Spatiotemporal Modeling of Resting State Brain Imaging Data for Functional Connectivity Analysis

BY

RAWAN ALLOZI RUPNOW
B.A., Washington University in St. Louis, 2009
M.S., Johns Hopkins University, 2012

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:

Dulal Bhaumik, Chair and Advisor
Supriya Mehta
Olusola Ajilore, Psychiatry
Runa Bhaumik, Psychiatry
Robert Krafty, University of Pittsburgh

To my family, who have supported me unconditionally along the way.

ACKNOWLEDGMENT

I would not have reached this point in my life without the endless love and support from numerous people. The journey here began years ago. I would like to thank my parents for putting me on that journey and acting as role models along the way. They always encouraged me and taught me to never settle, especially when it is the easier option. They believed in me when I did not. I would also like to thanks my siblings: Ala, Amal, Amer, and Renee. There is no doubt in my mind that positive childhood rivalries pushed me to achieve more. While I was motivated to obtain a PhD to further my career options, I also always knew one thing: I could not, and would not, be the only non-doctor in the family. When I married my husband, Scott, I was blessed to receive another champion in my life. He certainly believes in me and never lets me forget it. I am forever grateful for each of them.

I know that I would not be here without my advisor, Dr. Dulal Bhaumik. I am honored to have had the opportunity to work with him. He is extremely supportive of his students, and this does not go unnoticed. The work presented in this dissertation would not be possible without him, and for that I will always be thankful. I would also like to thank my committee members: Dr. Olusola Ajilore, Dr. Runa Bhaumik, Dr. Robert Krafty, and Dr. Supriya Mehta. The insights they provided have undoubtedly improved this work and offered new perspectives.

RAR

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABIDE	Autism Brain Imaging Data Exchange
AR(1)	Autoregressive of Order 1
AR(2)	Autoregressive of Order 2
ASD	Autism Spectrum Disorders
BOLD	Blood Oxygenation Level Dependent
CM-Step	Conditional Maximization Step
DMN	Default Mode Network
DSM	Diagnostic and Statistical Manual
E-Step	Expectation Step
EEG	Electroencephalography
EM	Expectation-Maximization
ECM	Expectation-Conditional Maximization
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonance Imaging
HRF	Hemodynamic Response Function
GEM	Generalized Expectation-Maximization
GMRF	Gaussian Markov Random Field

LIST OF ABBREVIATIONS (Continued)

M-Step	Maximization Step
MEG	Magnetoencephalography
MLEs	Maximum Likelihood Estimates
MSE	Mean Squared Error
PET	Positron Emission Tomography
ROI	Region of Interest
rsfMRI	Resting State fMRI
RMSE	Root Mean Square Error
SE	Standard Error
SVD	Singular Value Decomposition
WHO	World Health Organization

SUMMARY

In 2006, a World Health Organization report warned that neurological disorders are one of the greatest threats to global health. A detailed understanding of the effects of these conditions on the human brain network is therefore fundamental for early diagnosis and development of treatments. Among the available neuroimaging techniques used to visualize and evaluate neurological disorders, fMRI is valuable in describing how regions of the brain communicate. In particular, resting state fMRI studies are fundamental for measuring the natural communications that occur throughout the brain.

One of the main goals of resting state fMRI studies is to estimate functional connectivity. Functional connectivity is the spatial dependence in the signal of brain regions over time. A simple approach for analyzing the difference in functional connectivity between a disease group and healthy controls is to compare Pearson correlation coefficients of the fMRI time series for every pair of regions under study. However, fMRI measurements exhibit a non-negligible amount of temporal autocorrelation, violating the assumption of independence for inferences of Pearson correlation. fMRI data analysis is also complicated by noisy data, a complicated spatiotemporal structure, and large computational demands due to its high dimensional nature.

To make accurate inferences of functional connectivity, a deeper understanding of the observed data and covariance structure between and within brain regions is required. To address this, a spatiotemporal model for resting state fMRI data is introduced. The proposed model serves two main purposes. First, the model smooths the noisy data to estimate the underlying

true signal. A common challenge faced in this task is the large computational burden due to the high dimensionality of the data. We therefore explore the lower dimensional process common to all regions that dictates how fMRI signals change over time. This reduces the number of parameters required while still being able to estimate the connections between all regions under study. Second, using estimated outcomes from the spatiotemporal model, the temporal correlation that exists within each region's time series is removed without removing the spatial correlation across regions. Pairwise correlations of all brain regions are subsequently calculated using the uncorrelated time series. Fisher's z-transformation and Efron's local false discovery rate procedure are appropriately applied to make large sample inferences.

The proposed spatiotemporal model for resting state fMRI data has several notable properties. Statistical modeling of the fMRI signal must take spatial and temporal correlations into account. Hierarchical models are well suited for a dynamic process that consists of multiple levels of variation. Furthermore, to handle the large amount of data collected for each individual in a computationally feasible manner, dimension reduction is incorporated. Moran's I basis functions enable us to use a lower dimension latent factor dynamic model via incorporation of an adjacency matrix. To allow for both positive and negative correlations in spatial associations, the Bessel covariance function is used. This function is specified by a shape and range parameter. Furthermore, Bowman's functional distance measure is used to define the distance between every pair of regions for the adjacency matrix and Bessel function.

To estimate such a high dimensional data model, Bayesian methods are often the first choice. However, we illustrate that model estimation can be achieved in the traditional fre-

quentist framework. More specifically, estimation of the model parameters is performed by the Expectation-Conditional Maximization algorithm. This algorithm is an extension of the Expectation-Maximization algorithm and is useful when the maximization process becomes simpler by performing maximization with multiple steps under conditions placed on a subset of parameters of estimation. It is particularly suited for the proposed spatiotemporal model, which consists of unknown latent vectors and a subset of parameters without closed-form solutions.

Resting state fMRI data from the Autism Brain Imaging Data Exchange is used to illustrate the proposed method for analysis of functional connectivity. This initiative is a collaboration of laboratories around the world to create a large scale collection of functional and structural brain imaging data for autism spectrum disorder patients and healthy controls. Since autism spectrum disorder is a heterogenous condition, a large sample size is required. Our analysis consists of 162 autism subjects and 167 typically developing controls. Using the presented approach, 5,995 links between Harvard-Oxford brain regions are analyzed for disrupted functional connections in autism patients. We identify thirty-nine clinically relevant disrupted connections at a false discovery rate of 0.1.

CHAPTER 1

INTRODUCTION

In 2006, the World Health Organization (WHO) estimated that neurological disorders affect nearly one billion people worldwide (World Health Organization, 2006). Globally, neurological disorders are the leading cause of morbidity (as measured by disability adjusted life years) and second leading cause of mortality in adults, resulting in an estimated nine million deaths in 2016 (Neurological Disorders Collaborator Group, 2019). In the United States, neurological conditions are the leading cause of morbidity and mortality in children and "among the most serious acute pediatric illnesses" (Moreau et al., 2013). In light of these trends, WHO concluded that "there is ample evidence that pinpoints neurological disorders as one of the greatest threats to public health" (World Health Organization, 2006).

To combat the burden of neurological diseases, a detailed understanding of the differences in human brain networks between healthy and affected individuals is fundamental for early diagnosis and development of treatments. The use of sophisticated neuroimaging tools is vital toward understanding the complex brain network. Current widely used brain imaging techniques include electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). Electrical and magnetic brain activity are measured by EEG and MEG, respectively. While they provide highly accurate temporal measurements, they are unable to pinpoint the measured activity to a specific location. In contrast, PET and fMRI describe neuronal activity with high spatial reso-

lution. However, the temporal resolution of these modalities is relatively decreased (Lindquist, 2008).

Among the available neuroimaging technologies, fMRI has been particularly valuable for gaining a greater understanding of how regions of the brain communicate. After its introduction in 1990, fMRI quickly became "the tool of choice for visualizing neural activity in the human brain" (Kim and Bandettini, 2010). It is extremely versatile and has been used in studying a variety of clinical areas, including vision, motor, language, and cognition. One of the principal applications of fMRI in clinical research is describing functional connectivity. Functional connectivity is the dependence of fMRI signals between brain regions over time. It is well recognized that the study of functional connectivity via fMRI is "of high importance, providing new important insights in the core organization of the human brain" (VanDenHeuvel and Pol, 2008). The study of how brain connectivity may be altered by neurological diseases has consequently been greatly advanced with the advent of fMRI.

1.1 Functional Magnetic Resonance Imaging

The principal tool for mapping brain function is fMRI. Its dominance is attributed to its widespread availability, non-invasive nature, relatively low cost, and high spatial resolution. Although the temporal resolution of fMRI is decreased relative to PET and MEG, changes in neural activity can still be measured in the low frequency range (Leopold et al., 2003). In a typical fMRI study, a series of brain images is collected. Each image consists of uniformly spaced volumes, known as voxels, that partition the whole brain. The images are collected at

consecutive time points, which can vary between 100 and 2,000. This is repeated in multiple individuals, leading to a large number of observations to be analyzed (Lindquist, 2008).

Blood oxygenation level dependent (BOLD) contrast is used by fMRI to measure local brain activity (Ogawa et al., 1992). This contrast describes changes in deoxyhemoglobin within a space over time. The physiological mechanism measured by BOLD that signals activity is often referred to as the hemodynamic response function (HRF). An illustration of the HRF is provided in Figure 1. Oxygenated and deoxygenated hemoglobin have different magnetic properties and therefore produce different magnetic fields; while deoxyhemoglobin suppresses magnetic resonance, oxyhemoglobin does not. Increased metabolic demands due to localized neuronal activity lead to increased cerebral blood flow. More oxygen than consumed is supplied to the active space, resulting in an increase in oxyhemoglobin and thus an increase in the signal. However, right before activation, an initial dip in the BOLD signal occurs. An increase in the signal is observed one to two seconds after the onset of activity and reaches its peak at four to eight seconds. After reaching peak level, the BOLD signal decreases below baseline for roughly ten seconds. This phenomenon is referred to as the post-stimulus undershoot. Local brain activity over time is thus captured by fMRI via changes in local magnetic susceptibility (Lindquist, 2008).

1.2 Functional Connectivity

The use of fMRI is well suited for analysis of functional connectivity since it measures localized activity throughout the entire brain over time. The first type of study design using fMRI data to measure functional connectivity was task-based. In a task-based study, participants

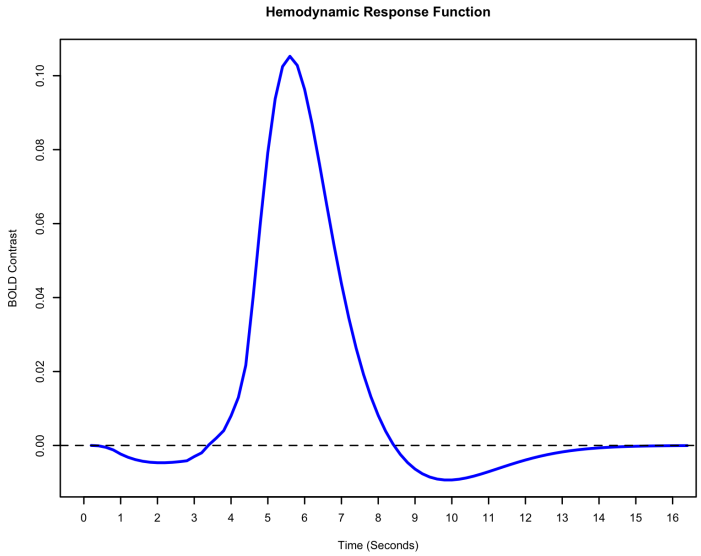


Figure 1: Hemodynamic response function.

are placed in the scanner and asked to alternate between performing an activity and rest. For example, Rao et al. (1993) had study participants perform self-paced simple and complex finger movements while scans were taken. During simple, self-paced movements, the authors identified functional changes only in the contralateral primary motor cortex. However, functional changes occurred in multiple areas during complex, self-paced movements in addition to the contralateral primary motor cortex, including the supplementary motor area, premotor cortex of both hemispheres, and contralateral somatosensory cortex (Rao et al., 1993). Activation maps that illustrate functional connections can be created after identifying areas of the brain that are simultaneously triggered by a task (Glover, 2011).

In 1995, Biswal and colleagues introduced a modified approach in study design for analyzing functional connectivity with data from fMRI. The authors had study participants remain at rest throughout the entire scan. Even though study subjects were not actively engaging in an assigned activity, they observed a temporal correlation in the BOLD signal pattern between the left and right sensorimotor cortices. This correlation is clinically meaningful as these regions are functionally related. The authors concluded that the series of images collected at rest therefore reflect the natural communication that occurs between brain regions (Biswal et al., 1995). Biswal later referred to this phenomenon as "resting state functional connectivity" (Biswal, 2012).

Following the discovery of resting state functional connectivity, numerous authors collected data from resting state fMRI (rsfMRI) to look into the validity and identification of resting state connections. Several studies uncovered resting state networks that correspond well to networks identified by task activation studies (Smith et al., 2009). For example, the dorsal and ventral attention networks have been observed in both study designs (Lee et al., 2013). Additional studies showed functional dependencies from rsfMRI signals across the hemispheres of the auditory and visual primary cortices. In 2006, Damoiseaux and colleagues confirmed the findings reported in Biswal and colleagues' original report (Damoiseaux et al., 2006). Moreover, the well established default mode network was identified using fMRI data by Greicius and colleagues (Greicius et al., 2003) and subsequently confirmed in separate research (Lee et al., 2013). Thus, across several analyses over the last twenty-five years, experimental evidence has

repeatedly pointed to the conclusion that the human brain is made up of functionally consistent networks.

1.3 Statistical Challenges and Literature Review

The introduction of rsfMRI has raised new statistical challenges and the need for new frameworks for analysis. One of the main goals of rsfMRI studies is to estimate functional connectivity by describing the dependencies that exist between brain regions over time. A simple approach that is often utilized is to compare Pearson correlation coefficients of the BOLD signals for every pair of regions under study. The major pitfall of this approach is that it incorrectly treats the signals collected over time within a region as independent. The signal within a region exhibits a non-negligible amount of temporal autocorrelation, violating the assumption of independence for using Fisher's z-transformation to make inferences on Pearson correlation. Since this assumption is not valid for fMRI data, results based on the Pearson correlation could be misleading (Caponera et al., 2018).

A deeper understanding of the observed data and covariance structure between and within brain regions is required to make accurate inferences for functional connectivity. Once the rsfMRI data and its dependencies are characterized, appropriate inferential tools for the analysis of functional connectivity can be utilized. However, unraveling the dependence structure that exists in the large number of observations from a rsfMRI study is challenging (Lindquist, 2008; Zhang et al., 2015). The measurements produced by fMRI exhibit "massive amounts of noisy data and a complicated spatiotemporal correlation structure" (Lindquist, 2008). Appro-

priate analysis of the data therefore requires implementation of a spatiotemporal model with a potentially large number of parameters and high computational burden (Lindquist, 2008).

Since task activation studies were the first to use fMRI for research, most statistical models that have been developed for fMRI data focus on identifying brain regions activated by an assigned activity. Initial strategies for modeling data from fMRI utilized a two-stage approach. Friston et al. (1995) proposed performing univariate time series analysis at each voxel in a first stage. While within-voxel temporal correlations are directly estimated in a time series model, spatial correlations are not. The authors proposed analyzing spatial correlations indirectly in a second stage. They suggested extracting voxel-specific test statistics from the time series models and applying the theory of statistical parametric maps to make spatial inferences. Under the statistical parametric map paradigm, it is assumed that each test statistic is a realization of a Gaussian random field. Analysis of the test statistics from the time series using a map is performed to determine which parts of the brain are activated following the stimulus (Friston et al., 1995).

Worsley et al. (2002) extended the first stage of Friston and colleagues' methodology by expanding it to two steps. The first step involves fitting within-subject time series of each voxel. They assumed the errors in the time series are autoregressive of order 1 (AR(1)). In the second step, the subject-specific results for each voxel are combined in a random-effects model to create a voxel-specific, group-level model. They recommend using the statistical parametric map methodology utilized by Friston and colleagues to make inferences about spatial correlations (Worsley et al., 2002).

These univariate approaches ignore spatial dependencies in the model formulation and are therefore not optimal for analysis of fMRI data. The test statistics from the voxel-specific models are not independent since spatial correlations between regions exist. Furthermore, due to the large dimensionality of the data, analysis of the vast number of test statistics leads to multiplicity issues (Zhang et al., 2015). Several authors have presented alternative frameworks for modeling fMRI data that directly incorporate spatial correlations in the statistical model. The majority of these approaches follow a Bayesian paradigm. Within the Bayesian framework for analysis, modeling temporal and spatial dependencies can be achieved by assuming a distribution for the measurement error and selecting a prior distribution for the parameters (Lee et al., 2014). A general linear model with autoregressive errors is commonly used to account for temporal correlations. Furthermore, spatial dependencies between brain voxels are often captured by imposing a Gaussian Markov random field (GMRF) prior on the model parameters. Several choices for the precision matrix of the GMRF prior have been considered (Zhang et al., 2015).

Gossl, Auer, and Fahrmeir (2001) argued that one of the main advantages of the Bayesian approach for analysis of spatiotemporal data is that spatial relationships can be easily introduced into time series models. They used voxel-specific time series models to account for temporal correlations. Spatial relationships are incorporated via the parameters of the voxel-specific models. Intrinsic Gaussian autoregressive priors are assigned to the model parameters, and first-order neighborhood information is included in the precision matrix of the spatial parameters. This combination of prior distribution and precision matrix can be interpreted as

a stochastic interpolation of adjacent neighbors. Their approach therefore differs from Friston et al. (1995) and Worsley et al. (2002) in that estimates of the voxel-specific time series are spatially smoothed (Gossl et al., 2001).

Katanoda, Matsuda, and Sugishita (2002) introduced a generalized least squares model to incorporate spatial autocorrelation. Similarly to the approach in Gossl, Auer, and Fahrmeir (2001), a key assumption of their approach is that spatially adjacent voxels are jointly stimulated by neural activity. Each voxel-specific model therefore includes the time series of the six neighboring voxels in three orthogonal directions in addition to its own. Spatial and temporal correlations are modeled assuming separability. The combined correlation structure can therefore be expressed as a product of the two. They chose to deal with their model in the frequency domain using the discrete Fourier transform at the Fourier frequencies for easier manipulation of the correlation structures. A nonparametric approach is used to estimate the spatial and temporal covariance matrices (Katanoda et al., 2002).

Gibbons et al. (2004) focused on directly modeling the BOLD response pattern to identify voxels activated by a stimulus. They utilized a hierarchical cubic polynomial model to fit the voxel-specific time series, where the four polynomial coefficients are random-effects. They assume the polynomial coefficients follow a multivariate normal distribution. Estimation is performed via an empirical Bayes approach. The strength of this approach is that the voxel-specific coefficients are estimated using the time series of all voxels. A clustering algorithm is then used to identify voxels that exhibit the HRF in response to the assigned task (Gibbons et al., 2004).

Penny, Trujillo-Barreto, and Friston (2005) used a Bayesian approach in their single-subject fMRI model. Temporal autocorrelation is taken into account by modeling the errors as an autoregressive process. Furthermore, spatial dependence is incorporated in the model via the precision matrix of the prior distribution for the generalized linear model regression coefficients. The Laplacian prior is assumed for the precision matrix, which enforces smoothness by penalizing differences between neighboring voxels. Local information of twelve neighboring voxels is incorporated by this prior. The authors note that multi-subject analysis can be done in a second stage by combining subject-specific effects into a separate model. This would require another set of spatial priors to account for between-subject differences (Penny et al., 2005).

Harrison et al. (2007) presented a generalization of Penny, Trujillo-Barreto, and Friston's (2005) model by incorporating diffusion based priors. For their generalized linear model, a diffusion kernel of a weighted graph-Laplacian is assigned as the prior for the covariance matrix of the model parameters. Their motivation was to represent the brain as an irregular graph defined by vertices, which correspond to voxels, and weighted edge sets, which define the neighbors of each voxel. Since diffusion based models require inversion of the $n \times n$ spatial covariance matrix, they proposed a three step strategy to make modeling computationally feasible: create a volume with a subset of voxels, partition the volume into segments, and fit a spatial model for each segment. In this approach, segments are not assumed to be homogeneous. A multivariate spatial model is used for every voxel in each segment (Harrison et al., 2007).

Another Bayesian spatiotemporal model for fMRI data was presented by Quiros, Diez, and Gamerman (2010). Like the Bayesian approaches described thus far, their model incorporates

an intrinsic GMRF for regression parameters to model spatial connectedness. In practice, they only considered first-order neighbors for each voxel. They parameterize the shape of the HRF to reflect the potential increase in signal with a subsequent exponential decay in the temporal dimension. The delay commonly seen in advance of the BOLD signal increasing is not fixed but modeled as an unknown parameter (Quiros et al., 2010).

Lee et al. (2014) also introduced a Bayesian hierarchical model that follows the standard Bayesian approach to analyzing fMRI data. A binary spatial Ising prior is used to incorporate spatial relationships. This prior was selected since it includes neighborhood information as well as spatial interactions between voxels. Different sets of neighbors can be specified. The authors note that commonly used neighborhood structures are based on four, eight, or twelve nearest neighbors. Spatial interaction between neighboring voxels are incorporated by prespecified weights. In their simulation study, they define the weights as the reciprocal of the pairwise Euclidean distances. The authors investigate an AR(1) process, second-order autoregressive (AR(2)) process, first-order autoregressive moving average process (ARMA(1,1)), and first-order moving average process (MA(1)) for the temporal covariance structure. They conclude that the AR(1) structure is most appropriate for modeling the temporal component as it "seems to be an effective compromise between inferential efficacy and computational efficiency" (Lee et al., 2014).

Musgrove, Hughes, and Eberly (2016) developed a spatial Bayesian variable selection procedure to identify voxels activated in response to a stimulus. A spike-and-slab mixture prior is used for the regression coefficients to reflect the prior belief that the coefficient for each brain

region is zero or non-zero. Latent indicator variables modeled by a sparse spatial generalized mixed model are included in the prior to induce spatial dependence. It is assumed that voxels are represented by the vertices of an underlying graph that reflects spatial relationships. The graph is represented by an adjacency matrix. They set the $(\mathtt{i},\mathtt{j})$ element of the matrix equal to 1 if voxels $\mathtt{i}$ and $\mathtt{j}$ are related and 0 otherwise. To account for temporal correlation, the model incorporates lagged prediction errors and a vague normal prior with zero mean for AR(2) coefficients (Musgrove et al., 2016).

The Bayesian spatiotemporal linear regression model in Zhang et al. (2016) was also designed for identification of activated brain regions following a stimulus. Zhang and colleagues assign a hierarchical Dirichlet process prior on the regression coefficients. This prior is selected since within-subject non-zero coefficients are assumed to come from a mixture model. Furthermore, the model incorporates correlation between the time series of voxels within and between subjects. This is done by creating clustering among voxels at two levels: one within a subject and another between subjects. An additional benefit of the Dirichlet process prior is that it incorporates spatial correlations between distant voxels. Spatial proximity of possible fMRI activations within a subject is taken into account via a Markov random field prior on the mixture model indicators (Zhang et al., 2016).

While the spatiotemporal models discussed thus far incorporate spatial relationships, the assumption that the BOLD signals of neighboring voxels are more similar than that of distant voxels is not valid. In fact, high correlation in the fMRI signal may exist between distant regions. Spatial correlations should therefore not be functions of physical distance or restricted to

contiguous locations. Bowman (2005) presented a two-stage hierarchical approach to address this limitation. In the first stage, voxel-specific time series are modeled under the assumption of spatial independence. Voxels are then partitioned into functionally related networks. The networks are identified from a descriptive cluster analysis. The second stage uses an autoregressive model to capture spatial dependencies within a network. A limiting assumption of this approach is that it assumes that correlations are present only between voxels within a network (Bowman, 2005). Motivated by this drawback, Bowman later proposed a single spatiotemporal model with separable temporal and spatial correlations for a pre-specified region of interest (ROI). In this work, he also introduced a functional distance metric that is not based on geometric location to describe the similarity between two regions. An exponential spatial covariance model is selected based on the empirical variogram. Furthermore, a compound symmetry structure is used to account for temporal correlations (Bowman, 2007).

The work presented in Bowman et al. (2008) and Castruccio, Ombao, and Genton (2018) also addressed limitations in the modeling approach in Bowman (2005). Bowman et al. (2008) introduced a more flexible Bayesian framework to capture correlations both within and between regions. The second stage of the two-stage hierarchical model in Bowman (2005) was further divided into between and within region components. An exchangeable correlation structure is used to capture the pairwise dependence of voxels within a region. Furthermore, the correlation between regions is modeled using an unstructured covariance matrix (Bowman et al., 2008). In Castruccio, Ombao, and Genton (2018), spatial dependence is modeled within and between anatomically defined ROIs. The second stage of their model includes both a local and regional

structure to capture the spatiotemporal dependencies across voxel-specific fMRI time series. To model within ROI relationships, a linear combination of independent Gaussian anisotropic processes is used. It is assumed that each Gaussian process has a Matern covariance structure. Temporal relationships are incorporated by assigning an AR(2) process to the time series error (Castruccio et al., 2018).

There have been few publications describing a single spatiotemporal model that incorporates all regions and does not limit spatial relationships to neighboring voxels or ROIs. Woolrich et al. (2004) presented a Bayesian approach in which spatiotemporal relationships are modeled in the noise process. The process is assumed to be a space-time vector autoregressive process, where separability is not presumed. The noise is modeled using a multivariate normal process with an $n\mathsf{T} \times n\mathsf{T}$ covariance matrix, where n is the number of regions and T the number of timepoints. Spatiotemporal relationships are parameterized via a spatial AR(1) model that is temporally fixed and a spatially varying general order temporal autoregressive model. As noted by the authors, a major obstacle to using this model in practice is the computational burden: it took six hours to analyze a single fMRI slice (Woolrich et al., 2004b). Due to the intensive process time, Woolrich, Behrens, and Smith (2004) modified this approach to only include dependencies between neighboring voxels with a conditional autoregressive model (Woolrich et al., 2004a).

Caponera et al. (2018) proposed the only known spatiotemporal model specifically for fMRI data from rsfMRI studies. They describe a Bayesian latent factor hierarchical spatiotemporal model for single-subject rsfMRI data analysis. A Gaussian process is specified for the temporal

component and latent factor models are utilized for the spatial component. The spatial latent model enables a low rank representation of a high dimensional process. The components of the latent process are assumed to be independent Gaussian, and a loading factor matrix measures dependence between regions. Temporal correlation is accounted for in the covariance function of components in the latent process. Their resulting cross-covariance function implies that the association between BOLD measurements is multiplicatively calibrated according to proximity in time (Caponera et al., 2018).

1.4 Proposed Model

Several approaches have been proposed for spatiotemporal modeling of fMRI data from task-based studies. However, these approaches are not suitable for modeling fMRI data from rsfMRI studies with the goal of analyzing functional connectivity. First, the spatiotemporal model covariance matrix often does not incorporate a proper spatial concept and temporal aspect. In the majority of past approaches, spatial dependencies are estimated by assigning priors and measuring joint activations. This does not yield an estimate of the between region covariance matrix needed to understand all spatial relationships without the presence of a stimulus. Second, a proper spatial covariance matrix for fMRI data has not been utilized. The exponential and Matern matrices have been used in models that parameterize a spatial covariance matrix (Bowman, 2007; Castruccio et al., 2018). These matrices only allow for positive correlations. However, negative correlations exist in fMRI data. In fact, Bowman noted a limitation of his exponential model is that it incorrectly treats negatively correlated voxels as uncorrelated (Bowman, 2007). These matrices are therefore inadequate for the analysis of

fMRI data. Third, most of the previous approaches incorrectly limit relationships to neighboring voxels. However, spatial dependences between physically distant voxels are expected. There is therefore a need for a computationally feasible spatiotemporal model within the traditional frequentist statistical framework that incorporates all brain regions. Finally, analysis using inference procedures for the Pearson correlation should not be performed on temporally correlated data. Fisher's z-transformation of the sample correlations is often applied to make inferences. However, the corresponding hypothesis test assumes independent data. Inferences based on correlated data could therefore be misleading.

This dissertation addresses some of the aforementioned limitations for better comparison of functional connectivity in two groups using fMRI data from rsfMRI studies. In this work, a spatiotemporal model that is suitable for modeling and denoising rsfMRI data is introduced. A natural approach for modeling the dynamic process inherent in rsfMRI is to take advantage of methodologies developed in the area of time series. More specifically, we use a dynamic spatiotemporal hierarchical model formulation (Xu and Wikle, 2007; Durbin and Koopman, 2012; Wikle, 2015). We assume separability of the spatial and temporal covariance matrix. This assumption has been employed in fMRI studies (Bowman, 2007; George and Aban, 2015; Caponera et al., 2018) and provides an interpretable result for the analysis of functional connectivity (Caponera et al., 2018). Furthermore, we use the Bessel function to model spatial relationships. Unlike the exponential and Matern covariance structures, the Bessel function has the flexibility to get both positive and negative correlation.

A key obstacle to implementing a full spatiotemporal model for fMRI studies has been the seemingly insurmountable computational cost due to the high dimensionality of the data. In fact, this is one of the main reasons researchers have resorted to Bayesian approaches for spatiotemporal applications (Xu and Wikle, 2007; Lee et al., 2014; Castruccio et al., 2018). However, a computationally efficient model that incorporates all brain regions has not been implemented. In this dissertation, a spatiotemporal model that does include all brain regions is achieved by incorporating dimension reduction via basis functions. This approach recasts the state vector to a lower dimension by including spatially referenced basis functions (Xu and Wikle, 2007; Wikle, 2015). In our proposed model, the dimensionality of the state process is reduced by projecting the process onto spectral basis functions derived from the Moran operator (Bradley et al., 2015). The operator incorporates spatial relationships since it is a function of the adjacency matrix, which represents the spatial structure of the data. Instead of simply assigning a 1 or 0 to indicate whether a relationship between regions exists, we propose a spatial semivariogram approach to estimate a continuous spatial weighting function. We also use Bowman's functional distance measure instead of geometric distance to accurately reflect the spatial relationships between brain regions.

Estimation of a dynamic spatiotemporal model is efficiently accomplished with a state-space framework via a Generalized Expectation-Maximization algorithm (Xu and Wikle, 2007; Durbin and Koopman, 2012). More specifically, an Expectation-Conditional Maximization algorithm is utilized for estimation. We use the Kalman smoother to estimate latent variables in the Expectation step. The maximization component of the algorithm consists of two conditional max-

imization steps: updating parameters with closed-form maximum likelihood estimates (MLEs) and parameters without closed-form MLEs via a single Newton-Raphson iteration. The results of this algorithm are sensitive to initial values and may possibly diverge away from the global maximum toward a local maximum. An algorithm for selecting starting values is proposed to reduce the likelihood of divergence.

To identify disrupted functional connections in a disease group relative to healthy controls, group-specific spatiotemporal models are used. A vector containing the estimated time series from the spatiotemporal model for every region in each individual is formed. The temporal autocorrelation of the observations within the vector is removed by whitening for extracting certain theoretical properties of the Pearson correlation coefficient. Fisher's z-transformation and the corresponding hypothesis test of equality of spatial correlation between the disease and control group is appropriately applied to the temporally uncorrelated time series. The large number of regions analyzed introduces multiplicity issues: for n regions, there are $n(n-1)/2$ hypothesis tests. We propose Efron's local false discovery rate as the most suitable for our data.

The rest of this dissertation is organized as follows. Detection of autism spectrum disorder is the clinical motivation of this dissertation and described in Chapter 2. In Chapter 2, we also introduce the data source used for application of our approach: the Autism Brain Imaging Data Exchange. The proposed spatiotemporal hierarchical model and details of its parameterization are discussed in Chapter 3. Chapter 4 describes the estimating procedure used for the unknown parameters and latent variables of the spatiotemporal model. The subsequent procedures for estimation and inference of functional connectivity, addressing multiplicity issues, is introduced

in Chapter 5. We detect disrupted links in autism spectrum disorder in Chapter 6. This dissertation concludes with a discussion in Chapter 7.

CHAPTER 2

MOTIVATING EXAMPLE: FUNCTIONAL CONNECTIVITY IN AUTISM SPECTRUM DISORDER

Autism spectrum disorder (ASD) is a pervasive and complex neurodevelopmental disorder. Diagnostic and Statistical Manual (DSM) 5 characterizes ASD by "persistent deficits in social communication and interaction" and "restricted, repetitive patterns of behavior, interests, or activities" (Centers for Disease Control and Prevention, 2019). These symptoms must appear in early development and cause significant impairment. Furthermore, ASD is a heterogenous condition with well recognized differences in presentation of behavioral features and social functioning across individuals. It is commonly associated with the presence of comorbidities, including social anxiety disorder, attention-deficit/hyperactivity disorder, depression, anxiety, immune system abnormalities, and gastrointestinal disorder (Masi et al., 2017).

In a 2014 report using data from eleven U.S. sites in the Autism and Developmental Disabilities Monitoring Network, the prevalence of ASD in children aged eight was estimated to be 1 in 59. Males were four times more likely than females to receive an ASD diagnosis, with approximately 1 in 37 males diagnosed compared to 1 in 151 females. Moreover, prevalence was higher in non-Hispanic white children compared to non-Hispanic black or Hispanic children (Baio et al., 2018). Although these prevalence rates are informative, they must be interpreted with caution. It is recognized that ASD is influenced by biological factors and is therefore charac-

terized as a developmental condition; however, social and cultural elements greatly influence diagnosis (Masi et al., 2017).

An ASD diagnosis is currently based on questionnaires and behavioral observations by parents and health professionals. Although research into more advanced diagnosis and treatment is ongoing, "the pace and clinical impact of the resulting findings have not kept up with the urgency to identify ways of determining the diagnosis at earlier ages, selecting optimal treatments, and predicting outcomes" (Besseling et al., 2018). This is partially attributed to the subjective nature of an ASD diagnosis. There is therefore an unmet need to identify a set of objective criteria for characterizing the condition (Uddin et al., 2017; Besseling et al., 2018). The continued study of functional connectivity using data from rsfMRI studies in ASD children and typically developing controls may yield the objective criteria necessary to fill this gap.

2.1 Functional Connectivity Studies in Autism Spectrum Disorder

The theory of functional connectivity has been enthusiastically applied to studies of ASD. As of 2013, over fifty disrupted connections in ASD patients had been described in over 200 published studies. The majority of abnormalities have been identified from task-based fMRI studies of adolescent and adult participants (Uddin et al., 2013b; Anderson et al., 2013). There is almost universal agreement across studies that alterations in brain connectivity are markers of ASD. However, the exact nature of the alterations continues to be debated (Uddin et al., 2013b; Uddin et al., 2017).

Several fMRI studies conducted in adults have concluded that functional connectivity between brain regions is decreased in ASD. This has lead to a theory of hypoconnectivity, or a

decrease in functional connectivity relative to controls, as a marker of ASD. Multiple task-based studies identified abnormal connections in the default mode network (DMN). More specifically, it has been observed that the DMN in ASD patients does not activate during "attentionally demanding tasks". These observations are clinically relevant, as they correspond to symptoms often seen in ASD patients. *Furthermore, the "virtually uniform" decrease in connectivity within the DMN is "particularly compelling in autism." The "behaviors associated with these brain regions (internal stimuli, internal narrative, self-focus) correspond to symptoms of autism in which individuals may exhibit internal reflection at the expense of awareness of the outside world"* (Anderson et al., 2013).

In contrast to the theory of hypoconnectivity, some studies conducted in children have lead to an opposing theory of hyperconnectivity in ASD. Hyperconnectivity is defined as an increase in functional connections relative to controls. DiMartino et al. (2011) analyzed rsfMRI data from twenty ASD subjects and twenty controls between the ages of seven and thirteen. They observed an increase in functional connections "between nearly all striatal subregions and heteromodal associative and limbic cortex previously implicated in the physiopathology of ASD" (DiMartino et al., 2011). In a 2013 study of twenty ASD subjects and twenty controls between the ages of seven and twelve, Uddin and colleagues identified hyperconnectivity in several major brain networks important to cognitive function. These include the default mode, salience, motor, frontotemporal, and visual networks. The salience network yielded the lowest misspecification among all networks examined when building a classifier (Uddin et al., 2013a). As these sample populations differ from previous studies, discrepancies in findings of

hypoconnectivity and hyperconnectivity in ASD might therefore be reconciled by accounting for age and pubertal stage in the selection of the population under study (Uddin et al., 2013b).

