2021

Statistical Approaches for Estimation and Comparison of Brain Functional Connectivity

Jifang Zhao
Virginia Commonwealth University

www.manaraa.com

STATISTICAL APPROACHES FOR ESTIMATION AND COMPARISON OF BRAIN FUNCTIONAL CONNECTIVITY

A Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at Virginia Commonwealth University.

by

JIFANG ZHAO

PhD Candidate in Systems Modeling & Analysis at Virginia Commonwealth University

Directors: Dr. Yanjun Qian, Assistant Professor, Department of Statistical
Sciences and Operations Research, VCU
Dr. Montserrat Fuentes, President of St. Edward's University
Dr. Qiong Zhang, Assistant Professor, Mathematical and Statistical Sciences,
Clemson University

Virginia Commonwewalth University
Richmond, Virginia
May, 2021

Acknowledgements

The process of earning a doctorate is academically rigorous - especially in such a difficult time. This dissertation becomes a reality with the kind support and helps of many individuals. It is a genuine pleasure to express my deep sense of appreciation to various people who have been journeyed with me in recent years as I have worked on this research.

Firstly I would like to express my gratitude to my directors Dr. Montserrat Fuentes, Dr. Qiong Zhang, and Dr. Yanjun Qian, for their timely advice, meticulous scrutiny, and scientific approach that helped me to a very great extent to accomplish this task.

Second, I would like to thank my dissertation committee members, Dr. Edward Boone and Dr. Liangsuo Ma, for their guidance, constant supervision, and support in completing this research.

I am incredibly thankful to the Department of Statistical Sciences and Operations Research and the Department of Mathematics and Applied Mathematics for giving me a chance for fulfilling my Ph.D. dream. And I sincerely thank all faculties and staff in both departments for their kind help and co-operation throughout my study period.

Finally, I would like to thank my family and all my friends for their love, support, and encouragement.

TABLE OF CONTENTS

iii

LIST OF TABLES

LIST OF FIGURES

Abstract

STATISTICAL APPROACHES FOR ESTIMATION AND COMPARISON OF BRAIN FUNCTIONAL CONNECTIVITY

By **Jifang Zhao**

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2021.

Director: Dr. Yanjun Qian, Assistant Professor, Department of Statistical Sciences and Operations Research, VCU

Dr. Montserrat Fuentes, President of St. Edward's University

Dr. Qiong Zhang, Assistant Professor, Mathematical and Statistical Sciences, Clemson University ,

,

Drug addiction can lead to many health-related problems and social concerns. Researchers are interested in the association between long-term drug usage and abnormal functional connectivity. Functional connectivity obtained from functional magnetic resonance imaging (fMRI) data promotes a variety of fundamental understandings in such association. Due to its complex correlation structure and large dimensionality, the modeling and analysis of the functional connectivity from neuroimage are challenging. By proposing a spatio-temporal model for multi-subject neuroimage data, we incorporate voxel-level spatio-temporal dependencies of whole-brain measurements to improve the accuracy of statistical inference. To tackle large-scale spatio-temporal neuroimage data, we develop a computational efficient algorithm to

estimate the parameters. Our method is used to first identify functional connectivity, and then detect the effect of cocaine use disorder (CUD) on functional connectivity between different brain regions. The functional connectivity identified by our spatio-temporal model matches existing studies on brain networks, and further indicates that CUD may alter the functional connectivity in the medial orbitofrontal cortex subregions and the supplementary motor areas. We further propose a method that jointly estimates the graphical models which share the common structure, while allowing for differences between categories in the data. By assigning different tuning parameters for the contrast of each categorical factor, our method could estimate the effects of multiple treatments or factors across brain regions more accurately, and achieve computational efficiency at the same time. Simulation studies suggest that our method achieves better accuracy in network estimation compared with the joint graphical lasso method. We apply our method to the cocaine-use disorder data and identify functional connectivity in brain affected by cocaine use disorder and gender.

CHAPTER 1

INTRODUCTION

Substance use disorders have been the concern of the public over decades. Abuse of drugs can cause serious harm to both physical and psychological health conditions. Some of the effects of drug abuse on a body include heart disease, liver damage, depression, or anxiety that last for years or even forever (Nestler, 2005). Ma et al. (2015) demonstrates that cocaine use is associated with altered connectivity between brain structures such as cortical-striatal regions and default mode network. Functional connectivity, which is a useful intermediary for studying the relationship between functional communication in brain with human behavior, has attracted attention from the neuroimaging community on the study of the effect of cocaine dependency (Smith et al., 2011). Functional connectivity is a manifestation of complex brain functions aiming at the interaction of brain regions and their behaviors towards different human activities and environmental stimulus. Traditionally, functional connectivity is defined as the temporal coherence in neural activity between spatially disconnected brain regions (Smith et al., 2011). From a statistical perspective, we adopt the definition in Friston (2011) and consider functional connectivity as groups of regions/voxels with high correlations.

In this work, we focus on the resting state functional magnetic resonance imaging (fMRI) data to study the effect of cocaine use disorder on functional connectivity. Resting state fMRI is a tool to explore functional connectivity and examine neurological or mental disorders (Fox and Raichle, 2007). During a typical resting state fMRI experiment, the subject is under a resting or a task-free state. The magnetic reso-

1

nance scanner captures the blood oxygenated level dependent (BOLD) signals which reflect oxygenation and deoxygenation in neurons every two to three seconds at several hundred time points. At each time point (referred to as a volume hereafter), the recorded BOLD signals form a three-dimensional brain image measuring brain activity at each brain voxel. The resulting fMRI datasets often consist of hundreds to thousands of voxels observed at hundreds of volumes for each subject. Its high dimensionality, massive size and complex spatio-temporal correlation structure make it difficult to establish a computationally feasible statistical model to accurately analyze voxel-level spatio-temporal characteristics of the brain activity (Huettel et al., 2014). To overcome this challenge, our goal is to develop a nonstationary computationally efficient spatio-temporal model at the voxel-level for multi-subject resting state fMRI data.

After estimated the voxel-level large-scale spatial structure of human brains, it is still challenging to jointly estimate the spatial dependence of multiple functional networks in brain across multiple factors, where the spurious signals identified from conditional covariances will significantly increase the estimation variation. Moreover, it's computational complexity is high when the numbers of factors, factor levels and the size of covariance matrices are large. In this study, we propose a method that jointly estimates the graphical models which share the common structure, while allowing for differences between categories in the data. By assigning different tuning parameters for the contrast of each categorical factor, our method could estimate the effects of multiple treatments or factors across brain regions more accurately, and achieve computational efficiency at the same time.

CHAPTER 2

A SPATIO-TEMPORAL MODEL FOR DETECTING THE EFFECT OF COCAINE USE DISORDER ON FUNCTIONAL CONNECTIVITY

2.1 Introduction

Substance use disorders have been the concern of the public over decades. Abuse of drugs can cause serious harm to both physical and psychological health conditions. Some of the effects of drug abuse on the body include heart disease, liver damage, depression or anxiety that last for years or even forever (Nestler, 2005). Ma et al., 2015 demonstrate that cocaine use is associated with altered functional connectivity between brain structures such as cortical-striatal regions and default mode network. Functional connectivity, which is a useful intermediary for studying the relationship between functional communication in brain with human behavior, has attracted attention from the neuroimaging community on the study of the effect of cocaine dependency (Smith et al., 2011). Functional connectivity is a manifestation of complex brain functions aiming at the interaction of brain regions and their behaviors towards different human activities and environmental stimulus. Traditionally, functional connectivity is defined as the temporal coherence in neural activity between spatially disconnected brain regions (Smith et al., 2011). From a statistical perspective, we adopt the definition in Friston, 2011 and consider functional connectivity as groups of regions/voxels with high correlations.

In this work, we focus on the resting state functional magnetic resonance imaging (fMRI) data to study the effect of cocaine use disorder on functional connectivity. Resting state fMRI is a tool to explore functional connectivity and examine neuro-

logical or mental disorders (Fox and Raichle, 2007). During a typical resting state fMRI experiment, the subject is under a resting or a task-free state. The magnetic resonance scanner captures the blood oxygenated level dependent (BOLD) signals, which reflect oxygenation and deoxygenation in neurons every two to three seconds at several hundred time points. At each time point (referred to as a volume hereafter), the recorded BOLD signals form a three-dimensional brain image measuring brain activity at each brain voxel. The resulting fMRI datasets often consist of hundreds to thousands of voxels observed at hundreds of volumes for each subject. Its high dimensionality, massive size and complex spatio-temporal correlation structure make it difficult to establish a computationally feasible statistical model to accurately analyze voxel-level spatio-temporal characteristics of the brain activity (Huettel et al., 2014). To overcome this challenge, our goal is to develop a nonstationary computationally efficient spatio-temporal model at the voxel-level for multi-subject resting state fMRI data.

Spatio-temporal models have been introduced to brain image research in recent years (Lindquist et al., 2008; Ombao et al., 2016). Woolrich et al. (2004) build simultaneous autoregressive models for adjacent voxels. Bowman (2005) aggregates data to regions of interest (ROIs), and then adopts a two-stage hierarchical Bayesian approach to estimate the locally-independent spatial activation patterns and the spatial-dependent within regional mean activity profiles. Kang et al. (2012) use a spatio-spectral mixed-effects model to capture pre-specified multi-scale spatial correlation. These methods analyze regional level activation, but ignores the voxel-level spatial dependence. Hyun et al. (2016) propose a spatio-temporal Gaussian process framework to delineate the developmental trajectories of brain structure and function. Reich et al. (2017) introduce a multi-resolution model and a computationally efficient methodology in the spectral domain to estimate whole-brain spatio-temporal

4

dependence structure. Their model is capable of detecting voxel-specific activation under a spatially non-stationary anisotropy setting, but only for single-subject time-independent fMRI data.

For multi-subject fMRI data, the existing methods are not able to incorporate the spatio-temporal dependence efficiently in the model and meanwhile make inference on the functional connectivity, due to its complex structure. In this chapter, we propose an efficient spatio-temporal model for multi-subject whole-brain resting state fMRI data. Our model incorporates the voxel-level spatio-temporal dependence of the whole-brain measurements to improve the accuracy in statistical inference. To achieve computational efficiency, we adopt ideas from fixed rank kriging (Cressie and Johannesson, 2008) to reduce the dimensionality of the covariance matrix in the spatio-temporal model, and develop an empirical Bayesian approach based on the EM algorithm to estimate the parameters. This approach is computationally efficient in processing massive spatio-temporal data. Based on our model, the functional connectivity is characterized by the correlation matrix across different brain voxels/regions. We apply our model to a multi-subject fMRI dataset to first identify functional connectivity, and then detect the effect of cocaine use disorder on the functional connectivity.

The rest of this chapter is organized as follows. Section 2.2 describes the multi-subject fMRI data used for the cocaine use disorder study. Section 2.3 proposes a spatio-temporal model to identify the functional connectivity. Section 2.4 develops a computationally efficient approach to estimate the unknown model parameters. A simulation study on the performance of model estimation is provided in Section 2.5. We apply our approach to the multi-subject fMRI data in Section 2.6. A discussion is given in Chapter 4.

2.2 Description of Multi-subject fMRI Data for Cocaine Use Disorder Study

This work is motivated by a multi-subject fMRI dataset collected from a cocaine use disorder study conducted by the Institute for Drug and Alcohol Studies at Virginia Commonwealth University. Subjects with current cocaine use disorder (DSM-IV, First et al., 1996) and non-drug-using normal controls were recruited through advertisements for research volunteers. All subjects underwent the Addiction Severity Index (McLellan et al., 1992) to document possible lifetime drug and alcohol use, and then been included in the cocaine use disorder group (22 subjects) or the control group (23 subjects). Four variables representing the demographic information of each subject are collected: 1. the gender; 2. the age (in years); 3. the years of education taken; 4. ethnicity, with AA indicating African American, C indicating White, and H indicating Hispanic or Latino. The summary statistics with respect to the treatment and those demographic factors are given in Table 2.1.

Table 2.1. A summary of the study population of cocaine use disorder (CUD) participants and control participants. (Note: AA stands for African American, C stands for White, and H stands for Hispanic or Latino)

Group	Age (years)	Education (years)	Ethnicity
CUD Group (22, 62.5% male)			
Mean	45	12.27	21 AA
Standard Deviation	8.52	2.42	1 C
Range(max-min)	32	10	
Control Group (23, 60.9% male)			
Mean	40.3	14.96	14 AA
Standard Deviation	10.52	2.58	8 C
Range(max-min)	35	10	1 H

For each subject, the resting state fMRI data were captured on a Philips Medical Systems (Best, Netherlands) Ingenia wide-bore dStream 3T MRI scanner. After acquiring the raw data, it were preprocessed using a standard pipeline for volume-

6

based analysis to realign the subject's functional images, and smooth it to remove redundant noise. After preprocessing, the BOLD signal is transformed to $2mm \times 2mm \times 2mm$ 3-dimensional fMRI sample size (voxel) on a $91 \times 108 \times 91$ regular lattice for each volume (time point). Each subject has 375 volumes. There are a total of $185,355$ voxels within the brain template space after registration across subjects. Further details on the participants and data preprocessing can be found in A.5.

2.3 A Spatio-temporal Model for Identifying Functional Connectivity via Multi-subject fMRI Data

To identify the functional connectivity for one group, we propose a spatio-temporal model for the fMRI data collected from N subjects. Let $y_i(\mathbf{v}, t)$ be the normalized fMRI response at voxel $\mathbf{v}$ and volume t for subject i. Associated with each subject is a p-dimensional vector of demographic covariates $\mathbf{x}_i$, which includes age, education, and race of the corresponding subject in our study. We assume that the spatial location $\mathbf{v}$ is allocated to a three-dimensional regular grid, and the total number of locations is M. The time volume t ranges from 1 to L, where L is the total number of volume in data. The model for the response of subject i is

$$y_i(\mathbf{v}, t) = \mu + \mathbf{x}_i^\top \boldsymbol{\beta} + \eta_i(\mathbf{v}, t) + e_i(\mathbf{v}, t). \qquad (2.1)$$

The terms in (2.1) are explained in turn as follows. The fixed effects $\mu + \mathbf{x}_i^\top \boldsymbol{\beta}$ contain a constant intercept μ, and linear additive effects associated with the demographic covariate vector $\mathbf{x}_i$, where $\boldsymbol{\beta}$ is a common linear coefficient vector across all subjects. The random effects $\eta_i(\mathbf{v}, t)$'s are independent over different subjects, and the spatial temporal dependence of the measurements from the same subjects are incorporated through this term. The random effect $e_i(\mathbf{v}, t)$'s are white noise processes, which are assumed to be independently and identically distributed across different subjects,

7

voxels, and volumes, i.e., $e_i(\mathbf{v}, t) \sim \mathcal{N}(0, \sigma^2)$. The common variance of the white noise is a popular assumption for the neruoimage data after using pre-processing procedures. Similar assumptions can be found in Zhu et al. (2014) and Reich et al. (2017).

Note that, the assumptions in (2.1) can be extended to more general cases for different applications. For example, the linear coefficient $\boldsymbol{\beta}$ can be extended to spatial or spatial-temporal varying coefficients. Also, the variance σ^2 of the white noise term can be assumed to be different for different subjects. This paper validates the assumptions in (2.1) to our application using the empirical data analysis in A.6.2 and A.6.1. Also, we include the spatial-temporal dependence between measurements with a space-time process $\eta_i(\mathbf{v}, t)$, which is a commonly used assumption for fMRI data due to the dense layout of spatial location and time frequency. However, it is worth to mention that, different from observational data from a spatial/temporal process, fMRI data in terms of time frequency and spatial location are discrete. Therefore, the purpose of spatial temporal modeling for this case is statistical inference to unknown parameters, but not prediction for measurements under unobserved spatial or time point.

We are interested in estimating correlation structure of the data over a large spatio-temporal domain, which leads the main computational challenge. To overcome this issue, we propose to use a low rank representation model to reduce the dimension of correlation matrix

$$\eta_i(\mathbf{v}, t) = \sum_{l=1}^{L} \omega_l(t) B(\mathbf{v})^\top \Gamma_{i,l} \tag{2.2}$$

where $\omega_l(t)$'s are known temporal basis functions at volume t; $B(\mathbf{v})$ is a vector of basis functions of size J; $\Gamma_{i,l}$'s are independent J-dimensional random vectors that characterize the random variation of brain networks between different subjects. Par-

ticularly, we assume that $\Gamma_{i,l}$'s are independent realizations from the multivariate normal distribution $N(0, \Sigma)$. Since J can be much smaller than the total number of voxels (denoted by M), we are able to reduce the dimensionality of the large-scale spatio-temporal observations. We do not reduce the temporal dimension. Thus, the number of temporal basis functions is the same as the total number of volumes L.

Under the proposed model in (2.1) and (2.2), the covariance between two space-time points of the i-th subject is written as

$$\text{Cov}\left\{y_i(\mathbf{v}, t), y_i(\mathbf{v}', t')\right\} = B(\mathbf{v})^\top \Sigma B(\mathbf{v}') \times \tau(t', t) + \sigma^2 I(\mathbf{v} = \mathbf{v}', t = t'),$$

where $\tau(t', t)$ is the temporal correlation functions constructed by $\omega_l(t)$'s, i.e., $\tau(t', t) = \sum_{l=1}^{L} \omega_l(t)\omega_l(t')$. This function implies a non-stationary space-time separable covariance structure, and separable covariance model assumption of the space and time domains, which has been used in various studies in neuroimaging study (see Benali et al. (1997), Katanoda et al. (2002), and Shvartsman et al. (2018) for example). In this model, the spatial dependence of the entire brain image is characterized by the covariance matrix of $\boldsymbol{B}\Sigma\boldsymbol{B}^\top$, where $\boldsymbol{B} = [B(\mathbf{v}_1), \ldots, B(\mathbf{v}_M)]^\top$ is an $M \times J$ matrix formed by the basis functions. Given the basis functions, the functional connectivity is determined by the relatively small $J \times J$ covariance matrix Σ under given basis functions. Therefore, statistical inference for the covariance matrix Σ enables the diagnostic and comparison of functional connectivity using the proposed model.

For our application on cocaine use disorder, the subjects belong to the cocaine use disorder group ($g = 1$) or the control group ($g = 2$). The covariance matrix of the cocaine and control groups are specified by Σ_1 and Σ_2, respectively. We model the response of subject from the g-th group by

$$y_{i,g}(\mathbf{v}, t) = \mu_g + \mathbf{x}_{i,g}^\top \boldsymbol{\beta} + \sum_{l=1}^{L} \omega_l(t) B(\mathbf{v})^\top \Gamma_{i,l,g} + e_{i,g}(\mathbf{v}, t), \tag{2.3}$$

9

where $\Gamma_{i,l,g}$'s are independent realizations from $\mathcal{N}(0, \Sigma_g)$. Based on the estimated spatial covariance matrices $\boldsymbol{B}\Sigma_1\boldsymbol{B}^\top$ and $\boldsymbol{B}\Sigma_2\boldsymbol{B}^\top$, we use hypothesis testing and graphical Lasso approaches to identify the differences between these two matrices and then assess the effects of the cocaine use disorder on functional connectivity in Section 2.6.

2.4 Model Estimation Using the EM Algorithm

For large scale spatio-temporal data, it is inefficient to implement the fully Bayesian inference. In this section, we develop an empirical Bayesian approach to assess the unknown parameters in (2.1) via the EM algorithm. Since the goal of this analysis is to estimate the covariance matrix Σ, an empirical Bayesian estimate can be obtained by maximizing the posterior (MAP) probability of Σ. We then develop computational efficient update formulas for the EM algorithm to maximize the posterior probability.

As our focus is to assess the spatial covariance, we adopt the isomorphic transformation in Zhang et al. (2016) and simplify the model by removing the temporal dependence. Let $\Sigma_T = \mathcal{W}\mathcal{W}^\top$ be the temporal correlation matrix, we choose $\mathcal{W}$ to be the solution of $U^\top \mathcal{W} = \sqrt{V}$, where U and V are the eigenvector matrix and eigenvalue matrix of Σ_T. The model in (2.1) can be transformed to

$$\tilde{Y}_{i,t} = \tilde{X}_{i,t}\boldsymbol{\beta} + \sqrt{\lambda_t}\boldsymbol{B}\Gamma_{i,t} + \tilde{E}_{i,t}, \tag{2.4}$$

where $\tilde{Y}_{i,t} = (U_t^\top \otimes I_M)Y_i$; $Y_i = \{Y_{i,1}, Y_{i,2}, \ldots, Y_{i,L}\}$ is a vector stacking all the responses of the i-th subject; U_t is the t-th column of U; $\otimes$ stands for the Kronecker product; $\tilde{X}_{i,t} = \underline{1}_M \otimes (\sum_{l=1}^{L} U_{t,l}^\top)x_i^\top$ is the covariates after this transformation; and $\tilde{E}_{i,t} = (U_t^\top \otimes I_M)E_{i,t}$ collects the residuals of subject i at volume t. After this trans-

formation, the variance-covariance matrix of $\tilde{Y}_{i,t}$ can be expressed by

$$\text{var}(\tilde{Y}_{i,t}) = \text{var}\left((U_t^\top \otimes I_M)Y_i\right) = \lambda_t \boldsymbol{B}\Sigma\boldsymbol{B}^\top + \sigma^2 I_M.$$

The expressions demonstrating the relationship between the original model (2.1) and transformed model (2.4) are deferred to the A.1.

