Clustered-Temporal Bayesian Models for Brain Connectivity in Neuroimaging Data

BY

NAIRITA GHOSAL
B.Sc., University of Calcutta, 2009
M.Sc., University of Kalyani, 2011
M.S., University of New Orleans, 2014

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2019

Chicago, Illinois

Defense Committee:
Sanjib Basu, Chair and Advisor
Saria Awadalla, Biostatistics
Dulal Bhaumik, Biostatistics
Runa Bhaumik, Computer Science
Michael L. Berbaum, Institute for Health Research and Policy

To my grandfather Mr. Ajay Chatterjee

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Professor Sanjib Basu for his continuous support and motivation. His advice on my research and on many facets of career and life has been invaluable. His immense knowledge and deep understanding was exceedingly helpful for my research. I am thankful to him for being patient and kind through out this journey. I could not have envision better advisor and mentor.

I am thankful to my committee members Dr. Saria Awadalla, Dr. Dulal Bhaumik, Dr. Runa Bhaumik and Dr. Michael Berbaum for their time and valuable suggestions.

I am extremely fortunate to have many wonderful teachers in my life. I am especially indebted to Dr. Sadhan Samar Maiti, Dr. Premadish Das, Dr. Linxiong Li, Dr. Tumulesh Solanky, Dr. Alan Polansky and Dr. Duchwan Ryu. My sincere gratitude goes to Dr. Bipasa Biswas, Dr. Ram Tiwari and Dr. James Travis for their guidance during my internships at FDA. I am thankful to Dr. Jyotirmoy Dey and his team at Abbvie for providing me the opportunity to learn greatly about clinical trials. Last but not least, I am very thankful to my high school mathematics teacher Samar Kumar Das, who is greatly responsible for my love of mathematics.

I also thank my friends Joydeep Sen, Patrick Stuchlik, Irandeep Kaur, Amy Elbrecht, Avik Roy Chowdhury, Shraddha Goyal and Sushmita Rajakutty for their incredible moral and emotional support.

ACKNOWLEDGMENTS (Continued)

I would like to thank my family members (too many to list here but you know who you are!) for their support. I especially like to thank my grandfather Ajay Chatterjee for showering me with love and affection.

Most importantly, I would like to thank my parents Samit Kumar Ghosal and Urmi Ghosal for their unconditional love and endless support.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF ABBREVIATIONS

ABIDE	Autism Brain Image Data Exchange
ADI-R	Autism Diagnostic Interview-Revised
AR	Autoregressive
ASD	Autism Spectrum Disorder
BOLD	Blood-oxygen-level-dependent
CAR	Conditional Auto Regressive
CBF	Cerebral Blood Flow
DLM	Dynamic linear model
DP	Dirichilet Process
DPM	Dirichilet Process Model
fMRI	Functional Magnetic Resonance Imaging
FPR	False Positive Rate
GLM	Generalized Linear Model
HIPAA	Health Insurance Portability and Accountability Act of 1996
HRF	Hemodynamic Response Function

LIST OF ABBREVIATIONS (Continued)

INDI	International Neuroimaging Datasharing Initiative
MCMC	Markov Chain Monte Carlo
MRF	Markov Random Field
OC	Over-connectivity
rs-fMRI	resting-state fMRI
ROI	Region of Interest
TNN	Task Negative Network
TPN	Task Positive Network
TPR	True Positive Rate

SUMMARY

The autism spectrum disorder (ASD) is a major public health concern in the United States. Neuroimaging analysis is becoming an important feature in ASD research. ASD has been associated with altered functional connectivity in brain. Functional connectivity can be measured by considering co-activation of brain regions in resting-state functional magnetic resonance imaging (fMRI). We developed Bayesian models to explore contrasting brain connectivity between subjects with ASD and controls using rs-fMRI data from Autism Brain Image Data Exchange (ABIDE) database. We considered two separate approaches for analyzing rs-fMRI data. We explored differential brain connectivity between subjects with ASD and controls using rs-fMRI data summarized over time by connectivity metrics. We developed three Bayesian models : the parametric model, Dirichlet Process Mixture (DPM) model and Neighborhood model on ROI pairs. Initially, we used these models to analyze cross-correlated ABIDE data set for all the sites separately. We observed that those links are not identical across all sites. Later, we combined the data set and reanalyze using above mentioned models. The combined analysis, models and identifies the links which were differentially connected across sites. Additionally, a regional-temporal model is proposed to directly model time sequence of rs-fMRI measurements at each brain region. We have applied dynamic linear model (DLM) to capture temporal structure. The potential correlation between connected regions was modeled using hidden Potts model with latent variable. An algorithm was developed for joint modeling of DLM and hidden Potts model with latent variable using fixed inverse temperature and hidden Potts model with

latent variable using pseudolikelihood approach. We applied the proposed approach to analyze time-dependent rs-fMRI ABIDE data set. We observed different patterns in heat maps between control and ASD subjects. Moreover, contrasting connectivity patterns can also be seen across sites. We can also notice differential connectivity between ROIs across lobes of brains of ASD subjects and controls. We observed dissimilarity in connectivity between ROIs across lobes across all the sites.

CHAPTER 1

INTRODUCTION

1.1 The Statistical Analysis of Functional Magnetic Resonance Imaging Data Set

Modern science have been fascinated with understanding human brain. Even in 2019, neuroscientists are still discovering crucial information about the brain. National Institute of Mental Health and the Library of Congress designated 1990s as "The Decade of the Brain". In the year 2003, the Nobel Prize in medicine was awarded to Paul C. Lauterbur and Peter Mansfield for their discoveries relating to "magnetic resonance imaging", which established the foundation of functional magnetic resonance imaging (fMRI) [Lazar, 2008]. Functional neuroimaging is an important tool to apprehend cognition, clinical and social psychology of mankind. The field of fMRI emerged in the year 1991 and and since then it has been expanding rapidly with an increasing number of scientific publications every year.

Statisticians and quantitative researchers play a crucial part in neuroscience and neuroimaging research. As data acquired from neuroimaging studies are complex and massive, the requirement for sophisticated statistical techniques have became increasingly prominent. The primary challenge in statistical inference arises due to small sample size in most of these studies. In the age of "Big Data", these studies are challenging researchers to develop statistical techniques which can be implemented in this scenario and are providing the opportunity to

make new discoveries about the human brain and understanding neurodegerative diseases and mental illnesses. According to [Van Horn and Toga, 2014] ""big data" can become "big" brain science".

1.2 Activated Brain Regions Detection

Functional magnetic resonance imaging (fMRI) depends on blood-oxygen-level-dependent (BOLD) signals and a time series of BOLD response can capture multiple simultaneous time points for each voxel of the brain. Each voxel represents a tiny cube of 3-dimensional brain tissue, which consists of a million or so brain cells.

One of the early statistical methods to model time series of BOLD response was proposed by [Friston et al., 1994] using generalized linear model (GLM). The BOLD response apprehends the vascular response and the stimulus in brain. The MR signal gets delayed hemodyanmically while evaluating the change in metabolism of BOLD contrast due as a result of an external stimulus. This type of hemodynamic response is called hemodynamic response function (HRF). Each input stimulus depends on the known external time-dependent stimulus function for a specific task and HRF, which depends on time-delay. [Friston et al., 1994] proposed a Poisson distribution to model HRF. Other researchers proposed Gaussian function [Friston et al., 1995] and Gamma functions [Lange and Zeger, 1997]. The Canonical HRF [Friston et al., 1998], which is defined as the difference of two gamma functions has also been used widely. [Lindquist and Wager, 2007] proposed three separate inverse logit functions.

Many models similar to the GLM model have been proposed under Bayesian framework. These models have hierarchical structures which makes modeling temporal and spatial features

of the data more adaptable. [Friston and Penny, 2003] presented a straightforward way for constructing posterior probability map using empirical Bayes method. This approach allows simple shrinkage priors based on variation between voxels. [Penny et al., 2003] used autoregressive (AR) error processes in the GLM and used variational Bayesian (VB) framework which approximates the true posterior density with a factorized density. [Flandin and Penny, 2007] described a spatially regularized GLM and proposed a spatial prior using sparse spatial basis functions (SSBFs).

1.2.1 Temporal Modeling

The temporal component of the HRF varies across brain voxels and most definitely across subjects. [Quirós et al., 2010] parameterized the HRF shape with a likely increase of signal and a subsequent exponential decay in the temporal dimension. [Woolrich et al., 2004] modeled temporal characteristic by assuming the parameters of the HRFs to be voxel-dependent. [Xia et al., 2009] modeled the HRF at each voxel non-parametrically.

There has been considerable work in apprehending temporal correlation in fMRI data through different possible error structure. [Friston et al., 1994] proposed to obtain estimate of the autocorrelation structure by prewhitening the fmri data, and then eliminating the correlation by applying a data-transformation. In Bayesian literature, researchers proposed to model the error terms by imposing an autoregressive structure of order q (AR(q)) [Lee et al., 2014]. Alternatively, wavelet transformations can be applied to filter noise. [Zhang et al., 2014] considered a general error structure and modeled the correlated noise.

1.2.2 Spatial Modeling

Voxel-level analysis of fMRI relies on spatial correlation as neighborhood voxels are likely to have similar response. Spatial dependence between voxels are captured using spatial priors on model parameters in Bayesian modeling. Gaussian Markov random field (GMRF) priors on regression coefficient vectors has been used by many researchers [Quirós et al., 2010]. [Harrison et al., 2008] used diffusion-based spatial priors on the regression coefficients. Conditional auto regressive (CAR) priors were used by [Harrison and Green, 2010]. [Flandin and Penny, 2007] used sparse spatial basis function (SSBF) priors on wavelet-based regression coefficients. An alternative approach is to select activated voxels, which can be seen as a variable selection problem. In Bayesian literature, a spike and slab on the regression coefficients were adopted to specify mixture distribution [Brown et al., 1998]. [Smith and Fahrmeir, 2007] considered a type of spatial Markov random field (MRF) prior to incorporate biological prior information while considering spatial interaction between voxels.

1.3 Brain Connectivity Modeling

Brain connectivity refers to networks that demonstrate relationship between brain regions. Interests in brain connectivity includes comparison of connectivity properties among multiple subgroups and between different scanning sessions. Connectivity patterns and their disruptions are useful to understand mental health disorders and neurodegenerative diseases. Two different types of connectivity are of interest, namely, functional connectivity and effective connectivity. A review of [Friston, 2011] and [Zhang et al., 2015a] is presented in this section.

1.3.1 Functional Connectivity

The objective of functional connectivity is to identify regions of the brain which present similar temporal characteristics. [Zalesky et al., 2012] proposed a simple approach of capturing the functional connectivity based on temporal correlations between regions of interest throughout the brain or between a "seed" region and other voxels throughout the brain. Principal component analysis (PCA) [Andersen et al., 1999] and independent component analysis (ICA) [Calhoun et al., 2001] have also been used to cluster the brain into sections that display similar temporal traits. [Cribben et al., 2012] proposed a data-driven technique called dynamic connectivity regression, which can detect change points over time in functional connectivity and estimates a graph or set of relationships between regions of interest. [Bhaumik et al., 2017] applied support vector machine (SVM) classifier with elastic net feature selection to differentiate resting-state functional connectivity of brain networks and to detect major depressive disorder. [Bhaumik et al., 2018] proposed a mixed-effects model, which can control false discovery rate while detecting disruptive connectivities for a neural network in brain.

[Patel et al., 2006] is one of the early works developing Bayesian methods to assess functional connectivity in which thresholding of fMRI time series data was used to identify presence or absence of enhanced activity. The relationship between pairs of specific brain regions was assessed by examining joint and marginal probabilities of elevated neural activity in hierarchical functional network. [Bowman et al., 2008] presented voxel-by-voxel modeling of the whole brain and region of interest (ROI) analysis within one integrated framework. Intra-regional correlations can be modeled by unstructed variance-covariance matrix and using an exchange-

able correlation structure. Spatially distant voxels with similar temporal characteristics were clustered using Dirichlet Process prior on parameters of long memory correlated error terms in [Zhang et al., 2014]. [Chen et al., 2016] proposed a Bayesian hierarchical model to model connectivity between voxel pairs from different brain regions. The connectivity is assumed to follow a mixture distribution with two components reflecting connected and non-connected voxel pairs. This modeling assumption derives from the "small-worldness" property, which assumes that two distant ROIs may be connected through a small percentage of highly connected nodes. A new statistic which measure the connectivity strength is introduced in this paper.

There is also a substantial literature on dynamic functional connectivity models, which study dynamic temporal interactions among brain regions. [Warnick et al., 2018] used a hidden Markov model to classify latent cognitive state and estimate time varying networks.

1.3.2 Effective Connectivity

Effective connectivity is defined as the explicit influence one neural system exerts over another. Unlike functional connectivity, it is dynamic (activity-dependent), and depends on a model of interactions. It typically generates models with small networks of connected brain regions. The most commonly used methods include structural equation models (SEM), dynamic causal models (DCM) and vector autoregressive (VAR) models.

1.4 Outline

The objective of this dissertation is to explore contrasting brain connectivity between subjects with Autism Spectrum Disorder (ASD) and controls using resting state fMRI (rs-fMRI) data. We develop Bayesian models and inference to explore differential connectivity using cross-

correlated functional connectivity between region of interest (ROI) pairs. We also propose a regional-temporal model to directly model time sequence of rs-fMRI measurements at each ROI. This dissertation is organized as follows. In Chapter 2, we discuss the basic concepts behind fMRI and functional connectivity as well as altered brain connectivity in individuals with ASD. We provide a detailed description on Autism Brain Image Data Exchange (ABIDE) data set and participant demographics. In Chapter 3, we explore differential functional connectivity between subjects with ASD and controls using rs-fMRI data summarized over time between ROI pairs. We develop three Bayesian models: the parametric model, Dirichlet Process Mixture (DPM) model and Neighborhood model on ROI pairs. Detailed simulation studies compare performance of the proposed models. We use these models to analyze cross-correlated ABIDE data set for all the sites separately. Later, we combine the data set and reanalyze using above mentioned models. In Chapter 4, we directly model time sequence of rs-fMRI measurement at each ROI. We propose a regional-temporal functional connectivity model. We develop a dynamic linear model (DLM) to capture temporal structure and the potential correlation between connected regions is modeled using hidden Potts model with latent variable. We develop Bayesian inference for this joint modeling of DLM and hidden Potts model. A detailed simulation study is performed to compare proposed approach to a comparable approaches. We apply the proposed approach to analyze time-dependent rs-fMRI ABIDE data set. Chapter 5 provides future works and conclusions.

CHAPTER 2

FUNCTIONAL CONNECTIVITY IN AUTISM

2.1 Introduction

The human brain consists of a network with anatomically separated regions that are continuously communicating with each other. Alteration in connectivity of brain networks has been associated with Alzheimer's disease, Parkinson's disease, depression, schizophrenia and others. In this chapter, we begin with a concise overview of the human brain. We then review basic concepts of fMRI and discuss altered brain connectivity in subjects with Autism Spectrum Disorder. We also discuss about Autism Brain Image Data Exchange database.

2.2 The Human Brain

The human brain is the central organ and the command center for the human nervous system. The extremely complex network of the brain consists of about 100 billion neurons and approximately 100 trillion connections. It weighs about three pounds and is divided into the left and right hemisphere, separated by the corpus callosum. The brain is comprised of three main parts: forebrain, midbrain and hindbrain. The forebrain consists of the cerebrum (also known as the cortex), the thalamus and the hypothalamus. The cerebrum is the largest part of the human brain. The thalamus is associated with sensory and motor function and the hypothalamus is involved with regulating thirst, hunger, circadian rhythm in the body and also in the modulation of reflex reaction and behavior, which are related to survival. The limbic

system, which supports emotion and behavior is made up of the thalamus, the hypothalamus, amygdala and the hippocampus. The hindbrain is comprised of cerebellum, pons and medulla. The cerebellum, which is LATIN for "little brain" plays an important role in motor control. Midbrain, pons and medulla make up the brain stem. The brain stem plays an important role in regulating cardiac and respiratory function. The cerebral cortex has four lobes namely frontal, parietal, occipital and temporal. The frontal lobe is engaged in higher functions like reasoning, planning as well as problem solving, emotion and motor control. The parietal lobe is responsible for recognition, perception and orientation. Visual processing is associated with the occipital lobe. The temporal lobe is associated with memory, speech and the identification of auditory stimuli.

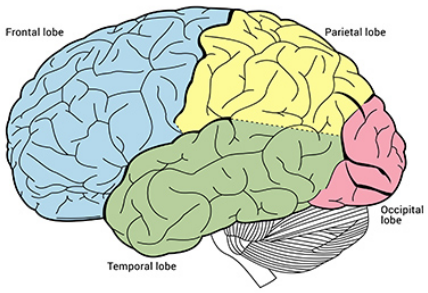


Figure 1: Lobes of the human brain

(Source: Queensland Brain Institute (https://qbi.uq.edu.au.))

2.3 Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) is a noninvasive neuroimaging method for measuring brain activity. Functional magnetic resonance imaging relies on BOLD signals that occur in response to neural activity. When an area of the brain becomes active in response to particular task or stimulus or even absence of any stimuli, the rate of blood flow increases and there is an increase in firing rate of neurons. Active neurons require significantly more oxygen, which causes an increase in oxygenated blood in adjacent blood vessels. fMRI measures the increase in metabolic demand of the active neurons. [Lazar, 2008]. The activated areas display shift in the ratio of oxygenated to deoxygenated blood, which is measured by hemodynamic response function (HRF).

Haemoglobin is an iron-containing oxygen transporting protein found in the red blood cell of almost every vertebrate. Haemoglobin can be saturated with oxygen molecules (oxyhaemoglobin) or desaturated with oxygen molecules (deoxyhaemoglobin). These two states have different magnetic properties, which can produce different magnetic fields as shown by [Pauling and Coryell, 1936]. Oxyhaemoglobin is diamagnetic and does not affect magnetic fields. On the other hand, deoxyhaemoglobin exhibits paramagnetism and causes distortion in the main magnetic fields. This disturbance can appear as a decay of transverse magnetization related to T_2 and T_2^* weighted MRI sequence ([Thulborn et al., 1982], [Ogawa et al., 1990]).

An increase in metabolic activity in a specific region of the cortex demands increased cerebral blood flow (CBF). This results in an increase in deoxyhemoglobin within the capillaries and an initial drop in oxyhaemoglobin causing the initial dip. After a lag of $2-6$ seconds, CBF

increases and delivers an excess of oxygenated haemoglobin and reduces deoxyhaemoglobin level. The response increases gradually and reaches to the peak at about $4-6$ seconds. If there is no more stimulus, the response will decline to an amplitude below baseline prior to the final return to baseline. This process takes approximately $15-20$ seconds to complete. Figure 2 shows a diagram of HRF for a voxel in an active brain region.

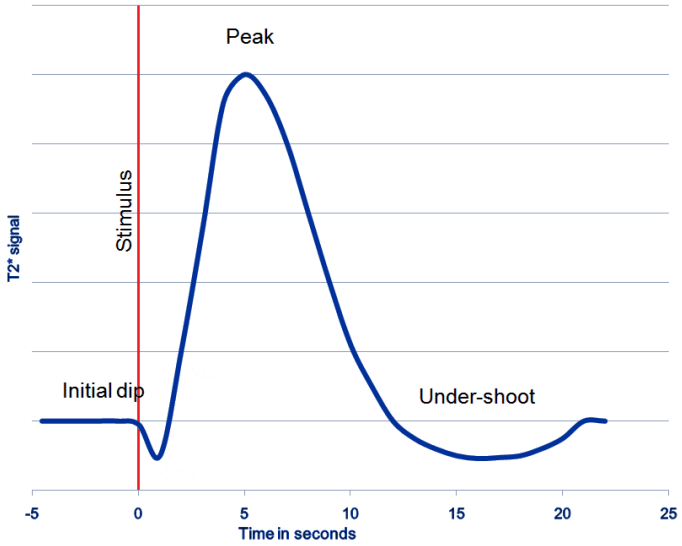


Figure 2: Blood Oxygen Level Dependent(BOLD) response.

(Source: https://radiopaedia.org)

A wide range of fMRI study designs is available to measure different neuronal activity.

- **Event-related design:** Stimulus are administered to the subject for a brief period mostly in random order. This produces BOLD responses which consist of short burst of activity.
- **Block design:** Multiple stimuli are administered like a block pattern or phase of $10-30$ seconds. This design helps to attain peak of BOLD signal for a longer interval, creating a plateau in the graph.
- **Resting-state design:** Resting state occurs when an explicit stimuli is not present. This design is also applied to understand the functional connectivity of the brain.

2.3.1 Functional Connectivity and Resting-state fMRI

The human brain consists of functionally linked and anatomically separated regions that are in continuous communication with each other. Functional connectivity is the temporal dependence of activation pattern of brain regions ([Aertsen et al., 1989], [Friston et al., 1993]). It is often analyzed by measuring co-activation of the brain regions during resting-state fMRI(rs-fMRI) [Lowe et al., 2000]. Resting-state fMRI can provide insight on alteration of functional connectivity due to neurodegenerative diseases [Greicius, 2008].

Multiple studies started to examine functional connectivity between brain regions during rs-fMRI around 15 years after the invention of fMRI [Van Den Heuvel and Pol, 2010]. Volunteers were instructed to relax and avoid thinking something in particular during these rs-fMRIs. Biswal et al.(1995 and 1997) showed that left and right hemispheric regions of the primary motor network demonstrated a high correlation between their fMRI BOLD time-series. Several other studies showed that there is high level of functional connectivity between primary visual

network, auditory network and higher order cognitive networks, including [Biswal et al., 1997]; [Cordes et al., 2002].

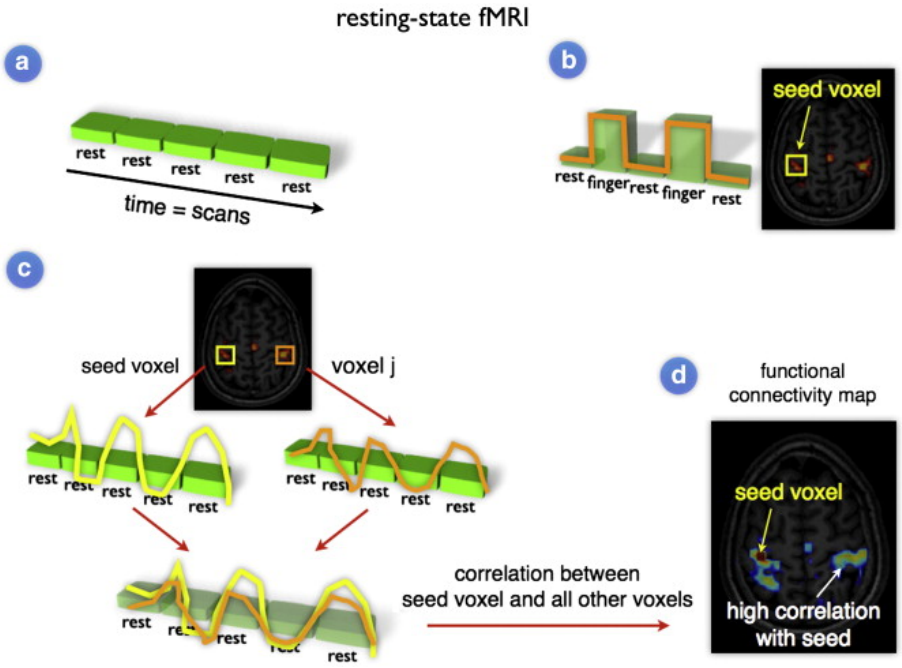


Figure 3: Functional connectivity and resting-state fMRI in Brain.

(Source: [Van Den Heuvel and Pol, 2010])

In Figure 3 panel a, BOLD rs-fMRI signal is measured during scans. In panel b, a seed region of interest is selected and its BOLD response was measured. Panel c shows the correlation between resting-state timeseries of voxel i and voxel j. The functional connectivity map (panel d) resulted from the timeseries of a voxel being highly correlated with time series of other voxels.

2.4 Autism Spectrum Disorder

The autism spectrum disorder (ASD) is a neurodevelopmental disorder which can cause restrictive and repetitive behavior and social, communication, behavioral challenges. Autism is known as a "spectrum" disorder due to the extensive variation in the type and severity of symptoms patients may experience.

According to DSM-IV, published by American Psychriatric [Association et al., 2000], ASD includes autistic disorder, pervasive developmental disorder not otherwise specified (PDD-NOS), Asperger syndrome, disintegrative disorder and Rett disorder. Pervasive developmental disorder not otherwise specified (PDD-NOS) is characterized by failure to obtain or early loss of communication and social interaction skills as well as reduced interest and repetitive behavior. Children with PDD also show poor motor coordination such as clumsiness. Rett disorder is a genetic disorder of postnatal brain development caused by a X-linked gene defect. This abnormal gene is responsible for death before or shortly after birth for most male fetuses therefore it is seen almost exclusively to affect girls. Normal functioning child up to age four starts showing sign of reduced social, verbal and cognitive development due to this disorder.

The autism spectrum disorder (ASD) is reported to prevail in all racial, ethnic and socioeconomic groups. It affects 1 in 59 children in the United States according to the Centers for Disease Control and Prevention's (CDC) Autism and Developmental Disabilities Monitoring (ADDM) [Baio J, 2018] network. It is also reported to occur about four times more among boys than among girls. Approximately, 31% of children identified with ASD also have intel-

lectual disability. The average prevalence of ASD is between 1% and 2% according to studies conducted in Asia, Europe and North America.

The causes of ASD have not yet been fully understood. It has a wide variety of both genetic and non-genetic causes. The rate of recurrence of ASD in siblings of affected subject is 2% to 8% which is much greater than prevalence rate in general population. Early studies of twins in the UK and Scandinavia reported that chances of being affected by autism for both monozygotic twins is $> 60\%$ whereas no concordance between dizygotic twins was found. The higher rate in monozygotic twins provides a stronger evidence for the robust influence of genetics in cause of autism. After reevaluating unaffected twins for broader autistic phenotypes such as communication skills and social disorders, the concordance for UK monozygotic twins increased to 92% and for UK dizygotic twins increased to 10% [Rebecca Muhle and Rapin, 2004]. [Marc G. Weisskopf and Roberts, 2015] discussed association with air pollution exposure during the third trimester of pregnancy and risk of ASD. Prenatal and first year of life exposure to nitrogen dioxide, particle matter, ozone, and near-roadway air pollution has been associated with ASD severity in children [Kerin et al., 2018]. Factors such as viral infections, medications or complications during pregnancy can also trigger ASD.

The link between autism and vaccination has received a close review over last two decades. Two studies had claimed that the mumps, measles and rubella (MMR) vaccine causes autism and were later found to be critically flawed. Wakefield et. al (1998) published an article in Lancet stating a hypothesis that MMR vaccine caused many side effects such as intestinal inflammation, entrance into the bloodstream of proteins harmful to the brain, and subsequent

development of autism. They also reported that 8 children developed autism within one month of receiving MMR vaccine. It was later discovered that those children had intestinal inflammation even before receiving vaccine. This study was later retracted as the study did not have a control group hence it was impossible to determine whether there is a causal relationship between MMR vaccine and autism. [Taylor et al., 2014] performed a meta analysis based on five cohort studies involving $1,256,407$ children and five case-control studies involving $9,920$ children. They found no relationship between developing ASD and vaccination, MMR vaccination, mercury and thimerosal.

2.5 Neuroimaging Analysis in Autism Spectrum Disorder

Neuroimaging analysis is becoming an increasingly important technique in ASD research. At present, ASD diagnosis solely depend on clinical assessment of an individual's verbal and physical behavior. Delay in diagnosis can lead an intellectually efficient children to social impairments. Recent advances of neuroimanging technology are making a path for identifying ASD based on the underlying brain activation patterns. However, classification of ASD can be complex because of heterogeneous nature of the disorder.

2.5.1 Functional Connectivity in Autism Spectrum Disorder

Researchers agree that ASD is associated with alterations in brain connectivity. But there is a growing debate about the precise nature of these alterations such as over-connectivity, under-connectivity or a combination of both. A review of functional connectivity ([Hull et al., 2017], [Uddin et al., 2013]) in ASD is provided in this section.

Under-connectivity is a statistically significant decrease in connectivity relative to a standard or normative reference value. A reduced inter-regional network mostly in frontal and posterior integration can be observed in task-based neuroimaging. Over-connectivity is reported when there is a statistically significant correlation in the affected subjects compared to control. More diffused connectivity of brain network in ASD has been reported recently such as stronger correlation with ectopic regions, which are outside of the generally defined circuit.

