SPECIAL TOPICS IN LATENT VARIABLE MODELS WITH SPATIALLY AND TEMPORALLY CORRELATED LATENT VARIABLES

Rachel C. Nethery

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2017

Approved by:

Young Truong

Amy Herring

Alana Campbell

Yun Li

Eric Bair

ABSTRACT

Rachel C. Nethery: Special Topics in Latent Variable Models with Spatially and Temporally Correlated Latent Variables
(Under the direction of Young Truong)

The term latent variable model (LVM) refers to any statistical procedure that utilizes information contained in a set of observed variables to construct a set of underlying latent variables that drive the observed values and associations. Independent component analysis (ICA) is a LVM that separates recorded mixtures of signals into independent source signals, called independent components (ICs). ICA is popular tool for separating brain signals of interest from artifacts and noise in electroencephalogram (EEG) data. Due to challenges in the estimation of uncertainties in ICA, standard errors are not generally estimated alongside ICA estimates and thus ICs representing brain signals of interest cannot be distinguished through a statistical hypothesis testing framework. In Chapter 2 of this dissertation, we propose a bootstrapping algorithm for ICA that produces bootstrap samples that retain critical correlation structures in the data. These are used to compute uncertainties for ICA parameter estimates and to construct a hypothesis test to identify ICs representing brain activity, which we demonstrate in the context of EEG functional connectivity. In Chapter 3, we extend this bootstrapping approach to accommodate pre-ICA dimension reduction procedures, and we use the resulting method to compare popular strategies for pre-ICA dimension reduction in EEG research.

In the final chapter, we turn our attention to another LVM, factor analysis, which utilizes the covariance structure of a set of correlated observed variables to model a smaller number of unmeasured underlying variables. A spatial factor analysis (SFA) model can be used to quantify the social vulnerability of communities based on a set of observed social variables. Current SFA methodology is ill-equipped to handle spatial misalignment in the observed variables. We propose a joint spatial factor analysis model that identifies a common set of latent variables underlying spatially misaligned observed variables and produces results at the level of the smallest spatial units, thereby minimizing loss of information. We apply this model to spatially misaligned data to construct an

index of community social vulnerability for Louisiana, which we integrate with Louisiana flood data to identify communities at high risk during natural disasters, based on both social and geographic features.

To my parents, whose constant love, support, and encouragement made this work possible.

ACKNOWLEDGEMENTS

I owe this accomplishment, along with anything else I have achieved in life, to my parents, whose unwavering support, through the best of times and the worst of times, has given me the courage, the confidence, and the inspiration to pursue my dreams. Mom and Dad, thank you for your love, encouragement, and generosity during these 22 years of school. You must have thought I was crazy at times; yet, you never failed to provide me with calm guidance and reassurance. This achievement is as much yours as mine. I love you. To the rest of my family, I am overwhelmed by the kindness and thoughtfulness you all show towards me, even though I've been away for eight years. I am incredibly fortunate to have such a large and strong support system, which has enabled me to persevere through the tough times. Thank you.

To James, thank you for all the time you've dedicated to help me with my work, for the encouragement you've given me, and, most importantly, for making my life so bright and happy. I am so lucky to have you in my life. I love you. I also want to thank my friends, both near and far, for bringing me joy and showing me support in a myriad of ways.

Over the past five years at UNC, innumerable people have contributed to my shaping as a statistician and have provided insights and inspiration for the work that follows. Of these, I would first like to thank my advisor and friend, Dr. Young Truong. You have inspired me both personally and professionally with your passion for statistics and your kind spirit. I am forever grateful for your invaluable statistical training and guidance, for your dedication to this project, and for your encouragement through the highs and lows of this process. You are a wonderful mentor and role model, and I look forward to continuing our work on this exciting topic in the years to come.

I also owe a huge debt of gratitude to Dr. Amy Herring for the countless ways in which she has supported me throughout graduate school. Amy, I am unspeakably grateful for your efforts not only to keep me funded but to find research projects that excited and engaged me. I leave UNC enthusiastic about working in biostatistics and hopeful about the impact this work can have, and this optimism is due in large part to the inspiration these projects have provided me.

I would like to say thanks to the remaining members of my dissertation committee, Dr. Alana Campbell, Dr. Yun Li, and Dr. Eric Bair, for your insights and contributions to this work. To Dr. Stephanie Engel, thank you for inviting me into your fantastic and exciting research, and for becoming a mentor and a friend in the process. I would also like to thank Dr. Richard Kwok, Dr. Dale Sandler, and everyone in the Epidemiology Branch at NIEHS. You welcomed me into your community last summer, initiated a project which turned into a paper in my dissertation, and offered tremendous guidance and encouragement.

To all the faculty, staff, and students in the UNC Biostatistics Department who have made my time here wonderful, thank you. In particular, I could never have made it through this process if not for my officemates (past and present). You are some of the most brilliant, thoughtful, and dedicated people I have ever known. You have taught me more about statistics than any professor and have selflessly dedicated your time to help me solve problems in my work, providing invaluable perspectives and finding solutions that I could never have found on my own. More importantly, though, you have kept me sane. Nearly every day of the past five years, you have made me laugh, commiserated with me, encouraged me, and shown incredible kindness and thoughtfulness that I can never repay. Thank you for everything.

This work was funded in part by the Intramural Research Program of the NIH, National Institute of Environmental Sciences (ZO1 ES 102945) and the NIH Summer Internship Program in Biomedical Research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
AR	Auto-regressive
BSS	Blind source separation
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
CICA	ColorICA
DIC	Deviance Information Criterion
EEG	Electroencephalogram
FBI	Federal Bureau of Investigation
FEMA	Federal Emergency Management Agency
fMRI	Functional magnetic resonance imaging
IC	Independent component
ICA	Independent component analysis
L0CC	Lag 0 cross-correlation
L1AC	Lag 1 auto-correlation
LAU	Large areal unit
LVM	Latent variable model
MCMC	Markov chain Monte Carlo
NFIP	National Flood Insurance Program
NP	Non-parametric
PBDR	Procedure based dimension reduction
PCA	Principal component analysis
PM	Posterior mean
SVD	Singular value decomposition
SE	Standard error
SFA	Spatial factor analysis
SNR	Signal-to-noise ratio
SP	Semi-parametric

TBDR	Theory based dimension reduction
UC	Uncorrelated component
UCR	Uniform Crime Report

CHAPTER 1: LITERATURE REVIEW

1.1 Introduction

With the rise of technology and the resulting advances in data collection and storage capabilities, so too have statistical procedures used to reduce the dimensionality of large datasets and dissect complicated data structures experienced a surge in popularity. The term latent variable model (LVM) refers to a broad class of statistical procedures that model a set of observed variables as a function of a set of unobserved or latent variables. Latent variable modeling procedures utilize information contained in the set of observed variables to construct a set of latent variables underlying the observed variables that drive their values and the connections between them. These latent variables, in conjunction with quantifiers of the relationships between the observed and latent variables which are estimated by latent variable modeling procedures, are used for a variety of purposes including dimension reduction and data summarization, identification of variable clustering structures, and unmixing of recorded signals. Hence, LVMs have a wide range of applications, particularly in big data settings.

In many application areas of LVMs, particularly in the biomedical and public health arenas, data are collected over time and/or space, and observations from the same variable contain temporal and/or spatial correlation, making standard LVMs invalid due to their assumption of independent observations within variables. Thus, extensions to the standard models have been developed to allow for spatial or temporal correlation in the latent variables, which induces the correlation in the observed variables. Though general methodologies exist for accounting for spatial and temporal correlation in latent variable models, they are often inadequate to address the challenges of real data and need to be extended further to increase their applicability. In this chapter, I review the literature on two types of LVMs, one of which allows for temporally correlated latent variables and one of which allows for spatially correlated latent variables. I then identify challenges presented by biomedical and public health data that cannot be sufficiently resolved with existing latent variable model methodology.

1.2 Independent Component Analysis

Blind source separation (BSS) techniques are a class of statistical and machine learning methods that can be applied to multivariate data generated by the linear mixing of signals in order to recover the original, unmixed signals, also known as latent sources of activity. Often, the use of BSS methods is motivated by a situation akin to the classic “cocktail party problem”, in which the mixing conversations of a roomful of mingling partygoers are being recorded throughout the party by some fixed number of devices/sensors placed in various locations across the room. Afterwards, the goal is to extract the speech of each individual partygoer from the mixed recordings and understand how the voice of each partygoer contributed to the recording from each sensor. In this situation, the recordings from the sensors are the multivariate mixture data, and the voices of the individual partygoers are the latent sources of activity. Many types of experiments across a variety of disciplines collect data in a fashion that results in mixtures of signals analogous to the recordings from the cocktail party, making BSS a topic of interest for many researchers.

1.2.1 The ICA Model

Independent component analysis (ICA) falls within the scope of BSS methods, distinguishing itself from its BSS kinfolk by its aim of uncovering latent sources of activity that are statistically independent of one another, also called independent components (ICs). ICA has proven to be a useful, data-driven tool for a diverse range of applications; in particular, it has become popular as a means of analyzing data collected from biomedical devices, which often record a mixture of signals originating from sources of interest, nuisance sources such as irrelevant bodily processes (known as artifacts), and noise. Functional magnetic resonance imaging (fMRI), electroencephalogram (EEG), and electrocardiogram research have all been fruitful areas of ICA application and development (Calhoun and Adali, 2006; He et al., 2006; Makeig et al., 1996; McKeown et al., 1997; Vigário et al., 2000; Wisbeck et al., 1998).

Let k denote the number of sensors/recording locations, T denote the number of discrete time points at which recordings are made. Then, the ICA model can be written mathematically as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \tag{1.1}$$

where $\boldsymbol{X}$ is the $k \times T$ matrix of observed mixtures with each row containing the recording from a given sensor over time (each column, thus, contains the recordings from all of the sensors at a single point in time). Each row of the $k \times T$ matrix $\boldsymbol{S}$ contains the unobserved values of a single IC over time. Finally, $\boldsymbol{A}$ is a $k \times k$ matrix of linear mixing coefficients which represent the contribution of each IC to the recording at each sensor. The columns of $\boldsymbol{A}$ are sometimes called spatial maps, as each column maps a single IC back onto the recording space. Henceforth, we refer to $\boldsymbol{X}$ as the data or mixture matrix, $\boldsymbol{A}$ as the mixing matrix, and $\boldsymbol{S}$ as the IC matrix. For convenience, we also define the unmixing matrix as $\boldsymbol{W} = \boldsymbol{A}^{-1}$. Note that this model formulation assumes that the number of sensors, k, is equal to the number of ICs. Below we address the procedure applied to relax this assumption when the number of ICs is assumed to be less than the number of sensors.

The primary goal of ICA is to obtain an estimate of the unmixing matrix, denoted $\hat{\boldsymbol{W}}$, in a manner that imposes as few assumptions as possible on the form and the distribution of the ICs. Given $\hat{\boldsymbol{W}}$, the estimates of the ICs can then be constructed by $\hat{\boldsymbol{S}} = \hat{\boldsymbol{W}}\boldsymbol{X}$. Most ICA estimation algorithms, such as the popular FastICA (Hyvarinen and Oja, 2000) and Infomax (Bell and Sejnowski, 1995) algorithms, take the approach of finding a $\hat{\boldsymbol{W}}$ that maximizes the statistical independence in the corresponding estimated ICs in $\hat{\boldsymbol{S}}$.

1.2.2 Popular Approaches to ICA Estimation

1.2.2.1 Entropy and Mutual Information

If the goal of ICA is to uncover maximally independent ICs, before embarking on the details of estimation schemes it is necessary to first understand how to measure statistical independence. A popular means of measuring the independence of random variables is through functions of the entropy. Entropy (also called the differential entropy when referring to continuous random variables) measures the degree of structure and predictability of a random variable, and larger entropy indicates less structure and predictability (Hyvarinen and Oja, 2000). The entropy of a continuous random vector $\boldsymbol{z}_1$ is defined as

$$H(\boldsymbol{z}_1) = -E_{p_{\boldsymbol{z}_1}}(\log(p_{\boldsymbol{z}_1}(\boldsymbol{z}_1))) \tag{1.2}$$

where $p_{\boldsymbol{z}_1}$ is the probability density function (pdf) of $\boldsymbol{z}_1$. The joint entropy of n random vectors, $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$ is defined as

$$H(\boldsymbol{z}_1, ..., \boldsymbol{z}_n) = -E_{p_{\boldsymbol{z}_1,...,\boldsymbol{z}_n}}(\log(p_{\boldsymbol{z}_1,...,\boldsymbol{z}_n}(\boldsymbol{z}_1, ..., \boldsymbol{z}_n))). \tag{1.3}$$

One of the most useful functions of entropy for ICA estimation is the multivariate mutual information, which is defined for a collection of vectors, $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$, as

$$I(\boldsymbol{z}_1, ..., \boldsymbol{z}_n) = E_{p_{\boldsymbol{z}_1,...,\boldsymbol{z}_n}}\left(\log\left(\frac{p_{\boldsymbol{z}_1,...,\boldsymbol{z}_n}(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{p_{\boldsymbol{z}_1}(\boldsymbol{z}_1)\cdots p_{\boldsymbol{z}_n}(\boldsymbol{z}_n)}\right)\right) = \sum_{j=1}^{n} H(\boldsymbol{z}_j) - H(\boldsymbol{z}_1, ..., \boldsymbol{z}_n). \tag{1.4}$$

The mutual information is equivalent to the Kullback-Leibler divergence, a statistical measure of distance, between the joint density and the product of the marginal densities of any collection of random vectors. It follows immediately, then, that mutual information could be an appropriate measure independence of random vectors, given that the product of the marginal densities is equivalent to the joint density when the random vectors are independent. The definition of entropy can easily be extended to accommodate stationary processes, as explained by Comon and Jutten (2010), and we will make no distinction in notation between the entropy and mutual information of a stationary process and those of a random vector.

To the best of our knowledge, the mutual information approach to ICA estimation was first proposed by Comon (1994). In the context of ICA, we want to uncover maximally independent ICs; thus, we would need to minimize the mutual information for the rows of $\boldsymbol{S}$. Because $\boldsymbol{S} = \boldsymbol{W}\boldsymbol{X}$, this can be done through a minimization of the mutual information with respect to $\boldsymbol{W}$. In doing this, we find a value for $\boldsymbol{W}$ such that the distance between the joint density and the product of the marginal densities of the ICs is minimized, resulting in approximately statistically independent ICs (Hyvarinen and Oja, 2000).

We denote the rows of $\boldsymbol{S}$ as $\boldsymbol{s}_1, ..., \boldsymbol{s}_k$ and the rows of $\boldsymbol{W}$ as $\boldsymbol{w}_1, ..., \boldsymbol{w}_k$. Comon and Jutten (2010) show that minimizing the mutual information between $\boldsymbol{s}_1, ..., \boldsymbol{s}_k$ is approximately the same as minimizing the following criteria with respect to $\boldsymbol{W}$:

$$C(\boldsymbol{W}) = \sum_{j=1}^{k} H(\boldsymbol{w}_j \boldsymbol{X}) - \log(\det(\boldsymbol{W})). \tag{1.5}$$

In practice, when using this estimation method it is typically assumed that the observations of the same IC over time are independent and identically distributed (i.i.d.), resulting in $C(\boldsymbol{W}) = \sum_{t=1}^{T}(\sum_{j=1}^{k} H(\boldsymbol{s}_j(t)) - \log(\det(\boldsymbol{W})))$. To minimize equation 1.5, we must also estimate the entropy. Entropy estimation is a complex topic and will not be addressed here. Details on common procedures used for this purpose can be found in Comon and Jutten (2010).

The minimization of the mutual information is a very intuitive means of finding ICs. The result of this minimization is an estimate of $\boldsymbol{W}$ with desirable statistical properties (Comon and Jutten, 2010). An advantage of this method of estimation is that it does not require that the data follow the ICA model in order to produce maximally independent components (Hyvarinen et al., 2001).

As in most classical ICA estimation methods, the assumption that the ICs are nongaussian distributed is needed for identifiability in the minimization of mutual information. More precisely, in order for the ICs to be blindly separable, they must not be gaussian with proportional covariance matrices (Comon and Jutten, 2010), but, because most classical methods assume that the ICs all have proportional covariance matrices, this reduces to the assumption that they are nongaussian. Hyvarinen and Oja (2000) explain that any orthogonal transformation of independent gaussian random variables has the same distribution as the original variables, and, due to this equivalence, the ICA model is only identifiable up to an orthogonal transformation if more than one of the random variables is gaussian. Nongaussianity is an assumption that is made in almost all popular ICA algorithms but is difficult to verify. Furthermore, in practice, methods which minimize mutual information to estimate ICA parameters typically assume that all the observations from the same IC are independent. In many applications, the observations from the same IC come from a process such as a time series and are highly correlated, making this assumption invalid. This assumption is relaxed in methods developed more recently.

1.2.2.2 Maximum Likelihood Estimation

Maximization of the likelihood is a common method of parametric statistical estimation, and, in spite of the fact that little is typically known about likelihoods in the ICA formulation, it can be applied to solve the ICA problem. To the best of our knowledge, maximum likelihood estimation was first applied to ICA by Gaeta and Lacoume (1990) and Pham et al. (1992). For the moment, consider the case where $T = 1$, making both $\boldsymbol{X}$ and $\boldsymbol{S}$ $k \times 1$ vectors. Let $p_i(s_i)$ denote the distribution of the i^{th} IC. Then, due to independence, the vector of ICs has joint distribution $p_{\boldsymbol{S}}(\boldsymbol{S}) = \prod_{i=1}^{k} p_i(s_i)$. By a simple transformation of variables, we can see that $p_{\boldsymbol{X}}(\boldsymbol{X}) = (\prod_{i=1}^{k} p_i(\boldsymbol{w}_i \boldsymbol{X}))\text{det}(\boldsymbol{W})$. Now, extending this to the case of $T > 1$ when the time points are assumed to be independent, $p_{\boldsymbol{X}}(\boldsymbol{X}) = \prod_{t=1}^{T}(\prod_{i=1}^{k} p_i(\boldsymbol{w}_i \boldsymbol{x}(t)))\text{det}(\boldsymbol{W})$, where $\boldsymbol{x}(t)$ is a column of $\boldsymbol{X}$. Therefore, we get the following likelihood for $\boldsymbol{W}$:

$$L(\boldsymbol{W}|\boldsymbol{X}) = \prod_{t=1}^{T} \left(\prod_{i=1}^{k} p_i(\boldsymbol{w}_i \boldsymbol{x}(t)) \right) \text{det}(\boldsymbol{W}) \tag{1.6}$$

which can be maximized (in practice the log likelihood is maximized) to estimate $\boldsymbol{W}$.

It should be noted that taking the expected value of the log likelihood renders the negative of the criteria given in equation 1.5, so that maximum likelihood is approximately equivalent to minimizing the mutual information when p_i is the true distribution of $\boldsymbol{s_i}$ (Hyvarinen and Oja, 2000). Thus, we see that, since the two methods are approximately equivalent, maximum likelihood estimation is also finding the maximally independent components. Moreover, the popular Infomax ICA algorithm (Bell and Sejnowski, 1995) is equivalent to maximum likelihood estimation under the default model specifications (Cardoso, 1997).

Although the formulation of the maximum likelihood method seems simple to this point, one can imagine how estimation using this method is complicated by the fact that the distributions of the ICs are typically unknown. Generally, estimation of densities is a computationally intensive problem because it must be done nonparametrically; however, Hyvarinen et al. (2001) prove that IC densities can be approximated using a simple family of densities, dramatically reducing the complexity of the problem, while retaining the local consistency of the maximum likelihood estimator. To use this approach, one only needs to specify whether the densities of the ICs are sub-gaussian or super-gaussian. Several algorithms are provided in Hyvarinen et al. (2001) that perform this approximation while simultaneously maximizing the likelihood.

The maximum likelihood estimation method for ICA is theoretically simple, as maximum likelihood is perhaps the most popular estimation method in statistics. The estimators also have desirable statistical properties under some mild conditions (Hyvarinen et al., 2001). Additionally, in the case that the distributions of the sources are known a priori, maximum likelihood should be preferred over the mutual information method, since it can take advantage of this added information (Comon and Jutten, 2010).

However, like many other ICA estimation methods, maximum likelihood requires that the ICs be nongaussian and assumes that the observations from the same IC are independent. Moreover, if the distributions of the ICs are unknown, the user must specify whether the ICs are sub- or super-gaussian in order to estimate the densities (Hyvarinen and Oja, 2000). This task is often difficult, and an incorrect choice compromises the good properties of the estimates (Hyvarinen et al., 2001).

1.2.2.3 Estimation by Maximizing Nongaussianity

One of the most popular ICA estimation procedures, FastICA (Hyvarinen and Oja, 2000), is based on the principle of finding rows of $\boldsymbol{W}$ such that the rows of $\boldsymbol{S} = \boldsymbol{W}\boldsymbol{X}$ are maximally nongaussian. We first seek to estimate $\boldsymbol{w}_1$ such that $\boldsymbol{w}_1\boldsymbol{X}$ equals one of the ICs. We note that computing $\boldsymbol{w}_1\boldsymbol{X}$ is equivalent to summing linear combinations of random variables. By the Central Limit Theorem, the sum of any two independent random variables is "more gaussian" than either of the two original variables; thus, if we maximize the nongaussianity of $\boldsymbol{w}_1\boldsymbol{X}$, we should arrive at a single random variable– specifically, one of the ICs (Hyvarinen and Oja, 2000). We repeat this procedure for $\boldsymbol{w}_2, ..., \boldsymbol{w}_k$. To ensure that we are detecting different ICs each time, after the estimation of each new component, we can constrain the search space to look exclusively for estimates that are uncorrelated with all the previous ICs, which is equivalent to orthogonalization under some conditions which can be imposed through preprocessing (Hyvarinen and Oja, 2000).

The primary challenge in this method is in finding a way to estimate gaussianity. Hyvarinen and Oja (2000) discuss several methods of quantifying gaussianity. Because gaussian variables are known to have the largest entropy in any set of random variables with equal variance (Cover and Thomas, 1991; Papoulis, 1991), certain functions of the entropy can also be used to estimate

$\boldsymbol{W}$ when approaching the problem from this perspective. FastICA maximizes nongaussianity by maximizing the negentropy, a function defined as the difference between the entropy of a gaussian random variable and the random variable of interest (Hyvarinen and Oja, 2000). Then, we wish to maximize

$$J(\boldsymbol{s}_j) = H(\boldsymbol{y}_j^{gauss}) - H(\boldsymbol{s}_j) = H(\boldsymbol{y}_j^{gauss}) - H(\boldsymbol{w}_j\boldsymbol{X}) \tag{1.7}$$

with respect to $\boldsymbol{w}_j$ for $j = 1, ..., m$, where $\boldsymbol{y}_j^{gauss}$ is a gaussian random vector with the same covariance matrix as $\boldsymbol{w}_j\boldsymbol{X}$. However, because the negentropy is difficult to estimate, an approximation is used in the fastICA algorithm, using methods described in detail in Hyvarinen and Oja (2000).

As implied by the name, one of the major advantages of the fastICA approach is its computational speed and simplicity and quick convergence (Giannakopoulos et al., 1999). It can also uncover ICs that are both sub-gaussian and super-gaussian, without the need for the user to specify this information, making it more user-friendly than the maximum likelihood approach (Hyvarinen et al., 2001). FastICA is also set apart from maximum likelihood estimation and minimization of mutual information by its ability to estimate the independent components one-by-one (Hyvarinen et al., 2001). However, it shares with the previous two estimation methods the difficult-to-verify assumptions of nongaussianity of the ICs and the independence of observations from the same IC.

FastICA is closely linked to maximum likelihood estimation and minimization of mutual information. The equivalence of maximum likelihood and mutual information was shown in the previous section. Hyvarinen et al. (2001) show further that fastICA is also equivalent to these methods when the estimates of the ICs are constrained to be uncorrelated (this constraint is built into the default procedure in the fastICA algorithm). Thus, with appropriate model specifications, which correspond to the default procedures in the estimation algorithms, these three estimation methods give identical estimates.

1.2.3 EEG Research and ICA

Since being proposed as a tool to separate artifact and brain activity signals in EEG data by Makeig et al. (1996), ICA has become wildly popular in EEG research. EEG data is collected using a helmet containing metal nodes called electrodes which record electrical signals at locations across the scalp. The recordings from the electrodes are mixtures of signals generated by brain activity, artifactual signals, and noise, and the use of ICA is highly recommended for separating brain activity signals from one another and from noise and artifacts prior to doing inference on EEG data (Delorme et al., 2001; Delorme and Makeig, 2004; Delorme et al., 2007; Flexer et al., 2005; Joyce et al., 2004; Jung et al., 1998, 2000; Makeig et al., 1996; Onton et al., 2006; Vigário, 1997).

After applying ICA to EEG data, both the spatial maps and the temporal structure of the ICs are analyzed to distinguish artifactual ICs from brain activity ICs and, where applicable, to determine what type of brain activity is reflected in an IC. Different types of brain activity are associated with electrical signals exhibiting distinct frequencies, also known as rhythms; thus, an analysis of the power of an IC over a range of frequencies provides insight into what type of brain activity it represents. Delta rhythms, with frequency in the range 0-4 Hertz (Hz), are typically observed in humans during sleep or while anesthetized (Schomer and Da Silva, 2012). Theta rhythms have frequency in the range 4-7 Hz and have been associated with navigation and memory tasks (Schomer and Da Silva, 2012). With frequency approximately 8-13 Hz, alpha rhythms arise when the brain is idle and visual attention is diminished, such as during rest with eyes closed (Schomer and Da Silva, 2012). Faster rhythms are associated with wakefulness and information processing (Schomer and Da Silva, 2012). Many of these rhythms are also distinguishable due to their prominence in a limited spatial region on the scalp in EEG recordings (Schomer and Da Silva, 2012).

An emerging question of interest in the EEG community that can be investigated using ICA relates to the functional connectivity of these EEG rhythms. Functional connectivity analyses aim to uncover functional dependencies between brain areas that are physically separated, i.e. identify physically separated brain areas that activate simultaneously either during rest or during the performance of tasks (Friston, 2011). For decades, the brain's functional connectivity has been assessed primarily through fMRI (Buckner et al., 2008), but the low temporal resolution of fMRI may result in high frequency connectivity patterns being missed. Recently, EEG functional connectivity analyses have become popular as a means of providing insight into the connectivity of high frequency brain activity (Chen et al., 2008). The recommended approach to EEG connectivity analyses is to first apply ICA to the data to extract ICs representing the rhythms of interest and then perform source localization and compute connectivity statistics for these ICs of interest (Chen et al., 2013; Delorme et al., 2002; Schoffelen and Gross, 2009).

As this discussion makes clear, the appeal of ICA as a tool to analyze EEG data lies in its potential to separate the data into ICs with recognizable temporal correlation structures representing various types of brain activity. Thus, popular algorithms which assume that the realizations from each IC are independent over time threaten to distort the features of interest in the EEG context. ICA algorithms that allow for temporal correlation within the ICs are needed in order to properly characterize the cyclic nature of the signals generated by brain activity.

1.2.4 ICA with Temporally Correlated Sources

Although they have not gained the popularity of FastICA and Infomax, ICA methods which account for temporal correlation in the ICs have been developed. Pham and Garat (1997) were the first to develop an ICA estimation procedure that models source autocorrelation, to the best of our knowledge. This method was followed by the colored ICA (CICA) method of Lee et al. (2011), which relaxes some of the assumptions of Pham and Garat. CICA is a semi-parametric ICA method that assumes that the ICs are auto-regressive (AR) time series processes. Rather than maximizing IC independence through higher order statistics, CICA estimates parameters by maximizing the Whittle likelihood (Whittle, 1952) of the AR ICs. CICA has been shown to out-perform its competitors which assume independence within ICs when applied to fMRI data (Lee et al., 2011). The authors also demonstrate that the parameter estimates have good statistical properties, namely consistency and asymptotic normality (Lee, 2011).

CICA is an estimation scheme based on exploiting the time series structure of the ICs, in the frequency domain, to estimate the unmixing matrix, $\boldsymbol{W}$, and, thereby, the IC matrix $\boldsymbol{S}$. The procedure views each IC as realizations from an AR process, i.e. for the j^{th} IC, $S_j(t)$,

$$S_j(t) = \mu + \sum_{h=1}^{p} \phi_h S_j(t-h) + \epsilon_j(t), \tag{1.8}$$

where p is the AR order chosen by model selection and $\epsilon_j(t)$ is an error term with unspecified distribution and variance σ_j^2. CICA utilizes the properties of AR processes from the frequency domain approach to perform estimation. The frequency domain approach treats a time series as a sinusoidal function, a system of periodic sines and/or cosines, and desires to estimate frequencies (number of sinusoid cycles/time) and explain the cycles in the system (Shumway and Stoffer, 2011).

The frequency domain analog of the covariance, called the spectral density, can be used to measure the power of a signal at a given frequency. The spectral densities of the ICs are critical in the estimation scheme of Lee et al. (2011). For an IC, $\boldsymbol{S_j}$ and a frequency, r, the spectral density matrix is defined as

$$g_{jj}(r;\boldsymbol{S}) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\boldsymbol{\gamma}(h;\boldsymbol{S_j})\exp(-irh) \tag{1.9}$$

where $\boldsymbol{\gamma}(h;\boldsymbol{S_j}) = \text{cov}(S_j(t), S_j(t+h))$. The condition needed for the existence of $\boldsymbol{g}(r;\boldsymbol{S_j})$ is $\sum_{h=-\infty}^{\infty}|\boldsymbol{\gamma}(h;\boldsymbol{S_j})| < \infty$ (Shumway and Stoffer, 2011).

The periodogram can be thought of as the sample spectral density estimator (Shumway and Stoffer, 2011). For $\boldsymbol{S_j}$, the periodogram can be defined as

$$\boldsymbol{I}(r_t,\boldsymbol{S_j}) = \frac{1}{2\pi T}\boldsymbol{\varphi}(r_t,\boldsymbol{S_j})\boldsymbol{\varphi}*(r_t,\boldsymbol{S_j}) \tag{1.10}$$

where $r_t = 2\pi t/T$ for $t = 0, ..., T-1$ are the Fourier frequencies, $\boldsymbol{\varphi}(r_t,\boldsymbol{S_j})$ is the discrete Fourier transform (DFT) of $S_j(t)$, and $\boldsymbol{\varphi}*$ is the conjugate transpose of $\boldsymbol{\varphi}$ (Lee et al., 2011).