Let $\tilde{\mathbf{Y}} = \{\tilde{Y}_{i,t}; i = 1, \ldots, N, t = 1, \ldots, L\}$ be a set containing all the transformed responses and $\boldsymbol{\Gamma}_i = \{\Gamma_{i,t}; i = 1, \ldots, N, t = 1, \ldots, L\}$ be the set of all the random factors for each subject. By inserting the prior distributions of the model parameters, the posterior distributions of the unknown parameters can be expressed by

$$p(\boldsymbol{\beta}, \Sigma, \sigma^2|\tilde{\mathbf{Y}}) \propto p(\tilde{\mathbf{Y}}|\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_N, \boldsymbol{\beta}, \Sigma, \sigma^2)p(\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_N|\boldsymbol{\beta}, \Sigma, \sigma^2)p(\Sigma, \sigma^2)p(\boldsymbol{\beta}). \qquad (2.5)$$

We obtain our parameter estimates by maximizing the posterior function. Due to the complicated function structure and large amount of parameters, directly maximizing the above posterior function is computationally challenging. To overcome this problem, we treat the random factors $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_N$ as hidden variables, and use the EM algorithm to obtain the MAP estimates.

In the $l + 1$ iteration of the EM algorithm, the expectation step calculates the expectation of the logarithm posterior probability function with respect to $\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_N\}$ given the parameter values $\boldsymbol{\beta}^l$, Σ^l, and σ^l at the previous iteration:

$$Q(\boldsymbol{\beta}, \Sigma, \sigma|\boldsymbol{\beta}^l, \Sigma^l, \sigma^l) = \text{E}_{\boldsymbol{\Gamma}}\left[-2\log p(\boldsymbol{\beta}, \Sigma, \sigma^2|\tilde{\mathbf{Y}})|\boldsymbol{\beta}^l, \Sigma^l, \sigma^l\right].$$

Then, the maximization step maximizes $Q(\boldsymbol{\beta}, \Sigma, \sigma|\boldsymbol{\beta}^l, \Sigma^l, \sigma^l)$ to obtain the updated parameters, $\boldsymbol{\beta}^{l+1}$, Σ^{l+1}, and σ^{l+1}. These three parameters will converge to the true MAP estimators, thanks to the convergence of EM algorithm (Wu et al., 1983). When closed-form update formulas are available, the EM algorithm can be computationally efficient. The prior settings and the close-form update formulas we used in the simu-

lation and data application is deferred to the A.2.

The order of the computational effort of the proposed update formulas in A.2 is $O(NLJ)$. In our application on the cocaine use disorder data, EM algorithm converges within 30 steps. Using this efficient update formulas, the empirical Bayesian inference on a 56.7 GB spatial temporal multi-subject fMRI data can be handled within 3 minutes on a desktop computer with a quad-core 4.00GHz processor and 16GB RAM space.

2.5   Simulation Studies

In this section, we illustrate the performance of our method using four simulation studies under different scenarios. We investigate the effects of model misspecification, different levels of the noise-to-signal ratio, and complex connectivity on the parameter estimation. We also provide a series of sensitivity analysis to obtain practical guidance for the choice of model parameters. In general, the responses are generated from model (2.1). Each simulated dataset includes N subjects and L volumes. The deterministic linear trend in (2.1) is specified to be $\mu + \beta_1 x_1 + \beta_2 x_2$ with $\mu = 0.1$. The covariates x_1 and x_2 are randomly sampled from uniform distributions $U(0,5)$ and $U(5,10)$, respectively. The temporal basis functions $\omega_l(t)$'s are specified as the Fourier basis functions (Konidaris et al., 2011). The random factors $\Gamma_{i,l}$'s in (2.2) are independently generated from $\mathcal{N}(0, \Sigma^0)$, where Σ^0 is a $J \times J$ matrix. In model misspecification and noise-to-signal ratio simuation studies, Σ^0 is a prespecified positive definite matrix. In the complex connectivity simulation study and sensitivity analysis, Σ^0 is estimated from the real data.

2.5.1 Model Misspecification

First, we explore the effect of model misspecification on parameter estimation. In this study, data are generated with $N = 25$ and $L = 400$. We use $\Sigma_{i,j}^0 = 0.5^{|i-j|}$, for $i = 1, \ldots, J$ and $j = 1, \ldots, J$. The $J = 125$ basis nodes $(\mathbf{v}_1, \ldots, \mathbf{v}_J)$ are expanded to $50 \times 50 \times 50$ spatial grids using basis vector $B(\mathbf{v}) = \{\phi(\mathbf{v}, \mathbf{v}_k), k = 1, \ldots, J\}$. The Gaussian kernel basis function $\phi(\mathbf{v}, \mathbf{v}_k) = \exp(-\|\mathbf{v} - \mathbf{v}_k\|_2^2 / 2h^2)$ is used as the basis function in this simulation. We choose h to be a half of the distance between two spatial basis nodes. The white noise $e_i(\mathbf{v}, t)$ are generated from a standard Normal distribution, $N(0, 1)$. The datasets are generated under three scenarios: no spatio-temporal dependence, spatial dependence exists but no temporal dependence, and spatio-temporal dependence. For each scenario, we estimate the model parameters using the proposed spatial-temporal model, and two alternative methods:

- Independent: fit a linear regression model on the two covariates without considering spatio-temporal dependence.

- Spatial: add spatial dependence to the linear model but ignore the temporal dependence.

For each method, we calculate the means and the standard errors of the estimation of each parameter over the 100 replications. We report the mean of squared difference between the true spatial covariance matrix and the estimate spatial covariance matrix over the total number of spatial locations as a test statistic (TS) to illustrate the accuracy of the estimation of the spatial dependence structure. The formula of the test statistic is expressed by

$$TS = \frac{\sum_{i,j} |(\boldsymbol{B}\Sigma^0 \boldsymbol{B}^\top - \boldsymbol{B}\hat{\Sigma}\boldsymbol{B}^\top)|_{i,j}}{M} \tag{2.6}$$

13

where $\hat{\Sigma}$ is the estimated spatial covariance matrix. The results are reported in Table 2.2. For the independent data, the performances of the methods incorporating spatial or spatial-temporal dependence are similar as the independent model. When the data is spatially correlated, the spatial model and our proposed model outperform the independent model. When the data is spatially and temporally dependent, our model greatly outperforms the independent and spatial model.

Table 2.2. Mean parameter estimates and standard error of mean (in parenthesis) of over 100 replications. The best value in each row is highlighted in blue color.

Data	True Value	Methods		
		Independent	Spatial	Spatio-Temporal
Independent	$\mu = 0.1$	0.0998 (1.706×10^{-4})	0.0996 (2.120×10^{-4})	0.0996 (2.142×10^{-4})
	$\beta_1 = 0.5$	0.500 (1.996×10^{-5})	0.500 (2.296×10^{-5})	0.500 (2.321×10^{-5})
	$\beta_2 = 1$	1.000 (1.735×10^{-5})	1.000 (2.053×10^{-5})	1.000 (2.077×10^{-5})
	$\sigma^2 = 1$	1.000 (6.698×10^{-5})	0.969 (6.525×10^{-5})	0.974 (6.530×10^{-5})
Spatial	$\mu = 0.1$	0.0850 (5.112×10^{-3})	0.0994 (4.427×10^{-4})	0.0992 (4.725×10^{-4})
	$\beta_1 = 0.5$	0.430 (2.479×10^{-2})	0.500 (4.563×10^{-5})	0.500 (2.362×10^{-4})
	$\beta_2 = 1$	0.860 (4.957×10^{-2})	1.000 (4.752×10^{-5})	1.000 (2.523×10^{-4})
	$\sigma^2 = 1$	1.701 (9.807×10^{-2})	1.000 (6.443×10^{-5})	0.999 (1.221×10^{-4})
	TS		9.509×10^{-5} (1.396×10^{-6})	9.509×10^{-5} (1.374×10^{-6})
Spatio-Temporal	$\mu = 0.1$	0.00555 (1.791×10^{-3})	0.00407 (8.206×10^{-4})	0.100 (5.570×10^{-4})
	$\beta_1 = 0.5$	0.0155 (2.260×10^{-4})	0.0176 (1.128×10^{-4})	0.500 (7.035×10^{-5})
	$\beta_2 = 1$	0.0314 (1.716×10^{-4})	0.0353 (8.127×10^{-5})	1.000 (4.674×10^{-5})
	$\sigma^2 = 1$	83.20 (8.007×10^{-3})	5.005 (2.057×10^{-4})	1.000 (6.235×10^{-5})
	TS		24.360 (7.497×10^{-4})	1.854×10^{-4} (2.457×10^{-6})

2.5.2 Noise to Signal Ratio Analysis

Next, we investigate the effect of different levels of noise to signal ratio on parameter estimation. We define the noise-signal ratio (NSR) by

$$NSR = \frac{J\sigma^2}{\sum_j^J \Sigma_{j,j}^0}.$$

We generate the simulation datasets the same way as the model misspecification simulation study, but the white noise generation scenarios vary by the variance σ^2. The means and the standard errors of the estimation of each parameter over 100 replications are given in Table 2.3. As we increase the noise to signal ratio, the proposed spatio-temporal model consistently gives accurate parameter estimates. The results demonstrate that the proposed method is robust with respect to large noise to signal ratio.

2.5.3 Model Performance on Complex Connectivity

To evaluate our model performance on a complex connectivity matrix, we generate the simulation dataset using the $\hat{\Sigma}$ estimated from the cocaine use disorder fMRI dataset in section 2.6. As none of the linear coefficients is significant in the real data, we use the same coefficients as simulation studies in Section 2.5.1. We report the mean estimates and standard error of the mean for both the linear regression model and our proposed model over 50 replications in Table A.4. The results show the stable performance of with covariance estimated using real data.

2.5.4 Sensitivity Analysis

To obtain practical guidance for the choice of model parameters, i.e., the number of subjects N, the number of volumes L, the number of spatial basis nodes J, and

Table 2.3. Mean parameter estimates and standard error of mean (in parenthesis) of over 100 replications under different noise-signal ratio.

Noise-Signal Ratio	True Value	Spatiao-Temporal Model
NSR=0.01	$\mu = 0.1$	$0.100\ (1.521\times10^{-4})$
	$\beta_1 = 0.5$	$0.500\ (1.661\times10^{-5})$
	$\beta_2 = 1$	$1.000\ (1.482\times10^{-5})$
	$\sigma^2 = 0.1$	$0.100\ (6.718\times10^{-6})$
	TS	$1.265\ \times10^{-4}\ (1.831\times10^{-6})$
NSR=0.1	$\mu = 0.1$	$0.100\ (5.570\times10^{-4})$
	$\beta_1 = 0.5$	$0.500\ (7.035\times10^{-5})$
	$\beta_2 = 1$	$1.000\ (4.674\times10^{-5})$
	$\sigma^2 = 1$	$1.000\ (6.235\times10^{-5})$
	TS	$1.854\ \times10^{-4}\ (2.457\times10^{-6})$
NSR=1	$\mu = 0.1$	$0.100\ (1.421\times10^{-3})$
	$\beta_1 = 0.5$	$0.500\ (1.896\times10^{-4})$
	$\beta_2 = 1$	$1.000\ (1.385\times10^{-4})$
	$\sigma^2 = 10$	$9.998\ (7.848\times10^{-4})$
	TS	$1.872\ \times10^{-4}\ (3.409\times10^{-6})$
NSR=10	$\mu = 0.1$	$0.0987\ (1.988\times10^{-3})$
	$\beta_1 = 0.5$	$0.500\ (2.320\times10^{-4})$
	$\beta_2 = 1$	$1.001\ (2.730\times10^{-4})$
	$\sigma^2 = 100$	$99.99\ (7.327\times10^{-3})$
	TS	$2.032\ \times10^{-4}\ (3.966\times10^{-6})$

Table 2.4. Simulation study mean parameter estimates and standard error of mean (in parenthesis) for data simulated from real data spatial covariance structure over 50 replications.

True Value	Model	
	Linear Regression	Spatio-Temporal Model
$\mu = 0.1$	$0.00314\ (2.313\times10^{-4})$	$0.0996\ (2.522\times10^{-4})$
$\beta_1 = 0.5$	$0.01577\ (3.157\times10^{-5})$	$0.5001\ (3.099\times10^{-5})$
$\beta_2 = 1$	$0.03146\ (2.213\times10^{-5})$	$1.000\ (2.143\times10^{-5})$
$\sigma^2 = 1$	$83.305\ (1.264\times10^{-3})$	$0.9997\ (7.626\times10^{-5})$
TS		$4.413\ \times10^{-6}\ (2.830\times10^{-8})$

16

different types of basis functions. The simulated datasets are generated in the same way as those in Section 2.5.3, with varying model parameters. We demonstrate the detailed experimental settings and results in A.7.

In our sensitivity analysis, we first verify that the parameter estimation of the proposed model is accurate and stable with small numbers of subjects N and volumes L. Second, we analyze the model misspecification with different numbers of nodes J and three types of basis functions: the Gaussian, exponential, and bisquare. We find that the Gaussian basis function and $J = 125$ work well even when the data are generated from other basis functions and different J. Thus we adopt this setting for the real data analysis in Section 2.6.

2.6   Application to Multi-Subject FMRI Data

We apply the proposed spatio-temporal model to the multi-subject FMRI data described in Section 2.2. The real data is standardized before processed to our model since the range of the data is very different across subjects. The modeling details and some model fitting results are given in the A.3. The scientific objective of this study is to detect the effect of cocaine use disorder on functional connectivity. Based on the proposed spatio-temporal model, it is equivalent to test whether or not the subjects from the cocaine use disorder group and the control group share the same spatial covariance structure. In this section, we first compare the functional connectivity identified by the proposed model with benchmark brain networks in the literature to confirm the effectiveness of our model. Then we use the estimated spatial covariance to identify the difference of functional connectivity between cocaine use disorder and control groups.

2.6.1 A Comparison with Benchmark Brain Networks

We compare the functional connectivity identified by the proposed spatio-temporal model with benchmark brain networks in the literature. The objective of this comparison is to demonstrate the effectiveness of using the proposed spatial model to identify functional connectivity. In this study, we use atlas of Intrinsic Connectivity of Homotopic Areas (AICHA, Joliot et al., 2015) to define regions of interest (ROIs), that partitions human brain into cortical and subcortical areas based on different structures and functions. The resulted atlas contains 384 ROIs.

Based on the proposed model, the functional connectivity is characterized by the covariance matrices of the large-scale spatial structure $f_{i,l}(\mathbf{v})$ in (2.1). When the spatial conditional covariance between two regions are higher than those between any pair of two arbitrary regions, we claim that those two regions have identifiable functional connectivity. Based on ROIs from the AICHA brain atlas, we average the estimated voxel-level brain spatial covariance values within the same region to create a 384×384 by-regional spatial covariance matrix S. To identify the region pairs with higher conditional covariance, we applied the graphical lasso method (Friedman et al., 2007) to obtain a sparse precision matrix $\hat{R}$ by solving

$$\hat{R} \in \operatorname{argmin}_{R \succeq 0} \left\{ \operatorname{tr}(SR) - \log \det(R) + \rho \sum_{i \neq j} |R_{ij}| \right\}, \qquad (2.7)$$

where ρ is the penalty parameter, and we specify it to be 0.3 in this study. A non-zero entry in $\hat{R}$ represents a high conditional covariance value between two regions, which indicates the functional connectivity.

We compare the functional connectivity identified from our model with four well-studied brain networks based on AICHA atlas:

- The default mode network (DM): it is usually active in resting state or mind-

18

wandering, and known to be involved in experience that are related to external task performance.

- The dorsal attention network (DA): it is one of two sensory orienting systems in the human brain, and known been related to voluntary orienting.

- The executive control (CE) network: it is involved in cognitive functions, problem solving, and decision making.

- The salience network (SA): it is involved in detecting and filtering salient stimuli.

These well-studied functional networks are known to have relatively stronger spatial dependence than random region pairs which do not belong to one of them. For each well-studied network, we compute the mean difference between the absolute values of conditional covariance in this network and those do not belong to any of these four networks, and use t-test to assess the significance of this mean difference. The results are given in Table 2.5. We see that the differences are all positive and significant at certain levels for both groups, which indicates that the conditional covariance values led by the proposed model match the results from those four well-studied networks (Smith et al., 2009).

Table 2.5. The mean differences and their significance between the absolute condition covariance values of each well-studied network and those do not belong to any of the well-studied networks.

	Cocaine Use Disorder Group		Control Group	
Networks	Difference	p-value	Difference	p-value
DM	0.054	<0.0001	0.055	<0.0001
DA	0.002	0.1029	0.003	0.0153
CE	0.022	<0.0001	0.023	<0.0001
SA	0.005	0.0008	0.005	0.0030

2.6.2 Statistical Comparisons on the Functional Connectivity between the Cocaine Use Disorder Group and the Control Group

In this section, we compare the differences of the functional connectivity between the cocaine use disorder group and the control group. According to model (2.3), the differences of functional connectivity are determined by the differences between the covariance matrices Σ_1 and Σ_2. Therefore, we first test the hypothesis that $H_0 : \Sigma_1 = \Sigma_2$. If H_0 is rejected, we conclude that the global difference on the functional connectivity between the two groups are significant.

Since the number of subjects is small and the number of unknown parameters is large, a parametric test procedure (i.e., goodness of fit test) is inaccurate in assessing the significance of the hypothesis. Alternatively, we propose to use a nonparametric test procedure to test the hypothesis. Let $\hat{\Sigma}_c$ be the estimated covariance matrix under H_0 (i.e., the two groups share the same covariance matrix), and $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ be the estimated covariance matrices of the cocaine use disorder group and the control group, respectively. Then the test statistic is constructed by

$$\sum_{g=1}^{2} \left\{ \sum_{j=1}^{J} (\lambda_{g,j} - \lambda_{c,j})^2 \right\}, \tag{2.8}$$

where $\{\lambda_{c,1}, \lambda_{c,2}, \ldots, \lambda_{c,J}\}$ and $\{\lambda_{g,1}, \lambda_{g,2}, \ldots, \lambda_{g,J}\}$ are ordered eigenvalues of $\hat{\Sigma}_c$ and $\hat{\Sigma}_g$ for $g = 1, 2$. If the value of the test statistic is large, we conclude that the difference of Σ_1 and Σ_2 is large. We permute the group indexes of the subjects to construct an empirical distribution of this test statistic. After 400 permutations, the p-value is 0.582, which shows that, the global difference on functional connectivity between the two groups is not significant. In addition, a simulation study is provided in the A.4 (see Table A.1 in the A.4) to verify the effectiveness of the test procedure on synthetic datasets.

20

To further explore the local-level differences of functional connectivity between the two groups, we apply the joint graphical lasso method (JGL, Danaher et al., 2013) with a group penalty to obtain sparse precision matrices (denoted by R_1 and R_2) based on the estimated spatial covariance matrices (Σ_1 and Σ_2) of the two groups by solving

$$\hat{R}_1, \hat{R}_2 \in \mathrm{argmax}_{R_1, R_2 \succeq 0} \left\{ \sum_{k=1}^{2} (\log \det R_k - \mathrm{tr}(\Sigma_k R_k)) - P(\{R_1, R_2\}) \right\}, \qquad (2.9)$$

$$P(\{R_1, R_2\}) = \lambda_1 \sum_{k=1}^{2} \sum_{i \neq j} |R_{k,i,j}| + \lambda_2 \sum_{i \neq j} \sqrt{\sum_{k=1}^{K} (R_{k,i,j})^2},$$

where λ_1 and λ_2 are two penalty parameters-associated with the Lasso penalty and the group penalty, respectively. The Lasso penalty encourages sparsity, while the group penalty will drive spatial covariance values in two groups to be the same. After obtaining $\hat{R}_1$ and $\hat{R}_2$ from (2.9), we take the difference between these two matrices and denote it as ΔR ($\Delta R = \hat{R}_1 - \hat{R}_2$). The region pairs that associate with non-zero values in ΔR are considered to have different spatial covariance between two groups. By specifying $\lambda_1 = 0.4$ and $\lambda_2 = 0.5$, the identified region pairs are provided in Table 2.6.

Table 2.6. The brain regions with large differences (ΔR) on functional connectivity (in terms of estimated spatial covariance) between the cocaine use disorder and control groups.