[Kennedy et al., 2006] compared default mode network (DMN) connectivity in ASD and control. The default mode network (DMN) is referred to as a "task-negative" network (TNN) as its activity reduce during performance of a cognitive task. Moreover, it is essential for social and emotional behavior. This study showed that TNN deactivation did not appear between resting-state and task-associated fMRI for ASD subjects. In a follow up study, [Kennedy and Courchesne, 2008] demonstrated under-connectivity within the TNN in ASD but it did not find any significant changes in the "task-positive" network (TPN). [Lee et al., 2016] investigated inter- and intra-hemispheric connectivity using Autism Brain Image Data Exchange (ABIDE) data set. This study suggested that under-connectivity in medial prefrontal cortex and posterior cingulate cortex may contribute to social impairments. Various studies reported under-connectivity in the brain regions which processes voice perception and language development such as posterior superior temporal sulcus. Under-connectivity was also observed in interhemispheric Broca's area, between the right cerebellar region and supratentorial language area and also in the dorsolateral prefrontal cortex (DLPFC). Communication challenges may have occurred due to these under-connectivity.

Over-connectivity (OC) have been reported in ASD using region of interest (ROI) seed-based analysis and as well as in graph theory and network analysis. [Di Martino et al., 2011] showed that OC can be observed between the striatum and right superior temporal gyrus and insular cortex and remarkably between striatum and pons, and more over between the pons and insular cortex. [Nebel et al., 2014] demonstrated increased and less segregated functional connectivity in primary motor cortex. Studies found correlation with degree of OC and of severity symptoms, which can suggest that ASD can be caused by the dysfunctional sensory connectivity.

In recent years, many literature demonstrated trends of both under and over connectivity in ASD, based on whether researchers are examining local or global networks. [Hahamy et al., 2015] found both under and over connectivity using a graph theory approach on the ABIDE data set. But this study had difficulty in finding a prevalent trend in ASD because of the uniqueness of the condition for each individual.

Many recent papers are focused on age-related changes in brain connectivity to understand developmental alternation on functional connectivity in ASD. This may explain the disagreement in hyper- and hypo- connectivity theory. [Uddin et al., 2013] proposed that young children with ASD may have OC of brain network, while under-connectivity can be observed more in adolescents and adults with ASD. [Padmanabhan et al., 2013] investigated the development of intrinsic functional connectivity with rs-fMRI measurements from $8 - 36$ years old subjects. Subjects with ASD exhibited increased connectivity in parietal cortex although decreased connectivity in prefronal cortex when subjects were controlled for age. Moreover, abnormal age-

related connectivity in the anterior part of cerebellum and posterior temporal regions can be observed in ASD subjects. [Nomi and Uddin, 2015] used a data driven approach to investigate within and between network functional connectivity in three age-cohorts of children, adolescents and adults. Children with ASD showed hyper-connectivity within brain networks and decreased between-network connectivity. Adolescents with ASD showed no difference in within network connectivity but showed decreased between-network connectivity compared to control. Interestingly, adults with ASD showed no difference in within or between-network compared to controls. [Lee et al., 2017] explored age-related changes in functional connectivity in the language network of three age groups. Broca's area in both children and adolescents with ASD showed decreased functional connectivity. Adults with ASD showed decreased degree centrality in Wernicke's area. An increased degree centrality was observed in left inferior parietal lobule and left middle temporal gyrus for every age group with ASD compared to control.

We still do not have a complete understanding of what biological phenomena may cause these differences. These disparities may be attributed from heterogeneity in conditions, small sample size and not considering difference among subjects and variation among sites, the design of resting-state scan, the pre-processing and methodology of analysis. Also gender representation in the literature is extremely uneven, which is a hindrance in understanding ASD in females.

2.6 Autism Brain Image Data Exchange

Autism Brain Image Data Exchange (ABIDE) is an alliance among medical centers across the USA and Europe and is the largest repository of functional and structural brain imaging data for autism. It is a part of the 1000 Functional Connectomes Project/ International Neuroimag-

ing Datasharing Initiative (INDI). The data sets are anonymous, with no protected health information according to Health Insurance Portability and Accountability (HIPAA) guidelines and the 1000 Functional Connectomes Project / INDI protocols. The ABIDE data set comprises of the two large-scale datasets: ABIDE I and ABIDE II. The ABIDE I data set includes 17 international sites. It contains structural and resting state fMRI data and also extensive phenotypic information from 539 individuals suffering from ASD and 573 controls. The data consists of approximately 90% male subjects. The ABIDE data were preprocessed by Connectome Computation System (CCS), the Configurable Pipeline for the Analysis of Connectomes (CPAC), the Data Processing Assistant for Resting-State fMRI (DPARSF) and the NeuroImaging Analysis Kit. The ABIDE data set has individuals with age, sex and IQ matched ASD and control group. The phenotypic information in this data set includes age at scan, sex, IQ and diagnostic information.

We used fMRI measurements from 7 sites; California Institute of Technology, New York University Langone Medical center, Olin institute of living at Hartford Hospital, University of Pittsburgh, San Diego State University, Social Brain Lab and Stanford University. The age at scan of subjects varies from 6 years minimum to 64 maximum. Full IQ (FIQ) was calculated by averaging performance and verbal IQ score for each individual. The FIQ has a sample mean of 111.06 and standard deviation of 14.43.

The Autism Diagnostic Interview-Revised (ADI-R) ([Lord et al., 1997], [Rutter et al., 2003]) is a comprehensive standardized clinical tool for diagnosis of autism in children and adults with mental age of at least 2 years. The interview addresses three areas: communication

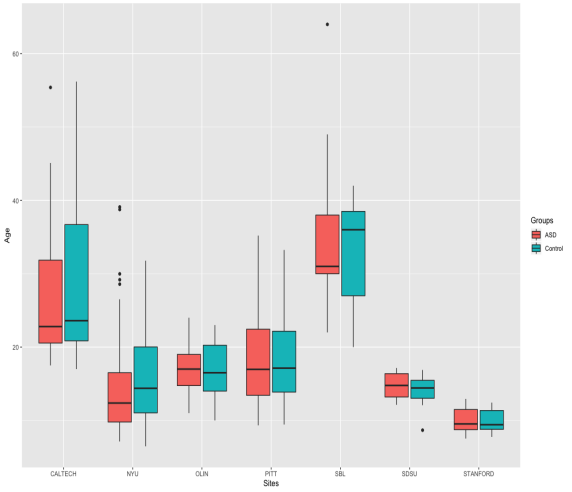
and language, social interaction and restrictive, repetitive and stereotyped behavior. It also considers abnormality of development observed by caregiver at or before 36 months. The interview consists of 93 questions and each question has a score which depends of severity of the condition. A total score is calculated for each of those areas. Most of the sites in ABIDE database provides ADI-R scores for ASD subjects.

Table I along with Figure 4 and Figure 5 represents participants demographics and characteristics for all sites. 84 regions of Broadmann areas of brain was applied to explore differentional cross-correlated functional connectivity between subjects with ASD and controls. Table II lists description for each region. The regional-temporal functional connectivity model considers 111 regions. Table III lists description for each region.

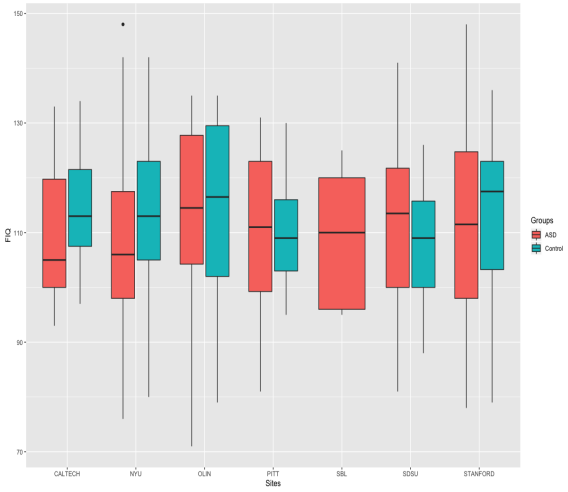
TABLE I: PARTICIPANTS DEMOGRAPHICS FOR ALL SITES

ASD							
	CALTECH	NYU	OLIN	PITT	SBL	SDSU	STANFORD
	(N = 19)	(N = 79)	(N = 20)	(N = 30)	(N = 15)	(N = 14)	(N = 20)
AGE	27.44(10.30)	14.52(6.97)	16.70(3.42)	18.93 (7.20)	35.00(10.43)	14.72(1.76)	9.96(1.59)
FIQ	108.17 (12.59)[a]	107.91(16.62)	113(17.87)[b]	109.97(14.38)	109.2 (13.63)[c]	111.36 (18.04)	112.55(17.80)
SEX							
Male	15 (79)	68 (86)	17 (85)	26 (87)	15 (100)	13 (93)	16 (80)
Female	4 (21)	11 (14)	3 (15)	4 (13)	0 (0)	1 (7)	4 (20)
Control							
	CALTECH	NYU	OLIN	PITT	SBL	SDSU	STANFORD
	(N = 19)	(N = 105)	(N = 16)	(N = 27)	(N = 15)	(N = 22)	(N = 20)
AGE	28.87(11.21)	15.81(6.25)	16.94(3.68)	18.88(6.64)	33.73(6.61)	14.22(1.90)	9.95 (1.60)
FIQ	114.21(9.65)	113.15(13.12)	114.94(16.54)	110.15(9.40)	–	108.09(10.50)	112.10 (15.37)
SEX							
Male	15 (79)	79 (75)	14 (88)	23 (85)	15 (100)	16 (73)	16 (80)
Female	4 (21)	26 (25)	2 (12)	4 (15)	0 (0)	6 (27)	4 (20)

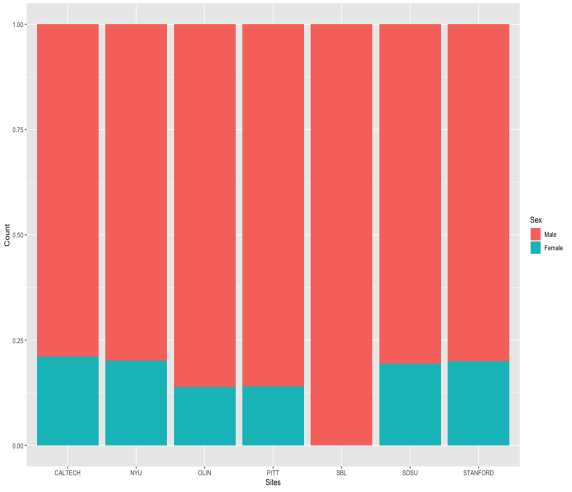
[a] Score missing for 1 participant, [b] Score missing for 2 participants, [c] Score missing for 10 participants



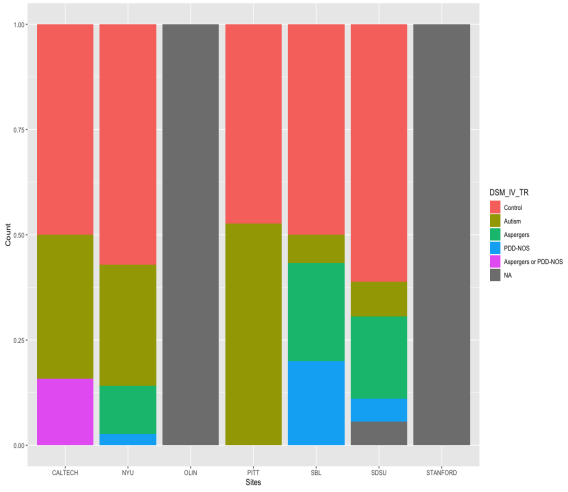
(a) Distribution of age at scan (in years) for individuals with ASD and Control per site



(b) Distribution of FIQ for individuals with ASD and Control per site

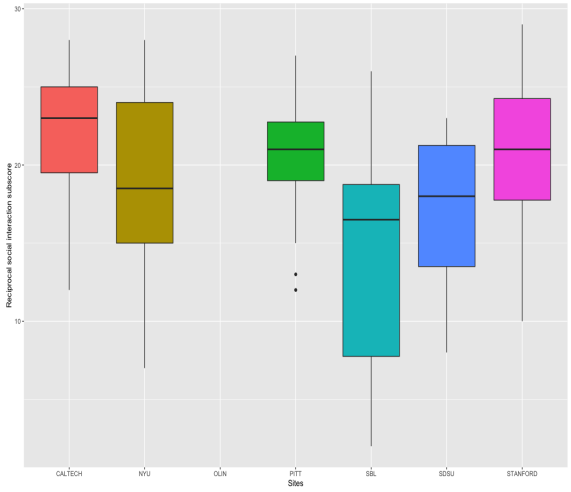


(c) Percentage of males and females for sites

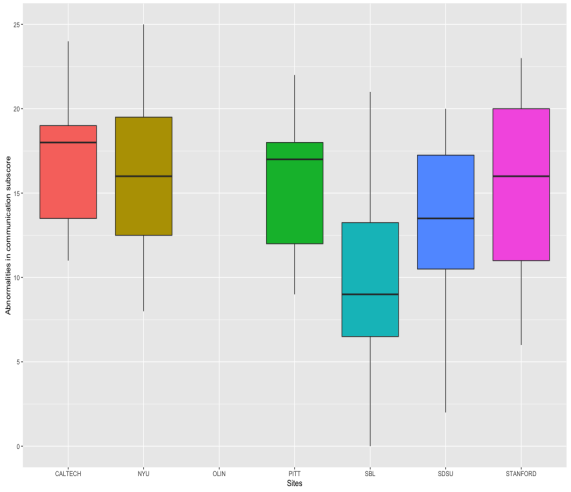


(d) Percentage of specific ASD diagnostic categories for all sites

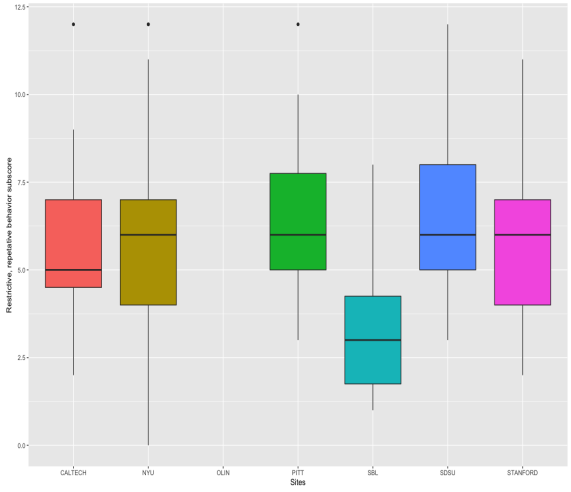
Figure 4: ABIDE participant characteristic



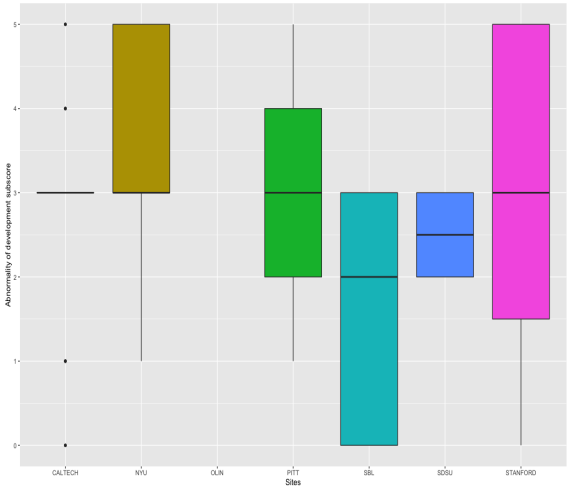
(a) The distribution of reciprocal social interaction subscore for all sites



(b) The distribution of abnormalities in Communication subscore for all sites



(c) The distribution of restrictive, repetitive and stereotyped behavior subscore for all sites



(d) The distribution of abnormality of development evident at or before 36 months subscore for all sites

Figure 5: Autism Diagnostic Interview-Revised (ADI-R) scores for ASD individuals for all sites

TABLE II: REGIONS OF INTEREST IN CROSS-CORRELATED FUNCTIONAL CONNECTIVITY MODEL

Region of Interest	
BA.1 (L & R). Primary Somatosensory Cortex	BA.33 (L & R). Anterior Cingulate Cortex
BA.10 (L & R). Anterior Prefrontal Cortex	BA.34 (L & R). Anterior Entorhinal Cortex
BA.11 (L & R). Orbitofrontal Cortex	BA.35 (L & R). Perirhinal cortex
BA.13 (L & R). Insular Cortex	BA.36 (L & R). Parahippocampal cortex
BA.17 (L & R). Primary Visual Cortex	BA.37 (L & R). Fusiform gyrus
BA.18 (L & R). Secondary Visual Cortex	BA.38 (L & R). Temporopolar Area
BA.19 (L & R). Associative Visual Cortex	BA.39 (L & R). Angular gyrus
BA.2 (L & R). Primary Somatosensory Cortex	BA.4 (L & R). Primary Motor Cortex
BA.20 (L & R). Inferior Temporal Gyrus	BA.40 (L & R). Supramarginal Gyrus
BA.21 (L & R). Middle Temporal Gyrus	BA.41 (L & R). Primary Auditory Cortex
BA.22 (L & R). Superior Temporal Gyrus	BA.42 (L & R). Primary Auditory Cortex
BA.23 (L & R). Ventral Posterior Cingulate Cortex	BA.43 (L & R). Subcentral Area
BA.24 (L & R). Ventral Anterior Cingulate Cortex	BA.44 (L & R). IFC pars opercularis
BA.25 (L & R). Subgenual cortex	BA.45 (L & R). IFC pars triangularis
BA.27 (L & R). Piriform Cortex	BA.46 (L & R). Dorsolateral Prefrontal Cortex
BA.28 (L & R). Posterior Entorhinal Cortex	BA.47 (L & R). Inferior Prefrontal Gyrus
BA.29 (L & R). Retrosplenial Cingulate Cortex	BA.5 (L & R). Somatosensory Association Cortex
BA.3 (L & R). Primary Somatosensory Cortex	BA.6 (L & R). Premotor Cortex
BA.30 (L & R). Cingulate Cortex	BA.7 (L & R). Somatosensory Association Cortex
BA.31 (L & R). Dorsal Posterior Cingulate Cortex	BA.8 (L & R). Dorsal Frontal Cortex
BA.32 (L & R). Dorsal anterior Cingulate Cortex	BA.9 (L & R). Dorsolateral Prefrontal Cortex

TABLE III: REGION OF INTEREST IN REGIONAL-TEMPORAL FUNCTIONAL CONNECTIVITY MODEL

Region of Interest	
(L & R). Thalamus	(L & R). Angular Gyrus
(L & R). Caudate	(L & R). Lateral Occipital Cortex; superior division
(L & R). Putamen	(L & R). Lateral Occipital Cortex; inferior division
(L & R). Pallidum	(L & R). Intracalcarine Cortex
(L & R). Hippocampus	(L & R). Frontal Medial Cortex
(L & R). Amygdala	(L & R). Juxtapositional Lobule Cortex
(L & R). Accumbens	(L & R). Subcallosal Cortex
(L & R). Frontal Pole	(L & R). Paracingulate Gyrus
(L & R). Insular Cortex	(L & R). Cingulate Gyrus; anterior division
(L & R). Superior Frontal Gyrus	(L & R). Cingulate Gyrus; posterior division
(L & R). Middle Frontal Gyrus	(L & R). Precuneous Cortex
(L & R). Inferior Frontal Gyrus; pars triangularis	(L & R). Cuneal Cortex
(L & R). Inferior Frontal Gyrus; pars opercularis	(L & R). Frontal Orbital Cortex
(L & R). Precentral Gyrus	(L & R). Parahippocampal Gyrus; anterior division
(L & R). Temporal Pole	(L & R). Parahippocampal Gyrus; posterior division
(L & R). Superior Temporal Gyrus; anterior division	(L & R). Lingual Gyrus
(L & R). Superior Temporal Gyrus; posterior division	(L & R). Temporal Fusiform Cortex; anterior division

TABLE III –REGION OF INTEREST IN REGIONAL-TEMPORAL FUNCTIONAL CONNECTIVITY MODEL(Continued)

Region of Interest	
(L & R). Middle Temporal Gyrus; anterior division	(L & R). Temporal Fusiform Cortex; posterior division
(L & R). Middle Temporal Gyrus; posterior division	(L & R). Temporal Occipital Fusiform Cortex
(L & R). Middle Temporal Gyrus; temporooccipital part	(L & R). Occipital Fusiform Gyrus
(L & R). Inferior Temporal Gyrus; anterior division	(L & R). Frontal Operculum Cortex
(L & R). Inferior Temporal Gyrus; posterior division	(L & R). Central Opercular Cortex
(L & R). Inferior Temporal Gyrus; temporooccipital part	(L & R). Parietal Operculum Cortex
(L & R). Postcentral Gyrus	(L & R). Planum Polare
(L & R). Superior Parietal Lobule	(L & R). Heschl's Gyrus(includes H1 and H2)
(L & R). Supramarginal Gyrus; anterior division	(L & R). Planum Temporale
(L & R). Supramarginal Gyrus; posterior division	(L & R). Supracalcarine Cortex
(L & R). Angular Gyrus	(L & R). Occipital Pole

CHAPTER 3

CROSS-CORRELATED FUNCTIONAL CONNECTIVITY MODEL

3.1 Introduction

In this chapter, we develop Bayesian models and inference to explore differential functional connectivity between subjects with Autism Spectrum Disorder (ASD) and controls using resting state fMRI (rs-fMRI) data, summarized over time by connectivity metrics, from Autism Brain Image Data Exchange (ABIDE) database.

3.2 Model Specification

The ABIDE pre-processed database provides rs-fMRI measurements at each time point from each ROI for every individual. Connectivity metrics are calculated to reduce dimension of data, which can help to decrease computational complexity [Turk-Browne, 2013]. These connectivity metrics can be obtained by taking cross-correlation between two ROI at lag μ [Cao et al., 1999]. The cross-correlation can be expressed as follows

$$Corr_{x,y}(\mu) = \frac{cov_{x,y}(\mu)}{\sqrt{var(x)var(y)}}$$

[Sun et al., 2004] proposed using cross spectral coherence, which measures the coherence between the time series sequence in two ROI.

$$Coh_{x,y}(\lambda) = \frac{|F_{x,y}(\lambda)|^2}{F_{x,x}(\lambda)F_{y,y}(\lambda)}$$

where $F_{x,y}(\lambda)$ is the cross spectrum, which is defined by the Fourier transformation as following

$$F_{x,y}(\lambda) = \sum_u cov_{x,y}(u) \times \exp^{\{-j\lambda u\}}$$

and the power spectrum $F_{x,x}(\lambda)$ and $F_{y,y}(\lambda)$ are defined as

$$F_{x,x}(\lambda) = \sum_u cov_{x,x}(u) \times \exp\{-j\lambda u\}$$

$$F_{y,y}(\lambda) = \sum_u cov_{y,y}(u) \times \exp\{-j\lambda u\}$$

The rs-fMRI network for each individuals with R many ROI was represented by $R \times R$ symmetric matrix. The upper triangular elements of the connectivity matrix are considered, which implies that for each subject we have $R \times (R-1)/2$ many functional connectivity measurements.

3.2.1 Model 1: Parametric Model

We use $Y_{i,rs}$ to denote the functional connectivity metric measurement between ROI r and s (where, $1 \leq r < s \leq R$) for subject i, where $i = 1, \cdots, n$. In our first model, we model these functional connectivity measurement as

$$
\begin{aligned}
Y_{i,rs} &\sim f(\cdot|\eta_{i,rs}) \\
g(\eta_{i,rs}) &= \beta_{1,rs} + \beta_{2,rs}X_i + \gamma_i + \delta_c
\end{aligned}
\tag{3.1}
$$

Here $f(\cdot)$ denotes a probability distribution, $g(\cdot)$ is a link function, $X_i = 0$ if subject i is control while $X_i = 1$ if the subject is ASD, γ_i is a subject-level random effect term and δ_c is a center-level random effect. Additionally, $\beta_{1,rs}$ is an ROI pair (r, s) specific intercept term to reflect different levels of connectivity among different brain regions, and $\beta_{2,rs}$ models the ROI pair (r, s) specific differential connectivity between normal and ASD subjects. The generalized linear model for the control group is

$$
g(\eta_{i,rs}) = \beta_{1,rs} + \gamma_i + \delta_c \tag{3.2}
$$

and for the ASD group is

$$
g(\eta_{i,rs}) = \beta_{1,rs} + \beta_{2,rs} + \gamma_i + \delta_c \tag{3.3}
$$

The parameters $\beta_{2,rs}$ model the differential connectivities of interest and we model them by a spike-and-slab prior ([Lempers, 1971], [Mitchell and Beauchamp, 1988])

$$
\beta_{2,rs} \overset{i.i.d}{\sim} p\,\delta_0(.) + (1-p)\,G(\cdot), \qquad 1 \le r < s \le R. \tag{3.4}
$$

Here $\delta_0(\cdot)$ denotes the degenerate distribution at 0 and represents the null case of no differential connectivity between region pair (r, s) or $\beta_{2,rs} = 0$. The second component of the mixture provides the distribution of $\beta_{2,rs}$ in the nun-null case when regions r and s have differential connectivity between control and ASD groups.

We assume that the mixture proportion p has Uniform $(0, 1)$ prior. The specification of our first Bayesian model, hereafter refered to as Model 1 or the parametric model, is completed by assuming a parametric functional form for G as $G = N(0, \tau^2)$. Spike and slab variable selection is extensively discussed in Section 3.3.

3.2.2 Model 2: Dirichlet Process Mixture (DPM) Model

We propose an alternative second Bayesian model where we utilize Bayesian non-parametric approach to provide flexibility in modeling the distribution $G(\cdot)$ of differential connectivities. We model $G(\cdot)$ as a random probability distribution with an associated Dirichlet Process (DP) probability measure [Ferguson, 1973], which defines a probability measure on the space of probability distributions and has an extensive literature that is ever-expanding. The support of DP is dense in the space of all probability distributions with support same as the base measure G_0. One important property of the Dirichlet Process is that it can induce ties and clustering. Dirichlet Process Mixture (DPM) model not only provide more flexibility but it can also allow clustering among the differentially connected regions. The DPM model can be expressed hierarchically as

$$
\begin{aligned}
Y_{i,rs} &\sim f(\cdot|\eta_{i,rs}) \\
g(\eta_{i,rs}) &= \beta_{1,rs} + \beta_{2,rs} X_i + \gamma_i + \delta_c \\
\beta_{2,rs} &\overset{i.i.d.}{\sim} p\,\delta_0(.) + (1-p)\,G(\cdot) \\
G(\cdot) &\sim DP(\kappa, G_0)
\end{aligned}
\tag{3.5}
$$

DPM models and Markov chain sampling methods for sampling from the posterior distribution of DPM models are reviewed in Section 3.4.

3.2.3 Model 3: Neighborhood Model on ROI Pairs

We introduce a third Bayesian model to capture dependence patterns among ROIs which are often indirectly connected through a small number of highly connected nodes [Achard et al., 2006]. We propose a conditionally defined model of dependence among local neighborhoods as is commonly used in spatial analysis. This first requires a neighborhood structure on the index set of region pairs $\{rs : 1 \leq r < s \leq R\}$. We define a neighborhood of rs as

$$\mathcal{N}(rs) = \{r's, r' \neq r\} \cup \{rs', s' \neq s\},$$

that is, links which are connected to rs at one of the vertices (either at r or at s). Let $\mathcal{N}$ denote the collection of these neighborhoods for all pairs rs.

The neighborhood model on ROI pairs is motivated by the postulate that differential connectivity at rs may be associated with connectivities in other links which are attached with rs

at one of its vertices. We propose a conditional auto-regressive (CAR) model with neighborhood structure $\mathcal{N}$ to incorporate potential dependencies in functional connectivity of regions rs and its neighbors.

$$\begin{aligned} Y_{i,rs} &\sim f(\cdot|\eta_{i,rs}) \\ g(\eta_{i,rs}) &= \beta_{1,rs} + \beta_{2,rs}X_i + \gamma_i + b_{rs} + \delta_c \\ \{b_{rs} : 1 \leq r < s \leq R\} &\sim \text{CAR}(\mathcal{N}) \end{aligned} \tag{3.6}$$

The CAR model is described in detail in Section 3.5.1.

3.3 Spike and Slab Variable Selection

All three models proposed above utilize the spike-and-slab prior described in (Equation 3.4). The spike and slab prior was introduced by [Lempers, 1971] and [Mitchell and Beauchamp, 1988]. The expression "spike and slab" is assigned to prior for $\beta_{2,rs}$ where $\beta_{2,rs}$ has a two-point mixture distribution made up with a degenerate distribution at zero (the spike) and a normal distribution (the slab). This prior helps to reduce $\beta_{2,rs}$ coefficients to zero that are truly zero by making posterior mean values small. The model can be specified by the following prior hierarchy

$$
\begin{aligned}
Y_{i,rs}|\cdot &\sim N(\mu_{i,rs}, \sigma^2) \\
\mu_{i,rs} &= \beta_{1,rs} + \beta_{2,rs} X_i + \gamma_i + \delta_c \\
\beta_{1,rs} &\sim N(\mu_{\beta_1}, \sigma^2_{\beta_1}) \\
\beta_{2,rs} &\overset{i.i.d}{\sim} p\,\delta_0(.) + (1-p)\,N(0, \tau^2) \\
\sigma^2 &\sim \pi(d\sigma^2)
\end{aligned}
\tag{3.7}
$$

where a uniform $(0, 1)$ prior is often assumed on p . $\pi(\cdot)$ is the prior distribution for σ^2. Small hypervariances can reduce the value of coefficient to zero. On the other hand, large values will inflate coefficients.

