omega} = \frac{r_0^2}{2} (1 + \cos^2 \theta) \cdot F^2(\chi, Z) \quad (4.9)$$

For the forced photons from Rayleigh scattering, the energy of the forced photons remained unchanged but the weight of each forced photon is adjusted according to

$$W_{\text{forced photon, Rayleigh}} = W_{\text{incident}} \cdot e^{-\mu_{\text{total}}(E) \cdot L} \cdot \frac{1}{\sigma_{\text{Rayleigh}}} \int \frac{d\sigma_{\text{Rayleigh}}}{d\Omega} d\Omega \quad (4.10)$$

where σ_{Rayleigh} is the total Rayleigh cross section and $\frac{d\sigma_{\text{Rayleigh}}}{d\Omega}$ is the differential Rayleigh cross section, both of them including the squared atomic form factor. The rest of the parameters in Equation (4.10) are similar to the Equation (4.7) and the integral is simplified the same way as Equation (4.7) by taking the differential cross section out of the integral and deriving the solid angle from the pre-calculated formula. The weight of the original, non-forced, final state photon after scattering is equal to its weight before interaction minus the sum of all the forced photon weights.

$$W_{\text{final state}} = W_{\text{incident}} - \sum W_{\text{forced photon}} \quad (4.11)$$

This final state photon then becomes the incident photon for the next scatter event.

The FD-MC was validated against the conventional MC (no forced detection). The validation was on the primary photons which reached the detector without interaction; both MC simulations must give similar results and are compared to an analytic calculation. The primary ratios for two sets of simulations with 100,000 events were calculated. The first set was the simulation without a dose compensator, and the second set included the dose compensator. We then calculated the ratio of the primary rays with the dose compensator to the primary without including the dose compensator for the FD-MC and the conventional MC, and compared it with the analytical calculation. The ratios are well aligned for all three models and are shown in Figure 4.3. The results are consistent with each other within the statistical accuracy of the simulations.

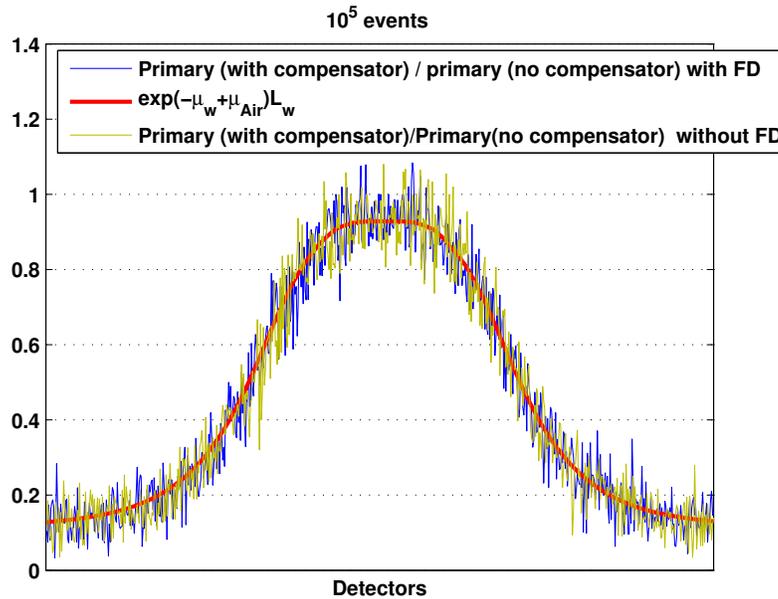


Figure 4.3. The ratio of primary with the dose compensator to the primary without the dose compensator for the FD, conventional MC, and analytic model are plotted for all detectors.

4.3 Dose compensator scatter kernel

For each detector a two dimensional dose compensator scatter kernel was obtained via FD-MC simulations. This kernel shows the likelihood of the scatter reaching the detector in each of the off-focal angles in two dimensions. Figure 4.4 shows the dose compensator scatter kernel for one of the detectors. This simulation was done with the FD technique and included 429,600 Compton and 52,400 Rayleigh scatters. The z-axis shows the likelihood of the scatter in each off-focal angle, and the x-axis and y-axis represent the off-focal angles in degrees. These calculated dose compensator scatter kernels can be used for correcting the dose compensator scatter radiation. The effects of the pre-patient collimator and the post-patient collimator on the dose compensator scatter kernels were accounted for by applying geometrically calculated aperture functions. For implementing the dose compensator scatter correction, we needed an estimate of the line integral of the rays from the dose compensator

