**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____     _____
Wenqiong Xue                                          Date

# Statistical Methods for Multi-modal Neuroimaging Data: Techniques for the Combined Analysis of Brain Function and Structure

By

Wenqiong Xue

Doctor of Philosophy

Biostatistics

---

F. DuBois Bowman, Ph.D.
Advisor

---

Jian Kang, Ph.D.
Committee Member

---

James K. Rilling, Ph.D.
Committee Member

---

Lance A. Waller, Ph.D.
Committee Member

Accepted:

---

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

---

Date

# Statistical Methods for Multi-modal Neuroimaging Data: Techniques for the Combined Analysis of Brain Function and Structure

By

Wenqiong Xue

M.S., Auburn University, 2008

B.S., Fudan University, 2006

Advisor: F. DuBois Bowman, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biostatistics

2013

Abstract

**Statistical Methods for Multi-modal Neuroimaging Data: Techniques for the Combined Analysis of Brain Function and Structure**

By Wenqiong Xue

Recent innovations in neuroimaging technology have provided opportunities for researchers to investigate human brain function and structure, improving our understanding of psychiatric disorders, clinical diagnosis, and neural networks. Brain imaging data have massive dimensionality and are marked by complex spatial and temporal correlations, which pose challenges for statistical modeling. Our objective is to develop novel statistical methods for high-dimensional neuroimaging data to explore the underlying complex neural processing in the human brain. Specifically, we propose three new statistical frameworks: (i) to determine the functional coherence and the associated hierarchical relationships between brain regions using combined functional and structural data; (ii) to robustly characterize co-activation patterns and functional networks using a novel meta-analytic approach; and (iii) to predict the disease status using imaging data from different modalities.

Rapid development in neuroimaging allows researchers to study the connectivity in the human brain by examining the anatomical circuitry as well as functional relationships between brain regions. We present a unified Bayesian framework for analyzing functional connectivity utilizing the knowledge of associated structural connections, which extends an approach by Patel et al. (2006a) that considers only functional data.We demonstrate the use of our Bayesian model using fMRI and DTI data from a study of auditory processing.

Meta analysis plays an important role in neuroimaging research. Several approaches have been developed to determine the consistency in activated brain regions for imaging studies. We focus on identifying the functional co-activation patterns and building a non-directed functional network in the human brain. We adopt a penalized likelihood approach to impose sparsity on the covariance matrix for region-level peak activations, which is used to construct a brain network. We apply our proposed method to a meta analysis of 162 functional neuroimaging studies on emotions.

Relating disease status to imaging data increases the clinical significance of neuroimaging studies. We propose a Bayesian hierarchical model to predict the disease status using both the functional and structural imaging scans. We consider a two-level brain parcellation, and take into account the correlations between voxels from different levels. We conduct both whole-brain and voxel-level prediction, and apply our model to a study of Parkinson's disease.

# Statistical Methods for Multi-modal Neuroimaging Data: Techniques for the Combined Analysis of Brain Function and Structure

By

Wenqiong Xue

M.S., Auburn University, 2008

B.S., Fudan University, 2006

Advisor: F. DuBois Bowman, Ph.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biostatistics

2013

# Acknowledgement

First and foremost, I would like to thank my advisor Dr. DuBois Bowman for his truly great guidance and boundless support throughout my dissertation research. Without his systematic training, insightful comments, and continuous encouragement, the work presented would not be completed. I have benefited so much from his broad knowledge, enthusiasm in imaging studies, as well as hours of extensive discussions with him.

My gratitude also goes to Dr. Jian Kang, Dr. Lance A. Waller, and Dr. James K. Rilling for their time and effort serving as my committee members and reviewing my dissertation. Their constructive comments and suggestions have improved this dissertation a lot.

In addition, I am deeply grateful to all the members in the Center of Biomedical Imaging Statistics (CBIS) at Emory for their dedicated help and inspiring discussions. I feel so fortunate to be a member of this wonderful group.

Moreover, I would like to extend my thanks to the faculties and staffs in the Department of Biostatistics and Bioinformatics. I have had an enjoyable experience during the past years in this department. Especially, I am thankful to Dr. Qi Long, and Mr. Kirk A. Easley for the opportunity and guidance in collaboration projects.

Last but not least, I am greatly indebted to my parents and my grandfather for their unconditional love and support throughout my entire Ph.D. study. Their devoted care and continuous encouragement is always the most powerful motivation for every step I move forward.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Overview

Recent innovations in neuroimaging technology have provided opportunities for researchers to investigate the anatomy as well as the function of the human brain, thus improving the understanding of brain activities which are related to psychiatric disorders, neurologic disorders, clinical diagnosis, and neural networks. The developments in the imaging technology include functional magnetic resonance imaging (fMRI), Positron Emission Tomograpy (PET), and diffusion tensor imaging (DTI). fMRI and PET reveal the functional information between brain regions, while DTI demonstrates the anatomical information. One common property of the brain imaging data is high-dimensionality and complex spatial as well as temporal correlation structure, which is challenging for statistical modeling. Our main objective is to develop novel statistical method for high-dimensional neuroimaging data to explore the complex neural processing characteristics.

The dissertation is organized as follows: the remainder of Chapter 1 provides a brief introduction to the organization of the human brain, a description of the neuroimaging modality we consider, and the current methods of data preprocessing

and statistical analysis, followed by the motivating examples and our proposed research objectives. Chapter 2 presents a Bayesian model for determining functional connectivity incorporating structural connectivity in the human brain. Chapter 3 describes a penalized multivariate Poisson model to identify the functional network with application to a meta analysis of emotions. Chapter 4 presents a novel Bayeisian spatial prediction model using imaging data from different modalities, and Chapter 5 summarizes our work and discusses some possible future work.

## 1.2  An Introduction to the Human Brain

Thoughts, feelings, perceptions, and acts - from the simplest movements to the most complex ideas - these are the products of the three-pound organ inside our head. The functional operating unit of the nervous system is neurons. Each neuron receives inputs from many other nerve cells, integrates those inputs, and then distributes the processed information to other neurons. This is the vital job that the brain must accomplish: the integration and analysis of information.

The dominating parts of the brain are the large cerebral hemispheres, which are the right and left halves of the forebrain. Within the cerebral hemispheres are the basal ganglia and the limbic system. The basal ganglia are very important in motor control, and the limbic system is involved in emotion and learning. Each of the hemispheres has an outer layer of grey matter, which is dominated by nerve cell bodies and dendrites, called the cerebral cortex that is supported by an inner layer of white matter consisting mostly of fiber tracts. In general, the cortex may be regarded as the seat of complex cognition; damage to the cortex may impair "higher" functions such as speech, memory, or visual processing. In contrast, "lower" parts of the brain regulate respiration, heart rate, and other basic functions. The cerebral convolutions are the result of elaborate folding together of tissue. The resulting

ridges of tissue are called gyri, and are separated from each other by furrow called sulci. Such folding enormously increases the cerebral surface area; about two-thirds of the cerebral surface is hidden in the depths of these folds.

The major sectors of the cerebral hemispheres are the frontal, parietal, temporal, and occipital lobes. The four lobes of the brain normally collaborate with one another to serve myriad different processes. In a general sense, however, we can identify categories of processing that are particularly associated with specific lobes. For example, the occipital lobes receive and process information from the eyes, giving rise to many aspects of the sense of vision. Auditory information is directed to the temporal lobes, and damage there can impair hearing (the temporal lobes are also particularly associated with the sense of smell, and with aspects of learning and memory).

The two hemispheres and four lobes provide us a general map of human brain; however, finer cerebral cortex parcellation is desired for in-depth study. One fine cerebral cortex parcellation is defined by Brodman areas (48 regions), which is based on cytoarchitecture, or organization of cells. Another pacellation is Automated Anatomical Labeling (AAL) regions (116 regions), which is constructed through the identification of major and minor sulci/gyri on a T1 MRI with subsequent labeling based on anatomical location (Tzourio-Mazayer et al., 2002). It is typically used in functional neuroimaging-based research to obtain neuroanatomical labels for the location in 3-dimensional space where the measurements of some aspect of brain function are captured. In other words, it projects the divisions in the brain atlas onto brain-shaped volumes of functional data.

Due to the three-dimensional structure of the nervous system, two-dimensional illustrations and diagrams cannot represent it completely. The brain is usually cut in one of three main planes to obtain a two-dimensional section from this three-dimensional object as illustrated in Figure 1.1. We call them axial, sagittal, and coronal view of brain scans.

Figure 1.1: Axial, sagittal, and coronal view of the human brain.
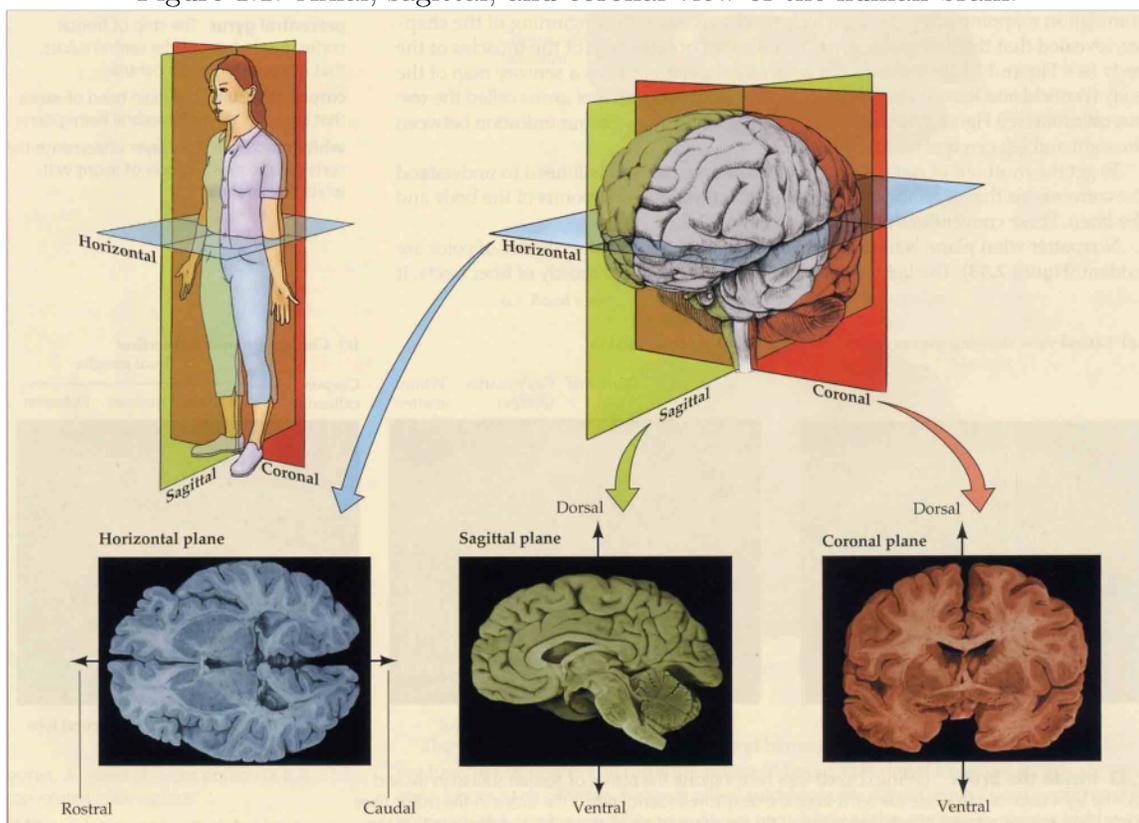
When group studies using brain imaging data are performed, the individual brain images are usually transformed into a common coordinate space to accommodate the between subject variation of brain size and orientation. The Talariach space and the Montreal Neurological Institute (MNI) space are the two most widely used standard spaces in the neuroscience community. The Talariach coordinate system is based on a stereotaxic atlas of the human brain published by Talairach and Tournoux (Talairach and Tournoux, 1988). It is related to anatomical landmarks and specifies locations relative to their distance from the midpoint of a brain structure called anterior commissure (AC) - a small but easy spot region. Each location is described by three numbers which describe the distance in millimeters from the AC: $x$ is the left/right dimension, $y$ is the posterior/anterior dimension, and $z$ is the ventral/dorsal dimension. In this atlas the axial slices are referred to by their $z$ coordinate, coronal by their $y$ coordinate, and sagittal by their $x$ coordinate. The atlas is based on a single brain of a 60-year-old French woman with mental disorder, whose brain is considerably smaller than the average brain by up to 10 millimeters in each dimension. The MNI defined a new standard brain by using a large series of MRI scans on normal controls (Evans et al., 1993). These atlases differ in shape and size, and are available in neuroimaging processing software.

## 1.3 Functional Neuroimaging

Functional neuroimaging of the brain improves our understanding of brain networks by determining brain activity and connectivity at distinct regions of the brain. There are many types of functional imaging scans, including Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), electroencephalography (EEG) and magnetoencephalography (MEG). PET and fMRI measure localized changes in cerebral blood flow that accompany neuronal activity with relatively high

spatial resolution, but with temporal resolution limited by the much slower rate of brain blood flow and blood oxygenation. In contrast, EEG and MEG map the underlying electrical activity of the brain, which allow high temporal resolution of neural processes, but have poor spatial resolution.

In my dissertation, we focus on developing statistical models for analyzing fMRI data. In the following sections, we give a brief introduction of this functional neuroimaging technique.

## 1.3.1 Functional Magnetic Resonance Imaging (fMRI)

Due to its low radiation exposure, low invasiveness, and relatively wide availability compared to PET or other functional neuroimaging techniques, fMRI has become widely used in the brain imaging studies. Before we discuss the principles of fMRI, we present how the MRI scans are acquired since fMRI is a type of specialized MRI scan used to measure the hemodynamic response related to neural activity in the brain.

MRI is an imaging technique based on the principles of nuclear magnetic resonance (NMR) to image nuclei of atoms inside the body. It uses magnetism, radio waves, and a computer to produce images of the body.

To obtain an MR image, a subject is first placed in a field of a large electromagnet from 1.5 to 4.0 Tesla. One Tesla is equal to 10,000 Gause; while the magnetic field of the earth is approximately 0.5 Gause. Therefore, 1 Tesla means that the magnetic field is 20,000 times as strong as that of earth. When the hydrogen proton in the human brain is placed in a large external magnetic field (B0), it will align in either a parallel or anti-parallel with the direction of the magnetic field due to its spin characteristics. A low-energy state refers to the state that the protons are in the parallel orientation, while a high-energy state refers to the state that the protons align in the anti-parallel orientation. The strength of B0 is positively related to the

number of spins align in the low-energy state. If the number of spins in the low-energy state exceeds the number of spins in the high-energy state, a net magnetization is formed and the tissue placed in the magnetic field becomes magnetized. When a pulse of radio frequency (RF) energy, which is oriented perpendicular to the main field, is injected from the scanner, some of the spins are raised out from their lower energy state to higher energy state due to the absorption of the energy from the RF. The angle that produced by RF rotation from the net magnetization toward the transverse plane is called flip angle (FA). If the RF excitation is discontinued, some of the protons that have been moved from the low-energy state to the high-energy state release their energy and fall back to the low-energy state, and hence producing MR signal. We are interested in three types of relaxation time which characterize the rates to distinguish among distinct tissue types. T1 relaxation time defines the rate at which longitudinal magnetization reappears. At the same time, T2 and T2* relaxation times measure the rate of decay of the transverse magnetization vector.

The T1 relaxation time is the time for the magnetization to return to 63% of its original strength. Two T1 times gives 86%, and three T1 times recovers 95% of the magnetization. Complete relaxation is happened after 3-5 T1 times. The length of T1 times is associated with the strength of the main magnetic field. T2 relaxation refers to the state that the spins in the high and low energy state exchange energy but do not loose energy, which results in the loss of the transverse magnetization. After one T2 period, 63% of the strength of the transverse magnetization decays. T2* dephasing, which is the loss of the signal due to one or more localizable sources, occurs at a faster rate than T2. The reason is that T2* relaxation is caused by magnetic field inhomogeneity that occurs in all magnets. The values of the different types of relaxation times construct scans called T1-weighted, T2-weighted and T2*-weighted images. fMRI is a T2* image.

Each of the 3D MRI scan is composed of 2D images. During the process of data

acquisition, the raw MR signals are stored in a temporary image space, called k-space. At the end of the scan, two successive one-dimensional Fourier transformations are applied to the rows and the columns of each 2D scans stored in the k-space to produce an MR image. Our analysis is performed on the transformed data. Each transformed image contains complex values, which stands for an amplitude and phase. In the fMRI analysis, the amplitude component of the complex number is considered. fMRI measures the neural activity of the brain by the changes in the blood flow. An increase in the blood flow is detected when the neuronal activity increases. At the same time, the demand for oxygen, which is carried by the hemoglobin in red blood cells, rises up in the regions of neural activity. Therefore, the human body will provide more hemoglobin than consumed energy to the active brain area, thus increasing the concentration of hemoglobin and decreasing the concentration of deoxygenated hemoglobin (dHb). Physiologically, the neurons require more glucose, which is the primary source of energy, and more oxygen to bring them back to their normal state when they are active. Different brain regions usually differ in the cerebral blood flow (CBF) relative to the consumed glucose in response to neural activity. Due to the paramagnetic properties of dHb, it has the effect of distorting the surrounding magnetic field. As a consequence, the nuclei around lose magnetization at a faster relaxation time T2*. Therefore, T2* signal is stronger in the area that the concentration of oxygen is high, and weaker in the area that the concentration of oxygen is low. fMRI uses the blood oxygen level dependent (BOLD) contrast as described to show the differences of blood oxygenation at different levels of neural activity. To characterize the time lag between the stimuli and the change of a BOLD signal, a hemodynamic response function (HRF) is applied.

When designing an fMRI study, there are a few imaging parameters that need to be considered for temporal and spatial resolution. Spatial resolution in fMRI scan is determined by the size of voxels. A voxel is a three-dimensional cuboid, whose size

ranges from 4 mm, 2 mm to 1 mm. The larger the size of the voxels is, the fewer the number of voxels that an fMRI scan contains, and the less time it takes to scan, thus increasing the temporal resolution. Repetition time (TR), the time between successive excitation pluses, determines the temporal resolution. TR is typically measured in seconds. Since successive excitations often do not have enough time for a full recovery, it is important to know TR to insight the possible influence it may have on the next excitation. Echo time (TE) is the time between excitation and the application of RF pluses. This time interval is usually measured in milliseconds. Echo time is determined by the T2* values for gray matter in the human brain.

## 1.3.2   Preprocessing

Preprocessing is required for fMRI for correcting non-task related variability in experimental data. The following steps are involved in the preprocessing: slice timing correction, which corrects for differences in acquisition time within a TR due to the fact that the 2D images that compose one 3D image are acquired at different time points, it is better to be performed before motion correction if the scans are acquired interleavely, and after motion correction if the scans are acquired sequentially; motion correction, which removes the effects of head move during the fMRI experiment and realigns all 3D scans to a common reference for each subject, usually the first or second scan of all the 3D images, by rigid body transformation with six degrees of freedom and minimization of some cost functions; spatial smoothing by convolving the functional scans with Gaussian kernel, which increases signal to noise ratio (SNR) with same shape/size as signal, reduces the number of comparisons, and improves comparison across subjects, however, spatial smoothing may reduce spatial resolution; temporal smoothing which filters low frequency and high frequency noises; registration which registers each individual's brain to a standard template brain using linear transformation with twelve degrees of freedom or nonlinear transformation,

it allows generalization of results to large population and improves comparison with other studies, however, it reduces spatial resolution; intensity normalization which scales each subject's 4D dataset by a single value to get the overall 4D mean to be the same for all subjects. Preprocessing steps and the order in which they are performed are important since they affect both the spatial and temporal correlation structure of the data.

Commonly used preprocessing tools for functional brain images include Statistical Parametric Mapping (SPM), FMRIB Software Library (FSL), and Analysis of Functional NeuroImages (AFNI).

### 1.3.3  Analysis of Functional Connectivity

#### 1.3.3.1  Definition

Functional connectivity is defined as temporal correlation between spatially distinct regions of the human brain. It allows us to characterize the neural processing of the human brain during motor, behavioral, and cognitive tasks, or simply from spontaneous brain activities during resting-state. Functional neuroimaging data are used to measure the functional properties of the brain. In my research, we use the time series of BOLD signal for the analysis of functional connectivity.

#### 1.3.3.2  Existing Approaches

*Seed Voxel Approach*

A region of interest (ROI) is preselected based on functional or anatomical knowledge from previous studies or selected using an independent data set approach (Hampson et al., 2002). A seed voxel approach is a method to evaluate the functional connectivity between an ROI and other brain regions.

Calculating the temporal correlation between the seed ROI and the remaining brain voxels is a traditional measure of functional connectivity used in the seed voxel

approach. However, the correlations that we observe may arise from neurophysiologic influences or non-neurophysiologic sources, such as those induced by the scanner and from image preprocessing. Some nonparametric wavelet-based methods for evaluating functional connectivity are proposed to account for the background spatial correlation inherent in imaging data. Wavelet-based nonparametric techniques (Breakspear et al., 2004; Patel et al., 2006a) are developed to determine whether the functional connectivity observed is significantly greater than the background correlation through spatiotemporal resampling of the data in the wavelet domain.

*Clustering*

Clustering analysis is a data-driven approach which uses the measures of distance to determine dissimilarity between voxels (clusters). Typically, people choose to use two easily computed and interpreted measures, Euclidean or Mahalonobis distances to quantify the dissimilarities. The primary objective of neuroimaging clustering is to identify collections of voxels (clusters) that exhibit similar brain activity patterns and reveal distinct neural response patterns between clusters.

Currently, several clustering algorithms are used in the neuroimaging studies, which generally fall into two categories: partitioning algorithm and hierarchical clustering algorithm. Partitioning algorithm, such as K-means approach (Balslev et al., 2002), fuzzy clustering (Baumgartner et al., 2000), are widely used in neuroimaging studies, partly due to the computational efficiency; however, K-means approach requires prior specification of the number of clusters and often no scientific basis exists for setting it in applications. Hierarchical clustering methods (Bowman et al., 2004; Filzmoser et al., 1999) begin with each voxel representing a separate cluster and proceeds with successive merges until all clusters unite. Specific hierarchical clustering algorithms use different functions to measure the distances between clusters.

*Independent Component Analysis (ICA)*

Independent component analysis identifies statistically independent spatial compo-

nents that are associated with time courses and hence offers a useful exploratory technique to investigate functional connectivity. There are two major types of ICA models: classic ICA model and probabilistic ICA model. The classic ICA procedure decomposes the observed time series into a product of two matrices, one representing the unknown realizations of spatial components, and the other containing the associated time courses. Probabilistic ICA (PICA) (Beckmann and Smith, 2004) includes an additional term, a Gaussian noise term in the classic ICA decomposition, which is a major distinction between probabilistic and classical ICA.

Making group level inferences from subject level analysis are often of interest in the neuroimaging studies. Several methods have been proposed for performing ICA on multisubject functional neuroimaging data. Beckman and Smith (2005) extend their single-session PICA to a higher dimension by adding an additional subject dimension, called tensor PICA. This approach decomposes observed time series from multiple subjects into a sum of products of spatial signals, and subject-specific components associated with each spatiotemporal process. A second group level ICA approach involves a temporal concatenation on each subject (Calhoun et al., 2001). When the interest is in finding a common spatial pattern; however, the assumption of consistency of the associated temporal response across subjects is violated, this would be appropriate approach.

*Bayesian Hierarchical Modeling*

Patel et al. (2006) propose a model-based analyses to quantify functional connectivity via a Bayesian hierarchical model. For each fMRI scan, the model defines joint activation between pairs of voxels based on the classification of activation or inactivation. The total number of times across all the subjects and all the scans that two voxels show elevated activities is defined as $Z_1$, while $Z_4$ is the number of times that neither voxels are active. $Z_2$ and $Z_3$ quantify the number of times that one voxel is active, and the other is not. A Bayesian model is formulated, in which they

assume that $Z_1, \cdots, Z_4$ follows a multinomial distribution with parameter $\theta_1, \cdots, \theta_4$. By imposing a prior distribution on $\theta_1, \cdots, \theta_4$, a posterior distribution conditional on the data can be derived. They also develop a measure of functional connectivity to describe to what degree two voxels are connected to each other. This measurement is based on the relative difference between the marginal probability of one voxel is active and the probability this voxel is active conditional on elevated activity of the other voxel. Given two voxels are functionally connected, a hierarchical relationship, called ascendancy, which compares the marginal activation probabilities of two voxels is developed. The first topic in my dissertation is an extension of this model.

### 1.3.3.3    Meta Analysis of Functional Neuroimaging

Recently, there has been a rapid increase in the functional neuroimaging studies using fMRI and PET. Due to the relatively high false positive rates in imaging studies compared to other studies that caused by small sample size or lack of multiple comparison correction, it is important to evaluate the consistency of the findings from different studies. Therefore, meta analysis of functional neuroimaging studies, which identify the task-related functionally activated or connected brain regions, has become of great interest.

Based on the reported locations of statistically significant regions of activation, called foci, i.e., coordinates in a template space, a number of coordinate based meta analysis methods have been proposed. Two commonly used meta analysis methods that investigate consistent activation across studies are activation likelihood estimate (ALE) (Turkeltaub et al., 2002) and Kernal density analysis (KDA) (Wager et al., 2003). Both of these two methods create a statistical map based on the activation areas of each study. They analyze the distributions of the number of peaks in an area using kernel-based methods. KDA smoothes the distribution with a spherical indicator kernel, while ALE uses a Gaussian kernel function.

Functional networks can be used to describe the functional connectivity maps of distinct spatial distribution of temporally correlated brain regions during tasks or resting states. Kober et al. (2008) use multivariate analyses to identify functional groups, in which consistent patterns of co-activation studies across studies are investigated. This approach applies a multilevel KDA (Wager et al., 2007) to the reported peak activations and forms a hierarchical clustering according to their patterns of co-activations. A structure learning of Bayesian network approach is proposed by Neumann et al. (2010) to investigate the functional connections between brain regions given the co-activation patterns across studies. Instead of functional or effective connectivity, probabilistic dependencies between brain regions are presented. In addition, a directionality of multivariate relations between functional brain regions is established.

## 1.4 Structural Neuroimaging

Structural neuroimaging studies describe the anatomy as well as the anatomical linkages between brain regions. The diffusion MRI technique allows for encoding the diffusion effects of molecular, mainly water, which can reflect the structures of the brain tissues in vivo. Diffusion tensor imaging (DTI) is widely used in the neuroimaging studies to reveal the anatomical connections in white matters.

### 1.4.1 Diffusion Tensor Imaging (DTI)

Diffusion tensor imaging (DTI) is another MRI technique. It measures the restricted diffusion of water in tissue to produce neural tracts between brain regions. The characteristics of each voxel in DTI are described by the diffusion tensor matrix, which estimates the diffusivity in six noncollinear directions. Therefore, the tensor matrix is a $3 \times 3$ symmetric matrix with six degrees of freedom. We use the eigenvectors and the

corresponding eigenvalues ($\lambda_1$, $\lambda_2$, $\lambda_3$) of the tensor matrix to describe the diffusivity and anisotropy of each voxel. Some commonly used method to summarize diffusivity includes axial diffusivity ($\lambda_1$), radial diffusivity (($\lambda_2 + \lambda_3$)/2), trace ($\lambda_1 + \lambda_2 + \lambda_3$), and mean diffusivity (($\lambda_1 + \lambda_2 + \lambda_3$)/3). The anisotropy describes the relative degree to which the three tensor eigenvalues differ from one another in a voxel. The most commonly used measurement, fractional anisotropy (FA), is the square root of the sum of squares of the diffusivity differences, divided by the sum of squares of the diffusivities. The effect of diffusion on the MRI signal is characterized by the echo attenuation, which is a function of the tensor matrix and the "$b$ matrix", whose elements describe the gradient pulses, including timing, amplitude, shape, used in the MRI sequence.

### 1.4.2   Preprocessing

Similar to fMRI data, preprocessing is also required for DTI data before performing any statistical analyses. Here are the steps that are involved in the preprocessing: eddy current correction, which corrects for distortions in diffusion MR images that are generated by different gradient directions; motion correction, which removes the effects of head move and realigns all 3D scans to a common reference for each subject, usually the first or second scan of all the 3D images, by rigid body transformation with six degrees of freedom and minimization of some cost functions. We implement these preprocessing steps by FSL.

### 1.4.3   Analysis of Structural Connectivity

Diffusion tensor tractograhpy (DTT) is used to reconstruct the white matter bundles between voxel pairs, thus describing the structural connections between them by demonstrating the neural tracts. There are two commonly used tracking algorithms: derterministic tractography and probabilistic tractography. Compared to determin-

istic tractography, probabilistic tractography is more computationally expensive, but it provides the probability that how probable two voxels are connected to each other. In our research, we adopt the probabilistic Bayesian DT-based tracking developed by Behrens et al. (2007). This algorithm estimates the probabilistic distribution of the orientation for each voxel by Markov Chain Monte Carlo sampling through a Bayesian framework, which determines the direction that each sample goes from a seed voxel based on certain criterion. The probability that two voxels are connected is calculated from the number of times that a streamline goes from a voxel to/through the other by the number of samples generated.

## 1.5   Prediction

Functional and structural neuroimaging play important roles in understanding the neurological basis for major psychiatric disorders or mental illness such as schizophrenia, depression, Alzheimer's diseases, and Parkinson's disease (PD). Several methods have been proposed to make the prediction of follow-up imaging scans based on the baseline scans (Guo et al., 2008; Derado et al., 2012). However, the classification or prediction of the disease status based on the imaging data has not been fully investigated. Though the diagnostic criterion may have been established for the above diseases, a prediction model can help to reveal the underlying neural basis of the diseases, thus informing the development of future treatments.

To predict the follow-up brain activity based on the baseline functional neuroimaging data, Guo et al. (2008) propose a Bayesian hierarchical model for functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) data. Bowman et al. (2008) develop a model which considers inter-regional and intraregional correlations for analyzing functional neuroimaging data. Derado et al. (2012) extends the model by introducing both spatial correlations between voxels and temporal cor-

relations between baseline and follow-up functional imaging scans. For structural data, Stonnington et al. (2010) propose a relevance vector regression (RVR) model to predict the clinical scores using MRI T1 scans.

The most commonly used approaches to predict a single outcome from high-dimensional data are LASSO (Tibshirant, 1996) and elastic-net (Zou and Hastie, 2005), which perform regularization and variable selection for regression models. Support vector machine (SVM) classify the data by constructing an optimal separating hyperplane in a high dimensional space in which the data are mapped to (Vapnik, 1995). As an alternative to SVM for prediction of high-dimensional data, Gaussian process (GP) find the posterior function distribution which is closest to the training data based on Bayesian theory (Marquand et al., 2010). Ham and Kwak (2012) propose a boosted-principal component analysis (PCA) algorithm for binary classification problems, which combines the procedure of feature selection and classification. However, these methods do not consider the existence of the spatial correlations in imaging scans, and build the statistical model directly for the prediction purpose rather than start from the imaging data.