Several studies have been published utilizing data from the Autism Brain Imaging Data Exchange (ABIDE). The ABIDE I initiative includes rsfMRI data for ASD subjects and controls between the ages of six and sixty-four across seventeen sites. Data from ABIDE I is used in this dissertation and described in Section 2.2. Nielson and colleagues (2013) used data from 964 subjects to build a generalized linear model based on a weighted average of connections. Regions that yielded the highest classification accuracy include the DMN, fusiform and parahippocampal gyri, the posterior middle and superior temporal gyrus, and intraparietal sulcus (Nielsen et al., 2013). Chen et al. (2015) used ABIDE data to build a classifier with random forest. They found that inclusion of the somatosensory, default mode, visual, and subcortical regions lead to the highest accuracy (Chen et al., 2015). Furthermore, Abraham et al. (2017) observed both hypoconnectivity and hyperconnectivity in their ABIDE analysis. They saw hypoconnectivity across symmetric regions of the temporo-parietal junctions, anterior insulae, and inferior parietal lobes and between the right middle temporal gyrus and left temporo-parietal junction. In contrast, hyperconnectivity was observed between the left middle temporal gyrus and right supramarginal gyrus (Abraham et al., 2017). Finally, Bhaumik et al. (2018) saw disrupted connectivities in auditory and visual cortex regions of the default mode and salience networks (Bhaumik et al., 2018a).

Several considerations must be made as disrupted functional connectivity in ASD continues to be explored. Functional connectivity in ASD has been studied in task-based and resting

state studies. Similar functional relationships have been observed from both study designs. However, the relationship between measurements from these two types of studies is not clear, as research has not been done on the relationship between both measurements in the same subject. Comparing results between these two study designs should thus proceed with caution (Uddin et al., 2013b). Moreover, better characterization of ASD in children is needed. Research conducted among younger children may enable a more accurate characterization of the neurophysiology of ASD for purposes of diagnosis and prognosis in a clinically relevant time frame. Third, larger sample sizes are required for increased power due to the heterogeneity of ASD and relatively large amount of noise in fMRI data. Fortunately, efforts to combine data across institutions have been established. Large scale efforts where data is combined and shared among researchers will help address some of these concerns and fill the gaps in our understanding of ASD (Uddin et al., 2013b; Anderson et al., 2013).

2.2 Autism Brain Imaging Data Exchange

The ABIDE initiative is a collaboration of laboratories around the world to create a large scale collection of functional and structural brain imaging data. This grass roots effort to combine data from worldwide research centers was started in recognition of the need for analyses with larger sample sizes. The first initiative of this effort is referred to as ABIDE I. The ABIDE I aggregation consists of rsfMRI, structural, and phenotypic data for 539 ASD subjects and 573 controls from twenty datasets collected from seventeen sites (O'Connor and Devoto, 2016). The Preprocessed Connectomes Project systematically preprocesses fMRI data using multiple pipelines for various data sharing initiatives, including ABIDE. We use rsfMRI data

TABLE I ABIDE I DESCRIPTIVE STATISTICS

Site	N	ASD, N (%)	Control, N (%)	Female, N (%)	Male, N (%)	Age, Mean (SD)
1.CALTECH	38	19 (50.0)	19 (50.0)	8 (21.1)	30 (78.9)	28.2 (10.6)
2.CMU	27	14 (51.9)	13 (48.1)	6 (22.2)	21 (77.8)	26.6 (5.7)
3.KKI	55	22 (40.0)	33 (60.0)	13 (23.6)	42 (76.4)	10.1 (1.3)
4.LEUVEN 1	29	14 (48.3)	15 (51.7)	0 (0.0)	29 (100.0)	22.6 (3.6)
5.LEUVEN 2	35	15 (42.9)	20 (57.1)	8 (22.9)	27 (77.1)	14.2 (1.4)
6.LUDWIG	57	24 (42.1)	33 (57.9)	7 (12.3)	50 (87.7)	26.2 (12.1)
7.NYU	184	79 (42.9)	105 (57.1)	37 (20.1)	147 (79.9)	15.3 (6.6)
8.OHSU	28	13 (46.4)	15 (53.6)	0 (0.0)	28 (100.0)	10.8 (1.9)
9.OLIN	36	20 (55.6)	16 (44.4)	5 (13.9)	31 (86.1)	16.8 (3.5)
10.PITT	57	30 (52.6)	27 (47.4)	8 (14.0)	49 (86.0)	18.9 (6.9)
11.SBL	30	15 (50.0)	15 (50.0)	0 (0.0)	30 (100.0)	34.4 (8.6)
12.SDSU	36	14 (38.9)	22 (61.1)	7 (19.4)	29 (80.6)	14.4 (1.8)
13.STANFORD	40	20 (50.0)	20 (50.0)	8 (20.0)	32 (80.0)	10.0 (1.6)
14.TRINITY	49	24 (49.0)	25 (51.0)	0 (0.0)	49 (100.0)	17.2 (3.6)
15.UCLA 1	82	49 (59.8)	33 (40.2)	11 (13.4)	71 (86.6)	13.2 (2.3)
16.UCLA 2	27	13 (48.1)	14 (51.9)	2 (7.4)	25 (92.6)	12.5 (1.5)
17.UM 1	110	55 (50.0)	55 (50.0)	26 (23.6)	84 (76.4)	13.4 (2.9)
18.UM 2	35	13 (37.1)	22 (62.9)	2 (5.7)	33 (94.3)	16.0 (3.3)
19.USM	101	58 (57.4)	43 (42.6)	0 (0.0)	101 (100.0)	22.1 (7.7)
20.YALE	56	28 (50.0)	28 (50.0)	16 (28.6)	40 (71.4)	12.7 (2.9)
TOTAL	1,112	539 (48.5)	573 (51.5)	164 (14.7)	948 (85.3)	17.0 (8.0)

preprocessed using the Configurable Pipeline for the Analysis of Connectomes in this study. After preprocessing, mean time series are created for several sets of ROIs. This study utilizes the Harvard-Oxford atlas, which includes 110 spatial regions for analysis. A description of the Harvard-Oxford regions is provided in Appendix A (Craddock et al., 2013).

Prior to data aggregation, participating sites agreed to a list of phenotypic variables to provide to the consortium. These variables were identified by overlapping measurements taken across sites and include age at scan, full IQ (FIQ) scores, sex, and diagnostic information (Di-Martino, 2014). Descriptive statistics by dataset are provided in Table I. Males make up the vast majority of ABIDE I, accounting for 85.3% of the sample. Each individual dataset is majority male, and five of the twenty datasets consist of only males. The data consists of subjects scanned at a wide range of ages, with a maximum age of sixty-four years and a minimum age around six years. A closer look at the between site differences in the distribution of age at scan can be seen in the boxplot in Figure 2. The median age at scan varies widely across studies. However, nearly 70% of scanned individuals are less than eighteen years old. A boxplot of FIQ scores by dataset can be seen in Figure 3. Although most sites provided FIQ scores, one of the two datasets from Leuven does not contain FIQ, and 83% of observations from SBL are missing FIQ. For the remaining sites, within site FIQ scores vary widely. Between site variation in FIQ is also evident.

2.3 Conclusion

The analysis of functional connectivity using fMRI data has been enthusiastically applied to the study of ASD. Disruptions in the DMN has been consistently identified as a marker of ASD

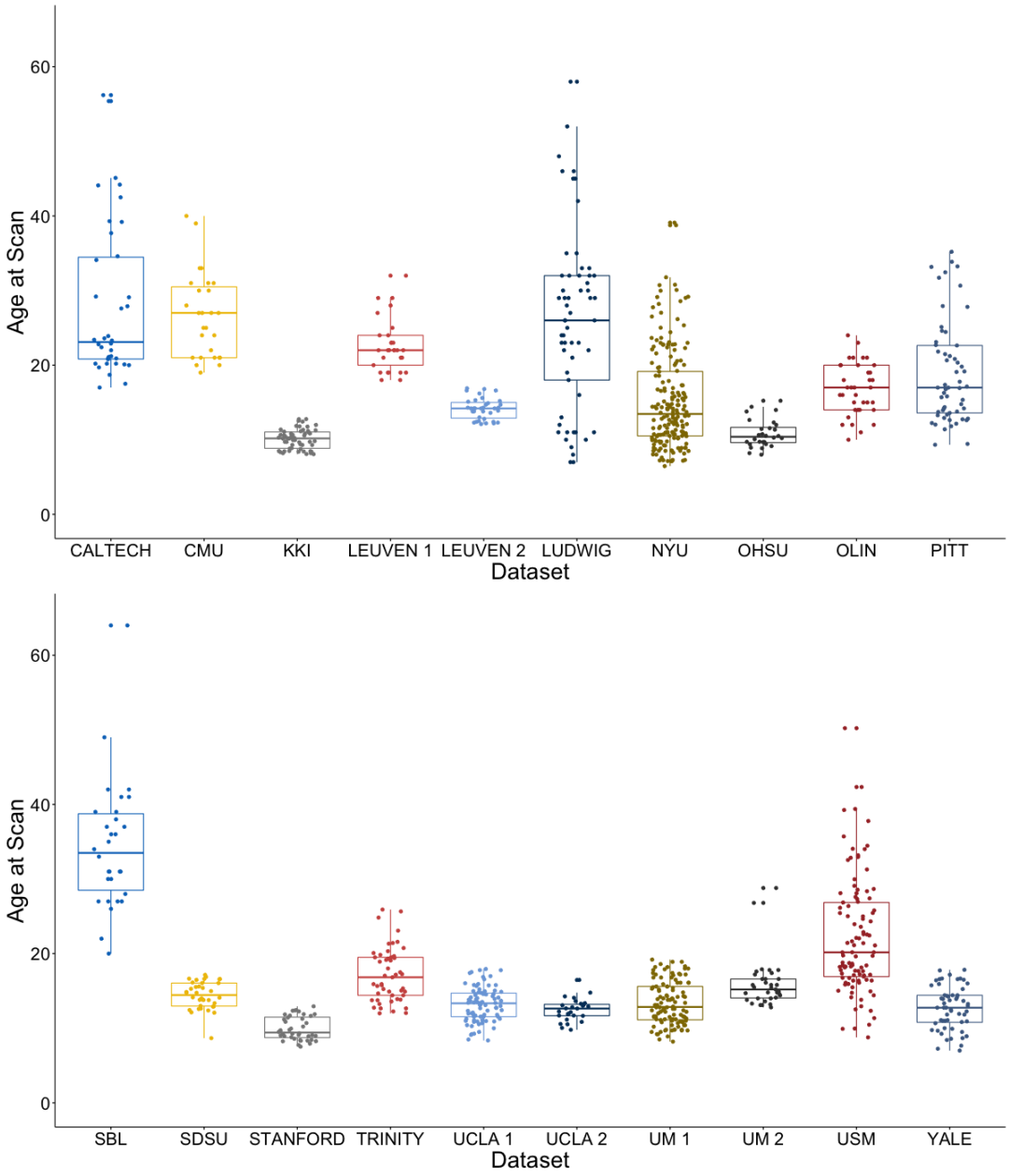


Figure 2: Boxplots of age at scan by dataset for ABIDE.

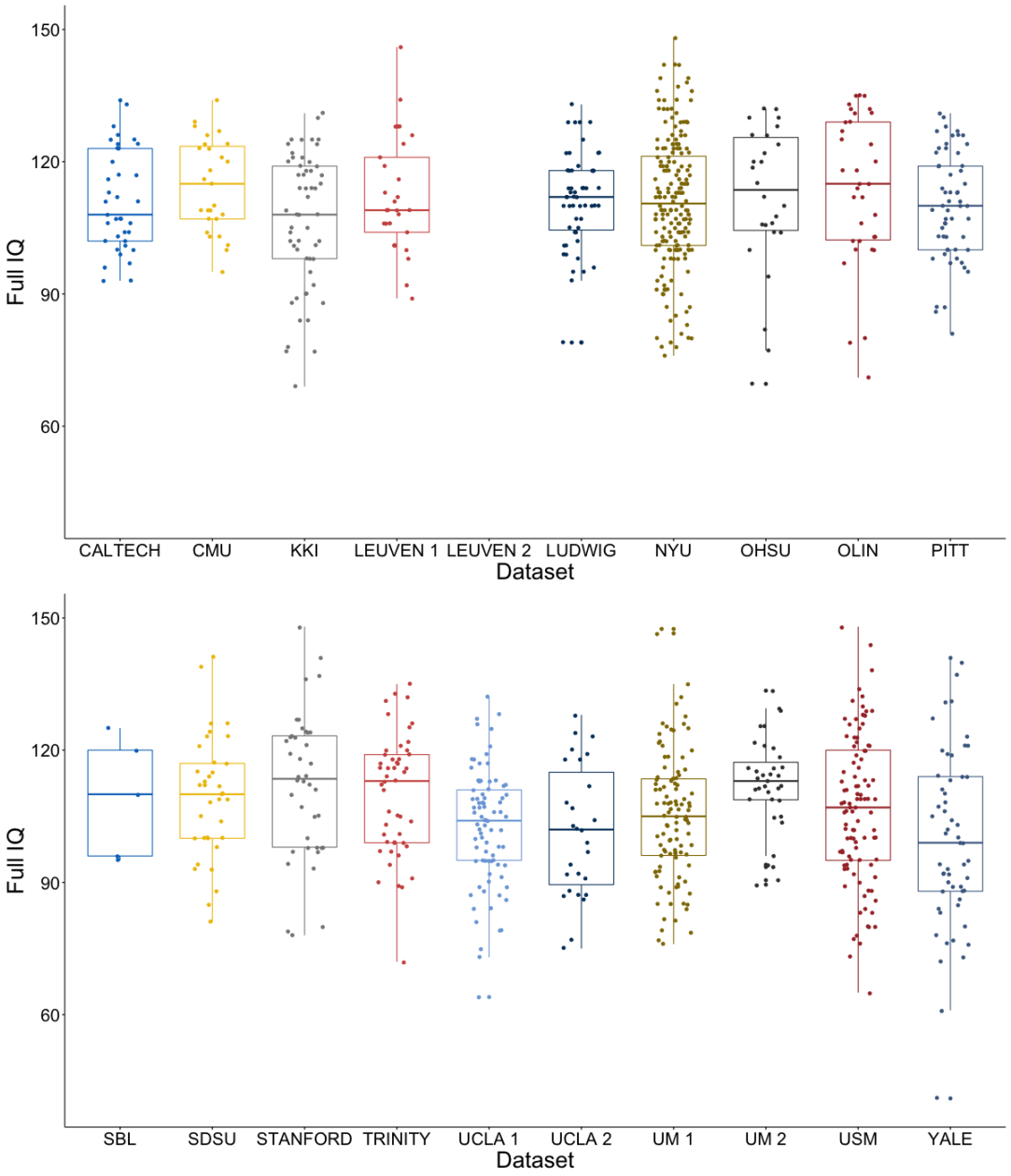


Figure 3: Boxplots of full IQ at scan by dataset for ABIDE.

across studies. As research in the area moves forward, more work needs to be done using larger samples sizes and data from children only. This dissertation applies the proposed modeling approach for analysis of functional connectivity described in Chapters 3 through 5 under this guidance. We use data from children between seven and fourteen years old in ABIDE I. Details of the analysis are provided in Chapter 6.

CHAPTER 3

SPATIOTEMPORAL MODEL

There are several considerations that must be made in specifying a spatiotemporal model for rsfMRI data. In an fMRI time series, the current value of the process evolves from past values. This evolutionary process is represented by a dynamic spatiotemporal hierarchical model that incorporates data from all regions to simultaneously describe how all spatial processes change over time (Cressie and Wikle, 2011). Furthermore, statistical modeling of the rsfMRI signal must take spatial and temporal correlations into account. Data from fMRI is collected across multiple regions within the brain, and spatial correlations due to functional connectivity exist between regions. In addition, multiple measurements from repeated scanning of a region over time exhibit temporal correlations. Hierarchical models are well suited to describe this dynamic process that consists of multiple levels of variation.

To further define the model, several specifications that account for characteristics of rsfMRI data must be made. In a study with data collected at T time points for $\mathfrak{n}$ regions in $\mathfrak{m}$ individuals, there are $\mathfrak{m} \times \mathfrak{n} \times \mathsf{T}$ signals to be analyzed. Dimension reduction is therefore incorporated to handle such a large amount of data in a computationally feasible manner. In fact, "it is often the case that the essential controlling dynamics for spatio-temporal processes reside on a relatively low-dimensional manifold" (Wikle, 2015). Incorporation of Moran's I basis functions enables us to use a lower dimension latent factor dynamic model. This class of basis functions measures spatial relatedness through the incorporation of an adjacency matrix. An

additional specification required is with regard to the spatial covariance matrix. To allow for both positive and negative correlations, the Bessel function is used. This function is specified by a shape and range parameter. As noted in Chapter 1, incorporating distance measures into the model based on geometric location is inappropriate for fMRI data. Thus, Bowman's functional distance measure is used to define the relatedness of each pair of regions in the adjacency matrix and spatial covariance function.

3.1 Hierarchical Model

The time series collected from multiple regions in an fMRI scan is described by a spatiotemporal dynamic process. Let $Z_i(s_j, t)$ denote the observed fMRI signal at time t in region j for individual i. Then, $\mathbf{Z}_{it} = \{Z_i(s_1, t), \ldots, Z_i(s_n, t)\}'$ denotes the n-dimensional vector containing individual i's fMRI measurements for all locations $s_1, \ldots, s_n$ at time t. Moreover, let $\mathbf{U}_{it}$ denote an unobserved $n \times 1$ vector representing the true process generating the observed data $\mathbf{Z}_{it}$. Let $\mathbf{Y}_{it} = \{Y_i(1, t), \ldots, Y_i(r, t)\}'$ denote an $r \times 1$ vector for the dynamic latent process at time t, where $r << n$. Furthermore, let $\mathbf{D}$ be an $n \times n$ matrix of pairwise distances for all regions and $\boldsymbol{\rho}(\mathbf{D})$ a stationary correlation matrix. The three-stage hierarchical model for individual $i, i = 1, \ldots, m$, at time t, $t = 1, \ldots, T$, is

$$\mathbf{Z}_{it} = \mathbf{U}_{it} + \boldsymbol{\varepsilon}_{it}, \boldsymbol{\varepsilon}_{it} \sim N\left(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n\right), \tag{3.1}$$

$$\mathbf{U}_{it} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{K}_i\mathbf{Y}_{it} + \boldsymbol{\omega}_{it}, \boldsymbol{\omega}_{it} \sim N\left(\mathbf{0}, \sigma_\omega^2 \boldsymbol{\rho}(\mathbf{D})\right), \tag{3.2}$$

$$\mathbf{Y}_{it} = \mathbf{G}_t\mathbf{Y}_{i(t-1)} + \boldsymbol{\eta}_{it}, \boldsymbol{\eta}_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \mathbf{Y}_{i0} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \tag{3.3}$$

This model can be divided into two broad components: a data model and process model. In the spatiotemporal literature, it is often assumed that the trend is stochastic and further defined by levels of the hierarchy. This is in contrast to other areas, where the trend is denoted by a deterministic function (Fasso and Cameletti, 2009). In this model formulation, Equation 3.1 is the data model. The data model describes the distribution of the observed data conditional on the true biological process, making the observations $\mathbf{Z}_{it}$ conditionally independent (Wikle, 2015). Then, $\boldsymbol{\varepsilon}_{it}$ is an $n \times 1$ vector of pure measurement error in the observed values $\mathbf{Z}_{it}$.

Equation 3.2 is the first level of the process model. The generating spatiotemporal process $\mathbf{U}_{it}$ is a smoothed version of the spatiotemporal observations $\mathbf{Z}_{it}$. $\mathbf{U}_{it}$ is a function of covariates with fixed-effects, a latent process $\mathbf{Y}_{it}$, and random error $\boldsymbol{\omega}_{it}$. $\mathbf{X}_i$ is the $n \times p$ matrix of p covariates for the n locations, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed-effects. $\mathbf{K}_i$ is a known $n \times r$ matrix that maps the observed data onto the reduced dimension latent process $\mathbf{Y}_{it}$. Moreover, $\sigma^2_\omega \boldsymbol{\rho}(\mathbf{D})$ is a time-constant spatial covariance matrix for the random error $\boldsymbol{\omega}_{it}$ (Xu and Wikle, 2007; Fasso and Cameletti, 2009).

Equation 3.3 is the second level of the process model. It is a first-order autoregressive model that describes the dynamic latent process. A first-order model was identified from the results of the temporal autocorrelation function from all regions in the ABIDE data used for functional connectivity analysis in ASD. Illustrations of the autocorrelation function for nine randomly selected regions are provided in Chapter 6. However, this equation is easily generalizable to an autoregressive model of higher order. $\mathbf{G}_t$ is the $r \times r$ transition or propagator matrix, defining the relationship between $\mathbf{Y}_{it}$ and $\mathbf{Y}_{i(t-1)}$. The dynamic process is common to all spatial locations

and analogous to factor analysis in multivariate statistics (Wikle, 2015). Since $r << n$, the number of parameters to be estimated is greatly reduced. $\boldsymbol{\eta}_{it}$ is the innovation error, or the difference between the observed value at time t and the optimal forecast (Fasso and Cameletti, 2009; Bradley et al., 2015). $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are nuisance parameters that initiate the latent process. The errors $\boldsymbol{\varepsilon}_{it}$, $\boldsymbol{\omega}_{it}$, and $\boldsymbol{\eta}_{it}$ are uncorrelated in time and mutually independent.

Estimation of a dynamic spatiotemporal model is best accomplished using a state-space framework (Xu and Wikle, 2007; Durbin and Koopman, 2012). Plugging Equation 3.2 into Equation 3.1 yields the two-stage hierarchical model given by

$$\mathbf{Z}_{it} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{K}_i\mathbf{Y}_{it} + \boldsymbol{\xi}_{it}, \boldsymbol{\xi}_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\xi), \tag{3.4}$$

$$\mathbf{Y}_{it} = \mathbf{G}_t\mathbf{Y}_{i(t-1)} + \boldsymbol{\eta}_{it}, \boldsymbol{\eta}_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \mathbf{Y}_{i0} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \tag{3.5}$$

In this formulation, Equation 3.4 is the observation equation and Equation 3.5 is the state equation. The rationale for the state-space model is that the dynamic process is determined by $\mathbf{Y}_{it}$ according to the state equation. However, $\mathbf{Y}_{it}$ is unobservable and analysis is based on observations $\mathbf{Z}_{it}$ (Durbin and Koopman, 2012). In the first stage of this two-stage model, the error $\boldsymbol{\xi}_{it} = \boldsymbol{\varepsilon}_{it} + \boldsymbol{\omega}_{it}$ follows a Gaussian distribution with variance-covariance matrix $\boldsymbol{\Sigma}_\xi = \sigma^2_\varepsilon\mathbf{I}_n + \sigma^2_\omega\boldsymbol{\rho}(\mathbf{D})$. This matrix can be written as $\boldsymbol{\Sigma}_\xi = \sigma^2_\omega\boldsymbol{\Gamma}(\mathbf{D})$, where

$$\boldsymbol{\Gamma}(\mathbf{D}) = \begin{cases} (1+\gamma)\mathbf{I}_n, & \text{for diagonal elements} \\ \boldsymbol{\rho}(\mathbf{D}), & \text{for off-diagonal elements} \end{cases}$$

and $\gamma = \sigma^2_\varepsilon / \sigma^2_\omega$. Scaling the covariance function in this way is preferred for positive-definiteness reasons (Xu and Wikle, 2007; Fasso and Cameletti, 2009).

3.2 Additional Model Specification

In the proposed model, it is assumed that $\mathbf{K}_i$ is a known matrix and must therefore be specified. Furthermore, functional forms for covariance matrix $\mathbf{\Sigma}_\eta$ and spatial correlation function $\boldsymbol{\rho}(\mathbf{D})$ must be selected.

3.2.1 Specification of $\mathbf{K}_i$

$\mathbf{K}_i$ is a known matrix of basis functions that account for the weights of the r-dimensional vector $\mathbf{Y}_{it}$ for each spatial location s_j (Bradley et al., 2015). The matrix is often derived in practice from a principle component decomposition (Fasso and Cameletti, 2009). It is possible to select the basis function coefficients so that they "exist in the 'spectral' space associated with the basis functions" (Wikle, 2015). Moran's I basis functions maintain this property and are therefore used to specify $\mathbf{K}_i$. These basis functions are derived from the Moran operator for $\mathbf{X}$. A key component of the Moran operator is the adjacency matrix. The utilization of the adjacency matrix permits a natural and dramatic dimension reduction. This speeds computation to make estimation feasible for the analysis of large datasets (Bradley et al., 2015).

3.2.1.1 Adjacency Matrix

The adjacency matrix, commonly referred to as a spatial weighting matrix, is a representation of the spatial structure of the data. Let $\mathbf{A}$ denote the $n \times n$ symmetric adjacency matrix for n regions. For $j = 1, 2, \ldots, n$ and $j' = 1, 2, \ldots, n$, $\mathbf{A}$ consists of elements $a_{jj'}$, $a_{jj'} \geq 0$, that

define the relationship between regions j and j'. The weights in the adjacency matrix can be based on contiguity or distance (Bailey and Gatrell, 1995).

Several options are available to construct spatial weighting matrices. Four common functions for constructing an adjacency matrix are threshold distance weights, power distance weights, double-power distance weights, and exponential distance weights. Illustrations of these functions are provided in Figure 4. Figure 4(a) presents an example of threshold distance weights. A value d is selected such that if the distance between two spaces is greater than d, then no relationship exists. All distances less than or equal to d are assigned equal weight. An illustration of power distance weights is provided in Figure 4(b). Power distance functions place relatively large weights on shorter distances and decay rapidly. The exponential distance weights shown in Figure 4(c) have a more gradual decline and converge to 0 as distance increases. The double-power distance weights in Figure 4(d) exhibit a bell-shaped taper (Bailey and Gatrell, 1995).

To guide selection of the adjacency matrix in practice, we propose an approach based on the empirical spatial semivariogram. The semivariogram is a function that describes the degree of spatial dependence of a stochastic process. For rsfMRI data, the classical estimator of the spatial semivariogram is

$$\hat{\zeta}(d) = \frac{1}{2N(d)} \sum_{j=1}^{N(d)} \sum_{i=1}^{m} \sum_{t=1}^{T} \{z_{it}(s_j) - z_{it}(s_j + d)\}^2 . \tag{3.6}$$

For individual i at time t, $z_{it}(s_j)$ is the observed value at location s_j and $z_{it}(s_j + d)$ is the observed value at a location separated by distance d from s_j. $N(d)$ is the number of pairs

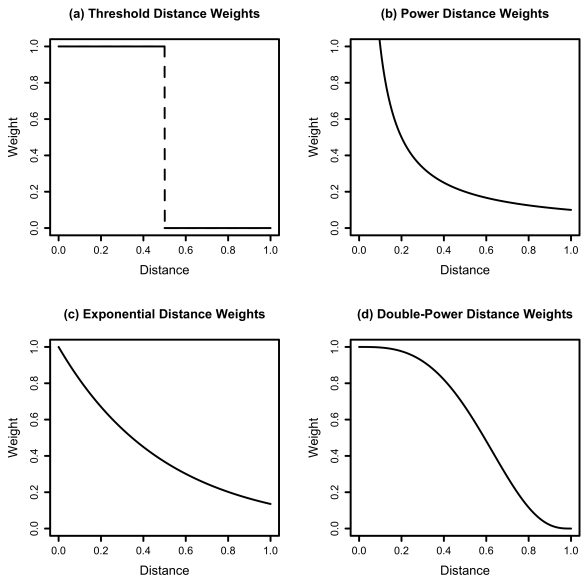


Figure 4: Spatial weighting function examples.

separated by distance $\mathbf{d}$. The lower the value of $\hat{\zeta}(\mathbf{d})$, the greater the dependence. The empirical estimates are scaled by the maximum, yielding values of dependence between 0 and 1. Complete spatial dependence is represented by 0 in $\hat{\zeta}(\mathbf{d})$ and 1 in the adjacency matrix. Therefore, $1 - \hat{\zeta}(\mathbf{d}) \left[\max\left\{\hat{\zeta}(\mathbf{d})\right\}\right]^{-1}$ is used as the spatial weight for pairs of regions separated by distance $\mathbf{d}$ to form the adjacency matrix.

3.2.1.2 Measuring Distance

A measure of distance must be selected for defining the adjacency matrix. Traditionally, measures of physical distance, such as Euclidean distance, are used to describe the relationship between two locations. These metrics assume that the correlation between two regions decreases as distance increases. However, this assumption is not suitable for describing functional

relationships; regions that are further apart may exhibit high correlations, whereas regions that are physically closer may have smaller correlations. A different measure of proximity for fMRI data must therefore be utilized (Bowman, 2007; Caponera et al., 2018).

Bowman (2007) presented an alternative measure of distance to describe the relatedness of functional data between regions within the brain. The proposed metric defines the similarity between two regions by the length of the difference between the mean fMRI measurements. For regions j and j', functional distance is defined as

$$d_{jj'} = \left[\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'} \right)' \mathbf{M} \left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'} \right) \right]^{1/2}, \tag{3.7}$$

where $\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'}$ is the T-dimensional vector of the difference between two mean activity profiles and $\mathbf{M}$ is a positive definite matrix. Possible choices for $\mathbf{M}$ include the identity matrix or covariance matrix of $\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'}$ (Bowman, 2007).

Figure 5 is presented in Bowman (2007) and illustrates the need for an alternative measure of distance when quantifying spatial relationships using neuroimaging data. The red area in Figure 5(a) highlights a region of the cerebellum that plays a role in motor function. Figure 5(b) displays Pearson correlations of least squares residuals between the location identified by the intersection of the cross-hairs and all other locations within the region. It is evident that the sample correlations do not decrease with increasing physical distance. In fact, locations that are highly correlated with the seed location are both physically close and far. In Figure 5(c), geometric distance is used to model the relationships between locations. These values do not accurately describe the empirical relationships seen in Figure 5(b). The use of scaled functional

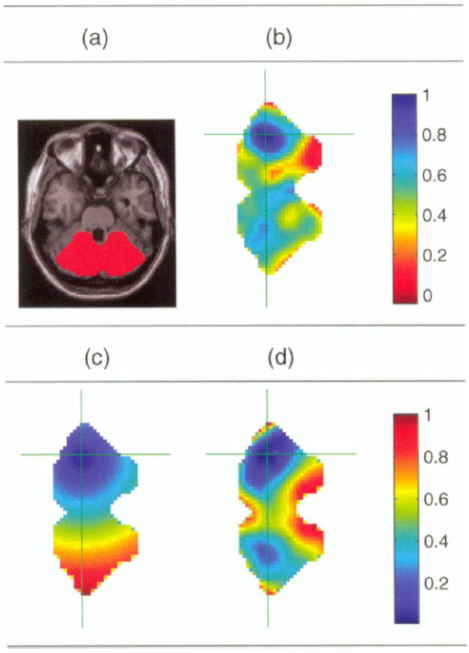


Figure 5: Functional brain activity and distance measures (Bowman, 2007).

distance is illustrated in Figure 5(d). In contrast, these measurements resemble the empirical relationships in Figure 5(b). Regions with a higher correlation with the seed region in Figure 5(b) have a smaller value of functional distance Figure 5(d) (Bowman, 2007). Thus, to accurately measure functional relationships, the adjacency matrix in the proposed spatiotemporal model utilizes functional distance as opposed to physical distance.

3.2.1.3 Moran Operator for X

In 1950, Moran introduced a non-parametric measure of spatial dependence (Moran, 1950). Let $\mathbf{I}_n$ denote the $n \times n$ identity matrix, $\mathbf{z} = (z_1, z_2, \ldots, z_n)'$ the outcomes for n spatial locations, and $\mathbf{A}$ the adjacency matrix. Moran's I statistic is

$$\mathrm{I}(\mathbf{A}) = \frac{\mathfrak{n}}{\mathbf{1}'\mathbf{A}\mathbf{1}} \frac{\mathbf{z}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathfrak{n})\,\mathbf{A}\,(\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathfrak{n})\,\mathbf{z}}{\mathbf{z}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathfrak{n})\,\mathbf{z}}. \tag{3.8}$$

$\mathrm{I}(\mathbf{A})$ is a measure of spatial autocorrelation, taking on values between -1 and 1. The numerator is the squared length of the orthogonal projection $(\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathfrak{n})$ "in the elliptical space" of adjacency matrix $\mathbf{A}$, while the denominator is the squared length of the orthogonal projection "in a spherical space" (Moran, 1950; Hughes and Haran, 2013).

The generalized form of Moran's I incorporates covariate matrix $\mathbf{X}$. Let $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, the projection onto the column space of $\mathbf{X}$. Then, the projection matrix orthogonal to $\mathbf{P}$ is $\mathbf{I} - \mathbf{P}$. The generalized Moran's I statistic is defined as

$$\mathrm{I}_{\mathbf{X}}(\mathbf{A}) = \frac{\mathfrak{n}}{\mathbf{1}'\mathbf{A}\mathbf{1}} \frac{\mathbf{z}'(\mathbf{I} - \mathbf{P})\,\mathbf{A}\,(\mathbf{I} - \mathbf{P})\,\mathbf{z}}{\mathbf{z}'(\mathbf{I} - \mathbf{P})\,\mathbf{z}}. \tag{3.9}$$

$(\mathbf{I} - \mathbf{P})\,\mathbf{A}\,(\mathbf{I} - \mathbf{P})$ in the numerator of the generalized form is referred to as the Moran operator for $\mathbf{X}$. It is interpreted as the squared length of the orthogonal projection in the elliptical space of the adjacency matrix $\mathbf{A}$ (Hughes and Haran, 2013).

3.2.1.4 Moran Basis Functions

For the proposed model, $\mathfrak{r}$ basis functions are collected from a subset of $\mathfrak{n}$ eigenvectors of the Moran operator. Incorporation of these basis functions reduces the residual spatial autocorrelation in the error. For individual $\mathfrak{i}$, let

$$\begin{aligned} \mathbf{M}(\mathbf{X}_{\mathfrak{i}}, \mathbf{A}) &= \left(\mathbf{I}_{\mathfrak{n}} - \mathbf{X}_{\mathfrak{i}}\left(\mathbf{X}_{\mathfrak{i}}'\mathbf{X}_{\mathfrak{i}}\right)^{-1*}\mathbf{X}_{\mathfrak{i}}'\right)\mathbf{A}\left(\mathbf{I}_{\mathfrak{n}} - \mathbf{X}_{\mathfrak{i}}\left(\mathbf{X}_{\mathfrak{i}}'\mathbf{X}_{\mathfrak{i}}\right)^{-1*}\mathbf{X}_{\mathfrak{i}}'\right) \\ &= \mathbf{E}_{\mathbf{X}_{\mathfrak{i}}}\mathbf{A}\mathbf{E}_{\mathbf{X}_{\mathfrak{i}}}, \end{aligned} \tag{3.10}$$

where $(\mathbf{X}_i'\mathbf{X}_i)^{-1*}$ is the inverse of the nearest positive definite matrix or the generalized inverse, as covariates are collected at the subject level and therefore $(\mathbf{X}_i'\mathbf{X}_i)^{-1}$ is not full rank. Both options yield similar matrices that are practically orthogonal to the column space of $\mathbf{X}_i$. The spectral decomposition of $\mathbf{M}(\mathbf{X}_i, \mathbf{A})$ is denoted

$$\mathbf{M}(\mathbf{X}_i, \mathbf{A}) = \mathbf{\Phi}_i \mathbf{\Lambda}_i \mathbf{\Phi}_i'. \tag{3.11}$$

$\mathbf{K}_i$ is derived from the $n \times r$ matrix formed from the first r columns of $\mathbf{\Phi}_i$, where $r << n$. This enables parsimonious fitting of the distribution of $\mathbf{Y}_{it}$, which can be computationally expensive for larger values of r (Bradley et al., 2015).