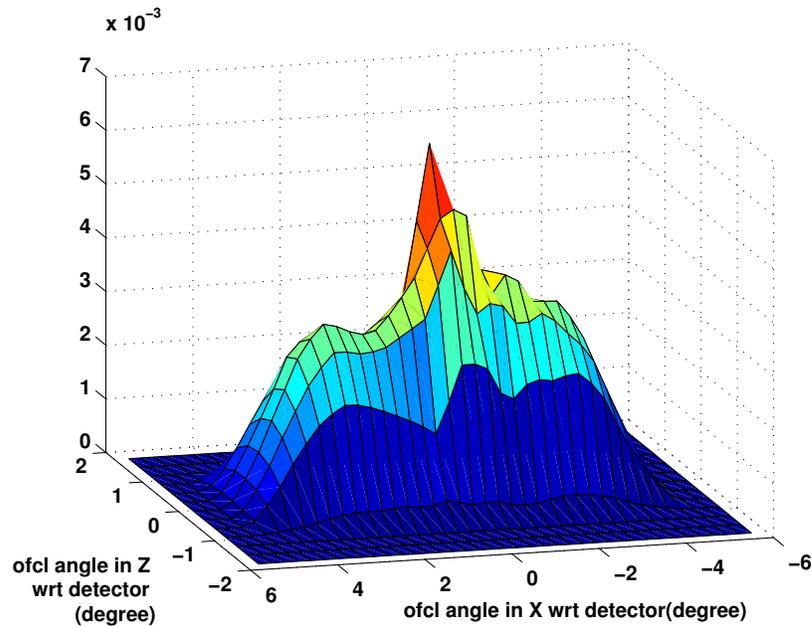


Figure 4.4. Dose compensator scatter kernel for one of the detectors with respect to off-focal angles in X and Z direction.

scatter location to the detector that we want to correct for the dose compensator scatter. We used readings from different detectors belonging to neighboring source fans.

4.4 Results

Since the dose compensator scattered radiation has a dependence in the z-direction, the scatter to primary ratio (SPR) varies with collimation due to the fact that the pre-patient collimator will block some of the radiation scattered in the dose compensator. Figure 4.5 shows the SPR for various collimations across all detectors including the effects of pre-patient and post-patient collimator indicates that the SPR reduces with narrower collimation as expected. The SPR for center detectors ranges from 0.5% to 1%, while for edge detectors it ranges from 3% to 11%, indicating the necessity to correct for the scattered radiation from the dose compensator. The results for SPR are based on the conventional MC simulation results, since calculating the SPR is not as straightforward with the FD.

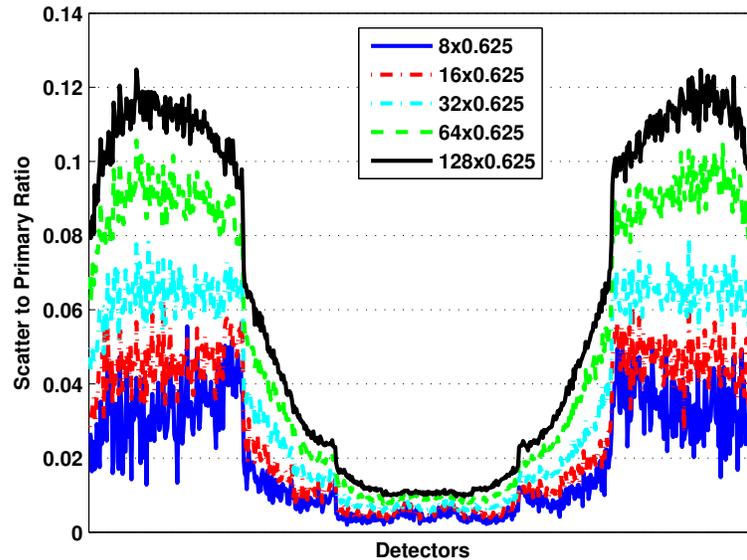


Figure 4.5. Effect of collimation on the scatter to primary ratio across the detectors for 8×0.625 , 16×0.625 , 32×0.625 , 64×0.625 and 128×0.625 mm collimations.

The estimated scatter kernels from FD simulation were used to evaluate the impact of dose compensator scatter on image quality using simulated test phantoms. The CT simulator provided us the ideal, noise free simulated attenuation line integrals. We simulated the effect of scatter radiation from the dose compensator by adding the scattered radiation to idealized, noise free data. The dose compensator scattered radiation was attenuated through the phantom before being added to the detector signal and the total scatter reaching each detector was the sum of all the scatter in the x and z off focal angles.

