Title
Bayesian Modeling for fMRI Brain Activation and Connectivity

Permalink
https://escholarship.org/uc/item/4000h7qg

Author
Yu, Zhe

Publication Date
2015

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, IRVINE

Bayesian Modeling for fMRI Brain Activation and Connectivity

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

Zhe Yu

Dissertation Committee:
Professor Hernando Ombao, Chair
Professor Raquel Prado
Associate Professor Babak Shahbaba

2015
DEDICATION

To

my family and friends,

for their selfless love, and generous support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>CURRICULUM VITAE</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT OF THE DISSERTATION</td>
<td>viii</td>
</tr>
</tbody>
</table>

1 Introduction 1

1.1 Scientific background 1

1.2 Current statistical approaches and their limitations 4

1.2.1 Activation modeling 4

1.2.2 Connectivity modeling 5

1.3 Proposed Bayesian models and their advantages 9

1.4 FMRI stroke study 10

1.4.1 Overview 10

1.4.2 Experimental design and fMRI data 12

2 Bayesian VAR Model for Single-Subject fMRI Data 15

2.1 Introduction 15

2.2 Methodology 18

2.2.1 Model 18

2.2.2 Prior specification 21

2.2.3 Inference 23

2.2.4 Computational and numerical issues 24

2.3 Simulation study 24

2.3.1 Simulation settings 25

2.3.2 Analysis 26

2.3.3 Results 29

2.4 Application to fMRI data 31

2.4.1 Description of the experiment and the data 31

2.4.2 Analysis 32

2.4.3 Results 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.4</td>
<td>Interpretation</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>Conclusion</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>Bayesian VAR Model for Multi-Subject fMRI Data</td>
<td>38</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>38</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Proposed approach</td>
<td>39</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Current statistical methods and their limitations</td>
<td>39</td>
</tr>
<tr>
<td>3.2</td>
<td>Methodology</td>
<td>41</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Hierarchical model: subject level</td>
<td>41</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Hierarchical model: group level</td>
<td>44</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Constrained linear basis for the HRFs</td>
<td>47</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Preliminary analysis</td>
<td>48</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Posterior inference</td>
<td>49</td>
</tr>
<tr>
<td>3.2.6</td>
<td>An alternative implementation by variational Bayes</td>
<td>53</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulation study</td>
<td>55</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Simulation settings</td>
<td>55</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Analysis</td>
<td>56</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Results</td>
<td>57</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Comparison between VB and MCMC</td>
<td>58</td>
</tr>
<tr>
<td>3.4</td>
<td>Application to fMRI data</td>
<td>60</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Analysis</td>
<td>60</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Results</td>
<td>62</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Discussion of the analysis</td>
<td>65</td>
</tr>
<tr>
<td>3.5</td>
<td>Conclusions</td>
<td>67</td>
</tr>
<tr>
<td>4</td>
<td>Bayesian State-Space Model for Single-Subject fMRI Data</td>
<td>70</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>70</td>
</tr>
<tr>
<td>4.2</td>
<td>Methodology</td>
<td>73</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Model</td>
<td>73</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Relation to the VAR model</td>
<td>78</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Prior specification</td>
<td>79</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Inference</td>
<td>80</td>
</tr>
<tr>
<td>4.3</td>
<td>Simulation study</td>
<td>83</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Simulation settings</td>
<td>83</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Analysis</td>
<td>85</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Results</td>
<td>89</td>
</tr>
<tr>
<td>4.4</td>
<td>Application to fMRI data</td>
<td>96</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Analysis</td>
<td>96</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Results</td>
<td>97</td>
</tr>
<tr>
<td>4.5</td>
<td>Extension to multi-subject modeling</td>
<td>100</td>
</tr>
<tr>
<td>4.6</td>
<td>Conclusion</td>
<td>102</td>
</tr>
<tr>
<td>5</td>
<td>Conclusion</td>
<td>104</td>
</tr>
<tr>
<td>5.1</td>
<td>Summary</td>
<td>104</td>
</tr>
<tr>
<td>5.2</td>
<td>Limitations and future work</td>
<td>105</td>
</tr>
</tbody>
</table>
A Appendix for Chapter 2
A.1 Detailed algorithm for posterior sample  
A.2 Parameter values for simulation  
A.3 Supplementary figures

B Appendix for Chapter 3
B.1 Supplementary figures  
B.2 Full conditional posterior distributions  
B.3 Sampling algorithm  
B.4 Parameter values used for simulation  
B.5 VB algorithm

C Appendix for Chapter 4
C.1 Detailed algorithm for posterior sample  
C.2 Supplementary figures
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Illustration of hemodynamic convolution. (a) Canonical HRF from SPM8. (b) Stimulus Indicator function (1 for “on” and 0 for “off”). (c) Expected BOLD response (= convolution of the HRF and stimulus indicator).</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Two types of connectivity. (a) Effective connectivity. (b) Functional connectivity.</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Locations of the five regions of interest: LM1 (left primary motor cortex), RM1 (right primary motor cortex), LPMd (left dorsal premotor cortex), RPMd (right dorsal premotor cortex), SMA (supplementary motor area).</td>
<td>12</td>
</tr>
<tr>
<td>1.4</td>
<td>Visual cues on the monitor inside the fMRI scanner, for indicating rest condition and task condition.</td>
<td>13</td>
</tr>
<tr>
<td>1.5</td>
<td>Block design of the experiment for each session.</td>
<td>13</td>
</tr>
<tr>
<td>1.6</td>
<td>Example fMRI time series at two regions from a healthy subject: LM1 (top) and RM1 (bottom); the blue wave indicates task condition (high value) and rest condition (low value). The time series of independent sessions are concatenated on the plot. The green dashed lines separate different sessions.</td>
<td>14</td>
</tr>
<tr>
<td>2.1</td>
<td>HRFs of four simulated ROIs.</td>
<td>25</td>
</tr>
<tr>
<td>2.2</td>
<td>One example of simulated data at one ROI. Blue (green) curves in the bottom indicates the starting time of a new epoch of stimulus condition (control condition), with height proportional to the response amplitude $\beta_{pk}$.</td>
<td>26</td>
</tr>
<tr>
<td>2.3</td>
<td>Examples of HRF from the first set of prior (ROI 1,2 &amp; 4).</td>
<td>27</td>
</tr>
<tr>
<td>2.4</td>
<td>Three HRFs used in the GLM analysis to assess performance on activation.</td>
<td>29</td>
</tr>
<tr>
<td>2.5</td>
<td>A typical posterior sample of HRF (normalized to unit peak height). Grey: posterior samples of HRFs; red: true HRF; blue: pointwise 95% credible bands.</td>
<td>31</td>
</tr>
<tr>
<td>2.6</td>
<td>The true and the estimated connectivity networks. (a) True network. Arrows represent effective connectivity; double lines represent functional connectivity. “+”/“-” signs indicate the property (positive or negative) of the connectivity. Signs within parentheses represent the control condition. (b) Mode of estimated network from BVAR-HRF. (c) Mode of estimated network from GLM. Dashed lines with red cross marks represent the missing connectivities in the estimated network (false negative); solid lines with red cross marks represent the connectivities in the estimated network but not in the true network (false positive).</td>
<td>31</td>
</tr>
<tr>
<td>2.7</td>
<td>Mean fit from real data analysis.</td>
<td>33</td>
</tr>
</tbody>
</table>
2.8 Estimated HRFs from prior 1. ........................................ 35
2.9 Histograms of $\phi_k(\ell)_{p,q}$. Blue dashed lines indicate position of zero. 35
2.10 Estimated connectivity networks. (a) Functional connectivity network. (b) Effective connectivity network under task condition. (c) Effective connectivity network under rest condition. Curves with ordinary arrow heads correspond to lag 1 and curves with oval arrow heads correspond to lag 2. The darker the sign is, the larger the connectivity measure is. .................. 36
3.1 A preliminary analysis of HRFs. (a) HRFs of all healthy subjects at an ROI. (b) HRFs of all stroke subjects at the same ROI. The black curve is canonical HRF. ....................................................... 46
3.2 Simulation settings. (a) Simulated mean connectivity network (numbers in parentheses represent the connectivity measure under rest condition; connectivities not shown in the graph are not present). (b) Simulated mean HRFs. 55
3.3 Examples of simulated fMRI time series at different ROIs from a single subject. Green lines separate different sessions. Blue lines indicate the task condition is on. .......................................................... 56
3.4 Posterior HRF results for ROIs 1, 2, and 5 in a given subject. Gray curves correspond to posterior samples of the HRFs, blue curves are the bounds of the 95% point-wise posterior credible bands, the red curves are the true HRFs, the black curves are the posterior median HRFs. .................. 59
3.5 Activation strength for the two groups and activation difference between the groups. Darker color indicates higher activation (or activation difference). Red indicates a positive value, and white represents zero. The black line outside the circle indicates that the corresponding 95% posterior CI does not include zero. The rectangular frame indicates the ROIs in the hemisphere corresponding to the moving hand. ................................. 62
3.6 The probabilities of the presence of connectivity for the two groups and the probability differences between the groups. A thicker arrow indicates a higher chance that connectivity is present in the group (plots (a) and (b)), or greater difference in the probability of connectivity between the two groups (plot (c)). In (a) and (b), those connectivities that are absent in all subjects in each group are not shown; a dashed arrow represents a connectivity whose presence probability does not exceed 0.5; the probabilities of the most likely and the least likely connectivities among the listed connectivities are also provided. In (c), red indicates that the connectivity is more likely to be present in the stroke group than the healthy group, and blue indicates that the connectivity is less likely to be present in the stroke group; only those differences whose 95% posterior CI does not contain 0 are included. .................. 63
3.7 Estimated HRF for each ROI and subject for the two groups. The vertical dashed line is at 5 sec and the horizontal dashed line is at the baseline value. The black curve is the median HRF of the corresponding group. .................. 64
3.8 Diagnostic plots. (a)$y_p^{(s,r)}$ (black) and $y_p^{s,(s,r)}$ (blue). (b) $\hat{c}_p^{(s,r)}(t)$. (c) PACFs of $\epsilon_p^{(s,r)}(t)$. The first row is for LM1 of a healthy subject, and the second row is for LM1 of a stroke patient. .................. 65
4.1 Illustration of the proposed state-space model for fMRI data.

4.2 Parameters of the proposed model. Parameters with a grey background are known quantities and parameters with a white background are unknown variables.

4.3 Examples of region-specific HRFs used in the simulations.

4.4 Examples of simulated fMRI time series of the three simulation scenarios. Black: simulated fMRI signals. Purple: mean of the BOLD signals (excluding the variation due to dynamic BOLD amplitudes). Red: total BOLD signal (including the variation due to dynamic BOLD amplitudes). Green vertical lines indicate the start/end of a fMRI scan session. Green horizontal lines indicate the stimulus condition is on.

4.5 Average ROC for the presence of connectivity, under different settings. The horizontal and vertical axes represent the sensitivity and specificity, respectively. The solid square represents the location of threshold $\lambda = 0.5$.

4.6 Examples of estimated HRFs from a simulated data. Red curve: true HRF. Black curve: estimated HRF. Blue curves: 95% credible band. Red vertical lines: true $T_{peek}$ and $T_{under}$. Blue vertical lines: estimated $T_{peek}$ and $T_{under}$. Grey curves: posterior samples of HRF. To better compare the shapes, the HRFs are rescaled to the same height.

4.7 Example results for $\beta_k^{(r)}(t)$. Red: true latent signal. Blue and black dashed: estimated signal. Blue dashed: point-wise 95% credible bands.

4.8 Estimated connectivity network between the five ROIs. Red: connectivity is positive. Blue: negative connectivity.

4.9 Estimated strength in the five ROIs. Black outline circle indicates the ROI is activated. The more red the background for a ROI is, the stronger the activation is.

4.10 Estimated HRFs from preliminary analysis. Blue vertical lines indicate $T_{peak}$ and $T_{under}$. The page number is not visible.
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Results comparing the proposed BVAR-HRF method with the standard GLM method. For bias terms, a negative sign indicates an underestimation.</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>The three prior distributions used in the analysis.</td>
<td>32</td>
</tr>
<tr>
<td>2.3</td>
<td>Analysis results from the three priors.</td>
<td>34</td>
</tr>
<tr>
<td>3.1</td>
<td>Comparison of results of VB and MCMC. $T_p$, $T_w$ and $T_u$ denote time to peak, FWHM, and time to post-stimulus undershoot (in seconds), respectively.</td>
<td>59</td>
</tr>
<tr>
<td>4.1</td>
<td>Six settings for the simulation study. “Scen.”: Scenario. “FixHRF”: the simplified model with HRFs fixed.</td>
<td>87</td>
</tr>
<tr>
<td>4.2</td>
<td>Model performance under different settings.</td>
<td>89</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Hernando Ombao, who has always been incredibly supportive of me and helpful to me and who sparked my interest in statistical models for neuroimaging data. Without his guidance this dissertation would not have been possible.

I would like to thank Professor Raquel Prado, who has been giving amazing advice on the Bayesian methodology and the computation, and thank Professor Babak Shahbaba, who has provided very valuable suggestions to this research.

I would also like to thank Professor Steven M. Cramer and Dr. Erin B. Quinlan, who are very generous to share their data and their insights and knowledge in neuroscience. And I would like to thank Professor Yaming Yu and Professor Daniel L. Gillen, who have taught me many important topics in statistics that have since greatly benefited my research.

This work is supported by UCI Department of Statistics, NSF grant (Division of Mathematical Sciences and Division of Social and Economic Sciences), and UCI Graduate Dean’s Dissertation Fellowship.
Zhe Yu

EDUCATION

Doctor of Philosophy in Statistics 2015
University of California, Irvine  Irvine, CA

Master of Science in Biostatistics 2009
Brown University  Providence, RI

Bachelor of Science in Information and Computing Science 2005
Fudan University  Shanghai, China

RESEARCH EXPERIENCE

Research Assistant 2011–2014
University of California, Irvine  Irvine, CA

Research Assistant 2009–2011
Brown University  Providence, RI

TEACHING EXPERIENCE

Teaching Assistant 2011–2014
University of California, Irvine  Irvine, CA

WORK EXPERIENCE

Statistician Intern Jul–Aug, 2010
Mayo Clinic  Rochester, MN

Data Scientist Intern Jul–Sep, 2014
Cylance, Inc  Irvine, CA
ABSTRACT OF THE DISSERTATION

Bayesian Modeling for fMRI Brain Activation and Connectivity

By

Zhe Yu

Doctor of Philosophy in Statistics

University of California, Irvine, 2015

Professor Hernando Ombao, Chair

Functional magnetic resonance imaging (fMRI) is a non-invasive technique that measures the associated changes in the cerebral blood flow, which reflects the neuronal activities in the brain. In this dissertation, novel Bayesian modeling approaches are developed for studying brain activation and connectivity using fMRI data. Brain activation modeling aims to identify the activated brain regions in response to the experimental condition; brain connectivity modeling aims to study the communications between brain regions. Findings from brain activation and connectivity analyses can be applied to medicine and neuroscience, as they may explain the pathological pathways of mental disorders and other neurological diseases, identify potential risk factors, explore hidden symptoms, and help improve disease diagnosis and treatment plans.

The proposed modeling approaches have the following advantages. (1) Local activation and global connectivity are simultaneously modeled in the framework of general linear model (GLM) and Granger-causality through vector autoregressive models or state-space models. (2) The effect of variation in the hemodynamic response functions (HRF) among brain regions, which could confound the estimation for connectivity, is accounted for by estimating region-specific HRFs. (3) Sparsity is imposed on connectivity to allow for convenient inference on connectivity network, without having to estimate multiple models of different
connectivity networks followed by model comparisons. (4) Connectivity is modeled to have the flexibility to vary across experimental conditions. (5) Using Bayesian approach, available scientifically relevant prior information can be incorporated. (6) The hierarchical modeling in the proposed multi-subject model allows the information to be shared across subjects to increase statistical power, and allows to compare differential activation and connectivity patterns across subject groups. Simulation studies show the proposed approaches perform well in detecting brain activation and inferring connectivity network. These models were applied to real fMRI data sets from the stroke study at UCI Neurorehabilitation Lab (PI: Cramer), and suggested potential compensatory effect that requires the involvement of the primary motor region from the unaffected brain hemisphere and the secondary motor regions to aid in executing the simple motor task using the stroke-affected hand.
Chapter 1

Introduction

1.1 Scientific background

Functional magnetic resonance image (fMRI) is able to measure, with excellent spatial resolution, the hemodynamic changes in the brain resulting from neuronal activities. The primary form of fMRI uses the blood-oxygen-level-dependent (BOLD) response [41]. When the neurons are activated, the metabolic requirement for oxygen of the local tissue increases. This leads to an increase in the cerebral blood flow and a decrease in the concentration of deoxyhemoglobin, which results in the change of the magnetic resonance signals of the brain [9]. Therefore it is an indirect measure of activity in a local neuronal population.

In this dissertation, I focus on one particular type of fMRI data: task-related fMRI data with block design. The fMRI experiment for this type of data usually consists of alternate sequence of blocks of a stimulus condition and a control condition. During the stimulus condition, some external stimulus is given to the participant to inform him/her to perform a particular task. During the control condition, the participant performs a control task or simply rests. One of the main goals for analyzing this type of fMRI data is to study the
association between the fMRI response and a cognitive or sensorimotor task, and hereby infer
the relation between neuronal activities and the brain functions related to the task. It is
commonly referred to as “brain activation analysis”. More specifically, it aims to identify the
localized brain regions that activate in response to the task and characterize the magnitude
and the pattern of the evoked activations. Through these analyses, neuroscientists hope to
map the regions that are associated with each cognitive or sensorimotor function.

The relation between the BOLD response and the sensory input is commonly characterized
by its transfer response function, the hemodynamic response function (HRF) [18]: “The
hemodynamic response function can be thought of as a temporal point spread function that
not only smooths sensory input but also applies a shift in time. In other words, these
functions describe the physiological response to a point (delta function or impulse) input,
if one were able to present such a stimulus.” A particular HRF, the canonical HRF [60], is
commonly used in fMRI studies. It is composed of a main rise and an undershoot afterwards
that persists for a considerable period (see Figure 1.1(a)).
The amplitude of BOLD response is often used to infer the magnitude of the neural activation. To identify the brain regions that respond to a stimulus, the amplitude for each region during the stimulus condition is compared to a control condition. It is assumed that when activation is evoked in a brain region by a specific stimulus, there is stronger activity during the stimulus condition than during the control condition. By identifying the brain regions that are activated under a stimulus related with a target brain function, scientists are able to locate the brain regions involved in the target brain function.

Beyond determining activation in each brain region, neuroscientists are becoming increasingly more interested in studying connectivity between distant and specialized brain regions. Brain connectivity is roughly described as the interaction and information transfer between different brain regions. Some studies have found connections between certain brain connectivity and certain diseases, such as autism, schizophrenia, stroke, Parkinson's disease, etc [3, 5, 24, 64]. In [51], fMRI connectivity is “a powerful, sensitive non-invasive tool for clinical and non-clinical investigation of brain structure, function, development and pathologies”.

There are two types of brain connectivity: functional connectivity and effective connectivity (Figure 1.2). Functional connectivity measures the undirected correlation among neurophysiological events in remote regions, which may result from various factors and cause problems in interpretation. Effective connectivity, on the other hand, measures the directed influence from one neural system on another, and therefore offers a more specific interpretation. This dissertation primarily focuses on effective connectivity.
By analyzing brain activation and brain connectivity through the observed BOLD signals, statistical models for fMRI data may aid in explaining the pathological pathways of mental disorders and other neurological diseases, in understanding the progression of those disorders and brain diseases, and in diagnosing these diseases more accurately. It may also help evaluate disease treatments and improve treatment plans.

1.2 Current statistical approaches and their limitations

1.2.1 Activation modeling

General linear model (GLM) is currently the most standard approach for modeling activation. It assumes a linear time invariant system. It models the fMRI signal as the superposition of the responses to each individual experimental condition.

\[ y_v(t) = \beta_{v0} + \beta_{v1}x_{v1}(t) + \beta_{v2}x_{v2}(t) + \epsilon_v(t), \quad \epsilon_v(t) \sim N(0, \sigma^2). \]  \hspace{1cm} (1.1)

Here, \( \{y_v(t)\} \) is the fMRI times series for \( v \), a voxel or a region of interest (ROI); \( x_{vk}(t) \) is the expected BOLD response shape for condition \( k \), with \( k = 1 \) denoting the stimulus condition and \( k = 2 \) denoting the control condition; \( \beta_{vk} \) is the corresponding amplitude of the BOLD response for condition \( k \); \( \epsilon_v(t) \) is the unexplained noise component, which may come from the neuronal level, hemodynamic level, and observation level. In this dissertation, I focus only on the analysis at the ROI level. In GLM, the expected BOLD response shape is the convolution between two components (a) HRF \( h(t) \) and (b) the binary condition indicator function \( c_k(t) \) representing the underlying neuronal activity, so that \( x_{vk}(t) = [h_v * c_k](t) \). The activation for the ROI can be inferred from the contrast \( \beta_{v,(c)} = \beta_{v1} - \beta_{v2} \): when it
is statistically significantly larger than 0, this ROI is activated in response to the stimulus. Although GLM does not model connectivity directly, it is utilized to identify the variation in the observed fMRI signal induced by the change of experimental conditions and help derive the specific quantities of interest on which connectivity analysis need to be based.

### 1.2.2 Connectivity modeling

The most prevalent approaches for modeling effective connectivity include: 1. Granger-causality based approaches, such as vector autoregressive models and state-space models; 2. structural equation modeling; and 3. dynamic causal modeling.

**Granger-causality (GC)**

Suppose \( \{x_t\} \) is the fMRI time series for ROI 1 and \( \{y_t\} \) is the fMRI time series for ROI 2. ROI 1 is said to “Granger-cause” ROI 2 if \( \{x_t\} \) helps improve predictions of \( \{y_t\} \) at future time points [25]. A mathematical definition is as follows [37]. Suppose \( \Omega_t \) is the information set containing all the relevant information in the universe available up to and including time \( t \). Let \( \hat{y}_{t+h}(\Omega_t) \) be the optimal (i.e. unbiased and least square criterion) \( h \)-step predictor of the process \( \{y_t\} \) at origin \( t \), based on the information in \( \Omega_t \). The corresponding forecast MSE is \( \sigma_y^2(h|\Omega_t) = E((y_{t+h} - \hat{y}_{t+h}(\Omega_t))^2) \). Define \( \Omega_t^{(-x)} = \Omega_t \setminus \{x_s|s \leq t\} \), which is the information set up to time \( t \) with all \( x(t) \) up to time \( t \) excluded. The process \( \{x_t\} \) is said to Granger-cause \( \{y_t\} \) if

\[
\sigma_y^2(h|\Omega_t) < \sigma_y^2(h|\Omega_t^{(-x)}).
\]

In reality, it is unlikely to obtain the “full” information set \( \Omega_t \), therefore only the information under study is considered. Moreover, the definition of Granger-causality is usually loosened to a linear sense: only linear predictors are considered and compared. The approaches in
this proposal will also be based on linear Granger-causality.

Measures based on Granger-causality through vector autoregressive models or state-space models are among the most prevalent procedures for modeling effective connectivity. For example, they were to study effective connectivity in [22, 27, 54, 56, 45, 28, 7, 14, 15, 24], among many others. One reason for this popularity is that GC does not require the direction and valence of effective connectivity to be pre-specified [53], and therefore is especially useful for exploring effective connectivity network.

**Vector autoregressive model.** Vector autoregressive models (VAR) model the multidimensional temporal dependence of the fMRI times series. The VAR coefficients, representing the temporal dependence, are used to measure effective connectivity, according to the concept of Granger-causality. Following the GLM model, the deterministic mean trend (the BOLD responses) is first estimated and removed from the observed fMRI time series. Then, VAR is applied to the model residuals.

**State-space model.** A general state-space model includes an observation equation that models the observed signal in terms of latent signals, and a state equation that models intrinsic Markov structure of the latent signals. In a state-space model for fMRI data, the observation equation is a general linear model that models brain activation, and the state equation models the multivariate temporal dependence in the underlying BOLD or neuronal signals typically using a VAR structure. Again, the VAR coefficients in the state equation are used to measure effective connectivity. The main difference between a state-space model and a VAR model is that a state-space model models effective connectivity through the dynamics of the underlying BOLD or neuronal signals, while the VAR approach models effective connectivity through the unexplained noise. Therefore, from a biological point of view, effective connectivity from state-space models is easier to interpret than that from VAR models.
There are limitations in the existing GC-based methods for fMRI data. For example, [27] did not take into account the systematic changes in the BOLD response induced by external stimuli. The method in [23] considered the BOLD response evoked by stimuli and also the differential connectivity across different experimental conditions. However, the BOLD response shape was based on a pre-fixed shape of HRF that is identical across all brain regions and subjects, which may lead to erroneous conclusions (see Chapter 2). The methods in [8] and [36] also took the experimental conditions into account, but modeled connectivity through a two-stage approach and used a pre-specified HRF. The method in [31] simultaneously analyzed activation and connectivity in the spectrum domain but also used a pre-specified HRF. The methods in [28, 7, 6] utilized state-space models to model the latent signal of the time-varying BOLD amplitude, and modeled effective connectivity through the multivariate temporal dependence of the latent signal; however, it also used a pre-specified HRF. Also, inference for effective connectivity in the above methods usually require computing models for different connectivity networks and then performing model comparisons. This could be computationally expensive if the number of ROIs are not small and when little information is available for which networks are most biologically plausible. More details about the existing methods can be found in the Chapter 2, 3, and 4.

**Dynamic causal model (DCM)**

Another way to model the effective connectivity is to use dynamic causal model [17]. DCM models interactions among brain regions directly at the neuronal level using fMRI time series from the hemodynamic level. However, DCM heavily relies on the correct specification of biological assumptions, such as how the neuronal states enter a region-specific hemodynamic model to produce the BOLD responses that are a complicated function of the region’s biophysical states reflecting deoxyhemoglobin content and venous volume. A primary limitation of DCM is that these assumptions are not easy to verify and might not be generalizable to
patients with certain diseases. Also, a set of plausible networks and the effects of stimulus inputs on the defined network need to be specified a-priori for DCM. However, the number of plausible models increases quadratically with the number of ROIs, and therefore DCM is not able to search over all plausible networks and is more appropriate for testing connectivity within certain set of networks. On the contrary, Granger-causality and VAR-based methods are exploratory techniques that do not make biological assumptions and fully extract information from data itself.

**Structural equation model (SEM)**

Structural equation model is also used to model connectivity network [38]. However, SEM assumes that random fluctuations in the fMRI signal change very slowly in relation to underlying physiology, which is not appropriate for electrophysiological and fMRI time series. In SEM, the covariance matrix of the ROIs is derived and a path model specifying the connections between brain regions is then fitted to this matrix. The strength of effective connectivity is measured by the path or structural coefficients in this method. This approach ignores the temporal correlation in the data which can lead to inaccurate standard errors and test statistics. Also, SEM is not sufficiently flexible for performing exploratory analysis because it is difficult for SEM to estimate reciprocal and cyclic connections efficiently. Similar to DCM, SEM requires specific information of the network structure.

In this dissertation, I adopt the general linear model to model activation and Granger-causality to model effective connectivity.
1.3 Proposed Bayesian models and their advantages

Chapter 2 develops a Bayesian VAR model for single-subject fMRI data, and Chapter 3 develops a hierarchical Bayesian VAR model for multi-subject fMRI data. In these VAR models, brain activation is modeled by general linear model, and effective connectivity is simultaneously modeled by VAR model on the unexplained noise term from the general linear model. Effect of variation in HRFs is accounted for by estimating the region-specific HRF shapes. Spike and slab priors are used to allow convenient inference on the presence and absence of effective connectivity, and effective connectivity is modeled to be able to vary across experimental conditions. Bayesian approach is used to allow for the incorporation of available scientifically relevant prior information. In addition, the multi-subject model in Chapter 3 uses hierarchical Bayesian modeling to allow information to be shared across subjects to increase statistical power, and to allow for the comparison of differential activation and effective connectivity patterns across subject groups.

Chapter 4 develops a Bayesian state-space model for single-subject fMRI data. In the state-space model, brain activation is modeled by general linear model in the observation equation, and effective connectivity is simultaneously modeled in the state equation by a VAR structure on the underlying dynamic BOLD amplitudes. Region-specific HRF, condition-specific effective connectivity and spike and slab priors are also included in the model as in the VAR models above.

The models developed in this dissertation have the following advantages. (1) They address the confounding effects from the variation in HRF by including region-specific HRF as unknown parameters in the model. (2) They are able to capture different effective connectivity across experimental conditions. (3) They allow one to easily pick the best effective connectivity network without having to estimate multiple models, by the use of spike and slab priors. (4) They are able to easily incorporate relevant information derived from other studies but,
unlike the dynamic causal modeling approach, they do not rely on biological assumption that are difficult to verify from the data. (5) In addition, the hierarchical VAR model for multi-subject data is able to borrow information across subjects via hierarchical modeling to increase power for group comparison; this is particularly relevant for data with low signal to noise ratio. (6) Particularly for the state-space model in Chapter 4, it allows the BOLD amplitudes to vary over time, which is more realistic and could be useful for situations such as learning effect during the experiment. (7) Furthermore, the state-space model models connectivity through the underlying dynamics of the BOLD signals, instead of through the noise term from GLM, and therefore gives more meaningful biological interpretations.

1.4 FMRI stroke study

In this dissertation, I developed the statistical methods to analyze the fMRI data set from the stroke study by UC Irvine Neurorehabilitation Lab (PI: Cramer). The broad scientific goal is to build models that help understand how brain motor function is altered after stroke, by studying the unique brain activation pattern and communications among brain motor regions for a stroke patient during the execution of a simple hand-motor task. This could potentially be helpful for neuroscientists and physicians to predict a patient’s ability to recover, determine the stage of recovery, predict the efficacy of a treatment for stroke, and design better treatment plans that target at the affected brain function.

1.4.1 Overview

Stroke is a disturbance in the blood supply to the brain that results in the death of neural tissue and subsequent behavioral deficits arising from the affected brain systems. One of the most commonly affected neural systems is the motor system, causing disability in people who
suffered a stroke. In order to compensate for the loss of brain function in the affected brain area, other areas within the motor system are often recruited (and, therefore, activated) to assist in the execution of movements in stroke-impaired subjects. Also, connectivity may be altered between regions within the motor network.

UC Irvine Neurorehabilitation Lab conducted a stroke study that focused on understanding how brain motor function is altered after stroke. It recruited 13 healthy subjects, and 29 stroke patients, including 16 with right brain hemisphere affected (left side of body affected), and 13 with left brain hemisphere affected (right side of body affected). The stroke patients had ischemic stroke 11 - 26 weeks prior to the study assessments and continued to have residual motor deficit on the right side of the body at the time of the experiment. Motor-task-related fMRI scans were acquired for all the participants. Five brain regions known to be implicated with motor function are considered for this study (see Figure 1.3): two primary motor regions including the left and right primary motor cortex (LM1 and RM1), and three secondary motor regions including the left and right dorsal premotor cortex (LPMd and RPMd), and a midline supplementary motor area (SMA). Note that a stroke patient with injury on the left brain hemisphere will have motor function deficits on the right side of the body; and vice versa. Typically, when a healthy subject performs a simple right-hand movement task, it is expected that only LM1 (M1 in the left hemisphere) is likely to be activated. RM1 is expected to not activate, since it is primarily responsible for motor function of the left side of the body. The secondary motor regions (SMA and the two PMd regions) are primarily responsible for more advanced motor function such as motor planning and bilateral coordination.
1.4.2 Experimental design and fMRI data

Design of the experiment. The fMRI experiment consisted of the two conditions: task condition (stimulus condition) and rest condition (control condition). A movie of two rotating sticks was shown on the monitor inside the fMRI scanner to inform the participants whether they were in the task condition or the rest condition (see Figure 1.4). During the task condition, the participants performed the hand grasp-release movement task following the green sticks (Figure 1.4(c)(d)). The stroke patients used their stroke-affected hand and the healthy subjects used their right hand. During the rest condition, the participants were resting both hands while watching the red sticks (Figure 1.4(a)(b)). The experiment was divided into three sessions for stroke patients and two sessions for healthy controls. Each session consisted of 48 consecutive scans, alternating between 12-scan task condition and 12-scan rest condition twice but always starting with rest condition (Figure 1.5).

FMRI data. The fMRI images were acquired using a T2*-weighted gradient-echo-planar imaging sequence with repetition time (TR) = 2000 ms. Functional data from all the sessions were preprocessed using the SPM8 software [60]. Preprocessing steps included realignment...
to the first image, co-registration to the mean image, normalization to the standard MNI EPI template, and spatial smoothing (FWHM = 8 mm). The time series for each ROI were obtained by averaging the fMRI signals recorded across the voxels in the region. The mean time series were detrended and centered to zero with time as the covariate to remove the drift effect. Finally, the detrended time series were scaled to have the same variance for all subjects and ROIs within each group.