Region Pair	Estimated functional connectivity		ΔR
	Cocaine	Control	
R MOFC subregion #2 & R MOFC subregion #1	0.9960	1.3349	-0.3389
L MOFC subregion #2 & R MOFC subregion #1	0.9243	1.2395	-0.3152
R MOFC subregion #2 & L MOFC subregion #2	0.8122	1.0948	-0.2826
R SMA subregion #1 & L SMA subregion #1	0.5336	0.6034	-0.0698
R SMA subregion #1 & R SMA subregion #3	0.3416	0.3870	-0.0454
R SMA subregion #1 & L cingulate cortex #2	0.3136	0.3512	-0.0376
R SMA subregion #1 & R cingulate cortex #2	0.3163	0.3551	-0.0388

In the top panel of Table 2.6, we see that the participants with cocaine use

21

disorder (CUD) had smaller functional connectivity (within the default mode network) between right medial orbitofrontal cortex (MOFC) subregion #2 and right MOFC subregion #1 (Cocaine-Control=-0.3389), between left MOFC subregion #2 and right MOFC subregion #1 (Cocaine-Control=-0.3152), and between right MOFC subregion #2 and left right MOFC subregion #2 (Cocaine-Control=-0.2826). The anatomical region MOFC largely overlays with the functional region ventromedial prefrontal cortex (VMPFC) in the DM network (Hiser and Koenigs, 2018). Thus, the results suggest that the CUD had altered VMPFC functional connectivity (i.e., reduced VMPFC function) than the controls. Consistently, the linkage between decreased VMPFC function and transition to cocaine use disorder has been well documented (see Bolla et al., 2003; Franklin et al., 2002; Lim et al., 2002 and Volkow et al., 1992). The VMPFC is critically involved in impulsivity (Ben-Shahar et al., 2012). Thus, reduced VMPFC function in the CUD groups could suggest the reduced ability of CUD individuals in inhibitory control over behavior.

In addition, the results in the bottom panel of Table 2.6 show that the CUD individuals have lower functional connectivity between the bilateral supplementary motor areas than controls, which have not been documented in the literature to the best of our knowledge. The supplementary motor area (SMA) plays important roles in the planning, initiation, and execution of motor acts (Amador and Fried, 2004). Impaired sensor-motor performance and altered functional connectivities involving supplementary motor area during finger tapping task have been reported in cocaine use disorder individuals (Lench et al., 2017). Thus, our finding on the supplementary motor area suggests that the connectivity between bilateral supplementary motor areas may be part of neuronal circuits underlying impaired sensor-motor performance. Given the sample size and relatively small group difference, these novel results need to be replicated in future studies with different sample.

A MULTI-LAYER REGULARIZATION APPROACH FOR GAUSSIAN GRAPHICAL MODELING WITH APPLICATION TO THE COMPARISON OF FUNCTIONAL CONNECTIVITY ACROSS DIFFERENT FACTORS

3.1   Introduction

Functional connectivity in brain is a representation of the complex relationship among billions of highly interactively brain neurons or regions (Bullmore and Sporns, 2009). Functional connectivity between neurons or brain region of interests (ROIs) is defined as the statistical correlation between the measures of their activities (Fox and Raichle, 2007). It is an effective tool for studying brain diseases, such as Alzheimer's disease (AD), Parkinson's disease, attention deficit hyperactivity disorder (ADHD), and substance use disorders (Filippi et al., 2019, Sutherland et al., 2012). In such studies, it is crucial to create sparse functional activity maps in brain under different treatment conditions and identify their difference.

In spatial statistics, the conditional correlation between different parts of human brains can be represented by the precision matrix using multivariate Gaussian models. Each element of the precision matrix is corresponding to the conditional dependency between a pair of brain neurons, voxels, or ROIs. Bayesian shrinkage methods have been long proposed for precision matrix estimation. A number of multiple graphs joint estimation approaches were introduced to brain network research in recent years (Carvalho et al., 2010, Piironen and Vehtari, 2017, Li et al., 2019). Yajima et al. (2012) proposed a framework and an efficient computational algorithm to study Gaussian

23

directed acyclic graphs. Peterson et al. (2015) used the Bayesian Markov random field approach for estimating multiple interaction networks. Piironen and Vehtari (2017) developed an integrative modeling approach for jointly modeling multiple brain networks, which allows flexible priors on the edge probabilities. Although these methods are very powerful in estimating functional connectivity in brain, they are often problematic in analyzing large-scale spatial-temporal data collected from multiple subjects because of their heavy computational burden (Cohen et al., 2017).

Many researchers use Gaussian graphical models to study functional connectivity in brain (Smith et al., 2011) by considering it as a graph. Borrowing ideas from graph theory, the functional network in brain can be represented by a graph, which is defined as the set of neurons as random variables and the set of their pairwise conditional dependencies, as edges among them (Butts, 2009). Gaussian graphical model (GGM) is a modeling technique that represents the relationships between a set of random variables through their joint (Gaussian) distribution. In recent years, numerous studies (for example Meinshausen and Bühlmann, 2006, Yuan and Lin, 2007, Friedman et al., 2008, Witten et al., 2009, Rothman et al., 2008, Zhang and Wang, 2010) have been published on learning differential networks under GGM due to its computational convenience and straightforward interpretation (Uhler, 2017). The majority of these studies are focused on the generalization of the graphical lasso method, which obtains symmetric positive definite sparse precision matrices of Gaussian graphical models via conditional distributions of random variables using L_1 penalty. For example, Zhang and Wang (2010) and Chiquet et al. (2010) used the neighborhood selection method with the fused lasso penalty, graphical intertwined lasso penalty, and cooperative-lasso penalty, respectively, to effectively extract structural changes in Gaussian graphical models. Zhao et al. (2014) directly estimated the difference between the precision matrices using a constrained L_1 regularization. Friedman et al.

24

(2008) introduced a coordinate descent procedure to efficiently solve the graphical lasso. Rothman et al. (2008) further studied the theoretical properties of the L_1 penalized maximum likelihood estimator for high-dimensional precision matrix, which can use the L_1 penalty to shrink some of the insignificant off-diagonal elements to zero. These methods estimate graphical models under different categories separately and use multiple hypothesis testing to compare estimated sparse precision matrices and obtain the difference between treatment conditions.

Based on the development of the Gaussian graphical model, methods that jointly estimate multiple Gaussian graphical models were proposed to improve the estimation for different categories sharing common features. Guo et al. (2011) used a hierarchical penalty to jointly estimate multiple graphical models in a non-convex setting. Yuan and Lin (2006) penalized the joint log-likelihood using the group penalty to encourage similarities of the estimated precision matrices. Danaher et al. (2014) imported the fused graphical lasso penalty and used the alternating direction method of multipliers (ADMM) to find the estimator. However, these methods suffer from a high false discovery rate (Liu, 2013). The main reason is their underlying assumption that treatment conditions have equivalent effects across the entries of precision matrices, which makes the spurious signals identified from conditional covariances significantly increase the estimation variation. Also, the computational costs of these methods increase dramatically while increasing the number of categories.

To accomplish an accurate and efficient joint estimation of multiple functional networks across multiple factors, we borrow ideas from the joint graphical lasso method (Danaher et al., 2014), and propose to use an additional joint hierarchical fused lasso penalty on the graphical lasso model. Our method assigns different tuning parameters for the contrast of each factor to allow different treatments or factors having different impacts across brain regions, and at the same time, achieves

25

computational efficiency.

3.2 Joint Graphical Lasso

Joint graphical lasso (JGL) is a sparse penalized maximum likelihood estimator for the precision matrix of multivariate distributions introduced by Danaher et al. (2014). The original JGL method is proposed to estimate Gaussian graphical models on multiple distinct classes of high-dimensional datasets. Consider K independently identically distributed multivariate Gaussian datasets. The covariance structure of each dataset is $\Sigma^{(k)}$. Let $S^{(k)}$ be the empirical estimation of $\Sigma^{(k)}$, and $\boldsymbol{\Theta}^{(k)}$ be the estimated sparse precision matrix for $\Sigma^{(k)}$, the JGL objective function is:

$$\max_{\boldsymbol{\Theta}} \left\{ \sum_{k=1}^{K} n_k \left(\log \det \boldsymbol{\Theta}^{(k)} - \left(S^{(k)} \boldsymbol{\Theta}^{(k)} \right) \right) - P(\boldsymbol{\Theta}) \right\} \tag{3.1}$$

subject to the constraint that $\boldsymbol{\Theta}^{(1)}, \ldots, \boldsymbol{\Theta}^{(K)}$ are positive definite. Here, $\{\boldsymbol{\Theta}\} = \{\boldsymbol{\Theta}^{(k)}; k = 1, \ldots, K\}$, and $P(\Theta)$ is a convex penalty function. Danaher et al. (2014) introduced two useful penalty functions with the original JGL method:

1. The fused penalty:

$$P(\{\boldsymbol{\Theta}\}) = \lambda_1 \sum_{k=1}^{K} \sum_{i \neq j} |\boldsymbol{\Theta}_{i,j}^{(k)}| + \lambda_2 \sum_{k < k'} \sum_{i,j} |\boldsymbol{\Theta}_{i,j}^{(k)} - \boldsymbol{\Theta}_{i,j}^{(k')}| \tag{3.2}$$

2. The group penalty:

$$P(\{\boldsymbol{\Theta}\}) = \lambda_1 \sum_{k=1}^{K} \sum_{i \neq j} |\boldsymbol{\Theta}_{i,j}^{(k)}| + \lambda_2 \sum_{i \neq j} \sqrt{\sum_{k=1}^{K} \left(\boldsymbol{\Theta}_{i,j}^{(k)} \right)^2} \tag{3.3}$$

For both of the penalty functions, λ_1 controls the sparsity of the estimated precision matrices, and λ_2 controls the similarity of the estimated precision matrices across factor levels. Danaher et al. (2014) used the ADMM algorithm to solve this problem. This problem has a closed-form solution for $K = 2$.

26

Our goal is to study the effect of factors, such as gender, ethnicity, and genetic phenotype, on functional connectivity in brain. As noted earlier, we define functional connectivity as groups of regions/voxels with high spatial correlations (Friston, 2011). To solve this problem, we need to estimate a sparse precision matrix for each level of each factor and then compare these matrices to identify similar or distinct covariance patterns across factors.

Suppose we have a dataset which includes a total of N fMRI brain image scans from N subjects. Each subject is associated P categorical factors, denoted by $z_1, \ldots, z_p,$ $\ldots, z_P$, and pth factor has K_p levels. Let Y_n be a tensor observation of four dimensions for the subject n, $Y_n \in \mathbb{R}^{r_1, \ldots, r_4}$, where r_1, r_2, and r_3 represent the three-dimensional brain space, and r_4 represents the time. The JGL method can not be directly used to estimate the sparse precision matrices for this type of data due to its high dimension. Instead, we can first estimate the spatial covariance matrix of this type of data and adopt the JGL method with slight modification to obtain their sparse precision matrices.

Assume that each combination of factor expression, $(z_1, \ldots, z_P)$, has a unique manifestation of the functional connectivity in brain. In this study, each combination of factors corresponds to a brain voxel spatial covariance matrix pattern, which is denoted by $\Sigma_{z_1, \ldots, z_P}$. For notational convenience, we denote a unique combination of factor expression $(z_1, \ldots, z_P)$ as l and simplify the notation of $\Sigma_{z_1, \ldots, z_P}$ as Σ_l. As a result, the total number of unique combination is $L = \prod_{p=1}^{P} K_p$. We can also denote $\boldsymbol{\Theta}_l$ as the corresponding sparse precision matrix of spatial covariance matrix Σ_l, and let $\boldsymbol{\Theta}_{l,i,j}$ be the entry of $\boldsymbol{\Theta}_l$ at location (i, j). The JGL objective function with a fused penalty can be expressed as follows:

$$\max_{\{\boldsymbol{\Theta}\}} \left\{ \sum_{l=1}^{L} \left(\log \det \boldsymbol{\Theta}_l - (\Sigma_l \boldsymbol{\Theta}_l) \right) - P(\boldsymbol{\Theta}) \right\} \tag{3.4}$$

27

$$P(\boldsymbol{\Theta}) = \lambda_1 \sum_{l=1}^{L} \sum_{i,j} |\boldsymbol{\Theta}_{l,i,j}| + \lambda_2 \sum_{l=1}^{L} \sum_{\substack{l'=1 \\ l>l'}}^{L} \sum_{i,j} |\boldsymbol{\Theta}_{l,i,j} - \boldsymbol{\Theta}_{l',i,j}|$$

Basically, the JGL method considers P factors with a total number of $\prod_{p=1}^{P} K_p$ levels. Following this approach, the sparse precision matrices estimation suffers two major drawbacks:

- Factors could impact functional connectivity in brain on different brain areas at different levels. In JGL, all entries of the precision matrices share the same penalty λ_2. On such a type of highly noisy data, it would possibly create false signals or ignore mild signals by using inappropriate λ_1 and λ_2.

  Consider a case in which factor A has big impacts on functional connectivity in brain. Different factor levels of A will create large signal differences on some parts of the brain between subjects, while factor B has significant impacts, but the signal differences it creates are not large. The JGL method will erase the impact of factor B if we try to verify the impact of factor A using large tuning parameters. Or we could create false signals on some parts of the brain which are affected by neither factor if we use small tuning parameters to embrace the effect of factor B.

- It's computational complexity is high when the number of factors, factor levels, and the size of covariance matrices is large. This method needs to process a total of $C(\prod_{p=1}^{P} K_p, 2)$ pairs of covariance matrices. Let J be the size of the covariance matrices, this computational complexity of this method will be $O(J^2(\prod_{p=1}^{P} K_p)^2)$.

3.3 Proposed Method

We propose a hierarchical penalty function which uses different penalties for each of the factors to estimate sparse precision matrices associated with multiple categorical factors. If we denote $\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P}$ as the corresponding sparse precision matrix of spatial covariance matrix $\Sigma_{z_1,\ldots,z_p=k_p,\ldots,z_P}$ when pth factor z_p is at level k_p, and $\{\boldsymbol{\Theta}\} = \{\boldsymbol{\Theta}_{z_1,\ldots,z_p,\ldots,z_P}; z_p = 1,\ldots,K_p, p = 1,\ldots,P\}$ is the set of all $\boldsymbol{\Theta}_{z_1,\ldots,z_P}$'s. This hierarchical penalty function can be expressed by

$$
\begin{aligned}
P(\boldsymbol{\Theta}) = & \lambda_0 \sum_{p=1}^{P} \sum_{k_p=1}^{K_p} \sum_{i,j} |\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P,i,j}| \\
& + \sum_{p=1}^{P} \lambda_p \left\{ \sum_{z_1=1}^{K_1} \cdots \sum_{z_{p-1}=1}^{K_{p-1}} \sum_{z_{p+1}=1}^{K_{p+1}} \cdots \sum_{z_P=1}^{K_P} \left(\sum_{k_p<k_p'} \sum_{i,j} |\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P,i,j} - \boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p',\ldots,z_P,i,j}| \right) \right\}.
\end{aligned}
$$
(3.5)

Here, λ_0 and λ_p's are nonnegative penalty parameters. This penalty will provide sparse estimations of $\{\boldsymbol{\Theta}\}$ when the tuning parameter λ_0 is large. λ_p's control the within-class similarity. Elements of $\boldsymbol{\Theta}_{z_1,\ldots,z_p,\ldots,z_P}$'s will be identical across classes of factor z_p when the tuning parameter λ_p is large. By assigning a different penalty parameter, λ_p, to each factor, we are able to obtain a better estimation of precision matrices by weighting different factors. Instead of comparing $C(\prod_{p=1}^{P} K_p, 2)$ pairs of covariance matrices as JGL method does, our method only compares a total of $\sum_{q=1}^{P}(\sum_{p=1}^{P} K_p - K_q)\binom{K_q}{2}$ pairs of covariance matrices. The computational complexity is reduced from $O(J^2(\prod_{p=1}^{P} K_p)^2)$ to $O(J^2 \sum_{p=1}^{P} K_p^2)$.

We modify the ADMM algorithm in Danaher et al. (2014) to solve the objective in (3.4). Under the constraint that $\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P}$ is positive definite for all $p = 1,\ldots,P$ and $k_p = 1,\ldots,K_p$, the scaled augmented Lagrangian (Boyd, 2011) for this problem is given by:

29

$$\mathcal{L}_\rho\left(\{\boldsymbol{\Theta}\},\{\mathbf{Z}\},\{\mathbf{U}\}\right) = -\sum_{p=1}^{P}\sum_{k_p=1}^{K_p}\left(\log\det\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P} - \left(\Sigma_{z_1,\ldots,k_p,\ldots,z_P}\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P}\right)\right)$$

$$+P(\{\mathbf{Z}\}) + \frac{\rho}{2}\sum_{p=1}^{P}\sum_{k_p=1}^{K_p}||\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P} - \mathbf{Z}_{z_1,\ldots,z_p=k_p,\ldots,z_P} + \mathbf{U}_{z_1,\ldots,z_p=k_p,\ldots,z_P}||_F^2, \qquad (3.6)$$

where $\mathbf{Z}_{z_1,\ldots,z_p=k_p,\ldots,z_P} = \boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P}$, $\{\mathbf{Z}\} = \{\mathbf{Z}_{z_1,\ldots,z_p=k_p,\ldots,z_P}\}$ for $k_p = 1,\ldots,K_p$. $\{\mathbf{U}\} = \{\mathbf{U}_{z_1,\ldots,z_p=k_p,\ldots,z_P}\}$ for $k_p = 1,\ldots,K_p$ are dual variables. The ADMM algorithm to solve our problem is given as follows:

(a) Initialize the variables: $\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P} = \mathbf{I}$, $\{\mathbf{Z}\} = \mathbf{0}$, $\{\mathbf{U}\} = \mathbf{0}$, for $p = 1,\ldots,p$ and $k_p = 1,\ldots,K_p$.

(b) Select a scalar $\rho > 0$.

(c) For $i = 1,2,3,\ldots$ until convergence:

   (i) For $p = 1,\ldots,P$ and $k_p = 1,\ldots,K_p$, update $\boldsymbol{\Theta}^{(i)}_{z_1,\ldots,z_p=k_p,\ldots,z_P}$ as the minimizer (with respect to $\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P}$) of

$$-\left(\log\det\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P} - \left(\Sigma_{z_1,\ldots,z_p=k_p,\ldots,z_P}\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P}\right)\right)$$
$$+\frac{\rho}{2}||\boldsymbol{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P} - \mathbf{Z}^{(i-1)}_{z_1,\ldots,z_p=k_p,\ldots,z_P} + \mathbf{U}^{(i-1)}_{z_1,\ldots,z_p=k_p,\ldots,z_P}||_F^2, \qquad (3.7)$$

   Let $\{\mathbf{VDV}\}^\top$ denotes the eigendecomposition of $\{\Sigma_{z_1,\ldots,z_p=k_p,\ldots,z_P} - \rho\mathbf{Z}^{(i-1)}_{z_1,\ldots,z_p=k_p,\ldots,z_P} + \rho\mathbf{U}^{(i-1)}_{z_1,\ldots,z_p=k_p,\ldots,z_P}\}$, the solution of $\boldsymbol{\Theta}^{(i)}_{z_1,\ldots,z_p=k_p,\ldots,z_P}$ is given (Witten et al., 2009) by $\{\mathbf{V\tilde{D}V}\}^\top$, where $\tilde{\mathbf{D}}$ is the diagonal matrix with jth diagonal element:

$$\frac{1}{2\rho}\left(-D_{jj} + \sqrt{D_{jj}^2 + 4\rho}\right) \qquad (3.8)$$

   (ii) Update $\{\mathbf{Z}^i\}$ as the minimizer (with respect to $\{\mathbf{Z}\}$) of

30

$$\frac{\rho}{2}||\mathbf{Z}_{z_1,\ldots,z_p=k_p,\ldots,z_P} - \left(\mathbf{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P}^{(i)} + \mathbf{U}_{z_1,\ldots,z_p=k_p,\ldots,z_P}^{(i-1)}\right)||_F^2 + P\left(\{\mathbf{Z}\}\right)$$
(3.9)

(iii) For $p = 1, \ldots, P$ and $k_p = 1, \ldots, K_p$, update

$$\mathbf{U}_{z_1,\ldots,k_p,\ldots,z_P}^{(i)} \leftarrow \mathbf{U}_{z_1,\ldots,z_p=k_p,\ldots,z_P}^{(i-1)} + \left(\mathbf{\Theta}_{z_1,\ldots,z_p=k_p,\ldots,z_P}^{(i)} - \mathbf{Z}_{z_1,\ldots,z_p=k_p,\ldots,z_P}^{(i)}\right)$$
(3.10)

.

The final $\left\{\hat{\mathbf{\Theta}}\right\}$ that result from this algorithm are the estimates of $\left\{\hat{\mathbf{\Sigma}}^{-1}\right\}$. This algorithm is guaranteed to converge to the global optimum (Boyd, 2011).