The CICA algorithm performs estimation by iteratively updating the unmixing matrix, through maximization of the Whittle likelihood, then updating the time series parameters corresponding to each IC, through standard time series model selection procedures, until convergence. Denoting the observed frequencies as $r_t = \frac{2\pi t}{T}$ for $t = 0, ..., T-1$, then the Whittle Likelihood is given by

$$\begin{aligned} L(\boldsymbol{G};\boldsymbol{S}) &= -\frac{1}{2}\sum_{j=1}^{k}\sum_{k=0}^{T-1}\left\{\frac{\boldsymbol{I}(r_t,\boldsymbol{S_j})}{g_{jj}(r_t;\boldsymbol{S})} + \log(g_{jj}(r_t;\boldsymbol{S}))\right\} \\ &= -\frac{1}{2}\sum_{j=1}^{k}\sum_{k=0}^{T-1}\left\{\frac{\boldsymbol{e_j^T}\boldsymbol{W}\boldsymbol{I}(r_t,\boldsymbol{X})\boldsymbol{W}^T\boldsymbol{e_j}}{g_{jj}(r_t;\boldsymbol{S})} + \log(g_{jj}(r_t;\boldsymbol{S}))\right\} + T\text{log}(\det(\boldsymbol{W})) \end{aligned} \tag{1.11}$$

where $\boldsymbol{W}$ is the unmixing matrix, $\boldsymbol{I}(r_t,\boldsymbol{X})$ is the periodogram for $\boldsymbol{X}$ and $\boldsymbol{I}(r_t,\boldsymbol{S_j})$ the periodogram for $\boldsymbol{S_j}$, and $\boldsymbol{e_j}$ is a $k \times 1$ matrix with j^{th} entry 1 and all other entries 0. $g_{jj}(r_t;\boldsymbol{S})$ is the spectral density of the j^{th} IC at frequency r_t, which can be written in terms of its AR time series parameters as $g_{jj}(r_t;\boldsymbol{S}) = \frac{\sigma_j^2}{2\pi|\Phi_j(e^{-ir_t})|^2}$, where σ_j^2 is the variance of the error term in the time series process and $\Phi_j(z) = 1 - \phi_{j,1}z^1 - ... - \phi_{j,p_j}z^{p_j}$ is the autoregressive polynomial. The spectral density matrix, which will be denoted $\boldsymbol{G}$, is a $k \times T$ matrix with the spectral density of IC j over all observed frequencies in row j. The estimate of $\boldsymbol{W}$ is constrained to be orthogonal through a penalty term added to the Whittle likelihood.

1.2.5 Pre-ICA Dimension Reduction

In some applications of ICA, including fMRI and EEG, the number of recording locations k, is often believed to be greater than the number of independent signals generating the data, and much of the activity in the observed data is believed to be pure noise. This scenario is referred to as overdetermined ICA (Winter et al., 2003), and it has been shown that applying ICA without regard for this problem can lead to overlearning (Särelä and Vigário, 2003). Moreover, due to the computationally intensive nature of ICA and the high dimensionality of the data in many of the popular ICA application areas, this assumption often introduces an insurmountable computational challenge. The mixing matrix, $\boldsymbol{A}$, is the target of estimation in ICA, and, as it is typically constrained to be orthogonal, it contains $k(k-1)/2$ free parameters (Hyvarinen and Oja, 2000). Thus, the computational burden of ICA increases dramatically as k increases and is less impacted by increasing T. In fMRI data, which have extremely high spatial resolution, k can range from hundreds of thousands to millions, making direct application of ICA computationally unfeasible (McKeown et al., 1998). EEG data, too, have recently begun to be collected at high enough spatial resolution ($k > 100$) to make ICA computationally burdenson.

In order to prevent overlearning and/or reduce the computational burden of ICA, it is often preceded by a dimension reduction step, which is accomplished using principal component analysis (PCA) (Pearson, 1901) or singular value decomposition (SVD). In applying these procedures, we reduce the dimensions of the data from $k \times T$ to $m \times T$, $m < k$. m can be chosen based on prior knowledge about an appropriate number of ICs (Xu et al., 2004), or a reasonable value for m can be inferred through exploratory data analyses (Calhoun et al., 2001b). ICA is then applied to the reduced data, resulting in the estimation of only m ICs.

Though mathematical justification of the combined use of dimension reduction methods and ICA is rarely provided in the literature, this procedure can be rationalized as using a linear transformation (PCA or SVD) to partition the data into a signal subspace and a noise subspace and applying ICA to transform the signal subspace into independent components. To formalize this concept, consider a $k \times T$ matrix of mixed signal data, $\boldsymbol{X}$, that has been row-centered. Then $\boldsymbol{X}$ can be decomposed using SVD as

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V'}, \tag{1.12}$$

where $\boldsymbol{U}$ is a $k \times k$ matrix of left singular vectors, $\boldsymbol{V}$ is a $T \times T$ matrix of right singular vectors, and $\boldsymbol{D}$ is a $k \times T$ diagonal matrix with singular values on the diagonal. If we assume that m components form the signal subspace of the data and the remaining components form a noise subspace, then we can partition each of these matrices into terms corresponding to the signal and noise subspace, denoted by S and N respectively, in the following way:

$$\begin{aligned} \boldsymbol{U} &= \begin{bmatrix} \boldsymbol{U_S} & \boldsymbol{U_N} \end{bmatrix} \\ \boldsymbol{D} &= \begin{bmatrix} \boldsymbol{D_S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D_N} \end{bmatrix} \\ \boldsymbol{V} &= \begin{bmatrix} \boldsymbol{V_S} & \boldsymbol{V_N} \end{bmatrix} \end{aligned} \tag{1.13}$$

where $\boldsymbol{U_S}$ is $k \times m$, $\boldsymbol{U_N}$ is $k \times (k - m)$, $\boldsymbol{D_S}$ is $m \times m$, $\boldsymbol{D_N}$ is $(k - m) \times (T - m)$, $\boldsymbol{V_S}$ is $T \times m$, and $\boldsymbol{V_N}$ is $T \times (T - m)$. Then, a reduced dataset with dimensions $m \times T$, representing only the signal subspace, can be formed by $\boldsymbol{X_R} = \boldsymbol{D_S}\boldsymbol{V'_S}$, and we can apply ICA to $\boldsymbol{X_R}$ to transform the signal subspace into independent components (Petersen et al., 2000). This approach is reasonable if the data are truly generated by a small number of signals that explain most of the variance in the data, but, because noise from the recording devices and artifacts can often explain more of the variation in fMRI and EEG data than the signals of interest, caution must be exercised in the choice of m so that the signals of interest are not removed during dimension reduction.

1.2.6 Uncertainty Estimation in ICA

The estimation of statistical uncertainties has always presented an obstacle for ICA users and researchers, which precludes the testing of statistical hypotheses related to ICA parameter estimates. One impediment to the estimation of uncertainties is the identifiability problem in ICA. If no constraints are placed on the model, $\boldsymbol{A}$ and $\boldsymbol{S}$ are identified only up to a permutation and scale factor, i.e.,

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} = [\boldsymbol{A}(\boldsymbol{P}\boldsymbol{D})]\left[(\boldsymbol{P}\boldsymbol{D})^{-1}\boldsymbol{S}\right] \tag{1.14}$$

for any permutation matrix, $\boldsymbol{P}$, and diagonal matrix, $\boldsymbol{D}$. Uncertainty estimates are not meaningful in a non-identifiable model. However, identifiability constraints can be applied to circumvent these problems. In CICA, the mixing matrix can be constrained to be orthogonal to resolve the scale ambiguity. After the model is fit, performing a permutation procedure, such as the one proposed by (Chen and Bickel, 2005), overcomes the permutation ambiguity. When such constraints are made, uncertainties can be estimated.

Although many ICA estimation techniques, including CICA, use maximum likelihood estimation or simplify under certain assumptions to maximum likelihood estimation (Cardoso, 1997; Hyvarinen and Oja, 2000) for which asymptotic theory has been developed (Comon and Jutten, 2010; Lee, 2011), the computation of the asymptotic standard errors for the mixing matrix proves to be very mathematically challenging. Moreover, the use of distributional and asymptotic theory for estimation is often seen as contrary to the spirit of BSS and ICA, which claim to be "blind" procedures, meaning they impose few or no assumptions on the data. Finally, if pre-ICA dimension reduction or other ICA pre-processing procedures are used, the variance in the data may be distorted so that asymptotic variance estimates are inaccurate. For these reasons, measures of uncertainty are rarely computed or used in practice.

1.3 Factor Analysis

Like ICA, factor analysis is a statistical procedure used to estimate latent variables underlying multivariate data; however, the assumptions, methodologies, and motivations surrounding factor analysis differ considerably from those of ICA. Factor analysis, which analyzes the covariance structure of a large set of observed variables to identify and estimate a small number of latent variables, called latent factors, that drive the values of all the observed data, is primarily employed for data reduction or summarization. It has experienced popularity particularly in research areas that rely on questionnaires to collect data, such as psychology, as questionnaires often ask participants many questions aimed at indirectly measuring the same underlying feature of interest, which can be uncovered by factor analysis.

1.3.1 The Factor Analysis Model

Using notation that facilitates the transition between classic factor analysis and spatial factor analysis (Banerjee et al., 2003), the classic factor analysis model takes the form

$$\boldsymbol{Y}(\boldsymbol{s}_i) = \boldsymbol{\Lambda}\boldsymbol{\eta}(\boldsymbol{s}_i) + \boldsymbol{\epsilon}(\boldsymbol{s}_i), \tag{1.15}$$

where $\boldsymbol{Y}(\boldsymbol{s}_i)$ is the $p \times 1$ vector of continuous, observed variables for the i^{th} unit of study, denoted $\boldsymbol{s}_i$, for $i = 1, ..., N$. Letting m be a prespecified number of latent factors, $\boldsymbol{\Lambda}$ represents the $p \times m$ matrix of factor loadings ($m \ll p$), $\boldsymbol{\eta}(\boldsymbol{s}_i)$ is the $m \times 1$ vector of latent factor scores for the i^{th} unit of study, and $\boldsymbol{\epsilon}(\boldsymbol{s}_i)$ represents the vector of errors for the i^{th} unit of study. It is assumed that the $\boldsymbol{\epsilon}(\boldsymbol{s}_i)$ vectors have independent and identically distributed multivariate normal distributions such that $\boldsymbol{\epsilon}(\boldsymbol{s}_i) \overset{\text{iid}}{\sim} \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a diagonal matrix with $(i,i)^{th}$ entry equal to σ_i^2, and $\boldsymbol{\Lambda}$ is constrained to be lower triangular with diagonal entries $\lambda_{ii} > 0$ for identifiability purposes (Bollen, 1989). In the standard model, the $m \times 1$ vectors of factor scores, $\boldsymbol{\eta}(\boldsymbol{s}_i)$, are assumed to be random vectors which are independent across units of study.

1.3.2 Bayesian Estimation in Factor Analysis

Here we choose to take a Bayesian approach to factor analysis parameter estimation. The observed data are assumed to be realizations of random variables following a specified statistical distribution, which is used to create a likelihood for the data. All model parameters are given prior distributions, and samples are drawn from the corresponding posterior (or full conditional) distributions using Markov Chain Monte Carlo (MCMC) sampling. These samples are then summarized to produce parameter estimates and credible intervals and to draw inference.

For our purposes, only continuous data will be considered, in which case it is standard use a multivariate normal likelihood. Following (Nethery et al., 2015), the vectors of data for each unit of study, conditional on the introduced model parameters, are independently distributed as

$$\boldsymbol{Y}\left(\boldsymbol{s}_i\right)|\boldsymbol{\Lambda}, \boldsymbol{\eta}\left(\boldsymbol{s}_i\right), \boldsymbol{\Sigma} \overset{\text{ind}}{\sim} \text{MVN}\left\{\boldsymbol{\Lambda}\boldsymbol{\eta}\left(\boldsymbol{s}_i\right), \boldsymbol{\Sigma}\right\}; i = 1, ..., n \tag{1.16}$$

which can also be written jointly as

$$\boldsymbol{Y}|\boldsymbol{\Lambda}, \boldsymbol{\eta}, \boldsymbol{\Sigma} \sim \text{MVN}(\boldsymbol{\Lambda}^{*}\boldsymbol{\eta}, \boldsymbol{\Sigma}^{*}) \tag{1.17}$$

where $\boldsymbol{Y} = \left\{\boldsymbol{Y}(\boldsymbol{s}_1)^T, \ldots, \boldsymbol{Y}(\boldsymbol{s}_N)^T\right\}^T$, $\boldsymbol{\Lambda}^{*}$ is an $np \times nm$ block diagonal matrix with $\boldsymbol{\Lambda}$ on the diagonal, $\boldsymbol{\Sigma}^{*}$ is an $np \times np$ block diagonal matrix with $\boldsymbol{\Sigma}$ on the diagonal, and $\boldsymbol{\eta} = \left\{\boldsymbol{\eta}(\boldsymbol{s}_1)^T, \ldots, \boldsymbol{\eta}(\boldsymbol{s}_N)^T\right\}^T$.

The standard Bayesian factor analysis prior distribution specifications (Ghosh and Dunson, 2009; Nethery et al., 2015; Rowe, 1998), which lead to semi-conjugacy, are as follows. The diagonal elements of the factor loadings matrix, $\boldsymbol{\Lambda}$, are given independent truncated normal prior distributions (truncated below by 0) with a common variance such that $\lambda_{jj} \overset{\text{iid}}{\sim} \text{TN}(0, \tau_1^2; \geq 0)$, $j = 1, \ldots, p$. The off-diagonal entries (below the diagonal) assume independent normal prior distributions with a common variance such that $\lambda_{jk} \overset{\text{iid}}{\sim} \text{N}(0, \tau_1^2)$, $j > k$. The variance parameters take independent and identically distributed inverse gamma prior distributions, such that $\sigma_j^2 \overset{\text{iid}}{\sim} \text{IG}(\alpha, \beta)$, $j = 1, \ldots, p$. Finally, the factor score vectors are assigned independent and identically distributed multivariate normal prior distributions, such that $\boldsymbol{\eta}(\boldsymbol{s}_i) \overset{\text{iid}}{\sim} \text{MVN}(\boldsymbol{0}, \boldsymbol{I}_m)$ where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix. In a standard factor analysis model, the factor scores are assumed to be independent both within and among locations.

Based on these prior distributions, full conditional distributions can be computed, and samples can be drawn from these distributions using a Gibbs sampler (Gelfand and Smith, 1990; Geman and Geman, 1984). The steps in the Gibbs sampler are as follows:

(1) Sample $\lambda_{jj}|\boldsymbol{\Sigma}, \boldsymbol{\eta}, \boldsymbol{Y}, \boldsymbol{\Lambda}\left(-j, -j\right)$ from TN $\left(\frac{\tau_1^2 \sum_{h=1}^{n} \gamma_{hj}\eta_j(\boldsymbol{s}_h)}{\tau_1^2 \sum_{h=1}^{n} \eta_j(\boldsymbol{s}_h)^2 + \sigma_j^2}, \frac{\sigma_j^2 \tau_1^2}{\tau_1^2 \sum_{h=1}^{n} \eta_j(\boldsymbol{s}_h)^2 + \sigma_j^2}; \geq 0\right)$ for $j = 1, \ldots, p$ where $\boldsymbol{\Lambda}\left(-j, -j\right)$ is the $\boldsymbol{\Lambda}$ matrix with the (j, j) element removed, $\gamma_{hj} = Y_j\left(\boldsymbol{s}_h\right) - \boldsymbol{\Lambda}(j, -j)^T \boldsymbol{\eta}_{-j}\left(\boldsymbol{s}_h\right)$, $\boldsymbol{\Lambda}(j, -j)$ is the j^{th} row of $\boldsymbol{\Lambda}$ with the j^{th} component removed, and $\boldsymbol{\eta}_{-j}\left(\boldsymbol{s}_h\right)$ is the set of factor scores for location $\boldsymbol{s}_h$ with the j^{th} component removed.

(2) Sample $\lambda_{jk}|\boldsymbol{\Sigma}, \boldsymbol{\eta}, \boldsymbol{Y}, \boldsymbol{\Lambda}\left(-j, -k\right)$ from N $\left(\frac{\tau_1^2 \sum_{h=1}^{n} \gamma_{hjk}\eta_k(\boldsymbol{s}_h)}{\tau_1^2 \sum_{h=1}^{n} \eta_k(\boldsymbol{s}_h)^2 + \sigma_j^2}, \frac{\sigma_j^2 \tau_1^2}{\tau_1^2 \sum_{h=1}^{n} \eta_k(\boldsymbol{s}_h)^2 + \sigma_j^2}\right)$ for $j > k$, $k = 1, \ldots, p-1$ where $\boldsymbol{\Lambda}\left(-j, -k\right)$ is the $\boldsymbol{\Lambda}$ matrix with the (j, k) element removed, $\gamma_{hjk} = Y_j\left(\boldsymbol{s}_h\right) - \boldsymbol{\Lambda}(j, -k)^T \boldsymbol{\eta}_{-k}\left(\boldsymbol{s}_h\right)$, $\boldsymbol{\Lambda}(j, -k)$ is the j^{th} row of $\boldsymbol{\Lambda}$ with the k^{th} component removed, and $\boldsymbol{\eta}_{-k}\left(\boldsymbol{s}_h\right)$ is the set of factor scores from location $\boldsymbol{s}_h$ with the k^{th} component removed.

(3) Sample $\sigma_j^2 | \boldsymbol{\Lambda}, \boldsymbol{\eta}, \boldsymbol{Y}, \boldsymbol{\Sigma}(-j,-j)$ from IG $\left(\frac{n}{2} + \alpha, \left(\frac{1}{2}\right) \sum_{h=1}^{n} \left\{ Y_j(\boldsymbol{s}_h) - \boldsymbol{\Lambda}_j^T \boldsymbol{\eta}(\boldsymbol{s}_h) \right\}^2 + \beta \right)$ for $j = 1, \ldots, p$ where $\boldsymbol{\Sigma}(-j,-j)$ is the $\boldsymbol{\Sigma}$ matrix with the (j,j) element removed and $\boldsymbol{\Lambda}_j$ is the j^{th} row of $\boldsymbol{\Lambda}$.

(4) Sample $\boldsymbol{\eta}(\boldsymbol{s}_i) | \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{Y}, \boldsymbol{\eta}(-\boldsymbol{s}_i)$ from MVN $\left(\left\{ \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \boldsymbol{I} \right\}^{-1} \left\{ \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Y}(\boldsymbol{s}_i) \right\}, \left\{ \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \boldsymbol{I} \right\}^{-1} \right)$ where $\boldsymbol{\eta}(-\boldsymbol{s}_i)$ is the complete vector of factor scores with those from location $\boldsymbol{s}_i$ removed.

A large number of samples may be collected in this way, and convergence is gauged by the user, generally through graphical representations of the samples, such as traceplots.

1.3.3 Social Vulnerability and Factor Analysis

After decades of use in psychology, where it is often applied to questionnaire and test data to quantify unmeasurable concepts such as intelligence, factor analysis has only more recently experienced popularity as a tool for quantifying latent variables in the context of epidemiology and public health. In particular, as climate change threatens to increase the frequency and severity of natural disasters, the public health community has become interested in measuring the “social vulnerability” of communities to natural disasters. Because the extent to which a community is able to prepare for and recover from disasters is largely determined by social factors, socially vulnerable areas may be more severely impacted; thus, identification of these areas is critical to disaster preparation (Cutter et al., 2003).

An index of community social vulnerability, which assigns relative vulnerability scores to each community across a region of interest, can be used to identify the most highly vulnerable areas. Community social vulnerability is not directly measurable, but an abundance of social indicator variables are available, many of which are highly correlated thanks to their common association with this broader concept of social vulnerability. Thus, an index of social vulnerability can be constructed as a latent factor (or set of latent factors) underlying a relevant set of observed social indicator variables, through the use of factor analysis (Cutter et al., 2003; Cutter and Finch, 2008). The standard factor analysis model, however, is potentially inappropriate because it fails to properly account for the spatial correlation that is likely present in the social indicator variables collected at the community level.

1.3.4 Factor Analysis with Spatially Correlated Factors

Spatial correlation occurs in data when the similarity between measures of interest collected from the units of study is dependent on the geographic distance and/or direction between the units of study. Spatial correlation often arises when the units of study are geographic regions, and interest lies in counts or averages of some measure within each region. Data collected in this fashion is referred to as spatially referenced data. Here we focus primarily on one type of spatially referenced data, areal data, which is data assembled at the level of blocks or regions formed by the partitioning of a space.

Spatial factor analysis (Wang and Wall, 2003) deviates from the standard factor analysis model above by its assumption that the latent factors, denoted above as $\boldsymbol{\eta}(\boldsymbol{s}_i)$, are Gaussian processes containing spatial correlation, i.e. the latent factors for a given unit of study are no longer independent of the latent factors for all other units of study, but are correlated based on some measure of the geographic distance between them. This correlation in the latent factors then induces spatial correlation in the observed data. Spatial factor analysis methodology in the Bayesian setting has been developed to accommodate a variety of data types and analysis goals and has been applied to a wide range of problems (Hogan and Tchernis, 2004; Liu et al., 2005; Lopes et al., 2008; Nethery et al., 2015; Mezzetti, 2012; Stakhovych et al., 2012; Wang and Wall, 2003). (Stakhovych et al., 2012) provides a detailed summary of these developments and applications.

The Bayesian model specification for the spatial factor analysis model is identical to that of the standard factor analysis model with the exception of the assumptions and prior distribution for the latent factors. The spatial model relies on the vectorized form of the latent factors, $\boldsymbol{\eta}$, as in (7). The prior distribution is now placed on $\boldsymbol{\eta}$ so that spatial correlation between the $\boldsymbol{\eta}(\boldsymbol{s}_i)$ may be introduced. The prior for $\boldsymbol{\eta}$ takes a multivariate normal distribution with a kronecker product form covariance matrix to account for the possibility of multiple latent factors represented in each $\boldsymbol{\eta}(\boldsymbol{s}_i)$. Let $\boldsymbol{\Sigma}_S$ be the $N \times N$ covariance matrix which controls the spatial correlation within the latent factors based on some measure of the known distance between units and an unknown spatial parameter, ϕ (the structure could of course be extended to include multiple spatial parameters). Then the prior has the form $\boldsymbol{\eta} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_m)$.

The form of the spatial covariance matrix, $\boldsymbol{\Sigma}_S$, may be determined by the user and should be guided by the type of spatially referenced data being analyzed, the goals of the analysis, and the available information. (Banerjee et al., 2003) provide a thorough consideration of the topic of valid spatial covariance functions, and Wang and Wall (2003) explain how some of these functions may be applied in the spatial factor analysis context. For data that are collected by counting or averaging some measure over pre-defined geographic regions, known as areal data, conditional auto-regressive model covariance functions (Besag, 1974) are commonly chosen (Hogan and Tchernis, 2004; Wang and Wall, 2003). They take the form (or a variant of the form) $\boldsymbol{\Sigma}_S = \boldsymbol{I} - \phi \boldsymbol{R}$, where $\boldsymbol{R}$ is an adjacency matrix, taking value 1 in the $(i,j)^{th}$ position if $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ share a boundary and value 0 otherwise. Analyses that rely on point-referenced spatial data, or data that are collected at geocoded points in space, might, instead, choose a distance-based spatial covariance matrix, so that spatial correlation increases as the distance between two units of study decreases. One example of a distance-based covariance matrix is the exponential covariance matrix, in which the $(i,j)^{th}$ entry of the matrix has the form $\boldsymbol{\Sigma}_S(i,j) = \exp\{-\phi ||\boldsymbol{s}_i - \boldsymbol{s}_j||\}$, where $||\boldsymbol{s}_i - \boldsymbol{s}_j||$ is the Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$.

In order to let the data inform the level of spatial correlation in the latent factors, a prior distribution is assigned to ϕ, and a step is added to the MCMC sampling algorithm to draw samples from its posterior distribution. The standard choice of prior is $\phi \sim$ Unif(a,b), where a and b are lower and upper bounds, respectively, whose values are determined by a combination of the data and the type of covariance structure used. Wang and Wall (2003) provide more detail about the computation of these bounds.

Steps (1)-(3) of the sampling algorithm for the spatial factor analysis model are identical to those of the standard factor analysis model, but step (4) must be revised to accommodate the updated prior on $\boldsymbol{\eta}$ (and corresponding full conditional distribution), and a 5^{th} step must be added to obtain samples of ϕ. Because the full conditional distribution for ϕ does not have a closed form, Nethery et al. (2015) recommend performing a transformation of ϕ and drawing samples of the transformed parameter using a Metropolis step (Hastings, 1970; Metropolis et al., 1953). This results in a sampler with steps (4) and (5) as follows:

(4) Sample $\boldsymbol{\eta}$ from

$$\text{MVN}\left(\left\{\boldsymbol{\Lambda}*^T\boldsymbol{\Sigma}*^{-1}\boldsymbol{\Lambda}*+(\boldsymbol{\Sigma}_S\otimes\boldsymbol{I}_m)^{-1}\right\}^{-1}\left\{\boldsymbol{\Lambda}*^T\boldsymbol{\Sigma}*^{-1}\boldsymbol{Y}\right\},\left\{\boldsymbol{\Lambda}*^T\boldsymbol{\Sigma}*^{-1}\boldsymbol{\Lambda}*+(\boldsymbol{\Sigma}_S\otimes\boldsymbol{I}_m)^{-1}\right\}^{-1}\right).$$

(5) Sample $\psi = \log\left(\frac{\phi-a}{b-\phi}\right) \in \mathbb{R}$ using a Metropolis sampler with a Normal proposal distribution. ϕ is obtained by transformation such that $\phi = \frac{\exp\{\psi\}b+a}{1+\exp\{\psi\}}$.

1.3.5 Spatial Misalignment in Factor Analysis

Though the Bayesian spatial factor analysis model allows for a good deal of flexibility and a wide range of spatial correlation structure specifications, it has not yet been extended to accommodate a common complication encountered in the analysis of areal data– spatial misalignment. Spatial misalignment occurs when spatially referenced variables intended for use in the same analysis originate from differing spatial levels (Banerjee et al., 2003). Spatial misalignment is common in areal data, because the geographic regions across which data are recorded may be incompatible for different measures of interest. Because many spatially referenced variables often need to be analyzed together using spatial factor analysis, misalignment of some variables is likely to present an obstacle to in this context.

Although we know of no instances of spatial misalignment addressed specifically in the context of spatial factor analysis, the topic of spatial misalignment in general has received a great deal of attention, as described by Gotway and Young (2002), and some of the general solutions may be able to be applied to the problem of areal misalignment in spatial factor analysis. A common approach to handling areal misalignment is to align all the variables to a common set of areal units prior to application of a statistical model (Banerjee et al., 2003). This can be done by choosing a single set of areal units and, for each variable not recorded at those units, assigning values to each of the chosen units in proportion to a value of that variable recorded in overlapping units. For example, consider the case where variables are collected at areal units of two different sizes, which we refer to as small areal units (SAUs) and large areal units (LAUs), and the SAUs are fully nested within the LAUs, i.e. each SAU is fully contained in a single LAU. In order to align the variables recorded at the LAUs to the SAUs, the value at each LAU could be directly assigned to each of its nested SAUs (for a rate or average variable) or an appropriate proportion of the value at each LAU could be assigned to each nested SAU, based on population or land area in the SAU (for a count variable). This method, however, imposes strong assumptions about the distributions of the aligned variables, which can distort patterns in the data that are critical to the performance of the factor analysis model.

Another general method of aligning the data prior to statistical analyses, proposed by Mugglin and Carlin (1998), is to construct a model to predict the values of each variable of interest over the desired set of units, using as predictors variables that are recorded at the desired units and are correlated with the variable of interest. However, this method may be inappropriate in the factor analysis context, because it may well be the case that any variable that is collected at the desired units and would be a reasonable predictor is also being included in the factor analysis. Given that factor analysis is fundamentally studying the relationships between observed variables in order to identify the proper latent variables, artificially constructing such a relationship between variables that will go on to be included in the factor analysis together results in a circular procedure.

CHAPTER 2: BOOTSTRAPPING MEASURES OF UNCERTAINTY FOR EEG RESTING STATE CONNECTIVITY STUDIES USING INDEPENDENT COMPONENT ANALYSIS

2.1 Introduction

Since the early 2000s, following a series of publications providing a theoretical justification for the study of the brains at-rest network, known as its default network (Gusnard et al., 2001; Gusnard and Raichle, 2001; Raichle et al., 2001), the study of resting state brain connectivity has exploded (Buckner et al., 2008). The resulting body of literature has demonstrated that not only is the characterization of resting state networks integral to the understanding of how tasks impact the brains functioning, but also that alterations in resting state networks are associated with a number of diseases, suggesting that resting state research will bring us closer to understanding some of the most perplexing psychological and neurological conditions (Buckner et al., 2008). For instance, autism (Assaf et al., 2010; Kennedy et al., 2006), attention deficit hyperactivity disorder (Tian et al., 2006), schizophrenia (Bluhm et al., 2007; Garrity et al., 2007), dementia (Greicius et al., 2004), and a number of other disorders, as described by Broyd et al. (2009), have been associated with default network abnormalities.

Resting state connectivity research has historically been dominated by functional magnetic resonance imaging (fMRI) studies (Broyd et al., 2009), a natural choice for identifying functionally connected brain region thanks to the fMRI's high spatial resolution. However, fMRIs suffer from low temporal resolution, and, as a consequence, high frequency resting state connectivity is likely to be missed by such studies. Electroencephalogram (EEG) recordings, which use metal electrodes to record scalp electrical activity at lower spatial resolution but very high temporal resolution, have more recently been recognized as a means to obtain insight into high frequency changes in resting state network activity (Britz et al., 2010; Laufs, 2010; Musso et al., 2010; Yuan et al., 2012). The default network for the brain's electrical activity, as characterized by EEG, was first proposed by Chen et al. (2008).

During rest, EEG scans record electrical signals produced by a variety of different types of brain activity, and these different activity types are distinguishable by the unique frequencies prominent in the resulting signals (Chen et al., 2008; Lusted and Knapp, 1996). Delta (0.5-3.5Hz), theta (4-7Hz), alpha (7.5-12Hz), beta (13-34Hz), and gamma (35-45Hz) activity have all been found to be present during rest (Chen et al., 2008). The goal of resting state EEG connectivity analyses is often to characterize the default networks for these different types of activity (Chen et al., 2008, 2013; Congedo et al., 2010).

EEG connectivity analyses commonly take one of two different approaches. The first, and simpler, of the two is to compute connectivity measures directly from the scalp recordings (Chen et al., 2008). We refer to this method as the "direct approach". The direct approach, while popular, is commonly criticized due to its neglect of the field spread issue and the EEG inverse problem, as described by Schoffelen and Gross (2009) and Delorme et al. (2002) respectively. Field spread refers to the inevitable EEG phenomenon in which the electrical signal from a single brain activity source will be recorded at multiple electrodes sources (Schoffelen and Gross, 2009), and, similarly, the EEG inverse problem arises because each electrode records a linear mixture of signals from a variety of different activity sources (Delorme et al., 2002). Due to these problems, performing connectivity analyses directly on the scalp recordings can lead to distorted results (Delorme et al., 2002; Schoffelen and Gross, 2009).

The second approach to EEG connectivity analyses is to first perform an unmixing procedure on the scalp recordings to recover the source signals and their corresponding scalp maps and compute connectivity measures using these source signals (Chen et al., 2013; Delorme et al., 2002; Schoffelen and Gross, 2009). Often, blind source separation procedures such as independent component analysis (ICA) are employed to perform the unmixing of the scalp signals; thus, we call this approach to connectivity analyses the "ICA approach". ICA (Bell and Sejnowski, 1995; Hyvarinen and Oja, 2000) is a multivariate statistical method that can be used to unmix recorded mixtures of signals to recover a set of independent source signals, called independent components (ICs).

Letting k denote the number of electrodes/scalp recording locations, and T denote the number of time points at which recordings are made, the ICA model has the form

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} \tag{2.18}$$

where $\boldsymbol{X}$ is the $k \times T$ matrix of EEG recordings with each row containing the recording from a given electrode over time, $\boldsymbol{S}$ is the $k \times T$ matrix of ICs with each row containing the values of a given IC over time, and $\boldsymbol{A}$ is a $k \times k$ matrix of linear mixing parameters which represent the contribution of each IC to the recording at each electrode. $\boldsymbol{A}$ is often assumed to be an orthogonal matrix, and, if we don't believe this to be true for the data, ICA pre-processing techniques are performed to enforce this assumption. ICA estimation is done by first estimating the mixing parameters in $\boldsymbol{A}$ in a manner that imposes as few assumptions as possible about the distributions of the ICs, and then obtaining the IC values by plugging in $\hat{\boldsymbol{S}} = \hat{\boldsymbol{A}}^{\boldsymbol{-1}}\boldsymbol{X}$.