The correction algorithm used the simulated scatter kernel. The rays coming from z direction were not available from the scanner data. The best available approximation to those rays was the line integral from that slice and neighboring slices. Figure 4.6 shows one slice of the 20 cm diameter water phantom for the ideal case and added dose compensator scatter with magnitude of 1% on central detector to the ideal data set. The corrected image for the scatter from dose compensator is shown in Figure 4.6 (c). The horizontal profile of these three images is shown in Figure 4.6 (d), zoomed to show the effect at the edge of the

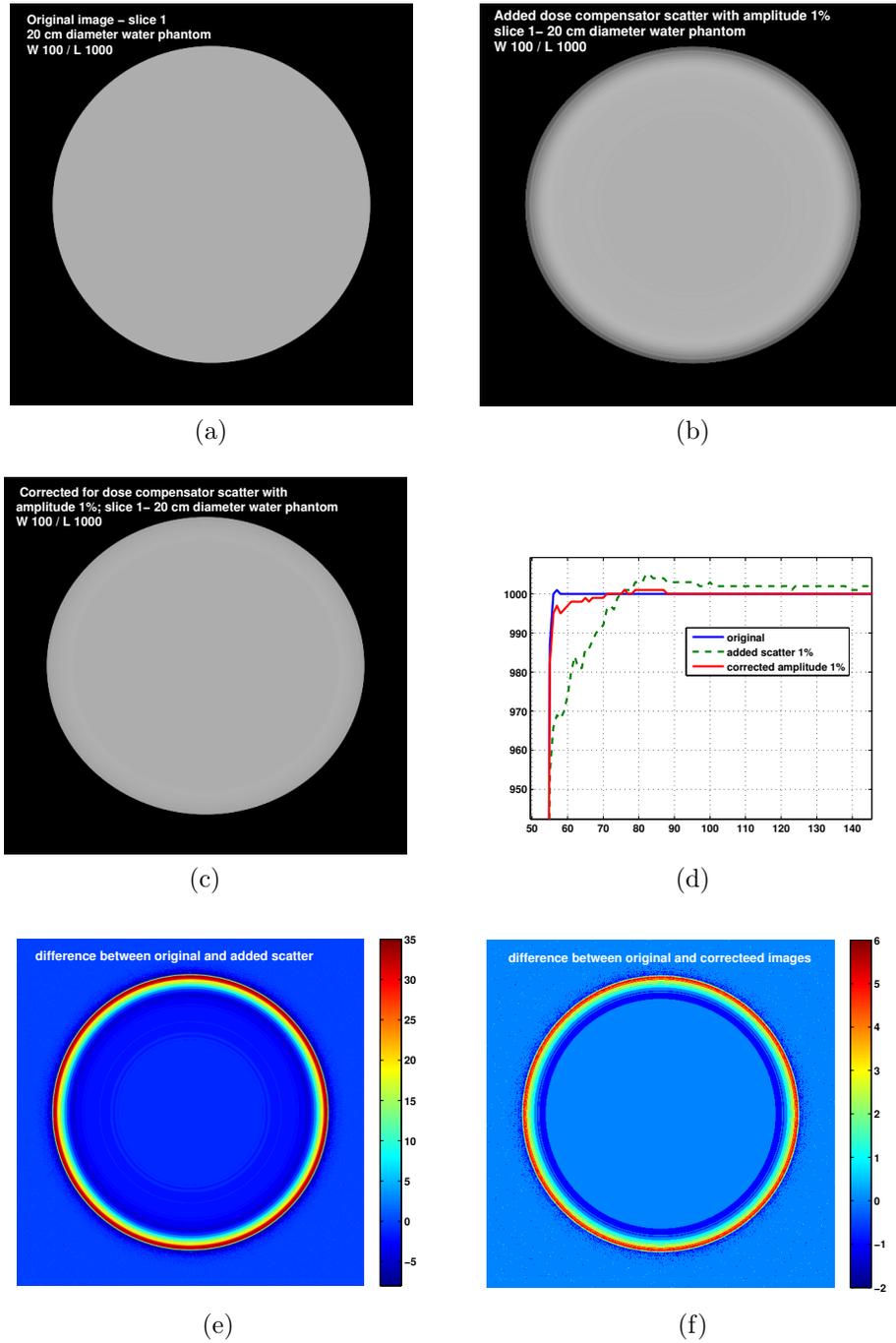


Figure 4.6. Slice 1 of the (a) simulated data, (b) added dose compensator scatter radiation, (c) corrected for compensator scatter. (d) The horizontal profiles. The difference between original and added scatter is shown in (e), the difference between original and corrected image is in (f).

phantom. The blurring artifact at the edge of the phantom is caused by the scatter from the dose compensator. The results show significant reduction of the compensator scatter artifact as seen by the corrected image in Figure 4.6 (c) and profiles in Figure 4.6 (d).

A 20 cm diameter solid water phantom located 14 cm above isocenter was also simulated. Horizontal profiles for slice 1 are plotted in Figure 4.7 (a) for the original image, added dose compensator scatter with amplitude 1% image and corrected image. The difference between original and added dose compensator images and difference image between original and corrected images are shown in Figure 4.7 (b) and Figure 4.7 (c), respectively. The effect of the dose compensator scatter radiation depends on the location of the phantom with respect to isocenter. For a centered phantom the effect is symmetric and has lower amplitude while for the uncentered phantom the effect is asymmetric and more farther from the center of the phantom. The reason for the larger artifact is that in the uncentered phantom is in a higher SPR region.