Selected time series from LM1 and RM1 regions in a healthy subject are shown in Figure 1.6. The block-shaped wave at the bottom indicates the times when the subject performed the motor task. In LM1 (top plot) the fMRI signal generally follows the block-shaped wave with some lag. It rises soon after the task condition begins and drops soon after the task condition ends, which indicates that this region is implicated in the motor task. However, in RM1 (bottom plot) the time series does not appear to be associated with the timing of the motor task.
Figure 1.6: Example fMRI time series at two regions from a healthy subject: LM1 (top) and RM1 (bottom); the blue wave indicates task condition (high value) and rest condition (low value). The time series of independent sessions are concatenated on the plot. The green dashed lines separate different sessions.
Chapter 2

Bayesian VAR Model for Single-Subject fMRI Data

2.1 Introduction

Motivated by the increasing interest in the brain activation patterns and connectivity between brain regions in the neuroscience literature, I develop a Bayesian approach for task-related fMRI data. The proposed approach, named BVAR-HRF, utilizes Bayesian vector autoregression (BVAR) to measure connectivity, gamma functions to model the unknown hemodynamic response functions (HRF), and linear regression to estimate activation. The goal here is to study local activation pattern and global connectivity simultaneously, while accounting for the non-negligible effects of the variation in HRFs across brain regions.

Measures based on vector autoregressive models (VAR) and Granger-causality [25] are among those simple and standard procedures for fMRI effective connectivity analysis. Since Granger-causality relies on temporal precedence, proper deconvolution of the observed BOLD signals and correct specification of HRF are key to the connectivity analysis. However, most ex-
isting studies for fMRI connectivity do not take this into account. For instance, in some studies, connectivity analysis is applied directly on the BOLD signals without deconvolution; in other studies, connectivity analysis is based on deconvolution with a pre-specified HRF (e.g., canonical HRF from [60]) that is assumed to be identical for all regions. In those situations, connectivity can be greatly confounded by incorrectly specified HRF.

Another way to model the effective connectivity is to use dynamic causal modeling (DCM) [17]. DCM models interactions among brain regions directly at the neuronal level using fMRI time series from the hemodynamic level. However, DCM heavily relies on the correct specification of biological assumptions, such as how the neuronal states enter a region-specific hemodynamic model to produce the BOLD responses that are a complicated function of the region’s biophysical states reflecting deoxyhemoglobin content and venous volume. These assumptions are not easy to verify and might not be generalizable to patients with certain diseases. Also, a set of plausible networks and the effects of stimulus inputs on the defined network need to be specified a-priori for DCM. However, the number of plausible models increases quadratically with the number of ROIs, and therefore DCM is not able to search over all plausible networks and is more appropriate for testing connectivity within certain set of networks. On the contrary, Granger-causality and VAR-based methods are exploratory techniques that do not make biological assumptions and fully extract information from data itself.

This study proposes a new method that estimates both HRF and connectivity (both effective and functional) simultaneously. The method uses a Bayesian paradigm, which can easily incorporate prior information of HRF and connectivity, and potentially increase the power of estimation when partial information is available. For instance, if the peak of the HRF of a particular region is known to take place around 5s and most likely in [3s, 7s] with 95% confidence, one can assign a normal prior $N(5,1)$ truncated on $\mathbb{R}^+$ for the peak of that HRF. Moreover, inference with Markov chain Monte Carlo (MCMC) technique is simple
and straightforward. Especially for effective connectivity parameters, the proposed method makes use of spike and slab priors, allowing a joint assessment and model comparison of the entire network without estimating multiple models. Furthermore, the proposed method is able to capture the difference in effective connectivity across experimental conditions, while existing methods assumed the effective connectivity to be identical across experimental conditions. A recent paper by [12] proposed a similar VAR-based method that also estimated both HRF and connectivity. However, the way they modeled HRF was based on DCM and therefore had the drawback mentioned above.

In a nutshell, I decompose the ROI-specific fMRI time series into the expected BOLD response and a noise component. The expected BOLD response is modeled as the convolution between the ROI-specific HRF and the indicator function of the stimulus input, estimate activation by computing the difference in magnitudes of HRFs between stimulus and control condition, and measure effective and functional connectivity by fitting a VAR to the noise component. The HRF is parameterized as a difference of two scaled gamma density functions. After specifying an appropriate prior for each model parameter, the full posterior density of the whole model is determined. MCMC and Gibbs sampler are employed to obtain the joint posterior samples of each model parameters based on posterior density function. The Bayesian inference is based on MCMC posterior samples.

The remainder of the chapter is organized as follows. In Section 2.2, I explain the model components, the prior distributions for the model parameters, and the procedures for model inference. Simulation studies are reported in Section 2.3, where I compare the performance of the proposed approach with a standard approach that uses the general linear model (GLM). I also include in Section 2.4 the results of the propose model and the results of the standard GLM applied to an fMRI dataset from a stroke study. Finally in Section 2.5, I discuss the advantages and the limitations of the proposed approach.
2.2 Methodology

2.2.1 Model

Suppose there are $P$ regions of interest (ROI), and two experimental conditions: stimulus and control. Let $y_p(t)$ be the motion corrected and detrended mean fMRI signal of ROI $p$ at time $t \Delta$ seconds, where $t = 1, 2, \ldots, T$, and $\Delta > 0$ is the TR (repetition time). Define the vector $\mathbf{y}(t) = [y_1(t), \ldots, y_P(t)]' \in \mathbb{R}^P$ to be the multivariate signal for all $P$ ROIs at time $t \Delta$. For simplicity I will omit $\Delta$ for the rest of the dissertation. Let $\mathbf{h}_p = [h_p(0), h_p(1), h_p(2), \ldots, h_p(T)]'$ be the HRF for region $p$. Let $\mathbf{c}_k = [c_k(1), c_k(2), \ldots, c_k(T)]'$ be the indicator for experimental condition $k$ with $k = 1$ for stimulus condition and $k = 2$ for control condition.

$\mathbf{y}(t)$ is decomposed into the mean structure (expected fMRI signal) and the total noise component: $\mathbf{y}(t) = \mathbf{M}(t) + \mathbf{u}(t)$. The mean structure $\mathbf{M}(t) = [M_1(t), \ldots, M_P(t)]' \in \mathbb{R}^P$ is modeled as:

$$M_p(t) = \beta_{p0} + \beta_{p1}x_{p1}(t) + \beta_{p2}x_{p2}(t), \quad p = 1, 2, \ldots, P,$$  

(2.1)

where the covariate $x_{pk}(t) = [h_p * \mathbf{c}_k](t), \quad p = 1, 2, \ldots, P, \quad k = 1, 2$, is the convolution between the HRF at region $p$ and the indicator for condition $k$, representing the shape of the BOLD response from neuronal activations in ROI $p$ at time $t$. The total noise $\mathbf{u}(t)$ accounts for the additional variation of the fMRI signals unexplained by the mean structure.

HRF. $h_p(t)$ is modeled using double gamma parameterization, a linear combination of two gamma density functions $h_p(t) = g(t; a_{p1}, b_{p1}) - g(t; a_{p2}, b_{p2})c_{p2}, \quad a_{p1}, a_{p2} > 1, b_{p1}, b_{p2}, c_{p2} > 0, \quad p = 1, 2, \ldots, P$, where $g(t; a, b) = \frac{b^a}{\Gamma(a)}t^{a-1}e^{-bt}I_{(0,\infty)}(t)$. The first gamma function models
the main rise of the HRF and the second models the post-stimulus undershoot of the HRF. As a matter of fact, the most widely used canonical HRF, is a function by the double gamma parameterization at the parameter values \((a_{p1}, b_{p1}, a_{p2}, b_{p2}, c_{p2}) = (6, 1, 16, 1, \frac{1}{6})\).

**Activation.** The regression coefficient \(\beta_{pk}\) represents the amplitude of the BOLD response, which reflects the strength of the neuronal activation. When the activation contrast at ROI \(p\), \(\beta_{p,c} = \beta_{p1} - \beta_{p2}\), between stimulus and control condition is statistically significantly greater than zero, the ROI is classified as “activated” (in response to the stimulus). Note that it makes more sense to compare the contrasts after normalizing the HRFs. Normalization can be carried out using different criteria, e.g. \(L_2\)-norm, and \(L_\infty\)-norm, etc. In this approach the normalizing constant is the maximum value: \(\tilde{h}_p(t) = \frac{h_p(t)}{\|h_p\|_{\text{max}}} = \frac{h_p(t)}{\max_t h_p(t)}\), and \(\tilde{\beta}_{p,c} = \beta_{p,c} \max_t h_p(t)\). This normalization ensures the height of HRF to be 1.

Following [24], the total noise \(u(t) \in \mathbb{R}^P\) is modeled as VAR with pre-determined order \(L\):

\[
u(t) = \sum_{\ell=1}^{L} \Phi^*(\ell, t)u(t - \ell) + \epsilon(t),\]

where

\[
\Phi^*(\ell, t) = \begin{cases} 
\Phi^*_1(\ell)c_1(t - \ell) + \Phi^*_2(\ell)c_2(t - \ell) \\
\Phi^*_1(\ell), \quad \text{if } c_1(t - \ell) = 1, \text{ i.e. stimulus condition at time } t - \ell, \\
\Phi^*_2(\ell), \quad \text{if } c_2(t - \ell) = 1, \text{ i.e. control condition at time } t - \ell,
\end{cases}
\]

and \(\Phi^*_k(\ell) = [\phi^*_k(\ell)_{p,q}]_{1 \leq p,q \leq P} \in \mathbb{R}^{P \times P}\); and \(\epsilon(t) \in \mathbb{R}^P\) is the temporally uncorrelated noise with mean 0 and an unstructured covariance matrix \(\Sigma_\epsilon\).

**Effective connectivity.** Denote the \((p,q)\) element of \(\Phi^*(\ell, t)\) as \(\phi^*(\ell, t)_{p,q}\). It is the coefficient of ROI \(q\) for predicting ROI \(p\) at \(\ell\) lags. When \(\phi^*(\ell, t)_{p,q} \neq 0\), fMRI signal at ROI \(q\) “Granger-causes” fMRI signal at ROI \(p\) with lag \(\ell\). This “causality” can also be interpreted as the information transfer from ROI \(q\) to ROI \(p\). The “information transfer” is initiated at time \(t - \ell\) and completed at time \(t\), so it is reasonable to assume the pattern of the information
transfer is fully determined by the condition at its initiation time, which results in Equation (2.2). Therefore $\phi_k^*(\ell)_{p,q}$ measures the effective connectivity from ROI $q$ to ROI $p$ at lag $\ell$ under condition $k$.

Functional connectivity. The inverse matrix $\Omega_\epsilon = \Sigma_\epsilon^{-1}$ is called the (zero-lag) precision matrix of the temporally uncorrelated noise. It relates to partial correlation between ROI $p$ and ROI $q$, $C_{\epsilon,(p,q)} = -\frac{\Omega_{\epsilon,(p,q)}}{\sqrt{\Omega_{\epsilon,(p,p)}\Omega_{\epsilon,(q,q)}}}$, $p \neq q$. Partial correlation (zero lag) is the undirected contemporaneous correlation between two ROIs conditional on all other ROIs. It can be used to measure the functional connectivity. When $\Omega_{\epsilon,(p,q)}$ is significantly different from zero, I conclude that there is functional connectivity between ROI $p$ and ROI $q$. Note that the function connectivity is derived from $\epsilon(t)$ which excludes the effect from the history.

Distribution. In this approach, I assume Gaussianity of the uncorrelated noise, i.e. $\epsilon(t) \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma_\epsilon)$.

Define $\beta = [\beta_{10}, \ldots, \beta_{P0}; \beta_{11}, \beta_{21}, \ldots, \beta_{P1}; \beta_{12}, \ldots, \beta_{P2}]^T \in \mathbb{R}^{3P}$, and

$$X(t) = \begin{bmatrix} I_P, \text{ diag}\{x_{11}(t), \ldots, x_{P1}(t)\}, \text{ diag}\{x_{12}(t), \ldots, x_{P2}(t)\} \end{bmatrix} \in \mathbb{R}^{P \times 3P}. $$

Then $\mathbf{M}(t)$ can be written as $\mathbf{M}(t) = X(t)\beta$. Define operator $\Phi_t(B) = I_P - \sum_{\ell=1}^{L} \Phi^*(\ell, t) B^\ell$, where $B$ is the backshift operator s.t. $BZ(t) = Z(t-1)$ for any time series $\{Z(t)\}$. Denote $\Phi_t(B)\left[\mathbf{y}(t)\right]$ as $\tilde{\mathbf{y}}(t)$, and $\Phi_t(B)[X(t)]$ as $\tilde{X}(t)$. It can be derived that $\tilde{\mathbf{y}}(t) \overset{i.i.d.}{\sim} \mathcal{N}(\tilde{\mathbf{X}}(t)\beta, \Sigma_\epsilon)$.

Define $\theta$ to be the collection containing all the parameters involved. Hereby I get the joint likelihood of all the parameters conditional on the first $L$ signals:

$$\mathcal{L}_\epsilon(\theta) \propto |\Omega_\epsilon|^{(T-L)/2} \exp \left\{ -\frac{1}{2} \sum_{t=L+1}^{T} \left( \tilde{\mathbf{y}}(t) - \tilde{\mathbf{X}}(t)\beta \right)' \Omega_\epsilon \left( \tilde{\mathbf{y}}(t) - \tilde{\mathbf{X}}(t)\beta \right) \right\}, \quad (2.3)$$

20
where $\Omega_\epsilon = \Sigma_\epsilon^{-1}$ is the precision matrix. Since $L$ is usually small compared to the number of observations $T$, I work on the conditional likelihood $L_c$ instead of the full likelihood.

### 2.2.2 Prior specification

**Prior for HRF.** I need partial prior information about the shape of HRF in order to specify a prior distribution that gives biologically plausible and reasonable HRFs. More specifically, I need to collect information about the time to peak $T_{p,(p)}$ and time to post-stimulus undershoot $T_{p,(u)}$ for each ROI, from literature. Since $m_{p1} = \frac{a_{p1}-1}{b_{p1}}$ and $m_{p2} = \frac{a_{p2}-1}{b_{p2}}$ are the respective modes for the two gamma functions, they respectively approximate the time to peak and time to post-stimulus undershoot when they are far apart. Suppose two reasonable intervals for these two quantities are $[l_1, r_1]$, $[l_2, r_2]$, respectively. I can “convert” the intervals into the 95% “confidence” intervals of appropriate normal distributions: $m_{p1} = a_{p1}-1 \sim N(\mu_{p,(p)}, \sigma_{p,(p)}^2)$, $m_{p2} = a_{p2}-1 \sim N(\mu_{p,(u)}, \sigma_{p,(u)}^2)$, truncated with $0 < m_{p1} < m_{p2}$, where $\mu_{p,(p)} = (l_1 + r_1)/2$, $\sigma_{p,(p)}^2 = ((r_1 - l_1)/4)^2$, $\mu_{p,(u)} = (l_2 + r_2)/2$, $\sigma_{p,(u)}^2 = ((r_2 - l_2)/4)^2$, based on that the interval of mean plus or minus 2 standard deviations roughly gives a 95% coverage.

Finally the prior distribution of the HRF parameters is:

$$p(\Theta_{HRF}) \propto \prod_{p=1}^P \left\{ \exp\left(-\frac{1}{2\sigma_{p,(p)}^2}(\frac{a_{p1}-1}{b_{p1}} - \mu_{p,(p)})^2 - \frac{1}{2\sigma_{p,(u)}^2}(\frac{a_{p2}-1}{b_{p2}} - \mu_{p,(u)})^2\right) \times \frac{a_{p1}-1}{b_{p1}} \frac{a_{p2}-1}{b_{p2}} \right\} \times I(1 < a_{p1}, a_{p2} < Q; 0 < b_{p1}, b_{p2}; 0 < c_{p2} < Q; 0 < \frac{a_{p1}-1}{b_{p1}} < \frac{a_{p2}-1}{b_{p2}}) \right\}, \quad (2.4)$$

where $\Theta_{HRF} = \{a_{p1}, b_{p1}, a_{p2}, b_{p2}, c_{p2}|p = 1, 2, \ldots, P\}$ is the collection of HRF parameters for all ROIs. $Q$ is a sufficiently large number.

**Prior for VAR Parameters.** I impose the spike and slab prior [40] on the VAR parameters
Φ}_k^*(\ell)'s for variable selection:

\[ \phi_k^*(\ell)_{p,q} = \phi_k(\ell)_{p,q} \xi_{p,q}(\ell), \]

\[ \phi_k(\ell)_{p,q} \overset{i.i.d.}{\sim} N(0, \sigma_0^2), \quad \sigma_0 \text{ is a large constant}, \]

\[ \xi_{p,q}(\ell) = 0 \text{ or } 1, \]

\[ \xi_{p,q} = [\xi_{p,q}(0), \ldots, \xi_{p,q}(L)] = [1_j, 0_{L-j}] \sim \text{Multi}(1, \pi_{0,(p,q)}), \quad j \in \{0, 1, \ldots, L\}, \tag{2.5} \]

Since \( \phi_k(\ell)_{p,q} \) follows a continuous distribution, I have \( \Pr(\phi_k^*(\ell)_{p,q} \neq 0) = \Pr(\xi_{p,q}(\ell) = 1) \).

Here I assume under the two conditions \( \phi_k^*(\ell)_{p,q} \) is either both present or both absent, but may be different in strength. As the lag for each pair of ROIs can be 0, 1, \ldots, or \( L \), I further constrain the binary indicators so that \( \xi_{p,q} = [\xi_{p,q}(1), \ldots, \xi_{p,q}(L)] \in \mathbb{R}^L \) can only take values in \( \{[1_j, 0_{L-j}]|j = 0, \ldots, L\} \). The prior distribution can thus be formed by Equation (2.5), and \( \xi_{p,q} \sim \text{Multi}(1, \pi_{0,(p,q)}) \), where the constant vector \( \pi_{0,(p,q)} = [\pi_{0,(p,q)}, \pi_{0,(p,q)}, \ldots, \pi_{0,(p,q),L}] \in \mathbb{R}^{L+1} \).

For \( \forall j \), constant \( \pi_{0,(p,q),j} = \Pr(\xi_{p,q} = [1_j, 0_{L-j}]) \) is the prior probability of the corresponding lag being \( j \). When there is no prior information about the lags, a flat prior is used

\[ \pi_{0,(p,q)} = \left[ \frac{1}{L+1}, \ldots, \frac{1}{L+1} \right]. \tag{2.6} \]

**Prior for \( \beta \).** Let \( \Sigma_{0,\beta} = \sigma_{0,\beta}^2 I_{3P} \) and \( \sigma_{0,\beta} \) be a large constant. A vague prior for the amplitudes \( \beta \) is used: \( \beta \sim N(0_{3P}, \Sigma_{0,\beta}) \).

**Prior for \( \Sigma_{\epsilon} \).** A improper vague prior is used: \( p(\Sigma_{\epsilon}) \propto |\Sigma_{\epsilon}|^{-(P+1)/2}. \)
2.2.3 Inference

Inference is based on the MCMC posterior samples of the model parameters. For parameters $\Phi_k(\ell)$, $\xi_{p,q}$, $\beta$, $\Sigma_\epsilon$, the full conditional posterior distributions (conditional on data and all other parameters) are known distributions and can be easily sampled by a built-in sampler in many statistical software packages; for the HRF parameters, I use Metropolis-Hasting (M-H) algorithm with random walk proposal. Details of the sampling algorithm can be found in Appendix A.1. **Activation contrast.** I calculate a $(1 - \alpha) \times 100\%$ credible interval (CI) of normalized version of $\beta_{p,c}$, $\tilde{\beta}_{p,c}$; if the interval is on the right side of zero, I conclude ROI $p$ is activated. **Effective connectivity.** I first calculate the empirical posterior probability $\pi_{\text{post},(p,q),j} = \Pr(\xi_{p,q} = [1_j, 0_{L-j}] | y)$ for each lag $j$ ($y$ denotes the set of all the data), then find the lag $j$ that gives the largest $\pi_{\text{post},(p,q),j}$, denoted as $\tilde{j}$, to be the lag supported by the model. If $\tilde{j}$ is larger than 0, it suggests that there exists effective connectivity from ROI $p$ to ROI $q$ under condition $k$ with certain lags, in which case I shall further examine the corresponding CIs of $\phi_k(\ell)_{p,q}$, $\ell = 1, \ldots, \tilde{j}$, to study the direction and strength of the effective connectivity for condition $k$. **Functional connectivity.** I calculate the CI of $C_{\epsilon,(p,q)}$ derived from $\Omega_{\epsilon,(p,q)}$ and compare it with zero. **HRF.** I derive the HRF samples from the HRF parameters posterior samples and calculate a rough pointwise $(1 - \alpha) \times 100\%$ credible band of the HRF. Neuroscientists are usually interested in the time to peak and duration of the peak measured by full width at half maximum (FWHM). Thus, I also calculate the CIs of these quantities.

In this approach I calculate the $(1 - \alpha) \times 100\%$ posterior CI for each continuous variable by finding the $(\alpha/2)$-th and $(1 - \alpha/2)$-th quantiles of its posterior sample. The point estimates can be obtained similarly by calculating the median of the corresponding posterior sample, except for $\xi_{p,q}$ which is estimated by the mode. I also calculate the joint posterior probability of the presence and absence of all the possible pairwise connectivities in the targeting network, and select the network with the highest joint probability as the best effective
connectivity network. Doing so, I avoid having to estimate multiple models.

### 2.2.4 Computational and numerical issues

Here I discuss the potential problems in the estimation process of the proposed model, and the solutions I use to deal with some of the problems.

1. **Local maximum**. The M-H algorithm can sometimes be trapped in local maxima because of non-linearity of the HRF parameterization, therefore I run a number of chains starting from different HRF initial values.

2. **Deconvolution issue**. When the signal-to-noise ratio is relatively high, different HRF shapes could have very similar performance in terms of mean fit. In such cases, it is hard to do deconvolution precisely and thus it may lead to erroneous inference on HRF and other quantities of interest.

3. **Identifiability issue in the HRF**. Different HRF parameters could produce very similar HRF shapes. This soft identifiability issue may cause the instability of the MCMC of the HRF parameters. However I am more interested in the HRF shape instead of the parameters themselves, therefore this could be of less concern.

4. **Heavy computation**. Due to the multivariate time series data and so many unknown parameters, computation becomes complicated and heavy. To reduce computation time, I program with C++ core using RcppArmadillo package [48].

### 2.3 Simulation study

In this section, I compare the performance of the proposed method to the standard GLM method.
2.3.1 Simulation settings

I simulated 30 independent datasets, according to Section 2.2.1. Each dataset has 4 ROIs, 4 independent sessions with the 4 consecutive blocks: stimulus condition, control condition, stimulus condition, control condition. Each block of condition lasts for 16 time points, with TR=1s. The model parameter are set as follows:

HRF: for ROI 1 and 2, I used the SPM canonical HRF ($T_p = 5s$, $T_u = 15s$); for ROI 3 and 4, I used different HRFs that peak a little later. The HRFs can be found in Figure 2.1. Activation: $\beta_{p1}$’s were generated independently from folded $N(240,6400)$, and $\beta_{p2}$’s were generated independently from folded $N(0.3\beta_{p1}, 6400)$, except that for ROI 2, $\beta_{p2} = \beta_{p1}$ (not activated). Connectivity: both the effective and functional connectivity network are summarized in Figure 2.6(a).

The parameter values are provided in Appendix A.2. At last, mean trend was removed from all the time series by subtracting the average for each ROI in each dataset. An example of the generated time series of a ROI can be found in Figure 2.2.

![HRF](image)

Figure 2.1: HRFs of four simulated ROIs.
2.3.2 Analysis

Assuming I had partial knowledge that the HRFs at ROI 1, 2 and 4 were similar to the SPM canonical HRF, and the HRF at ROI 3 may have a little later peak and undershoot, I used the following prior constants for the HRFs: $\mu_{p,(p)} = 5$, $\sigma^2_{p,(p)} = 6.25$, $\mu_{p,(u)} = 15$, $\sigma^2_{p,(u)} = 25$, $p = 1, 2, 4$, and $\mu_{3,(p)} = 6$, $\sigma^2_{3,(p)} = 9$, $\mu_{3,(u)} = 17$, $\sigma^2_{3,(u)} = 36$. It contained the prior belief that with 95% confidence, the time to peak was between 0.1s and 9.9s and centered around 5s for ROI 1, 2 and 4, and was between 0.1s and 11.9 and centered around 6s for ROI 3; and with 95% confidence, the time to undershoot was between 5.2s and 24.8s and centered around 15s for ROI 1, 2 and 4, and was between 5.2s and 28.8s and centered around 17s. I also constrained $a_{p1}, a_{p2}, c_{p2}$ to be smaller than $Q = 100$, $p = 1, \cdots, 4$. Examples of HRF from the first set of HRF prior (ROI 1, 2 & 4) are shown in Figure 2.3.

For simplicity I used order $L = 1$. As a result, the first signal for each of the 4 sessions needed to be discarded (4 signals in total). Assuming I did not have any information about the two types of connectivity or the activations, I used vague priors as discussed in Section 2.2.2 with the following prior constants: $\sigma_{0,\phi} = 100$, $\pi_{0,(p,q)} = [0.5, 0.5]$, $p, q = 1, \cdots, 4$ (see Equation (2.6)); $\Sigma_{0,\beta} = 10^7I_{3P}$.
Note that although each dataset consisted of multiple independent sessions, I just needed to change Equation (2.3) to be the product of the likelihood for each session. The performance of the procedure was assessed in terms of power, type I error, 95% CI coverage, and bias. For effective connectivity, “network correctness” was also assessed, which was defined as the proportion of the estimated network being equal to the true network. The details of the measures are as follows:

1. Ability to recover HRF: (i) bias (point estimate minus true value) in time to peak; (ii) bias in FWHM as a measure of duration of peak.

2. Ability to recover activation: (i) power, the proportion of the 95% CI of $\tilde{\beta}_{p,(c)}$, $p = 1, 3, 4$, lying entirely on the right side of 0 (ROI 1, 3 and 4 were more activated in the stimulus condition); (ii) type I error, the proportion of the 95% CI of $\tilde{\beta}_{2c}$ not containing 0 (ROI 2 was equally activated in the two conditions); (iii) 95% CI coverage, the proportion of the 95% CI of $\tilde{\beta}_{p,(c)}$ containing the true value; (iv) relative bias, calculated as bias divided by average true values of $\tilde{\beta}_{p,(c)}$.

3. Ability to recover effective connectivity: (i) power, the proportion of the estimated $\xi_{p,q}(1)$ being 1 when the true value is 1; (ii) type I error, the proportion of $\xi_{p,q}(1)$ being
1 when the true value is zero; (iii) “network correctness”, proportion of the estimated network being the true network; (iv) 95% CI coverage, the proportion of the 95% CI of \( \phi_k(1)_{p,q} \) containing the true value when the true value is non-zero; (v) relative bias (divided by average true values of \( \phi_k^*(1)_{p,q} \)).

4. Ability to recover functional connectivity: (i) power, the proportion of the 95% CI of \( C_{\epsilon,(p,q)} \), \( p \neq q \) not containing 0 when the true value is non-zero; (ii) type I error, the proportion of the 95% CI not containing 0 when the true value is zero; (iii) 95% CI coverage, the proportion of the 95% CI containing the true value; (iv) relative bias (divided by average true \( C_{\epsilon,(p,q)} \) in absolute values).

We also applied the standard GLM on the simulated data using the SPM canonical HRF for all ROIs:

\[
y_p(t) = \beta_{p0} + \beta_{p1}(h_c \ast c_1)(t) + \beta_{p2}(h_c \ast c_2)(t) + u_p(t), \tag{2.7}
\]

where \( h_c(t) = g(t; 6, 1) - \frac{1}{6}g(t; 16, 1) \) is the canonical HRF. The least square estimator was used to get the point estimates \( \hat{\beta}_{p0}, \hat{\beta}_{p1}, \hat{\beta}_{p2} \) and the residuals \( \hat{u}_p(t) \). After that, I ran a VAR(1) model on the multivariate residuals \( \hat{u}(t) = [\hat{u}_1(t), \ldots, \hat{u}_P(t)] \), assuming \( \epsilon(t) \overset{i.i.d.}{\sim} N(0, \Sigma_\epsilon) \):

\[
\hat{u}(t) = \hat{u}(t - 1)\Phi(1) + \epsilon(t). \tag{2.8}
\]

The maximum likelihood estimator \( \hat{\Phi} \) and \( \hat{\Sigma}_\epsilon \) (based on the conditional likelihood \( L_c \)) can be computed through conditional maximization algorithm.

The point estimator for activation contrast can be obtained by \( \hat{\beta}_{p1} - \hat{\beta}_{p2} \), the point estimator for effective connectivity can be obtained by \( \hat{\Phi} \), and the point estimator for functional con-
nectivity can be obtained by calculating the partial correlation from the estimated covariance matrix $\hat{\Sigma}_e$.

Inference for the standard GLM method was based on the corresponding 95% CI with normal approximations and the standard errors of the parameters. (The abbreviation CI stands for credible interval when the context is the proposed Bayesian method, and stands for confidence interval when the context is standard GLM.) Performance in terms of connectivity was assessed similarly as the proposed BVAR-HRF method, except that here the true value of effective connectivity was considered as the average effective connectivity of two conditions, since it could only estimate the overall effective connectivity. The estimated effective network was composed of the significantly non-zero effective connectivity. To assess the performance in terms of activation, I also manipulated the shape of HRF by decreasing or increasing time to peak by 1s (HRF 1 and HRF 2, see Figure 2.4) to see how this would affect the inference on activations. Finally, I compared the performances of the two methods.

![Figure 2.4: Three HRFs used in the GLM analysis to assess performance on activation.](image)

### 2.3.3 Results

Numerical results from the two methods are summarized in Table 2.1. Note that the values are averaged across ROIs. A typical result for estimated HRF curve is presented in Figure 2.5. Figure 2.6 in the end of this subsection shows the resulting networks from the two methods.
<table>
<thead>
<tr>
<th>Performance</th>
<th>BVAR-HRF</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HRF</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias in time to peak (sec)</td>
<td>0.056</td>
<td>N/A</td>
</tr>
<tr>
<td>Bias in FWHM (sec)</td>
<td>-0.023</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Activation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Type I error</td>
<td>0.067</td>
<td>0.033</td>
</tr>
<tr>
<td>Type I error (HRF 1)</td>
<td>N/A</td>
<td>0.967</td>
</tr>
<tr>
<td>Type I error (HRF 2)</td>
<td>N/A</td>
<td>0.900</td>
</tr>
<tr>
<td>Joint correctness</td>
<td>0.933</td>
<td>0.967</td>
</tr>
<tr>
<td>Recovering activation</td>
<td>0.858</td>
<td>0.483</td>
</tr>
<tr>
<td>Relative bias</td>
<td>2.3%</td>
<td>18.3%</td>
</tr>
<tr>
<td><strong>Effective connectivity</strong></td>
<td>0.856</td>
<td>0.717</td>
</tr>
<tr>
<td>Power</td>
<td>0.000</td>
<td>0.263</td>
</tr>
<tr>
<td>Joint correctness</td>
<td>0.216</td>
<td>0.000</td>
</tr>
<tr>
<td>Recovering effective connectivity</td>
<td>0.950</td>
<td>0.685</td>
</tr>
<tr>
<td>Relative bias</td>
<td>-2.8%</td>
<td>39.2%</td>
</tr>
<tr>
<td><strong>Functional connectivity</strong></td>
<td>0.883</td>
<td>0.650</td>
</tr>
<tr>
<td>Power</td>
<td>0.067</td>
<td>0.583</td>
</tr>
<tr>
<td>Joint correctness</td>
<td>0.433</td>
<td>0.000</td>
</tr>
<tr>
<td>Recovering functional connectivity</td>
<td>0.911</td>
<td>0.461</td>
</tr>
<tr>
<td>Relative bias</td>
<td>3.9%</td>
<td>12.2%</td>
</tr>
</tbody>
</table>

Table 2.1: Results comparing the proposed BVAR-HRF method with the standard GLM method. For bias terms, a negative sign indicates an underestimation.

Table 2.1 shows BVAR-HRF method did a great job in estimating the position and the duration of the peak for HRF. Both methods performed well in determining whether the ROI was activated or not (power, type I error). BVAR-HRF outperformed the standard approach in terms of relative bias and the coverage of 95% CIs for activation. Also, after manipulating the HRF used for the standard GLM by decreasing or increasing 1s in time to peak, the type I error went up dramatically. This implies GLM tends to be highly vulnerable for false positive when the assumed HRF was even slightly different from the truth.

In terms of connectivity, BVAR-HRF performed relatively well in all aspects, while the standard GLM method performed poorly due to the erroneous residuals to start the connectivity analysis with.

Figure 2.6 gives the most frequent network out of the 30 estimated networks for each of the two methods. For BVAR-HRF, I further obtained the joint probability of the network, which was 0.7452 on average, confirming it was indeed the best network suggested by the
2.4 Application to fMRI data

2.4.1 Description of the experiment and the data

To demonstrate utility of the proposed model, I analyzed the fMRI data from a stroke patient with some residual motor deficit on the left side of the body. For simplicity of demonstration
and limited amount of data for a single subject analysis, I focused only on three of the ROIs known to be involved in this task after stroke: left and right dorsal premotor cortex (LPMd and RPMd), and a midline supplementary motor area (SMA) (Figure 1.3).