The primary rationale for using Moran's I eigenvector approach is in its interpretation. The eigenvectors that are taken from a transformed adjacency matrix exhibit distinct spatial patterns. Incorporating a linear combination of a subset of these eigenvectors can therefore capture hidden spatial patterns, making the eigenvectors a proxy for the underlying process common to all regions (Tiefelsdorf and Griffith, 2007). In fact, Boots and Tiefelsdorf (2000) illustrate that the eigenvectors of the operator "comprise all possible mutually distinct patterns of clustering residual to $\mathbf{X}$" while accounting for the underlying spatial relationships (Boots and Tiefelsdorf, 2000; Hughes and Haran, 2013). Furthermore, patterns explained by the eigenvectors are filtered from the error. Thus, by choosing orthogonal patterns and adding them to the spatiotemporal model, spatial dependence present in the residuals after accounting for fixed-effects is incorporated into the model. A key issue to address in practice is how many

eigenvectors to include (Tiefelsdorf and Griffith, 2007). We propose looking at a combination of the variance explained and likelihood-ratio tests for selection of r.

3.2.2 Specification of $\boldsymbol{\Sigma_\eta}$

Selection of $\boldsymbol{\Sigma_\eta}$ becomes simple with this choice of $\mathbf{K}_i$. Since the columns of $\mathbf{K}_i$ are orthogonal, the elements of $\mathbf{Y}_{it}$ are "approximately a posteriori uncorrelated" (Musgrove et al., 2019). It can therefore be assumed that $\boldsymbol{\Sigma_\eta}$ is a diagonal matrix. Under the assumption of equal variance, $\boldsymbol{\Sigma_\eta}$ is given as

$$\boldsymbol{\Sigma_\eta} = \sigma_\eta^2 \mathbf{I}_r, \tag{3.12}$$

where $\mathbf{I}_r$ is the r-dimensional identity matrix. Thus, only one parameter needs to be estimated.

3.2.3 Specification of $\boldsymbol{\rho}(\mathbf{D})$

Any model of fMRI data must account for the spatial correlations that exist between regions. A parametric covariance function that maps the relationship between distance and dependence of each pair of regions is needed. Commonly used covariance structures in spatial statistics include the exponential, spherical, and Matern matrices. However, these functions only allow positive correlations, and negative functional connectivity has been reported from rsfMRI data (Chen et al., 2011). The use of these structures would treat negatively correlated regions as uncorrelated, making them unsuitable for the proposed model.

To allow for greater flexibility in the covariance matrix, the Bessel function is used to define $\boldsymbol{\rho}(\mathbf{D})$. For regions j and j', the Bessel function is given by

$$B_\nu\left(\frac{d_{jj'}}{s}\right) = 2^\nu\Gamma(\nu+1)\left(\frac{d_{jj'}}{s}\right)^{-\nu} J_\nu\left(\frac{d_{jj'}}{s}\right), \tag{3.13}$$

where $d_{jj'} \in \mathbb{R}$ is functional distance between regions j and j', $\nu \geq -1/2$ is a shape parameter, $s > 0$ is a range parameter, and Γ denotes the gamma function. $J_\nu\left(\frac{d_{jj'}}{s}\right)$ is the Bessel function of the first kind, given as

$$J_\nu\left(\frac{d_{jj'}}{s}\right) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(k+\nu+1)}\left(\frac{d_{jj'}}{2s}\right)^{2k+\nu}. \tag{3.14}$$

The Bessel function is a rotation invariant function since its value does not change with arbitrary rotations. The corresponding random field is therefore weakly stationary. Note that $B_\nu\left(\frac{d_{jj'}}{s}\right)$ is a positive definite function on $\mathbb{R}$ for any $\nu \geq -1/2$. Hence, the Bessel matrix is positive semi-definite (Schlather, 2012)

Illustrations of the Bessel function are provided in Figure 6. The top graph provides examples of the function for a fixed $s = 0.04$ and changing ν. As ν decreases, the decline in correlation is more rapid and negative correlations are more likely. The bottom graph shows various functions for a fixed $\nu = 0.6$ and changing s. Decreasing s also yields a more rapid decline in correlation. However, the minimum correlation is the same across all values of s.

3.3 Conclusion

In this chapter, a hierarchical spatiotemporal model for rsfMRI data incorporating all brain regions is presented. We assume that the observed data are a function of a true process and measurement error. The true process incorporates the effects of covariates and a lower dimension first-order dynamic latent process. The proposed model requires specification of the matrix $\mathbf{K}_i$,

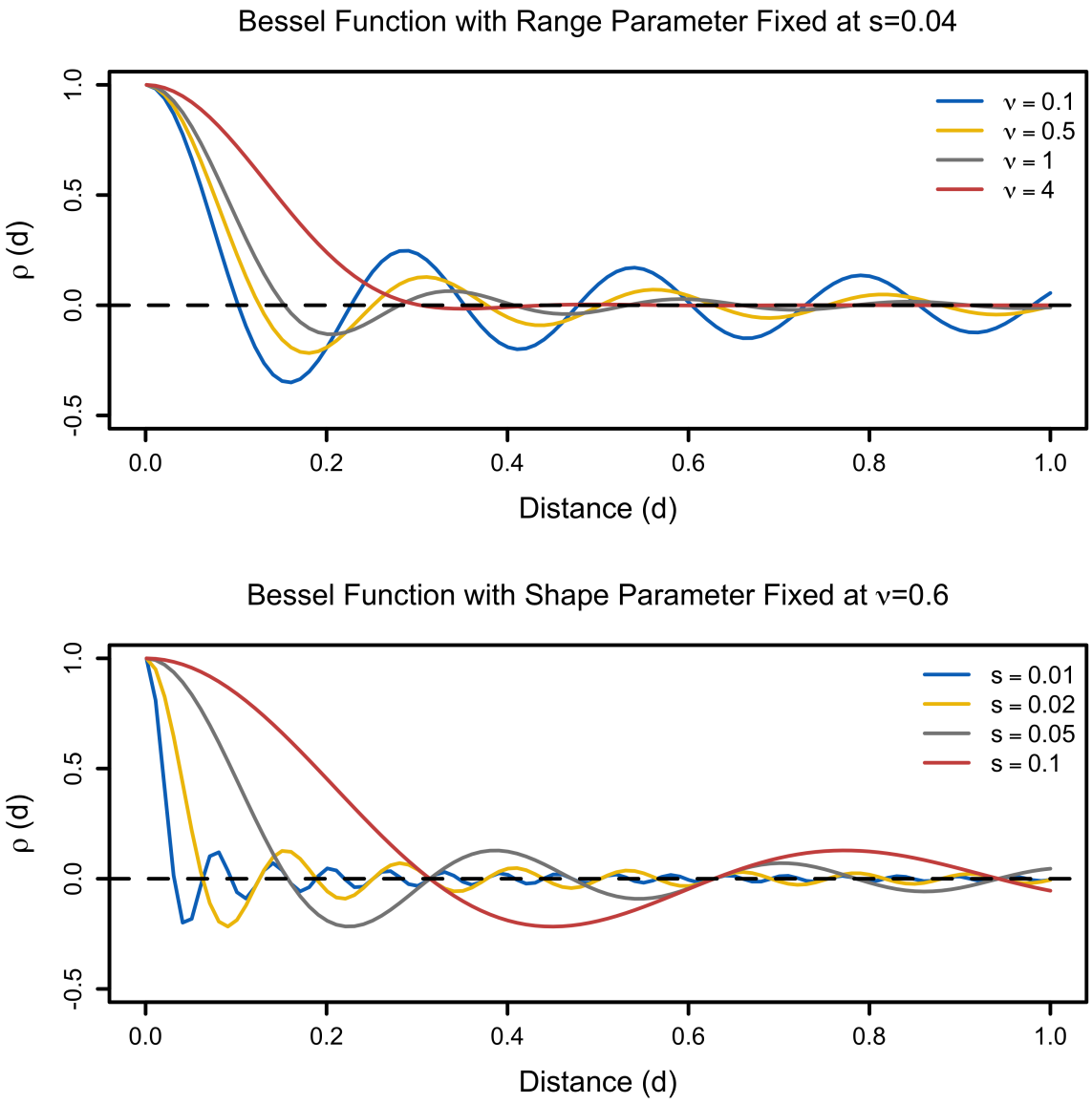


Figure 6: Bessel function examples.

variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$, and spatial correlation matrix $\boldsymbol{\rho}(\mathbf{D})$. Moran's I basis functions are used to derive $\mathbf{K}_{\mathrm{i}}$, leading to a natural dimension reduction technique while maintaining spatial interpretation. The orthogonal nature of the basis functions leads to a simple specification of $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ as a diagonal matrix. To allow for both negative and positive spatial covariances, the Bessel function is used. With a fully specified model, estimation of the unknown parameter vector is subsequently performed.

CHAPTER 4

SPATIOTEMPORAL MODEL ESTIMATION

The unknown parameter vector to be estimated for the proposed spatiotemporal model is $\boldsymbol{\Psi} = \left(\boldsymbol{\mu}_0, \mathbf{G}_t, \sigma^2_\eta, \sigma^2_\omega, \boldsymbol{\beta}, \gamma, \nu, s\right)$, where $\gamma = \sigma^2_\epsilon / \sigma^2_\omega$. Since $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are nuisance parameters and cannot be estimated simultaneously, the lower dimensional vector $\boldsymbol{\mu}_0$ is updated (Xu and Wikle, 2007). We use the Expectation-Conditional Maximization (ECM) algorithm for estimation. This algorithm was proposed by Meng and Rubin in 1993 and is an extension of the popular Expectation-Maximization (EM) algorithm (Meng and Rubin, 1993). In the ECM algorithm, the single maximization step of the EM algorithm is replaced with multiple conditional maximization steps to achieve a computationally simpler process (Meng and Rubin, 1993; McLachlan and Krishnan, 1997). This approach is particularly suited for the proposed spatiotemporal model, which consists of unknown latent vectors and a subset of parameters without closed-form solutions. A procedure for selecting initial values to increase the likelihood of convergence to the global maximum is introduced. Our simulation study illustrates that the proposed estimation procedure yields unbiased estimates with large sample properties. We also prove that the conditions required for convergence in the likelihood are satisfied.

4.1 Expectation-Maximization Algorithm

The EM algorithm is an iterative procedure proposed by Dempster, Laird, and Rubin in 1977 (Dempster et al., 1977). It is designed for computing maximum likelihood estimates

(MLEs) that would be easily obtained if the data were complete but cannot be due to missing data. The definition of missing data is broad; it includes unobserved values and variables that can never be observed, such as latent variables. The incomplete data likelihood is maximized indirectly through iterative maximization of the expected value of the complete data log likelihood function. Each iteration includes an expectation step (E-step) and maximization step (M-step). Let $\mathbf{z}$ denote the observed data, $\mathbf{y}$ the missing data, and $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_d)$ the vector of unknown parameters in parameter space $\boldsymbol{\Omega}$. Moreover, let $L(\boldsymbol{\Psi}|\mathbf{z})$ denote the incomplete data likelihood and $L_c(\boldsymbol{\Psi}|\mathbf{z}, \mathbf{y})$ the complete data likelihood if $\mathbf{z}$ and $\mathbf{y}$ were observed. For iterations $k = 0, 1, 2, \ldots$, the expected value of the complete data log likelihood function with respect to $\mathbf{y}$ conditional on $\mathbf{z}$ and $\boldsymbol{\Psi}^{(k)}$ is calculated in the E-step. This expectation is often referred to as the Q-function and denoted

$$Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right) = E_{\mathbf{y}}\left[\log L_c\left(\boldsymbol{\Psi}|\mathbf{z}, \mathbf{y}\right)|\mathbf{z}\right]_{\boldsymbol{\Psi}=\boldsymbol{\Psi}^{(k)}}. \tag{4.1}$$

The Q-function $Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right)$ is subsequently maximized in the M-step. The updated estimates $\boldsymbol{\Psi}^{(k+1)}$ are selected such that

$$Q\left(\boldsymbol{\Psi}^{(k+1)}; \boldsymbol{\Psi}^{(k)}\right) \geq Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right) \forall\ \boldsymbol{\Psi} \in \boldsymbol{\Omega}. \tag{4.2}$$

This iterative procedure continues until a prespecified convergence criteria is satisfied (McLachlan and Krishnan, 1997).

Dempster, Laird, and Rubin described key properties of the algorithm with its introduction. One of the most notable features of the EM algorithm is that the incomplete data likelihood function monotonically increases at each iteration, such that

$$\mathrm{L}\left(\boldsymbol{\Psi}^{(\mathrm{k}+1)}|\mathbf{z}\right) \geq \mathrm{L}\left(\boldsymbol{\Psi}^{(\mathrm{k})}|\mathbf{z}\right). \tag{4.3}$$

This leads to the following result:

Result 4.1.1. *For an iterative sequence* $\left\{\boldsymbol{\Psi}^{(\mathrm{k})}\right\}$*, if* $\mathrm{Q}\left(\boldsymbol{\Psi}^{(\mathrm{k}+1)};\boldsymbol{\Psi}^{(\mathrm{k})}\right) \geq \mathrm{Q}\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(\mathrm{k})}\right)$ *for any* $\boldsymbol{\Psi} \in \boldsymbol{\Omega}$*, then* $\mathrm{L}\left(\boldsymbol{\Psi}^{(\mathrm{k}+1)}|\mathbf{z}\right) \geq \mathrm{L}\left(\boldsymbol{\Psi}^{(\mathrm{k})}|\mathbf{z}\right)$.

Thus, an algorithm that satisfies Equation 4.2 yields a monotonically increasing likelihood (Dempster et al., 1977; McLachlan and Krishnan, 1997).

4.2 Generalized Expectation-Maximization Algorithm

The solutions in the M-step of the EM algorithm often exist in closed-form. However, when closed-form solutions are not attainable, finding the value of $\boldsymbol{\Psi}$ that globally maximizes the Q-function at each iteration is challenging. For this scenario, Dempster, Laird, and Rubin defined a generalized EM (GEM) algorithm.

Definition 4.2.1. *The GEM algorithm is a type of EM algorithm and consists of an M-step that selects* $\boldsymbol{\Psi}^{(\mathrm{k}+1)}$ *satisfying* $\mathrm{Q}\left(\boldsymbol{\Psi}^{(\mathrm{k}+1)};\boldsymbol{\Psi}^{(\mathrm{k})}\right) \geq \mathrm{Q}\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(\mathrm{k})}\right)$.

Thus, for every iteration in a GEM algorithm, the Q-function is monotonically increasing. Global maximization over all $\boldsymbol{\Psi} \in \boldsymbol{\Omega}$ is therefore not necessary. By Result 4.1.1, this is sufficient to ensure a monotonically increasing likelihood (McLachlan and Krishnan, 1997).

One iteration of the Newton-Raphson algorithm is a popular choice for maximization in the M-step of a GEM algorithm. For iteration $k+1$, Rai and Matthews (1993) propose using the updating equation

$$\boldsymbol{\Psi}^{(k+1)} = \boldsymbol{\Psi}^{(k)} - a^{(k)} \left[\frac{\partial^2 Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'} \right]^{-1}_{\boldsymbol{\Psi}=\boldsymbol{\Psi}^{(k)}} \left[\frac{\partial Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right)}{\partial \boldsymbol{\Psi}} \right]_{\boldsymbol{\Psi}=\boldsymbol{\Psi}^{(k)}}, \tag{4.4}$$

where $0 < a^{(k)} < 1$. Let $\mathbf{I}_d$ denote the d-dimensional identity matrix,

$$\tilde{\mathbf{I}}_c^{(k)} = - \left[\frac{\partial^2 Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'} \right]_{\boldsymbol{\Psi}=\boldsymbol{\Psi}^{(k)}}, \tag{4.5}$$

and

$$\mathbf{I}_c^{-1}\left(\boldsymbol{\Psi}^{(k)}\right) = \left[-\frac{\partial^2 \log L_c\left(\boldsymbol{\Psi}\right)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'} \right]^{-1}. \tag{4.6}$$

It can be shown that Equation 4.4 satisfies Definition 4.2.1 for a GEM sequence if the matrix

$$\mathbf{A}^{(k)} = \mathbf{I}_c^{-1}\left(\boldsymbol{\Psi}^{(k)}\right) \left\{ \mathbf{I}_d - \frac{1}{2} a^{(k)} \tilde{\mathbf{I}}_c^{(k)} \mathbf{I}_c^{-1}\left(\boldsymbol{\Psi}^{(k)}\right) \right\} \tag{4.7}$$

is positive-definite. Choosing a constant $a^{(k)}$ sufficiently small will yield a positive-definite $\mathbf{A}^{(k)}$ (Rai and Matthews, 1993; McLachlan and Krishnan, 1997).

An important property of the GEM algorithm is convergence of the likelihood. In 1983, Wu described the conditions to be satisfied for convergence of the likelihood (Wu, 1983). He first described the conditions for a sequence to converge to some point L^*, yielding the following result stated without proof:

Result 4.2.1. *Let* $\mathbf{\Omega}$ *denote the parameter space of* $\mathbf{\Psi}$*. Suppose that (i)* $\mathbf{\Omega}$ *is a* d*-dimensional subset of* $\mathbb{R}^{\mathrm{d}}$*, (ii)* $\mathbf{\Omega}_{\mathbf{\Psi}_{\mathrm{o}}} = \{\mathbf{\Psi} \in \mathbf{\Omega} : \mathrm{L}(\mathbf{\Psi}) \geq \mathrm{L}(\mathbf{\Psi}_{\mathrm{o}})\}$ *is compact for any* $\mathrm{L}(\mathbf{\Psi}_{\mathrm{o}}) > -\infty$*, (iii)* $\mathrm{L}(\mathbf{\Psi})$ *is continuous in* $\mathbf{\Omega}$ *and differentiable in the interior of* $\mathbf{\Omega}$*. Then any sequence* $\left\{\mathrm{L}\left(\mathbf{\Psi}^{(\mathrm{k})}\right)\right\}$ *is bounded above for any vector of initial values* $\mathbf{\Psi}^{(0)} \in \mathbf{\Omega}$ *such that* $\mathrm{L}\left(\mathbf{\Psi}^{(0)}\right) > -\infty$ *and, hence, the sequence converges to some point* L^* *(McLachlan and Krishnan, 1997).*

Assuming Result 4.2.1, Wu presented his main convergence theorem for a GEM. Let $\mathrm{M}\left(\mathbf{\Psi}^{(\mathrm{k})}\right)$ denote the point-to-set map of a GEM, or the choice of $\mathbf{\Psi}^{(\mathrm{k}+1)}$ such that $\mathrm{Q}\left(\mathbf{\Psi}^{(\mathrm{k}+1)}; \mathbf{\Psi}^{(\mathrm{k})}\right) \geq \mathrm{Q}\left(\mathbf{\Psi}; \mathbf{\Psi}^{(\mathrm{k})}\right)$. The following result is stated without proof:

Result 4.2.2. *"Let* $\left\{\mathbf{\Psi}^{(\mathrm{k})}\right\}$ *be an instance of a GEM sequence generated by* $\mathbf{\Psi}^{(\mathrm{k}+1)} \in \mathrm{M}\left(\mathbf{\Psi}^{(\mathrm{k})}\right)$*. Suppose that (i)* $\mathrm{M}\left(\mathbf{\Psi}^{(\mathrm{k})}\right)$ *is closed over the complement of* S*, the set of stationary points in the interior of* $\mathbf{\Omega}$ *and that (ii)* $\mathrm{L}\left(\mathbf{\Psi}^{(\mathrm{k}+1)}\right) \geq \mathrm{L}\left(\mathbf{\Psi}^{(\mathrm{k})}\right)$ *for all* $\mathbf{\Psi}^{(\mathrm{k})} \notin \mathrm{S}$*. Then all the limit points of* $\left\{\mathbf{\Psi}^{(\mathrm{k})}\right\}$ *are stationary points and* $\mathrm{L}\left(\mathbf{\Psi}^{(\mathrm{k})}\right)$ *converges monotonically to* $\mathrm{L}^* = \mathrm{L}(\mathbf{\Psi}^*)$ *for some stationary point* $\mathbf{\Psi}^* \in \mathrm{S}$*" (McLachlan and Krishnan, 1997).*

By Result 4.2.2, the GEM algorithm converges to a stationary point, defined as a saddle point, local maximum, or global maximum. In practice, one should compare results with different sets of starting values. However, McLachlan and Krishnan (1997) state that in many applications, the algorithm will converge to a local or global maximum and not a saddle point. If the sequence $(\mathbf{\Psi})$ is "trapped" at a saddle point $\mathbf{\Psi}^*$, "a small random perturbation of $\mathbf{\Psi}$ away from the saddle point $\mathbf{\Psi}^*$ will cause the GEM algorithm to diverge from the saddle point" (McLachlan and Krishnan, 1997). Thus, for any arbitrary small ϵ, the convergence criteria will not be satisfied at a saddle point since successive iterations will likely result in

a large change in the convergence criteria. The algorithm will therefore not stop at a saddle point and, hence, the stopping point will not be a saddle point. In order to address the concern that $\boldsymbol{\Psi}^*$ is not a local maximum, the algorithm should be implemented with different sets of initial values and the likelihood for each set should be calculated. If different initial values yield different limiting points, the likelihood values should be compared to identify the maximum. Additional sets of starting values should then be used to increase the likelihood that a global maximum has been attained.

4.3 Expectation-Conditional Maximization Algorithm

One of the most flexible GEM algorithms is the ECM algorithm. In the ECM algorithm, a complicated M-step can be divided into several conditional maximization (CM) steps. Each CM-step maximizes the Q-function subject to multiple constraints on $\boldsymbol{\Psi}$. The individual steps are performed over a smaller dimensional space and are thus simpler, faster, and more stable than a single maximization step (McLachlan and Krishnan, 1997).

Let S denote the number of CM steps. Furthermore, let $\boldsymbol{\Psi}^{(k+s/S)}$ denote the value of $\boldsymbol{\Psi}$ in the sth step of iteration $k+1$. In the sth CM-step, $\boldsymbol{\Psi}^{(k+s/S)}$ maximizes $Q\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(k)}\right)$ subject to the constraint

$$g_s\left(\boldsymbol{\Psi}\right) = g_s\left(\boldsymbol{\Psi}^{\left(k+\frac{s-1}{S}\right)}\right), \tag{4.8}$$

where $\{g_s\left(\boldsymbol{\Psi}\right), s = 1,\ldots,S\}$ is the set of S prespecified functions. This maximization procedure satisfies

$$Q\left(\boldsymbol{\Psi}^{(k+s/S)};\boldsymbol{\Psi}^{(k)}\right) \geq Q\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(k)}\right) \text{ for all } \boldsymbol{\Psi} \in \Omega_s\left(\boldsymbol{\Psi}^{(k+(s-1)/S)}\right), \tag{4.9}$$

where $\Omega_s\left(\boldsymbol{\Psi}^{(k+(s-1)/S)}\right) \equiv \left\{\boldsymbol{\Psi} \in \Omega : g_s(\boldsymbol{\Psi}) = g_s\left(\boldsymbol{\Psi}^{\left(k+\frac{s-1}{S}\right)}\right)\right\}$. By Equation 4.9,

$$\begin{aligned} Q\left(\boldsymbol{\Psi}^{(k+1)};\boldsymbol{\Psi}^{(k)}\right) &\geq Q\left(\boldsymbol{\Psi}^{(k+\frac{S-1}{S})};\boldsymbol{\Psi}^{(k)}\right) \\ &\geq Q\left(\boldsymbol{\Psi}^{(k+\frac{S-2}{S})};\boldsymbol{\Psi}^{(k)}\right) \\ &\vdots \\ &\geq Q\left(\boldsymbol{\Psi}^{(k)};\boldsymbol{\Psi}^{(k)}\right). \end{aligned} \tag{4.10}$$

By Result 4.1.1, this is a sufficient condition for the algorithm to yield a monotonically increasing likelihood. The ECM is therefore a GEM by Definition 4.2.1 (McLachlan and Krishnan, 1997).

In many applications of the ECM algorithm, the parameter vector $\boldsymbol{\Psi}$ is partitioned into S subvectors $\boldsymbol{\Psi} = \left(\boldsymbol{\Psi}_1', \ldots, \boldsymbol{\Psi}_S'\right)'$. The Q-function is maximized with respect to the subvector $\boldsymbol{\Psi}_s$ in the sth CM-step with the other $S-1$ subvectors fixed at the current values. The constraints $g_s(\boldsymbol{\Psi})$ are therefore equal to the vector containing all subvectors of $\boldsymbol{\Psi}$ except $\boldsymbol{\Psi}_s$ (McLachlan and Krishnan, 1997).

4.4 Spatiotemporal Expectation-Conditional Maximization

Due to its simplicity and key properties, the ECM algorithm is used for estimation. Let $\mathbf{G}_t = \mathbf{G}$. The unknown parameter vector to be estimated for the proposed model is $\boldsymbol{\Psi} = \left(\mu_0, \mathbf{G}, \sigma_\eta^2, \sigma_\omega^2, \boldsymbol{\beta}, \gamma, \nu, s\right)$. Let $\boldsymbol{\Psi} = \left(\boldsymbol{\Psi}_1', \boldsymbol{\Psi}_2'\right)'$, where $\boldsymbol{\Psi}_1 = \left(\mu_0, \mathbf{G}, \sigma_\eta^2, \sigma_\omega^2, \boldsymbol{\beta}\right)$ is the subset of parameters with closed-form MLEs and $\boldsymbol{\Psi}_2 = (\gamma, \nu, s)$ is the subset without closed-form MLEs. The CM-step consists of two steps: conditional maximization of $\boldsymbol{\Psi}_1$ and conditional

maximization of Ψ_2 via one iteration of Newton-Raphson. The algorithm can be expressed as follows:

1. Calculate $\Psi_1^{(k+1)}$ as the values of Ψ_1 that maximizes $Q\left(\Psi; \Psi^{(k)}\right)$ conditional on $\Psi_2 = \Psi_2^{(k)}$.
2. Calculate $\Psi_2^{(k+1)}$ as the values of Ψ_2 that maximizes $Q\left(\Psi; \Psi^{(k)}\right)$ conditional on $\Psi_1 = \Psi_1^{(k+1)}$.

Since $Q\left(\Psi_1^{(k)}, \Psi_2^{(k)}\right) \geq Q\left(\Psi_1^{(k)}, \Psi_2^{(k-1)}\right) \geq Q\left(\Psi_1^{(k-1)}, \Psi_2^{(k-1)}\right)$, the conditional maximization approach satisfies Definition 4.2.1 for a GEM algorithm (Xu and Wikle, 2007). The algorithm converges when

$$\frac{\left|L^{(k+1)} - L^{(k)}\right|}{\left|L^{(k)}\right|} < \epsilon \tag{4.11}$$

for some small $\epsilon > 0$.

4.4.1 Expectation-Step

The first step of the ECM algorithm is the E-Step. Let $\mathbf{z}_i = (\mathbf{z}_{i1}', \mathbf{z}_{i2}', \ldots, \mathbf{z}_{iT}')'$ and $\mathbf{y}_i = (\mathbf{y}_{i1}', \mathbf{y}_{i2}', \ldots, \mathbf{y}_{iT}')'$. For the spatiotemporal model, the complete data joint distribution for individual $i, i = 1, ..., m$, is $f(\mathbf{z}_i, \mathbf{y}_i) = f(\mathbf{z}_i|\mathbf{y}_i) f(\mathbf{y}_i)$. For m independent individuals with fMRI observations at T time points, the complete data likelihood is

$$\begin{aligned} L_c(\Psi; \mathbf{z}, \mathbf{y}) &= \prod_{i=1}^{m} f(\mathbf{z}_i|\mathbf{y}_i) f(\mathbf{y}_i) \\ &= \prod_{i=1}^{m} \left\{ \left\{ \prod_{t=1}^{T} f(\mathbf{z}_{it}|\mathbf{y}_{it}) \right\} \times \left\{ f(\mathbf{y}_{i0}) \prod_{t=1}^{T} f\left(\mathbf{y}_{it}|\mathbf{y}_{i(t-1)}\right) \right\} \right\} \\ &\propto \prod_{i=1}^{m} |\Sigma_0|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{y}_{i0} - \mu_0)' \Sigma_0^{-1} (\mathbf{y}_{i0} - \mu_0) \right\} \end{aligned} \tag{4.12}$$

$$\times \prod_{i=1}^{m} \prod_{t=1}^{T} |\boldsymbol{\Sigma}_{\eta}|^{-1/2} \exp\left\{-\frac{1}{2}\left(\mathbf{y}_{it} - \mathbf{G}\mathbf{y}_{i(t-1)}\right)' \boldsymbol{\Sigma}_{\eta}^{-1}\left(\mathbf{y}_{it} - \mathbf{G}\mathbf{y}_{i(t-1)}\right)\right\}$$
$$\times \prod_{i=1}^{m} \prod_{t=1}^{T} |\boldsymbol{\Sigma}_{\xi}|^{-1/2} \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}\right)' \boldsymbol{\Sigma}_{\xi}^{-1}\left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}\right)\right\}$$

The complete data log likelihood is thus

$$\begin{aligned}
\log L_c(\boldsymbol{\Psi}; \mathbf{z}, \mathbf{y}) \quad \propto \quad & -mT \log|\boldsymbol{\Sigma}_{\xi}| - m \log|\boldsymbol{\Sigma}_0| - mT \log|\boldsymbol{\Sigma}_{\eta}| \\
& - \sum_{i=1}^{m} \left(\mathbf{y}_{i0} - \boldsymbol{\mu}_{i0}\right)' \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{y}_{i0} - \boldsymbol{\mu}_{i0}\right) \\
& - \sum_{i=1}^{m} \sum_{t=1}^{T} \left(\mathbf{y}_{it} - \mathbf{G}\mathbf{y}_{i(t-1)}\right)' \boldsymbol{\Sigma}_{\eta}^{-1} \left(\mathbf{y}_{it} - \mathbf{G}\mathbf{y}_{i(t-1)}\right) \\
& - \sum_{i=1}^{m} \sum_{t=1}^{T} \left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}\right)' \boldsymbol{\Sigma}_{\xi}^{-1} \left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}\right).
\end{aligned} \tag{4.13}$$

The Q-function is the expectation of this function (Xu and Wikle, 2007; Fasso and Cameletti, 2009).

4.4.1.1 Q-Function

Let $Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right)$ denote $E_{\mathbf{y}}\left[\log L_c(\boldsymbol{\Psi}|\mathbf{z}, \mathbf{y}) | \mathbf{z}, \boldsymbol{\Psi}^{(k)}\right]_{\boldsymbol{\Psi}=\boldsymbol{\Psi}^{(k)}}$, the expectation of the complete data log likelihood with respect to $\mathbf{y}$. The Q-function is

$$\begin{aligned}
Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right) \quad = \quad & -mT \log|\boldsymbol{\Sigma}_{\xi}| - m \log|\boldsymbol{\Sigma}_0| - mT \log|\boldsymbol{\Sigma}_{\eta}| \\
& - \sum_{i=1}^{m} E_{\mathbf{y}_{it}}\left[\left(\mathbf{y}_{i0} - \boldsymbol{\mu}_0\right)' \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{y}_{i0} - \boldsymbol{\mu}_0\right) | \mathbf{z}.\boldsymbol{\Psi}^{(k)}\right] \\
& - \sum_{i=1}^{m} \sum_{t=1}^{T} E_{\mathbf{y}_{it}}\left[\left(\mathbf{y}_{it} - \mathbf{G}\mathbf{y}_{i(t-1)}\right)' \boldsymbol{\Sigma}_{\eta}^{-1} \left(\mathbf{y}_{it} - \mathbf{G}\mathbf{y}_{i(t-1)}\right) | \mathbf{z}, \boldsymbol{\Psi}^{(k)}\right] \\
& - \sum_{i=1}^{m} \sum_{t=1}^{T} E_{\mathbf{y}_{it}}\left[\left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}\right)' \boldsymbol{\Sigma}_{\xi}^{-1} \left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}\right) | \mathbf{z}, \boldsymbol{\Psi}^{(k)}\right]
\end{aligned} \tag{4.14}$$

$$= \quad -mT \log|\boldsymbol{\Sigma}_{\xi}| - m \log|\boldsymbol{\Sigma}_0| - mT \log|\boldsymbol{\Sigma}_{\eta}| \tag{4.15}$$

$$
\begin{aligned}
&-\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{m}\mathsf{E}_{\mathbf{y}_{it}}\left[\left(\mathbf{y}_{i0}-\boldsymbol{\mu}_0\right)\left(\mathbf{y}_{i0}-\boldsymbol{\mu}_0\right)'|\mathbf{z},\boldsymbol{\Psi}^{(k)}\right]\right\}\\
&-\mathrm{tr}\left\{\boldsymbol{\Sigma}_{\eta}^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathsf{E}_{\mathbf{y}_{it}}\left[\left(\mathbf{y}_{it}-\mathbf{G}\mathbf{y}_{i(t-1)}\right)\left(\mathbf{y}_{it}-\mathbf{G}\mathbf{y}_{i(t-1)}\right)'|\mathbf{z},\boldsymbol{\Psi}^{(k)}\right]\right\}\\
&-\mathrm{tr}\left\{\boldsymbol{\Sigma}_{\xi}^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathsf{E}_{\mathbf{y}_{it}}\left[\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}\right)\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}\right)'|\mathbf{z},\boldsymbol{\Psi}^{(k)}\right]\right\}.
\end{aligned}
$$

Calculation of the Q-function requires estimates of $\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}\mathbf{y}_{i(t-1)}'|\mathbf{z},\boldsymbol{\Psi}^{(k)}\right]$, $\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}\mathbf{y}_{it}'|\mathbf{z},\boldsymbol{\Psi}^{(k)}\right]$, and $\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}|\mathbf{z},\boldsymbol{\Psi}^{(k)}\right]$. The Kalman filter and smoother algorithm is used to derive these estimates and described in Section 4.4.1.2. Let $\mathbf{y}_{it}^{T}$ and $\mathbf{P}_{it}^{T}$ denote the Kalman smoother mean and variance, respectively, of $\mathbf{y}_{it}$ conditional on the observations $\mathbf{z}$ and kth iteration parameter estimates $\boldsymbol{\Psi}^{(k)}$. Similarly, let $\mathbf{P}_{i(t,t-1)}^{T}$ denote the Kalman smoother estimate of the lag-one covariance between $\mathbf{y}_{it}^{T}$ and $\mathbf{y}_{i(t-1)}^{T}$. Using the Kalman smoother estimates, the expectation $\mathsf{E}_{\mathbf{y}_{it}}\left[\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}\right)\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}\right)'|\mathbf{z}\right]$ is simplified to

$$
\begin{aligned}
&\mathsf{E}_{\mathbf{y}_{it}}\left[\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}\right)\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}\right)'|\mathbf{z}\right] \qquad (4.16)\\
=\;&\mathbf{z}_{it}\mathbf{z}_{it}'-\mathbf{z}_{it}\boldsymbol{\beta}'\mathbf{X}_i'-\mathbf{z}_{it}\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}'|\mathbf{z}\right]\mathbf{K}'\\
&-\mathbf{X}_i\boldsymbol{\beta}\mathbf{z}_{it}'+\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}_i'+\mathbf{X}_i\boldsymbol{\beta}\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}'\middle|\,\mathbf{z}\right]\mathbf{K}_i'\\
&-\mathbf{K}_i\mathsf{E}\left[\mathbf{y}_{it}|\mathbf{z}\right]\mathbf{z}_{it}'+\mathbf{K}_i\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}|\mathbf{z}\right]\boldsymbol{\beta}'\mathbf{X}_i'+\mathbf{K}_i\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}\mathbf{y}_{it}'|\mathbf{z}\right]\mathbf{K}_i'\\
=\;&\mathbf{z}_{it}\mathbf{z}_{it}'-\mathbf{z}_{it}\boldsymbol{\beta}'\mathbf{X}_i'-\mathbf{z}_{it}\mathbf{y}_{it}^{T'}\mathbf{K}_i'-\mathbf{X}_i\boldsymbol{\beta}\mathbf{z}_{it}'+\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}_i'+\mathbf{X}_i\boldsymbol{\beta}\mathbf{y}_{it}^{T'}\mathbf{K}_i'\\
&-\mathbf{K}_i\mathbf{y}_{it}^{T}\mathbf{z}_{it}'+\mathbf{K}_i\mathbf{y}_{it}^{T}\boldsymbol{\beta}'\mathbf{X}_i'+\mathbf{K}_i\mathbf{y}_{it}^{T}\mathbf{y}_{it}^{T'}\mathbf{K}_i'+\mathbf{K}_i\mathbf{P}_{it}^{T}\mathbf{K}_i'\\
=\;&\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}^{T}\right)\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}^{T}\right)'+\mathbf{K}_i\mathbf{P}_{it}^{T}\mathbf{K}_i',
\end{aligned}
$$