Now consider a simple case when $P = 2$ and $K_p = 2$ for $p = 1, 2$. The penalty function in (3.5) takes the form:

$$\min_{\{\mathbf{Z}\}}\left\{ \sum_{p=1}^{2}\sum_{q=1}^{2}||\mathbf{Z}_{p,q} - \mathbf{A}_{p,q}||_F^2 + \lambda_0 \sum_{p=1}^{2}\sum_{q=1}^{2}\sum_{i,j,i\neq j}|Z_{p,q,i,j}| \right.$$
$$\left. + \sum_{p=1}^{2}\lambda_1\sum_{i,j}|Z_{p,2,i,j} - Z_{p,1,i,j}| + \sum_{q=1}^{2}\lambda_2\sum_{i,j}|Z_{2,q,i,j} - Z_{1,q,i,j}| \right\}.$$
(3.11)

(3.11) is completely separable with respect to each pair of matrix elements (i, j), then this function can be further simplified for each (i, j) as:

$$\min_{\{\mathbf{Z}\}}\left\{ \sum_{p=1}^{2}\sum_{q=1}^{2}(Z_{p,q,i,j} - A_{p,q,i,j})^2 + \lambda_0 \sum_{p=1}^{2}\sum_{q=1}^{2}|Z_{p,q,i,j}| \right.$$
$$\left. + \sum_{p=1}^{2}\lambda_1|Z_{p,2,i,j} - Z_{p,1,i,j}| + \sum_{q=1}^{2}\lambda_2|Z_{2,q,i,j} - Z_{1,q,i,j}| \right\}.$$
(3.12)

Note that, under $\lambda_0 = 0$, the solution to this problem does not have a closed-form expression. Here, we use a numerical optimizer to solve this problem and then use the soft-thresholding operator defined as $S(x, c) = sgn(x)(|x| - c)_+$ to find the initial solution of $\{\mathbf{Z}\}$.

31

3.4   Simulation Studies

In this section, we conduct a simulation to compare the performance of our method and the JGL method. We consider a simple case that includes two factors, each with two levels. We generate four precision matrices of size 10×10 with pre-specified covariance structures (denote by $\boldsymbol{\Theta}^0_{(1,1)}$, $\boldsymbol{\Theta}^0_{(2,1)}$, $\boldsymbol{\Theta}^0_{(1,2)}$ and $\boldsymbol{\Theta}^0_{(2,2)}$) to represent the true spatial precision matrices of functional connectivity in brain. Each precision matrix associates with one level combination. We take the inversion of these matrices to obtain the covariance matrices (denote by $\Sigma^0_{(1,1)}$, $\Sigma^0_{(2,1)}$, $\Sigma^0_{(1,2)}$ and $\Sigma^0_{(2,2)}$). To mimic the noise contained in real data, we generate 2000 samples from each $\mathcal{N}(0, \Sigma^0_{z_1,z_2})$ with a standard normal white noise as samples from each level of the factors, and calculate their empirical covariance matrices. We then apply the original JGL method and our method to this data to estimate the sparse precision matrices for this simulation dataset.

We are interested in the accuracy of the estimation on the areas with prespecified covariance structures, so we use the summation of absolute differences of elements within these prespecified areas between the true precision matrix and the estimated sparse precision matrix as the test statistic (TS) to choose the best combinations of λ's as the final results of our method and the JGL method. A smaller TS means the estimated precision matrix is closer to the true one in targeting areas. We also use this TS to compare the performance of our method with JGL.

Four precision matrices we generated as the true spatial precision matrices are:

$$
\Sigma^{1,1} = \begin{cases} \phi_1^{|i-j|} & \text{if } i,j \in [1,3] \\ \phi_2^{|i-j|} & \text{if } i,j \in [7,10] \\ 1 & \text{on diagonal} \\ 0 & \text{otherwise} \end{cases} \qquad \Sigma^{1,2} = \begin{cases} \phi_1^{|i-j|} & \text{if } i,j \in [1,3] \\ 1 & \text{on diagonal} \\ 0 & \text{otherwise} \end{cases}
$$

$$
\Sigma^{2,1} = \begin{cases} \phi_2^{|i-j|} & \text{if } i,j \in [7,10] \\ 1 & \text{on diagonal} \\ 0 & \text{otherwise} \end{cases} \qquad \Sigma^{2,2} = \begin{cases} 1 & \text{on diagonal} \\ 0 & \text{otherwise} \end{cases} \qquad (3.13)
$$

We vary ϕ_1 and ϕ_2 from 0.1 to 0.5 to test the performance of our method and JGL when two factors have a different level of impact on the covariance. The structure of the precision matrices generated when $\phi_1 = 0.1$ and $\phi_2 = 0.5$ and their associated spatial covariance matrix is shown in Figure 2.1.

We report the mean and standard error of the best (smallest) test statistic for each combination of ϕ's over 10 replications as long as associating λ's. The results are shown in Table 3.1.

These results show that our method provides better estimations of the true precision matrices in terms of our test statistics than JGL when two factors have a very different level of impact on the covariance structure. And when two factors have similar impacts on the covariance structure, our method performs as good as JGL.

3.5   Case Study

In this section, we evaluate the method on the cocaine-use disorder data. This data is a multi-subject resting-state functional magnetic resonance imaging (fMRI) dataset collected by the Institute for Drug and Alcohol Studies at Virginia Common-wealth University. This data includes a total of 45 subjects, in which 22 subjects were identified as having current cocaine use disorder and 23 non-drug-using subjects were recruited as the control group. Researchers also collected their gender (male or

33
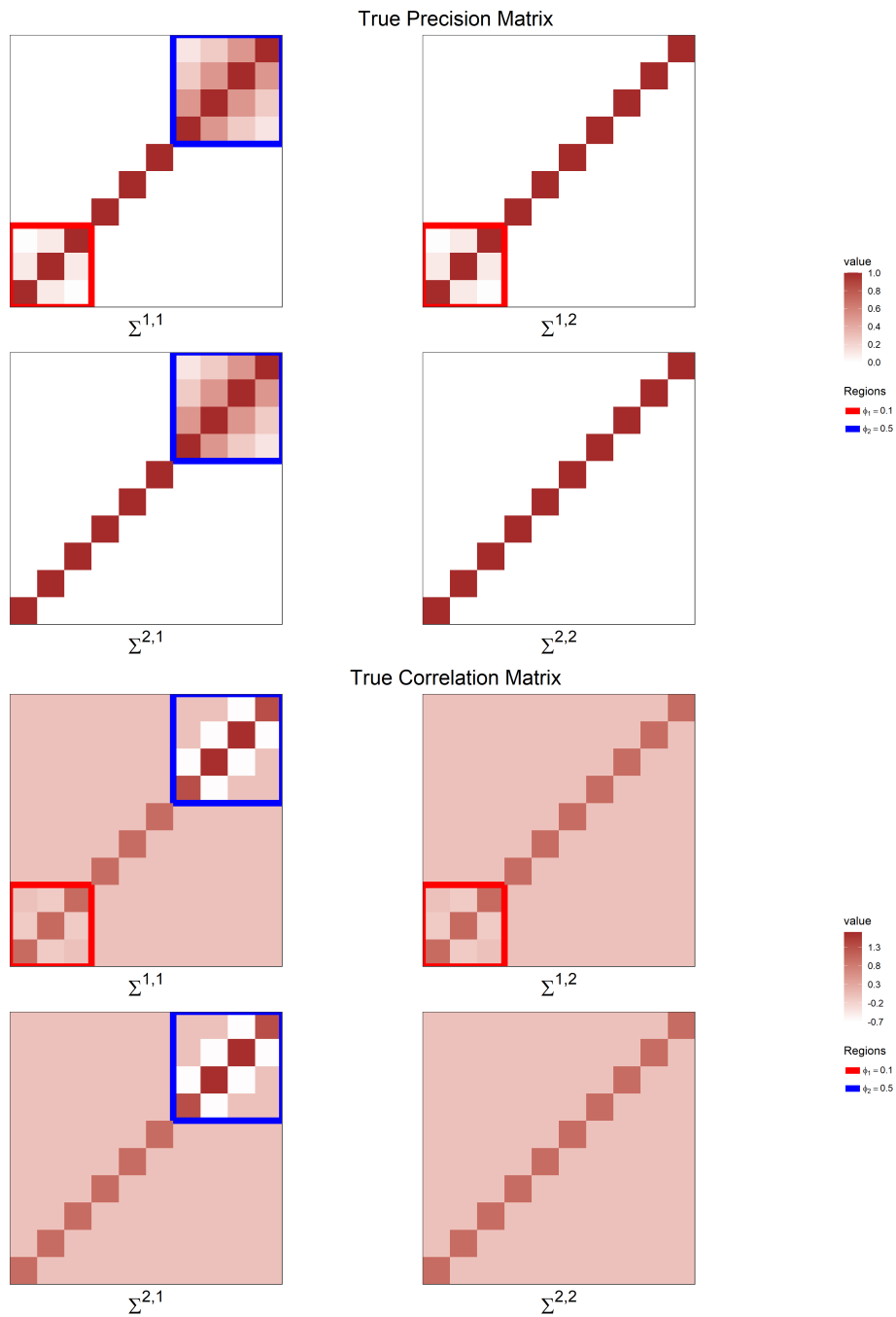
Fig. 2.1. Structure of the spatial precision matrices and covariance matrices generated when $\phi_1 = 0.1$ and $\phi_2 = 0.5$.

Table 3.1. Mean and standard error of mean (in parenthesis) of the best (smallest) TS under each scenario and associating λ combinations over 10 simulation replications.

Φ_1	Φ_2	Method	λ_1	λ_2	λ_3	Test Statistic
0.1	0.5	Proposed	0.0026 (0.0026)	0.0331 (0.0092)	0.0036 (0.0034)	1.4573 (0.1886)
		JGL	0.0016 (0.0024)	0.0026 (0.0026)	NA	1.7799 (0.2436)
0.2	0.5	Proposed	0.0021 (0.0026)	0.0326 (0.0072)	0.0016 (0.0034)	1.5320 (0.1434)
		JGL	0.0026 (0.0026)	0.0006 (0.0016)	NA	1.8101 (0.1760)
0.3	0.5	Proposed	0.0031 (0.0035)	0.0196 (0.0072)	0.0016 (0.0024)	1.6575 (0.2116)
		JGL	0.0031 (0.0035)	0.0011 (0.0022)	NA	1.8553 (0.1990)
0.4	0.5	Proposed	0.0006 (0.0016)	0.0171 (0.0109)	0.0031 (0.0042)	1.6682 (0.3358)
		JGL	0.0016 (0.0034)	0.0011 (0.0021)	NA	1.8938 (0.2459)
0.5	0.5	Proposed	0.0006 (0.0016)	0.0106 (0.0060)	0.0076 (0.0035)	1.6313 (0.1284)
		JGL	0.0041 (0.0034)	0.0001 (0.0000)	NA	1.8956 (0.1917)
0.5	0.4	Proposed	0.0016 (0.0024)	0.0196 (0.0083)	0.0031 (0.0068)	1.6769 (0.2265)
		JGL	0.0036 (0.0032)	0.0006 (0.0016)	NA	1.8163 (0.2632)
0.5	0.3	Proposed	0.0021 (0.0026)	0.0151 (0.0058)	0.0136 (0.0047)	1.4133 (0.1370)
		JGL	0.0036 (0.0041)	0.0031 (0.0026)	NA	1.5641 (0.1254)
0.5	0.2	Proposed	0.0016 (0.0024)	0.0081 (0.0075)	0.0191 (0.0074)	1.6537 (0.2100)
		JGL	0.0031 (0.0048)	0.0031 (0.0026)	NA	1.8024 (0.2070)
0.5	0.1	Proposed	0.0026 (0.0026)	0.0141 (0.0081)	0.0221 (0.0079)	1.4892 (0.2300)
		JGL	0.0041 (0.0039)	0.0041 (0.0021)	NA	1.6520 (0.2513)

female) and age (in years). Each subject in this data has 375 volumes of brain fMRI scan image, and each volume is partitioned into a $91 \times 109 \times 91$ rectangular lattice (voxels) of size $2 \times 2 \times 2$mm. The data were preprocessed using the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012a).

We are interested in studying the potential effect of cocaine dependence on brain

connectivities as long as the genders. Following the previous studies in cocaine usage disorder, we focus on five well-studied functional networks (WSFNs) that are closely related to cocaine addiction. The five WSFNs studied is (1) the default mode network (DM); (2) the dorsal attention network (DA); (3) the executive control network (CE); (4) the salience network (SA), and (5) the sensory motor network (SM). Brain regions in these WSFNs play important roles in external task performance, voluntary orienting, cognitive functions, problem-solving, and decision making. Similar to Razi et al. (2017), 27 regions of interest were selected from the five WSFNs. Each region of interest was defined as a sphere with a radius of 12 mm, and the center of each sphere was determined based on previous studies (Razi et al., 2017). These WSFNs contain a total of 6154 voxels. See Table 3.2 for the details of these regions of interest.

To investigate the effect of cocaine dependence and gender on brain connectivities, we first divide subjects into four groups by their cocaine usage status and gender and label them from group 1 to group 4 (Group 1: cocaine male. Group 2: cocaine female. Group 3: control male. Group 4 control female). Group 1-4 contains 16, 6, 14, and 9 subjects, respectively. The differences of spatial correlation between Group 1/Group 3 and Group 2/Group 4 illustrate the effect of cocaine dependence on brain connectivity. Similarly, the differences of spatial correlation between Group 1/Group 2 and Group 3/Group 4 illustrate the effect of gender on brain connectivity.

Before model fitting, we first estimate the whole-brain voxel-level spatial correlation structure for each group from the cocaine use disorder data as the representation of functional connectivity map in brain using the method introduced in Zhao et al. (2021). Then we collect all voxel-to-voxel spatial correlations which belong to five WSFNs. We calculate mean voxel-to-voxel spatial correlations between each pair of brain regions and create four 27×27 spatial correlation matrices for each group. We then apply the proposed method to these four matrices and obtain group-level sparse

36

Table 3.2. The detailed information about the 27 spherical regions. DM = default mode, DA = dorsal attention, EC = executive control, SA = salience, SM = sensory motor network. IPS = inferior parietal sulcus, MT = middle temporal, PFC = prefrontal cortex. L = left, R = right.

Node #	Abbreviation	Name	Center Coordinates [x y z]
1	DM_PCP	Posterior cingulate/Precuneus	0, 52, 12
2	DM_mPFC	Medial Prefrontal	2, 60, 16
3	DM_L_LP	Left lateral parietal	38, 74, 30
4	DM_R_LP	Right lateral parietal	46, 64, 26
5	DA_L_FEF	Left frontal eye field	32, 6, 60
6	DA_R_FEF	Right frontal eye field	32, 6, 58
7	DA_L_pIPS	Left posterior IPS	22, 54, 52
8	DA_R_pIPS	Right posterior IPS	24, 56, 56
9	DA_L_aIPS	Left anterior IPS	36, 40, 58
10	DA_R_aIPS	Right anterior IPS	36, 38, 58
11	DA_L_MT	Left MT	48, 68, 0
12	DA_R_MT	Right MT	50, 60, 12
13	EC_DMPFC	Dorsal medial PFC	2, 22, 48
14	EC_L_aPFC	Left anterior PFC	30, 50, 4
15	EC_R_aPFC	Right anterior PFC	38, 46, 4
16	EC_L_SP	Left superior parietal	60, 36, 34
17	EC_R_SP	Right superior parietal	60, 30, 34
18	SA_dACC	Dorsal anterior cingulate	4, 38, 36
19	SA_L_aPFC	Left anterior PFC	46, 20, 26
20	SA_R_aPFC	Right anterior PFC	42, 22, 38
21	SA_L_INS	Left insula	32, 20, 8
22	SA_R_INS	Right insula	36, 20, 8
23	SA_L_LP	Left lateral parietal	42, 66, 36
24	SA_R_LP	Right lateral parietal	52, 50, 40
25	SM_L_MC	Left motor cortex	40, 30, 56
26	SM_R_MC	Right motor cortex	42, 28, 46
27	SM_SMA	Supplementary motor area	2, 0, 52

precision matrices for different factors at different levels.

A major challenge in the study of the graphical lasso method is the choice of tuning parameters. In the literature, parameter tuning is often done by splitting the data into training and testing sets and using the model misspecification rate as the criterion to guide parameter selection. Other works choose tuning parameters on experience or conventions. Given the lack of prior knowledge on tuning these parameters and the

37

small sample size, we use a p-value based method to choose all three tuning parameters (one for the L_1 penalty and two for fused penalties) in this study. WSFNs are sets of brain regions that perform cognitive tasks by networks (Smith et al., 2011). They are known to have relatively stronger within-network spatial dependency than out-of-network spatial dependency. So, we assume all five WSFNs we studied have higher within-network spatial correlations. We can evaluate the significance of these WSFNs with respect to the estimated sparse precision matrices and find the tuning parameter combination. First, we estimate sparse precision matrices for four groups under numerous combinations of tuning parameters ($\{\lambda_1, \lambda_2, \lambda_3\} = \{0 <= \lambda_i <= 1, i = 1, 2, 3\}$ for our proposed method, and $\{\lambda_1, \lambda_2\} = \{0 <= \lambda_i <= 1, i = 1, 2\}$ for the JGL method) using the proposed method and the JGL method. For each combination of lambdas, we aggregate all estimated sparse precision matrix values from all four groups as the dependent variable and define the regions within or out of each network as a binary indicator. We fit five linear regression models, each for every network, and record the p-value for five WSFNs. Then, we choose the combination of lambdas which has all five WSFNs significant under the smallest significance level as the best choice of tuning parameters.

For our proposed method, the tuning parameters we choose are $\lambda_1 = 0.1$, $\lambda_2 = 0$ and $\lambda_3 = 0.2$. The estimated sparse precision matrices under the chosen combination of tuning parameters are visualized in Figure 2.2. For the JGL method, the tuning parameters we choose are $\lambda_1 = 0.1$ and $\lambda_2 = 0.1$. The estimated sparse precision matrices using the JGL method under the chosen combination of tuning parameters are visualized in Figure 2.3.

To find the effect of cocaine dependence and gender on brain connectivities, we further take the differences of estimated sparse precision matrices between Group 1/Group 3, Group 2/Group 4, Group 1/Group 2, and Group 3/Group 4 and mark the

38
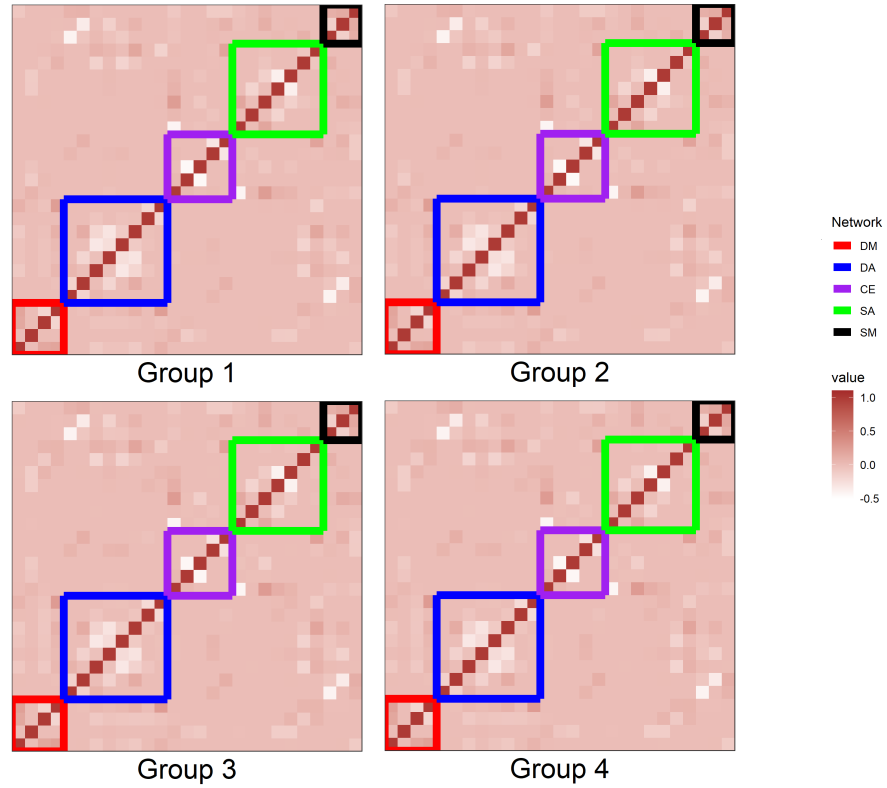
Fig. 2.2. Sparse precision matrices for four groups estimated by our proposed method when $\lambda_1 = 0.1$, $\lambda_2 = 0$ and $\lambda_3 = 0.2$. The square areas along the diagonal are within network correlation structure for WSFNs. DM = default mode, DA = dorsal attention, EC = executive control, SA = salience, SM = sensory motor network.
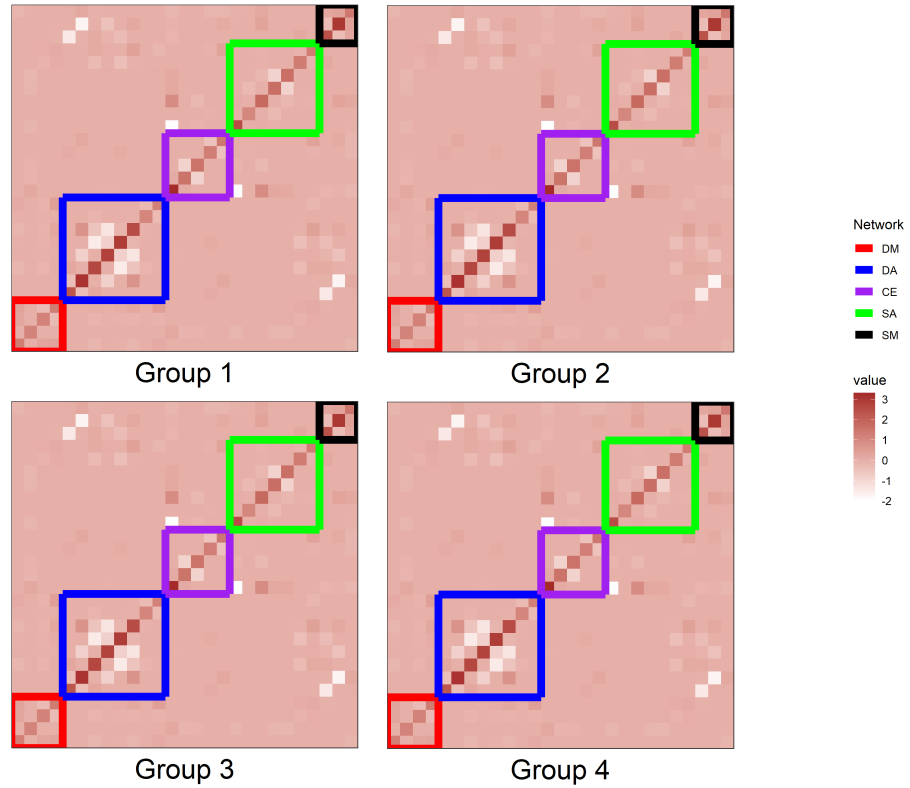
Fig. 2.3. Sparse precision matrices for four groups estimated by the JGL method when $\lambda_1 = 0.1$ and $\lambda_2 = 0.1$. The square areas along the diagonal are within network correlation structure for WSFNs. DM = default mode, DA = dorsal attention, EC = executive control, SA = salience, SM = sensory motor network.

region pairs with higher between-group difference values than a prespecified threshold as regions strongly affected by a factor. We visualize these regions provided by our proposed method in Figure 2.4. By specifying $\lambda_1 = 0.1$, $\lambda_2 = 0$ and $\lambda_3 = 0.2$, the identified region pairs are provided in Table 3.3 and Table 3.4.