## 1.6 Motivating Examples

### 1.6.1 An fMRI Study on Auditory Spatial-Cueing Task

We consider data from an fMRI study of 32 right-handed adults (15 males, 17 females) who were scanned while performing an auditory spatial-cueing task. The auditory stimuli consist of a series of paired tone pips. In each trial, the first auditory tone (2,000 Hz) may occur in either the left or the right ear and serves as a spatial cue, and the second tone (1,000 Hz) is a target tone and may also occur in either ear. Valid trials are those in which both the cue and the target tones sound in the same ear; otherwise, we call them invalid trials. The chance of either happening is 50%.

Targets are equally likely to sound in the right or the left ear, and the order of trials is randomized. A key-press device is positioned under the subject's right hand to record responses. Subjects are instructed to make a key press with their right middle finger for targets arising in the right ear and right index finger for targets arising in the left ear. There are a total of 84 valid and 84 invalid trials presented across three imaging runs. Prior to the start of the experiment, subjects are informed that the cues did not contain any useful information to predict the location of the target. Subjects fixate on a white visual cross on a black background during the entire session. Several objectives of the auditory spatial-cueing task have been addressed previously by Mayer et al. (2009). Our objective here is to identify the functionally connected brain regions associated with the neural processing stemming from an auditory task (both spatially valid and invalid), which also involves motor and visual functions. Note that our analysis assesses the combined contributions of valid and invalid trials and makes no attempt to dissociate the effects of spatial priming.

## 1.6.2 A Meta Analysis of Emotions

We consider findings from 162 functional neuroimaging studies, including 57 positron emission tomography (PET) and 105 functional magnetic resonance imaging (fMRI) studies. Collectively, these studies yield 439 contrasts (e.g., happy vs. neutral) as in Kober et al. (2008). These studies are published in English from 1990 to 2005. We consider seven different emotions: sadness, happiness, anger, fear, disgust, surprise, and affective. These studies all meet the following criteria: (1) All the subjects included in the studies are healthy; (2) All the studies measure regional cerebral blood flow (PET) or blood oxygenation (fMRI) across the entire brain instead of regions of interest; (3) The activation coordinates are determined using the image subtraction methodology; (4) Standard Talairach or Montreal Neurological Institute (MNI) coordinates are provided to ensure the results are spatially normalized to

standard coordinate systems, thus allowing for comparison of findings across different studies. For our analysis, all data were converted into MNI space. For each study, the activation locations for these contrasts are included when they meet the criteria of significance defined each individual study. A total of 2478 activated coordinates are reported for 439 contrasts.

### 1.6.3  Parkinson's Disease Data

A total of 20 subjects, 11 of which are diagnosed as PD patients, and the rest are healthy controls, are included in the study. The average age is 66 ($\pm$ 11) years old, and 12 of them are males. Resting-state fMRI scans, T1-weighted MRI scans, and diffusion tensor imaging (DTI) scans are obtained.

We extract voxel-level information from these three types of imaging scans, including fractional amplitude of low-frequency fluctuation (fALFF) from resting-state fMRI scans, voxel base morphometry (VBM) from T1-weighted MRI scans, and fractional anisotropy (FA) from DTI scans. fALFF reflects the amplitude of spontaneous signal fluctuations of each brain region. VBM measures the localized intensity value for each voxel after spatially normalizing all the images to a standard space, and extracting white or grey matter from the normalized images (Ashburner and Friston, 1999). FA has a single value for each voxel which measures the difference in directions along different axes of the random motion of water molecules in the brain.

The imaging preprocessing are performed by statistical parametric mapping (SPM) (Wellcome Department of Cognivite Neurology, http://www.fil.ion.ucl.ac.uk/spm) and FMRIB (Functional Magnetic Resonance Imaging of the Brain) Software Library (FSL) (Smith et al. 2004).

## 1.7 Proposed Research

### 1.7.1 Modeling Functional Connectivity with Incorporation of Structural Connectivity

Recent innovations in neuroimaging technology have provided opportunities for researchers to investigate connectivity in the human brain by examining the anatomical circuitry as well as functional relationships between brain regions. Existing statistical approaches for connectivity generally examine resting-state or task-related functional connectivity (FC) between brain regions or separately examine structural linkages. We present a unified Bayesian framework for analyzing FC utilizing the knowledge of associated structural connections, which extends an approach by Patel et al. (2006a) that considers only functional data. Our FC measure rests upon assessments of functional coherence between regional brain activity identified from functional magnetic resonance imaging (fMRI) data. Our structural connectivity (SC) information is drawn from diffusion tensor imaging (DTI) data, which is used to quantify probabilities of SC between brain regions. We formulate a prior distribution for FC that depends upon the probability of SC between brain regions, with this dependence adhering to structure-function links revealed by our fMRI and DTI data. We further characterize the functional hierarchy of functionally connected brain regions by defining an ascendancy measure that compares the marginal probabilities of elevated activity between regions. Posterior estimation is performed using Markov Chain Monte Carlo (MCMC) techniques. We demonstrate the use of our Bayesian model using fMRI and DTI data from a study of auditory processing. We further illustrate the advantages of our method by comparisons to methods that only incorporate functional information.

### 1.7.2 A Graphical Model for Count Data: A Meta Analysis of Functional Co-activation Patterns in Imaging Studies

Meta analysis plays an important role in neuroimaging research. Several approaches have been developed to determine the consistency in activated brain regions for a particular type of task, cognition, emotion or social process. Here, we focus on identifying the functional co-activation patterns and building a non-directed functional network in the human brain. We adopt a penalized likelihood approach to impose sparsity on the covariance matrix for region-level peak activations based on an extended multivariate Poisson model. The sparse covariance matrix is in turn used to construct a brain network. We obtain the penalized maximum likelihood estimates via the expectation-maximization (EM) algorithm and optimize an associated tuning parameter by maximizing the predictive log-likelihood. We conduct permutation tests on the brain co-activation pattern network. The proposed approach achieves small bias and coverage rate that is close to 95% in terms of estimation. We also discuss the choice of the penalty term and its impact on identifying the network from simulation studies. We apply our proposed method to a meta analysis of 162 functional neuroimaging studies on emotions. Our model identifies a functional network that consists of regions from the basal ganglia, limbic system, and other emotion related brain regions.

### 1.7.3 A Bayesian Spatial Model to Predict Disease Status Using Imaging Data from Different Modalities

Relating disease status to imaging data increases the clinical significance of neuroimaging studies. We propose a Bayesian hierarchical model to predict the disease status using both the functional and structural imaging scans. Our approach mod-

els the imaging data and performs prediction using posterior predictive probabilities. We consider a two-level brain parcellation, and take into account the correlations between voxels from different levels. The posterior estimations are performed by Markov Chain Monte Carlo (MCMC) via Gibbs sampling. We evaluate our method by examining the prediction accuracy rates based on leave-one-out cross validation and reduce the computational time by importance sampling strategy. We conduct both whole-brain and voxel-level prediction, and identify the brain regions that are highly associated with the disease status based on the voxel-level prediction results. We apply our model to a study of Parkinson's disease, and the whole-brain prediction indicates 100% accuracy rate.

# Chapter 2

# Modeling Functional Connectivity in the Human Brain with Incorporation of Structural Connectivity

## 2.1 Introduction

Applications of functional and structural neuroimaging have provided novel insights into brain connectivity. Friston et al. (1993) define functional connectivity (FC) as the "temporal correlations between spatially remote neurophysiological events", which establishes an important goal of many neuroimaging statistical analyses. This non-directional association may pertain to resting-state brain activity or to neural activity stemming from cognitive, emotional, visual, and behavioral tasks. By contrast, structural connectivity (SC) refers to the underlying white-matter structural links between different brain regions. By enabling the transmission of electrical signals that pass along axons (Hendelman, 2000), SC provides a mechanism for functional

relationships in neural activity. Yet, FC and SC properties typically are evaluated separately.

The most common approach to determine FC is based on Pearson (or partial) correlation between the temporal activity profiles from pairs of spatially distinct brain regions (Hampson et al., 2002; Marrelec et al., 2006). Related approaches use different measures of association, e.g. mutual information (Cover and Thomas, 1991), or consider transformations of the time series data, e.g. to examine associations in the frequency domain (Fiecas et al., 2010) or in the wavelet domain (Patel et al., 2006). Patel et al. (2006a) develop a Bayesian model that assesses the functional connectivity based on the relative probabilities of elevated activity. In addition, partitioning algorithms establish networks on the basis of distinct patterns of neural activities, with each defined network containing brain regions exhibiting similar neural processing traits. Independent component analysis (ICA) is a very popular partitioning method (Calhoun et al., 2001; Guo and Pagnoni, 2008), and cluster analysis is another commonly used approach (Bowman et al., 2004).

There are alternative methods for determining connectivity from functional neuroimaging data such as structural equation modeling (SEM) (McIntosh and Gonzalaz-Lima, 1994) and dynamic causal modeling (DCM) (Friston et al., 2003). However, these methods seek a stronger relationship between brain regions than we consider, specifically attempting to determine the influence of one brain region on another, also known as effective connectivity. Quantifying these stronger directional associations typically requires the advanced specification of a few potential networks to be hypothesized and compared, and they are not able to search across the full range of possible network topologies.

SC refers to the presence of white-matter fiber tracts (bundles of axons) directly connecting different brain regions. Diffusion tensor imaging (DTI) is a magnetic resonance imaging (MRI) technique for describing the orientation of axonal fiber

bundles based on the diffusivity of water molecules, since water will tend to diffuse more rapidly in the direction aligned with the white-matter fibers and more slowly as it moves perpendicular to this structure. Thus, DTI images contain information about white-matter structure that is important for fiber tracking (Kollias, 2009). Probabilistic diffusion tensor tractography (DTT) is a technique that uses DTI data to empirically reconstruct fiber tracts by quantifying the likelihood of white-matter SC (Behrens et al., 2007).

A notable limitation of the aforementioned procedures for determining FC is that they do not consider any information about the associated SC. There have been a few recent attempts to examine both FC and SC. These studies generally either examine the correspondence between SC and localized (voxel-level) analysis of functional MRI (fMRI) data or assess SC and FC sequentially and use SC to guide region selection for FC evaluation (or vice versa) (Rykhlevskaia et al., 2008). From such analyses, Morgan et al. (2009) suggest that FC is supported by SC along the language pathways. Also, Greicius et al. (2009) and van den Heuval et al. (2009) indicate that resting-state FC reflects SC to a large degree. Bowman et al. (2012) present a novel framework that simultaneously considers fMRI and DTI data to determine functional networks. However, their approach is descriptive and does not directly permit statistical inferences. As the association between brain structure and function is revealed, an important next step is to develop unified, model-based statistical frameworks that incorporate both sources of information simultaneously, and permit direct statistical inferences.

We present a novel multimodal approach to assess FC, which incorporates SC as supplementary information. We extend the previously developed model by Patel et al. (2006a) that determines FC by examining the concurrence of elevated activity in pairs of brain locations. Our Bayesian model utilizes DTT information as a supplement to fMRI to make inferences regarding task-related functional coherence and may

be applied as a seed-based approach at the voxel level or to evaluate FC globally between all pairs of defined regions of interest. We determine the hierarchy among functionally connected pairs of brain regions based on the associated probabilities of elevated activity for each node. We perform estimation using Markov Chain Monte Carlo (MCMC) techniques. We apply our method to an auditory spatial-cueing task data set and conduct simulation studies to evaluate the performance of our approach.

## 2.2 Data

### 2.2.1 Experimental Data

We consider data from an fMRI study of 32 right-handed adults (15 males, 17 females) who were scanned while performing an auditory spatial-cueing task. The auditory stimuli consist of a series of paired tone pips. In each trial, the first auditory tone (2,000 Hz) may occur in either the left or the right ear and serves as a spatial cue, and the second tone (1,000 Hz) is a target tone and may also occur in either ear. Valid trials are those in which both the cue and the target tones sound in the same ear; otherwise, we call them invalid trials. The chance of either happening is 50%. Targets are equally likely to sound in the right or the left ear, and the order of trials is randomized. A key-press device is positioned under the subject's right hand to record responses. Subjects are instructed to make a key press with their right middle finger for targets arising in the right ear and right index finger for targets arising in the left ear. There are a total of 84 valid and 84 invalid trials presented across three imaging runs. Prior to the start of the experiment, subjects are informed that the cues did not contain any useful information to predict the location of the target. Subjects fixate on a white visual cross on a black background during the entire session. Several objectives of the auditory spatial-cueing task have been addressed previously by Mayer et al. (2009). Our objective here is to identify the functionally connected

brain regions associated with the neural processing stemming from an auditory task (both spatially valid and invalid), which also involves motor and visual functions. Note that our analysis assesses the combined contributions of valid and invalid trials and makes no attempt to dissociate the effects of spatial priming.

## 2.2.2 Image Acquisition and Data Preprocessing

A total of $T^* = 483$ fMRI scans were collected for each of the $N = 32$ subjects, 161 for each of three runs. Two DTI scans were acquired for each subject. We performed several standard preprocessing steps to the functional images using FM-RIB (Functional Magnetic Resonance Imaging of the Brain) Software Library (FSL) (Smith et al., 2004). Specifically, we performed slice timing corrections that align all of the slices from a single 3D image to the same point in time. We performed motion correction. Also, we applied spatial normalization, which realigns images to the standard Montreal Neurological Institute (MNI) space to enable group-level analyses. Pre-whitening was conducted to remove the temporal correlations between scans from the same subject by iteratively estimating the autocorrelation matrix of the residuals to achieve independence through the whole time series (Woolrich et al., 2001).

## 2.2.3 Determining Regional Activity

We consider region-to-region connectivity, rather than voxel-to-voxel. We begin with 90 brain regions defined by the automated anatomic labeling (AAL) system (Tzourio-Mazoyer et al., 2002) excluding the cerebellum. Each region that we consider contains more than 200 voxels. The neural activity throughout each AAL region is typically not uniform. Therefore, we summarize neural activity profiles across defined subregions, rather than across the entire AAL region. The subregion centers are selected as the voxel within each AAL region that is most involved with the auditory task, determined on the basis of a standardized statistic (mean/standard deviation), and

a small sphere containing the roughly 150 closest voxels is grown around each of the selected voxels. The subregion size of 150 voxels corresponds roughly to a sphere with a 6mm radius surrounding the most active voxel, although we do not strictly require a spherical shape, e.g., to address boundary constraints. To obtain a single fMRI temporal profile representing each subregion for each subject, we perform a singular value decomposition (SVD) in the time domain to a $T^* \times 150$ matrix. We extract the first right singular vector, yielding a single temporal profile reflecting the most dominant temporal trend in that subregion. Since singular vectors are unique up to multiplication by a unit phase factor, we compare the singular vector to the subregion mean signal to ensure that the selected signal represents the region correctly, and we apply a sign change to our extracted signal, if necessary.

### 2.2.4 Determining Structural Connectivity

We employ the widely used approach of Behrens et al. (2007), implemented in FSL, to perform probabilistic DTT. We define subregions within the AAL system for DTT, which are centered in white matter proximal to the fMRI-based subregions. Probabilistic tractography successively initiates streams, which are intended to follow or trace the paths of white matter tracts in the brain. The probabilistic tractography implemented in FSL creates a distribution of possible fiber tract pathways, weighted by their likelihood, and according to stopping rules for the streams. A given number of streams (3000 in our analysis) are sent from the seed voxel, and each stream chooses a path based on the principle diffusion direction of the underlying white matter at each voxel and ceases according to a stopping rule. The probabilistic DTT for each pair of regions initially yields voxel-level counts (out of 3000 trials) indicating the likelihood of a fiber tract extending from the voxel in the seed region to (or through) the target region. We use the 90th percentile of the voxel-level counts connecting voxels in the seed region to voxels in the target region to reflect the strongest anatomical connectiv-

Figure 2.1: Histogram of functional joint activation counts at a lower and higher level of structural voxel-level counts for two subjects. Note that the functional joint activation values tend to be larger between region pairs exhibiting high SC relative to low SC.
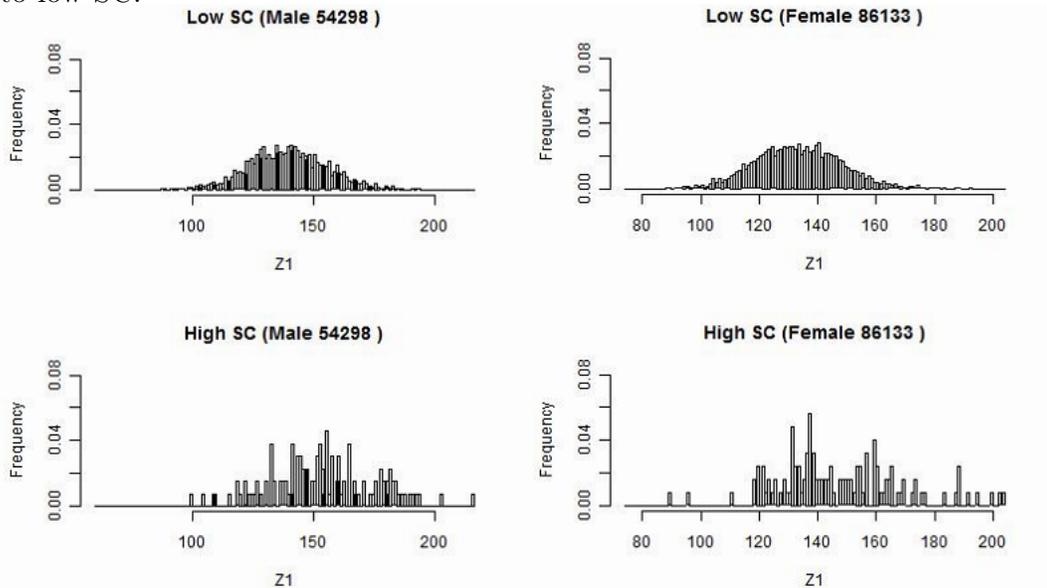


ity between pairs of regions. The voxel-level counts connecting two regions are usually asymmetric; yet for our purpose, we regard FC as a symmetric measure. Therefore, we impose symmetry of SC between two regions by calculating the maximum of the two directional counts for each region pair. To reduce the noise, we analyze both DTI scans and average the resulting SC counts. We adjust the SC counts by the corresponding geometric distances between region centers by fitting a log-linear model on voxel-level SC counts adjusted for the geometric distance between region centers. The adjusted counts are easily converted into probabilities when dividing them by the adjusted number of streams.

## 2.3    Methods

We introduce a statistic $\kappa$ to capture the functional coherence between region pairs and an associated ascendancy measure $\tau$ to quantify the hierarchy of identified coherent regions. A preliminary look at our data reveals that higher levels of SC counts

tend to have associated larger values of functional coherence (see Figure 2.1). Specifically, we examine the distribution of functional coherence at lower and higher levels of voxel-level SC for each subject across 4005 region pairs based on 90 brain regions. Figure 2.1 illustrates the results for selected subjects, but the results across all the subjects reveal similar patterns. Therefore, we build our Bayesian model based on the observation that increased SC is generally associated with higher functional coherence. SC is a static property, whereas FC is a transient characteristic that may vary with the performance of different tasks. Therefore, we do not make the link too strong in our model since high SC may exist without corresponding elevated FC during the auditory task.

### 2.3.1 Joint Activation and Structural Connectivity

Define $A_{gnt} = I\left(R_{gnt} > c \times \sigma_{gn}\right)$, where $R_{gnt} = Y_{gnt} - \hat{\mu}_{gn}$ is the mean-adjusted level of neural activity for region $g$, subject $n$, and scan $t$; $c$ is a constant; and $\sigma_{gn}^2$ is the variance of $Y_{gnt}$, with $n = 1, \cdots, N$ and $t = 1, \cdots, T^*$. Thus, $A_{gnt}$ serves as an indicator of elevated regional brain activity at time $t$. We choose $c = 0.01$ when analyzing the auditory spatial-cueing task.

The joint activation between two regions $a$ and $b$ for subject $n$ can be expressed as:

$$
\begin{aligned}
Z_1^* &= \sum_{t=1}^{T^*} I(A_{ant} = 1, A_{bnt} = 1) \\
Z_2^* &= \sum_{t=1}^{T^*} I(A_{ant} = 1, A_{bnt} = 0) \\
Z_3^* &= \sum_{t=1}^{T^*} I(A_{ant} = 0, A_{bnt} = 1) \\
Z_4^* &= \sum_{t=1}^{T^*} I(A_{ant} = 0, A_{bnt} = 0).
\end{aligned}
\tag{2.1}
$$

$Z_{1n}^*$ is interpreted as the number of times that both regions $a$ and $b$ experience an elevated fMRI signal for subject $n$. The subscripts for subjects are then dropped to simply the notations. We assume $\mathbf{Z}^* = (Z_1^*, \cdots, Z_4^*)'$ follows a multinomial distribution with parameters $T^*$ and $\theta = (\theta_1, \cdots, \theta_4)'$, where

$$
\begin{aligned}
\theta_1 &= P\left(A_{ant} = 1, A_{bnt} = 1\right) \\
\theta_2 &= P\left(A_{ant} = 1, A_{bnt} = 0\right) \\
\theta_3 &= P\left(A_{ant} = 0, A_{bnt} = 1\right) \\
\theta_4 &= P\left(A_{ant} = 0, A_{bnt} = 0\right)
\end{aligned}
\tag{2.2}
$$

To facilitate interpretations across different analyses, we standardize $Z_i^*$ ($i = 1, \cdots, 4$) by scaling it with a specified number of scans. We set the scaling unit to $T = 100$ for our data, so our standardized measure, $Z_i$, is the average number of times that $a$ and $b$ are coherent per one hundred scans. $\mathbf{Z}$ also follows a multinomial distribution with parameters $T$ and $\theta$.

For the anatomical data, we denote the DTT counts between regions $a$ and $b$ for subject $n$ by $S^*$ (omitting the subscripts for simplicity). Let $M^*$ be the number of trials for probabilistic DTT fiber tracking from the voxels in the seed region. We assume that $S^*$ follows a binomial distribution with parameters $M^*$ and $\pi$, where $\pi$ is the probability of SC between regions $a$ and $b$ for any subject. Using similar scaling applied to $Z_i^*$, we generate scaled counts $S$, which follow a binomial distribution with parameters $M$ and $\pi$, where here we choose $M = 1000$.

## 2.3.2 Functional Coherence and Ascendancy

We extend the agreement measure of Patel et al. (2006a), which is based on Cohen's Kappa (Cohen, 1960), to describe functional coherence between pairs of brain regions.

Table 2.1: Joint activation probabilities for regions $a$ and $b$.

|  |  | Region a | |  |
|---|---|---|---|---|
|  |  | Active | Inactive |  |
| Region b | Active | $\theta_1$ | $\theta_3$ | $\theta_1 + \theta_3$ |
|  | Inactive | $\theta_2$ | $\theta_4$ | $\theta_2 + \theta_4$ |
|  |  | $\theta_1 + \theta_2$ | $\theta_3 + \theta_4$ | 1 |

Figure 2.2: Functional ascendancy relationship of four regions $a$, $b$, $c$, and d at different time points. Shading for a given region indicates elevated activity. Region $a$ is functionally connected with regions $b$, $c$ and $d$. Also, region $a$ is ascendant to regions $c$ and $d$.
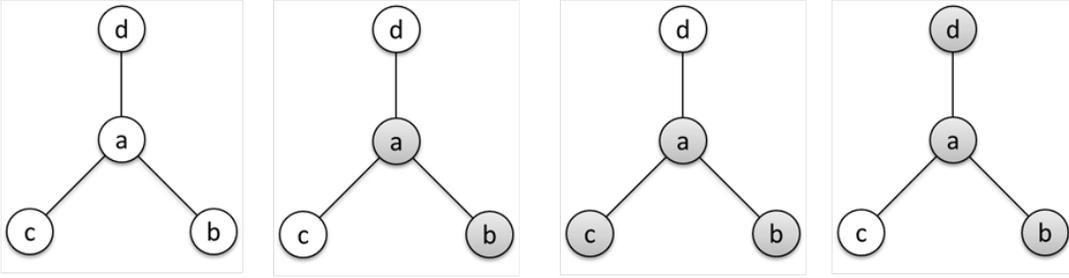


Considering Table 2.1, our functional coherence measure $\kappa$ is defined as:

$$\kappa = \begin{cases} \frac{\theta_1 + \theta_4 - E}{1 - E} & \text{if } \theta_1\theta_4 > \theta_2\theta_3 \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

where $E = (\theta_1 + \theta_2)(\theta_1 + \theta_3) + (\theta_3 + \theta_4)(\theta_2 + \theta_4)$. The numerator of $\kappa$ measures the difference between the probability of coherence and the expected probability of coherence under independence. We restrict our attention to nonnegative values of $\kappa$, so our measure of agreement ranges from 0 to 1. $\kappa$ equals 1 when the probability of joint activation and deactivation $\theta_1$ and $\theta_4$ sums to 1, and hence $\theta_2$ and $\theta_3$ are 0, which indicates complete coherence. If there is no agreement between regions $a$ and $b$, $\kappa = 0$.

Given that $a$ and $b$ are functionally connected, i.e. $\kappa$ exceeds a specified threshold (say $e_\kappa$) with high probability, we define a measure of ascendancy to determine the

hierarchical relationship between the regions. Our ascendancy measure, $\tau_{ab}$, is based on the ratio of $P(A_a = 1)/(1 - P(A_a = 1))$ and $P(A_b = 1)/(1 - P(A_b = 1))$ and takes the following form:

$$\tau_{ab} = \frac{\theta_1 + \theta_2}{\theta_3 + \theta_4} \bigg/ \frac{\theta_1 + \theta_3}{\theta_2 + \theta_4}. \tag{2.4}$$

$\tau_{ab}$ ranges from 0 to $\infty$ and is interpreted as the odds of region $a$ being active relative to the odds of region $b$ being active. Figure 2.2 presents a hierarchical relationship among four regions. We use shading to denote a region exhibiting elevated activity. If two regions $a$ and $b$ become active together and inactive together, we consider them as functionally connected; however, an ascendancy relationship does not necessarily exist between them. Given that region $a$ and region $b$ are functionally connected, we say that $a$ is ascendant to $b$ whenever the marginal odds of activation of $a$ are larger than that of $b$. As shown in Figure 2.2, while regions $a$, $b$, $c$, and $d$ are functionally connected, regions $c$ and $d$ exhibit elevated activity for a subset of the stimuli for which $a$ exhibits elevated activity, suggesting that $a$ is ascendant to $c$ and $d$. Given $\kappa > e_\kappa$, $\tau_{ab} > 1$ indicates that $a$ is ascendant to $b$, while $\tau_{ab} < 1$ indicates that $b$ is ascendant to $a$.

### 2.3.3 Bayesian Model

For any pair of regions $a$ and $b$, the likelihood function for our data takes the form:

$$p(\mathbf{Z}, S | \theta, \pi) \propto \prod_{i=1}^{4} \theta_i^{\sum_{n=1}^{N} Z_{in}} \pi^{\sum_{n=1}^{N} S_n} (1 - \pi)^{N \times M - \sum_{n=1}^{N} S_n}. \tag{2.5}$$

Following the approach by Patel et al. (2006a), we assume that each repeated measure on the same region pair is independent across subjects. We also assume that each repeated measure on the same region pair is independent over time since we have performed pre-whitening in the pre-processing step to remove the temporal correlations between scans from the same subject. In addition, given both probability measure-

ments $\boldsymbol{\theta}$ and $\pi$, $S$ is independent of $\mathbf{Z}$ because we build structure-function dependence in the distribution of $[\boldsymbol{\theta}|\pi]$. This is a conditional independence assumption between $S$ and $\mathbf{Z}$, but marginally our model still captures the dependence between $S$ and $\mathbf{Z}$ through the corresponding parameters $\boldsymbol{\theta}$ and $\pi$.

Using a Bayesian formulation, we express our prior belief about structural connection probabilities $\pi$ by defining a beta prior which takes the form:

$$p(\pi) \propto \pi^{\alpha_0-1}(1-\pi)^{\beta_0-1}. \tag{2.6}$$

We specify a flat prior for our DTT data by setting $\alpha_0 = \beta_0 = 1$ for each region pair, suggesting no available prior information regarding the SC of any region pair. In our simulation studies, we evaluate the performance of our method under different combinations of the hyperparameters $\alpha_0$ and $\beta_0$.

The prior for $\boldsymbol{\theta}$ is taken to follow a Dirichlet distribution with parameters $(\alpha(\pi) + \alpha_1, \alpha_2, \alpha_3, \alpha_4)'$, where $\alpha(\pi)$ is a function of $\pi$ and reflects the assumed relationship between FC and SC. Specifically, we assume that

$$p(\theta|\pi) \propto \frac{\Gamma(\alpha(\pi) + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)}{\Gamma(\alpha(\pi) + \alpha_1)} \theta_1^{\alpha(\pi)+\alpha_1-1} \theta_2^{\alpha_2-1} \theta_3^{\alpha_3-1} \theta_4^{\alpha_4-1}. \tag{2.7}$$

We set $\alpha_1 = 5$, $\alpha_2 = \alpha_3 = \alpha_4 = 10$, and $\alpha(\pi) = 1/\left(\frac{9}{\ln(10)} - 1\right) \times 10^\pi - 1/\left(\frac{9}{\ln(10)} - 1\right)$, so the average value of $\alpha(\pi)$ on $\pi \in [0,1]$ is 10. In this case, $\theta_2, \ldots, \theta_4$ have the same expected values and moderate variances. Our prior is based on the observation that weak SC corresponds to relatively few joint functional activations, and extremely strong SC is assumed to yield an expected value of $\theta_1$ to be around 0.5, which is approximately the highest maximum likelihood estimate of $\theta_1$. When $\alpha(\pi)$ is an increasing function, the expected value of $\theta_1$, which takes the form of $(\alpha(\pi) + \alpha_1)/(\alpha(\pi) + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)$ is also an increasing function with respect to $\pi$; thus, matching our observation from the data. Later, we present results from a

sensitivity analysis of our choice of $\alpha(\pi)$, which shows that our results do not change much with respect to different functions of $\alpha(\pi)$.