4.5 Conclusion

The scatter radiation from the dose compensator was modeled using Monte Carlo simulation and found to be comparable to the off focal radiation from the x-ray tube. The forced detection technique was implemented in a GEANT4 application and reduced the computation time by a factor of 42,800. The impact of the scatter radiation on image quality was studied on simulated data. A scatter correction algorithm was implemented which used dose compensator scatter kernels in the correction process.

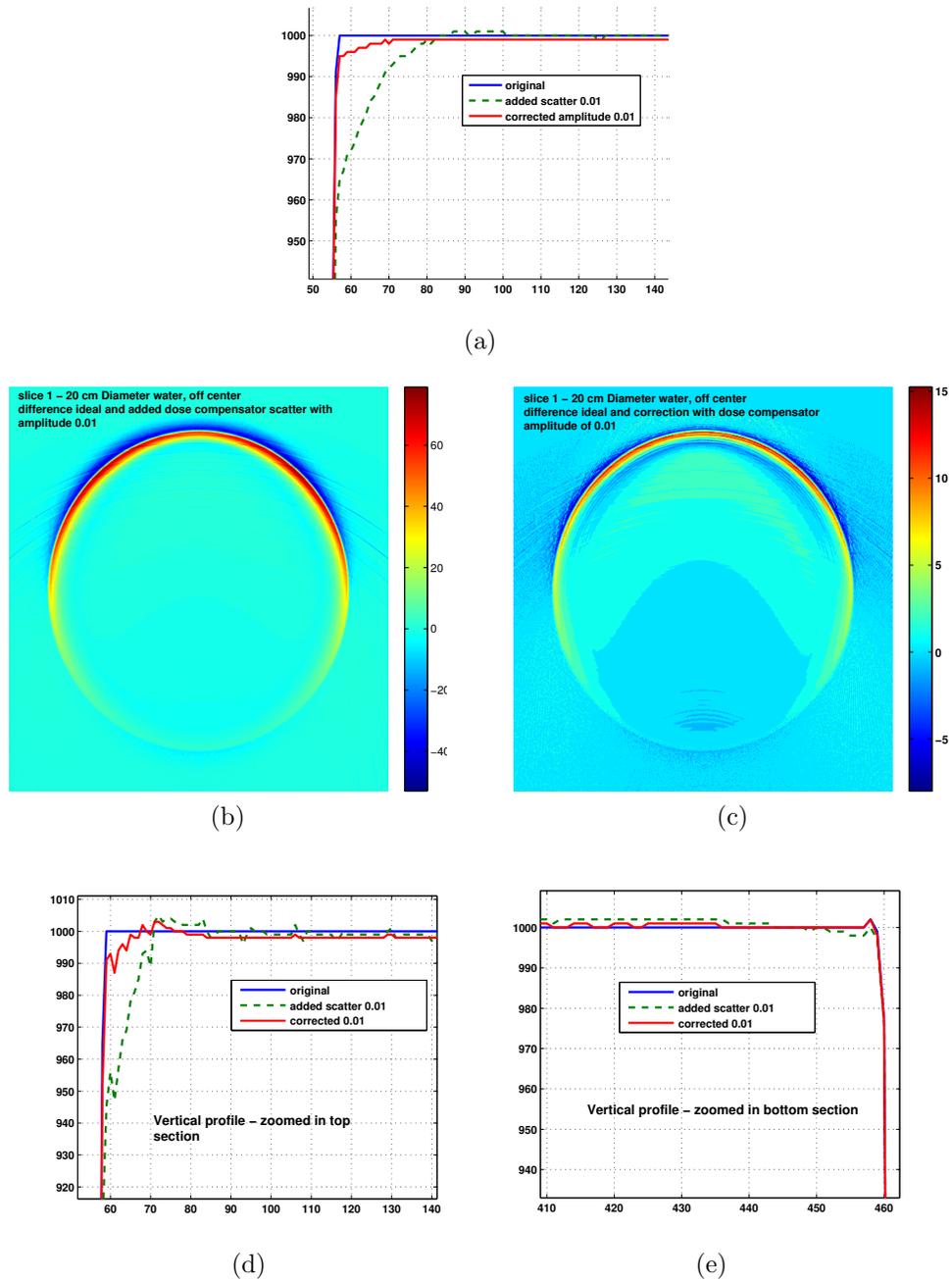


Figure 4.7. Simulated uncentered 20 cm diameter water phantom. (a) Horizontal profiles for original, added scatter with amplitude 1%, corrected for dose compensator, (b) difference between the original and added dose compensator scatter radiation, (c) difference between the original and corrected images and (d) and (e) Vertical profiles for original, added scatter with amplitude 1%, corrected for dose compensator zoomed in top and bottom section of image.

CHAPTER 5

CONCLUSION

5.1 Contributions

The contributions of this dissertation were developing algorithms for modeling signal and artifact components in fMRI and Computed Tomography. Accurate modeling of signal components affect image analysis.