The expectation $\mathsf{E}_{\mathbf{y}_{it}}\left[\left(\mathbf{y}_{i0}-\boldsymbol{\mu}_0\right)\left(\mathbf{y}_{i0}-\boldsymbol{\mu}_0\right)'\middle|\,\mathbf{z}\right]$ is simplified to

$$
\begin{aligned}
& \mathsf{E}_{\mathbf{y}_{it}}\left[(\mathbf{y}_{i0}-\boldsymbol{\mu}_0)(\mathbf{y}_{i0}-\boldsymbol{\mu}_0)'\middle|\,\mathbf{z}\right] \qquad (4.17)\\
= \;& \mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{i0}\mathbf{y}_{i0}'|\mathbf{z}\right]-\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{i0}|\mathbf{z}\right]\boldsymbol{\mu}_0'-\boldsymbol{\mu}_0\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{i0}'|\mathbf{z}\right]+\boldsymbol{\mu}_0\boldsymbol{\mu}_0'\\
= \;& \mathbf{y}_{i0}^{\mathsf{T}}\mathbf{y}_{i0}^{\mathsf{T}'}+\mathbf{P}_{i0}^{\mathsf{T}}-\mathbf{y}_{i0}^{\mathsf{T}}\boldsymbol{\mu}_0'-\boldsymbol{\mu}_0\mathbf{y}_{i0}^{\mathsf{T}'}+\boldsymbol{\mu}_0\boldsymbol{\mu}_0'\\
= \;& \left(\mathbf{y}_{i0}^{\mathsf{T}}-\boldsymbol{\mu}_0\right)\left(\mathbf{y}_{i0}^{\mathsf{T}}-\boldsymbol{\mu}_0\right)'+\mathbf{P}_{i0}^{\mathsf{T}}.
\end{aligned}
$$

The expectation $\mathsf{E}_{\mathbf{y}_{it}}\left[\left(\mathbf{y}_{it}-\mathbf{G}\mathbf{y}_{i(t-1)}\right)\left(\mathbf{y}_{it}-\mathbf{G}\mathbf{y}_{i(t-1)}\right)'|\mathbf{z}\right]$ is simplified to

$$
\begin{aligned}
& \mathsf{E}_{\mathbf{y}_{it}}\left[\left(\mathbf{y}_{it}-\mathbf{G}\mathbf{y}_{i(t-1)}\right)\left(\mathbf{y}_{it}-\mathbf{G}\mathbf{y}_{i(t-1)}\right)'|\mathbf{z}\right] \qquad (4.18)\\
= \;& \mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}\mathbf{y}_{it}'|\mathbf{z}\right]-\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}\mathbf{y}_{i(t-1)}'|\mathbf{z}\right]\mathbf{G}'-\mathbf{G}\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{i(t-1)}\mathbf{y}_{it}'|\mathbf{z}\right]\\
& +\mathbf{G}\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{i(t-1)}\mathbf{y}_{i(t-1)}'\right]\mathbf{G}'\\
= \;& \left(\mathbf{y}_{it}^{\mathsf{T}}\mathbf{y}_{it}^{\mathsf{T}'}+\mathbf{P}_{it}^{\mathsf{T}}\right)-\left(\mathbf{y}_{it}^{\mathsf{T}}\mathbf{y}_{i(t-1)}^{\mathsf{T}'}+\mathbf{P}_{i(t,t-1)}^{\mathsf{T}}\right)\mathbf{G}'-\mathbf{G}\left(\mathbf{y}_{i(t-1)}^{\mathsf{T}}\mathbf{y}_{t}^{\mathsf{T}'}+\mathbf{P}_{i(t,t-1)}^{\mathsf{T}}\right)'\\
& +\mathbf{G}\left(\mathbf{y}_{i(t-1)}^{\mathsf{T}}\mathbf{y}_{i(t-1)}^{\mathsf{T}'}+\mathbf{P}_{i(t-1)}^{\mathsf{T}}\right)\mathbf{G}'
\end{aligned}
$$

The Q-function to be maximized is thus

$$
\begin{aligned}
\mathrm{Q}\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(k)}\right) = \;& -mT\log|\boldsymbol{\Sigma}_\xi|-mT\log|\boldsymbol{\Sigma}_\eta|-m\log|\boldsymbol{\Sigma}_0| \qquad (4.19)\\
& -\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{m}\left[\left(\mathbf{y}_{i0}^{\mathsf{T}}-\boldsymbol{\mu}_0\right)\left(\mathbf{y}_{i0}^{\mathsf{T}}-\boldsymbol{\mu}_0\right)'+\mathbf{P}_{i0}^{\mathsf{T}}\right]\right\}\\
& -\mathrm{tr}\left\{\boldsymbol{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\left[\mathbf{S}_{it(11)}-\mathbf{S}_{it(10)}\mathbf{G}'-\mathbf{G}\mathbf{S}_{it(10)}'+\mathbf{G}\mathbf{S}_{it(00)}\mathbf{G}'\right]\right\}\\
& -\mathrm{tr}\left\{\boldsymbol{\Sigma}_\xi^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\left[\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}^{\mathsf{T}}\right)\left(\mathbf{z}_{it}-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{K}_i\mathbf{y}_{it}^{\mathsf{T}}\right)'+\mathbf{K}_i\mathbf{P}_{it}^{\mathsf{T}}\mathbf{K}_i'\right]\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{S}_{it(11)} &= \mathbf{y}_{it}^{T}\left(\mathbf{y}_{it}^{T}\right) + \mathbf{P}_{it}^{T}, \\
\mathbf{S}_{it(10)} &= \mathbf{y}_{it}^{T}\left(\mathbf{y}_{i(t-1)}^{T}\right)' + \mathbf{P}_{i(t,t-1)}^{T}, \\
\mathbf{S}_{it(00)} &= \mathbf{y}_{i(t-1)}^{T}\left(\mathbf{y}_{i(t-1)}^{T}\right)' + \mathbf{P}_{i(t-1)}^{T}.
\end{aligned}
$$

4.4.1.2 Kalman Filter and Smoother

Simplification of the Q-function requires estimates of $\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}|\mathbf{z}, \boldsymbol{\Psi}^{(k)}\right]$, $\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}\mathbf{y}_{it}'|\mathbf{z}, \boldsymbol{\Psi}^{(k)}\right]$, and $\mathsf{E}_{\mathbf{y}_{it}}\left[\mathbf{y}_{it}\mathbf{y}_{i(t-1)}'|\mathbf{z}, \boldsymbol{\Psi}^{(k)}\right]$. Under the state-space formulation of the spatiotemporal model given in Equation 3.4 and Equation 3.5, $\mathbf{Y}_{it}$ is defined as a latent variable that dictates the observable dynamic process $\mathbf{Z}_{it}$. For maximization of the Q-function, the unobserved variable $\mathbf{y}_{it}$ is estimated given information in observed data $\mathbf{z}_{it}$. It is often of interest to estimate the value of the state at a particular time t conditional on a sequence of data. If $\mathbf{y}_{it}$ is estimated conditional on data from the past, the process is referred to as prediction or forecasting. If $\mathbf{y}_{it}$ is estimated conditional on data from the past and present, the process is referred to as filtering. Furthermore, if $\mathbf{y}_{it}$ is estimated conditional on data from the past, present, and future, the process is called smoothing (Shumway and Stoffer, 2017).

Calculation of the Q-function requires estimates of expectations of functions of $\mathbf{y}_{it}$ given all observed data $\mathbf{z}_{it}$ up to time T. Smoothed estimates of $\mathbf{y}_{it}$ are therefore required for the ECM algorithm. In time series analysis by state-space methods, the Kalman filter and smoother are often used to estimate the conditional mean and variance of $\mathbf{y}_{it}$ and the lag-one covariance of $\mathbf{y}_{it}$ and $\mathbf{y}_{i(t-1)}$. The smoother is a function of the filter, which estimates the conditional distribution of $\mathbf{y}_{it}$ given observed data up to time t. The estimating procedure is recursive, producing an

updated estimate of the dynamic system each time a new observation is observed. Since $\mathbf{z}_{it}$ is assumed to be normal, the conditional distributions of $\mathbf{y}_{it}$ given any subset of observed data are also normal. The estimating procedure therefore yields minimum variance linear unbiased estimates. Furthermore, conditional distributions of $\mathbf{y}_{it}$ obtained from a frequentist approach and posterior densities of $\mathbf{y}_{it}$ derived from the standpoint of Bayesian inference yield the same estimates of the mean vectors and variance matrices (Kalman, 1960; Durbin and Koopman, 2012).

Definitions for notation used for the Kalman estimates and updating equations are provided in Table II. The first step of the process is calculation of the Kalman filter. Let $\mathbf{y}_{it}^{t}$ and $\mathbf{P}_{it}^{t}$ denote the mean and variance, respectively, of the conditional distribution of $\mathbf{y}_{it}$ given observed data up to time t. For $t = 1, ..., T$, the Kalman filtered values are

$$\begin{aligned} \mathbf{y}_{it}^{t} &= \mathbf{y}_{it}^{t-1} + \mathbf{A}_{it}\left(\mathbf{z}_{it} - \mathbf{X}_{i}\boldsymbol{\beta} - \mathbf{K}_{i}\mathbf{y}_{it}^{t-1}\right), \qquad (4.20) \\ \mathbf{P}_{it}^{t} &= \mathbf{P}_{it}^{t-1} - \mathbf{A}_{it}\mathbf{K}_{i}\mathbf{P}_{it}^{t-1}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{y}_{it}^{t-1} &= \mathbf{G}\mathbf{y}_{i(t-1)}^{t-1}, \qquad (4.21) \\ \mathbf{P}_{it}^{t-1} &= \mathbf{G}\mathbf{P}_{i(t-1)}^{t-1}\mathbf{G}' + \boldsymbol{\Sigma}_{\eta}, \\ \mathbf{A}_{it} &= \mathbf{P}_{it}^{t-1}\mathbf{K}_{i}'\left(\mathbf{K}_{i}\mathbf{P}_{it}^{t-1}\mathbf{K}_{i}' + \boldsymbol{\Sigma}_{\xi}\right)^{-1}. \end{aligned}$$

$\mathbf{y}_{i0}^{0}$ and $\mathbf{P}_{i0}^{0}$ initiate the recursive procedure and are considered nuisance parameters (Xu and Wikle, 2007; Fasso and Cameletti, 2009).

TABLE II KALMAN NOTATION

	Notation	Mathematical Expression
Predicted Value	$\mathbf{y}_{it}^{t-1}$	$\mathrm{E}\left[\mathbf{y}_{it}\middle\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{(t-1)}\right]$
	$\mathbf{P}_{it}^{t-1}$	$\mathrm{var}\left(\mathbf{y}_{it}\middle\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{(t-1)}\right)$
	$\mathbf{P}_{i(t,t-1)}^{t-1}$	$\mathrm{cov}\left(\mathbf{y}_{it}, \mathbf{y}_{i(t-1)}\middle\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{(t-1)}\right)$
Filtered Value	$\mathbf{y}_{it}^{t}$	$\mathrm{E}[\mathbf{y}_{it}\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_t]$
	$\mathbf{P}_{it}^{t}$	$\mathrm{var}(\mathbf{y}_{it}\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_t)$
	$\mathbf{P}_{i(t,t-1)}^{t}$	$\mathrm{cov}\left(\mathbf{y}_{it}, \mathbf{y}_{i(t-1)}\middle\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_t\right)$
Smoothed Value	$\mathbf{y}_{it}^{T}$	$\mathrm{E}[\mathbf{y}_{it}\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T]$
	$\mathbf{P}_{it}^{T}$	$\mathrm{var}(\mathbf{y}_{it}\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T)$
	$\mathbf{P}_{i(t,t-1)}^{T}$	$\mathrm{cov}\left(\mathbf{y}_{it}, \mathbf{y}_{i(t-1)}\middle\vert \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T\right)$

Using the filter values, the Kalman smoother estimates the conditional distribution of $\mathbf{y}_{it}$ given data observed at all times T. Let $\mathbf{y}_{it}^{T}$ and $\mathbf{P}_{it}^{T}$ denote the mean and variance, respectively, of the conditional distribution of $\mathbf{y}_{it}$ given data observed at all times T. Furthermore, let $\mathbf{P}_{i(t,t-1)}^{T}$ denote the covariance of $\left(\mathbf{y}_{it}, \mathbf{y}_{i(t-1)}\right)$ given data observed at all times T. To calculate the Kalman smoother, a backward recursion is used. For $t = T, T-1, \ldots, 1$, the Kalman smoother values $\mathbf{y}_{i(t-1)}^{T}$ and $\mathbf{P}_{i(t-1)}^{T}$ are

$$\begin{aligned} \mathbf{y}_{i(t-1)}^{T} &= \mathbf{y}_{i(t-1)}^{t-1} + \mathbf{J}_{i(t-1)}\left(\mathbf{y}_{it}^{T} - \mathbf{G}\mathbf{y}_{i(t-1)}^{t-1}\right), \qquad (4.22) \\ \mathbf{P}_{i(t-1)}^{T} &= \mathbf{P}_{i(t-1)}^{t-1} + \mathbf{J}_{i(t-1)}\left(\mathbf{P}_{it}^{T} - \mathbf{P}_{it}^{t-1}\right)\mathbf{J}_{i(t-1)}', \end{aligned}$$

where

$$\mathbf{J}_{i(t-1)} = \mathbf{P}_{i(t-1)}^{t-1}\mathbf{G}'\left(\mathbf{P}_{it}^{t-1}\right)^{-1}. \qquad (4.23)$$

The lag-one covariance $\mathbf{P}^{T}_{i(t-1,t-2)}$ is also estimated by a backward recursion. For $t = T, T-1, \ldots, 2$, the Kalman smoother value for the lag-one covariance is

$$\mathbf{P}^{T}_{i(t-1,t-2)} = \mathbf{P}^{t-1}_{i(t-1)}\mathbf{J}'_{i(t-2)} + \mathbf{J}_{i(t-1)}\left(\mathbf{P}^{T}_{i,(t,t-1)} - \mathbf{G}\mathbf{P}^{t-1}_{i(t-1)}\right)\mathbf{J}'_{i(t-2)}. \tag{4.24}$$

The smoothed estimates are used in Equation 4.19 for each iteration of the algorithm (Xu and Wikle, 2007; Fasso and Cameletti, 2009).

4.4.2 Conditional Maximization Step

Once the E-Step is complete, the CM-step is performed.

4.4.2.1 Closed-Form Solutions

The first CM-step in iteration $k+1$ of the ECM algorithm maximizes the Q-Function in Equation 4.19 with respect to $\boldsymbol{\Psi}_1 = \left(\boldsymbol{\mu}_0, \mathbf{G}, \sigma^2_\eta, \sigma^2_\omega, \boldsymbol{\beta}\right)$ conditional on $\boldsymbol{\Psi}_2 = \boldsymbol{\Psi}_2^{(k)}$. The solutions are derived by taking partial derivatives of the Q-function with respect to each parameter and setting to 0. The estimate for $\boldsymbol{\mu}_0$, the mean of $\mathbf{Y}_{i0}$, is derived as follows:

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\mu}_0} &= -\frac{\partial}{\partial \boldsymbol{\mu}_0}\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{m}\left[\left(\mathbf{y}^{T}_{i0} - \boldsymbol{\mu}_0\right)\left(\mathbf{y}^{T}_{i0} - \boldsymbol{\mu}_0\right)' + \mathbf{P}^{T}_{i0}\right]\right\} \\
&= \sum_{i=1}^{m}\left[\frac{\partial}{\partial \boldsymbol{\mu}_0}\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\mathbf{y}^{T}_{i0}\boldsymbol{\mu}'_0\right\} + \frac{\partial}{\partial \boldsymbol{\mu}_0}\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\mathbf{y}^{T\prime}_{i0}\right\} - \frac{\partial}{\partial \boldsymbol{\mu}_0}\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\boldsymbol{\mu}'_0\right\}\right] \\
&= \sum_{i=1}^{m}\left[\boldsymbol{\Sigma}_0^{-1}\mathbf{y}^{T}_{i0} + \boldsymbol{\Sigma}_0^{-1}\mathbf{y}^{T}_{i0} - \left(\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_0^{-1}\right)\boldsymbol{\mu}_0\right] \\
&= 2\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{m}\left[\mathbf{y}^{T}_{i0}\right] - 2m\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 \\
0 &= \boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{m}\left[\mathbf{y}^{T}_{i0}\right] - m\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{\mu}}_0 \\
\hat{\boldsymbol{\mu}}_0 &= \frac{1}{m}\sum_{i=1}^{m}\mathbf{y}^{T}_{i0}.
\end{aligned} \tag{4.25}$$

The estimate of the transition matrix $\mathbf{G}$ is derived as follows:

$$\begin{aligned}
\frac{\partial Q}{\partial \mathbf{G}} &= -\frac{\partial}{\partial \mathbf{G}} tr\left\{\mathbf{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\left[\mathbf{S}_{it(11)} - \mathbf{S}_{it(10)}\mathbf{G}' - \mathbf{G}\mathbf{S}_{it(10)}' + \mathbf{G}\mathbf{S}_{it(00)}\mathbf{G}'\right]\right\} \qquad (4.26)\\
&= \frac{\partial}{\partial \mathbf{G}} tr\left\{\mathbf{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{S}_{it(10)}\mathbf{G}'\right\} + \frac{\partial}{\partial \mathbf{G}} tr\left\{\mathbf{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{G}\mathbf{S}_{it(10)}'\right\}\\
&\quad -\frac{\partial}{\partial \mathbf{G}} tr\left\{\mathbf{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{G}\mathbf{S}_{it(00)}\mathbf{G}'\right\}\\
&= \mathbf{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{S}_{it(10)} + \mathbf{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{S}_{it(10)}' + \mathbf{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{G}\mathbf{S}_{it(00)}'\\
&\quad +\mathbf{\Sigma}_\eta^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{G}\mathbf{S}_{it(00)}\\
0 &= 2\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{S}_{it(10)} + 2\sum_{i=1}^{m}\sum_{t=1}^{T}\hat{\mathbf{G}}\mathbf{S}_{it(00)}\\
\hat{\mathbf{G}} &= \sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{S}_{it(10)}\left(\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{S}_{it(00)}\right)^{-1}.
\end{aligned}$$

Let $\mathbf{E} = \sum_{i=1}^{m}\sum_{t=1}^{T}\left[\mathbf{S}_{it(11)} - \mathbf{S}_{it(10)}\mathbf{G}' - \mathbf{G}\mathbf{S}_{it(10)}' + \mathbf{G}\mathbf{S}_{it(00)}\mathbf{G}'\right]$. The estimate of the diagonal element of $\mathbf{\Sigma}_\eta$, σ_η^2, is derived as follows:

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma_\eta^2} &= \frac{\partial}{\partial \sigma_\eta^2}\left\{-mT\log|\mathbf{\Sigma}_\eta| - tr\left\{\mathbf{\Sigma}_\eta^{-1}\mathbf{E}\right\}\right\} \qquad (4.27)\\
&= -\frac{\partial}{\partial \sigma_\eta^2} mT\log|\sigma_\eta^2\mathbf{I}_r| - \frac{\partial}{\partial \sigma_\eta^2}\frac{1}{\sigma_\eta^2} tr\left\{\mathbf{I}_r^{-1}\mathbf{E}\right\}\\
&= -mT\times tr\left\{\left(\sigma_\eta^2\mathbf{I}_r\right)^{-1}\frac{\partial}{\partial \sigma_\eta^2}\sigma_\eta^2\mathbf{I}_r\right\} + \frac{1}{\left(\sigma_\eta^2\right)^2} tr\{\mathbf{E}\}\\
&= -mT\times\frac{1}{\sigma_\eta^2} tr\left\{\mathbf{I}_r^{-1}\mathbf{I}_r\right\} + \frac{1}{\left(\sigma_\eta^2\right)^2} tr\{\mathbf{E}\}\\
0 &= -\frac{mrT}{\hat{\sigma}_\eta^2} + \frac{1}{\left(\hat{\sigma}_\eta^2\right)^2} tr\{\mathbf{E}\}\\
\hat{\sigma}_\eta^2 &= \frac{tr\{\mathbf{E}\}}{mrT}.
\end{aligned}$$

Let $\mathbf{W} = \sum_{i=1}^{m}\sum_{t=1}^{T}\left[\left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}^{T}\right)\left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}^{T}\right)' + \mathbf{K}_i\mathbf{P}_{it}^{T}\mathbf{K}_i'\right]$. The estimate of σ_ω^2 is derived as follows:

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma_\omega^2} &= \frac{\partial}{\partial \sigma_\omega^2}\left\{-mT\log|\boldsymbol{\Sigma}_\xi| - tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{W}\right\}\right\} \qquad (4.28)\\
&= -\frac{\partial}{\partial \sigma_\omega^2}mT\log|\boldsymbol{\Sigma}_\xi| - \frac{\partial}{\partial \sigma_\omega^2}tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{W}\right\}\\
&= -\frac{\partial}{\partial \sigma_\omega^2}mT\log|\sigma_\omega^2\boldsymbol{\Gamma}| - \frac{\partial}{\partial \sigma_\omega^2}tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\mathbf{W}\right\}\\
&= -mT\times tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\frac{\partial}{\partial \sigma_\omega^2}\sigma_\omega^2\boldsymbol{\Gamma}\right\} + \frac{1}{\left(\sigma_\omega^2\right)^2}tr\left\{\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\}\\
&= -mT\times\frac{1}{\sigma_\omega^2}tr\left\{\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}\right\} + \frac{1}{\left(\sigma_\omega^2\right)^2}tr\left\{\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\}\\
0 &= -\frac{mnT}{\hat{\sigma}_\omega^2} + \frac{1}{\left(\hat{\sigma^2}_\omega\right)^2}tr\left\{\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\}\\
\hat{\sigma}_\omega^2 &= \frac{tr\left\{\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\}}{mnT}.
\end{aligned}$$

Finally, estimates of fixed-effect parameters $\boldsymbol{\beta}$ are derived as follows:

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\beta}} &= -\frac{\partial}{\partial \boldsymbol{\beta}}tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}\left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}^{T}\right)\left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{y}_{it}^{T}\right)' \qquad (4.29)\right.\\
&\quad \left.+\mathbf{K}_i\mathbf{P}_{it}^{T}\mathbf{K}_i'\right\}\\
&= \sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\beta}}\left(tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{z}_{it}\boldsymbol{\beta}'\mathbf{X}_i'\right\} + tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{X}_i\boldsymbol{\beta}\mathbf{z}_{it}'\right\} - tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}_i'\right\}\right.\\
&\quad \left.-tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{X}_i\boldsymbol{\beta}\mathbf{y}_{it}^{T'}\mathbf{K}_i'\right\} - tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{K}_i\mathbf{y}_{it}^{T}\boldsymbol{\beta}'\mathbf{X}_i'\right\}\right)\\
&= \sum_{i=1}^{m}\sum_{t=1}^{T}\left(\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{z}_{it} + \mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{z}_{it} - \mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{X}_i\boldsymbol{\beta} - \mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{X}_i\boldsymbol{\beta} - \mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{K}_i\mathbf{y}_{it}^{T}\right.\\
&\quad \left.-\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{K}_i\mathbf{y}_{it}^{T}\right)\\
&= \sum_{i=1}^{m}\sum_{t=1}^{T}\left(2\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{z}_{it} - 2\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{X}_i\boldsymbol{\beta} - 2\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{K}_i\mathbf{y}_{it}^{T}\right)
\end{aligned}$$

$$
\begin{aligned}
0 &= \sum_{i=1}^{m}\sum_{t=1}^{T}\left(\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{z}_{it} - \mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{K}_i\mathbf{y}_{it}^{T}\right) - \sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{X}_i\hat{\boldsymbol{\beta}} \\
\hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{m}\sum_{t=1}^{T}\mathbf{X}_i'\boldsymbol{\Sigma}_\xi^{-1}\left(\mathbf{z}_{it} - \mathbf{K}_i\mathbf{y}_{it}^{T}\right)\right).
\end{aligned}
$$

4.4.2.2 Newton-Raphson

The second CM-step of iteration $k+1$ consists of maximizing $\boldsymbol{\Psi}_2 = (\gamma, \nu, s)$ conditional on $\boldsymbol{\Psi}_1 = \boldsymbol{\Psi}_1^{(k+1)}$ via one Newton-Raphson iteration. In this stage, the Newton-Raphson updating equation is

$$
\begin{bmatrix} \gamma^{(k+1)} \\ \nu^{(k+1)} \\ s^{(k+1)} \end{bmatrix} = \begin{bmatrix} \gamma^{(k)} \\ \nu^{(k)} \\ s^{(k)} \end{bmatrix} - a \begin{bmatrix} \frac{\partial^2 Q}{\partial\gamma^2} & \frac{\partial^2 Q}{\partial\gamma\partial\nu} & \frac{\partial^2 Q}{\partial\gamma\partial s} \\ \frac{\partial^2 Q}{\partial\gamma\partial\nu} & \frac{\partial^2 Q}{\partial\nu^2} & \frac{\partial^2 Q}{\partial\nu\partial s} \\ \frac{\partial^2 Q}{\partial\gamma\partial s} & \frac{\partial^2 Q}{\partial\nu\partial s} & \frac{\partial^2 Q}{\partial s^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial Q}{\partial\gamma} \\ \frac{\partial Q}{\partial\nu} \\ \frac{\partial Q}{\partial s} \end{bmatrix},
$$

where $0 < a < 1$. This updating equation requires the gradient and Hessian matrix of the Q-function with respect to parameters in $\boldsymbol{\Psi}_2$. The first partial derivative of the Q-function with respect to γ is derived as follows:

$$
\begin{aligned}
\frac{\partial Q}{\partial\gamma} &= -mT\frac{\partial}{\partial\gamma}\log|\boldsymbol{\Sigma}_\xi| - \frac{\partial}{\partial\gamma}tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{W}\right\} \qquad (4.30) \\
&= -mT\frac{\partial}{\partial\gamma}\log|\sigma_\omega^2\boldsymbol{\Gamma}| - \frac{\partial}{\partial\gamma}tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\mathbf{W}\right\} \\
&= -mT\times tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\frac{\partial\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)}{\partial\gamma}\right\} + tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\frac{\partial\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)}{\partial\gamma}\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\mathbf{W}\right\} \\
&= -mT\times tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\sigma_\omega^2\mathbf{I}\right\} + tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\sigma_\omega^2\mathbf{I}\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\mathbf{W}\right\} \\
&= -mT\times tr\left\{\boldsymbol{\Gamma}^{-1}\right\} + \frac{1}{\sigma_\omega^2}tr\left\{\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\}.
\end{aligned}
$$

The first partial derivative with respect to ν is derived as follows:

$$\begin{aligned}
\frac{\partial Q}{\partial \nu} &= -mT\frac{\partial}{\partial \nu}\log|\boldsymbol{\Sigma}_\xi| - \frac{\partial}{\partial \nu}tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{W}\right\} \qquad (4.31)\\
&= -mT\frac{\partial}{\partial \nu}\log|\sigma_\omega^2\boldsymbol{\Gamma}| - \frac{\partial}{\partial \nu}tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\mathbf{W}\right\}\\
&= -mT\times tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\frac{\partial\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)}{\partial \nu}\right\} + tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\frac{\partial\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)}{\partial \nu}\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\mathbf{W}\right\}\\
&= -mT\times tr\left\{\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial \nu}\right\} + \frac{1}{\sigma_\omega^2}tr\left\{\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial \nu}\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\},
\end{aligned}$$

where

$$\frac{\partial\boldsymbol{\Gamma}}{\partial \nu} = \begin{cases} \mathbf{0}, & \text{for diagonal elements}\\ \frac{\partial\boldsymbol{\rho}(\mathbf{D})}{\partial \nu}, & \text{for off-diagonal elements.}\end{cases}$$

This requires the first partial derivative of $\boldsymbol{\rho}(\mathbf{D})$ with respect to ν, which is equal to

$$\frac{\partial\boldsymbol{\rho}}{\partial \nu} = \left\{\left(\frac{2s}{d}\right)^\nu\Gamma(\nu+1)\right\}\times\frac{\partial}{\partial \nu}J_\nu\left(\frac{d}{s}\right) + J_\nu\left(\frac{d}{s}\right)\times\frac{\partial}{\partial \nu}\left\{\left(\frac{2s}{d}\right)^\nu\Gamma(\nu+1)\right\},$$

where

$$\begin{aligned}
\frac{\partial}{\partial \nu}\left\{\left(\frac{2s}{d}\right)^\nu\Gamma(\nu+1)\right\} &= \left\{\left(\frac{2s}{d}\right)^\nu\times\frac{\partial}{\partial \nu}\Gamma(\nu+1)\right\} + \left\{\Gamma(\nu+1)\times\frac{\partial}{\partial \nu}\left(\frac{2s}{d}\right)^\nu\right\}\\
&= \left(\frac{2s}{d}\right)^\nu\Gamma(\nu+1)\psi^{(0)}(\nu+1) + \Gamma(\nu+1)\left(\frac{2s}{d}\right)^\nu\log\left(\frac{2s}{d}\right)\\
&= \left(\frac{2s}{d}\right)^\nu\Gamma(\nu+1)\left\{\psi^{(0)}(\nu+1) + \log\left(\frac{2s}{d}\right)\right\}
\end{aligned}$$

and

$$\frac{\partial}{\partial \nu}J_\nu\left(\frac{d}{s}\right) = \sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\left[\frac{1}{\Gamma(k+\nu+1)}\times\frac{\partial}{\partial \nu}\left(\frac{d}{2s}\right)^{2k+\nu} + \left(\frac{d}{2s}\right)^{2k+\nu}\frac{\partial}{\partial \nu}\frac{1}{\Gamma(k+\nu+1)}\right]$$

$$
\begin{aligned}
&= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left[\frac{1}{\Gamma(k+\nu+1)} \times \left(\frac{d}{2s}\right)^{2k+\nu} \log\left(\frac{d}{2s}\right) - \left(\frac{d}{2s}\right)^{2k+\nu} \frac{\psi^{(0)}(k+\nu+1)}{\Gamma(k+\nu+1)} \right] \\
&= J_\nu\left(\frac{d}{s}\right) \log\left(\frac{d}{2s}\right) - \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(k+\nu+1)} \psi^{(0)}(k+\nu+1) \left(\frac{d}{2s}\right)^{2k+\nu}.
\end{aligned}
$$

$\psi^{(0)}$ denotes the digamma function. The first partial derivative of the Q-function with respect to s is:

$$
\begin{aligned}
\frac{\partial Q}{\partial s} &= -mT \frac{\partial}{\partial s} \log|\boldsymbol{\Sigma}_\xi| - \frac{\partial}{\partial s} tr\left\{\boldsymbol{\Sigma}_\xi^{-1}\mathbf{W}\right\} \qquad (4.32) \\
&= -mT \frac{\partial}{\partial s} \log|\sigma_\omega^2\boldsymbol{\Gamma}| - \frac{\partial}{\partial s} tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\mathbf{W}\right\} \\
&= -mT \times tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1} \frac{\partial\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)}{\partial s}\right\} + tr\left\{\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1} \frac{\partial\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)}{\partial s}\left(\sigma_\omega^2\boldsymbol{\Gamma}\right)^{-1}\mathbf{W}\right\} \\
&= -mT \times tr\left\{\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial s}\right\} + \frac{1}{\sigma_\omega^2} tr\left\{\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial s}\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\},
\end{aligned}
$$

where

$$
\frac{\partial\boldsymbol{\Gamma}}{\partial s} = \begin{cases} \mathbf{0}, & \text{for diagonal elements} \\ \frac{\partial\boldsymbol{\rho}(\mathbf{D})}{\partial s}, & \text{for off-diagonal elements} \end{cases}
$$

This requires the first partial derivative of $\boldsymbol{\rho}(\mathbf{D})$ with respect to s, derived as:

$$
\begin{aligned}
\frac{\partial\boldsymbol{\rho}}{\partial s} &= \Gamma(\nu+1)\left[\left(\frac{2s}{d}\right)^\nu \times \frac{\partial}{\partial s} J_\nu\left(\frac{d}{s}\right) + J_\nu\left(\frac{d}{s}\right) \times \frac{\partial}{\partial s}\left(\frac{2s}{d}\right)^\nu\right] \\
&= \Gamma(\nu+1)\left[\left(\frac{2s}{d}\right)^\nu \times \frac{\partial}{\partial s} J_\nu\left(\frac{d}{s}\right) + J_\nu\left(\frac{d}{s}\right) \nu\left(\frac{2s}{d}\right)^\nu s^{-1}\right] \\
&= \Gamma(\nu+1)\left(\frac{2s}{d}\right)^\nu \left[\frac{\partial}{\partial s} J_\nu\left(\frac{d}{s}\right) + \nu s^{-1} J_\nu\left(\frac{d}{s}\right)\right],
\end{aligned}
$$

where

$$\begin{aligned}
\frac{\partial}{\partial s}J_\nu\left(\frac{d}{s}\right) &= \sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\times\frac{\partial}{\partial s}\left(\frac{d}{2s}\right)^{2k+\nu} \\
&= \sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\left[-(2k+\nu)\left(\frac{d}{2s}\right)^{2k+\nu-1}\frac{d}{2}\left(\frac{1}{s}\right)^2\right] \\
&= -\frac{d}{2s^2}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\left[(2k+\nu)\left(\frac{d}{2s}\right)^{2k+\nu-1}\right].
\end{aligned}$$

The second partial derivative of the Q-function with respect to γ is:

$$\begin{aligned}
\frac{\partial^2 Q}{\partial\gamma^2} &= \frac{\partial}{\partial\gamma}\left\{-mT\times tr\left\{\Gamma^{-1}\right\}+\frac{1}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\Gamma^{-1}\mathbf{W}\right\}\right\} \qquad (4.33)\\
&= -mT\times\frac{\partial}{\partial\gamma}tr\left\{\Gamma^{-1}\right\}+\frac{1}{\sigma_\omega^2}\frac{\partial}{\partial\gamma}tr\left\{\Gamma^{-1}\Gamma^{-1}\mathbf{W}\right\} \\
&= mT\times tr\left\{-\Gamma^{-1}\frac{\partial\Gamma}{\partial\gamma}\Gamma^{-1}\right\}+\frac{1}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\times-\Gamma^{-1}\frac{\partial\Gamma}{\partial\gamma}\Gamma^{-1}\mathbf{W}-\Gamma^{-1}\frac{\partial\Gamma}{\partial\gamma}\Gamma^{-1}\times\Gamma^{-1}\mathbf{W}\right\} \\
&= mT\times tr\left\{\Gamma^{-1}\Gamma^{-1}\right\}-\frac{2}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\Gamma^{-1}\Gamma^{-1}\mathbf{W}\right\}.
\end{aligned}$$