### 2.3.4 Posterior Sampling

We regard the model parameters corresponding to different region pairs as independent, so for each region pair, we express the posterior distribution as:

$$p(\boldsymbol{\theta}, \pi | \mathbf{Z}, S) \propto \frac{\Gamma(\alpha(\pi) + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)}{\Gamma(\alpha(\pi) + \alpha_1)} \tag{2.8}$$

$$\pi^{\sum_{n=1}^{N} S_n + \alpha_0 - 1} (1 - \pi)^{N \times M - \sum_{n=1}^{N} S_n + \beta_0 - 1} \theta_1^{\alpha(\pi)} \prod_{i=1}^{4} \theta_i^{\sum_{n=1}^{N} Z_{in} + \alpha_i - 1}.$$

The full conditional posterior distributions take the form:

$$p(\pi | \boldsymbol{\theta}, \mathbf{Z}, S) \propto \frac{\Gamma(\alpha(\pi) + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)}{\Gamma(\alpha(\pi) + \alpha_1)} \pi^{\sum_{n=1}^{N} S_n + \alpha_0 - 1} (1 - \pi)^{N \times M - \sum_{n=1}^{N} S_n + \beta_0 - 1} \theta_1^{\alpha(\pi)}$$

$$\tag{2.9}$$

and

$$p(\theta | \pi, \mathbf{Z}, S) \sim \text{Dirichlet}(\sum_{n=1}^{N} Z_{1n} + \alpha(\pi) + \alpha_1, \sum_{n=1}^{N} Z_{2n} + \alpha_2, \sum_{n=1}^{N} Z_{3n} + \alpha_3, \sum_{n=1}^{N} Z_{4n} + \alpha_4).$$

$$\tag{2.10}$$

Estimation of $\kappa$ and $\tau$ are based on the posterior distribution, $p(\boldsymbol{\theta}|\mathbf{Z}, S)$, as we are able to estimate $p(\kappa|\mathbf{Z}, S)$ and $p(\tau|\mathbf{Z}, S)$ by sampling $p(\boldsymbol{\theta}, \pi|\mathbf{Z}, S)$. We set effect sizes for $\kappa$, denoted $e_\kappa$, and $\tau$, denoted $e_\tau$, in our analysis. These are used to compute exceedance probabilities $P(\kappa > e_\kappa \mid \mathbf{Z}, S)$ and $P(\tau > e_\tau \mid \mathbf{Z}, S)$ from our modeling framework. For $\kappa$, we choose $e_\kappa = 0.3$, which reflects fair to moderate agreement (Landis and Koch, 1977) and is above the 95th percentile across all region pairs. Since the range of $\tau$ is extremely narrow, we explore the ascendancy relationship when $\tau$ is above 75th percentile across all region pairs. We conduct inference on $\kappa$ and $\tau$ by estimating $P(\kappa > e_\kappa \mid \mathbf{Z}, S) > p_\kappa$ and $P(\tau > e_\tau \mid \mathbf{Z}, S) > p_\tau$, respectively.

We set $p_\kappa$ as 0.95, which is also above 95th percentile among $P(\kappa > e_\kappa \mid \mathbf{Z}, S)$, and we choose $p_\tau$ as 0.5, which is above 80th percentile of $P(\tau > e_\tau \mid \mathbf{Z}, S)$. The choices of effect sizes and probability thresholds are made to reflect characteristics of functional coherence and ascendancy, rather than on the basis of statistical properties. For example, $e_\kappa$ may be set to 0.3 as in our case to reflect moderate coherence or to a higher value, e.g., above 0.6 to reflect substantial agreement. For the probability thresholds, we suggest selections between the 80th percentile and 95th percentile. Ideally, these thresholds should be determined before performing the data analysis. However, due to the complexity and variability of neuroimaging data, the user has the flexibility to investigate connections between regions at different levels of these thresholds.

As revealed by the conditional posterior distribution of $\theta$, our prior belief has an impact on the posterior through $\alpha(\pi)$, but does not drive the direction of results. As the SC $\pi$ becomes stronger, the expected value of $\theta_1$ increases, and the expected values of $\theta_2, \cdots, \theta_4$ decrease accordingly, but at a slower rate.

We estimate our Bayesian hierarchical model using MCMC methods, implemented via the Gibbs sampler with an embedded Metropolis step. The parameter $\boldsymbol{\theta}$ is updated from a Dirichlet distribution with $\pi$ specified from the previous step. The parameter $\pi$ is updated by $\pi^*$, which is sampled from a Normal jumping distribution $J_t(\pi^* | \pi^{t-1}) = N(\pi^* | \pi^{t-1}, \tau^2)$ at time $t$, with probability $\min(r, 1)$, where $r$ is the ratio of the conditional densities from time $t-1$ to the conditional densities of the proposed value with respect to $\boldsymbol{\theta}$,

$$r = \frac{p(\pi^* | \theta, Z, S)}{p(\pi^{t-1} | \theta, Z, S)}. \tag{2.11}$$

The variance $\tau^2$ in the Normal jumping distribution is adjusted based on the data, which ensures the acceptance ratio close to 25% to achieve the optimal efficiency of Metropolis algorithm (Gelman et al., 1995).

## 2.4 Results

### 2.4.1 Auditory Data Results

We apply our Bayesian model to the auditory processing fMRI data to determine functionally connected regions and also to examine the corresponding ascendancy relationship to illustrate the neural integration underlying auditory and related processing. We find strongly connected regions within the auditory cortex with associated hierarchical relationships between these functionally connected pairs of regions. For example, Heschl's gyrus (both left and right) is ascendant to the left superior temporal gyrus, and the right pole of the superior temporal gyrus is ascendant to both the middle temporal gyrus (both left and right) and the right pole of the middle temporal gyrus, with $P\left(\kappa > 0.3 \mid \mathbf{Z}, S\right) > 0.95$ and $P\left(\tau > e_\tau \mid \mathbf{Z}, S\right) > 0.5$ in each case (Figure 2.3). We also detect high posterior probabilities of functional coherence and ascendancy between regions within the primary auditory cortex and auditory-speech encoding regions. Specifically, we find that Heschl's gyrus (both left and right) is functionally connected and ascendant to the left Rolandic operculum, which may reflect the use by many people of language encoding areas during auditory processing involving perception of the spatial location of a tone. The design of the spatial-cueing auditory experiment calls for subjects to momentarily remember the location of the target tone before they press the button. Many people make use of visualization to aid the memory process, especially spatial memory, which may in turn activate the visual cortex (Ungerleider, 1995). Our findings support this relationship by revealing strong FC between the memory related regions and the visual cortex, especially the visual-spatial system. For example, the right lingual gyrus is ascendant bilaterally to the posterior cingulate gyrus, right hippocampus, and right parahippocampal gyrus. Our results also reflect the neural processing related to button presses in response to the target cue, revealing strong functional coherence between the right pallidum and

left thalamus as well as between the right thalamus and the right posterior cingulate gyrus, all of which are included in motor and sensory systems. In addition, high probabilities of functional coherence are detected within the visual cortex, e.g. between left and right lingual gyrus, left and right middle occipital region, and also right lingual and right precuneus gyrus, which identify associations in neural processing regions when the subjects perform visual fixation.

We also detect relatively strong SC underlying the identified functional networks. The estimated probabilities of SC for each of the functionally connected region pairs above exceeds the 93rd percentile of SC for all 4005 brain region pairs, with most of them having SC probabilities greater than 0.5. Compared to a naive estimator that quantifies the proportion of simultaneously active or inactive states among all scans and across all subjects, weak SC in our model tends to reduce the functional coherence, while strong SC is more likely accompanied by a stronger functional relationship, though the difference is relatively small. Although little evidence of SC reduces the chances of having associated FC in our model, it does not preclude detection of FC. Similarly, strong SC does not necessarily imply that two regions exhibit functionally connectivity.

For comparison, we apply the approach of Patel et al. (2006a) to fMRI data from the auditory spatial cueing task. The Patel method is not able to capture the ascendancy relationship between left superior temporal gyrus and left Heschl's gyrus, which is an important region pair within the auditory cortex. Another major difference is that more region pairs within the visual cortex are detected by the original approach. Although visualization is an essential component of the experimental design, we do not expect to see the high degree of connectivity within the visual cortex because the task only involves fixation on a cross hair, with more engaging auditory processing. This leads us to conjecture that - without structural information, more false positives are generated while the major findings may go undetected.

Figure 2.3: Ascendancy maps with $P(\kappa > 0.30) > 0.95$ and $P(\tau > e_\tau) > 0.50$ from three different views. Arrow from region $a$ to $b$ means that region $a$ is ascendant to $b$. Auditory cortex, visual cortex, memory systems and sensory systems are included in the maps.
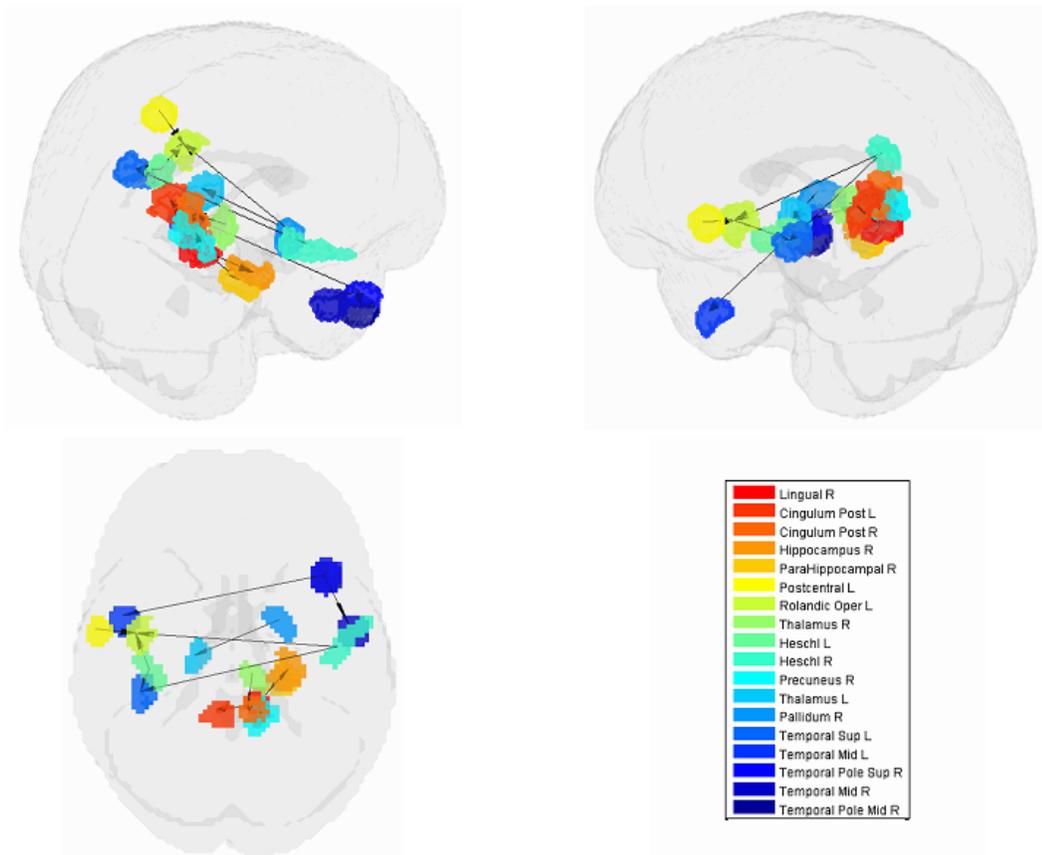
## 2.4.2 Simulation Results

We conduct two simulation studies, one to compare our method with the approach of Patel et al. (2006a) and the second to compare our approach to a traditional correlation-based analysis. In addition, we conduct a sensitivity analysis to evaluate the impact of various choices of the function $\alpha(\pi)$, which links FC and SC, on the performance of our approach.

We first discuss simulation results from the comparison our combined fMRI and DTI approach (labeled as the *FC with SC* method) with the approach by Patel et al. (2006a) utilizing fMRI data only (labeled as the *FC only* method). We generate data $\mathbf{Z}$ and $S$ from our model with different settings of hyperparameters $\alpha_0$ and $\beta_0$ in the prior distribution of $\pi$. We evaluate the methods by comparing the bias of the corresponding posterior means of $\boldsymbol{\theta}$, $\kappa$ and $\tau$. For each simulation setting, 10 $\pi$'s are generated, and 10 $\theta$'s are simulated for each $\pi$, 100 data sets are generated from each set of $\theta$'s. Therefore, a total of 10,000 data sets are simulated to compute the mean biases. The results indicate that our method performs better in all settings with smaller bias. Table 2.2 shows that the mean bias of $\theta_i$'s from our method is smaller than that from *the FC only* method in every case, which indicates that incorporating structural information improves the estimation of FC. Since our definitions of $\kappa$ and $\tau$ are both functions of $\theta_i$'s, it follows that our model also outperforms *the FC only* method for estimating these measures as defined in (2.3) and (2.4). Alternatively, we compare estimation performance of $\kappa$ and $\tau$ from *the FC only* method based on the original definitions from Patel et al. (2006a) with estimation of our extended definitions of $\kappa$ and $\tau$ under our *FC with SC* approach. We contrast how these two methods address functional coherence/association and ascendancy. The standard deviation of the bias also yields similar conclusions favoring our combined *FC with SC* approach over *the FC only* method (see Table 2.3). We examine the performance of our method using samples sizes of 15, 30 and 100 subjects. Although the difference

Table 2.2: Comparison of mean of bias between two Bayesian methods. The table reveals the improvements of FC with SC from FC only in terms of the mean of bias.

| $\alpha_0$ | $\beta_0$ | $E(\pi)$ | FC with SC$(\times 10^{-3})$ | | | | | | FC only$(\times 10^{-3})$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\kappa$ | $\tau$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\kappa$ | $\tau$ |
| N=30 | | | | | | | | | | | | | | |
| 1 | 100 | 0.01 | 4.925 | 6.550 | 6.394 | 6.391 | 2.185 | 30.759 | 4.963 | 6.587 | 6.432 | 6.416 | 233.875 | 116.591 |
| 2 | 18 | 0.1 | 5.127 | 6.361 | 6.282 | 6.469 | 2.791 | 29.577 | 5.173 | 6.427 | 6.307 | 6.508 | 188.989 | 117.824 |
| 2 | 5 | 0.3 | 5.979 | 6.292 | 6.206 | 6.211 | 5.702 | 29.458 | 6.017 | 6.330 | 6.247 | 6.247 | 116.983 | 127.239 |
| 2 | 2 | 0.5 | 6.271 | 6.044 | 6.045 | 5.952 | 7.556 | 30.224 | 6.321 | 6.106 | 6.077 | 5.998 | 112.727 | 119.231 |
| 5 | 2 | 0.7 | 7.007 | 5.777 | 5.653 | 5.692 | 12.085 | 29.761 | 7.078 | 5.829 | 5.711 | 5.756 | 57.719 | 123.946 |
| 18 | 2 | 0.9 | 7.150 | 5.387 | 5.330 | 5.390 | 13.531 | 28.887 | 7.242 | 5.434 | 5.376 | 5.442 | 80.766 | 139.150 |
| N=100 | | | | | | | | | | | | | | |
| 1 | 100 | 0.01 | 2.704 | 3.536 | 3.597 | 3.534 | 1.120 | 17.098 | 2.710 | 3.541 | 3.597 | 3.545 | 232.029 | 115.003 |
| 2 | 18 | 0.1 | 2.996 | 3.483 | 3.481 | 3.532 | 2.103 | 16.705 | 3.006 | 3.492 | 3.485 | 3.538 | 181.038 | 114.898 |
| 2 | 5 | 0.3 | 3.139 | 3.435 | 3.377 | 3.380 | 2.717 | 16.048 | 3.150 | 3.446 | 3.389 | 3.385 | 153.390 | 126.220 |
| 2 | 2 | 0.5 | 3.540 | 3.362 | 3.266 | 3.315 | 4.950 | 16.250 | 3.541 | 3.362 | 3.273 | 3.327 | 93.892 | 127.077 |
| 5 | 2 | 0.7 | 3.748 | 3.172 | 3.122 | 3.066 | 6.080 | 15.228 | 3.761 | 3.177 | 3.131 | 3.077 | 78.071 | 161.000 |
| 18 | 2 | 0.9 | 3.939 | 2.938 | 3.024 | 2.920 | 7.495 | 16.320 | 3.959 | 2.949 | 3.037 | 2.931 | 84.696 | 144.013 |
| N=15 | | | | | | | | | | | | | | |
| 1 | 100 | 0.01 | 6.882 | 8.906 | 9.006 | 9.078 | 2.874 | 44.578 | 6.951 | 9.011 | 9.154 | 9.183 | 221.313 | 106.883 |
| 2 | 18 | 0.1 | 7.39 | 9.073 | 8.851 | 8.822 | 4.442 | 42.553 | 7.461 | 9.162 | 8.993 | 8.914 | 204.43 | 122.885 |
| 2 | 5 | 0.3 | 8.565 | 8.861 | 8.544 | 8.58 | 8.545 | 41.804 | 8.667 | 9.038 | 8.636 | 8.689 | 120.509 | 126.061 |
| 2 | 2 | 0.5 | 9.239 | 8.333 | 8.484 | 8.499 | 13.308 | 41.421 | 9.297 | 8.44 | 8.63 | 8.644 | 71.363 | 128.054 |
| 5 | 2 | 0.7 | 9.798 | 8.146 | 8.086 | 8.196 | 16.415 | 40.797 | 9.862 | 8.301 | 8.174 | 8.316 | 59.667 | 136.606 |
| 18 | 2 | 0.9 | 10.056 | 7.379 | 7.52 | 7.313 | 19.624 | 41.791 | 10.233 | 7.503 | 7.68 | 7.457 | 88.121 | 136.543 |

Table 2.3: Comparison of mean of bias between two Bayesian methods. The table reveals the improvements of FC with SC from FC only in terms of the standard deviation of bias.

| $\alpha_0$ | $\beta_0$ | $E(\pi)$ | FC with SC$(\times 10^{-3})$ | | | | | | FC only$(\times 10^{-3})$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\kappa$ | $\tau$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\kappa$ | $\tau$ |
| N=30 | | | | | | | | | | | | | | |
| 1 | 100 | 0.01 | 3.817 | 4.930 | 4.889 | 4.906 | 6.049 | 25.263 | 3.857 | 4.950 | 4.911 | 4.944 | 191.969 | 64.374 |
| 2 | 18 | 0.1 | 4.022 | 4.817 | 4.806 | 4.961 | 7.070 | 24.206 | 4.066 | 4.851 | 4.827 | 4.994 | 160.240 | 62.968 |
| 2 | 5 | 0.3 | 4.581 | 4.730 | 4.767 | 4.772 | 9.269 | 24.365 | 4.600 | 4.757 | 4.806 | 4.783 | 131.978 | 69.942 |
| 2 | 2 | 0.5 | 4.806 | 4.602 | 4.620 | 4.636 | 10.160 | 24.665 | 4.858 | 4.655 | 4.647 | 4.669 | 113.153 | 76.524 |
| 5 | 2 | 0.7 | 5.277 | 4.381 | 4.375 | 4.411 | 10.745 | 23.596 | 5.340 | 4.417 | 4.412 | 4.438 | 45.638 | 80.886 |
| 18 | 2 | 0.9 | 5.411 | 4.133 | 4.060 | 4.183 | 10.798 | 22.775 | 5.481 | 4.171 | 4.105 | 4.213 | 48.692 | 88.015 |
| N=100 | | | | | | | | | | | | | | |
| 1 | 100 | 0.01 | 2.148 | 2.726 | 2.753 | 2.680 | 3.299 | 14.407 | 2.156 | 2.734 | 2.758 | 2.688 | 183.179 | 64.847 |
| 2 | 18 | 0.1 | 2.315 | 2.683 | 2.637 | 2.691 | 4.586 | 13.796 | 2.318 | 2.687 | 2.640 | 2.694 | 157.503 | 67.204 |
| 2 | 5 | 0.3 | 2.428 | 2.628 | 2.631 | 2.589 | 4.949 | 13.279 | 2.436 | 2.632 | 2.636 | 2.596 | 129.687 | 69.334 |
| 2 | 2 | 0.5 | 2.729 | 2.560 | 2.500 | 2.570 | 5.964 | 13.173 | 2.738 | 2.564 | 2.502 | 2.583 | 109.046 | 76.622 |
| 5 | 2 | 0.7 | 2.817 | 2.440 | 2.430 | 2.354 | 5.991 | 12.225 | 2.821 | 2.442 | 2.433 | 2.359 | 64.481 | 79.986 |
| 18 | 2 | 0.9 | 2.992 | 2.274 | 2.337 | 2.248 | 6.085 | 13.011 | 3.001 | 2.282 | 2.347 | 2.257 | 48.929 | 87.404 |
| N=15 | | | | | | | | | | | | | | |
| 1 | 100 | 0.01 | 5.457 | 6.792 | 6.827 | 6.954 | 8.146 | 36.287 | 5.536 | 6.890 | 6.952 | 7.011 | 170.042 | 65.855 |
| 2 | 18 | 0.1 | 5.783 | 6.945 | 6.757 | 6.852 | 10.037 | 34.913 | 5.854 | 6.976 | 6.869 | 6.936 | 166.338 | 70.194 |
| 2 | 5 | 0.3 | 6.576 | 6.733 | 6.537 | 6.431 | 13.389 | 33.654 | 6.610 | 6.824 | 6.633 | 6.502 | 122.500 | 70.714 |
| 2 | 2 | 0.5 | 7.072 | 6.271 | 6.575 | 6.482 | 14.457 | 33.382 | 7.131 | 6.385 | 6.704 | 6.594 | 78.412 | 75.823 |
| 5 | 2 | 0.7 | 7.339 | 6.267 | 6.200 | 6.253 | 15.060 | 32.658 | 7.407 | 6.354 | 6.268 | 6.393 | 44.905 | 78.162 |
| 18 | 2 | 0.9 | 7.649 | 5.681 | 5.792 | 5.631 | 15.049 | 32.852 | 7.762 | 5.807 | 5.886 | 5.728 | 51.656 | 90.954 |

in bias between the two approaches is relatively small, our FC with SC approach outperforms the FC only approach in every case that we consider.

Our second simulation study compares our method to a traditional correlation analysis. We use the same simulated $\boldsymbol{\theta}$ from the previous simulation study to generate the neural activity profiles $Y_{ant}$ and $Y_{bnt}$, for regions $a$ and $b$, respectively, from a bivariate normal distribution with variances $\sigma_a^2$ and $\sigma_b^2$ and correlation $\rho$. Thus, the mean adjusted level of neural activity profiles $R_{ant}$ and $R_{bnt}$ also follow a bivariate

Figure 2.4: Relationship between $\kappa$ and $\rho$. A positive linear relationship is detected for three cases with different sample sizes.
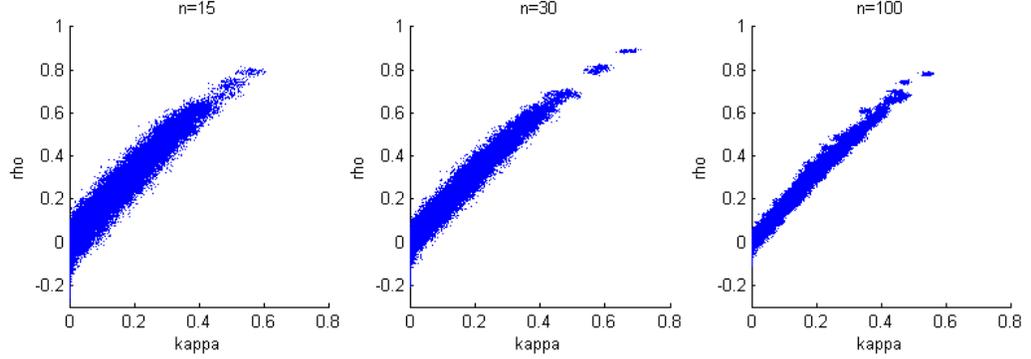


normal distribution. We derive the expectation of $Z_i$'s as follows:

$$E(Z_1) = P(Z_1 = 1) = P(R_{ant} > c_a, R_{bnt} > c_b) = \theta_1$$

$$E(Z_2) = P(Z_2 = 1) = P(R_{ant} > c_a, R_{bnt} < c_b) = \theta_2$$

$$E(Z_3) = P(Z_3 = 1) = P(R_{ant} < c_a, R_{bnt} > c_b) = \theta_3 \tag{2.12}$$

$$E(Z_4) = P(Z_4 = 1) = P(R_{ant} < c_a, R_{bnt} < c_b) = \theta_4.$$

We solve for $c_a$ and $c_b$ using the marginal probabilities of $R_{ant}$ and $R_{bnt}$, which are functions of $\boldsymbol{\theta}$, and we subsequently solve for $\rho$ using any of the above equations. We estimate the Pearson correlation coefficient $\rho$ from our simulated data and compare it to the functional coherence $\kappa$. We expect to see substantial correspondence between these two measures since they both capture the functional associations between two regions.

We generate results for all combinations of hyperparameters specified in the first simulation study and again consider sample sizes of 15, 30 and 100. The accuracy of the estimation is not heavily influenced by the variance of the bivariate normal distribution; therefore, we present results when $\sigma_a^2 = \sigma_b^2 = 0.2$, which corresponds to estimates from our experimental data. We find a positive linear relationship between $\kappa$ and $\rho$ (Figure 2.4), while larger sample size yields smaller variability in the estimates.

Table 2.4: The bias of $\boldsymbol{\theta}$ from estimations of different functions of $\alpha(\pi)$. Note that not much difference is detected across different functions with the same structural connectivity. The bias is calculated from the sum of the bias in all $\theta_i$'s. Here, $f(a) = 10 \times (a+1) \times \pi^a$, $g(a) = 10 \big/ \left( \frac{a-1}{\ln(a)} - 1 \right) \times a^\pi - 10 \big/ \left( \frac{a-1}{\ln(a)} - 1 \right)$

| | Generate Data from $g(10)$ | | | | | Generate Data from $g(0.01)$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\pi$ | $g(10)$ | $f(1.5)$ | $f(0.5)$ | $g(0.01)$ | $\pi$ | $g(10)$ | $f(1.5)$ | $f(0.5)$ | $g(0.01)$ |
| 0.2045 | 0.0024 | 0.0026 | 0.0047 | 0.0051 | 0.0325 | 0.0018 | 0.0017 | 0.0023 | 0.0021 |
| 0.4626 | 0.0034 | 0.0035 | 0.004 | 0.0042 | 0.164 | 0.0035 | 0.0034 | 0.0026 | 0.0027 |
| 0.6894 | 0.0041 | 0.0038 | 0.0043 | 0.0043 | 0.4724 | 0.0038 | 0.0037 | 0.0035 | 0.0034 |
| 0.8987 | 0.0066 | 0.0073 | 0.0086 | 0.0090 | 0.7031 | 0.0059 | 0.0058 | 0.0061 | 0.0059 |

Finally, we examine the impact of $\alpha(\pi)$ on the estimation of $\boldsymbol{\theta}$. As the data suggests, the functional coherence tends to increase as the SC increases. Therefore, we use four different increasing functions for the parameter $\alpha(\pi)$ as a parameter of the prior distribution of $\boldsymbol{\theta}$, based on power functions and exponential functions. Figure 2.5 shows the functions that we consider in the posterior simulations. As each function has the same integration over the interval [0,1], the expected values of all the $\theta_1$'s simulated from each function are the same. We choose specific forms of functions $f$ and $g$ to satisfy the above criterion. We consider both convex functions, i.e., $g\,(10)$ and $f\,(1.5)$, and concave functions, i.e., $f\,(0.5)$ and $g\,(0.01)$. Table 2.4 summarizes the biases of $\boldsymbol{\theta}$ estimated using different functions, where the true values are generated from the two most extreme cases $g\,(10)$ and $g\,(0.01)$. Here, the bias is calculated from the sum of the biases in all $\theta_i$'s. We also vary the probability of SC, $\pi$, from weak to strong in the simulation study. The results indicate that the biases of $\boldsymbol{\theta}$ across all tested functions are comparable. Thus, we conclude that our method is not very sensitive to the choice of $\alpha(\pi)$, among those considered.

## 2.5 Discussion

We build a unified Bayesian framework that provides a novel approach to combine functional and structural brain imaging for an integrated assessment of FC. Joint

Figure 2.5: Functions that are used in the sensitivity analysis of $\alpha(\pi)$. All of them are increasing functions with respect to $\pi$ and have the same area under curve.
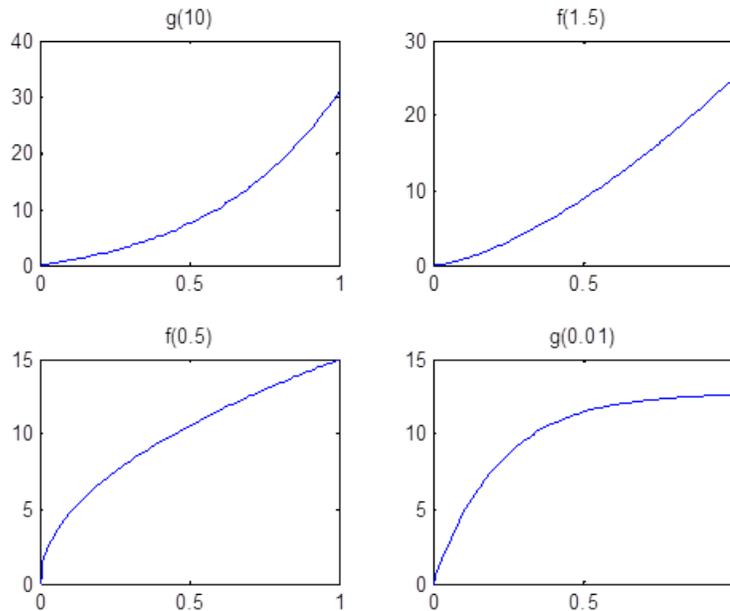


analysis of both imaging modalities is an important tool to gain a better understanding of sensory and cognitive functions in the human brain as well as pathophysiology associated with psychiatric or neurologic disorders. Different from other methods that examine both FC and SC in a descriptive way, our method incorporates SC information into the model and allows for making statistical inferences. Our Bayesian model utilizes SC information to estimate the functional coherence between pairs of regions, yet our model does not allow the structural information to unduly drive the functional results. Our method is a purely data-driven, hypothesis-unconstrained approach, which can efficiently search across all pairs of defined brain regions of interest. We develop two measures, $\kappa$ and $\tau$, to capture the functional coherence and degree of ascendancy, between the brain regions. $\kappa$ is based on the probability of joint activation and deactivation, while $\tau$ assesses the ascendancy between functionally connected regions, enabling us to construct a directed brain network.

Our method permits analyses examining all 4005 possible brain region pairs, be-

tween 90 AAL regions, to construct FC networks. We conduct a whole-brain analysis rather than requiring a pre-defined network or regions of interest. We use one temporal neural activity profile for each AAL region. Unlike other methods that average data across an entire region, our method is based on subregions centered on the most active voxel, yielding neural activity profiles that are representative of the brain activity within the small spherical subregions. Alternatively, we may allow multiple subregions from each AAL region, which would provide more complete coverage in our whole-brain analyses. It takes several hours to complete the computations required to implement of our method for 90 regions. We anticipate that it will readily extend to accommodate more regions, e.g. from permitting multiple subregions. However, as the number of subregions increases toward the number of voxels, computations would become more challenging, given the massive number of subregion pairs.

Our model assumes that the functional connection between one region pair is independent of the connections between other region pairs. This is a common assumption in imaging studies. Although there maybe departures from this assumption reflected in the data, it is difficult to model these potential correlations between pairs of region pairs since the specific nature of the dependence structure is unclear and because computation may become prohibitive as it would require estimation of approximately $R^4/8$ correlations, which grows quickly as the number of regions $R$ increases.