FMRI contrasted affected (left) hand grasp-release movements (stimulus condition) with rest condition (control condition). The experiment was divided into 3 sessions, each with the same experiment design as in Figure 1.5. The total time series points considered in the analysis is $T = 3 \times 48 = 144$, and $TR = 2s$. Data was pre-processed using SPM8 software. For each ROI, the time series of the included voxels were summarized into the mean time series. Finally, the mean time series was detrended with time as the covariate to remove the linear drift effect.

### 2.4.2 Analysis

In order to test for sensitivity to prior specification, I applied the BVAR-HRF model using three different prior distributions for HRF (see Table 2.2). ACF plots showed insufficiency of VAR(1) but sufficiency of VAR(2), so I chose VAR(2) as the final model. The prior distributions for the other parameters, $\Phi_k(\ell)$, $\xi_{p,q}$ and $\beta$, were the same as those in the simulation study in Section 2.3.2, except for $\xi_{p,q}$ the flat prior became $\pi_{0,(p,q)} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$. 

<table>
<thead>
<tr>
<th>Prior 1</th>
<th>Prior 2</th>
<th>Prior 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{p1} \sim \text{N}(\mu_{p,(p)}, \sigma_{p,(p)}^2)$</td>
<td>N(5, 9)</td>
<td>N(5, 6.25)</td>
</tr>
<tr>
<td>$m_{p2} \sim \text{N}(\mu_{p,(u)}, \sigma_{p,(u)}^2)$</td>
<td>N(15, 36)</td>
<td>N(15, 25)</td>
</tr>
</tbody>
</table>

Table 2.2: The three prior distributions used in the analysis.
2.4.3 Results

Results from BVAR-HRF are summarized in Table 2.3. The results were similar across the three sets of prior distributions. Estimated HRF curves and diagnostic plots are shown only for prior 1 as an example. Figure 2.7 shows the estimated mean followed the data well. The trace plot of the unnormalized log posterior density shows the chain was stable (see Figure A.1 in Appendix A); The ACF and PACF plots for the estimated $\epsilon(t)$ show no further lag will be needed to be included in the model (see Figure A.2 and A.3 in Appendix A).

![Figure 2.7: Mean fit from real data analysis.](image)

HRF: median HRF curves and HRF samples for each ROI are given in Figure 2.8. RPMd peaked later than LPMd and had a smaller duration of peak.

Activation: in all three ROIs, $\tilde{\beta}_{p,c}$ was significantly greater than zero, suggesting LPMd, RPMd and SMA were all activated during task condition.

Effective connectivity: Table 2.3 (the 4th row group) and Figure 2.9 suggested there was connectivity from LPMd to SMA with lag 1, and connectivity from LPMd to LPMd, and from RPMd to SMA with lag 2. Only connectivity from SMA to RPMd at lag 1 under task condition was estimated to be negative. However, zero was close to the estimate and was well covered by the 95% CI of $\phi_1(1)_{3,2}$. This was also the case for the connectivity under rest condition. Effective connectivity from SMA to RPMd under both conditions was probably negligible until lag 2. Other non-zero connectivities were positive and generally higher under
<table>
<thead>
<tr>
<th></th>
<th>Prior 1</th>
<th>Prior 2</th>
<th>Prior 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>95% CI</td>
<td>Median</td>
</tr>
<tr>
<td>Time to peak (sec): LPMd</td>
<td>3.1 [1.6, 4.6]</td>
<td>3.3 [1.7, 4.6]</td>
<td>3.2 [1.8, 4.6]</td>
</tr>
<tr>
<td>Time to peak (sec): RPMd</td>
<td>5.9 [5.0, 7.0]</td>
<td>5.9 [5.0, 7.0]</td>
<td>6.1 [5.0, 7.0]</td>
</tr>
<tr>
<td>Time to peak (sec): SMA</td>
<td>4.4 [3.1, 5.9]</td>
<td>4.5 [3.4, 5.7]</td>
<td>4.5 [3.4, 5.7]</td>
</tr>
<tr>
<td>FWHM (sec): LPMd</td>
<td>6.3 [4.9, 8.5]</td>
<td>6.6 [5.3, 8.7]</td>
<td>6.3 [5.2, 8.5]</td>
</tr>
<tr>
<td>FWHM (sec): RPMd</td>
<td>2.1 [1.3, 5.3]</td>
<td>1.7 [1.3, 3.7]</td>
<td>2.1 [1.5, 3.0]</td>
</tr>
<tr>
<td>FWHM (sec): SMA</td>
<td>5.7 [4.0, 7.0]</td>
<td>5.6 [4.4, 6.8]</td>
<td>5.6 [4.4, 6.7]</td>
</tr>
<tr>
<td>Activation: LPMd</td>
<td>3.6 [2.5, 4.8]</td>
<td>3.5 [2.5, 4.6]</td>
<td>3.5 [2.5, 4.7]</td>
</tr>
<tr>
<td>Activation: SMA</td>
<td>5.6 [4.4, 7.2]</td>
<td>5.5 [4.3, 6.9]</td>
<td>5.6 [4.4, 6.9]</td>
</tr>
<tr>
<td>Effective (task, lag1): LPMd to LPMd</td>
<td>0.32 [0.12, 0.52]</td>
<td>0.32 [0.12, 0.52]</td>
<td>0.32 [0.12, 0.52]</td>
</tr>
<tr>
<td>Effective (rest, lag1): LPMd to LPMd</td>
<td>0.39 [0.18, 0.59]</td>
<td>0.38 [0.17, 0.59]</td>
<td>0.39 [0.17, 0.58]</td>
</tr>
<tr>
<td>Effective (task, lag1): SMA to RPMd</td>
<td>-0.09 [-0.41, 0.24]</td>
<td>-0.09 [-0.40, 0.22]</td>
<td>-0.09 [-0.40, 0.22]</td>
</tr>
<tr>
<td>Effective (rest, lag1): SMA to RPMd</td>
<td>0.14 [-0.18, 0.44]</td>
<td>0.14 [-0.17, 0.44]</td>
<td>0.14 [-0.18, 0.45]</td>
</tr>
<tr>
<td>Effective (task, lag1): LPMd to SMA</td>
<td>0.74 [0.55, 0.92]</td>
<td>0.73 [0.56, 0.93]</td>
<td>0.72 [0.53, 0.92]</td>
</tr>
<tr>
<td>Effective (rest, lag1): LPMd to SMA</td>
<td>0.55 [0.30, 0.81]</td>
<td>0.55 [0.30, 0.80]</td>
<td>0.54 [0.29, 0.80]</td>
</tr>
<tr>
<td>Effective (task, lag2): LPMd to LPMd</td>
<td>0.55 [0.35, 0.74]</td>
<td>0.54 [0.35, 0.74]</td>
<td>0.55 [0.36, 0.74]</td>
</tr>
<tr>
<td>Effective (rest, lag2): LPMd to LPMd</td>
<td>0.44 [0.26, 0.63]</td>
<td>0.44 [0.26, 0.63]</td>
<td>0.44 [0.26, 0.63]</td>
</tr>
<tr>
<td>Effective (task, lag2): SMA to RPMd</td>
<td>0.83 [0.49, 1.15]</td>
<td>0.82 [0.49, 1.13]</td>
<td>0.83 [0.52, 1.15]</td>
</tr>
<tr>
<td>Effective (rest, lag2): SMA to RPMd</td>
<td>0.40 [0.09, 0.71]</td>
<td>0.40 [0.11, 0.73]</td>
<td>0.39 [0.09, 0.70]</td>
</tr>
<tr>
<td>Functional: LPMd and RPMd</td>
<td>0.51 [0.37, 0.63]</td>
<td>0.51 [0.36, 0.63]</td>
<td>0.51 [0.36, 0.62]</td>
</tr>
<tr>
<td>Functional: LPMd and SMA</td>
<td>0.65 [0.54, 0.74]</td>
<td>0.65 [0.53, 0.74]</td>
<td>0.65 [0.54, 0.74]</td>
</tr>
<tr>
<td>Functional: RPMd and SMA</td>
<td>-0.38 [-0.53, -0.20]</td>
<td>-0.37 [-0.51, -0.20]</td>
<td>-0.38 [-0.53, -0.21]</td>
</tr>
</tbody>
</table>

Table 2.3: Analysis results from the three priors.

the task condition than under the rest condition.

Functional connectivity: Table 2.3 (the 5th row group) shows evidence that there existed positive functional connectivity between LPMd and LPMd, LPMd and SMA, but negative functional connectivity between RPMd and SMA.

The resulting network from BVAR-HRF is provided in Figure 2.10. There appeared to be effective connectivity from LPMd to RPMd through SMA.

2.4.4 Interpretation

Here I summarize findings from the proposed BVAR-HRF method: 1) a delayed HRF in the PMd of the lesioned (right) hemisphere, 2) activation in all three secondary motor areas
Figure 2.8: Estimated HRFs from prior 1.

Figure 2.9: Histograms of $\phi_k(\ell)_{p,q}$. Blue dashed lines indicate position of zero.

included in the analysis, and 3) brain connectivities that are also supported by other studies. Using the proposed method, I observed a notable difference in the HRF in the lesioned versus unlesioned hemisphere: a delayed time to peak in ipsilesional PMd (RPMd) compared to contralesional PMd (LPMd). This is similar to the findings in [2] that the time to peak increased in the stroke-affected hemisphere compared to the unaffected hemisphere as subjects progressed beyond the acute phase. These results support adjusting the HRF (across ROIs and between subjects) to optimize studies of BOLD activation and connectivity derived from fMRI.

The proposed method also suggested that all three secondary motor areas in the analysis
Figure 2.10: Estimated connectivity networks. (a) Functional connectivity network. (b) Effective connectivity network under task condition. (c) Effective connectivity network under rest condition. Curves with ordinary arrow heads correspond to lag 1 and curves with oval arrow heads correspond to lag 2. The darker the sign is, the larger the connectivity measure is.

(namely, SMA, LPMd and RPMd) were activated during simple motor movements. It is interesting that these regions are not typically activated in healthy individuals during simple motor tasks. This finding is consistent with the fact that PMd is known to be important for higher order functions involved in motor planning [46].

Finally, the proposed method observed effective connectivity from contralesional PMd to SMA, which is in-line with previous studies that report the compensatory role of contralesional PMd in paretic arm movement, e.g. [30]. The proposed method also captured effective connectivity from SMA to ipsilesional PMd. An influence of SMA onto ipsilesional PMd has also been found in a study of effective connectivity by [49]. The excitatory functional connectivity between bilateral PMd regions found by BVAR-HRF may represent an adaptation supporting post-stroke recovery, as these connectivities positively correlate with measures of motor impairment [57].
2.5 Conclusion

The simulation study has demonstrated the advantages of the proposed BVAR-HRF model: (1) It ruled out the confounding effects from HRF by including HRF as unknown parameters in the model. (2) The proposed BVAR-HRF model captured different effective connectivity across experimental conditions. (3) Bayesian paradigm in BVAR-HRF made the inference straightforward; especially with the use of spike and slab prior, one could easily pick the best effective connectivity network without having to estimate multiple models. (4) Bayesian method also made it possible to include prior information into the model.

There are still several limitations with the BVAR-HRF model: (1) Computation for the method could be quite heavy. In future study, I would like to try using non-linear optimization algorithm such as Levenberg-Marquardt in the M-H sampling step to improve the efficiency and save total computation time. (2) Conclusions made from the model is based on the hemodynamics of the brain instead of the neuronal activities directly, so I need to be cautious when extending the conclusions from the hemodynamic level to the neuronal level. (3) The model only handles single subject analysis, however, this can be easily generalized to the hierarchical setting that is able to make inference at group level and measure inter-subject variability, which is introduced in the next Chapter.
Chapter 3

Bayesian VAR Model for Multi-Subject fMRI Data

3.1 Introduction

As an imaging modality, fMRI is able to indirectly measure neuronal activity in the brain through the hemodynamic response: higher level of neuronal activity at a localized region requires a greater amount of oxygen, which results in higher level of the blood-oxygen-level-dependent (BOLD) contrast fMRI signal at that location. By analyzing fMRI data, it is possible to study local neuronal activation as well as inter-regional connectivity. The shape of the hemodynamic response following a point stimulus is called the hemodynamic response function (HRF). The HRF is likely to vary across brain regions and subjects. Thus, it is very important to correctly estimate the HRF for each region and subject in order to correctly infer information regarding activation and connectivity because neural activity is indirectly measured through the hemodynamic response.
3.1.1 Proposed approach

Here I developed a Bayesian approach that jointly estimates activation and connectivity. The proposed approach provides estimates of region/subject-specific hemodynamic response functions (HRFs). It utilizes the general linear model (GLM) to describe activation, a Bayesian vector autoregression (BVAR) to measure connectivity, and a constrained linear basis set to model the unknown HRFs. The proposed model provides a hierarchical framework that handles group level activation, connectivity and their variability among subjects. With the hierarchical structure, subject-specific estimates for activation and connectivity are obtained by pooling information from other subjects. Spike and slab priors are placed on the model parameters that describe the connectivity network. This allows us to explore the full posterior distribution of possible connectivity networks. Additionally, I allow condition-specific connectivity measures, thus making it possible to detect differences in connectivity between experimental conditions. My goal is to build a model that studies local activation and connectivity between brain regions while controlling for variation in the HRFs, and compares the activation and connectivity patterns between different groups of subjects, in order to explore the effects of grouping factor.

3.1.2 Current statistical methods and their limitations

In the past 20 years, there has been an increasing number of papers on statistical methods for fMRI data. Detailed reviews of the existing methods are provided by, for example, [33] and [67]. Methods for brain activation are proposed in [18, 63, 50, 12, 26, 65, 13]. Methods for brain connectivity are developed in [17, 27, 8, 11, 31, 23, 36]. While methods for activation usually do not consider connectivity estimation, [12] in fact considered connectivity estimation but uses a two-stage approach that estimates connectivity from the residuals of the activation analysis, and hence is not optimal for properly assessing the uncertainty associated
to those estimates. The method in [27] modeled connectivity using vector autoregressive models, but did not take into account the systematic changes in the BOLD response induced by external stimuli. In [11] dynamic connectivity regression was developed which models the changes in connectivity, but again does not take into account the systematic changes of the BOLD response. The method in [23] considered the BOLD response evoked by stimuli and also the differential connectivity across different experimental conditions. However, the BOLD response shape is based on a pre-fixed shape of HRF that is identical across all brain regions and subjects, which may lead to erroneous conclusions (see Chapter 2). In [8] and [36] the experimental conditions were also taken into account, but connectivity were modeled through a two-stage approach and used a pre-specified HRF. The method in [31] simultaneously analyzed activation and connectivity in the spectrum domain but also used a pre-specified HRF. Dynamic causal modeling (DCM) was proposed in [17], which addresses ROI-specific HRF and connectivity simultaneously. However, DCM heavily relies on the correct specification of the biological assumptions, and needs a set of plausible connectivity networks a-priori which are difficult to verify from the observed data.

There are many ways to parameterize the HRFs. Non-linear parameterization approaches that use gamma functions, cosine functions [66], or inverse logit functions [34] can be considered. Linear basis approaches such as spline basis functions, finite impulse response bases [42] and semi-parametric models [68] are also available. Non-linear parameterization takes more computational effort and is more prone to the local maxima problem and numerical stability issues. This problem was also seen in Chapter 2. On the other hand, linear parameterization is easier from the inferential viewpoint, but often requires more parameters and constraints in order to provide enough flexibility in the HRF shapes. Woolrich proposed constrained linear basis sets which are able to provide both flexible and reasonable HRF shapes with a relatively small number of HRF basis functions while keeping the computation cost low [62]. Therefore, in this proposed model I adopt the constrained linear basis approach to parameterize the HRFs.
In a nutshell, I propose a model that has the following advantages over currently available methods: (1) it simultaneously estimates activation, connectivity and HRFs; (2) it provides ROI-specific and subject-specific HRFs, as well as condition-specific connectivity measures; (3) it increases the power in detecting group differences by pooling information across subjects via hierarchical modeling; (4) it provides full posterior distributions of all the model parameters, including group-specific activation parameters and brain connectivity networks; (5) it can incorporate available scientifically relevant prior information.

The remainder of the chapter is organized as follows. In Section 3.2, I develop the approach including the model, the prior distributions and the inferential procedure. In Section 3.3, I present a simulation study. In Section 3.4, I present the analysis of the fMRI data from the stroke study. In Section 3.5, I summarize the approach and discuss future research directions.

3.2 Methodology

In this section I describe the proposed Bayesian hierarchical modeling approach and related inferential procedures. As previously stated, the proposed approach allows to infer activation and connectivity at the group level and at the same time takes into account subject-specific variability in the HRFs.

3.2.1 Hierarchical model: subject level

I follow the general linear model to describe the fMRI signal for a given ROI and subject:

\[
y_p^{(s,r)}(t) = \beta_{p0}^{(s)} + x_{p1}^{(s,r)}(t)\beta_{p1}^{(s)} + x_{p2}^{(s,r)}(t)\beta_{p2}^{(s)} + u_p^{(s,r)}(t),
\]

\[
x_{pk}^{(s)}(t) = [h_p^{(s)} * c_k](t),
\]
with ROI \( p = 1, 2, \ldots, P \); subject \( s = 1, 2, \ldots, S \); session \( r = 1, 2, \ldots, R \); and time (in a session) \( t = 1, 2, \ldots, T \). The response variable \( y_{p}^{(s,r)}(t) \) is the preprocessed fMRI signal; the covariate \( x_{pk}^{(s,r)}(t) \) represents the expected shape of BOLD response which is the convolution between the subject and region specific HRF \( h_{p}^{(s)}(t) \) and the condition indicator \( c_{k}(t) \); the regression coefficient \( \beta_{pk}^{(s)} \) represents the amplitude of the BOLD response which indirectly reflects the strength of neuronal activity; \( c_{k}(t) \) is the known binary condition indicator for whether condition \( k \) is on (\( k = 1 \) for task condition and \( k = 2 \) for rest condition; see Figure 2 for the timing of the conditions); and \( u_{p}^{(s,r)}(t) \) is the unexplained noise. The effect of the convolution has been illustrated in Figure 1.1. The expected BOLD response is no longer block-shaped as the stimulus input, but is a smooth and delayed transformation of the input. In this model, I have imposed the following assumptions: (1) the BOLD amplitude remains the same within a condition and across sessions; (2) the HRF shape within each ROI and subject remains constant across conditions and sessions.

On modeling activation. Neuroscientists are interested in estimating the difference between activity strength of two conditions as measured by a region-specific contrast \( \beta_{p,(c)}^{(s)} = \beta_{p1}^{(s)} - \beta_{p2}^{(s)} \). When ROI \( p \) has stronger activity during the task condition than during the rest condition, it suggests that this ROI is implicated in executing the task. Therefore, when \( \beta_{p,(c)}^{(s)} > 0 \), ROI \( p \) is labeled as active in subject \( s \), and \( \beta_{p,(c)}^{(s)} \) is used to measure the strength of the local activation.

On modeling connectivity. Following [23], the total unexplained noise \( u^{(s,r)}(t) \in \mathbb{R}^{1 \times P} \) (row vector) is assumed to follow a Gaussian vector autoregressive model (VAR) with predetermined order \( L \) and condition-dependent VAR coefficients, i.e.,

\[
\mathbf{u}^{(s,r)}(t) = \left[ u_{1}^{(s,r)}(t), \ldots, u_{P}^{(s,r)}(t) \right] = \sum_{\ell=1}^{L} \Phi^{(s)}(\ell, t) \mathbf{u}^{(s,r)}(t - \ell) + \mathbf{e}^{(s,r)}(t), \tag{3.3}
\]
where \( \epsilon^{(s)}(t) = [\epsilon_1^{(s,r)}(t), \ldots, \epsilon_P^{(s,r)}(t)]^T \overset{\text{ind}}{\sim} N(0, \Sigma^{(s)}) \), and

\[
\Phi^{(s)}(\ell, t) = c_1(t - \ell)\Phi_1^{(s)}(\ell) + c_2(t - \ell)\Phi_2^{(s)}(\ell).
\] (3.4)

Denote the \((p,q)\)-th element of \(\Phi^{(s)}(\ell, t)\) as \(\Phi_k^{(s)}(\ell)_{p,q}\). This describes the linear relation between ROI \(q\) and ROI \(p\) with time lag \(\ell\): when \(\Phi_k^{(s)}(\ell)_{p,q} \neq 0\), ROI \(q\) at time \(t - \ell\) helps predict ROI \(p\) at the current time point \(t\). The temporal precedence of ROI \(q\) in this relationship suggests the interpretation of \(\Phi_k^{(s)}(\ell)_{p,q}\) as a measure of the directed influence from ROI \(q\) to ROI \(p\). This type of influence is termed as “effective connectivity” in neuroscience. Therefore, \(\Phi_k^{(s)}(\ell)_{p,q}\) can be used to measure effective connectivity from ROI \(q\) to ROI \(p\). For simplicity, in the remainder of this chapter I refer to this as “connectivity”. Also, it is possible that the manner in which an ROI influences another varies under different conditions. Hereby, I allow \(\Phi_k^{(s)}(\ell)_{p,q}\) to depend on the condition at the corresponding past time point as specified in Equation (3.4).

I use a Bayesian variable selection approach [40, 21, 29] to determine which elements of \(\Phi_k^{(s)}(\ell, t)_{p,q}\) are different from zero, and therefore infer the connectivity network across ROIs. More specifically, I impose a “spike and slab” structure on \(\Phi_k^{(s)}(\ell)\) given by

\[
\Phi_k^{(s)}(\ell)_{p,q} = \Phi_k^{(s)}(\ell)_{p,q} \xi_{p,q}^{(s)}(\ell), \ \text{with} \ \Phi_k^{(s)}(\ell)_{p,q} \in \mathbb{R}, \ \xi_{p,q}^{(s)}(\ell) \in \{0, 1\},
\] (3.5)

so that \(\Phi_k^{(s)}(\ell)_{p,q} = 0 \iff \xi_{p,q}^{(s)}(\ell) = 0 \) (a.s.); \(\Pr(\Phi_k^{(s)}(\ell)_{p,q} \neq 0) = \Pr(\xi_{p,q}^{(s)}(\ell) = 1)\), and \(\Phi_k^{(s)}(\ell)_{p,q} = \Phi_k^{(s)}(\ell)_{p,q} \iff \xi_{p,q}^{(s)}(\ell) = 1 \) (a.s.).

When \(\Phi_k^{(s)}(\ell)_{p,q} = 0\), or equivalently when \(\xi_{p,q}^{(s)}(\ell) = 0\) for all \(\ell\), it is claimed that there is no
connectivity from ROI \( q \) to ROI \( p \). Otherwise, there is connectivity from ROI \( q \) to ROI \( p \).

On modeling the HRFs. A linear basis for the representation of the HRFs was obtained based on the approach in [62]. In particular, using the linear basis, the HRF of subject \( s \) and region \( p \) can be written in terms of \( J \) basis vectors as 

\[
h^{(s)}_{p}(t) = \sum_{j=1}^{J} H(t, j) d^{(s)}_j \equiv H(t, :) d^{(s)}_p,
\]

where \( H \) is the basis matrix with each column being a basis vector, and \( d^{(s)}_p \in \mathbb{R}^J \) is the vector of HRF basis coefficients for ROI \( p \) and subject \( s \). For identifiability, I impose a normalization constraint on \( d^{(s)}_p \) so that \( \sum_{t=1}^{T} h^{(s)}_{p}(t) = 1 \). Further details on the basis representation and how to choose \( J \) are given in Section 3.2.3.

I use “\( A(i,:) \)”, “\( A(:,j) \)”, and “\( A(i,j) \)” to denote, respectively, the \( i \)-th row, the \( j \)-th column, and the \((i,j)\)-th element of any matrix \( A \). Then, \( x^{(s)}_{pk}(t) \) can be written as 

\[
x^{(s)}_{pk}(t) = [H * c_k(t)] d^{(s)}_p := \Lambda_k(t, :) d^{(s)}_p, \quad t = 1, 2, \ldots, T,
\]

where the matrix \( \Lambda_k \in \mathbb{R}^{T \times J} \) consists of the convolutions between each linear basis vector \( H(:,j) \), and the condition indicator \( c_k(t) \).

Therefore, combining this with Equation (3.1) leads to the following equation:

\[
y^{(s,r)}_{p}(t) = \beta^{(s)}_{p0} + \Lambda_1(t, :) d^{(s)}_p \beta^{(s)}_{p1} + \Lambda_2(t, :) d^{(s)}_p \beta^{(s)}_{p2} + u^{(s,r)}_{p}(t).
\] (3.6)

### 3.2.2 Hierarchical model: group level

Since the signal to noise ratio is usually low in fMRI data (typically below 5%), it would be beneficial to borrow information across subjects within a relatively homogeneous group. In order to do this, I use a Bayesian hierarchical modeling approach. Suppose there are \( S \) subjects. Each subject belongs to a single group, denoted by \( g_s \), from a total of \( G \) groups. Define: \( \beta^{(s)}_p = [\beta^{(s)}_{p0}, \beta^{(s)}_{p1}, \beta^{(s)}_{p2}]^\prime, \quad p = 1, 2, \ldots, P \), \( \beta^{(s)} = [(\beta^{(s)}_1)^\prime, \ldots, (\beta^{(s)}_P)^\prime] \); \( \phi^{(s)}_k = [\text{vec}(\Phi^{(s)}_k(1))^\prime, \ldots, \text{vec}(\Phi^{(s)}_k(L))^\prime]^\prime \); \( \phi^{(s)} = [(\phi^{(s)}_1)^\prime, \ldots, (\phi^{(s)}_P)^\prime]^\prime \); \( \xi^{(s)}_{pq} = [\xi^{(s)}_{pq}(1), \ldots, \xi^{(s)}_{pq}(L)] \); \( d^{(s)} = [(d^{(s)}_1)^\prime, \ldots, (d^{(s)}_P)^\prime]^\prime \); and \( \Omega^{(s)} = (\Sigma^{(s)})^{-1} \). At the subject level, I use a structure that takes
into account the groups, i.e., for $s = 1, \ldots, S$,

$$
\begin{align*}
\beta^{(s)} & \sim ind \ N(\mu_{\beta}^g, \Sigma_{\beta}^g), \quad d^{(s)} \sim ind \ N(\mu_{d}^g, \Sigma_d^g), \quad \phi^{(s)} \sim ind \ N(\mu_{\phi}^g, \Sigma_{\phi}^g), \\
\xi^{(s)}_{pq} & \sim ind \ Multinomial(1, \pi_{pq}^g), \quad \Omega^{(s)} \sim Wishart(\Omega^g, \nu_{\Omega}).
\end{align*}
$$

Here $Pr(\xi^{(s)}_{pq} = \ell_j) = \pi_{pq}^g(j)$, with $\ell_j = [1, \ldots, 1, 0, \ldots, 0]$, $j = 0, 1, \ldots, L$. The $j$ in $\xi^{(s)}_{pq} = \ell_j$ can be interpreted as the largest lag at which there is connectivity from ROI $q$ to $p$.

Therefore, all the subjects belonging to the same group (say $g$), will have the same group parameters $\mu_{\beta}^g, \Sigma_{\beta}^g, \mu_{d}^g, \Sigma_d^g, \mu_{\phi}^g, \Sigma_{\phi}^g, \pi_{pq}^g,$ and $\Omega^g$. These group level parameters can be interpreted as follows: (1) $\mu_{\beta}^g$ represents the mean BOLD amplitudes and intercepts for group $g$; (2) $\mu_{d}^g$ represents the mean HRF basis coefficients for group $g$; (3) $\nu_{\Omega} \cdot \Omega^g$ represents the mean precision matrix for the temporally uncorrelated noise; (4) $\pi_{pq}^g(\ell)$ represents the overall probability that connectivity from ROI $q$ to $p$ exists among the group up to $\ell$ lags (this is the probability that connectivity from ROI $q$ to $p$ is present).

For simplicity, $\nu_{\Omega}$ is taken as a fixed constant across groups, and $\Sigma_{\beta}^g, \Sigma_d^g$ and $\Sigma_{\phi}^g$ are assumed to be diagonal.

The prior distributions for the group parameters of any group $g$ are listed below:

$$
\begin{align*}
\mu_{\beta}^g & \sim ind \ N(\mu_{\beta,0}, \Sigma_{\beta,0}), \quad \Sigma_{\beta}^g(i,i)^{-1} \sim Gamma(a_{\beta}, b_{\beta}), \\
\mu_{d}^g & \sim ind \ N(\mu_{d,0}, \Sigma_{d,0}), \quad \Sigma_d^g(i,i)^{-1} \sim Gamma(a_d, b_d), \\
\mu_{\phi}^g & \sim ind \ N(\mu_{\phi,0}, \Sigma_{\phi,0}), \quad \Sigma_{\phi}^g(i,i)^{-1} \sim Gamma(a_{\phi}, b_{\phi}), \\
\pi_{pq}^g & \sim ind \ Dir(\alpha_{\pi}), \\
p(\Sigma^g) & = p((\Omega^g)^{-1}) \propto |\Sigma^g|^{-\frac{1+P}{2}}.
\end{align*}
$$
Figure 3.1: A preliminary analysis of HRFs. (a) HRFs of all healthy subjects at an ROI. (b) HRFs of all stroke subjects at the same ROI. The black curve is canonical HRF.

\( \mu_{\mu\beta,0}, \Sigma_{\mu\beta,0}, \mu_{\mu\phi,0}, \Sigma_{\mu\phi,0}, a_\beta, b_\beta, a_\phi, b_\phi, a_\delta, b_\delta, \alpha_\pi, \Sigma_0 \) and \( \nu_0 \) are pre-determined constants. The choices of these constants will be discussed in Section 3.3 and 3.4. When there are several groups, another level can be added to the hierarchy by imposing additional priors on the group parameters for activation, i.e., \( \mu_{\mu\beta,0}, \Sigma_{\mu\beta,0}, a_\beta \) and \( b_\beta \), and similarly on the group parameters for connectivity and HRF.

The Bayesian hierarchical model is now fully specified. The quantities of interest are as follows. \textbf{Activation} for ROI \( p \): \( \beta_{p1}^{(s)} - \beta_{p2}^{(s)} \) is the activation strength for subject \( s \), and \( \mu_{\beta,p1}^g - \mu_{\beta,p2}^g \) is the mean activation strength for the group. Here \( \mu_{\beta,pk}^g \) denotes the element in \( \mu_{\beta}^g \) corresponding to \( \beta_{pk}^{(s)} \). \textbf{Connectivity} from ROI \( p \) to ROI \( q \) under condition \( k \) of lag \( \ell \): \( \Pr(\xi_{qp}^{(s)}(\ell) = 1) \) is the probability that such connectivity is present for subject \( s \); \( \Phi_k^{(s)}(\ell)_{qp}, \ell = 1, \ldots, L \) contain the connectivity measures for subject \( s \); \( \pi_{qp}^g(\ell) \) is the overall probability that connectivities are present at all \( \ell \) lags for the group, and \( \mu_{\Phi_k(\ell),qp}^g \) is the mean connectivity for the group. \textbf{HRF} for ROI \( p \): \( h_p^{(s)} = Hd_p^{(s)} \) is the HRF of ROI \( p \) for subject \( s \) and \( \mu_{h_p}^g = H\mu_{d_p}^g \) is the HRF of ROI \( p \) for group \( g \).

In the particular case of the stroke study, based on some preliminary analyses (see Section 3.2.4), it is found that there is a large within-group variability in the shape of the HRFs across subjects (see Figure 3.1). Averaging curves of very different shapes within a group is likely to produce meaningless overall group-specific shapes due to the non-linearity involved. Hence, it is prudent to not pool the highly heterogeneous shapes of the HRFs across subjects.
Therefore, specifically for the analysis of the stroke data, I fix $\mu_d^{\text{gs}}$ as a constant. However, information on activation and connectivity is still combined across subjects.

### 3.2.3 Constrained linear basis for the HRFs

The method in [62] obtained the constrained linear basis sets through applying PCA on a large sample of reasonably shaped HRFs randomly generated by a non-linear parameterization. The sample of HRFs is used to approximate the space of reasonable HRF shapes. The method then obtains the “constraints” for the basis coefficients through a Gaussian approximation of the calculated basis coefficients of the generated HRFs. Such constrained linear basis representation is flexible in capturing time to peak, width of peak, time to undershoot, and also penalizes shapes that deviate from the space of reasonable shapes.