3.4 Non-parametric Bayesian Models

We begin this section with a review of Dirichlet Process(DP) and Dirichlet Process Mixture (DPM) models which are primary components of our proposed model 2. Dirichlet Process defines a probability measure on the space of probability measures.

Definition 3.4.1 ([Ferguson, 1973]). *Let $\alpha > 0$, $G_0(.)$ be a specified probability measure and let $G(\cdot)$ (a random probability measure) be a stochastic process. G is a Dirichlet process, or $G \sim \mathcal{DP}(\alpha G_0(\cdot))$, if for every finite measurable partition $\{B_1, ..., B_m\}$, the random vector $(G(B_1), ..., G(B_m))$ has a Dirichlet distribution with parameter $(\alpha G_0(B_1), ..., \alpha G_0(B_m))$.*

The random distribution G is almost surely a discrete distribution. Dirichlet process has the following Bayesian conjugate property.

Property 1. If $G \in DP(\alpha G_0(\cdot))$ and if $Y_1, \ldots, Y_n \overset{iid}{\sim} G$, then the posterior of $G|\, Y_1, \ldots, Y_n$ is $DP(\alpha\ G_0 + \sum_1^n \delta_{Y_i})$, where δ_Y is the Dirac measure.

3.4.1 Dirichlet Process Mixture

Dirichlet process Mixture (DPM) models are reviewed in [Ferguson, 1973], [Antoniak, 1974], [Escobar and West, 1995] and many other literature. The data $(Y_1, \ldots, Y_n)$ is regarded as part of an indefinite exchangeable sequence which implies that they are independently drawn from an unknown distribution G by di Finneti's theorem. The DPM model can be hierarchically defined as

$$\begin{aligned} Y_i|\theta_i &\sim F(\theta_i) \\ \theta_i &\overset{i.i.d}{\sim} G \\ G &\sim DP(\alpha, G_0) \end{aligned} \tag{3.8}$$

The distributions of F and G_0 can depend on additional hyperparameters. According to [Ferguson, 1983], these models are considered as countably infinite mixtures. The conditional distribution of θ_i follows the following Polya urn representation

$$\theta_i|\theta_1, \cdots, \theta_{i-1} \sim \frac{\alpha}{\alpha + i - 1} G_0(\cdot) + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{\theta_j}(\cdot) \tag{3.9}$$

where $\delta_{\theta_j}(\cdot)$ is the degenerate distribution at the point θ_j.

An alternate way to represent is based on "latent class", which is indicated by c_i and it is related with observation y_i. Here the labeling of c_i has no significance and the parameters ϕ_c

determine the distribution of the observations from a particular label c. The model [Neal, 2000] can be defines as follows

$$\begin{aligned} Y_i|c_i, \phi &\sim F(\phi_{c_i}) \\ c_i|\mathbf{p} &\sim Discrete(p_1, \ldots, p_K) \\ \phi_c &\sim G_0 \\ \mathbf{p} &\sim Dirichlet(\alpha/K, \ldots, \alpha/K) \end{aligned} \tag{3.10}$$

where the mixing proportion for the classes are defined by $\mathbf{p} = (p_1, \ldots, p_K)$. Here $\mathbf{p}$ is a symmetric DP, with concentration parameter α/K, which approaches to zero as $K \to \infty$.

3.4.1.1 Pólya Urn Scheme

Posterior computation under DPM initially seemed problematic as the mixing measure G is characterized by infinitely many parameters. The posterior of G given $(y_1, \ldots, y_n)$ also has a convoluted form. A way to deal with this issue is to marginalize out G to obtain an induced prior distribution on parameters $\theta_1, \ldots, \theta_n$. We obtain the Pólya urn predictive rule by marginalizing out G. In the Pólya urn scheme, it is assumed that there is a opaque urn which contained colored balls and balls are drawn randomly from it. When a ball is drawn, we inspect its color and return to the urn and also insert an another ball of the same color. [Blackwell et al., 1973] used a similar scheme for construction of Dirichlet Process. The color of the ball placed in the urn is $\theta_{n+1} \sim G_0$. There are already previously seen balls in the urn. In the starting, the urn is empty and we pick a color drawn from G_0 i.e. $\theta_1 \sim G_0$ and add a ball of that particular color in the urn. Then $(n+1)^{th}$ step, we either pick a new color (i.e. draw $\theta_{n+1} \sim G_0$) with

probability $\frac{\alpha}{\alpha+n}$ and add a ball with that color into the urn or with probability $\frac{n}{\alpha+n}$ pick out a ball randomly from the urn, pick a ball of same color and drop both balls back into the urn. This scheme produces a sequence of θ_i with conditional probabilities

$$\theta_{n+1}|\theta_1,\ldots,\theta_n \sim \frac{\alpha}{\alpha+n}G_0(\theta_{n+1}) + \frac{\sum_{i=1}^{n}\delta_{\theta_{n+1}=\theta_i}}{\alpha+n} \tag{3.11}$$

The simple form for the conditional distribution in Equation 3.9 generate an useful idea for posterior computation. As the values of draws θ_k are repeated, let $\theta_1^*,\ldots,\theta_m^*$ be the unique values among $\theta_1,\ldots,\theta_n$, and n_k be the number of repeats of θ_k^*. Then we can write Equation 3.11 as follows

$$\theta_{n+1}|\theta_1,\ldots,\theta_n \sim \frac{\alpha}{\alpha+n}G_0(\theta_{n+1}) + \frac{\sum_{k=1}^{m}n_k\delta_{\theta_k^*}}{\alpha+n} \tag{3.12}$$

3.4.1.2 Stick-breaking Process

[Sethuraman, 1994] developed a constructive way of forming G known as "stick breaking". This construction is more straightforward and general than existing DP model. The Stick-breaking representation allows us to induce $G \sim DP(\alpha, G_0)$ by letting

$$
\begin{aligned}
G &= \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*} \\
V_k &\sim Beta(1, \alpha) \\
\pi_1 &= V_1 \\
\pi_k &= V_k \prod_{l=1}^{k-1} (1 - V_l), \quad k = 2, 3, \ldots \\
\theta_k^* &\overset{i.i.d}{\sim} G_0
\end{aligned}
\tag{3.13}
$$

where δ_{θ^*} denotes a degenerate distribution. The atoms θ_k^* are generated independently from base distribution G_0. Here π_k is the probability mass at atom θ_k^* and these probability masses guarantees that the weights sum to 1.

In this representation, there is a stick of unit length, which represent the total probability to be allocated to all the atoms. The stick is broken off into a random piece of length V_1, where V_1 is generated from $Beta(1, \alpha)$ distribution and allocate $\pi_1 = V_1$ to the randomly generated first atom θ_1^* . Now generate $V_2 \sim Beta(1, \alpha)$ of the remaining $1 - V_1$ stick and the probability $\pi_2 = V_2(1 - V_1)$ allocated the second atom θ_2^*. We repeat the same process to obtain $\pi_3, \pi_4, \ldots$ and we get an explicit construction of G. Here, $E(V_h) = \frac{1}{1+\alpha}$, values close to zero lead to high weight on the first couple atoms and small probabilities are assigned to the remaining atoms.

3.4.2 Posterior Simulations for Dirichlet Process Mixture Model

We review inference algorithms from [Neal, 2000] for DPM models in this section. There are various simulation-based methods to get inference on posterior distribution for conjugate

and non-conjugate priors. Gibbs sampling can be applied for conjugate prior models. It cannot be applied on non-conjugate priors because of complexity in numerical integration. A Monte Carlo based approaches to handle the non-conjugate priors were developed by [MacEachern and Müller, 1998] and [West and Escobar, 1993]. Here we have discussed simulation based methods for conjugate priors.

3.4.2.1 Conjugate Priors

The first algorithm was introduced by [Escobar, 1994] and by [Escobar and West, 1995]. This is the most direct approach to sample θ_i from the distribution in Equation 3.9.

$$\theta_i|\theta_{-i}, y_i \sim \sum_{j \neq i} q_{i,j}\delta(\theta_j) + r_i H_i \tag{3.14}$$

Here, H_i is the posterior for θ based on the prior G_0 and the likelihood $F(y_i, \theta)$. $q_{i,j}$ and r_i are defined as the following

$$\begin{aligned} q_{i,j} &= bF(y_i, \theta_j) \\ r_i &= b\alpha \int F(y_i, \theta)\, dG_0(\theta), \end{aligned} \tag{3.15}$$

where b is a normalizing constant, such that $\sum_{j \neq i} q_{i,j} + r_i = 1$. The algorithm can be summarized as follows

Algorithm 3.1 Let the Markov chain consists of $\theta = (\theta_1, \ldots, \theta_n)$

- Initialize θ.
- Draw a new value from $\theta_i|\theta_{-i}$ from Equation 3.14.

This algorithm has a slow convergence rate and sampling may be inefficient due to that. It can occur because sometime a class of observations are associated with same θ with high probability. As this algorithm can only change one at a time, a change to the θ values for observations can occur rarely in such a class. This kind of change requires low probability transition.

This can be avoided by applying Gibbs sampling to the model Equation 3.10. This method is used by [Bush and MacEachern, 1996]. In every Gibbs sampling, a new value for each c_i is selected from the conditional distribution $c_i|y_i,\ \phi_c,\ c_j$. Then a new value for each ϕ_c is selected from its conditional distribution given the y_i for which $c_i = c$. The conditional probabilities for c_i can be computed

$$P(c_i = c|c_{-i}, y_i, \phi) = b\frac{n_{i,c} + \alpha/K}{n - 1 + \alpha}F(y_i, \phi_c), \tag{3.16}$$

where $n_{i,c}$ is the number of c_j for $j \neq i$ that are equal to c and b is the normalizing constant. When K tends to infinity, Gibbs sampling is done in certain ϕ_c that are associated with certain observation. Hence, the posterior distribution over c_i can be computed

$$P(c_i = c|c_{-i}, y_i, \phi) = \begin{cases} b\frac{n_{i,c}+\alpha}{n-1+\alpha}F(y_i, \phi_c), & \text{if} \quad c = c_j \quad \text{for some} \quad j \neq i \\ b\frac{\alpha}{n-1+\alpha}\int F(y_i, \phi_c)\, dG_0(\phi), & \text{if} \quad c \neq c_j \quad \text{for all} \quad j \neq i \end{cases} \tag{3.17}$$

where $n_{i,c}$ is the number of c_j for $j \neq i$ that are equal to c, b is the appropriate normalizing constant and ϕ is the set of ϕ_c. The second sampling method as follows:

Algorithm 3.2 Let the Markov chain consists of $c = (c_1, \ldots, c_n)$ and $\phi = (\phi_c : c \in \{c_1, \ldots, c_n\})$

- Initialize $(c_1, \ldots, c_n)$ and ϕ.
- For $i = 1, \ldots, n$: Remove ϕ_{c_i} from the state if current value of c_i is not related to observation (i.e $n_{-i,c_i} = 0$). Draw a new value for c_i as defined by Equation 3.17. If the new c_i is not related with any other observation then draw a value of ϕ_{c_i} from H_i.
- For all $c \in \{c_1, \ldots, c_n\}$): Draw a new value from $\phi_c | y_i$, $\forall i$ for which $c_i = c$.

In conjugate context, we can use the Markov chain consists only of c_i by integrating over ϕ_c and eliminating it from the algorithm. This algorithm is discussed in [MacEachern, 1994] and [Neal, 1992]. The posterior probabilities is as follows

$$P(c_i = c | c_{-i}, y_i, \phi) = \begin{cases} b \frac{n_{i,c}+\alpha}{n-1+\alpha} \int F(y_i, \phi)\, dH_{-i,c}(\phi), & \text{if } c = c_j \text{ for some } j \neq i \\ b \frac{\alpha}{n-1+\alpha} \int F(y_i, \phi)\, dG_0(\phi), & \text{if } c \neq c_j \text{ for all } j \neq i \end{cases} \tag{3.18}$$

where $dH_{-i,c}(\phi)$ is the posterior distribution of ϕ based on the prior G_0 and for all y_j for which $c_j = c$. The Gibbs sampling method is the following

Algorithm 3.3 Let the Markov chain consists of $c = (c_1, \ldots, c_n)$.

- Initialize $(c_1, \ldots, c_n)$.
- For $i = 1, \ldots, n$: Draw a new value from $c_i | c_{-i}, y_i$ by Equation 3.18.

This algorithm provides an efficient sampler as provides a precise estimate of the likelihood.

3.4.3 Blocked Gibbs Sampling for Stick-breaking Priors

[Ishwaran and James, 2001] introduced blocked Gibbs sampling method for fitting Bayesian non-parametric hierarchical model based on stick-breaking priors. This approach allows direct sampling of the non-parametric posterior, which leads to many computational and inferential benefits. The observed data $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and the sequence of unobserved random elements $c_1, \cdots, c_n$ with conditional distribution G sampled from stick-breaking prior. The hierarchical model is

$$\begin{aligned} (Y_i | c_i, \theta) &\overset{ind}{\sim} \pi(Y_i | c_i, \theta), \qquad i = 1, \ldots, n, \\ (c_i | G) &\overset{iid}{\sim} G \\ \theta &\sim \pi(\theta) \\ G &\sim DP(\alpha, G_0) \end{aligned} \tag{3.19}$$

We are assuming $c_i = Z_{K_i}$ where the K_i is the classification variable to match Z_K associated with each c_i. With finite dimensional Stick-breaking prior, the Bayesian semiparametric model Equation 3.19 can be rewritten as

$$
\begin{aligned}
(Y_i|Z,K,\theta) &\overset{ind}{\sim} \pi(Y_i|Z_{K_i},\theta), \qquad i=1,\ldots,n,\\
(K_i|p) &\overset{i.i.d.}{\sim} \sum_{k=1}^{N} p_k \delta_k(\cdot),\\
(p,Z) &\sim \pi(p)\times H^N(Z),\\
\theta &\sim \pi(\theta)
\end{aligned}
\tag{3.20}
$$

where $K=(K_1,\cdots,K_n)$, $Z=(Z_1,\cdots,Z_N)$, $p=(p_1,\cdots,p_N)$ and $Z_k \overset{iid}{\sim} H$. Here $\theta \in \mathcal{R}^d$ represents a finite dimensional parameter. The following algorithm will allow to draw values from the conditional distributions of the blocked variables.

Algorithm 3.4 Let m unique values of $\mathbf{K}$ is denoted by $\{K_1^*,\cdots,K_m^*\}$.

- Conditional for $\mathbf{Z}$: For each $k \in \mathbf{K} - K_1^*,\ldots,K_m^*$ simulate $Z_k \sim H$ and also draw $(Z_{K_j^*}|\mathbf{K},\theta,\mathbf{Y})$ from the density

$$
f(Z_{K_j^*}|\mathbf{K},\theta,\mathbf{Y}) \propto H(\,dZ_{K_j^*}) \prod_{\{i:K_i=K_j^*\}} f(Y_i|Z_{K_j^*},\theta), j=1,\ldots,m.
$$

- Conditional for $\mathbf{K}$: Draw K_i from

$$
(K_i|\mathbf{Z},\mathbf{p},\theta,\mathbf{Y}) \overset{ind}{\sim} \sum_{k=1}^{N} p_{j,i}\delta_k(.), i=1,\ldots,n.
$$

where $(p_{1,i},\ldots,p_{N,i}) \propto (p_1 f(Y_i|Z_1,\theta),\ldots,p_N f(Y_i|Z_N,\theta))$.

Algorithm 3.4 (Continued)

- Conditional for **p**: As the generalized Dirichlet distribution is conjugate to Multinomial sampling. Then draws are following

$$p_1 = V_1^*, \qquad p_k = (1 - V_1^*)(1 - V_2^*) \dots (1 - V_{k-1}^*)V_k^*, \quad k = 2, \dots, N-1,$$

  where $V_k^* \overset{ind}{\sim} Beta(a_k + M_k, b_k + \sum_{l=k+1}^{N} M_l)$ for $k = 1, \dots, N-1$ and M_k is the number of K_i values that equal k.

- Conditional for θ: θ will be drawn from the density

$$f(\theta|\mathbf{Z}, \mathbf{K}, \mathbf{Y}) \propto \pi(\, d\theta) \prod_{i=1}^{n} f(Y_i|c_i, \theta)$$

3.4.4 Posterior Analysis in Model 2: DPM Model

In section 3.2.2, we propose a DPM model for differential functional connectivity by modeling the connectivity measurement $Y_{i,rs}$ between ROI pairs (r, s) for subject i as

$$\begin{aligned} Y_{i,rs} &\sim f(\cdot|\eta_{i,rs}) \\ g(\eta_{i,rs}) &= \beta_{1,rs} + \beta_{2,rs} X_i + \gamma_i + \delta_c \\ \beta_{2,rs} &\overset{i.i.d.}{\sim} p\, \delta_0(.) + (1-p)\, G(\cdot) \\ G(\cdot) &\sim DP(\alpha, G_0) \\ G_0 &= N(0, \sigma_\theta^2), \quad \sigma_\theta^2 \sim \pi(d\sigma_\theta^2) \end{aligned} \tag{3.21}$$

For posterior computation of this model, we utilize finite truncation of Sethuraman's stick-breaking representation [Sethuraman, 1994] of the random G and its finite truncation to write the model as follows.

$$
\begin{aligned}
G &= \sum_{k=1}^{N} \pi_k \delta_{\theta_k^*} \quad , N \geq 1 \\
V_k &\sim Beta(1, \alpha) \quad , k \geq 2 \\
\pi_1 &= V_1 \\
\pi_k &= V_k \prod_{l=1}^{k-1} (1 - V_l), \quad k = 2, \cdots, N \\
\theta_k^* &\overset{i.i.d}{\sim} N(0, \sigma_\theta^2) \\
\sigma_\theta^2 &\sim \pi(d\sigma_\theta^2)
\end{aligned}
\tag{3.22}
$$

where $V_N = 1$ so that $\sum_{k=1}^{N} \pi_k = 1$. The DP precision parameter α controls the prior on the quantity of clusters. We used $\alpha \sim Gamma(a_\alpha, b_\alpha)$ to keep it updating during the MCMC sampling.

The posterior computation for this model used an algorithm similar to **Algorithm 3.4**. OpenBUGS [Surhone et al., 2010] and R are used to generate MCMC samples for the Bayesian posterior inference.

3.5 Neighborhood Model on ROI Pairs

3.5.1 Conditional Auto Regressive Model

Conditional autoregressive (CAR) models have been used extensively for analysis of spatial data in many areas, such as ecology, demography, epidemiology, economics to model spatially

autocorrelated data based on neighborhood relationships. CAR models are a class of Gaussian Markov random fields [Carlin et al., 2014] and early influential works include [Besag, 1974]. This subsection introduces the CAR model utilizing general spatial notations. The application of the CAR model in modeling differential functional connectivities is described in the next section.

We consider a process $Y(\cdot)$, with finite dimensional joint distribution $(Y(s_1), \ldots, Y(s_r))$ defined on the lattice $D_s \equiv \{s_1, \cdots, s_r\} \in \mathbb{R}^d$. The neighborhood of location s_k is defined as

$$N(s_k) = \{s_j : s_j \text{ is a neighbor of } s_k\}.$$

Note that the neighborhood relationships can differ with respect to the area of interest. The neighborhood $N(s_k)$ of the kth location s_k can be formally defined as

$$[Y(s_k)|Y_{-k}] = [Y(s_k)|\mathbf{Y}(N(s_k))], \qquad k = 1, \cdots, r,$$

where $Y(N(s_k)) \equiv \{Y(u) : u \in N(s_k)\}$.

Lemma 3.5.1. *[Cressie and Wikle, 2015] Let $Y(\cdot)$ and $W(\cdot)$ be two processes. Brook's Lemma relates the joint probability with conditional probability as following*

$$\frac{P(y)}{P(w)} = \prod_{k=1}^{r} \frac{P(y(s_k)|y(s_1), \cdots, y(s_{k-1}), w(s_{k+1}), \cdots, w(s_r))}{P(w(s_k)|y(s_1), \cdots, y(s_{k-1}), w(s_{k+1}), \cdots, w(s_r))}, \qquad y, w \in \Omega \tag{3.23}$$

The neighborhood structure can be represented by $B = \{b_{kj}\}$ such that $b_{kk} = 0$ and $b_{kj} = 0$ for $s_j \notin N(s_k)$, $k = 1, \cdots, r$. CAR model for any exponential family can be expressed as following

$$p(y_k|y_j, j \neq k) \propto \exp\Big\{\psi(\theta_k y_k - \chi(\theta_k))\Big\}, \tag{3.24}$$

where, canonical link $\theta_k = \sum_{j \neq k} b_{kj} y_j$ and ψ is a non-negative dispersion parameter.
For Gaussian (or autonormal) case, we assume that [Carlin et al., 2014]

$$Y_k|(Y_j, j \neq k) \sim N(\sum_{j \in N(s_k)} b_{kj} y_j, \tau_k^2), \quad k = 1. \cdots, r. \tag{3.25}$$

By applying Brook's Lemma we can obtain

$$p(y_1, \cdots, y_r) \propto \exp\{-\frac{1}{2} y' D^{-1}(I - B)y\} \tag{3.26}$$

where $B = \{b_{kj}\}$ and D is a diagonal matrix with $D_{kk} = \tau_k^2$. $Y = (y_1, \cdots, y_r)$ follows a Multivariate Normal distribution with mean 0 and variance matrix $\Sigma_y = (I - B)^{-1} D$ if following two conditions are satisfied

- $D^{-1}(I - B)$ is symmetric, which is equivalent to $\frac{b_{kj}}{\tau_k^2} = \frac{b_{jk}}{\tau_j^2}$ for all k, j.
- $D^{-1}(I - B)$ is positive definite.

In the case where B is not symmetric, b_{kj} is assumed to depend on symmetric proximity matrix $W = \{w_{kj}\}$. The entries w_{kj} in W connect units k and j, where $w_{kj} = 1$ if k and j share same neighborhood and assume that $w_{kk} = 0$. w_{kj}'s may be standardized by $\sum_j w_{kj} = w_{k+}$.

The condition $\frac{b_{kj}}{\tau_k^2} = \frac{b_{jk}}{\tau_j^2}$ satisfied if we assume $b_{kj} = w_{kj}/w_{k+}$ and $\tau_k^2 = \tau^2/w_{k+}$. We can rewrite Equation 3.25 as

$$p(y_k|y_j, j \neq k) \sim N(\sum_j w_{kj}y_j/w_{k+}, \tau^2/w_{k+})$$

and then applying Brook's Lemma the joint distribution can be expressed as

$$p(y_1, \cdots, y_r) \propto \exp\{-\frac{1}{2\tau^2}y'(D_w - W)y\} \tag{3.27}$$

where D_w is a diagonal matrix with $(D_w)_{kk} = w_{k+}$. The distribution in Equation 3.27 is improper as $\Sigma_y = (D_w - W)1 = 0$. Equation 3.27 can be rewritten as following

$$p(y_1, \cdots, y_r) \propto \exp\{-\frac{1}{2\tau^2}\sum_{k\neq j} w_{kj}(y_k - y_j)^2\} \tag{3.28}$$

This joint distribution is improper but has proper full conditionals. Equation 3.28 is referred as an intrinsically autoregressive (IAR) model.

We can get a proper CAR model by redefining $\Sigma_y = D_w - \rho W$. Σ_y^{-1} is non-singular if $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(r)})$, where $\lambda_{(1)} < \lambda_{(2)} < \cdots < \lambda_{(r)}$ are eigen values of ${D_w}^{-1/2}W{D_w}^{-1/2}$. The full conditional can be written as

$$p(y_k|y_j, j \neq k) \sim N(\rho\sum_j w_{kj}y_j/w_{k+}, \tau^2/w_{k+}) \tag{3.29}$$

However, a proper CAR model may have limited spread of spatial pattern. An improper CAR model allows a wide scope for posterior spatial pattern.

3.5.2 Model 3: Neighborhood Model on ROI Pairs

We utilize the CAR model described in the previous section to incorporate dependence between connectivity of the link rs and an appropriately defined neighborhood structure. We define the neighbors of link rs as

$$\mathcal{N}(rs) = \{r's, r' \neq r\} \cup \{rs', s' \neq s\},$$

that is, links which are connected to rs at one of the vertices (either at r or at s). Let $\mathcal{N}$ denote the collection of these neighborhoods for all pairs rs.

The model can be specified by the following prior hierarchy

$$\begin{aligned}
Y_{i,rs}|\cdot &\sim N(\mu_{i,rs}, \sigma^2) \\
\mu_{i,rs} &= \beta_{1,rs} + \beta_{2,rs} X_i + \gamma_i + b_{rs} + \delta_c \\
\beta_{1,rs} &\sim N(\mu_{\beta_1}, \sigma^2_{\beta_1}) \\
\beta_{2,rs} &\overset{i.i.d}{\sim} p\,\delta_0(.) + (1-p)\,N(0, \tau^2) \\
\sigma^2 &\sim \pi(d\sigma^2) \\
\{b_{rs} : 1 \leq r \leq s \leq R\} &\sim \text{CAR Normal}(\mathcal{N}(rs), \sigma_b^2) \\
\sigma_b^2 &\sim \pi(d\sigma_b^2)
\end{aligned} \tag{3.30}$$

The neighborhood structure $\mathcal{N}(rs)$ is associated with weights w_{rs}. These weights are assumed to be symmetric and usually it is set to be equal to 1 [Besag et al., 1991].

3.6 Performance Evaluation

We evaluate the performances of (1) Model 1: Parametric model, (2) Model 2: DPM model and (3) Model 3: Neighborhood model on multiple simulation studies. These simulation studies represent both correct analysis model and incorrect analysis model situations. Each simulation study evaluates performances of the three analysis models over 100 replicated data sets. To keep the Bayesian computational complexity over these replicated data sets manageable, each data set considers 20 ROI's resulting in $\binom{20}{2} = 190$ connectivities of rs pairs. Among these 190 links, 20 are specified to be differentially connected and remaining 170 are non-differentially connected in the data generation models. We considered a sample size of $n_1 = 30$ in the control group and $n_2 = 30$ in the ASD group.

3.6.1 Simulation Study 1

3.6.1.1 Data Generation Model

For this first simulation study, the data generation model is specified as follows

$$\begin{aligned} Y_{i,rs} &\sim N(\mu_{i,rs}, \sigma^2) \\ \mu_{i,rs} &= \beta_{1,rs} + \beta_{2,rs} X_i + \gamma_i \end{aligned}$$

where we generate $\gamma_i \sim \text{Normal}(0, \sigma^2_\gamma)$, with $\sigma^2_\gamma = 0.05^2$ and $\sigma^2 = 0.10^2$. We specify the intercept term $\beta_{1,rs} = 0.15$ for all the 190 links. We specified connected and non-connected

regions in parameter $\beta_{2,rs}$. We considered the first 20 region pairs to be differentially connected for which $\beta_{2,rs} = 0.055$. $\beta_{2,rs}$ for non-differentially connected regions are set to be $= 0$.

$$\beta_{2,rs} = (\underbrace{0.055, \cdots, 0.055}_{20}, \underbrace{0, \cdots, 0}_{170})$$

3.6.1.2 Results from Analysis Models 1, 2, and 3

We utilize the data-generating model specified above to generate 100 replicated data sets, each of size $(n_1 = 30, n_2 = 30)$ for the control and ASD groups respectively. We then analyzed each data set using each of the three analysis models described. Note that analysis model 1 is the correctly specified model here and analysis model 2 (DPM) is, in some sense, a super-model.