We introduced applying the MDL principle in the fMRI analysis, for removing the drift and estimating the hemodynamic response function in Chapter 2. We developed an algorithm for simultaneously estimating drift and activation level. The advantage of the proposed method was that no model or structure were assumed for the drift component and we let the data select the best model for the drift term. We evaluated the performance of the proposed algorithm against the polynomial and spline detrending methods on simulated and real fMRI datasets. The results showed better estimation and removal of the drift via the MDL-based detrending.

The MDL principle was also used to design adaptive methods for HRF and activation level estimation for a region of interest. The proposed iterative algorithm described in Chapter 2 provided a nonparametric approach to HRF estimation without any explicit assumptions. The estimated HRF was compared with the benchmark, the Gamma and Poisson HRF models. The results showed the ability of the MDL-based algorithm to estimate the benchmark HRF with good approximation. Statistical analysis based on the estimated HRF by the proposed method showed better estimates of the activation level and t-tests compared with that of fixed HRF models. The performance of the algorithm was compared with AFNI on a real fMRI dataset and showed smoother HRF estimates and comparable activation maps.

A maximum likelihood method was implemented in Chapter 3 for jointly estimating the activation levels and the hemodynamic response function for a region of interest in fMRI. A rank-one approximation was used for finding the HRF in a computationally tractable manner. A smoothness constraint was imposed on the HRF estimate to improve the results under limited observation data. Both block design and event related paradigms can be used with this method. We investigated the effect of colored noise. Our results showed that accurate modeling of noise autocorrelation did not significantly affect the HRF estimation, which was consistent with the results reported earlier in the literature. The algorithms were tested using simulated and real fMRI datasets.

For the computed tomography modality, as described in Chapter 4, we modeled the scatter radiation from the dose compensator in a Monte Carlo simulation and found it to be comparable to the off focal radiation from the x-ray tube. The forced detection technique was implemented in a GEANT4 application and reduced the computation time by a factor of more than 40,000. We obtained two dimensional dose compensator scatter kernels for each of the detectors in a multislice detector CT. The dose compensator scatter to primary ratio was calculated based on the MC simulations. The impact of the scatter radiation on image quality was studied on simulated data and showed that the effect was significant and varied with the size of the phantom and its location with respect to isocenter. A dose compensator scatter correction algorithm was implemented which used dose compensator scatter kernels of each detector in the correction process.

REFERENCES

- [1] Abdi, H. (2007). *Bonferroni and Sidak corrections for multiple comparisons*, pp. 103–107. N.J. Salkind (Ed.): Encyclopedia of Measurement and Statistics, Thousand Oaks(CA): Sage.
- [2] Afonso, D., J. Sanches, and M. Lauterbach (2008, October). Robust brain activation detection in functional MRI. In *15th IEEE International Conference on Image Processing (ICIP)*, San Diego, CA, pp. 2960–2963.
- [3] Agostinelli, S., J. Allison, and et al (2003). Geant4-a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506(3), 250 – 303.
- [4] Aguirre, G., E. Zarahn, and M. D’Esposito (1998). The variability of human BOLD hemodynamic response. *Neuroimage* 8(4), 360–369.
- [5] Ardekani, B., J. Kershaw, K. Kashikura, and I. Kanno (1999). Activation detection in functional MRI using subspace modeling and maximum likelihood estimation. *IEEE Transactions Medical Imaging* 18(2), 101–114.
- [6] Bandettini, P., A. Jesmanowics, E. Wong, and J. Hyde (1993). Processing strategies for time-course data sets in functional MRI of human brain. *Mag. Res. Med.* 30, 161–173.
- [7] Bazargani, N., R. A. Thompson, and Y. Yoad (2012, February). Modeling scattered radiation from dose compensator in CT by forced detection Monte Carlo simulation. In *Proc. SPIE 8313*, San Diego. 83132L.
- [8] Beck, J. W., R. J. Jaszczak, R. E. Coleman, C. F. Starmer, and L. W. Nolte (1982). Analysis of SPECT including Scatter and Attenuation Using Sophisticated Monte Carlo Modeling Methods. *IEEE Transactions on Nuclear Science* 29, 506–511.
- [9] Beckmann, C. and S. Smith (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging* 23(2), 137–152.
- [10] Boone, J. M. (2010). Method for evaluating bow tie filter angle-dependent attenuation in CT: Theory and simulation results. *Medical Physics* 37(1), 40–48.
- [11] Bootsma, G. J., F. Verhaegen, and D. A. Jaffray (2011). The effects of compensator and imaging geometry on the distribution of x-ray scatter in cbct. *Medical Physics* 38(2), 897–914.