We dichotomize the time series data to define functional joint activations, from which we evaluate the functional connectivity using our proposed $\kappa$ metric that differs from the traditional correlation approach. There is no scale for the fMRI signal in the human brain that lends itself to natural interpretations of the level of neural activity. We attempt to define high and low neural activity from the fMRI signal based on a selected threshold. A possible extension to our current method is to use finer categories such as ordinal or even continuous measures to define joint functional activations.

In the context of dichotomizing the data, our modeling framework depends on $c$ to declare elevated and inactive states of neural activity. Drastic changes to $c$ will have a direct impact on the data that are produced and input to our model, e.g., by setting $c$ sufficiently high (or low), all the regions will become inactive (or active). We choose an arbitrary value of $c$ when determining the indicator of elevated brain activity for regional fMRI profiles. We use a small value of $c$ ($c = 0.01$) because the indicator is based on the mean adjusted level of brain activity, and we would like to capture as much information as possible; therefore, we set a low threshold. As a result, about 45% of the time points are determined as active states. We conduct sensitivity analysis of choice of $c$, and find that when $c$ fluctuates within a small range, i.e. for $c$ up to 0.1, the major findings in our data application do not change.

We propose a functional coherence measurement that builds on Cohen's $\kappa$-statistic, which evaluates the levels of agreement adjusted for chance. Here, the chance agreement is defined as $(\theta_1 + \theta_2)(\theta_1 + \theta_3) + (\theta_3 + \theta_4)(\theta_2 + \theta_4)$. In addition to Cohen's $\kappa$, we may consider other agreement measures, e.g., Scott's $\pi$-statistic (Scott, 1955), in which the chance agreement is obtained by $[((\theta_1 + \theta_2) + (\theta_1 + \theta_3))/2]^2 + [((\theta_3 + \theta_4) + (\theta_2 + \theta_4))/2]^2$; Fleiss' $\kappa$-statistic (Fleiss, 1971), which is a generalization of Scott's $\pi$; and other alternative chance-corrected statistics (Gwet, 2002). The major difference among these statistics is the way they calculate the chance agreement. Some researchers (Gwet, 2002) argue that the conditions that Cohen's $\kappa$ requires, e.g., the chance-agreement probability is less than 0.5, are not always met in practice. In our case, however, the sum of the marginal probabilities $P(A_a = 1) + P(A_b = 1)$ is close to 1, which ensures that the chance-agreement probability does not exceed 0.5. In addition, other statisticians (Strijbos et al., 2006) believe that when fewer categories are included, Cohen's $\kappa$ is a more conservative measurement of agreement. Therefore, we use this more strict measurement in our case.

The study of functional connections in the human brain is important to understand

basic cognition, mental and neurological disorders, and response to treatments for these disorders. Moreover, the structural circuitry underlying functional connections may offer additional insights. We develop a Bayesian model that combines both functional and structural information to help characterize FC networks. Leveraging SC to quantify FC, our model yields more accurate and more informative results than considering solely functional data.

# Chapter 3

# Identifying Functional Co-activation Patterns in Neuroimaging Studies via Poisson Graphical Models

## 3.1  Introduction

It is fundamental to our understanding of brain function to study the interactions among different brain regions. Neuroscience has established functional activation as a principle of brain organization in humans. The integration of activation regions has been proven to be more changeling to assess. One question of interest is to estimate the statistical dependences between activation regions. This characterizes functional connectivity of brain and leads to an estimate of the brain's functional network. There are various analytical methods of functional connectivity for functional neuroimaging studies. Typically, for a functional magnetic resonance imaging (fMRI) study, one can estimate the spatial correlation, partial correlation or mutual information between the

time series of blood oxygenate level dependent (BOLD) signals. In this article, we perform functional connectivity analysis from a different perspective by identifying functional co-activation patterns across independently performed functional imaging studies.

The number of functional imaging studies has risen dramatically in recent years. However, due to the relatively small sample size of most imaging studies, many problems such as inflated false positive rates and low reproducibility (Thirion et al. 2007) have emerged. Meta-analysis, which provides solutions to these problems by combing the results from independent studies, addresses this issue in a unique and important way. Functional neuroimaging studies rarely report entire statistical parametric maps (SPMs), but instead only report the peak activation coordinates. We refer to these as foci, or a single focus. Thus, the most popular approach is coordinate-based meta-analysis (Fox et al., 1997; Nielsen and Hansen, 2002; Turkeltaub et al., 2002; Wager et al., 2004; Kober et al., 2008; Eickhoff et al., 2009; Radua and Mataix-Cols, 2009; Kang et al., 2011), where multilevel density kernel analysis (MKDA) (Wager et al., 2007a) and modified activation likelihood estimation (ALE) (Eickhoff et al., 2009) are two commonly used approaches. Both of these construct a statistical map for each study and produce the consistency of activations across studies. Kang et al. (2011) provide a much richer inference on the population level activation locations from a Bayesian perspective.

In addition to identifying the consistency of activations in imaging studies, a few approaches have been proposed to address another interesting problem, i.e., the evaluation of co-activation patterns, which provide information about functional connectivity between brain regions. Nielsen and Hansen (2004) proposed a matrix factorization algorithm, similar to principal component analysis, to identify the distributed patterns in the brain. Kober et al. (2008) proposed a functional grouping approach, which analyzes the spatial density of reported foci using MKDA, and then combines

non-metric multidimensional scaling and cluster analysis to group regions based on their co-activation patterns. A structural learning approach for the Bayesian network is developed to construct a directed functional network (Neumann et al., 2010). Instead of functional or effective connectivity, a probabilistic dependency between brain regions is presented. However, all of the above methods suffer one or more of the following shortcomings. First, the functional co-activation patterns are by-products of approaches which primarily target for problems of consistent activations; thus lacking sounding statistical modeling. Second, the directed networks produced by some of the methods lead to difficulty in the interpretation. Third, no formal statistical tests can be performed.

To fill this gap, in this article, we propose a Poisson graphical model to estimate co-activation patterns, and thus identify the functional network in the meta-analysis of functional neuroimaging studies. We characterize region-level co-activation patterns using the covariance of the number of peak activations in different regions, where these region-specific activation counts are jointly modeled by a multivariate Poisson distribution (Kawamura 1979) (a review of the multivariate Poisson distribution is available in Appendix A). We impose sparsity on the covariance function by assuming that only some region pairs are co-activated for a particular brain function. We propose a penalized likelihood approach to efficiently estimate the sparse covariance function. Specifically, we introduce a set of latent variables to facilitate obtaining the penalized maximum likelihood estimates (PMLE) of the covariance via the expectation-maximization (EM) algorithm. The latent variables explicitly model the expected number of co-activation foci between regions. The non-directed functional network is then determined by the region-level estimated covariance between the numbers of foci. The proposed shrinkage method is tuned to reproduce the sparsity found in brain networks. The shrinkage tuning parameter is optimized based on the predictive log-likelihoods. We apply our approach to a meta-analysis of emotion

studies and conduct simulations studies to evaluate its performance.

The novel contributions of our models are several-fold. First, we are among the very first to propose a generative model for the sparse brain network analysis using functional neuroimaging meta-analysis count data. Second, our model provides more interpretable results by explicitly modeling the strength of functional co-activations compared with existing methods. Third, we propose a fast computational algorithm for model parameter estimation as well as a feasible permutation testing procedure to assess the significance of the identified brain network.

## 3.2 Data

We consider findings from 162 functional neuroimaging studies, including 57 positron emission tomography (PET) and 105 fMRI studies. Collectively, these studies yield 439 contrasts (e.g., happy vs. neutral) as in Kober et.al (2008). These studies are published in English from 1990 to 2005. We consider seven different emotions: sadness, happiness, anger, fear, disgust, surprise, and affective. These studies all meet the following criteria: (1) All the subjects included in the studies are healthy; (2) All the studies measure regional cerebral blood flow (PET) or blood oxygenation (fMRI) across the entire brain instead of regions of interest; (3) The activation coordinates are determined using image subtraction methodology; (4) Standard Talairach or Montreal Neurological Institute (MNI) coordinates are provided to ensure the results are spatially normalized to standard coordinate systems, thus allowing for comparison of findings across different studies. For our analysis, all data were converted into MNI space. For each study, the activation locations for these contrasts are included when they meet the criteria of significance defined each individual study. A total of 2478 activated coordinates are reported for 439 contrasts.

## 3.3 Methods

### 3.3.1 The Bivariate Model

We start with a model for any two regions $i$ and $j$ in the brain. For simplicity, we use region 1 and 2 as an example in this section. Suppose we collect foci data from $n$ contrasts. For $k = 1, \ldots, n$, let $X_{1,k}$ and $X_{2,k}$ represent the number of foci in region 1 and 2 respectively. We assume that $(X_{1,k}, X_{2,k})'$ follows a bivariate Poisson distribution with parameter $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{22}, \lambda_{12})$ (Kocherlakota and Kocherlakota, 1992). The joint probability function is

$$
\begin{aligned}
&P(X_{1,k} = x_{1,k}, X_{2,k} = x_{2,k}) \\
&= e^{-(\lambda_{11}+\lambda_{22}+\lambda_{12})} \frac{\lambda_{11}^{x_{1,k}}}{x_{1,k}!} \frac{\lambda_{22}^{x_{2,k}}}{x_{2,k}!} \sum_{s=0}^{\min(x_{1,k},x_{2,k})} \binom{x_{1,k}}{s}\binom{x_{2,k}}{s} s! \left(\frac{\lambda_{12}}{\lambda_{11}\lambda_{22}}\right)^s.
\end{aligned} \tag{3.1}
$$

This implies that $X_{1,k}$ and $X_{2,k}$ are marginally Poisson distributed with parameters $\lambda_{11} + \lambda_{12}$ and $\lambda_{22} + \lambda_{12}$, respectively. Also, the covariance between $X_{1,k}$ and $X_{2,k}$ is $\lambda_{12}$, which can be interpreted in our model as the strength of co-activation between regions 1 and 2. A covariance $\lambda_{12} = 0$ implies that there is no statistical dependence between the two regions.

We also impose sparsity on the brain network in our model. This implies that $\lambda_{12} = 0$ for many region pairs. It is natural to introduce a penalty term into the likelihood (3.1) to efficiently estimate the sparse co-activations between the two regions. In particular, we minimize the following penalized negative log-likelihood with

respect to $\boldsymbol{\lambda}$ given $\theta$:

$$- l_{\mathrm{obs}}(\boldsymbol{\lambda}; \mathbf{X}_1, \mathbf{X}_2) + \theta\lambda_{12}$$

$$= \sum_{k=1}^{n} \left\{ \sum_{i=1}^{2} [\lambda_{ii} - X_{i,k}\log(\lambda_{ii})] + \lambda_{12} - \log \left[ \sum_{s=0}^{\min(x_1,x_2)} \binom{X_{1,k}}{s}\binom{X_{2,k}}{s} s! \left( \frac{\lambda_{12}}{\lambda_{11}\lambda_{22}} \right)^s \right] \right\}$$

$$+ \theta\lambda_{12},$$

$$(3.2)$$

where the parameter $\theta$ controls the degree of sparsity. Larger values of $\theta$ will tend to shrink the covariance parameters toward zero, reflecting more sparsity in the brain network. The joint probability function of $(\mathbf{X}_1, \mathbf{X}_2)$ is complicated, especially when the number of dimensions is large. Kano and Kawamura (1991) derived a recursive scheme for constructing the probability function of a multivariate Poisson distribution; however, the computational intensity as well as the errors induced by recursion increase with the number of dimensions. Thus, a maximum likelihood estimation method without calculating the probability function is desired. Kalis (2003) proposed an EM algorithm (Dempster et al., 1977; Meng and Van Dyk, 1997; McLachlan and Krishnan, 1997) based on the multivariate reduction derivation of the multivariate Poisson distribution for estimation.

In our case, to simplify the computation, for each contrast $k$, we introduce three latent Poisson variables $Y_{11,k}, Y_{22,k}$ and $Y_{12,k}$ to represent $X_{1,k}$ and $X_{2,k}$. Specifically, we have

$$\begin{cases} X_{1,k} = Y_{11,k} + Y_{12,k}, \\ X_{2,k} = Y_{22,k} + Y_{12,k}, \end{cases} \qquad (3.3)$$

and

$$Y_{*,k} \overset{i.i.d}{\sim} \mathrm{Poisson}(\lambda_*), \ \text{for} \ * \in \{11, 22, 12\}, \qquad (3.4)$$

where $Y_{11,k}$ and $Y_{22,k}$ represent the number of localized foci in regions 1 and 2 respec-

tively. And $Y_{12,k}$ denotes the number of co-activations between the two regions. Note that the expectation of $Y_{12,k}$ is $\lambda_{12}$ which characterizes the covariance between $X_{1,k}$ and $X_{2,k}$.

Model representation (3.3) involves the unobserved data $Y_{12,k}$, $Y_{11,k}$ and $Y_{22,k}$ for $k = 1, \ldots, n$. Note that $Y_{11,k} = X_{1,k} - Y_{12,k}$ and $Y_{22,k} = X_{2,k} - Y_{12,k}$ based on (3.3). Since we observe $X_{1,k}$ and $X_{2,k}$, the only missing data is $Y_{12,k}$, and the complete data is $(Y_{12,k}, X_{1,k}, X_{2,k})$ for $k = 1, \ldots, n$. Write $\mathbf{Y}_{12} = (Y_{12,1}, \ldots, Y_{12,n})'$. The penalized complete negative log-likelihood is given by

$$
\begin{aligned}
& -l_{\text{comp}}(\boldsymbol{\lambda}; \mathbf{Y}_{12}, \mathbf{X}_1, \mathbf{X}_2) + \theta\lambda_{12} \\
& = \sum_{k=1}^{n} \left\{ \sum_{i=1}^{2} [\lambda_{ii} - (X_{i,k} - Y_{12,k})\log(\lambda_{ii})] + \lambda_{12} - Y_{12,k}\log(\lambda_{12}) \right\} + \theta\lambda_{12}.
\end{aligned}
\tag{3.5}
$$

We propose to use an EM-algorithm to minimize (5). The E-step calculates the conditional expectation of unobserved data $\mathbf{Y}_{12}$ given the observed data $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ using the current estimates of the parameters. And then we minimize the penalized complete negative log-likelihood in the M-step.

The EM algorithm is described as follow, with an initial value of $\boldsymbol{\lambda}^{(0)}$, for $t = 0, \cdots, T - 1$, in the $(t + 1)$th step,

- E-step, compute

$$
\begin{aligned}
Y_{12,k}^{(t+1)} &= E[Y_{12,k}|X_{1,k}, X_{2,k}; \boldsymbol{\lambda}^{(t)}] \\
&= \sum_{y_{12,k}=0}^{\min(x_{1,k}, x_{1,k})} \frac{y_{12,k} P(Y_{12,k}, X_{1,k}, X_{2,k}; \boldsymbol{\lambda}^{(t)})}{\sum_{y_{12,k}=0}^{\min(x_{1,k}, x_{2,k})} P(Y_{12,k}, X_{1,k}, X_{2,k}; \boldsymbol{\lambda}^{(t)})},
\end{aligned}
\tag{3.6}
$$

- M-step, update the estimates by

$$\lambda_{12}^{(t+1)} = \frac{\sum_{k=1}^{n} Y_{12,k}^{(t+1)}}{\theta + n},$$

$$\lambda_{ii}^{(t+1)} = \frac{1}{n} \sum_{k=1}^{n} X_{i,k} - \frac{\theta + n}{n} \lambda_{12}^{(t+1)}. \tag{3.7}$$

The iteration proceeds between the E-steps and M-steps and stops when specified convergence criteria are attained. The joint probability function $P(Y_{12,k}, X_{1,k}, X_{2,k})$ in (3.6) is derived in Appendix B.

The algorithm described above is a special case of the approach of Kalis (2003), which considered multivariate cases with the same covariance for all pairs of observations. Below, we consider a more general case that assumes different covariances.

## 3.3.2 The Multivariate Model

We consider a general extension to the bivariate model. In theory, a $n$-dimensional Poisson model might include $n$-way interactions. This leads to a very complicated model and might result in extremely high computational cost. Thus, in practice, a $m$-way interaction model is adequate for application, for a relatively smaller $m$. In our case, the two-way interaction, i.e., the covariance between $\mathbf{X}_i$ and $\mathbf{X}_j$ for $1 \leq i, j \leq p$, is sufficient to construct a brain network which is of primary interest in this paper. Therefore, we ignore interactions between three or more $\mathbf{X}_i$'s.

Suppose for $k = 1, \cdots, n$, we observe the number of foci in $p$ regions, $\mathbf{X}_k = (X_{1,k}, \cdots, X_{p,k})'$. We assume that $\mathbf{X}_k$ follows a multivariate Poisson distribution with parameters $\boldsymbol{\lambda} = (\lambda_{ij})_{1 \leq i,j \leq p}$. Similar to (3.3), we have

$$X_{i,k} = \sum_{j=1}^{p} Y_{ij,k}, \text{ for } i = 1, \cdots, p, \tag{3.8}$$

where $\mathbf{Y}_k = (Y_{ij,k})_{1 \leq i \leq j \leq p}$ is a collection of independent Poisson random variables,

and each $Y_{ij,k}$ follows a Poisson distribution with parameter $\lambda_{ij}$ respectively. Also, each $X_{i,k}$ follows a Poisson distribution with parameter $\sum_{j=1}^{p} \lambda_{ij}$. To simplify the notation, we let $Y_{ij,k} = Y_{ji,k}$ and $\lambda_{ij} = \lambda_{ji}$, for $1 \leq i < j \leq p$. The observed number of foci in region $i$, $X_{i,k}$ can be decomposed into $p$ parts: $Y_{ii,k}$, the number of localized activations in region $i$, and $\{Y_{ij,k}\}_{j \neq i}$, the number of co-activations in region $i$ and the remaining regions.

To incorporate sparsity in the covariance structure of the brain network, we utilize the following penalized observed log-likelihood:

$$-l_{\text{obs}}(\boldsymbol{\lambda}; \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n) + \theta \sum_{i=1}^{p} \sum_{j=i+1}^{p} \lambda_{ij}. \tag{3.9}$$

Note that it is computationally intensive to evaluate the observed data likelihood. Similar to the bivariate model, however, we can easily compute the complete data likelihood for $(\widetilde{\mathbf{Y}}_k, \mathbf{X}_k)$, where $\widetilde{\mathbf{Y}}_k = \{Y_{ij,k}\}_{1 \leq i < j \leq p}$ contains all the information of co-activations patterns. Specifically, we consider the following penalized complete data negative log-likelihood

$$
\begin{aligned}
& -l_{\text{comp}}(\boldsymbol{\lambda}; \widetilde{\mathbf{Y}}_1, \cdots, \widetilde{\mathbf{Y}}_n, \mathbf{X}_1, \cdots, \mathbf{X}_n) + \theta \sum_{i=1}^{p} \sum_{j=i+1}^{p} \lambda_{ij} \\
& = \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{j=i}^{p} [\lambda_{ij} - Y_{ij,k} \log(\lambda_{ij})] + \theta \sum_{i=1}^{p} \sum_{j=i+1}^{p} \lambda_{ij}.
\end{aligned} \tag{3.10}
$$

To minimize the observed negative log-likelihood (3.9) using the EM algorithm, we will need to calculate the conditional expectation of each $Y_{ij,k}$ given the observed data and current estimates of the unknown parameters. Note that the conditional probability of $Y_{ij,k}$ given the observed data only depends on $X_{i,k}$ and $X_{j,k}$, i.e., $E(Y_{ij,k}|\mathbf{X}_k; \boldsymbol{\lambda})) = E(Y_{ij,k}|X_{i,k}, X_{j,k}; \boldsymbol{\lambda})$.

Given the initial value $\boldsymbol{\lambda}^{(0)}$, in the $(t+1)$th step, for $t = 0, \cdots, T-1$,

- E-step, for $i = 1, \cdots, p$, $j = i+1, \cdots, p$, compute

$$Y_{ij,k}^{(t+1)} = E[Y_{ij,k}|X_{i,k}, X_{j,k}; \lambda^{(t)}] \tag{3.11}$$

$$= \sum_{y_{ij,k}=0}^{\min(x_{i,k}, x_{j,k})} \frac{y_{ij,k} P(Y_{ij,k}, X_{i,k}, X_{j,k}; \lambda^{(t)})}{\sum_{y_{ij,k}=0}^{\min(x_{i,k}, x_{j,k})} P(Y_{ij,k}, X_{i,k}, X_{j,k}; \lambda^{(t)})}$$

- M-step, update the estimates by

$$\lambda_{ij}^{(t+1)} = \frac{\sum_{k=1}^{n} Y_{ij,k}^{(t+1)}}{\theta + n}, \text{ for } 1 \le i < j \le p$$

$$\lambda_{ii}^{(t+1)} = \frac{1}{n}\sum_{k=1}^{n} X_{i,k} - \frac{\theta + n}{n} \sum_{j \neq i} \lambda_{ij}^{(t+1)} \tag{3.12}$$

The joint probability function $P(Y_{ij,k}, X_{i,k}, X_{j,k})$ in (3.11) is derived as shown in Appendix C.

### 3.3.3   Tuning Parameter

We consider the predictive log-likelihood as the criteria to determine the optimal value of the tuning parameter $\theta$. In our upcoming simulation studies, we also examine the mean-squared error as a supplementary tool to verify our findings.

Letting $\hat{\lambda}_{ij}(\theta)$ denote the estimate of $\lambda_{ij}$ derived from the EM algorithm, given $\theta$, the predictive log-likelihood given $\hat{\boldsymbol{\lambda}}(\theta) = \{\hat{\lambda}_{ij}(\theta)\}_{1 \le i,j \le p}$ is defined as:

$$l(\hat{\boldsymbol{\lambda}}(\theta); \mathbf{X}) = \sum_{k=1}^{n} l(\hat{\boldsymbol{\lambda}}(\theta); \mathbf{X}_k). \tag{3.13}$$

The probability function of $\mathbf{X}_k$ is given by

$$P(\mathbf{X}_k|\hat{\boldsymbol{\lambda}}(\theta)) = \sum_{\mathbf{y}_k} P(\mathbf{Y}_k, \mathbf{X}_k|\hat{\boldsymbol{\lambda}}(\theta)) = \sum_{\mathbf{y}_k} P(\mathbf{X}_k|\mathbf{Y}_k, \hat{\boldsymbol{\lambda}}(\theta)) P(\mathbf{Y}_k|\hat{\boldsymbol{\lambda}}(\theta))$$

$$= \sum_{\mathbf{y}_k} \frac{P(\mathbf{Y}_k|\hat{\boldsymbol{\lambda}}(\theta))}{P(\mathbf{Y}_k|\mathbf{X}_k, \hat{\boldsymbol{\lambda}}(\theta))} P(\mathbf{X}_k|\mathbf{Y}_k, \hat{\boldsymbol{\lambda}}(\theta)) P(\mathbf{Y}_k|\mathbf{X}_k, \hat{\boldsymbol{\lambda}}(\theta)) \tag{3.14}$$

where $P(\mathbf{X}_k|\mathbf{Y}_k, \hat{\boldsymbol{\lambda}}(\theta)) = 1$, if $\sum_{j=1}^p Y_{ij,k} = X_{i,k}$, for all $i \in \{1, \cdots, p\}$; otherwise $P(\mathbf{X}_k|\mathbf{Y}_k, \hat{\boldsymbol{\lambda}}(\theta)) = 0$. Therefore, the probability function can be expressed as

$$P(\mathbf{X}_k|\hat{\boldsymbol{\lambda}}(\theta)) = E_{\mathbf{Y}_k|\mathbf{X}_k} \left[ \frac{P(\mathbf{Y}_k|\hat{\boldsymbol{\lambda}}(\theta))}{P(\mathbf{Y}_k|\mathbf{X}_k, \hat{\boldsymbol{\lambda}}(\theta))} \right]. \tag{3.15}$$

Details of the conditional probability in (3.15) are presented in Appendix D.

We select the value of $\theta$ that yields the maximum predictive log-likelihood, where the predictive likelihood can be obtained as previously described. We use five-fold cross validation to choose the optimal estimation of $\theta$ in the simulation studies and ten-fold cross validation in the data application.

To verify the findings from the predictive log-likelihood, we also use the mean-squared error (MSE) in the simulation studies, defined as

$$\text{MSE}(\theta) = \sum_{i \leq j} (\lambda_{ij} - \hat{\lambda}_{ij}(\theta))^2. \tag{3.16}$$

We select the value of $\theta$ that gives the smallest MSE. In our data application, we only use the predictive log-likelihood measure for optimizing the tuning parameter, since the MSE depends on the true value of $\lambda$. In Section 4, we show that the MSE and the predictive log-likelihood strategies select very similar values of $\theta$.

We perform two independent grid searches when determining the optimal value of $\theta$, from a more coarse grid to a finer grid in a certain range for these two approaches. The value that we find achieves the maximum predictive log-likelihood and relatively small MSE within the searching range and at the level of precision which equals to the finer grid. It is possible that the optimal value is beyond the searching range or the level of precision is smaller than the finer grid. The simulation studies show that the selected $\theta$ yields the estimated covariance matrix with small MSE.

### 3.3.4 Statistical Testing

Our modeling framework enables us to perform statistical testing on the $\lambda_{ij}$'s to make inferences on the co-activation patterns between regions and the associated functional network. We conduct permutation tests to detect the existence of connections between regions and check whether the identified network is statistically significant. First, we say that two regions are not connected if the estimated $\lambda_{ij}$ obtained with the penalty term $\theta$ is below a fixed cutoff, here $10^{-3}$. This implies that on average there are no co-activating foci reported on the two regions across 1000 independent studies (in our application, we only have around 500 studies). Then we apply the algorithm again for more accurate estimates without the penalty term as our final estimates for the model.

We build the functional network based on the estimated non-zero $\lambda_{ij}$'s. Let $\Phi$ denote the set of region pairs within the identified network, we consider the following hypothesis for the network identification:

$$H_0 : \lambda_{ij} = 0, \forall \{ij\} \in \Phi \quad \text{vs.} \quad H_a : \exists \{ij\} \in \Phi \text{ s.t. } \lambda_{ij} > 0.$$

For each region $i$, we denote a permuted data of $\mathbf{X}_i$ by $\widetilde{\mathbf{X}}_i = (X_{i,\tau_1}, \ldots, X_{i,\tau_n})$, where $(\tau_1, \ldots, \tau_n)$ is a permutation of $(1, \ldots, n)$. We compute the estimated $\lambda_{ij}$ for each permuted dataset and then obtain a null distribution of $\hat{\lambda}_{ij}$ given that there is no connection between region $i$ and $j$.

Also, we perform a permutation test to examine $H_0 : \lambda_{ij} = 0$ vs. $H_a : \lambda_{ij} > 0$, correcting for multiple comparisons using the false discovery rate (FDR) approach by Benjamini and Hochberg (1995). Suppose we have $m$ tests $H_1, H_2, \cdots, H_m$ and the corresponding p-values $P_1, P_2, \cdots, P_m$ are ordered as $P_{(1)} \leq P_{(2)} \leq \cdots \leq P_m$. Let $h$ be the largest $i$ such that $P_{(i)} \leq \frac{i}{m}\alpha$, where $\alpha$ is the significance level, then all $H_{(i)}$ for $i = 1, 2, \cdots, h$ are rejected.

### 3.3.5 Graph Theoretical Properties of the Network

We perform graph theoretical analyses of co-activation patterns to demonstrate topological properties of the brain network. Each network is composed of $n$ nodes and $k$ edges, which represent $n$ brain regions and $k$ co-activation connections, respectively, in the brain imaging analyses. Several distance metrics have been developed to describe the relationships between nodes. For example, the clustering coefficient $C$ measures the average likelihood of connecting neighbors. For each node $i$, it is defined as $C_i = 2E_i/k_i(k_i - 1)$, where $k_i$ is the degree of node $i$, and $E_i$ is the number of direct links connecting neighbors of node $i$. The path length $L$ is the average minimum number of connections to link two nodes. Network topology is described as a *small-world network* (Watts and Strogatz 1998) if comparing to a *similar random network*, the small-world index $\sigma = (C/C_{\mathrm{random}})/(L/L_{\mathrm{random}}) > 1$ (Humphries et al. 2006). Here, a similar random network is defined as a network with same number of nodes, same number of edges and same degree distributions. We conduct statistical testing on the small-worldness property of identified functional network by performing the following hypothesis test using permutation approach: $H_0 : \sigma \leq 1$ vs. $H_a : \sigma > 1$.

Hubs in the network are of high importance since they serve as the common connections between other edges. The nodes with high degree (D) or high centrality (CEN) are defined as hubs. The centrality of node $i$ is the number of shortest paths between any nodes that pass through node $i$ (Sporns and Zwi 2004). We examine the high-degree or high-centrality nodes, i.e., the nodes with a degree or centrality at least one standard deviation above the network mean, in the network (Sporns et al. 2007).

### 3.3.6 Initial Values of EM Algorithm

In the EM algorithm, the choice of initial values is of great importance as it may have a substantial impact on convergence time. A number of methods have been proposed

for addressing the problem of choosing the initial values, e.g., a grid search for setting the initial values (Laird 1978) and forming clusters with supplementary information and using means as initial values (Leroux 1992). We use another natural choice of the starting points for our estimations, namely estimates obtained by the method of moments. Specifically, we let $\lambda_{ij}^{(0)} = \widehat{\text{cov}}(\mathbf{X_i}, \mathbf{X_j})$, i.e., the sample covariance. By setting reasonable initial values, we achieve relatively fast convergence of the algorithm.

## 3.4   Results

### 3.4.1   A Meta Analysis of Functional Neuroimaging Studies

We apply our proposed method to a meta analysis of emotion functional neuroimaging studies. Our data collects the findings of 439 contrasts from 162 emotion related functional neuroimaging studies. A total of seven different emotions are considered, including sad (45 contrasts), happy (36 contrasts), anger (26 contrasts), fear (68 contrasts), disgust (44 contrasts), surprise (2 contrasts) and "affective" (175 contrasts), a placeholder term for a blend of emotions elicited by complex stimuli such as aversive pictures. For each contrast, the coordinates of the activation locations are reported, and the reported points are assigned to different brain regions based on the coordinates. The number of the reported points in each region is our data input. On average, approximately 6 activated coordinates are reported for each contrast. In our analysis, we use the GlaxoSmithKline Clinical Imaging Centre (CIC) (Tziortzi et al. 2011) brain atlas based on Harvard-Oxford atlas (Makris et al. 2006) and consider 19 regions of interest (ROIs) related to emotion processing, yielding a $19 \times 439$ data matrix. The objectives of this study include estimation of the co-activation patterns and the corresponding functional network for emotion process, statistical testing for the connections between regions, and testing for the identified brain network.