The hyperparameters for the model 1 (parametric model) in the Equation 3.7 is specified as following

$$\mu_{\beta_1} \sim \mathrm{N}(0.5, 1), \quad \sigma^2_{\beta_1} \sim \text{Inverse-Gamma}(0.01, 0.01)$$

$$\sigma^2 \sim \text{Inverse-Gamma}(0.001, 0.001), \quad \tau^2 \sim \text{Inverse-Gamma}(0.01, 0.01)$$

$$p \sim \text{Uniform}(0, 1)$$

We have considered the initial numbers of clusters to be 10 in the model 2 (DPM). Hyperparameters in Equation 3.22 are specified as following

$$\alpha \sim \text{Gamma}(0.1, 0.1), \quad \sigma^2_\theta \sim \text{Inverse-Gamma}(0.01, 0.01)$$

Model 3 (neighborhood model) hyperparameters in Equation 3.30 are specified to be

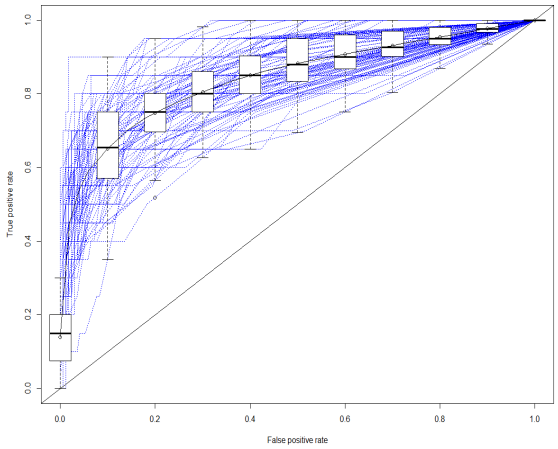
$$\{b_{rs} : 1 \leq r < s \leq 20\} \sim \text{CAR Normal}(\mathcal{N}_{rs}, \sigma_b^2), \quad \sigma_b^2 \sim \text{Inverse-Gamma}(0.01, 0.01).$$

Here weights $\{w_{rs}\}$, which are associated with $\mathcal{N}_{rs}$ are set to be equal to 1. We generated 100 replicated data sets and apply the different analysis model. The following table shows mean and standard error of parameters $\beta_{1,\cdot\cdot} = \frac{1}{190}\sum_{rs=1}^{190}\beta_{1,rs}$, σ^2 and σ_γ^2 obtained from simulation study. We notice in Table IV that the estimated values of $\beta_{1,\cdot\cdot}$, σ^2 and σ_γ^2 are approaching to their respective true values.

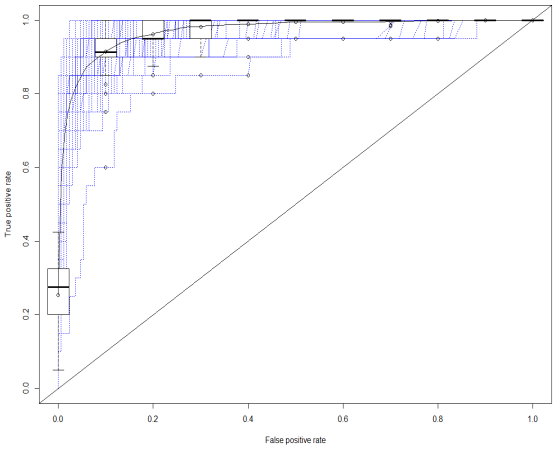
TABLE IV: SIMULATION STUDY 1: AVERAGE VALUES AND STANDARD ERRORS FROM THREE ANALYSIS MODELS

		Analysis Models		
Parameter	Data generation	Model 1 Parametric	Model 2 DPM	Model 3 Neighborhood
$\beta_{1,\cdot\cdot}$	0.15	0.1512(0.0117)	0.1493(0.0105)	0.1508(0.0095)
σ^2	0.10	0.1002 (0.0006)	0.0999 (0.0006)	0.1000(0.0006)
σ_γ^2	0.05	0.0499(0.0041)	0.0499(0.0041)	0.0500 (0.0041)

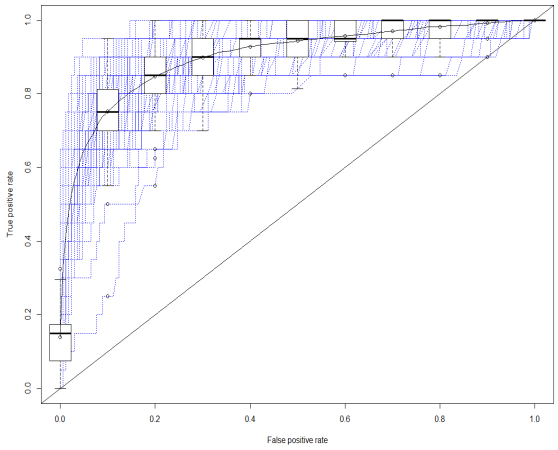
The panel of ROC curves in Figure 6 plot the false positive rate (FPR) versus the true positive rate (TPR) for the parametric, Dirichlet process and neighborhood models. We notice that DPM model performs strongly compared to the other two models.



(a) Parametric Model



(b) DPM Model



(c) Neighborhood model

Figure 6: Each panel shows 100 separate ROC curves for 100 replicated data sets. The average curve over the 100 replications is shown in black

3.6.2 Simulation Study 2

3.6.2.1 Data Generation Model

The data set are generated from a Conditional Autoregressive (CAR) Normal model for the second simulation study. The dependence and neighborhood structures are as defined before, that is, the neighborhood of rs are all links that include either the region r or the region s.

$$Y_{i,rs} \sim N(\mu_{i,rs}, \sigma^2)$$

$$\mu_{i,rs} = \beta_{1,rs} + \beta_{2,rs} X_i + \gamma_i + b_{rs}$$

where we generate $\gamma_i \sim \text{Normal}(0, \sigma_\gamma^2)$, with $\sigma_\gamma^2 = 0.05^2$ and $\sigma^2 = 0.10^2$. Here we are considering $\beta_{1,rs} = 0.15$ for all the links. Here b_{rs} is the random effect coming from neighborhood links and $b_{rs} \sim$ CAR Normal$(\mathcal{N}_{rs}, \sigma_b^2)$ with $\sigma_b^2 = 0.7^2$. $\beta_{2,rs}$ for non-connected regions are set to be zero. We specified connected and non-connected regions in parameter $\beta_{2,rs}$. We considered first 20 region pairs to be differentially connected for which $\beta_{2,rs} = 0.055$. $\beta_{2,rs}$ for non-differentially connected regions are set to be $= 0$.

$$\beta_{2,rs} = (\underbrace{0.055, \cdots, 0.055}_{20}, \underbrace{0, \cdots, 0}_{170})$$

3.6.2.2 Results from Analysis Models 1, 2, and 3

We utilize the data-generating model specified above to generate 100 replicated data set, each of size $(n_1 = 30, n_2 = 30)$ for the control and ASD groups respectively. We then analyzed

each data set using each of the three analysis models described. Note that analysis model 3 (neighborhood model) is the correctly specified model and the analysis model 2 (DPM) is, in some sense, a super-model.

The hyperparameters for the model 1 (parametric model) in the Equation 3.7 is specified as following

$$\mu_{\beta_1} \sim \mathrm{N}(0.5, 1), \quad \sigma^2_{\beta_1} \sim \text{Inverse-Gamma}(0.01, 0.01)$$

$$\sigma^2 \sim \text{Inverse-Gamma}(0.001, 0.001), \quad \tau^2 \sim \text{Inverse-Gamma}(0.01, 0.01)$$

$$p \sim \text{Uniform}(0, 1)$$

We have considered initial numbers of clusters to be 10 in the model 2 (DPM). Hyperparameters in Equation 3.22 is specified as following

$$\alpha \sim \text{Gamma}(0.1, 0.1), \quad \sigma^2_\theta \sim \text{Inverse-Gamma}(0.01, 0.01)$$

Model 3 (neighborhood model) hyperparameters in the Equation 3.30 are specified to be

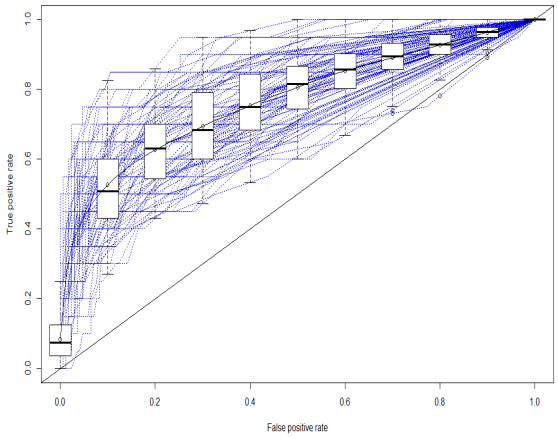
$$\{b_{rs} : 1 \leq r \leq 20\} \sim \text{CAR Normal}(\mathcal{N}_{rs}, \sigma^2_b), \quad \sigma^2_b \sim \text{Inverse-Gamma}(0.01, 0.01)$$

here weights $\{w_{rs}\}$, which are associated with $\mathcal{N}_{rs}$ are set to be equal to 1. We generated 100 replicated data sets and apply the different analysis model. The following table shows mean and standard error of parameters $\beta_{1,\cdot\cdot} = \frac{1}{190}\sum_{rs=1}^{190}\beta_{1,rs}$, σ^2 and σ^2_γ obtained from simulation study. We notice in Table V that estimated values of $\beta_{1,\cdot\cdot}$, σ^2 and σ^2_γ are approaching to their respective true values.

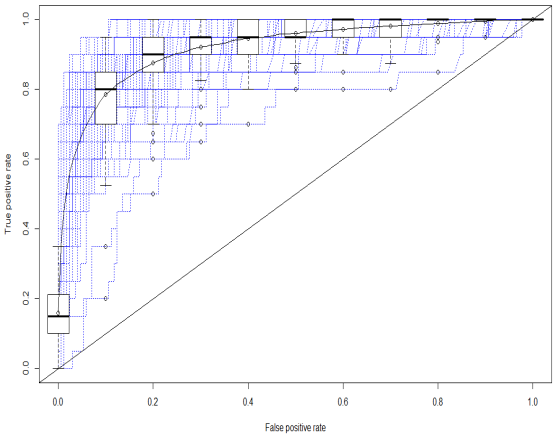
TABLE V: SIMULATION STUDY 2: AVERAGE VALUES AND STANDARD ERRORS FROM THREE ANALYSIS MODELS

		Analysis Models		
Parameter	Data generation	Model 1 Parametric	Model 2 DPM	Model 3 Neighborhood
$\beta_{1,\cdot\cdot}$	0.15	0.1519(0.1197)	0.1501(0.1194)	0.1517(0.0796)
σ^2	0.10	0.1003 (0.0007)	0.0999 (0.0007)	0.0999 (0.0007)
σ^2_γ	0.05	0.0487(0.0048)	0.0487(0.0048)	0.0487 (0.0048)

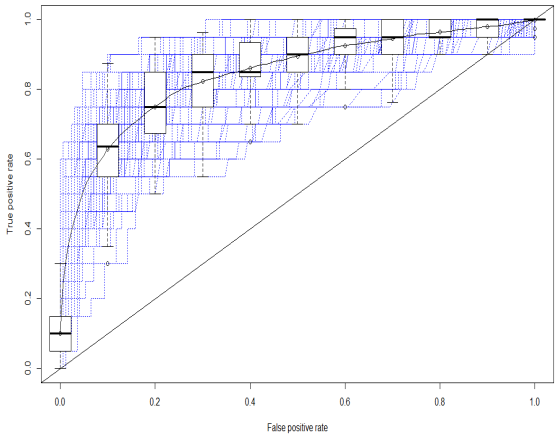
Figure 7 represents the multiple sets of ROC curve which plots the false positive rate (FPR) on the x-axis and the true positive rate (TPR) for parametric, Dirichlet process method and neighborhood model for the parameter $\beta_{2,rs}$. We notice that DPM performs exceptionally well even better than neighborhood model (the correctly specified model). The parametric model has performed really poorly in this simulation study.



(a) Parametric Model



(b) DPM Model



(c) Neighborhood model

Figure 7: Each panel shows 100 separate ROC curves for 100 replicated data sets. The average curve over the 100 replications is shown in black

3.7 ABIDE Data Analysis

Autism Brain Image Data Exchange (ABIDE) is an alliance among medical centers across the USA and Europe and is the largest repository of functional and structural brain imaging data for autism. It is a part of 1000 Functional Connectomes Project/ International Neuroimaging Datasharing Initiative (INDI). The data sets are anonymous, with no protected health information according to Health Insurance Portability and Accountability (HIPAA) guidelines and 1000 Functional Connectomes Project / INDI protocols. The ABIDE data were preprocessed by Connectome Computation System (CCS), the Configurable Pipeline for the Analysis of Connectomes (CPAC), the Data Processing Assistant for Resting-State fMRI (DPARSF) and the NeuroImaging Analysis Kit.

We used fMRI measurements from 84 regions of Broadmann areas of brain (see Table II) from 7 sites for investigating functional connectivity. There are a total of 314 subjects (see Table VI) including 164 control and 150 subjects with ASD. Each sites have sample of size less than 60 except New York University.

Connectivity metrics in this data set are obtained by computing Pearson correlation between two ROI pairs over all time points and then applying Fisher's z-transformation to the correlation [Bhaumik et al., 2018]. There are total $\binom{84}{2} = 3486$ many region pairs.

TABLE VI: NUMBER OF SUBJECTS AT EACH SITE FOR ABIDE DATA SET

Site	Control	ASD	Total
California Institute of Technology	21	16	37
New York University	42	35	77
Olin Institute of living	16	20	36
University of Pittsburgh	28	30	58
Social Brain Lab	15	15	30
San Diego State University	22	14	36
Stanford University	20	20	40
Total	164	150	314

3.7.1 Analysis on Single Site Data Set

We analyzed rs-fMRI data from 7 sites to explore differentially connected region between ASD and control subjects.

3.7.1.1 Preliminary Analysis

We perform a marginal analysis to understand the pattern of the data. We have randomly selected two subjects from California Institute of Technology, one from ASD and another one from control and plot their cross-correlated functional connectivity over 3486 links in Figure 8.

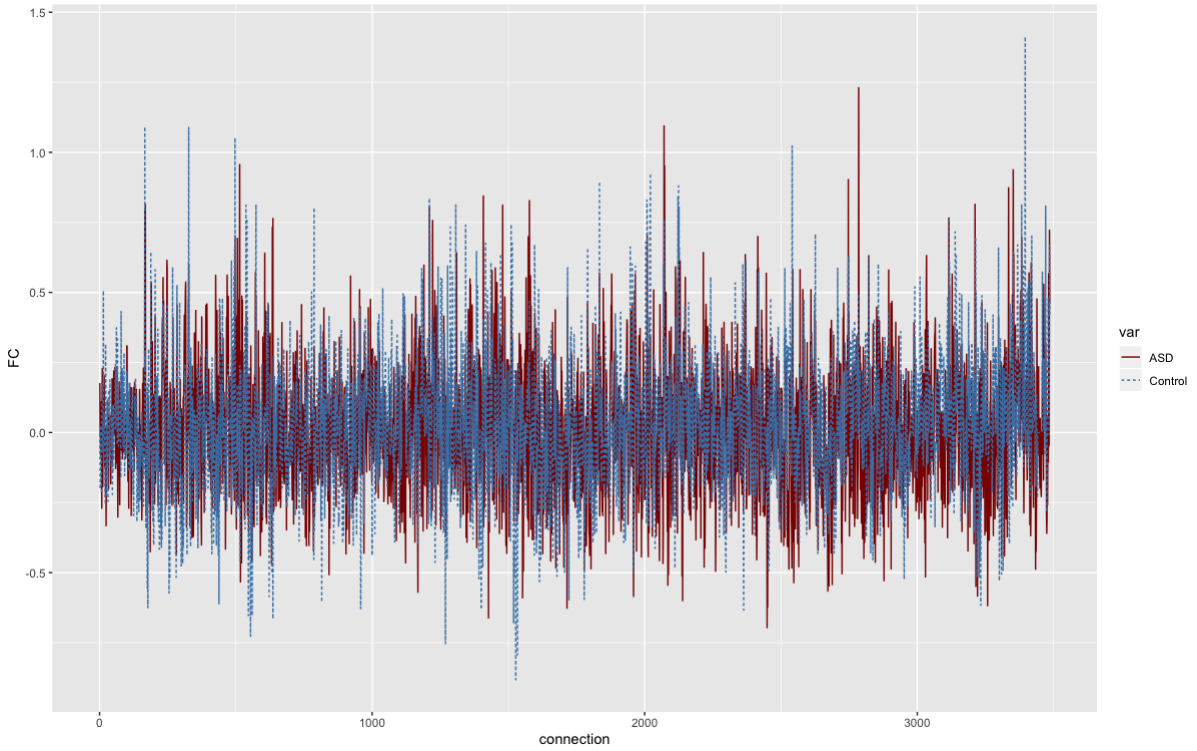


Figure 8: Cross-correlated Functional Connectivity from randomly selected subject of site California Institute of Technology

Figure 9 illustrates cross-correlated functional connectivity measurements, which are averaged over subjects from ASD and control group for all sites. Difference in patterns in averaged cross-correlated functional connectivity can be observed among sites.

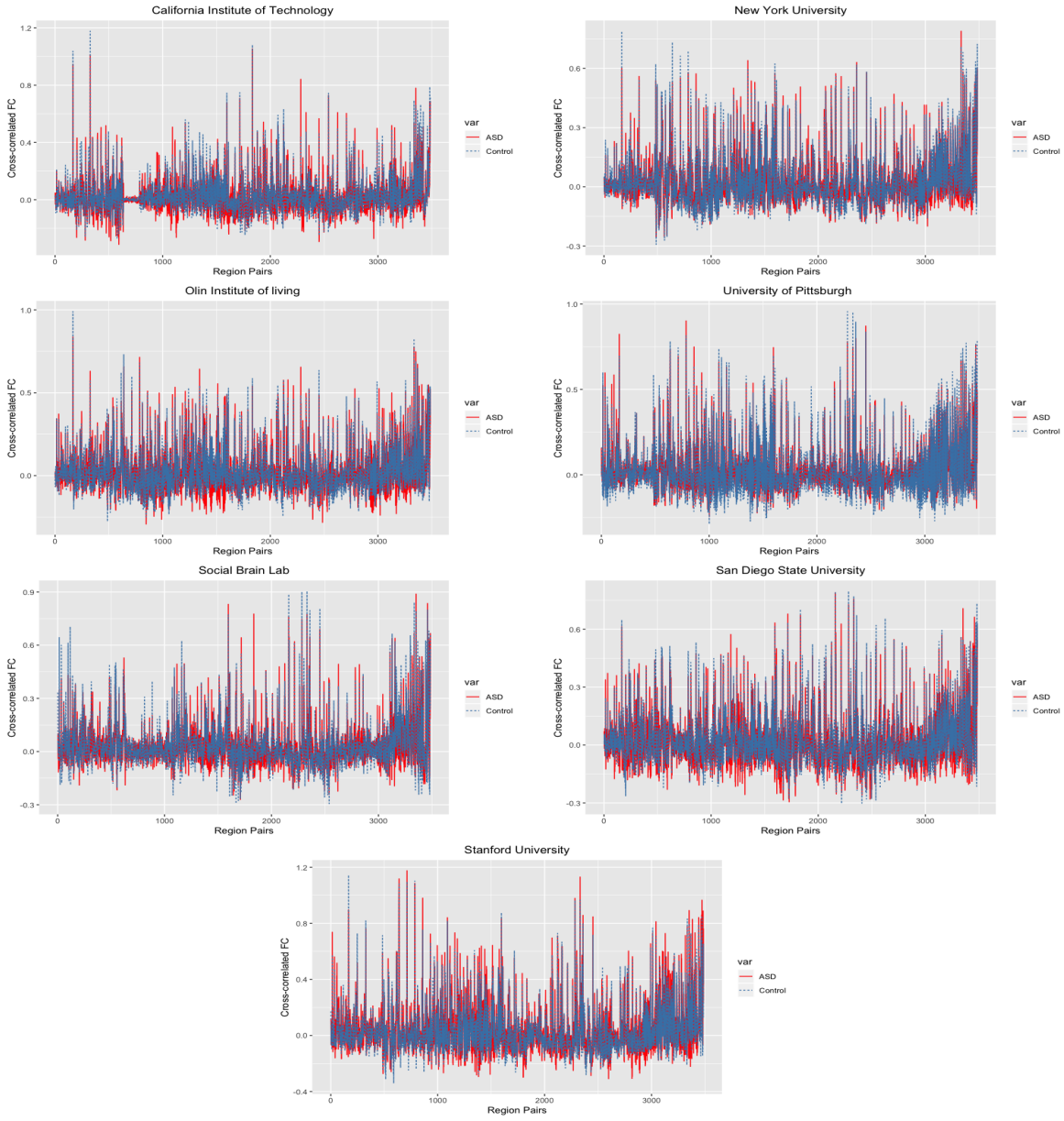


Figure 9: Average cross-correlated functional connectivity for all sites

We further proceed to perform a two sample t-test on cross-correlated functional connectivity between ASD and control group for each site. Numbers of significantly different ROI pair connection observed using two sample t-test is reported in Table VII.

TABLE VII: NUMBERS OF SIGNIFICANTLY DIFFERENT REGION PAIR CONNECTION OBSERVED USING TWO SAMPLE T-TEST

Site	P-value < 0.05
California Institute of Technology	199
New York University	192
Olin Institute of living	134
University of Pittsburgh	242
Social Brain Lab	200
San Diego State University	166
Stanford University	203

3.7.1.2 Results from Analysis Models 1, 2, and 3

Let $Y_{i,rs}$ is functional connectivity measurement between ROI from Broadmann areas of brain r and s (where, $1 \leq r < s \leq 84$) for subject. i, $i = 1, \cdots, n$.

$$\begin{aligned} Y_{i,rs}|\cdot &\sim N(\mu_{i,rs}, \sigma^2) \\ \mu_{i,rs} &= \beta_{1,rs} + \beta_{2,rs} X_i + \gamma_i \\ \sigma^2 &\sim \pi(d\sigma^2) \end{aligned} \tag{3.31}$$

where $X_i = 0$ if subject i is control while $X_i = 1$ if the subject is ASD, γ_i is a subject-level random effect term and $\sigma^2 \sim$ Inverse-Gamma$(0.001, 0.001)$.

The parametric model can be expressed as

$$\begin{aligned} \beta_{1,rs} &\sim N(\mu_{\beta_1}, \sigma^2_{\beta_1}), \quad \text{where } \mu_{\beta_1} \sim \text{N}(0.5, 1), \sigma^2_{\beta_1} \sim \text{Inverse-Gamma}(0.01, 0.01) \\ \beta_{2,rs} &\overset{i.i.d}{\sim} p\,\delta_0(.) + (1-p)G(\cdot), \quad \text{where } p \sim \text{Uniform}(0, 1) \\ G(\cdot) &\sim N(0, \tau^2), \quad \text{where } \tau^2 \sim \text{Inverse-Gamma}(0.01, 0.01), \\ \gamma_i &\sim N(0, \sigma^2_\gamma), \quad \text{where } \sigma^2_\gamma \sim \text{Inverse-Gamma}(0.001, 0.001) \end{aligned}$$

We consider a DPM extension of the parametric model by assuming $G(\cdot) \sim DP(\alpha, G_0)$, where the base measure $G_0 = N(0, \sigma^2_\theta)$. We are using Sethuraman's finite stick-breaking representation for posterior computation of the random G. The initial number of clusters are set to be $k = 12$ and precision parameter $\alpha \sim Gamma(0.1, 0.1)$. The complexity of the computation goes up

when DPM model is being used. As a result, DPM model has significantly higher computation time compared to parametric model.

The neighborhood model on ROI pairs is motivated by the assumption that differential connectivity at rs may be associated with connectivities in other links which are attached with rs at one of its vertices. We propose a conditional auto-regressive (CAR) model with neighborhood structure $\mathcal{N}$ to incorporate potential dependencies in functional connectivity of regions rs and its neighbors.

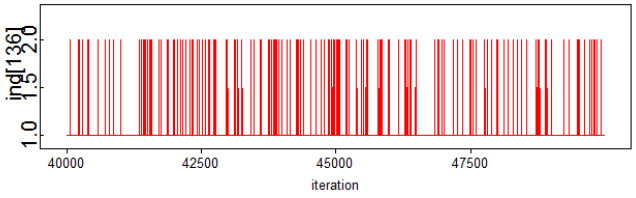
$$Y_{i,rs} \sim f(\cdot|\eta_{i,rs})$$

$$g(\eta_{i,rs}) = \beta_{1,rs} + \beta_{2,rs}X_i + \gamma_i + b_{rs}$$

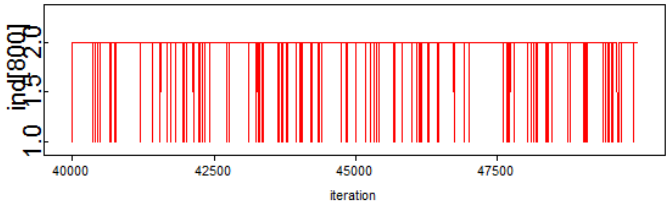
$$\{b_{rs} : 1 \leq r < s \leq 84\} \sim \text{CAR Normal}(\mathcal{N}, \sigma_b^2), \quad \text{where } \sigma_b^2 \sim \text{Inverse-Gamma}(0.001, 0.001)$$

The results that we present below are based on $50,000$ post-burn-in Markov Chain samples from the posterior distribution of each of the Bayesian models after an initial burn-in of $25,000$.

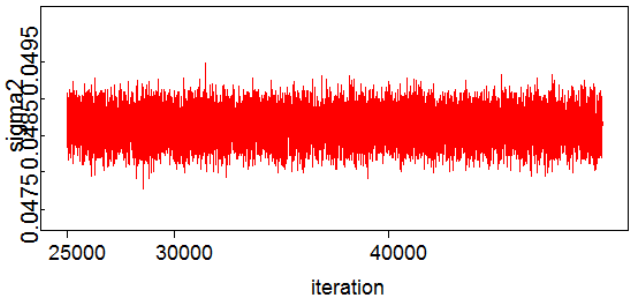
Figure 10 and Figure 11 illustrates trace plots of the indicator variable of links 136, 800, σ^2 and deviance $(-2 * \log(Y_{i,rs}|\cdot))$ from site California Institute of Technology by applying parametric model (model 1) and DPM model (model 2).



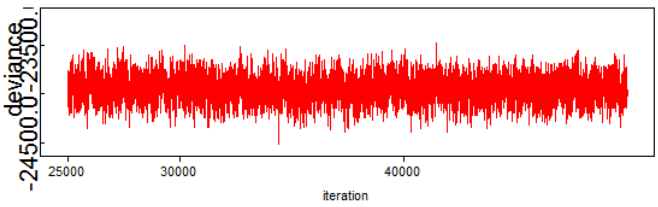
(a) Trace plot of Link 136



(b) Trace plot of Link 800

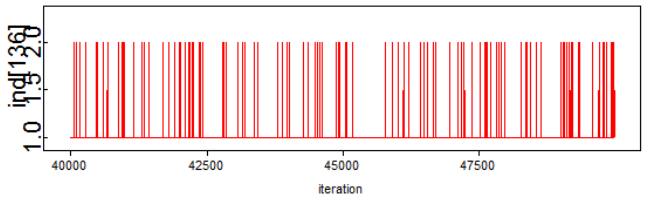


(c) Trace plot of σ^2

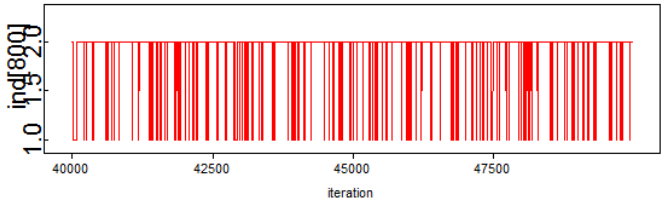


(d) Trace plot of deviance

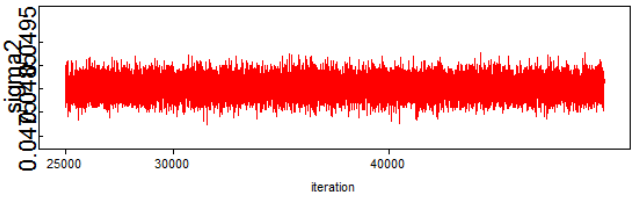
Figure 10: Trace plots by applying parametric model for data set from site California Institute of Technology



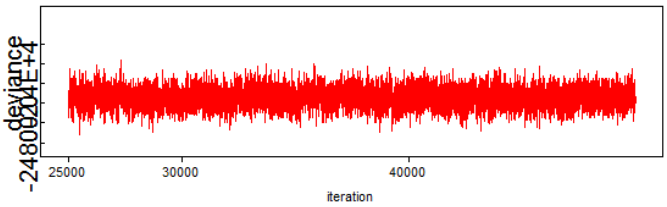
(a) Trace plot of Link 136



(b) Trace plot of Link 800



(c) Trace plot of σ^2



(d) Trace plot of deviance

Figure 11: Trace plots by applying DPM model for data set from site California Institute of Technology

The following table summarizes significant links discovered from these models.

TABLE VIII: NUMBERS OF SIGNIFICANT DIFFERENTIAL CONNECTED LINKS AT EACH SITE

Site	Parametric	DPM	Neighborhood
California Institute of Technology	68	89	95
New York University	264	207	341
Olin Institute of living	85	100	135
University of Pittsburgh	184	160	275
Social Brain Lab	174	166	226
San Diego State University	143	137	197
Stanford University	119	117	178

The significant links are not identical across all sites. The number of significant links also varies across sites. There might be several reasons for these disparities. The samples size for most of the sites are quite small. Moreover, there is an unequal distribution of age across sites (Table I) which can contribute to these disparities. [Uddin et al., 2013], [Padmanabhan et al., 2013], [Nomi and Uddin, 2015] investigated age-related differences in brain connectivity in ASD.