Here I briefly describe how to obtain the constrained linear basis for the HRFs. Following the general framework of [62], one can simulate $n$ discrete HRF vectors based on the half-cosine parameterization. The parameters involved in the parameterization are resolution $\delta$, duration parameters $h_1, h_2, h_3, h_4$, and depth parameters $f_1$ and $f_2$ (see Figure B.1(a) in Appendix B). Using principal component analysis (PCA) one can obtain $J$ linear basis vectors for the HRFs, and the corresponding basis coefficients (loadings of the top $J$ components), denoted as $d_i$, for the $i$-th simulated HRF, $i = 1, 2, ..., n$. For the simulation study and the fMRI data analysis I used $n = 1000$, $\delta = 0.1$, $h_1 \sim \text{Unif}(0, 2)$, $h_2 \sim \text{Unif}(2, 7)$, $h_3 \sim \text{Unif}(2, 8)$, $h_4 \sim \text{Unif}(2, 12)$, $f_1 = 0$, $f_2 \sim \text{Unif}(0, 0.5)$, and $J = 5$. The top $J = 5$ principal components for the HRF basis set, i.e., the $J = 5$ columns of the matrix $H$, explain about 99% of the total variability in the simulated HRFs (see Figure B.2(a) in Appendix B). Note also that the first HRF basis vector (red) looks very similar to the canonical HRF (see Figure B.2(b)).

As a next step, I compute the empirical mean $\mu_d^* \in \mathbb{R}^J$ and variance $\Sigma_d^* \in \mathbb{R}^{J \times J}$ of the basis coefficients $\{d_i\}_{i=1}^n$. I then use a Gaussian distribution, $d^{(s)} \sim N(\mu_d^*, \Sigma_d^*)$, as the prior
distribution for the HRF basis coefficients. This prior distribution provides enough flexibility in the HRF shapes while simultaneously constraining them to a reasonable range. Note that the prior beliefs about the HRFs are adopted through the choice of the duration parameters and the depth parameters. The normal distribution just provides a convenient approximation of such prior beliefs. In order to determine if the prior structure is reasonable, visual checks of the quality of the basis set can be performed by generating prior samples of the HRFs using the basis vectors and corresponding coefficients randomly sampled from the normal distribution described above. Figure B.2(c) displays examples of randomly generated HRFs from the HRF basis set used in the analysis. The HRFs sampled from the prior indicate that HRF basis representation is sufficiently flexible in capturing time to peak, width of peak, time to undershoot, and lead to reasonable HRF shapes.

3.2.4 Preliminary analysis

Preliminary analyses can be helpful to obtain decent initial values for the MCMC algorithm in order to reduce burn-in time. The preliminary analysis in this work is done as follows. The noise term \( u^{(s,r)}(t) \) is set to be i.i.d. across time and sessions, and independent across ROIs. The model degenerates to

\[
\begin{align*}
    y_p^{(s,r)}(t) &= \beta_p^{(s)} + \{H \ast c_1\}(t)d_p^{(s)}\beta_{p1}^{(s)} + \{H \ast c_2\}(t)d_p^{(s)}\beta_{p2}^{(s)} + \epsilon_p^{(s,r)}(t), \\
    \epsilon_p^{(s,r)}(t) &\overset{i.i.d.}{\sim} N(0, \Sigma_{pp})
\end{align*}
\]

For the unknown group parameters, set \( \Sigma_\phi(i, i) \) and \( \Sigma_\beta(i, i) \) to relatively large values, and fix \( \Omega^y = I \). Posterior modes for \( \beta_{pk}^{(s)}, k = 0, 1, 2, d_p^{(s)} \) and \( \Sigma^{(s)} \) can then be computed separately for each subject, via conditional maximization, and be used as their preliminary estimates.
3.2.5 Posterior inference

The hierarchical structure of the model described in the previous section is such that information is pooled across subjects within a group and not across groups. This combined with the fact that the group-specific hyperparameters are fixed and known simplifies the inference, as the full posterior distribution can be written as a product of group-specific components. This implies that sampling from the joint posterior distribution of all the parameters is equivalent to separately sampling from the group-specific joint posterior distributions. Therefore, I only discuss single group estimation and drop the group indicator in this section for convenience.

The subject level model can essentially be decomposed into three multivariate linear regression (MLR) components when conditioning on all the remaining components. I now describe each component. Define \( y_{s,r}^{(s,r)}(t) = [y_{1}^{(s,r)}(t), \ldots, y_{p}^{(s,r)}(t)]' \), \( x_{p}^{(s)}(t) = [1, x_{p1}^{(s)}, x_{p2}^{(s)}] \), and \( X^{(s)}(t) = \text{Bdiag}\{x_{1}^{(s)}(t), \ldots, x_{p}^{(s)}(t)\} \). Also define \( y_{0}^{(s,r)}(t) = y^{(s,r)}(t) - [\beta_{10}^{(s)}, \ldots, \beta_{p0}^{(s)}]' \), and \( Z^{(s)}(t) = \text{Bdiag}\{A_{1}^{(s)}, \ldots, A_{1}^{(s)}\beta_{11}^{(s)} + A_{2}^{(s)}\beta_{12}^{(s)} + \ldots + A_{1}^{(s)}\beta_{p1}^{(s)} + A_{2}^{(s)}\beta_{p2}^{(s)}\} \). Then,

\[
\begin{align*}
    y^{(s,r)}(t) &= X^{(s)}(t)\beta^{(s)} + u^{(s,r)}(t) \\
    y_{0}^{(s,r)}(t) &= Z^{(s)}(t)d^{(s)} + u^{(s,r)}(t).
\end{align*}
\] (3.9)

Note that the noise terms \( u^{(s,r)}(t) \) are correlated over time due to the VAR structure, and therefore an extra step is needed to calculate the full conditionals: “whitening” the data. Define \( \Phi^{(s)}(B) = I_{p} - \sum_{\ell=1}^{L} \Phi^{(s)}(\ell, t) B^{\ell} \), where \( B \) is the backshift operator, and

\[
\begin{align*}
    \tilde{y}^{(s,r)}(t) &= \Phi^{(s)}(B)[y^{(s,r)}(t)] = y^{(s,r)}(t) - \sum_{\ell=1}^{L} \Phi^{(s)}(\ell, t)y^{(s,r)}(t - \ell), \\
    \tilde{y}_{0}^{(s,r)}(t) &= \Phi^{(s)}(B)[y_{0}^{(s,r)}(t)], \\
    \tilde{X}^{(s)}(t) &= \Phi^{(s)}(B)[X^{(s)}(t)], \\
    \tilde{Z}^{(s)}(t) &= \Phi^{(s)}(B)[Z^{(s)}(t)].
\end{align*}
\] (3.10-3.13)
Then, the equations in (3.9) can be written in terms of the temporally uncorrelated noise terms, leading to the following two MLR components, one for the BOLD amplitude parameters $\beta^{(s)}$ and another for the HRF parameters $d^{(s)}$:

\[
\tilde{y}^{(s,r)}(t) = \tilde{X}^{(s)}(t)\beta^{(s)} + \epsilon^{(s,r)}(t), \tag{3.14}
\]
\[
\tilde{y}_0^{(s,r)}(t) = \tilde{Z}^{(s)}(t)d^{(s)} + \epsilon^{(s,r)}(t). \tag{3.15}
\]

For computational efficiency, I work with the likelihood that is conditional on the initial $L$ observations in each session, instead of the full likelihood.

The third MLR component can be written in terms of the connectivity measures $\phi$. Using the conditional likelihood and conditioning on all other parameters, VAR estimation simplifies to an MLR problem: $u^{(s)}(t) = y^{(s,r)}(t) - X^{(s)}(t)\beta^{(s)}$ can be calculated and treated as the response and the covariates (with proper lags) in Equation (3.3). More specifically, define

\[
u_k^{(s,r)}(t) = [c_k(t-1)u^{(s,r)}(t-1)' \otimes I_P, \ldots, c_k(t-L)u^{(s,r)}(t-L)' \otimes I_P], \tag{3.16}
\]
\[
W^{(s,r)}(t) = [w_1^{(s,r)}(t), w_2^{(s,r)}(t)], \quad \tilde{W}^{(s,r)}(t)_{p,i} = W^{(s,r)}(t)_{p,i}\xi_{pq}^{(s)}(\ell), \tag{3.17}
\]
\[
\tilde{\phi}^{(s)}(i) = \phi^{(s)}(i)\xi_{pq}^{(s)}(\ell), \text{ with } \xi_{pq}^{(s)} = \ell_j, \quad \forall p,q,\ell,k, \text{ and } i = J(p,q,\ell,k) := p + (q-1)P + (\ell-1)P^2 + (k-1)P^2L.
\]

Then, the third MLR can be expressed as:

\[
u^{(s,r)}(t) = \tilde{W}^{(s,r)}(t)\tilde{\phi}^{(s)} + \epsilon^{(s,r)}(t). \tag{3.18}
\]
A MCMC algorithm was implemented to obtain samples from the joint posterior distribution of the model parameters. The fact that I have three MLR components simplifies the steps in the MCMC algorithm. In particular, combining Equations (3.14), (3.15) and (3.18) above with the Gaussian distributions for $\beta^{(s)}$, $d^{(s)}$ and $\phi^{(s)}$, I have that the full conditional distributions for these parameters are also Gaussian, resulting in Gibbs steps. Similarly, the full conditional distributions for $\mu^{(g)}_{\beta}$, $\mu^{(g)}_{d}$ and $\mu^{(g)}_{\phi}$ are also Gaussian. The MCMC algorithm also requires steps for sampling the covariance matrices, the indicators and the probabilities. The diagonal entries of the covariance matrices, $\Sigma_{\beta}(i, i)^{(g)}$, $\Sigma_{\phi}(i, i)^{(g)}$ and $\Sigma_{d}(i, i)^{(g)}$, are sampled from inverse Gamma distributions; the noise covariance matrices $\Sigma^{(s)}$, $\Sigma^{g}$ are sampled from inverse Wishart distributions; sampling for the indicators $\xi^{(s)}_{qp}$ follows [32] for stochastic search variable selection; the probabilities $\pi^{(g)}_{qp}$ are sampled from Dirichlet distributions. The choice of conjugate priors for the model parameters largely simplifies the structure of the MCMC algorithm for posterior inference. The full conditional posterior distributions and all the steps of the MCMC algorithm are detailed in Appendix B.2 and B.3.

The point estimates for the quantities of interest are based on medians of the MCMC posterior samples. Inference for each quantity of interest is as follows (credible level = $(1 - \alpha) \times 100\%$):

**Activation.** I calculate a $(1 - \alpha) \times 100\%$ posterior credible interval, referred to as CI, for the group level contrast $\mu^{g}_{\beta, p,(c)} = \mu^{g}_{\beta, p1} - \mu^{g}_{\beta, p2}$ of each ROI $p$. If the interval is entirely on the right side of zero, it indicates that ROI $p$ is activated under the task condition for group $g$. For each specific subject, I calculate the $(1 - \alpha) \times 100\%$ posterior CI of the subject level contrast $\beta^{(s)}_{p,(c)}$. If the interval is entirely on the right side of zero, it indicates that ROI $p$ is activated for subject $s$. In the analysis I calculated the $(1 - \alpha) \times 100\%$ CI as the interval between the $\alpha/2$-th and $(1 - \alpha/2)$-th quantiles obtained using the MCMC posterior samples.

**Connectivity.** For connectivity from ROI $p$ to $q$ at the subject level, I estimate the posterior mode for $\xi^{(s)}_{qp}$. Suppose $\ell_{j}$ is the posterior mode. If $j = 0$, there is no connectivity from ROI
p to q for subject s; otherwise there is connectivity. Posterior medians and CIs of $\Phi^{(s)}_k(\ell)_{qp}$ are calculated to measure the connectivity strength at lag $\ell = 1, 2, \ldots, j$ if $j > 0$. At the group level, I calculate the posterior medians of $\pi^{\theta}_{qp}(j)$ as the overall presence probability of connectivity with $j$ lags among the group, for $j > 0$. $j = \arg\max_{\ell=0,\ldots,L} \pi^{\theta}_{qp}(j)$ is the estimated lag for connectivity from p to q, at the group level. Further examination of the corresponding posterior CIs of $\mu^{\theta}_{\Phi_k(\ell),qp}$, $\ell = 1, \ldots, j$, provides inference on the direction and strength of the connectivity for each condition $k$.

**HRFs.** I derive the posterior samples of the HRFs by calculating $Hd^{(s)}_p$ for each posterior sample of $d^{(s)}_p$; point-wise credible bands of the HRFs can hence be calculated. Group-level HRF estimates and credible bands can also be obtained similarly from $H\mu^{\theta}_{d_p}$.

The algorithm is implemented in R. Since there are a lot of parameters involved in the model, the computation is heavy. In order to increase the computation efficiency, I implement the sampling for the subject parameters in C++ core using R package RcppArmadillo [16]. I also implement parallel computing so that the subject-level parameters are sampled simultaneously for all subjects. Furthermore, the spike and slab prior causes high sampling autocorrelation for the subject parameters. Therefore, when I update the subject parameters I update it for 4 additional iterations and only store the last one in the MCMC sample, i.e., thinning for only the subject parameters. By doing so, I efficiently reduce the autocorrelation and thus increase the effective sample size without having to increase the iterations for the entire set of parameters. The implementation is available on https://github.com/soapless/Hierarchical_BVAR_fMRI_multisub.

For a dataset that contains many more regions of interest, the current algorithm could become much slower. It is mainly due to (1) the large matrix multiplications and inversions involved the estimation of the connectivity parameters $\Phi^{(s)}$, and (2) the evaluations for each of the probabilities $\Pr(\xi^{(s)}_{pq} = \ell_j)$ ($j = 0, \ldots, L; \ p, q = 1, \ldots, P$) in the spike and slab structure. It is possible to further speed up the computation after the burn-in period by splitting the
posterior sample into samples with smaller sizes and running parallel chains to obtain these smaller samples. Computation time can also be reduced by imposing special structures on the sparsity (e.g., row-wise sparsity) instead of element-wise sparsity for the spike and slab prior, at the cost of the flexibility on the sparsity of the connectivity. Another way to speed up the estimation is to use the approximate but much faster algorithms such as variational Bayes (VB), which I briefly discuss in the next subsection.

### 3.2.6 An alternative implementation by variational Bayes

I first discuss the VB algorithm for a single subject $s$ with group parameters treated as fixed prior constants. For simplicity the superscript for subject $s$ will be dropped unless otherwise stated. The log posterior density is $\log p(\Theta|Y) = \text{const} + \log L(\beta, d, \phi, \Omega) + \log p(\beta) + \log p(d) + \log p(\phi) + \sum_{p,q} \log p(\xi_{pq}) + \log p(\Omega)$. The approximation function I use is of the form $q(\Theta) = q(\beta)q(d)q(\phi)\prod_{p,q} q(\xi_{pq})q(\Omega)$. In order to obtain $\log q(\Theta_i)$, the approximate log posterior density for the $i$-th parameter set, that minimizes the KL divergence $KL(q||p)$, I need to calculate $\mathbb{E}_{q}(\Theta - i) \log p(\Theta)$ [20]. Let notation $\|x\|_A^2$ denote $x' A^{-1} x$. It is easy to see that $q(\Theta_i)$’s are all conjugate: when $\Theta_i = \beta, d$ or $\phi$, the approximation $\log q(\Theta_i)$ can be written as $\text{const} - \frac{1}{2} \| \Theta_i - \tilde{\mu}_{\Theta_i} \|_{\tilde{\Sigma}_{\Theta_i}}^2$, following the Gaussian kernel for $N(\tilde{\mu}_{\Theta_i}, \tilde{\Sigma}_{\Theta_i})$; when $\Theta_i = \xi_{pq}$, $\log q(\Theta_i) = \text{const} + \sum_{j=0}^L I(\Theta_i = \ell_j) \log \tilde{\pi}_{j,\Theta_i}$ (s.t. $\sum_{j=0}^L \tilde{\pi}_{j,\Theta_i} = 1$), following the multinomial kernel for $\text{Multinomial}(1, \{\tilde{\pi}_{j,\Theta_i}\}_{j=0}^L)$; and when $\Theta_i = \Omega$, $\log q(\Theta_i) = (\tilde{\nu}_{\Theta_j} - P - 1) \log |\Theta_i| - tr((\tilde{\mu}_{\Theta_j}/\tilde{\nu}_{\Theta_j})^{-1} \Theta_i)$ following the Wishart kernel for $\text{Wishart}(\tilde{\nu}_{\Theta_j}, \tilde{\Psi}_{\Theta_j})$.

Let $\tilde{\mu}_{\xi_{pq}}, \tilde{\Sigma}_{\xi_{pq}}$ denote the mean and the covariance matrix for $\xi_{pq} = [\xi_{pq}(1), \ldots, \xi_{pq}(L)]$ under $q(\cdot)$, and let $\tilde{\pi}_{\xi_{pq}} = [\tilde{\pi}_{0,\xi_{pq}}, \ldots, \tilde{\pi}_{L,\xi_{pq}}]$. I further simplify the calculations by approximating $\mathbb{E}_q \phi \phi'$, $\mathbb{E}_q \xi_{pq} \phi_q$, $\mathbb{E}_q \beta \beta'$, $\mathbb{E}_q d d'$ using $\mathbb{E}_q \phi \mathbb{E}_q \phi'$, $\mathbb{E}_q \xi_{pq} \mathbb{E}_q \xi_{pq}'$, $\mathbb{E}_q \beta \mathbb{E}_q \beta'$, $\mathbb{E}_q d \mathbb{E}_q d'$, i.e., dropping the covariance terms $\tilde{\Sigma}_\beta, \tilde{\Sigma}_d, \tilde{\Sigma}_\phi, \tilde{\Sigma}_{\xi_{pq}}$. Except that for $\phi, \xi_{pq}$ and $\Omega$, I will keep $\tilde{\Sigma}_\phi$ and $\tilde{\Sigma}_{\xi_{pq}}$ in order to get more accurate results for the connectivity selection. Now the calculations
simplify to the following form:

\[
\begin{align*}
\tilde{\mu}_\beta &= f_{\tilde{\mu}_\beta}(\tilde{\mu}_d, \tilde{\mu}_\phi, \tilde{\mu}_{\xi, pq}; \tilde{\mu}_\Omega; p, q = 1, \ldots, P), \\
\tilde{\Sigma}_\beta &= f_{\tilde{\Sigma}_\beta}(\tilde{\mu}_d, \tilde{\mu}_\phi, \tilde{\mu}_{\xi, pq}; \tilde{\mu}_\Omega; p, q = 1, \ldots, P), \\
\tilde{\mu}_d &= f_{\tilde{\mu}_d}(\tilde{\mu}_\beta, \tilde{\mu}_\phi, \tilde{\mu}_{\xi, pq}; \tilde{\mu}_\Omega; p, q = 1, \ldots, P), \\
\tilde{\Sigma}_d &= f_{\tilde{\Sigma}_d}(\tilde{\mu}_\beta, \tilde{\mu}_\phi, \tilde{\mu}_{\xi, pq}; \tilde{\mu}_\Omega; p, q = 1, \ldots, P), \\
\tilde{\mu}_\phi &= f_{\tilde{\mu}_\phi}(\tilde{\mu}_{\xi, pq}, \tilde{\Sigma}_{\xi, pq}; \tilde{\mu}_d, \tilde{\mu}_\beta, \tilde{\mu}_\Omega; p, q = 1, \ldots, P), \\
\tilde{\Sigma}_\phi &= f_{\tilde{\Sigma}_\phi}(\tilde{\mu}_d, \tilde{\mu}_\beta, \tilde{\mu}_{\xi, pq}, \tilde{\Sigma}_{\xi, pq}; \tilde{\mu}_\Omega; p, q = 1, \ldots, P), \\
\tilde{\pi}_{\xi, pq} &= f_{\tilde{\pi}_{\xi, pq}}(\tilde{\mu}_\phi, \tilde{\Sigma}_\phi, \tilde{\mu}_{\xi, q’ q’}; \tilde{\mu}_d, \tilde{\Sigma}_d, \tilde{\mu}_\beta, \tilde{\Sigma}_\beta; \tilde{\mu}_\Omega; (p’, q’) \neq (p, q)), \\
\tilde{\mu}_\Omega &= f_{\tilde{\mu}_\Omega}(\tilde{\mu}_\beta, \tilde{\mu}_d; \tilde{\mu}_\phi, \tilde{\Sigma}_\phi, \tilde{\mu}_{\xi, pq} \tilde{\Sigma}_{\xi, pq}; \tilde{\mu}_\Omega; p, q = 1, \ldots, P), \\
\tilde{\nu}_\Omega &= \nu_\Omega + (T - L)R.
\end{align*}
\]

The VB algorithm will work by first specifying the initial values for \(\tilde{\mu}_\beta, \tilde{\Sigma}_\beta, \tilde{\mu}_d, \tilde{\Sigma}_d, \tilde{\mu}_\phi, \tilde{\Sigma}_\phi, \tilde{\pi}_{\xi, pq}, \tilde{\mu}_\Omega\) and \(\tilde{\nu}_\Omega\), and then updating their values according to Equations (3.19)-(3.26) repeatedly until convergence. Details of the VB algorithm is provided in Appendix B.5. The posterior distributions can be approximated by \(\beta \sim N(\bar{\mu}_\beta, \bar{\Sigma}_\beta), \ d \sim N(\bar{\mu}_d, \bar{\Sigma}_d), \ \phi \sim N(\bar{\mu}_\phi, \bar{\Sigma}_\phi), \ \xi_{p,q} \sim Multinomial(1, \bar{\pi}_{\xi, pq}), \) and \(\Omega \sim Wishart(\bar{\nu}_\Omega, \bar{\Psi}_\Omega).\)

The VB algorithm mentioned above could be easily extended to the full multi-subject model. The approximation function becomes \(q(\Theta) = \prod_{s=1}^{S} q(\Theta^{(s)}) \cdot q(\mu_\beta^{(g)}) q(\mu_\phi^{(g)}) q(\pi_{\xi, pq}^{(g)}) q(\Sigma_\beta^{(g)}) q(\Sigma_\phi^{(g)}) q(\Sigma_{\xi, pq}) q(\Sigma_{\xi, pq}) \cdot q(\Sigma^{(g)}) q(\Sigma^{(g)})),\) where \(q(\Theta^{(s)})\) is the same \(q(\cdot)\) for the single subject model before and \(\Theta^{(s)}\) is the set of the subject-level parameters for subject \(s\). The calculations for \(q(\Theta^{(s)})\) will be similar to Equations (3.19)-(3.26), except that the fixed group parameters are replaced with their expectations under \(q(\cdot)\). The calculations for \(q(\cdot)\) of the group parameters are also conjugate and standard. I have that \(\mu_\beta^{(g)}, \mu_\phi^{(g)} \sim Gaussian\) distribution, \(\pi_{\xi, pq}^{(g)} \sim Dirichlet\) distribution, \(\Sigma_\beta^{(g)}(i, i), \Sigma_\phi^{(g)}(i, i) \sim Gamma\) distribution, and \(\Sigma^{(g)} \sim Wishart\) distribution.
Figure 3.2: Simulation settings. (a) Simulated mean connectivity network (numbers in parentheses represent the connectivity measure under rest condition; connectivities not shown in the graph are not present). (b) Simulated mean HRFs.

3.3 Simulation study

In this section, I show the performance of the proposed models in a simulation study, and compare the performance of VB and MCMC.

3.3.1 Simulation settings

I simulated 30 datasets of a single group of 30 subjects, 2 conditions, 5 ROIs, 3 sessions with 48 time points for each independent session, and TR = 2 seconds. The dataset for each simulation was independently generated according to the model described in the previous section with \( L = 1 \). The BOLD amplitudes, connectivity parameters and the HRFs for each subject were generated independently using multivariate normal distributions. ROI 2 was simulated to be not activated. The BOLD amplitudes for the two conditions in this ROI were the same, and therefore had a low signal to noise ratio due to a flat BOLD response shape. The connectivity network and connectivity strength implied by the mean connectivity parameter \( \mu_\phi \) is summarized in Figure 3.2(a). I also set the connectivities that are not present in Figure 3.2(a) to be zero for all the subjects. This setting assumes that the presence/absence of connectivity is the same across all subjects. The mean HRFs for
the five ROIs are shown in Figure 3.2(b). ROIs 3 and 4 have the same HRF shapes. $\Sigma^{(s)}$ was constrained to be the same across the subjects. The exact values of the simulation parameters are listed in Appendix B.4. Selected simulated time series for ROIs 1, 2, and 3 are shown in Figure 3.3.

3.3.2 Analysis

Here I discuss the prior distributions for the group parameters used in the analysis of the simulated data. $\nu_{\Omega}$, the prior constant in $\Omega^{(s)} \sim \text{Wishart}(\Omega^{gs}, \nu_{\Omega})$ is set to $P + 1$, the smallest value that gives a proper Wishart distribution. $d^{(s)} \overset{\text{ind}}{\sim} N(\mu_{d}^{*}, \Sigma_{d}^{*})$, which comes from the constrained linear basis described in Section 3.2.3 in order to help to produce HRF estimates that have reasonable shapes. $\mu_{g}^{(s)} \overset{\text{ind}}{\sim} N(0, 750 \cdot I_{3P})$, which is chosen to match the scale of the data and more specifically, the variance, is set to be a little larger than the maximum of all $\mu_{g}^{2}$. $\mu_{g}^{(s)} \overset{\text{ind}}{\sim} N(0, 750 \cdot I_{3P})$, as $\mu_{g}$ is not expected to be much larger than 1. $\Sigma_{g}(i, i)^{-1} \overset{\text{ind}}{\sim} \text{Gamma}(1, 1)$, and $\Sigma_{g}(i, i)^{-1} \overset{\text{ind}}{\sim} \text{Gamma}(1, 1)$; they are chosen so that $\Sigma_{g}(i, i)$ and $\Sigma_{g}(i, i)$ are allowed to have large variability. $\pi_{q}^{(s)} \overset{\text{ind}}{\sim} \text{Dir}(0.1 \cdot 1_{L+1})$, so that $\pi_{q}^{(s)}$ has a quite large variability but does not favor in any number of lags $j \in \{0, 1, 2, ..., L\}$. The maximum number of lags allowed in the model was set to $L = 1$.

A MCMC sample of size 5,000 was obtained, after a burn-in period of 3,000 iterations. No
convergence problems were detected. Posterior summaries were computed from the MCMC samples. To save storage space and also to reduce autocorrelation of the samples, I only stored every 5th iteration, and the final MCMC sample of each parameter has a size of 5,000.

### 3.3.3 Results

**Activation.** All the ROIs were correctly identified as activated or not activated for all 30 datasets. The average bias in the group-level activation contrast is $\mu_{\beta, p, (c)}$, is 0.48; the average relative bias (i.e., the bias divided by the true value) is 2.9% for activated regions, and specifically $[0.079, 0.041, 0.033, 0.044] \times 100\%$ for ROIs 1, 3, 4, and 5. The coverage of the 95% posterior CI for the contrast is 86.7% on average, and $[1.00, 1.00, 0.80, 0.67, 0.87] \times 100\%$ for ROI 1-5. While ROI 3 has relatively low coverage, it also has relatively small bias compared to other ROIs, which suggests that the variability of the activation parameter at this ROI could be underestimated.

**Connectivity.** At the subject level, the connectivity network was correctly inferred for 96.7% individuals across the datasets. Even for connectivity from ROI 5 to ROI 3, which has non-zero $\phi$ for the task condition and zero $\phi$ for the rest condition, the presence of connectivity was also correctly inferred. The overall sensitivity for the presence of connectivity is 99.5%, and the overall specificity is 100%. These results are based on a 0.5 threshold, i.e., using $\Pr(\xi_{qp}^{(s)} = 1) > 0.5$ to determine the presence of connectivity. Other thresholds ranging from 0.05 to 0.95 were also applied, and the results were similar. At the group level, the bias in connectivity presence probability $\pi_{qp}^{g}(1)$ was 0.0019, averaged across all $1 \leq p, q \leq P$. Specifically, this number is 0.0080 for connectivities that are present, and $2 \times 10^{-5}$ for connectivities that are absent. For the connectivities that exist in the simulated network, the average coverage of the 95% posterior CI of the corresponding $\phi$ parameters is 100%,
with average bias around $4.69 \times 10^{-5}$, about $1.2\%$ of the average non-zero $\mu_\phi$.

**HRFs.** The coverage of point-wise 95% posterior credible bands for subject and ROI specific HRFs is around 94.5% overall, and $[0.911, 1.000, 0.958, 0.935, 0.922] \times 100\%$ on average for each ROI. Point-wise 95% posterior CIs of average HRFs at ROI 1, 2, and 5 for a specific dataset are shown in Figure 3.4. As expected, ROI 2, the non-activated ROI, shows a larger variability in HRF estimates than the other ROIs, due to low signal to noise ratio.

Overall, the proposed approach is able to correctly infer activation, connectivity and the HRF shapes. It also has relatively small biases and good coverages of the 95% posterior CIs in general, except that the variation of the activation parameters is sometimes a little underestimated. Moreover, it is very robust to the choice of threshold for determining whether the connectivity is present or not at the subject level.

### 3.3.4 Comparison between VB and MCMC

I ran a simulation study and compared VB with MCMC for a simple case. I simulated the fMRI data for 15 subjects, using the same simulation settings described in Section 3.3.1. The prior distributions (the group parameters) used in the simulation are $\beta \sim N(0, 50^2 I)$, $d \sim N(\mu_d, \Sigma_d)$, $\phi \sim N(0, 0.5^2 I)$, $\pi_{pq} = [.5, .5]$, $\nu_\Omega = P + 1$, $\Omega_\mu = I$. I applied both VB and MCMC on each of the 15 subjects. For MCMC, I used burn-in period of 5,000 iterations, and obtained a posterior sample of size 1,000, thinned at every 5th iteration. VB algorithm converged within 150 iterations for all subjects, and took 2 seconds; the MCMC algorithm took 104 seconds.

The results in terms of bias, relative bias ("rbias"), 95% CI coverage ("cover"), true positive rate ("tp") and true negative rate ("tn") are summarized in Table 3.1. Overall, VB provides reasonable approximation to the posterior distributions, especially in terms of bias. It only
Figure 3.4: Posterior HRF results for ROIs 1, 2, and 5 in a given subject. Gray curves correspond to posterior samples of the HRFs, blue curves are the bounds of the 95% point-wise posterior credible bands, the red curves are the true HRFs, the black curves are the posterior median HRFs.

has a little bigger bias than MCMC for most of the parameters, and even a little smaller bias for FWHM $T_w$ and time to post-stimulus undershoot $T_u$. Although in terms of 95% CI, VB seems to yield intervals with smaller coverage than MCMC. Also, for inferring whether a connectivity is present or whether a region is activated, VB provides a little smaller true negative rate, i.e., VB seems to report more false positives.

Overall, VB approximations could greatly reduce the computation time, at the cost of lower specificity and smaller coverage of the 95% CIs. In the case where false positive is of less concern and MCMC sampling is too slow, VB method would be a very good alternative.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$.bias</th>
<th>$\beta$.rbias</th>
<th>$\beta$.tp</th>
<th>$\beta$.tn</th>
<th>$\beta$.cover</th>
<th>$\Omega$.bias</th>
<th>$\Omega$.rbias</th>
<th>$\Omega$.cover</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB</td>
<td>-0.011</td>
<td>-0.0344</td>
<td>0.856</td>
<td>0.828</td>
<td>0.972</td>
<td>0.004</td>
<td>0.106</td>
<td>0.884</td>
</tr>
<tr>
<td>MCMC</td>
<td>-0.006</td>
<td>-0.0183</td>
<td>0.889</td>
<td>0.898</td>
<td>0.961</td>
<td>0.002</td>
<td>0.062</td>
<td>0.929</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>hrf.cover</th>
<th>$T_p$.bias</th>
<th>$T_p$.cover</th>
<th>$T_w$.bias</th>
<th>$T_w$.cover</th>
<th>$T_u$.bias</th>
<th>$T_u$.cover</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB</td>
<td>0.876</td>
<td>-0.156</td>
<td>0.907</td>
<td>-0.239</td>
<td>0.893</td>
<td>-1.095</td>
<td>0.893</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.967</td>
<td>-0.053</td>
<td>0.987</td>
<td>-0.403</td>
<td>0.947</td>
<td>-1.165</td>
<td>0.973</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\phi$.bias</th>
<th>$\phi$.rbias</th>
<th>$\phi$.tp</th>
<th>$\phi$.tn</th>
<th>$\phi$.cover</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB</td>
<td>0.191</td>
<td>0.0156</td>
<td>0.983</td>
<td>0.867</td>
<td>0.773</td>
</tr>
<tr>
<td>MCMC</td>
<td>-0.010</td>
<td>-0.0012</td>
<td>0.983</td>
<td>1.000</td>
<td>0.960</td>
</tr>
</tbody>
</table>

Table 3.1: Comparison of results of VB and MCMC. $T_p$, $T_w$ and $T_u$ denote time to peak, FWHM, and time to post-stimulus undershoot (in seconds), respectively.
3.4 Application to fMRI data

In this section I applied the proposed model to analyze the fMRI data from the stroke study. Detailed information about the dataset was provided in Chapter 1. Here I only selected right-handed healthy subjects and right-handed stroke patients whose right side of body was affected (left brain hemisphere affected), so that it is safe to compare the two groups since they all used their right hand and also the dominant hand to perform the task. In the end, there are 12 healthy subjects and 15 stroke patients selected in the analysis. The data contains 144 time points (in 3 fMRI sessions) for each stroke patient and 96 time points (in 2 fMRI sessions) for each healthy subject. All the five ROIs, LM1, RM1, LPMd, RPMd, and SMA, are included in the analysis.