The second partial derivative of the Q-function with respect to ν is:

$$\begin{aligned}
\frac{\partial^2 Q}{\partial\nu^2} &= -mT\times\frac{\partial}{\partial\nu}tr\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right\}+\frac{1}{\sigma_\omega^2}\frac{\partial}{\partial\nu}tr\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\mathbf{W}\right\} \qquad (4.34)\\
&= -mT\times tr\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu^2}-\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right\}-\frac{1}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\times\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\mathbf{W}\right\} \\
&\quad+\frac{1}{\sigma_\omega^2}tr\left\{\left[\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu^2}-\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right]\times\Gamma^{-1}\mathbf{W}\right\} \\
&= -mT\times tr\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu^2}-\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right\}-\frac{1}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\mathbf{W}\right\} \\
&\quad+\frac{1}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu^2}\Gamma^{-1}\mathbf{W}\right\}-\frac{1}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\mathbf{W}\right\} \\
&= -mT\times tr\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu^2}-\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right\}+\frac{1}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu^2}\Gamma^{-1}\mathbf{W}\right\} \\
&\quad-\frac{2}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\mathbf{W}\right\},
\end{aligned}$$

where

$$\frac{\partial^2 \mathbf{\Gamma}}{\partial \nu^2} = \begin{cases} \mathbf{0}, & \text{for diagonal elements} \\ \frac{\partial^2 \boldsymbol{\rho}(\mathbf{D})}{\partial \nu^2}, & \text{for off-diagonal elements.} \end{cases}$$

This requires the second derivative of $\boldsymbol{\rho}(\mathbf{D})$ with respect to s, derived as:

$$\begin{aligned}
\frac{\partial^2 \boldsymbol{\rho}}{\partial \nu^2} &= \frac{\partial}{\partial \nu}\left[\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\} \times \frac{\partial}{\partial \nu} J_{\nu}\left(\frac{d}{s}\right) + J_{\nu}\left(\frac{d}{s}\right) \times \frac{\partial}{\partial \nu}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\}\right] \\
&= \left[\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1) \times \frac{\partial^2}{\partial \nu^2} J_{\nu}\left(\frac{d}{s}\right)\right] + \left[\frac{\partial}{\partial \nu} J_{\nu}\left(\frac{d}{s}\right) \times \frac{\partial}{\partial \nu}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\}\right] \\
&\quad + \left[J_{\nu}\left(\frac{d}{s}\right) \times \frac{\partial^2}{\partial \nu^2}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\}\right] + \left[\frac{\partial}{\partial \nu}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\}\frac{\partial}{\partial \nu} J_{\nu}\left(\frac{d}{s}\right)\right] \\
&= \left[\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1) \times \frac{\partial^2}{\partial \nu^2} J_{\nu}\left(\frac{d}{s}\right)\right] + 2\left[\frac{\partial}{\partial \nu} J_{\nu}\left(\frac{d}{s}\right) \times \frac{\partial}{\partial \nu}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\}\right] \\
&\quad + \left[J_{\nu}\left(\frac{d}{s}\right) \times \frac{\partial^2}{\partial \nu^2}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\}\right] \\
&= \left[\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1) \times \frac{\partial^2}{\partial \nu^2} J_{\nu}\left(\frac{d}{s}\right)\right] + \left[J_{\nu}\left(\frac{d}{s}\right) \times \frac{\partial^2}{\partial \nu^2}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\}\right] \\
&\quad + 2\left[\frac{\partial}{\partial \nu} J_{\nu}\left(\frac{d}{s}\right)\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\left\{\psi^{(0)}(\nu+1) + \log\left(\frac{2s}{d}\right)\right\}\right] \\
&= \left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\left[\frac{\partial^2}{\partial \nu^2} J_{\nu}\left(\frac{d}{s}\right) + 2\frac{\partial}{\partial \nu} J_{\nu}\left(\frac{d}{s}\right)\left\{\psi^{(0)}(\nu+1) + \log\left(\frac{2s}{d}\right)\right\}\right] \\
&\quad + J_{\nu}\left(\frac{d}{s}\right)\frac{\partial^2}{\partial \nu^2}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial^2}{\partial \nu^2} J_{\nu}\left(\frac{d}{s}\right) &= \frac{\partial}{\partial \nu}\left\{J_{\nu}\left(\frac{d}{s}\right)\log\left(\frac{d}{2s}\right) - \sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\psi^{(0)}(k+\nu+1)\left(\frac{d}{2s}\right)^{2k+\nu}\right\} \\
&= \log\left(\frac{d}{2s}\right)\frac{\partial}{\partial \nu} J_{\nu}\left(\frac{d}{s}\right) + \sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\left[\frac{\partial}{\partial \nu}\frac{1}{\Gamma(k+\nu+1)}\left(\frac{d}{2s}\right)^{2k+\nu}\right] \\
&= \log\left(\frac{d}{2s}\right)\frac{\partial}{\partial \nu} J_{\nu}\left(\frac{d}{s}\right) + \sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\left[\frac{\partial}{\partial \nu}\frac{1}{\Gamma(k+\nu+1)} \times \left(\frac{d}{2s}\right)^{2k+\nu}\log\left(\frac{d}{2s}\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&+\left(\frac{d}{2s}\right)^{2k+\nu}\frac{\partial^2}{\partial\nu^2}\frac{1}{\Gamma(k+\nu+1)}\Bigg] \\
&= \log\left(\frac{d}{2s}\right)\frac{\partial}{\partial\nu}J_\nu\left(\frac{d}{s}\right)+\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\left[-\frac{\psi^{(0)}(k+\nu+1)}{\Gamma(k+\nu+1)}\left(\frac{d}{2s}\right)^{2k+\nu}\log\left(\frac{d}{2s}\right)\right. \\
&\left.+\left(\frac{d}{2s}\right)^{2k+\nu}\frac{\psi^{(0)}(k+\nu+1)^2-\psi^{(1)}(k+\nu+1)}{\Gamma(k+\nu+1)}\right] \\
&= \log\left(\frac{d}{2s}\right)\frac{\partial}{\partial\nu}J_\nu\left(\frac{d}{s}\right)+\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\left(\frac{d}{2s}\right)^{2k+\nu}\left[-\psi^{(0)}(k+\nu+1)\log\left(\frac{d}{2s}\right)\right. \\
&\left.+\psi^{(0)}(k+\nu+1)^2-\psi^{(1)}(k+\nu+1)\right].
\end{aligned}
$$

$\psi^{(1)}$ denotes the first derivative of the digamma function. Furthermore,

$$
\begin{aligned}
\frac{\partial^2}{\partial\nu^2}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\} &= \frac{\partial}{\partial\nu}\left[\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\left\{\psi^{(0)}(\nu+1)+\log\left(\frac{2s}{d}\right)\right\}\right] \\
&= \left[\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\times\frac{\partial}{\partial\nu}\left\{\psi^{(0)}(\nu+1)+\log\left(\frac{2s}{d}\right)\right\}\right] \\
&\quad+\left[\left\{\psi^{(0)}(\nu+1)+\log\left(\frac{2s}{d}\right)\right\}\times\frac{\partial}{\partial\nu}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\right\}\right] \\
&= \left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\psi^{(1)}(\nu+1)+\left\{\psi^{(0)}(\nu+1)+\log\left(\frac{2s}{d}\right)\right\} \\
&\quad\times\left[\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\left\{\psi^{(0)}(\nu+1)+\log\left(\frac{2s}{d}\right)\right\}\right].
\end{aligned}
$$

The second partial derivative of the Q-function with respect to s is derived as:

$$
\begin{aligned}
\frac{\partial^2 Q}{\partial s^2} &= -mT\times tr\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial s^2}-\Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\frac{\partial\Gamma}{\partial s}\right\}+\frac{1}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial s^2}\Gamma^{-1}\mathbf{W}\right\} \\
&\quad-\frac{2}{\sigma_\omega^2}tr\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\mathbf{W}\right\},
\end{aligned}
\tag{4.35}
$$

where

$$\frac{\partial^2 \boldsymbol{\Gamma}}{\partial s^2} = \begin{cases} \mathbf{0}, & \text{for diagonal elements} \\ \frac{\partial^2 \boldsymbol{\rho}(\mathbf{D})}{\partial s^2}, & \text{for off-diagonal elements.} \end{cases}$$

This requires the second derivative of $\boldsymbol{\rho}(\mathbf{D})$, which is derived as:

$$\begin{aligned} \frac{\partial^2 \rho}{\partial s^2} &= \frac{\partial}{\partial s}\left\{\Gamma(\nu+1)\left(\frac{2s}{d}\right)^{\nu}\left[\frac{\partial}{\partial s}J_{\nu}\left(\frac{d}{s}\right)+\nu s^{-1}J_{\nu}\left(\frac{d}{s}\right)\right]\right\} \\ &= \Gamma(\nu+1)\left(\frac{2}{d}\right)^{\nu}\left[\frac{\partial}{\partial s}\left\{s^{\nu}\times\frac{\partial}{\partial s}J_{\nu}\left(\frac{d}{s}\right)+\nu s^{\nu-1}J_{\nu}\left(\frac{d}{s}\right)\right\}\right] \\ &= \Gamma(\nu+1)\left(\frac{2}{d}\right)^{\nu}\left[s^{\nu}\times\frac{\partial^2}{\partial s^2}J_{\nu}\left(\frac{d}{s}\right)+\nu s^{\nu-1}\frac{\partial}{\partial s}J_{\nu}\left(\frac{d}{s}\right)+\nu s^{\nu-1}\frac{\partial}{\partial s}J_{\nu}\left(\frac{d}{s}\right)\right. \\ &\quad \left.+\nu(\nu-1)s^{\nu-2}J_{\nu}\left(\frac{d}{s}\right)\right] \\ &= \Gamma(\nu+1)\left(\frac{2}{d}\right)^{\nu}\left[s^{\nu}\times\frac{\partial^2}{\partial s^2}J_{\nu}\left(\frac{d}{s}\right)+2\nu s^{\nu-1}\frac{\partial}{\partial s}J_{\nu}\left(\frac{d}{s}\right)+\nu(\nu-1)s^{\nu-2}J_{\nu}\left(\frac{d}{s}\right)\right], \end{aligned}$$

where

$$\begin{aligned} \frac{\partial^2}{\partial s^2}J_{\nu}\left(\frac{d}{s}\right) &= \frac{\partial}{\partial s}\left\{-\frac{d}{2s^2}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\left[(2k+\nu)\left(\frac{d}{2s}\right)^{2k+\nu-1}\right]\right\} \\ &= -\frac{d}{2}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\left[(2k+\nu)\left(\frac{d}{2}\right)^{2k+\nu-1}\frac{\partial}{\partial s}\left(\frac{1}{s}\right)^{2k+\nu+1}\right] \\ &= \frac{d}{2}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\left[(2k+\nu)(2k+\nu+1)\left(\frac{d}{2}\right)^{2k+\nu-1}\left(\frac{1}{s}\right)^{2k+\nu+2}\right]. \end{aligned}$$

The second partial derivative of the Q-function with respect to γ and ν is derived as:

$$\begin{aligned} \frac{\partial^2 Q}{\partial\gamma\partial\nu} &= -m\tau\times\frac{\partial}{\partial\nu}\mathrm{tr}\left\{\boldsymbol{\Gamma}^{-1}\right\}+\frac{1}{\sigma_{\omega}^2}\times\frac{\partial}{\partial\nu}\mathrm{tr}\left\{\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\} \\ &= -m\tau\times\frac{\partial}{\partial\nu}\mathrm{tr}\left\{\boldsymbol{\Gamma}^{-1}\right\}+\frac{1}{\sigma_{\omega}^2}\mathrm{tr}\left\{-\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial\nu}\boldsymbol{\Gamma}^{-1}\mathbf{W}-\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial\nu}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}^{-1}\mathbf{W}\right\} \end{aligned} \tag{4.36}$$

$$= \quad m\mathsf{T} \times \mathrm{tr}\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\right\} - \frac{2}{\sigma_\omega^2}\mathrm{tr}\left\{\Gamma^{-1}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\mathbf{W}\right\}.$$

The second partial derivative of the Q-function with respect to γ and s is derived as:

$$\begin{aligned}\frac{\partial^2 Q}{\partial\gamma\partial s} &= \frac{\partial}{\partial s}\left\{-m\mathsf{T} \times \mathrm{tr}\left\{\Gamma^{-1}\right\} + \frac{1}{\sigma_\omega^2}\mathrm{tr}\left\{\Gamma^{-1}\Gamma^{-1}\mathbf{W}\right\}\right\} \\ &= -m\mathsf{T} \times \frac{\partial}{\partial s}\mathrm{tr}\left\{\Gamma^{-1}\right\} + \frac{1}{\sigma_\omega^2}\frac{\partial}{\partial s}\mathrm{tr}\left\{\Gamma^{-1}\Gamma^{-1}\mathbf{W}\right\} \\ &= m\mathsf{T} \times \mathrm{tr}\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\right\} - \frac{2}{\sigma_\omega^2}\mathrm{tr}\left\{\Gamma^{-1}\Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\mathbf{W}\right\}.\end{aligned} \tag{4.37}$$

The second partial derivative of the Q-function with respect to ν and s is derived as:

$$\begin{aligned}\frac{\partial^2 Q}{\partial\nu\partial s} &= \frac{\partial}{\partial s}\left\{-m\mathsf{T} \times \mathrm{tr}\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right\} + \frac{1}{\sigma_\omega^2}\mathrm{tr}\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\mathbf{W}\right\}\right\} \\ &= -m\mathsf{T} \times \mathrm{tr}\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu\partial s} - \Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right\} - \frac{1}{\sigma_\omega^2}\mathrm{tr}\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\mathbf{W}\right\} \\ &\quad +\mathrm{tr}\left\{\left[\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu\partial s} - \Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right]\Gamma^{-1}\mathbf{W}\right\} \\ &= -m\mathsf{T} \times \mathrm{tr}\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu\partial s} - \Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\right\} + \frac{1}{\sigma_\omega^2}\mathrm{tr}\left\{\Gamma^{-1}\frac{\partial^2\Gamma}{\partial\nu\partial s}\Gamma^{-1}\mathbf{W}\right\} \\ &\quad -\frac{2}{\sigma_\omega^2}\mathrm{tr}\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\nu}\Gamma^{-1}\frac{\partial\Gamma}{\partial s}\Gamma^{-1}\mathbf{W}\right\},\end{aligned} \tag{4.38}$$

where

$$\frac{\partial^2\Gamma}{\partial\nu\partial s} = \begin{cases} 0, & \text{for diagonal elements} \\ \frac{\partial^2\rho(\mathbf{D})}{\partial\nu\partial s}, & \text{for off-diagonal elements}\end{cases}$$

and

$$\begin{aligned}\frac{\partial^2 \boldsymbol{\rho}(\mathbf{D})}{\partial \nu \partial s} &= \frac{\partial}{\partial s}\left\{\left(\frac{2s}{d}\right)^{\nu}\Gamma(\nu+1)\left\{\frac{\partial}{\partial \nu}J_{\nu}\left(\frac{d}{s}\right)+J_{\nu}\left(\frac{d}{s}\right)\left[\psi^{(0)}(\nu+1)+\log\left(\frac{2s}{d}\right)\right]\right\}\right\} \\ &= \Gamma(\nu+1)\left[\frac{\partial}{\partial s}\left\{\left(\frac{2s}{d}\right)^{\nu}\frac{\partial}{\partial \nu}J_{\nu}\left(\frac{d}{s}\right)\right\}+\psi^{(0)}(\nu+1)\frac{\partial}{\partial s}\left\{\left(\frac{2s}{d}\right)^{\nu}J_{\nu}\left(\frac{d}{s}\right)\right\}\right. \\ &\quad \left.+\frac{\partial}{\partial s}\left\{\left(\frac{2s}{d}\right)^{\nu}J_{\nu}\left(\frac{d}{s}\right)\log\left(\frac{2s}{d}\right)\right\}\right].\end{aligned}$$

This is further defined as

$$\frac{\partial}{\partial s}\left\{\left(\frac{2s}{d}\right)^{\nu}\frac{\partial}{\partial \nu}J_{\nu}\left(\frac{d}{s}\right)\right\} = \left(\frac{2s}{d}\right)^{\nu}\frac{\partial^2}{\partial \nu \partial s}J_{\nu}\left(\frac{d}{s}\right)+\nu s^{\nu-1}\left(\frac{2}{d}\right)^{\nu}\frac{\partial}{\partial \nu}J_{\nu}\left(\frac{d}{s}\right),$$

where

$$\begin{aligned}\frac{\partial^2 J_{\nu}\left(\frac{d}{s}\right)}{\partial \nu \partial s} &= \frac{\partial}{\partial s}\left\{J_{\nu}\left(\frac{d}{s}\right)\log\left(\frac{d}{2s}\right)-\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\psi^{(0)}(k+\nu+1)\left(\frac{d}{2s}\right)^{2k+\nu}\right\} \\ &= \log\left(\frac{d}{2s}\right)\frac{\partial}{\partial s}\left\{J_{\nu}\left(\frac{d}{s}\right)\right\}+\frac{1}{s}J_{\nu}\left(\frac{d}{s}\right) \\ &\quad -\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\psi^{(0)}(k+\nu+1)\left(\frac{d}{2}\right)^{2k+\nu}\frac{\partial}{\partial s}\left(\frac{1}{s}\right)^{2k+\nu} \\ &= \log\left(\frac{d}{2s}\right)\frac{\partial}{\partial s}\left\{J_{\nu}\left(\frac{d}{s}\right)\right\}+\frac{1}{s}J_{\nu}\left(\frac{d}{s}\right) \\ &\quad +\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+\nu+1)}\psi^{(0)}(k+\nu+1)\left(\frac{d}{2}\right)^{2k+\nu}(2k+\nu)\left(\frac{1}{s}\right)^{2k+\nu+1}.\end{aligned}$$

Furthermore,

$$\frac{\partial}{\partial s}\left\{\left(\frac{2s}{d}\right)^{\nu}J_{\nu}\left(\frac{d}{s}\right)\right\} = \left(\frac{2s}{d}\right)^{\nu}\frac{\partial}{\partial s}J_{\nu}\left(\frac{d}{s}\right)+\nu s^{\nu-1}\left(\frac{2}{d}\right)^{\nu}J_{\nu}\left(\frac{d}{s}\right)$$

and

$$\frac{\partial}{\partial s}\left\{\left(\frac{2s}{d}\right)^{\nu}J_{\nu}\left(\frac{d}{s}\right)\log\left(\frac{2s}{d}\right)\right\} = \frac{1}{s}\left(\frac{2s}{d}\right)^{\nu}J_{\nu}\left(\frac{d}{s}\right)+\log\left(\frac{2s}{d}\right)\frac{\partial}{\partial s}\left\{\left(\frac{2s}{d}\right)^{\nu}J_{\nu}\left(\frac{d}{s}\right)\right\}.$$

4.4.3 Initial Values

Iterative procedures such as the ECM require initial values to start the algorithm. As previously noted, the ECM algorithm converges to a stationary point, which can be a global maximum, local maximum, or saddle point. It is often advised to check results by running the algorithm with several sets of values. However, running the algorithm several times comes with challenges since spatiotemporal models require large computation times. To increase the likelihood that the derived solution is a global maximum, an algorithm for selecting initial values is proposed.

4.4.3.1 Selection of Initial Values

The method of moments is used to select initial values to achieve faster convergence and increase the likelihood that the algorithm converges to the global maximum (Xu and Wikle, 2007). For parameters $\boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \mathbf{G}, \sigma^2_\omega, \sigma^2_\varepsilon, \sigma^2_\eta$, the following steps are used to select initial values:

1. $\boldsymbol{\beta}^{(0)}$

   (a) Run the linear regression model $\mathbf{z}_{ij} = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_{ij}$ for each fMRI region $j = 1, 2, \ldots, n$ to obtain the estimate $\hat{\boldsymbol{\beta}}_j^{(0)}$.

   (b) Set $\boldsymbol{\beta}^{(0)} = \frac{1}{n}\sum_{j=1}^{n} \hat{\boldsymbol{\beta}}_j^{(0)}$.

2. $\boldsymbol{\mu}_0^{(0)}$

   (a) Estimate the latent variable at $t = 1$ by $\mathbf{y}_{i1}^{(0)} = \mathbf{K}_i^{-1}\left(\mathbf{z}_{i1} - \mathbf{X}_i\boldsymbol{\beta}^{(0)}\right)$.

(b) Set $\boldsymbol{\mu}_0^{(0)} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{y}_{i1}^{(0)}$.

3. $\boldsymbol{\Sigma}_0^{(0)}$

(a) Set $\boldsymbol{\Sigma}_0^{(0)} = \text{diag}\left(\sum_{i=1}^{m}\left(\mathbf{y}_{i1}^{(0)} - \boldsymbol{\mu}_0^{(0)}\right)\left(\mathbf{y}_{i1}^{(0)} - \boldsymbol{\mu}_0^{(0)}\right)'\right)$.

4. $\mathbf{G}^{(0)}$

(a) For $t = 2, \ldots, T$, estimate the latent variable at time t by $\mathbf{y}_{it}^{(0)} = \mathbf{K}_i^{-1}\left(\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta}^{(0)}\right)$.

Run a linear regression model $\mathbf{y}_{it}^{(0)} = \mathbf{G}_t\mathbf{y}_{i(t-1)}^{(0)} + \boldsymbol{\eta}_{it}$ and obtain $\widehat{\mathbf{G}}_t^{(0)}$.

(b) Set $\mathbf{G}^{(0)} = \frac{1}{T-1}\sum_{t=2}^{T}\widehat{\mathbf{G}}_t^{(0)}$.

5. $\sigma_\omega^{2(0)}, \sigma_\varepsilon^{2(0)}$

(a) For $t = 1, \ldots, T$, calculate $\boldsymbol{\xi}_{it}^{(0)} = \mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta}^{(0)} - \mathbf{K}_i\mathbf{y}_{it}$.

(b) Calculate $\bar{\boldsymbol{\xi}}^{(0)} = \frac{1}{mT}\sum_{i=1}^{m}\sum_{t=1}^{T}\boldsymbol{\xi}_{it}^{(0)}$.

(c) Let $\boldsymbol{\Sigma}_\xi^{(0)} = \frac{1}{mT}\sum_{i=1}^{m}\sum_{t=1}^{T}\left(\boldsymbol{\xi}_{it}^{(0)} - \bar{\boldsymbol{\xi}}^{(0)}\right)\left(\boldsymbol{\xi}_{it}^{(0)} - \bar{\boldsymbol{\xi}}^{(0)}\right)'$.

(d) Set $\sigma_\omega^{2(0)} = \sigma_\varepsilon^{2(0)} = \frac{1}{2n}\sum_{j=1}^{n}\boldsymbol{\Sigma}_{\xi(jj)}^{(0)}$.

6. $\boldsymbol{\sigma}_\eta^{2(0)}$

(a) For $t = 2, \ldots, T$, calculate $\boldsymbol{\eta}_{it}^{(0)} = \mathbf{y}_{it} - \mathbf{G}^{(0)}\mathbf{y}_{i(t-1)}$.

(b) Calculate $\bar{\boldsymbol{\eta}}^{(0)} = \frac{1}{m(T-1)}\sum_{i=1}^{m}\sum_{t=2}^{T}\boldsymbol{\eta}_{it}^{(0)}$.

(c) Set $\boldsymbol{\Sigma}_\eta^{(0)} = \frac{1}{m(T-1)}\sum_{i=1}^{m}\sum_{t=2}^{T}\left(\boldsymbol{\eta}_{it}^{(0)} - \bar{\boldsymbol{\eta}}^{(0)}\right)\left(\boldsymbol{\eta}_{it}^{(0)} - \bar{\boldsymbol{\eta}}^{(0)}\right)'$.

(d) Set $\sigma_\eta^{2(0)} = \frac{1}{r}\sum_{k=1}^{r}\boldsymbol{\Sigma}_{\eta(kk)}^{(0)}$.

Since there is no way to measure the relative contribution of $\sigma_{\omega}^{2(0)}$ and $\sigma_{\varepsilon}^{2(0)}$ to the diagonal elements of $\boldsymbol{\Sigma}_{\xi}$, they are each assigned $1/2$ of the mean of the diagonal elements of $\boldsymbol{\Sigma}_{\xi}^{(0)}$. Furthermore, it is difficult to derive closed-form moment based estimators for Bessel function parameters ν and s. To obtain initial values for these two parameters, the sample mean, variance, skewness, and kurtosis of the upper triangle elements of the matrix $\boldsymbol{\rho}$ are estimated. The sum of squared differences between the sample and empirical moments for various combinations of ν and s are calculated. The parameters of the Bessel function with the minimum sum of squared differences is selected as the starting values.

A simulation study was performed to evaluate the accuracy of these initial values. Results are provided in Table III. All method of moments estimators are close to the true values; however, the estimates for $\nu^{(0)}$, $s^{(0)}$, and $\gamma^{(0)} = \sigma_{\varepsilon}^{2(0)}/\sigma_{\omega}^{2(0)}$ are not. These parameters are estimated by Newton-Raphson in step 2 of CM. It is well known that the convergence of Newton's method depends on the shape of the function maximized and initial values. If initial values are far from the true values, performance of Newton-Raphson is poor. Performance may sometimes be remedied by taking smaller steps when the gradient, denoted $\frac{\partial Q\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(k)}\right)}{\partial \boldsymbol{\Psi}}$, is large. Smaller steps are achieved by modifying the updating equation to

$$\boldsymbol{\Psi}^{(k+1)} = \boldsymbol{\Psi}^{(k)} - a \left[\frac{\partial^2 Q\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(k)}\right)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'} + \frac{\partial Q\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(k)}\right)}{\partial \boldsymbol{\Psi}} \right]^{-1}_{\boldsymbol{\Psi}=\boldsymbol{\Psi}^{(k)}} \left[\frac{\partial Q\left(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(k)}\right)}{\partial \boldsymbol{\Psi}} \right]_{\boldsymbol{\Psi}=\boldsymbol{\Psi}^{(k)}}, \quad (4.39)$$

where $0 < a < 1$ (Givens and Hoeting, 2013). This modification was incorporated into step 2 of CM. However, consistent convergence toward the global maximum was not observed.

4.4.3.2 Updating Initial Values

The initial values for ν, s and γ are further updated to improve performance of the estimating procedure. The updated initial values are then used as the starting values for the ECM algorithm. Initial values are obtained by applying a modified ECM algorithm: the procedure is applied in multiple iterations with two out of the three parameters estimated via Newton-Raphson fixed. For a given c, the algorithm is as follows:

1. With ν and s fixed at $\nu^{(0)}$ and $s^{(0)}$, respectively, update remaining parameters using the ECM algorithm until $\left|\frac{\partial Q}{\partial \gamma^{(k)}}\right| < c$. Denote the final estimates as $\boldsymbol{\Psi}^{(0_1')}$.
2. With γ and ν fixed at $\gamma^{(0_1')}$ and $\nu^{(0)}$, respectively, update remaining parameters using the ECM algorithm until $\left|\frac{\partial Q}{\partial s^{(k)}}\right| < c$. Denote the final estimates as $\boldsymbol{\Psi}^{(0_2')}$.
3. With γ and s fixed at $\gamma^{(0_1')}$ and $s^{(0_2')}$, respectively, update remaining parameters using the ECM algorithm until $\left|\frac{\partial Q}{\partial \nu^{(k)}}\right| < c$. Denote the final estimates as $\boldsymbol{\Psi}^{(0_3')}$.
4. Set the updated initial values to $\boldsymbol{\Psi}^{(0')} = \left\{\boldsymbol{\beta}^{(0_3')}, \boldsymbol{\mu}_0^{(0_3')}, \mathbf{G}^{(0_3')}, \sigma_\omega^{2(0_3')}, \sigma_\varepsilon^{2(0_3')}, \sigma_\eta^{2(0_3')}, \gamma^{(0_1')}, s^{(0_2')}, \nu^{(0_3')}\right\}$.

A simulation study of the estimating procedure illustrates that convergence to the global maximum is observed when $\boldsymbol{\Psi}^{(0)} = \boldsymbol{\Psi}^{(0')}$.

4.4.4 Simulation Studies

A simulation study was performed to evaluate the accuracy and reliability of the parameter estimation procedure. Simulations were performed in R version 3.6.1 and code was modified from functions in the Stem package (Cameletti, 2012). 500 datasets were simulated for sample

TABLE III INITIAL VALUES, DIAGONAL COVARIANCE

Parameter	True Value	Mean (SE)	Middle 95%
β_0	3.65	3.6498 (0.0478)	(3.5633,3.7501)
β_1	0.046	0.0460 (0.0017)	(0.0426,0.0491)
β_2	0.904	0.9039 (0.0098)	(0.8864,0.9233)
β_3	-2	-2.0004 (0.0197)	(-2.0348,-1.9600)
σ^2_ϵ	0.3	0.3961 (0.0028)	(0.3906,0.4014)
σ^2_ω	0.5	0.3961 (0.0028)	(0.3906,0.4014)
σ^2_η	0.25	0.2502 (0.0028)	(0.2450,0.2560)
ν	0.6	1.4495 (0.4751)	(0.4745,2.0000)
s	0.2	0.1768 (0.0225)	(0.1500,0.2300)
G_{11}	0.4	0.4007 (0.0116)	(0.3786,0.4251)
G_{12}	0.2	0.1991 (0.0115)	(0.1776,0.2192)
G_{13}	0.2	0.2005 (0.0114)	(0.1765,0.2243)
G_{21}	0.2	0.2016 (0.0108)	(0.1805,0.2205)
G_{22}	0.4	0.3996 (0.0112)	(0.3762,0.4199)
G_{23}	0.2	0.2000 (0.0119)	(0.1755,0.2223)
G_{31}	0.2	0.2004 (0.0109)	(0.1814,0.2210)
G_{32}	0.2	0.2007 (0.0111)	(0.1774,0.2219)
G_{33}	0.4	0.3997 (0.0112)	(0.3765,0.4198)

TABLE IV ESTIMATING PROCEDURE SIMULATION RESULTS

Parameter	True Value	m	Mean (SE)	Bias	RMSE	Middle 95%
β_0	3.65	100	3.6495 (0.0288)	-0.0005	0.0288	(3.5908,3.7039)
		200	3.6483 (0.0215)	-0.0017	0.0216	(3.6066,3.6879)
		400	3.6492 (0.0150)	-0.0008	0.0150	(3.6186,3.6781)
β_1	0.046	100	0.0460 (0.0011)	3.77e-05	0.0011	(0.0439,0.0481)
		200	0.0460 (0.0008)	3.45e-05	0.0008	(0.0446,0.0476)
		400	0.0460 (0.0005)	3.00e-05	0.0005	(0.0450,0.0472)
β_2	0.904	100	0.9040 (0.0066)	4.17e-05	0.0066	(0.8900,0.9157)
		200	0.9039 (0.0042)	-0.0001	0.0042	(0.8954,0.9118)
		400	0.9038 (0.0030)	-0.0002	0.0030	(0.8980,0.9096)
β_3	-2	100	-1.9999 (0.0126)	0.0001	0.0126	(-2.0215,-1.9764)
		200	-1.9992 (0.0080)	0.0008	0.0080	(-2.0158,-1.9829)
		400	-2.0001 (0.0063)	-0.0001	0.0063	(-2.0126,-1.9885)
G_{11}	0.4	100	0.3879 (0.0487)	-0.0121	0.0501	(0.2939,0.4815)
		200	0.3969 (0.0354)	-0.0031	0.0355	(0.3225,0.4652)
		400	0.3925 (0.0255)	-0.0075	0.0265	(0.3369,0.4378)
G_{12}	0.2	100	0.2023 (0.0366)	0.0023	0.0366	(0.1307,0.2702)
		200	0.1990 (0.0272)	-0.0010	0.0272	(0.1461,0.2568)
		400	0.2005 (0.0182)	0.0005	0.0182	(0.1637,0.2385)
G_{13}	0.2	100	0.2003 (0.0313)	0.0003	0.0312	(0.1396,0.2596)
		200	0.1992 (0.0224)	-0.0008	0.0224	(0.1592,0.2448)
		400	0.1993 (0.0163)	-0.0007	0.0163	(0.1665,0.2309)
G_{21}	0.2	100	0.2028 (0.0374)	0.0028	0.0374	(0.1328,0.2735)
		200	0.1984 (0.0274)	-0.0016	0.0274	(0.1463,0.2533)
		400	0.2004 (0.0187)	0.0004	0.0187	(0.1648,0.2415)
G_{22}	0.4	100	0.3923 (0.0491)	-0.0077	0.0496	(0.2929,0.4773)
		200	0.3957 (0.0349)	-0.0043	0.0352	(0.3245,0.4614)
		400	0.3948 (0.0237)	-0.0052	0.0242	(0.3433,0.4371)

TABLE IV ESTIMATING PROCEDURE SIMULATION RESULTS (Continued)

Parameter	True Value	m	Mean (SE)	Bias	RMSE	Middle 95%
G_{23}	0.2	100	0.1980 (0.0316)	-0.0020	0.0317	(0.1418,0.2570)
		200	0.1989 (0.0216)	-0.0011	0.0216	(0.1575,0.2448)
		400	0.1990 (0.0159)	-0.0010	0.0159	(0.1715,0.2331)
G_{31}	0.2	100	0.2026 (0.0330)	0.0026	0.0331	(0.1370,0.2655)
		200	0.2018 (0.0241)	0.0018	0.0242	(0.1596,0.2491)
		400	0.2016 (0.0179)	0.0016	0.0179	(0.1692,0.2370)
G_{32}	0.2	100	0.1990 (0.0336)	-0.0010	0.0336	(0.1365,0.2717)
		200	0.2011 (0.0232)	0.0011	0.0232	(0.1583,0.2467)
		400	0.2000 (0.0176)	-0.0000	0.0176	(0.1674,0.2330)
G_{33}	0.4	100	0.3952 (0.0386)	-0.0048	0.0389	(0.3178,0.4662)
		200	0.3973 (0.0285)	-0.0027	0.0286	(0.3410,0.4471)
		400	0.3981 (0.0209)	-0.0019	0.0210	(0.3605,0.4394)
ν	0.6	100	0.6254 (0.1686)	0.0254	0.1703	(0.3278,0.9790)
		200	0.5992 (0.1128)	-0.0008	0.1127	(0.4063,0.8392)
		400	0.6014 (0.0777)	0.0014	0.0777	(0.4597,0.7732)
s	0.2	100	0.1992 (0.0114)	-0.0008	0.0114	(0.1787,0.2210)
		200	0.2004 (0.0078)	0.0004	0.0078	(0.1852,0.2150)
		400	0.2001 (0.0055)	0.0001	0.0055	(0.1890,0.2105)
σ^2_ϵ	0.3	100	0.2999 (0.0027)	-0.0001	0.0027	(0.2947,0.3053)
		200	0.2999 (0.0018)	-0.0001	0.0018	(0.2964,0.3034)
		400	0.3001 (0.0013)	0.0001	0.0013	(0.2974,0.3024)
σ^2_η	0.25	100	0.2522 (0.0155)	0.0022	0.0156	(0.2227,0.2819)
		200	0.2498 (0.0105)	-0.0002	0.0105	(0.2283,0.2721)
		400	0.2511 (0.0076)	0.0011	0.0077	(0.2354,0.2672)
σ^2_ω	0.5	100	0.4999 (0.0041)	-0.0001	0.0041	(0.4925,0.5074)
		200	0.4998 (0.0029)	-0.0002	0.0029	(0.4944,0.5053)
		400	0.4997 (0.0019)	-0.0003	0.0019	(0.4960,0.5032)

sizes of 100, 200, and 400. Each individual's simulated observations consisted of thirty regions across sixty timepoints and were generated according to Equation 3.4 and Equation 3.5. Three covariates and an intercept were included, where $X_1 \sim N(26.59, 32.33)$ represents age, $X_2 \sim N(0, 1)$ represents standardized IQ scores, and $X_3 \sim Bin(0.5)$ represents gender. These parameters were derived from the CMU dataset in ABIDE I. For the latent process, the dimension was reduced to $r = 3$, or 10% of the spatial dimension. The elements of the adjacency matrix were set to $\exp\left(-0.5 \times d_{jj'}\right)$, where $d_{jj'}$ is the functional distance between regions j and j' for $j = 1, 2, \ldots, n$ and $j' = 1, 2, \ldots, n$. Functional distances were also derived from the CMU dataset.

The performance of the estimating procedure is evaluated using bias, standard error (SE), root mean squared error (RMSE), and the 0.025 and 0.975 quantiles of the estimates. The results of the simulation are provided in Table IV. Bias across all parameters are small. As a result, the RMSE is very close to the SE. As the sample size increases, bias decreases, RMSE decreases, and the 0.025 and 0.975 quantiles of the estimates are closer to the true value for all parameters. This simulation study therefore demonstrates the accuracy of the ECM algorithm for the proposed model.