Among 19 ROIs, the dorsolateral frontal cortex (DLFC) is the region that is

Table 3.1: The distribution of number of activations reported in DLFC and cuneus across all the contrasts included in the study.

| ROIs | Number of Activations | | | | | | |
|------|------|------|------|------|------|------|------|
|      | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| DLFC | 312 (71.07%) | 77 (17.54%) | 26 (5.92%) | 14 (3.19%) | 6 (1.37%) | 2 (0.46%) | 2 (0.46%) |
| GP_R | 436 (99.32%) | 3 (0.68%) | | | | | |

reported most often (217 from 439 contrasts), and the right globus pallidus (GP_R) is reported the least often (3 out of 439 contrasts). On average, each region is found to be associated with the neural processing of emotions 61 times out of 439 contrasts. Table 3.1 shows the distribution of number of activations reported in DLFC and GP_R across all the contrasts included in the study.

We perform cross validations to choose the value of $\theta$ that yields the largest predictive likelihoods, and we conduct estimation of the covariance parameters based on the selected $\theta$. Following the steps described in section 3.3.4, we identify the functional network and perform statistical testing on the network and marginal distributions of co-activations between regions. Among the 19 ROIs included in the analysis, the total number of possible region pairs is 171. After estimation, we detect a network including 17 ROIs with 79 connections.

We find strong functional co-activation patterns within the limbic system, basal ganglia and other most reported emotion related brain regions, as shown in Figure 3.1 and Figure 3.2. The anterior cingulate cortex (ACC) is thought to be involved in reward and other diverse affective/motivational processes. It is found to be connected with other 11 regions in the network. Especially, among the first 6 region pairs with highest covariance estimates, ACC appears 4 times, while the rest two region pairs are bilateral regions. For example, ACC is functionally connected with orbitofrontal cortex (OFC), which is one of the major centers for affective proceses ($\hat{\lambda}_{ij} = 0.023$, $p <$

Figure 3.1: The functional network identified from 162 functional neuroimaging studies with 439 contrasts. 17 ROIs are included in the network. The size of each node represents the degree of the node. Ins – Insular, ATP – Anterior Temporal Pole, para-HCMP – Parahippocampal, Amy – Amygdala, HCMP – Hippocampus, DLFC – DorsoLateral Frontal Cortex, MFC – Medial Frontal Cortex, frOP – Frontal Operculum, PCC – Posterior Cingulate Cortex.
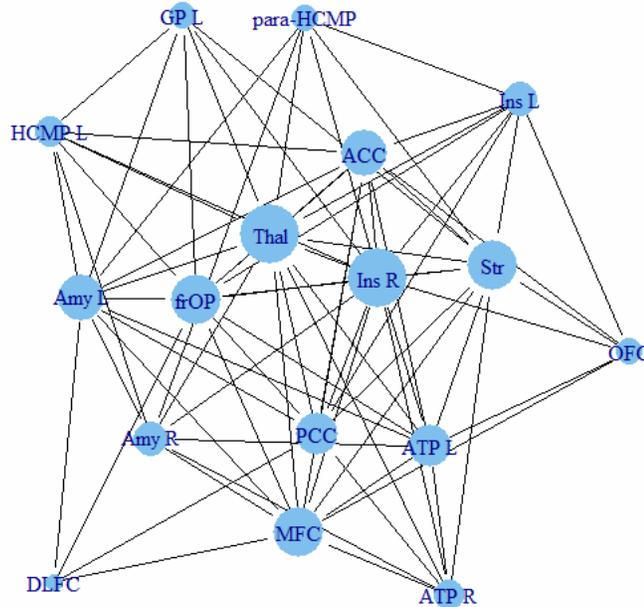


$0.005^*$ [1]). Strong co-activation is also identified between ACC and striatum (Str) ($\hat{\lambda}_{ij} = 0.018$, $p < 0.005^*$); ACC and thalamus (Thal) ($\hat{\lambda}_{ij} = 0.013$, $p < 0.005^*$), as well as ACC and frontal operculum (frOP) ($\hat{\lambda}_{ij} = 0.012$, $p < 0.005^*$). Overall, we build a functional network with brain regions that are involved in the neural processing of different emotions as shown in Figure 3.1 and 3.2 ($p < 0.005^*$). Also, almost all of the marginal connections between each region pair is significant after FDR correction($p < 0.005^*$).

We also examine graph theoretical properties of the identified network. The clustering coefficient of the identified network $C = 0.710$, and the path length $L = 1.129$. Comparing to the average of 1000 random networks ($C_{\text{random}} = 0.693$, $L_{\text{random}} =$

---

[1]$\lambda_{ij}$ with a p-value less than or equal to 0.01 is considered to be a significant connection, denoted by *.

Figure 3.2: Three different views of the functional network identified from 162 functional neuroimaging studies with 439 contrasts. 17 ROIs are included in the network.
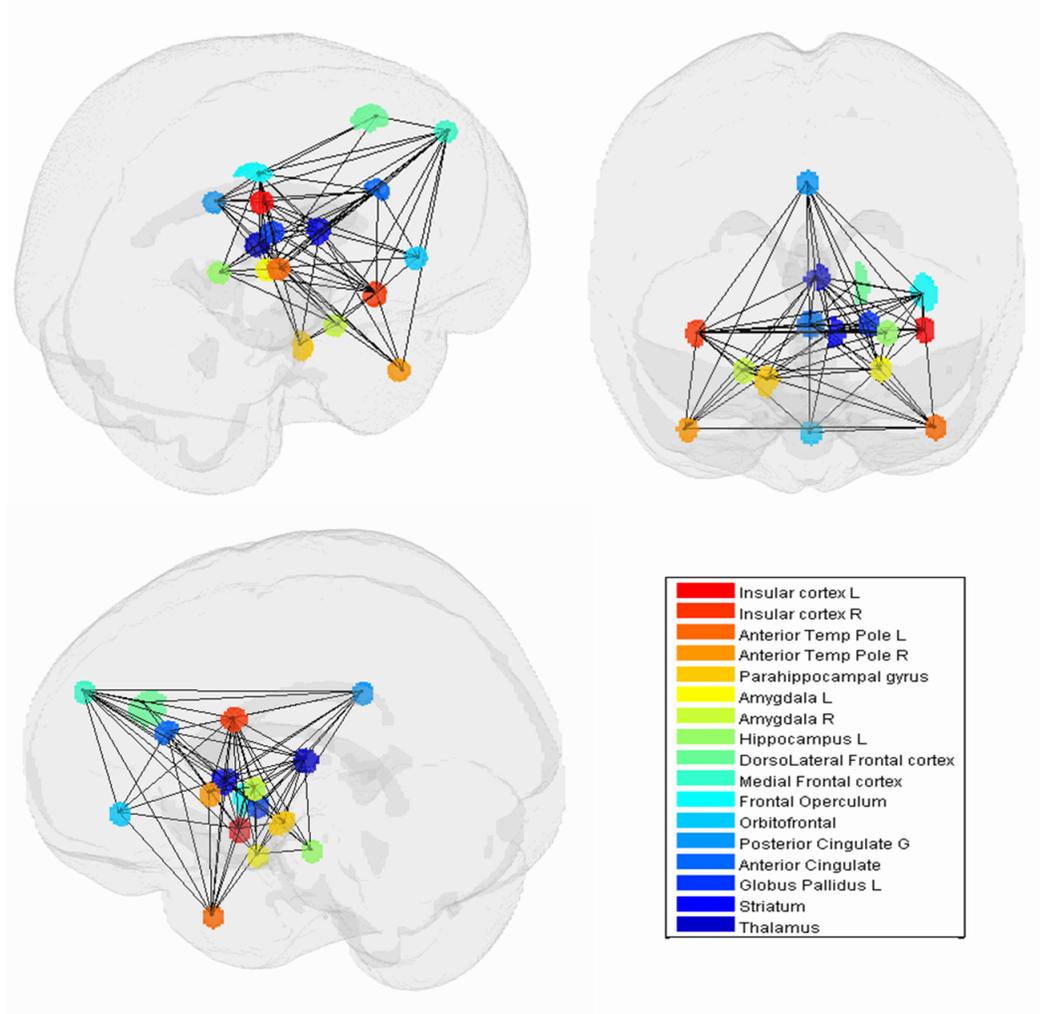


1.131), the corresponding small-world index $\hat{\sigma} = 1.027$ is significantly greater than 1 ($p < 0.005$), which indicates that the identified network has properties which are consistent with a small-world network. We use degree and centrality measures to identify the network hubs as described in the previous section. We find several regions which play important roles in the network, e.g., the right insular (Ins_R) (D=14, CEN=18.34), Thal (D=14, CEN=14.69), the left amygdala (Amy_L) (D=11, CEN=14.08) and the medial frontal cortex (MFC) (D=12, CEN=13.07). Ins, Thal and Amy are among the most reported emotion-related regions, and MFC is involved

Figure 3.3: The functional networks identified from 162 functional neuroimaging studies with 439 contrasts for negative emotions – anger, disgust, fear, and sad.
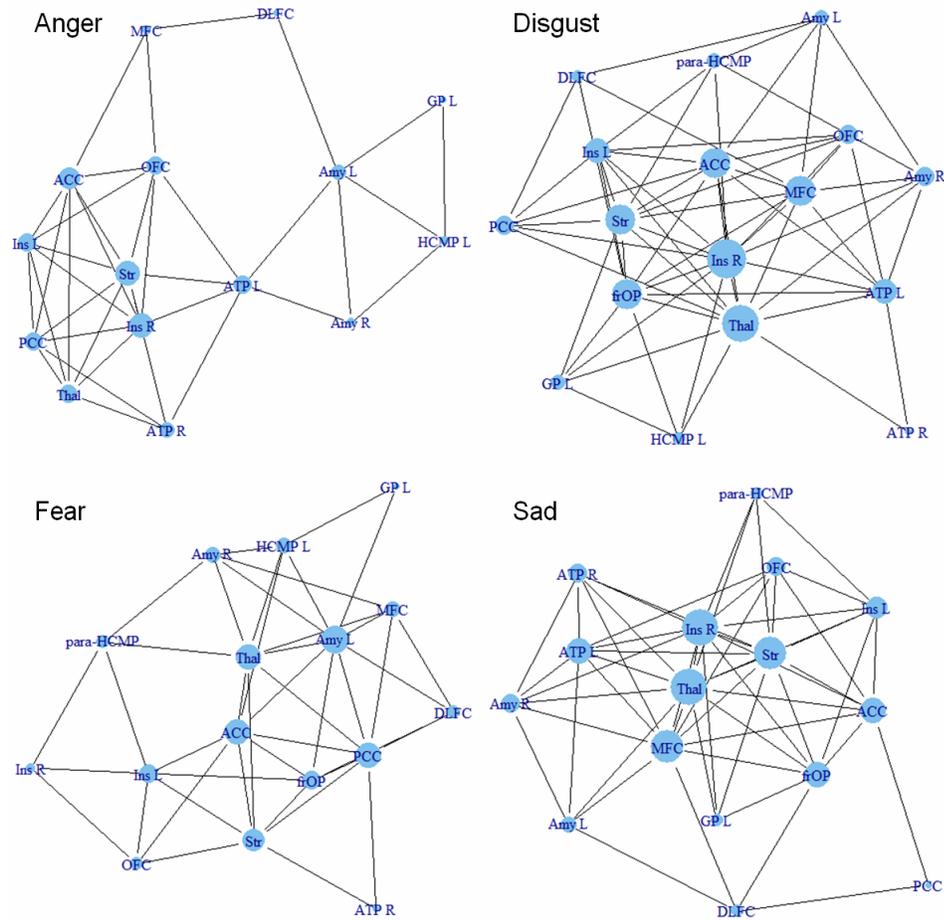


in cognitive control and related processes in a variety of settings, and may reflect some of the cognitive "ingredients" of the emotion generation process or, alternatively, play a more direct role in the generation of emotional feelings.

We further examine the subnetwork within the identified network separately for each emotion, especially for the negative emotions due to the restricted number of positive emotions and relatively small number of studies for positive emotions. For all the region pairs within the identified network, we count the number of times that both regions have at least one peak activation coordinates reported for different emotions, and they are considered as connected region pairs in the subnetwork. We expect to see more sparse networks for distinct emotions as compared to the network

Table 3.2: The region pairs with top frequencies of co-activations for anger, disgust, fear and sad.

| Emotion | Region | Region | Frequency (%) |
|---------|--------|--------|---------------|
| Anger   | ATP_L  | Amy_L  | 3 (11.54%)    |
|         | DLFC   | MFC    | 3 (11.54%)    |
|         | OFC    | ACC    | 3 (11.54%)    |
| Disgust | OFC    | ACC    | 8 (18.18%)    |
|         | Ins_R  | OFC    | 5 (11.36%)    |
| Fear    | Amy_L  | Amy_R  | 6 (8.82%)     |
|         | ACC    | Str    | 5 (8.62%)     |
| Sad     | DLFC   | MFC    | 10 (22.22%)   |
|         | MFC    | Thal   | 6 (13.33%)    |

containing all emotions (see Figure 3.3). The region pairs with top frequencies for anger, disgust, fear, and sad are reported in Table 3.2. Within each network, we focus on the region pair that co-activates most and find that some of the emotions share same co-activation patterns. For instance, the region pair DLFC and MFC, both of which are involved in cognitive control, is identified by anger and sad emotions. The region pair OFC and ACC is found to activate a lot in anger and disgust, which indicates these two emotions may stimulate similar brain activities. In addition, as a primary region in the processing of emotions, bilateral co-activation in amygdala is detected in fear. The subnetwork analysis show that although different types of emotions have their own contributions to the network identified from all emotions, similar emotions may share same co-activation patterns.

## 3.4.2   Simulation Studies

### 3.4.2.1   Simulated Data Sets

*Data set 1*

The first simulation setting includes three regions and six non-zero parameters. Specif-

ically, we let

$$\boldsymbol{\lambda} = \begin{pmatrix} 1 & 3 & 1 \\ & 2 & 5 \\ & & 3 \end{pmatrix}.$$

A total of 300 data sets are generated, and for each simulation, 100 samples are simulated from a bootstrap resampling strategy to evaluate the accuracy of the estimations.

*Data set 2*

The second simulation setting builds a network including eight regions. We assume the existence of co-activations for a total of eight region pairs. For example, region 1 and 2 ($\lambda_{12} = 3$), region 1 and 5 ($\lambda_{15} = 4$), region 1 and 6 ($\lambda_{16} = 2$), region 2 and 7 ($\lambda_{27} = 2$), region 3 and 6 ($\lambda_{36} = 3$), region 4 and 8 ($\lambda_{48} = 4$), region 5 and 7 ($\lambda_{57} = 5$), region 7 and 8 ($\lambda_{78} = 1$). We generate 500 data sets for this setting.

### 3.4.2.2 Accuracy of the Estimations

First, we consider the situation that $\theta = 0$. In a simulation study using simulated data set 1, we estimate the six non-zero parameters $\lambda_{ij}$'s by applying our penalized multivariate Poisson model, calculate the variance of each parameter by a bootstrap resampling method, and examine the coverage rate of the estimations. Table 3.3 shows the average bias, with the percentage change in parenthesis, and the coverage rate of the estimation from 300 simulations. The average bias is 0.008 with an average of 0.46% change over six parameters. Also the average coverage rate is 94.17%. The results indicate that our method can accurately estimate the parameters of interest in the model with coverage rate close to 95%.

We compare our proposed method to the method of moments approach in which $\lambda_{ij}$ is estimated by calculating the covariance between $\mathbf{X}_i$ and $\mathbf{X}_j$. The results in Table 1 indicate that our penalized multivariate Poisson model substantially improves

Table 3.3: Comparison of the bias, with the percentage changes in parenthesis and the coverage rates of $\hat{\boldsymbol{\lambda}}$ for networks with three regions from 300 simulations between proposed method and covariance approach.

| Penalized Multivariate Poisson Model | | | | | |
|---|---|---|---|---|---|
| Bias (%) | | | Coverage Rate | | |
| 0.0020 (0.20%) | 0.0019 (0.06%) | 0.0115 (1.15%) | 93.33% | 95.00% | 90.67% |
| | 0.0142 (0.71%) | 0.0024 (0.05%) | | 96.00% | 95.00% |
| | | 0.0169 (0.56%) | | | 95.00% |

| Covariance Method | | | | | |
|---|---|---|---|---|---|
| Bias (%) | | | Coverage Rate | | |
| 0.1657 (16.57%) | 0.0624 (2.08%) | 0.1381 (13.81%) | 92.00% | 92.33% | 91.67% |
| | 0.0576 (2.88%) | 0.1457 (2.91%) | | 96.33% | 91.67% |
| | | 0.2747 (9.16%) | | | 92.00% |

the performance by decreasing the biases (0.008 vs. 0.141) with percentage changes (0.46% vs. 7.9%) and increasing the coverage rates (94.17% vs. 92.67%).

### 3.4.2.3 Impact of $\theta$ on Networks

We conduct a simulation study to examine the impact of $\theta$ on the number of connections in the network. Here, we use the simulated data set 2. In this network, we have eight regions, resulting in 28 connections, eight of which are non-zero values as indicated in the simulation settings. Therefore, the number of zeros is 20.

We consider different values of the penalty term $\theta$ on a natural log scale from -1 to 6 which ranges $\theta$ from 0.37 to 403. We set $\lambda_{ij} = 0$, if the estimated value is below $10^{-3}$. Generally speaking, as $\theta$ increases, the number of zeros also increases as shown in Figure 3.4. When $\theta$ varies within a small range, the change on the number of

Figure 3.4: The change of average number of zeros detected from 100 simulations vs. $\ln(\theta)$. The blue line indicates the true number of zeros.
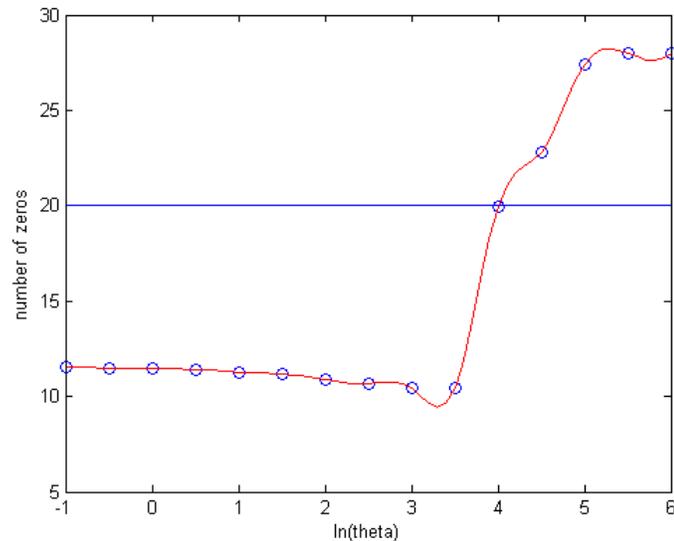


zero connections is small. As $\theta$ gets closer to 200, all the connections shrink to zeros, which shows the impact of the penalty term on the estimation of the network.

### 3.4.2.4   The Choice of $\theta$

Next, we perform simulation studies to choose the optimal tuning parameter $\theta$ based on the predictive log-likelihood from (3.13) and the mean-squared error using (3.16). We consider the same brain network as described in the second simulation setting. From coarse to finer parcellation of $\theta$, we find that $\theta = 3$ yields the largest average predictive log-likelihood -12.039 from five-fold cross validation. The smallest MSE, which equals to 0.5497, is achieved when $\theta = 2.8$. We notice that when $\theta$ varies from 2.0 to 3.6, the difference between the calculated MSE and smallest MSE is less than 0.01. In addition, the mean-squared error is the sum of the squared difference between all $\lambda_{ij}$'s and $\hat{\lambda}_{ij}$'s; as a result, it is the sum of all 28 squared-differences in our case. Therefore, 0.01 difference in MSE leads to an average of less than 0.02 difference between $\lambda_{ij}$ and $\hat{\lambda}_{ij}$. We do not restrict the range of $\theta$ during the grid search of the MSE approach to allow it to search for the best solution. Figure 3.5 shows the trend

Figure 3.5: The relationship between the predictive log-likelihood and $\theta$ (left); the predictive log-likelihood achieves the smallest value when $\theta = 3$. The relationship between the mean-squared error and $\theta$ (right); the MSE achieves smallest value when $\theta = 2.8$.
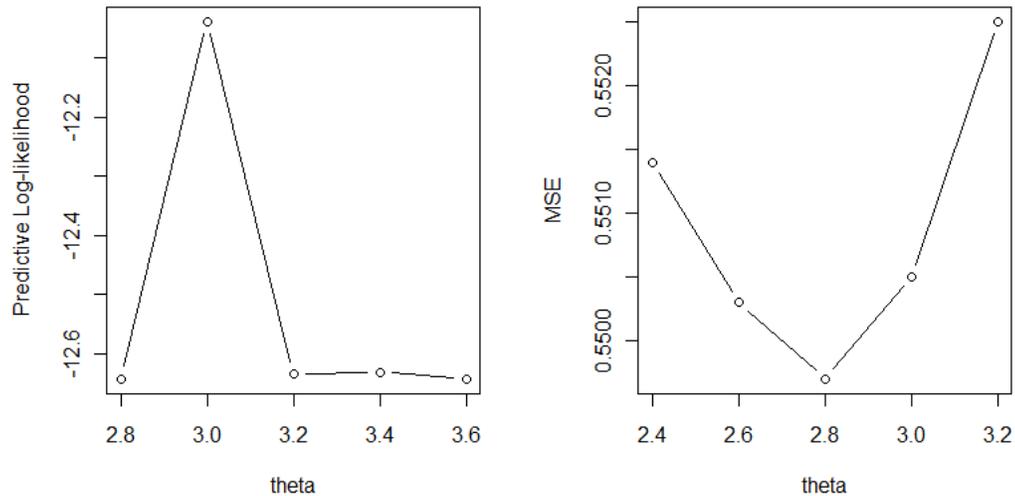


of the predictive log-likelihoods and the change of the mean-squared errors within a small range of $\theta$ identified by simulation studies. We can see that the optimal $\theta$ obtained from these two criterion are very close to each other.

## 3.5  Discussion

We propose a Poisson graphical model to identify functional co-activation patterns and produce undirected brain networks in the meta analyses of neuroimaging studies. Our method jointly models the region-level numbers of foci that are reported by different independently performed studies. The estimated sparse covariance matrix between regions is used to construct the undirected brain network associated with a particular brain function. We also perform a permutation test to assess the significance of the functional connectivity between regions. We extend the original multivariate Poisson model by including a penalty term to account for the sparsity of

the brain network and perform estimation using the EM algorithm. The simulation studies show that our method achieves about 95% coverage rate. Also, we select the tuning parameter by optimizing the predictive log-likelihood and the MSE. The results show that the shrinkage method produces more accurate estimates of the co-variance and reduces the computation time. We show that the two criteria choose a similar optimal value in simulation studies.

Meta-analysis techniques for brain imaging studies increase the accuracy and power compared to single analyses. Using the meta-analysis data, our method provides a systematic framework to estimate the co-activation patterns, which can be used to test for specific relationships among brain regions of interest or to establish groups of contiguous voxels that show similar functional characteristics and may be treated as prior information in the future studies. We address this problem by extending a well-developed model, which however, has never been applied in the field of brain imaging studies. Our penalized multivariate Poisson model can also be applied to other public health or statistical problems.

One limitation of our model is that it is based on the pre-defined parcellation of the brain. However, this is a common problem in other meta-analytic approaches for functional networks, and more broadly for a range of region-based functional connectivity approaches. Possible remedies to this problem include starting the analysis from voxel level and then reducing the dimension of the data by the singular value decomposition (SVD). Another possible extension to our model is to account for the over-dispersion problem for the Poisson distribution by setting

$$X_{i,k} = a_i \sum_{j=1}^{p} Y_{ij,k}, \text{ for } i = 1, \cdots, p, \tag{3.17}$$

where $a_i$ is a positive integer.

In our meta-analysis of 162 functional neuroimaging emotion studies, we only

use foci reported from different studies as the dataset. However, those studies may have different imaging modalities, sample sizes, criteria and thresholds for testing of statistical significance. To obtain a set of standardized count data, we can consider a Poisson covariance regression model estimating the co-activations adjusted for the significance levels and other covariates, which will increase the flexibilities of our proposed method in application.

# Chapter 4

# A Bayesian Spatial Model to Predict Disease Status Using Imaging Data from Various Modalities

## 4.1 Introduction

Functional and structural neuroimaging play important roles in understanding the neurological basis for major psychiatric disorders or mental illness such as schizophrenia, depression, Alzheimer's diseases, and Parkinson's disease (PD). Several methods have been proposed to make the prediction of follow-up imaging scans based on the baseline scans (Guo et al., 2008; Derado et al., 2012). However, the classification or prediction of the disease status based on the imaging data has not been fully investigated. Though the diagnostic criterion may have been established for the above diseases, a prediction model can help to reveal the underlying neural basis of the diseases, thus informing the development of future treatments; in addition, it can be

used for diagnoses of other diseases which have not been fully studied yet.

To predict the follow-up brain activity based on the baseline functional neuroimaging data, Guo et al. (2008) propose a Bayesian hierarchical model for functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) data; Bowman et al. (2008) develop a model which considers inter-regional and intraregional correlations for analyzing functional neuroimaging data; and Derado et al. (2012) extends the model by introducing both spatial correlations between voxels and temporal correlations between baseline and follow-up functional imaging scans. For structural data, Stonnington et al. (2010) propose a relevance vector regression (RVR) model to predict the clinical scores using MRI T1 weighted scans.

The most commonly used approaches to predict a single outcome from high-dimensional data are LASSO (Tibshirant, 1996) and elastic-net (Zou and Hastie, 2005), which perform regularization and variable selection for regression models. Support vector machine (SVM) classifies the data by constructing an optimal separating hyperplane in a high dimensional space in which the data are mapped to (Cortes and Vapnik, 1995). As an alternative to SVM for prediction of high-dimensional data, Gaussian process (GP) finds the posterior function distribution which is closest to the training data based on Bayesian theory (Marquand et al., 2010). Ham and Kwak (2012) propose a boosted-principal component analysis (PCA) algorithm for binary classification problems, which combines the procedure of feature selection and classification. However, these methods do not consider the existence of the spatial correlations in imaging scans or the associations between different imaging modalities, and build the statistical model directly for the prediction purpose rather than starting from the imaging data.

We propose a novel Bayesian hierarchical model to predict the disease status using imaging scans of different modalities in both grey and white matter to reflect the functional as well as the structural properties of the brain. We consider a two-level

brain parcellation and assume different spatial correlation structures between voxels within a subregion, within a region, and in different regions. We perform Markov Chain Monte Carlo (MCMC) estimations via Gibbs sampling. The predictions for disease status are conducted based on the predictive posterior probabilities. Both whole-brain and voxel-level prediction are performed using leave-one-out cross validation (LOOCV). Also, feature selections are conducted to identify the regions that are associated with the disease based on the voxel-level prediction results. We apply our approach to a PD study and conduct simulation studies to evaluate its performance.

## 4.2   Parkinson's Disease Data

A total of 20 subjects, 11 of which are diagnosed as PD patients, and the rest are healthy controls, are included in the study. The average ($\pm$ standard deviation) age is 66 ($\pm$ 11) years old, and 12 of them are males. Resting-state fMRI scans, T1-weighted MRI scans, and diffusion tensor imaging (DTI) scans are obtained.

We extract voxel-level information from these three types of imaging scans, including fractional amplitude of low-frequency fluctuation (fALFF) from resting-state fMRI scans, voxel base morphometry (VBM) from T1-weighted MRI scans, and fractional anisotropy (FA) from DTI scans. fALFF reflects the amplitude of spontaneous blood-oxygen-level-dependent (BOLD) signal fluctuations of each voxel. VBM measures the localized intensity value of each voxel after spatially normalizing all the images to a standard space, and extracting white or grey matter from the normalized images (Ashburner and Friston, 1999). FA has a single value for each voxel which measures the difference in directions along different axes of the random motion of water molecules in the brain. In summary, fALFF provides the functional information, while FA and VBM describe the structural properties of the brain.

The imaging preprocessing are performed by statistical parametric mapping (SPM) (Wellcome Department of Cognivite Neurology, http://www.fil.ion.ucl.ac.uk/spm) and FMRIB (Functional Magnetic Resonance Imaging of the Brain) Software Library (FSL) (Smith et al., 2004).

## 4.3 Methods

We propose a novel Bayesian hierarchical model to predict the disease status using imaging data of different modalities, including fALFF, VBM, and FA. For resting-state fMRI scans and DTI scans, the functional and structural information lie in grey matter and white matter, respectively. Most VBM analysis focus on grey matter; however, VBM in white matter has also been found to be associated psychiatric diseases such as Alzheimer's diseases and schizophrenia (Li et al., 2011; Di et al., 2009). Thus, our prediction model involves the voxels in grey or/and white matter for different imaging modalities.

### 4.3.1 Model and Estimation

We consider a two-level brain parcellation consisting of $G$ brain regions defined by the automated anatomic labeling (AAL) system (Tzourio-Mazoyer et al., 2002). Each region $g$ consists of $L_g$ subregions, which are built based on the brain parcellation algorithms described in Appendix E. Each subregion $l$ is composed of $V_l$ voxels. Let $X_{ilg}(v)$, $Y_{ilg}(v)$ and $Z_{ilg}(v)$ respectively denote the observed fALFF, FA and VBM for subject $i$ at voxel $v$ in subregion $l$ region $g$, for $i = 1, \ldots, n$, $v = 1, \ldots, V_l$, $l = 1, \ldots, L_g$, $g = 1, \ldots, G$. Let $N_g(l) \subseteq \{1, \ldots, L_g\}$ denote the neighbors of subregion $l$ in region $g$, and $n_{lg}$ be the number of members in $N_g(l)$. In our model, all the subregions in region $g$ are considered as neighbors of subregion $l$; therefore, we have $N_g(l) = \{1, \ldots, L_g\}$, and $n_{lg} = L_g$. Let $D_i \in \{0, 1\}$ denote the disease status, where 0 means non-disease;

and $\mathbf{W}_i = (W_{i1}, \cdots, W_{iQ})$ denote the covariates, where $Q$ is the number of covariates in the model. Let $\mathcal{B}$, $\mathcal{W}$ and $\mathcal{G}$ respectively represent the whole brain region, the white matter region and the gray matter region.