3.4.1 Analysis

I use \( g = (\text{healthy}), (\text{stroke}) \) to denote, respectively, the healthy group and the stroke group. Both \( L = 1 \) and \( L = 2 \) was used for the maximum lag. The results on connectivity network, activation and HRF shapes at the group level for \( L = 1 \) were the same as the those for \( L = 2 \) (see Appendix B.1). In addition, although for certain subjects \( L = 2 \) might still be needed, the estimated lag at the group level for connectivity from ROI \( p \) to \( q \) for each pair of \((p, q)\) was all no more than 1. The above results suggested \( L = 1 \) was sufficient for the purpose of group analysis for this dataset, besides its advantage of less computation time and less variation in the estimated results. Therefore, \( L = 1 \) was chosen over \( L = 2 \). In the rest of the section, only \( L = 1 \) will be discussed. Diagnostic plots were generated to further check the lag and other aspects of model fit, which will be discussed in the next subsection.

I used the same prior distributions used for the simulation study; except for activation strength I used \( \mu_{\beta}^g \sim N(0, 25 \cdot I) \), so that the standard deviation is about the largest \( \mu_{\beta}^g \).
obtained from the preliminary data analysis. In addition, in order to check the sensitivity of the results to the choice of prior distributions, the following alternative prior distributions were also considered. Sensitivity with respect to the choice of $\alpha \pi$: I considered two alternative prior settings with $\alpha \pi = [0.2, 0.1]$ and $\alpha \pi = [0.1, 0.2]$. The first one assumes that the presence probability of connectivity is mostly likely to be around $\frac{2}{3}$, the second assumes the probability is mostly likely to be around $\frac{1}{3}$; while the original prior, $\alpha \pi = [0.1, 0.1]$, assumes the probability is equally likely to be $<0.5$ or $>0.5$. Sensitivity with respect to the priors on $\mu^g_{\beta}$ and $\mu^g_{\phi}$: I considered two alternative sets of priors, namely (1) $\mu^g_{\beta} \sim N(0, 25 \times 25 I_{3 P})$, $\mu^g_{\phi} \sim N(0, 25 I_{2 L P^2})$, and (2) $\mu^g_{\beta} \sim N(0, \frac{1}{4} \times 25 I_{3 P})$, $\mu^g_{\phi} \sim N(0, \frac{1}{4} I_{2 L P^2})$. The first set has a larger variability compared to the original set of priors and the second set has a smaller variability. All of these priors led to very similar posterior results.

In addition to studying the posterior distributions of the activation and connectivity parameters within each group, I also obtained posterior inference on the difference in the presence probabilities of connectivity from $p$ to $q$ for stroke and healthy subjects, i.e., $\pi_{qp}^{(\text{stroke})}(\ell) - \pi_{qp}^{(\text{healthy})}(\ell)$, and also the difference in connectivity strength, $\mu_{\Phi_k(\ell), qp}^{(\text{stroke})} - \mu_{\Phi_k(\ell), qp}^{(\text{healthy})}$, for $\ell = 1$. The 95% posterior CIs of these quantities were computed. If the 95% posterior CI of the difference of a given quantity does not include 0, then it is claimed that there is sufficient evidence to claim that such quantity is different for the stroke group and the healthy group.

An posterior sample of size 5,000 was obtained, after a burn-in period of 3,000 iterations. The analysis was carried out on Linux platform in a computer with 64 x86-64 processors and 256G memory. The total computation time was 50 min.
Figure 3.5: Activation strength for the two groups and activation difference between the groups. Darker color indicates higher activation (or activation difference). Red indicates a positive value, and white represents zero. The black line outside the circle indicates that the corresponding 95% posterior CI does not include zero. The rectangular frame indicates the ROIs in the hemisphere corresponding to the moving hand.

3.4.2 Results

Activation. It was found that all the ROIs were activated in response to the task condition for the stroke patients, and all the ROIs were activated for the healthy subjects except for RM1 (see Figure 3.5(b)). In both groups, LM1, which is primarily responsible for the motor function of the right side of the body, was found to be highly activated. A formal comparison of the activations between the groups is shown in Figure 3.5(c). Red indicates a positive difference. The plot shows there is sufficient evidence that the stroke group is more activated than the healthy group in RPMd, RM1 and SMA. Interestingly, even for the healthy subjects, some ROIs in addition to LM1 are also activated. This indicates compensatory effect that requires M1 from the unaffected brain hemisphere and secondary motor regions to aid in executing the motor task.

Connectivity. Plots 3.6(a) and (b) list all the connectivities that are present in at least one subject for each group, and describe how likely is the presence of connectivity in each group in terms of the thickness of the arrow. Connectivity from LM1 to LPMd in healthy participants has the highest presence probability, around 1.00. Connectivity from RPMd to RM1 in stroke patients has the lowest presence probability, around 0.08, apart from the connectivities that are not listed. The solid arrows represent those connectivities that are predominant (i.e.,
Figure 3.6: The probabilities of the presence of connectivity for the two groups and the probability differences between the groups. A thicker arrow indicates a higher chance that connectivity is present in the group (plots (a) and (b)), or greater difference in the probability of connectivity between the two groups (plot (c)). In (a) and (b), those connectivities that are absent in all subjects in each group are not shown; a dashed arrow represents a connectivity whose presence probability does not exceed 0.5; the probabilities of the most likely and the least likely connectivities among the listed connectivities are also provided. In (c), red indicates that the connectivity is more likely to be present in the stroke group than the healthy group, and blue indicates that the connectivity is less likely to be present in the stroke group; only those differences whose 95% posterior CI does not contain 0 are included.

its presence probability exceeds 0.5: $\pi_{gqp}^{(1)} > 0.5$, while the dashed arrows correspond to connectivities that exist at least in one subject in the group (Pr($\xi_{qp}^{(s)}(1) = 1 | Y$) > 0.5 for at least one $s$) but are not predominant in the group ($\pi_{gqp}^{(1)} \leq 0.5$). All the listed connectivities have positive $\mu_{\phi_{k(1),qp}}^{g}$ values, indicating a positive lagged association: the larger the current signal at ROI $p$ is, the larger the future signal at ROI $q$ is expected to be. In both groups, connectivity within LM1 is present in the majority of the subjects. This suggests that there may be positive feedback within M1 on the hemisphere corresponding to the moving hand. In the healthy group, the predominant connectivities are mostly interregional and all start from LM1. However, in the stroke group, the predominant connectivities are all intraregional.

Figure 3.6(c) lists all the differences in connectivity presence probability between the two groups ($\pi_{gqp}^{(stroke)}(1) - \pi_{gqp}^{(healthy)}(1)$) whose 95% posterior CI does not include 0. The red connectivities are more likely in the stroke group than the healthy group, while the blue connectivities are more likely in the healthy group than the stroke group. For the stroke group, there is more connectivity from secondary motor regions (LPMd, SMA) to themselves.
Figure 3.7: Estimated HRF for each ROI and subject for the two groups. The vertical dashed line is at 5 sec and the horizontal dashed line is at the baseline value. The black curve is the median HRF of the corresponding group.

or to other regions; while for the healthy group, there is more connectivity from LM1, the region primarily responsible for the right hand movement, to other regions.

I also calculated the posterior samples of the difference in connectivity strength between the groups, \( \mu_{\Phi_{1,q}}^{(\text{stroke})} - \mu_{\Phi_{1,q}}^{(\text{healthy})} \). However, all of the 95% posterior CIs of the difference in connectivity strength cover 0; i.e., there is not sufficient evidence for differences in the strength of connectivities across different groups.

HRFs. The estimated HRF for each region and subject is shown in Figure 3.7. There do not seem to be obvious differences between the two groups, except that in general there seems to be more interindividual variability in the subject-specific HRFs from the stroke group.

I checked the goodness of fit by examining the estimated mean, denoted by \( \hat{y}^{(s,r)}(t) \), and the temporally uncorrelated noise, \( \epsilon^{(s,r)}(t) \). For each subject \( s \) and session \( r \), I obtained posterior samples of \( \epsilon^{(s,r)}(t) \) by calculating \( \epsilon^{(s,r)}(t) = \left[ X^{(s)}(t) \right]^{(m)} (\beta^{(s)})^{(m)} - \left[ X^{(s)}(t) \right]^{(m)} (\beta^{(s)})^{(m)} \) for each iteration \( m = 1, 2, ..., M \) in the MCMC algorithm; and calculated a posterior estimate, \( \hat{\epsilon}^{(s,r)}(t) \), by taking the median of the posterior samples \( \left\{ \epsilon^{(s,r)}(t) \right\}_{m=1}^{M} \). Similarly, I obtained posterior samples of \( \epsilon^{(s,r)}(t) \) by calculating \( \epsilon^{(s,r)}(t) = \left[ y^{(s,r)}(t) \right]^{(m)} - \left[ y^{(s,r)}(t) \right]^{(m)} \) for each iteration \( m = 1, 2, ..., M \) in the MCMC algorithm; and calculated a posterior estimate, \( \hat{\epsilon}^{(s,r)}(t) \), by taking the median. I inspected the time series plot of each ROI of \( \hat{\epsilon}^{(s,r)}(t) \), and did not find any obvious temporal patterns in most of these
Figure 3.8: Diagnostic plots. (a) $\hat{y}_{p}^{(s,r)}$ (black) and $\hat{y}_{p}^{*(s,r)}$ (blue). (b) $\hat{\epsilon}_{p}^{(s,r)}(t)$. (c) PACFs of $\hat{\epsilon}_{p}^{(s,r)}(t)$. The first row is for LM1 of a healthy subject, and the second row is for LM1 of a stroke patient.

time series plots. In addition, PACF plots also indicated that $L = 1$ was sufficient for most subjects, sessions and regions, but some of these plots suggest that a model with $L = 2$ may be required for a small number of subjects. Finally, plots of $\hat{y}^{*(s,r)}(t)$ also show that the group model captures the mean trend of the data, especially for the ROIs that have relatively strong activations (e.g., LM1). Figure 3.8 shows examples of time series plots of $\hat{y}^{*(s,r)}(t)$ and $\epsilon^{(s,r)}(t)$, and the PACF plots of $\hat{\epsilon}^{(s,r)}(t)$, for LM1 and two subjects, one healthy and the other one a stroke patient.

3.4.3 Discussion of the analysis

Using the proposed model, I obtained interesting results about the relative activation patterns in stroke patients compared to control subjects. First, it is found that, as expected, there was strong activation within LM1 for both the stroke and control groups. This is consistent with the fact that LM1 is primarily responsible for motor function of the right hand. All the ROIs were activated in response to the task condition (see Figure 3.5(a)) for
the stroke patients; and all the ROIs are activated for healthy subjects expect for RM1 (see Figure 3.5(b)), which is primarily responsible for the motor function of the left side of the body.

Activation of secondary motor areas in healthy controls initially seems counterintuitive. However, the control subjects were age matched to the stroke group and, therefore, were of 50+ years of age. It is established that these regions are involved in the control of unilateral movement [4] and with aging increasing involvement of secondary brain regions is seen to support simple motor movements [58].

Second, it is found that greater activation strength in the 3 motor regions (LPMd, SMA, RM1 and RPMd) for stroke patients compared to healthy controls. This difference in activation strength was notably greater in RM1, RPMd, and SMA. The literature confirms this finding. After stroke, such secondary motor regions are routinely more activated to support execution of post-stroke movements, particularly in more impaired patients [59].

The connectivity modeling produced some unexpected results. After stroke, fMRI connectivity studies show that there are excitatory and inhibitory connections between primary motor cortex in the stroke-affected hemisphere and secondary motor regions within the affected and unaffected hemispheres [44]. Therefore, Figure 3.6(b) showing a lack of connectivity between LM1 and other regions like RM1 and RPMd is surprising.

Stroke is a very heterogeneous disorder and perhaps in the stroke group there was sufficient noise and variation across the patients leaving the current statistical approach unable to detect predominant between-region connections. There is a dearth of literature describing how activity within a region predicts subsequent activation in the region. A recent study from the UCI space-time models group did identify that this intraregional activation prediction is present for LM1, LPMd, SMA, and RPMd [23], largely confirming what I observed in the current analyses. Together the data suggest possible positive feedforward connectivity in
these regions. Furthermore, the greater likelihood of the connectivity being present in stroke patients versus healthy control may lend support for these areas being more important to guiding and coordinating movement after stroke. But why the model does not see an interaction between these regions and LM1 remains uncertain.

In healthy control subjects, the proposed method identified probable connections from LM1 to the secondary motor regions. Although the exact neurobiological underpinnings of the connectivity cannot be determined from these analyses, considering the older age of the control group, the predominant connection from LM1 to LPMd is perhaps suggestive of recruitment of a secondary motor area in the hemisphere responsible for movement of the right hand.

These special patterns of stroke patients might further help neuroscientists measure the degree of post-stroke recovery, help evaluate the effect of a treatment, and help predict the chance of full recovery from a treatment to make medical decisions.

### 3.5 Conclusions

The proposed approach presents several advantages with respect to currently available methods: (1) it is able to simultaneously infer activation, connectivity and HRFs; (2) it is able to provide ROI-specific and subject-specific HRFs, as well as condition-specific connectivity; (3) it is able to borrow information across subjects via hierarchical modeling to increase power for group comparison; this is particularly relevant for data with low signal to noise ratio; (4) it is able to easily incorporate relevant information derived from other studies but, unlike the dynamic causal modeling approach, it does not rely on biological assumption that are difficult to verify from the data.

As a note of caution, the connectivity obtained using the proposed approach is at the hemo-
dynamic level instead of the neuronal level, i.e., connectivity from ROI $p$ to ROI $q$ here means there is an association between the current hemodynamic activity in ROI $p$ to the future hemodynamic activity in ROI $q$. This type of connectivity is more difficult to interpret compared to neuronal-level connectivity mainly due to two limitations. One limitation is that the information transmission happening between the neurons is much faster than the fMRI sampling frequency, and spurious connectivity may be produced by VAR models due to downsampling. It was illustrated in Fig.7 in [55]. One solution is to use vector autoregressive moving average models (VARMA) which are more robust to this issue, although they involve more complicated computation than VAR models. In fact, if the uncorrelated noise term $\epsilon^{(s,r)}(t)$ is moved from the VAR structure (Equation 3.3) to the GLM component (Equation 3.1), the proposed model can be written into a VARMA representation:

\[ y^{(s,r)}(t) = (I - \Phi^*(s) B)M^{(s)}(t) + \Phi^*(s)y^{(s,r)}(t-1) + (I - \Phi_t B)\epsilon^{(s,r)}(t), \]

where $M^{(s)}(t)$ is the mean component with the $p$-th element equal to $M_p(t) = \beta_0^{(s)} + \sum_k \beta_k^{(s)} x_{pk}^{(s)}(t)$ for $p = 1, \ldots, P$, and $\Phi^*(s) = \sum_k \Phi_k^{*(s)} c_k(t-1)$. The second limitation is that the sparsity of the neuronal-level connectivity may be distorted when transferred to the hemodynamic level through convolution [55, 52]. A solution to this limitation is to model the neuronal dynamics through a state-space model [43]. In the state-space model, the fMRI signals are treated as the observed variables and the neuronal dynamics as the hidden states. Neuronal-level connectivity can thus be modeled via the hidden states. This type of model can be solved using Kalman filter and Kalman smoother, and can be implemented efficiently by variational Bayes method [47, 35]. However, the hidden states are modeled at a much lower resolution than the actual neuronal activity. This leads us to suspect that such connectivity might still suffer from the downsampling limitation. Despite the limitations, the proposed model well captured the differential patterns between the stroke patients and the healthy subjects in this stroke study; I speculate it is partially due to the fact that I modeled
the region-specific HRFs.

Besides the limitations and possible solutions mentioned above, I also want to make improvements to my statistical approach in the following ways. In the current approach I assume independence across groups and analyze the groups separately; in the future I want to incorporate grouping factors into the model, instead of estimate the model separately for each group. This may further help increase the power of detecting the differences between groups since I can borrow information from more subjects. The current approach also assumes the activation strength is the same across sessions; in the future I want to extend the proposed model to take into account the variation in the activation strength among different sessions.
Chapter 4

Bayesian State-Space Model for Single-Subject fMRI Data

4.1 Introduction

This chapter discusses another type of general linear model and Granger-causality based approach for modeling activation and effective connectivity – the state-space modeling. The term “effective connectivity” will be abbreviated as “connectivity” if not otherwise specified. The aforementioned vector autoregressive (VAR) modeling approaches and the state-space modeling approach share the same model for brain activation, the general linear model (GLM). Compared to the vector autoregressive modeling approaches, the state-space modeling approach models effective connectivity through the dynamics of the underlying BOLD or neuronal signals which are modeled as latent signals through the state equation. While the VAR approach models effective connectivity through the unexplained noise. Therefore, from a biological point of view, a state-space model has the advantage that the effective connectivity from state-space models is easier to interpret than that from VAR models.
In [28] a state-space model was developed for modeling brain activation and brain connectivity. This model consists of the activation equation and the connectivity equation. The activation equation is the observation equation that follows a GLM, and it models the observed fMRI signals as a function of the BOLD response that potentially represents the brain activities. The connectivity equation is the state equation for the time-varying amplitudes of BOLD responses, which models the lagged dependence between the regions of interest (ROIs) by modeling the BOLD amplitude of each ROI at the $t$-th time point in terms of the BOLD responses of other ROIs at the previous time point. The model in [28] is illustrated in Figure 4.1. By including the latent process (discrete-time), this model captures the dynamics of the region-specific brain activities, and describes brain connectivity through the lagged between-region dependence among the BOLD amplitudes in different regions. Note that the connectivity for this model remains at the hemodynamic level since it is based on the BOLD amplitudes. However, this model does not account for possible non-stationarities in connectivity due to alternating experimental conditions. Bhattacharya et al. extended the work by treating connectivity also as time-varying, and specified a random walk [7] and a Dirichlet process [6] on the connectivity, respectively.

All of the above methods assumed the same shape of HRF across all the regions and may have produce spurious results of connectivity. Also, without imposing possible sparsity on the connectivity parameters, inference for connectivity requires computing models for different connectivity networks and then performing model comparisons. This could be computationally expensive if the number of ROIs are not small and when little information is available for which networks are most biologically plausible.

All the above state-space models study brain connectivity at the hemodynamic level. The direct interpretation for brain connectivity from ROI $p$ to ROI $q$ from these models is as follows: the higher the BOLD amplitude in ROI $p$ is at the current time point, the higher the BOLD amplitude in ROI $q$ is expected to be (for positive connectivity) at the next time point,
or the lower the BOLD amplitude in ROI $q$ is expected to be (for negative connectivity) at the next time point. We need to be cautious that this connectivity is the lagged correlation between ROIs at the hemodynamic level but not necessarily the causal relationship between the ROIs at the neuronal level.

Ryali et al. developed a different state-space model to model effective connectivity [47]. Instead of modeling the dynamic BOLD amplitude as the latent signal, it models the underlying neuronal activity as the latent signal (before hemodynamic convolution), and models effective connectivity based on the lagged dependence among the neuronal signals from different ROIs. The neuronal signal connects to the observed signal through convolution with (basis functions of) the HRF. Luessi et al. largely enhanced its computational efficiency by introducing an additional layer of hidden signals and by using variational Bayes (VB) inference [35]. According to the above models, the estimated connectivity is at the neuronal level. As is known, neuronal activities happen at a very fast pace compared to the sampling rate of fMRI scans. However, the latent neuronal signal was modeled with the same temporal resolution as the fMRI observations, and thus has a much lower resolution than the actual neuronal activity. The “neuronal signals” in the model are likely to be the result of accu-
mulation of many neuronal activities. As the connectivity is inferred from such “neuronal signals”, it is still not safe to directly interpret it as the neuronal level connectivity.

In this study I adopt the hemodynamic-level based connectivity model in [28], as it is more consistent with the low-temporal resolution of fMRI data. To address the aforementioned limitations of this state-space model, a model with the following improvements is proposed. (1) A region-specific HRF will is explicitly modeled to account for its potential effect on connectivity. (2) Sparsity is imposed on connectivity through spike and slab priors, to allow for direct inference on the connectivity network. (3) Connectivity is modeled as condition specific by introducing the separate processes of BOLD amplitudes for different experimental conditions.

The rest of the chapter is organized as follows. Section 4.2 describes the proposed model and its inference procedure; Section 4.3 discusses the model performance from the simulations studies; Section 4.4 shows the application of this model to an actual fMRI data; Section 4.5 discusses possible extension to multi-group analysis; and Section 4.6 concludes the study.

### 4.2 Methodology

#### 4.2.1 Model

Here I propose a new state-space model for modeling fMRI brain activation and connectivity.

\[
\mathbf{y}^{(r)}(t) = \mathbf{\Upsilon}_0 + X_1(t)\tilde{\mathbf{\beta}}_1^{(r)}(t) + X_2(t)\tilde{\mathbf{\beta}}_2^{(r)}(t) + \mathbf{\epsilon}^{(r)}(t), \tag{4.1}
\]

\[
\tilde{\mathbf{\beta}}_k^{(r)}(t) = \mathbf{\Upsilon}_k + \mathbf{\beta}_k^{(r)}(t), \tag{4.2}
\]

\[
\mathbf{\beta}_k^{(r)}(t) = \Phi_k X_k(t - 1)\tilde{\mathbf{\beta}}_k^{(r)}(t - 1) + \mathbf{w}_k^{(r)}(t), \tag{4.3}
\]
where

\[
X_k(t) = \begin{bmatrix}
x_{1k}(t) & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & x_{Pk}(t)
\end{bmatrix}, \quad \Phi_k^* = \begin{bmatrix}
\phi_{11,k} & \ldots & \phi_{1P,k} \\
\vdots & \ddots & \vdots \\
\phi_{P1,k} & \ldots & \phi_{PP,k}
\end{bmatrix}.
\]

Equation (4.1) and Equation (4.3) correspond to the activation equation and the connectivity equation in [28]. In the activation equation, the variable \( y^{(r)}(t) \in \mathbb{R}^P \) is the vector of the observed fMRI signals for the \( P \) ROIs at time \( t \). It is composed of the total BOLD response \( X_k(t) \tilde{\beta}_k^{(r)}(t) \) from two experimental conditions \( k = 1 \) (stimulus condition) and \( k = 2 \) (control condition), and the noise term \( \epsilon^{(r)}(t) \in \mathbb{R}^P \). The total BOLD response is the product of the BOLD response shapes, \( X_k(t) \), and the dynamic BOLD amplitudes, \( \tilde{\beta}_k^{(r)}(t) \in \mathbb{R}^P \). \( \tilde{\beta}_k^{(r)}(t) \) is further decomposed to the mean component \( \Upsilon_k \), and the stochastic component \( \beta_k^{(r)}(t) \) whose elements are all centered at zero. \( \beta_k^{(r)}(t), k = 1, 2 \) are the latent processes in this state-space model, for conditions \( k = 1, 2 \), respectively. The two processes can also be seen as a single joint process by stacking them together: \( \beta^{(r)}(t) = [\beta_1^{(r)}(t), \beta_2^{(r)}(t)]' \). Note that for a more general experimental design with \( K \) experimental conditions, the model can be generalized to have \( K \) separate processes for the \( K \) conditions. The estimation and inference for \( K > 2 \) will be similar to \( K = 2 \). In this work I only discuss the simplest case where \( K = 2 \).

In the connectivity equation, the BOLD amplitudes depend on the history of the BOLD responses, including the history of each ROI itself and also the other ROIs. Based on Granger-causality, the within-region and between-region dependence with temporal lag allows us to examine the potential influence from one region to another, and therefore can help us model the brain connectivity from the fMRI data. Note that the BOLD amplitudes
here are modeled to only depend on the history of one lag, i.e., $\beta_k^{(r)}(t)$ only depends on the BOLD response at time $t - 1$. However, it is possible to include more lags by rewriting the latent processes, which will be discussed in the Section 4.6.

Based on the model, brain activation and connectivity can be defined as follows:

**Activation:** when the contrast between the mean BOLD amplitudes of the two conditions, $\Upsilon_{p,(c)} = \Upsilon_{p1} - \Upsilon_{p2}$, is larger than 0, then ROI $p$ is said to be activated. The magnitude of the activation contrast $\Upsilon_{p,(c)}$ represents the strength of the activation. Note that it is uncommon but still possible that $\Upsilon_{p,(c)} < 0$, and ROI $p$ is commonly called “deactivated”.

**Connectivity:** when the $(p, q)$-th element of the transition matrix $\Phi_k^*$, denoted $\phi_{pq,k}^*$, is not equal to 0, then one concludes that there exists connectivity from ROI $q$ to ROI $p$ under condition $k$; the magnitude of $\phi_{pq,k}^*$ represents the strength of the connectivity. Note that this connectivity is a type of effective connectivity but not functional connectivity. In this chapter, I focus on only the effective connectivity, and the term “effective connectivity” will be abbreviated as “connectivity”.

In the proposed model, region-specific HRFs are also modeled to account for the variability of HRFs across regions and the possible confounding effects on the lagged association between the regions. A spike and slab prior is put on the connectivity parameters so that inference on the whether connectivity is present or not can be made by directly calculating $\Pr(\phi_{pq,k}^* \neq 0)$, for $p, q = 1, \ldots, P$, instead of doing model comparison for all possible combinations of connectivity. The above features are incorporated in the model in the following equations.

\[
x_{pk}(t) = [h_p * c_k](t),
\]

\[
h_p = H d_p, \quad H \in \mathbb{R}^{T_h \times J}, \quad d_p \in \mathbb{R}^J, \quad \text{subject to } \sum_t h_p(t) = 1,
\]

\[
c_k(t) = 1 \quad \text{when condition } k \text{ is on}, \quad c_k(t) = 0 \quad \text{when condition } k \text{ is off},
\]

\[
\phi_{pq,k}^* = \phi_{pq,k} \times \xi_{pq}, \quad \phi_{pq,k} \in \mathbb{R}, \quad \xi_{pq} \in \{0, 1\}.
\]
$h_p(t)$ is the HRF for ROI $p$, it is the linear combination of the linear HRF basis matrix $H$ of $J$ columns, and the vector of the HRF basis coefficients, $d_p \in \mathbb{R}^J$. $c_k(t)$ is the binary indicator for condition $k$. Condition $k = 1$ is stimulus condition, and $k = 2$ is control condition. The BOLD response shape $x_{pk}(t)$ is then the linear convolution between the HRF $h_p(t)$ and the condition indicator $c_k(t)$. Equation (4.7) allows for sparsity and direct inference on $\Pr(\phi_{pq,k}^{\ast} \neq 0)$ through the binary variable $\xi_{pq}$. In fact, it is a computationally convenient implementation for the spike and slab prior, which will be explained in Section 4.2.3. In this chapter, the same HRF basis as the previous chapter is used. It is a constrained linear basis of size 5, obtained using the approach in [62]. Details can be found in Section 3.2.3 in Chapter 3.

Furthermore, Gaussian assumptions are imposed on the observation noise, the state noise and the initial point of the latent process (Equations 4.8–4.10), so the model is a Gaussian state-space model. $\beta^{(r)}(0)$, $\epsilon^{(r)}(t)$ and $w^{(r)}_k(t)$ are also assumed to be independent with each other. The posterior distributions of the latent processes at each time point conditional on all other model parameters can be obtained by forward filtering backward smoothing algorithm [10, 19] (FFBS) based on Kalman filter and Kalman smoother.

$$\beta^{(r)}(0) \overset{i.i.d.}{\sim} N(0_{2P}, I_{2P}),$$  \hspace{1cm} (4.8)

$$\epsilon^{(r)}(t) \sim N(0, V),$$ \hspace{1cm} (4.9)

$$w^{(r)}_k(t) \sim N(0, Q), \ k = 1, 2.$$ \hspace{1cm} (4.10)

All the unknown variables to be estimated in the model are illustrated in Figure 4.2 (the parameters with a white background). The new notations appear in the Figure are defined as:

$\Upsilon = [\Upsilon_0', \Upsilon_1', \Upsilon_2']' = [\Upsilon_{10}, \ldots, \Upsilon_{P0}, \Upsilon_{11}, \ldots, \Upsilon_{P1}, \Upsilon_{12}, \ldots, \Upsilon_{P2}]' \in \mathbb{R}^{3P}$, $d = [d_1, \ldots, d_P] \in \mathbb{R}^{JP}$, and $\Phi_k = [\phi_{pq,k}]_{p,q} \in \mathbb{R}^{P \times P}$. 

76
There are several additional assumptions inherent in this model besides the linear convolution assumption and the Gaussian assumption. First, the mean global BOLD amplitude for each ROI is assumed to be the same across different sessions \( r, r = 1, 2, \ldots, R \). Second, the HRF shape for each ROI is also assumed to be the same across different sessions, but is allowed to vary across ROIs. Third, the two latent processes of BOLD amplitudes for the two experimental conditions are assumed to be independent of each other. Fourth, by specifying \( \xi_{pq} \) to be independent with \( k \), it is assumed that the connectivity from ROI \( q \) to \( p \) exists either for both conditions, or for neither of the conditions; however, it is allowed that for different conditions, the connectivity strength can vary, i.e., \( \phi_{pq,k} \) depends on \( k \). Last, the covariance for the dynamic BOLD amplitudes is the same across conditions.

Compared to the state-space model in [28], this new model has its improvements in: (1) allowing the dynamic BOLD amplitudes to be different stochastic processes across different conditions \( (\beta_1^{(r)}(t) \text{ and } \beta_2^{(r)}(t)) \), and thus allow for the modeling of condition-specific connectivity, \( \Phi_k^* \); (2) accounting for the regional variation in the HRFs by modeling \( d_p \), which is crucial for fMRI data modeling; (3) allowing for the incorporation of available scientifically relevant prior information through a Bayesian framework, which will be discussed in detail.
in the Section 4.2.3; (4) allowing for inference on the connectivity network without having to do model comparison for all possible networks, through the spike and slab prior on $\Phi_k^*$. 

### 4.2.2 Relation to the VAR model

In general, a VAR model can be seen as a special case of a state-space model, and can be written in a state-space representation. For the fMRI Bayesian VAR models proposed in this dissertation, they can also be written into a state-space representation as the following equations (ignoring superscripts for subject and fMRI sessions, and assuming lag 1, for notational simplicity):

$$y(t) = \Upsilon_0 + \sum_k X_k(t)\Upsilon_k + u(t) + \epsilon(t), \quad \text{observation equation}$$

$$u(t) = \sum_k \Phi_k^* c_k(t-1)u(t-1) + w(t), \quad \text{state equation}$$

where $\epsilon(t) \equiv 0$.

Here the unexplained noise $u(t)$ is considered as the latent process in the state-space model, instead of the observation noise; while the observation noise is assumed to be 0 ($\epsilon(t) \equiv 0$).

Notice that this representation of the VAR model looks very similar to the proposed state-space model, and for both models connectivity is inferred from the latent process. One major difference is that, in the VAR model, the relationship between the observed signal, $y(t)$, and the latent process, $u(t)$, is not tied with the BOLD response. However, in the proposed state-space model, the latent process relates to the observed signal through the BOLD response ($\sum_k X_k(t)\beta_k(t)$). Because of this difference, the latent process in the state-space model seems to be more biologically meaningful. The other major difference is that in the VAR model, it is assumed that there is no observation noise, i.e., the observed signal can
be fully explained by the BOLD response and the latent process; while in the state-space model additional noise is allowed in the observed signals, which is more realistic.

### 4.2.3 Prior specification

The prior distributions for the model parameters are specified as follows. All the constant hyperparameters of the prior distributions are illustrated in Figure 4.2 (the parameters with having grey background).

\[
\mathbf{Y} \sim N(\mu_Y, \Sigma_Y), \quad (4.11)
\]
\[
d \sim N(\mu_d, \Sigma_d), \quad (4.12)
\]
\[
\phi_k = \text{vec}\{\Phi_k\} \sim N(\mu_\phi, \Sigma_\phi), \quad (4.13)
\]
\[
\xi_{pq} \sim \text{Bernoulli}(\pi_{\xi,pq}), \quad (4.14)
\]
\[
V^{-1} \sim \text{Wishart}(\Psi_V, \nu_V), \quad (4.15)
\]
\[
Q^{-1} \sim \text{Wishart}(\Psi_Q, \nu_Q). \quad (4.16)
\]

\(\mu_Y, \mu_d\) and \(\mu_\phi\) represent the mean/median of the mean BOLD amplitudes, the HRF basis coefficients and the connectivity strength, respectively, in the prior knowledge. \(\Sigma_Y, \Sigma_d\) and \(\Sigma_\phi\) are assumed to be diagonal matrices, and the diagonal elements represent the uncertainty in the prior knowledge of the corresponding quantities. Let \(\phi\) denote \([\phi_1, \phi_2]\). One way to elicit \(\Sigma_{\theta,ii}(\theta = \mathbf{Y}, \mathbf{d}, \text{or } \phi)\) is to obtain a \(\alpha\)-th quantile for \(\theta_i\), \(F_\alpha\), from prior studies, and set \(\Sigma_{\theta,ii} = (F_\alpha - \mu_{\theta,i})/qnorm(\alpha)\). If no prior information is available, \(\mu_{\theta,i}\) is set to 0 and \(\Sigma_{\theta,ii}\) is set to be large value to represent the high uncertainty in this variable. \(\pi_{\xi,pq}\) represents the probability for the presence of connectivity from ROI \(q\) to ROI \(p\) based on available prior knowledge. If no prior information is available, \(\pi_{\xi,pq}\) can be set to 0.5. \(\nu_V\Psi_V\) and \(\nu_Q\Psi_Q\)
represents the prior mean of the observation precision matrix and the state precision matrix. \( \nu_V \) and \( \Psi_V \) (or \( \nu_Q \) and \( \Psi_Q \)) control for the uncertainty of \( V^{-1} \) and \( Q^{-1} \); and the larger \( \nu_V \) (or \( \nu_Q \)) is, the more certain the prior belief is about the precision matrix \( V^{-1} \) (or \( Q^{-1} \)) (and thus the more certain about \( V \) (or \( Q \))), when the prior mean \( \nu_V \Psi_V \) (or \( \nu_Q \Psi_Q \)) is held constant.