Additional simulation studies were performed to evaluate the impact of misspecification of r, the dimension of the first-order latent dynamic model. Data was simulated based on the specifications described for the previous simulation study for sample sizes of 100 and 200. However, for estimation, r was underestimated at 2 and overestimated at 4. The results of the simulation are provided in Table V and Table VI. When r is underestimated, the estimates

TABLE V UNDERESTIMATING LATENT DIMENSION SIMULATION

Parameter	True Value	m	Mean (SE)	Bias	RMSE	Middle 95%
β_0	3.65	100	3.6509 (0.0315)	0.0009	0.0315	(3.5961,3.7127)
		200	3.6494 (0.0230)	-0.0006	0.0230	(3.6052,3.6948)
β_1	0.046	100	0.0460 (0.0011)	-0.0000	0.0011	(0.0439,0.0480)
		200	0.0460 (0.0008)	0.0000	0.0008	(0.0445,0.0475)
β_2	0.904	100	0.9036 (0.0070)	-0.0004	0.0070	(0.8901,0.9198)
		200	0.9039 (0.0048)	-0.0001	0.0048	(0.8934,0.9119)
β_3	-2	100	-2.0001 (0.0129)	-0.0001	0.0129	(-2.0260,-1.9765)
		200	-1.9999 (0.0092)	0.0001	0.0092	(-2.0182,-1.9815)
G_{11}	0.4	100	0.4557 (0.0449)	0.0557	0.0716	(0.3611,0.5372)
		200	0.4554 (0.0328)	0.0554	0.0643	(0.3867,0.5164)
G_{12}	0.2	100	0.2664 (0.0351)	0.0664	0.0751	(0.1969,0.3309)
		200	0.2653 (0.0238)	0.0653	0.0695	(0.2242,0.3122)
G_{21}	0.2	100	0.2657 (0.0360)	0.0657	0.0749	(0.1983,0.3374)
		200	0.2672 (0.0256)	0.0672	0.0719	(0.2183,0.3212)
G_{22}	0.4	100	0.4568 (0.0452)	0.0568	0.0726	(0.3648,0.5421)
		200	0.4560 (0.0313)	0.0560	0.0641	(0.3922,0.5140)
ν	0.6	100	1.4888 (0.2887)	0.8888	0.9344	(1.0377,2.1507)
		200	1.4779 (0.1868)	0.8779	0.8975	(1.1565,1.8932)
s	0.2	100	0.1544 (0.0095)	-0.0456	0.0466	(0.1351,0.1718)
		200	0.1543 (0.0063)	-0.0457	0.0461	(0.1415,0.1664)
σ^2_ϵ	0.3	100	0.2858 (0.0029)	-0.0142	0.0145	(0.2796,0.2915)
		200	0.2856 (0.0020)	-0.0144	0.0145	(0.2819,0.2896)
σ^2_η	0.25	100	0.2715 (0.0244)	0.0215	0.0325	(0.2277,0.3195)
		200	0.2725 (0.0173)	0.0225	0.0284	(0.2407,0.3074)
σ^2_ω	0.5	100	0.5342 (0.0048)	0.0342	0.0345	(0.5250,0.5432)
		200	0.5346 (0.0034)	0.0346	0.0348	(0.5282,0.5410)

TABLE VI OVERESTIMATING LATENT DIMENSION SIMULATION

Parameter	True Value	m	Mean (SE)	Bias	RMSE	Middle 95%
β_0	3.65	100	3.6508 (0.0294)	0.0008	0.0294	(3.5974,3.7081)
		200	3.6488 (0.0218)	-0.0012	0.0218	(3.6051,3.6909)
β_1	0.046	100	0.0460 (0.0010)	-0.0000	0.0010	(0.0440,0.0479)
		200	0.0460 (0.0008)	0.0000	0.0008	(0.0445,0.0476)
β_2	0.904	100	0.9033 (0.0063)	-0.0007	0.0063	(0.8914,0.9153)
		200	0.9039 (0.0046)	-0.0001	0.0046	(0.8947,0.9115)
β_3	-2	100	-2.0001 (0.0123)	-0.0001	0.0123	(-2.0244,-1.9776)
		200	-1.9996 (0.0092)	0.0004	0.0092	(-2.0174,-1.9823)
G_{11}	0.4	100	0.4712 (0.0611)	0.0712	0.0938	(0.3230,0.5702)
		200	0.4721 (0.0393)	0.0721	0.0821	(0.3902,0.5420)
G_{12}	0.2	100	0.1850 (0.0452)	-0.0150	0.0475	(0.1065,0.2770)
		200	0.1802 (0.0293)	-0.0198	0.0353	(0.1349,0.2420)
G_{13}	0.2	100	0.1803 (0.0362)	-0.0197	0.0412	(0.1193,0.2556)
		200	0.1822 (0.0248)	-0.0178	0.0305	(0.1390,0.2359)
G_{14}	0	100	0.0002 (0.0328)	0.0002	0.0327	(-0.0594,0.0688)
		200	-0.0010 (0.0234)	-0.0010	0.0234	(-0.0431,0.0472)
G_{21}	0.2	100	0.1819 (0.0459)	-0.0181	0.0493	(0.1059,0.2817)
		200	0.1801 (0.0305)	-0.0199	0.0364	(0.1288,0.2441)
G_{22}	0.4	100	0.4702 (0.0549)	0.0702	0.0891	(0.3391,0.5632)
		200	0.4752 (0.0372)	0.0752	0.0839	(0.3977,0.5377)
G_{23}	0.2	100	0.1807 (0.0358)	-0.0193	0.0406	(0.1165,0.2596)
		200	0.1803 (0.0239)	-0.0197	0.0310	(0.1322,0.2261)
G_{24}	0	100	-0.0023 (0.0329)	-0.0023	0.0330	(-0.0628,0.0562)
		200	0.0003 (0.0234)	0.0003	0.0234	(-0.0443,0.0458)
G_{31}	0.2	100	0.1758 (0.0398)	-0.0242	0.0465	(0.1044,0.2554)

TABLE VI OVERESTIMATING LATENT DIMENSION SIMULATION (Continued)

Parameter	True Value	m	Mean (SE)	Bias	RMSE	Middle 95%
		200	0.1785 (0.0257)	-0.0215	0.0335	(0.1260,0.2355)
G_{32}	0.2	100	0.1789 (0.0369)	-0.0211	0.0425	(0.1161,0.2593)
		200	0.1778 (0.0249)	-0.0222	0.0333	(0.1320,0.2271)
G_{33}	0.4	100	0.4793 (0.0439)	0.0793	0.0906	(0.3889,0.5557)
		200	0.4766 (0.0321)	0.0766	0.0830	(0.4092,0.5302)
G_{34}	0	100	0.0012 (0.0306)	0.0012	0.0306	(-0.0556,0.0607)
		200	-0.0006 (0.0216)	-0.0006	0.0216	(-0.0428,0.0394)
G_{41}	0	100	-0.0014 (0.0423)	-0.0014	0.0423	(-0.0823,0.0835)
		200	-0.0006 (0.0304)	-0.0006	0.0303	(-0.0551,0.0561)
G_{42}	0	100	-0.0032 (0.0439)	-0.0032	0.0440	(-0.0917,0.0829)
		200	5.9e-05 (0.0309)	0.0001	0.0308	(-0.0603,0.0617)
G_{43}	0	100	0.0021 (0.0437)	0.0021	0.0437	(-0.0771,0.0854)
		200	0.0002 (0.0298)	0.0002	0.0298	(-0.0611,0.0631)
G_{44}	0	100	-0.0004 (0.0428)	-0.0004	0.0428	(-0.0765,0.0819)
		200	0.0016 (0.0291)	0.0016	0.0291	(-0.0495,0.0544)
ν	0.6	100	0.6326 (0.1684)	0.0326	0.1713	(0.3714,0.9898)
		200	0.6218 (0.1143)	0.0218	0.1162	(0.4271,0.8505)
s	0.2	100	0.1987 (0.0112)	-0.0013	0.0113	(0.1762,0.2181)
		200	0.1991 (0.0078)	-0.0009	0.0078	(0.1842,0.2126)
σ^2_ϵ	0.3	100	0.2994 (0.0026)	-0.0006	0.0026	(0.2938,0.3041)
		200	0.2994 (0.0018)	-0.0006	0.0019	(0.2958,0.3027)
σ^2_η	0.25	100	0.1757 (0.0114)	-0.0743	0.0751	(0.1533,0.1978)
		200	0.1771 (0.0079)	-0.0729	0.0733	(0.1625,0.1933)
σ^2_ω	0.5	100	0.5005 (0.0040)	0.0005	0.0040	(0.4922,0.5081)
		200	0.5007 (0.0029)	0.0007	0.0030	(0.4949,0.5065)

of $\boldsymbol{\beta}$ are unbiased. However, the transition matrix $\mathbf{G}$, σ^2_{ω}, and σ^2_{η} are overestimated while σ^2_{ϵ} and s are underestimated. The parameter most affected by underestimation of r is $\boldsymbol{\nu}$; the mean estimate of $\boldsymbol{\nu}$ is far from the true value. When r is overestimated, the estimates of $\boldsymbol{\beta}$ are again unbiased. σ^2_{ϵ}, σ^2_{ω}, and s are also unbiased in this scenario. The elements of the transition matrix introduced by the inclusion of the extra latent variable are correctly estimated to be close to 0. For the true latent variables, the diagonal elements of the transition matrix $\mathbf{G}$ are overestimated while the off-diagonal estimates are underestimated. Furthermore, σ^2_{η} is underestimated, and the estimate of $\boldsymbol{\nu}$ is much closer to the truth but slightly overestimated. Thus, misspecification caused by overestimating or underestimating r leads to subsets of biased parameters.

In practice, misspecification can be avoided by comparing likelihood values. In an additional simulation study, 500 datasets were simulated based on the same specifications for a sample size of 100. Estimation was performed with the correct value of $r = 3$, r underestimated at 2, and r overestimated at 4. For every simulation, the likelihood was highest when the latent dimension r was correctly set to 3. Proper selection of r by comparing likelihood values therefore resolves the problems introduced by over and under estimation.

4.4.5 Convergence Theorem

We show theoretical convergence in the likelihood for the proposed spatiotemporal model. We follow the incomplete data likelihood presented by Shumway and Stoffer (2017), where the likelihood is computed using prediction error. Define prediction error as $\boldsymbol{\epsilon}_{it}(\boldsymbol{\Psi}) = \mathbf{z}_{it} - \mathsf{E}\left[\mathbf{z}_{it}|\mathbf{z}_{(t-1)}\right]$, where

$$\begin{aligned} E\left[\mathbf{z}_{it}|\mathbf{z}_{(t-1)}\right] &= E\left[\mathbf{X}_i\boldsymbol{\beta} + \mathbf{K}_i\mathbf{Y}_{it} + \boldsymbol{\xi}_{it}|\mathbf{z}_{(t-1)}\right] \\ &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{K}_i E\left[\mathbf{Y}_{it}|\mathbf{z}_{(t-1)}\right] \\ &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{K}_i\mathbf{G}\mathbf{y}_{i(t-1)}^{t-1}. \end{aligned} \tag{4.40}$$

$\boldsymbol{\epsilon}_{it}$ follows a Normal distribution with mean $\mathbf{0}$ and variance

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{it}}(\boldsymbol{\Psi}) &= \text{var}\left[\mathbf{z}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{G}\mathbf{y}_{i(t-1)}^{t-1}|\mathbf{z}_{(t-1)}\right] \\ &= \text{var}\left[\mathbf{X}_i\boldsymbol{\beta} + \mathbf{K}_i\left(\mathbf{G}_t\mathbf{Y}_{i(t-1)} + \boldsymbol{\eta}_{it}\right) + \boldsymbol{\xi}_{it} - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{K}_i\mathbf{G}\mathbf{y}_{i(t-1)}^{t-1}|\mathbf{z}_{(t-1)}\right] \\ &= \text{var}\left[\mathbf{K}_i\boldsymbol{\eta}_{it} + \boldsymbol{\xi}_{it} - \mathbf{K}_i\mathbf{G}\mathbf{y}_{i(t-1)}^{t-1}|\mathbf{z}_{(t-1)}\right] \\ &= \mathbf{K}_i\left(\mathbf{G}\text{var}\left[\mathbf{y}_{i(t-1)}^{t-1}|\mathbf{z}_{(t-1)}\right]\mathbf{G}' + \text{var}\left[\boldsymbol{\eta}_{it}|\mathbf{z}_{(t-1)}\right]\right)\mathbf{K}_i' + \text{var}\left[\boldsymbol{\xi}_{it}|\mathbf{z}_{(t-1)}\right] \\ &= \mathbf{K}_i\left(\mathbf{G}\mathbf{P}_{i(t-1)}^{t-1}\mathbf{G}' + \boldsymbol{\Sigma}_\eta\right)\mathbf{K}_i + \boldsymbol{\Sigma}_\xi. \end{aligned} \tag{4.41}$$

The likelihood is thus

$$L(\boldsymbol{\Psi}) = (2\pi)^{-1/2}\prod_{i=1}^{m}\prod_{t=1}^{T}|\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{it}}(\boldsymbol{\Psi})|^{-1/2}\exp\left\{\boldsymbol{\epsilon}_{it}(\boldsymbol{\Psi})'\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{it}}(\boldsymbol{\Psi})^{-1}\boldsymbol{\epsilon}_{it}(\boldsymbol{\Psi})\right\}. \tag{4.42}$$

We show that the sequence $\left\{L\left(\boldsymbol{\Psi}^{(k)}\right)\right\}$ from the GEM algorithm for the spatiotemporal model converges to some point $L^* = L(\boldsymbol{\Psi}^*)$ for some stationary point $\boldsymbol{\Psi}^* \in S$, where S is the set of stationary points in $\boldsymbol{\Omega}$.

Lemma 4.4.1. *Let $\boldsymbol{\Omega}$ be a compact set and L a continuous function in $\boldsymbol{\Omega}$. Let $\boldsymbol{\Psi}_o \in \boldsymbol{\Omega}$, such that $L(\boldsymbol{\Psi}_o) > -\infty$. Define $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o} = \{\boldsymbol{\Psi} \in \boldsymbol{\Omega} : L(\boldsymbol{\Psi}) \geq L(\boldsymbol{\Psi}_o)\}$. Then, $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$ is a closed and bounded set and, hence, is a compact set.*

Proof: Note that $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$ *is a subset of* $\boldsymbol{\Omega}$ *i.e.* $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o} \subseteq \boldsymbol{\Omega}$. *As* $\boldsymbol{\Omega}$ *is bounded, hence* $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$ *is also bounded.*

Let $\left\{\boldsymbol{\Psi}^d\right\}$ *be sequence in* $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$ *and* $\boldsymbol{\Psi}^d \to \overline{\boldsymbol{\Psi}}^d$. *We want to show that* $\overline{\boldsymbol{\Psi}}^d \in \boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$. *As* $\boldsymbol{\Psi}^d \in \boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$ *and* $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o} \subseteq \boldsymbol{\Omega}$, *hence,* $\boldsymbol{\Psi}^d \in \boldsymbol{\Omega}$. *Thus,* $\overline{\boldsymbol{\Psi}}^d \in \boldsymbol{\Omega}$, *as* $\boldsymbol{\Omega}$ *is compact. For* $\overline{\boldsymbol{\Psi}}^d$, *either (i)* $L(\overline{\boldsymbol{\Psi}}^d) \geq L(\boldsymbol{\Psi}_o)$ *or (ii)* $L(\overline{\boldsymbol{\Psi}}^d) < L(\boldsymbol{\Psi}_o)$. *(i) implies that* $\overline{\boldsymbol{\Psi}}^d \in \boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$. *If (ii) holds, then* $L(\overline{\boldsymbol{\Psi}}^d) < L(\boldsymbol{\Psi}_o) \leq L(\boldsymbol{\Psi}^d), \forall\ \boldsymbol{\Psi}^d \in \boldsymbol{\Psi}_o$. *Let* $\delta = L(\boldsymbol{\Psi}_o) - L(\overline{\boldsymbol{\Psi}}^d)$. *Thus,* $\left|L(\boldsymbol{\Psi}^d) - L(\overline{\boldsymbol{\Psi}}^d)\right| > \delta, \forall\ \boldsymbol{\Psi}^d \in \boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$, *which contradicts the fact that* L *is a continuous function and* $\boldsymbol{\Psi}^d \to \overline{\boldsymbol{\Psi}}^d$. *This means (ii) cannot hold. Thus,* $\overline{\boldsymbol{\Psi}}^d \in \boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$, *which means* $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$ *is closed. Hence,* $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o}$ *is compact.*

The combination of Result 4.2.1 and Lemma 4.4.1 yield the following Lemma:

Lemma 4.4.2. *Assume that* $\boldsymbol{\Omega}$ *is a compact* d*-dimensional subset of* $\mathbb{R}^d$. $L(\boldsymbol{\Psi})$ *is continuous in* $\boldsymbol{\Omega}$ *and differentiable in the interior of* $\boldsymbol{\Omega}$ *and* $\boldsymbol{\Omega}_{\boldsymbol{\Psi}_o} = \{\boldsymbol{\Psi} \in \boldsymbol{\Omega} : L(\boldsymbol{\Psi}) \geq L(\boldsymbol{\Psi}_o)\}$ *is a compact set. Then any sequence* $\left\{L\left(\boldsymbol{\Psi}^{(k)}\right)\right\}$ *is bounded above for any vector of initial values* $\boldsymbol{\Psi}^{(0)} \in \boldsymbol{\Omega}$ *such that* $L\left(\boldsymbol{\Psi}^{(0)}\right) > -\infty$ *and, hence, the sequence converges to some point* L^*

Proof: The regularity conditions of Result 4.2.1 are satisfied. The convergence of $\left\{L\left(\boldsymbol{\Psi}^{(k)}\right)\right\}$ *to* L^* *follows.*

By Lemma 4.4.2, Result 4.2.2 applies to our GEM estimating procedure for the proposed spatiotemporal model, yielding the following theorem:

Theorem 4.4.1. The sequence $\left\{L\left(\boldsymbol{\Psi}^{(k)}\right)\right\}$ from the ECM algorithm for the spatiotemporal model converges to some point $L^* = L(\boldsymbol{\Psi}^*)$ for some stationary point $\boldsymbol{\Psi}^* \in S$, where S is the set of stationary points in $\boldsymbol{\Omega}$.

Proof: Lemma 4.4.2 satisfies the convergence to some point L^*. $Q\left(\boldsymbol{\Psi}_1^{(k)}, \boldsymbol{\Psi}_2^{(k)}\right) \geq Q\left(\boldsymbol{\Psi}_1^{(k)}, \boldsymbol{\Psi}_2^{(k-1)}\right) \geq Q\left(\boldsymbol{\Psi}_1^{(k-1)}, \boldsymbol{\Psi}_2^{(k-1)}\right)$, *satisfying Definition 4.2.1 of a GEM algorithm. By Result 4.2.2, the convergence of* $\left\{L\left(\boldsymbol{\Psi}^{(k)}\right)\right\}$ *to* $L^* = L(\boldsymbol{\Psi}^*)$ *for some stationary point* $\boldsymbol{\Psi}^* \in S$, *where* S *is the set of stationary points in* $\boldsymbol{\Omega}$, *follows.*

4.5 Conclusion

The ECM algorithm is a flexible estimating procedure that utilizes the well known principles of the EM algorithm. Estimation of parameters for the spatiotemporal model is performed via the ECM algorithm to account for the missing latent variable and optimization of parameters without closed-form solutions. An algorithm for selecting initial values is also proposed to increase the likelihood that the estimating procedure converges to the global maximum. Our simulation study shows that estimates of the ECM algorithm for the spatiotemporal model are unbiased with decreasing RMSE as the sample size increases. Furthermore, we show that model selection for specifying r can be performed using the likelihood. Estimates for the model parameters are therefore reliable for use in analysis of functional connectivity.

CHAPTER 5

ESTIMATING FUNCTIONAL CONNECTIVITY

Analysis of functional connectivity is extremely important for detecting disrupted connections in individuals affected with a neurological condition. A deeper understanding of the disease specific abnormalities in communication between brain regions may lead to early diagnosis and targeted treatment interventions (Bhaumik et al., 2018a). The simplest approach to measuring functional connectivity is by calculating pairwise correlations between regions of interest. However, inferences made using Fisher's z-transformation on sample correlations are not valid due to the temporal correlation of observations within a region. Alternative approaches to estimating functional connectivity include multivariate methods such as principal components analysis and independent components analysis. These methodologies can help identify connectivity patterns without having to make any assumptions regarding functional form (Lindquist, 2008). However, it is not clear how the false positive rate is controlled for multiple comparisons using these methods (Bhaumik et al., 2018a).

In order to make accurate inferences about correlations, the denoised, temporally correlated estimated outcomes within a region from the spatiotemporal model are decorrelated. From the uncorrelated time series, estimates of the spatial correlations are derived. Through this transformation, application of Fisher's z-transformation and the associated inference procedures are statistically valid. For n regions under study, there are $L = n(n-1)/2$ simultaneous tests of functional connections. We control the false discovery rate to address the issue of multiple

comparisons introduced by the large number of pairwise correlations under study. Efron's local false discovery rate is selected as the most suitable choice for our data.

5.1 Estimating Spatial Correlation

To estimate functional connectivity, a separate spatiotemporal model is fit for each group. Let g denote the group of individual i, where $g = 0$ for controls and $g = 1$ for cases. Furthermore, let m_g denote the sample size for group g. For the jth region, $j = 1, \ldots, n$, at time t, $t = 1, \ldots, T$ in individual i, $i = 1, \ldots, m_g$, the estimated outcome from the spatiotemporal model is

$$\hat{z}_{gijt} = \mathbf{X}_{gij}\hat{\boldsymbol{\beta}}_g + \mathbf{K}_{gij}\mathbf{y}_{git}^{t-1}, \tag{5.1}$$

where $\mathbf{X}_{gij}$ is the p-dimensional row vector from the jth row of the matrix $\mathbf{X}_i$ and $\mathbf{K}_{gij}$ is the r-dimensional row vector from the jth row of the matrix $\mathbf{K}_{gi}$. $\mathbf{y}_{git}^{t-1} = \hat{\mathbf{G}}_g\mathbf{y}_{gi(t-1)}^{t-1}$, where $\mathbf{y}_{gi(t-1)}^{t-1}$ is the Kalman filter given the final parameter estimates. Then, $\hat{\mathbf{z}}_{gij} = (\hat{z}_{gij1}, \hat{z}_{gij2}, \ldots, \hat{z}_{gijT})'$ is the $T \times 1$ vector containing all estimated fMRI measurements for region j across time. If the elements of $\hat{\mathbf{z}}_{gij}$ are uncorrelated, inferences using the Pearson correlation can be appropriately made.

Uncorrelation of $\hat{\mathbf{z}}_{gij}$ is achieved by whitening. Whitening, also referred to as sphering, is a linear transformation of a T-dimensional vector into a new uncorrelated vector of the same dimension. Let $\boldsymbol{\Sigma}$ denote the positive-definite $T \times T$ covariance matrix of the dependent vector $\mathbf{z}$. The T-dimensional random vector

$$\mathbf{z}^* = (z_1^*, \ldots, z_n^*)' = \mathbf{W}\mathbf{z} \tag{5.2}$$

has variance-covariance matrix equal to the identity matrix $\mathbf{I}_T$. The $T \times T$ matrix $\mathbf{W}$ is the whitening matrix. The resulting transformation produces orthogonality among random variables, which simplifies multivariate data analysis computationally and statistically (Kessy et al., 2018).

The transformation requires a suitable choice for $\mathbf{W}$. For a given $\mathbf{\Sigma}$, there are an infinite number of whitening matrices that produce orthogonal but different sphered random variables. Common approaches to selecting $\mathbf{W}$ include zero-phase component analysis, principal component analysis, and the Cholesky decomposition. The Choleksy factorization is the most suitable for decorrelating rsfMRI time series data as it implicitly assumes an ordering of the variables. The Cholesky factorization of the inverse of the covariance matrix is

$$\mathbf{\Sigma}^{-1} = \mathbf{L}\mathbf{L}', \tag{5.3}$$

where $\mathbf{L}$ is a unique lower triangular $T \times T$ matrix with positive diagonal values. This yields a whitening matrix of $\mathbf{W} = \mathbf{L}'$ (Kessy et al., 2018).

To uncorrelate the time series, an estimate of temporal covariance is needed. Let $\hat{\mathbf{z}}_{git} = (\hat{z}_{gi1t}, \hat{z}_{gi2t}, \ldots, \hat{z}_{gint})'$ denote the $n \times 1$ vector of observations across all regions at time t for individual i in group g. Furthermore, denote the covariance between $\hat{\mathbf{z}}_{git}$ and $\hat{\mathbf{z}}_{git'}$ for $t = 1, 2, \ldots, T$ and $t' = 1, 2, \ldots, T$ as $\text{cov}\left(\hat{\mathbf{z}}_{git}, \hat{\mathbf{z}}_{git'}\right) = \sigma_{\hat{\mathbf{z}}_{gitt'}}$. The $T \times T$ temporal covariance matrix is

$$\mathbf{\Sigma}_{\hat{\mathbf{z}}_{giT}} = \begin{bmatrix} \sigma_{\hat{\mathbf{z}}_{gi11}} & \sigma_{\hat{\mathbf{z}}_{gi12}} & \sigma_{\hat{\mathbf{z}}_{gi13}} & \dots & \sigma_{\hat{\mathbf{z}}_{gi1T}} \\ \sigma_{\hat{\mathbf{z}}_{gi21}} & \sigma_{\hat{\mathbf{z}}_{gi22}} & \sigma_{\hat{\mathbf{z}}_{gi23}} & \dots & \sigma_{\hat{\mathbf{z}}_{gi2T}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\hat{\mathbf{z}}_{giT1}} & \sigma_{\hat{\mathbf{z}}_{giT2}} & \sigma_{\hat{\mathbf{z}}_{giT3}} & \dots & \sigma_{\hat{\mathbf{z}}_{giTT}} \end{bmatrix}.$$

Let $\mathbf{L}'_{\mathbf{\Sigma}_{\hat{\mathbf{z}}_{giT}}}$ denote the unique lower triangular matrix from the Cholesky decomposition of $\mathbf{\Sigma}^{-1}_{\hat{\mathbf{z}}_{giT}}$. The temporally uncorrelated time series for region j of the ith subject nested within the gth group is

$$\mathbf{z}^*_{gij} = \mathbf{L}'_{\mathbf{\Sigma}_{\hat{\mathbf{z}}_{giT}}} \hat{\mathbf{z}}_{gij}. \tag{5.4}$$

An estimate of spatial covariance is derived from this uncorrelated time series. Let $\sigma_{\mathbf{z}^*_{gijj'}}$ denote $\text{cov}\left(\mathbf{z}^*_{gij}, \mathbf{z}^*_{gij'}\right)$ for $j = 1, 2, \dots, n$ and $j' = 1, 2, \dots, n$ for individual i in group g. The $n \times n$ spatial covariance matrix is

$$\mathbf{\Sigma}_{\mathbf{z}^*_{gin}} = \begin{bmatrix} \sigma_{\mathbf{z}^*_{gi11}} & \sigma_{\mathbf{z}^*_{gi12}} & \sigma_{\mathbf{z}^*_{gi13}} & \dots & \sigma_{\mathbf{z}^*_{gi1n}} \\ \sigma_{\mathbf{z}^*_{gi21}} & \sigma_{\mathbf{z}^*_{gi22}} & \sigma_{\mathbf{z}^*_{gi23}} & \dots & \sigma_{\mathbf{z}^*_{gi2n}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\mathbf{z}^*_{gin1}} & \sigma_{\mathbf{z}^*_{gin2}} & \sigma_{\mathbf{z}^*_{gin3}} & \dots & \sigma_{\mathbf{z}^*_{ginn}} \end{bmatrix}.$$

The jj' element of the corresponding correlation matrix derived from $\mathbf{\Sigma}_{\mathbf{z}^*_{gin}}$ is the estimate of functional connectivity for regions j and j'. We denote the estimated correlation matrix of the

ith subject nested within the gth group by R_{ig}, where $R_{ig} = \left(\left(\text{corr}\left(z^*_{gij}, z^*_{gij'}\right)\right)\right)$. R_{ig} is a subject and group specific correlation matrix based on temporally uncorrelated but spatially correlated transformed variables.

5.2 Fisher's z-Transformation

A transformation of the sample correlation is necessary to make asymptotic inference about the value of a population correlation coefficient. R. A. Fisher described this transformation in his book "Statistical Methods for Research Workers" published in 1925. Let r denote the sample correlation coefficient calculated from T pairs of independent observations. When the transformation

$$z = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right) \tag{5.5}$$

is applied to the sample correlation coefficient, the sampling distribution of the resulting variable z is approximately normal. This is commonly referred to as Fisher's z-transformation. The standard error of z is

$$\sigma_z = \frac{1-\rho^2}{\sqrt{T-1}}, \tag{5.6}$$

where ρ denotes the true correlation coefficient. ρ is unknown in practice and replaced with its estimate r. For large samples, the standard error of z is

$$\hat{\sigma}_z = \frac{1}{T-3}. \tag{5.7}$$

By applying this transformation, hypotheses about the population correlation coefficient ρ can be performed using large-sample inferential procedures for the normal distribution (Fisher, 1925).

5.3 Hypothesis Testing

A z-test is used for analysis of functional connectivity. For all regions $j = 1, 2, \ldots, n$, $j' = 1, 2, \ldots, n$, $j \neq j'$, let $\rho_{1jj'}$ and $\rho_{0jj'}$ denote the population correlation coefficient of regions j and j' in the disease and control groups, respectively. The hypothesis test for equality of the correlation coefficients is

$$H_{0jj'} : \rho_{1jj'} = \rho_{0jj'}, \tag{5.8}$$

$$H_{Ajj'} : \rho_{1jj'} \neq \rho_{0jj'}.$$

Let m_1 denote the number of individuals in the case group and m_0 the number in the control group. Then, $z_{1jj'} = \left\{z_{11jj'}, z_{12jj'}, \ldots, z_{1m_1jj'}\right\}$ and $z_{0jj'} = \left\{z_{01jj'}, z_{02jj'}, \ldots, z_{0m_0jj'}\right\}$ denote the sets of Fisher's z-transformations of the sample correlations $r_{1jj'} = \left\{r_{11jj'}, r_{12jj'}, \ldots, r_{1m_1jj'}\right\}$ and $r_{0jj'} = \left\{r_{01jj'}, r_{02jj'}, \ldots, r_{0m_0jj'}\right\}$, respectively. Let $\bar{z}_{1jj'}$ and $\bar{z}_{0jj'}$ denote the mean of $z_{1jj'}$ and $z_{0jj'}$, respectively. Using this transformation, testing procedures for the normal distribution can be used. The z-test statistic is given by

$$z_{jj'} = \frac{\bar{z}_{1jj'} - \bar{z}_{0jj'}}{\sqrt{\frac{1}{m_1(T-3)} + \frac{1}{m_0(T-3)}}} \tag{5.9}$$

for all $j = 1, 2, \ldots, n$ and $j' = 1, 2, \ldots, n, j \neq j'$.

5.4 False Discovery Rate

In a study where multiple links are tested for disrupted connectivities, multiple comparison issues arise. The most common correction procedures for multiple comparisons control the familywise error rate, or the probability of committing any type I error. As an alternative method, Benjamini and Hochberg (1995) suggested controlling the "expected proportion of errors among rejected hypotheses," which they call the false discovery rate (FDR). Among the set of null hypotheses rejected, let V denote the number of hypotheses that were truly null and S the number of hypotheses that are truly not null. The proportion of hypotheses that were rejected but are truly null is denoted Q. The FDR is the expectation of Q, or

$$E[Q] = E\left[\frac{V}{V+S}\right]. \tag{5.10}$$

If $V + S = 0$, then Q is set to 0. Let $P_{(1)}, P_{(2)}, \ldots, P_{(L)}$ represent the ordered p-values corresponding to hypotheses $H_{0(1)}, H_{0(2)}, \ldots, H_{0(L)}$. To control the FDR at level q^*, let k be the largest i for which

$$P_{(i)} \leq \frac{i}{L} q^*. \tag{5.11}$$

The set of hypotheses to reject include all $H_{(i)}, i \leq k$ (Benjamini and Hochberg, 1995).

5.5 Local False Discovery Rate

Benjamini and Hochberg's original FDR procedure requires independent test statistics to ensure control of the FDR. Several improvements to the FDR have been proposed since its introduction. In a comparison of the adaptive Benjamini and Hochberg procedure, Cai and

Sun's oracle and adaptive compound decision procedure, and Efron's local FDR, Bhaumik et al. (2018) identified Efron's local FDR as "the most suitable for neuroconnectivity studies" (Bhaumik et al., 2018b). Moreover, it has also been recognized that multiple hypotheses often "share a natural underlying group, hierarchical, nested, or network structure of dependence" that should be "utilized in performing multiple comparisons" (Sankaran and Holmes, 2014). In many applications, prior information is available for formation of groups from the set of hypotheses being tested. Efron notes that ignoring this information and applying a single FDR procedure to the entire set of hypotheses may lead to conservative or liberal results within a subgroup (Efron, 2008).

Currently there is not a suitable approach that incorporates group information for multiple hypothesis testing for the data used in this dissertation. In a study with a large sample size and large number of time points observed in the rsfMRI time series, Equation 5.9 yields large z-test statistics and therefore small p-values. Methods of controlling the FDR that rely on p-values therefore do not identify differences in functional connectivity that are both clinically and statistically significant. Liu, Sarkar, and Zhao (2016) and Efron (2008) have proposed group methods based on Efron's local FDR. However, select equations in the approach of Liu, Sarkar, and Zhao (2016) cannot be applied when there is a large number of links in a group (Liu et al., 2016). Furthermore, Efron notes that his grouped approach is statistically valid when small groups can be formed (Efron, 2008). Since small subgroups of tests cannot be created for our data, Efron's local FDR approach is the most suitable for the proposed procedure at this time.

Unlike the original FDR, which relies on p-values and is thus based on traditional frequentist hypothesis testing approaches, the local FDR is based on an empirical Bayes approach. The number of simultaneous tests must be "at least in the hundreds" to implement the procedure. Let $z_1, z_2, \ldots, z_L$ be the set of test statistics corresponding to the null hypotheses $H_{01}, H_{02}, \ldots$, H_{0L}. The test statistics need not be independent. The L tests can be divided into two classes of null or nonnull, each occurring with prior probabilities p_0 and $p_1 = 1 - p_0$, respectively. Let $f_0(z)$ and $f_1(z)$ denote the densities of the null and nonnull test statistics, respectively. $f_1(z)$ does not need to be specified; however, it assumed to be longer-tailed than $f_0(z)$. The mixture density of the test statistics is

$$f(z) = p_0 f_0(z) + p_1 f_1(z). \tag{5.12}$$

Then, the Bayes posterior probability that a hypothesis is null given test statistic z is

$$q_{loc}(z) = \frac{p_0 f_0(z)}{f(z)}. \tag{5.13}$$

This posterior probability is the local FDR. For a preselected threshold q^*, the null hypothesis is rejected if $q_{loc}(z) < q^*$ (Efron, 2007).

Efron notes that "the literature has not reached consensus on a standard choice of q...the equivalent of 0.05 for single tests" (Efron, 2007). However, the Bayes factor can be used to offer some insight into selecting a cutoff value of q for determining significance. For the two competing hypotheses of nonnull vs. null, the posterior odds ratio in favor of the nonnull hypothesis is related to the prior odds ratio by

$$\begin{aligned} \frac{\Pr(\text{nonnull}|z)}{\Pr(\text{null}|z)} &= \frac{\Pr(z|\text{nonnull})}{\Pr(z|\text{null})} \times \frac{\Pr(\text{nonnull})}{\Pr(\text{null})} \\ &= K \times \frac{\Pr(\text{nonnull})}{\Pr(\text{null})}. \end{aligned} \tag{5.14}$$

K is the Bayes factor and quantifies the relative evidence of the data z in favor of the nonnull compared to the null. When a large number of tests are being conducted, it is usually assumed that a large proportion are null, such that $p_0 \geq 0.9$. If a cutoff value of 0.2 is used, then

$$\frac{0.8}{0.2} = K \times \frac{0.1}{0.9},$$

yielding a Bayes factor of 36 in favor of the nonnull versus the null. A q-value less than 0.2 is therefore recommended as a cutoff for statistical significance (Efron, 2007).