Following the application of ICA, one challenge to characterizing resting state connectivity is identifying the type(s) of resting state brain activity reflected in each IC (Congedo et al., 2010). Although the power spectrum of the ICs can be assessed to determine what frequencies are most powerful in a signal, providing some insight into the type(s) of brain activity that generated it, the spectrum of a single IC may exhibit power peaks in multiple frequency ranges of interest, as demonstrated by Congedo et al. (2010). Moreover, some of these peaks may be small, making it difficult to determine which type(s) of activity are reflected in the IC. This type of imperfect separation of brain activity signals is a result of multiple activity types demonstrating similar spatial activity across the scalp and, therefore, being grouped together into a single IC.

In connectivity analyses, decisions about the type(s) of activity represented in an IC are typically made simply by eyeballing power spectrum plots (Chen et al., 2013; Congedo et al., 2010), a strategy that could easily lead to misplaced inference, because it fails to account for the uncertainties associated with the IC estimates. Statistical hypothesis testing, which would provide a natural solution to this problem, is obstructed by the difficulty in computing uncertainties for ICA parameters. Although asymptotic theory is available for some ICA estimation procedures (Comon and Jutten, 2010), it is very complex and requires strong assumptions that are often undesirable given that one of the selling points of ICA is the minimal assumptions it imposes.

In this paper, we propose a semi-parametric bootstrapping algorithm which invokes ICA estimates in order to create bootstrap samples of either single subject or group EEG scalp data. By bootstrapping from the independent auto-regressive (AR) time series residuals of each estimated IC and reconstructing the data using the ICA parameter estimates, we are able to preserve the cross-correlation between EEG channels and the auto-correlation within EEG scalp channels, critical features for assessing connectivity, in these bootstrap samples. Bose (1988) showed that bootstrapping from the residuals in AR models approximates the distributions of the parameter estimates with $o(\sqrt{T})$ accuracy, which improves on the error of the normal approximation, under some mild assumptions. A review of popular methods for bootstrapping time series is provided by Li and Maddala (1996). Bootstrapping from the ICs in a manner that preserves temporal structure was proposed by Meinecke et al. (2002) to gauge the separation performance of the ICA algorithm. We introduce a new method for bootstrapping from the ICs, which preserves their temporal structure, and we demonstrate that the bootstrap samples of EEG data produced by re-mixing the bootstrapped ICs can be used to form confidence intervals and perform hypothesis tests on connectivity-related parameters.

If one is using the direct approach to connectivity analyses, standard errors (SEs) and confidence intervals (CIs) for scalp channel spectral coherences and other connectivity measures can be formed using the bootstrap samples. In the ICA approach to connectivity analyses (recommended), these bootstrap samples allow for the computation of SEs and CIs for the IC power spectra and related quantities. Using these CIs, we propose a novel hypothesis testing framework that can be used to detect the presence of various types of brain activity in each IC. In particular, confidence limits on a variation of the IC power spectra are used to test for significant peaks in the frequency ranges of brain activity of interest. Such a test allows for a more statistically rigorous approach to analyzing the default networks of each brain activity type. Though Congedo et al. (2010) proposed the use of non-parametric IC power spectra CIs to identify peaks representing abnormal brain functioning in the context of resting state EEG, their method can only be used to test a patient against a normative database. Our method allows any type of brain activity to be detected in any of the ICs within a single subject or group of subjects.

In Section 2.2, we introduce the bootstrapping algorithm and the corresponding approach to hypothesis testing for brain activity within an IC. In Section 2.3, simulation studies are used to demonstrate the effectiveness and utility of our bootstrap approach. The bootstrap algorithm and hypothesis test are applied to single subject resting state EEG data in Section 2.4, and using this analysis, we illustrate why it is critical to consider the variability in the IC-related estimates when drawing conclusions about the type(s) of brain activity an IC contains. Finally, the results and the impact of this method are discussed in Section 2.5.

2.2 Methods

While the most popular ICA algorithms assume that the ICs contain no auto-correlation (Bell and Sejnowski, 1995; Hyvarinen and Oja, 2000), Lee et al. (2011) developed a semi-parametric ICA algorithm called colorICA (CICA) that models the ICs as AR time series processes, i.e. for the j^{th} IC, $S_j(t)$,

$$S_j(t) = \mu + \sum_{h=1}^{p} \phi_h S_j(t-h) + \epsilon_j(t), \tag{2.19}$$

where p is the AR order chosen by model selection and $\epsilon_j(t)$ is an error term with unspecified distribution and variance σ_j^2. A run of CICA produces estimates of the mixing matrix and ICs, $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{S}}$, a collection of estimated AR coefficients for each IC, $\hat{\boldsymbol{\phi}}$, a collection of estimated time series variances for each IC, $\hat{\boldsymbol{\sigma^2}}$, and a matrix of smoothed power spectra estimates for each IC, $\hat{\boldsymbol{G}}$. Allowing for auto-correlation within ICs is critical in the analysis of resting state EEG data, because the electrical signals emitted by resting state brain activity are known to be cyclic processes. Hence, we focus the development of our bootstrapping algorithm around CICA.

The recorded EEG scalp channels contain both cross-correlation and auto-correlation, and each of these features are critical to properly characterizing connectivity. Thus, any useful bootstrapping algorithm must preserve both the channel cross-correlation and auto-correlation in the bootstrapped datasets. In order to do so, our algorithm must take into consideration both the mixing of signals, which induces the cross-correlation between the channels, and the temporal correlation in these signals, which induces the auto-correlation within the channels.

We propose the following semi-parametric ICA-based procedure for creating a bootstrapped resting state EEG dataset, which is analogous to the semi-parametric procedure recommended for bootstrapping in a linear model framework. First, CICA should be applied to the matrix of resting state EEG data, $\boldsymbol{X}$.Because the ICs are independent, we can construct a bootstrap sample of each of one and mix them to obtain a bootstrap sample of the original data that preserves the channel cross-correlations. To create a bootstrap sample of each IC that retains its time series structure, its estimated AR model residuals should be resampled with replacement and plugged into the estimated AR model, as described by Efron and Tibshirani (1986), initializing the bootstrapped time series using the block initialization method of Stine (1987). Finally, to construct the bootstrap sample of the data, $\tilde{\boldsymbol{X}}$, the bootstrapped ICs should be concatenated into a matrix $\tilde{\boldsymbol{S}}$ and multiplied by $\hat{\boldsymbol{A}}$, i.e. $\tilde{\boldsymbol{X}} = \hat{\boldsymbol{A}}\tilde{\boldsymbol{S}}$. A large number, B, of bootstrapped datasets can be constructed by repeating this process B times. A summary of this method can be found in Table 2.1.

After constructing B bootstrap samples in this manner, direct connectivity statistics, such as the squared coherence between channels, may be estimated for each bootstrap sample. As explained in Efron and Tibshirani (1986), the standard deviation of these bootstrap estimates can be used as SEs for each connectivity statistic, and CIs can be formed by applying the percentile method.

To compute SEs and CIs for the ICA parameters, which are needed for the ICA approach to connectivity, more involved computations using the bootstrap samples are needed. In particular, CICA must be performed on each bootstrapped dataset, to obtain B bootstrap estimates for each CICA parameter. Thus, using * notation to denote the bootstrap parameter estimates, we have $\boldsymbol{A}^*_{\boldsymbol{1}}, ..., \boldsymbol{A}^*_{\boldsymbol{B}}$, $\boldsymbol{S}^*_{\boldsymbol{1}}, ..., \boldsymbol{S}^*_{\boldsymbol{B}}$, $\boldsymbol{\phi}^*_{\boldsymbol{1}}, ..., \boldsymbol{\phi}^*_{\boldsymbol{B}}$, $\boldsymbol{\sigma^2}^*_{\boldsymbol{1}}, ..., \boldsymbol{\sigma^2}^*_{\boldsymbol{B}}$, and $\boldsymbol{G}^*_{\boldsymbol{1}}, ..., \boldsymbol{G}^*_{\boldsymbol{B}}$. One final complication obstructs the computation of SEs and CIs for the ICA parameters from these bootstrap estimates– namely, the IC permutation ambiguity in ICA.

In ICA, the ICs are not estimated in any consistent order (unlike principal component analysis, in which the components are estimated in order of the amount of the variability in the observed data they explain). Thus, due to the jittering of the observed data through bootstrapping, the ordering of the ICs may be different in the bootstrap estimates from each bootstrapped dataset. Then, the ICs estimated from all the bootstrapped datasets (and their corresponding parameters) must be aligned or matched prior to computing SEs or CIs to ensure that bootstrap parameters estimates for corresponding ICs are being summarized. To achieve a common permutation of the ICs in all the bootstrap estimates, the bootstrap estimated ICs (and their corresponding parameters) should be placed in the same order as the original estimated ICs.

Although the cross-correlation between the original estimated ICs ($\hat{\boldsymbol{S}}$) and the bootstrap estimated ICs ($\boldsymbol{S}^*_{\boldsymbol{1}}, ..., \boldsymbol{S}^*_{\boldsymbol{B}}$) might seem like a natural measure to use to perform this permutation, the cross-correlation between these time series may not, in fact, be a relevant measure of their similarity. This is a result of the fact that there may be little to no cross-correlation between a time series and a bootstrap sample of it, due to differences in the starting values of the two series. Instead, the permutation of the bootstrap estimates should be performed based on the magnitude of the correlation between the original estimated power spectra of the ICs ($\hat{\boldsymbol{G}}$) and the bootstrap estimated power spectra of the ICs ($\boldsymbol{G_1^*}, ..., \boldsymbol{G_B^*}$), as the power spectra of a time series is unaffected by its starting value.

Then, the permutation of the bootstrap estimates should proceed as follows. For each set of bootstrap estimates, choose the row of $\boldsymbol{G}^*$ with the highest magnitude of correlation with the first row of $\hat{\boldsymbol{G}}$, say row i, and make row i of $\boldsymbol{G}^*$ the first row of the new permuted power spectra matrix, $\boldsymbol{G}^*_{perm}$. Repeat this matching process for the second row of $\hat{\boldsymbol{G}}$, removing row i of $\boldsymbol{G}^*$, which was chosen in the first iteration, from consideration, and now placing the most highly correlated row of $\boldsymbol{G}^*$ into the second row of $\boldsymbol{G}^*_{perm}$. Continue this process for each of the k rows of $\hat{\boldsymbol{G}}$, removing a row of $\boldsymbol{G}^*$ from consideration in all future repetitions after it has been chosen, so that each row of $\boldsymbol{G}^*$ appears as exactly one row of $\boldsymbol{G}^*_{perm}$. Upon completion of this process, the estimated power spectrum for a given IC should be in the same row in $\hat{\boldsymbol{G}}$ and $\boldsymbol{G}^*_{perm}$. Of course, all other bootstrap parameter estimates in the set must be permuted accordingly.

After each set of bootstrap estimates has been permuted in this way, SEs may be computed for the IC AR parameters by taking the standard deviation of the bootstrap estimates, and CIs may be formed using the percentile method (Efron and Tibshirani, 1986). These uncertainty measures can be computed pointwise for the IC power spectra in $\boldsymbol{G}$. This procedure is summarized in Table 2.2. Finally, the SEs and CIs can be used to test statistical hypotheses about the ICs.

In particular, to form a test for the presence of a certain type of brain activity in an IC, we adapt a method commonly used in the time series literature to test whether a peak in a power spectrum is significant. In this method, a lower 95% confidence limit is computed for the power spectrum in the frequency range around the peak, and if that confidence limit exceeds a chosen "baseline" value for the power spectrum, the null hypothesis of no significant peak is rejected (Shumway and Stoffer, 2011). This method, to our knowledge, has not previously been applied in the EEG setting.

In resting state EEG data, we want to know whether the power spectrum for a given IC contains a significant peak in the frequency range of a certain type of brain activity. Thus, we will test whether the IC's power spectrum significantly exceeds its AR noise level (our chosen baseline value) anywhere in the frequency range of that activity type. This is equivalent to testing whether the difference in the power spectrum and the AR model noise is significantly greater than zero anywhere in that frequency range. Thus, for IC j with AR variance σ_j^2 and spectrum value $g_j(r_h)$ at frequency

r_h, we test the hypothesis

$$H_0 : g_j(r_h) \leq \sigma_j^2 \tag{2.20}$$

$$g_j(r_h) - \sigma_j^2 \leq 0 \tag{2.21}$$

To perform this test, we can compute a one-sided lower 95% bootstrap confidence limit for the difference in the estimated power spectrum (at each frequency of interest) and the AR noise, and, using zero as our critical value, we reject the null only if this lower confidence limit exceeds zero. The significance level can be Bonferroni corrected for multiple comparisons if many frequencies are being considered. A rejected null hypothesis implies that the IC under consideration exhibits "significant" brain activity of the tested type.

2.3 Simulation Studies

2.3.1 Preservation of Correlation Structures in Bootstrap Samples

All simulations are carried out in R statistical software (R Core Team, 2016). In this section, we intend to demonstrate through simulations that key cross-correlation and auto-correlations structures in EEG data are preserved in the bootstrap samples of the data constructed using our bootstrapping algorithm. We also compare these properties in bootstrap samples constructed using simpler bootstrapping procedures. Simulation structures in this section are informed by the simulations of Lee et al. (2011).

Given fixed values of k and T, mixed signal data, $\boldsymbol{X}$, are simulated by first generating T realizations from each of k independent signals with AR time series structures (corresponding to the ICs), concatenating these signals into the rows of a matrix, $\boldsymbol{S}$, and then mixing the signals using a fixed, orthogonal matrix of mixing parameters, $\boldsymbol{A}$, i.e., $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$. Here we consider $k = \{2, 5\}$ and $T = 1,000$. For the $k = 2$ simulation, the AR structures used for the ICs are as follows:

- IC 1: AR(2), $\phi_{11} = 1, \phi_{12} = -.21$ with random error from Unif($-\sqrt{3}$, $\sqrt{3}$).
- IC 2: AR(1), $\phi_{21} = .3$ with random error from Normal(0,1).

For $k = 5$, the first two ICs are simulated as above, and the remaining ICs are generated from the following AR processes:

- IC 3: AR(2), $\phi_{31} = 1.3, \phi_{32} = -0.7$ with random error from Unif($a = -\sqrt{9}, b = \sqrt{9}$).
- IC 4: AR(1), $\phi_{41} = -0.8$ with random error from Laplace($\mu = 0, b = 1$).
- IC 5: AR(2), $\phi_{51} = 0.5, \phi_{52} = 0.2$ with random error from Logistic($\mu = 0, s = 1.5$).

These are chosen to test a range of time series structures and error distribution shapes.

Our bootstrapping algorithm, which we refer to as semi-parametric CICA bootstrapping, is applied to each simulated dataset to collect 1,000 bootstrap samples. We also collect 1,000 bootstrap samples of the data using each of three simpler bootstrapping methods. The first, which we call non-parametric data bootstrapping, resamples non-parametrically directly from the observed signals, i.e. the rows of $\boldsymbol{X}$. The second, called semi-parametric data bootstrapping, resamples from the AR residuals from AR models fit for each of the observed signals and plugs back into the AR model to construct a bootstrap sample of the data (analogous to the way that semi-parametric CICA bootstrapping constructs a bootstrap sample of the ICs). We also consider bootstrapping non-parametrically from the CICA estimated ICs, a method we call non-parametric CICA bootstrapping. To do so, we simply apply CICA to the simulated data, resample non-parametrically from the estimated ICs, and construct a bootstrap sample of the data by multiplying the resampled ICs by the estimated mixing matrix.

To examine the preservation of the data correlation structures among these bootstrapping methods, we compare the average lag 1 auto-correlation of each of the signals across bootstrap samples with the lag 1 auto-correlations in the observed data, and we compare the average lag 0 cross-correlation between each pair of signals across bootstrap samples with the lag 0 cross-correlations in the observed data. Table 2.3 provides the results of this simulation. The results indicate that, as expected, the semi-parametric CICA bootstrap is the only method considered that preserves both the auto-correlation and the cross-correlation structures in the bootstrap samples. By focusing our bootstrapping at the IC level and mixing the bootstrapped ICs, cross-correlation structures are retained in the bootstrap samples, and by semi-parametrically bootstrapping from the AR models of the ICs, auto-correlation structures are retained. Each of the simpler approaches to bootstrapping fails to preserve one or both of these types of correlation, making them unsuitable for use in the connectivity analysis setting.

2.3.2 CICA Confidence Interval Coverage Rates

In this section, we conduct simulations to test whether the bootstrap CIs formed for the CICA parameters using our method attain the appropriate coverage rates. We simulate mixed signal data using the same procedure as described in the previous section and the same IC structures, although we now consider a range of T values for each k: $T = \{500, 1000, 5000, 10000\}$. For each combination of k and T, we simulate 2,000 mixed signal datasets and apply our bootstrapping algorithm with $B = 1,000$ to each dataset to obtain 95% CIs for the CICA parameters. Coverage rates are computed as the percentage of simulated datasets for which the 95% CI contains the parameter's true value.

Coverage rates for IC AR model parameters for the $k = 2$ and $k = 5$ simulations can be found in Table 2.4. Plots of the coverage rates of the pointwise IC spectra CIs for the $k = 2$ simulations can be found in Figure 2.1. Parameter coverage rates improve overall as T increases, with 28% of parameters achieving 95% or greater coverage at $T = 500$ and 100% of parameters achieving 95% or greater coverage at $T = 10,000$. While these coverage rates provide compelling evidence for the reliability of the bootstrap CIs with T large, making them well-suited for use on EEG data, some caution should be exercised in applying these CIs in other settings, such as fMRI data, in which T is typically small. Refer to the discussion section for additional discussion of the applicability of these CIs in fMRI data.

2.3.3 Hypothesis Testing for Power Spectra Peaks

To demonstrate the use of the lower confidence limits of the IC spectra for detecting significant activity peaks, we now simulate data from signals with cyclic properties. We create four-channel mixed signal data recorded at a rate of 200Hz for a total of 15 seconds ($T = 3000$). Four ICs are used to generate the mixed signal data. The first is a pure noise signal with high variability. High variability noise ICs are common in EEG data. ICs 2, 3, and 4 are the "brain activity" ICs, containing delta, alpha, and beta activity, respectively. The ICs can be summarized as follows:

- IC 1: $S_1(t) = \epsilon(t)$
- IC 2: $S_2(t) = \sum_{z_1}(\sin(2\pi * (z_1/200) * (t + 1000))) + \gamma w(t)$
- IC 3: $S_3(t) = \sum_{z_2}(\sin(2\pi * (z_2/200) * (t + 1000))) + \gamma w(t)$

- IC 4: $S_4(t) = \sum_{z_3}(\sin(2\pi * (z_3/200) * (t + 1000))) + \gamma w(t)$

where z_1 is a sequence of 25 frequencies between 2.5 and 3.5 Hz, z_2 is a sequence of 25 frequencies between 9.5 and 10.5 Hz, and z_3 is a series of 25 frequencies between 16.5 and 17.5 Hz. $\epsilon(t)$ is random noise from a $N(0, 5)$, $w(t)$ is an AR(1) process with $\phi = 0.3$ and noise from a $N(0, 1)$, and γ is a scalar used to control the signal to noise ratio (Lee et al., 2011). Figure 2.2 shows the first 500 time points of the sinusoidal signals (ICs 2-4) with no noise and with signal-to-noise ratios (SNRs) of 0.25 and 0.50. For SNRs of 0.200, 0.225, 0.250, and 0.500, we generate 500 replicates of the ICs and used them, along with a fixed orthogonal mixing matrix, to create 500 mixed signal datasets.

We investigate the power of our hypothesis testing method at each SNR by applying the bootstrap algorithm to each simulated dataset and finding the proportion of times our hypothesis test correctly detects a single IC with each of delta, alpha, and beta activity. Figure 2.3 plots the power at each SNR. These results demonstrate that our hypothesis testing method is highly powered for detecting brain activity in the ICs, even in the presence of strong noise, making it well-suited for EEG data, which are known to be highly noisy.

2.4 Resting State EEG Analysis

The following analyses are performed using R (R Core Team, 2016), MATLAB (The MathWorks Inc., 2015), and EEGLAB (Delorme and Makeig, 2004) softwares. Resting state, eyes open EEG data were recorded from a single subject using an EEG cap with 32 electrodes, including vertical and horizontal electro-oculograms. Channels were referenced to the right mastoid (M2). Samples were collected at a rate of 500Hz with a 0.1570Hz bandpass recording filter. A 1Hz high pass filter was applied to the data to remove low frequency activity, such as slow drift, and the data were thinned to include only 10,000 time points.

We apply the CICA bootstrap to these data, including a pre-whitening step in the CICA algorithm. Pre-whitening is a common ICA pre-processing technique (Hyvarinen and Oja, 2000), which we apply to ensure that our assumption that the mixing matrix is orthogonal is met. Figure 2.4 and Figure 2.5 contain the resulting lower 95% bootstrap confidence limits for the difference in the spectrum and the AR error variance for ICs 1-16 and ICs 17-32, respectively. Figure 2.6 provides the topographical maps ($\hat{\boldsymbol{A}}$) corresponding to each of the ICs. While many of the IC spectra exhibit a "bump" in the alpha range, the lower 95% confidence limit for the difference in the spectrum and the AR error variance, which can be used for hypothesis testing, suggest that these bumps are not indicative of significant alpha activity in most of the ICs. Only nine of the ICs– 3, 4, 6, 8, 9, 10, 11, 12, and 14 – have a 95% confidence limit exceeding zero in the alpha range. Thus, we conclude that only these ICs contain significant alpha activity. The topographical maps for the ICs identified by our method as containing alpha activity largely agree with previous research, with alpha activity most prominent in the posterior regions during rest (Barry et al., 2007).

This analysis suggests that conclusions drawn from our method about which ICs contain brain activity may differ dramatically from the conclusions one might make by simply "eyeballing" IC spectra plots. Consider, for instance, IC 15. While the spectrum for this IC exhibits a spike in alpha range and the spatial map exhibits high values in the posterior region of the head, which would likely lead us to assume this IC contains alpha activity without a formal testing mechanism, our test shows that, after appropriately accounting for the variability in the IC, this peak is not statistically significant. Erroneous conclusions about the type(s) of activity contained in an IC could result in misleading connectivity inference; thus, our formal hypothesis test for brain activity in the ICs is needed in order to increase the reliability of EEG connectivity studies.

2.5 Discussion

In this paper, we proposed a semi-parametric bootstrapping algorithm for constructing bootstrap samples of resting state EEG data and creating confidence intervals for CICA parameters, which can be used in resting state EEG connectivity analyses to detect brain activity in ICs. We demonstrated how the bootstrap samples created with this algorithm preserve correlation structures in the data that are critical to assessing connectivity, while simpler bootstrapping methods do not preserve these features. We also constructed simulations to demonstrate the reliable performance of the confidence intervals for the IC time series parameters and to confirm that our hypothesis testing approach has high power, even when SNRs in the ICs are low. Finally, we applied the hypothesis testing method to an EEG resting state dataset to identify ICs containing significant alpha activity. This analysis revealed that a formal hypothesis testing mechanism like ours is needed in order to take into account the variability in the IC-related estimates when using them to make a decision about the presence of brain activity, otherwise erroneous conclusions could easily be made. Such erroneous decisions threaten the validity of results and inference made in downstream connectivity analyses.

The use of uncertainties in ICA has previously been limited because asymptotics for these methods are difficult and unappealing. This often leads to ad hoc and subjective decision making based on ICA results. To our knowledge, ours is the first attempt to develop a bootstrapping approach that can be used to measure uncertainty, create CIs, and perform hypothesis tests on either single subject or group ICA parameters. Because group ICA is typically performed by simply concatenating the data across subjects into a single matrix and applying ICA to all the data simultaneously, our bootstrap approach could easily be applied in this setting. While we have focused on an application to EEG resting state connectivity, the potential demonstrated by our method to accurately capture uncertainty in ICA parameters could have much more far-reaching effects. Variations of this approach could be used to construct CIs and hypothesis tests for task-based EEG and fMRI analyses.

One limitation to our approach that is crucial to address in order to extend the applicability of this method is that it cannot yet accommodate pre-ICA dimension reduction procedures. Pre-ICA dimesion reduction can be achieved using principal component analysis or singular value decomposition. Such procedures are extremely common in fMRI applications (McKeown et al., 1998), where the high spatial resolution can make direct application of ICA computationally untenable, and are increasingly appearing in EEG analyses as well, as the number of recording channels increases (De Vos et al., 2011; Dyrholm et al., 2007; Kachenoura et al., 2008; McMenamin et al., 2010; Xu et al., 2004). Future work will investigate an extension of this approach to account for pre-ICA dimension reduction.

2.6 Tables and Figures

Table 2.1: Steps to create a bootstrap sample of EEG data.

1. Run CICA on the observed EEG scalp channel data, $\boldsymbol{X}$
2. For each IC, resample from its estimated AR model residuals, $\hat{\epsilon_j}(t)$
3. Plug resampled residuals back into the fitted AR model to get a bootstrap sample of the IC, $\tilde{S}_j(t)$
4. Concatenate the $\tilde{S}_j(t)$ into a matrix $\tilde{\boldsymbol{S}}$
5. Create a bootstrap sample, $\tilde{\boldsymbol{X}}$, of the data by plugging in $\tilde{\boldsymbol{X}} = \hat{\boldsymbol{A}}\tilde{\boldsymbol{S}}$

Table 2.2: Steps to form bootstrap uncertainties for CICA parameters.

1. Form a large number, B, of bootstrap samples of the data using the method described in Table 2.1
2. Run CICA on each bootstrap sample to get B bootstrap estimates of all parameters
3. Permute each set of bootstrap estimates to order the ICs and corresponding parameters in the same way they are ordered in the original estimates (based on the correlation in the IC spectra in the original and bootstrap estimates)
4. Using the permuted bootstrap estimates, compute bootstrap SEs and apply the percentile method to create CIs for the IC AR parameters and power spectra (pointwise)

Table 2.3: True lag 1 auto-correlation (L1AC) and lag 0 cross-correlation (L0CC) for observed signals from two simulated datasets ($k = 2; T = 1,000$ and $k = 5; T = 1,000$) and average L1AC and L0CC across 1,000 bootstrap samples of each dataset for each of the following four bootstrapping methods: non-parametric data bootstrapping (NP Data), semi-parametric data bootstrapping (SP Data), non-parametric CICA bootstrapping (NP CICA), and semi-parametric CICA bootstrapping (SP CICA).

	Measure	Signal(s)	Truth	NP Data	SP Data	NP CICA	SP CICA
$k = 2$	L1AC	1	0.36	-0.00	0.36	-0.00	0.37
		2	0.82	-0.00	0.82	-0.00	0.82
	L0CC	1 and 2	0.31	-0.00	0.00	0.31	0.31
$k = 5$	L1AC	1	0.59	-0.00	0.59	-0.00	0.60
		2	0.70	-0.00	0.70	-0.00	0.70
		3	0.75	-0.00	0.74	-0.00	0.74
		4	-0.75	-0.00	-0.74	-0.00	-0.75
		5	0.69	-0.00	0.68	-0.00	0.69
	L0CC	1 and 2	-0.12	0.00	0.00	-0.10	-0.10
		1 and 3	0.72	0.00	-0.00	0.72	0.72
		1 and 4	0.10	0.00	0.00	0.11	0.11
		1 and 5	-0.30	-0.00	0.00	-0.26	-0.26
		2 and 3	0.25	0.00	0.00	0.27	0.27
		2 and 4	0.03	0.00	0.00	0.03	0.03
		2 and 5	0.56	0.00	0.00	0.57	0.58
		3 and 4	-0.16	0.00	0.00	-0.15	-0.15
		3 and 5	-0.09	-0.00	0.00	-0.06	-0.06
		4 and 5	-0.01	0.00	0.00	-0.01	-0.01

Table 2.4: Coverage rates of semi-parametric CICA bootstrap 95% CIs for IC AR time series parameters over 2,000 simulations.

		$T = 500$	$T = 1,000$	$T = 5,000$	$T = 10,000$
$k = 2$	ϕ_{11}	0.90	0.93	0.94	0.95
	ϕ_{12}	0.95	0.96	0.96	0.95
	ϕ_{21}	0.94	0.94	0.95	0.95
	σ_1^2	0.96	0.95	0.96	0.95
	σ_2^2	0.93	0.93	0.95	0.95
$k = 5$	ϕ_{11}	0.87	0.92	0.96	0.95
	ϕ_{12}	0.94	0.97	0.98	0.97
	ϕ_{21}	0.94	0.95	0.95	0.95
	ϕ_{31}	0.89	0.94	0.96	0.95
	ϕ_{32}	0.92	0.97	0.97	0.96
	ϕ_{41}	0.94	0.97	0.98	0.98
	ϕ_{51}	0.97	0.97	0.97	0.96
	ϕ_{52}	0.93	0.95	0.97	0.95
	σ_1^2	0.95	0.97	0.98	0.96
	σ_2^2	0.91	0.91	0.95	0.96
	σ_3^2	0.95	0.96	0.94	0.95
	σ_4^2	0.91	0.93	0.94	0.95
	σ_5^2	0.89	0.93	0.96	0.95

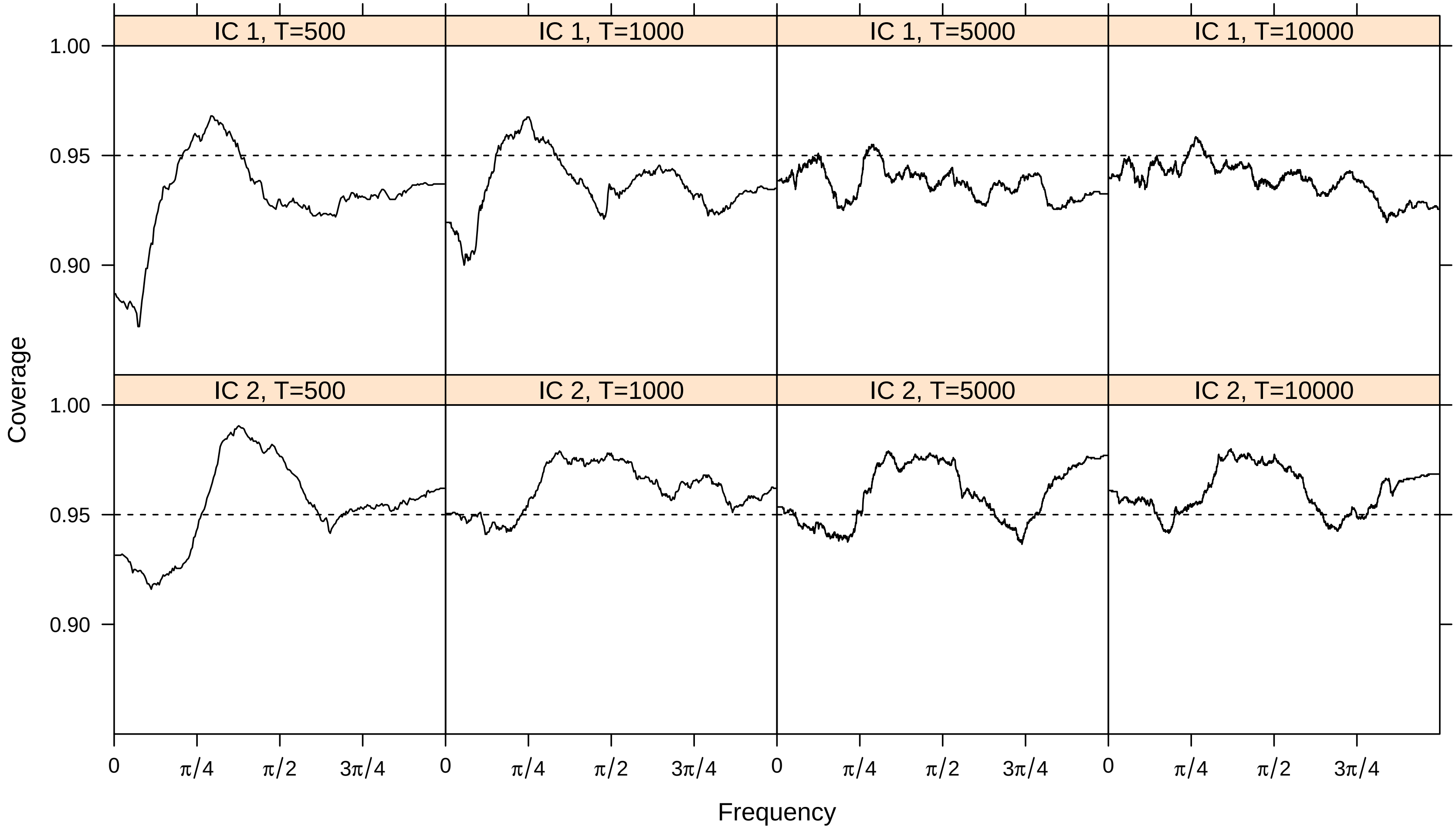


Figure 2.1: Pointwise spectra CI coverage rates from $k = 2$ simulations.