- [12] Boynton, G., S. Engel, G. Glover, and D. Heeger (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience* 16(3), 4207–4221.
- [13] Büchel, C. and K. J. Friston (1997). Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* 7(8), 768–778.
- [14] Bullmore, E., M. Brammer, S. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham (1996). Statistical methods of estimation and inference for functional MR image analysis. *Mag. Reson. Med.* 35, 261–277.
- [15] Buxton, R., K. Uluda, D. J. Dubowitz, and T. T. Liu (2004). Modeling the hemodynamic response to brain activation. *NeuroImage* 23, S220–S233.
- [16] Calhoun, V., T. Adali, G. Pearlson, and J. Pekar (2001). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Human Brain Mapping* 13(1), 43–53.
- [17] Casanova, R., S. Ryalib, J. Serences, L. Yanga, R. Kraft, P. Laurientia, and J. Maldjian (2008). The impact of temporal regularization on estimates of the bold hemodynamic response function: A comparative analysis. *Neuroimage* 40, 1606–1618.
- [18] Chaari, L., T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu (2013). Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach. *Medical Imaging, IEEE Transactions on* 32(5), 821–837.
- [19] Chen, S., C. Bouman, and M. Lowe (2004). Clustered components analysis for functional MRI. *IEEE Transactions on Medical Imaging* 23(1), 85–98.
- [20] Ciuciu, P., J.-B. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali (2003). Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Transactions on Medical Imaging* 22(10), 1235–1251.
- [21] Cohen, M. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage* 6(2), 93–103.
- [22] Cover, T. and J. Thomas (1991). *Elements of Information Theory*. Wiley and Sons, New York.
- [23] Cox, R. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research* 29, 162–173.
- [24] Cox, R. (1997). Improving the task-to-signal correlations in fMRI by detrending slow components. In *ISMRM*, pp. 1668.

- [25] Cox, R., A. Jesmanowics, and J. Hyde (1995). Real time functional magnetic resonance imaging. *Magn.Reson.Med.* 33, 230–236.
- [26] Dale, A. M. and R. L. Buckner (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping* 5, 329–340.
- [27] De Beenhouwer, J., S. Staelens, S. Vandenberghe, and I. Lemahieu (2008). Acceleration of GATE SPECT simulations. *Medical Physics* 35(4), 1476–1485.
- [28] Donnet, S., M. Lavielle, and J.-B. Poline (2006). Are fMRI event-related response constant in time? a model selection answer. *NeuroImage* 31(3), 1169 – 1176.
- [29] Donoho, D. and I. Johnstone (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81, 425–455.
- [30] Donoho, D. and I. Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of Amer. Statist. Assoc.* 90(432), 1200–1224.
- [31] Eckart, C. and G. Young (1936). The approximation of one matrix by another of lower rank. *Psychometrika* 1(3), 211–218.
- [32] Fadili, J. and E. Bullmore (2005). Penalized partially linear models using sparse representations with an application to fMRI time series. *IEEE Transactions Signal Processing* 53(9), 3436–3448.
- [33] Figueiredo, M. and J. Leitaõ (1997). Unsupervised image restoration and edge location using compound gauss-markov random fields and the MDL principle. *IEEE Transactions on Image Processing* 6, 1089–1102.
- [34] Frackowiak, R., J. T. Ashburner, W. Penny, S. Zeki, K. Friston, C. Frith, R. Dolan, and C. Price (2004). *Human Brain Function*. Elsevier Science. Second edition.
- [35] Friman, O., M. Borga, P. Lundberg, and H. Knutsson (2003, July). Adaptive analysis of fMRI data. *NeuroImage* 19(3), 837–845.
- [36] Friston, K., P. Jezzard, and R. Turner (1994). Analysis of functional MRI time-series. *Human Brain Mapping* 1(2), 153–171.
- [37] Friston, K., O. Josephs, G. Rees, and R. Turner (1998). Nonlinear event-related responses in fMRI. *Magnetic Resonance in Medicine* 39(1), 41–52.
- [38] Friston, K., O. Josephs, E. Zarahn, A. Holmes, S. Rouquette, and J. Poline (2000). To smooth or not to smooth? bias and efficiency in fMRI time-series analysis. *NeuroImage* 12(2), 196–208.