5.5.1 Mixture Density Estimation

Calculation of the local FDR requires estimates of the mixture density $f(z)$, null density $f_0(z)$, and prior probability of a null test-statistic p_0. Lindsay's method is used to estimate the mixture density $f(z)$. This method reconstructs density estimation problems in terms of Poisson regression models, resulting in efficient and flexible parametric density estimation (Efron, 2007). Suppose the L z-values are binned, yielding bin counts $s_1, s_2, \ldots, s_K$, where $\sum_{k=1}^{K} s_k = L$. Let $z_{(k)}$ denote the the midpoint value of the kth bin. It is assumed that the counts are independent Poisson, such that

$$s_k \sim \text{Po}(\mu_k(\beta)), \tag{5.15}$$

where

$$\mu_k(\beta) = c_\beta \exp\left(\sum_{j=1}^{p} \beta_j z_{(k)}^j\right). \tag{5.16}$$

For a given p, $\mu_k(\beta)$ is fit using maximum likelihood estimation. The estimate of $f(z)$ is

$$\hat{f}(z) = \hat{c}_\beta \exp\left(\sum_{j=1}^{p} \hat{\beta}_j z^j\right), \tag{5.17}$$

where $\hat{c}_\beta$ is the constant that makes $\hat{f}(z)$ integrate to one (Efron and Tibshirani, 1996; Efron, 2007).

5.5.2 Null Density Estimation

The null density $f_0(z)$ plays an important role in the calculation of the local FDR. A null distribution is often assumed when conducting a statistical test. However, when a large number of hypothesis tests are conducted simultaneously, there is the possibility of detecting deviations of the distribution of test statistics from the theoretical null. This may be due to failed assumptions, unobserved covariates, and correlations between multivariate outcomes. In cases where there is clear evidence against the theoretical null, an empirical null distribution can be estimated (Efron, 2007).

When the theoretical null density is assumed to be standard normal, the empirical null distribution is also assumed to be normal. However, instead of specifying a known mean and variance, it is assumed that

$$f_0(z) \sim N\left(\delta_0, \sigma_0^2\right). \tag{5.18}$$

δ_0, σ_0^2 and p_0 are estimated from the distribution of test statistics near $z = 0$ by maximum likelihood estimation of a truncated normal model. Let

$$f_1(z) = 0 \text{ for } z \in [-x_0, x_0], \tag{5.19}$$

implying that the nonnull density is supported outside the interval $[-x_0, x_0]$. Furthermore, $I_0 = \{i : z_i \in [-x_0, x_0]\}$, N_0 equals the number of $z_i \in [-x_0, x_0]$, $\mathbf{z}_0 = \{z_i, i \in I_0\}$,

$$H_0(\delta_0, \sigma_0) = \Phi\left(\frac{x_0 - \delta_0}{\sigma_0}\right) - \Phi\left(-\frac{x_0 - \delta_0}{\sigma_0}\right), \tag{5.20}$$

and

$$\varphi(z) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2}\left(\frac{z - \delta_0}{\sigma_0}\right)^2\right\}. \tag{5.21}$$

The probability that $z_i \in [-x_0, x_0]$ is

$$\theta = p_0 \times H_0(\delta_0, \sigma_0). \tag{5.22}$$

$(\delta_0, \sigma_0, \theta)$ are estimated by maximum likelihood. The likelihood function of the observed data $(N, \mathbf{z}_0)$ is

$$f(N, \mathbf{z}_0) = \left[\theta^{N_0}(1-\theta)^{N-N_0}\right] \prod_{I_0} \frac{\varphi_{\delta_0,\sigma_0}(z_i)}{H_0(\delta_0, \sigma_0)}. \tag{5.23}$$

Let $\left(\hat{\delta}_0, \hat{\sigma}_0, \hat{\theta}\right)$ denote the maximum likelihood estimates of $(\delta_0, \sigma_0, \theta)$. The maximum likelihood estimate of the proportion of null test statistics is

$$\hat{p}_0 = \frac{\hat{\theta}}{H_0\left(\hat{\delta}_0, \hat{\sigma}_0\right)}. \tag{5.24}$$

Then,

$$q\left(z\right) = \frac{\hat{p}_0 \hat{f}_0\left(z\right)}{\hat{f}\left(z\right)}. \tag{5.25}$$

is the estimated q-value for a given z-statistic z (Efron, 2007).

5.6 Conclusion

Analysis of functional connectivity is fundamental for a deeper understanding of neurological conditions. Although principal components analysis and independent components analysis are optional statistical tools for understanding connectivity patterns, inference based on the sample correlation is simple, well understood, and allows for control of the false discovery rate. In our proposed approach, the spatiotemporal model introduced in Chapters 3 and 4 is used to denoise the data. The denoised data within a region is subsequently decorrelated. Functional connectivity is thus based on temporally uncorrelated but spatially correlated transformed variables. This enables valid implementation of Fisher's z-transformation to sample correlations with the goal of identifying disrupted connectivity patterns. This method is applicable to a wide range of neurological conditions and may lead to early diagnosis and targeted treatment.

CHAPTER 6

ANALYSIS OF FUNCTIONAL CONNECTIVITY IN AUTISM SPECTRUM DISORDER

Analysis of functional connectivity in ASD is crucial for our understanding of the neurodevelopmental disorder. A deeper understanding of the disease will ultimately lead to the development of more objective diagnosis criteria and targeted treatment options. In this chapter, subjects from the ABIDE I dataset described in Chapter 2 are used to illustrate the proposed methodologies. A separate spatiotemporal model is fit for the ASD and control groups, and the within-region time series of the estimated outcome is decorrelated. Spatial correlations are estimated from the temporally uncorrelated observations. This allows appropriate application of Fisher's z-transformation for statistical inference. The fMRI data consists of 110 brain regions, yielding $(110 \times 109)/2 = 5,995$ links to be analyzed. Descriptions of the brain regions are provided in Appendix A.

6.1 Analysis of Autism Brain Imaging Data Exchange

As described in Section 2.2, the ABIDE I dataset consists of 539 ASD subjects and 573 controls from twenty datasets collected from seventeen sites. The subjects are scanned at a wide range of ages, with a minimum of 6.47 and a maximum of 64.00. This analysis is restricted to individuals between the ages of seven and fourteen due to the need for better characterization of ASD in children and the practical implications of understanding the condition at a younger

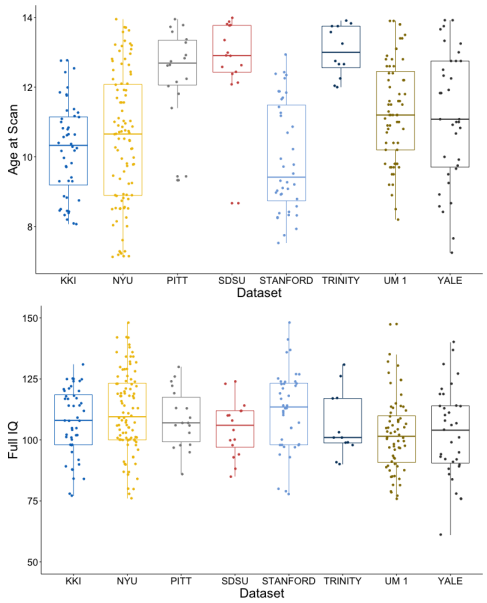


Figure 7: Boxplots of age at scan and FIQ by dataset for analysis population.

age. Children from ABIDE I sites with at least ten subjects in this age range and more than 140 fMRI time points observed per region are included in this analysis. Although the "Levuen 1" dataset satisfies these criteria, the phenotypic dataset does not include full IQ scores. It was therefore excluded from the sample population.

Table VII and Table VIII describe the typically developing controls and ASD subjects in the analysis population, respectively. This analysis consists of 162 ASD subjects and 167 controls from eight ABIDE I sites. Males make up the vast majority of the sample; 83.3% of ASD subjects and 74.9% of controls are male. Boxplots of age at scan and full IQ are provided in Figure 7. Although the population has been restricted to individuals between seven and

TABLE VII ABIDE I CONTROL SAMPLE DESCRIPTIVE STATISTICS

Site	N	Female, N (%)	Male, N (%)	Age, Mean (SD)	Full IQ, Mean (SD)
1.KKI	31	9 (29.0)	22 (71.0)	10.2 (1.3)	113.0 (9.5)
2.NYU	47	13 (27.7)	34 (72.3)	10.7 (1.9)	114.6 (14.2)
3.PITT	8	2 (25.0)	6 (75.0)	12.5 (1.4)	107.4 (10.7)
4.SDSU	10	4 (40.0)	6 (60.0)	12.6 (1.5)	104.6 (10.2)
5.STANFORD	20	4 (20.0)	16 (80.0)	10.0 (1.6)	112.1 (15.4)
6.TRINITY	6	0 (0.0)	6 (100.0)	12.8 (0.6)	108.2 (16.3)
7.UM 1	27	6 (22.2)	21 (77.8)	11.2 (1.5)	105.4 (10.2)
8.YALE	18	4 (22.2)	14 (77.8)	11.1 (2.0)	109.1 (16.2)
TOTAL	167	42 (25.1)	125 (74.9)	10.9 (1.8)	110.7 (13.2)

TABLE VIII ABIDE I ASD SAMPLE DESCRIPTIVE STATISTICS

Site	N	Female, N (%)	Male, N (%)	Age, Mean (SD)	Full IQ, Mean (SD)
1.KKI	17	3 (17.6)	14 (82.4)	10.3 (1.5)	97.1 (15.1)
2.NYU	49	4 (8.2)	45 (91.8)	10.3 (1.9)	108.6 (18.0)
3.PITT	10	4 (40.0)	6 (60.0)	12.4 (1.3)	110.0 (13.3)
4.SDSU	6	1 (16.7)	5 (83.3)	13.0 (0.8)	104.8 (14.4)
5.STANFORD	20	4 (20.0)	16 (80.0)	10.0 (1.6)	112.5 (17.8)
6.TRINITY	6	0 (0.0)	6 (100.0)	13.3 (0.8)	104.0 (10.9)
7.UM 1	37	6 (16.2)	31 (83.8)	11.4 (1.4)	100.0 (17.7)
8.YALE	17	5 (29.4)	12 (70.6)	11.2 (1.9)	96.4 (18.7)
TOTAL	162	27 (16.7)	135 (83.3)	10.9 (1.8)	104.4 (17.7)

fourteen, between-site variability in age still exists. Full IQ scores exhibit substantial within-site variability.

6.2 Model Specifications

As discussed in Chapter 3, several data-specific specifications for the spatiotemporal model must be made. This includes identifying the appropriate order for the temporal process, adjacency matrix, and dimension of the latent factor dynamic model, r.

6.2.1 Temporal Process

The spatiotemporal model developed for rsfMRI data contains a first-order autoregressive model for the temporal process. A first-order model was identified from the results of the temporal autocorrelation function of rsfMRI data for the analysis population. The temporal autocorrelation plots of nine regions are provided in Figure 8. For each of the regions, the lag-one correlation is high, indicating that the observed data at time t is highly dependent on the observation at time $t-1$. However, for lags greater than one, the correlation between observations is small. A first-order model was therefore identified as the most appropriate for this dataset.

6.2.2 Adjacency Matrix

As described in Section 3.2.1, we derive the dimension reducing matrix $\mathbf{K}_i$ from the Moran operator of $\mathbf{X}_i$. The operator requires an adjacency matrix. The approach described in Section 3.2.1.1 is used. For each group, nineteen bins were created between the maximum and minimum functional distance. Within each bin, the mean of the estimated semivariogram was taken and subtracted from one. Among the candidate spatial weighting function, the residuals

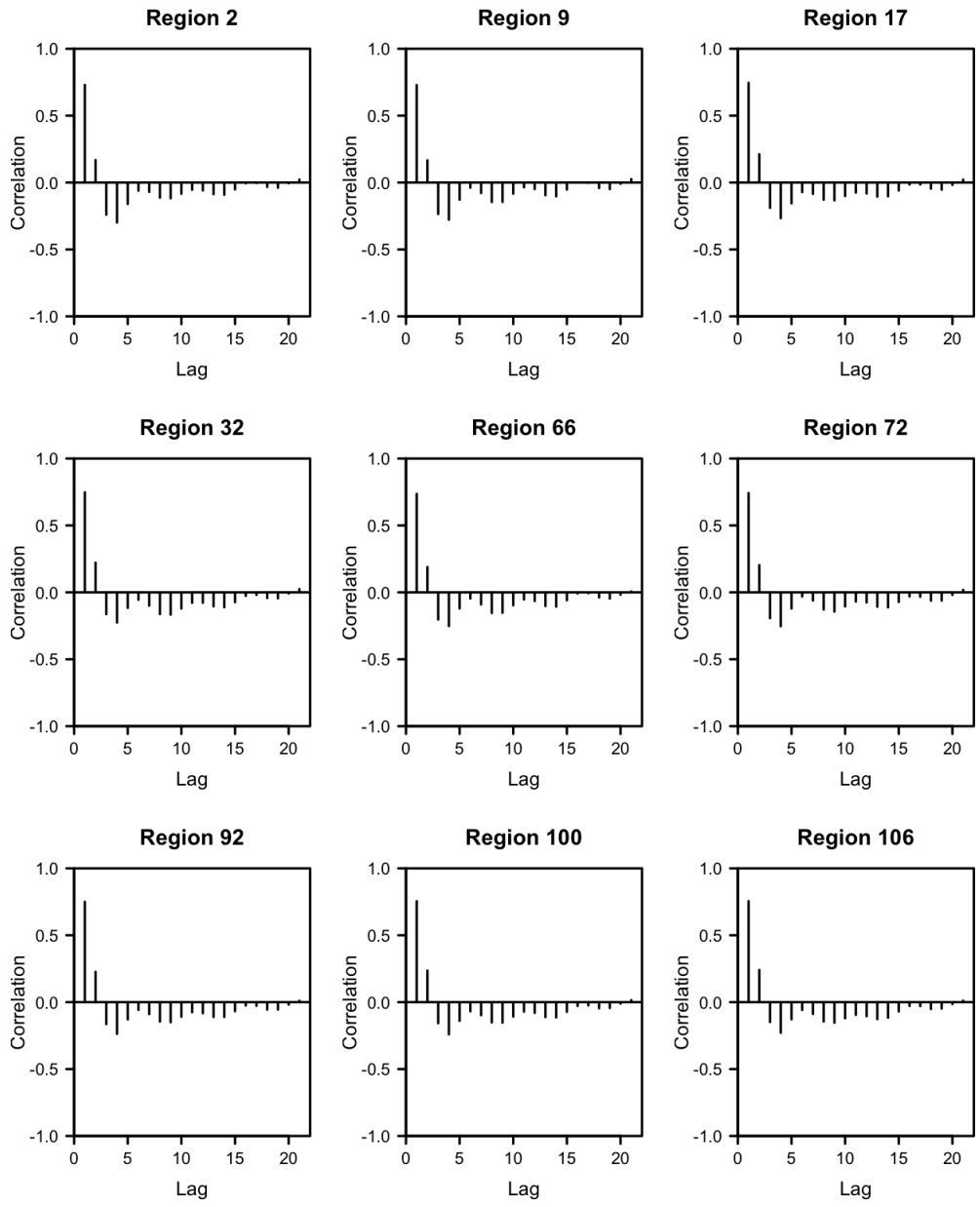


Figure 8: Temporal autocorrelation functions for ABIDE analysis.

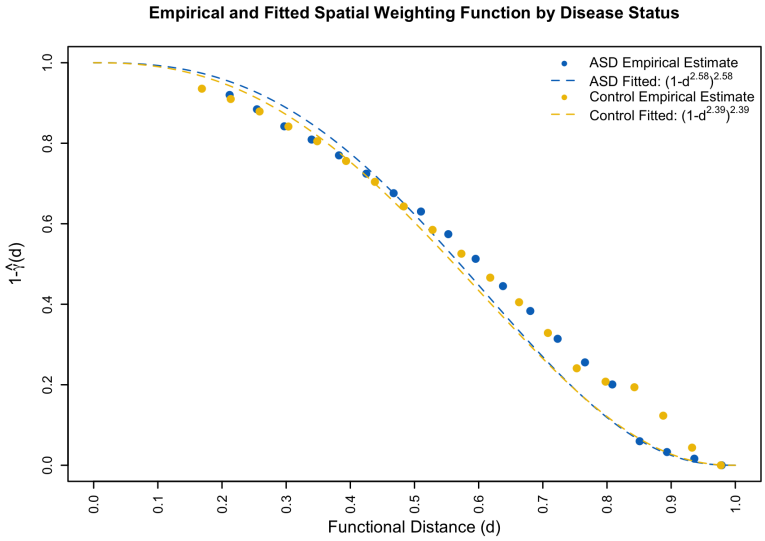


Figure 9: Empirical and estimated spatial weighting functions.

were minimized for cases and controls with the double-power distance weights function seen in Figure 4. The double-power distance weights function is given as $(1 - d^p)^p$, where d is the distance and p is a fitted parameter. We use nonlinear regression on the double-power weights function to estimate p for each group. Plots with estimated and fitted spatial weighting functions by group are provided in Figure 9. Inspecting these plots, we find a similar pattern for both ASD and controls.

6.2.3 Reduced Dimension

As described in Section 3.2.1, a lower dimension r is selected from the SVD of the Moran operator of $\mathbf{X}$. Plots of the spatial variability explained by reduced dimensions are provided in Figure 10. For both ASD and controls, drastic dimension reduction is achieved while explain-

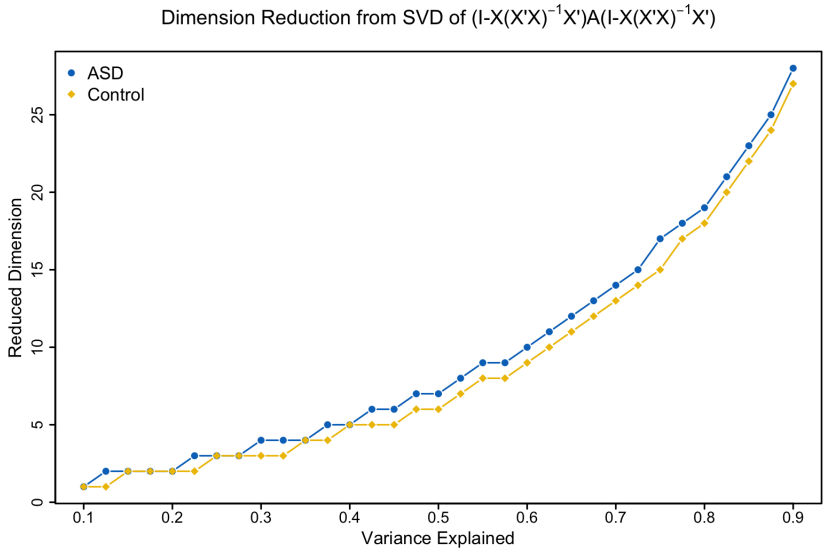


Figure 10: SVD of Moran operator for selection of $\mathtt{r}$.

ing a larger percent of spatial variability. For both groups, the likelihood ratio test rejected the null model of 65% total spatial variability in favor of 70%, but failed to reject the null model of 70% total spatial variability versus 75%. Thus, we use spatiotemporal models with 70% total variance explained, with reduced dimensions $\mathtt{r} = 14$ and $\mathtt{r} = 13$ for ASD and controls, respectively.

6.3 Results

We fit separate spatiotemporal models as described in Chapter 3 and Chapter 4 for ASD subjects and healthy controls. Analysis of functional connectivity as described in Chapter 5 is subsequently performed.

6.3.1 Spatiotemporal Model Estimates

Age at scan (X_1), FIQ (X_2), and sex (X_3) were included as covariates. For ASD, the estimated parameters are $\hat{\boldsymbol{\beta}}_1 = (0.048, -0.005, -0.012, -0.006)'$, $\hat{\sigma}^2_{\epsilon 1} = 277.997$, $\hat{\sigma}^2_{\omega 1} = 160.635$, $\hat{\sigma}^2_{\eta 1} = 1002.943$, $\hat{\nu}_1 = 1.517$, $\hat{s}_1 = 0.036$, and 14×14 matrix $\hat{\mathbf{G}}_1$ equals

$$\begin{pmatrix}
0.766 & -1e-5 & 0.051 & 0.128 & -0.017 & -0.212 & -0.170 & -0.234 & -0.004 & -0.286 & 0.034 & -0.197 & 0.434 & -0.357 \\
-0.024 & 0.813 & -0.055 & -0.027 & -0.051 & 0.119 & 0.122 & 0.061 & -0.149 & 0.393 & 0.003 & 0.305 & -0.275 & 0.277 \\
0.023 & 0.011 & 0.749 & 0.030 & -0.031 & -0.077 & -0.054 & 0.008 & 0.062 & -0.208 & -0.026 & -0.026 & 0.009 & -0.031 \\
-0.013 & -0.011 & 0.016 & 0.758 & 0.040 & -0.038 & -0.021 & 1e-4 & 0.014 & -0.036 & -0.030 & -0.071 & 0.073 & -0.062 \\
0.009 & 0.027 & 0.008 & -0.023 & 0.743 & -0.017 & -0.001 & -0.024 & -0.069 & 0.006 & -0.032 & 0.053 & 0.138 & -0.062 \\
0.030 & -0.009 & 0.041 & 0.044 & 0.007 & 0.677 & 0.015 & -0.038 & -0.047 & -0.059 & -0.021 & -0.002 & 0.157 & -0.067 \\
0.038 & 0.004 & -0.004 & 0.016 & 0.019 & -0.015 & 0.697 & 0.020 & -0.017 & 0.028 & 0.014 & 0.059 & 0.021 & -0.018 \\
0.047 & 0.006 & 0.013 & -0.008 & -0.049 & -0.029 & -0.004 & 0.693 & 0.002 & -0.023 & -0.015 & -0.007 & -0.027 & 0.058 \\
0.001 & 0.031 & 0.009 & -0.007 & 0.018 & 0.018 & 0.050 & 0.020 & 0.662 & 0.010 & -0.006 & 0.044 & -0.012 & 0.015 \\
-4e-5 & -0.047 & 0.049 & -0.003 & 0.014 & 0.009 & -0.043 & 0.022 & 0.040 & 0.647 & -0.028 & -0.042 & 0.002 & 0.003 \\
-0.011 & 0.022 & 0.032 & 0.002 & -0.003 & 0.004 & 0.003 & 0.028 & -0.013 & 0.030 & 0.656 & 0.030 & 0.001 & 0.019 \\
0.020 & -0.033 & 0.002 & 0.016 & 0.003 & -0.019 & -0.021 & -0.013 & -0.017 & -0.032 & -0.023 & 0.597 & 0.050 & -0.023 \\
-0.052 & 0.026 & -0.017 & -0.032 & -0.047 & 0.012 & 0.019 & 0.010 & -0.005 & 0.023 & -0.027 & -0.034 & 0.595 & 0.061 \\
0.026 & -0.036 & -0.010 & 0.022 & 0.001 & -0.034 & 0.024 & -0.034 & 0.008 & -0.070 & -0.008 & -0.060 & 0.044 & 0.601
\end{pmatrix}.$$

For controls, the estimated parameters are $\hat{\boldsymbol{\beta}}_0 = (-0.017, 0.001, 1.5e-4, 0.016)'$, $\hat{\sigma}^2_{\epsilon 0} = 166.287$, $\hat{\sigma}^2_{\omega 0} = 49.93$, $\hat{\sigma}^2_{\eta 0} = 479.548$, $\hat{\nu}_0 = 0.837$, $\hat{s}_0 = 0.038$, and 13×13 matrix $\hat{\mathbf{G}}_0$ equals

$$\begin{pmatrix}
0.805 & 0.094 & -0.028 & -0.177 & -0.032 & -0.197 & -0.065 & -0.139 & -0.719 & 0.214 & -0.142 & 0.142 & 1.179 \\
-0.014 & 0.816 & -0.073 & -0.223 & 0.232 & -0.002 & -0.117 & 0.227 & -0.312 & -0.073 & 0.441 & -0.699 & 0.760 \\
-0.009 & 0.017 & 0.762 & 0.017 & 0.009 & 0.071 & -0.036 & 0.052 & 0.056 & -0.150 & 0.187 & -0.231 & -0.136 \\
0.001 & 0.026 & -0.024 & 0.730 & 0.034 & 0.011 & -0.007 & 0.037 & 0.050 & -0.001 & 0.014 & -0.008 & -0.021 \\
0.033 & -0.057 & 0.038 & 3.8e-4 & 0.704 & 0.081 & 0.036 & -0.027 & 0.089 & -0.039 & 0.069 & 0.037 & -0.231 \\
0.010 & -0.003 & -0.027 & -0.001 & 0.016 & 0.744 & 0.014 & 0.029 & 0.083 & -0.010 & 0.039 & -0.094 & -0.137 \\
0.009 & -0.019 & 0.012 & 0.009 & -0.018 & 0.028 & 0.741 & -0.018 & 0.054 & 0.006 & -0.043 & 0.161 & -0.105 \\
0.016 & -0.011 & 0.012 & -0.024 & -0.037 & -0.028 & 0.027 & 0.674 & -0.103 & 0.041 & -0.055 & 0.105 & 0.191 \\
0.040 & 0.009 & -0.029 & -0.029 & -0.006 & -0.050 & -0.018 & 0.009 & 0.626 & 0.030 & -0.005 & -0.008 & 0.133 \\
-0.008 & -0.013 & 0.036 & 0.016 & 0.009 & 0.032 & 0.010 & -0.028 & 0.039 & 0.650 & 0.002 & 0.039 & -0.144 \\
0.002 & -0.028 & -0.036 & 0.002 & -0.013 & -0.016 & 0.009 & -0.022 & -0.018 & 0.029 & 0.644 & -0.024 & 0.024 \\
-0.017 & 0.042 & 0.035 & -0.008 & 0.017 & 0.034 & -0.031 & 0.034 & 0.008 & -0.028 & 0.066 & 0.555 & 0.010 \\
-0.062 & -0.058 & -0.008 & 0.035 & 0.002 & 0.005 & 0.023 & -0.015 & 0.107 & 0.002 & -0.033 & 0.061 & 0.510
\end{pmatrix}.$$

There is more variability in rsfMRI measurements in ASD children relative to controls, as indicated by higher values of $\hat{\sigma}^2_\epsilon$, $\hat{\sigma}^2_\omega$, and $\hat{\sigma}^2_\eta$. This is consistent with the current understanding of ASD as a heterogeneous condition across patients. Plots of the estimated Bessel functions are provided in Figure 11. This function explains the remaining 30% of variability not incorporated into the second level of the three-level hierarchical model via spectral decomposition of the Moran operator. The estimated range parameters s are almost equal between the two groups. However, the lower estimate of $\hat{\nu}_0$ compared to $\hat{\nu}_1$ yields more negative correlation in controls. The transition matrix $\hat{\mathbf{G}}_1$ for the ASD group indicates dependencies of the first two latent variables on several latent variables from the previous time point. The remaining twelve latent variables show within-variable dependencies only from the previous time. In controls, the transition matrix $\hat{\mathbf{G}}_0$ indicates dependencies of the first three latent variables on several latent variables from the previous time point. Similarly to ASD subjects, the remaining variables only show within-variable dependencies.

6.3.2 Functional Connectivity

The time series of the estimated fMRI signal within each subject were decorrelated. We used the uncorrelated time series to calculate pairwise spatial correlations, and Fisher's z-transformation was performed on subject specific correlations. Analysis of functional connectivity was performed on the difference in means of the transformed scores. Using a local FDR level of 0.1, thirty-nine disrupted hypoconnected links were identified in ASD subjects compared to controls.

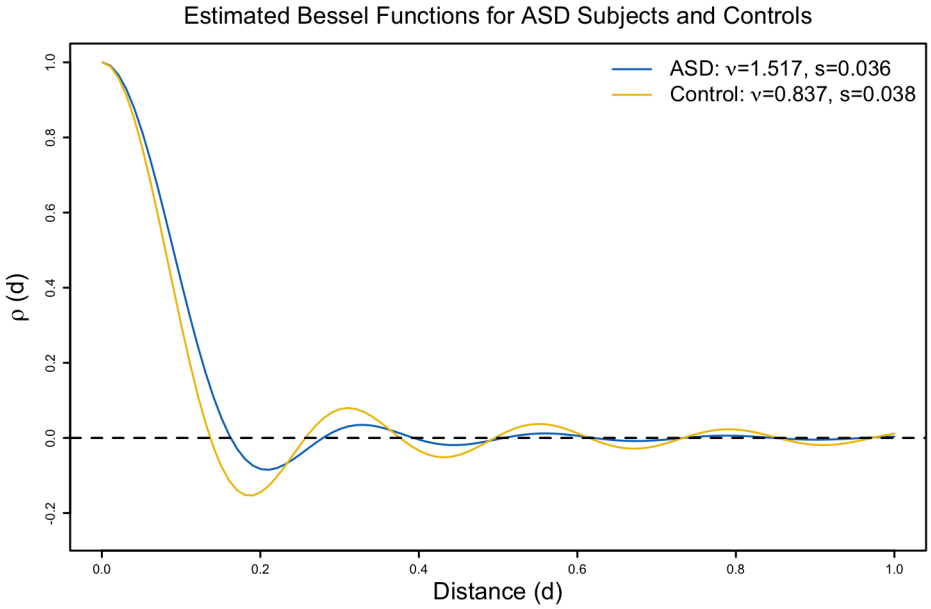


Figure 11: Estimated Bessel functions.

Disrupted connections are provided in Table IX and a network hub plot is provided in Figure 12. The brain region associated with each number is provided in Appendix A. As discussed in Section 2.1, the DMN has been consistently identified as a marker of ASD when analyzing fMRI data. We identified brain regions in this study as part of the DMN as defined by Andrews-Hanna and colleagues (Andrews-Hanna et al., 2014). The numeric value in the DMN column in Table IX corresponds to the number of regions within the link that are part of the DMN. Fourteen (36%) links involve at least one region in the DMN.

The areas of the brain that are involved in the most disrupted connections include the left hemisphere and the temporal lobe. Out of the thirty-nine disruptions, thirty-eight (97%) involve the left hemisphere. More specifically, seventeen (44%) are within the left hemisphere and

TABLE IX SIGNIFICANT LINKS AT FDR LEVEL 0.1

Region 1	Region 2	DMN[a]
Left Frontal Medial Cortex	Right Middle Temporal Gyrus; posterior division	2
Left Planum Polare	Left Temporal Fusiform Cortex; anterior division	0
Left Inferior Temporal Gyrus; posterior division	Left Inferior Temporal Gyrus; anterior division	0
Left Temporal Fusiform Cortex; anterior division	Right Superior Temporal Gyrus; posterior division	1
Left Frontal Medial Cortex	Right Middle Temporal Gyrus; anterior division	2
Left Subcallosal Cortex	Left Frontal Pole	0
Left Inferior Temporal Gyrus; posterior division	Left Middle Temporal Gyrus; anterior division	1
Left Angular Gyrus	Left Inferior Frontal Gyrus; pars triangularis	2
Left Temporal Occipital Fusiform Cortex	Right Temporal Fusiform Cortex; posterior division	0
Right Planum Temporale	Left Parietal Operculum Cortex	0
Left Temporal Fusiform Cortex; anterior division	Left Superior Temporal Gyrus; anterior division	0
Right Middle Temporal Gyrus; posterior division	Left Frontal Pole	1
Left Supracalcarine Cortex	Left Central Opercular Cortex	0
Right Planum Temporale	Left Central Opercular Cortex	0
Right Frontal Orbital Cortex	Left Frontal Pole	0
Left Temporal Pole	Left Inferior Frontal Gyrus; pars triangularis	1
Left Parietal Operculum Cortex	Right Central Opercular Cortex	0
Left Inferior Temporal Gyrus; temporooccipital part	Right Inferior Temporal Gyrus; posterior division	0
Left Cingulate Gyrus; anterior division	Left Frontal Pole	1
Left Juxtapositional Lobule Cortex	Left Frontal Pole	0
Left Occipital Fusiform Gyrus	Left Lingual Gyrus	0
Left Temporal Fusiform Cortex; anterior division	Left Superior Temporal Gyrus; posterior division	1
Left Inferior Temporal Gyrus; anterior division	Right Middle Temporal Gyrus; posterior division	1
Right Temporal Pole	Left Frontal Pole	0
Right Temporal Fusiform Cortex; anterior division	Right Inferior Temporal Gyrus; posterior division	0
Left Planum Temporale	Right Planum Temporale	0
Left Subcallosal Cortex	Right Temporal Pole	0
Left Angular Gyrus	Left Temporal Pole	1
Right Frontal Orbital Cortex	Left Subcallosal Cortex	0
Left Inferior Temporal Gyrus; posterior division	Right Middle Temporal Gyrus; posterior division	1
Right Frontal Orbital Cortex	Left Temporal Pole	0
Left Subcallosal Cortex	Left Temporal Pole	0
Left Occipital Pole	Right Lateral Occipital Cortex; inferior division	0
Left Middle Temporal Gyrus; posterior division	Right Middle Temporal Gyrus; posterior division	2
Left Temporal Pole	Left Frontal Pole	0
Left Temporal Fusiform Cortex; anterior division	Left Subcallosal Cortex	0
Left Inferior Temporal Gyrus; temporooccipital part	Left Inferior Frontal Gyrus; pars opercularis	1
Left Lingual Gyrus	Right Lingual Gyrus	0
Left Juxtapositional Lobule Cortex	Right Pallidum	0

[a]The number of regions within the link in the DMN.

twenty-one (54%) are between the left and right hemispheres. This is consistent with previous studies using Diffusion Tensor Imaging (DTI) in ASD that suggest greater impairment in the left hemisphere (Perkins et al., 2014). Furthermore, twenty-seven (69%) disruptions involve areas in the temporal lobe. This is an extremely clinically relevant finding, since "temporal lobe abnormality in autism is a likely candidate because core symptoms of the disorder center on deficits in language and social behavior, which are frequently accompanied by intellectual impairment" (Bigler et al., 2003). Multiple subregions within the temporal gyrus and temporal fusiform cortex were involved in disrupted links. The middle, inferior, and superior temporal gyrus were involved in seven, seven, and three disruptions, respectively. Moreover, seven disrupted links were identified in the temporal fusiform cortex. As previously noted, Nielson and colleagues identified the posterior middle temporal gyrus and fusiform gyri among regions contributing to the highest classification accuracy in their functional connectivity analysis using data from ABIDE (Nielsen et al., 2013). The fusiform gyrus and middle temporal gyrus are key regions of the social brain network, which "plays an important role in social cognition" defined as "the accumulation of cognitive processes required to comprehend and interact with others" (Kim et al., 2015). The identified disruptions in these key regions may therefore be associated with disruptions in social cognition in ASD.

The left anterior temporal fusiform cortex (#69) is an important region for facial processing and involved in five disrupted connections. It has been noted that failure to incorporate facial expressions into social interactions is "among the most characteristic socialcommunicative impairments in ASD" (Hadjikhani et al., 2007). One disruption identified is between the left

anterior temporal fusiform cortex and the right posterior inferior temporal gyrus (#17). The inferior temporal gyrus is involved in visual processing. Impaired facial processing in ASD may therefore result from disruptions in visual processing. Disruptions in visual processing where identified in Bhaumik et al. (2018a). Furthermore, four disrupted links are identified between the left anterior temporal fusiform cortex and regions of the superior temporal gyrus, including the right posterior superior temporal gyrus (#2), left anterior superior temporal gyrus (#110), left posterior superior temporal gyrus (#3), and the left planum polare (#89). The superior temporal gyrus is involved in sound processing and language comprehension, and disruptions involving this region were also identified in Bhaumik et al. (2018a). The disruptions between the left anterior temporal fusiform cortex and superior temporal gyrus may therefore manifest in an inability to integrate facial queues with verbal communications.