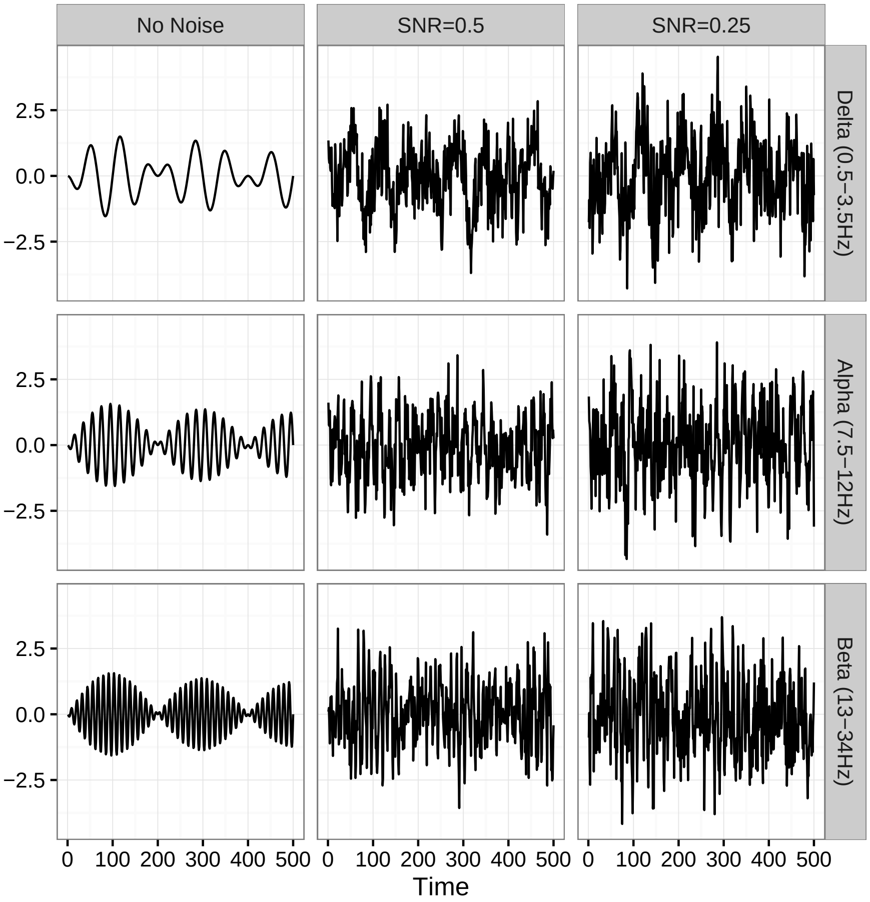


Figure 2.2: Delta, alpha, and beta signals without noise and with SNRs of 0.5 and 0.25.

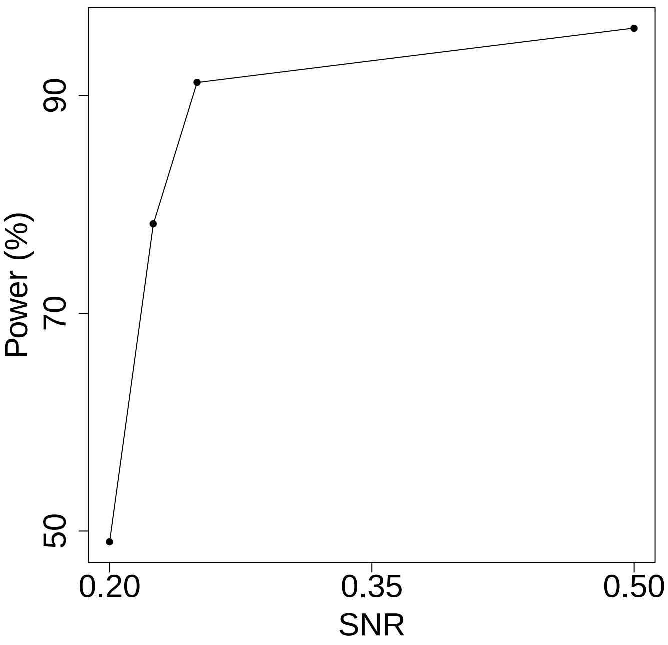


Figure 2.3: Power of our hypothesis test to detect a single IC with each of delta, alpha, and beta activity at various SNRs.

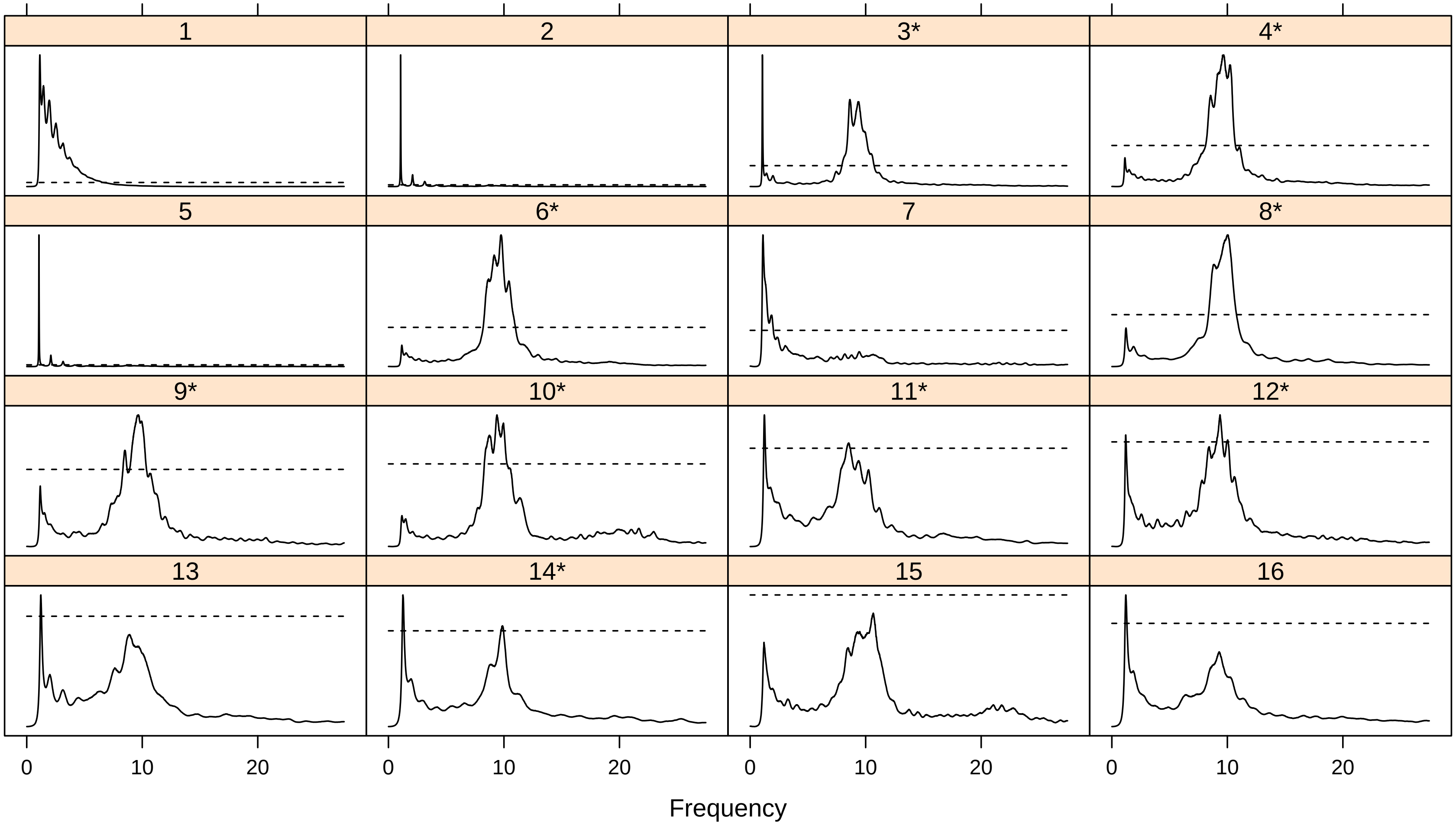


Figure 2.4: Lower 95% confidence limits for the difference in the spectrum and the AR error variance (solid line) for ICs 1-16 with zero indicated by a dotted line. ICs containing significant alpha activity are labeled with a *.

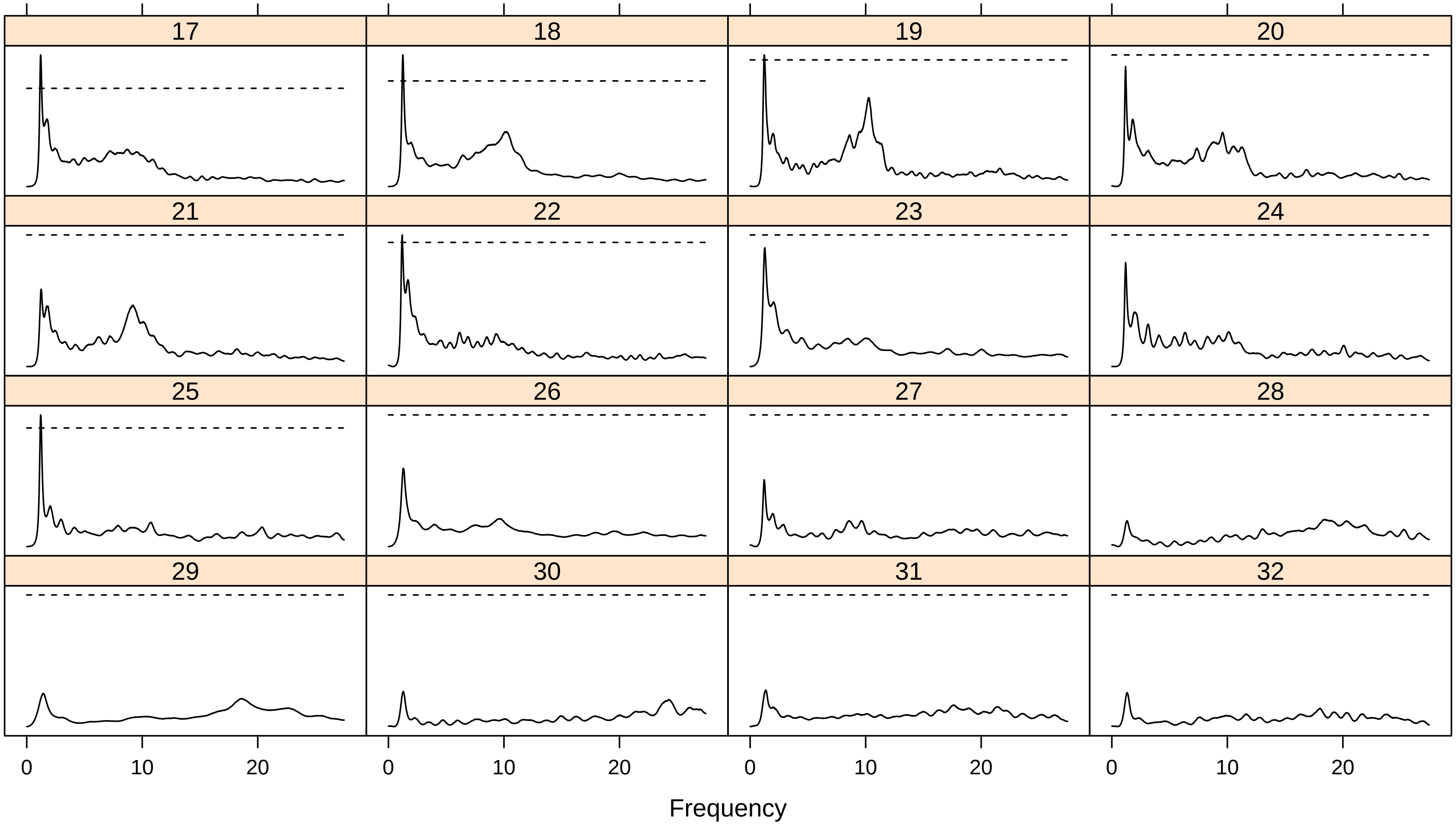


Figure 2.5: Lower 95% confidence limits for the difference in the spectrum and the AR error variance (solid line) for ICs 17-32 with zero indicated by a dotted line. ICs containing significant alpha activity are labeled with a *.

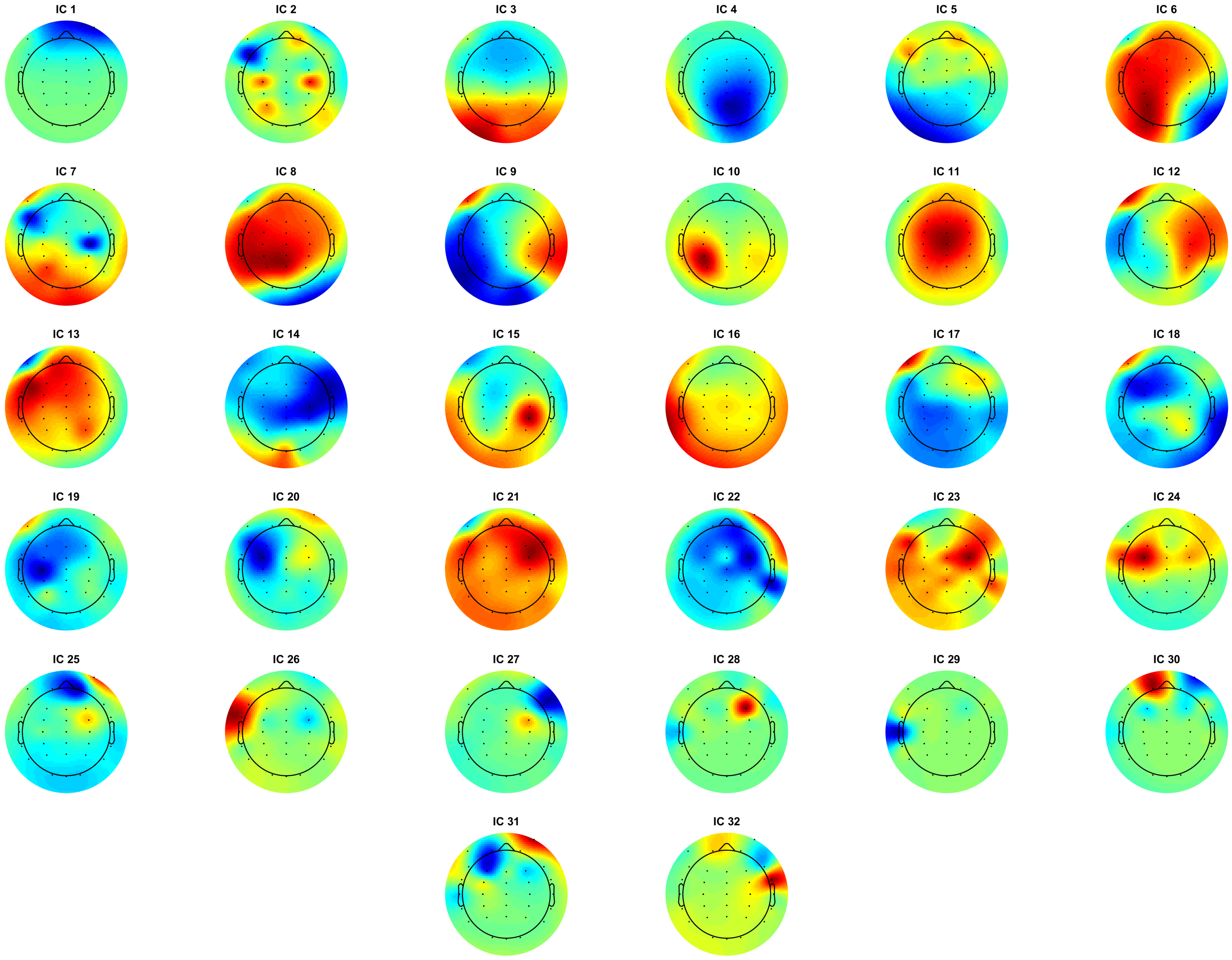


Figure 2.6: Topographical maps for each IC.

CHAPTER 3: A BOOTSTRAP APPROACH TO COMPARE METHODS FOR EEG DATA DIMENSION REDUCTION PRIOR TO INDEPENDENT COMPONENT ANALYSIS

3.1 Introduction

Independent component analysis (ICA) is a multivariate statistical method that can be used to decompose recorded mixtures of signals into independent source signals, also called independent components (ICs). ICA is commonly used to unmix signal mixtures recorded by biomedical devices, such as functional magnetic resonance imaging (fMRI) and electroencephalogram (EEG) scanners. EEG scanners use metal electrodes positioned on the scalp to record electrical signals. The signal recorded by each electrode may be a mixture of the brain signals of interest, artifacts, and noise, and separation of the brain signals from one another and from the artifacts/noise is often desired for inference (Makeig et al., 1996).

Letting k denote the number of electrodes/recording locations and T denote the number of time points at which recordings are made, the ICA model has the form

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} \tag{3.22}$$

where $\boldsymbol{X}$ is a $k \times T$ matrix of observed data with each row containing the recording from a given electrode over time, $\boldsymbol{S}$ is a $k \times T$ matrix of ICs with each row containing the values of a given IC over time, and $\boldsymbol{A}$ is a $k \times k$ matrix of linear mixing parameters which represent the contribution of each IC to the recording at each electrode. Note that this formulation of the ICA model is often called temporal ICA, and, while other formulations of the model are possible, we use ICA to refer exclusively to temporal ICA. ICA estimation is performed by first estimating the mixing parameters in $\boldsymbol{A}$ in a manner that imposes as few assumptions as possible about the distributions of the ICs, and then predicting the IC values by plugging in $\hat{\boldsymbol{S}} = \hat{\boldsymbol{A}}^{\boldsymbol{-1}}\boldsymbol{X}$. Most ICA algorithms constrain $\boldsymbol{A}$ to be orthogonal in order to reduce the number of parameters being estimated and improve convergence.

One limitation of ICA in EEG, as well as a number of other applications, is its assumption that the number of ICs is equal to the number of electrodes, k. In recent years, as the push to improve the spatial resolution of EEG data has led to increasing numbers of electrodes, researchers often believe that the number of independent source signals generating EEG data is smaller than the number of electrodes and that much of the observed data are pure noise, an assumption adopted from fMRI applications of ICA (Beckmann and Smith, 2004; Calhoun et al., 2001a,b; Chawla, 2011; Cordes and Nandy, 2006; De Vos et al., 2011; Delorme et al., 2007; Dyrholm et al., 2007; Kachenoura et al., 2008; McMenamin et al., 2010; Varoquaux et al., 2010; Xu et al., 2004). This scenario is referred to in the literature as overdetermined ICA (Winter et al., 2003). ICA model fitting in the overdetermined setting has been shown to lead to overlearning (also called overfitting), which can often be recognized by the estimation of a single large peak in each IC (Särelä and Vigário, 2003).

Not only does application of ICA in the overdetermined case lead to poor model performance, but it can also introduce an enormous computational burden, only to estimate many ICs that reflect uninteresting noise. Because $\boldsymbol{A}$, which contains $k(k-1)/2$ free parameters (Hyvarinen and Oja, 2000), is the primary target of estimation, the number of parameters to be estimated and, consequently, the computational burden of ICA increases dramatically as k increases. The computational burden is less impacted by the increase of T. The heavy computational burden of ICA when k is very large has plagued users of ICA with fMRI for decades, as k in fMRI data can be greater than 1,000,000, leading the fMRI community to develop procedures to reduce the dimensionality of the data prior to the application of ICA (McKeown et al., 1998). While the low spatial resolution of EEG scanners has historically produced data to which ICA can be directly applied without major computational obstacles, improvements in spatial resolution have recently introduced these concerns for users of ICA on EEG data.

Thus far, the EEG community's response to the challenge of large k and overdetermined ICA has been somewhat ad hoc. Following the fMRI community's lead, EEG users have applied methods to reduce the dimensionality of the data prior to ICA. This pre-ICA dimension reduction is typically performed using one of two approaches. The first, which we call "theory based" dimension reduction (TBDR), takes into consideration EEG theory and the goals of the analysis and, based on these considerations, simply discards data from electrodes that are believed to be irrelevant and applies ICA to the remaining data (Frank and Frishkoff, 2007; Lau et al., 2012). While integration of subject matter knowledge is critical to proper statistical analyses, this method could preclude novel findings by limiting the analysis to only investigation of the expected associations. Moreover, this approach wastes resources and potentially valuable information.

The second approach to pre-ICA EEG dimension reduction, which we call "procedure based" dimension reduction (PBDR), has been adopted from the fMRI literature. In it, statistical dimension reduction procedures, such as principal component analysis (PCA) or singular value decomposition (SVD), are applied to the observed data to reduce its dimensionality to m ($m < k$) and ICA applied to the dimension reduced data (Calhoun et al., 2001b; Delorme et al., 2007; Petersen et al., 2000). In some cases, authors claim to know a reasonable value for m a priori based on subject matter knowledge (Xu et al., 2004). In other cases, a reasonable value for m may be inferred through exploratory analyses (Calhoun et al., 2001b). While this approach has been widely used in the context of fMRI, its effects have not been sufficiently investigated in the EEG setting. Furthermore, to our knowledge, these two approaches to pre-ICA dimension reduction in EEG have not been compared. In this paper, we propose a method that can be used to aid in the comparison of results from these two approaches.

According to Petersen et al. (2000), pre-ICA PBDR with SVD can be performed using the following procedure, which yields equivalent results to a PCA on the covariance matrix of the data. Considering a $k \times T$ matrix of mixed signal data, $\boldsymbol{X}$, that has been row-centered, decompose $\boldsymbol{X}$ as

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}', \tag{3.23}$$

where $\boldsymbol{U}$ is a $k \times k$ matrix of left singular vectors, $\boldsymbol{V}$ is a $T \times T$ matrix of right singular vectors, and $\boldsymbol{D}$ is a $k \times T$ diagonal matrix with singular values on the diagonal. If we assume that m components form the signal subspace of the data and the remaining components form a noise subspace, then we can partition each of these matrices into terms corresponding to the signal and noise subspace, denoted by S and N, respectively, in the following way:

$$\begin{aligned} \boldsymbol{U} &= \begin{bmatrix} \boldsymbol{U_S} & \boldsymbol{U_N} \end{bmatrix} \\ \boldsymbol{D} &= \begin{bmatrix} \boldsymbol{D_S} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D_N} \end{bmatrix} \\ \boldsymbol{V} &= \begin{bmatrix} \boldsymbol{V_S} & \boldsymbol{V_N} \end{bmatrix} \end{aligned} \tag{3.24}$$

where $\boldsymbol{U_S}$ is $k \times m$, $\boldsymbol{U_N}$ is $k \times (k-m)$, $\boldsymbol{D_S}$ is $m \times m$, $\boldsymbol{D_N}$ is $(k-m) \times (T-m)$, $\boldsymbol{V_S}$ is $T \times m$, and $\boldsymbol{V_N}$ is $T \times (T-m)$.

Then, a reduced dataset with dimensions $m \times T$, $\boldsymbol{X_R}$, can be formed by $\boldsymbol{X_R} = \boldsymbol{D_S}\boldsymbol{V_S'}$. We use the terms reduced dataset and signal subspace interchangeably to refer to $\boldsymbol{X_R}$. Note that, while the terms signal and noise subspace are adopted to be consistent with the literature, these terms could be misleading. Applied in this way, SVD finds a set of uncorrelated basis vectors of the data in which the first m components explain the maximal amount of variability in the data. Thus, pure noise components with no interesting temporal structure, if they account for a large amount of the variance in the data, will be retained in the signal subspace after SVD. Hence, while intended to remove noise, this method could result in the removal of important signals if they account for very little of the variability in the data (Särelä and Vigário, 2003). Theoretical details on the assumptions of pre-ICA dimension reduction are discussed at length by Comon and Jutten (2010).

A salient feature of SVD is that it not only produces a reduced dataset but it also solves much of the problem of ICA by finding uncorrelated components (UCs) of the original data (in the rows of $\boldsymbol{X_R}$). While the UCs are not generally equivalent to ICs, they are typically "closer" to the ICs than the original data, so that only a small orthogonal rotation of the UCs is needed to recover the ICs (Hyvarinen and Oja, 2000). We demonstrate this property through simulations in Section 3.3.1. This partial solving of the ICA problem, along with the reduced data size, can substantially lighten the computational burden of ICA when applied after SVD.

ICA is performed on $\boldsymbol{X_R}$, so that only m ICs are estimated. The resulting $m \times m$ estimated mixing matrix, $\hat{\boldsymbol{A}}$, is multiplied by $\boldsymbol{U_S}$ to produce a $k \times m$ matrix of mixing coefficients, $\hat{\boldsymbol{C}}$. In $\hat{\boldsymbol{C}}$, each column contains the set of mixing coefficients that quantify the relationships between a given

IC and each electrode, also known as the spatial map of the IC; thus, in this case $\hat{\boldsymbol{C}}$ is of interest rather than the ICA estimated mixing matrix.

Because much of the brain's electrical activity is known to demonstrate cyclic patterns (Schomer and Da Silva, 2012), an ICA algorithm that allows for temporal correlation in the ICs is needed in this context; however, many popular ICA algorithms require the ICs to be i.i.d. over time (Bell and Sejnowski, 1995; Hyvarinen and Oja, 2000). Thus, we focus instead on a semi-parametric algorithm called colorICA (CICA) (Lee et al., 2011), which assumes that the ICs are autoregressive (AR) time series processes with unspecified error distributions, i.e. for the j^{th} IC, $S_j(t)$,

$$S_j(t) = \mu + \sum_{h=1}^{p} \phi_h S_j(t-h) + \epsilon_j(t), \tag{3.25}$$

where p is the AR order chosen by model selection and $\epsilon_j(t)$ is an error term with unspecified distribution and variance σ_j^2. We consider an approach that applies SVD to reduce the observed data, followed by CICA to separate the reduced data into ICs. We refer to this procedure as SVD-CICA.

Uncertainty estimation in ICA is typically foregone, because asymptotic theory is complex and requires restrictive assumptions that sacrifice the highly appealing flexibility of ICA. Without uncertainties, hypothesis testing and inference cannot be carried out; thus, proper consideration cannot be given to the question of how different pre-ICA dimension reduction approaches impact results and inference. In Chapter 2, we proposed a bootstrapping approach to allow for inference and hypothesis testing in ICA (Efron and Tibshirani, 1986). However, when procedure based pre-ICA dimension reduction is performed, this approach must be extended to accommodate the additional uncertainty introduced into the estimation by the application of two statistical procedures rather than one. Fortunately, the flexibility of the bootstrap allows us to estimate uncertainties that take into account the variability of parameter estimates produced by applying these two procedures sequentially.

In this paper, we introduce a semi-parametric bootstrapping algorithm to estimate standard errors (SEs) and create confidence intervals (CIs) for the IC time series parameters and spectral densities from SVD-CICA applied to resting state EEG data. This method is an extension of the one presented in Chapter 2 to integrate the SVD step into the semi-parametric creation of the bootstrap samples and the construction of empirical distributions for the ICA parameters. The uncertainties formed from these empirical distributions can then be used for hypothesis testing on the ICs. For our purposes, the benefits of this method are two-fold. First, because both ICA and bootstrapping are computationally intensive and EEG data are typically large, allowing for pre-ICA dimension reduction, and thereby reducing computation time, dramatically increases the usability of this bootstrapping approach compared to that of Chapter 2. Second, combined with the ICA bootstrap proposed in Chapter 2, this method provides a novel opportunity to formally compare the results of no pre-ICA dimension reduction, TBDR, and PBDR in EEG analyses.

In Section 3.2, we introduce the bootstrapping algorithm to compute uncertainties for and perform hypothesis tests on SVD-CICA estimates. We simulate data in Section 3.3 in order to test the performance of our bootstrapping approach and to compare the bootstrap uncertainties from SVD-CICA with the bootstrap uncertainties resulting from running CICA on the full data. In Section 3.4, we apply the SVD-CICA bootstrap as well as the CICA bootstrap with TBDR to eyes open and eyes closed resting state EEG data to demonstrate how the resulting bootstrap hypothesis tests aid in the comparison of the results from these two methods. Finally, we summarize and discuss our method and findings in Section 3.5.

3.2 Methods

We refer the reader back to Table 2.1 and Table 2.2 in Chapter 2 for descriptions of how to create bootstrap samples of the data and how to estimate SEs and create CIs for the CICA parameters in the setting without dimension reduction. We use the same approach with a few additional steps to create bootstrap samples and CIs in the presence of dimension reduction. Again, useful bootstrap samples must preserve both the cross-correlation and auto-correlation in the observed data, which are a result of the mixing of the ICs and the temporal correlation in the ICs, respectively; thus, we must resample from the independent AR residuals of the ICs in order to retain these properties. However, in order to properly characterize the uncertainty in the estimates, we must also integrate the dimension reduction step into our bootstrap algorithm.

To create a bootstrap sample of the row-centered mixed signal data, $\boldsymbol{X}$, when pre-ICA dimension reduction is needed, we first reduce the data using SVD, as described in the previous section, to get an $m \times T$ matrix $\boldsymbol{X_R}$. We then apply CICA to $\boldsymbol{X_R}$ and create a bootstrap sample of each IC that retains its temporal structure by resampling with replacement from its estimated AR residuals and plugging the resampled residuals into the estimated AR model (Efron and Tibshirani, 1986). As before, the block initialization method of Stine (1987) can be used to initialize the bootstrapped time series. The bootstrapped ICs should then be concatenated into a matrix, $\tilde{\boldsymbol{S}}$, and a bootstrap sample of the reduced data, or signal subspace, can be constructed as $\tilde{\boldsymbol{X_R}} = \hat{\boldsymbol{A}}\tilde{\boldsymbol{S}}$.

Finally, to create a bootstrap sample that fully captures the noisiness of the original data, we must mix the removed noise subspace with the bootstrap sample of the signal subspace. Thus, the bootstrap sample can be constructed as

$$\tilde{\boldsymbol{X}} = \boldsymbol{U} \begin{bmatrix} \tilde{\boldsymbol{X_R}} \\ \boldsymbol{D_N V_N'} \end{bmatrix}. \tag{3.26}$$

This procedure is summarized in Table 3.5.