- [39] Friston, K., A. Mechelli, R. Turner, and C. Price (2000). Nonlinear responses in fMRI: The balloon model, volterra kernels, and other hemodynamics. *NeuroImage* 12(4), 466–477.
- [40] Genovese, C. R. (1997). A time-course model for fMRI data. In *Proc., ISMRM, 5th Annual Meeting*, pp. 1669.
- [41] Glover, G. (1999). Deconvolution of impulse response in event-related bold fMRI. *Neuroimage* 9, 416–429.
- [42] Golub, G., M. Heath, and G. Wahba (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21(2), 215–223.
- [43] Gössl, C., D. Auer, and L. Fahrmeir (2001). Bayesian spatiotemporal inference in Functional Magnetic Resonance Imaging. *Biometrics* 57, 554–562.
- [44] Gössl, C., L. Fahrmeir, and D. Auer (2001). Bayesian modeling of the hemodynamic response function in bold fMRI. *NeuroImage* 14(1), 140–148.
- [45] Goutte, C., F. Nielsen, and L. Hansen (2000). Modeling the haemodynamic response in fMRI using smooth FIR filters. *IEEE Transactions on Medical Imaging* 19(12), 1188–1201.
- [46] Handwerker, D., J. Ollinger, and M. D’Esposito (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21(4), 1639–1651.
- [47] Henson, R., C. Price, M. Rugg, R. Turner, and K. Friston (2002). Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *Neuroimage* 15(1), 83–97.
- [48] Holmes, A. P., O. Josephs, C. Bchel, and F. K. J. (1997). Statistical modeling of low-frequency confounds in fMRI. *Neuroimage* 5, S480.
- [49] Hsieh, J. and S. (Society) (2009). *Computed tomography: principles, design, artifacts, and recent advances*. Press Monograph. SPIE.
- [50] Hu, J., J. Tian, and Yang (2006). Functional feature embedded space mapping of fMRI data. In *28th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, EMBS ’06*, pp. 1014–1017.
- [51] Hubbell, J. H., W. J. Veigele, E. A. Briggs, R. T. Brown, D. T. Cromer, and R. J. Howerton (1975). Atomic form factors, incoherent scattering functions, and photon scattering cross sections. *Journal of Physical and Chemical Reference Data* 4(3), 471–538.

- [52] Jezzard, P., P. Matheews, and S. Smith (2002). *Functional MRI, an Introduction to Methods*. Oxford University Press.
- [53] Josephs, O., R. Turner, and K. Friston (1997). Event-related fMRI. *Human Brain Mapping* 5(4), 243–248.
- [54] Knoll, G. (2000). *Radiation detection and measurement*. Wiley. 3rd Edition, 50-51.
- [55] Krim, H. and I. Schick (1999). Minimax description length for signal denoising and optimized representation. *IEEE Transactions on Information Theory* 45, 898–908.
- [56] Kruggel, F., D. Von Cramon, and X. Descombes (1999, November). Comparison of filtering methods for fMRI datasets. *NeuroImage* 10(5), 530–543.
- [57] Lange, N. and S. Zeger (1997). Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Journal of the Royal Statistical Society. Series C: Applied Statistics* 46(1), 1–29.
- [58] Liao, C., K. Worsley, J. Poline, J. Aston, G. Duncan, and A. Evans (2002). Estimating the delay of the fMRI response. *Neuroimage* 16(3 Pt 1), 593–606.
- [59] Lindquist, M. and T. Wager (2007). Validity and power in Hemodynamic Response Modeling: A comparison study and a new approach. *Human Brain Mapping* 28(8), 764–784.
- [60] Liu, X., A. Srivastava, H.-K. Lee, and J. Hsieh (2013). Monte Carlo simulation of bowtie filter scatter on a wide-cone low-dose CT system. In *Proc. SPIE*, Volume 8668, pp. 86685W–86685W–7.
- [61] Ljungberg, M. and S. E. Strand (1991). Attenuation and scatter correction in SPECT for sources in a nonhomogeneous object: A Monte Carlo study. *Journal of Nuclear Medicine* 32(6), 1278–1284.
- [62] Lowe, M. and D. Russell (1999). Treatment of baseline drifts in fMRI time series analysis. *Journal of Computer Assisted Tomography* 23(3), 463–473.
- [63] Mainegra-Hing, E. and I. Kawrakow (2010). Variance reduction techniques for fast Monte Carlo CBCT scatter correction calculations. *Physics in Medicine and Biology* 55(16), 4495.
- [64] Makni, S., P. Ciuciu, J. Idier, and J. Poline (2005). Joint detection-estimation of brain activity in Functional MRI: A multichannel deconvolution solution. *IEEE Transactions Signal Processing* 53(9), 3448–3502.

- [65] Makni, S., J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu (2008). A fully bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage* 41, 941–969.
- [66] Marchini, J. and B. Ripley (2000). A new statistical approach to detecting significant activation in functional MRI. *NeuroImage* 12(4), 366–380.
- [67] Marrelec, G., H. Benali, P. Ciuciu, M. Pelegrini Issac, and J. Poline (2003). Robust estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. *Human brain mapping* 19, 1–17.
- [68] Marrelec, G., P. Ciuciu, M. Pelegrini-Issac, and H. Benali (2003). Estimation of the hemodynamic response function in event-related functional MRI: directed acyclic graphs for a general bayesian inference framework. *Inf Process Med Imaging* 18, 635–646.
- [69] McKeown, M., S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping* 6(3), 160–188.
- [70] Menser, B., J. Wiegert, S. Wiesner, and M. Bertram (2010). Use of beam shapers for cone-beam CT with off-centered flat detector. In *Proc. SPIE*, Volume 7622, pp. 762233–762233–12.
- [71] Meyer, F. G. (2003). Wavelet-based estimation of a semiparametric generalized linear model of fMRI time-series. *IEEE Transactions Medical Imaging* 22(3), 315–322.
- [72] Ndili, U., R. Nowak, and M. Figueiredo (2001). Coding theoretic approach to image segmentation. In *IEEE International Conference on Image Processing*, Volume 3, pp. 78–81.
- [73] Ogawa, S., T. Lee, A. S. Nayak, and P. Glynn (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine* 14(1), 68–78.
- [74] Ollinger, J., G. Shulman, and M. Corbetta (2001). Separating processes within a trial in event-related functional MRI. *Neuroimage* 13(1), 210–217.
- [75] Pedregosa, F., M. Eickenberg, P. Ciuciu, A. Gramfort, and B. Thirion (2014). Data-driven HRF estimation for encoding and decoding models. *CoRR abs/1402.7015*.
- [76] Rajapakse, J., F. Kruggel, J. Maisog, and D. Von Cramon (1998). Modeling hemodynamic response for analysis of functional MRI time-series. *Human Brain Mapping* 6(4), 283–300.