The middle temporal gyrus is a critical component of the social brain and DMN. The right posterior middle temporal gyrus (#10) was involved in five disrupted connections with the left frontal medial cortex (#42), left frontal pole (#5), left anterior inferior temporal gyrus (#16), left posterior inferior temporal gyrus (#18), and left posterior middle temporal gyrus (#11). It is theorized that the posterior middle temporal gyrus plays a role in verbal and non-verbal semantic cognition (Hoffman et al., 2011). Semantic cognition is defined as "our ability to use, manipulate and generalize knowledge that is acquired over the lifespan to support innumerable verbal and non-verbal behaviours" (Ralph et al., 2017). The DSM-5 specifically states "deficits in nonverbal communicative behaviors used for social interaction" as a diagnostic criteria of ASD (Centers for Disease Control and Prevention, 2019). The disruption across the

brain between the left and right posterior middle temporal gyrus may manifest in impaired semantic cognition. The frontal medial cortex is involved several cognitive processes, such as decision-making, response conflict, reward, and action monitoring (Moreira et al., 2016). The disruption observed between the right posterior middle temporal gyrus with the left frontal medial cortex may lead to impairments in using knowledge acquired over the lifespan to make decisions impacting behavior. Furthermore, the observed disrupted connections between the right posterior middle temporal gyrus and regions of the inferior temporal gyrus, which are involved in visual processing, may be associated the "deficits in understanding and use of gestures" when communicating with others (Centers for Disease Control and Prevention, 2019).

The left frontal pole (#5) is involved in seven disrupted links, the most observed in this study. The frontal pole is the frontal part of the prefrontal region of the human brain and is "functionally correlated with the default mode network...which is involved in internally focused tasks" (Liu et al., 2013). Disruptions between the left frontal pole and the right temporal pole (#107), left temporal pole (#108), left anterior cingulate cortex (#51), and right frontal orbital cortex (#60) may be associated with the "deficits in social-emotional reciprocity" (Centers for Disease Control and Prevention, 2019) observed in ASD. It is theorized that regions in the temporal poles are involved in social and emotional processing (Olson et al., 2007). Moreover, the anterior cingulate cortex, which is part of the limbic system, is also involved in emotional processing and vocalization of emotions. The disruptions between these regions and the left frontal pole may therefore manifest in internal reflection at the expense of social and emotional interactions with others. The frontal orbital cortex is also part of the limbic system and is critical

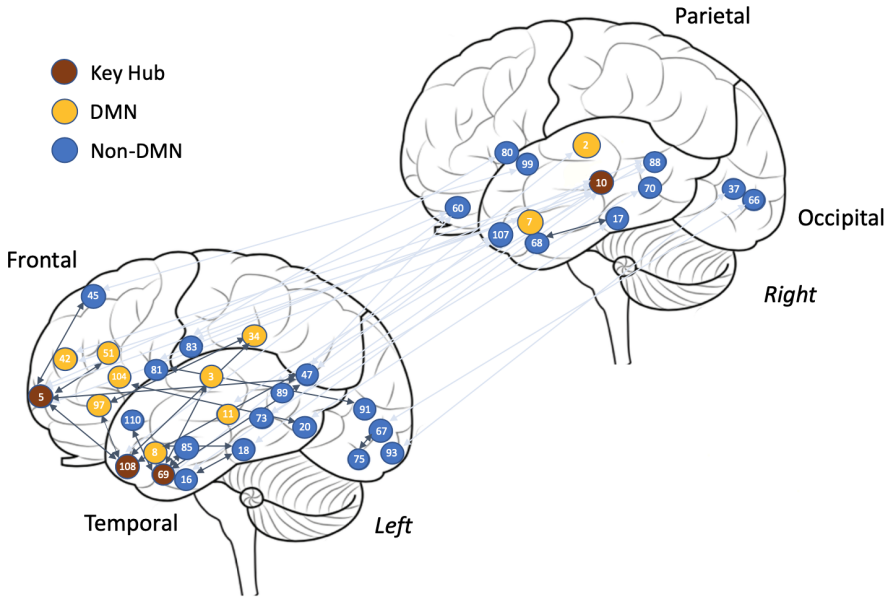


Figure 12: Disrupted connections network.

to multiple psychological functions, including emotional and cognitive processing, learning, and social behavior. A disruption between the left frontal pole and right frontal orbital cortex may affect the behavior of ASD subjects with regard to social interactions and cognitive processes.

6.3.3 Conclusion

In this chapter, we apply the proposed approach to analysis of functional connectivity in ABIDE I, a large multisite dataset of rsfMRI data for ASD subjects and typically developing controls. The sample population consists of 162 ASD subjects and 167 controls between the ages of seven and fourteen from eight ABIDE sites. The Harvard-Oxford atlas was used to define 110 brain regions, yielding 5,995 hypothesis tests of functional connections. We identified thirty-nine disrupted connections at an FDR level of 0.1. Key regions identified include the temporal

lobe, left frontal pole, temporal fusiform gyrus, and middle temporal gyrus. These regions play key roles in decision-making, semantic cognition, social behavior, and emotional processing and correspond well to the DSM-5 definition of ASD. These findings may therefore fill the gap in creating objective criteria necessary for an accurate ASD diagnosis.

CHAPTER 7

DISCUSSION AND CONCLUSION

Neurological diseases contribute substantially to morbidity and mortality worldwide. A deeper understanding of these conditions for early diagnosis and treatment is required to combat their debilitating symptoms. Among the neuroimaging techniques available for visualizing and evaluating these conditions, fMRI has been valuable in describing how regions of the brain communicate. Resting state fMRI studies have played a vital role in evaluating how neurological conditions affect the natural communications that occur within the brain while an individual is at rest. However, application of inference procedures for comparing correlation coefficients in a disease group have ignored the correlation within a time series, violating a key assumption and leading to the risk of misleading results.

In this dissertation, a spatiotemporal model that denoises the observed data and is subsequently used for functional connectivity analysis of rsfMRI data is introduced. Unlike previously proposed model, the model incorporates all brain regions under analysis. Unknown parameters are estimated in a computationally feasible manner using the ECM algorithm. This is partially attributed to incorporation of dimension reduction techniques using Moran's operator. As seen in analysis of functional connectivity of ASD in Chapter 6, the dimension of the dynamic latent model in practice can be substantially reduced. Furthermore, the rsfMRI model is novel in the use of the Bessel covariance function. This function produces greater flexibility in its ability to incorporate negative correlations in addition to positive correlations. In the application of

our approach to ASD, the estimated Bessel function does yield negative correlations. The estimated outcomes within region time series from the spatiotemporal model are subsequently decorrelated, yielding temporally uncorrelated time series for appropriate application of spatial inference procedures using Fisher's z-transformation. We use Efron's local FDR to adjust for the thousands of tests performed simultaneously. We identified thirty-nine clinically meaningful disrupted connectivities in ASD subjects using the proposed approach in a subset of data from the ABIDE initiative.

One of the main motivations for the spatiotemporal model is to denoise the data. A natural followup is to see what the result would be if modeling was not performed and the raw observations were used for analysis of functional connectivity. Thus, for the analysis dataset described in Section 6.1, we decorrelated the raw observations and performed hypothesis testing at the same FDR level of 0.1. Using this approach, none of the links were significant. We therefore conclude the spatiotemporal model has achieved its objective of extracting the true underlying signal from the noise.

While the proposed approach provides improvements in modeling of rsfMRI data, future research should be explored. It has recently been identified that rsfMRI functional connectivity is highly variable across time (Valsasina et al., 2019). This model can be modified to allow for dynamic functional connectivity. One option is to explore the computational feasibility of estimating the transition matrix $\mathbf{G}$ at every time $\mathbf{t}$. Furthermore, in studies containing data from multiple sites, between-site variabilities arise. These multisite collaborations are vital for grasping a greater understanding of heterogenous conditions such as ASD. However, they

introduce additional variability and require proper statistical care. A dynamic spatiotemporal model that incorporates an additional hierarchical level for site should be explored. Meta-analysis has been proposed as a solution to accounting for multisite variability (Bhaumik et al., 2020). Third, the data used for this analysis was restricted to the same number of time points across individuals. Data for subjects with more time points is therefore excluded. Missing data techniques should be explored to improve power and incorporate all available information.

An additional area of research for analysis of functional connectivity would be development of a model that includes multimodal imaging. Multimodal imaging is the combination of different imaging techniques. For example, fMRI and DTI can be performed on a patient at rest at the same time and can improve our understanding of functional connectivity (Tagliazucchi and Laufs, 2015). A model that incorporates data from both techniques may increase power for detection of disruptions due to a disease. Moreover, a model that directly estimates a $\mathsf{T} \times \mathsf{T}$ temporal covariance matrix to use in decorrelation while also accounting for spatial relationships should be explored. Finally, although the approach described in this dissertation was designed for fMRI data from rsfMRI studies, it may be applied to other areas where spatiotemporal data is collected. There are numerous sources of spatiotemporal data, such as microbiome data, mortality and morbidity rates, and economic data, which may benefit from the approach outlined in this work.

APPENDICES

Appendix A

DESCRIPTION OF ROIS

TABLE X DESCRIPTIONS OF HARVARD OXFORD REGIONS

Region Number	Region Description
1	Left Thalamus
2	Right Superior Temporal Gyrus; posterior division
3	Left Superior Temporal Gyrus; posterior division
4	Right Frontal Pole
5	Left Frontal Pole
6	Left Caudate
7	Right Middle Temporal Gyrus; anterior division
8	Left Middle Temporal Gyrus; anterior division
9	Left Putamen
10	Right Middle Temporal Gyrus; posterior division
11	Left Middle Temporal Gyrus; posterior division
12	Left Pallidum
13	Right Middle Temporal Gyrus; temporooccipital part
14	Left Middle Temporal Gyrus; temporooccipital part
15	Right Inferior Temporal Gyrus; anterior division
16	Left Inferior Temporal Gyrus; anterior division
17	Right Inferior Temporal Gyrus; posterior division
18	Left Inferior Temporal Gyrus; posterior division
19	Right Inferior Temporal Gyrus; temporooccipital part
20	Left Inferior Temporal Gyrus; temporooccipital part
21	Left Hippocampus
22	Right Postcentral Gyrus
23	Left Postcentral Gyrus
24	Left Amygdala
25	Right Superior Parietal Lobule
26	Left Superior Parietal Lobule
27	Right Supramarginal Gyrus; anterior division
28	Left Supramarginal Gyrus; anterior division
29	Right Supramarginal Gyrus; posterior division
30	Left Supramarginal Gyrus; posterior division
31	Right Insular Cortex

Appendix A (Continued)

TABLE X DESCRIPTIONS OF HARVARD OXFORD REGIONS (Continued)

Region Number	Region Description
32	Left Insular Cortex
33	Right Angular Gyrus
34	Left Angular Gyrus
35	Right Lateral Occipital Cortex; superior division
36	Left Lateral Occipital Cortex; superior division
37	Right Lateral Occipital Cortex; inferior division
38	Left Lateral Occipital Cortex; inferior division
39	Right Intracalcarine Cortex
40	Left Intracalcarine Cortex
41	Right Frontal Medial Cortex
42	Left Frontal Medial Cortex
43	Left Accumbens
44	Right Juxtapositional Lobule Cortex
45	Left Juxtapositional Lobule Cortex
46	Right Subcallosal Cortex
47	Left Subcallosal Cortex
48	Right Paracingulate Gyrus
49	Left Paracingulate Gyrus
50	Right Cingulate Gyrus; anterior division
51	Left Cingulate Gyrus; anterior division
52	Right Cingulate Gyrus; posterior division
53	Left Cingulate Gyrus; posterior division
54	Right Superior Frontal Gyrus
55	Left Superior Frontal Gyrus
56	Right Precuneous Cortex
57	Left Precuneous Cortex
58	Right Cuneal Cortex
59	Left Cuneal Cortex
60	Right Frontal Orbital Cortex
61	Left Frontal Orbital Cortex
62	Right Parahippocampal Gyrus; anterior division
63	Left Parahippocampal Gyrus; anterior division
64	Right Parahippocampal Gyrus; posterior division
65	Left Parahippocampal Gyrus; posterior division
66	Right Lingual Gyrus
67	Left Lingual Gyrus
68	Right Temporal Fusiform Cortex; anterior division

Appendix A (Continued)

TABLE X DESCRIPTIONS OF HARVARD OXFORD REGIONS (Continued)

Region Number	Region Description
69	Left Temporal Fusiform Cortex; anterior division
70	Right Temporal Fusiform Cortex; posterior division
71	Left Temporal Fusiform Cortex; posterior division
72	Right Temporal Occipital Fusiform Cortex
73	Left Temporal Occipital Fusiform Cortex
74	Right Occipital Fusiform Gyrus
75	Left Occipital Fusiform Gyrus
76	Right Middle Frontal Gyrus
77	Left Middle Frontal Gyrus
78	Right Frontal Operculum Cortex
79	Left Frontal Operculum Cortex
80	Right Central Opercular Cortex
81	Left Central Opercular Cortex
82	Right Parietal Operculum Cortex
83	Left Parietal Operculum Cortex
84	Right Planum Polare
85	Left Planum Polare
86	Right Heschl's Gyrus (includes H1 and H2)
87	Left Heschl's Gyrus (includes H1 and H2)
88	Right Planum Temporale
89	Left Planum Temporale
90	Right Supracalcarine Cortex
91	Left Supracalcarine Cortex
92	Right Occipital Pole
93	Left Occipital Pole
94	Right Thalamus
95	Right Caudate
96	Right Inferior Frontal Gyrus; pars triangularis
97	Left Inferior Frontal Gyrus; pars triangularis
98	Right Putamen
99	Right Pallidum
100	Right Hippocampus
101	Right Amygdala
102	Right Accumbens
103	Right Inferior Frontal Gyrus; pars opercularis
104	Left Inferior Frontal Gyrus; pars opercularis
105	Right Precentral Gyrus

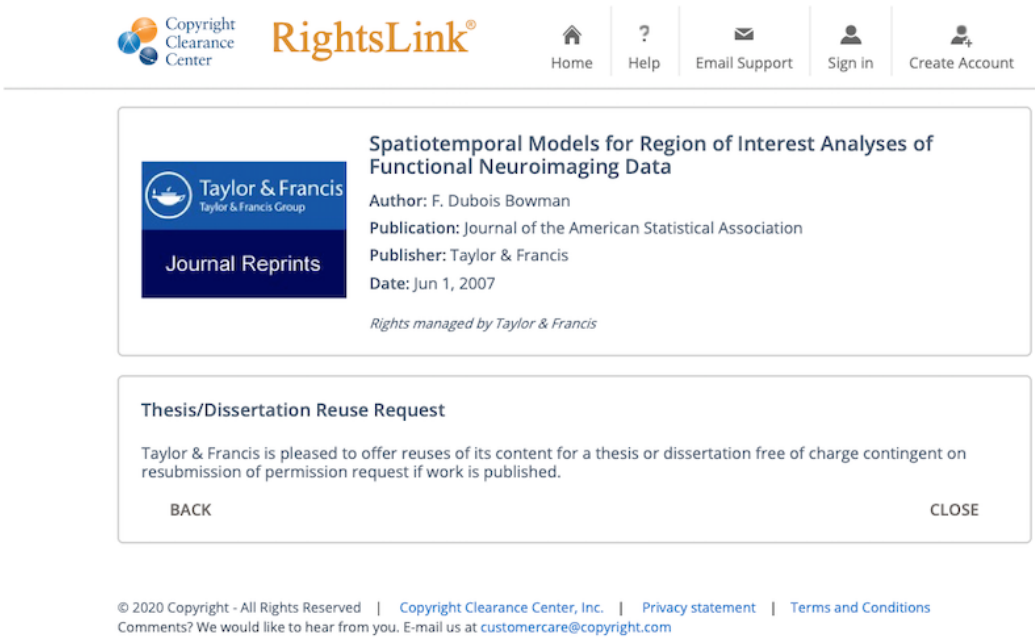
Appendix A (Continued)

TABLE X DESCRIPTIONS OF HARVARD OXFORD REGIONS (Continued)

Region Number	Region Description
106	Left Precentral Gyrus
107	Right Temporal Pole
108	Left Temporal Pole
109	Right Superior Temporal Gyrus; anterior division
110	Left Superior Temporal Gyrus; anterior division

Appendix B

DISSERTATION POLICY FOR TAYLOR AND FRANCIS CONTENT



CITED LITERATURE

Abraham, A., Milham, M., DiMartino, A., Craddock, R., Samaras, D., Thirion, B., and Varoquaux, G.: Deriving reproducible biomarkers from multi-site resting-state data, an autism-based example. Neuroimage, 147:736–745, 2017.

Anderson, J. S., Ferguson, M. A., and Nielsen, J. A.: Functional mri in autism. In Imaging the Brain in Autism, eds. M. F. Casanova, A. S. El-Baz, and J. S. Suri, pages 289–323. Springer Science and Business Media, New York, 2013.

Andrews-Hanna, J. R., Smallwood, J., and Spreng, R. N.: The default network and self-generated thought: component processes, dynamic control, and clinical relevance. Ann N Y Acad Sci, 1316(1):29–52, 2014.

Bailey, T. C. and Gatrell, A. C.: Interactive Spatial Data Analysis. Routledge, 1995.

Baio, J., Wiggins, L., Christensen, D., et al.: Prevalence of autism spectrum disorder among children aged 8 years, autism and developmental disabilities monitoring network, 11 sites, united states, 2014. MMWR Surveill Summ 2018, 67(SS-6):1–23, 2018.

Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B, 57:289–300, 1995.

Besseling, R., Lamerichs, R., Michels, B., Heunis, S., deLouw, A., Tijhuis, A., Bergmans, J., and Aldenkamp, B.: Functional network abnormalities consistent with behavioral profile in autism spectrum disorder. Psychiatry Research: Neuroimaging, 275:43–48, 2018.

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

Bhaumik, D. K., Song, Y., Zhao, D., and Ajilore, O.: Controlling false disrupted connectivities in neuroimaging studies. Journal of Biostatistics and Biometrics, 108(100008), 2018.

Bhaumik, R., Jie, F., and Rupnow, R. A.: Discovering disrupted connectivities in neuroimaging studies: A meta analytic approach. Annals of Biostatistics and Biometric Applications, 3(5):1–10, 2020.

Bigler, E. D., Tate, D. F., Neeley, E. S., Wolfson, L. J., Miller, M. J., Rice, S. A., Cleavinger, H., Anderson, C., Coon, H., Ozonoff, S., Johnson, M., Dinh, E., Lu, J., McMahon, W., and Lainhart, J. E.: Temporal lobe, autism, and macrocephaly. American Journal of Neuroradiology, 24(10):2066–2076, 2003.

Biswal, B.: Resting state fmri: A personal history. NeuroImage, 62:938–944, 2012.

Biswal, B., Yetkin, F., Haughton, V., and Hyde, J.: Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magnetic Resonance in Medicine, 34:537–541, 1995.

Boots, B. and Tiefelsdorf, M.: Global and local spatial autocorrelation in bounded regular tessellations. J Geog Syst, 2:319, 2000.

Bowman, F. D.: Spatio-temporal modeling of localized brain activity. Biostatistics, 6(4):558–575, 2005.

Bowman, F. D.: Spatiotemporal models for region of interest analyses of functional neuroimaging data. Journal of the American Statistical Association, 102(478):442–453, 2007.

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

Bradley, J. R., Holan, S. H., and Wikle, C. K.: Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. The Annals of Applied Statistics, 9(4):1761–1791, 2015.

Cameletti, M.: Stem: Spatio-temporal models in R, 2012. R package version 1.0.

Caponera, A., Denti, F., Rigon, T., Sottosanti, A., and Gelfand, A.: Hierarchical spatio-temporal modeling of resting state fmri data. In Studies in Neural Data Science, eds. A. Canale, D. Durante, L. Paci, and B. Scarpa, pages 111–130. Springer, 2018.

Castruccio, S., Ombao, H., and Genton, M. G.: A scalable multi-resolution spatio-temporal model for brain activation and connectivity in fmri data. Biometrics, 74:823–833, 2018.

Centers for Disease Control and Prevention: Autism spectrum disorder (asd) diagnostic criteria. `https://www.cdc.gov/ncbddd/autism/hcp-dsm.html`, 2019. Accessed: 2020-02-25.

Chen, C. P., Keowna, C. L., Jahedi, A., Naira, A., Pflieger, M. E., Bailey, B. A., and Muller, R.-A.: Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. Neuroimage Clinical, 8:238–245, 2015.

Chen, G., Chen, G., Xie, C., and Li, S.-J.: Negative functional connectivity and its dependence on the shortest path length of positive network in the resting-state human brain. Brain Connectivity, 1(22), 2011.

Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., Khundrakpam, B. S., Lewis, J. D., Li, Q., Milham, M., Yan, C., and Bellec, P.: The neuro bureau preprocessing initiative, open sharing of preprocessed neuroimaging data and derivatives. Frontiers in Neuroinformatics, 7, 2013.

Cressie, N. and Wikle, C. K.: Statistics for Spatio-Temporal Data. John Wiley and Sons, 2011.

Damoiseaux, J., Rombouts, S. A., Barkhof, F., Scheltens, P., Stam, C., Smith, S., and Beckmann, C. F.: Consistent resting-state networks across healthy subjects. Proceedings of the Natural Academy of Sciences, 103(37):13838–13853, 2006.

Dempster, A., Laird, N., and Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B, 39(1):1–38, 1977.

DiMartino, A.: The autism brain imaging data exchange: Towards large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry, 19(6):659–667, 2014.

DiMartino, A., Kelly, C., Grzadzinski, R., Zuo, X.-N., Mennes, M., Mairena, M. A., Lord, C., Castellanos, F. X., and Milham, M. P.: Aberrant striatal functional connectivity in children with autism. Biological Psychiatry, 69:847–856, 2011.

Durbin, J. and Koopman, S.: Time Series Analysis by State Space Methods. Oxford University Press, 2012.

Efron, B.: Size, power, and false discovery rates. The Annals of Statistics, 35(4):1251–1377, 2007.

Efron, B.: Simultaneous inference, when should hypothesis testing problems be combined? The Annals of Applied Statistics, 2(1):197–223, 2008.

Efron, B. and Tibshirani, R.: Using specially designed exponential families for density estimation. Annals of Statistics, 24:2431–2461, 1996.

Fasso, A. and Cameletti, M.: The em algorithm in a distributed computing environment for modelling environmental space-time data. Environmental Modelling and Software, 24:1027–1035, 2009.

Fisher, R.: Statistical Methods for Research Workers. Oliver and Boyd, 1925.

Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C., and Frackowiak, R.: Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping, 2:189–210, 1995.

George, B. and Aban, I.: Selecting a separable parametric spatiotemporal covariance structure for longitudinal imaging data. Statistics in Medicine, 34:145–161, 2015.

Gibbons, R. D., Lazar, N. A., Bhaumik, D. K., Sclove, S. L., Chen, H. Y., Thulborn, K. R., Sweeney, J. A., Hur, K., and Patterson, D.: Estimation and classification of fmri hemodynamic response patterns. Neuroimage, 2:804–814, 2004.

Givens, G. H. and Hoeting, J. A.: Computational Statistics. John Wiley and Sons, 2013.

Glover, G. H.: Overview of functional magnetic resonance imaging. Neurosurgery Clinics of North America, 22(2):133–139, 2011.

Gossl, C., Auer, D., and Fahrmeir, L.: Bayesian spatiotemporal inference in functional magnetic resonance imaging. Biometrics, 57:554–562, 2001.

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V.: Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. Proceedings of the Natural Academy of Sciences, 100(1):253–258, 2003.

Hadjikhani, N., Joseph, R. M., Snyder, J., and Tager-Flusberg, H.: Abnormal activation of the social brain during face perception in autism. Human Brain Mapping, 28:441–449, 2007.

Harrison, L., Penny, W., Ashburner, J., Trujillo-Barreto, N., and Friston, K.: Diffusion-based spatial priors for imaging. NeuroImage, 38(4):677–695, 2007.

Hoffman, P., Pobric, G., Drakesmith, M., and Ralph, M. A. L.: Posterior middle temporal gyrus is involved in verbal and non-verbal semantic cognition: Evidence from rtms. Aphasiology, 26:1119–1130, 2011.

Hughes, J. and Haran, M.: Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 75:139–159, 2013.

Kalman, R.: A new approach to linear filtering and prediction problems. Transactions of the ASME-Journal of Basic Engineering, 82:35–45, 1960.

Katanoda, K., Matsuda, Y., and Sugishita, M.: A spatio-temporal regression model for he analysis of functional mri data. Neuroimage, 17(3):1415–1428, 2002.

Kessy, A., Lewin, A., and Strimer, K.: Optimal whitening and decorrelation. The American Statistician, 72(4):309–314, 2018.

Kim, S. G. and Bandettini, P. A.: Principles of functional mri. In BOLD fMRI: A Guide to Functional Imaging for Neuroscientists, eds. S. H. Faro and F. B. Mohamed, pages 3–22. Springer Science + Business Media, LLC, 2010.

Kim, S.-Y., Choi, U.-S., Park, S.-Y., Oh, S.-H., Yoon, H.-W., Koh, Y.-J., Im, W.-Y., Park, J.-I., Song, D.-H., Cheon, K.-A., and Lee, C.-U.: Abnormal activation of the social brain network in children with autism spectrum disorder, an fmri study. Psychiatry Investigation, 12(1):37–45, 2015.

Lee, K.-J., Jones, G. L., Caffo, B. S., and Bassett, S. S.: Spatial bayesian variable selection models on functional magnetic resonance imaging time-series data. Bayesian Analysis, 9(3):699–732, 2014.

Lee, M. H., Smyser, C. D., and Shimony, J. S.: Resting state fmri: a review of methods and clinical applications. American Journal of Neuroradiology, 34(10):1866–1872, 2013.

Leopold, D. A., Murayama, Y., and Logothetis, N. K.: Very slow activity fluctuations in monkey visual cortex: Implications for functional brain imaging. Cerebral Cortex, 13:422–433, 2003.

Lindquist, M. A.: The statistical analysis of fmri data. Statistical Science, 33(4):439–464, 2008.

Liu, H., Qin, W., Li, W., Fan, L., Wang, J., Jiang, T., and Yu, C.: Connectivity-based parcellation of the human frontal pole with diffusion tensor imaging. The Journal of Neuroscience, 33(16):6782–6790, 2013.

Liu, Y., Sarkar, S. K., and Zhao, Z.: A new approach to multiple testing of grouped hypotheses. Journal of Statistical Planning and Inference, 179:1–14, 2016.

Masi, A., DeMayo, M. M., Glozier, N., and Guastella, A. J.: An overview of autism spectrum disorder, heterogeneity and treatment options. Neuroscience Bulletin, 33(2):183–193, 2017.

McLachlan, G. J. and Krishnan, T.: The EM Algorithm and Extensions. John Wiley and Sons, Inc., 1997.

Meng, X. L. and Rubin, D. B.: Maximum likelihood estimation via the ecm algorithm: a general framework. Biometrika, 80(2):267–278, 1993.

Moran, P.: Notes on continuous stochastic phenomena. Biometrika, 37:17–23, 1950.

Moreau, J. F., Find, E. L., Hartman, M. E., Angus, D. C., Bell, M. J., Linde-Zwirble, W. T., and Watson, R. S.: Hospitalizations of children with neurological disorders in the united states. Pediatric Critical Care Medicine, 14(8):801–810, 2013.

Moreira, P. S., Marques, P., and Magalhaes, R.: Identifying functional subdivisions in the medial frontal cortex. The Journal of Neuroscience, 36(44):11168–11170, 2016.

Musgrove, D., Young, D. S., Hughes, J., and Eberly, L. E.: A sparse areal mixed model for multivariate outcomes with an application to zero-inflated census data. In Modern Statistical Methods for Spatial and Multivariate Data, ed. N. Diawara, pages 51–74. Springer, 2019.

Musgrove, D. R., Hughes, J., and Eberly, L. E.: Fast, fully bayesian spatiotempral inference for fmri data. Biostatistics, 17(2):291–303, 2016.

Neurological Disorders Collaborator Group, G. .: Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurology, 18(5):459–480, 2019.

Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., Lainhart, J. E., and Anderson, J. S.: Multisite functional connectivity mri classification of autism, abide results. Frontiers in Human Neuroscience, 7(599):1–12, 2013.

O'Connor, D. and Devoto, A.: Autism brain imaging data exchange, abide i. `http://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html`, 2016. Accessed: 2019-08-30.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, Seong-Gi, K., Merkle, H., and Ugurbil, K.: Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci, 89:5951–5955, 1992.

Olson, I. R., Plotzker, A., and Ezzyat, Y.: The enigmatic temporal pole: a review of findings on social and emotional processing. Brain, 130(7):1718–1731, 2007.

Penny, W., Trujilllo-Barreto, N., and Friston, K.: Bayesian fmri time series analysis with spatial priors. Neuroimage, 24(2):350–362, 2005.

Perkins, T. J., Stokes, M. A., McGillivray, J. A., Mussap, A. J., Cox, I. A., Maller, J. J., and Bittar, R. G.: Increased left hemisphere impairment in high-functioning autism: A tract based spatial statistics study. Psychiatry Research: Neuroimaging, 224:119–123, 2014.

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

Rai, S. N. and Matthews, D. E.: Improving the em algorithm. Biometrics, 49(2):587–591, 1993.

Ralph, M. A. L., Jefferies, E., Patterson, K., and Rogers, T. T.: The neural and computational bases of semantic cognition. Nature Reviews Neuroscience, 18:42–55, 2017.

Rao, S., Binder, J., Bandettini, P., Hammeke, T., Yetkin, F., Jesmanowicz, A., Lisk, L., Morris, G., Mueller, W., Estkowski, L., Wong, E., Haughton, V., and Hyde, J.: Functional magnetic resonance imaging of complex human movements. Neurology, 43(11):2311, 1993.

Sankaran, K. and Holmes, S.: structssi, simultaneous and selective inference for grouped or hierarchically structured data. Journal of Statistics Software, 59(13):1–21, 2014.

Schlather, M.: Construction of covariance functions and unconditional simulation of random fields. In Advances and Challenges in Space-time Modelling of Natural Events, eds. P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, and S. Zegar, pages 25–54. Springer, 2012.

Shumway, R. H. and Stoffer, D. S.: Time Series Analysis and Its Applications. Springer, 2017.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., and Beckmann, C. F.: Correspondence of the brain's functional architecture during activation and rest. PNAS, 106(31):13040–13045, 2009.

Tagliazucchi, E. and Laufs, H.: Multimodal imaging of dynamic functional connectivity. Frontiers in Neurology, 6(10):1–9, 2015.

Tiefelsdorf, M. and Griffith, D. A.: Semiparametric filtering of spatial authocorrelation: the eigenvector approach. Environment and Planning, 39:1193–1221, 2007.

Uddin, L., Dajani, D., Voorheis, W., Bednarz, H., and Kana, R.: Progress and roadblocks in the search for brain-based biomarkers of autism and attention-deficit,hyperactivitydisorder. Translational Psychiatry, 7:e1218, 2017.

Uddin, L. Q., Supekar, K., Lynch, C. J., Khouzam, A., Phillips, J., Feinstein, C., Ryali, S., and Menon, V.: Salience network-based classification and prediction of symptom severity in children with autism. JAMA Psychiatry, 70(8):869–879, 2013.

Uddin, L. Q., Supekar, K., and Menon, V.: Reconceptualizing functional brain connectivity in autism from a developmental perspective. Frontiers in Human Neuroscience, 7(458):195–206, 2013.

Valsasina, P., DeLaCruz, M. H., Filippi, M., and Rocca, M. A.: Characterizing rapid fluctuations of resting state functional connectivity in demyelinating, neurodegenerative, and psychiatric conditions: From static to time-varying analysis. Frontiers in Neuroscience, 13(618):1–29, 2019.

VanDenHeuvel, M. P. and Pol, H. E.: Exploring the brain network: A review on resting-state fmri functional connectivity. Statistical Science, 33(4):439–464, 2008.

Wikle, C. K.: Modern perspectives on statistics for spatio-temporal data. Wiley Periodicals, Inc, 7:86–98, 2015.

Woolrich, M., Behrens, T., and Smith, S.: Constrained linear basis sets for hrf modelling using variational bayes. Neuroimage, 21:1748–1761, 2004.

Woolrich, M. W., Jenkinson, M., Brady, J. M., and Smith, S. M.: Fully bayesian spatio-temporal modeling of fmri data. IEEE Transactions on Medical Imaging, 23(2):213–231, 2004.

World Health Organization: Neurological disorders, public health challenges. https://www.who.int/mental_health/neurology/neurodiso/en/, 2006. Accessed: 2020-05-28.

Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., and Evans, A.: A general statistical analysis for fmri data. Neuroimage, 15:1–15, 2002.

Wu, C. J.: On the convergence of the em algorithm. The Annals of Statistics, 11(1):95–103, 1983.

Xu, K. and Wikle, C. K.: Estimation of parameterized spatio-temporal dynamic models. Journal of Statistical Planning and Inference, 137:567–588, 2007.

Zhang, L., Guindani, M., and Vannucci, M.: Bayesian models for functional magnetic resonance imaging data analysis. Wiley Periodicals, 7:21–41, 2015.

Zhang, L., Guindani, M., Versace, F., Engelmann, J. M., and Vannucci, M.: A spatiotemporal non-parametric bayesian model of multi-subject fmri data. Annals of Applied Statistics, 10(2):638–666, 2016.

VITA

NAME Rawan Allozi Rupnow

EDUCATION B.A., Anthropology and Public Health, Washington University in St. Louis, St. Louis, MO, 2009

M.S., Biostatistics, Johns Hopkins University, Baltimore, MD, 2012

Ph.D., Biostatistics, University of Illinois, Chicago, IL, 2020

PUBLICATIONS Bhaumik, R., Jie, F., and Rupnow, R.A.: Discovering disrupted connectivities in neuroimaging studies: a meta analytic approach. Annals of Biostatistics and Biometric Applications, 3(5):1-10, 2020.

Jie, F., Rupnow, R.A., Bhaumik, D.K., Bhaumik, R., and Sinha, B.K.: Analysis of functional connectivity data by hierarchical bayesian models and applications in autism study. International Journal of Statistical Sciences, 16:51-68, 2018.

Ekoue, D.N., Ansong, E., Hong, L.K., Nonn, L., Macias, V., Deaton, R., Rupnow, R., Gann, P.H., Kajdacsy-Ball, A., and Diamond, A.M.: GPX1 localizes to the nucleus in prostrate epithelium and its levels are not associated with prostate cancer recurrence. Antioxidants (Basel), 7:E167, 2018.

Demirtas H., Allozi R., Hu Y., Inan G., and Ozbek L.: Joint generation of binary, ordinal, count, and normal data with specified marginal and association structures in monte-carlo simulations. In Monte-Carlo Simulation-Based Statistical Modeling, eds. D.G. Chen and J.D. Chen, pages 3-15, Springer, 2018.

Oh, J.H., Craft, J., Allozi, R., Vaidya, M., Meng, Y., Deasy, J.O., Bradley, J.D., and El Naqa, I.: A Bayesian network approach for modeling local failure in lung cancer. Physics in Medicine and Biology 56(6):1635-1651, 2011.

Allozi, R., Li, X.A., White, J., and Apte, A., and Tai, A., and Michalski J.M., and Bosch W.R., and El Naqa, I.: Tools for consensus analy-

sis of experts' contours for radiotherapy structure definitions. Radiotherapy and Oncology, 97(3):572-578, 2010.

Apte, A., Allozi, R., Pereira, G., Johnson, M., Mansur, D., and El Naqa, I.: A graphical tool and methods for assessing margin definition from daily image deformations. Journal of Radiation Oncology Informatics, 2(1):9-19, 2010.

Oh, J.H., Apte, A., Allozi, R., and El Naqa, I.: Towards prediction of radiation pneumonitis arising from lung cancer patients using machine learning approaches. Journal of Radiation Oncology Informatics, 1(1):30-43, 2009.