When a large number, B, of bootstrap samples of the data are created in this manner, we can perform SVD-CICA on each of them and use the resulting set of bootstrap parameter estimates to assess the uncertainty in the estimates. Let $\boldsymbol{C_1^*}, ..., \boldsymbol{C_B^*}$, $\boldsymbol{S_1^*}, ..., \boldsymbol{S_B^*}$, $\boldsymbol{\phi_1^*}, ..., \boldsymbol{\phi_B^*}$, $\boldsymbol{\sigma^2}_{\boldsymbol{1}}^{\boldsymbol{*}}, ..., \boldsymbol{\sigma^2}_{\boldsymbol{B}}^{\boldsymbol{*}}$, and $\boldsymbol{G_1^*}, ..., \boldsymbol{G_B^*}$ denote the B estimates for each parameter resulting from the application of SVD-CICA to each of the bootstrapped datasets. Then, due to the ICA permutation ambiguity discussed in Chapter 2, these bootstrap estimates must be permuted in order to ensure that parameters associated with corresponding ICs are located in the same position in each set of bootstrap estimates. To do this, we apply the same procedure described in Chapter 2. We identify matching ICs in the original and bootstrap estimates based on the correlation in the IC spectra estimates ($\hat{\boldsymbol{G}}$ and $\boldsymbol{G_b^*}$), and we use this information to order the bootstrap estimated ICs and corresponding bootstrap parameter estimates in the same order as the ICs are originally estimated.

After the bootstrap estimates have been permuted appropriately, SEs can be computed pointwise for the IC AR parameters and spectra or any linear combination of these parameters by simply computing the standard deviation over all the bootstrap estimates (Efron and Tibshirani, 1986). CIs can also be created pointwise using the percentile method (Efron and Tibshirani, 1986). This procedure for computing uncertainties for the SVD-CICA parameters is summarized in Table 3.6. In the analysis of EEG data, we can use this approach to perform hypothesis tests to detect brain activity in the ICs by constructing a lower 95% bootstrap confidence limit for the difference in the spectrum and the AR error variance of each IC and determining whether the CI exceeds zero in the frequency range of interest. This procedure is described in greater detail in the context of the CICA bootstrap in Chapter 2 and can be applied equivalently with the SVD-CICA bootstrap.

3.3 Simulation Studies

3.3.1 How SVD-CICA Works

All simulations were carried out using R statistical software (R Core Team, 2016). In this section, we present a simple simulated scenario to demonstrate how SVD and CICA can work together to recover the independent signals underlying a dataset, while decreasing computing time over ICA alone by removing low variance noise. With $T = 3000$, we generate realizations from an IC containing delta activity, an IC containing alpha activity, and an IC that is white noise with variance lower than that of the signals using the following procedure:

- IC 1: $S_1(t) = \sum_{z_1}(\sin(2\pi * (z_1/200) * (t + 1000))) + \gamma w(t)$
- IC 2: $S_2(t) = \sum_{z_2}(\sin(2\pi * (z_2/200) * (t + 1000))) + \gamma w(t)$
- IC 3: $S_3(t) = \epsilon(t)$

where z_1 is a sequence of 25 frequencies between 2.5 and 3.5 Hz, z_2 is a sequence of 25 frequencies between 9.5 and 10.5 Hz, $\epsilon(t)$ is random noise from a $N(0, 0.5)$, $w(t)$ is an AR(1) process with $\phi = 0.3$ and noise from a $N(0, 1)$, and γ is a scalar used to set the signal to noise ratio to 0.5 (Lee et al., 2011). We then mix these signals using an orthogonal mixing matrix, apply SVD to reduce the data to $m = 2$, and perform CICA on the reduced dataset.

Figure 3.7 provides the spectra of the original ICs, the mixed signals, the SVD decomposed UCs, and the SVD-CICA estimated ICs. This figure illustrates that SVD removes the noise IC from the data while retaining the signals and also that, by finding UCs, SVD alone solves much of the problem of separating our components of interest. However, the UCs from SVD are clearly somewhat mixed versions of the original ICs, with each component containing a considerable bump in each of the delta and alpha ranges. Further application of CICA is needed to achieve full separation of the ICs.

Although SVD-CICA preserves the signals of interest effectively when the variance of the pure noise is low, as in the above example, we emphasize that, in the presence of noise or artifacts with higher variance than the signals of interest, the signals of interest may be at risk of being distorted or removed by SVD if the situation is not properly accounted for when choosing m. To illustrate this, we repeat the above simulation while increasing the variance of IC 3 to 1, 1.1, and 1.2, so that it has variance comparable to the variances of IC 1 and IC 2. The resulting SVD-CICA estimated IC spectra are plotted in Figure 3.8. While the signals are still preserved at variance 1, they are somewhat distorted at variance 1.1. At variance 1.2, the noise IC is fully preserved while the signals are partially removed by SVD, with what remains of the two signals forced into a single IC. Because EEG data is known to be corrupted by high variance noise and artifacts, it is critical to allow for more ICs in SVD-CICA than the number of anticipated brain activity ICs, otherwise critical brain activity could be removed by SVD, in which the first few components may be dominated by noise and artifacts.

3.3.2 Bootstrap Performance when Removing Low Variance Noise

To test the performance of the SVD-CICA bootstrap, we simulate overdetermined mixed signal datasets, i.e. mixed signal datasets in which the number of signal ICs is less than the number of observed recording locations (rows of $\boldsymbol{X}$). The datasets are generated by mixing m signal ICs and $k-m$ pure noise ICs using an orthogonal mixing matrix. In particular, we used two signal ICs ($m = 2$) and eight noise ICs ($k = 10$). The signal ICs are given below.

- IC 1: AR(2), $\phi_{11} = 1.3, \phi_{12} = -.7$ with random error from Unif($-\sqrt{9}$, $\sqrt{9}$).
- IC 2: AR(2), $\phi_{21} = .5, \phi_{12} = .2$ with random error from Logistic(0,1.5).

In the first set of simulations, we generate the eight noise ICs from distributions with variances considerably smaller than the variance of the signal ICs. In particular, four of the noise ICs come from Normal distributions with zero mean and variances $\{0.001, 0.072, 0.144, 0.215\}$ and four come from exponential distributions with variances $\{0.286, 0.357, 0.429, 0.500\}$. We simulated 500 mixed signal datasets using these specifications for each of $T = \{2500, 5000\}$.

The SVD-CICA bootstrap with $m = 2$ and $B = 1,000$ was applied to each of the simulated datasets. Because the variance of each of the noise ICs is smaller than the variance of the signals, SVD separates the noise into the noise subspace, which is removed, and the signals are retained in the reduced dataset, to which CICA is applied. Thus, our bootstrap algorithm provides SEs and CIs for the AR parameters for IC 1 and IC 2. The 95% CI coverage rates for each of these parameters is given in Table 3.7.

For each simulated dataset, the CICA bootstrap is also performed on the full data. The coverage rates for the IC AR parameters can be seen to be similar to those from the SVD-CICA bootstrap (Table 3.7). Moreover, we want to compare the SEs from the SVD-CICA bootstrap with the SEs from the application of the CICA bootstrap to the full data. This provides insight into how the variability of ICA estimates is impacted by pre-ICA dimension reduction. Figure 3.9 provides boxplots of the SEs for each signal IC AR parameter from each simulation (outliers omitted). The distribution of the SEs across simulations are highly similar for SVD-CICA and CICA on the full data, with SVD-CICA producing only marginally higher distributions for several parameters. This indicates that very little precision in the parameter estimation is lost by applying SVD prior to ICA. We also note that the SEs from the $T = 5000$ simulations demonstrate a substantial decrease compared to the $T = 2500$ simulations for both SVD-CICA and CICA, suggesting good performance of these bootstrap SEs.

3.3.3 Bootstrap Performance when Noise Variance Exceeds Signal Variance

The simulated data above, while providing a simple and insightful example, may not reflect real biomedical signal data, because such data may contain noise that has variance greater than the variances of the signals of interest. This is particularly common in EEG data, where noise ICs often account for a large portion of the variability in the data. In this section, we test the SVD-CICA bootstrap on mixed signal data generated using noise ICs with variances both larger and smaller than the variance of the signal ICs. When noise explains more of the variability in the data than the signals, it will be retained in the signal subspace during SVD; thus, unless we make m large enough to allow for it, it will distort the estimation of the ICs. We wish to investigate if and how the performance of SVD-CICA is impacted by this high variance noise when it is accounted for in the model by increasing m appropriately.

We simulate data as in Section 3.3.2, by mixing two signal ICs with eight noise ICs using an orthogonal mixing matrix. The signal ICs are generated from the same AR processes as in Section 3.3.2. Now, however, one of the eight noise ICs has variance greater than the variance of the signal ICs.

The four Normally distributed noise ICs have mean zero and variances $\{10, 0.001, 0.084, 0.167\}$ and the four exponential distributed noise ICs have variances $\{0.251, 0.334, 0.417, 0.500\}$. Again, we create 500 datasets for each of $T = \{2500, 5000\}$. We apply the SVD-CICA bootstrap to each dataset, with $B = 1,000$ and $m = 3$, to allow for the high variance noise IC and both signal ICs to be retained in the signal subspace.

As before, we consider the coverage rates (Table 3.8) and SEs (Figure 3.10) for the signal IC AR parameters from the SVD-CICA bootstrap and from the application of the CICA bootstrap to the full data. We see that, having allowed for the high variance noise to remain in the data and separated it into its own IC, the bootstrap CIs for the signal AR parameters perform just as well as above. Again, SVD-CICA and CICA on the full data produce SEs that are remarkably similar.

3.4 Resting State EEG Analysis

The following analyses are performed using R (R Core Team, 2016), MATLAB (The MathWorks Inc., 2015), and EEGLAB (Delorme and Makeig, 2004) softwares. In this section, we demonstrate the use of bootstrap hypothesis tests for comparing the results of EEG analyses using different pre-ICA dimension reduction techniques. To compare results, we test hypotheses about the presence of brain activity in the ICs estimated by each technique, using the approach described in Chapter 2, i.e. we form a lower 95% confidence limit on the difference in each IC spectrum and its AR error variance and, if it exceeds zero in the frequency range of the brain activity type of interest, we reject the null hypothesis of no activity.

In particular, we use these hypothesis tests to detect alpha activity (frequency 7.5-12 Hz) in ICs estimated with TBDR and PBDR from both eyes open and eyes closed resting state EEG scans. It is well known that alpha activity is prominent during rest with eyes closed and is suppressed when the eyes are opened, a phenomenon observed in the very early days of EEG research and often called "alpha desynchronization" (Chen et al., 2013; Klimesch et al., 2000; Pollen and Trachtenberg, 1972). In accordance with this phenomenon, we expect to detect alpha activity in more ICs from the eyes open scans than eyes closed scan. Furthermore, alpha activity is known to originate primarily from the posterior regions of the head.

The EEG data analyzed were collected by Schalk et al. (2004) using BCI2000 software (Schalk Lab, 2017) and are publicly available through PhysioNet (Goldberger et al., 2000). We consider the eyes closed and the eyes open scan from Subject 1. In each scan, data are recorded from 64 channels at a speed of 160 Hz for 2 minutes, with the electrodes positioned on the scalp in accordance with the international 10-10 system. An earlobe channel was used as the reference. Further information about the data collection procedures can be found at `https://www.physionet.org/pn4/eegmmidb/`. Prior to analysis, we applied a 1Hz high pass filter to the data to remove low frequency activity.

We first consider an analysis using PBDR, by applying the SVD-CICA bootstrap to these datasets. The dimensionality of the data is reduced from 64 to 20 in the SVD step ($k = 64$, $m = 20$). These 20 components retain 98.9% and 98.8% of the variance in the full data for the eyes closed and eyes open scans, respectively. To speed up computation and improve convergence, we pre-whiten the data in the SVD step as well, by constructing the reduced data using only $\boldsymbol{V}_{\boldsymbol{S}}'$ rather than $\boldsymbol{D}_{\boldsymbol{S}}\boldsymbol{V}_{\boldsymbol{S}}'$ and adjusting the bootstrapping algorithm accordingly.

Figure 3.11 and Figure 3.12 provide the bootstrap lower 95% confidence limit for the difference in the IC spectrum and the AR error variance for each of the 20 ICs from the eyes closed and eyes open scans, respectively. The corresponding spatial maps are provided in Appendix A, Figure 20 and Figure 21. In the eyes closed scan, our hypothesis testing method detects alpha activity in 14 out of 20 ICs. The spatial maps indicate that the six ICs not containing alpha activity may be eye and muscle artifacts. On the other hand, during the eyes open scan, our method only detects alpha activity in 11 out of 20 ICs. This is consistent with Barry et al. (2007), who find reductions in alpha activity between resting state eyes closed and eyes open scans.

Because alpha activity originates primarily in the posterior region of the head (Barry et al., 2007), we investigate IC alpha activity with TBDR by restricting our analysis to electrodes in this region. For the eyes open and eyes closed scans, we select 20 electrodes in this region for analysis, so that results are comparable with the PBDR above. We note that this approach will only allow us to uncover ICs from which activity has been recorded at the selected electrodes; thus, the resulting set of 20 ICs is likely to look quite different from the SVD-CICA ICs. Moreover, IC spatial maps will be restricted to the selected electrodes.

The CICA bootstrap hypothesis test plots for alpha activity in the ICs from eyes closed and eyes open scans are contained in Figure 3.13 and Figure 3.14, respectively, with corresponding spatial maps in Appendix A, Figures 22 and 23. 18 ICs from the eyes closed data contain alpha activity, according to the bootstrap hypothesis tests with TBDR, considerably more than the 14 that contain alpha activity in the eyes closed data using PBDR. The spatial maps from TBDR, while they may restrict some types of inference by being confined to a small area, do provide a more detailed glimpse of how alpha activity is distributed in the posterior region of the head. In the eyes closed data, we see that many of the ICs containing alpha rhythms are most active on the right side of the head in the posterior region (ICs 2, 3, 6, 7, 9, and 10). TBDR makes this trend much more obvious than PBDR. However, Barry et al. (2007) show that, particularly during rest with eyes closed, alpha activity can be detected widely across the scalp and is not restricted to only the posterior regions, inference that is missed by TBDR.

Surprisingly, while TBDR leads to more ICs containing alpha activity in the eyes closed data, our hypothesis tests suggest that many fewer ICs from the eyes open data contain alpha activity when using TBDR compared to PBDR. Significant alpha activity is detected in only four ICs from TBDR, compared to 11 from PBDR. One possible explanation for this phenomenon relates to the fact that eyes open data are likely to contain both more noise and more artifactual activity than eyes closed data, due to increased eye and muscle movement. Compared to TBDR, PBDR is better able to remove noise, and, by preserving the spatial map across the entire scalp, PBDR can more easily separate artifacts like eye and muscle movement. Thus, in the eyes open data, the improved reduction of noise and separation of artifacts provided by PBDR may enhance our ability to detect alpha activity in the ICs when compared with TBDR. This explanation is supported by the spatial maps, many of which demonstrate prominent activity on the right side of the head, as in the eyes closed data, but we are unable to detect significant alpha activity in these ICs, possibly due to the increased noise.

3.5 Discussion

Due to the high dimensional nature of data in the primary application areas of ICA, dimension reduction procedures are often needed prior to ICA in order to ensure that the assumptions of ICA are met and the application of ICA is computationally feasible. Asymptotic variance estimation in ICA is challenging and distastefully restrictive even in the simplest of scenarios. The use of pre-processing procedures such as dimension reduction further complicates this situation by adding another source of variability to the ICA estimates.

In response to these challenges, we developed a semi-parametric bootstrapping approach to estimate uncertainties for CICA parameters in the presence of procedure based pre-ICA dimension reduction. The resulting uncertainties take into account the variability in the parameter estimates introduced by both ICA and the pre-ICA dimension reduction step without imposing restrictive parametric assumptions, and, compared to computing bootstrap uncertainties for CICA on the full data, this method provides a substantial decrease in the required computation time. Moreover, these uncertainties allow for novel hypothesis tests to be performed on the ICA parameters to detect the presence of brain activity in the ICs. Using these inferential tools, we are able to formally compare approaches to pre-ICA dimension reduction in resting state EEG, where such procedures are relatively new and unexplored.

In Section 3.3, we performed simulations to demonstrate that the CIs formed by our bootstrapping approach achieve appropriate coverage. We also used simulations to compare the bootstrap standard errors for signal ICs of interest from CICA on the full data to those from SVD-CICA. We found that procedure based pre-ICA dimension reduction adds surprisingly little variability to these parameter estimates, with the distribution of the standard errors from SVD-CICA shifted up only slightly, if at all, compared to the distribution of the standard errors from CICA on the full data.

Finally, we performed a novel comparison of the results of pre-ICA TBDR and PBDR using single subject resting state EEG data with eyes open and eyes closed by applying both the CICA bootstrap (on a relevant subset of the EEG channels) and the SVD-CICA bootstrap. In the eyes closed data, both TBDR and PBDR produced many ICs containing alpha activity, with PBDR allowing for the analysis of this activity across the entire scalp while TBDR provides a more detailed picture of the distribution of alpha activity in the posterior region of the head, where it primarily originates. Thus, with eyes closed data, the preferable approach to pre-ICA dimension reduction may depend on the goals of the analysis. On the other hand, due to increased noise in the eyes open data, PBDR seemed to yield better results than TBDR, allowing for the detection of alpha activity in more ICs thanks to its reduction of noise and easier separation of artifacts.

Over the past decade, EEG has begun to play an increasing role in our understanding of the brain's functionality, thanks to the high temporal resolution of EEG data compared to other imaging techniques like fMRI. Many studies now take a multi-modal approach to investigating the brain, recording fMRI and EEG simultaneously and analyzing the data together in a manner that takes advantage of the high spatial resolution of fMRI and the high temporal resolution of EEG (Huster et al., 2012). As these multi-modal studies, which collect huge quantities of data, become increasingly popular, dimension reduction procedures become increasingly critical; thus, we must improve our understanding of how such procedures may impact conclusions. To improve our understanding of the impact of different pre-ICA dimension reduction techniques on inference in EEG data analysis, comparisons such as the ones presented here should be considered in a diverse range of settings.

3.6 Tables and Figures

Table 3.5: Steps to create a bootstrap sample of mixed signal data using SVD-CICA.

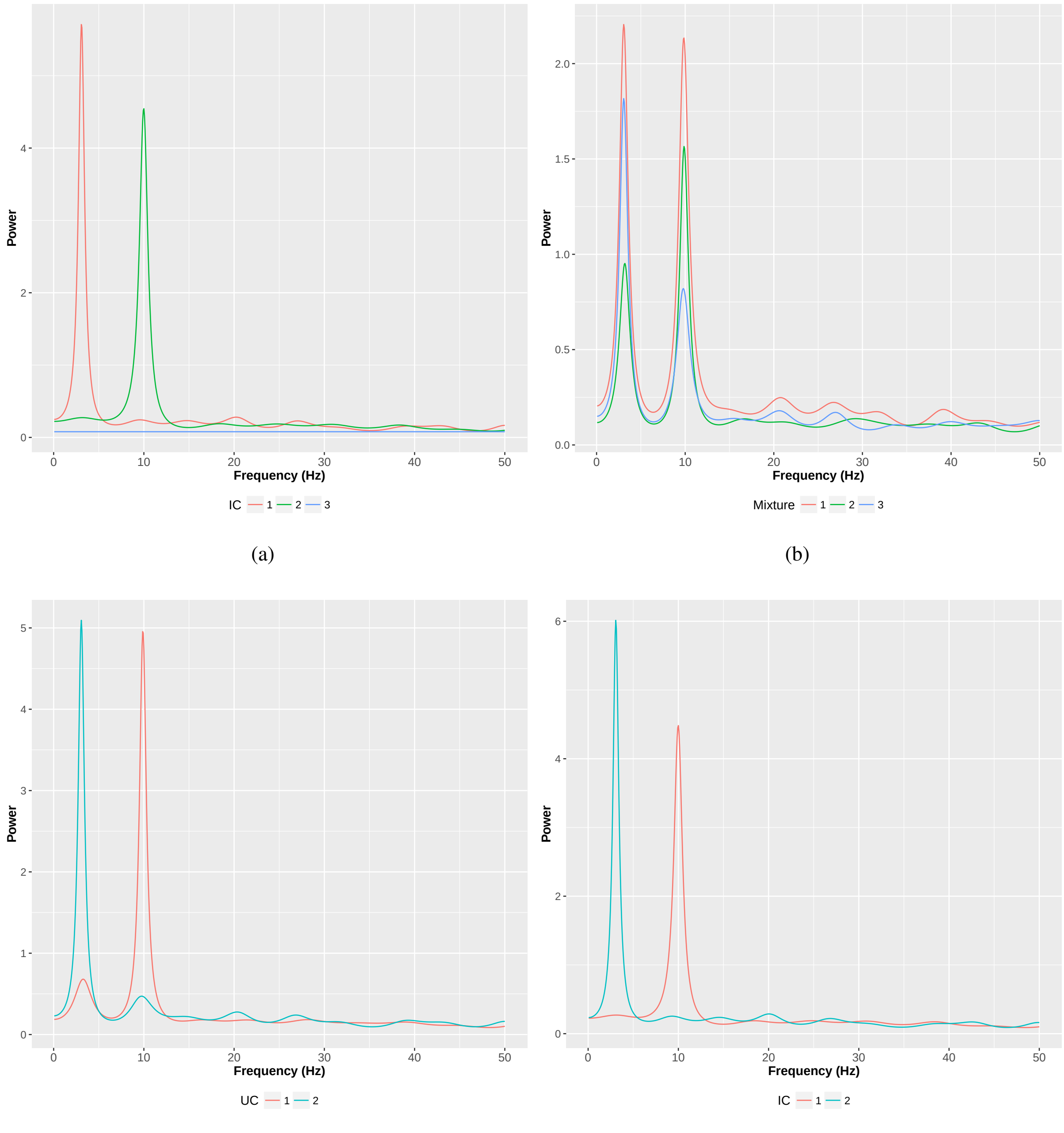
1. Reduce the observed data, $\boldsymbol{X}$, using SVD to get an $m \times T$ dataset $\boldsymbol{X_R}$
2. Run CICA on $\boldsymbol{X_R}$
3. For each IC, resample from its estimated AR model residuals, $\hat{\epsilon_j}(t)$
4. Plug resampled residuals back into the fitted AR model to get a bootstrap sample of the IC, $\tilde{S_j}(t)$
5. Concatenate the $\tilde{S_j}(t)$ into a matrix $\tilde{\boldsymbol{S}}$
6. Create a bootstrap sample of the signal subspace, $\tilde{\boldsymbol{X_R}} = \hat{\boldsymbol{A}}\tilde{\boldsymbol{S}}$
7. Form a bootstrap sample of the original data by mixing the noise subspace with the bootstrap sample of the signal subspace, i.e. $\tilde{\boldsymbol{X}} = \boldsymbol{U} \left[\begin{array}{c} \tilde{\boldsymbol{X_R}} \\ \boldsymbol{D_N V_N'} \end{array} \right]$

Table 3.6: Steps to compute SEs and form bootstrap CIs for SVD-CICA parameters

1. Form a large number, B, of bootstrap samples of the data using the method described in Table 3.5
2. Run SVD-CICA on each bootstrap sample to get B bootstrap estimates of all parameters
3. Permute each set of bootstrap estimates to order the ICs and corresponding parameters in the same way they are ordered in the original estimates (based on the correlation in the IC spectra in the original and bootstrap estimates)
4. Compute pointwise standard errors for IC AR parameters and spectra by finding the standard deviation over all bootstrap estimates or form CIs using the percentile method

Table 3.7: 95% bootstrap confidence interval coverage rates for IC AR parameters from low variance noise simulations with SVD-CICA and CICA on the full data.

	SVD-CICA		CICA	
	$T = 2500$	$T = 5000$	$T = 2500$	$T = 5000$
ϕ_{11}	0.96	0.97	0.96	0.97
ϕ_{12}	0.98	0.97	0.98	0.98
ϕ_{21}	0.96	0.95	0.96	0.96
ϕ_{22}	0.95	0.95	0.95	0.95
σ_1^2	0.95	0.96	0.95	0.95
σ_2^2	0.93	0.94	0.93	0.95



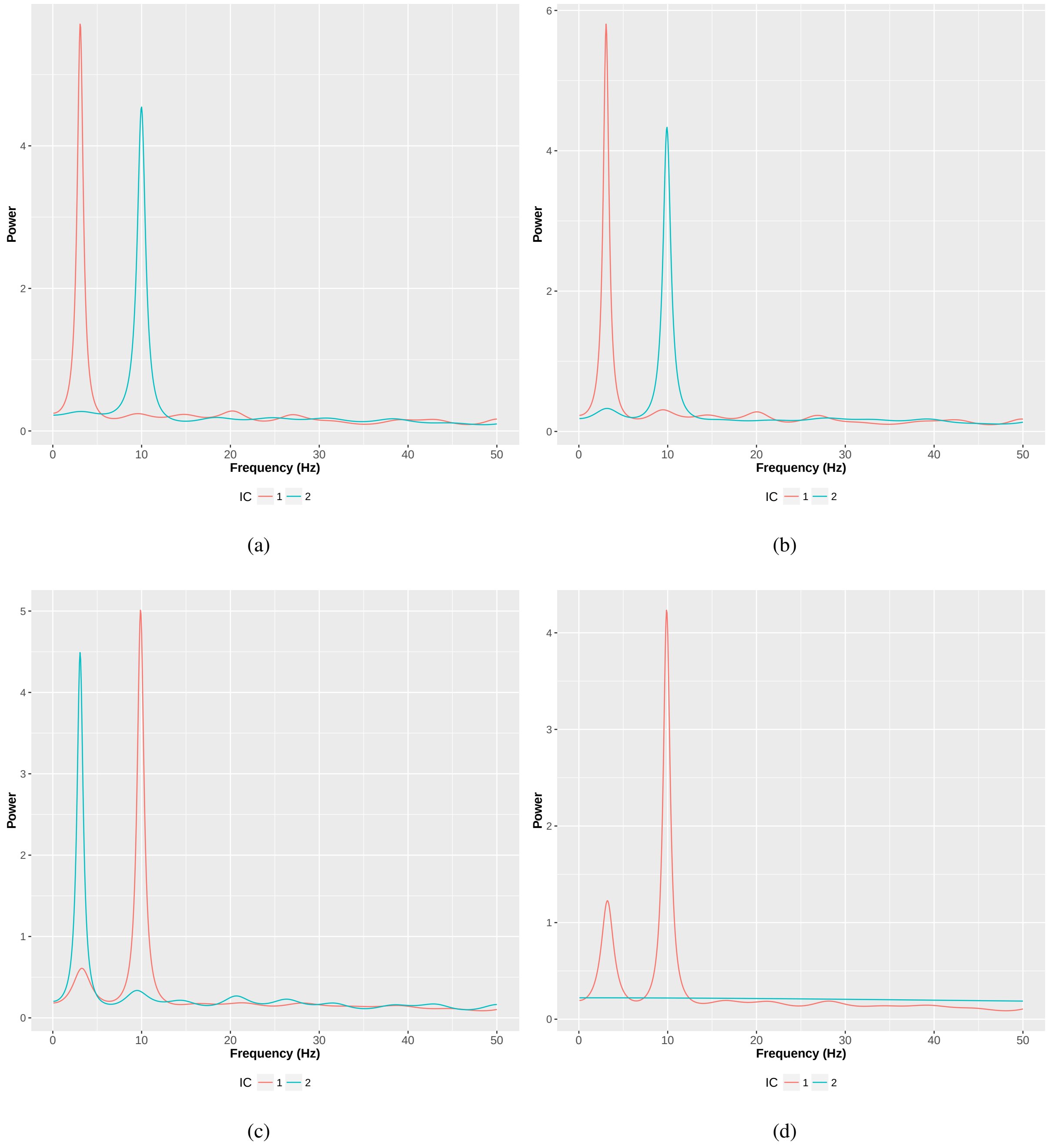
(a)

(b)

(c)

(d)

Figure 3.7: Spectra of the original ICs (a), the mixed signals (b), the SVD UCs (c), and the SVD-CICA estimated ICs (d).



(a)

(b)

(c)

(d)

Figure 3.8: Spectra of the original brain activity ICs (a) and the estimated ICs with noise variance 1 (b), noise variance 1.1 (c), and noise variance 1.2 (d).

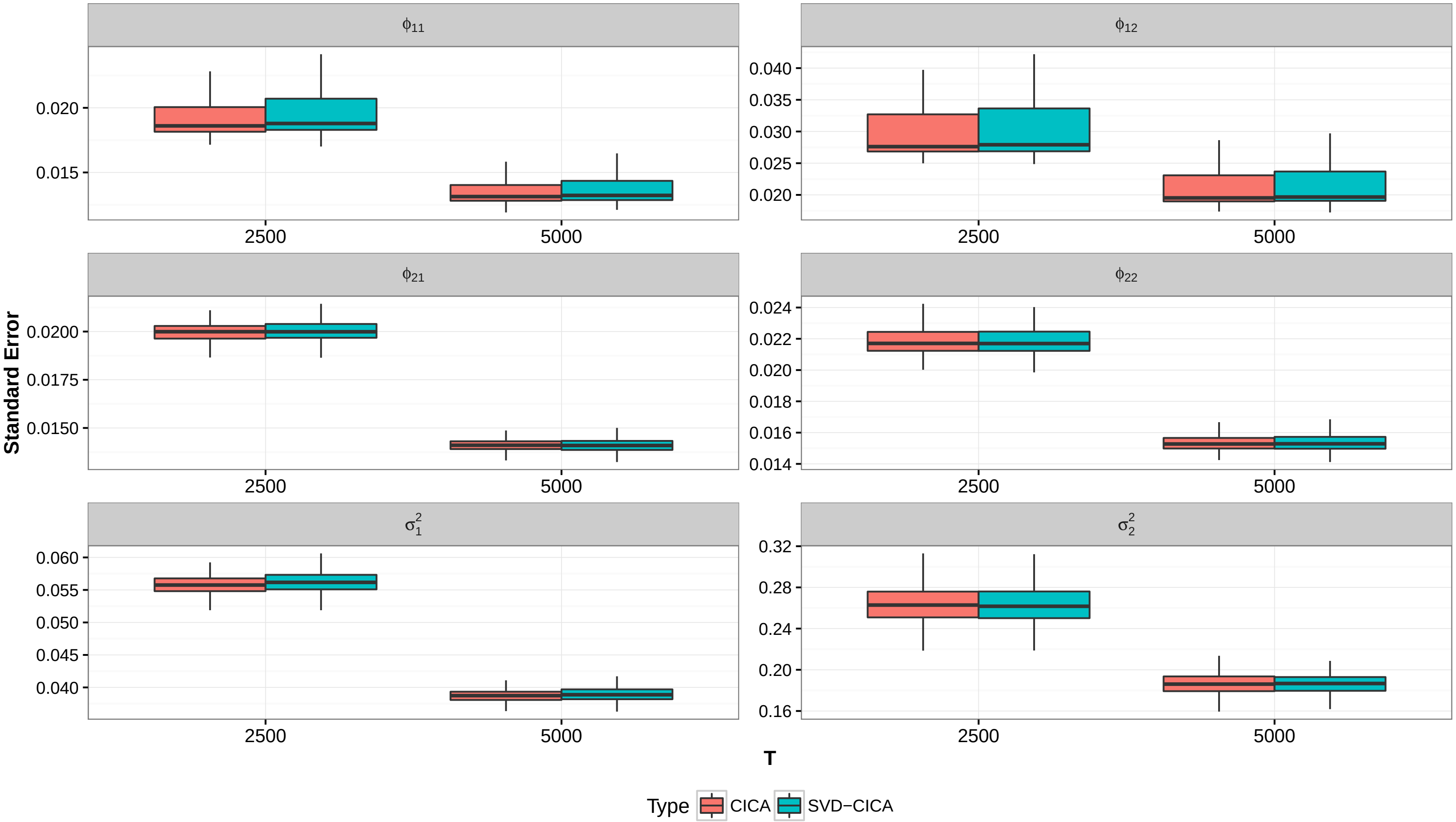


Figure 3.9: Standard errors for the signal IC AR parameters from the low variance noise simulations for both SVD-CICA and CICA on the full data.