We propose a model that accounts for the spatial correlations between voxels within a same subregion, between subregions within a same region, and between regions. First, we assume consistent correlations between voxels in a same subregion. Then the spatial correlations between subregions within a same AAL region is described by a conditional autoregressive (CAR) model, which allows the estimations at subregion level borrow strength from their neighbors within the same AAL region. In addition, we introduce unstructured spatial correlations between AAL regions. The proposed model has the following hierarchical structure:

For any $v \in \mathcal{G}$,

$$[X_{ilg}(v) \mid Z_{ilg}(v), D_i, \mathbf{W}_i, \bullet\,]$$

$$\sim N\left\{\sum_{k=0,1}[c_{klg}^{xz}(v)(Z_{ilg}(v) - \bar{Z}_{lg}(v)) + \mathbf{W}_i\boldsymbol{\gamma}_{klg}^x(v) + \beta_{klg}^x(v) + \alpha_{ilg}^x + \eta_{kg}^x]I(D_i = k), \delta_{lg}^{xz}\right\},$$

for any $v \in \mathcal{W}$,

$$[Y_{ilg}(v) \mid Z_{ilg}(v), D_i, \mathbf{W}_i, \bullet\,]$$

$$\sim N\left\{\sum_{k=0,1}[c_{klg}^{yz}(v)(Z_{ilg}(v) - \bar{Z}_{lg}(v)) + \mathbf{W}_i\boldsymbol{\gamma}_{klg}^y(v) + \beta_{klg}^y(v) + \alpha_{ilg}^y + \eta_{kg}^y]I(D_i = k), \delta_{lg}^{yz}\right\},$$

for any $v \in \mathcal{B}$,

$$[Z_{ilg}(v) \mid D_i, \mathbf{W}_i, \bullet\,] \sim N\left\{\sum_{k=0,1}(\mathbf{W}_i\boldsymbol{\gamma}_{klg}^z(v) + \beta_{klg}^z(v) + \alpha_{ilg}^z + \eta_{kg}^z)I(D_i = k), \delta_{lg}^z\right\},$$

$$c_{klg}^{xz}(v) \sim N(\zeta_{klg}^{xz}, \omega_{klg}^{xz}), \qquad \zeta_{klg}^{xz} \sim N(a_\zeta, b_\zeta), \qquad \omega_{klg}^{xz} \sim \text{InvG}(a_\omega, b_\omega),$$

$$c_{klg}^{yz}(v) \sim N(\zeta_{klg}^{yz}, \omega_{klg}^{yz}), \qquad \zeta_{klg}^{yz} \sim N(a_\zeta, b_\zeta), \qquad \omega_{klg}^{yz} \sim \text{InvG}(a_\omega, b_\omega),$$

$$\gamma_{klgq}^{x}(v) \sim \mathrm{N}(0, s_{klg}^{x}), \qquad\qquad s_{klg}^{x} \sim \mathrm{InvG}(a_s, b_s),$$

$$\gamma_{klgq}^{y}(v) \sim \mathrm{N}(0, s_{klg}^{y}), \qquad\qquad s_{klg}^{y} \sim \mathrm{InvG}(a_s, b_s),$$

$$\gamma_{klgq}^{z}(v) \sim \mathrm{N}(0, s_{klg}^{z}), \qquad\qquad s_{klg}^{z} \sim \mathrm{InvG}(a_s, b_s),$$

$$\beta_{klg}^{x}(v) \sim \mathrm{N}\{\beta_{klg}^{x}, \lambda_{klg}^{x}\}, \qquad\qquad \lambda_{klg}^{x} \sim \mathrm{InvG}(a_\lambda, b_\lambda),$$

$$\beta_{klg}^{y}(v) \sim \mathrm{N}\{\beta_{klg}^{y}, \lambda_{klg}^{y}\}, \qquad\qquad \lambda_{klg}^{y} \sim \mathrm{InvG}(a_\lambda, b_\lambda),$$

$$\beta_{klg}^{z}(v) \sim \mathrm{N}\{\beta_{klg}^{z}, \lambda_{klg}^{z}\}, \qquad\qquad \lambda_{klg}^{z} \sim \mathrm{InvG}(a_\lambda, b_\lambda),$$

$$[\beta_{klg}^{x} \mid \{\beta_{kl'g}^{x}\}_{l'\neq l}, \bullet\,] \sim \mathrm{N}\left\{\frac{\rho_g^x}{L_g}\sum_{l'\in N_g(l)}\beta_{kl'g}^{x}, \frac{\phi_g^x}{L_g}\right\},$$

$$[\beta_{klg}^{y} \mid \{\beta_{kl'g}^{y}\}_{l'\neq l}, \bullet\,] \sim \mathrm{N}\left\{\frac{\rho_g^y}{L_g}\sum_{l'\in N_g(l)}\beta_{kl'g}^{y}, \frac{\phi_g^y}{L_g}\right\},$$

$$[\beta_{klg}^{z} \mid \{\beta_{kl'g}^{z}\}_{l'\neq l}, \bullet\,] \sim \mathrm{N}\left\{\frac{\rho_g^z}{L_g}\sum_{l'\in N_g(l)}\beta_{kl'g}^{z}, \frac{\phi_g^z}{L_g}\right\},$$

$$\rho_g^x \sim \mathrm{U}(\{0, 0.05, 0.1\cdots, 0.8, 0.81, \cdots, 0.9, 0.91, \cdots, 0.99\}), \quad \phi_g^x \sim \mathrm{InvG}(a_\phi, b_\phi),$$

$$\rho_g^y \sim \mathrm{U}(\{0, 0.05, 0.1\cdots, 0.8, 0.81, \cdots, 0.9, 0.91, \cdots, 0.99\}), \quad \phi_g^y \sim \mathrm{InvG}(a_\phi, b_\phi),$$

$$\rho_g^z \sim \mathrm{U}(\{0, 0.05, 0.1\cdots, 0.8, 0.81, \cdots, 0.9, 0.91, \cdots, 0.99\}), \quad \phi_g^z \sim \mathrm{InvG}(a_\phi, b_\phi),$$

$$\alpha_{ilg}^{x} \sim \mathrm{N}(0, \tau_{lg}^{x}), \qquad\qquad \tau_{lg}^{x} \sim \mathrm{InvG}(a_\tau, b_\tau),$$

$$\alpha_{ilg}^{y} \sim \mathrm{N}(0, \tau_{lg}^{y}), \qquad\qquad \tau_{lg}^{y} \sim \mathrm{InvG}(a_\tau, b_\tau),$$

$$\alpha_{ilg}^{z} \sim \mathrm{N}(0, \tau_{lg}^{z}), \qquad\qquad \tau_{lg}^{z} \sim \mathrm{InvG}(a_\tau, b_\tau),$$

$$\boldsymbol{\eta}_k^x = (\eta_{k1}^x, \ldots, \eta_{kG}^x)' \sim \mathrm{N}(0, \boldsymbol{\Sigma}_k^x), \qquad\qquad \boldsymbol{\Sigma}_k^x \sim \mathrm{InvW}(\boldsymbol{\Lambda}, \nu),$$

$$\boldsymbol{\eta}_k^y = (\eta_{k1}^y, \ldots, \eta_{kG}^y)' \sim \mathrm{N}(0, \boldsymbol{\Sigma}_k^y), \qquad\qquad \boldsymbol{\Sigma}_k^y \sim \mathrm{InvW}(\boldsymbol{\Lambda}, \nu),$$

$$\boldsymbol{\eta}_k^z = (\eta_{k1}^z, \ldots, \eta_{kG}^z)' \sim \mathrm{N}(0, \boldsymbol{\Sigma}_k^z), \qquad\qquad \boldsymbol{\Sigma}_k^z \sim \mathrm{InvW}(\boldsymbol{\Lambda}, \nu),$$

$$\delta_{lg}^{xz} \sim \mathrm{InvG}(a_\delta, b_\delta),$$

$$\delta_{lg}^{yz} \sim \mathrm{InvG}(a_\delta, b_\delta),$$

$$\delta_{lg}^{z} \sim \mathrm{InvG}(a_\delta, b_\delta).$$

We assume that the probability of disease status $P(D_i = k_i)$ is a constant, and independent of all the parameters. Also, we assume conditional independency among voxels and different modalities of data. Our model reflects the assumption that for each voxel $v$ in the grey matter, the fALFF value $X_{ilg}(v)$ follows a normal distribution conditioning on the VBM value $Z_{ilg}(v)$; for each voxel $v$ in the white matter, the FA value $Y_{ilg}(v)$ follows a normal distribution conditioning on the VBM value $Z_{ilg}(v)$; and for each voxel $v$ included in the analysis, the VBM value $Z_{ilg}(v)$ follows a normal distribution. The mean structure of the model is composed of several parameters given different diseases status, which is denoted by subscript $k$: $c_{klg}(v)$ is the slope term for centered VBM values; $\boldsymbol{\gamma}_{klg}(v) = (\gamma_{klg1}(v), \cdots, \gamma_{klgQ}(v))'$ is the covariates' parameters; $\beta_{klg}(v)$, $\alpha_{ilg}$, and $\eta_{kg}$ are the voxel-level intercept term, subregion level random effect, and region level intercept term, respectively. Each imaging modality is assumed to have a same subregion-level variance $\delta_{lg}$ for both disease status.

The prior belief of the parameters included in the likelihood function is expressed in the second or lower level of the model. The slope term $c_{klg}(v)$ follows a normal distribution, whose mean and variance are from noninformative hyperpriors. Each covariate parameter $\gamma_{klgq}(v)$ is assumed to arise from a normal mean-zero distribution with variance $s_{klg}$, which has a noninformative hyperprior distribution. $\beta_{klg}(v)$ within a same subregion is assumed to follow a normal distribution with a same mean $\beta_{klg}$, whose distribution will be discussed later, and variance $\lambda_{klg}$, whose hyperprior is noninformative. $\boldsymbol{\eta}_k$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_k$ which models the spatial dependence between AAL regions by assuming a unstructured covariance matrix. Spatial associations between voxels within each subregion are introduced by individual random effect term $\alpha_{ilg}$, which follows a mean-zero normal distribution with variance $\tau_{lg}$, thus assuming same spatial correlations between voxels in the same subregion.

By assuming a subregion level CAR model, we capture the spatial dependence

between subregions within each AAL region. In the model, $\rho_g$ represents the overall degree of spatial dependence in region $g$ and $\frac{\phi_g}{L_g}$ is the conditional variance of $\beta_{klg}$. The neighborhood of subregion $l$, which is in region $g$, is defined as all the other subregions in region $g$ in our analysis. The spatial neighborhood effect $\rho_g$ is assumed to follow a discrete uniform distribution (Gelfand and Vounatsou, 2003). As we would like to identify the similarity of the neighboring subregions, we impose $0 \leq \rho_g < 1$. Specifically, equal mass is put on the following 36 values: 0, 0.05, 0.1, ..., 0.8, 0.81, 0.82, ..., 0.90, 0.91, 0.92, ..., 0.99, which favors the upper range of $\rho_g$ The other hyperpriors are specified accordingly for our model.

It is easy to show that for any diseases status $k$, the covariance between the voxels within a same subregion $l$ in region $g$ is contributed by the variance from three components: $\beta_{klg}$, $\alpha_{ilg}$, and $\eta_{kg}$; the covariance between the voxels in two subregions $l$ and $l'$, but the same AAL region $g$, comes from the covariance between $\beta_{klg}$ and $\beta_{kl'g}$, and the variance of $\eta_{kg}$; and the covariance between the voxels in two AAL regions $g$ and $g'$ is determined by the covariance of $\eta_{kg}$ and $\eta_{kg'}$.

We perform estimation using Markov chain Monte Carlo (MCMC) implemented via Gibbs sampling. The full conditional posterior distributions are derived as shown in Appendix F.

## 4.3.2 Prediction

### 4.3.2.1 Whole Brain Prediction

The objective of our model is to predict the disease status of a subject given imaging data and other covariates. To achieve this goal, we use the posterior samples from the estimation step to calculate the posterior predictive probability.

Let $\boldsymbol{\theta}$ denote the parameter space, $\mathbf{B}_i = (\mathbf{X}_{ilg}, \mathbf{Y}_{ilg}, \mathbf{Z}_{ilg})$ denote the observed imaging data for subject $i$, $\mathbf{A}_i = (\mathbf{B}_i, D_i)$ denote the combination of the imaging data and the disease status. Suppose we have $n$ training subjects, and want to predict the

disease status of the $n + 1^{th}$ subject $D_{n+1}$. The posterior predictive distribution for $D_{n+1}$ is given by

$$
\begin{aligned}
&P(D_{n+1} = k \mid \mathbf{B}_{n+1}, \{\mathbf{A}_i\}_{i=1}^n) \\
&= \frac{P(D_{n+1} = k, \mathbf{B}_{n+1} \mid \{\mathbf{A}_i\}_{i=1}^n)}{\sum_{k'=0,1} P(D_{n+1} = k', \mathbf{B}_{n+1} \mid \{\mathbf{A}_i\}_{i=1}^n)} \\
&= \frac{P(D_{n+1} = k) \int_{\boldsymbol{\theta}} P(\mathbf{B}_{n+1} \mid D_{n+1} = k, \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \{\mathbf{A}_i\}_{i=1}^n) d\boldsymbol{\theta}}{\sum_{k'=0,1} P(D_{n+1} = k') \int_{\boldsymbol{\theta}} P(\mathbf{B}_{n+1} \mid D_{n+1} = k', \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \{\mathbf{A}_i\}_{i=1}^n) d\boldsymbol{\theta}},
\end{aligned} \tag{4.1}
$$

where

$$
\begin{aligned}
P(\mathbf{B}_{n+1} \mid D_{n+1} = k, \boldsymbol{\theta}) = &\prod_{v \in \mathcal{G}} P(X_{n+1}(v) \mid Z_{n+1}(v), D_{n+1} = k, \boldsymbol{\theta}) P(Z_{n+1}(v) \mid D_{n+1} = k, \boldsymbol{\theta}) \\
&\prod_{v \in \mathcal{W}} P(Y_{n+1}(v) \mid Z_{n+1}(v), D_{n+1} = k, \boldsymbol{\theta}) P(Z_{n+1}(v) \mid D_{n+1} = k, \boldsymbol{\theta}),
\end{aligned} \tag{4.2}
$$

Suppose we draw a total number of $T$ posterior samples $\boldsymbol{\theta}^{(t)}$ from $P(\boldsymbol{\theta} \mid \{\mathbf{A}_i\}_{i=1}^n)$, for $t = 1, \cdots, T$. Let $\pi_k^{(t)} = P(\mathbf{B}_{n+1} \mid D_{n+1} = k, \boldsymbol{\theta}^{(t)})$, the posterior predictive probability can be expressed by

$$
\hat{P}(D_{n+1} = k \mid \mathbf{B}_{n+1}, \{\mathbf{A}_i\}_{i=1}^n) = \frac{P(D_{n+1} = k) \sum_{t=1}^T \pi_k^{(t)}}{\sum_{k'=0,1} P(D_{n+1} = k') \sum_{t=1}^T \pi_k^{(t)}}. \tag{4.3}
$$

Then the prediction of $D_{n+1}$ is given by

$$
\hat{D}_{n+1} = \arg\max_k \left( P(D_{n+1} = k) \sum_{t=1}^T \pi_k^{(t)} \right). \tag{4.4}
$$

To evaluate the performance of our method, we calculate the accuracy rate of prediction using LOOCV due to the small sample size. Specifically, we use data $\mathbf{A}_{-i} = \{\mathbf{A}_u\}_{u \neq i}$ to make a prediction of $D_i$, which is denoted by $\hat{D}_i$. The accuracy rate is defined as the proportion of correct predictions. To obtain $\hat{D}_i$, we compute

the LOOCV predictive probabilities for $i = 1, \cdots, n$:

$$P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}) = \int_{\boldsymbol{\theta}} P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}, \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \mathbf{A}_{-i}) d\boldsymbol{\theta}, \qquad (4.5)$$

which can be estimated via MCMC simulations as in (4.3).

However, the LOOCV is very computational expensive because it involves multiple posterior simulations with tens of thousands voxels included in the analysis. Therefore, we employ an importance sampling approach to reduce the computation of LOOCV of our model (Gelfand et al., 1992; Gelfand, 1996; Alqallaf and Gustafson, 2001; Vehtari and Lampinen, 2002). Specifically, the LOOCV predictive probabilities can be expressed by

$$P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}) = \frac{P(D_i = k) Q_{kd_i}}{\sum_{k'=0,1} P(D_i = k') Q_{k'd_i}}, \qquad (4.6)$$

where

$$Q_{kd_i} = \int \frac{P(\mathbf{B}_i \mid D_i = k, \boldsymbol{\theta})}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta})} P(\boldsymbol{\theta} \mid \{\mathbf{A}_i\}_{i=1}^n) d\boldsymbol{\theta}, \qquad (4.7)$$

and $d_i$ is the observed disease status for subject $i$. Next, we provide the details of how $Q_{kd_i}$ is derived. The posterior predictive probability can be written as follows:

$$
\begin{aligned}
&P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}) \\
&= \int P(D_i = k \mid \mathbf{B}_i, \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \mathbf{B}_i, \mathbf{A}_{-i}) d\boldsymbol{\theta} \\
&= \int P(D_i = k \mid \mathbf{B}_i, \boldsymbol{\theta}) \frac{P(\boldsymbol{\theta} \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i})} P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i}) d\boldsymbol{\theta}.
\end{aligned}
\qquad (4.8)
$$

Note that

$$\frac{P(\boldsymbol{\theta} \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i})} = \frac{P(\boldsymbol{\theta}, \mathbf{B}_i, \mathbf{A}_{-i})P(\mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i})}{P(\mathbf{B}_i, \mathbf{A}_{-i})P(\boldsymbol{\theta}, \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i})} = \frac{P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(D_i = d_i \mid \mathbf{B}_i, \boldsymbol{\theta})}.$$

$$(4.9)$$

Therefore,

$$\begin{aligned}
\frac{P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i})} &= \int \frac{P(D_i = k \mid \mathbf{B}_i, \boldsymbol{\theta})}{P(D_i = d_i \mid \mathbf{B}_i, \boldsymbol{\theta})} P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i}) d\boldsymbol{\theta} \\
&= \int \frac{P(\mathbf{B}_i \mid D_i = k, \boldsymbol{\theta})P(D_i = k)}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta})P(D_i = d_i)} P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i}) d\boldsymbol{\theta} \\
&:= \frac{P(D_i = k)}{P(D_i = d_i)} Q_{kd_i}.
\end{aligned}$$

$$(4.10)$$

By using the fact that $\sum_{k=0,1} P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}) = 1$, we have

$$P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i}) = \frac{P(D_i = d_i)}{\sum_{k=0,1} P(D_i = k)Q_{kd_i}}, \qquad (4.11)$$

thus leading to the above LOOCV predictive probability (4.6). For $i = 1, \cdots, n$ and $k = 0, 1$, we compute

$$\hat{Q}_{kd_i} = \frac{1}{T} \sum_{t=1}^{T} \frac{P(\mathbf{B}_i \mid D_i = k, \boldsymbol{\theta}^{(t)})}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)})}. \qquad (4.12)$$

The estimate of $D_i$ is

$$\hat{D}_i = \arg\max_{k} \left( P(D_i = k)Q_{kd_i} \right). \qquad (4.13)$$

Since there are only two possible values for $D_i$, we only need to calculate $P(\mathbf{B}_i \mid D_i = k, \boldsymbol{\theta}^{(t)})$ and $P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)})$, where $k \neq d_i$, for each subject $i$.

### 4.3.2.2 Voxel-Level Prediction

Now we use the imaging data $\mathbf{B}_i(v) = (X_{ilg}(v), Y_{ilg}(v), Z_{ilg}(v))$ for subject $i$ at voxel $v$ and all the other data $\mathbf{A}_{-i} = \{\mathbf{A}_u\}_{u \neq i}$ to predict the disease status $D_i$ for subject $i$. Similar to (4.6), the voxel-level LOOCV predictive probabilities can be expressed by

$$P(D_i = k \mid \mathbf{B}_i(v), \mathbf{A}_{-i}) = \frac{P(D_i = k)Q_{kd_i}}{\sum_{k'=0,1} P(D_i = k')Q_{k'd_i}}, \tag{4.14}$$

where

$$Q_{kd_i} = \int \frac{P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}) / \sum_{k'=0,1} P(\mathbf{B}_i(v) \mid D_i = k', \boldsymbol{\theta})P(D_i = k')}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}) / \sum_{k'=0,1} P(\mathbf{B}_i \mid D_i = k', \boldsymbol{\theta})P(D_i = k')} P(\boldsymbol{\theta} \mid \{\mathbf{A}_i\}_{i=1}^n)d\boldsymbol{\theta}, \tag{4.15}$$

which is estimated by

$$\hat{Q}_{kd_i} = \frac{1}{T} \sum_{t=1}^{T} \frac{P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}^{(t)}) / \sum_{k'=0,1} P(\mathbf{B}_i(v) \mid D_i = k', \boldsymbol{\theta}^{(t)})P(D_i = k')}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)}) / \sum_{k'=0,1} P(\mathbf{B}_i \mid D_i = k', \boldsymbol{\theta}^{(t)})P(D_i = k')}. \tag{4.16}$$

Then the estimate of $D_i$ is

$$\hat{D}_i = \arg\max_k \left( P(D_i = k)Q_{kd_i} \right), \tag{4.17}$$

which is equivalent to

$$\hat{D}_i = \arg\max_k \left( P(D_i = k)\frac{1}{T} \sum_{t=1}^{T} P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}^{(t)}) \right). \tag{4.18}$$

$Q_{kd_i}$ is derived in the similar way as in the whole brain analysis. Let $\mathbf{B}_i(\bar{v})$ denote all the imaging data in the brain except voxel $v$. First, we write out the posterior

predictive probability:

$$P(D_i = k \mid \mathbf{B}_i(v), \mathbf{A}_{-i})$$

$$= \int P(D_i = k \mid \mathbf{B}_i(v), \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \mathbf{B}_i(v), \mathbf{A}_{-i}) d\boldsymbol{\theta}$$

$$= \int P(D_i = k \mid \mathbf{B}_i(v), \boldsymbol{\theta}) \frac{P(\boldsymbol{\theta} \mid \mathbf{B}_i(v), \mathbf{B}_i(\bar{v}), \mathbf{A}_{-i})}{P(\boldsymbol{\theta} \mid \mathbf{B}_i(v), \mathbf{B}_i(\bar{v}), D_i = d_i, \mathbf{A}_{-i})} P(\boldsymbol{\theta} \mid \mathbf{B}_i(v), \mathbf{B}_i(\bar{v}), D_i = d_i, \mathbf{A}_{-i}) d\boldsymbol{\theta},$$

$$(4.19)$$

in which

$$\frac{P(\boldsymbol{\theta} \mid \mathbf{B}_i(v), \mathbf{B}_i(\bar{v}), \mathbf{A}_{-i})}{P(\boldsymbol{\theta} \mid \mathbf{B}_i, \mathbf{B}_i(\bar{v}), D_i = d_i, \mathbf{A}_{-i})} = \frac{P(\boldsymbol{\theta}, \mathbf{B}_i, \mathbf{A}_{-i}) P(\mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i})}{P(\mathbf{B}_i, \mathbf{A}_{-i}) P(\boldsymbol{\theta}, \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i})} = \frac{P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(D_i = d_i \mid \mathbf{B}_i, \boldsymbol{\theta})}.$$

$$(4.20)$$

Thus,

$$\frac{P(D_i = k \mid \mathbf{B}_i(v), \mathbf{A}_{-i})}{P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i})}$$

$$= \int \frac{P(D_i = k \mid \mathbf{B}_i(v), \boldsymbol{\theta})}{P(D_i = d_i \mid \mathbf{B}_i, \boldsymbol{\theta})} P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i}) d\boldsymbol{\theta}$$

$$= \int \frac{P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}) P(D_i = k)/P(\mathbf{B}_i(v) \mid \boldsymbol{\theta})}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}) P(D_i = d_i)/P(\mathbf{B}_i \mid \boldsymbol{\theta})} P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i}) d\boldsymbol{\theta}$$

$$:= \frac{P(D_i = k)}{P(D_i = d_i)} Q_{kd_i},$$

$$(4.21)$$

which leads to (4.14).

The voxel-level prediction result can be used as a way to select the regions that are highly associated with the disease status if the accuracy rate of prediction is high in these regions. A more sophisticated way to perform feature selection using our model is discussed in section 4.5.

## 4.4 Results

### 4.4.1 Parkinson's Disease Data

We applied our proposed Bayesian spatial model to PD data, which has T1 and resting-state fMRI images available; therefore, our model is reduced to the model which includes two imaging modalities, VBM and fALFF, and only considers data in the grey matter. Our goal is to make a whole-brain prediction of the disease status using the imaging data and discuss the voxel-level prediction as well. By evaluating the accuracy rate of the prediction at each voxel, we will be able to identify the regions that are highly associated with Parkinson's disease, as an alternative way to perform feature selection.

In the estimation procedure, the hyperparameters for prior distribution are set to provide vague information to ensure that the results are dominated by the information in the data. Specifically, all the hyperparameters in the inverse-gamma distribution are set to $10^{-3}$ (Spiegelhalter et al., 1994, 2003), the normal prior for $\zeta_{klg}$ is assumed to have mean $a_\zeta = 0$ and variance $b_\zeta = 10^5$. In the inverse-Wishart distribution, the degrees of freedom $\nu$ should be greater than $G - 1$ to build a proper distribution, so we set $\nu = G$, which provides the least information based on our data. The scale matrix $\mathbf{\Lambda}$ is set as $10^{-3} \times \mathbf{I}_G$, where $\mathbf{I}_G$ is a $G \times G$ identity matrix.

We perform a total of 10,000 MCMC iterations including 5,000 burn-in iterations, and store the results thinning by 10. Due to the huge amount of parameters included in our model, it is unrealistic to check the convergency of each parameter. We randomly check the trace plots for the parameters at voxel-level, subregion-level, and region-level, respectively, and some examples are presented in Appendix G.

After estimating the model parameters, we perform a whole-brain and voxel-level prediction using posterior samples based on procedures described in section 4.3.2. Here, we have a total of 500 posterior samples. By assuming an equal probability for

Figure 4.1: The distribution of average accuracy rates for prediction across subjects for all the voxels included in the analyses.
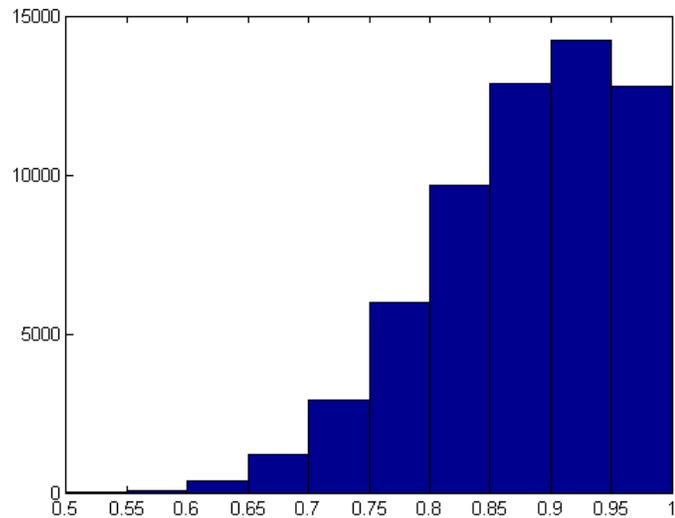


Table 4.1: Summary of average accuracy rates for prediction across subjects.

| Accuracy rate | [80%, 85%) | [85%, 90%) | [90%, 95%) | [95%, 100%) | 100% |
|---|---|---|---|---|---|
| Number of voxels | 5993 | 9663 | 12878 | 14236 | 12764 |
| (Percentage) | (9.97%) | (16.07%) | (21.42%) | (23.68%) | (21.23%) |

disease and non-disease, our model achieves 100% accuracy rate in the whole-brain prediction based on LOOCV.

The results from voxel-level prediction provide interesting information as well. The highest voxel-level accuracy rate is 100%, and the lowest is 50%. Figure 4.1 shows the distribution of the average accuracy rate across subjects for all the voxels included in the analysis, and Table 4.1 gives the number of voxels that achieves high accuracy rates. Also, an average whole-brain prediction map based on the results from voxel-level prediction across subjects are presented in Figure 4.2.

To identify the regions which are predictive for the disease status, we compute the average accuracy rates across voxels within a region. The right rectus, which is associated with cognitive impairment in PD patient, and is shown to have different

Figure 4.2: The average prediction map based on the voxel-level prediction results across subjects.
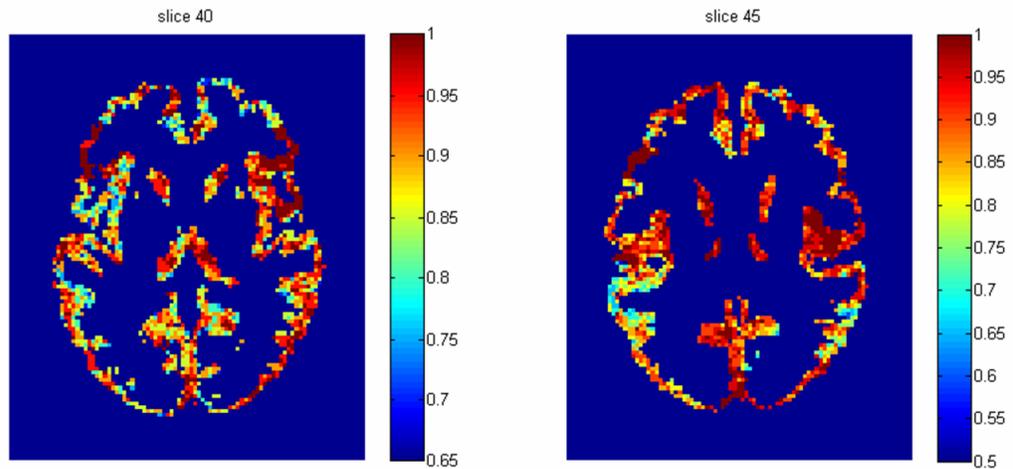


grey matter density between PD and controls (Nagano-Saito et al., 2005), is identified in our analysis. Not surprisingly, we find the bilateral caudate and the left putamen as regions with accurate predictions since PD is caused by the loss of dopaminergic neurons in the substantia nigra that project to striatal neurons in the caudate nucleus and putamen (Spencer et al., 1992). The right fusiform gyrus, which is believed to related to impaired ability to correctly identify negative facial expressions (Geday et al., 2006), and the left inferior parietal lobule which is involved in the perception of emotions in facial stimuli, may play a role of differentiating healthy controls and PD patients as well. In addition to fusiform and inferior parietal lobule, other regions which are involved in face perception such as the right mid-temporal pole is also identified. The left postcentral gyrus, the left superior parietal lobule, and the right superior medial frontal gyrus also standout since all of them are parts of the sensory system. A region-level prediction map based on the average accuracy rates across voxels within a region is shown in Figure 4.3.