Note that in practice, using values no less than \( P + 1 \) for \( \nu_V \) and \( \nu_Q \) will be preferred since it ensures the existence of the posterior mean of \( V \) and \( Q \).

Here the prior distributions 4.13, 4.14 and Equation 4.7 together define a spike and slab prior on the connectivity parameters \( \phi_{pq,k}^{*} \) (marginally):

\[
\phi_{pq,k}^{*} \sim \pi_{pq} N \left( \mu_{\phi,pq,k}, \Sigma_{\phi,pq,k} \right) + \left( \text{1} - \pi_{pq} \right) \{0\},
\]

where \( \mu_{\phi,pq,k} \) and \( \Sigma_{\phi,pq,k} \) are the element in the vector \( \mu_{\phi} \) and the diagonal element in the matrix \( \Sigma_{\phi} \) corresponding to the parameter \( \phi_{pq,k}^{*} \).

### 4.2.4 Inference

#### MCMC algorithm

It is easy to see that the prior distributions are all conditionally conjugate, so I can use a Gibbs sampling algorithm to obtain posterior samples for the latent processes and the other unknown parameters. Let \( \Theta \) denote the set of all the parameters except those describing the latent processes, \( \beta \) denote the set of all latent processes, and \( Y \) denote the set of all the observations. The algorithm works in the following steps.

1. Set initial values for all the unknown parameters except the latent processes.
2. Use FFBS algorithm to sample the joint latent process \( \beta^{(r)}(t) = [\beta_{1}^{(r)}(t), \beta_{2}^{(r)}(t)] \), for \( r = 1, \ldots, R \), based on the latest iteration of the other unknown parameters.
3. Sample the rest of the unknown parameters based on the latest iteration of the latent processes and the other parameters:

(a) Sample $\phi|\{Y, \beta, \Theta \setminus \phi\}$. (Multivariate Gaussian).

(b) Sample $Q^{-1}|\{Y, \beta, \Theta \setminus Q\}$ (Wishart distribution)

(c) Sample $\xi_{pq}|\{Y, \beta, \Theta \setminus \{\xi_{pq}, \phi_{pq,k}, k = 1, 2\}\}$, for $p, q = 1, \ldots, P$ in a random order. (Bernoulli)

(d) Sample $d|\{Y, \beta, \Theta \setminus d\}$. (Multivariate Gaussian)

(e) Sample $\Upsilon|\{Y, \beta, \Theta \setminus \Upsilon\}$. (Multivariate Gaussian)

(f) Sample $R^{-1}|\{Y, \beta, \Theta \setminus R\}$. (Wishart)

4. Repeat steps 2 and 3 until the desired size of the posterior sample is reached.

The FFBS algorithm was implemented in C++ core using R package RcppArmadillo [16]. The rest of the sampling algorithm was implemented in R. The detailed full conditional distributions and the sampling algorithm is provided in Appendix C.

When the signal to noise ratio in the data is low, there could be very large variability in the estimated parameters, especially in the HRF basis coefficients. This could possibly cause the HRF shapes to become unreasonable since it is modeled simply as the linear combination of the basis vectors, with no hard constraints on the shape. For this reason, I implement an simplified version of model which treats the HRFs as fixed but still region-specific. More specifically, region-specific HRFs are pre-estimated in a preliminary analysis prior to the Bayesian state-space model; the HRFs are fixed at these the pre-estimated shapes during the posterior estimation for this simplified model. The sampling algorithm is the same as that for the full version of the model except that Step 3d is dropped. In the next section, the effects of fixing HRFs at pre-estimated shapes will be discussed.
Preliminary analysis

Preliminary analysis can be helpful to obtain decent initial values for the MCMC algorithm in order to reduce burn-in time, and the estimate for \( d \) can be used for the simplified version of the model mentioned above, when needed. Here, the preliminary analysis is done as follows. The dynamic component of the BOLD response is ignored, i.e., \( \beta(t) \) is set to \( 0 \). The observation noise is set to be i.i.d. across time and sessions, and independent across ROIs, i.e., \( V \) is assumed to be diagonal. The model degenerates to

\[
\begin{align*}
y^{(r)}_p(t) &= \Upsilon_{p0} + \{H * c_1\}(t)d_p\Upsilon^{(r)}_{p1}(t) + \{H * c_2\}(t)d_p\Upsilon^{(r)}_{p2}(t) + \epsilon^{(r)}(t), \\
\epsilon^{(r)}_p(t) &\sim N(0, V_{pp}),
\end{align*}
\]

(4.17) 

(4.18)

\( t = 1, 2, \ldots, T; \ r = 1, 2, \ldots, R. \)

The prior distributions remain the same. Posterior modes for \( \Upsilon_{pk} \), \( k = 0, 1, 2, d_p \) and \( V \) can then be computed via conditional maximization and be used as their preliminary estimates.

Inference for brain activation and connectivity

Inference on whether a ROI is activated or not, and whether there is connectivity from one ROI to another ROI can be made from the posterior samples obtained through the MCMC algorithm. **Activation:** if the \( (1 - \alpha) \) credible interval (CI) of the contrast of the BOLD amplitudes, \( \Upsilon_{p(c)} = \Upsilon_{p1} - \Upsilon_{p2} \) is entirely on the right side of 0, then it is claimed that the ROI \( p \) is activated; otherwise the ROI is not activated. **Connectivity:** if \( \Pr(\xi_{pq} \neq 0) > \lambda \), where \( \lambda \) is a pre-specified threshold (\( \lambda = 0.5 \) is used by default for this study), then it is claimed that the connectivity from ROI \( q \) to ROI \( p \) is present; otherwise the connectivity is absent. In addition, if the \( (1 - \alpha) \) CI for \( \phi^{(r)}_{pq,k} \) is entirely on the right side of 0, then it is
claimed that ROI $q$ has positive connectivity to ROI $p$; and if it entirely on the left side of 0, then it is claimed that ROI $q$ has negative connectivity to ROI $p$.

For a HRF curve, a point estimate can be obtained by multiplying the HRF basis matrix $H$ with the point estimate of the HRF basis coefficients; and a point-wise $(1 - \alpha)$ credible band can be obtained by first calculating the posterior samples of the HRF curve as the product of $H$ with each sample of the basis coefficients, and then obtaining the $(1 - \alpha)$ CI for each time point to form a credible band. In addition, for HRFs, the following three shape quantities are often of interest, namely, the time to peak, the with of the peak – usually measured by full width at half maximum (FWHM) – and time to the post-stimulus undershoot. The point estimates and $(1 - \alpha)$ CI for these quantities can also be similarly computed, using the posterior samples of the HRFs.

The $(1 - \alpha)$ CIs mentioned above can be obtained by finding the empirical $\alpha/2$-th and $(1 - \alpha/2)$-th quantile from the corresponding posterior samples. For this study, I use 95% CIs for inference. The point estimates can be obtained by finding the empirical posterior median.

### 4.3 Simulation study

In this section the properties of the proposed model are examined through simulation studies. The first simulation study evaluates the model performance in terms of inferring connectivity, activation and HRF shapes in details; the second simulation study investigates the effects of certain simulation settings, prior distributions and posterior algorithms on the model performance.

#### 4.3.1 Simulation settings
For the first simulation study, I simulated 50 datasets with 3 ROIs based on the proposed model described in Section 4.1. The simulated datasets have 2 independent fMRI sessions each containing 4 consecutive blocks; each block consists of 16 TRs alternating between stimulus condition and control condition starting with a block of stimulus condition. TR equals to 1 second. For connectivity, the parameters were generated as follows:

\[
\xi_{pq} \sim \text{Bernoulli}(0.16), \\
\phi_{pq,1}|\eta_{pq} \sim \text{sign}(\eta_{pq} - 1/2) \times \text{Unif}(0.6, 1.1), \\
\phi_{pq,2}|\eta_{pq} \sim \text{sign}(\eta_{pq} - 1/2) \times \text{Unif}(0.4, 0.7), \quad \text{where} \\
\eta_{pq} \sim \text{Bernoulli}(2/3).
\]

For activation, the mean BOLD amplitudes were fixed at \(\Upsilon_1 = [6, 4, 2]'\) under the stimulus condition, \(\Upsilon_2 = [2, 4, 6]'\) under the control condition, and thus the contrasts were \(\Upsilon_{(c)} = [4, 0, -4]'\); the intercepts were fixed at \(\Upsilon_0 = [-1, -3, -5]\). For HRF, the HRF shapes were generated from the constrained linear basis functions with basis size \(J = 10\), larger than the actual basis size used for model estimation \((J = 5)\). Examples of generated HRFs are shown in Figure 4.3; they vary in time to peak, width of the peak, and the time to the post-stimulus undershoot. The observation covariance matrix was fixed at \(V = I_P\), and the state covariance matrix was fixed at \(Q = \frac{1}{8}\text{diag}\{3, 4, 5\}\). The above settings for the data generation will be referred to as Scenario 1. Here

For the second simulation study, I simulated 50 datasets under each of the two additional scenarios described below: Scenario 2: \(Q = \frac{1}{2}\text{diag}\{3, 4, 5\}\) with all other settings identical to Scenario 1; in this scenario, the state noise increases and consequently, the overall magnitude of the latent processes who are driven by the state noise also increases. Scenario 3: 8 fMRI sessions with all other settings identical to Scenario 1; in this scenario, the amount of fMRI
Figure 4.3: Examples of region-specific HRFs used in the simulations.

time series increases and there is potentially more information available for estimating the parameters. Examples of simulated fMRI signals under all scenarios are shown in Figure 4.4.

4.3.2 Analysis

For the first simulation study, the posterior samples were computed for the simulated datasets using the algorithm described in 4.2.4. The posterior samples for each dataset are of size 1,000 thinned at every 5th iteration, after a burn-in period of 1,000 iterations. The constant hyperparameters for the prior distributions used in the estimation are:

\[
\mu_Y = 0_{3P}, \quad \Sigma_Y = 10I_{3P},
\]
\[
\mu_\phi = 0_{P^2}, \quad \Sigma_\phi = I_{P^2}, \quad \pi^{(s)}_{pq} = 0.5,
\]
\[
\mu_d = 1_P \otimes [1.6814, 0.4377, 0.0213, -0.0343, 0.0615]',
\]
Figure 4.4: Examples of simulated fMRI time series of the three simulation scenarios. Black: simulated fMRI signals. Purple: mean of the BOLD signals (excluding the variation due to dynamic BOLD amplitudes). Red: total BOLD signal (including the variation due to dynamic BOLD amplitudes). Green vertical lines indicate the start/end of a fMRI scan session. Green horizontal lines indicate the stimulus condition is on.
The prior distribution for \( d \) was obtained using the approach in [62]. This set of prior distributions is referred to as Prior 1.

The second simulation study investigated the effects of certain simulation settings, prior distributions and posterior algorithms on the model performance. In particular, it compared the model performance under the following 6 settings in Table 4.1. Prior 2 is the same as Prior 1 except that \( \phi \sim N(0, 4I) \) instead of \( \phi \sim N(0, I) \), and Prior 3 is the same as Prior 1 except that \( \phi \sim N(0, I/4) \). Setting 1, or the default setting, is the same set of settings used for the first simulation study.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Scen.1</th>
<th>Scen.2</th>
<th>Scen.3</th>
<th>FixHRF</th>
<th>Prior 1</th>
<th>Prior 2</th>
<th>Prior 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting 1</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Setting 2</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Setting 3</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Setting 4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Setting 5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Setting 5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Six settings for the simulation study. “Scen.”: Scenario. “FixHRF”: the simplified model with HRFs fixed.

Effects of increasing the overall magnitude of the latent signals were checked by comparing Setting 2 to Setting 1; effects of increasing the amount of fMRI time series were checked by comparing Setting 3 to Setting 1; effects of fixing the HRFs at the pre-estimated shapes
(using pre-estimated $d$) were checked by comparing Setting 4 to Setting 1; and finally, effects of increasing or decreasing $\Sigma_\phi$—the prior variance of the slab part of the prior for the connectivity parameters—were checked by comparing Setting 5 or 6 to Setting 1.

Model performance was assessed by the following measures.

1. Connectivity. (a) ROC curve and area under the ROC curve (AUC) will be used to assess the accuracy of classification for whether a connectivity is present. (b) Average sensitivity (the probability that connectivity is claimed to be present given that connectivity is actually present) and specificity (the probability that connectivity is claimed to be not present given that connectivity is absent) at threshold $\lambda = 0.5$ for claiming the presence of connectivity (“Sens.” and “Spec.”). (c) Average coverage of 95% CIs for $\phi_{pq,k}^*$ for all present connectivity (“cover”). (d) Average bias of point estimates for $\phi_{pq,k}^*$ of all present connectivity (“bias”).

2. Activation. (a) Average proportion that the 95% CI for $\Upsilon_{p,(c)}$ not containing 0 given that $\Upsilon_{p,(c)} \neq 0$ (“power”). Note that this includes the detection for both activated ROIs and “deactivated” ROIs. (b) Average proportion of the 95% CI containing 0 given that $\Upsilon_{p,(c)} = 0$ (“spec”). (c) Average coverage of 95% CI for $\Upsilon_{p,(c)}$ (“cover”). Average bias of point estimates for $\Upsilon_{p,(c)}$ (“bias”).

3. HRF. (a) Average coverage of 95% credible band for HRFs (“avg.cover”). (b) Average coverage of 95% CI for the time to peak $T_{peak}$, FWHM $T_{wid}$, and time to undershoot $T_{under}$ ($T_{peak,cover}, T_{wid,cover}, T_{under,cover}$). (c) Average bias of point estimates for $T_{peak}, T_{wid}$, and $T_{under}$ ($T_{peak,bias}, T_{wid,bias}, T_{under,bias}$). (d) Average standard deviation of the derived posterior samples $T_{peak}$, FWHM $T_{wid}$, and time to undershoot $T_{under}$ ($T_{peak,sd}, T_{wid,sd}, T_{under,sd}$). For $T_{peak}, T_{wid}$, and $T_{under}$, the bias and the standard deviation are both in the unit of second.
4.3.3 Results

The quantitative measures for the model performance for all six settings are listed in Table 4.2, and the ROC curves for connectivity are shown in Figure 4.5. Figure 4.7 shows some examples of the estimated signals. The estimated signals well follow the trend of the true dynamic latent signals. Example trace plots for various parameters are provided in Appendix C. The posterior chains look stable.

<table>
<thead>
<tr>
<th>Setting:</th>
<th>1 (Default)</th>
<th>2 (Scen.2)</th>
<th>3 (Scen.3)</th>
<th>4 (FixHRF)</th>
<th>5 (Prior2)</th>
<th>6 (Prior3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>power</td>
<td>0.9800</td>
<td>0.9200</td>
<td>1.0000</td>
<td>0.9900</td>
<td>0.9700</td>
<td>0.9900</td>
</tr>
<tr>
<td>spec</td>
<td>0.7200</td>
<td>0.5400</td>
<td>0.6400</td>
<td>0.5400</td>
<td>0.7800</td>
<td>0.6400</td>
</tr>
<tr>
<td>cover</td>
<td>0.4000</td>
<td>0.4667</td>
<td>0.4400</td>
<td>0.3667</td>
<td>0.4467</td>
<td>0.4533</td>
</tr>
<tr>
<td>bias</td>
<td>0.3475</td>
<td>0.2880</td>
<td>0.4563</td>
<td>-0.0936</td>
<td>0.7422</td>
<td>0.1284</td>
</tr>
<tr>
<td>Connectivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.9380</td>
<td>0.9774</td>
<td>0.9909</td>
<td>0.9358</td>
<td>0.8626</td>
<td>0.9397</td>
</tr>
<tr>
<td>sens</td>
<td>0.8939</td>
<td>0.9947</td>
<td>1.0000</td>
<td>0.8380</td>
<td>0.9218</td>
<td>0.8380</td>
</tr>
<tr>
<td>spec</td>
<td>0.8413</td>
<td>0.7462</td>
<td>0.7612</td>
<td>0.8672</td>
<td>0.5535</td>
<td>0.8856</td>
</tr>
<tr>
<td>cover</td>
<td>0.9832</td>
<td>0.9763</td>
<td>0.9505</td>
<td>0.9553</td>
<td>0.9860</td>
<td>0.9302</td>
</tr>
<tr>
<td>bias</td>
<td>-0.0909</td>
<td>-0.0682</td>
<td>-0.0462</td>
<td>-0.1501</td>
<td>-0.0256</td>
<td>-0.1358</td>
</tr>
<tr>
<td>HRF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg.cover</td>
<td>0.6561</td>
<td>0.6086</td>
<td>0.6273</td>
<td>NA</td>
<td>0.4827</td>
<td>0.8346</td>
</tr>
<tr>
<td>$T_{peak}$.cover</td>
<td>0.9200</td>
<td>0.9333</td>
<td>0.8800</td>
<td>NA</td>
<td>0.9467</td>
<td>0.9133</td>
</tr>
<tr>
<td>$T_{wid}$.cover</td>
<td>0.9467</td>
<td>0.9467</td>
<td>0.8400</td>
<td>NA</td>
<td>0.9800</td>
<td>0.9333</td>
</tr>
<tr>
<td>$T_{under}$.cover</td>
<td>0.9000</td>
<td>0.8333</td>
<td>0.8267</td>
<td>NA</td>
<td>0.6933</td>
<td>0.9533</td>
</tr>
<tr>
<td>$T_{peak}$.bias</td>
<td>-0.2933</td>
<td>-0.4327</td>
<td>-0.3447</td>
<td>0.1340</td>
<td>-0.2793</td>
<td>-0.2007</td>
</tr>
<tr>
<td>$T_{wid}$.bias</td>
<td>-0.2673</td>
<td>-0.0547</td>
<td>0.5093</td>
<td>0.2527</td>
<td>0.0560</td>
<td>-0.0807</td>
</tr>
<tr>
<td>$T_{under}$.bias</td>
<td>-0.8633</td>
<td>-1.4613</td>
<td>-0.5693</td>
<td>0.0987</td>
<td>-2.3600</td>
<td>-0.1107</td>
</tr>
<tr>
<td>$T_{peak}$.sd</td>
<td>0.6236</td>
<td>0.6883</td>
<td>0.6568</td>
<td>NA</td>
<td>0.7261</td>
<td>0.6015</td>
</tr>
<tr>
<td>$T_{wid}$.sd</td>
<td>1.0813</td>
<td>1.2609</td>
<td>1.0892</td>
<td>NA</td>
<td>2.0702</td>
<td>0.8444</td>
</tr>
<tr>
<td>$T_{under}$.sd</td>
<td>2.8677</td>
<td>2.9457</td>
<td>1.8946</td>
<td>NA</td>
<td>2.8321</td>
<td>2.0820</td>
</tr>
</tbody>
</table>

Table 4.2: Model performance under different settings.

Simulation study 1: model performance under Setting 1
Figure 4.5: Average ROC for the presence of connectivity, under different settings. The horizontal and vertical axes represent the sensitivity and specificity, respectively. The solid square represents the location of threshold $\lambda = 0.5$. 

Set.1 (Default)  
Set.2 (Scen.2)  
Set.3 (Scen.3)  
Set.4 (Approx.)  
Set.5 (Prior2)  
Set.6 (Prior3)
Activation. Both the power to detect an activated (or deactivated) ROI is relatively high ($power = 0.9800$), but the ability to identify a non-activated ROI is much lower than 0.95 ($spec = 0.7200$). Coverage of the 95% CIs for the activation contrast is also much lower than 0.95 ($cover = 0.4000$). Bias is not extremely small but reasonable ($bias = 0.4600$). The above indicate the underestimation of variability in the activation contrast $Y$.

Connectivity. The classification accuracy for inferring the presence of connectivity is relatively high ($AUC = 0.9380$). At threshold 0.5, sensitivity and specificity are both reasonably well ($Sens = 0.8939$, $Spec = 0.8413$, also shown as the solid square in Figure 4.5). The coverage of 95% is relatively high (0.9832). The average bias in $\phi_{pq,k}^*$ is $-0.0909$.

HRF. An example of the estimated HRFs for the 3 ROIs are shown in Figure 4.6. The average coverage of 95% point-wise credible band for the HRF curves on the interval of $[0, 25]$ (sec.) is 0.6561, which is relatively low, however, the coverages of the 95% CIs for $T_{peak}$, $T_{wid}$ and $T_{under}$ are relatively high (0.9200, 0.9467, and 0.9000, respectively). The biases in $T_{peak}$, $T_{wid}$ and $T_{under}$ are relatively small ($-0.2933$, $-0.2673$, and $-0.8633$ seconds, respectively). The estimated HRFs are reasonably reliable, with the standard deviation being 0.6236, 1.0813, and 2.8677, respectively.

Simulation study 2: effects of different settings

Effects of magnitude of the latent signal. When the magnitude of the latent signal increases (Setting 2 versus Setting 1), AUC for connectivity increases; sensitivity increases but specificity decreases, when the threshold is fixed at $\lambda = 0.5$. However, using a little larger threshold, e.g., $\lambda = 0.7$, actually improves both quantities ($sens = 0.9263$, $spec = 0.8769$, not shown in the table). Coverage of 95% CI for the connectivity parameters $\phi_{pq,k}$ remains similar. Bias becomes a little smaller; this is expected, as the increase in the magnitude of $\beta_k(t)$ with the same observation covariance matrix increases the proportion of the informa-
Figure 4.6: Examples of estimated HRFs from a simulated data. Red curve: true HRF. Black curve: estimated HRF. Blue curves: 95% credible band. Red vertical lines: true $T_{\text{peak}}$ and $T_{\text{under}}$. Blue vertical lines: estimated $T_{\text{peak}}$ and $T_{\text{under}}$. Grey curves: posterior samples of HRF. To better compare the shapes, the HRFs are rescaled to the same height.
Figure 4.7: Example results for $\beta_k^{(r)}(t)$. Red: true latent signal. Blue and black dashed: estimated signal. Blue dashed: point-wise 95% credible bands.
tion about $\beta_k(t)$ contained in the observed signals. However, due to the same fact, both power and specificity for activation detection decreases, and the standard deviations of the HRF shape quantities increase. The other measurements remain similar.

Effects of amount of fMRI time series. The effects of increasing the amount of fMRI time series (Setting 3 versus Setting 1) are very similar to those of increasing the magnitude of the latent signal, in terms of connectivity; using a little larger threshold also helps improve both sensitivity and specificity ($sens = 0.9780$, $spec = 0.8806$). For a real data with total number of time series much larger than 128, if one wants to improve specificity with sensitivity maintained at a reasonable level, using a threshold a little bigger than $\lambda = .5$ could help. However, there is little effect on the performance for activation. For HRF, the standard deviations of the shape quantities decrease; however the coverage of 95% CIs decreases as well with little change in biases, indicating potential underestimation of variability in HRF.

Effects of fixing HRFs at the pre-estimated shapes. When the HRFs are fixed at the pre-estimated shapes during posterior sampling (Setting 4 versus Setting 1), sensitivity for connectivity becomes a little smaller. Power to detect activation remains similar, however, specificity for activation decreases, possibly due to the underestimation of the variability in $\Upsilon$. This is expected, because $\Upsilon$ is closely tied with HRF basis coefficients $d$ in the observation equation component of the model, and fixing $d$ results in the underestimation of variability in $\Upsilon$.

Effects of the prior variance $\Sigma_{\phi}$. When the prior variance for the slab part of connectivity parameter $\phi$ increases (from $\Sigma_{\phi} = I$ to $\Sigma_{\phi} = 4I$, Setting 5 versus Setting 1), the prior distribution for $\phi$ becomes flatter and therefore imposes less shrinkage towards 0 on the posterior distribution of $\phi$. Sensitivity for connectivity goes up a little but specificity decreases substantially. This indicates in here a flatter prior distribution on $\phi$ tends to be less conservative with the same threshold, i.e., it favors the presence of connectivity. Bias in $\phi$ decreases since there is less shrinkage toward 0 from the prior distribution. When
the prior variance decreases (from $\Sigma_{\phi} = I$ to $\Sigma_{\phi} = I/4$, Setting 6 versus Setting 1), the prior distribution for $\phi$ becomes more concentrated around the prior mean $0$ and therefore imposes more shrinkage towards $0$. Specificity for connectivity increases while sensitivity decreases, although not by much; here a more concentrated prior distribution on $\phi$ tends to be more conservative and favors the absence of connectivity with the same threshold. Bias in $\phi$ increases since there is more shrinkage toward $0$ from the prior distribution, and it also leads to smaller coverage of $95\%$ CI for $\phi$. Although increasing the prior variance to $\Sigma = 4I$ provides better sensitivity, the overall classification accuracy drops (from $AUC = 0.9380$ to $AUC = 0.8626$); while the overall classification accuracy is not affected when a smaller prior variance $\Sigma = I/4$ is used.

For activation, bias and coverage of $95\%$ CI are affected in a similar way as with connectivity; but power and specificity remain similar. For HRF, there do not seem to be obvious and systematic changes caused by the increase or decrease in the prior variance of $\phi$.

In conclusion, the model performs well in detecting activation, although the variability of the activation contrasts tends to be underestimated, thus leading to lower specificity for activation, i.e., ability to identify a ROI is not activated. The model also performs reasonably well in inferring the presence or absence of connectivity and estimating the connectivity strength under the threshold $\lambda = 0.5$. It also well captures the shapes of the HRFs.

The comparisons of the model performance under different settings suggest the follows. First, the overall classification accuracy for connectivity improves when the latent signal has a larger magnitude, with the side effect of decreasing the power to detect activation and increasing the variability in the estimation of HRF. Second, increasing the amount of observations can also improve the overall inference on connectivity with the main side effect of underestimating the variability in the estimated HRFs. Furthermore, fixing the HRFs at the pre-estimated shapes tends to decrease sensitivity for connectivity and underestimate the variability in the activation contrasts. For a real data, when the HRF results from the
full version of the model has high variability or has biologically unreasonable shapes, fixing the HRFs at the pre-estimated shapes could help. Finally, inference on connectivity could be sensitive to the choice of prior variance of the spike and slab prior in this single-subject state-space model. It seems that using a prior variance that matches the range of $\phi$ might be a reasonable choice. In some cases one might be able to know the range of $\phi$ from previous studies with similar experiments; however, if such information is not available, one can choose a relatively large or small $\Sigma_\phi$ depending on whether sensitivity or specificity is more important.

4.4 Application to fMRI data

4.4.1 Analysis

The proposed model is also applied to the fMRI data from a stroke patient in the stroke study by UCI Neurorehabilitation lab (PI: Cramer). The subject is right-handed, had stroke in the left brain hemisphere, still had residual motor deficits on the right side of the body during the study, and performed the hand grasp and release task using the right hand and also the dominant hand during the fMRI experiment. The fMRI design of the experiment is shown in Figure 1.5. More details can be found in Chapter 1. The data contains 144 time points in total, from 3 fMRI sessions. All five ROIs, LM1, RM1, LPMd, RPMd, and SMA, are included in the model. The fMRI data was preprocessed in the same way as in Chapter 2 and Chapter 3.

Prior 1 from the simulation study (Section 4.3.2) is used in for the fMRI data. A posterior sample of size 5,000 was computed for each parameter, after a burn-in period of 1,000 iterations. However, the estimated HRF shapes had a lot of oscillations and did not look biologically reasonable (see the black curves in Figure C.6 in Appendix C), when the full version of the model was applied; this was likely to be caused by the low signal to noise ratio.
in the fMRI data. In contrast, the estimated HRF shapes from the preliminary analysis appeared much more biologically meaningful (see the green curves in Figure C.6), i.e., consisting of a main rise followed by an relatively shallow undershoot. Therefore, the simplified version of the model was used instead and the HRFs were fixed at the estimates from the preliminary analysis described in 4.2.4. The estimation took about 30 min to run for the simplified model and about 60 min for the full version of the model.

4.4.2 Results

Here I show the model results of connectivity, activation and HRF for the stroke patient. The mean signal estimated by the model well fits the observed signal (see Figure C.5 in Appendix C). Connectivity. The connectivity network is shown in Figure 4.8. There is within-ROI connectivity in LM1 with $\phi_{pq,k} > 0$ $k = 1, 2$, and between-region connectivity from LM1 to SMA with $\phi_{pq,k} > 0$ $k = 1, 2$, and from LM1 to RM1 with $\phi_{pq,k} < 0$ $k = 1, 2$.
Figure 4.9: Estimated strength in the five ROIs. Black outline circle indicates the ROI is activated. The more red the background for a ROI is, the stronger the activation is.

Therefore, under both task condition and rest condition, the higher the BOLD amplitude in LM1 at the current time is, the higher the BOLD amplitude in LM1 and SMA at the next time point is expected to be, and the lower the BOLD amplitude in RM1 at the next time point is expected to be. This implies that there might be communications within LM1, from LM1 to SMA, and from LM1 to RM1 that compensate the damage in the left brain hemisphere, in order for the patient to perform the simple motor task of hand grasp and release movement.

**Activation.** The activation result is shown in Figure 4.9. Only LM1, SMA and RPMd are claimed to be activated by the model. The point estimates for the activation contrasts in LPMd and RM1 are positive, but are not significantly larger than 0 based on the 95% CI of $\Upsilon_{(c)}$. The model suggests the involvement of the secondary motor regions SMA and RPMd, besides the primary motor region LM1, is required for this patient to perform the simple motor task.
Figure 4.10: Estimated HRFs from preliminary analysis. Blue vertical lines indicate \( T_{\text{peak}} \) and \( T_{\text{under}} \).

HRF. The HRF shapes are shown in Figure 4.10. For this stroke patient, the HRFs in the ROIs in the left hemisphere where stroke occurred appear to peak 1 to 2 seconds later than the HRFs in the rest of the brain. This might be caused by the damage in the left hemisphere by stroke. Note that the HRF shapes are obtained from the preliminary analysis using a non-Bayesian approach. The credible bands are not provided here, but it is possible to approximately compute the credible bands in the following way: (1) estimate the variance covariance matrix of \( d_p \), \( \hat{\text{Var}} \, d_p \), through the observed Fisher information; (2) randomly generate a relatively large sample of \( d_p^{(b)} \)'s based on the asymptotic Gaussian distribution \( N(d_p, \hat{\text{Var}} \, d_p) \); (3) calculate the HRF samples by \( h_p^{(b)} = H d_p^{(b)} \); (4) obtain the empirical point-wise confidence band for the vector \( h_p \) based on \( h_p^{(b)} \).

Since this stroke patient was also analyzed by the hierarchical Bayesian vector autoregressive model in Chapter 3, it is also interesting to see the difference between the results from the two models. The results of activation, connectivity network and HRFs of the two models are listed together for comparison in Figure C.7, C.8, and C.9 in Appendix C. The activation results
from the two models are generally inline with each other. This is not surprising since they share the same activation modeling, the GLM. The connectivity results look quite different between the two models. The only overlap is the within-ROI connectivity in LM1, the region that is the most responsible for executing the experiment task. The state-space model found connectivity from LM1 to SMA and RM1, however, the VAR model only found within-ROI connectivity in SMA and LPMd. This is not surprising, since the two models adopt two different concepts of connectivity. However, which one is closer to the true connectivity at the neuronal level remains unclear, and needs further confirmation from studies using other neuroimaging modalities such as EEG and PET. For HRF, the estimated shapes from the hierarchical Bayesian VAR model are very similar to those from the preliminary analysis and used for the simplified state-space model, except that for LM1, the estimated HRF from the former model has a little delayed and wider peak compared to the latter model. This is also expected, since both models used the same activation modeling approach, the GLM, for estimating the shapes of HRF.