3.7.2 Analysis on Combined Data Set

In the previous section, analysis was done for each of the 7 sites separately and we notice that significant links are not identical across all the sites. This might have occurred due to small sample size and unequal age group distribution in the sites. We combine data across all the site, which provide considerable sample size. The combined data set has total of 314 subjects including 164 control and 150 subjects with ASD. We consider a random effect $\delta_c \sim \mathrm{N}(0, 0.5^2)$ to account for heterogeneity across sites as described in Equation 3.1.

The results that we present below are based on $100,000$ post-burn-in Markov Chain samples from the posterior distribution of each of the Bayesian models after an initial burn-in of $50,000$. The following table summarizes significant links discovered from these models.

TABLE IX: NUMBERS OF SIGNIFICANT DIFFERENTIAL CONNECTED LINKS FOR COMBINED DATA SET

Parametric	DPM	Neighborhood
218	135	203

The following table demonstrates the shared differentially connected regions obtained using each of the Bayesian models from combined data set of all 7 sites of ABIDE data and the probability of each links to be in connected regions using above mentioned methods.

TABLE X: DIFFERENTIAL CONNECTED REGIONS IN COMBINED DATA

Region-1	Region-2	SS	DPM	Neighborhood
(R). Primary Somatosensory Cortex	(R). Secondary Visual Cortex	0.98	0.96	0.98
(L). Anterior Prefrontal Cortex	(R). Secondary Visual Cortex	0.89	0.83	0.87
(L). Anterior Prefrontal Cortex	(L). Angular gyrus	0.83	0.78	0.83
(L). Anterior Prefrontal Cortex	(R). Supramarginal Gyrus	0.86	0.77	0.83
(L). Anterior Prefrontal Cortex	(R). Inferior Prefrontal Gyrus	0.86	0.78	0.85
(R). Anterior Prefrontal Cortex	(R). Primary Motor Cortex	0.92	0.88	0.90
(L). Orbitofrontal Cortex	(L). Primary Somatosensory Cortex	0.85	0.79	0.87
(L). Orbitofrontal Cortex	(R). Dorsal Frontal Cortex	0.95	0.95	0.96
(R). Orbitofrontal Cortex	(L). Primary Visual Cortex	0.98	0.97	0.99
(R). Orbitofrontal Cortex	(L). Inferior Temporal Gyrus	1.00	1.00	1.00
(R). Insular Cortex	(R). Anterior Cingulate Cortex	0.94	0.92	0.95
(R). Insular Cortex	(R). Angular gyrus	0.95	0.93	0.95
(R). Insular Cortex	(R). Premotor Cortex	0.95	0.90	0.96
(L). Primary Visual Cortex	(R). Associative Visual Cortex	0.86	0.77	0.83
(L). Primary Visual Cortex	(L). Primary Somatosensory Cortex	0.96	0.95	0.97
(L). Primary Visual Cortex	(L). Ventral Anterior Cingulate Cortex	0.95	0.93	0.95
(L). Primary Visual Cortex	(L). Piriform Cortex	0.92	0.86	0.92
(L). Secondary Visual Cortex	(L). Subgenual cortex	1.00	1.00	1.00
(L). Secondary Visual Cortex	(R). Subcentral Area	0.95	0.94	0.96
(L). Secondary Visual Cortex	(R). IFC pars triangularis	0.89	0.76	0.87
(R). Secondary Visual Cortex	(R). Angular gyrus	0.88	0.86	0.90
(R). Secondary Visual Cortex	(R). Somatosensory Association Cortex	0.98	0.95	0.98
(R). Secondary Visual Cortex	(L). Premotor Cortex	1.00	1.00	1.00
(R). Secondary Visual Cortex	(R). Premotor Cortex	0.90	0.80	0.90
(L). Associative Visual Cortex	(L). Inferior Prefrontal Gyrus	1.00	1.00	1.00
(R). Associative Visual Cortex	(R). Ventral Posterior Cingulate Cortex	0.90	0.85	0.89
(R). Associative Visual Cortex	(R). Perirhinal cortex	0.99	0.98	0.99

TABLE X – DIFFERENTIAL CONNECTED REGIONS IN COMBINED DATA (Continued)

Region-1	Region-2	SS	DPM	Neighborhood
(R). Associative Visual Cortex	(R). Dorsolateral Prefrontal Cortex	0.95	0.94	0.96
(L). Primary Somatosensory Cortex	(L). Primary Auditory Cortex	0.97	0.96	0.98
(R). Primary Somatosensory Cortex	(R). Angular gyrus	1.00	1.00	1.00
(L). Inferior Temporal Gyrus	(R). Subgenual cortex	0.85	0.79	0.85
(L). Inferior Temporal Gyrus	(R). Posterior Entorhinal Cortex	0.86	0.79	0.85
(L). Inferior Temporal Gyrus	(R). IFC pars opercularis	0.85	0.76	0.84
(L). Middle Temporal Gyrus	(R). Anterior Cingulate Cortex	0.89	0.83	0.88
(L). Middle Temporal Gyrus	(L). IFC pars triangularis	0.90	0.88	0.90
(L). Middle Temporal Gyrus	(L). Dorsolateral Prefrontal Cortex	0.90	0.77	0.89
(R). Middle Temporal Gyrus	(R). Dorsal Posterior Cingulate Cortex	0.94	0.87	0.93
(R). Middle Temporal Gyrus	(R). Premotor Cortex	0.92	0.82	0.90
(L). Superior Temporal Gyrus	(L). Ventral Posterior Cingulate Cortex	0.93	0.86	0.93
(L). Superior Temporal Gyrus	(R). Retrosplenial Cingulate Cortex	0.94	0.88	0.93
(L). Superior Temporal Gyrus	(L). Primary Somatosensory Cortex	0.99	0.99	0.99
(L). Superior Temporal Gyrus	(L). Subcentral Area	0.89	0.82	0.86
(R). Superior Temporal Gyrus	(R). Dorsal anterior Cingulate Cortex	0.92	0.82	0.92
(R). Superior Temporal Gyrus	(L). Anterior Cingulate Cortex	0.86	0.78	0.83
(R). Superior Temporal Gyrus	(L). Parahippocampal cortex	0.87	0.80	0.84
(L). Ventral Posterior Cingulate Cortex	(L). Anterior Cingulate Cortex	0.95	0.94	0.95
(L). Ventral Posterior Cingulate Cortex	(R). Fusiform gyrus	0.93	0.89	0.92
(R). Ventral Posterior Cingulate Cortex	(L). Posterior Entorhinal Cortex	0.88	0.84	0.89
(R). Ventral Posterior Cingulate Cortex	(L). Primary Somatosensory Cortex	0.93	0.90	0.92
(R). Ventral Posterior Cingulate Cortex	(R). Cingulate Cortex	0.88	0.82	0.88
(R). Ventral Posterior Cingulate Cortex	(R). IFC pars triangularis	0.93	0.89	0.93
(L). Ventral Anterior Cingulate Cortex	(R). Premotor Cortex	0.88	0.76	0.88
(L). Ventral Anterior Cingulate Cortex	(L). Dorsal Frontal Cortex	0.93	0.90	0.90
(R). Ventral Anterior Cingulate Cortex	(L). Parahippocampal cortex	0.86	0.79	0.84
(R). Ventral Anterior Cingulate Cortex	(L). Temporopolar Area	1.00	0.99	0.99

TABLE X – DIFFERENTIAL CONNECTED REGIONS IN COMBINED DATA (Continued)

Region-1	Region-2	SS	DPM	Neighborhood
(R). Ventral Anterior Cingulate Cortex	(R). IFC pars triangularis	0.89	0.79	0.89
(R). Ventral Anterior Cingulate Cortex	(R). Dorsolateral Prefrontal Cortex	0.87	0.79	0.88
(L). Subgenual cortex	(R). Primary Somatosensory Cortex	0.97	0.95	0.97
(R). Subgenual cortex	(L). Retrosplenial Cingulate Cortex	0.87	0.83	0.87
(R). Subgenual cortex	(L). Dorsal Frontal Cortex	1.00	1.00	1.00
(R). Subgenual cortex	(R). Dorsolateral Prefrontal Cortex	0.97	0.97	0.97
(L). Piriform Cortex	(R). Piriform Cortex	0.99	0.97	0.99
(L). Piriform Cortex	(L). Somatosensory Association Cortex	1.00	0.99	0.99
(R). Piriform Cortex	(L). Primary Somatosensory Cortex	0.94	0.92	0.94
(R). Piriform Cortex	(R). Cingulate Cortex	1.00	0.99	0.99
(R). Piriform Cortex	(R). Dorsolateral Prefrontal Cortex	0.86	0.77	0.86
(R). Piriform Cortex	(L). Inferior Prefrontal Gyrus	0.95	0.88	0.93
(R). Piriform Cortex	(L). Dorsolateral Prefrontal Cortex	0.96	0.92	0.96
(R). Posterior Entorhinal Cortex	(R). Temporopolar Area	0.93	0.87	0.93
(L). Retrosplenial Cingulate Cortex	(R). Supramarginal Gyrus	0.96	0.94	0.96
(R). Retrosplenial Cingulate Cortex	(L). Fusiform gyrus	0.94	0.90	0.93
(R). Retrosplenial Cingulate Cortex	(L). IFC pars triangularis	0.98	0.96	0.98
(R). Retrosplenial Cingulate Cortex	(R). Inferior Prefrontal Gyrus	0.89	0.84	0.88
(L). Primary Somatosensory Cortex	(R). Supramarginal Gyrus	0.99	0.99	0.99
(R). Primary Somatosensory Cortex	(R). Somatosensory Association Cortex	0.89	0.79	0.88
(L). Cingulate Cortex	(L). Primary Auditory Cortex	1.00	1.00	1.00
(L). Cingulate Cortex	(L). IFC pars triangularis	0.90	0.89	0.92
(L). Cingulate Cortex	(R). Dorsal Frontal Cortex	0.93	0.88	0.93
(R). Cingulate Cortex	(L). IFC pars opercularis	0.90	0.85	0.90
(L). Dorsal Posterior Cingulate Cortex	(R). Temporopolar Area	0.97	0.94	0.97
(L). Dorsal Posterior Cingulate Cortex	(R). Primary Motor Cortex	0.96	0.92	0.95
(R). Dorsal Posterior Cingulate Cortex	(R). IFC pars triangularis	0.92	0.88	0.93
(L). Dorsal anterior Cingulate Cortex	(R). Parahippocampal cortex	0.99	0.99	0.99

TABLE X – DIFFERENTIAL CONNECTED REGIONS IN COMBINED DATA (Continued)

Region-1	Region-2	SS	DPM	Neighborhood
(L). Dorsal anterior Cingulate Cortex	(R). Subcentral Area	1.00	1.00	1.00
(L). Dorsal anterior Cingulate Cortex	(L). IFC pars opercularis	1.00	1.00	1.00
(L). Dorsal anterior Cingulate Cortex	(R). Dorsolateral Prefrontal Cortex	0.85	0.79	0.85
(R). Dorsal anterior Cingulate Cortex	(R). Inferior Prefrontal Gyrus	1.00	0.99	1.00
(L). Anterior Cingulate Cortex	(L). Somatosensory Association Cortex	0.92	0.84	0.92
(R). Anterior Cingulate Cortex	(L). IFC pars triangularis	0.97	0.94	0.96
(R). Anterior Cingulate Cortex	(R). Inferior Prefrontal Gyrus	0.92	0.90	0.93
(L). Anterior Entorhinal Cortex	(R). Fusiform gyrus	0.86	0.78	0.87
(L). Anterior Entorhinal Cortex	(L). Temporopolar Area	0.99	0.99	1.00
(L). Anterior Entorhinal Cortex	(R). Premotor Cortex	0.84	0.77	0.83
(L). Anterior Entorhinal Cortex	(R). Somatosensory Association Cortex	0.86	0.77	0.85
(R). Anterior Entorhinal Cortex	(R). Parahippocampal cortex	1.00	1.00	1.00
(R). Anterior Entorhinal Cortex	(L). Temporopolar Area	0.93	0.89	0.90
(R). Anterior Entorhinal Cortex	(R). Primary Motor Cortex	0.87	0.80	0.86
(R). Anterior Entorhinal Cortex	(L). Somatosensory Association Cortex	0.99	0.98	0.99
(L). Perirhinal cortex	(R). Premotor Cortex	1.00	1.00	1.00
(L). Perirhinal cortex	(R). Somatosensory Association Cortex	0.94	0.92	0.94
(L). Perirhinal cortex	(R). Dorsolateral Prefrontal Cortex	0.96	0.95	0.95
(L). Parahippocampal cortex	(L). Primary Motor Cortex	0.94	0.90	0.94
(L). Parahippocampal cortex	(R). Dorsal Frontal Cortex	0.90	0.89	0.90
(R). Parahippocampal cortex	(L). Fusiform gyrus	0.95	0.94	0.95
(R). Parahippocampal cortex	(R). Primary Motor Cortex	0.97	0.96	0.97
(R). Parahippocampal cortex	(L). Primary Auditory Cortex	0.99	0.99	0.99
(R). Parahippocampal cortex	(R). Subcentral Area	0.87	0.79	0.87
(R). Parahippocampal cortex	(R). IFC pars triangularis	0.94	0.90	0.94
(L). Fusiform gyrus	(L). Primary Auditory Cortex	0.99	0.99	0.98
(L). Fusiform gyrus	(R). Primary Auditory Cortex	0.95	0.90	0.95
(R). Fusiform gyrus	(R). Temporopolar Area	0.99	0.97	0.99

TABLE X – DIFFERENTIAL CONNECTED REGIONS IN COMBINED DATA (Continued)

Region-1	Region-2	SS	DPM	Neighborhood
(L). Temporopolar Area	(R). Primary Motor Cortex	0.96	0.95	0.96
(L). Temporopolar Area	(L). Primary Auditory Cortex	0.98	0.97	0.98
(L). Temporopolar Area	(R). Somatosensory Association Cortex	0.93	0.90	0.93
(R). Temporopolar Area	(L). Somatosensory Association Cortex	0.98	0.96	0.98
(R). Temporopolar Area	(R). Premotor Cortex	0.90	0.80	0.90
(R). Temporopolar Area	(L). Dorsolateral Prefrontal Cortex	0.89	0.78	0.90
(L). Angular gyrus	(R). Primary Auditory Cortex	0.92	0.90	0.90
(R). Angular gyrus	(R). Supramarginal Gyrus	0.93	0.85	0.93
(R). Angular gyrus	(L). Somatosensory Association Cortex	0.90	0.84	0.89
(L). Primary Motor Cortex	(L). Subcentral Area	0.99	0.97	0.98
(L). Primary Motor Cortex	(R). Inferior Prefrontal Gyrus	0.87	0.82	0.87
(L). Primary Motor Cortex	(L). Somatosensory Association Cortex	0.94	0.92	0.95
(R). Primary Motor Cortex	(R). Subcentral Area	0.89	0.80	0.89
(L). Supramarginal Gyrus	(L). IFC pars triangularis	0.97	0.98	0.97
(R). Supramarginal Gyrus	(R). IFC pars opercularis	0.95	0.90	0.95
(R). Supramarginal Gyrus	(R). Dorsal Frontal Cortex	0.90	0.85	0.90
(L). Primary Auditory Cortex	(L). Dorsolateral Prefrontal Cortex	0.87	0.80	0.86
(L). Primary Auditory Cortex	(R). IFC pars triangularis	0.98	0.99	0.99
(L). Dorsolateral Prefrontal Cortex	(R). Dorsal Frontal Cortex	0.98	0.96	0.98
(R). Dorsolateral Prefrontal Cortex	(R). Dorsal Frontal Cortex	0.99	0.99	1.00
(R). Inferior Prefrontal Gyrus	(L). Dorsolateral Prefrontal Cortex	0.99	0.98	0.99
(R). Inferior Prefrontal Gyrus	(R). Dorsolateral Prefrontal Cortex	0.87	0.80	0.85
(L). Somatosensory Association Cortex	(R). Somatosensory Association Cortex	0.99	1.00	1.00
(R). Premotor Cortex	(L). Somatosensory Association Cortex	0.92	0.90	0.90

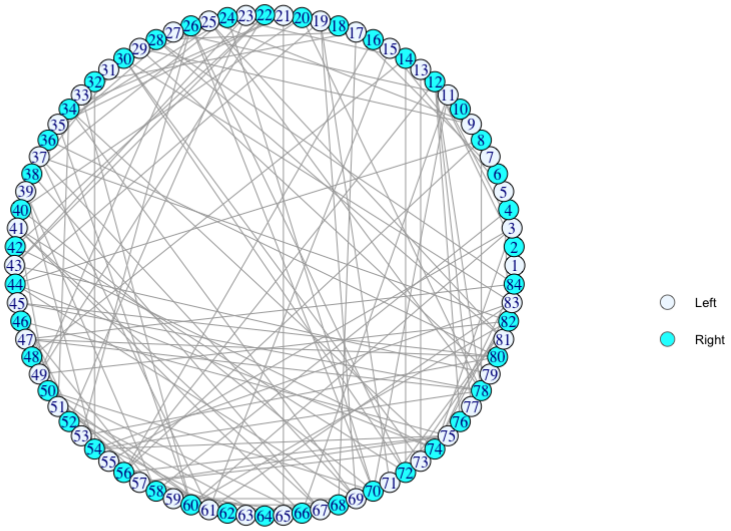


Figure 12: Connectivity between differentially connected regions across brain hemisphere using combined data set of all 7 sites of ABIDE data

3.7.3 Conclusion

We analyze summarized functional connectivity measurements between ROI pairs from Broadmann areas of brain using proposed Bayesian models. The analysis is done using data from 7 sites separately and then again combining the data across all the sites. The combined analysis models and identifies the links which are differentially connected across sites. The

single site data analysis identifies differentially connected links at each individual site. We observe that those links are not identical across all sites.

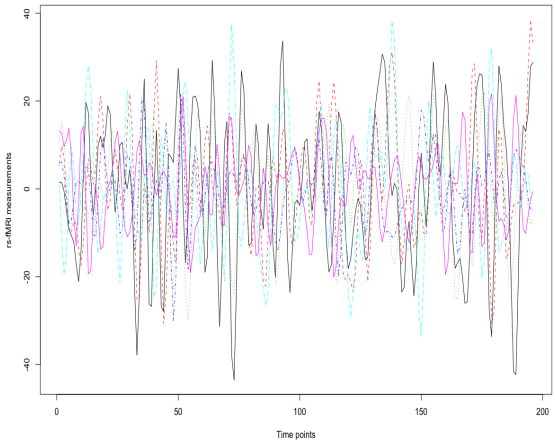
CHAPTER 4

REGIONAL-TEMPORAL FUNCTIONAL CONNECTIVITY MODEL

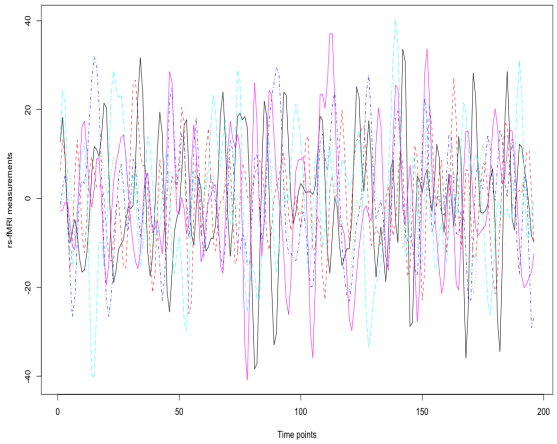
4.1 Introduction

In the previous chapter, the outcome variable was a connectivity metric summarized over time between ROI pairs. In this chapter, instead of considering measures summarized over time, we directly model the time sequence of rs-fMRI measurement at each ROI. The time-dependent model for functional connectivity jointly considers the time sequence of fMRI measurements and the modularity structure of brain regions. We have applied dynamic linear model to capture the temporal structure of data and the potential correlation between connected regions is modeled using the Hidden Potts model with latent variable. As an application we used these models to analyze functional connectivity in the Autism Brain Imaging Data Exchange (ABIDE).

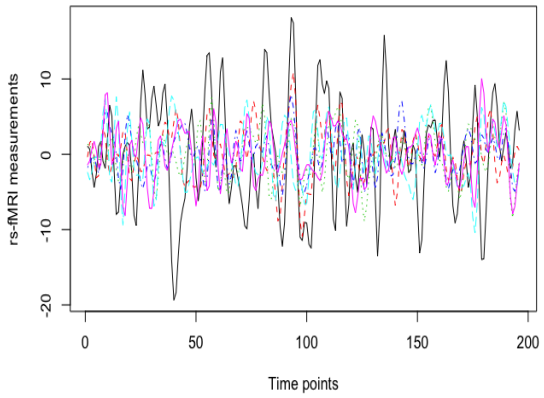
Figure 13 illustrates rs-fMRI measurements over 196 time points for 6 randomly selected ROIs with two randomly selected subjects, one from ASD subject and other from control.



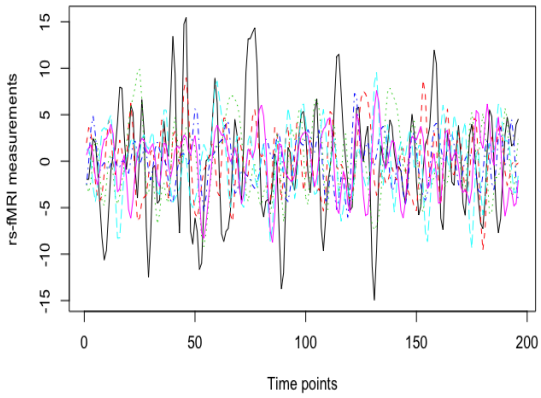
(a) rs-fMRI measurement for randomly selected control subject



(b) rs-fMRI measurement for randomly selected ASD subject



(c) Average rs-fMRI measurement for all control subject



(d) Average rs-fMRI measurement for all ASD subject

Figure 13: rs-fMRI measurement for randomly selected ROIs $(6, 17, 32, 40, 55, 56)$ from randomly selected site (Social Brain Lab)

4.2 Model Specification

We have rs-fMRI measurements Y_{itr} for $i = 1, \cdots, n$ subjects, at $t = 1, \cdots, T$ are discrete time points, and $r = 1, \cdots, R$ regions We model Y_{itr} as following

$$Y_{itr} = G(\theta_{itr}, b_r) + \varepsilon_{itr}, \tag{4.1}$$

where θ_{itr} is the temporal component of the model and b_r is a random effect which brings in potential correlation between connected regions. We assume that errors $\varepsilon_{itr} \overset{i.i.d}{\sim} N(0, \sigma^2)$.

We utilize Dynamic Linear Model (DLM) [West and Harrison, 2006] to model the temporal component θ_{itr}. Dynamic Linear Models are state-space models where we specify an auto-regressive structure on the latent parameters θ_{itr}.

$$\begin{aligned} \text{Observational Equation} \quad & Y_{itr} = G(\theta_{itr}, b_r) + \varepsilon_{itr} \\ \text{System Equation of Temporal Component} \quad & \theta_{itr} = \beta_{ir}\theta_{i(t-1)r} + \eta_{itr} \end{aligned} \tag{4.2}$$

where, $\eta_{itr} \overset{\text{i.i.d}}{\sim} N(0, w^2)$ is the error term of system equation and the coefficient β_{ir} depends on subject i and region r. We assume a normal prior for $\beta_{ir} \sim N(\mu_{\beta_r}, \sigma^2_{\beta_r})$, where the hyperpriors μ_{β_r} and $\sigma^2_{\beta_r}$ are shared across subjects and specify the initial distribution as $\theta_{i0r} \sim N(m_0, C_0)$, for all i and r.

The region-level random effects $\{b_r\}_{r=1}^R$ are modeled using a hidden Potts model which creates stochastic clustering among $\{b_r\}_{r=1}^R$ with same cluster b_r values sharing identical values. Let z_r denote the (random) cluster label of b_r. The random effects $\{b_r\}_{r=1}^R$ can then be labeled

as $\{b^*_{z_1}, \cdots, b^*_{z_R}\}$, where $b^*_1, b^*_2, \ldots$ denote the unique values associated with the clusters. The hidden Potts model allocates small weights to large cluster, and therefore generates various smaller clusters [Zhang et al., 2015b].

4.3 State-space Models

We describe the state-space model for the temporal components θ_{itr} in Equation 4.2 for observations Y_{itr}. For notational simplicity, we suppress subscripts i and r in this section and denote them as θ_t and Y_t respectively. We also suppress the dependence of Y_{itr} on the region-level random effects b_r in this section.

State-space models have become popular in time series analysis in the last two decades. Univariate and multivariate time series can be modeled using state-space models. The observed time series Y_t is related to an underlying unobservable process called state process (θ_t). The following description of state-space models follows [West and Harrison, 2006], and [Petris et al., 2009].

1. $\{\theta_t, t = 0, 1, \cdots\}$ is a (first-order) stationary Markov process. The distribution of θ_t depends on the past history through θ_{t-1}. The density of the initial θ_0 is given by $p_0(\theta_0)$ and the transition density of the stationary process is given by $p(\theta_t|\theta_{t-1})$.

2. Observation Y_t depends on the state process $\{\theta_t\}$ only through θ_t. For any $T \geq 1$, $(Y_1, \ldots, Y_t)$ are conditionally independent so that the joint density of $(Y_1, \cdots, Y_T)|(\theta_1, \cdots, \theta_T)$ is given by $\prod_{t=1}^{T} f(Y_t|\theta_t)$.

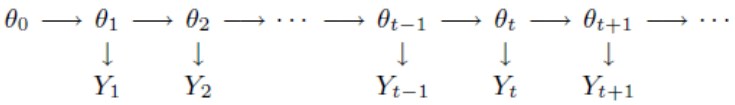


Figure 14: state-space model.

(Source: [Petris et al., 2009])

In this model state variables are continuous unlike hidden Markov model, where state variables are discrete. The process $\{(\theta_t, Y_t), t = 1, 2, \cdots\}$ is Markovian. The joint density of (θ_t, Y_t) can be obtained as

$$\begin{aligned}(\theta_0, \theta_1, \cdots, \theta_T, Y_1, \cdots, Y_T) &\sim p_0(\theta_0) \prod_{t=1}^{T} p(\theta_t, Y_t | \theta_0, \theta_1, \cdots, \theta_{t-1}, Y_1, \cdots, Y_{t-1}) \\ &= p_0(\theta_0) \prod_{t=1}^{T} f(Y_t | \theta_0, \theta_1, \cdots, \theta_t, Y_1, \cdots, Y_{t-1}) \\ &\quad p(\theta_t | \theta_0, \theta_1, \cdots, \theta_{t-1}, Y_1, \cdots, Y_{t-1}) \\ &= p_0(\theta_0) \prod_{t=1}^{T} f(Y_t | \theta_t) p(\theta_t | \theta_{t-1}) \end{aligned} \tag{4.3}$$

Equation 4.3 implies that Y_t is conditionally independent from the past observations $(Y_1, \cdots, Y_{t-1})$ given θ_t. The density of $(Y_1, \cdots, Y_T)$ can be obtained by marginalizing the joint density in Equation 4.3 over θ, however the marginal density does not have a closed form expression.

4.3.1 Dynamic Linear Model

One of the most important classes of state-space model is Gaussian linear state-space model, also known as dynamic linear model (DLM).

Definition 4.3.1. *[West and Harrison, 2006] The general dynamic linear model (DLM) is defined by a set of quadruples for each time point t,* $\{F, G, V, W\}_t = \{F_t, G_t, V_t, W_t\}$ *where* F_t *is a known* $(n \times r)$ *design matrix,* G_t *is a known* $(n \times n)$ *transfer or state matrix,* V_t *is a known* $(r \times r)$ *variance matrix and* W_t *is a known* $(n \times n)$ *variance matrix. Let* Y_t *be an* $(r \times 1)$ *vector observation over times* $t = 1, 2, \cdots$*, which is associated with state vector* θ_t *through the set of quadruples at time* t*.*

$$
\begin{aligned}
(Y_t|\theta_t) &\sim N[F_t'\theta_t, V_t] \\
(\theta_t|\theta_{t-1}) &\sim N[G_t\theta_{t-1}, W_t]
\end{aligned}
\tag{4.4}
$$

Equation 4.4 also conditional on D_{t-1}*, where* $D_{t-1} = \{Y_{t-1}, D_{t-2}\}$ *the information set available prior to time* t*.*

Alternatively, the equation can be defined as follows

$$
\begin{aligned}
Y_t &= F_t\theta_t + \nu_t, \quad \nu_t \sim N[0, V_t] \\
\theta_t &= G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t]
\end{aligned}
\tag{4.5}
$$

where ν_t *is the observational error and* ω_t *is the system or evolution error. These error sequences are internally and mutually independent.*

The general univariate DLM is defined by Definition 4.3.1 with $r = 1$.