Table 3.8: 95% bootstrap confidence interval coverage rates for IC AR parameters from high variance noise simulations with SVD-CICA and CICA on the full data.

	SVD-CICA		CICA	
	$T = 2500$	$T = 5000$	$T = 2500$	$T = 5000$
ϕ_{11}	0.97	0.97	0.97	0.96
ϕ_{12}	0.97	0.98	0.98	0.98
ϕ_{21}	0.95	0.95	0.96	0.95
ϕ_{22}	0.95	0.95	0.95	0.95
σ_1^2	0.96	0.96	0.96	0.97
σ_2^2	0.94	0.93	0.95	0.93

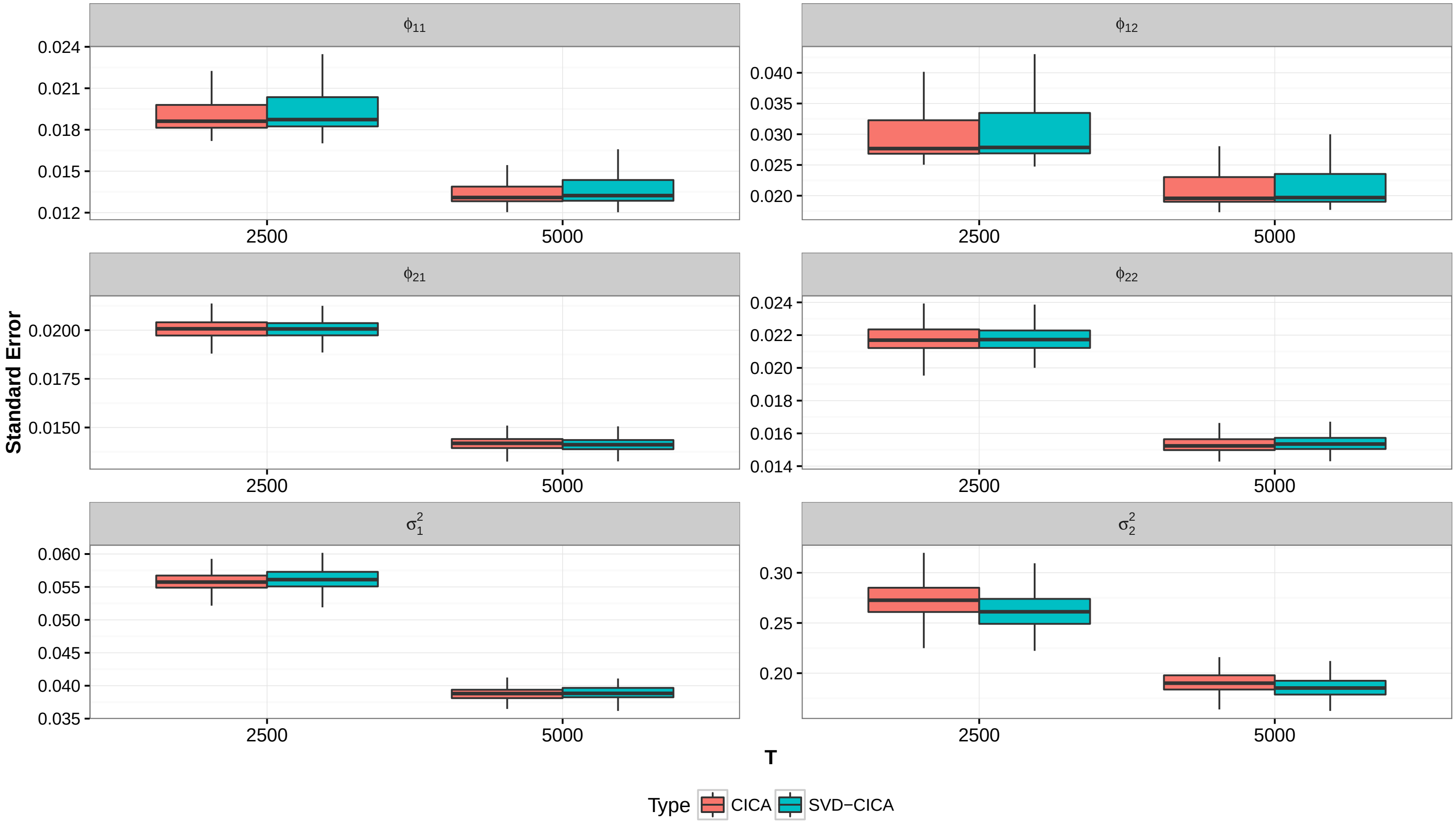


Figure 3.10: Standard errors for the signal IC AR parameters from the high variance noise simulations for both SVD-CICA and CICA on the full data.

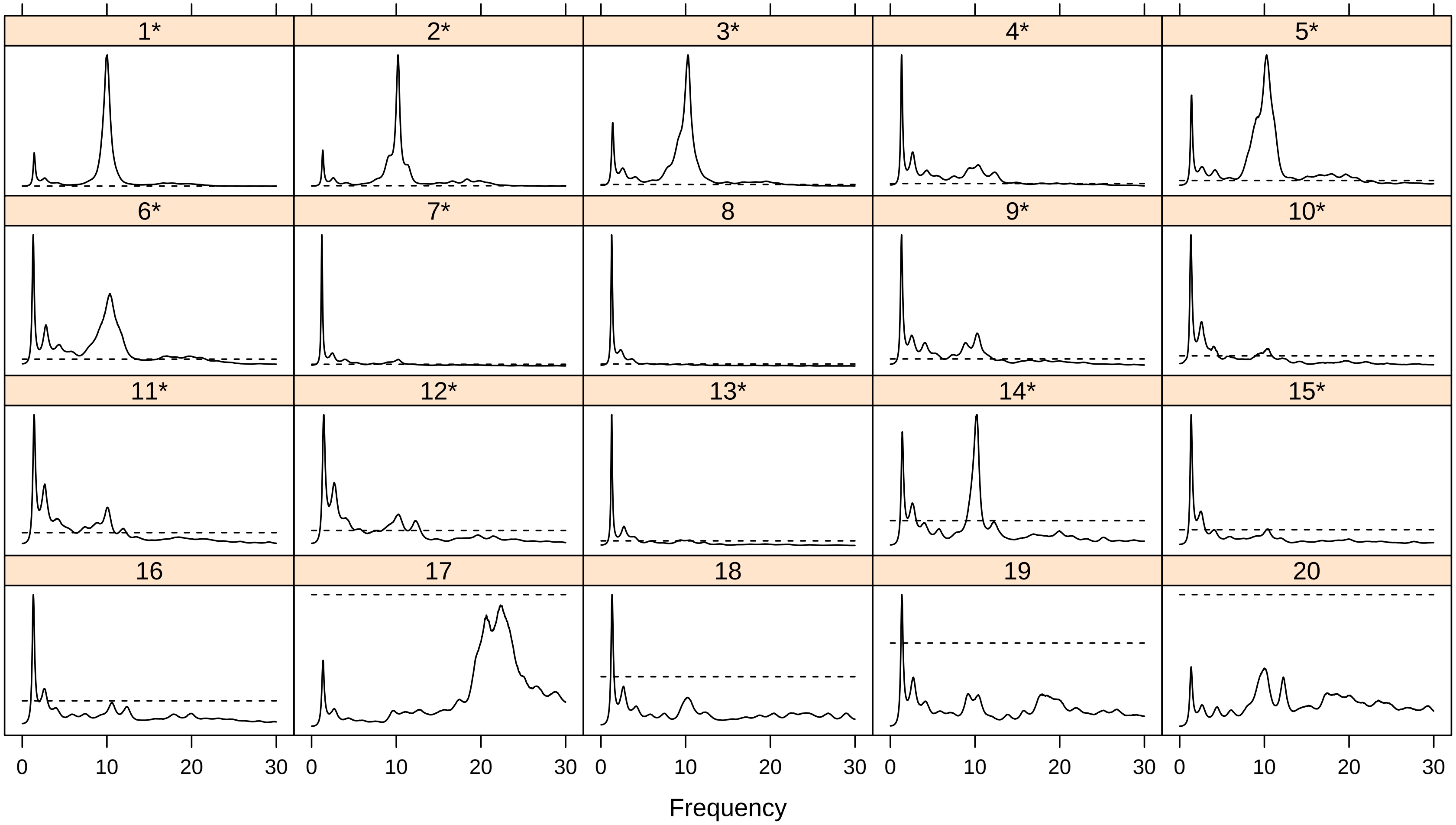


Figure 3.11: Spectrum brain activity hypothesis tests (solid line) for ICs 1-20 from the SVD-CICA bootstrap on the eyes closed data with a dotted line at zero indicating the null hypothesis. ICs containing significant alpha activity are labeled with a *.

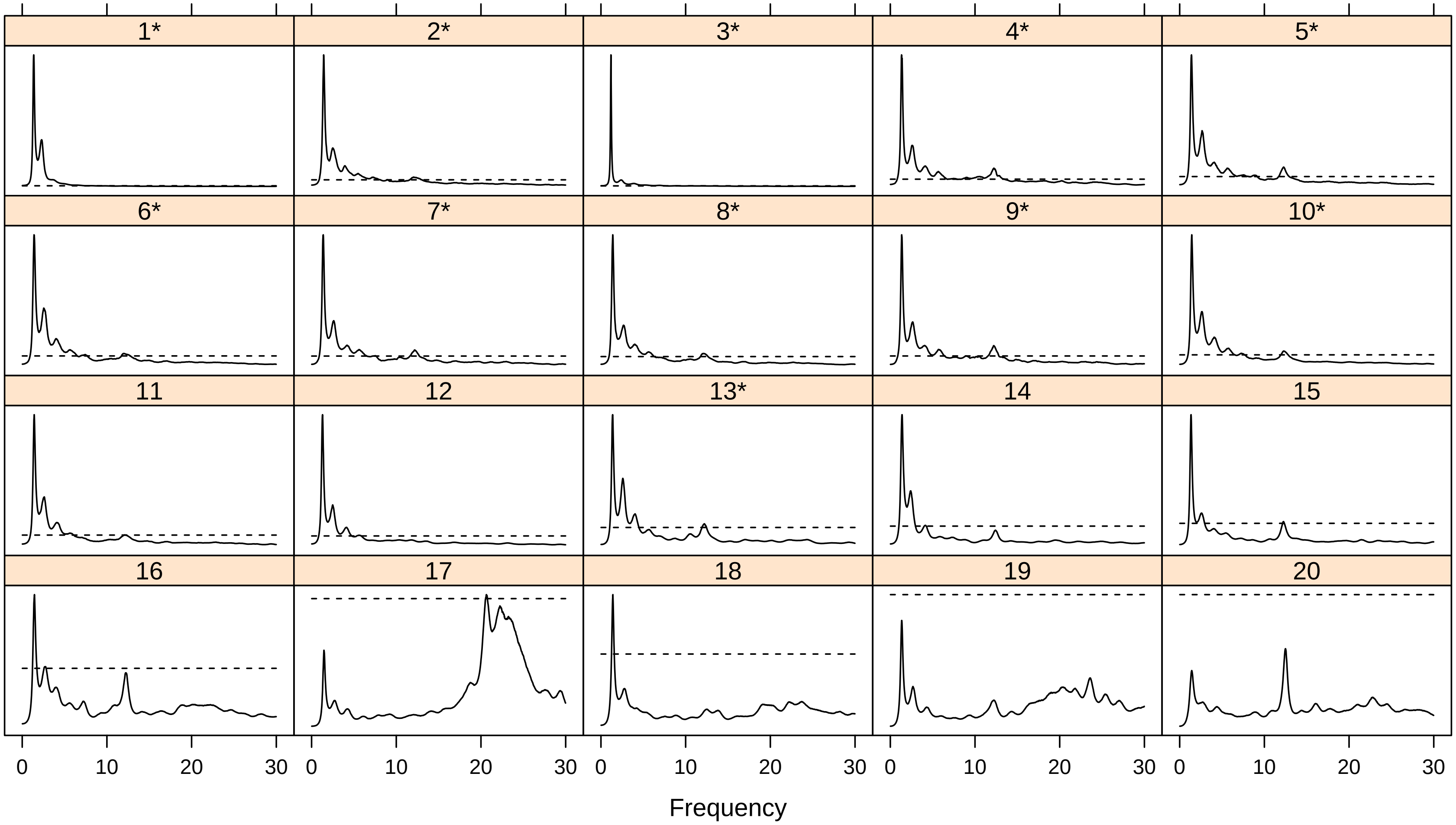


Figure 3.12: Spectrum brain activity hypothesis tests (solid line) for ICs 1-20 from the SVD-CICA bootstrap on the eyes open data with a dotted line at zero indicating the null hypothesis. ICs containing significant alpha activity are labeled with a *.

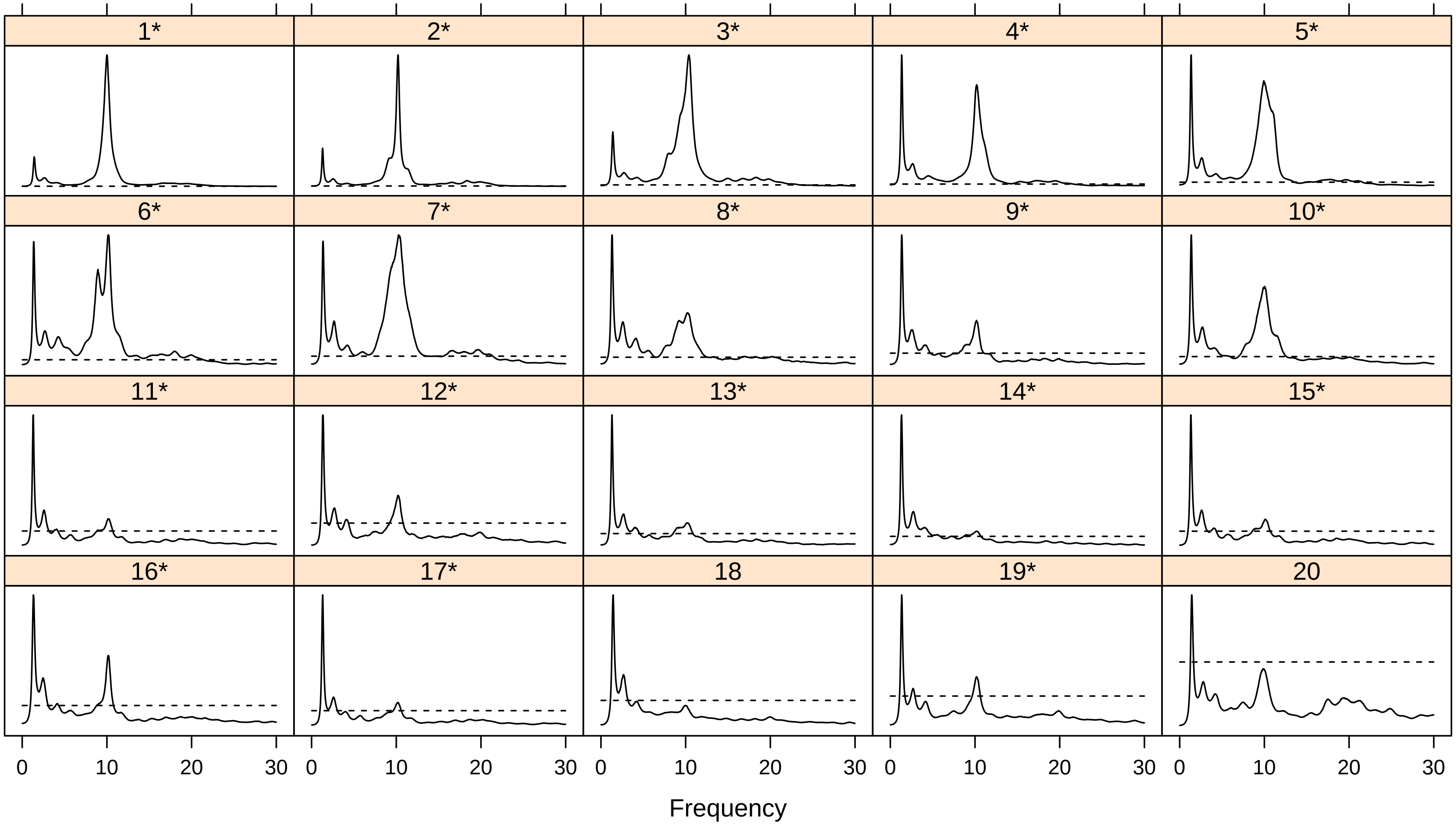


Figure 3.13: Spectrum brain activity hypothesis tests (solid line) for ICs 1-20 from the CICA bootstrap with pre-ICA TBDR on the eyes closed data with a dotted line at zero indicating the null hypothesis. ICs containing significant alpha activity are labeled with a *.

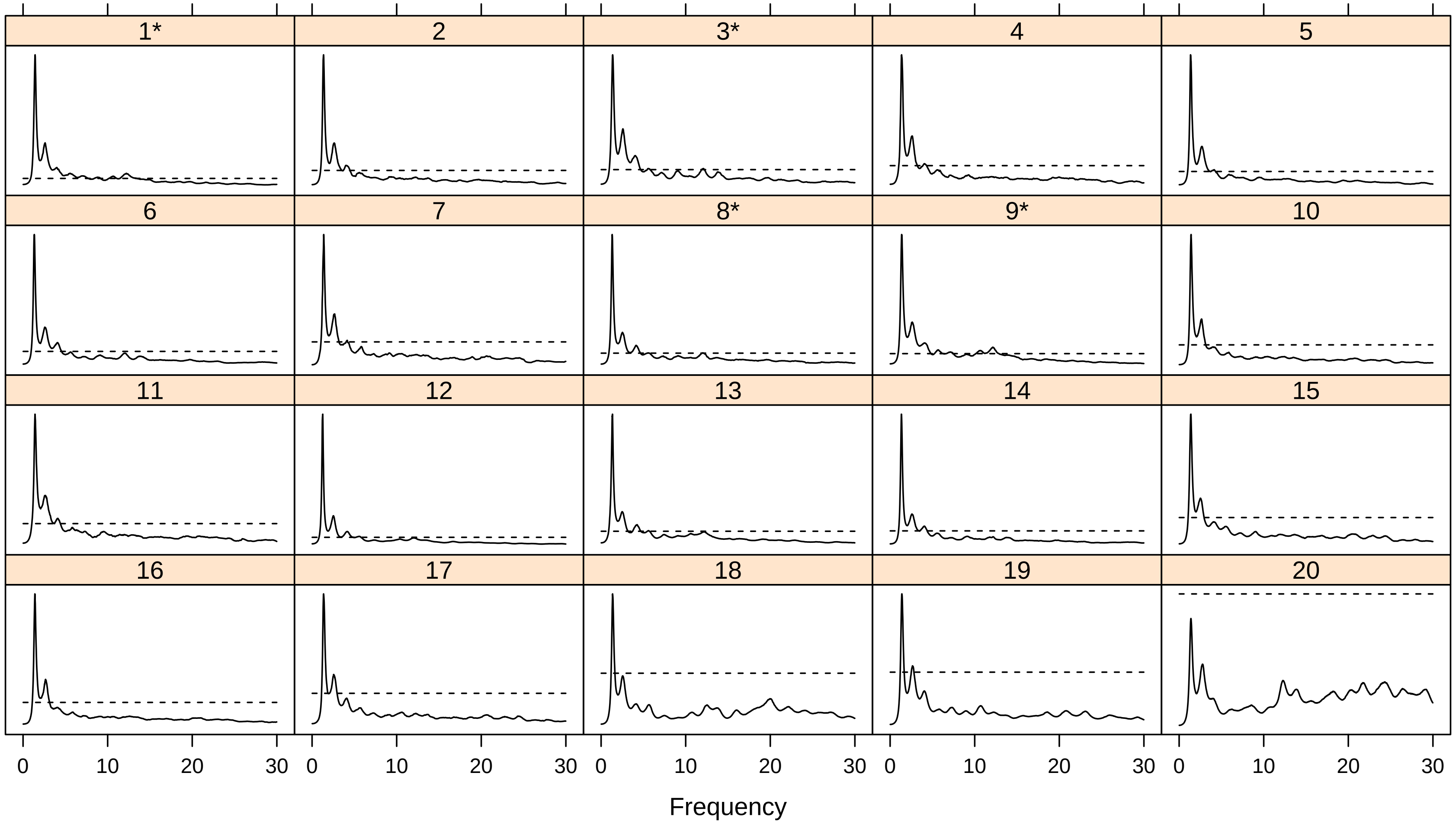


Figure 3.14: Spectrum brain activity hypothesis tests (solid line) for ICs 1-20 from the CICA bootstrap with pre-ICA TBDR on the eyes open data with a dotted line at zero indicating the null hypothesis. ICs containing significant alpha activity are labeled with a *.

CHAPTER 4: A JOINT SPATIAL FACTOR ANALYSIS MODEL TO ACCOMMODATE DATA FROM MISALIGNED NESTED AREAL UNITS WITH APPLICATION TO LOUISIANA SOCIAL VULNERABILITY

4.1 Introduction

In recent years, advances in global positioning system and geographic information system technologies have simplified the process of collecting and analyzing spatially referenced data. The consequent explosion of spatially based research led to an increased capacity to holistically characterize places and communities. The assessment of social indicators across space is a topic that has a long and rich academic history (Duncan, 1974; Smith, 1973, 1981; Taylor and Hudson, 1970), but, recently, the mining of enormous quantities of data on social indicators has propelled this topic beyond the realm of the purely academic (Cutter et al., 2003). One obstacle to the analysis of trends in social indicators across space is the need to accommodate spatially referenced variables collected at differing spatial levels– a problem known as spatial misalignment (Banerjee et al., 2003).

Much of the academic community's recent interest in social indicators has focused on quantifying the "social vulnerability" of places/communities to climate change and natural disasters (Cutter et al., 2003). Because the extent to which a community is able to prepare for and recover from disasters is largely determined by social factors, socially vulnerable areas may be more severely impacted; thus, identification of these areas is critical to disaster preparation (Cutter et al., 2003). Several groups have developed indices of social resilience/vulnerability to natural disasters and environmental changes and have estimated them across the US (Cutter et al., 2003; Cutter and Finch, 2008; Cutter et al., 2008).

Community social vulnerability is not directly measurable, but an abundance of social indicator variables are available, many of which are highly correlated thanks to their common association with this broader concept of social vulnerability. Thus, an index of social vulnerability can be constructed as a latent factor (or set of latent factors) underlying a relevant set of observed social indicator variables, through the use of factor analysis or principal component analysis (Cutter et al., 2003; Cutter and Finch, 2008). The standard factor analysis model, however, is potentially inappropriate because it fails to properly account for the spatial correlation that is likely present in spatially referenced data. We propose instead using spatial factor analysis to create social vulnerability indices, and we extend spatial factor analysis to address the issue of spatially misaligned data.

When the units of study in an analysis are spatially or geographically defined, units that are closer together are likely to be more similar than units further apart, leading to spatial correlation. In this case, spatial factor analysis (Wang and Wall, 2003) should be used instead of the standard factor analysis to appropriately account for the spatial correlation and improve estimation and model fit (Nethery et al., 2015; Wang and Wall, 2003). The spatial factor analysis model takes the form

$$\boldsymbol{Y}(s_i) = \boldsymbol{\Lambda}\boldsymbol{\eta}(s_i) + \boldsymbol{\epsilon}(s_i), \tag{4.27}$$

where $\boldsymbol{Y}(s_i)$ is the $p \times 1$ vector of continuous, observed variables for the i^{th} geographic location/region, denoted s_i, for $i = 1, ..., N$ (Banerjee et al., 2003). Letting m be a prespecified number of latent factors ($m \ll p$), $\boldsymbol{\eta}(s_i)$ is the $m \times 1$ vector of latent factor scores for s_i, $\boldsymbol{\Lambda}$ represents the $p \times m$ matrix of factor loadings, and $\boldsymbol{\epsilon}(s_i) \overset{\text{i.i.d.}}{\sim} \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ represents the vector of errors for s_i. $\boldsymbol{\Sigma}$ is a diagonal matrix with $(i,i)^{th}$ entry equal to σ_i^2, and $\boldsymbol{\Lambda}$ is constrained to be lower triangular with diagonal entries $\lambda_{ii} > 0$ for identifiability purposes (Bollen, 1989). After applying spatial factor analysis to the observed data, the predicted latent factors quantify the latent constructs underlying the observed data, and, in our setting, are used to construct community social vulnerability scores.

Here, we focus on a Bayesian approach. Spatial factor analysis methodology in the Bayesian setting has been extended to accommodate various data types and analysis goals and has been applied to a wide range of problems, including the reduction of social indicator data (Hogan and Tchernis, 2004; Liu et al., 2005; Lopes et al., 2008; Nethery et al., 2015; Mezzetti, 2012; Rowe, 1998; Stakhovych et al., 2012; Wang and Wall, 2003). The standard Bayesian factor analysis prior distribution specifications, which lead to semi-conjugacy, are discussed by Ghosh and Dunson (2009) and Rowe (1998). In the spatial factor analysis model, spatial correlation is introduced in the factor scores through the prior distribution placed on $\boldsymbol{\eta} = [\boldsymbol{\eta}(s_1)' \cdots \boldsymbol{\eta}(s_N)']'$ (Wang and Wall, 2003). The prior distribution has the form $\boldsymbol{\eta} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_m)$, where $\boldsymbol{\Sigma}_S$ is the $N \times N$ spatial covariance matrix, containing a spatial parameter, ϕ (Banerjee et al., 2003; Wang and Wall, 2003).

Although the Bayesian spatial factor analysis model allows for a good deal of flexibility and a wide range of spatial correlation structures, it has not yet been extended to accommodate spatial misalignment, a common obstacle to the analysis of social indicator data. Henceforth, we will consider a single type of spatial data, known as areal data. Areal data are counts or averages of a measure collected over geographically defined regions formed by the partitioning of the area of interest. Areal social indicator data are often spatially misaligned (Cressie, 1996), as indicators may be collected by different organizations over distinct geographic partitions that correspond to each organization's specific goals. Social vulnerability research, due to its reliance on areal social indicator data, may be dramatically impacted by spatial misalignment.

Our objective is to assess the social vulnerability of census tracts. Census tracts are areal units defined by the US Census Bureau that contain approximately 2,500-8,000 residents and are subsets of counties (US Bureau of the Census, 1994). Although many social vulnerability indices are developed at the county level (Cutter et al., 2003; Cutter and Finch, 2008), effects are commonly obscured at this level, particularly for counties that contain large cities, where many communities with different vulnerability levels may be grouped together. Thus, performing inference over smaller regions with more homogeneity in population size, such as census tracts, may provide more insight.

In pursuit of this goal, we wish to analyze a standard set of social indicators, including socioeconomic, demographic, and crime data (Diener and Suh, 1997). The US Census Bureau collects socioeconomic and demographic data for each census tract in the US and makes these data publicly available (US Bureau of the Census, 1994). The US Federal Bureau of Investigation (FBI) makes crime data publicly available for each county in the US through its Uniform Crime Report (UCR) (US Federal Bureau of Investigation, 2010), but crime data are not consistently collected at any finer spatial level. To use both the census tract level socioeconomic and demographic data and the county level crime data together in a spatial factor analysis to create social vulnerability scores for each census tract, it is imperative to address the issue of spatial misalignment.

While there are limited instances of spatial misalignment addressed specifically in the context of spatial factor analysis, the topic of spatial misalignment in general has received a great deal of attention (Gotway and Young, 2002). Some of the general approaches for handling misalignment may be applied in the spatial factor analysis setting. We focus on the case where variables are collected at areal units of two different sizes, which we refer to as small areal units (SAUs) and large areal units (LAUs), with the SAUs fully nested within the LAUs, as in the county/census tract example. One approach to resolving misalignment, which we call pre-analysis data alignment, is to align the data to a common set of areal units prior to application of a statistical model. For inference at the SAU level, data collected at LAUs must be aligned to the SAUs. This can be done by allocating the value at a larger unit into its component smaller units either directly (for variables that are averages or rates), by assigning each SAU the exact value of its LAU, or proportionally to the population/land area in the smaller units (for variables that are counts or sums) (Banerjee et al., 2003). This method imposes the assumption that, for a variable collected at the LAUs, its distribution is the same across all SAUs within a LAU. For some variables, such as crime, this type of constancy would likely be an unreasonable assumption.

In this paper, we propose a joint spatial factor analysis model in the Bayesian setting that can accommodate a set of spatially referenced variables recorded at misaligned nested areal units. The model identifies and predicts a common set of latent factors underlying all the data from two (and possibly more) levels of nested areal units by sharing information between spatial factor analysis models constructed for each spatial level. The model allows prediction of factor scores and inference at the SAU level. In Section 4.2, we introduce the joint model, the necessary assumptions, and the recommended approach to model fitting. We test our method on simulated data in Section 4.3 and compare those results to results from the pre-analysis alignment method. In Section 4.4, we apply the joint model to a set of misaligned social indicator data from the state of Louisiana to create a social vulnerability index, and we combine it with Louisiana flood vulnerability data to identify the highest risk communities in a region prone to natural disasters. Finally, we discuss results and possible extensions in Section 4.5.

4.2 Methods

Let s_i denote LAU $i, i = 1, ..., N$, and s_{ij} denote SAU j nested within LAU $i, j = 1, ..., n_i$. We denote the total number of SAUs as N_T ($N_T = \sum_{i=1}^{N} n_i$). Consider a set of p_1 variables, $\boldsymbol{Y}$, that is recorded at SAUs such that $\boldsymbol{Y}(s_{ij})$ is a $p_1 \times 1$ vector of measurements at s_{ij}. Let $\boldsymbol{Z}$ represent a set of p_2 variables measured at LAUs such that $\boldsymbol{Z}(s_i)$ is a $p_2 \times 1$ vector of measurements at s_i. Our goal is to construct spatial factor analysis models that relate the observed variables at both the SAUs and the LAUs to a common set of spatially correlated latent factors at the SAUs. In other words, we want to write both $\boldsymbol{Y}(s_{ij})$ and $\boldsymbol{Z}(s_i)$ in terms of latent factors $\boldsymbol{\eta}(s_{ij})$. Then these models can be fit jointly.

Of course, we can model the $\boldsymbol{Y}(s_{ij})$ as a function of the $\boldsymbol{\eta}(s_{ij})$ using the standard spatial factor analysis model; however, constructing a model relating the $\boldsymbol{Z}(s_i)$ to the $\boldsymbol{\eta}(s_{ij})$ is less straightforward. To do so, we rely on the assumption that the variables recorded at LAU i ($\boldsymbol{Z}(s_i)$) are weighted sums of the same unobserved variables at all the SAUs contained in LAU i, i.e.,

$$\boldsymbol{Z}(s_i) = \sum_{j=1}^{n_i} w_{ij} \boldsymbol{Z}(s_{ij}) \tag{4.28}$$

for $i = 1, ..., N$, where $\boldsymbol{Z}(s_{ij})$ is a $p_2 \times 1$ vector of the unobserved values of the variables in $\boldsymbol{Z}$ at s_{ij} and the w_{ij} are known, scalar weights. These weights might be assumed to be 1 if $\boldsymbol{Z}$ contains sum/count variables or to be the proportion of the land area or population in s_i that is contained in s_{ij} if $\boldsymbol{Z}$ contains average or rate variables.