- [77] Riera, J., J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, T. Ozaki, and et al. (2004). A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *Neuroimage* 21(2), 547–567.
- [78] Rissanen, J. (1978). Modeling by shortest data description. *Automatica* 14(5), 465–471.
- [79] Rissanen, J. (2000). MDL denoising. *IEEE Transactions Information Theory* 46(7), 2537–2543.
- [80] Roos, T., P. Myllymaki, and J. Rissanen (2009, Sept). MDL denoising revisited. *Signal Processing, IEEE Transactions on* 57(9), 3347–3360.
- [81] Saito, N. (1994). Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion. In *WAVELETS IN GEOPHYSICS*, pp. 299–324. Academic Press.
- [82] Shen, H. and J. Huang (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis* 99, 1015–1034.
- [83] Shtarkov, Y. (1987). Universal sequential coding of single messages. *Problems of Information Transmission* 23, 175–86.
- [84] Skudlarski, P., R. Constable, and J. Gore (1999). ROC analysis of statistical methods used in functional MRI: individual subjects. *Neuroimage* 9(3), 311–329.
- [85] Smith, A. M., B. K. Lewis, U. E. Ruttimann, F. Q. Ye, T. M. Sinnwell, Y. Yang, J. H. Duyn, and J. A. Frank (1999, May). Investigation of low frequency drift in fMRI signal. *NeuroImage* 9(5), 526–533.
- [86] Strother, S. (2006, March). Evaluating fMRI preprocessing pipeline. *IEEE Engineering in Medicine and Biology Magazine* 25(2), 27–41.
- [87] Tanabe, J., D. Miller, J. Tregellas, R. Freedman, and F. G. Meyer (2002, April). Comparison of detrending methods for optimal fMRI preprocessing. *Neuroimage* 15(4), 902–907.
- [88] Tikhonov, A. and V. Arsenin (1977). *Solution of ill-posed problems*. W.H. Winston & Sons, Washington DC.
- [89] Vakorin, V., R. Borowsky, and G. Sarty (2007). Characterizing the functional MRI response using Tikhonov regularization. *Statistics in Medicine* 26(21), 3830–3844.
- [90] Woolrich, M., T. Behrens, and S. Smith (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. *Neuroimage* 21(4), 1748–1761.

- [91] Woolrich, M., M. Jenkinson, J. Brady, and S. Smith (2004). Fully bayesian spatio-temporal modeling of fMRI data. *IEEE Transactions on Medical Imaging* 23(2), 213–231.
- [92] Woolrich, M., B. Ripley, M. Brady, and S. Smith (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage* 14, 1370–1386.
- [93] Worsley, K., C. Liao, J. Aston, V. Petre, G. Duncan, F. Morales, and A. Evans (2002). A general statistical analysis for fMRI data. *NeuroImage* 15(1), 1–15.
- [94] Zarahn, E. (2002). Using larger dimensional signal subspaces to increase sensitivity in fMRI time series analyses. *Human Brain Mapping* 17(1), 13–16.
- [95] Zhang, C., Y. Jiang, and T. Yu (2007). A comparative study of one-level and two-level semiparametric estimation of hemodynamic response function for fMRI data. *Statistics in Medicine* 26(21), 3845–3861.

VITA

Negar Bazargani received the B.S. and M.S. degrees in Biomedical Engineering at the Amirkabir University of Technology, Tehran, Iran in 2001 and 2004, respectively. She received a M.S. degree in Electrical Engineering from the Erik Jonsson School of Engineering and Computer Science, The University of Texas at Dallas (UTD) in 2008, after which she pursued her Ph.D. degree. During her studies at UTD, she was a Teaching Assistant and a Research Assistant in the Electrical Engineering department at the Multimedia Communications Lab. Her research areas mainly include medical signal processing, statistical analysis of the functional magnetic resonance images (fMRI) and Computed Tomography data modeling. She joined Philips Healthcare in September 2009 for an internship, where she is continuing her career in the CT Engineering Department as a full time research scientist since 2011.