Definition 4.3.2. *[West and Harrison, 2006] The general univariate DLM for each timepoint t is defined by*

$$\begin{aligned} \textit{Observational equation:} \quad & Y_t = F_t'\theta_t + \nu_t, \quad \nu_t \sim N(0, V), \\ \textit{System equation:} \quad & \theta_t = G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t), \\ \textit{Initial information:} \quad & (\theta_0|D_0) \sim N(m_0, C_0), \end{aligned}$$

where ν_t and ω_t are internally and mutually independent and also independent of $(\theta_0|D_0)$

Theorem 4.3.1. *[West and Harrison, 2006] The one-step forecast and the posterior distribution for any time $t > 0$ can be obtained as following*

1. *Posterior for θ_{t-1}:*

$$\begin{gathered} (\theta_{t-1}|D_{t-1}) \sim N[m_{t-1}, C_{t-1}], \\ \textit{for some mean } \ m_{t-1} \ \textit{ and variance matrix } \ C_{t-1}. \end{gathered}$$

2. *Prior for θ_t :*

$$\begin{gathered} (\theta_t|D_{t-1}) \sim N[a_t, R_t], \\ \textit{where } \ a_t = G_t m_{t-1} \ \textit{ and } \ R_t = G_t C_{t-1} G_t' + W_t. \end{gathered}$$

3. *1-step forecast:*

$$\begin{gathered} (Y_t|D_{t-1}) \sim N[f_t, Q_t], \\ \textit{where } \ f_t = F_t' a_t \ \textit{ and } \ Q_t = F_t' R_t F_t + V_t. \end{gathered}$$

4. *Posterior for θ_t:*

$$\begin{gathered} (\theta_t|D_t) \sim N[m_t, C_t], \\ \textit{with } \ m_t = a_t + A_t e_t \ \textit{ and } \ C_t = R_t - A_t Q_t A_t', \\ \textit{where } \ A_t = R_t F_t Q_t^{-1}, \ \textit{ and } \ e_t = Y_t - f_t. \end{gathered}$$

The posterior probability of regression coefficient Equation 4.2 can be calculated as following

$$\text{Prior Probability: } \beta_{ir} \sim N(\mu_{\beta_r}, \sigma^2_{\beta_r})$$

$$\text{Posterior Probability: } (\beta_{ir}|\theta_{1:T}, D_T) \sim \pi(\beta_{ir}) * \prod_{t=1}^{T} p(\theta_t|\beta_{ir}, D_t)$$

$$\sim N(\mu_{\beta_{ir}}, \sigma^2_{\beta_{ir}})$$

$$\text{with } \mu_{\beta_{ir}} = \frac{\left(\frac{\sum_{t=1}^{T}\theta_t}{w^2} + \frac{\mu_{\beta_r}}{\sigma^2_{\beta_r}}\right)}{\left(\frac{T}{w^2} + \frac{1}{\sigma^2_{\beta_r}}\right)} \text{ and } \sigma^2_{\beta_{ir}} = \left(\frac{T}{w^2} + \frac{1}{\sigma^2_{\beta_r}}\right)^{-1}$$

The priors of μ_{β_r} and $\sigma^2_{\beta_r}$ are the following

$$\mu_{\beta_r} \sim N(\mu_{b_0}, \sigma^2_{b_0})$$

$$\sigma^{-2}_{\beta_r} \sim Gamma(a_0, b_0)$$

4.4 Modeling Dependence among ROIs

We now return to model in Equation 4.2 and focus on the model for the region level random effects b_r. Markov random field (MRF) models has been used extensively in image analysis. It is applied towards image correction and pattern detection.

4.4.1 Grids and Lattices

The lattice is a multidimensional object on which spatial structures may be constructed. [Besag, 1974] introduced such spatial structures in plant ecology and agricultural experiments. Locations of the pixels are random and these pixels are related by spatial proximity. However,

the neighborhood relation can certainly have other connection beside geographical proximity. [Marin and Robert, 2014] argued probabilistic dependence between two points of lattice can be interpreted as the neighborhood relationship.

The lattice is associated with an image, which is a $N \times M$ array made of pairs (i, j) where $i = 1, \cdots, N,\ j = 1, \cdots, M$. The adjacent neighbors can be four entries such as $(i, j-1)$, $(i, j+1)$, $(i-1, j)$ and $(i+1, j)$. Let us assume, $i \in \mathcal{I}$. Here $\sim$ denotes neighborhood relationship on a lattice $\mathcal{I}$. $i \sim j$, where $i, j \in \mathcal{I}$ are considered to be neighbors. We can associate these lattice $\mathcal{I}$ with probability distribution on z, where $z = (z_i)_{i \in \mathcal{I}}$. Two components z_i and z_j are correlated if the sites $i \sim j$. The neighborhood relation is symmetric if $i \sim j$ then $j \sim i$ and i is not a neighbor of itself. Three common types of neighborhood structure on a regular grid are illustrated in Figure 15.

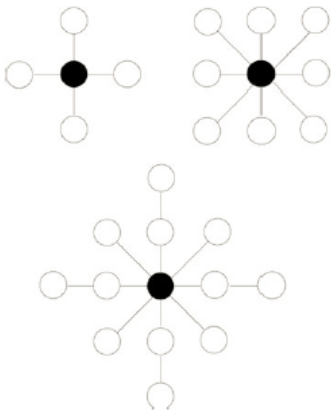


Figure 15: Common neighborhood structures with four neighbors, eight neighbors and twelve neighbors in imaging.

(Source: [Marin and Robert, 2014])

4.4.2 The Markov Random Field

A random field can be specified on a lattice $\mathcal{I}$ and there exists random variables $\{z_i : i \in \mathcal{I}\}$, where each random variable z_i belongs to finite set $\mathcal{X}$ and z_i's may have dependent neighborhood structure on $\mathcal{I}$. Let $n(i)$ denote the set of neighbors of region i and $z_{n(i)}$ is the collection of values taken by the neighborhood region of i. Such a random field is called a Markov random field (MRF) if the conditional distribution of region i given the other regions only depends on the neighbors of region $i \in \mathcal{I}$ i.e

$$\pi(z_i|z_{-i}) = \pi(z_i|z_{n(i)}) \tag{4.6}$$

Markov random field has been used avidly in imaging because of its stabilizing properties in modeling. It is often difficult to directly characterize the global properties of an image through a probability distribution which makes construction of joint prior distribution extremely complication. The simple way around this problem is to use the full conditional distribution, which concentrates on properties of z_i when the neighborhood structure is known.

Let us assume, on a lattice $\mathcal{I} = \{1, \ldots, n\}$ the joint density can be written as the following

$$\pi(z) = \pi(z_1|z_{-1})\pi(z_{-1})$$

and for a fixed value z^* we can rewrite the expression as

$$\pi(z) = \frac{\pi(z_1|z_{-1})}{\pi(z_1^*|z_{-1})}\pi(z_1^*, z_{-1})$$

The representation for all terms in the lattice is the following

$$\frac{\pi(z)}{\pi(z^*)} = \prod_{i=0}^{n-1} \frac{\pi(z_{i+1}|z_1^*, \ldots, z_i^*, z_{i+2}, \ldots, z_n)}{\pi(z_{i+1}^*|z_1^*, \ldots, z_i^*, z_{i+2}, \ldots, z_n)} \tag{4.7}$$

This enables to express the joint density as a product of ratios of its full conditionals. This representation is known as Hammersley-Clifford theorem [Marin and Robert, 2014]. Equation 4.7 must follow every representation of a lattice $\mathcal{I}$ if there exists a joint density such that full conditionals are not canceling and for every fixed value z^*.

A clique is a maximal subset of a lattice $\mathcal{I}$ consists of regions which are all neighbors. The joint distribution can be expressed in terms of its system of cliques [Cressie, 1993].

$$\pi(x) \propto exp(-\sum_{C \in \mathcal{C}} \Phi_C(x_C)) \tag{4.8}$$

where $\mathcal{C}$ is the collection of all cliques.

We now review specific MRFs such as Ising model, the Potts model and the hidden Potts model.

4.4.2.1 The Ising Model

The Ising model has been extensively used for binary images and other spatial models. The random variable z_i can take binary values. The distribution of a region is conditional on its neighboring region is assumed to be Bernoulli with

$$\pi(z_i = j|z_{-i}) \propto exp(\beta n_{i,j}), \quad \beta > 0, \tag{4.9}$$

where $n_{i,j} = \sum_{l \in n(i)} (\mathbb{I}_{z_l} = j)$ is the number of neighbors of region z_i and $n(i)$ denotes the neighbors of region i. The full conditionals of Ising model is

$$\pi(z_i = 1 | z_{-i}) = \frac{\exp(\beta n_{i,1})}{\exp(\beta n_{i,0}) + \exp(\beta n_{i,1})}$$

and the joint distribution therefore satisfies

$$\pi(z) \propto \exp\left(\beta \sum_{j \sim i} \mathbb{I}_{z_j = z_i}\right), \tag{4.10}$$

where the summation is taken over all pairs (i, j) of neighbors [Marin and Robert, 2014].

The normalizing constant $Z(\beta)$ of the joint probability distribution in Equation 4.10 is intractable in most of the cases except for very small lattices $\mathcal{I}$, which makes statistical inference about β problematic, for example, using Markov chain sampling. **Algorithm 4.1** [Marin and Robert, 2014] assumes fixed β for the Gibbs sampler.

Algorithm 4.1 Isling Gibbs sampler

- Initialize: $i \in \mathcal{I}$ and generate $z_i^{(0)} \sim Bernoulli(0.50)$.
  For iteration (≥ 1):
- Generate $u = \{u_i\}_{i \in \mathcal{I}}$ to randomly order the components of $\mathcal{I}$.
- For $l = 1, \cdots, |\mathcal{I}|$, update $n_{u_l,0}^{(t)}$ and $n_{u_l,1}^{(t)}$ and generate

$$z_{u_l}^{(t)} \sim Bernoulli\left(\frac{exp(\beta n_{u_l,0}^{(t)})}{exp(\beta n_{u_l,0}^{(t)}) + exp(\beta n_{u_l,1}^{(t)})}\right)$$

The selection of β can be crucial. Gibbs sampler encounters difficulty to alter the value of a single pixel if β is large.

4.4.2.2 The Potts Model

The Potts model is a generalization of the Ising model. There are G ($G > 2$) regions and for g ($g = 1, \cdots, G$).Numbers of neighbors of i is $n_{i,g} = \sum_{j \sim i} \mathbb{I}_{z_j = g}$. In the Potts model, the full conditional distribution of z_i satisfies

$$\pi(z_i = g | z_{-i}) \propto \exp(\beta n_{i,g})$$

The joint distribution of the Potts model can be described as [Marin and Robert, 2014]

$$\pi(z) \propto \exp\left(\beta \sum_{j \sim i} \mathbb{I}_{z_j = z_i}\right) \tag{4.11}$$

where, the normalizing constant of the joint distribution is intractable. The following Metropolis-Hastings Sample [Marin and Robert, 2014] can be used for Markov chain sampling from the Potts model with fixed β.

Algorithm 4.2 Potts Metropolis-Hastings Sampler

- Initialize: $i \in \mathcal{I}$ and generate $z_i^{(0)} \sim \mathcal{U}(1, \ldots, G)$.
  For iteration (≥ 1):
- Generate $u = \{u_i\}_{i \in \mathcal{I}}$ to randomly order the components of $\mathcal{I}$.
- For $l = 1, \cdots, |\mathcal{I}|$ generate

$$\tilde{z}_{u_l} \sim \mathcal{U}(1, 2, \ldots, z_{u_l}^{(t-1)} - 1, z_{u_l}^{(t-1)} + 1, \ldots, G)$$

  and compute $n_{u_l,g}^{(t)}$ and compute the probability

$$\rho_l = \exp(\beta n_{u_l, \tilde{z}_{u_l}}) / \exp(\beta n_{u_l, z_{u_l}}^{(t)}) \wedge 1$$

  and $\tilde{z}_{u_l}$ is equal to z_{u_l} with probability ρ_l.

4.4.3 The Hidden Potts Model

Spatial dependence between adjacent neighbours can be modeled by a hidden Potts model [Potts, 1952] which is a latent MRF on discrete states. The hidden Potts model is a hierarchical model where there are observed $y_i \in \mathbb{R}$, $i = 1, \cdots, n$ and finite set of discrete states z, where $z \in \{1, \ldots, k\}$. The Potts model is defined on these z's. The likelihood equation that relates observed pixel values with corresponding latent labels will be

$$p(y|z, \theta) = \prod_{i=1}^{n} p(y_i|z_i, \theta_{z_i}) \tag{4.12}$$

here θ_{z_i} are the parameter that regulate the distribution of the y_i with label z_i.

The conditional probabilities of the labels z_i's follow a Potts model as

$$p(z_i|z_{-i},\beta) = \frac{\exp(\beta \sum_{i\sim l} \mathbb{I}(z_i, z_l))}{\sum_{j=1}^{k} \exp(\beta \sum_{i\sim l} \mathbb{I}(j, z_l))}, \tag{4.13}$$

where β is often termed as inverse temperature, [Geman and Geman, 1984], [Alston et al., 2007] and many other assume that y_i with latent label j follows a Gaussian distribution with a common mean μ_j and variance σ_j^2.

$$(y_i|z_i = j, \mu_j, \sigma_j^2) \sim Normal(\mu_j, \sigma_j^2) \tag{4.14}$$

There is a sufficient statistic for this model as Normal distribution is a member of the exponential family. The sufficient statistics given by [Grelaud et al., 2009] is $S(z) = \sum_{i\sim l} I(z_i, z_l))$, which represents the total number of neighbors that share same label. This enables us to factorize the likelihood $p(y, z|\theta, \beta)$ into $p(y|z, \theta)p(S(z)|\beta)$. The posterior is the following

$$\begin{aligned} p(\theta, \beta, z|y) &\propto p(y|z, \theta)\pi(\theta)p(S(z)|\beta)\pi(\beta) \\ &\propto p(y|z, \theta)\pi(\theta)\frac{p(S(z)|\beta)\pi(\beta)}{\mathcal{C}(\beta)} \end{aligned} \tag{4.15}$$

where $\pi(\theta)$ is the joint prior for θ and $\pi(\beta)$ is the prior for the inverse temperature β. The posterior normalizing constant $\mathcal{C}(\beta) = \sum_{z\in Z} exp\{\beta S(z)\}$ is intractable as it involves a sum over all available combinations of labels with computational complexity of $\mathcal{O}(nk^n)$. Therefore, sampling from the posterior distribution has no exclusive solution.

The conditional expectation of $S(z)$ given β can also be expressed in terms of normalizing constant

$$E_{z|\beta}(S(z)) = \frac{d}{d\beta}\log\{\mathcal{C}(\beta)\} \tag{4.16}$$

All y_i are almost surely allocated to same label if $\beta \to \infty$ as expectation of $S(z)$ converges asymptotically to the total number of combinations of labels and the variance approaches to zero. The labels z_i are independent and uniformly-distributed when $\beta = 0$.

[Potts, 1952] showed that the model transition from disordered to an ordered state at a certain critical value of β. For a regular 2D lattice, the critical value can be calculated as following

$$\beta_{crit} = \log\{1 + \sqrt{k}\} \tag{4.17}$$

4.4.3.1 Bayesian Computational Methods for Hidden Potts Model

We provided a review of major algorithms [Moores et al., 2018] used for hidden Potts model. These provides alternative methods to simulate parameter value β without computing the normalizing constant. These methods include Markov Chain Monte Carlo (MCMC) with pseudolikelihood and Approximate Bayesian Computation(ABC)-MCMC algorithm.

Pseudolikelihood is the fastest and easiest method to deal with intractable likelihood in Image analysis. According to, [Rydén and Titterington, 1998] the intractable distribution could be approximated using product of conditional density given by Equation 4.13.

$$p(z|\beta) \approx \prod_{i=1}^{n} p(z_i|z_{-i}, \beta) \tag{4.18}$$

The conditional distributions for each pixels are being computed individually. The algorithm is following

Algorithm 4.3 Hidden Potts Model MCMC with pseudolikelihood

- Initialize β_0, μ_0, σ_0^2, z_0.
  For iteration $t(t \geq 1)$:
- Update the labels: $z_i \sim p(y_i|z_i, \mu_{z_i}, \sigma_{z_i}^2)p(z_i|z_{-i}),\ \ \forall i \in \{1, \ldots, n\}$.
- Calculate the sufficient statistics: $\bar{y_j}$, s_j^2, $\forall z_i = j$, $\forall j \in \{1, \ldots, k\}$.
- Update the noise parameters $\mu_j, \sigma_j^2 \sim p(y_j, s_j^2|\mu_j, \sigma_j^2)\pi(\mu_j|\sigma_j^2)\pi(\sigma_j^2)$.
- Draw proposed parameter value $\beta' \sim q(\beta'|\beta_{t-1})$.
- Approximate $p(z|\beta')$ and $p(z|\beta_{t-1})$ by using : $p(z|\beta) \approx \prod_{i=1}^{n} p(z_i|z_{-i}, \beta)$.
- Calculate the Metropolis-Hastings ratio $\rho = min\left(1, \frac{p(z|\beta')\pi(\beta')q(\beta_{t-1}|\beta'))}{p(z|\beta_{t-1})\pi(\beta_{t-1})q(\beta'|\beta_{t-1}))}\right)$.
- Draw $u \sim Uniform(0, 1)$ if $u < \rho$ then $\beta_t = \beta'$. Else $\beta_t = \beta_{t-1}$.

In ABC algorithm, an auxiliary variable w is introduced to determine whether to accept or reject the proposed value of β'. The summary statistics of the auxiliary variable $S(w)$ and observed data $S(z)$ are compared [Grelaud et al., 2009]. The proposal is accepted if $\parallel S(w) - S(z) \parallel \leq \epsilon$. The labels z are observed without error if the following approximation is true.

$$p(\beta|y, z, \theta) \approx \pi_\epsilon(\beta| \parallel S(w) - S(z) \parallel \leq \epsilon) \tag{4.19}$$

The ABC approximation Equation 4.19 approaches to the true posterior as $n \to \infty$ and $\epsilon \to 0$. The Hidden Potts Model ABC-MCMC algorithm follows.

Algorithm 4.4 Hidden Potts Model ABC-MCMC

- Initialize β_0, μ_0, σ_0^2, z_0.
  For iteration $t(t \geq 1)$:
- Update the labels: $z_i \sim p(y_i|z_i, \mu_{z_i}, \sigma_{z_i}^2)p(z_i|z_{-i}),\ \ \forall i \in \{1, \ldots, n\}$.
- Calculate the sufficient statistics: $\bar{y_j}$, s_j^2, $\forall z_i = j$, $\forall j \in \{1, \ldots, k\}$.
- Update the noise parameters $\mu_j, \sigma_j^2 \sim p(y_j, s_j^2|\mu_j, \sigma_j^2)\pi(\mu_j|\sigma_j^2)\pi(\sigma_j^2)$.
- Draw proposed parameter value $\beta' \sim q(\beta'|\beta_{t-1})$.
- Generate $w|\beta'$ by sampling from Equation 4.13.
- Draw $u \sim Uniform(0, 1)$.
- if $u < \frac{\pi(\beta')q(\beta_{t-1}|\beta')}{\pi(\beta_{t-1})q(\beta'|\beta_{t-1})}$ and $\parallel S(w) - S(z) \parallel \leq \epsilon$ then $\beta_t = \beta'$. Else $\beta_t = \beta_{t-1}$.

4.5 A Region-temporal Model for Functional Connectivity

As we describe before, we consider resting state fMRI measurements Y_{itr} for $i = 1, \cdots, n$ subjects, at $t = 1, \cdots, T$ are discrete time points, and $r = 1, \cdots, R$ regions We model Y_{itr} as following

$$Y_{itr} = G(\theta_{itr}, b_r) + \varepsilon_{itr}, \tag{4.20}$$

where θ_{itr} is the temporal component of the model and b_r is the region-level random effect to brings in potential correlation between connected regions. We assume that errors $\varepsilon_{itr} \overset{i.i.d}{\sim} N(0, \sigma^2)$.

As we describe in the previous sections, we utilize dynamic Linear Model (DLM) to model the temporal component θ_{itr}.

$$\begin{aligned} \text{Observational Equation} \quad & Y_{itr} = G(\theta_{itr}, b_r) + \varepsilon_{itr} \\ \text{System Equation of Temporal Component} \quad & \theta_{itr} = \beta_{ir}\theta_{i(t-1)r} + \eta_{itr} \end{aligned} \tag{4.21}$$

where, $\eta_{itr} \overset{\text{i.i.d}}{\sim} N(0, w^2)$ is the error term of system equation and the coefficient β_{ir} depends on subject i and region r. We assume a normal prior for $\beta_{ir} \sim N(\mu_{\beta_r}, \sigma^2_{\beta_r})$, where the hyperpriors μ_{β_r} and $\sigma^2_{\beta_r}$ are shared across subjects and specify the initial distribution as $\theta_{i0r} \sim N(m_0, C_0)$, for all i and r.

The region-level random effects $\{b_r\}_{r=1}^R$ are modeled using a hidden Potts model which creates stochastic clustering among $\{b_r\}_{r=1}^R$. Regions r which are in the same cluster share identical values of b_r. Let z_r denote the (random) cluster label of b_r. The random effects $\{b_r\}_{r=1}^R$ can then be labeled as $\{b^*_{z_1}, \cdots, b^*_{z_R}\}$, where $b^*_1, b^*_2, \ldots$ denote the unique values associated with the clusters.

As we have described in the previous sections we assume a Potts model on the cluster label $\{z_r\}_{r=1}^R$. In particular, the conditional probabilities of z_r is given as follows

$$p(z_r | z_{-r}, \beta) \propto \exp\{-\beta \sum_{r \sim l} \delta(z_r, z_l)\}, \tag{4.22}$$

The regions or ROI's which are in the same cluster c thus share the random effect b_c^* and are thus dependent or functionally connected. This clustering is stochastic as the Potts model provides a probability distribution over the space of all possible clustering.

4.5.1 Posterior Computation for Region-temporal Model

The likelihood equation that relates observed values with corresponding latent labels z_r and random effect $b_{z_r}^*$ can be expressed as

$$p(Y_{itr}|z, b_z^*, \theta) = \prod_{r=1}^{R} p(Y_{itr}|z_r, b_{z_r}^*, \theta_{z_r}), \tag{4.23}$$

here θ_{z_r} are the parameter that regulate the distribution of the pixel values with label z_r. We are assuming that Y_{itr} with label j has a common mean μ_j and noise σ_j^2.

$$(Y_{itr}|z_r = j, b_{z_r}^* = b_j^*, \mu_j, \sigma_j^2) \sim Normal(\mu_j, \sigma_j^2) \tag{4.24}$$

The conditional distribution of random effect b_r is the following

$$p(b_r|Y_{itr}, z_r, \theta_{z_r}) \propto \pi(b_r) \prod_{\{h:b_{z_h}^* = b_{z_j}^*\}} p(Y_{ith}|z_h, \theta_{z_h}, b_{z_h}^*), \tag{4.25}$$

where $\pi(b_r)$ is the prior for b.

$$\begin{aligned} p(\theta, \beta, z, b_z^*|y) &\propto p(y|z, b_z^*, \theta)\pi(\theta)\pi(b_z^*)p(S(z)|\beta)\pi(\beta) \\ &\propto p(y|z, b_z^*, \theta)\pi(\theta)\pi(b_z^*)\frac{p(S(z)|\beta)\pi(\beta)}{\mathcal{C}(\beta)} \end{aligned} \tag{4.26}$$

where $\pi(\theta)$ is the joint prior for θ and $\pi(\beta)$ is the prior for the inverse temperature β. $S(z) = \sum_{i\sim l}\delta(z_i, z_l))$, which represents the total number of neighbors that share same label. The posterior normalizing constant $\mathcal{C}(\beta) = \sum_{z\in Z} exp\{\beta S(z)\}$ is intractable as it involves a sum over all available combinations of labels with computational complexity of $\mathcal{O}(nk^n)$. Therefore, sampling from the posterior distribution has no exclusive solution. We use conjugate priors for $\mu \sim N(\mu_0, \sigma_0^2)$ and $\sigma^2 \sim$ Inverse Gamma(A_s, B_s).

The Markov chain sampling for the joint modeling method for dynamic linear model and hidden Potts model with latent variable with fixed β and pseudolikelihood approach can be described as the following

$\theta_{\mathbf{itr}}|\mathbf{b_r}$ can be updated by applying one-step forecast and the posterior distribution on the general univariate DLM for each timepoint t (**Theorem 4.3.1**).

- Updating $\theta_{\mathbf{itr}}|\mathbf{b_r}$ and $\beta_{\mathbf{ir}}$ for every individual i and region r

  1. **Initialize:** m_{i0r}, C_{i0r}.
  2. **Posterior for** $\theta_{i(t-1)r}$**:** $(\theta_{i(t-1)r}|D_{i(t-1)r}) \sim N[m_{i(t-1)r}, C_{i(t-1)r}]$.
  3. **Prior for** θ_{itr} **:** $(\theta_{itr}|D_{i(t-1)r}) \sim N[a_{itr}, R_{itr}]$,

     where $a_{itr} = \beta_{ir} m_{i(t-1)r}$ and $R_{itr} = \beta_{ir}^2 C_{t-1} + w^2$.
  4. **1-step forecast :** $(Y_{itr}|D_{i(t-1)r}) \sim N[f_{itr}, Q_{itr}]$,

     where $f_{itr} = a_{itr}$ and $Q_{itr} = R_{itr} + \sigma^2$.

5. **Posterior for** θ_{itr} **:** $(\theta_{itr}|D_{itr}) \sim N[m_{itr}, C_{itr}]$,

   with $m_{itr} = a_{itr} + A_{itr}e_{itr}$ and $C_{itr} = R_{itr} - A_{itr}^2 Q_{itr}$,

   where $A_{itr} = R_{itr}/Q_{itr}$, and $e_{itr} = Y_{itr} - f_{itr}$.

6. **Posterior for** β_{ir} **:** $(\beta_{ir}|\cdot) \sim N(\mu_{\beta_{ir}}, \sigma^2_{\beta_{ir}})$

   with $\mu_{\beta_{ir}} = \dfrac{\left(\frac{\sum_{t=1}^{T}\theta_{itr}}{w^2} + \frac{\mu_{\beta_r}}{\sigma^2_{\beta_r}}\right)}{\left(\frac{T}{w^2} + \frac{1}{\sigma^2_{\beta_r}}\right)}$ and $\sigma^2_{\beta_{ir}} = \left(\frac{T}{w^2} + \frac{1}{\sigma^2_{\beta_r}}\right)^{-1}$.

$\mathbf{b_r}|\theta_{\mathbf{itr}}$ can be updated using two different hidden Potts based methods. The first method depends on fixing β values to avoid the intractable likelihood altogether. The second method relies on Markov chain Monte Carlo with pseudolikelihood. We consider similar algorithm for MCMC with Pseudolikelihood described in [Moores et al., 2018].

Model 1: Hidden Potts model with latent variable using fixed β values

- Updating $\mathbf{b_r}|\theta_{\mathbf{itr}}$

  1. Set a fixed β value.
  2. Initialize μ_0, σ_0^2, z_0.
  3. Collapse the data as conditional on θ, the averages $\overline{d}_{\cdot\cdot r} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}(Y_{itr} - \theta_{itr})$ are sufficient statistics.

     For iteration(≥ 1):

  4. Update random effect b_r: $b_r \sim \pi(b_r)\prod_{\{h:b^*_{z_h}=b^*_{z_j}\}} p(\overline{d}_{\cdot\cdot h}|z_h, b^*_{z_h})$.
  5. Update the labels: $z_r \sim p(\overline{d}_{\cdot\cdot r}|z_r, b_r, \mu_{z_r}, \sigma^2_{z_r})p(z_r|z_{-r}, \beta)$, $\forall r \in \{1, \ldots, R\}$.
  6. Update parameters of b_r: $\mu_b, \sigma_b^2 \sim p(b_r|\mu_b, \sigma_b^2)\pi(\mu_b|\sigma_b^2)\pi(\sigma_b^2)$.

7. Calculate the sufficient statistics: $\overline{d}_{\cdot\cdot j}$, s_j^2, $\forall z_r = j$, $\forall j \in \{1, \ldots, k\}$.

8. Update the noise parameters $\mu_j, \sigma_j^2 \sim p(\overline{d}_{\cdot\cdot j}, s_j^2 | \mu_j, \sigma_j^2)\pi(\mu_j|\sigma_j^2)\pi(\sigma_j^2)$.

Model 2: Hidden Potts model with latent variable using pseudolikelihood approach

- Updating $\mathbf{b_r}|\theta_{\mathbf{itr}}$

1. Initialize β_0, μ_0, σ_0^2, z_0.

2. Collapse the data as conditional on θ, the averages $\overline{d}_{\cdot\cdot r} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}(Y_{itr} - \theta_{itr})$ are sufficient statistics.