Now, if the $\boldsymbol{Z}(s_{ij})$ were observed, we could fit a spatial factor analysis model at the SAU level as in (1):

$$\begin{bmatrix} \boldsymbol{Y}(s_{ij}) \\ \boldsymbol{Z}(s_{ij}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_2 \end{bmatrix} \boldsymbol{\eta}(s_{ij}) + \begin{bmatrix} \boldsymbol{\epsilon}_1(s_{ij}) \\ \boldsymbol{\epsilon}_2(s_{ij}) \end{bmatrix} \tag{4.29}$$

where $\boldsymbol{\Lambda}_1$ is a $p_1 \times m$ matrix of factor loadings corresponding to the variables in $\boldsymbol{Y}$, $\boldsymbol{\Lambda}_2$ is a $p_2 \times m$ matrix of factor loadings corresponding to the variables in $\boldsymbol{Z}$, $\boldsymbol{\eta}(s_{ij})$ is an $m \times 1$ vector of the common factors at location s_{ij} (these factors are correlated across locations), and $\boldsymbol{\epsilon}_1(s_{ij})$ and $\boldsymbol{\epsilon}_2(s_{ij})$ are location-specific error vectors for the variables in $\boldsymbol{Y}$ and $\boldsymbol{Z}$, respectively. It is assumed that $\boldsymbol{\epsilon}_1(s_{ij}) \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{\Sigma}_1)$ and $\boldsymbol{\epsilon}_2(s_{ij}) \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{\Sigma}_2)$.

We now return to the construction of our two models. The top half of equation 4.29 provides the spatial factor analysis model for $\boldsymbol{Y}(s_{ij})$. For $\boldsymbol{Z}(s_i)$, we must use the model specification in equation 4.29 combined with our assumption that $\boldsymbol{Z}(s_i)$ is a weighted sum of the $\boldsymbol{Z}(s_{ij}), j = 1, ..., n_i$ to write

$$\boldsymbol{Z}(s_i) = \sum_{j=1}^{n_i} w_{ij} \boldsymbol{Z}(s_{ij}) = \sum_{j=1}^{n_i} w_{ij} \boldsymbol{\Lambda}_2 \boldsymbol{\eta}(s_{ij}) + \sum_{j=1}^{n_i} w_{ij} \boldsymbol{\epsilon}_2(s_{ij}). \tag{4.30}$$

In this way, $\boldsymbol{Y}(s_{ij})$ and $\boldsymbol{Z}(s_i)$ can be modeled in terms of a common set of latent factors at the SAU level, allowing us to specify our joint model.

Borrowing notation from equation 4.29, the two models can be written separately as

Model 1:

$$\boldsymbol{Y}(s_{ij}) = \boldsymbol{\Lambda}_1 \boldsymbol{\eta}(s_{ij}) + \boldsymbol{\epsilon}_1(s_{ij}) \tag{4.31}$$

Model 2:

$$\boldsymbol{Z}(s_i) = \boldsymbol{\Lambda}_2 \sum_{j=1}^{n_i} w_{ij} \boldsymbol{\eta}(s_{ij}) + \boldsymbol{\epsilon}_2^*(s_i) \tag{4.32}$$

where $\boldsymbol{\epsilon}_2^*(s_i) = \sum_{j=1}^{n_i} w_{ij} \boldsymbol{\epsilon}_2(s_{ij})$ and $\boldsymbol{\epsilon}_2^*(s_i) \sim$ MVN$(\boldsymbol{0}, \boldsymbol{\Sigma}_2^*)$. These models can then be written jointly in the following way:

$$\boldsymbol{X}(s_i) = \boldsymbol{\Lambda}(s_i) \boldsymbol{\eta}(s_i) + \boldsymbol{\epsilon}^*(s_i) \tag{4.33}$$

where $\boldsymbol{\epsilon}^*(s_i) \sim$ MVN$(\boldsymbol{0}, \boldsymbol{\Sigma}(s_i))$ and

$$\begin{gathered}
\boldsymbol{X}(s_i) = \begin{bmatrix} \boldsymbol{Y}(s_{i1}) \\ \boldsymbol{Y}(s_{i2}) \\ \vdots \\ \boldsymbol{Y}(s_{in_i}) \\ \boldsymbol{Z}(s_i) \end{bmatrix}, \quad \boldsymbol{\Lambda}(s_i) = \begin{bmatrix} \boldsymbol{\Lambda}_1 & 0 & \cdots & \cdots & 0 \\ 0 & \boldsymbol{\Lambda}_1 & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \boldsymbol{\Lambda}_1 & 0 \\ w_{i1}\boldsymbol{\Lambda}_2 & w_{i2}\boldsymbol{\Lambda}_2 & \cdots & \cdots & w_{in_i}\boldsymbol{\Lambda}_2 \end{bmatrix}, \\
\boldsymbol{\eta}(s_i) = \begin{bmatrix} \boldsymbol{\eta}(s_{i1}) \\ \vdots \\ \vdots \\ \boldsymbol{\eta}(s_{in_i}) \end{bmatrix}, \quad \boldsymbol{\epsilon}^*(s_i) = \begin{bmatrix} \boldsymbol{\epsilon}_1(s_{i1}) \\ \vdots \\ \vdots \\ \boldsymbol{\epsilon}_1(s_{in_i}) \\ \boldsymbol{\epsilon}_2^*(s_i) \end{bmatrix}, \quad \boldsymbol{\Sigma}(s_i) = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \boldsymbol{\Sigma}_1 & 0 \\ 0 & \cdots & 0 & \boldsymbol{\Sigma}_2 \end{bmatrix}.
\end{gathered} \tag{4.34}$$

To finalize the model structure, the number of latent factors, m, must be specified. Although model selection techniques could be integrated into the estimation procedure, we forego the discussion of such an extension here and assume that the user has specified m in advance. Methods to guide the choice of m are discussed later. Having specified m, estimation can proceed.

Define λ_{1kl} as the element in the (k, l) position of $\boldsymbol{\Lambda}_1$, λ_{2kl} as the element in the (k, l) position of $\boldsymbol{\Lambda}_2$, σ_{1k}^2 as the element in the (k, k) position of $\boldsymbol{\Sigma}_1$, and σ_{2k}^2 as the element in the (k, k) position of $\mathbf{\Sigma}_2$. As in the classic spatial factor analysis model, this model allows for spatial correlation in the factor scores through a spatial covariance matrix in the prior distribution for the factors, with spatial parameter ϕ embedded in this matrix. Using this notation, we provide the data likelihood and recommended prior distributions for Bayesian estimation in Appendix B, Section B.1 and Section B.2, respectively. Markov chain Monte Carlo (MCMC) samples are drawn from the full conditional distribution of each parameter using a Gibbs sampler with a Metropolis step (Geman and Geman, 1984; Gelfand and Smith, 1990; Hastings, 1970; Metropolis et al., 1953). The full conditional distributions and sampling algorithm corresponding to the recommended prior distributions are provided in Appendix B, Section B.3. Finally, posterior means and credible intervals for each parameter can be computed from the MCMC output.

4.3 Simulation Studies

4.3.1 Assessment of Joint Model Performance

Simulations are carried out using R (R Core Team, 2014) and MATLAB (The MathWorks Inc., 2015) statistical software. Misaligned areal data are simulated from a lattice with 125 LAUs and 625 nested SAUs, with each LAU containing exactly 5 SAUs, as shown in Figure 4.15 ($N = 125$, $n_i = 5$ for all i, $N_T = 625$). We construct this lattice so that the simulated data reflect anticipated real data, with many fewer LAUs than SAUs. A single latent factor is simulated over the SAUs from a normal distribution, with spatial correlation introduced by assigning a covariance of 0.15 to the factor scores of any pairs of SAUs sharing a boundary ($m = 1$, $\phi = -0.15$).

The latent factor and fixed loadings, along with randomly generated error terms from normal distributions with fixed variances (Table 4.9), are used to generate six observed variables at the SAUs ($p_1 = 6$), by plugging into equation 4.31. Six distinct observed variables are generated at the LAUs ($p2 = 6$) by plugging the latent factor, fixed loadings, randomly generated errors from normal distributions, and a common set of weights, $w_i = \{.2, .15, .3, .1, .25\}$ into equation 4.32 (Table 4.9). This process is repeated 1,000 times to generate 1,000 simulated misaligned datasets. We apply the Gibbs sampler described above, with hyperparameters chosen to create non-informative prior distributions, to each dataset for 100,000 iterations. We discard the first 50,000 samples as burn-in and retain 50,000.

Our primary interest is in accurate prediction of the latent factor scores, as these will form the social vulnerability index. Across the 1,000 simulations, the average correlation between the true and predicted factor scores is 0.99, and the average coverage rate of the 95% credible interval for each factor score across all simulations is 0.93, indicating that this model performs extremely well for our purposes. Average parameter estimates and 95% credible interval coverage rates are provided in Table 4.9. All parameters are estimated consistently with credible interval coverage very close to the desired 95%, with the exception of the parameters impacted by the identifiability constraint, i.e. λ_{11}, λ_{21}, σ^2_{11}, σ^2_{21}, whose estimation suffer somewhat due to this constraint. However, extensive investigation into this issue has revealed that the sub-optimal performance of these parameter estimates has little, if any, impact on the performance of the other parameter estimates and the factor score predictions. While the effects appear to be minor, it may be wise to order the variables such that the first variable at each spatial level is one not believed to carry great importance or to run the model with different variable orderings to ensure results are not highly impacted.

This joint model works by sharing information from both the SAU variables and the LAU variables to predict the values of the common latent factor(s). The unique parameters for each model are estimated using the data corresponding to that model as well as these predicted factor values; thus, the parameter estimation also incorporates information from both the SAU and LAU data. In order to justify prediction of the latent factor(s) at the SAU level, intuition suggests that some data from the SAUs are needed to provide information about characteristics at the SAU level. Although the development of theoretical results for this model will be left for future work, we note that, in accordance with existing factor analysis theory, we need $p_1 \geq m$ to predict latent factors at the SAU level (Ghosh and Dunson, 2009).

Then, a pertinent question for users of this method might be related to model performance when the number of SAU variables is small. To test this performance, we conduct 1,000 simulations with misaligned data constructed using the same procedure as described above, except with $p_1 = 3$ rather than $p_1 = 6$. While the model identifies the correct factor and performs comparably to the previous simulation in parameter estimation, precision in the latent factor prediction is reduced, as expected when less information is available at the SAU level. For instance, average correlation between the true latent factor and the predicted latent factor declines from 0.99 in the previous simulation to 0.73. Thus, we believe this joint model will be most advantageous when used on data involving many SAU variables. Similar results to those presented in this section are observed in simulations with two latent factors as well ($m = 2$).

4.3.2 Comparison of the Joint Model with Pre-Analysis Alignment

We also want to compare our joint model to the simple method of pre-analysis data alignment described in Section 4.1. To the best of our knowledge, alignment methods like this one would typically be the only viable competitors to our model that would allow for factor scores to be predicted and inference performed at the level of the SAUs. We implement pre-analysis alignment on the first 1,000 simulated datasets described above ($p_1 = 6$). To do so, for each variable collected at LAUs, we assign the value of each LAU to all of its nested SAUs, and we use these constructed variables at the SAUs together with the variables originally collected at the SAUs in a standard spatial factor model. A spatial factor model with the number of factors correctly specified and prior distributions analogous to those used in the joint model is applied to each aligned dataset using a Gibbs sampler with a Metropolis step. The sampler is run for 100,000 iterations with the first 50,000 samples discarded as burn-in.

When $m = 1$, the pre-analysis alignment model is able to identify the factor underlying the simulated data by relying on the information contained in the SAU data (average correlation of 0.99 between the true and predicted factor across the simulations); however, the estimated factor loadings matrices demonstrate that the information contained in the LAU data has little to no impact on the factor score predictions. The average posterior means of the LAU factor loadings are $\Lambda_2 = [0.09 \quad 0.13 \quad 0.11 \quad -0.05 \quad 0.27 \quad 0.07]'$. The very small estimates of these loadings compared to their true values (Table 4.9) show that the LAU variables, likely distorted through the alignment procedure, no longer contribute appropriately to the factor score predictions, rendering this approach ineffective for integrating information from misaligned data for factor score prediction.

The superiority of the joint model over pre-analysis alignment is even more evident when a multi-factor simulation is assessed. Thus, we simulate 10 misaligned datasets using the same lattice described above, now with $m = 2$. We again fit both the joint model and the pre-analysis alignment model to each dataset. Because the factors can be estimated in a different order in each simulation, we investigate factor score predictive performance using the maximum of the pairwise correlations between each of the predicted factors and each of the true factors. For the joint model, the average maximum correlation across the 10 simulations was 0.80 and 0.81 for factors one and two, respectively. For the pre-analysis model, the average maximum correlation across the 10 simulations was 0.70 and 0.24 for factors one and two, respectively. This demonstrates that, while the pre-analysis alignment model can again use the SAU variables to identify one factor reasonably well, the distortion of the LAU variables misguides the model in the identification of the second factor.

4.3.3 Choosing the Number of Latent Factors

Finally, a well-known and highly contentious issue in factor analysis is how to choose m, the number of latent factors to model. This question may become even more complicated in the spatial misalignment setting, when variables at the LAUs and variables at the SAUs do not have identical dimensions. Popular criteria for choosing m in a standard factor analysis include the eigenvalue criterion and/or scree tests as described by (Ledesma and Valero-Mora, 2007); however, these cannot be applied directly to the misaligned data due to dimensional discrepancies. We simulate a single set of misaligned data with two latent factors using the procedure described in Section 4.3.2. We then pre-align these simulated data at the LAU level by taking weighted sums of the SAU data over the corresponding LAUs, and we test these methods on the LAU aligned data. After applying this procedure to accommodate the misalignment, both the eigenvalue criterion and the scree test identified the correct number of factors for our simulated dataset ($m = 2$) (Figure 4.16).

Model selection criteria, such as the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) which is appropriate in Bayesian settings, have also been used to select the number of factors (Nethery et al., 2015; Wang and Wall, 2003), though in some cases this method may be less desirable than those previously identified due to the computation time required for fitting multiple models. The DIC, an adaptation of the frequentist model selection criteria Akaike Information Criteria (AIC) (Akaike, 1998), does not require specification of the number of parameters in the model, making it particularly well-suited for Bayesian and hierarchical models where this number is often ambiguous. As with the AIC, the DIC is a relative measure that only has value for comparing models, which can either be nested or non-nested, with smaller DIC values indicating better fit. We fit the joint model to the simulated $m = 2$ data described above three times, once correctly specifying m ($m = 2$) and twice misspecifying it ($m = 1$, $m = 3$). DIC also correctly chooses the model with $m = 2$, with DIC values of 5172327.51, 5171898.41, and 5173833.12 for $m = 1$, $m = 2$, and $m = 3$, respectively.

Thus, while the best method of choosing the number of factors remains a contentious topic in the literature, this simulation suggests that standard methods may perform well when applied in the misaligned data setting.

4.4 Application to Louisiana Social Vulnerability

Since the US Gulf Coast Region is particularly susceptible to environmental disasters, which could be compounded by climate change, there is a great deal of interest in developing indices of social vulnerability for this region in order to identify high risk communities and implement measures to reduce the impact of future disasters (Oxfam America Inc, 2009). Oxfam America Inc (2009) has analyzed the Social Vulnerability Index created by Cutter et al. (2003) in the Gulf Coast Region and has integrated it with an indicator of susceptibility to climate hazards. The US National Oceanic and Atmospheric Administration, in combination with other institutions, is developing a Coastal Resilience Index to measure the social vulnerability of disaster-prone coastal communities in the region (Simpier et al., 2010). The US Centers for Disease Control and Prevention (CDC) have also constructed a Social Vulnerability Index, which ranks the social vulnerability of the census tracts in each state to disaster and disease and can be used to identify the most vulnerable communities in each state in the Gulf Coast Region (Agency for Toxic Substances and Disease Registry, 2014). However, none of these indices take into account crime, which is an important component of social vulnerability to environmental hazards (Perdikaris, 2014). In this section, we present an example of how our joint spatial factor analysis model can be applied to misaligned socioeconomic, demographic, and crime data from Louisiana to create a social vulnerability score for each census tract in the state.

Louisiana socioeconomic and demographic data, collected by the US Census Bureau, and crime data, collected by the US FBI, were accessed using Social Explorer (US Bureau of the Census, 2016; US Federal Bureau of Investigation, 2016). Socioeconomic and demographic measures come from the 2008-2012 American Community Survey (US Bureau of the Census, 2013) and are obtained at the census tract level. The crime data being used are from the FBI's UCR in the year 2010 and are obtained at the county level (counties are referred to as parishes in Louisiana), as they are unavailable at the census tract level, as described in Section 1. The names and descriptions of all variables being included in the factor analysis appear in Table 4.10.

Census tracts for which some or all variables are missing are excluded from the analysis. Crime data are not missing for any parish. The final dataset contains a total of 1,110 census tracts and 64 parishes ($N_T = 1,110$, $N = 64$). Adhering to common practice in factor analysis, each variable is centered and scaled before applying the model (Ghosh and Dunson, 2009; Nethery et al., 2015). Hyperparameters are chosen to create vague prior distributions, and we let $m = 1$ in order to construct a single measure of vulnerability. Finally, the MCMC sampler for the joint model is run for 100,000 iterations, discarding the first 90,000 iterations as burn-in. Larger burn-ins are often required in the analysis of real data as compared to simulated data, as real data typically exhibit more messiness, causing samplers to converge more slowly. Model convergence was assessed through traceplots of the remaining 10,000 samples and found to be acceptable. Posterior means are used as estimates for all parameters.

Figure 4.17a provides a map of the predicted latent factor scores for each census tract in Louisiana, with zooming for only Orleans Parish, Louisiana, which contains much of the city of New Orleans, in Figure 4.17b. Census tracts mapped in white were eliminated from the analysis due to missing data. The pattern in the factor loadings, shown in Figure 4.17c, suggests that census tracts with higher predicted latent factor scores may, indeed, have higher levels of social vulnerability.

As further evidence that the latent factor from the joint model, which we call the joint model index, measures the social vulnerability of the Louisiana census tracts, we have plotted it against the CDC's social vulnerability rankings for the census tracts in Louisiana from 2014 in Figure 4.18. This plot demonstrates that the joint model index is correlated with a respected social vulnerability index (correlation of 0.63), but it also illustrates the effect of including crime in the scores. The ordinary least squares (OLS) regression line from the regression of the joint model index on the CDC's social vulnerability ranking is included on the plot. Census tracts from counties with high crime rates, defined as having both violent and property crime rates above the 75th percentile for the state, largely fall above the OLS regression line, indicating that these tracts are typically assigned relatively higher vulnerability scores from our index compared to the CDC's index. Thus, as we hoped, evidence suggests that our index appropriately integrates information about crime and gives higher scores to communities with higher crime rates when compared with existing social vulnerability indices which fail to account for crime. This application of the joint model demonstrates how it can be used to incorporate data at different spatial scales to improve on existing social vulnerability indices.

Given that these results suggest that the joint model index constitutes an index of social vulnerability for Louisiana, we now provide an example of how it can be integrated with historic natural disaster data for Louisiana to identify the communities that are at high risk geographically for natural disasters and are highly socially vulnerable, which exacerbates the impacts of such disasters. During future natural disasters, this type of information can be consulted to help allocate resources in a way that alleviates the burden on the highest risk communities. We choose to focus on geographic vulnerability to flooding, because Louisiana has been historically hard hit by tropical storms and floods. To measure flood vulnerability, we employ data from the Federal Emergency Management Agency (FEMA). FEMA offers low-cost flood insurance to property owners nationwide through its National Flood Insurance Program (NFIP), and it makes historic policy and claims data available for each county in the US, summarized January 1, 1978-January 30, 2017 (Federal Emergency Management Agency, 2017a).

As a proxy for flood vulnerability, we investigate the rate of losses closed with payment (per 100,000 residents) from NFIP for each county in Louisiana. This variable is chosen because a loss closed with payment indicates that flood damages to property were confirmed by multiple sources– both the property owner and the insurance assessor– making it a more reliable measure of flood vulnerability than other NFIP statistics such as rate of policies or rate of claims made (Federal Emergency Management Agency, 2017b). Finally, using thresholds corresponding to the 75^{th} percentile of both the flood vulnerability and social vulnerability measures for Louisiana, we create binary classifiers of flood risk and social vulnerability, i.e., any county above the 75^{th} percentile of losses closed with payment is considered high flood risk and any census tract above the 75^{th} percentile of social vulnerability is considered high social vulnerability.

The interaction between these classifiers is mapped across Louisiana, with zooming for Orleans Parish (Figure 4.19). This indicator shows that the city neighborhoods in New Orleans are both highly socially vulnerable and also highly geographically vulnerable to flooding. This is not a surprising finding given the extent of the flooding damages in New Orleans following Hurricane Katrina in 2005, which exacerbated the already high social vulnerability in many of these city neighborhoods.

4.5 Discussion

Given the vast amount of spatially referenced data now available, the issue of using these data to provide meaningful and concise characterizations of places and communities is of great interest to researchers and policy makers alike. However, these data may be spatially misaligned. For this reason, we have developed a joint spatial factor analysis model to handle data from misaligned nested areal units, which can be used to produce a common set of latent factors underlying two (or more) nested spatial levels. The ability of this model to provide results and inference at the smallest spatial level is critical, as this prevents loss of information and potential obscuring of results.

In Section 4.3, we demonstrated the effectiveness of the model and its superiority over naive methods for dealing with spatial misalignment in factor analysis. We also made recommendations, based on our observations in simulations, that this joint model be applied to misaligned data containing a reasonably large number of SAU variables and that classic methods be used to determine the number of latent factors to model. Finally, the model was applied to misaligned social indicator data from Louisiana to develop social vulnerability scores for each census tract in the state. We provided evidence that the joint model produces a social vulnerability index for Louisiana that improves on existing indices by incorporating information from different spatial levels while yielding high spatial resolution results.

When integrated with information about past or future vulnerability to environmental disasers, our social vulnerability index can be used to help policy makers and disaster responders identify communities likely to need the most assistance in future disasters, as we demonstrated through a joint assessment of flood vulnerability and social vulnerability in Louisiana. Future work could combine our index with climate change disaster predictions to prepare for impacts on the population in the high risk US Gulf Coast Region. Data from the Gulf Long Term Follow-Up Study (Kwok et al., 2017), a study tracking the long term health of workers on the 2010 Deepwater Horizon oil spill, when combined with our social vulnerability index, provide an opportunity to investigate whether people living in socially vulnerable communities were differentially impacted by the oil spill.

We have demonstrated the use of this model in a relatively simple form. However, methodological extensions could be implemented to accommodate additional data problems and allow for greater model flexibility. For instance, three or more models could be fit jointly to allow for more nested spatial levels. When $m > 1$, separate spatial parameters can be specified for each latent factor in order to allow the latent factors to contain different degrees of spatial correlation. This method could also be integrated with a model-based approach to choosing m, such as the reversible jump MCMC method of Lopes and West (2004). Although non-nested misaligned data cannot be accommodated by this model in its current form, future work will investigate a similar approach for dealing with non-nested misaligned spatial data in factor analysis.

The applications of this method extend well beyond the social vulnerability application emphasized here. For example, environmental toxicant and pollutant data are often measured at different spatial levels. A model like the one presented here could be used to reduce misaligned toxicant and pollutant data to develop environmental exposure scores across a region of interest.

4.6 Tables and Figures

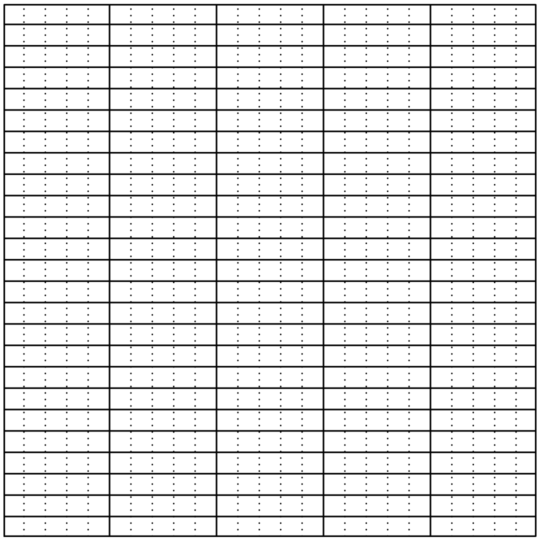


Figure 4.15: Lattice for simulations with large areal units bounded by solid lines ($N = 125$) and nested small areal units bounded by dotted and solid lines ($n_i = 5$).

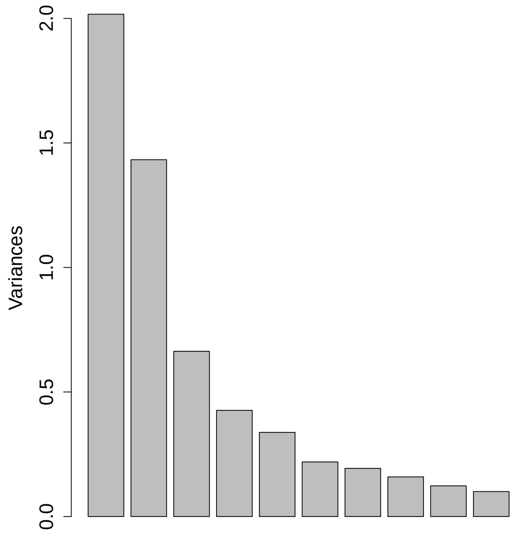


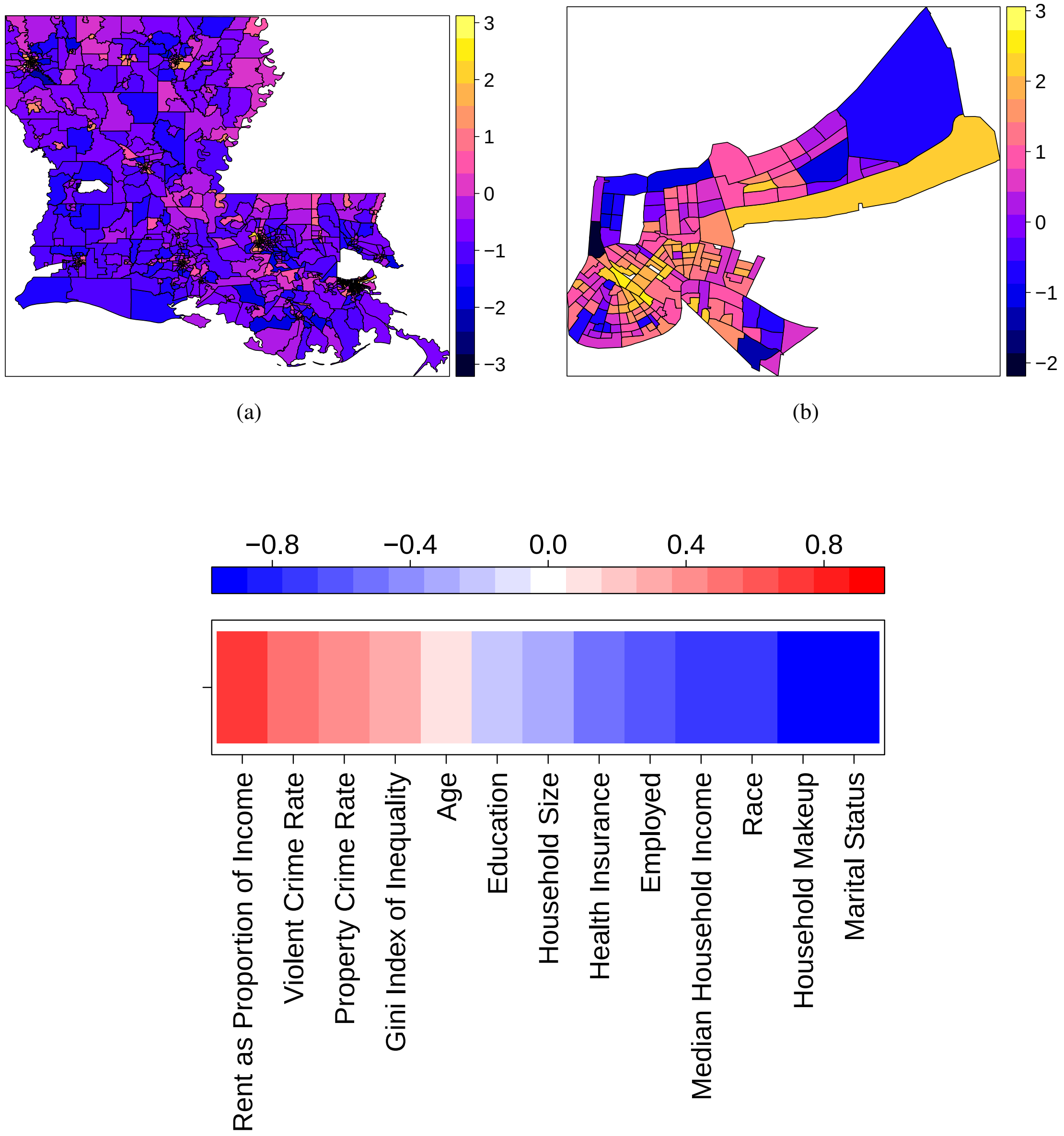
Figure 4.16: Scree plot for simulated data with the small areal units pooled over the large areal units.

Table 4.9: True values for all parameters in the misaligned data simulation and average posterior means (PM) and 95% credible interval coverage rates from the joint model across 1,000 simulations.

	Truth	Average PM	Coverage
λ_{111}	0.50	0.40	0.78
λ_{121}	0.14	0.14	0.95
λ_{131}	-0.63	-0.63	1.00
λ_{141}	1.20	1.21	0.99
λ_{151}	0.25	0.25	0.95
λ_{161}	-0.62	-0.62	0.98
λ_{211}	0.37	0.30	0.74
λ_{221}	0.55	0.56	0.94
λ_{231}	0.43	0.44	0.96
λ_{241}	-0.23	-0.23	0.95
λ_{251}	1.13	1.16	0.95
λ_{261}	0.29	0.30	0.95
σ^2_{11}	0.27	0.32	0.75
σ^2_{12}	0.39	0.39	0.94
σ^2_{13}	0.01	0.01	0.96
σ^2_{14}	0.38	0.39	0.95
σ^2_{15}	0.87	0.87	0.95
σ^2_{16}	0.34	0.34	0.95
σ^2_{21}	0.48	0.50	0.94
σ^2_{22}	0.60	0.61	0.95
σ^2_{23}	0.49	0.50	0.95
σ^2_{24}	0.19	0.19	0.94
σ^2_{25}	0.83	0.84	0.94
σ^2_{26}	0.67	0.68	0.94
ϕ	-0.15	-0.11	0.96

Table 4.10: Variables included in the joint spatial factor analysis model.

Variable Name	Definition
Census Tract Level	
Age	Proportion of the population under 65 years old
Race	Proportion of the population caucasian/white
Education	Proportion of the population 25 or older with a bachelor's degree or higher
Marital Status	Proportion of the population 15 or older married
Employed	Proportion of the civilian population 16 or older in the labor force employed
Median Household Income	Median household income in 2012 inflation adjusted dollars
Rent as Proportion of Income	Median gross rent in 2012 inflation adjusted dollars as a proportion of median household income in 2012 inflation adjusted dollars
Household Size	Average household size
Health Insurance	Proportion of the population with health insurance
Gini Index of Inequality	Measure of income inequality developed by Gini (1912)
Household Makeup	Proportion of the population living in married family households
County Level	
Violent Crime Rate	Rate of violent crimes reported per 100,000 population (violent crimes include murders, rapes, robberies, and aggravated assaults)
Property Crime Rate	Rate of property crimes reported per 100,000 population (property crimes include burglaries, larcenies, and motor vehicle thefts)



(a)

(b)

(c)

Figure 4.17: Predicted factor scores for each census tract from the joint model mapped across all of Louisiana (a) and Orleans Parish only (b) and a heat map of the estimated factor loadings (c).