4.5 Extension to multi-subject modeling

The model here addresses only the single-subject fMRI data. As demonstrated in the previous chapter, fMRI data analysis could be a very powerful tool for studying effects of disease on human brain, when applied to multi-subject data. Below I describe how the proposed state-space model can be extended to multi-subject/multi-group analysis. Suppose there are $S$ subjects in total. Each subject $s$ ($s = 1, 2, \ldots, S$) belongs to one and only one of the group, $g$ ($g = 1, 2, \ldots, G$). For example for the stroke study, $g = 1$ (stroke), and $g = 2$ (healthy). Let $z^{(s)}$ be the vector of dummy variables representing the group for subject $^{(s)}$, i.e., $z^{(s)} = e_j$ means subject $s$ belongs to group $g$. $z^{(s)}$ can also be generalized to be any characteristics of the subjects, numerical or categorical.
The subject-level model stays the same. The prior distributions for the subject parameters become

\[ \xi_{pq}^{(s)} \sim \text{Bernoulli}(\pi_{pq}^{(s)}), \]
\[ \phi^{i.i.d.} \sim N(\mu_{\phi}^{(s)}, \text{diag}\{\sigma_{\phi,i}^2\}), \]
\[ \Phi^{(s)} \sim N(\mu_{\Phi}^{(s)}, \text{diag}\{\sigma_{\Phi,i}^2\}), \]
\[ d_{(s)}^{i.i.d.} \sim N(\mu_d, \Sigma_d), \]
\[ (V^{(s)})^{-1} \sim \text{Wishart}(\Psi_V, \nu_V), \]
\[ (Q^{(s)})^{-1} \sim \text{Wishart}(\Psi_Q, \nu_Q). \]

The parameters in the prior distributions are further modeled by the grouping factor (or other attributes of the subjects) \( z^{(s)}): \)

\[ \mu_{\Psi}^{(s)} \sim N(\mu_{\Psi} + z^{(s)} \gamma_{\Psi}, \frac{1}{S} \cdot \text{diag}\{\sigma_{\Psi,i}^2\})^{3P}, \]
\[ \text{logit}(\pi_{pq}) \sim N(\mu_{\pi, pq} + z^{(s)} \gamma_{\pi}, \sigma_{\pi}^2), \]
\[ \mu_{\phi}^{(s)} \sim N(\mu_{\phi} + z^{(s)} \gamma_{\phi}, \frac{1}{S} \cdot \text{diag}\{\sigma_{\phi,i}^2\})^{2P^2}, \]
\[ \sigma_{\phi,i}^{-1} \sim \text{Gam}(a_{\phi}, b_{\phi}), \ i = 1, 2, \ldots, 2P^2, \]
\[ \sigma_{\Psi,i}^{-1} \sim \text{Gam}(a_{\Psi}, b_{\Psi}), \ i = 1, 2, \ldots, 3P. \]

Additional prior distributions are required for the hyperparameters used above.

\[ \mu_{\Psi} \sim N(0, \Sigma_{\mu_{\Psi}}), \]
\[
\gamma_{\Upsilon,i} \sim 0.5\{0\} + 0.5N(0, \sigma_{\gamma,\Upsilon}^2),
\]
\[
\zeta_{\Upsilon} \sim \text{Gam}(a_{\zeta,\Upsilon}, b_{\zeta,\Upsilon}),
\]
\[
\mu_{\pi,pq} \sim N(0, \sigma_{\mu,\pi,pq}^2),
\]
\[
\gamma_{\pi} \sim 0.5\{0\} + 0.5N(0, \sigma_{\gamma,\pi}^2),
\]
\[
\mu_{\mu,\phi} \sim N(0, \Sigma_{\mu,\phi}),
\]
\[
\gamma_{\phi,i} \sim 0.5\{0\} + 0.5N(0, \sigma_{\gamma,\phi}^2),
\]
\[
\zeta_{\phi} \sim \text{Gam}(a_{\zeta,\phi}, b_{\zeta,\phi}).
\]

Here \(\Sigma_{\mu,\Upsilon}, \sigma_{\gamma,\Upsilon}^2, a_{\zeta,\Upsilon}, b_{\zeta,\Upsilon}, \sigma_{\mu,\pi,pq}^2, \sigma_{\gamma,\pi}^2, \Sigma_{\mu,\phi}, \sigma_{\gamma,\phi}^2, a_{\zeta,\phi}, \) and \(b_{\zeta,\phi}\) are pre-specified constants.

This hierarchical Bayesian modeling allows us to examine the effects of the subject covariate \(z^{(s)}\) on activation, connectivity structure, and connectivity strength, through \(\gamma_{\Upsilon}\), \(\gamma_{\pi}\), and \(\gamma_{\phi}\), respectively.

### 4.6 Conclusion

In short, the state-space model has shown the following advantages. (1) It allows the BOLD amplitudes to vary over time, which is more realistic than the VAR models. It could be useful for situations such as learning effect, due to which the activation strength extenuates over the course of the experiment [39]. (2) It models connectivity through the underlying dynamics of the BOLD signals, instead of through the noise term from GLM, and therefore is biologically more interpretable. (3) It takes into account the region-specific HRF, and thus different shapes of HRFs, especially different time to peak, is less likely to confound the temporal dependence in the BOLD signal. (4) It uses spike and slab priors to aid in inference on connectivity network, without having to estimate multiple models and conduct
model comparisons.

The proposed model still has some other limitations and can be improved in the future, besides the multi-subject model extension. (1) Currently the model only addresses lag 1 dependence in the state equation. It could be easily extended to higher order of lags by concatenating the dynamic BOiD amplitudes of multiple lags into one vector and using this multi-lag signal as the latent processes. The posterior sampling algorithm will be very similar to the current algorithm. (2) Due the downsampling limitation of fMRI technique, spurious connectivity may be inferred from low temporal resolution fMRI observations. In the future, VARMA structure, which performs more sensibly under downsampling, could be used instead of VAR structure to model the temporal dependence among ROIs. (3) Currently, the implementation for the proposed model is not sufficiently efficient in terms of computation time. Especially when the number of ROIs gets larger, the computation time could increase more than cubically with the number of ROIs, mainly due to the large matrix inversion in the calculation of the full conditional posterior covariance matrices of $d$, $\phi$ and $\Upsilon$ at each iteration. To improve computational efficiency, stochastic gradient descent MCMC approach [61] could be used to sample the posterior distributions of the unknown parameters conditional on the latent process $\beta^{(r)}(t)$, since it only needs the gradients of the parameters to calculate the new posterior sample at each iteration and thus avoids the large matrix inversions. Computation time could be further saved by partitioning the fMRI time series into smaller batches of time points and applying parallel computing techniques among these batches, such as the distributed stochastic gradient MCMC approach described in [1]. Moreover, variational inference could be developed for the model and used as an approximate but much faster implementation.
Chapter 5

Conclusion

5.1 Summary

Chapter 2 developed a novel vector autoregressive model for brain activation and connectivity in single-subject fMRI data. The simulation study has demonstrated the advantages of the proposed model: (1) It ruled out the confounding effects from HRF by including HRF as unknown parameters in the model. (2) The proposed model captured different effective connectivities across experimental conditions. (3) The use of spike and slab prior allows one to easily pick the best effective connectivity network without having to estimate multiple models. (4) It is able to easily incorporate relevant information derived from other studies but, unlike the dynamic causal modeling approach, it does not rely on biological assumptions that are difficult to verify from the data.

Based on the model in Chapter 2, Chapter 3 further develops a hierarchical Bayesian model for multi-subject fMRI data. Through simulation study, the approach has presented all the advantages of the single-subject model in Chapter 2. In addition, it is able to borrow information across subjects via hierarchical modeling to increase power for group comparison;
this is particularly relevant for data with low signal to noise ratio.

Chapter 4 developed a novel state space model that has shown its advantages in the following ways. (1) It allows the BOLD amplitudes to vary over time, which is more realistic. It could be useful for situations such as learning effect during the experiment. (2) It models connectivity through the underlying dynamics of the BOLD signals, instead of through the noise term from GLM, and therefore is biologically more interpretable. (3) It takes into account the region-specific HRF, and thus different shapes of HRFs, especially different time to peak, is less likely to confound the temporal dependence in the BOLD signal. (4) It uses spike and slab priors to aid in inference on connectivity network, without having to estimate multiple models and conduct model comparisons.

Application of all the proposed models to the stroke fMRI data also consistently found compensatory effect that requires the involvement of the primary motor region from the unaffected brain hemisphere and the secondary motor regions to aid in executing the simple motor task using the stroke-affected hand. This depicts the utility of the proposed Bayesian models in this dissertation.

5.2 Limitations and future work

There are still several limitations with the proposed models. (1) Computation for these models is still relatively heavy: 50 min for the multi-subject fMRI data analysis in Chapter 3 (27 subjects) and 30 min for the single-subject fMRI data analysis in Chapter 4. In the future, more efficient sampling algorithm needs to be developed and further parallel computing needs to be implemented. Also, for the state space model, variational Bayes method could be developed and used as an approximate but much faster implementation. (2) In the multi-subject VAR model, different groups are assumed to be independent and are
analyzed separately. In the future, grouping factors can be incorporated into the model. This may further help increase the power of detecting the differences between groups since I can borrow information from more subjects. (3) The current state space model only addresses connectivity with one lag. It could be extended to higher order of lags by concatenating the dynamic BOLD amplitudes of multiple lags into one vector and using this multi-lag signal as the latent variable. (4) Due the downsampling limitation of fMRI technique, spurious connectivity may be inferred from low temporal resolution fMRI observations. In the future, VARMA, which performs more sensibly under downsampling, could be used instead of VAR to model the temporal dependence among ROIs. Finally, I want to emphasize again that conclusions made from these models is based on the hemodynamics of the brain instead of the neuronal activities directly, so we need to be cautious when extending the conclusions from the hemodynamic level to the neuronal level.
Bibliography


[60] Wellcome Trust Center for Neuroimaging, UCL, 2009. SPM8 Software.


Appendix A

Appendix for Chapter 2

A.1 Detailed algorithm for posterior sample

We first define some notations:

- $\Theta$: collection of all the parameters involved in the model.
- $\Theta_{\Sigma_{\epsilon}}$: collection of all the parameters involved in $\Sigma_{\epsilon}$.
- $\Theta_{\Phi}$: collection of all the parameters involved in $\Phi_1$, $\Phi_2$.
- $\Theta_{HRF,p}$: collection of all the parameters involved in $h_p$.
- $Y$: collection of all the observations in the multivariate time series.
- Superscript ($i$): to denote iteration $i$. It will be omitted for simplicity of notation unless necessary for clarity.

Note that each element of $\Theta$, $\Theta_{\Sigma_{\epsilon}}$, $\Theta_{\Phi}$ and $\Theta_{HRF,p}$ takes its most current value.

Updating algorithm:
1. Sample $\beta|Y, \Theta \setminus \beta \sim N(\tilde{\beta}, \Sigma_{\beta})$, where $\tilde{\beta} = \left(\sum_{t=L+1}^{T}\tilde{y}(t)\Omega_c\tilde{X}(t)\right)'\Sigma_{\beta}$, $\Sigma_{\beta} = (\Sigma_{0,\beta}^{-1} + \Sigma_{\beta}^{-1})^{-1}$, and $\Sigma_{\beta}^{-1} = \sum_{t=L+1}^{T}\tilde{X}(t)\Omega_c\tilde{X}(t)'$. Update $M(t)$ and $u(t)$ accordingly.

2. Sample $\Sigma_{\epsilon}|Y, \Theta \setminus \Theta_{\Sigma_c} \sim \text{InvWishart}\left((T - L) - 1, \sum_{t=L+1}^{T}\epsilon(t)'\epsilon(t)\right)$ where $\epsilon(t) = \tilde{y}(t) - \beta\tilde{X}(t)$. Update $\Omega_{\epsilon}$ accordingly.

3. Define the index conversion function $J(p, q, \ell, k) = p + (q - 1)P + (\ell - 1)P^2 + (k - 1)P^2L$. Note that it is a bijection from $\{1, \ldots, P\} \times \{1, \ldots, P\} \times \{1, \ldots, L\} \times \{1, 2\}$ to $\{1, 2, \ldots, 2P^2L\}$. Define the row vector $b \in \mathbb{R}^{1 \times 2LP^2}$ such that its $J(p, q, \ell, k)$-th element is $\phi_k(\ell)_{p,q}$. Sample all $\phi_k(\ell)_{p,q}$'s simultaneously by sampling $b$: $b|Y, \Theta \setminus \Theta_a \sim N(\tilde{b}, \Sigma_b) \in \mathbb{R}^{1 \times 2LP^2}$, where

$$
\tilde{b} = \left(\sum_{t=L+1}^{T}u(t)\Omega_c\tilde{A}(t)\right)'\Sigma_b,$$

$$
\Sigma_b = (\Sigma_{0,b}^{-1} + \Sigma_{b}^{-1})^{-1},
$$

$$
\Sigma_{b}^{-1} = \sum_{t=L+1}^{T}\tilde{A}(t)\Omega_c\tilde{A}(t)',$$

$$
A(t) = \begin{bmatrix}
1_P \otimes \text{diag}\{c_1(t-1)u(t-1)\} \\
\vdots \\
1_P \otimes \text{diag}\{c_1(t-L)u(t-L)\} \\
1_P \otimes \text{diag}\{c_2(t-1)u(t-1)\} \\
\vdots \\
1_P \otimes \text{diag}\{c_2(t-L)u(t-L)\}
\end{bmatrix} \in \mathbb{R}^{2LP^2 \times P},
$$

and $\tilde{A}(t)$ is the same as $A(t)$ except that the elements corresponding to $\xi_{p,q}(\ell) = 0$ are set to 0, i.e., $\tilde{A}_{J(p,q,\ell,k), p}(t) = 0$, for all $(p, q, \ell)$ s.t. $\xi_{p,q}(\ell) = 0, k = 1, 2$. 

113
4. (a) Sample $\xi_{p,q}|Y, \Theta \setminus \xi_{p,q} \sim Multi(1, \pi_{p,q})$, where the $j$-th element of $\pi_{p,q}$

$$\tilde{\pi}_{(p,q),j} = \pi_{0,(p,q),j} \exp \left\{ -\frac{1}{2} \sum_{t=L+1}^{T} \left( u(t) - b_{(j)} A(t) \right) \Omega_{\epsilon} \left( u(t) - b_{(j)} A(t) \right) \right\},$$

$j = 0, 1, \ldots, L$. $b_{(j)}$ is equal to $b^*$ except that the $J(p,q,\ell,k)$-th element equals 0, for $\forall \ell = 1, 2, \ldots, j$, $k = 1, 2$, and that the $J(p,q,m,k)$-th element equals the $J(p,q,m,k)$-th element of $b$, for $\forall m = j + 1, \ldots, L$, $k = 1, 2$; and $b^*$ is equal to $b$ except that all the elements corresponding to $\xi_{p,q}(\ell)$ is 0 set to 0.

(b) For $\forall \ell$ s.t. $\xi_{p,q}(\ell) = 1$, update $b^*_{J(p,q,\ell,k)}$ to equal $b_{J(p,q,\ell,k)}$; for $\forall \ell$ s.t. $\xi_{p,q}(\ell) = 0$, where $k = 1, 2$, update $b^*_{J(p,q,\ell,k)}$ to equal 0. Note that when we update the Gibbs sample for $\xi_{p,q}$’s, it is preferred to update them in a random order [32], i.e., each time, randomly pick $(p,q)$ without replacement and update $\xi_{p,q}$ until all $P^2 \xi_{p,q}$’s have been updated. By then, the zeros and non-zeros of $b^*$ will have been fully determined, and thus $\Phi^*_k(\ell)$’s can be updated according to that $\phi^*_k(\ell)_{p,q}$ equals the $J(p,q,\ell,k)$-th element of $b^*$. Update $\tilde{y}(t)$, $\tilde{X}(t)$ accordingly.

5. Sample $\Theta_{HRF,p}$ using Metropolis-Hasting algorithm, looping from $p = 1$ to $P$.

(a) $a_{p1}$: sample $a'_{p1} \sim Unif \left( \max(1, a_{p1}^{(i-1)} - \delta_a), \min(a_{p1}^{(i-1)} + \delta_a, b_{p1}^{(i-1)} m_{p2} + 1) \right)$.

Define $\Theta'_{HRF,p}$ to be the same as $\Theta_{HRF,p}$ and $\Theta'$ to be the same as $\Theta$ except that the value of $a_{p1}$ is replaced with $a'_{p1}$. Calculate $p(\Theta_{HRF,p})$ as in Equation 2.4, and $L_c(\Theta')$ as in Equation 2.3. Calculate the acceptance rate $r = \min(1, \frac{f'_p q(a_{p1}^{(i-1)} | a'_{p1})}{f_p q(a_{p1}^{(i-1)} | a'_{p1})})$ where

$$f'_p = L_c(\Theta') p(\Theta'_{HRF,p})$$

is proportional to the full conditional posterior density for $a_{p1}$ being $a'_{p1}$,

$$q(a_{p1}^{(i-1)} | a'_{p1}) = \left( \max(1, a_{p1}^{(i-1)} + \delta_a, b_{p1}^{(i-1)} m_{p2} + 1) - \max(1, a_{p1}^{(i-1)} - \delta_a) \right)^{-1} \times$$

$$I \left( \max(1, a_{p1}^{(i-1)} - \delta_a) < a'_{p1} < \min(a_{p1}^{(i-1)} + \delta_a, b_{p1}^{(i-1)} m_{p2} + 1) \right),$$
(we used 0.5 for the step size $\delta_a$)

\[
q(a_{p1}^{(i-1)}|a_{p1}^i) = \left(\min(a_{p1}^i + \delta_a, \ b_{p1}^{(i-1)}m_{p2} + 1) - \max(1, \ a_{p1}^i - \delta_a)\right)^{-1} \times \\
I\left(\max(1, \ a_{p1}^i - \delta_a) < a_{p1}^{(i-1)} < \min(a_{p1}^i + \delta_a, \ b_{p1}^{(i-1)}m_{p2} + 1)\right).
\]

Note that $f_p$ is available from the previous iteration. Accept $a_{p1}^{(i)} = a_{p1}^i$ with probability $r$, and otherwise keep the old $a_{p1}$ value so that $a_{p1}^{(i)} = a_{p1}^{(i-1)}$. Update $\tilde{X}(t)$ accordingly; update $m_{p1}$ to be $(a_{p1}^{(i)} - 1)/b_{p1}^{(i-1)}$, and $f_p$ to be $f_p'$.

(b) $b_{p1}$: sample $b_{p1}' \sim Unif\left(\max(b_{p1}^{(i-1)} - \delta_b, \ (a_{p1}^{(i)} - 1)/m_{p2}), \ b_{p1}^{(i-1)} + \delta_b\right)$; define $\Theta'_{HRF,p}$ to be the same as $\Theta_{HRF,p}$ and $\Theta'$ to be the same as $\Theta$ except that the value of $b_{p1}$ is replaced with $b_{p1}'$; similarly as before but with $a_{p1}$ replaced, calculate the new $f_p'$ and $r = \min(1, \ f_p'q(b_{p1}^{(i-1)}|b_{p1}^i))$ where

\[
q(b_{p1}'|b_{p1}^{(i-1)}) = \left(b_{p1}' + \delta_b - \max(b_{p1}' - \delta_b, \ (a_{p1}^{(i)} - 1)/m_{p2})\right)^{-1} \times \\
I\left(\max(b_{p1}' - \delta_b, \ (a_{p1}^{(i)} - 1)/m_{p2}) < b_{p1}' < b_{p1}^{(i-1)} + \delta_b\right),
\]

(we used 0.025 for the step size $\delta_b$)

\[
q(b_{p1}^{(i-1)}|b_{p1}') = \left(b_{p1}' + \delta_b - \max(b_{p1}' - \delta_b, \ (a_{p1}^{(i)} - 1)/m_{p2})\right)^{-1} \times \\
I\left(\max(b_{p1}' - \delta_b, \ (a_{p1}^{(i)} - 1)/m_{p2}) < b_{p1}' < b_{p1}^{(i-1)} + \delta_b\right);
\]

accept $b_{p1}^{(i)} = b_{p1}'$ with probability $r$, and otherwise $b_{p1}^{(i)} = b_{p1}^{(i-1)}$. Update $\tilde{X}(t)$ accordingly; update $m_{p1}$ to be $(a_{p1}^{(i)} - 1)/b_{p1}'$, and $f_p$ to be $f_p'$.

(c) $a_{p2}$: sample $a_{p2}' \sim Unif\left(\max(1, \ a_{p2}^{(i-1)} - \delta_a, \ b_{p2}^{(i-1)}m_{p1} + 1), \ a_{p2}^{(i-1)} + \delta_a\right)$; define $\Theta'_{HRF,p}$ and $\Theta'$ similarly as before but with $a_{p2}$ replaced; calculate the new $f_p'$ and $r = \min(1, \ f_p'q(a_{p2}^{(i-1)}|a_{p2}^i))$ where

\[
q(a_{p2}'|a_{p2}^{(i-1)}) = \left(a_{p2}'^{(i-1)} + \delta_a - \max(1, \ a_{p2}'^{(i-1)} - \delta_a, \ b_{p2}'^{(i-1)}m_{p1} + 1)\right)^{-1} \times \\
I\left(\max(1, \ a_{p2}' - \delta_a, \ b_{p2}'^{(i-1)}m_{p1} + 1) < a_{p2}' < a_{p2}^{(i-1)} + \delta_a\right),
\]

115
\[
q(a_{p2}^{(i-1)} | a_{p2}') = \left( a_{p2}' + \delta_a - \max(1, a_{p2}' - \delta_a, b_{p2}^{(i-1)} m_{p1} + 1) \right)^{-1} \times
I \left( \max(1, a_{p2}' - \delta_a, b_{p2}^{(i-1)} m_{p1} + 1) < a_{p2}^{(i-1)} < a_{p2}' + \delta_a \right);
\]

accept \(a_{p2}^{(i)} = a_{p2}'\) with probability \(r\), and otherwise \(a_{p2}^{(i)} = a_{p2}^{(i-1)}\). Update \(\bar{X}(t)\) accordingly; update \(m_{p2}\) to be \((a_{p2}^{(i)} - 1)/b_{p2}^{(i-1)}\), and \(f_p\) to be \(f_p'\).

(d) \(b_{p2}\): sample \(b_{p2}' \sim Unif \left( \max(b_{p2}^{(i-1)} - \delta_b, 0), \ \min(b_{p2}^{(i-1)} + \delta_b, (a_{p2}^{(i)} - 1)/m_{p1}) \right)\); define \(\Theta'_{HRF,p}\) and \(\Theta'\) similarly as before but with \(b_{p2}\) replaced; calculate the new \(f_p'\) and \(r = \min(1, \frac{f_p' q(b_{p2}'|b_{p2}^{(i-1)})}{f_p q(b_{p2}'|b_{p2}^{(i-1)})})\) where

\[
q(b_{p2}'|b_{p2}^{(i-1)}) = \left( \min(b_{p2}^{(i-1)} + \delta_b, (a_{p2}^{(i)} - 1)/m_{p1}) - \max(b_{p2}^{(i-1)} - \delta_b, 0) \right)^{-1} \times
I \left( \max(b_{p2}^{(i-1)} - \delta_b, 0) < b_{p2}' < \min(b_{p2}^{(i-1)} + \delta_b, (a_{p2}^{(i)} - 1)/m_{p1}) \right);
\]

\[
q(b_{p2}^{(i-1)} | b_{p2}') = \left( \min(b_{p2}' + \delta_b, (a_{p2}' - 1)/m_{p1}) - \max(b_{p2}' - \delta_b, 0) \right)^{-1} \times
I \left( \max(b_{p2}' - \delta_b, 0) < b_{p2}^{(i-1)} < \min(b_{p2}' + \delta_b, (a_{p2}' - 1)/m_{p1}) \right);
\]

accept \(b_{p2}^{(i)} = b_{p2}'\) with probability \(r\), and otherwise \(b_{p2}^{(i)} = b_{p2}^{(i-1)}\). Update \(\bar{X}(t)\) accordingly; update \(m_{p1}\) to be \((a_{p2}^{(i)} - 1)/b_{p2}^{(i-1)}\), and \(f_p\) to be \(f_p'\).

(e) \(c_{p2}\): sample \(c_{p2}' \sim Unif \left( \max(c_{p2}' - \delta_c, 0), \ c_{p2}' + \delta_c \right)\); define \(\Theta'_{HRF,p}\) and \(\Theta'\) similarly as before but with \(c_{p2}\) replaced; calculate the new \(f_p'\) and \(r = \min(1, \frac{f_p' q(c_{p2}'|c_{p2}^{(i-1)})}{f_p q(c_{p2}'|c_{p2}^{(i-1)})})\) where

\[
q(c_{p2}'|c_{p2}^{(i-1)}) = \left( c_{p2}^{(i-1)} + \delta_c - \max(c_{p2}^{(i-1)} - \delta_c, 0) \right)^{-1} \times
I \left( \max(c_{p2}^{(i-1)} - \delta_c, 0) < c_{p2}' < c_{p2}^{(i-1)} + \delta_c \right);
\]

\[
q(c_{p2}^{(i-1)}|c_{p2}') = \left( c_{p2}' + \delta_c - \max(c_{p2}' - \delta_c, 0) \right)^{-1} \times
I \left( \max(c_{p2}' - \delta_c, 0) < c_{p2}^{(i-1)} < c_{p2}' + \delta_c \right);
\]

(we used 0.025 for the step size \(\delta_c\))
accept $c_{p2}^{(i)} = c_{p2}'$ with probability $r$, and otherwise $c_{p2}^{(i)} = c_{p2}^{(i-1)}$. Update $\tilde{X}(t)$ accordingly; update $f_p$ to be $f_p'$.

(f) Update $x_{pk}(t)$ accordingly.

A.2 Parameter values for simulation

**HRF:**

ROI 1 & 2: $a_{p1} = 6$, $b_{p1} = 1$, $a_{p2} = 16$, $b_{p2} = 1$, $c_{p2} = 1/6$, $p = 1, 2$;

ROI 3: $a_{31} = 9$, $b_{31} = 1$, $a_{32} = 21$, $b_{32} = 1$, $c_{32} = .4$;

ROI 4: $a_{41} = 6.6$, $b_{41} = 1$, $a_{42} = 18.4$, $b_{42} = 1$, $c_{42} = .5$.

$\beta_{pk}$:

ROI 1: $\beta_{10} = 0$, $\beta_{11} = 231.70$, $\beta_{12} = 115.36$;

ROI 2: $\beta_{20} = 0$, $\beta_{21} = 289.03$, $\beta_{22} = 289.03$;

ROI 3: $\beta_{30} = 0$, $\beta_{31} = 237.24$, $\beta_{32} = 23.81$;

ROI 4: $\beta_{40} = 0$, $\beta_{41} = 203.90$, $\beta_{42} = 56.51$.

$\Phi_1(1)$:

$$
\begin{bmatrix}
0 & 1.141 & 0 & 1.409 \\
0 & 0 & 0 & 0 \\
0 & 0.838 & 0 & 0 \\
0 & 0.167 & 0.165 & 0.244
\end{bmatrix}
$$

$\Phi_2(1)$:
\[
\begin{bmatrix}
0 & 0.141 & 0 & 0.223 \\
0 & 0 & 0 & 0 \\
0 & -0.303 & 0 & 0 \\
0 & 0.083 & 0.146 & -0.067
\end{bmatrix}
\]

\[
\Omega_\epsilon:
\begin{bmatrix}
1.63E-2 & 2.53E-2 & 0 & -3.08E-3 \\
2.53E-2 & 6.99E-2 & 0 & -9.67E-3 \\
0 & 0 & 8.51E-3 & 1.38E-3 \\
-3.08E-3 & -9.67E-3 & 1.38E-3 & 1.03E-3
\end{bmatrix}
\]

Or equivalently, \(\Sigma_\epsilon:\)
\[
\begin{bmatrix}
140.45 & -51.66 & 1.08 & -6.68 \\
-51.66 & 35.50 & -2.97 & 18.35 \\
1.08 & -2.97 & 120.51 & -18.69 \\
-6.68 & 18.35 & -18.69 & 115.31
\end{bmatrix}
\]
A.3 Supplementary figures

Figure A.1: Trace plot of unnormalized log posterior density.

Figure A.2: ACF of estimated $\epsilon$ from fMRI data analysis.
Figure A.3: PACF of estimated $\epsilon$ from fMRI data analysis.
Appendix B

Appendix for Chapter 3

B.1 Supplementary figures

Figure B.1: (a) Half-cosine parameterization of the HRF. (b) Simulated HRFs.
Figure B.2: (a) Variability explained by the basis vectors; (b) HRF basis set; (c) HRFs generated by the constrained linear basis.

Figure B.3: Comparison of activation results for models using $L = 2$ and $L = 1$, for both groups.
(a) lag 1, $L = 2$  
(b) lag 2, $L = 2$  
(c) lag 1, $L = 1$

Figure B.4: Comparison of connectivity results for models using $L = 2$ and $L = 1$ for the healthy group. Only the point estimate of connectivity are shown here.

(a) lag 1, $L = 2$  
(b) lag 2, $L = 2$  
(c) lag 1, $L = 1$

Figure B.5: Comparison of connectivity results for models using $L = 2$ and $L = 1$ for the stroke group.
Figure B.6: HRF results for models using $L = 2$ and $L = 1$, for both groups.