Figure 4.3: The region-level prediction map based on the average accuracy rates across voxels within a region.
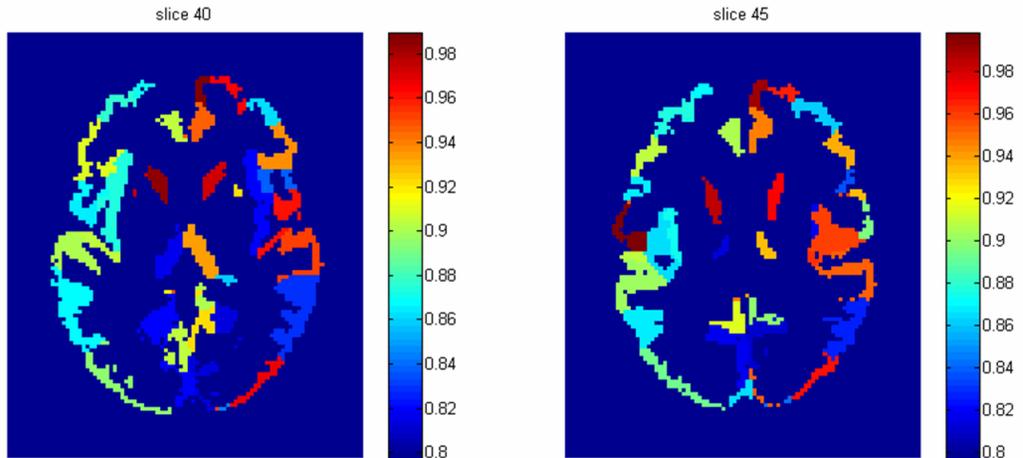


## 4.4.2  Simulation Studies

We conduct a simulation study to evaluate the performance of our proposed model. The purpose of this simulation study is to show that the MCMC samplers generated from our model is close to the true values, also the whole-brain prediction is accurate. In addition, our model can distinguish the regions that are predictive for disease status.

We assume that the imaging data is generated from the likelihood function of our model. We simulate data for 25 subjects from three AAL regions, the number of subregions within an AAL region has a mean and variance of 3, and the number of voxels within a subregion has a mean and variance of 50. Here, we specify the true values for the parameters in the likelihood function, i.e., $c_{klg}(v)$, $\gamma_{klg}(v)$, $\beta_{klg}(v)$, $\alpha_{ilg}$, $\eta_{kg}$, and $\delta_{lg}$, which are the most relevant parameters for voxel-level inference and future prediction. In this way, we can compare our posterior estimations with specified true values. All the other parameters are updated from the posterior distributions. And the hyperparameters are set as same as in the real data application. We select some of the subregions to be the regions that are associated with the disease, and

a region is classified into this category if it contains those selected subregions. We set different true values of parameters for disease and non-disease group if they are within the pre-specified regions, otherwise assume same true values for two groups. A total of 100 data sets are drawn in the simulation study.

First, we evaluate the posterior estimations by comparing the posterior means to the true values. Instead of examining a total of five thousand parameters which have known true values separately, we calculate the mean structure and variance of the likelihood function from posterior samplings and compare them to the truth since they are the most essential parts for inferences and predictions. The average bias (percentage change) in mean structure is $3.52 \times 10^{-2}$ (0.54%), and in variance is $1.04 \times 10^{-5}$ (1.04%). Secondly, we calculate the accuracy rate of a whole-brain prediction. The LOOCV achieves 100% for the whole-brain prediction for all 100 simulated data sets. Thirdly, we identify the regions that are highly associated with the disease status by evaluating the voxel-level accuracy rates for prediction. We compare the average accuracy rates for voxel-level prediction between the pre-specified regions and the others. Within the pre-specified regions, the average accuracy rate is 99.8%; for voxels which are in the other regions, the average accuracy rate is 71.7%. Here, we can see an improvements in prediction when voxels are from the pre-specified regions.

In summary, our model accurately perform posterior estimation with small bias, perfectly predict the disease status with the whole-brain imaging data, and correctly identify the regions that are highly associated with disease.

## 4.5 Discussion

We propose a Bayesian spatial model to predict the disease status of a subject using different modalities of imaging data, including fALFF, VBM and FA in grey and white

matter. Our framework perform voxel-level estimation for imaging data and conduct the whole-brain and voxel-level prediction of disease status based on the posterior predictive probabilities. Our model demonstrates the ability of estimating the mean and variance structures of the imaging data, predicting the disease status using the whole-brain imaging data, and identifying the regions which are highly associated with the disease based on voxel-level prediction results.

In our framework, we consider the spatial correlations at voxel level, subregion level, and region level, and specify different correlation structures such as exchangeable, CAR, and unstructured correlation matrices for them. Specifically, the intra-subregion correlation is described by a single value within each subregion; the inter-subregion correlation is modeled by a CAR model which borrows information from the subregions within a same AAL region; the inter-region correlation is assumed to have a unstructured correlation matrix.

We derive the posterior predictive probability using the whole brain data and data from a single voxel. Due to the complexity of computation, we adopt an importance sampling strategy to conduct LOOCV. We evaluate the accuracy rate of the whole-brain prediction and identify the regions that are predictive for disease based on the results from voxel-level accuracy rates.

One weakness of our method is the computational time. Since we are performing voxel-level estimation and the huge number of voxels included in our studies leads to a total of millions of parameters to estimate. However, by applying the importance sampling strategy, we only need to perform the posterior estimation once, and then the posterior predictive probabilities can be computed fairly efficiently.

Comparing to the existing feature selection methods, e.g., LASSO or elastic-net, our model uses different modeling strategy and criteria for selections. LASSO and elastic-net approaches model the probability of disease status, while our method starts from the imaging data. Also, we use posterior predictive probability as the criteria

to select the features, which is the exact target of the prediction problems; on the other hand, LASSO and elastic-net, from Bayesian perspective, use posterior mode to perform feature selections.

In our proposed method, we select the features based on the posterior predictive probability of each single voxel; ideally, we would like to identify the voxels $v \in \mathcal{V}$ s.t.

$$P(D_i = k \mid \{\mathbf{B}_i(v)\}_{v \in \mathcal{V}}, \mathbf{A}_{-i}) = P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}), \tag{4.22}$$

which could be a possible extension of our proposed approach.

For Parkinson's disease, there have been established criterion to perform diagnostics. Although prediction model may not serve as a practical way to diagnose the patient, it helps to understand the underlying neurological information of the disease, and to develop the effective treatment based on the findings from our model. Also, this model can be applied to the diseases which have not been fully investigated yet. Moreover, our model can be applied to the problem of predicting clinical responses to treatment based on different sources of imaging scans.

# Chapter 5

# Summary and Future Work

## 5.1 Summary

In this dissertation, we propose three statistical models for two important problems in the neuroimaging studies: functional connectivity and prediction.

In the first part of this dissertation, a Bayesian model is proposed to evaluate the functional connectivity in the human brain with additional information from structural imaging scans. The structural connectivity extracted from the DTI scans is used as a prior for estimating the functional connectivity, which is measured by the concurrence of elevated brain activities for BOLD signal from fMRI scans. In addition, we develop two measurements: functional coherence and ascendancy, one to capture the functional connectivity, the other to describe the functional hierarchy relationship between regions. Also, we make task-related inferences based on these two metrics.

The second model addresses another type of functional connectivity problem: the co-activation patterns in the meta analysis of functional neuroimaging studies. We model the number of activated coordinates in each region reported by different papers using an extension of the multivariate Poisson model by including a penalty to ac-

count for the sparsity of the functional network in the human brain. The co-activation patterns are characterized by the covariance between each region pair, and the functional network is built based on the estimated covariance. We perform statistical testings using permutation tests to evaluate the statistical significance of identified functional network and marginal co-activations for region pairs.

In the last part of this dissertation, we propose a Bayesian spatial model to predict the disease status using imaging containing both functional and structural information. We consider a two-level brain parcellation, and take into account three levels of spatial correlations. We perform prediction of disease status based on the estimation of imaging data using posterior predictive probabilities. Both whole-brain and voxel-level prediction results are obtained via importance sampling strategy. The voxel-level results are used as a criteria to identify the regions that are highly associated with the disease.

## 5.2  Future Work

In the first part of this dissertation, we perform analysis at region level which includes 90 AAL regions. A possible extension is to define a finer parcellation to create subregions within each AAL region. Also, we use a fixed function to link the structural and functional information of the data together. Alternatively, we can think about estimating function either in a empirical way or in a model framework.

In the penalized multivariate Poisson model, we may take into account the overdispersion problem for Poisson distribution by multiplying a coefficient to the sum of the independent Poisson variables. Also, we can work on the algorithms to simplify it in order to reduce the computational time.

For the prediction model, we identify the regions that are highly associated with the disease based on the voxel-level prediction results. A more statistically precise

way to perform feature selections is to start from the posterior predictive probabilities, and select the subset of the voxels which achieves closest probability to the predictive probability which uses the whole brain data. Also, we would like to extend our model to predict post-treatment clinical responses from the imaging data.

# Appendices

## Chapter 3 Appendices

### Appendix A: A Review of the Multivariate Poisson Distribution

Suppose we have a random vector composed of $2^p-1$ random variables $\mathbf{Y} = \{\mathbf{Y}_m\}_{m \in \mathcal{S}_p}$, where $\mathcal{S}_p = (\{1\}, \cdots \{p\}, \{12\}, \cdots \{12 \cdots p\})$, and $\mathbf{Y}_m$, for $m \in \mathcal{S}_p$, independently follow Poisson distributions with nonnegative parameter $\lambda_m$, respectively. Then a multivariate Poisson random vector $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p)$ can be represented as $\mathbf{X} = \mathbf{AY}$, where $\mathbf{A}$ is a $p \times (2^p - 1)$ matrix with zeros and ones. Denote $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_p]$, where $\mathbf{A}_i$ is a matrix of dimensions $p \times C_p^i$, and each column has $i$ ones and $p - i$ zeros. Each $\mathbf{X}_i$ marginally follows a Poisson distribution. For example, $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ have a multivariate Poisson distribution with parameters $\{\lambda_m\}_{m \in \mathcal{S}_3}$ in which

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{A}_2 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \text{and } \mathbf{A}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

The corresponding relationship between $\mathbf{X}$ and $\mathbf{Y}$ can be described as

$$\mathbf{X}_1 = \mathbf{Y}_1 + \mathbf{Y}_{12} + \mathbf{Y}_{13} + \mathbf{Y}_{123}$$

$$\mathbf{X}_2 = \mathbf{Y}_2 + \mathbf{Y}_{12} + \mathbf{Y}_{23} + \mathbf{Y}_{123} \qquad (6.1)$$

$$\mathbf{X}_3 = \mathbf{Y}_3 + \mathbf{Y}_{13} + \mathbf{Y}_{23} + \mathbf{Y}_{123},$$

where $\{\mathbf{Y}_m\}_{m \in \mathcal{S}_3}$ are independent Poisson random variables with parameters $\{\lambda_m\}_{m \in \mathcal{S}_3}$. In this trivariate case, $\lambda_{ij}$ is the covariance parameter between $\mathbf{X}_i$ and $\mathbf{X}_j$, and $\lambda_{123}$ is a 3-way interaction parameter. The mean vector and covariance matrix can be written as $E(\mathbf{X}) = \mathbf{AM}$ and $Cov(\mathbf{X}) = \mathbf{A\Sigma A}'$, where $\mathbf{M} = (\lambda_1, \cdots, \lambda_{123})'$ is the mean vector of $\mathbf{Y}$ and $\mathbf{\Sigma} = \text{diag}\{\lambda_1, \cdots, \lambda_{123}\}$ is the covariance matrix of $\mathbf{Y}$ (Mahamunulu, 1967; Loukas and Kemp, 1983; Johnson et al., 1997; Krummenauer 1998). In our model specifications, we make small modifications of the notions that are described above, $\mathbf{Y}_i$ is equivalent to $\mathbf{Y}_{ii}$, and $\lambda_i$ is equivalent to $\lambda_{ii}$.

## Appendix B: Derivation of the joint probability function in (3.6)

Since elements in $\mathbf{Y}_k$ are independent, the joint probability function $P(Y_{12,k}, X_{1,k}, X_{2,k})$ in (3.6) can be expressed as

$$P(Y_{12,k} = y_{12,k}, X_{1,k} = x_{1,k}, X_{2,k} = x_{2,k})$$

$$= P(Y_{12,k} = y_{12,k}, Y_{11,k} = x_{1,k} - y_{12,k}, Y_{22,k} = x_{2,k} - y_{12,k}) \qquad (6.2)$$

$$= e^{-\lambda_{12}} \frac{\lambda_{12}^{y_{12,k}}}{y_{12,k}!} e^{-\lambda_{11}} \frac{\lambda_{11}^{x_{1,k}-y_{12,k}}}{(x_{1,k} - y_{12,k})!} e^{-\lambda_{22}} \frac{\lambda_{22}^{x_{2,k}-y_{12,k}}}{(x_{2,k} - y_{12,k})!}$$

This probability function can also be solved using the following recursive expression

$$P(Y_{12,k}, X_{1,k}, X_{2,k}) = \frac{\lambda_{12}}{\lambda_{11}\lambda_{22}} \frac{\prod_{i=1}^{2}(x_{i,k} - y_{12,k} + 1)}{y_{12,k}} P(Y_{12,k} - 1, X_{1,k}, X_{2,k}), \qquad (6.3)$$

which reduces the computation time.

## Appendix C: Derivation of the joint probability function in (3.11)

We use the relation revealed in (3.8) and the independence among $Y_{ij,k}$'s to obtain the joint probability function $P(Y_{ij,k}, X_{i,k}, X_{j,k})$ in (3.11) as follows:

$$
\begin{aligned}
&P(Y_{ij,k} = y_{ij,k}, X_{i,k} = x_{i,k}, X_{j,k} = x_{j,k}) \\
&= P\left(Y_{ij,k} = y_{ij,k}, \sum_{s \neq j} Y_{is,k} = x_{i,k} - y_{ij,k}, \sum_{s \neq i} Y_{js,k} = x_{j,k} - y_{ij,k}\right) \\
&= e^{-\lambda_{ij}} \frac{\lambda_{ij}^{y_{ij,k}}}{y_{ij,k}!} e^{-\sum_{s \neq j} \lambda_{is}} \frac{\left(\sum_{s \neq j} \lambda_{is}\right)^{x_{i,k} - y_{ij,k}}}{(x_{i,k} - y_{ij,k})!} e^{-\sum_{s \neq i} \lambda_{js}} \frac{\left(\sum_{s \neq i} \lambda_{js}\right)^{x_{j,k} - y_{ij,k}}}{(x_{j,k} - y_{ij,k})!}
\end{aligned}
\tag{6.4}
$$

Similarly, we can find the following recursive relationship to reduce the computation time.

$$
\begin{aligned}
&P(Y_{ij,k}, X_{i,k}, X_{j,k}) \\
&= \frac{\lambda_{ij}}{\left(\sum_{s \neq j} \lambda_{is}\right)\left(\sum_{s \neq i} \lambda_{js}\right)} \frac{(x_{i,k} - y_{ij,k} + 1)(x_{j,k} - y_{ij,k} + 1)}{y_{ij,k}} P(Y_{ij,k} - 1, X_{i,k}, X_{j,k}),
\end{aligned}
\tag{6.5}
$$

## Appendix D: Derivation of the conditional probability in (3.15)

The conditional probability in (3.15) extends that in (3.11) because here we consider the probability of all the components of $\mathbf{Y}_k$, conditioning on the entire set of elements in $\mathbf{X}_k$. Using Bayes' theorem, we can express the conditional probability $P(\mathbf{Y}|\mathbf{X})$ as

follows:

$$P(\mathbf{Y}|\mathbf{X}) = P(Y_{12}, \cdots, Y_{1p}, Y_{23}, \cdots, Y_{2p}, \cdots, Y_{(p-1),p}|X_1, \cdots, X_p)$$

$$= \prod_{i<j} P(Y_{ij}|\{Y_{ij'}\}_{j'>j}, \{Y_{i'j}\}_{i'>i,i'\neq j}, X_i, X_j)$$

$$= \frac{P(Y_{ij} = y_{ij}, \sum_{j'<j} Y_{ij'} = x_i - \sum_{j'\geq j} y_{ij'}, \sum_{i'<i} Y_{i'j} + Y_{jj} = x_j - \sum_{i'\geq i,i'\neq j} y_{i'j})}{\sum_{y_{ij}} P(Y_{ij} = y_{ij}, \sum_{j'<j} Y_{ij'} = x_i - \sum_{j'\geq j} y_{ij'}, \sum_{i'<i} Y_{i'j} + Y_{jj} = x_j - \sum_{i'\geq i,i'\neq j} y_{i'j})},$$

$$(6.6)$$

where we drop the contrast index $k$ for simplicity. The joint probability function in the numerator of (6.6) can be simplified as follows:

$$P(Y_{ij} = y_{ij}, \sum_{j'<j} Y_{ij'} = x_i - \sum_{j'\geq j} y_{ij'}, \sum_{i'<i} Y_{i'j} + Y_{jj} = x_j - \sum_{i'\geq i,i'\neq j} y_{i'j})$$

$$= \exp(-\lambda_{ij} - \sum_{j'<j} \lambda_{ij'} - \lambda_{jj} - \sum_{i'<i} \lambda_{i'j}) \qquad (6.7)$$

$$\times \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \frac{(\sum_{j'<j} \lambda_{ij'})^{x_i - \sum_{j'\geq j} y_{ij'}}}{(x_i - \sum_{j'\geq j} y_{ij'})!} \frac{(\lambda_{jj} + \sum_{i'<i} \lambda_{i'j})^{x_j - \sum_{i'\geq i,i'\neq j} y_{i'j}}}{(x_j - \sum_{i'\geq i,i'\neq j} y_{i'j})!},$$

and further if we denote $P(Y_{ij}, \sum_{j'<j} Y_{ij'}, \sum_{i'<i} Y_{i'j} + Y_{jj})$ by $q(Y_{ij})$, then we obtain the recursive expression

$$q(Y_{ij}) = \frac{\lambda_{ij}}{y_{ij}} \frac{(x_i - \sum_{j'>j} y_{ij'} - y_{ij} + 1)}{\sum_{j'<j} \lambda_{ij}} \frac{(x_j - \sum_{i'>I,i'\neq j} y_{i'j} - y_{ij} + 1)}{\lambda_{jj} + \sum_{i'<i} \lambda_{i'j}} q(Y_{ij} - 1). \quad (6.8)$$

# Chapter 4 Appendices

## Appendix E: Brain parcellation

The brain parcellation is based on the AAL regions, of which the first 90 are of interest. We treat each region independently, and utilize a hierarchical clustering algorithm with average linkage to construct subregions. The number of clusters in each region is chosen based on the desired granularity of the sub-parcellation. We

consider both white and gray matter voxels, with a goal of developing a parcellation such that each subregion consists of both tissue-types. To enforce the restriction that subregions should be composed of physically contiguous voxels, we consider a distance matrix based on the Manhattan distance between voxels. However, using this criterion alone will yield numerous ties. Therefore, we use information derived from resting-state fMRI and DTI data to inform tie-breaking. Specifically, assume $d_{ig}$ represents the Manhattan distance between voxels $i$ and $j$. Then, we can define our distance matrix as:

$$
D_{ij} = \begin{cases} d_{ij} + (1 - \mathrm{FC}_{ij}) & \text{if } i \in \mathcal{G}, j \in \mathcal{G} \\ d_{ij} + (1 - \mathrm{SC}_{ij}) & \text{if } i \in \mathcal{W}, j \in \mathcal{W} \\ d_{ij} & \text{otherwise} \end{cases} \tag{6.9}
$$

Specifically, we allow gray matter voxel pairs to be informed by their functional connectivity, which is derived from the Pearson correlation between voxel time series. For white matter voxel pairs, we utilize DTI information; we consider the Pearson correlation between "connectivity profiles", which are based on probabilistic tractography implemented through FSL. For voxel pairs composed of different tissue types, we cannot use fMRI or DTI directly. Instead, we ignore the issue of ties for these voxels; as a consequence, the algorithm is partial to fusing voxels composed of different tissue types, and we tend to get resulting subregions with a greater balance of tissue composition.

## Appendix F: Posterior distributions

We present the full conditional posterior distributions for Gibbs sampling. Here, we take the parameters with superscript $xz$ as examples, parameters with other superscripts can be derived similarly. Let $\overline{\xi}$ denote all the parameters except $\xi$. Denote $\tilde{Z}_{ilg}(v) = Z_{ilg}(v) - \bar{Z}_{lg}(v)$. Then the full conditional posterior distributions can be

derived as follows:

(1) $[c_{klg}(v) \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{c}_{klg}(v)]$

$\propto [\mathbf{X}_{lg}(v) \mid \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, c_{klg}(v), \overline{c}_{klg}(v)] \, [c_{klg}(v) \mid \overline{c}_{klg}(v)]$

$$\propto \exp\left\{ -\frac{1}{2\delta_{lg}} \sum_{i=1}^{n} \left\{ X_{ilg}(v) - \sum_{k=0,1} [c_{klg}(v)\tilde{Z}_{ilg}(v) + \mathbf{W}_i\boldsymbol{\gamma}_{klg} + \beta_{klg}(v) + \alpha_{ilg} + \eta_{kg}]I(D_i = k) \right\}^2 \right.$$

$$\left. -\frac{1}{2\omega_{klg}}\left(c_{klg}(v) - \zeta_{klg}\right)^2 \right\}$$

$$\propto \mathrm{N}\left(\mu_{c_{klg}(v)}, \sigma^2_{c_{klg}(v)}\right)$$

$$\mu_{c_{klg}(v)} = \frac{\omega_{klg}\sum_{i=1}^{n} I(D_i = k)\tilde{Z}_{ilg}(v)\left\{X_{ilg}(v) - \mathbf{W}_i\boldsymbol{\gamma}_{klg} - \beta_{klg}(v) - \alpha_{ilg} - \eta_{kg}\right\} + \delta_{lg}\zeta_{klg}}{\omega_{klg}\sum_{i=1}^{n} I(D_i = k)\tilde{Z}_{ilg}(v)^2 + \delta_{lg}}$$

$$\sigma^2_{c_{klg}(v)} = \frac{\delta_{lg}\omega_{klg}}{\omega_{klg}\sum_{i=1}^{n} I(D_i = k)\tilde{Z}_{ilg}(v)^2 + \delta_{lg}}$$

(2) $[\gamma_{klgq}(v) \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\gamma}_{klgq}(v)]$

$\propto [\mathbf{X}_{lg}(v) \mid \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \gamma_{klgq}(v), \overline{\gamma}_{klgq}(v)] \, [\gamma_{klgq}(v) \mid \overline{\gamma}_{klgq}(v)]$

$$\propto \exp\left\{ -\frac{\gamma^2_{klgq}(v)}{2s_{klg}} \right.$$

$$\left. -\frac{1}{2\delta_{lg}}\sum_{i=1}^{n}\left\{ X_{ilg}(v) - \sum_{k=0,1}[c_{klg}(v)\tilde{Z}_{ilg}(v) + \sum_{q=1}^{Q} W_{iq}\gamma_{klgq}(v) + \beta_{klg}(v) + \alpha_{ilg} + \eta_{kg}]I(D_i = k) \right\}^2 \right\}$$

$$\propto \exp\left\{ -\frac{1}{2\delta_{lg}}\sum_{i=1}^{n} I(D_i = k)\left\{ W_{iq}^2\gamma^2_{klqg}(v) \right.\right.$$

$$\left.\left. -2W_{iq}\gamma_{klgq}(v)[X_{ilg}(v) - c_{klg}(v)\tilde{Z}_{ilg}(v) - \beta_{klg}(v) - \alpha_{ilg} - \eta_{kg} - \sum_{q'\neq q} W_{iq'}\gamma_{klgq'}(v)] \right\} - \frac{\gamma^2_{klgq}(v)}{2s_{klg}} \right\}$$

$$\propto \mathrm{N}\left(\mu_{\gamma}, \sigma^2_{\gamma}\right)$$

$$\mu_{\gamma} = \frac{s_{klg}\sum_{i=1}^{n} I(D_i = k)\left(W_{iq}[X_{ilg}(v) - c_{klg}(v)\tilde{Z}_{ilg}(v) - \beta_{klg}(v) - \alpha_{ilg} - \eta_{kg} - \sum_{q'\neq q} W_{iq'}\gamma_{klgq'}(v)]\right)}{s_{klg}\sum_{i=1}^{n} I(D_i = k)W_{iq}^2 + \delta_{lg}}$$

$$\sigma^2_{\gamma} = \frac{\delta_{lg}s_{klg}}{s_{klg}\sum_{i=1}^{n} I(D_i = k)W_{iq}^2 + \delta_{lg}}$$

(3) $[\beta_{klg}(v) \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\beta}_{klg}(v) \,]$

$\propto [\mathbf{X}_{lg}(v) \mid \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \beta_{klg}(v), \overline{\beta}_{klg}(v) \,] \, [\beta_{klg}(v) \mid \overline{\beta}_{klg}(v) \,]$

$$\propto \exp\left\{ -\frac{1}{2\delta_{lg}} \sum_{i=1}^{n} \left\{ X_{ilg}(v) - \sum_{k=0,1} [c_{klg}(v)\tilde{Z}_{ilg}(v) + \mathbf{W}_i\boldsymbol{\gamma}_{klg} + \beta_{klg}(v) + \alpha_{ilg} + \eta_{kg}]I(D_i = k) \right\}^2 \right.$$
$$\left. -\frac{(\beta_{klg}(v) - \overline{\beta}_{klg}(v))^2}{2\lambda_{klg}} \right\}$$

$\propto \mathrm{N}\left( \mu_{\beta_{klg}(v)}, \sigma^2_{klg}(v) \right)$

$$\mu_{\beta_{klg}(v)} = \frac{\lambda_{klg} \sum_{i=1}^{n} I(D_i = k)[X_{ilg}(v) - c_{klg}(v)\tilde{Z}_{ilg}(v) - \mathbf{W}_i\boldsymbol{\gamma}_{klg} - \alpha_{ilg} - \eta_{kg}] + \overline{\beta}_{klg}\delta_{lg}}{\lambda_{klg} \sum_{i=1}^{n} I(D_i = k) + \delta_{lq}}$$

$$\sigma^2_{\beta_{klg}(v)} = \frac{\delta_{lg}\lambda_{klg}}{\lambda_{klg} \sum_{i=1}^{n} I(D_i = k) + \delta_{lg}}$$

(4) $[\alpha_{ilg} \mid X_{ilg}(v), Y_{ilg}(v), Z_{ilg}(v), D_i, \mathbf{W}_i, \overline{\alpha}_{ilg} \,]$

$\propto \prod_{v \in l} [X_{ilg}(v) \mid Z_{ilg}(v), D_i, \mathbf{W}_i, \alpha_{ilg}, \overline{\alpha}_{ilg} \,] \, [\alpha_{ilg} \mid \overline{\alpha}_{ilg} \,]$

$$\propto \exp\left\{ -\frac{1}{2\delta_{lg}} \sum_{v \in l} \left\{ X_{ilg}(v) - \sum_{k=0,1} [c_{klg}(v)\tilde{Z}_{ilg}(v) + \mathbf{W}_i\boldsymbol{\gamma}_{klg} + \beta_{klg}(v) + \alpha_{ilg} + \eta_{kg}]I(D_i = k) \right\}^2 \right\}$$

$$\exp\left( -\frac{\alpha_{ilg}^2}{2\tau_{lg}} \right)$$

$$\propto \mathrm{N}\left\{ \frac{\tau_{lg} \sum_{v \in l}[X_{ilg}(v) - c_{klg}(v)\tilde{Z}_{ilg}(v) - \mathbf{W}_i\boldsymbol{\gamma}_{klg} - \beta_{klg}(v) - \eta_{kg}]}{\tau_{lg}V_{lg} + \delta_{lg}}, \frac{\delta_{lg}\tau_{lg}}{\tau_{lg}V_{lg} + \delta_{lg}} \right\}$$

$\propto \mathrm{N}\left( \mu_{\alpha_{ilg}}, \sigma^2_{\alpha_{ilg}} \right)$

(5) $[\delta_{lg} \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\delta}_{lg}]$

$\propto \prod_{v \in l} [\mathbf{X}_{lg}(v) \mid \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \delta_{lg}, \overline{\delta}_{lg}] [\delta_{lg} \mid \overline{\delta}_{lg}]$

$\propto \exp \left\{ -\frac{1}{2\delta_{lg}} \sum_{i=1}^{n} \sum_{v \in l} \left\{ X_{ilg}(v) - \sum_{k=0,1} [c_{klg}(v)\tilde{Z}_{ilg}(v) + \mathbf{W}_i \boldsymbol{\gamma}_{klg} + \beta_{klg}(v) + \alpha_{ilg} + \eta_{kg}] I(D_i = k) \right\}^2 \right.$

$\left. -\frac{b_\delta}{\delta_{lg}} \right\} \times \delta_{lg}^{-\frac{nV_{lg}}{2} - a_\delta - 1}$

$\propto \mathrm{InvG}\,(a, b)$

$a = a_\delta + \dfrac{nV_{lg}}{2}$

$b = b_\delta + \dfrac{\sum_{i=1}^{n} \sum_{v \in l} \left\{ X_{ilg}(v) - [c_{klg}(v)\tilde{Z}_{ilg}(v) + \mathbf{W}_i \boldsymbol{\gamma}_{klg} + \beta_{klg}(v) + \alpha_{ilg} + \eta_{kg}] \right\}^2}{2}$

(6) $[\zeta_{klg} \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\zeta}_{klg}]$

$\propto \prod_{v \in l} [c_{klg}(v) \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \zeta_{klg}, \overline{\zeta}_{klg}] [\zeta_{klg} \mid \overline{\zeta}_{klg}]$

$\propto \exp \left\{ -\frac{1}{2\omega_{klg}} \sum_{v \in l} (c_{klg}(v) - \zeta_{klg})^2 - \frac{1}{2b_\zeta} (\zeta_{klg} - a_\zeta)^2 \right\}$

$\propto \mathrm{N} \left( \dfrac{b_\zeta \sum_{v \in l} c_{klg}(v) + \omega_{klg} a_\zeta}{V_{lg} b_\zeta + \omega_{klg}}, \dfrac{\omega_{klg} b_\zeta}{V_{lg} b_\zeta + \omega_{klg}} \right)$

(7) $[\omega_{klg} \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\omega}_{klg}]$

$\propto \prod_{v \in l} [c_{klg}(v) \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \omega_{klg}, \overline{\omega}_{klg}] [\omega_{klg} \mid \overline{\omega}_{klg}]$

$\propto \exp \left\{ -\frac{1}{2\omega_{klg}} \sum_{v \in l} (c_{klg}(v) - \zeta_{klg})^2 - \frac{b_\omega}{\omega_{klg}} \right\} \omega_{klg}^{-\frac{V_{lg}}{2} - a_\omega - 1}$

$\propto \mathrm{InvG} \left( a_\omega + \dfrac{V_{lg}}{2}, b_\omega + \dfrac{\sum_{v \in l} (c_{klg}(v) - \zeta_{klg})^2}{2} \right)$

(8) $[s_{klg} \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \bar{s}_{klg} ]$

$$\propto \prod_{v \in l} [\boldsymbol{\gamma}_{klg}(v) \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, s_{klg}, \bar{s}_{klg} ] [s_{klg} \mid \bar{s}_{klg} ]$$

$$\propto \exp\left\{ -\frac{1}{2 s_{klg}} \sum_{v \in l} \sum_{q=1}^{Q} \gamma_{klgq}^2(v) - \frac{b_s}{s_{klg}} \right\} s_{klg}^{-Q \times V_{lg}/2 - a_s - 1}$$

$$\propto \mathrm{InvG}\left( a_s + \frac{Q \times V_{lg}}{2}, b_s + \frac{\sum_{v \in l} \sum_{q=1}^{Q} \gamma_{klgq}^2(v)}{2} \right)$$