   For iteration(≥ 1):

3. Update random effect b_r: $b_r \sim \pi(b_r)\prod_{\{h:b^*_{z_h}=b^*_{z_j}\}} p(\overline{d}_{\cdot\cdot h}|z_h, \theta_{z_h}, b^*_{z_h})$.

4. Update the labels: $z_r \sim p(\overline{d}_{\cdot\cdot r}|z_r, b_r, \mu_{z_r}, \sigma^2_{z_r})p(z_r|z_{-r}, \beta)$, $\forall r \in \{1, \ldots, R\}$.

5. Update random effect b_r: $b_r \sim \pi(b_r)\prod_{\{h:b^*_{z_h}=b^*_{z_j}\}} p(\overline{d}_{\cdot\cdot h}|z_h, b^*_{z_h})$.

6. Calculate the sufficient statistics: $\overline{d}_{\cdot\cdot j}$, s_j^2, $\forall z_r = j$, $\forall j \in \{1, \ldots, k\}$.

7. Update the noise parameters $\mu_j, \sigma_j^2 \sim p(\overline{d}_{\cdot\cdot j}, s_j^2 | \mu_j, \sigma_j^2)\pi(\mu_j|\sigma_j^2)\pi(\sigma_j^2)$.

8. Draw proposed parameter value $\beta' \sim q(\beta'|\beta_{t-1})$.

9. Approximate $p(z|\beta')$ and $p(z|\beta_{t-1})$ by using : $p(z|\beta) \approx \prod_{r=1}^{R} p(z_r|z_{-r}, \beta)$.

10. Calculate the Metropolis-Hastings ratio $\rho = min\left(1, \frac{p(z|\beta')\pi(\beta')q(\beta_{t-1}|\beta'))}{p(z|\beta_{t-1})\pi(\beta_{t-1})q(\beta'|\beta_{t-1}))}\right)$.

11. Draw $u \sim Uniform(0, 1)$ if $u < \rho$ then $\beta_t = \beta'$. Else $\beta_t = \beta_{t-1}$.

4.6 Performance Evaluation

We evaluate the performances of three different clustering methods: Blocked Dirichlet process, hidden Potts model with latent variable (fixed inverse temperature) and hidden Potts model with latent variable (pseudolikelihood approach). We describe Blocked Dirichlet process in Equation 3.19. Each simulation study evaluates performances of the clustering methods over 100 replicated data sets. To keep the Bayesian computational complexity over these replicated data sets manageable, each data set considers $R = 50$ regions. The first 30 of regions are specified to be inter-connected and remaining regions are non-connected. We assumed a AR(2) structure for temporal part. A sample size of $n = 20$ individuals and $T = 30$ time points has been considered.

4.6.1 Data Generation Model

The data sets are generated from the following model

$$Y_{itr} = G(\theta_{itr}, b_r) + \epsilon_{itr}$$

$$\theta_{itr} = \beta_1 \theta_{i(t-1)r} + \beta_2 \theta_{i(t-2)r}$$

where $\epsilon_{itr} \sim N(0, 0.5^2)$ and $i = 1, \cdots, 20,\ t = 1, \cdots, 30,\ r = 1, \cdots, 50$.

Out of 30 connected regions, we have assumed that there exists total of 9 clusters with different sizes and different levels of connectivity: cluster 1=$\{1, \ldots, 5\}$ and cluster 2 = $\{6, \ldots, 10\}$ are assumed to be strongly connected, cluster 3=$\{11, \ldots, 14\}$, 4=$\{15, \cdots, 17\}$ and 5=$\{19, \cdots, 22\}$ are moderately connected. Then we have clusters 6, 7, 8 and 9 with only 2 regions each: $23-24$,

$25-26$, $27-28$ and $29-30$. The first two clusters are specified to be moderately connected and last two are highly connected.

4.6.2 Results from Clustering Methods

We perform a joint modeling method by applying dynamic linear model with three clustering methods to assess performance of each model.

$\theta_{\mathbf{itr}}|\mathbf{b_r}$ is updated by applying one-step forecast and the posterior distribution on the general univariate DLM for each time point t (**Theorem 4.3.1**). The hyperparameters for the DLM is specified as following

$$\mu_{\beta_r} \sim \mathrm{N}(0.1, 0.05^2), \quad \sigma^2_{\beta_r} \sim \text{Inverse-Gamma}(10, 1)$$

$\mathbf{b_r}|\theta_{\mathbf{itr}}$ is updated using DPM and two different hidden Potts based models. We initialized the numbers of clusters to be 30 for all three clustering methods. For both hidden Potts models with latent variable, we are assuming inverse temperature to be $\beta = 1.25$.

We follow **Algorithm 3.9** for blocked Gibbs sampling method based on stick-breaking prior. For Blocked DPM model, the hyperparameters are the following

$$\alpha \sim \text{Gamma}(10, 2), \quad \sigma^2 \sim \text{Inverse-Gamma}(100, 2)$$

We follow posterior computation in 4.5.1 for hidden Potts model with latent variable using fixed β values. The hyperparameters are described below

$$\mu_b \sim \mathrm{N}(0.1, 0.02^2), \quad \sigma_b^2 \sim \text{Inverse-Gamma}(1, 0.5)$$

$$\sigma^2 \sim \text{Inverse-Gamma}(100, 1)$$

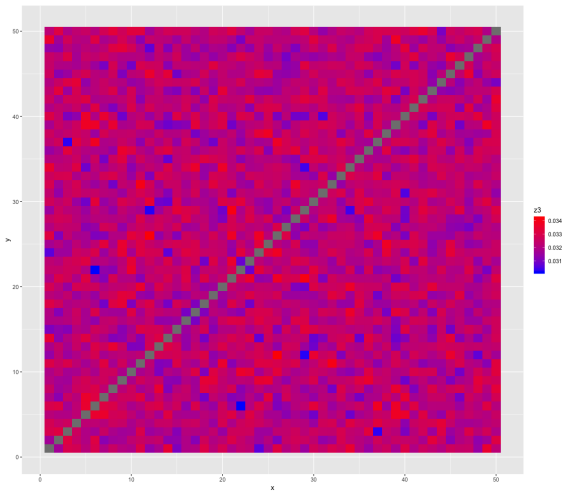
We follow posterior computation in 4.5.1 hidden Potts model with latent variable using pseudolikelihood approach. The hyperparameters are the following

$$\mu_b \sim \mathrm{N}(0.1, 0.2^2), \quad \sigma_b^2 \sim \text{Inverse-Gamma}(5, 2),$$

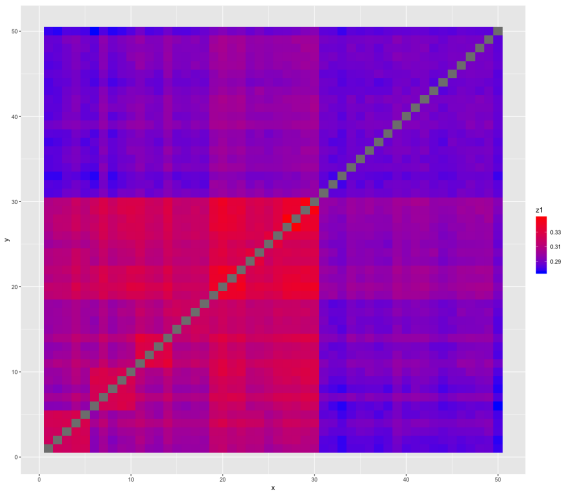
$$\sigma^2 \sim \text{Inverse-Gamma}(100, 1)$$

and the proposed parameter value $\beta' \sim$ Log Normal$(\beta, 0.25^2)$.

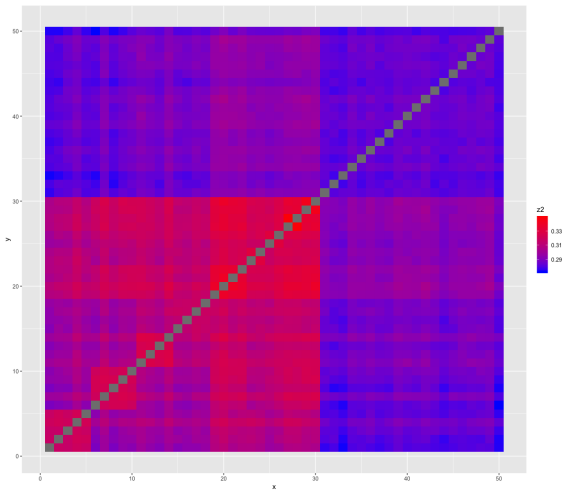
Figure 16 displays the heat map of clustering for joint modeling DLM with blocked Dirichlet process (Figure 16a), DLM with hidden Potts model with latent variable (model 1: fixed inverse temperature) (Figure 16b) and DLM with hidden Potts model with latent variable (model 2: pseudolikelihood approach) (Figure 16c). Figure 16a is not capable to capture the clustering structure of the simulated data. Figure 16b and Figure 16c picked up the clustering structure of the simulated data appropriately. However, we do not see any significant difference in heat maps of hidden Potts model with latent variable (fixed inverse temperature) and hidden Potts model with latent variable (pseudolikelihood approach) methods.



(a) Joint modeling DLM and blocked Dirichlet Prior Process



(b) Joint modeling DLM and hidden Potts model with latent variable (fixed inverse temperature)



(c) Joint modeling DLM and hidden Potts model with latent variable (pseudolikelihood approach)

Figure 16: Simulation Study heat map of clustering

4.7 ABIDE Data Analysis

Autism Brain Image Data Exchange(ABIDE) is an alliance among medical centers across the USA and Europe and is the largest repository of functional and structural brain imaging data for autism. It is a part of 1000 Functional Connectomes Project/ International Neuroimaging Datasharing Initiative (INDI). The data sets are anonymous, with no protected health information according to Health Insurance Portability and Accountability (HIPAA) guidelines and 1000 Functional Connectomes Project / INDI protocols. We have analyzed 5 sites of ABIDE database separately using joint modeling of DLM and hidden Potts with latent variable. There are total of 324 subjects including 177 controls and 147 subjects with ASD (see Table XI). Each of these data sets have different number of time points (see Table XII).

TABLE XI: NUMBER OF SUBJECTS AT EACH SITE FOR ABIDE DATA SET

Site	Control	ASD	Total
California Institute of Technology	19	19	38
New York University	105	79	184
Olin Institute of living	16	20	36
Social Brain Lab	15	15	30
San Diego State University	22	14	36
Total	177	147	324

TABLE XII: NUMBERS OF TIME POINTS AT EACH SITE

Site	Numbers of time points
California Institute of Technology	146
New York University	176
Olin Institute of living	206
Social Brain Lab	196
San Diego State University	176

4.7.1 Results from Analysis using Region-temporal Model

Recall that, Y_{itr} is a rs-fMRI measurement from ith individual at time t from ROI r (where $i = 1, \cdots, n$ $t = 1, \cdots, T$, $r = 1, \cdots, 111$). Y_{itr} is modeled as following

$$Y_{itr} = G(\theta_{itr}, b_r) + \varepsilon_{itr},$$

where θ_{itr} is the temporal component of the model and b_r is a random effect which brings in potential correlation between connected regions. We assume that errors $\varepsilon_{itr} \overset{i.i.d}{\sim} N(0, \sigma^2)$.

We utilize Dynamic Linear Model (DLM) to model the temporal component θ_{itr}. Dynamic Linear Models are state-space models where we specify an auto-regressive structure on the latent parameters θ_{itr}.

$$\text{Observational Equation} \quad Y_{itr} = G(\theta_{itr}, b_r) + \varepsilon_{itr}$$

$$\text{System Equation of Temporal Component} \quad \theta_{itr} = \beta_{ir}\theta_{i(t-1)r} + \eta_{itr}$$

where, β_{ir} depends on subject i and region r . We assume that $\beta_{ir} \sim N(\mu_{\beta_r}, \sigma^2_{\beta_r})$, where the hyperiors $\mu_{\beta_r} \sim N(0.1, 0.05^2)$ and $\sigma^2_{\beta_r} \sim$ Inverse-Gamma$(10, 1)$. The error term for the system equation $\eta_{itr} \overset{iid}{\sim} N(0, 0.5^2)$.

The region-level random effect $\{b_r\}_{r=1}^R$ are modeled using a hidden Potts model with latent variable. We perform joint modeling method to fit DLM and hidden Potts model with latent variable. We follow posterior computation in 4.5.1 for model 1 (hidden Potts model with latent variable) using fixed β values. The hyperparameters are described below

$$\mu_b \sim \text{N}(0.1, 0.02^2), \quad \sigma^2_b \sim \text{Inverse-Gamma}(1, 0.5)$$

$$\sigma^2 \sim \text{Inverse-Gamma}(100, 1)$$

We follow posterior computation in 4.5.1 for in model 2 (hidden Potts model with latent variable using pseudolikelihood approach). The hyperparameters are the following

$$\mu_b \sim \mathrm{N}(0.1, 0.2^2), \quad \sigma_b^2 \sim \text{Inverse-Gamma}(5, 2),$$

$$\sigma^2 \sim \text{Inverse-Gamma}(100, 1)$$

and the proposed parameter value $\beta' \sim \text{Log Normal}(\beta, 0.25^2)$.

The results we present below are based on $50,000$ post burn-in Markov Chain samples from the posterior distribution after an initial burn-in of $25,000$. Here we consider the initial $\beta = 1.1$.

Figure 17 is a trace plot of β value from regional component for ASD subjects of Social Brain Lab using hidden Potts model with latent variable (pseudolikelihood approach) for iteration $25,001$ to $50,000$.

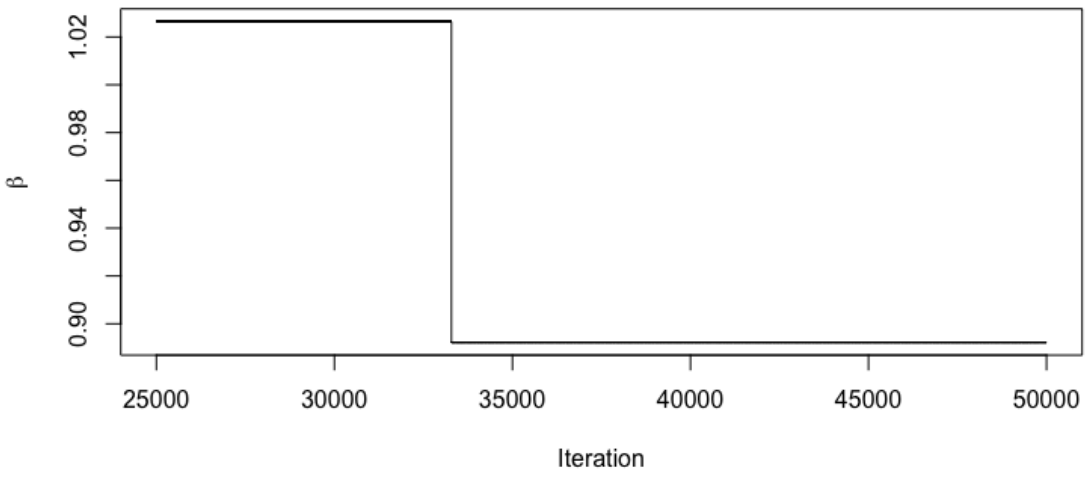
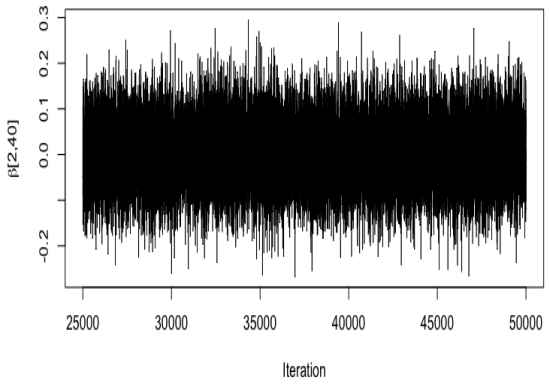
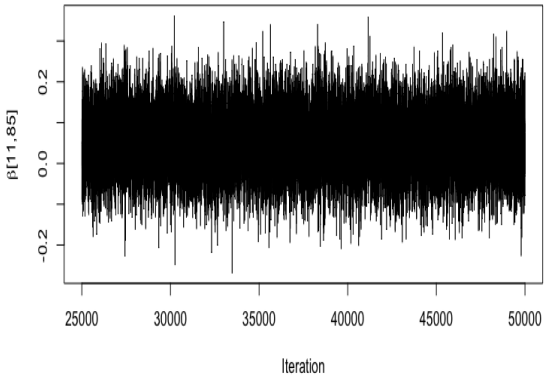


Figure 17: Trace plots for regional component β from ASD subjects of Social Brain Lab using hidden Potts model with latent variable (pseudolikelihood approach)

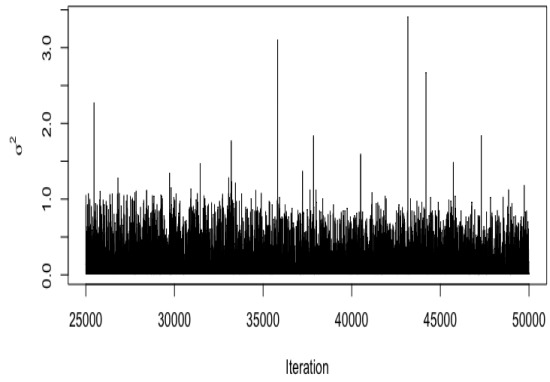
Figure 18 illustrates trace plots for randomly selected temporal components $\beta_{2,40}$ and $\beta_{11,85}$ and σ^2 for iteration $25,001$ to $50,000$.



(a) Trace plot $\beta_{2,40}$



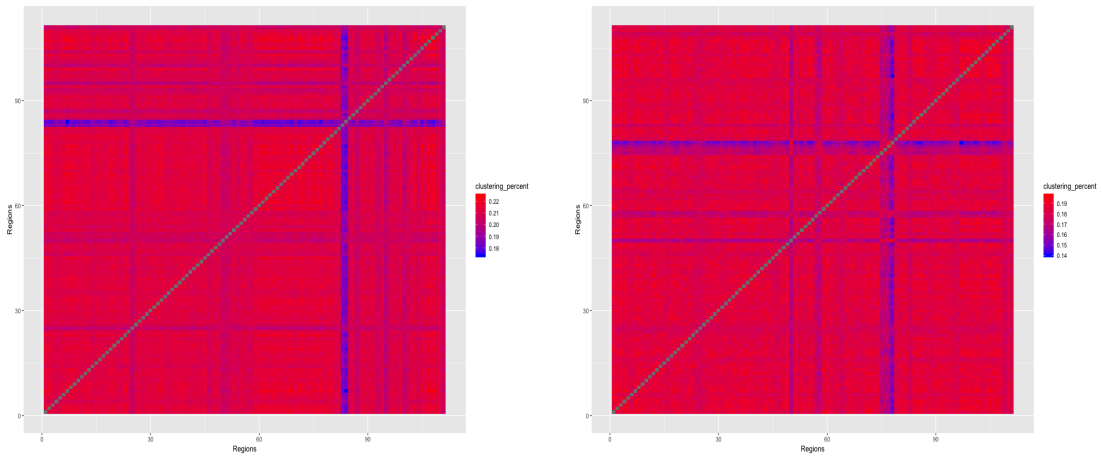
(b) Trace plot $\beta_{11,85}$



(c) Trace plot of σ^2

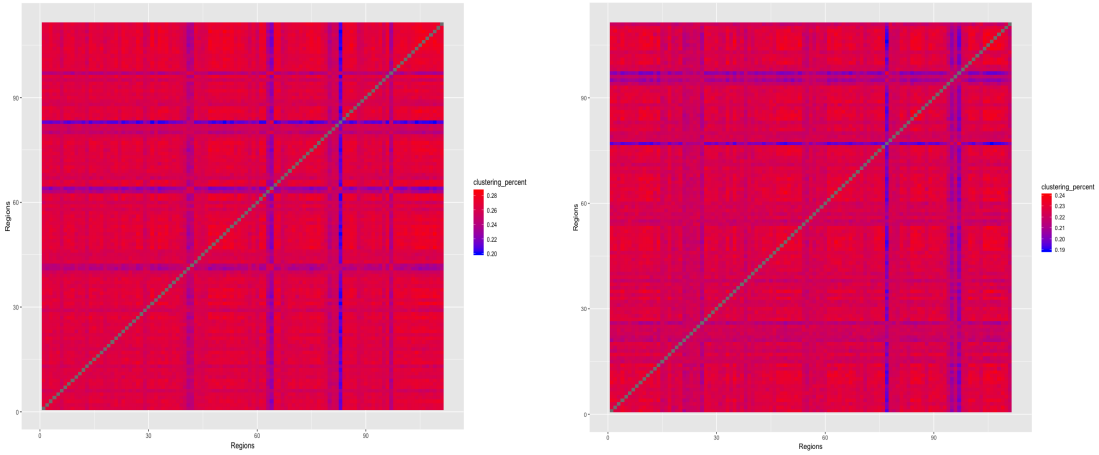
Figure 18: Trace plots for temporal component and σ^2 from ASD subjects of Social Brain Lab using hidden Potts model with latent variable (pseudolikelihood approach)

The heat maps of clustering by joint modeling DLM and hidden Potts model with latent variable for each sites will be displayed in Figures 19-22. The regions of interests (see Table III) are categorized into Frontal lobe, Temporal lobe, Parietal lobe, Occipital lobe, Insular cortex, Cingulate cortex and subcortical grey matter (SCGM). Figures 24-28 illustrate connectivity between regions of interest across lobes of brain for each sites.



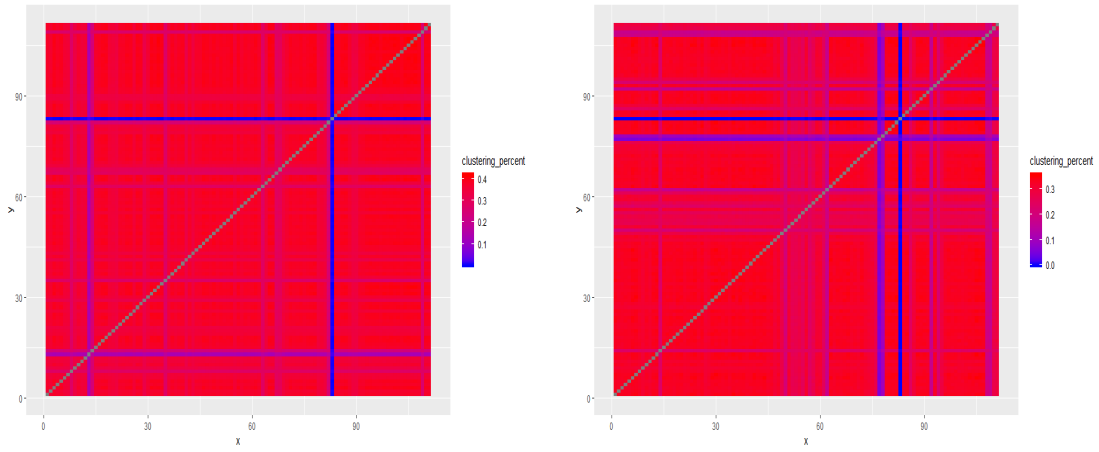
(a) Control (b) ASD Subject

Figure 19: Heat map of clustering of brain regions in California Institute of Technology



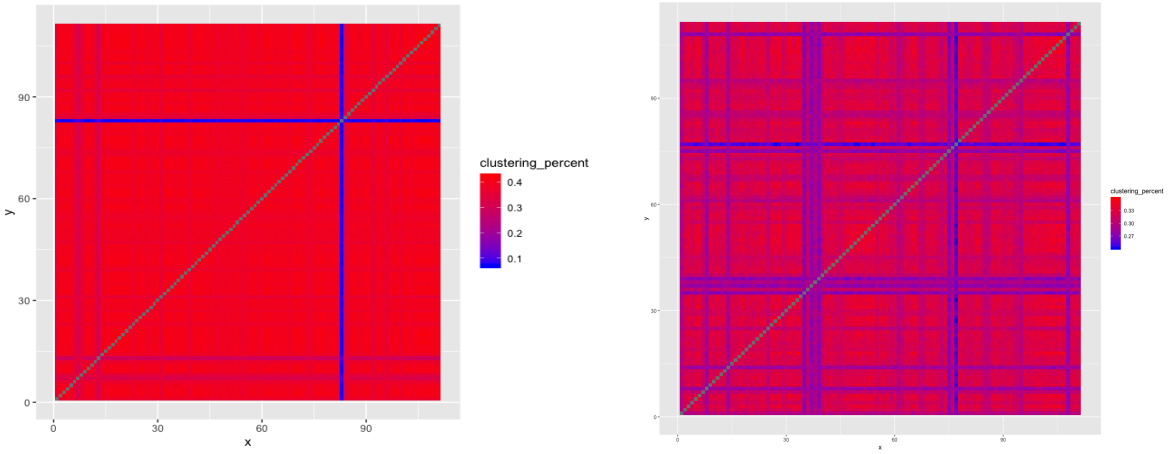
(a) Control (b) ASD subject

Figure 20: Heat map of clustering of brain regions in New York University



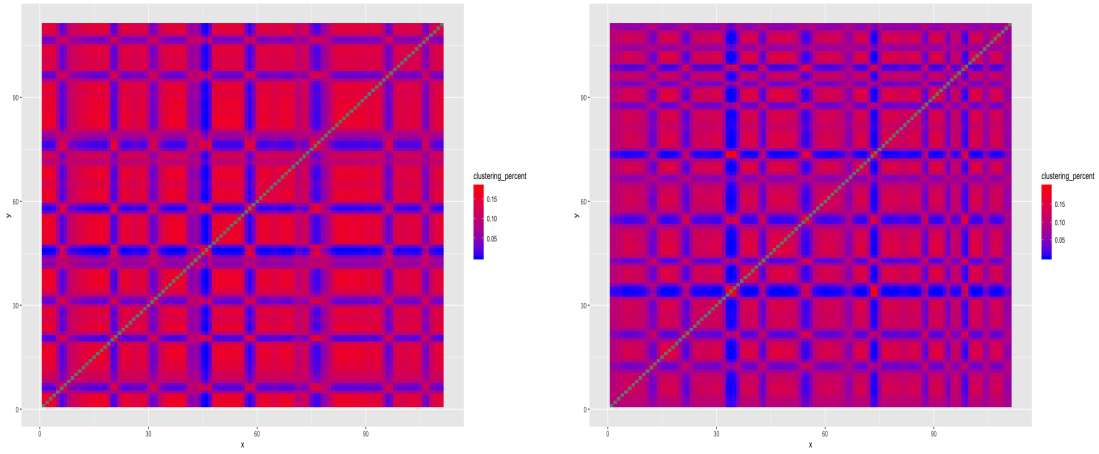
(a) Control (b) ASD subject

Figure 21: Heat map of clustering of brain regions in Olin Institute of living



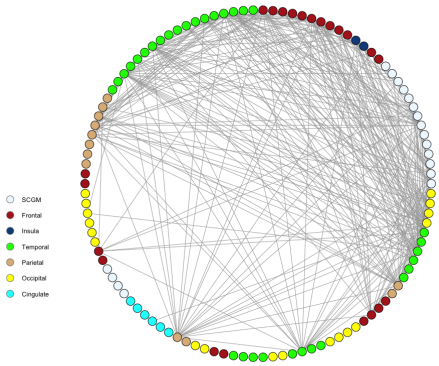
(a) Control (b) ASD subject

Figure 22: Heat map of clustering of brain regions in Social Brain Lab

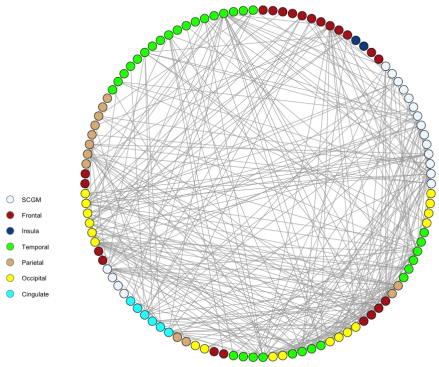


(a) Control (b) ASD subject

Figure 23: Heat map of clustering of brain regions in San Diego State University

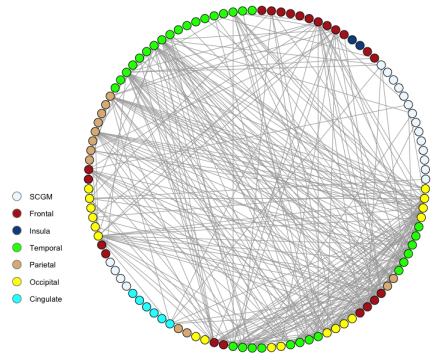


(a) Control

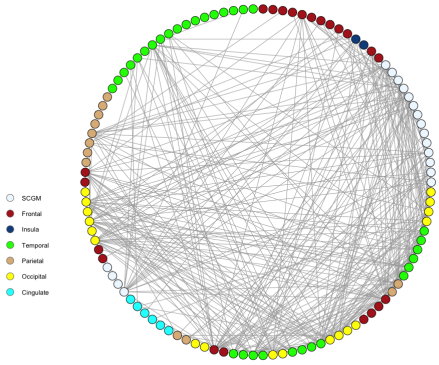


(b) ASD Subject

Figure 24: Connectivity between ROIs across lobes of brain in California Institute of Technology