The JGL method fail to identify the effect of both factors. The precision matrices estimated using the JGL method are identical across groups.

Table 3.3. Effect of cocaine use disorder on functional connectivity in terms of difference of estimated spatial correlation between group 1/group 3 and group 2/group 4 in regions detected by our proposed model. Illustrating the effect of cocaine use disorder on functional connectivity in brain. IPS = inferior parietal sulcus, MT = middle temporal, PFC = prefrontal cortex.

Region Pair	Difference of Estimated Spatial Correlation	
	Group 1 - Group 3	Group 2 - Group 4
Posterior Cingulate/Precuneus vs. Right lateral parietal	4.472e-02	4.471e-02
Right Anterior IPS vs. Left MT	-8.501e-02	-8.511e-02
Left MT vs. Right MT	-5.116e-02	-5.119e-02
Left Superior Parietal vs. Right Superior Parietal	4.660e-02	4.658e-02
Right Anterior PFC vs. Right Insula	2.266e-02	2.248e-02
Right Anterior PFC vs. Left Lateral Parietal	8.846e-02	8.859e-02
Right Insula vs. Left Lateral Parietal	-6.308e-02	-6.324e-02
Left Lateral Parietal vs. Right Lateral Parietal	9.020e-02	9.028e-02

Following the definition of functional connectivity in brain in Chapter 2, we need to take a close look at the spatial correlation in regions affected by cocaine use disorder (CUD) and gender in order to study the effect of these two factors on the functional connectivity. We first take the spatial correlations associated with regions identified by our proposed method for all four groups, and then take difference between groups.

Table 3.3 shows the difference of Estimated Spatial Correlation between group 1/group 2 and group 3/group 4 in regions detected by our proposed method. We

Table 3.4. Effect of gender on functional connectivity in terms of difference of estimated spatial correlation between group 1/group 2 and group 3/group 4 in regions detected by our proposed model. Illustrating the effect of gender on functional connectivity in brain. IPS = inferior parietal sulcus.

Region Pair	Difference of Estimated Spatial Correlation	
	Group 1 - Group 2	Group 3 - Group 4
Left Posterior IPS vs. Left Anterior IPS	-2.052e-04	-5.685e-04
Right Posterior IPS vs. Left Anterior IPS	-1.091e-04	-7.529e-05
Left Anterior IPS vs. Right Anterior IPS	-2.041e-04	-5.287e-04

can see that the participants with CUD had larger functional connectivity in terms of spatial correlation than the control participants between posterior cingulate and right lateral parietal (4.472e-02 for male and 4.471e-02 for female), between left superior parietal and right superior parietal (4.660e-02 for male and 4.658e-02 for female), between right anterior PFC and right insula (2.266e-02 for male and 2.248e-02 for female), between right anterior PFC and left lateral parietal (8.846e-02 for male and 8.859e-02 for female), and between left lateral parietal and right lateral parietal (9.020e-02 for male and 9.028e-02 for female). The CUD participants also had smaller functional connectivity than the control participants between right anterior IPS and left MT (-8.501e-02 for male and -8.511e-02 for female), between left MT and right MT (-5.116e-02 for male and -5.119e-02 for female) and between right insula and left lateral parietal (-6.308e-02 for male and -6.324e-02 for female). This result is also consistent with previous clinical studies on cocaine use in Chapter 2, validating the clinical reliability of our proposed method.

In Table 3.4, we see that the male participants had smaller functional connectivity in terms of spatial correlation than female participants between left posterior IPS and left anterior IPS (-2.052e-04 for cocaine group and -5.685e-04 for control group),

between right posterior IPS and left anterior IPS (-1.091e-04 for cocaine group and -7.529e-05 for control group), and between left anterior IPS and right anterior IPS (-2.041e-04 for cocaine group and -5.287e-04 for control group).

The functional connectivity differences between cocaine use disorder subjects and control subjects identified by our method in the precuneus and the lateral parietal area are consistent with previous clinical studies on cocaine use in Chapter 2, validating the clinical reliability of our proposed method.
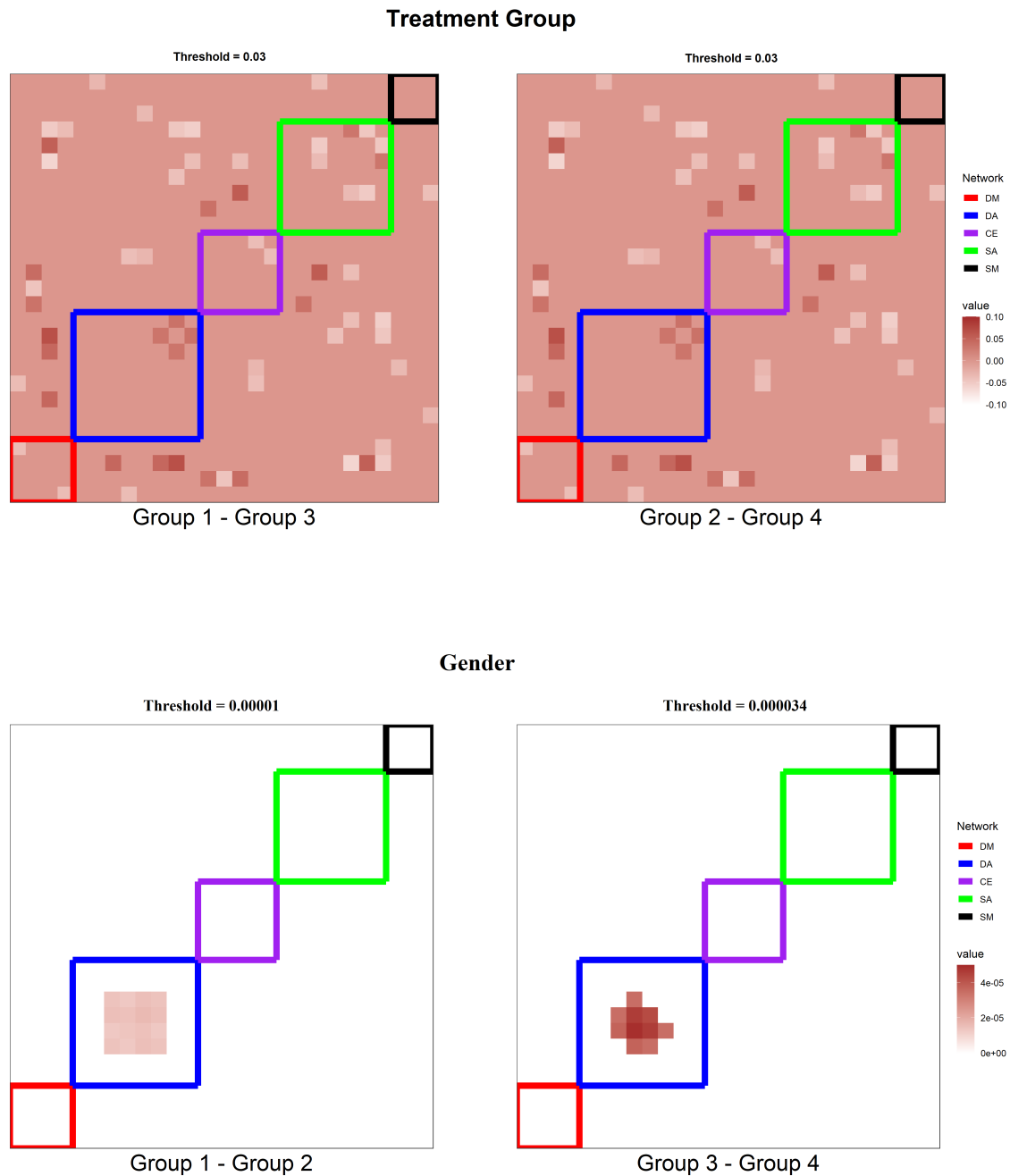
Fig. 2.4. Difference of estimated sparse precision matrices between groups to show the effect of cocaine use disorder and gender when $\lambda_1 = 0.1$, $\lambda_2 = 0$ and $\lambda_3 = 0.2$ under prespecified threshold. The cubic areas along the diagonal are within network correlation structure for WSFNs. DM = default mode, DA = dorsal attention, EC = executive control, SA = salience, SM = sensory motor network.

CHAPTER 4

CONCLUSION

In Chapter 2, we proposed a Spatio-temporal model for multi-subject fMRI data that efficiently incorporates voxel-level Spatio-temporal dependencies of whole-brain measurements to improve the accuracy of statistical inferences. This model addresses common challenges in fMRI brain imaging data, including large-scale nonstationary spatial covariance and complex Spatio-temporal dependence. From the simulated study, we show that properly accounting for spatial and temporal correlation is critical for efficient estimation and valid inference from the simulated study. Using our efficient estimation method, we successfully reduce the computational effort to analyze multi-subject fMRI data from days to minutes. We also developed a permutation test to test the brain global spatial dependence difference between two groups and a graphical-lasso based method to identify the local difference. We applied our method to the cocaine use disorder data and identified multiple brain region pairs with significant connectivity difference between the cocaine and control groups. Our results are consistent with literatures that cocaine use disorder may be associated with the functional connectivity in the medial orbitofrontal cortex subregions and the supplementary motor areas. Particularly, the cocaine use disorder individuals showed lower functional connectivity between the bilateral supplementary motor areas than controls. Given the sample size and relatively small group difference, the novel results need to be replicated in future studies with a different sample. Specifically, it is critical to improve the design and/or modeling with respect to the demographic covariates, such as increasing the number of subjects in each demographic group or

including confound modeling (Alfaro-Almagro et al., 2021).

In Chapter 3, we propose a lasso-based hierarchical penalty function for estimating sparse precision matrices for large-scale multi-subject fMRI data with multiple categorical factors. To improve the estimation accuracy and reduce computational effort, we assign different penalties for each of the multiple distinct level factors. We also propose a significance-level-based approach to choose the best combination of tuning parameters. We further apply with applications to the cocaine-use disorder fMRI data. The simulation studies show that our method is more accurate in estimating sparse precision matrices than the JGL method. The real data application demonstrates the effectiveness of our method in studying the effect of multi-level factors in fMRI data. We identify regions with different spatial correlations between the cocaine use disorder group and the control group. Our results are consistent with the literature that the cocaine use disorder individuals show lower functional connectivity between the bilateral supplementary motor areas than controls. The main limitation of this study is the computational time. The average computational time for the JGL method is 1.276 seconds on a desktop computer with a quad-core 4.00GHz processor and 16GB RAM space. The average computational time for our proposed method is 150 seconds on the same system. This result does not meet our expectation in section 3.3 which is based on the analysis of the algorithm cost. This difference is because (1) we use a numerical optimizer to solve for the $\mathbf{Z}$ matrix instead of the generalized fused lasso approximator used by the JGL R package; and (2) we conduct simulation studies on a simple case that includes two factors, each with two levels. In the future study, we will adopt a more efficient optimization procedure to improve the computational time of our method and then conduct simulation studies on a larger size of factors to further evaluate the performance of the proposed method.

<center>**Appendix A**</center>

<center>**APPENDICES**</center>

A.1 Transformation from model (2.1) to (2.4)

Let $Y_{i,t}$ be the vector that collects all the responses of subject i at volume t and let $Y_i = [Y_{i,1}^\top, Y_{i,2}^\top, \ldots, Y_{i,L}^\top]^\top$ stack all the responses of subject i. Following the notation in Section 2.3 and Section 2.4, we can rewrite our model in matrix form as:

$$Y_i = X_i\boldsymbol{\beta} + (\mathcal{W} \otimes B)\Gamma_i + E_i, \tag{A.1}$$

where $\otimes$ shows the Kronecker product, $X_i = \underline{1}_{ML} \otimes x_i^\top$, $x_i^\top = [1, \mathbf{x}_i^\top]$, $\underline{1}_{ML}$ is the vector of all 1's of size $ML \times 1$, $\boldsymbol{\beta} = [\mu(v,t), \beta]$, E_i stacks all the random errors. The variance-covariance matrix of Y_i can be expressed by

$$\text{var}(Y_i) = \Sigma_T \otimes B\Sigma B^\top + \sigma^2 I_{ML}, \tag{A.2}$$

where I_{ML} is the identity matrix of size $ML \times ML$.

Let $\tilde{Y}_i = (U^\top \otimes I_M)Y_i$ be the transformed response vector of subject i at volume t, and we will have $\tilde{Y}_{i,t} = (U_t^\top \otimes I_M)Y_i$, where $U_t^\top = [U_{t,1}, U_{t,2}, \ldots, U_{t,L}]$ is the t-th row of matrix $U^\top$. Then, the variance-covariance matrix of $\tilde{Y}_i$ is

$$\begin{aligned}
\text{var}(\tilde{Y}_i) &= \text{var}\left((U^\top \otimes I_M)Y_i\right) \\
&= (U^\top \otimes I_M)^\top(\Sigma_T \otimes B\Sigma B^\top + \sigma^2 I_{ML})(U \otimes I_M) \\
&= U^\top \Sigma_T U \otimes B\Sigma B^\top + \sigma^2(U^\top U \otimes I_M) \\
&= V \otimes B\Sigma B^\top + \sigma^2 I_{ML}, \tag{A.3}
\end{aligned}$$

<center>47</center>

where V is an $L \times L$ diagonal matrix with element λ_t, $t = 1, \ldots, L$. We can see that $\tilde{Y}_{i,t}$'s are independent after this transformation, and the variance-covariance matrix of $\tilde{Y}_{i,t}$ is

$$\text{var}(\tilde{Y}_{i,t}) = \lambda_t B \Sigma B^\top + \sigma^2 I_M. \tag{A.4}$$

Now, the transformed model can be expressed by

$$
\begin{aligned}
\tilde{Y}_{i,t} &= (U_t^\top \otimes I_M) Y_i \\
&= (U_t^\top \otimes I_M) \left(X_i \boldsymbol{\beta} + (\mathcal{W} \otimes B) \Gamma_i + E_i \right) \\
&= (U_t^\top \otimes I_M) \left(\underline{1}_{ML} \otimes x_i^\top \boldsymbol{\beta} + (\mathcal{W} \otimes B) \Gamma_i + E_i \right) \\
&= \left(\underline{1}_M \otimes (\sum_{l=1}^L U_{t,l}^\top) x_i^\top \right) \boldsymbol{\beta} + (U_t^\top \mathcal{W} \otimes B) \Gamma_i + (U_t^\top \otimes I_M) E_{i,t}.
\end{aligned}
\tag{A.5}
$$

If we choose $\mathcal{W}$ to be the solution of $U^\top \mathcal{W} = \sqrt{V}$, we will have $U_t^\top \mathcal{W}$ to be a row vector whose tth element is $\sqrt{\lambda_t}$ and all other elements are 0's. Then the transformed model can be further simplified as

$$\tilde{Y}_{i,t} = \tilde{X}_{i,t} \boldsymbol{\beta} + \sqrt{\lambda_t} B \Gamma_{i,t} + \tilde{E}_{i,t} \tag{4}$$

where $\tilde{X}_{i,t} = \underline{1}_M \otimes (\sum_{l=1}^L U_{t,l}^\top) x_i^\top$, and $\tilde{E}_{i,t} = (U_t^\top \otimes I_M) E_{i,t}$.

A.2 Prior Settings and The Close-form Update Formulas

In this section, we derive the updating formula of each parameters using the EM algorithm. Following the notation setting in Section 2.3, Section 2.4 and A.1, our model is

$$\tilde{Y}_{i,t} = \tilde{X}_{i,t}^\top \boldsymbol{\beta} + \sqrt{\lambda_t} B \Gamma_{i,t} + \tilde{E}_{i,t}. \tag{4}$$

48

The prior settings we use to develop the update formulas for the EM algorithm are

$$\Gamma_{i,t}|\Sigma \sim \mathcal{N}(0,\Sigma), \text{ for } i = 1, \ldots, N \text{ and } t = 1, \ldots, L$$

$$\Sigma|\Psi, \nu \sim \mathcal{IW}(\Psi, \nu)$$

$$\sigma^2|a, b \sim \mathcal{IG}(a, b)$$

$$\boldsymbol{\beta}|\mu_\beta, R_\beta \sim \mathcal{N}(\mu_\beta, \sigma^2 R_\beta) \tag{A.6}$$

By inserting the prior distributions for the model parameters, the posterior distributions of the unknown parameters can be expressed by

$$p(\boldsymbol{\beta}, \Sigma, \sigma^2|\tilde{\mathbf{Y}}) \propto p(\tilde{\mathbf{Y}}|\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_N, \boldsymbol{\beta}, \Sigma, \sigma^2) p(\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_N|\boldsymbol{\beta}, \Sigma, \sigma^2) p(\Sigma, \sigma^2) p(\boldsymbol{\beta})$$

$$\propto \prod_{i=1}^{N} \left\{ \prod_{t=1}^{L} \left[p(\tilde{Y}_{i,t}|\Gamma_{i,t}, \boldsymbol{\beta}, \sigma^2) \right] p(\Gamma_{i,t}|\Sigma) \right\} p(\Sigma) p(\sigma^2) p(\boldsymbol{\beta}) \tag{A.7}$$