(9) $[\beta_{klg} \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\beta}_{klg} ]$

$$\propto \prod_{v \in l} [\beta_{klg}(v) \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \beta_{klg}, \overline{\beta}_{klg} ] [\beta_{klg} \mid \overline{\beta}_{klg} ]$$

$$\propto \exp\left\{ -\frac{1}{2\lambda_{klg}} \sum_{v \in l} (\beta_{klg}(v) - \beta_{klg})^2 - \frac{n_{lg}}{2\phi_g} \left( \beta_{klg} - \frac{\rho_g}{n_{lg}} \sum_{l' \in N_g(l)} \beta_{kl'g} \right)^2 \right\}$$

$$\propto \mathrm{N}\left( \frac{\phi_g \sum_{v \in l} \beta_{klg}(v) + \lambda_{klg} \rho_g \sum_{l' \in N_g(l)} \beta_{kl'g}}{V_{lg} \phi_g + n_{lg} \lambda_{klg}}, \frac{\lambda_{klg} \phi_g}{V_{lg} \phi_g + n_{lg} \lambda_{klg}} \right)$$

(10) $[\lambda_{klg} \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\lambda}_{klg} ]$

$$\propto \prod_{v \in l} [\beta_{klg}(v) \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \lambda_{klg}, \overline{\lambda}_{klg} ] [\lambda_{klg} \mid \overline{\lambda}_{klg} ]$$

$$\propto \exp\left\{ -\frac{1}{2\lambda_{klg}} \sum_{v \in l} (\beta_{klg}(v) - \beta_{klg})^2 - \frac{b_\lambda}{\lambda_{klg}} \right\} \lambda_{klg}^{-\frac{V_{lg}}{2} - a_\lambda - 1}$$

$$\propto \mathrm{InvG}\left( a_\lambda + \frac{V_{lg}}{2}, b_\lambda + \frac{\sum_{v \in l} (\beta_{klg}(v) - \beta_{klg})^2}{2} \right)$$

(11) $[\tau_{lg} \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\tau}_{lg} ]$

$$\propto [\boldsymbol{\alpha}_{lg} \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \tau_{lg}, \overline{\tau}_{lg} ] [\tau_{lg} \mid \overline{\tau}_{lg} ]$$

$$\propto \exp\left\{ -\frac{1}{2\tau_{lg}} \sum_{i=1}^{n} \alpha_{ilg}^2 - \frac{b_\tau}{\tau_{lg}} \right\} \tau_{lg}^{-\frac{n}{2} - a_\tau - 1}$$

$$\propto \mathrm{InvG}\left( a_\tau + \frac{n}{2}, b_\tau + \frac{\sum_{i=1}^{n} \alpha_{ilg}^2}{2} \right)$$

(12) $[\phi_g \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\phi}_g]$

$$\propto \prod_{l \in g} [\boldsymbol{\beta}_{0lg}, \boldsymbol{\beta}_{1lg} \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \phi_g, \overline{\phi}_g] \, [\phi_g \mid \overline{\phi}_g]$$

$$\propto \exp \left\{ -\frac{1}{2\phi_g} \sum_{k=0,1} \sum_{l \in g} n_{lg} \left( \beta_{klg} - \frac{\rho_g}{n_{lg}} \sum_{l' \in N_g(l)} \beta_{kl'g} \right)^2 - \frac{b_\phi}{\phi_g} \right\} \phi_g^{-L_g - a_\phi - 1}$$

$$\propto \mathrm{InvG} \left( a_\phi + L_g, b_\phi + \frac{\sum_{k=0,1} \sum_{l \in g} n_{lg} \left( \beta_{klg} - \frac{\rho_g}{n_{lg}} \sum_{l' \in N_g(l)} \beta_{kl'g} \right)^2}{2} \right)$$

(13) $[\rho_g \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\rho}_g]$

$$\propto \prod_{l \in g} [\boldsymbol{\beta}_{0lg}, \boldsymbol{\beta}_{1lg} \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \rho_g, \overline{\rho}_g] \, [\rho_g \mid \overline{\rho}_g]$$

$$\propto \exp \left\{ -\frac{1}{2\phi_g} \sum_{k=0,1} \sum_{l \in g} n_{lg} \left( \beta_{klg} - \frac{\rho_g}{n_{lg}} \sum_{l' \in N_g(l)} \beta_{kl'g} \right)^2 \right\}$$

$$\propto \exp \left\{ -\frac{1}{2\phi_g} \left( \rho_g^2 \sum_{k=0,1} \sum_{l \in g} \frac{\left( \sum_{l' \in N_g(l)} \beta_{kl'g} \right)^2}{n_{lg}} - 2\rho_g \sum_{k=0,1} \sum_{l \in g} \beta_{klg} \sum_{l' \in N_g(l)} \beta_{kl'g} \right) \right\}$$

$$\propto L(\rho_g)$$

Suppose $\rho_g$ follows a discrete uniform distribution. In our analysis, $\rho_g$ can take 36 values $\{m_1, \cdots m_n\}$. Denote the posterior probability of $\rho_g$ as $\{p_1, \cdots, p_n\}$, then

$$p_j = \frac{L(\rho_g = m_j)}{L(\rho_g = m_1) + \cdots + L(\rho_g = m_n)}.$$

(14) Let $\boldsymbol{\beta}_{kg} = \left( \beta_{k1g}(v_1), \cdots, \beta_{k1g}(v_{V_{1g}}), \beta_{k2g}(v_1) \cdots, \beta_{kL_gg}(v_{V_{L_gg}}) \right)^T$ denote all the voxels in region $g$; define $\mathbf{c}_{kg}$, $\boldsymbol{\gamma}_{kg}$, $\mathbf{X}_{ig}$ and $\mathbf{Z}_{ig}$ in the same way; $V_g = \sum_{l=1}^{L_g} V_{lg}$ denote the number of voxels in region $g$; $\boldsymbol{\alpha}_{ig} = (\alpha_{i1g}, \cdots, \alpha_{iL_gg})^T$; $\boldsymbol{\Delta}_g = diag(\boldsymbol{\Delta}_{1g}, \cdots, \boldsymbol{\Delta}_{L_gg})$, where $\boldsymbol{\Delta}_{lg} = \mathbf{I}_{V_{lg}} \otimes \delta_{lg}$; $\odot$ denote Hadamard product.

$$[\boldsymbol{\eta}_k \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\boldsymbol{\eta}}_k]$$

$$\propto [\mathbf{X}_{lg}(v) \mid \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \boldsymbol{\eta}_k, \overline{\boldsymbol{\eta}}_k] \, [\boldsymbol{\eta}_k \mid \overline{\boldsymbol{\eta}}_k]$$

$$\propto \prod_{i=1}^{n} N \left( \sum_{k=0,1} [c_{klg}(v)\tilde{Z}_{ilg}(v) + \mathbf{W}_i\boldsymbol{\gamma}_{klg}(v) + \beta_{klg}(v) + \alpha_{ilg} + \eta_{kg}]I(D_i = k), \delta_{lg} \right) N(0, \boldsymbol{\Sigma}_k)$$

$$\propto \exp\left\{ -\frac{1}{2}\boldsymbol{\eta}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\eta}_k \right\}$$

$$\exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \sum_{g=1}^{G} \left\{ \mathbf{X}_{ig} - \sum_{k=0,1} \left\{ \mathbf{c}_{kg} \odot \tilde{\mathbf{Z}}_{ig} + \mathbf{W}_i\boldsymbol{\gamma}_{kg} + \boldsymbol{\beta}_{kg} + \left(\mathbf{1}_{V_{1g}}^T, \cdots, \mathbf{1}_{V_{L_gg}}^T\right)^T \otimes \boldsymbol{\alpha}_{ig} + \mathbf{1}_{V_g}\eta_{kg} \right\} \right.$$

$$\left. I(D_i = k) \right\}^T \boldsymbol{\Delta}_g^{-1}$$

$$\left\{ \mathbf{X}_{ig} - \sum_{k=0,1} \left\{ \mathbf{c}_{kg} \odot \tilde{\mathbf{Z}}_{ig} + \mathbf{W}_i\boldsymbol{\gamma}_{kg} + \boldsymbol{\beta}_{kg} + \left(\mathbf{1}_{V_{1g}}^T, \cdots, \mathbf{1}_{V_{L_gg}}^T\right)^T \otimes \boldsymbol{\alpha}_{ig} + \mathbf{1}_{V_g}\eta_{kg} \right\} I(D_i = k) \right\} \right\}$$

$$\propto \exp\left\{ -\frac{1}{2}\boldsymbol{\eta}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\eta}_k \right\}$$

$$\exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \sum_{g=1}^{G} \left\{ \sum_{k=0,1} \mathbf{1}_{V_g}\eta_{kg}I(D_i = k) - \right. \right.$$

$$\left. \underbrace{\left( \mathbf{X}_{ig} - \sum_{k=0,1} \left\{ \mathbf{c}_{kg} \odot \tilde{\mathbf{Z}}_{ig} + \mathbf{W}_i\boldsymbol{\gamma}_{kg} + \boldsymbol{\beta}_{kg} + \left(\mathbf{1}_{V_{1g}}^T, \cdots, \mathbf{1}_{V_{L_gg}}^T\right)^T \otimes \boldsymbol{\alpha}_{ig} \right\} I(D_i = k) \right)}_{=:\mathbf{t}_g} \right\}^T$$

$$\left. \boldsymbol{\Delta}_g^{-1} \left( \sum_{k=0,1} \mathbf{1}_{V_g}\eta_{kg}I(D_i = k) - \mathbf{t}_g \right) \right\}$$

$$\propto \exp\left\{-\frac{1}{2}\boldsymbol{\eta}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\eta}_k\right\}$$

$$\exp\left\{-\frac{1}{2}\underbrace{\sum_{i=1}^{N}\sum_{g=1}^{G}\left\{\sum_{k=0,1}\mathbf{1}_{V_g}\eta_{kg}I(D_i=k)\right\}^T \boldsymbol{\Delta}_g^{-1}\left\{\sum_{k=0,1}\mathbf{1}_{V_g}\eta_{kg}I(D_i=k)\right\}}_{I}\right.$$

$$\left.-\underbrace{\sum_{i=1}^{N}\sum_{g=1}^{G}\left\{\sum_{k=0,1}\mathbf{1}_{V_g}\eta_{kg}I(D_i=k)\right\}^T \boldsymbol{\Delta}_g^{-1}\mathbf{t}_g}_{II}\right\}$$

$$I = \boldsymbol{\eta}_k^T diag\left\{\underbrace{\sum_{i=1}^{n}I(D_i=k)\sum_{l=1}^{L_1}[\delta_{l1}^{-1}V_{l1}], \cdots, \sum_{i=1}^{n}I(D_i=k)\sum_{l=1}^{L_G}[\delta_{l1}^{-1}V_{lG}]}_{\boldsymbol{\Omega}_{\boldsymbol{\eta}_k}}\right\}\boldsymbol{\eta}_k$$

$$II = \boldsymbol{\eta}_k^T \underbrace{\left(P_{\boldsymbol{\eta}_{k1}}, \cdots, P_{\boldsymbol{\eta}_{kG}}\right)'}_{\mathbf{P}_{\boldsymbol{\eta}_k}}$$

$$P_{\boldsymbol{\eta}_{kg}} = \sum_{i=1}^{n}I(D_i=k)\sum_{l=1}^{L_g}[\delta_{lg}^{-1}\sum_{v=1}^{V_{lg}}\left(X_{ilg}(v) - c_{klg}(v)\tilde{Z}_{ilg}(v) - \mathbf{W}_i\boldsymbol{\gamma}_{klg} - \beta_{klg}(v) - \alpha_{ilg}\right)]$$
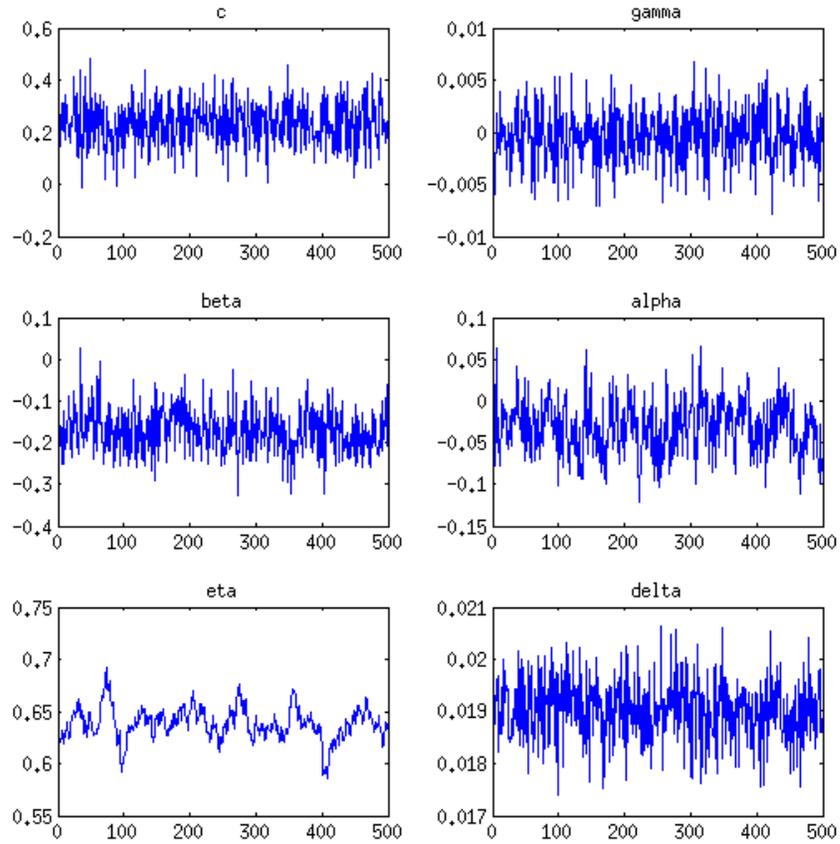
Then

$$[\boldsymbol{\eta}_k \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\boldsymbol{\eta}}_k]$$

$$\propto \exp\left\{-\frac{1}{2}\boldsymbol{\eta}_k^T(\boldsymbol{\Sigma}_k^{-1} + \boldsymbol{\Omega}_{\boldsymbol{\eta}_k})\boldsymbol{\eta}_k - 2\boldsymbol{\eta}_k\mathbf{P}_{\boldsymbol{\eta}_k}\right\}$$

$$\propto N\left((\boldsymbol{\Sigma}_k^{-1} + \boldsymbol{\Omega}_{\boldsymbol{\eta}_k})^{-1}\mathbf{P}_{\boldsymbol{\eta}_k}, (\boldsymbol{\Sigma}_k^{-1} + \boldsymbol{\Omega}_{\boldsymbol{\eta}_k})^{-1}\right)$$

(15) $[\boldsymbol{\Sigma}_k \mid \mathbf{X}_{lg}(v), \mathbf{Y}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \overline{\boldsymbol{\Sigma}}_k]$

$$\propto [\boldsymbol{\eta}_k \mid \mathbf{X}_{lg}(v), \mathbf{Z}_{lg}(v), \mathbf{D}, \mathbf{W}, \boldsymbol{\Sigma}_k, \overline{\boldsymbol{\Sigma}}_k] [\boldsymbol{\Sigma}_k \mid \overline{\boldsymbol{\Sigma}}_k]$$

$$\propto |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}\boldsymbol{\eta}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\eta}_k \right\} |\boldsymbol{\Sigma}_k|^{-\frac{\nu+G+1}{2}} \exp\left\{ -\frac{1}{2} tr(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_k^{-1}) \right\}$$

$$\propto |\boldsymbol{\Sigma}_k|^{-\frac{\nu+G+2}{2}} \exp\left\{ -\frac{1}{2} tr\left(\boldsymbol{\Lambda} + \boldsymbol{\eta}_k\boldsymbol{\eta}_k^T\right) \boldsymbol{\Sigma}_k^{-1} \right\}$$

$$\propto \text{InvW}\left(\boldsymbol{\Lambda} + \boldsymbol{\eta}_k\boldsymbol{\eta}_k^T, \nu+1\right)$$

## Appendix G: Trace plots

Figure 6.1: Trace plots for selected voxel-level, subregion-level, and region-level parameters from posterior sampling.

# Bibliography

Alqallaf, F., and Gustafson, P. (2001). On cross-validation of Bayesian models. *Canadian Journal of Statistics* **29**, 333-340.

Ashburner, J. and Friston, K.J. (2000). Voxel-based morphometry – the methods. *NeuroImage* **11**, 805-821.

Balslev, D., Nielsen, F.Å., Frutiger, S.A., Sidtis, J.J., Christiansen, T.B., Svarer, C., Strother, S.C., Rottenberg, D.A., Hansen, L.K., Paulson, O.B. and Law, I. (2002). Cluster analysis of activity-time series in motor learning. *Human brain mapping* **15**, 135-145.

Baumgartner, R., Ryner, L., Richter, W., Summers, R., Jarmasz, M. and Somorjai, R. (2000). Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. *Magnetic Resonance Imaging* **18**, 89-94.

Beckmann, C.F. and Smith, S.M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *Medical Imaging* **23**, 137-152.

Beckmann, C.F. and Smith, S.M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *NeuroImage* **25**, 294-311.

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* **34**, 144-155.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical*

*Society. Series B (Methodological)* **57**, 289-300.

Bowman, F. D., Patel, R., Lu, C. (2004). Methods for detecting functional classifications in neuroimaging data. *Human Brain Mapping* **23**, 109-119.

Bowman, F.D., Caffo, B.S., Spear Bassett, S., and Kilts, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* **39**, 146-156.

Breakspear, M., Brammer, M.J., Bullmore, E.T., Das, P. and Williams, L.M. (2004). Spatiotemporal wavelet resampling for functional neuroimaging data. *Human Brain Mapping* **23**, 1-25.

Calhoun, V.D., Adali, T., Pearlson, G.D., and Pekar, J.J. (2001). A Method for Making Group Inferences from Functional MRI Data Using Independent Component Analysis. *Human Brain Mapping* **14**,140-151.

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **41**, 687-699.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine learning* **20**, 273-297.

Cover, T. M. and Thomas, J. A. (1991). Elements of information theory. *New York: John Wiley & Sons.*

Cressie, N.A.C. (1993) *Statistics for Spatial Data.* New York: Wiley.

Dempster, A.P. and Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 1-38.

Derado, G., Bowman, F.D., and Zhang, L. (2012). Predicting brain activity using a Bayesian spatial model. *Statistical Methods in Medical Research* **0**, 1-16.

Di, X., Chan, R.C.K., and Gong, Q. (2009). White matter reduction in patients with schizophrenia as revealed by voxel-based morphometry: an activation likelihood estimation meta-analysis. *Progress in neuro-psychopharmacology and biological psychiatry* **33**, 1390-1394.

Dušek, P., Jech, R., Sieger, T., Vymazal, J., Růžička, E., Wackermann, J., and Mueller, K. (2012). Abnormal Activity in the Precuneus during Time Perception in Parkinson's Disease: An fMRI Study. *PloS one* **7**, e29635.

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K. and Fox, P.T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping* **30**, 2907-2926.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E.D., Kelly, R. L. and Peters, T. M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *Nuclear Science Symposium and Medical Imaging Conference*, 1813-1817.

Fiecas, M., Ombao, H., Linkletter, C., Thompson, W. and Sanes, J. (2010). Functional connectivity: shrinkage estimation and randomization test. *NeuroImage* **49**, 3005-3014.

Filzmoser, P., Baumgartner, R. and Moser, E. (1999). A hierarchical clustering method for analyzing functional MR images. *Magnetic Resonance Imaging* **17**, 817-826.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* **76**, 378-382.

Fox, P.T., Lancaster, J.L., Parsons, L.M., Xiong, J.H. and Zamarripa, F. (1997). Functional volumes modeling: Theory and preliminary assessment. *Human Brain Mapping* **5**, 306-311.

Friston, K. J., Frith, C. D., Liddle, P. F. and Frackowiak, R. S. (1993). Functional connectivity: the principal component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* **13**, 5-14.

Friston, K. J. and Penny, W. (2003). Dynamic causal modeling. *NeuroImage* **19**, 1273-1302.

Geday, J., Ostergaard, K., and Gjedde, A. (2006). Stimulation of subthalamic nucleus

inhibits emotional activation of fusiform gyrus. *NeuroImage* **33**, 706-714.

Gelfand, A.E. (1996). Inference and monitoring convergence. *Markov Chain Monte Carlo in Practice*, eds. Gilks, W.R., Richardson, S., and Spiegelhalter, D.J., Chapman & Hall, pp. 131-144.

Gelfand, A.E., Dey, D.K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). *Bayesian Statistics,* eds. Bernado, J.M., Berger, J.O., Dawid, A.P., and Smith, A.F.M., Oxford University Press, Vol.4, pp. 147-167.

Gelfand, A.E., and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**, 11-15.

Gelman, A., Rubin, D. B., Carlin, J., and Stern, H. (1995). Bayesian data analysis. *London: Chapman & Hall.*

Greicius, M. D., Supekar, K., Menon, V. and Dougherty, R.F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex* **19**, 72-78.

Guo, Y., Bowman, F.D., and Kilts, C. (2008). Predicting the brain response to treatment using a Bayesian hierarchical model with application to a study of schizophrenia. *Human brain mapping* **29**, 1092-1109.

Guo, Y. and Pagnoni, G. (2008). A Unified Framework for Group Independent Component Analysis for Multi-Subject fMRI Data. *NeuroImage* **42**, 1078-1093

Gwet, K. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods for Inter-rater Reliability Assessment* **1**, 1-6.

Ham, S.L. and Kwak, N. (2012). Boosted-PCA for binary classification problems. *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, 1219-1222.

Hampson, M., Peterson, B., Skudlarski, P., Gatenby, C. and Gore, J. (2002). Detection of Functional Connectivity Using Temporal Correlations in MRI images. *Human*

*Brain Mapping* **15**, 247-262.

Hendelman, W. (2000). Atlas of Functional Neuroanatomy. *CRC Press.*

Johnson, N.L. and Kotz, S. and Balakrishnan, N. (1997). Discrete multivariate distributions. *Recherche* **67**, 2.

Kang, J., Johnson, T.D., Nichols, T.E. and Wager, T.D. (2011). Meta Analysis of Functional Neuroimaging Data via Bayesian Spatial Point Processes. *Journal of the American Statistical Association* **106**, 124-134.

Kano, K. and Kawamura, K. (1991). On recurrence relations for the probability function of multivariate generalized Poisson distribution. *Communications in statistics-theory and methods* **20**, 165-178.

Kawamura, K. (1979). The structure of multivariate Poisson distribution. *Kodai Mathematical Journal* **2**, 337-345.

Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K. and Wager, T.D. (2008). Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *NeuroImage* **42**, 998-1031.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate discrete distributions* (New York, Marcel Dekker).

Kollias, S. (2009). Parcelation of the white matter using DTI: Insights into the functional connectivity of the brain. *The Neuroradiology Journal* **22** (Suppl. I).

Krummenauer, F. (1998). Limit theorems for multivariate discrete distributions. *Metrika* **47**, 47-69.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**, 805-811.

Landis, J.R. and Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**, 159-174.

Leroux, B.G. (1992). Consistent estimation of a mixing distribution. *Annals of Statistics* **20**, 1350-1360.

Li, J.P., Pan, P.L., Huang, R., and Shang, H.F. (2011). A meta-analysis of voxel-based morphometry studies of white matter volume alterations in Alzheimer's disease. *Neuroscience & Biobehavioral Reviews* **36**, 757-763.

Ling, J., Merideth F., Caprihan, A., Pena, A., Teshiba, T., and Mayer, A.R. (2012). Head injury or head motion? Assessment and quantification of motion artifacts in diffusion tensor imaging studies. *Human Brain Mapping* **33**, 50-62.

Loukas, S. and Kemp, C.D. (1983). On computer sampling from trivariate and multi-variate discrete distributions. *Journal of Statistical Computation and Simulation* **17**, 113-123.

Maddock, R.J., Garrett, A.S., and Buonocore, M.H. (2002). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human brain mapping* **18**, 30-41.

Mahamunulu, D.M. (1967). A note on regression in the multivariate Poisson distribution. *Journal of the American Statistical Association* **61**, 251-258.

Makris, N., Goldstein, J.M., Kennedy, D., Hodge, S.M., Caviness, V.S., Faraone, S.V., Tsuang, M.T. and Seidman, L.J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr Res.* **83**, 155-71.

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., and Mourão-Miranda, J. (2010). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *NeuroImage* **49**, 2178-2189.

Marrelec, G., Krainik, A., Duffau, H., Pelegrini-Issac, M., Lehericy, S., Doyon, J. and Benali, H. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage* **32**, 228-237.

Mayer, A. R., Franco, A. R. and Harrington D. L. (2009). Neuronal modulation of auditory attention by informative and uninformative spatial cues. *Human Brain Mapping* **30**, 1652-1666.

McIntosh A. R. and Gonzalez-Lima F. (1994). Structural equation modeling and its

application to network analysis in functional brain imaging. *Human Brain Mapping* **2**, 2-22.

McLachlan, G.J. and Krishnan, T. (1997). *The EM algorithm and extensions* (Wiley New York).

Meng, X.L. and Van Dyk, D. (1997). The EM Algorithm – an Old Folk-song Sung to a Fast New Tune. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**, 511-567.

Morgan, V. L., Mishra, A., Newton, A.T., Gore, J.C. and Ding, Z. (2009). Integrating functional and diffusion magnetic resonance imaging for analysis of structure-function relationship in the human language network. *PLoS ONE* **4(8)**, e6660.

Nagano-Saito, A., Washimi, Y., Arahata, Y., Kachi, T., Lerch, J.P., Evans, A.C., Dagher, A., and Ito, K. (2005). Cerebral atrophy and its relation to cognitive impairment in Parkinson disease. *Neurology* **64**, 224-229.

Neumann, J., Fox, P.T., Turner, R. and Lohmann, G. (2010). Learning partially directed functional networks from meta-analysis imaging data. *NeuroImage* **49**, 1372-1384.

Nielsen, F.Å. and Hansen, L.K. (2002). Modeling of activation data in the BrainMap database: Detection of outliers. *Human Brain Mapping* **15**, 146-156.

Nielsen, F.Å., Hansen, L.K. and Balslev, D. (2004). Mining for associations between text and brain activation in a functional neuroimaging database. *Neuroinformatics* **2**, 369-379.

Olson, I.R., Plotzker, A. and Ezzyat Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* **130**, 1718-1731.

Patel, R., Bowman, F. D. and Rilling, J. K. (2006a). A Bayesian approach to determining connectivity of the human. *Human Brain Mapping* **27**, 267-276.

Patel, R., Van De Ville, D., and Bowman, F. D. (2006). Determining Significant Connectivity by 4D Spatiotemporal Wavelet Packet Resampling of Functional Neu-

roimaging Data. *NeuroImage* **31**: 1142-1155.

Radua, J. and Mataix-Cols, D. (2009). Voxel-wise meta-analysis of grey matter changes in obsessive–compulsive disorder. *The British Journal of Psychiatry* **195**, 393-402.

Rammsayer, T., and Classen, W. (1997). Impaired temporal discrimination in Parkinson's disease: temporal processing of brief durations as an indicator of degeneration of dopaminergic neurons in the basal ganglia. *International Journal of Neuroscience* **91**, 45-55.

Rykhlevskaia, E., Gratton, G. and Fabiani, M. (2008). Combining structural and functional neuroimaging data for studying brain connectivity: A review. *Psychophysiology* **45**, 173-187.

Scott, W. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly* **19**, 321-325.

Smith, S. M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R., Saunders, J. , Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M. and Matthews, P.M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23(S1)**, 208-219.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., and Woolrich, M. W. (2010). Network modelling methods for FMRI. *NeuroImage* **54**, 875-891.

Spencer, D.D., Robbins, R.J., Naftolin, F., Marek, K.L., Vollmer, T., Leranth, C., Roth, R.H., Price, L.H., Gjedde, A., Bunney, B.S., Sass, K.J., Elsworth, J.D., Kier, E.L., Makuch, R., Hoffer, P.B., and Redmond, D.E. (1992). Unilateral transplantation of human fetal mesencephalic tissue into the caudate nucleus of patients with Parkinson's disease. *New England Journal of Medicine* **327**, 1541-1548.

Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., and Lunn, D. (1994,

2003). BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, England. www.mrc-bsu.cam.ac.uk/bugs/

Stonnington, C.M., Chu, C., Klöppel, S., Jack Jr, C.R., Ashburner, J., Frackowiak, R.S.J., et al. (2010). Predicting Clinical Scores from Magnetic Resonance Scans in Alzheimer's Disease. *NeuroImage* **51**, 1405-1413.

Strijbos, J. W. and Martens, R. L. and Prins, F. J. and Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education* **46**, 29-48.

Sun, D., Tsutakawa, R.K., Kim, H., and He, Z. (2000). Bayesian analysis of mortality rates with disease maps. *Statistics in Medicine* **19**, 2015-2035.

Talairach, J. and Tournoux, P. (1988) Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. *Thieme.*

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267-288.

Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S. and Poline, J.B. (2007). Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage* **35**, 105-120.

Turkeltaub, P.E., Eden, G.F., Jones, K.M. and Zeffiro, T.A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage* **16**, 765–780.

Tziortzi, A. C., Searle, G. E., Tzimopoulou, S., Salinas, C., Beaver, J. D., Jenkinson, M., Laruelle, M., et al. (2011). Imaging dopamine receptors in humans with [(11)C]-(+)-PHNO: Dissection of D3 signal and anatomy. *NeuroImage* **54**, 264–277.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer and B., M, J. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273-289.

Ungerleider, L. G. (1995). Functional Brain Imaing Studies of Cortical Mechanisms

for Memory. *Science* **270**, 769-774.

van den Heuvel, M. P., Mandl, R. C. W., Kahn, R. S. and Hulshoff Pol, H. E. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping* **30**, 3127-3141.

Vehtari, A., and Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation* **14**, 2439-2468.

Wager, T.D., Barrett, L.F., Bliss-Moreau, E., Lindquist, K., Duncan, S., Kober, H., Joseph, J., Davidson, M., and Mize, J. (2008). The neuroimaging of emotion. In Handbook of Emotion, M. Lewis, J.M. Haviland-Jones, and L.F. Barrett, eds. *New York: Guilford Press*, 249-271.

Wager, T.D., Hernandez, L., Jonides, J. and Lindquist, M. (2007a). Elements of functional neuroimaging, In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (Eds.), *Handbook of Psychophysiology*, 4th ed. Cambridge University Press, Cambridge, 19-55.

Wager, T.D., Jonides, J. and Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage* **22**, 1679-1693.

Wager, T.D., Phan, K.L., Liberzon, I. and Taylor, S.F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *NeuroImage* **19**, 513-531.

Woolrich, M. W., Ripley, B. D., Brady, M., Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage* **14**, 1370-1386.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301-320.