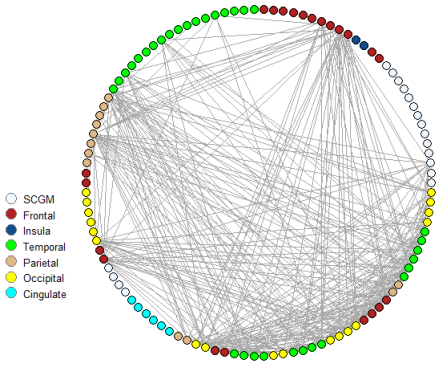


(a) Control

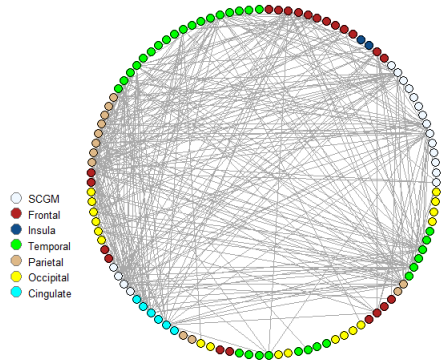


(b) ASD subject

Figure 25: Connectivity between ROIs across lobes of brain in New York University

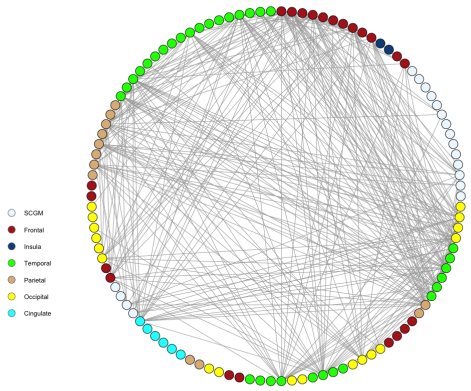


(a) Control

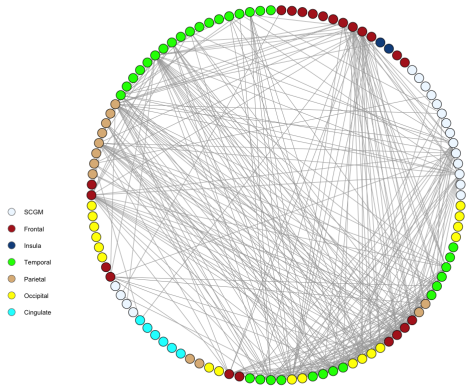


(b) ASD subject

Figure 26: Connectivity between ROIs across lobes of brain in Olin Institute of living

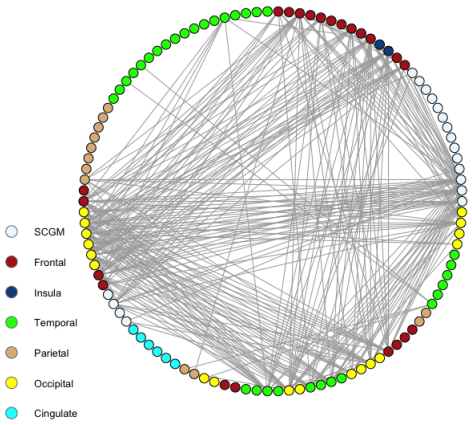


(a) Control

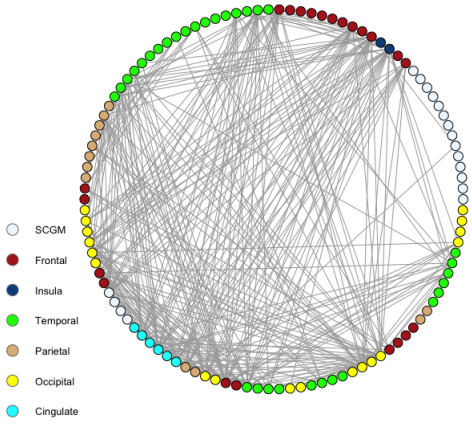


(b) ASD subject

Figure 27: Connectivity between ROIs across lobes of brain in Social Brain Lab



(a) Control



(b) ASD subject

Figure 28: Connectivity between ROIs across lobes of brain in San Diego State University

4.7.2 Conclusion

We analyzed the time-sequence of rs-fMRI at each brain region. We have implemented dynamic linear model to capture temporal structure of the data and the potential correlation between connected regions is modeled using hidden Potts model with latent variable. We notice different patterns in heat map in control and ASD subject. Contrasting connectivity patterns can also be observed across sites. Connectivity between ROIs across lobes of brain suggests that difference in connectivity exists in ASD individuals and control. We can also notice dissimilarity in connectivity between ROIs across lobes of brain over all the sites.

CHAPTER 5

FUTURE WORK AND SUMMARY

5.1 Introduction

Functional connectivity can be determined by considering co-activation of brain region during resting-state fMRI. Connectivity patterns and their disruptions are immensely helpful to understanding mental health disorders and neurodegenerative diseases. The characteristic of altered brain connectivity in Autism Spectrum Disorder has been debatable. There are many theories associated with alterations such as under-connectivity, over-connectivity and a combination of both. Additionally, researchers found age-related changes in functional connectivity in with ASD. These disparities may be associated with heterogeneity in conditions, small sample size, the design of resting-state scan, the pre-processing and data analysis approach.

5.2 Future Work

Many recent papers ([Uddin et al., 2013], [Padmanabhan et al., 2013], [Nomi and Uddin, 2015]) are focused on age-related changes in brain connectivity to understand developmental alternation on functional connectivity in ASD. This may explain the disagreement in hyper- and hypo- connectivity theory.

We perform a preliminary sub group analysis based on cross-correlated functional connectivity data using Bayesian models mentioned in Chapter 3. In the analysis, the subjects are divided into three subgroups based on age such as children, teenager and adults. The subgroups

are: children (less than 13 year), teenagers (ages 13-19) and adults (over 19). We consider the rs-fMRI data from New York University, which is the only site which has subjects from each subgroup.

TABLE XIII: NUMBERS OF SUBJECTS AT EACH SUBGROUP

Age group	Control	ASD	Total
Children	42	43	85
Teenager	33	19	52
Adult	30	17	47

We use the parametric model, Dirichlet process mixture model and neighborhood model on ROI pairs to analyze three subgroups from New York University. The results that we present below are based on $50,000$ post burn-in Markov Chain samples from the posterior distribution of the Bayesian models after an initial burn-in of $25,000$.

TABLE XIV: NUMBERS OF SIGNIFICANT DIFFERENTIAL CONNECTIVITY FOR NEW YORK UNIVERSITY DATA

Age Group	Parametric	DPM	Neighborhood
Children	194	81	108
Teen	459	72	437
Adult	0	0	0

The common differential connectivity shared between children and teenager subgroups are following: (R). Primary Somatosensory Cortex and (R). Dorsal Posterior Cingulate Cortex, (L). Inferior Temporal Gyrus and (L). Cingulate Cortex, (L). Ventral Posterior Cingulate Cortex and (L). Primary Auditory Cortex. and, (R). Perirhinal cortex and (L). Somatosensory Association Cortex. Adults with ASD show no differential functional connectivity compared to controls. This result is similar to [Nomi and Uddin, 2015].

This research can be extended in several aspects. We notice that there is age-related difference in brain connectivity. Bayesian models mentioned in Chapter 3 can be extended by adding the coefficient age to explore change in brain connectivity that may occur when individuals with ASD age. We use $Y_{i,rs}$ to denote the functional connectivity metric measurement between ROI

r and s (where, $1 \leq r < s \leq R$) for subject i, where $i = 1, \cdots, n$ and each subject i is nested within group Z_g, such that

$$Z_g = \begin{cases} 1, \text{ if age } < 13 \\ 2, \text{ if } 13 \leq \text{age} \leq 19 \\ 3, \text{ if age} > 19 \end{cases}$$

We can rewrite the parametric model from Chapter 3 as the following

$$\begin{aligned} Y_{i,rs} &\sim f(\cdot|\eta_{i,rs}) \\ g(\eta_{i,rs}) &= \beta_{1,rs} + \beta_{2,rs} X_i + \gamma_i + \delta_c \\ \beta_{2,rs} &\overset{i.i.d.}{\sim} p_g\, \delta_0(.) + (1 - p_g)\, G(\cdot) \\ \log\left(\frac{p_g}{1-p_g}\right) &= \alpha_1 + \alpha_2 Z_g \end{aligned} \tag{5.1}$$

Here, we assume that the mixture proportion p_g depends on age group. The above mentioned model can be extended to the Dirichlet process mixture model and neighborhood model.

5.3 Summary

We developed Bayesian models to explore contrasting brain connectivity between subjects with ASD and controls using rs-fMRI data from Autism Brain Image Data Exchange (ABIDE) database. We considered two separate approaches for analyzing rs-fMRI data.

We explored differential brain connectivity between subjects with ASD and controls using rs-fMRI data summarized over time by connectivity metrics. We developed three Bayesian models : parametric model, Dirichlet Process Mixture (DPM) model and Neighborhood model

on ROI pairs. Detailed simulation studies were implemented to compare performance of proposed models. Simulation study results indicated that DPM model performs exceptionally well compared to others. Initially, we used these models to analyze cross-correlated ABIDE data set for all the sites separately. We observed that those links are not identical across all sites. Later, we combined the data set and reanalyze using above mentioned models. The combined analysis, models and identifies the links which were differentially connected across sites.

We directly modeled time sequence of rs-fMRI measurement at each region of interest. We proposed a regional-temporal functional connectivity model. We have applied dynamic linear model (DLM) to capture temporal structure. The potential correlation between connected regions was modeled using hidden Potts model with latent variable. An algorithm was developed for joint modeling of DLM and hidden Potts model with latent variable using fixed inverse temperature and hidden Potts model with latent variable using pseudolikelihood approach. A detailed simulation study was performed to compare proposed approaches to blocked Dirichlet process. We observed that DLM with blocked Dirichlet process was unable capable capture the clustering structure of the simulated data. We did not observe any significant difference in performance between hidden Potts model with latent variable using fixed inverse temperature and hidden Potts model with latent variable using pseudolikelihood approach. We applied the proposed approach to analyze time-dependent rs-fMRI ABIDE data set. We observed different patterns in heat maps between control and ASD subjects. Moreover, contrasting connectivity patterns can also be seen across sites. We can also notice differential connectivity between ROIs

across lobes of brains of ASD subjects and controls. We observed dissimilarity in connectivity between ROIs across lobes across all the sites.

CITED LITERATURE

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72.

Aertsen, A., Gerstein, G., Habib, M., and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of" effective connectivity". *Journal of neurophysiology*, 61(5):900–917.

Alston, C. L., Mengersen, K. L., Robert, C. P., Thompson, J., Littlefield, P., Perry, D., and Ball, A. (2007). Bayesian mixture models in a longitudinal setting for analysing sheep cat scan images. *Computational Statistics & Data Analysis*, 51(9):4282–4296.

Andersen, A. H., Gash, D. M., and Avison, M. J. (1999). Principal component analysis of the dynamic response measured by fmri: a generalized linear systems framework. *Magnetic resonance imaging*, 17(6):795–815.

Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, pages 1152–1174.

Association, A. P. et al. (2000). Diagnostic and statistical manual of mental disorders. 4. *Text Revision: DSM-IV-TR*.

Baio J, Wiggins L, C. D. e. a. (2018). Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 sites, united states, 2014. *MMWR Surveill Summ*, 67((6)):1–23.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1):1–20.

Bhaumik, D., Jie, F., Nordgren, R., Bhaumik, R., and Sinha, B. K. (2018). A mixed-effects model for detecting disrupted connectivities in heterogeneous data. *IEEE transactions on medical imaging*, 37(11):2381–2389.

Bhaumik, R., Jenkins, L. M., Gowins, J. R., Jacobs, R. H., Barba, A., Bhaumik, D. K., and Langenecker, S. A. (2017). Multivariate pattern analysis strategies in detection of remitted major depressive disorder using resting state functional connectivity. *NeuroImage: Clinical*, 16:390–398.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541.

Biswal, B. B., Kylen, J. V., and Hyde, J. S. (1997). Simultaneous assessment of flow and bold signals in resting-state functional connectivity maps. *NMR in Biomedicine*, 10(4-5):165–170.

Blackwell, D., MacQueen, J. B., et al. (1973). Ferguson distributions via pólya urn schemes. *The annals of statistics*, 1(2):353–355.

Bowman, F. D., Caffo, B., Bassett, S. S., and Kilts, C. (2008). A bayesian hierarchical framework for spatial modeling of fmri data. *NeuroImage*, 39(1):146–156.

Brown, P. J., Vannucci, M., and Fearn, T. (1998). Multivariate bayesian variable selection and prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):627–641.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186.

Bush, C. A. and MacEachern, S. N. (1996). A semiparametric bayesian model for randomised block designs. *Biometrika*, 83(2):275–285.

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. (2001). A method for making group inferences from functional mri data using independent component analysis. *Human brain mapping*, 14(3):140–151.

Cao, J., Worsley, K., et al. (1999). The geometry of correlation fields with an application to functional connectivity of the brain. *The Annals of Applied Probability*, 9(4):1021–1057.

Carlin, B. P., Gelfand, A. E., and Banerjee, S. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Chen, S., Bowman, F. D., and Mayberg, H. S. (2016). A bayesian hierarchical framework for modeling brain connectivity for neuroimaging data. *Biometrics*, 72(2):596–605.

Cordes, D., Haughton, V., Carew, J. D., Arfanakis, K., and Maravilla, K. (2002). Hierarchical clustering to measure connectivity in fmri resting-state data. *Magnetic resonance imaging*, 20(4):305–317.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Cressie, N. A. (1993). Statistics for spatial data. john willy and sons. *Inc., New York*.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., and Lindquist, M. A. (2012). Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage*, 61(4):907–920.

Di Martino, A., Kelly, C., Grzadzinski, R., Zuo, X.-N., Mennes, M., Mairena, M. A., Lord, C., Castellanos, F. X., and Milham, M. P. (2011). Aberrant striatal functional connectivity in children with autism. *Biological psychiatry*, 69(9):847–856.

Escobar, M. D. (1994). Estimating normal means with a dirichlet process prior. *Journal of the American Statistical Association*, 89(425):268–277.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics*, pages 287–302. Elsevier.

Flandin, G. and Penny, W. D. (2007). Bayesian fmri data analysis with sparse spatial basis function priors. *NeuroImage*, 34(3):1108–1125.

Friston, K., Frith, C., Liddle, P., and Frackowiak, R. (1993). Functional connectivity: the principal-component analysis of large (pet) data sets. *Journal of Cerebral Blood Flow & Metabolism*, 13(1):5–14.

Friston, K. and Penny, W. (2003). Posterior probability maps and spms. *Neuroimage*, 19(3):1240–1249.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., and Turner, R. (1998). Event-related fmri: characterizing differential responses. *Neuroimage*, 7(1):30–40.

Friston, K. J., Holmes, A. P., Poline, J., Grasby, P., Williams, S., Frackowiak, R. S., and Turner, R. (1995). Analysis of fmri time-series revisited. *Neuroimage*, 2(1):45–53.

Friston, K. J., Jezzard, P., and Turner, R. (1994). Analysis of functional mri time-series. *Human brain mapping*, 1(2):153–171.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721.

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology*, 21(4):424–430.

Grelaud, A., Robert, C. P., Marin, J.-M., Rodolphe, F., Taly, J.-F., et al. (2009). Abc likelihood-free methods for model choice in gibbs random fields. *Bayesian Analysis*, 4(2):317–335.

Hahamy, A., Behrmann, M., and Malach, R. (2015). The idiosyncratic brain: distortion of spontaneous connectivity patterns in autism spectrum disorder. *Nature neuroscience*, 18(2):302.

Harrison, L. M. and Green, G. G. (2010). A bayesian spatiotemporal model for very large data sets. *NeuroImage*, 50(3):1126–1141.

Harrison, L. M., Penny, W., Daunizeau, J., and Friston, K. J. (2008). Diffusion-based spatial priors for functional magnetic resonance images. *Neuroimage*, 41(2):408–423.

Hull, J. V., Jacokes, Z. J., Torgerson, C. M., Irimia, A., and Van Horn, J. D. (2017). Resting-state functional connectivity in autism spectrum disorders: A review. *Frontiers in psychiatry*, 7:205.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173.

Kennedy, D. P. and Courchesne, E. (2008). The intrinsic functional organization of the brain is altered in autism. *Neuroimage*, 39(4):1877–1885.

Kennedy, D. P., Redcay, E., and Courchesne, E. (2006). Failing to deactivate: resting functional abnormalities in autism. *Proceedings of the National Academy of Sciences*, 103(21):8275–8280.

Kerin, T., Volk, H., Li, W., Lurmann, F., Eckel, S., McConnell, R., and Hertz-Picciotto, I. (2018). Association between air pollution exposure, cognitive and adaptive function, and asd severity among children with autism spectrum disorder. *Journal of autism and developmental disorders*, 48(1):137–150.

Lange, N. and Zeger, S. L. (1997). Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(1):1–29.

Lazar, N. (2008). *The statistical analysis of functional MRI data*. Springer Science & Business Media.

Lee, J. M., Kyeong, S., Kim, E., and Cheon, K.-A. (2016). Abnormalities of inter-and intra-hemispheric functional connectivity in autism spectrum disorders: a study using the autism brain imaging data exchange database. *Frontiers in neuroscience*, 10:191.

Lee, K.-J., Jones, G. L., Caffo, B. S., and Bassett, S. S. (2014). Spatial bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Analysis (Online)*, 9(3):699.

Lee, Y., Park, B.-y., James, O., Kim, S.-G., and Park, H. (2017). Autism spectrum disorder related functional connectivity changes in the language network in children, adolescents and adults. *Frontiers in human neuroscience*, 11:418.

Lempers, F. B. (1971). Posterior probabilities of alternative linear models.

Lindquist, M. A. and Wager, T. D. (2007). Validity and power in hemodynamic response modeling: a comparison study and a new approach. *Human brain mapping*, 28(8):764–784.

Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folstein, S., Fombonne, E., Leboyer, M., and Minshew, N. (1997). Diagnosing autism: analyses of data from the autism diagnostic interview. *Journal of autism and developmental disorders*, 27(5):501–517.

Lowe, M. J., Dzemidzic, M., Lurito, J. T., Mathews, V. P., and Phillips, M. D. (2000). Correlations in low-frequency bold fluctuations reflect cortico-cortical connections. *Neuroimage*, 12(5):582–587.

MacEachern, S. N. (1994). Estimating normal means with a conjugate style dirichlet process prior. *Communications in Statistics-Simulation and Computation*, 23(3):727–741.

MacEachern, S. N. and Müller, P. (1998). Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics*, 7(2):223–238.

Marc G. Weisskopf, M.-A. K. and Roberts, A. L. (2015). Air pollution and autism spectrum disorders: Causal or confounded? *Curr Environ Health Rep.*, 2((4)):430–439.

Marin, J.-M. and Robert, C. P. (2014). *Bayesian essentials with R*, volume 48. Springer.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.

Møller, J., Pettitt, A. N., Reeves, R., and Berthelsen, K. K. (2006). An efficient markov chain monte carlo method for distributions with intractable normalising constants. *Biometrika*, 93(2):451–458.

Moores, M., Nicholls, G., Pettitt, A., Mengersen, K., et al. (2018). Scalable bayesian inference for the inverse temperature of a hidden potts model. *Bayesian Analysis*.

Müller, P., Erkanli, A., and West, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika*, 83(1):67–79.

Neal, R. M. (1992). Connectionist learning of belief networks. *Artificial intelligence*, 56(1):71–113.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265.

Nebel, M. B., Joel, S. E., Muschelli, J., Barber, A. D., Caffo, B. S., Pekar, J. J., and Mostofsky, S. H. (2014). Disruption of functional organization within the primary motor cortex in children with autism. *Human brain mapping*, 35(2):567–580.

Nomi, J. S. and Uddin, L. Q. (2015). Developmental changes in large-scale network connectivity in autism. *NeuroImage: Clinical*, 7:732–741.

Ogawa, S., Lee, T.-M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872.

Padmanabhan, A., Lynn, A., Foran, W., Luna, B., and O'Hearn, K. (2013). Age related changes in striatal resting state functional connectivity in autism. *Frontiers in human neuroscience*, 7:814.

Patel, R. S., Bowman, F. D., and Rilling, J. K. (2006). A bayesian approach to determining connectivity of the human brain. *Human brain mapping*, 27(3):267–276.

Pauling, L. and Coryell, C. D. (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, 22(4):210–216.

Penny, W., Kiebel, S., and Friston, K. (2003). Variational bayesian inference for fmri time series. *NeuroImage*, 19(3):727–741.

Petris, G., Petrone, S., and Campagnoli, P. (2009). Dynamic linear models. In *Dynamic Linear Models with R*, pages 31–84. Springer.

Potts, R. B. (1952). Some generalized order-disorder transformations. In *Mathematical proceedings of the cambridge philosophical society*, volume 48, pages 106–109. Cambridge University Press.

Quirós, A., Diez, R. M., and Gamerman, D. (2010). Bayesian spatiotemporal model of fmri data. *NeuroImage*, 49(1):442–456.

Rebecca Muhle, S. V. T. and Rapin, I. (2004). The genetics of autism. *Pediatrics*, 113:e472–e486.

Robert, C. P., Ryden, T., and Titterington, D. M. (2000). Bayesian inference in hidden markov models through the reversible jump markov chain monte carlo method. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):57–75.

Rutter, M., Le Couteur, A., Lord, C., et al. (2003). Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*, 29:30.

Rydén, T. and Titterington, D. (1998). Computational bayesian analysis of hidden markov models. *Journal of Computational and Graphical Statistics*, 7(2):194–211.

Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica sinica*, pages 639–650.

Smith, M. and Fahrmeir, L. (2007). Spatial bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association*, 102(478):417–431.

Sun, F. T., Miller, L. M., and D'esposito, M. (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fmri data. *Neuroimage*, 21(2):647–658.

Surhone, L. M., Tennoe, M. T., and Henssonow, S. F. (2010). Openbugs.

Taylor, L. E., Swerdfeger, A. L., and Eslick, G. D. (2014). Vaccines are not associated with autism: an evidence-based meta-analysis of case-control and cohort studies. *Vaccine*, 32(29):3623–3629.

Thulborn, K. R., Waterton, J. C., Matthews, P. M., and Radda, G. K. (1982). Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 714(2):265–270.

Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science*, 342(6158):580–584.

Uddin, L. Q., Supekar, K., and Menon, V. (2013). Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Frontiers in human neuroscience*, 7:458.

Van Den Heuvel, M. P. and Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fmri functional connectivity. *European neuropsychopharmacology*, 20(8):519–534.

Van Horn, J. D. and Toga, A. W. (2014). Human neuroimaging as a "big data" science. *Brain imaging and behavior*, 8(2):323–331.

Warnick, R., Guindani, M., Erhardt, E., Allen, E., Calhoun, V., and Vannucci, M. (2018). A bayesian approach for estimating dynamic functional network connectivity in fmri data. *Journal of the American Statistical Association*, 113(521):134–151.

West, M. and Escobar, M. D. (1993). *Hierarchical priors and mixture models, with application in regression and density estimation.* Institute of Statistics and Decision Sciences, Duke University.

West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2004). Multilevel linear modelling for fmri group analysis using bayesian inference. *Neuroimage*, 21(4):1732–1747.

Xia, J., Liang, F., and Wang, Y. M. (2009). fmri analysis through bayesian variable selection with a spatial prior. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 714–717. IEEE.

Zalesky, A., Fornito, A., and Bullmore, E. (2012). On the use of correlation as a measure of network connectivity. *Neuroimage*, 60(4):2096–2106.

Zhang, L., Guindani, M., and Vannucci, M. (2015a). Bayesian models for functional magnetic resonance imaging data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(1):21–41.

Zhang, L., Guindani, M., Versace, F., and Vannucci, M. (2014). A spatio-temporal nonparametric bayesian variable selection model of fmri data for clustering correlated time courses. *NeuroImage*, 95:162–175.

Zhang, T., Wu, J., Li, F., Caffo, B., and Boatman-Reich, D. (2015b). A dynamic directional model for effective brain connectivity using electrocorticographic (ecog) time series. *Journal of the American Statistical Association*, 110(509):93–106.

VITA

NAIRITA GHOSAL

Education

- **Doctor of Philosophy**, 2019
  Biostatistics
  University of Illinois at Chicago, Chicago, Illinois, USA
- **Master of Science**, 2014
  Mathematical Science
  University of New Orleans, New Orleans, Louisiana, USA
- **Master of Science**, 2011
  Statistics
  University of Kalyani, Kalyani, West Bengal, India
- **Master of Science**, 2009
  Statistics
  University of Calcutta, Kolkata, West Bengal, India

Research Experience

- **Center for Clinical and Translational Sciences, University of Illinois at Chicago**
  Research Assistant, August 2018–May 2019.
  - Providing statistical support regarding data analysis, study design, power analysis, sample size calculation and grant preparation.
- **Center for Drug Evaluation and Research, FDA**
  Summer Intern, June 2018–August,2018.
  - Applied and evaluated Bayesian Hierarchical models to estimate subgroup treatment effects across multiple levels.
  - Applied Bayesian approach for joint modeling of benefit and risk in clinical trial.
- **AbbVie Inc.**
  Research Extern at Abbvie, North Chicago Site, August 2017–May 2018.
  - Applied missing data handling techniques to obtain bias corrected estimate.
  - Analyzed Patient Reported Outcome (PRO) data.
  - Provided statistical support for Phase III clinical trial.

- **Center for Devices and Radiological Health, FDA**
  Summer Intern, June 2017–August 2017.
  - Examined multiple methods to generate reference interval and applied them to Optical Coherence Tomography Database.
- **Institute for Health Research and Policy, University of Illinois at Chicago**
  Research Assistant, August 2016–May 2017.
  - Worked on building R Shiny App for accessing REDCap database and produced R markdown documents.
  - Provided statistical support regarding longitudinal analysis, multilevel modeling, missing data issues, power analysis, sample size calculation to UIC researcher community.

Teaching Experience

- **School of Public Health, University of Illinois at Chicago**
  Teaching Assistant, August 2016–May 2017.
  - Biostatistics II (Spring 2017), Longitudinal Data Analysis (Fall 2016).
- **Department of Mathematics, Northern Illinois University**
  Teaching Assistant, August 2014–August 2016.
  - *Recitation Instructor*: Core Competency in Mathematics (Summer 2016),Calculus for Business and Social Science (Spring 2016), Elementary Statistics (Fall 2015).
  - *Graduate course grader*: Introduction to Probability Theory (Fall 2015).
  - *Tutor in the mathematics assistant center*: Various undergraduate courses (Fall 2014 and Spring 2015).
- **Department of Mathematics, University of New Orleans**
  Teaching Assistant, August 2012–May 2014.
  - *Full responsibility Instructor*: Introductory Statistics (Fall 2013 and Spring 2014).
  - *Teaching assistant*: Precalculus (Spring 2013).
  - *Tutor in the mathematics assistant center*: Various undergraduate and some graduate courses(Fall 2012–Spring 2014).

Awards

- FDA Research Fellowship, awarded by Center for Drug Evaluation and Research (CDER), Food and Drug Administration, 2018.
- FDA Research Fellowship, awarded by Center for Devices and Radiological Health (CDRH), Food and Drug Administration, 2017.

- Honorable Mention Award, awarded by Department of Mathematics, Northern Illinois University, 2016.
- Mathematics Faculty Award, awarded by College of Sciences, University of New Orleans, 2013.
- Conference Travel Awards
  - International Indian Statistical Association Conference 2018, University of Florida, FL, 2018.
  - BiostatMCW 2017, Medical College of Wisconsin, Milwaukee, WI, 2017.

Presentations

- FDA ORISE Statistical Conference, Silver Spring, Maryland, August 17, 2018.
- Joint Statistical Meeting 2018, Vancouver, BC, July 28-August 1, 2018.
- International Indian Statistical Association Conference 2018 at University of Florida, FL, May 17-20, 2018.
- BiostatMCW 2017 at the Medical College of Wisconsin in Milwaukee, WI, September 28-30, 2017.

Publication

1. Bipasa Biswas, **Nairita Ghosal**. " Comparison of Methods to Generate Reference Limits". *In JSM Proceedings, Biopharmaceutical Section, Vancouver, BC* (2018).
2. Ellen Manieri, Michelle De Lima, **Nairita Ghosal**. " Testing for success: A logistic regression analysis to determine which pre-admission exam best predicts success in an associate degree in nursing program", *Teaching and Learning in Nursing*, 10 (2015), 25–29.

Professional Membership

- American Statistical Association.
- International Indian Statistical Association.