Taking $\tilde{Y}_{i,t}$ as observed data and $\Gamma_{i,t}$'s as missing data. At iteration $l + 1$, EM algorithm expectation step is to find

$$
\begin{aligned}
Q(\boldsymbol{\beta}, \Sigma, \sigma | \boldsymbol{\beta}^l, \Sigma^l, \sigma^l) =& \mathrm{E}_\Gamma \left[-2 \log p(\boldsymbol{\beta}, \Sigma, \sigma | Y) | \boldsymbol{\beta}^l, \Sigma^l, \sigma^l \right] \\
=& -2\mathrm{E}_\Gamma \Bigg[\sum_{i=1}^N \sum_{t=1}^L \log p(\tilde{Y}_{i,t} | \Gamma_{i,t}, \boldsymbol{\beta}, \sigma^2) + \sum_{i=1}^N \sum_{t=1}^L \log p(\Gamma_{i,t} | \Sigma) \\
& + \log p(\Sigma) + \log p(\sigma^2) + \log p(\boldsymbol{\beta}) \Big| \boldsymbol{\beta}^l, \Sigma^l, \sigma^l \Bigg] \\
=& -2\mathrm{E}_\Gamma \Bigg[\sum_{i=1}^N \sum_{t=1}^L \log \Bigg\{ (2\pi)^{-\frac{M}{2}} |\sigma^2 I_M|^{-\frac{1}{2}} \exp \Bigg\{ -\frac{1}{2} (\tilde{Y}_{i,t} - \tilde{X}_{i,t}\boldsymbol{\beta} - \sqrt{\lambda_t} B \Gamma_{i,t})^\top \\
& \qquad\qquad\qquad\qquad (\sigma^2 I_M)^{-1} \left(\tilde{Y}_{i,t} - \tilde{X}_{i,t}\boldsymbol{\beta} - \sqrt{\lambda_t} B \Gamma_{i,t} \right) \Bigg\} \Bigg\} \\
& + \sum_{i=1}^N \sum_{t=1}^L \log \Bigg\{ (2\pi)^{-\frac{J}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \Gamma_{i,t}^\top \Sigma \Gamma_{i,t} \right\} \Bigg\} \\
& + \log p(\Sigma) + \log p(\sigma^2) + \log p(\boldsymbol{\beta}) \Big| \boldsymbol{\beta}^l, \Sigma^l, \sigma^l \Bigg] \\
=& (NML + NLJ) \log 2\pi + NML \log \sigma^2 + NL \log |\Sigma| \\
& + \frac{1}{\sigma^2} \sum_{i=1}^N \sum_{t=1}^L \tilde{Y}_{i,t}^\top \tilde{Y}_{i,t} - \frac{2}{\sigma^2} \sum_{i=1}^N \sum_{t=1}^L \tilde{Y}_{i,t}^\top \tilde{X}_{i,t} \boldsymbol{\beta} + \frac{1}{\sigma^2} \boldsymbol{\beta}^\top \left(\sum_{i=1}^N \sum_{t=1}^L \tilde{X}_{i,t}^\top \tilde{X}_{i,t} \right) \boldsymbol{\beta} \\
& - \frac{2}{\sigma^2} \sum_{i=1}^N \sum_{t=1}^L \sqrt{\lambda_t} \tilde{Y}_{i,t}^\top B \, \mathrm{E} \left[\Gamma_{i,t} | \tilde{Y}_{i,t}, \boldsymbol{\beta}^l, \Sigma^l, \sigma^l \right] \\
& + \frac{2}{\sigma^2} \sum_{i=1}^N \sum_{t=1}^L \sqrt{\lambda_t} \boldsymbol{\beta}^\top \tilde{X}_{i,t}^\top B \, \mathrm{E} \left[\Gamma_{i,t} | \tilde{Y}_{i,t}, \boldsymbol{\beta}^l, \Sigma^l, \sigma^l \right] \\
& + \sum_{i=1}^N \sum_{t=1}^L \mathrm{tr} \left\{ \left(\frac{\lambda_t B^\top B}{\sigma^2} + \Sigma^{-1} \right) \mathrm{E} \left[\Gamma_{i,t} \Gamma_{i,t}^\top | \tilde{Y}_{i,t}, \boldsymbol{\beta}^l, \Sigma^l, \sigma^l \right] \right\} \\
& - 2 \log p(\Sigma) - 2 \log p(\sigma^2) - 2 \log p(\boldsymbol{\beta}).
\end{aligned}
\tag{A.8}
$$

We need to find the distribution of $\Gamma_{i,t} | \tilde{Y}_{i,t}, \boldsymbol{\beta}^l, \Sigma^l, \sigma^l$. From the Bayesian theorem, we have

$$
p(\Gamma_{i,t} | \tilde{Y}_{i,t}, \boldsymbol{\beta}^l, \Sigma^l, \sigma^l) = \frac{p(\tilde{Y}_{i,t} | \Gamma_{i,t}, \boldsymbol{\beta}^l, \sigma^l) p(\Gamma_{i,t} | \Sigma^l)}{\int_{\Gamma_{i,t}} p(\tilde{Y}_{i,t} | \Gamma_{i,t}, \boldsymbol{\beta}^l, \sigma^l) p(\Gamma_{i,t} | \Sigma^l)}
\tag{A.9}
$$

and

$$
\begin{aligned}
p(\tilde{Y}_{i,t}|\Gamma_{i,t},\boldsymbol{\beta}^l,\sigma^l)p(\Gamma_{i,t}|\Sigma^l) =&(2\pi)^{-\frac{M}{2}}|(\sigma^2)^l I_M|^{-\frac{1}{2}}\exp\Big\{-\frac{1}{2}\big(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l-\sqrt{\lambda_t}B\Gamma_{i,t}\big)^{\top}\\
&((\sigma^2)^l I_M)^{-1}\left(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l-\sqrt{\lambda_t}B\Gamma_{i,t}\right)\Big\}\\
&(2\pi)^{-\frac{J}{2}}|\Sigma^l|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\Gamma_{i,t}^{\top}(\Sigma^l)^{-1}\Gamma_{i,t}\right\}\\
=&(2\pi)^{-\frac{M+J}{2}}|(\sigma^2)^l I_M|^{-\frac{1}{2}}|\Sigma^l|^{-\frac{1}{2}}\\
&\exp\Big\{-\frac{1}{2(\sigma^2)^l}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)^{\top}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)\\
&+\frac{1}{(\sigma^2)^l}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)^{\top}\sqrt{\lambda_t}B\Gamma_{i,t}\\
&-\frac{1}{2}\Gamma_{i,t}^{\top}\left(\frac{\lambda_t B^{\top}B}{(\sigma^2)^l}+(\Sigma^l)^{-1}\right)\Gamma_{i,t}\Big\}.
\end{aligned}
\tag{A.10}
$$

Let

$$
\Sigma_t^l=\left[\frac{\lambda_t B^{\top}B}{(\sigma^2)^l}+(\Sigma^l)^{-1}\right]^{-1}, \qquad\qquad t=1,\ldots,L \tag{A.11}
$$

$$
\mu_{i,t}^l=\left[\frac{\lambda_t B^{\top}B}{(\sigma^2)^l}+(\Sigma^l)^{-1}\right]^{-1}\frac{\sqrt{\lambda_t}B^{\top}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)}{(\sigma^2)^l}, \qquad t=1,\ldots,L,\quad i=1,\ldots,N. \tag{A.12}
$$

We have

$$
\begin{aligned}
p(\tilde{Y}_{i,t}|\Gamma_{i,t},&\boldsymbol{\beta}^l,\sigma^l)p(\Gamma_{i,t}|\Sigma^l)=\\
&(2\pi)^{-\frac{M+J}{2}}|(\sigma^2)^l I_M|^{-\frac{1}{2}}|\Sigma^l|^{-\frac{1}{2}}\exp\Big\{-\frac{1}{2(\sigma^2)^l}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)^{\top}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)\Big\}\\
&\exp\Big\{-\frac{1}{2}(\Gamma_{i,t}-\mu_{i,t}^l)^{\top}(\Sigma_t^l)^{-1}(\Gamma_{i,t}-\mu_{i,t}^l)\Big\}\exp\Big\{\frac{1}{2}(\mu_{i,t}^l)^{\top}\Sigma_t^{-1}\mu_{i,t}^l\Big\}.
\end{aligned}
\tag{A.13}
$$

Then

$$\int_{\Gamma_{i,t}} p(\tilde{Y}_{i,t}|\Gamma_{i,t},\boldsymbol{\beta}^l,\sigma^l)p(\Gamma_{i,t}|\Sigma^l) = \int_{\Gamma_{i,t}} (2\pi)^{-\frac{M+J}{2}}|(\sigma^2)^l I_M|^{-\frac{1}{2}}|\Sigma^l|^{-\frac{1}{2}}$$

$$\exp\left\{-\frac{1}{2(\sigma^2)^l}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)^\top(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)\right\}$$

$$\exp\left\{-\frac{1}{2}(\Gamma_{i,t}-\mu_{i,t}^l)^\top(\Sigma_t^l)^{-1}(\Gamma_{i,t}-\mu_{i,t}^l)\right\}\exp\left\{\frac{1}{2}(\mu_{i,t}^l)^\top(\Sigma_t^l)^{-1}\mu_{i,t}^l\right\}$$

$$=(2\pi)^{-\frac{M}{2}}|(\sigma^2)^l I_M|^{-\frac{1}{2}}|\Sigma^l|^{-\frac{1}{2}}|\Sigma_t^l|^{\frac{1}{2}}$$

$$\exp\left\{-\frac{1}{2(\sigma^2)^l}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)^\top(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)\right\}$$

$$\exp\left\{\frac{1}{2}(\mu_{i,t}^l)^\top(\Sigma_t^l)^{-1}\mu_{i,t}^l\right\}$$

$$\int_{\Gamma_{i,t}}(2\pi)^{-\frac{J}{2}}|\Sigma_t^l|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\Gamma_{i,t}-\mu_{i,t}^l)^\top(\Sigma_t^l)^{-1}(\Gamma_{i,t}-\mu_{i,t}^l)\right\}$$

$$=(2\pi)^{-\frac{M}{2}}|(\sigma^2)^l I_M|^{-\frac{1}{2}}|\Sigma^l|^{-\frac{1}{2}}|\Sigma_t^l|^{\frac{1}{2}}$$

$$\exp\left\{-\frac{1}{2(\sigma^2)^l}(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)^\top(\tilde{Y}_{i,t}-\tilde{X}_{i,t}\boldsymbol{\beta}^l)\right\}$$

$$\exp\left\{\frac{1}{2}(\mu_{i,t}^l)^\top(\Sigma_t^l)^{-1}\mu_{i,t}^l\right\} \tag{A.14}$$

Then we have

$$p(\Gamma_{i,t}|\tilde{Y}_{i,t},\boldsymbol{\beta}^l,\Sigma^l,\sigma^l) = \frac{p(\tilde{Y}_{i,t}|\Gamma_{i,t},\boldsymbol{\beta}^l,\sigma^l)p(\Gamma_{i,t}|\Sigma^l)}{\int_{\Gamma_{i,t}}p(\tilde{Y}_{i,t}|\Gamma_{i,t},\boldsymbol{\beta}^l,\sigma^l)p(\Gamma_{i,t}|\Sigma^l)}$$

$$=(2\pi)^{-\frac{J}{2}}|\Sigma_t^l|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\Gamma_{i,t}-\mu_{i,t}^l)^\top(\Sigma_t^l)^{-1}(\Gamma_{i,t}-\mu_{i,t})\right\} \tag{A.15}$$

Now, we know that $\Gamma_{i,t}|\tilde{Y}_{i,t},\boldsymbol{\beta}^l,\Sigma^l,\sigma^l \sim N(\mu_{i,t}^l,\Sigma_t^l)$. Then, $\mathrm{E}\left[\Gamma_{i,t}|\tilde{Y}_{i,t},\boldsymbol{\beta}^l,\Sigma^l,\sigma^l\right] = \mu_{i,t}^l$ and $\mathrm{E}\left[\Gamma_{i,t}\Gamma_{i,t}^\top|\tilde{Y}_{i,t},\boldsymbol{\beta}^l,\Sigma^l,\sigma^l\right] = \Sigma_t^l + \Omega_{i,t}^l$, where $\Omega_{i,t}^l = \mu_{i,t}^l(\mu_{i,t}^l)^\top$.

Finally, the EM algorithm expectation step objective function is

$$Q(\boldsymbol{\beta},\Sigma,\sigma|\boldsymbol{\beta}^l,\Sigma^l,\sigma^l) = (NML+NLJ)\log 2\pi + NML\log\sigma^2 + NL\log|\Sigma|$$

$$+\frac{1}{\sigma^2}\sum_{i=1}^N\sum_{t=1}^L\tilde{Y}_{i,t}^\top\tilde{Y}_{i,t} - \frac{2}{\sigma^2}\sum_{i=1}^N\sum_{t=1}^L\tilde{Y}_{i,t}^\top\tilde{X}_{i,t}\boldsymbol{\beta} + \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\left(\sum_{i=1}^N\sum_{t=1}^L\tilde{X}_{i,t}^\top\tilde{X}_{i,t}\right)\boldsymbol{\beta}$$

$$-\frac{2}{\sigma^2}\sum_{i=1}^N\sum_{t=1}^L\sqrt{\lambda_t}\tilde{Y}_{i,t}^\top B\mu_{i,t}^l + \frac{2}{\sigma^2}\sum_{i=1}^N\sum_{t=1}^L\sqrt{\lambda_t}\boldsymbol{\beta}^\top\tilde{X}_{i,t}^\top B\mu_{i,t}^l$$

$$+\sum_{i=1}^N\sum_{t=1}^L\mathrm{tr}\left\{\left(\frac{\lambda_t B^\top B}{\sigma^2}+\Sigma^{-1}\right)(\Sigma_t^l+\Omega_{i,t}^l)\right\}$$

$$-2\log p(\Sigma) - 2\log p(\sigma^2) - 2\log p(\boldsymbol{\beta}). \tag{A.16}$$

In the M-step of the EM algorithm, we need to find $\boldsymbol{\beta}^{l+1}$, Σ^{l+1}, σ^{l+1} to minimize $Q(\boldsymbol{\beta}, \Sigma, \sigma | \boldsymbol{\beta}^l, \Sigma^l, \sigma^l)$. We first take derivative of $Q(\boldsymbol{\beta}, \Sigma, \sigma | \boldsymbol{\beta}^l, \Sigma^l, \sigma^l)$ with respect to each parameters

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\beta}, \Sigma, \sigma | \boldsymbol{\beta}^l, \Sigma^l, \sigma^l)}{\partial \beta} = & 2\left((\sigma^2)^l R_\beta\right)^{-1} \boldsymbol{\beta} - 2\left((\sigma^2)^l R_\beta\right)^{-1} \mu_\beta - \frac{2}{(\sigma^2)^l}\left(\sum_{i=1}^{N}\sum_{t=1}^{L} \tilde{Y}_{i,t}^\top \tilde{X}_{i,t}\right) \\
& + \frac{2}{(\sigma^2)^l}\left(\sum_{i=1}^{N}\sum_{t=1}^{L} \tilde{X}_{i,t}^\top \tilde{X}_{i,t}\right)\boldsymbol{\beta} + \frac{2}{(\sigma^2)^l}\sum_{i=1}^{N}\sum_{t=1}^{L}\sqrt{\lambda_t}\tilde{X}_{i,t}^\top B\mu_{i,t}^l
\end{aligned} \tag{A.17}
$$

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\beta}, \Sigma, \sigma | \boldsymbol{\beta}^l, \Sigma^l, \sigma^l)}{\partial \sigma^2} = & \frac{2(a+1)}{\sigma^2} - \frac{2b}{(\sigma^2)^2} - \frac{1}{(\sigma^2)^2}(\boldsymbol{\beta}^{l+1})^\top R_\beta^{-1}\boldsymbol{\beta}^{l+1} + \frac{2}{(\sigma^2)^2}(\boldsymbol{\beta}^{l+1})^\top R_\beta^{-1}\mu_\beta \\
& - \frac{1}{(\sigma^2)^2}\mu_\beta^\top R_\beta^{-1}\mu_\beta + \frac{NML}{\sigma^2} - \frac{1}{(\sigma^2)^2}\sum_{i=1}^{N}\sum_{t=1}^{L}\tilde{Y}_{i,t}^\top \tilde{Y}_{i,t} \\
& + \frac{2}{(\sigma^2)^2}\sum_{i=1}^{N}\sum_{t=1}^{L}\tilde{Y}_{i,t}^\top \tilde{X}_{i,t}\boldsymbol{\beta}^{l+1} - \frac{1}{(\sigma^2)^2}(\boldsymbol{\beta}^{l+1})\left(^\top\sum_{i=1}^{N}\sum_{t=1}^{L}\tilde{X}_{i,t}^\top \tilde{X}_{i,t}\right)\boldsymbol{\beta}^{l+1} \\
& + \frac{2}{(\sigma^2)^2}\sum_{i=1}^{N}\sum_{t=1}^{L}\sqrt{\lambda_t}\tilde{Y}_{i,t}^\top B\mu_{i,t}^l - \frac{2}{(\sigma^2)^2}(\boldsymbol{\beta}^{l+1})\sum_{i=1}^{N}\sum_{t=1}^{L}\sqrt{\lambda_t}\tilde{X}_{i,t}^\top B\mu_{i,t}^l \\
& - \frac{1}{(\sigma^2)^2}\sum_{i=1}^{N}\sum_{t=1}^{L}\operatorname{tr}\left\{\lambda_t B^\top B(\Sigma_t^l + \Omega_{i,t}^l)\right\}
\end{aligned} \tag{A.18}
$$

$$
\frac{\partial Q(\boldsymbol{\beta}, \Sigma, \sigma | \boldsymbol{\beta}^l, \Sigma^l, \sigma^l)}{\partial \Sigma} = (\nu + J + 1)\Sigma^{-1} - \Sigma^{-1}\Psi\Sigma^{-1} + NL\Sigma^{-1} - \sum_{i=1}^{N}\sum_{t=1}^{L}\Sigma^{-1}\left(\Sigma_t^l + \Omega_{i,t}^l\right)\Sigma^{-1} \tag{A.19}
$$

By letting each partial derivatives be 0, we solve $\boldsymbol{\beta}^{l+1}$, Σ^{l+1}, σ^{l+1} as:

$$
\hat{\boldsymbol{\beta}}^{l+1} = \left[R_\beta^{-1} + \sum_{i=1}^{N}\sum_{t=1}^{L}\tilde{X}_{i,t}^\top \tilde{X}_{i,t}\right]^{-1}\left[R_\beta^{-1}\mu_\beta + \sum_{i=1}^{N}\sum_{t=1}^{L}\tilde{Y}_{i,t}^\top \tilde{X}_{i,t} - \sum_{i=1}^{N}\sum_{t=1}^{L}\sqrt{\lambda_t}\tilde{X}_{i,t}^\top B\mu_{i,t}^l\right] \tag{9}
$$

$$
\begin{aligned}
(\hat{\sigma}^2)^{l+1} = & \frac{1}{NML + 2a + 2}\left\{2b + \mu_\beta^\top R_\beta^{-1}\mu_\beta + \sum_{i=1}^{N}\sum_{t=1}^{L}\tilde{Y}_{i,t}^\top \tilde{Y}_{i,t} - 2\sum_{i=1}^{N}\sum_{t=1}^{L}\sqrt{\lambda_t}\tilde{Y}_{i,t}^\top B\mu_{i,t}^l\right. \\
& + \sum_{i=1}^{N}\sum_{t=1}^{L}\operatorname{tr}\left\{\lambda_t B^\top B\left[\Sigma_t^l + \Omega_{i,t}^l\right]\right\} + \left(\boldsymbol{\beta}^{l+1}\right)^\top R_\beta^{-1}\left(\boldsymbol{\beta}^{l+1}\right) \\
& - 2\left(\boldsymbol{\beta}^{l+1}\right)^\top R_\beta^{-1}\mu_\beta - 2\left(\sum_{i=1}^{N}\sum_{t=1}^{L}\tilde{Y}_{i,t}^\top \tilde{X}_{i,t}\right)\boldsymbol{\beta}^{l+1} \\
& \left. + \left(\boldsymbol{\beta}^{l+1}\right)^\top\left(\sum_{i=1}^{N}\sum_{t=1}^{L}\tilde{X}_{i,t}^\top \tilde{X}_{i,t}\right)\boldsymbol{\beta}^{l+1} + 2\left(\boldsymbol{\beta}^{l+1}\right)^\top\left(\sum_{i=1}^{N}\sum_{t=1}^{L}\sqrt{\lambda_t}\tilde{X}_{i,t}^\top B\mu_{i,t}^l\right)\right\}
\end{aligned} \tag{10}
$$

$$
\hat{\Sigma}^{l+1} = \frac{1}{NL + \nu + J + 1}\left[\Psi + \sum_{i=1}^{N}\sum_{t=1}^{L}(\Sigma_t^l + \Omega_{i,t}^l)\right] \tag{11}
$$

53

A.3 Modeling Detail in Sections 2.5 and 2.6

In this section, we list all prior settings we use to conduct the simulation study and analyze the cocaine use disorder data in this research.

We denote the cocaine use disorder group as group 1 and the control group as group 2. Following the transformed model in (2.4), we set:

- U_t is the t-th column of U, the eigenvector matrix from the eigenvalue decomposition $(\Sigma_T = \mathcal{W}\mathcal{W}^\top = UVU^\top)$, of the temporal covariance matrix Σ_T.

- λ_t is the tth diagonal element of V.

- Σ_T is the time series covariance matrix of a AR_2 time series model:

$$Y_i(\mathbf{v}, t) = \phi_1 Y_i(\mathbf{v}, t-1) + \phi_2 Y_i(\mathbf{v}, t-2) + e_i(\mathbf{v}, t) \tag{A.20}$$

$$\epsilon_i(\mathbf{v}, t) \sim N(0, \sigma_\epsilon^2) \tag{A.21}$$

In this research, we use $\phi_1 = 1.723$, $\phi_2 = -0.904$ and $\sigma_\epsilon^2 = 3.121 \times 10^{-2}$. These values are estimated from the cocaine use disorder data. More details on the parameter choice can be found in Appendix A.6.3. If we ignore the spatial dependence, the data for each voxel and each subject is a time series. We fit a time series model for each of these time series. We use AIC to select the model of the best order and estimate their parameters. As a result, autoregressive model of order 2 has the best AIC value for almost all the time series, and the variations of all three parameters estimated are very small. So, we choose the median of these parameters estimated as the final value we use for the real data.