B.2 Full conditional posterior distributions

Define notation $v(i_1 : i_2)$ to be the subset of vector $v$ from elements $i_1$ to $i_2$. Subscript “post” is used to indicate the quantity is based on posterior distribution. Without loss of generality, I only need to list the posterior distributions for a single group. For notational convenience, I drop all the superscripts for group. Subject level:

\[ \beta^{(s)} \sim N(\mu_{\beta,post}^{(s)}, \Sigma_{\beta,post}^{(s)}) \]  
(B.1)

\[ \phi^{(s)} \sim N(\phi_{\phi,post}^{(s)}, \Sigma_{\phi,post}^{(s)}) \]  
(B.2)

\[ d^{(s)} \sim N(\mu_{d,post}^{(s)}, \Sigma_{d,post}^{(s)}) \]  
(B.3)

\[ \xi_{pq}^{(s)} \sim \text{Multinomial}(1, \pi_{pq,post}^{(s)}) \]  
(B.4)

\[ \Omega^{(s)} \sim \text{Wishart}(\Omega_{\Omega,post}^{(s)}, \nu_{\Omega,post}^{(s)}) \]  
(B.5)
where

\[ \Sigma_{\beta, post}(s) = (R \sum_{t=L+1}^{T} (\tilde{X}^{(s)}(t))' \Omega^{(s)} \tilde{X}^{(s)}(t) + (\Sigma_{\beta}^{q})^{-1})^{-1} \]  \hspace{1cm} (B.6) \\
\[ \mu_{\beta, post}^{(s)} = \Sigma_{\beta, post}^{(s)} \sum_{r=1}^{T} \sum_{t=L+1}^{T} (\tilde{X}^{(s)}(t))' \Omega^{(s)} \tilde{y}^{(s,r)}(t) + (\Sigma_{\beta}^{q})^{-1} \mu_{\beta}^{q} \]  \hspace{1cm} (B.7) \\
\[ \Sigma_{d, post}^{(s)} = (R \sum_{t=L+1}^{T} (\tilde{Z}^{(s)}(t))' \Omega^{(s)} \tilde{Z}^{(s)}(t) + (\Sigma_{d}^{q})^{-1})^{-1} \]  \hspace{1cm} (B.8) \\
\[ \mu_{d, post}^{(s)} = \Sigma_{d, post}^{(s)} \sum_{r=1}^{T} \sum_{t=L+1}^{T} (\tilde{Z}^{(s)}(t))' \Omega^{(s)} \tilde{y}^{(s,r)}_{0}(t) + (\Sigma_{d}^{q})^{-1} \mu_{d}^{q} \]  \hspace{1cm} (B.9) \\
\[ \Sigma_{\phi, post}^{(s)} = (\sum_{r=1}^{T} \sum_{t=L+1}^{T} (\tilde{W}^{(s,r)}(t))' \Omega^{(s)} \tilde{W}^{(s,r)}(t) + (\Sigma_{\phi}^{q})^{-1})^{-1} \]  \hspace{1cm} (B.10) \\
\[ \mu_{\phi, post}^{(s)} = \Sigma_{\phi, post}^{(s)} \sum_{r=1}^{T} \sum_{t=L+1}^{T} (\tilde{W}^{(s,r)}(t))' \Omega^{(s)} \tilde{u}^{(s,r)}(t) + (\Sigma_{\phi}^{q})^{-1} \mu_{\phi}^{q} \]  \hspace{1cm} (B.11) \\
\[ \pi_{pq, post}^{(s)}(j) \propto \pi_{pq}^{q}(j) I(\xi_{pq}^{q} = \ell_{j}) \times \exp\left\{ -\frac{1}{2} \sum_{r=1}^{R} \sum_{t=L+1}^{T} (\tilde{u}^{(s,r)}(t) - W^{(s,r)}(t) \phi^{(s)}_{(pq,j)})' \right. \] \\
\[ \left. \times \Omega^{(s)}(\tilde{u}^{(s,r)}(t) - W^{(s,r)}(t) \phi^{(s)}_{(pq,j)}) \right\} \] \\
\[ j = 0, \ldots, L; \text{ and } \sum_{j=0}^{L} \pi_{pq, post}^{(s)}(j) = 1 \]  \hspace{1cm} (B.12) \\
\[ \Omega_{post}^{(s)} = (SSLE^{(s)} + (\Omega^{q})^{-1})^{-1} \]  \hspace{1cm} (B.13) \\
\[ \nu_{\Omega, post}^{(s)} = R \times (T - L) + \nu_{\Omega} \]  \hspace{1cm} (B.14) \\
\[ SSE^{(s)} = \sum_{r=1}^{R} \sum_{t=L+1}^{T} (\tilde{y}^{(s,r)}(t) - \tilde{X}^{(s)}(t) \theta^{(s)})(\tilde{y}^{(s,r)}(t) - \tilde{X}^{(s)}(t) \theta^{(s)})' \]  \hspace{1cm} (B.15) \\

Group level:

\[ \mu_{\beta}^{0} \sim N(\mu_{\mu_{\beta, post}}, \Sigma_{\mu_{\beta, post}}) \]  \hspace{1cm} (B.16) \\
\[ \mu_{\phi}^{0} \sim N(\mu_{\mu_{\phi, post}}, \Sigma_{\mu_{\phi, post}}) \]  \hspace{1cm} (B.17)
\[ \mu_{g}^{\beta} \sim N(\mu_{g,\text{post}}, \Sigma_{g,\text{post}}) \] (B.20)

\[ (\Sigma_{g}(i, i))^{-1} \sim \text{Gamma}(a_{g,\text{post}}, b_{g,\text{post}}), \quad i = 1, \ldots, P \] (B.21)

\[ (\Sigma_{g}(i, i))^{-1} \sim \text{Gamma}(a_{\phi,i,\text{post}}, b_{\phi,i,\text{post}}), \quad i = 1, \ldots, 2LP^{2} \] (B.22)

\[ (\Sigma_{g}(i, i))^{-1} \sim \text{Gamma}(a_{d,i,\text{post}}, b_{d,i,\text{post}}), \quad i = 1, \ldots, PJ \] (B.23)

\[ \pi_{pq}^{\beta} \sim \text{Dir}(\alpha_{\pi,\text{post}}) \] (B.24)

\[ (\Omega^{g})^{-1} \sim \text{Wishart}(\Sigma_{\text{post}}, \nu_{\text{post}}) \] (B.25)

where

\[
\begin{align*}
\mu_{g,\text{post}} &= \Sigma_{g,\text{post}}(\Sigma_{\beta}^{-1} \sum_{s=1}^{S} \beta^{(s)} + \Sigma_{\mu,0}^{-1} \mu_{\mu,0}) \\
\Sigma_{g,\text{post}} &= (S\Sigma_{\beta}^{-1} + \Sigma_{\mu,0}^{-1})^{-1} \\
\mu_{\phi,\text{post}} &= \Sigma_{\phi,\text{post}}(\Sigma_{\phi}^{-1} \sum_{s=1}^{S} \phi^{(s)} + \Sigma_{\mu,0}^{-1} \mu_{\mu,0}) \\
\Sigma_{\phi,\text{post}} &= (S\Sigma_{\phi}^{-1} + \Sigma_{\mu,0}^{-1})^{-1} \\
\mu_{d,\text{post}} &= \Sigma_{d,\text{post}}(\Sigma_{d}^{-1} \sum_{s=1}^{S} d^{(s)} + \Sigma_{\mu,0}^{-1} \mu_{\mu,0}) \\
\Sigma_{d,\text{post}} &= (S\Sigma_{d}^{-1} + \Sigma_{\mu,0}^{-1})^{-1} \\
a_{g,\text{post}} &= \frac{1}{2} S + a_{g}, \quad b_{g,\text{post}} = \frac{1}{2} \sum_{s=1}^{S} (\beta^{(s)})^{2} + b_{g} \\
a_{\phi,i,\text{post}} &= \frac{1}{2} S + a_{\phi}, \quad b_{\phi,i,\text{post}} = \frac{1}{2} \sum_{s=1}^{S} (\phi^{(s)})^{2} + b_{\phi} \\
\alpha_{\pi,\text{post}}(j) &= \sum_{s=1}^{S} I(\xi^{(s)}_{pq} = \ell(j)) + \alpha_{\pi} \\
\Sigma_{\text{post}} &= (\sum_{s=1}^{S} \Omega^{(s)} + \Sigma_{0}^{-1})^{-1}, \quad \nu_{\text{post}} = S\nu_{\Omega} + \nu_{0}
\end{align*}
\]
B.3 Sampling algorithm

1. Initialize:

(a) Set the initial values for group-level parameters: $\Sigma_{\phi}(i, i)$ and $\Sigma_{\beta}(i, i)$ set to relatively large values, $\Omega^{g} = I$.

(b) Initialize the subject-level parameters: initialize $\beta^{(s)}$ and $d^{(s)}$ based on the point estimates from the preliminary analysis. Apply VAR($L$) on the resulting residuals and calculate the point estimates $\hat{\phi}^{(s)}$ and $\hat{\Omega}^{(s)}$; initialize $\phi^{(s)}$ and $\xi^{pq}_{\ell}(\ell)$ as $[\hat{\phi}^{(s)}, \hat{\phi}^{(s)}]$ and $\ell_j$ s.t. $\ell = argmax_{0 \leq j' \leq L} \{\phi^{pq}_{\ell}(\ell) < 0.1, \forall \ell = 1, \ldots, j'\}$, and initialize $\Omega^{(s)}$ as $\hat{\Omega}$.

2. Sample the subject parameters for each subject $s$ separately: based on the latest value of all other parameters,

(a) Sample from posterior (B.2), (B.4). Update $\phi^{(s)}$, $\tilde{y}^{(s,r)}(t)$ and $\tilde{Z}^{(s)}(t)$ according to Equation (3.5), (3.10), and (3.13) in the paper.

(b) Sample from posterior (B.3) and rescale $d^{(s)}$ s.t. $\sum_t H(t,: \cdot) d^{(s)} = 1$. Update $\tilde{X}^{(s)}(t)$ according to Equation (3.12) in the paper.

(c) Sample from posterior (B.1). Update $\tilde{y}^{(s,r)}(t)$, $U^{(s,r)}(t)$, $\tilde{W}^{(s,r)}(t)$ and $SSE^{(s)}$ according to Equation (3.11), (3.16), and (3.17) in the paper, and Equation (B.17).

(d) Sample from posterior (B.5).

3. Refine the initial values for the subject parameters by repeating only Step 2 for a certain amount of iterations (pilot estimation), and continue to the next steps (formal estimation) with the latest iteration.

4. Aggregate the latest information across subjects and update for group level: Sample from posterior (B.18), (B.19), (B.21), (B.22), (B.19), (B.25).
5. Repeat Step 2 for all the subjects for 5 iterations, and then repeat Step 4.

6. Repeat the previous step until the chain reaches the pre-specified size (e.g. 5,000) after a certain burn-in period of Step 5.

B.4 Parameter values used for simulation

- Mean and covariance matrix for generating $\beta^{(s)}$:

$$
\mathbb{E}\beta = [27, 3, 18, 30, 33; 15, 3, 23, 4, 26; 0_5]' \\
\text{Cov}(\beta_{pk}, \beta_{qj}) = \\
\begin{cases} 
3.556, & \text{if } p = q, p \neq 2, k = j \\
0.5000, & \text{if } p = q = 2, k = j \\
\frac{1}{5}\sqrt{\text{Cov}(\beta_{pk}, \beta_{pk}) \text{Cov}(\beta_{pj}, \beta_{pj})}, & \text{if } p = q, k \neq j \\
0, & \text{if } p \neq q 
\end{cases}
$$

- Covariance matrix for generating $\phi^{(s)}$: $0.05^2I_{2LP^2}$

- HRF basis coefficient $d^{(s)}$:

$$
\mathbb{E}d_1^{(s)} = [1.9783, 2.4983, 1.7715, -0.7298, 0.3259] \\
\mathbb{E}d_2^{(s)} = [2.5797, -0.2253, 0.1054, 0.5102, -0.1120] \\
\mathbb{E}d_3^{(s)} = \mathbb{E}d_4^{(s)} = [2.4341, 0.4343, 0.5088, 0.2099, -0.0059] \\
\mathbb{E}d_5^{(s)} = [2.7224, -0.8712, -0.2898, 0.8043, -0.2158] \\
\text{Cov}(d_p) = \text{diag}\{0.05587, 0.2529, 0.1547, 0.1152, 0.04066\} \\
d_p \perp d_q, \ p \neq q
$$
• Covariance matrix of $\text{e}$ (the same across subjects)

\[
(\Sigma^{(s)})^{-1} = \Omega^{(s)} = \begin{bmatrix}
0.0625 & -0.00117 & 0 & 0 & 0.00117 \\
-0.00117 & 0.0625 & 0 & 0 & 0 \\
0 & 0 & 0.0625 & 0 & 0 \\
0 & 0 & 0 & 0.0625 & 0 \\
0.00117 & 0 & 0 & 0 & 0.0625 \\
\end{bmatrix}
\]

## B.5 VB algorithm

I first introduce some notations. Let $\odot$ denote element-wise matrix or vector multiplication, and $\text{Bdiag}\{\mathbf{x}\}$ denote $\text{Bdiag}\{\mathbf{x}, \mathbf{x}\}$. $\tilde{\mu}^{(s)}_{\phi,k}(\ell)$ denotes the matrix formed by the corresponding elements in the vector $\tilde{\mu}^{(s)}_{\phi}$ for lag $\ell$ and condition $k$; $\tilde{\mu}^{(s)}_{\xi_{pq}(\ell)}$ denotes $\mathbb{E}_{q}[\xi_{pq}(\ell); \tilde{\mu}^{(s)}_{\xi}]$; $\tilde{\Sigma}^{(s)}_{\xi_{pq}(\ell)}$ denotes $\text{Cov}_{q}(\xi_{pq}(\ell), \xi_{pq'}(\ell'))$; $\tilde{\mu}^{(s)}_{\xi}$ and $\tilde{\Sigma}^{(s)}_{\xi}$ denotes the mean and the covariance matrix under $q(\cdot)$ for $\xi^{(s)} = [\text{vec}{\xi}^{(s)}(1)], \ldots, \text{vec}{\xi}^{(s)}(L)]'$ and they are formed by the elements of $\tilde{\mu}^{(s)}_{\xi}(\ell)$ and $\tilde{\Sigma}^{(s)}_{\xi_{pq}(\ell):\xi_{pq'}(\ell')}$; $\tilde{\mu}^{(s)}_{\beta}$ denotes the element in $\tilde{\mu}^{(s)}_{\beta}$ for ROI $p$ and condition $k$; $\tilde{\mu}^{(s)}_{\beta,0}$ denotes the element in $\tilde{\mu}^{(s)}_{\beta}$ for all the intercepts; $\tilde{\mu}^{(s)}_{d,p}$ denotes the elements in $\tilde{\mu}^{(s)}_{d}$ for ROI $p$; $\mathbf{i}_{\ell}$ denotes the index vector $[1, \ldots, \ell, L+1, \ldots, L+\ell]$; $\mathbf{i}_{pq}^{\phi}$ denotes the index vector $[p+(q-1)P+(1-1)P^{2}, \ldots, p+(q-1)P+(L-1)P^{2}, p+(q-1)P+(1-1)P^{2}+LP^{2}, \ldots, p+(q-1)P+(L-1)P^{2}+LP^{2}']$ and $-\mathbf{i}_{pq}^{\phi}$ denotes the complement of $\mathbf{i}_{pq}^{\phi}$ with respective to 1 : $(2LP^{2})$; $\mathbf{i}_{pq}^{\xi}$ denotes the index vector $[p+(q-1)P+(1-1)P^{2}, \ldots, p+(q-1)P+(L-1)P^{2}]'$ and $-\mathbf{i}_{pq}^{\xi}$ denotes the complement of $\mathbf{i}_{pq}^{\xi}$ with respective to 1 : $(LP^{2})$; $\mathbf{x}(\mathbf{i})$ denotes the vector formed by $\mathbf{x}$’s elements ($\mathbf{i} = [\mathbf{i}(1), \mathbf{i}(2), \ldots, \mathbf{i}(n)]$); $A(:, \mathbf{i})$ (or $A(\mathbf{i}, :)$) denotes the matrix formed by $A$’s $\mathbf{i}$(1)-th, $\mathbf{i}$(2)-th, ..., and $\mathbf{i}$(n)-th columns (or rows); $\tilde{\mu}_{\xi,pq}$ denotes $\tilde{\mu}_{\xi}(\mathbf{i}_{pq})$, and $\tilde{\mu}_{\xi,-pq}$ denotes $\tilde{\mu}_{\xi}(-\mathbf{i}_{pq})$.

The VB algorithm iteratively updates $\tilde{\mu}^{(s)}_{\beta}$, $\tilde{\Sigma}^{(s)}_{\beta}$, $\tilde{\mu}^{(s)}_{d}$, $\tilde{\Sigma}^{(s)}_{d}$, $\tilde{\mu}^{(s)}_{\phi}$, $\tilde{\Sigma}^{(s)}_{\phi}$, $\tilde{\mu}_{pq}^{(s)}$, $\tilde{\nu}_{\Omega}^{(s)}$ and $\tilde{\mu}_{\Omega}^{(s)}$ until
convergence, according to the following equations:

\[
\Phi_t^{(s)}(B) = I_P - \sum_{\ell=1}^{L} \sum_{k=1}^{2} c_k(t-\ell)\tilde{\Phi}_{\Phi,k}(\ell) \otimes \tilde{\mu}_{\xi,\ell}^{(s)}(\ell)B^\ell
\]

\[
\tilde{Z}^{(s)}(t) = \Phi_t^{(s)}(B)[\text{Bdiag}\{\sum_{k=1}^{2}\Lambda_k(t,:)\tilde{\mu}_{\beta,1k}^{(s)}; \ldots; \sum_{k=1}^{2}\Lambda_k(t,:)\tilde{\mu}_{\beta,pk}^{(s)}\}]
\]

\[
\tilde{y}_0^{(s,r)}(t) = \Phi_t^{(s)}(B)[y^{(s,r)}(t) - \tilde{\mu}_{\beta,0}]
\]

\[
\tilde{\Sigma}_d = (R \sum_{t=1}^{T} \tilde{Z}^{(s)}(t)'\tilde{\mu}_{\Omega}^{(s)}\tilde{Z}^{(s)}(t) + (\Sigma_d^{-1})^{-1}
\]

\[
\tilde{\mu}_d = \tilde{\Sigma}_d(R \sum_{r=1}^{R} \sum_{t=1}^{T} \tilde{\mu}_{\Omega}^{(s)}\tilde{y}_0^{(s,r)}(t) + (\Sigma_d^{-1})\mu_d)
\]

\[
X^{(s)}(t) = \text{Bdiag}\{1, \Lambda_1(t,:), \Lambda_2(t, :)\tilde{\mu}_{d,1}, \ldots, 1, \Lambda_1(t,:)\tilde{\mu}_{d,p}, \Lambda_2(t,:)\tilde{\mu}_{d,p}\}
\]

\[
\tilde{X}^{(s)}(t) = \Phi_t^{(s)}(B)[X^{(s)}(t)]
\]

\[
\tilde{y}^{(s,r)}(t) = \Phi_t^{(s)}(B)[y^{(s,r)}(t)]
\]

\[
\tilde{\Sigma}_d = (R \sum_{t=1}^{T} \tilde{X}^{(s)}(t)'\tilde{\mu}_{\Omega}^{(s)}\tilde{X}^{(s)}(t) + (\Sigma_d^{-1})^{-1}
\]

\[
\tilde{\mu}_d = \tilde{\Sigma}_d(R \sum_{t=1}^{T} \tilde{X}^{(s)}(t)'\tilde{\mu}_{\Omega}^{(s)}\tilde{y}^{(s,r)}(t) + (\Sigma_d^{-1})\mu_d)
\]

\[
\tilde{u}^{(s,r)}(t) = y^{(s,r)}(t) - X^{(s)}(t)\tilde{\mu}_d
\]

\[
W_t^{(s,r)} = [c_1(t-1)\tilde{u}^{(s,r)}(t)' \otimes I_P, \ldots, c_1(t-L)\tilde{u}^{(s,r)}(t)' \otimes I_P;
\]

\[
c_2(t-1)\tilde{u}^{(s,r)}(t)' \otimes I_P, \ldots, c_2(t-L)\tilde{u}^{(s,r)}(t)' \otimes I_P]
\]

\[
\tilde{\Sigma}_d^{(s)} = (R \sum_{t=1}^{T} (\tilde{W}_t^{(s,r)})'\tilde{\mu}_{\Omega}^{(s)}\tilde{W}_t^{(s,r)} + \text{Bdiag}_2(\tilde{\mu}_{\xi}^{(s)})' + \tilde{\Sigma}_d^{(s)}) + (\Sigma_d^{-1})^{-1}
\]

\[
\tilde{\mu}_d^{(s)} = \tilde{\Sigma}_d^{(s)}(\text{diag}\{\tilde{\mu}_{\xi}^{(s)}, \tilde{\mu}_{\xi}^{(s)}\}) R \sum_{t=1}^{T} (\tilde{W}_t^{(s,r)})'\tilde{\mu}_{\Omega}^{(s)}\tilde{u}^{(s,r)}(t) + (\Sigma_d^{-1})\mu_d)
\]

Loop over \((p, q) \in \{1, 2, \ldots, P\}^2\) in a random order:

\[
A_1 = R \sum_{t=1}^{T} \sum_{r=1}^{R} (\tilde{W}_t^{(s,r)}(:, i_{pq}^{(s)})'\tilde{\mu}_{\Omega}^{(s)}\tilde{W}_t^{(s,r)}(:, i_{pq}^{(s)})(\tilde{\mu}_{\phi}^{(s)}(i_{pq}^{(s)})\tilde{\mu}_{\phi}^{(s)}(i_{pq}^{(s)})) + \tilde{\Sigma}_d^{(s)}(i_{pq}^{(s)}))
\]

\[
A_2 = R \sum_{t=1}^{T} \sum_{r=1}^{R} [\tilde{W}_t^{(s,r)}(:, i_{pq}^{(s)})'\tilde{\mu}_{\Omega}^{(s)}\tilde{W}_t^{(s,r)}(:, -i_{pq}^{(s)})\text{diag}\{\tilde{\mu}_{\xi,-p,q}^{(s)}, \tilde{\mu}_{\xi,-p,q}^{(s)}\}]
\]
Finally, obtain

\[ \beta^{(s)} \sim N(\tilde{\mu}_\beta, \tilde{\Sigma}_\beta), \]  
\[ d^{(s)} \sim N(\tilde{\mu}_d, \tilde{\Sigma}_d), \]  
\[ \phi^{(s)} \sim N(\tilde{\mu}_\phi, \tilde{\Sigma}_\phi), \]  
\[ \xi_{p,q}^{(s)} \sim Multinomial(1, \tilde{\pi}_{pq}), \]  
\[ \Omega^{(s)} \sim Wishart(\tilde{\nu}_\Omega^{(s)}, \tilde{\mu}_\Omega^{(s)}/\tilde{\nu}_\Omega^{(s)}). \]
Appendix C

Appendix for Chapter 4

C.1 Detailed algorithm for posterior sample

Let \( \xi \) denote the matrix \( \{\xi_{pq}\}_{p,q} \in \mathbb{R}^{P \times P} \), \( \odot \) denote element-wise matrix multiplication, the superscript \( \text{old} \) denote the value of the variable before being updated, \( \Sigma_{d,-J} \) denote the matrix \( \Sigma_d \) with the \( J \)-th, \( 2J \)-th, \ldots, and \( PJ \)-th rows and columns removed, \( \mu_{d,-J} \) denote the vector \( \mu_d \) with the \( J \)-th, \( 2J \)-th, \ldots, and \( PJ \)-th elements removed, \( d_{-J} \) denote the vector \( d \) with the \( J \)-th, \( 2J \)-th, \ldots, and \( PJ \)-th elements removed, and \( [A](i), [A](i,j), [A](i,:), [A](:,j) \) denote the \( i \)-th element, the \( (i,j) \)-th element, the \( i \)-th row, and the \( j \)-th column of the matrix \( A \), respectively. Denote \( \text{diag}\{A\} \) to be the vector formed by the diagonal elements of \( A \) when \( A \) is a matrix, and \( \text{diag}\{x\} \) the diagonal matrix whose diagonal elements are the elements of \( x \) when \( x \) is a vector.

First, calculate the following constant quantities, \( \Sigma^{-1}_\phi \), \( \Sigma^{-1}_\phi \mu_\phi \), \( \Sigma^{-1}_{d,-J} \), \( \Sigma^{-1}_{d,-J} \mu_{d,-J} \), \( \Sigma^{-1}_{\Upsilon} \), \( \nu_{Q,\text{post}} \), \( \nu_{V,\text{post}} \), \( \Psi^{-1}_V \), \( \Psi^{-1}_Q \), \( H^{\text{conv}}_k = H \ast c_k \) (convolution), \( H^{\text{sum}} = \sum_t[H](t,:) \), 
\[
\begin{bmatrix}
\bar{H}^{\text{conv}}_k
\end{bmatrix}
\begin{bmatrix}
(\cdot,j)
\end{bmatrix}
= 
\begin{bmatrix}
\bar{H}^{\text{conv}}_k
\end{bmatrix}
\begin{bmatrix}
(\cdot,j)
\end{bmatrix}
- 
\begin{bmatrix}
\bar{H}^{\text{conv}}_k
\end{bmatrix}
\begin{bmatrix}
(\cdot,J)
\end{bmatrix}
\frac{H^{\text{sum}}_j}{H^{\text{sum}}_J},
\quad j = 1, \ldots, J - 1,
\quad \bar{H}^0_k = \frac{H^{\text{conv}}_J}{H^{\text{sum}}_J},
\quad \nu_{Q,\text{post}} = 2T \times R + \nu_Q,
\quad \text{and}
\quad \nu_{V,\text{post}} = T \times R + \nu_V.
\]
Then, obtain the posterior samples according to the following algorithm:

1. Set initial values for all the unknown parameters except the latent processes. Calculate $X_k(t)$ from the initial value of $d$. The initial values for $\Upsilon$ and $d$ can be obtained from the preliminary analysis in Section 4.2.4. The initial values for $\xi_{pq}$ can be set to 1’s for all $p, q$ or 0’s for all $p, q$, or, can be generated randomly according to $Bernoulli(0.5)$. The initial value for $\phi$ can be set to 0. The initial values for $V$ and $Q$ can be be set to $I_P$.

2. Use FFBS [10, 19] to sample the joint latent process $[\beta_1^{(r)}(t), \beta_2^{(r)}(t)]$ for $r = 1, \ldots, R$, based on the latest iteration of the other unknown parameters.

3. Sample the rest of the unknown parameters based on the latest iteration of the latent processes and the other parameters:

   (a) Sample $\phi_k \sim N(\mu_{\phi_k,post}, \Sigma_{\phi_k,post})$. Calculate the residual $\hat{w}_k^{(r)}(t)$, for $r = 1, \ldots, R, t = 1, \ldots, T$.

   (b) Sample $Q^{-1} \sim Wishart(\Psi_{Q,post}, \nu_{Q,post})$. Calculate the Cholesky decomposition of $Q$ and denote its transpose as $Q^{-1/2}$.

   (c) Calculate the orthogonalized residual $\tilde{w}_k^{(r)}(t) = Q^{-1/2}\hat{w}_k^{(r)}(t)$.

   (d) Loop over $(p, q) \in \{1, \ldots, P\}^2$ in a random order to sample $\xi_{pq}$:

      i. recalculate $\tilde{z}_k^{(r)}(t)$ and $\tilde{\beta}_k^{(r)}(t)$.

      ii. sample $\xi_{pq} \sim Bernoulli(\pi_{pq,post})$.

      iii. recalculate $\tilde{w}_k^{(r)}(t)$.

   (e) Sample $\Upsilon \sim N(\mu_{\Upsilon,post}, \Sigma_{\Upsilon,post})$.

   (f) Sample $d$:

      i. sample $d_{-J} \sim N(\mu_d,post, \Sigma_d,post)$,

      ii. calculate $d_{p,J}$ for $p = 1, \ldots, P$ according to Equation C.1
The variables that appear in the algorithm above are calculated as follows.

4. Repeat steps 2 and 3 until the desired size of the posterior sample is reached.

The variables that appear in the algorithm above are calculated as follows.

\[
X_k(t) = \text{diag}\{(I_P \otimes H_k^{conv}(t,.))d\},
\]

\[
z_k^{(r)}(t) = X_k(t-1)\beta_k^{(r)}(t-1),
\]

\[
Z_k^{(r)}(t) = z_k^{(r)}(t)' \otimes I_P,
\]

\[
\tilde{Z}_k^{(r)}(t) = Z_k^{(r)}(t)\text{diag}\{\text{vec}\{\xi\}\},
\]

\[
\Sigma_{\phi,k,post} = \left(\sum_{r=1}^R \sum_{t=1}^T \tilde{Z}_k^{(r)}(t)'(Q)^{-1}\tilde{Z}_k^{(r)}(t) + \Sigma_{\phi}^{-1}\right)^{-1},
\]

\[
\mu_{\phi,k,post} = \Sigma_{\phi,k,post} \left(\sum_{r=1}^R \sum_{t=1}^T \tilde{Z}_k^{(r)}(t)'(Q)^{-1}\beta_k^{(r)}(t) + \Sigma_{\phi}^{-1}\mu_{\phi}\right),
\]

\[
\tilde{w}_k^{(r)}(t) = \beta_k^{(r)}(t) - \tilde{Z}_k^{(r)}(t)\phi_k,
\]

\[
\Psi_{Q,post} = \sum_{k=1}^K \sum_{r=1}^R \sum_{t=1}^T \tilde{w}_k^{(r)}(t)Q^{-1}\tilde{w}_k^{(r)}(t)' + \Psi_Q^{-1},
\]

\[
\tilde{z}_k^{(r)}(t) = [Q^{-1/2}(.,:)][z_k^{(r)}(t)](q),
\]

\[
\tilde{\beta}_k^{(r)}(t) = \tilde{w}_k^{(r)}(t) + \xi_{pq}\Phi_{pq,k}z_k^{(r)}(t),
\]

\[
i_k = p + (q-1)P + (k-1)P^2,
\]

\[
\tilde{\rho}_{\phi,pq,k} = \sum_{r,t} \tilde{z}_k^{(r)}(t)'\tilde{z}_k^{(r)}(t) + [\Sigma_{\phi}^{-1}](i_k,i_k),
\]

\[
\tilde{\mu}_{\phi,pq,k} = (\tilde{\rho}_{\phi,pq,k})^{-1}(\sum_{r,t} \tilde{z}_k^{(r)}(t)'\tilde{\beta}_k^{(r)}(t) + [\Sigma_{\phi}^{-1}\mu_{\phi}](i_k)),
\]

\[
\eta_{pq} = \frac{1}{2} \sum_k \left[-\log(\tilde{\rho}_{\phi,pq,k}) + \log(\tilde{\rho}_{\phi,pq,k}\tilde{\mu}_{\phi,pq,k}^2) - \log([\Sigma_{\phi}^{-1}](i_k,i_k)([\mu_{\phi}](i_k))^2)\right],
\]

\[
\pi_{pq,post} = 1/(1 + \exp\{-\eta_{pq} - \text{logit}(\pi_{pq})\}),
\]

\[
\tilde{w}_k^{(r)}(t) = \tilde{w}_k^{(r)}(t)^{odd} - (\xi_{pq} - \xi_{pq}^{odd}) * [\phi](i_k) * \tilde{z}_k^{(r)}(t),
\]

\[
134
\]
\begin{align*}
\tilde{y}_d^{(r)}(t) &= y^{(r)}(t) - \sum_k X_k(t)\beta_k^{(r)}(t), \\
Z_{\mathbf{y}}(t) &= [I_p, X_1(t), X_2(t)], \\
\Sigma_{\mathbf{Y,post}} &= \left( R \sum_t Z_{\mathbf{y}}(t)'V^{-1}Z_{\mathbf{y}}(t) + \Sigma_{\mathbf{Y}}^{-1} \right)^{-1}, \\
\mu_{\mathbf{Y,post}} &= \Sigma_{\mathbf{Y,post}} \left( \sum_{r,t} Z_{\mathbf{y}}(t)'V^{-1}y_{\mathbf{y}}^{(r)}(t) + \Sigma_{\mathbf{Y}}^{-1} \mu_{\mathbf{Y,-J}} \right), \\
\tilde{y}_d^{(r)}(t) &= y^{(r)}(t) - Y_0 - \sum_k (\Sigma_k + \beta_k^{(r)}(t)) \left[ \tilde{H}_k^0 \right](t), \\
\tilde{\beta}_{d,k}^{(r)}(t) &= \beta_k^{(r)}(t) - (\xi \odot \Phi_k)\beta_k^{(r)}(t-1) \left[ \tilde{H}_k^0 \right](t-1), \\
Z_{d}^{(r)}(t) &= \sum_k \text{diag}\{(\Sigma_k + \beta_k^{(r)}(t)) \otimes [H_k^{\text{conv}}(t,:)]', \\
W_{d,k}^{(r)}(t) &= (\xi \odot \Phi_k)\text{diag}\{\beta_k^{(r)}(t-1) \otimes [H_k^{\text{conv}}](t-1,:)' \\
\Sigma_{d,post} &= \left( \sum_{r,t} \left( Z_{d}^{(r)}(t)'V^{-1}Z_{d}^{(r)}(t) + \sum_k W_{d,k}^{(r)}(t)'Q^{-1}W_{d,k}^{(r)}(t) \right) + \Sigma_{d,-J}^{-1} \right)^{-1}, \\
\mu_{d,post} &= \Sigma_{d,post} \left( \sum_{r,t} Z_{d}^{(r)}(t)'V^{-1}y_{d}^{(r)}(t) + \sum_k W_{d,k}^{(r)}(t)'Q^{-1}\tilde{\beta}_{d,k}^{(r)}(t) + \Sigma_{d,-J}^{-1} \mu_{d,-J} \right), \\
\hat{e}^{(r)}(t) &= \tilde{y}_d^{(r)}(t) - Z_{d}^{(r)}(t)d, \\
\Psi_{V,post}^{-1} &= \sum_{r,t} \hat{e}^{(r)}(t)V^{-1}\hat{e}^{(r)}(t)' + \Psi_V^{-1}.
\end{align*}

Notice that in the algorithm only the part \(d_{-J}\) is randomly sampled, instead of \(d\). This is because that the rank of \(d_p\) is only \(J-1\) instead of \(J\) due to the identifiability constraint \(\sum_t h_p(t) = 1\). The reason is that the constraint \(\sum_t h_p(t) = 1\) for \(\forall \ p\) can be rewritten as \(\sum_t h_p(t) = \sum_t h_p(t) = \sum_t H[t,:]d_p = \sum_t \sum_{j=1}^J H[t,j]d_{pj} = \sum_{j=1}^J (\sum_t H[t,j])d_{pj} \overset{\text{def.}}{=} \sum_{j=1}^J H_j^{\text{sum}}d_{pj} = 1\). Therefore, the \(J\)-th basis coefficient can be written as the linear combination of the first \(J-1\) coefficients:

\[ d_{pJ} = \frac{1}{H_j^{\text{sum}}} (1 - \sum_{j=1}^{J-1} H_j^{\text{sum}}d_{pj}) \quad \text{(C.1)} \]
Therefore, the constraint reduces the HRF linear basis down to $J - 1$ basis vectors. When I sample $d$, I only need to sample $d_{-j}$ and then calculate $d_{jp}$ accordingly.

### C.2 Supplementary figures

![Trace plots](image)

Figure C.1: Example trace plots for $\Phi_k^*$. 

136
Figure C.2: Example trace plots for $\mathbf{Y}_{(c)}$.

Figure C.3: Example trace plots for $\mathbf{V}$.

Figure C.4: Example trace plots for $\mathbf{Q}$. 

Figure C.5: Estimated mean signal versus the observed signal. Black: observed fMRI signal. Blue: point estimate for $\Upsilon_0 + \sum_k X_k(t) \tilde{\beta}_k(t)$ (total BOLD signal). Yellow: point estimate for $\Upsilon_0 + \sum_k X_k(t) \Upsilon_k$ (mean BOLD signal with the variation due to dynamic BOLD amplitudes excluded).
Figure C.6: Estimated HRFs from the original algorithm. Black: estimated HRF using the original algorithm. Blue dashed curves: 95% credible band. Blue vertical lines indicate $T_{peak}$ and $T_{under}$. Green: estimated HRF from preliminary analysis.
Figure C.7: Comparison of the estimated activation results for the same subject, from respectively the state-space model and from the hierarchical BVAR model in Chapter 3.

Figure C.8: Comparison of the estimated connectivity network for the same subject, from respectively the state-space model and from the hierarchical BVAR model.
Figure C.9: Comparison of the estimated HRFs used in the state-space model and the estimated HRFs from the hierarchical BVAR model. Green: the HRFs used in the state-space model. Black: estimated HRF from the hierarchical BVAR model.