- We use the isotropic Gaussian kernel as the basis function over the 3-dimensional

voxel space

$$\phi(\mathbf{v}, \mathbf{v}_k) = \exp(-\|\mathbf{v} - \mathbf{v}_k\|_2^2/2h^2). \qquad \text{(A.22)}$$

We conduct an empirical analysis with respect to the isotropic Gaussian kernel we used in real data analysis through empirical variogram to show that the spatial variability in the real data does not vary along with three-dimensional spatial directions. We fit variogram models for each subject at each volume. For each volume, we fix a dimension and fit a two-dimension variogram for the other 2 spatial dimensions and then switch to other dimensions. We record all the fitting results. Figure A.1 shows the distributions of sill and range of all these models. We can see that both sill and range are roughly at the same range across three directions. Thus, the isotropic Gaussian kernel basis functions over 3-dimensional voxel space is adequate.
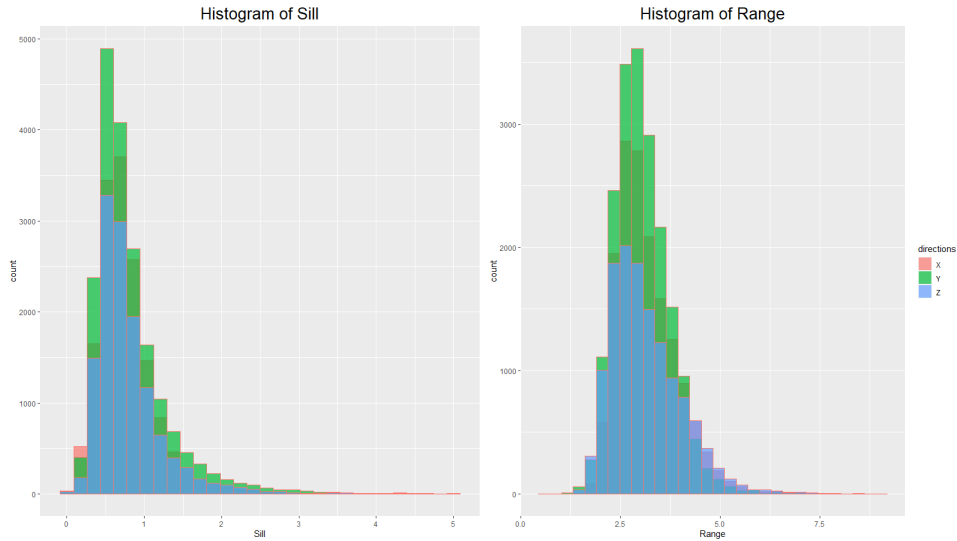
Fig. A.1. Histograms of sill and range parameters on each directions.

- In this research, we choose h to be a half of the distance between two spatial

basis nodes.

- We use, $J = 5 \times 5 \times 5$, a total of 125 equally spaced spatial basis nodes for the whole brain. Thus, the size of the spatial dependence matrix Σ is 125×125.

For the real data in Section 2.6, the linear factors described in Table 2.1 are not significant to the response via a z-test with significant level 0.0001.

A.4   A Simulation Study on the Permutation Test in Section 2.6.2

We use synthetic datasets to assess the effectiveness of the permutation test for $H_0 : \Sigma_1 = \Sigma_2$ in Section 2.6.2. We generate our simulation dataset the same way as in Section 2.5, but instead of generating one set of data, for each replication, we generate two sets of data as two treatment groups. For each group, we also use exponential Toeplitz covaraince matrix Σ_g^0 with parameter θ_g as the true reduced spatial dependence matrix:

$$\Sigma^0{}_{g,i,j} = \theta_g^{|i-j|}, i = 1, \ldots, J, j = 1, \ldots, J. \tag{A.23}$$

By changing the value of θ_g for each group, we are able to assess the our proposed test statistic under different level of variation. This test statistic is always positive, and the larger it is, more different Σ_1 and Σ_2 are. In this study, we choose θ_g from 0.1 to 0.7, where a large θ_g indicates strong spatial dependence.

We use permutation test over 100 permutation samples to test the power of this test statistics. For each combination of θ_g's, we run 50 replications. The significance of our test is determined by the p-value, which is defined as the number of permutation samples that has larger test statistic value than the permutation population over 100 permutation samples. Here, we report the ratio of the number of replications with significant p-value (smaller than 0.05) over 50 replications for each combination of

56

θ_g's. The result is shown in Table A.1.

Table A.1. Median, first and third quantiles of the p-values of the permutation test over 50 replications.

	θ_1	θ_2	First Quantile	Median	Third Quantile
H_0 is True	0.100	0.100	0.010	0.060	0.275
	0.300	0.300	0.140	0.420	0.730
	0.500	0.500	0.180	0.335	0.610
	0.700	0.700	0.188	0.455	0.697
H_0 is Not True	0.100	0.300	<0.0001	<0.0001	<0.0001
	0.100	0.500	<0.0001	<0.0001	<0.0001
	0.100	0.700	<0.0001	<0.0001	<0.0001
	0.300	0.500	<0.0001	<0.0001	<0.0001
	0.300	0.700	<0.0001	<0.0001	<0.0001
	0.500	0.700	<0.0001	<0.0001	<0.0001

A.5   Participants and Data Preprocessing

This study was approved by the institutional review board at the Virginia Commonwealth University and was performed in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Subjects with current cocaine use disorder (DSM-IV) and non-drug-using normal controls were recruited through advertisements for research volunteers. Informed consent was obtained from all subjects.

All subjects were screened for psychiatric disorders using the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edi-

tion (DSM-IV; First et al., 1996). All subjects underwent physical examination, and their medical history was obtained. Subjects also underwent the Addiction Severity Index (McLellan et al., 1992) to document possible lifetime drug and alcohol use. All female subjects underwent a urine pregnancy test immediately before MRI scanning. Each subject's urine was screened for amphetamine, barbiturates, buprenorphine, benzodiazepines, cocaine, methamphetamine, MDMA (ecstasy), methadone, opiates, oxycodone, phencyclidine (PCP), propoxyphene (PPX), tricyclic antidepressants, marijuana (THC) using a device called 14 PANEL URINE DRUG SCREENS | T-CUP TDOA-1145A3 (Wondfo Biotech Co., Ltd, Guangzhou, China), and each subject was screened for alcohol using Alco-Sensor FST (Intoximeters, St. Louis, MO) immediately before MRI scanning.

Subject inclusion criteria were: (1) 18-55 years old; (2) free of alcohol (per breath alcohol screen) at the time of MRI scanning; (3) cocaine use disorder (CocUD) subjects met Diagnostic and Statistical Manual Fourth Edition (DSM-IV) (American Psychiatric Association, 2000) criteria for current cocaine use disorder based on Structured Clinical Interview for DSM-IV (First et al., 1996) , and (4) normal control subjects had no current or lifetime history of any DSM-IV substance use or psychiatric disorder. Exclusion criteria were: (1) left-handed; (2) CocUD subjects who met current or past DSM- IV Axis I disorder other than substance abuse or substance dependence; (3) medical disorders or taking medication that may affect the central nervous system; (4) claustrophobia experienced during MRI or MRI simulator sessions; (5) any definite or suspected clinically significant abnormalities of the brain on Fluid Attenuated Inversion Recovery (FLAIR) MRI scans, as read prior to data analysis by a board certified radiologist; (6) positive urine drug screen for control subjects; (7) positive pregnancy test result. After applying for the inclusion and exclusion criteria, a total of 22 CocUD subjects (cocaine group) and 23 control subjects

(control group) were included in this study.

The fMRI data were preprocessed using the CONN toolbox version 17.f (`http://www.nitrc.org/projects/conn`) (Whitfield-Gabrieli and Nieto-Castanon, 2012b), based on Statistical Parametric Mapping 12 (SPM12) software (`http://www.fil.ion.ucl.ac.uk/spm/`), and Matlab R2015a (Mathworks Inc. Sherborn MA, USA). The T1-weighted MPRAGE image and the resting state fMRI images were input to CONN as the structural image and functional images respectively. All preprocessing steps were conducted using the pipeline for volume-based analysis (to Montreal Neurological Institute or MNI space). Specifically, each subject's functional images were realigned to the first volume. Then they were unwarped, slice-timing corrected (ascending-sequential or foot to head), co-registered with structural image, and spatially normalized into the standard MNI space. The final images were smoothed using a Gaussian kernel of 8 mm isotropic full width at half maximum. After these, default aComCor fMRI denoising (Behzadi et al., 2007) and fMRI band-pass filtering (0.008-0.10 Hz) were applied as pre-processing.

A.6   Model Justification

A.6.1   Common Variance Assumption

In this section, we justify the common variance assumption in our proposed model (2.1). We estimate the white noise variance for each subject in the cocaine use disorder multi-subject FMRI data separately using our proposed method. Mean and standard deviation of the white noise variance of each subject separately are compared with the estimated white noise variance when assuming homogeneity in Table A.2 and in histogram A.6.1. The results show that the homogeneous assumption makes little difference in the estimation. It is worth noting that We standardized the data to

range (-1, 1) for each subject before feed it into our model since the range of the real data is very different across subjects. Then we can conclude that the common variance assumption is appropriate for this data.

Table A.2. Mean and standard deviation of white noise variances for each subject in the cocaine use disorder multi-subject FMRI, and the estimated white noise variance when assuming homogeneity.

	Mean	Standard Deviation
Single Subject Estimation	0.7825	3.107×10^{-2}
All subjects assuming Homogeneous Variance	0.7824	
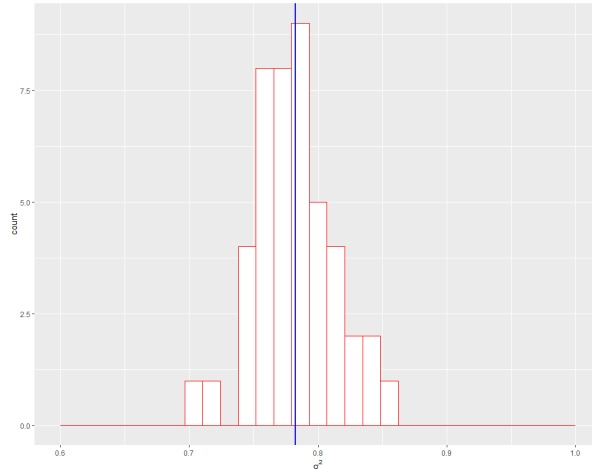
Fig. A.2. Histogram of separately estimated white noise variances for each subject in the cocaine use disorder multi-subject FMRI. The blue vertical line is the estimated white noise variance when assuming homogeneity.

A.6.2   Common Linear Coefficients

In this section, we explore the significance of linear coefficients ($\boldsymbol{\beta}$). We fit a linear regression model on each voxel at each volume using the standardized Bold signal from all subjects as the dependent variable and four demographic variables

as independent variables. We record p-value's for all coefficients. The results show that all coefficients have p-value greater than 0.9999 at all voxel locations. This demonstrates that the demo-graphical covariates are insignificant to Bold signal, and it is not necessary to make further exploration to more complex structure of the linear covariates.

A.6.3 Choice of Temporal Basis Function and Spatio-Temporal Separability

The fMRI data is a 4-dimensional spatio-temporal data. There is a time series at each spatial (voxel) location. Before analyzing the real data, we first explore the temporal nature of the cocaine use disorder dataset. We fit multiple time series model for each voxel for each subject in the dataset including AR(1), AR(2), AR(3), MA(1) and MA(2). We record all the parameter estimates and BIC value for each of the model at each location. We found that AR(2) model fits the best for almost all time series according to the BIC value. Histograms A.6.3 show all parameter estimates of AR(2) model at all voxel locations, and table A.3 reports the mean, standard deviation of the mean, p-value and standard deviation of p-values of these AR(2) model estimates. The results show that the time series fitted at different spatial location for different subjects are very similar. Thus, we use a homogeneous AR(2) temporal basis function with mean estimates as parameters from the real data model, to represent the temporal dependence of the cocaine use disorder data. We can also conclude from these results that the space and time domain in our data are separable, since the temporal correlation does not affected by spatial locations of the data.
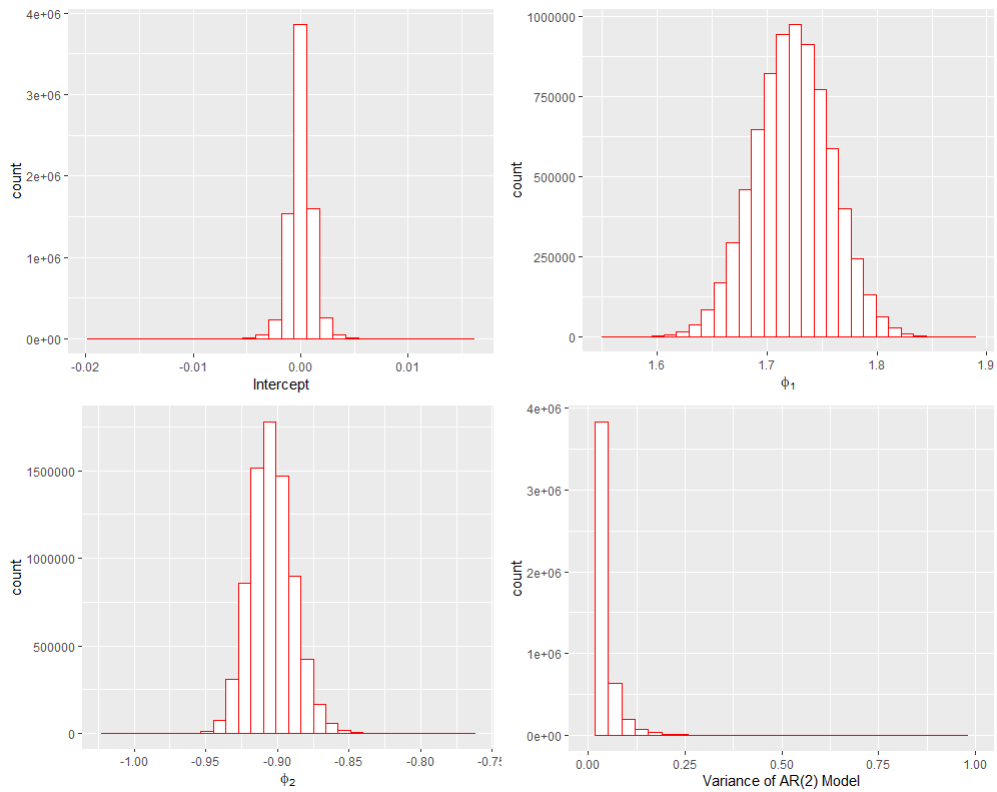
Fig. A.3. Histogram of all parameter estimates of AR(2) model at all voxel locations for real data AR(2) time series models.

Table A.3. Mean parameter estimates, standard deviation of mean (in parenthesis), p-value and standard deviation of p-value (in parenthesis) for real data AR(2) time series models.

	mean estimates (standard deviation of the mean)	p-value (standard deviation of p-values)
Intercept	1.852×10^{-5} (1.062×10^{-3})	0.9259 (5.719×10^{-2})
ϕ_1	1.723 (3.469×10^{-2})	< 0.0001 (< 0.00001)
ϕ_2	-0.904 (1.514×10^{-2})	< 0.0001 (< 0.00001)
σ^2	3.121 (3.640×10^{-2})	

A.6.4 Model Performance on Complex Connectivity

In section 2.5, the random factors $\Gamma_{i,l}$'s in (2.2) are independently generated from $\mathcal{N}(0, \Sigma^0)$, where Σ^0 is a $J \times J$ matrix with $\Sigma^0_{i,j} = 0.5^{|i-j|}$. In this section, we generate the simulation dataset using the $\hat{\Sigma}$ of the cocaine group estimated from the cociane use disorder fMRI dataset to study the performance of our model when the true Σ^0 are more complex. Since none of the linear coefficients is significant in the real data, we still use the same coefficients as simulation studies in section 2.5. We also report the mean estimates and standard error of mean for both linear regression model and our proposed model over 50 replications. Table A.4 reports the estimation results. The results show that our model performs very well for real connectivity.

A.6.5 Isotropic Gaussian Kernel

Here, we conduct an empirical analysis with respect to the isotropic Gaussian kernel we used in real data analysis through empirical variogram to show that the spatial variability in the real data does not vary along with three-dimensional spatial directions. We fit variogram models for each subject at each volume. For each volume, we fix a dimension and fit a two-dimension variogram for the other 2 spatial dimensions and then switch to other dimensions. We record all the fitting results.

63

Table A.4. Simulation study mean parameter estimates and standard error of mean (in parenthesis) for data simulated from real data spatial covariance structure over 50 replications.

True Value	Model	
	Linear Regression	Spatio-Temporal Model
$\mu = 0.1$	$0.00314\ (2.313 \times 10^{-4})$	$0.0996\ (2.522 \times 10^{-4})$
$\beta_1 = 0.5$	$0.01577\ (3.157 \times 10^{-5})$	$0.5001\ (3.099 \times 10^{-5})$
$\beta_2 = 1$	$0.03146\ (2.213 \times 10^{-5})$	$1.000\ (2.143 \times 10^{-5})$
$\sigma^2 = 1$	$83.305\ (1.264 \times 10^{-3})$	$0.9997\ (7.626 \times 10^{-5})$
TS		$4.413 \times 10^{-6}(2.830 \times 10^{-8})$

A.7 Sensitivity Analysis

In this section, we conduct a series of simulation studies under different model settings to obtain practical guidance regarding the choice of number of subjects N, number of spatial basis nodes J, number of volumes L and different types of basis functions on the real data.

The simulation datasets are generated the same way as in section 2.5. We use the covariance matrix estimated from the cocaine use disorder group of the real data in section 2.6 as the true Σ_0 matrix. We vary the N, L, J in the simulated datasets to explore the effects of parameter misspecification on the inference. We also study the performance of our model when data is generated using different basis functions. We applied our proposed model to each of the simulation dataset, and report the mean, standard error of the mean and the test statistic (2.6) proposed in section 2.5 over 50 replications under each scenario. We still use the same uniform distributions for the covariates x_1 and x_2 as we did in section 2.5 in the sensitivity analysis.

A.7.1 Number of Subjects (N)

First, we would like to study the performance of our model when the number of subjects are small. We generate simulation dataset with number of subjects from 5 to 100. The results are shown in table A.5. The parameter estimation is accurate and stable even when there is only 5 subjects in the simulated dataset.

Table A.5. Simulation study mean parameter estimates and standard error of mean (in parenthesis) for data with different number of subjects over 50 replications.

True Value	N				
	5	10	25	50	100
$\mu = 0.1$	0.09913 (6.779 $\times 10^{-4}$)	0.1002 (1.034 $\times 10^{-3}$)	0.1004 (4.905 $\times 10^{-4}$)	0.1006 (4.031 $\times 10^{-4}$)	0.1000 (2.235 $\times 10^{-4}$)
$\beta_1 = 0.5$	0.500 (3.075 $\times 10^{-4}$)	0.4999 (1.337 $\times 10^{-4}$)	0.5000 (7.092 $\times 10^{-5}$)	0.4999 (4.993 $\times 10^{-5}$)	0.5000 (3.094 $\times 10^{-5}$)
$\beta_2 = 1$	1.000 (1.683 $\times 10^{-4}$)	1.000 (7.996 $\times 10^{-5}$)	1.000 (5.473 $\times 10^{-5}$)	1.000 (4.083 $\times 10^{-5}$)	1.000 (2.721 $\times 10^{-5}$)
$\sigma^2 = 1$	0.9996 (1.329 $\times 10^{-4}$)	0.9996 (1.011 $\times 10^{-4}$)	0.9997 (6.188 $\times 10^{-4}$)	1.000 (5.111 $\times 10^{-4}$)	1.000 (4.024 $\times 10^{-5}$)
TS	2.002 $\times 10^{-4}$ (3.981 $\times 10^{-6}$)	2.154 $\times 10^{-4}$ (3.511 $\times 10^{-6}$)	1.747 $\times 10^{-4}$ (2.415 $\times 10^{-6}$)	1.145 $\times 10^{-4}$ (2.347 $\times 10^{-6}$)	6.886 $\times 10^{-5}$ (1.088 $\times 10^{-6}$)

A.7.2 Number of Total Volumes (L)

Next, we vary the number of total volumes (L) of each subject from 25 to 400 to see if our model can estimate parameters and spatial covariance matrix accurately when L is small. The results are shown in table A.6. We can see that when each subject have only 25 volumes our model still performance very well.

A.7.3 Number of Spatial Basis Nodes (J)

The number of spatial basis nodes (J) is a very important feature in Kriging methods. Using a small number of basis nodes could significantly reduced the computational time for very large spatial datasets. However, the smaller J we use, the more details we lose in the data. Choosing J for a dataset is truly a balance between accuracy and computational cost. Through this study, we would like to find a J value that can achieve computational efficiency, at the same time, without losing too much

Table A.6. Simulation study mean parameter estimates and standard error of mean (in parenthesis) for data with different number of volumes over 50 replications.

True Value	L		
	25	100	400
$\mu = 0.1$	0.1019 (2.437×10^{-3})	0.1005 (1.004×10^{-3})	0.1004 (4.905×10^{-4})
$\beta_1 = 0.5$	0.4994 (3.310×10^{-4})	0.5001 (1.990×10^{-4})	0.5000 (7.092×10^{-5})
$\beta_2 = 1$	1.000 (2.631×10^{-4})	0.9998 (1.533×10^{-5})	1.000 (5.473×10^{-5})
$\sigma^2 = 1$	0.9991 (2.412×10^{-4})	0.9995 (1.258×10^{-4})	0.9997 (6.188×10^{-4})
TS	1.127×10^{-4} (2.168×10^{-6})	2.164×10^{-4} (3.674×10^{-6})	1.816×10^{-4} (2.080×10^{-6})

information.

We first estimate the covariance matrix of the cocaine use disorder group in section 2.6 using $J_0 = 8$ (2^3), 125 (5^3) and 1000 (10^3) from the real data. We then generate data from it, and estimate parameters using $J = 8$ (2^3), 27 (3^3) and 125 (5^3). The results are shown in table A.7. We can see that even though we can achieve the most accuracy by using the same basis nodes in estimation as that in generating the simulation datasets, $J = 125$ (5^3) always provides good estimation results. Hence, we choose $J = 125$ (5^3) in this study to estimate the spatial covariance structure for the cocaine use disorder fMRI data.

A.7.4 Basis Function

The types of basis functions will also affect the performance of dimension reduction methods. Some basis functions imply strong relationships among close spatial locations, while some others have impacts across large spatial domain.

The Gaussian basis function is widely used for large scale spatial datasets due to its computational convenience and flexibility. In all simulation studies above, we generate simulation datasets using the Gaussian basis function and explore the model

66

Table A.7. Simulation study mean parameter estimates and standard error of mean (in parenthesis) for data with different number of spatial basis nodes J under different true J_0 over 50 replications.

J_0	True Value	J		
		$2 \times 2 \times 2$	$3 \times 3 \times 3$	$5 \times 5 \times 5$
	$\mu = 0.1$	0.0997 (2.896×10^{-4})	0.1017 (1.459×10^{-3})	0.100 (3.327×10^{-4})
	$\beta_1 = 0.5$	0.5000 (2.688×10^{-5})	0.4999 (2.064×10^{-4})	0.500 (3.010×10^{-5})
$2 \times 2 \times 2$	$\beta_2 = 1$	1.000 (3.879×10^{-5})	0.9998 (2.309×10^{-4})	1.000 (4.030×10^{-5})
	$\sigma^2 = 1$	1.000 (6.473×10^{-5})	1.0438 (1.585×10^{-4})	0.9944 (6.773×10^{-5})
	TS	6.047×10^{-5} (6.342×10^{-6})	1.789×10^{-2} (1.557×10^{-4})	4.647×10^{-3} (1.028×10^{-4})
	$\mu = 0.1$	0.1028 (1.758×10^{-2})	0.1005 (1.384×10^{-3})	0.1001 (4.772×10^{-4})
	$\beta_1 = 0.5$	0.4997 (2.029×10^{-4})	0.5000 (1.730×10^{-4})	0.5000 (6.592×10^{-5})
$5 \times 5 \times 5$	$\beta_2 = 1$	0.9997 (2.093×10^{-4})	1.000 (1.675×10^{-4})	1.000 (7.139×10^{-5})
	$\sigma^2 = 1$	1.603 (8.583×10^{-4})	1.355 (4.760×10^{-4})	0.9997 (5.841×10^{-5})
	TS	2.225×10^{-2} (2.693×10^{-5})	1.700×10^{-2} (7.802×10^{-5})	1.729×10^{-4} (2.459×10^{-6})
	$\mu = 0.1$	0.0997 (1.401×10^{-3})	0.0994 (7.767×10^{-4})	0.1007 (8.662×10^{-4})
	$\beta_1 = 0.5$	0.4999 (1.511×10^{-4})	0.5001 (8.875×10^{-5})	0.4999 (8.961×10^{-5})
$10 \times 10 \times 10$	$\beta_2 = 1$	1.000 (1.548×10^{-4})	1.000 (9.645×10^{-5})	1.000 (9.454×10^{-5})
	$\sigma^2 = 1$	1.902 (4.776×10^{-4})	1.784 (4.193×10^{-4})	1.496 (2.621×10^{-4})
	TS	1.930 (8.845×10^{-4})	1.234 ($86..659 \times 10^{-4}$)	2.944×10^{-2} (2.655×10^{-4})

performance using the same basis function. The simulation results are satisfying. However, we don't not know the truth for the real data. We would like to investigate if Gaussian basis function would work well under different scenarios. In this simulation study, we use the Gaussian basis function to estimate the model when the datasets are generated from the exponential basis function and the bisquare basis function. The expression of Gaussian basis function is

$$\phi(\Delta) = \exp\left(-\frac{||\Delta||^2}{2\sigma^2}\right), \tag{A.24}$$

67

exponential basis function is

$$\phi(\Delta) = \exp\left(-\frac{||\Delta||}{\tau}\right), \tag{A.25}$$

and bisquare basis function is

$$\phi(\Delta) = \left(1 - \left(\frac{||\Delta||}{R}\right)^2\right)^2 I(||\Delta|| < R), \tag{A.26}$$

where Δ is the Euclidean distance between a pair of spatial locations. The parameters σ, τ, R are scale arguments. We choose $\tau = 5$ and 10 for the exponential basis function and $R = 5$ and 10 for the bisquare basis function to test the performance of Gaussian basis function in more complex scenarios. We use $J = 125$ (5^3) in both data generating and estimating. In model estimating, we use $\sigma = 5$ as the scale parameter for the Gaussian basis function.

The simulation results are shown in table A.8. We can see that the Gaussian basis function works very well even when the data are generated from exponential and bisquare basis functions with different scale parameters. Thus, we are confident to use Gaussian basis function to analyze the real data.

68

Table A.8. Simulation study mean parameter estimates and standard error of mean (in parenthesis) when using Gaussian basis function to estimate exponential and bisquare data over 50 replications.

Basis Functions	Scale Parameter	True Value	Estimates
Exponential	$\tau = 5$	$\mu = 0.1$	$0.0992\ (8.217 \times 10^{-4})$
		$\beta_1 = 0.5$	$0.4999\ (8.532 \times 10^{-5})$
		$\beta_2 = 1$	$1.000\ (9.298 \times 10^{-5})$
		$\sigma^2 = 1$	$1.016\ (7.344 \times 10^{-5})$
		TS	$1.790 \times 10^{-3}\ (1.757 \times 10^{-5})$
	$\tau = 10$	$\mu = 0.1$	$0.0990\ (1.295 \times 10^{-3})$
		$\beta_1 = 0.5$	$0.5001\ (1.571 \times 10^{-4})$
		$\beta_2 = 1$	$1.000\ (1.588 \times 10^{-4})$
		$\sigma^2 = 1$	$1.017\ (7.265 \times 10^{-5})$
		TS	$8.333 \times 10^{-3}\ (1.125 \times 10^{-4})$
Bisquare	$R = 5$	$\mu = 0.1$	$0.1002\ (4.970 \times 10^{-4})$
		$\beta_1 = 0.5$	$0.5000\ (5.585 \times 10^{-5})$
		$\beta_2 = 1$	$0.9999\ (4.750 \times 10^{-5})$
		$\sigma^2 = 1$	$1.089\ (1.107 \times 10^{-4})$
		TS	$4.223 \times 10^{-3}\ (1.297 \times 10^{-5})$
	$R = 10$	$\mu = 0.1$	$0.0997\ (4.640 \times 10^{-4})$
		$\beta_1 = 0.5$	$0.5000\ (4.360 \times 10^{-5})$
		$\beta_2 = 1$	$1.000\ (6.073 \times 10^{-5})$
		$\sigma^2 = 1$	$1.018\ (6.767 \times 10^{-5})$
		TS	$1.854 \times 10^{-3}\ (5.883 \times 10^{-6})$

REFERENCES

[1] Fidel Alfaro-Almagro et al. "Confound modelling in UK Biobank brain imaging". In: *NeuroImage* 224 (2021), p. 117002.

[2] Nelly Amador and Itzhak Fried. "Single-neuron activity in the human supplementary motor area underlying preparation for action". In: *Journal of Neurosurgery* 100 (Feb. 2004), pp. 250–259. (Visited on 04/19/2020).

[3] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, Fourth Edition, Text revision.* 2000.

[4] Yashar Behzadi et al. "A component based noise correction method (CompCor) for BOLD and perfusion based fMRI". In: *NeuroImage* 37 (Aug. 2007), pp. 90–101. (Visited on 03/28/2020).

[5] Osnat M. Ben-Shahar et al. "Extended access to cocaine self-administration results in reduced glutamate function within the medial prefrontal cortex". In: *Addiction Biology* 17 (Feb. 2012), pp. 746–757. (Visited on 03/28/2020).

[6] H. Benali et al. "Space-time statistical model for functional MRI image sequences". In: *Lecture Notes in Computer Science* (1997), 285–298. DOI: `10.1007/3-540-63046-5_22`.

[7] K.I Bolla et al. "Orbitofrontal cortex dysfunction in abstinent cocaine abusers performing a decision-making task". In: *NeuroImage* 19 (July 2003), pp. 1085–1094. (Visited on 02/27/2020).

[8] F. Dubois Bowman. "Spatio-temporal modeling of localized brain activity". In: *Biostatistics* 6.4 (2005), 558–575. DOI: `10.1093/biostatistics/kxi027`.

[9] Stephen Boyd. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". In: *Foundations and Trends® in Machine Learning* 3 (2011), pp. 1–122. DOI: 10.1561/2200000016. (Visited on 10/10/2019).

[10] Ed Bullmore and Olaf Sporns. "Erratum: Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nature Reviews Neuroscience* 10 (Mar. 2009), pp. 312–312. DOI: 10.1038/nrn2618. URL: https://www.nature.com/articles/nrn2618 (visited on 12/09/2019).

[11] Carter T. Butts. "Revisiting the Foundations of Network Analysis". In: *Science* 325.5939 (2009), 414–416. DOI: 10.1126/science.1171022.

[12] C. M. Carvalho, N. G. Polson, and J. G. Scott. "The horseshoe estimator for sparse signals". In: *Biometrika* 97 (Apr. 2010), pp. 465–480. DOI: 10.1093/biomet/asq017. (Visited on 10/16/2020).

[13] Julien Chiquet, Yves Grandvalet, and Christophe Ambroise. "Inferring multiple graphical structures". In: *Statistics and Computing* 21 (June 2010), pp. 537–553. DOI: 10.1007/s11222-010-9191-2. (Visited on 10/16/2020).

[14] Jonathan D Cohen et al. "Computational approaches to fMRI analysis". In: *Nature Neuroscience* 20.3 (2017), 304–313. DOI: 10.1038/nn.4499.

[15] Noel Cressie and Gardar Johannesson. "Fixed rank kriging for very large spatial data sets". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.1 (2008), pp. 209–226.

[16] Patrick Danaher, Pei Wang, and Daniela M. Witten. "The joint graphical lasso for inverse covariance estimation across multiple classes". In: *Journal of the*

Royal Statistical Society: Series B (Statistical Methodology) 76 (Aug. 2013), pp. 373–397. (Visited on 04/07/2020).

[17]   Patrick Danaher, Pei Wang, and Daniela M. Witten. "The joint graphical lasso for inverse covariance estimation across multiple classes". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76 (Aug. 2014), pp. 373–397. DOI: `10.1111/rssb.12033`.

[18]   Massimo Filippi et al. "Resting State Dynamic Functional Connectivity in Neurodegenerative Conditions: A Review of Magnetic Resonance Imaging Findings". In: *Frontiers in Neuroscience* 13 (2019). DOI: `10.3389/fnins.2019.00657`.

[19]   M. First et al. "Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Editions (SCID-I/P, Version 2.0)". In: *Biometrics Research Department, New York State Psychiatric Institute, New York* (1996).

[20]   Michael D. Fox and Marcus E. Raichle. "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging". In: *Nature Reviews Neuroscience* 8 (Sept. 2007), pp. 700–711. URL: `https://www.nature.com/articles/nrn2201` (visited on 06/05/2019).

[21]   Teresa R Franklin et al. "Decreased gray matter concentration in the insular, orbitofrontal, cingulate, and temporal cortices of cocaine patients". In: *Biological Psychiatry* 51 (Jan. 2002), pp. 134–142. (Visited on 02/22/2020).

[22]   J. Friedman, T. Hastie, and R. Tibshirani. "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9.3 (2007), pp. 432–441. DOI: `10.1093/biostatistics/kxm045`.

[23] J. Friedman, T. Hastie, and R. Tibshirani. "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9 (Dec. 2008), pp. 432–441. DOI: 10.1093/biostatistics/kxm045.

[24] Karl J. Friston. "Functional and effective connectivity: a review". In: *Brain Connectivity* 1.1 (2011), 13–36. DOI: 10.1089/brain.2011.0008.

[25] J. Guo et al. "Joint estimation of multiple graphical models". In: *Biometrika* 98 (Feb. 2011), pp. 1–15. DOI: 10.1093/biomet/asq060. (Visited on 04/16/2020).

[26] Jaryd Hiser and Michael Koenigs. "The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology". In: *Biological Psychiatry* 83 (Apr. 2018), pp. 638–647. (Visited on 04/29/2019).

[27] Scott A Huettel, Allen W Song, and Gregory Mccarthy. *Functional magnetic resonance imaging.* Sinauer Associates, Inc., Publishers, 2014.

[28] Jung Won Hyun et al. "STGP: Spatio-temporal Gaussian process models for longitudinal neuroimaging data". In: *Neuroimage* 134 (2016), pp. 550–562.

[29] Marc Joliot et al. "AICHA: An atlas of intrinsic connectivity of homotopic areas". In: *Journal of Neuroscience Methods* 254 (2015), 46–59. DOI: 10.1016/j.jneumeth.2015.07.013.

[30] Hakmook Kang et al. "Spatio-Spectral Mixed-Effects Model for Functional Magnetic Resonance Imaging Data". In: *Journal of the American Statistical Association* 107.498 (2012), 568–577. DOI: 10.1080/01621459.2012.664503.

[31] Kota Katanoda, Yasumasa Matsuda, and Morihiro Sugishita. "A Spatio-temporal Regression Model for the Analysis of Functional MRI Data". In: *NeuroImage* 17.3 (2002), 1415–1428. DOI: 10.1006/nimg.2002.1209.

73

[32]   G. D. Konidaris, S. Osentoski, and P. S. Thomas. "Value function approxima-
       tion in reinforcement learning using the Fourier basis". In: *Proceedings of the
       Twenty Fifth Conference on Artificial Intelligence* (2011), 380–395.

[33]   Daniel H. Lench, William DeVries, and Colleen A. Hanlon. "The effect of task
       difficulty on motor performance and frontal-striatal connectivity in cocaine
       users". In: *Drug and Alcohol Dependence* 173 (Apr. 2017), pp. 178–184. (Visited
       on 04/19/2020).

[34]   Yunfan Li, Bruce A. Craig, and Anindya Bhadra. "The Graphical Horseshoe
       Estimator for Inverse Covariance Matrices". In: *Journal of Computational and
       Graphical Statistics* 28 (Apr. 2019), pp. 747–757. DOI: 10.1080/10618600.
       2019.1575744. (Visited on 10/16/2020).

[35]   Kelvin O Lim et al. "Reduced frontal white matter integrity in cocaine depen-
       dence: a controlled diffusion tensor imaging study". In: *Biological Psychiatry*
       51 (June 2002), pp. 890–895. (Visited on 03/04/2019).

[36]   Martin A Lindquist et al. "The statistical analysis of fMRI data". In: *Statistical
       science* 23.4 (2008), pp. 439–464.

[37]   Weidong Liu. "Gaussian graphical model estimation with false discovery rate
       control". In: *The Annals of Statistics* 41.6 (2013), 2948–2978. DOI: 10.1214/
       13-aos1169.

[38]   Liangsuo Ma et al. "Effect of cocaine dependence on brain connections: clinical
       implications". In: *Expert Review of Neurotherapeutics* 15 (Oct. 2015), pp. 1307–
       1319. (Visited on 05/18/2020).

[39] A.Thomas McLellan et al. "The fifth edition of the addiction severity index". In: *Journal of Substance Abuse Treatment* 9 (June 1992), pp. 199–213. (Visited on 03/28/2020).

[40] Nicolai Meinshausen and Peter Bühlmann. "High-dimensional graphs and variable selection with the Lasso". In: *The Annals of Statistics* 34 (June 2006), pp. 1436–1462. DOI: 10.1214/009053606000000281. (Visited on 09/01/2019).

[41] Eric Nestler. "The Neurobiology of Cocaine Addiction". In: *Science & Practice Perspectives* 3 (Dec. 2005), pp. 4–10.

[42] Hernando Ombao et al. *Handbook of neuroimaging data analysis*. CRC Press, 2016.

[43] Christine Peterson, Francesco C. Stingo, and Marina Vannucci. "Bayesian Inference of Multiple Gaussian Graphical Models". In: *Journal of the American Statistical Association* 110 (Jan. 2015), pp. 159–174. DOI: 10.1080/01621459.2014.896806. (Visited on 04/07/2020).

[44] Juho Piironen and Aki Vehtari. "Sparsity information and regularization in the horseshoe and other shrinkage priors". In: *Electronic Journal of Statistics* 11 (2017), pp. 5018–5051. DOI: 10.1214/17-ejs1337si. (Visited on 10/16/2020).

[45] Adeel Razi et al. "Large-scale DCMs for resting-state fMRI". In: *Network Neuroscience* 1.3 (2017), 222–241. DOI: 10.1162/netn_a_00015.

[46] Brian J. Reich et al. "Fully Bayesian spectral methods for imaging data". In: *Biometrics* 74.2 (2017), 645–652. DOI: 10.1111/biom.12782.

[47] Adam J. Rothman et al. "Sparse permutation invariant covariance estimation". In: *Electronic Journal of Statistics* 2 (2008), pp. 494–515. DOI: 10.1214/08-ejs176.

75

[48]   Michael Shvartsman et al. "Matrix-normal models for fMRI analysis". In: *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. Ed. by Amos Storkey and Fernando Perez-Cruz. Vol. 84. Proceedings of Machine Learning Research. PMLR, 2018, pp. 1914–1923. URL: `http://proceedings.mlr.press/v84/shvartsman18a.html`.

[49]   S. M. Smith et al. "Correspondence of the brains functional architecture during activation and rest". In: *Proceedings of the National Academy of Sciences* 106.31 (2009), 13040–13045. DOI: `10.1073/pnas.0905267106`.

[50]   Stephen M. Smith et al. "Network modelling methods for fMRI". In: *NeuroImage* 54.2 (2011), pp. 875–891. DOI: `10.1016/j.neuroimage.2010.08.063`.

[51]   Matthew T. Sutherland et al. "Resting state functional connectivity in addiction: Lessons learned and a road ahead". In: *NeuroImage* 62.4 (2012), 2281–2295. DOI: `10.1016/j.neuroimage.2012.01.117`.

[52]   Caroline Uhler. "Gaussian Graphical Models: An Algebraic and Geometric Perspective". In: *arXiv preprint* (2017). DOI: `arXiv:1707.04345`.

[53]   Nora D. Volkow et al. "Long-Term frontal brain metabolic changes in cocaine abusers". In: *Synapse* 11 (July 1992), pp. 184–190. (Visited on 11/24/2019).

[54]   Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. "Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks". In: *Brain Connectivity* 2.3 (2012), 125–141. DOI: `10.1089/brain.2012.0073`.

[55]   Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. "Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks". In: *Brain Connectivity* 2 (June 2012), pp. 125–141. (Visited on 09/27/2019).

[56] D. M. Witten, R. Tibshirani, and T. Hastie. "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis". In: *Biostatistics* 10.3 (2009), 515–534. DOI: `10.1093/biostatistics/kxp008`.

[57] M.w. Woolrich et al. "Fully Bayesian Spatio-Temporal Modeling of fMRI Data". In: *IEEE Transactions on Medical Imaging* 23.2 (2004), 213–231. DOI: `10.1109/tmi.2003.823065`.

[58] CF Jeff Wu et al. "On the convergence properties of the EM algorithm". In: *The Annals of statistics* 11.1 (1983), pp. 95–103.

[59] Masanao Yajima et al. "Differential Patterns of Interaction and Gaussian Graphical Models". In: *COBRA Preprint Series* (2012), p. 91.

[60] M. Yuan and Y. Lin. "Model selection and estimation in the Gaussian graphical model". In: *Biometrika* 94 (Feb. 2007), pp. 19–35. DOI: `10.1093/biomet/asm018`.

[61] Ming Yuan and Yi Lin. "Model selection and estimation in regression with grouped variables". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (Feb. 2006), pp. 49–67. DOI: `10.1111/j.1467-9868.2005.00532.x`. URL: `https://www.stat.wisc.edu/~myuan/papers/glasso.final.pdf` (visited on 05/02/2019).

[62] Bai Zhang and Yue Wang. "Learning structural changes of Gaussian graphical models in controlled experiments". In: *In Proc. of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence* UAI 2010 (2010), 701–708.

[63] Lin Zhang et al. "Functional CAR models for large spatially correlated functional datasets". In: *Journal of the American Statistical Association* 111.514 (2016), pp. 772–786.

[64] S. D. Zhao, T. T. Cai, and H. Li. "Direct estimation of differential networks". In: *Biometrika* 101 (May 2014), pp. 253–268. DOI: 10.1093/biomet/asu009. (Visited on 10/16/2020).

[65] Hongtu Zhu, Jianqing Fan, and Linglong Kong. "Spatially Varying Coefficient Model for Neuroimaging Data With Jump Discontinuities". In: *Journal of the American Statistical Association* 109.507 (2014), 1084–1098. DOI: 10.1080/01621459.2014.881742.