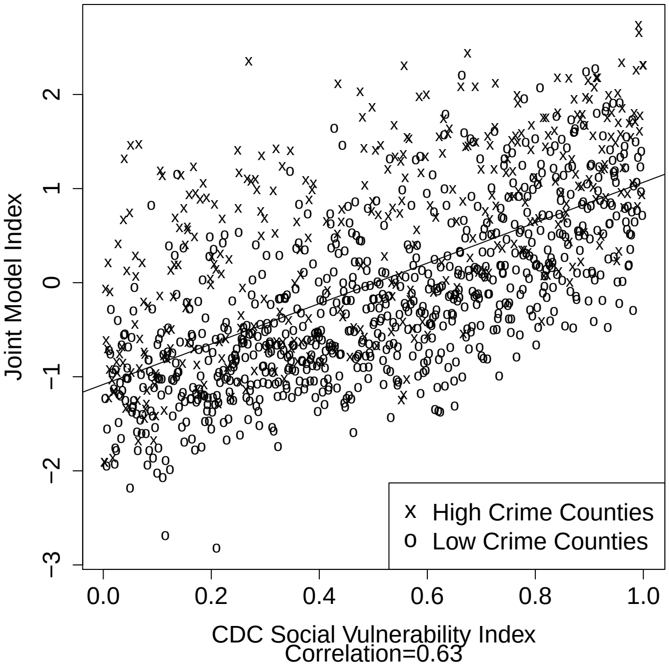
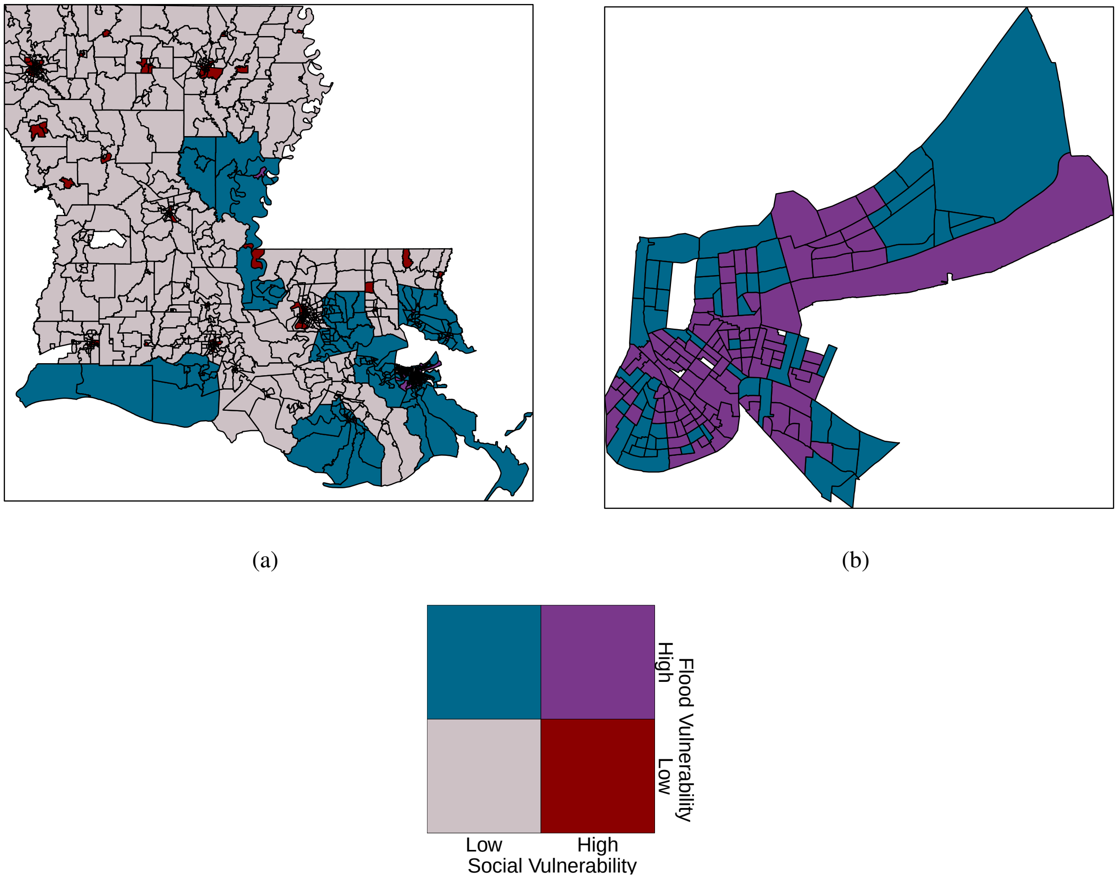


Figure 4.18: The index created by the joint model plotted against the US Centers for Disease Control and Prevention's Social Vulnerability Index for Louisiana census tracts. Points plotted as 'x' represent census tracts from counties with both violent and property crime rates above the 75th percentile for the state. All other census tracts are plotted with an 'o'. The ordinary least squares regression line from the regression of the joint model index on the CDC's social vulnerability index is included.



(a)

(b)

Figure 4.19: The interaction between flood vulnerability and social vulnerability classifiers mapped across all of Louisiana (a) and Orleans Parish only (b).

APPENDIX A: SUPPLEMENTARY MATERIALS FOR CHAPTER 3

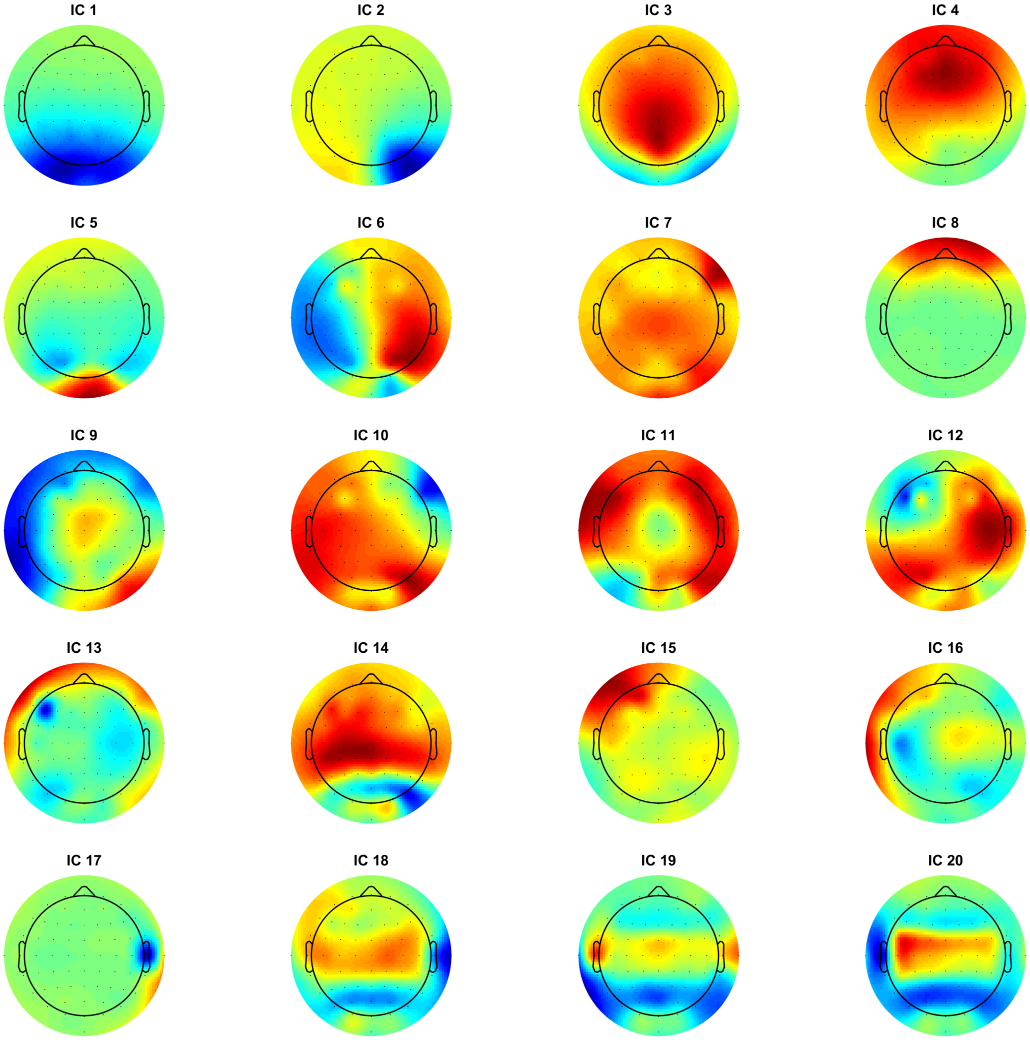


Figure 20: Topographical maps for each IC from PBDR on eyes closed data.

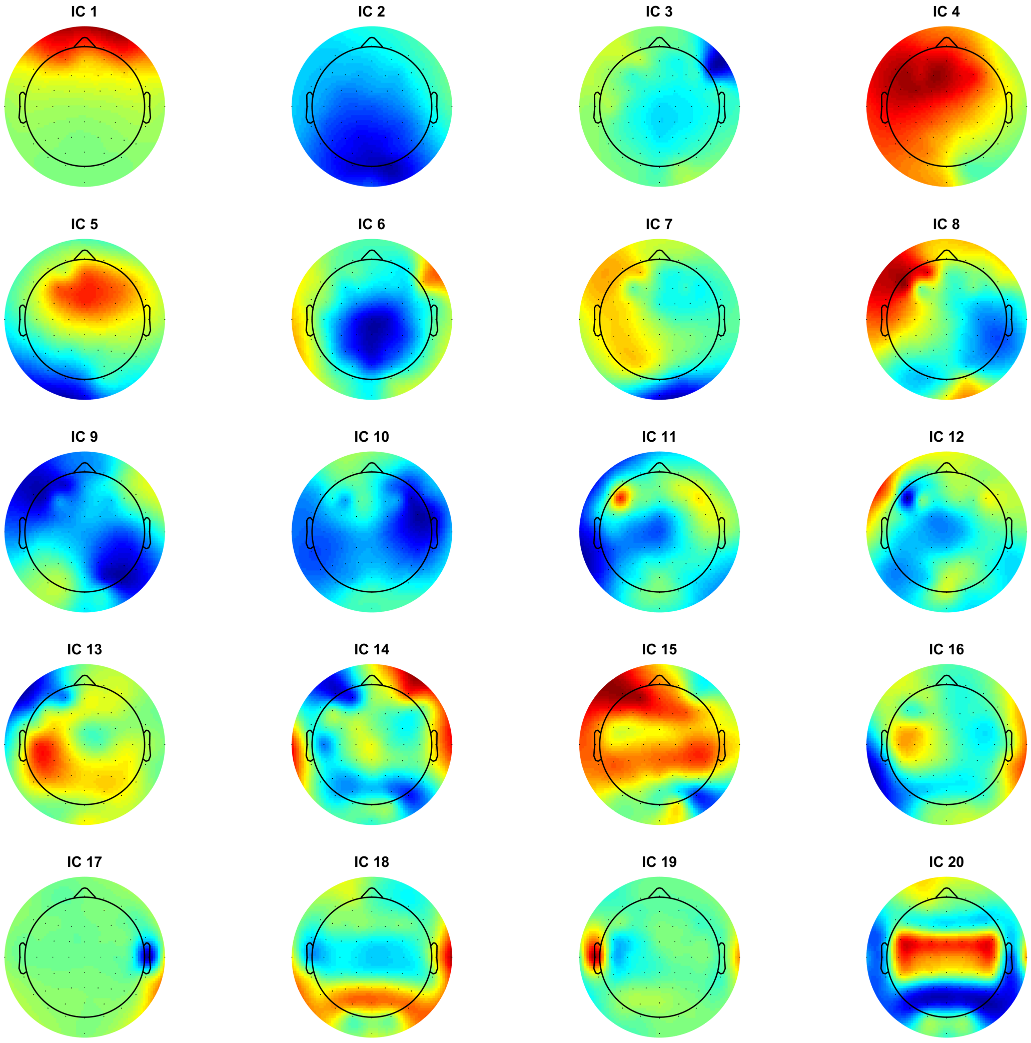


Figure 21: Topographical maps for each IC from PBDR on eyes open data.

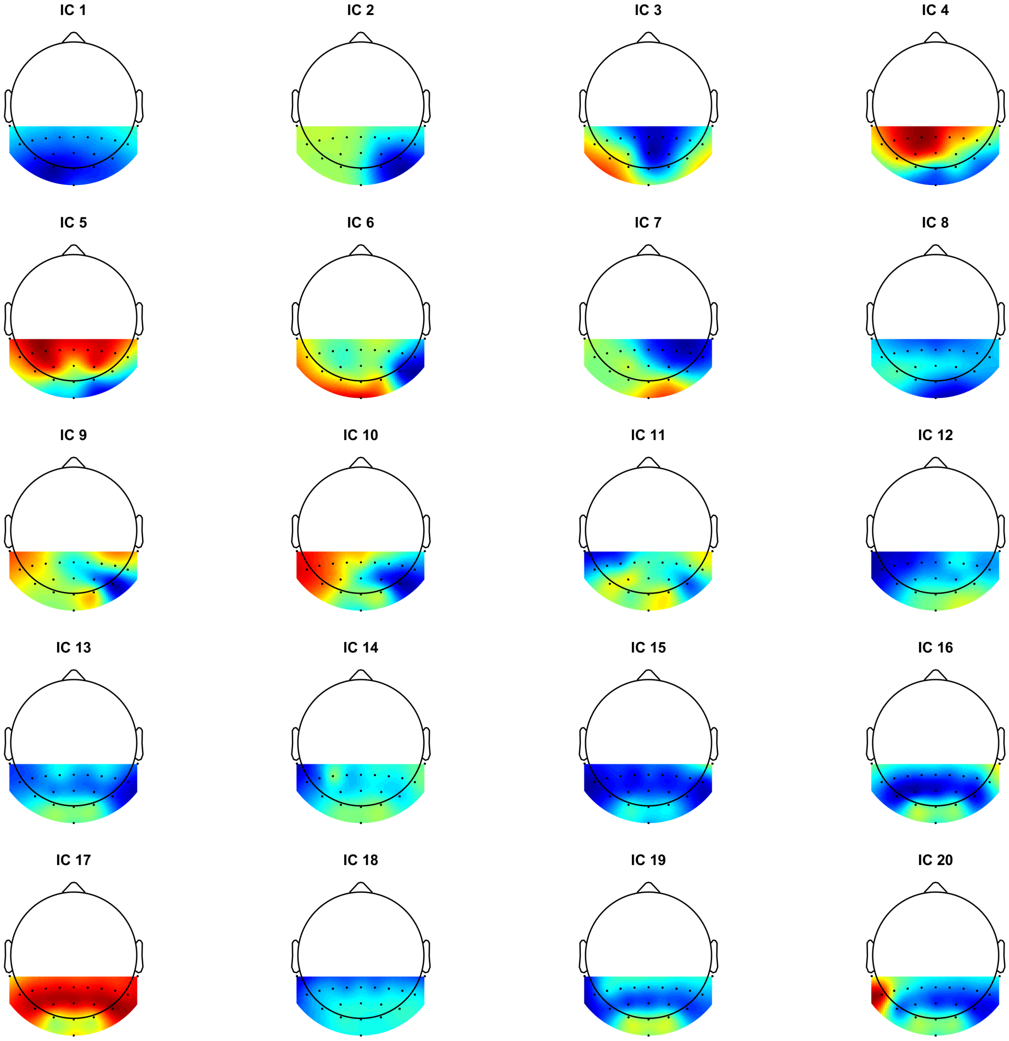


Figure 22: Topographical maps for each IC from TBDR on eyes closed data.

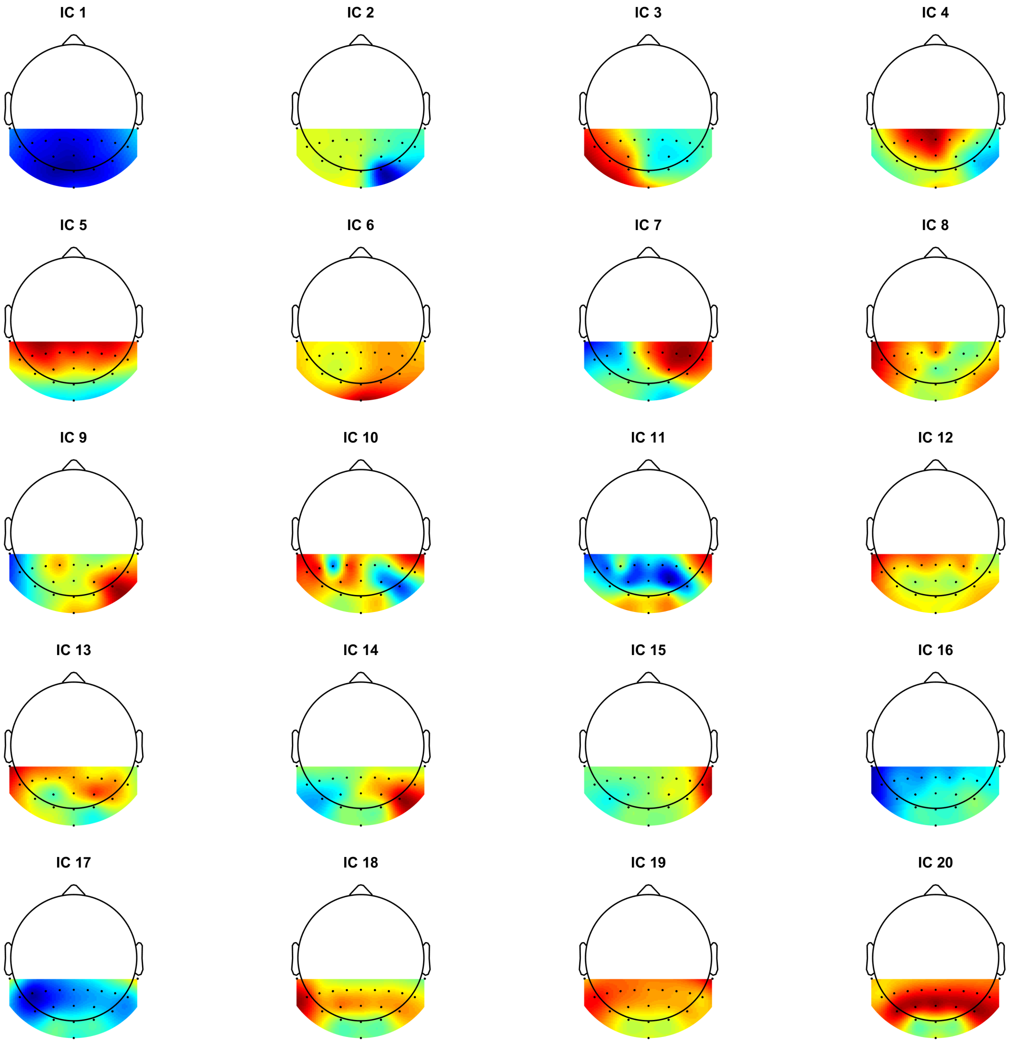


Figure 23: Topographical maps for each IC from TBDR on eyes open data.

APPENDIX B: SUPPLEMENTARY MATERIALS FOR CHAPTER 4

B.1 Data Likelihood

The data likelihood can be written as $\boldsymbol{X}(s_i)|\boldsymbol{\Lambda}(s_i), \boldsymbol{\eta}(s_i), \boldsymbol{\Sigma}(s_i) \sim \text{MVN}(\boldsymbol{\Lambda}(s_i)\boldsymbol{\eta}(s_i), \boldsymbol{\Sigma}(s_i))$ or in vector form, to facilitate modeling of the spatial correlation, as $\boldsymbol{X}|\boldsymbol{\Lambda}, \boldsymbol{\eta}, \boldsymbol{\Omega} \sim \text{MVN}(\boldsymbol{\Lambda\eta}, \boldsymbol{\Omega})$ where

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}(s_1) \\ \vdots \\ \boldsymbol{X}(s_N) \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}(s_1) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{\Lambda}(s_N) \end{bmatrix},$$

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}(s_1) \\ \vdots \\ \boldsymbol{\eta}(s_N) \end{bmatrix}, \quad \boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma}(s_1) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{\Sigma}(s_N) \end{bmatrix}.$$

B.2 Prior Distributions

The prior structure for this model is informed by the default, semi-conjugate factor analysis prior structure introduced by Ghosh and Dunson (2009) and extended to spatial factor analysis by Wang and Wall (2003). As in the standard Bayesian spatial factor analysis model, an identifiability constraint of lower triangularity must be placed on the factor loadings matrices, so we force $\lambda_{1kl} = 0$ for $k < l$ and $\lambda_{2kl} = 0$ for $k < l$. The prior for each of the remaining model parameters is given below.

$\lambda_{1kk} \sim \text{TN}(0, \tau_1^2), \quad k = 1, ..., p1$

$\lambda_{1kl} \sim \text{N}(0, \tau_1^2), \quad k = 1, ..., p1, \quad l = 1, ..., m$

$\lambda_{2kk} \sim \text{TN}(0, \tau_2^2), \quad k = 1, ..., p2$

$\lambda_{2kl} \sim \text{N}(0, \tau_2^2), \quad k = 1, ..., p2 \quad l = 1, ..., m$

$\sigma_{1k}^2 \sim \text{IG}(\alpha_1, \beta_1), \quad k = 1, ..., p1$

$\sigma_{2k}^2 \sim \text{IG}(\alpha_2, \beta_2), \quad k = 1, ..., p2$

$\boldsymbol{\eta} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_s \otimes \boldsymbol{I}_m), \quad \boldsymbol{\Sigma}_s = \boldsymbol{I}_N - \phi \boldsymbol{R}, \quad$ (**R** is an adjacency matrix)

$\phi \sim \text{Unif}(a, b)$

Note that $\tau_1^2, \tau_2^2, \alpha_1, \beta_1, \alpha_2$, and β_2 are hyper-parameters to be selected based on prior knowledge or, more likely, selected to make the prior distributions vague and flat in order to give equal weight to a wide range of values. a and b should be chosen to be the inverses of the minimum and maximum eigenvalues of $\boldsymbol{R}$, respectively, to ensure positive definiteness of the covariance matrix of $\boldsymbol{\eta}$ (Hogan and Tchernis, 2004). If $m > 1$, a unique spatial parameter can be specified for each factor to allow for different amounts of spatial correlation in each factor, which will increase the flexibility and robustness of the model.

B.3 Sampling Algorithm and Full Conditional Distributions

(1) Sample $\lambda_{1kk} | \boldsymbol{\Lambda}_1(-k,-k), \boldsymbol{\Lambda}_2, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}$ from

$$\text{TN}\left(\frac{\tau_1^2 \sum_i \sum_j \boldsymbol{\gamma}_{kk}(s_{ij}) \boldsymbol{\eta}_k(s_{ij})}{\tau_1^2 \sum_i \sum_j \boldsymbol{\eta}_k^2(s_{ij}) + \sigma_{1k}^2}, \frac{\sigma_{1k}^2 \tau_1^2}{\tau_1^2 \sum_i \sum_j \boldsymbol{\eta}_k^2(s_{ij}) + \sigma_{1k}^2}; \geq 0 \right),$$

for $k = 1, ..., p_1$ where $\boldsymbol{\Lambda}_1(-k,-k)$ is the matrix $\boldsymbol{\Lambda}_1$ with the (k,k) element removed, $\boldsymbol{\eta}_k(s_{ij})$ is the k^{th} element of the vector $\boldsymbol{\eta}(s_{ij})$, $\gamma_{kk}(s_{ij}) = \mathbf{Y}_k(s_{ij}) - \boldsymbol{\Lambda}_1(k,-k)^T \boldsymbol{\eta}_{-k}(s_{ij})$, $\boldsymbol{\Lambda}_1(k,-k)$ is the k^{th} row of $\boldsymbol{\Lambda}_1$ with the k^{th} element removed, and $\boldsymbol{\eta}_{-k}(s_{ij})$ is $\boldsymbol{\eta}(s_{ij})$ with the k^{th} component removed.

(2) Sample $\lambda_{1kl} | \boldsymbol{\Lambda}_1(-k,-l), \boldsymbol{\Lambda}_2, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}$ from

$$\text{N}\left(\frac{\tau_1^2 \sum_i \sum_j \boldsymbol{\gamma}_{kl}(s_{ij}) \boldsymbol{\eta}_l(s_{ij})}{\tau_1^2 \sum_i \sum_j \boldsymbol{\eta}_l^2(s_{ij}) + \sigma_{1k}^2}, \frac{\sigma_{1k}^2 \tau_1^2}{\tau_1^2 \sum_i \sum_j \boldsymbol{\eta}_l^2(s_{ij}) + \sigma_{1k}^2} \right),$$

for $k > l$ where $\boldsymbol{\Lambda}_1(-k,-l)$ is the matrix $\boldsymbol{\Lambda}_1$ with the (k,l) element removed, $\boldsymbol{\eta}_l(s_{ij})$ is the l^{th} element of the vector $\boldsymbol{\eta}(s_{ij})$, $\gamma_{kl}(s_{ij}) = \mathbf{Y}_k(s_{ij}) - \boldsymbol{\Lambda}_1(k,-l)^T \boldsymbol{\eta}_{-l}(s_{ij})$,$\boldsymbol{\Lambda}_1(k,-l)$ is the k^{th} row of $\boldsymbol{\Lambda}_1$ with the l^{th} element removed, and $\boldsymbol{\eta}_{-l}(s_{ij})$ is $\boldsymbol{\eta}(s_{ij})$ with the l^{th} component removed.

(3) Sample $\lambda_{2kk}|\mathbf{\Lambda}_1, \mathbf{\Lambda}_2(-k,-k), \boldsymbol{Y}, \boldsymbol{Z}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \boldsymbol{\eta}$ from

$$\text{TN}\left(\frac{\tau_2^2 \sum_i \boldsymbol{\gamma}_{kk}(s_i)\boldsymbol{\eta}_k^*(s_i)}{\tau_2^2 \sum_i \boldsymbol{\eta}_k^{*2}(s_i) + \sigma_{2k}^2}, \frac{\sigma_{2k}^2 \tau_2^2}{\tau_2^2 \sum_i \boldsymbol{\eta}_k^{*2}(s_i) + \sigma_{2k}^2}\right),$$

for $k = 1, ..., p_2$ where $\mathbf{\Lambda}_2(-k,-k)$ is the matrix $\mathbf{\Lambda}_2$ with the (k,k) element removed, $\boldsymbol{\eta}_k^*(s_i)$ the k^{th} element of the vector $\boldsymbol{\eta}^*(s_i) = \sum_i w_{ij}\boldsymbol{\eta}(s_{ij})$, $\boldsymbol{\gamma}_{kk}(s_i) = \boldsymbol{Z}_k(s_i) - \mathbf{\Lambda}_2(k,-k)^T \boldsymbol{\eta}_{-k}^*(s_i)$, $\mathbf{\Lambda}_2(k,-k)$ is the k^{th} row of $\mathbf{\Lambda}_2$ with the k^{th} element removed, and $\boldsymbol{\eta}_{-k}^*(s_i)$ is $\boldsymbol{\eta}^*(s_i)$ with the k^{th} component removed.

(4) Sample $\lambda_{2kl}|\mathbf{\Lambda}_1, \mathbf{\Lambda}_2(-k,-l), \boldsymbol{Y}, \boldsymbol{Z}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \boldsymbol{\eta}$ from

$$\text{N}\left(\frac{\tau_2^2 \sum_i \boldsymbol{\gamma}_{kl}(s_i)\boldsymbol{\eta}_l^*(s_i)}{\tau_2^2 \sum_i \boldsymbol{\eta}_l^{*2}(s_i) + \sigma_{2k}^2}, \frac{\sigma_{2k}^2 \tau_2^2}{\tau_2^2 \sum_i \boldsymbol{\eta}_l^{*2}(s_i) + \sigma_{2k}^2}\right),$$

for $k = 1, ..., p_2$ where $\mathbf{\Lambda}_2(-k,-l)$ is the matrix $\mathbf{\Lambda}_2$ with the (k,l) element removed, $\boldsymbol{\eta}_l^*(s_i)$ the l^{th} element of the vector $\boldsymbol{\eta}^*(s_i) = \sum_i w_{ij}\boldsymbol{\eta}(s_{ij})$, $\boldsymbol{\gamma}_{kl}(s_i) = \boldsymbol{Z}_k(s_i) - \mathbf{\Lambda}_2(k,-l)^T \boldsymbol{\eta}_{-l}^*(s_i)$, $\mathbf{\Lambda}_2(k,-l)$ is the k^{th} row of $\mathbf{\Lambda}_2$ with the l^{th} element removed, and $\boldsymbol{\eta}_{-l}^*(s_i)$ is $\boldsymbol{\eta}^*(s_i)$ with the l^{th} component removed.

(5) Sample $\sigma_{1k}^2|\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \boldsymbol{Y}, \boldsymbol{Z}, \mathbf{\Sigma}_1(-k,-k), \mathbf{\Sigma}_2, \boldsymbol{\eta}$ from

$$\text{IG}\left(\frac{N_T}{2} + \alpha_1, \frac{\sum_i \sum_j (\boldsymbol{Y}_k(s_{ij}) - \mathbf{\Lambda}_1(k,\cdot)^T \boldsymbol{\eta}(s_{ij}))^2}{2} + \beta_1\right),$$

for $k = 1, ..., p_2$, where $\mathbf{\Sigma}_1(-k,-k)$ is $\mathbf{\Sigma}_1$ with the (k,k) element removed and $\mathbf{\Lambda}_1(k,\cdot)$ is the k^{th} row of $\mathbf{\Lambda}_1$.

(6) Sample $\sigma_{2k}^2|\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \boldsymbol{Y}, \boldsymbol{Z}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2(-k,-k), \boldsymbol{\eta}$ from

$$\text{IG}\left(\frac{N}{2} + \alpha_2, \frac{\sum_i \sum_j (\boldsymbol{Z}_k(s_i) - \mathbf{\Lambda}_2(k,\cdot)^T \boldsymbol{\eta}^*(s_{ij}))^2}{2} + \beta_2\right),$$

for $k = 1, ..., p_2$, where $\mathbf{\Sigma}_2(-k,-k)$ is $\mathbf{\Sigma}_2$ with the (k,k) element removed and $\mathbf{\Lambda}_2(k,\cdot)$ is the k^{th} row of $\mathbf{\Lambda}_2$.

(7) Sample $\boldsymbol{\eta}|\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ from

$$\text{MVN}\left((\boldsymbol{\Lambda}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda} + (\boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_m)^{-1})^{-1}(\boldsymbol{\Lambda}^T\boldsymbol{\Omega}^{-1}\boldsymbol{X}), (\boldsymbol{\Lambda}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda} + (\boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_m)^{-1})^{-1}\right).$$

(8) Sample $\psi = \log\left(\frac{\phi-a}{b-\phi}\right) \in \mathbb{R}$ using a Metropolis sampler with a Normal proposal distribution. ϕ is obtained by transformation such that $\phi = \frac{\exp\{\psi\}b+a}{1+\exp\{\psi\}}$.

BIBLIOGRAPHY

Agency for Toxic Substances and Disease Registry (2014). The Social Vulnerability Index. The Centers for Disease Control and Prevention. `https://svi.cdc.gov/Index.html`. Online; accessed 22-December-2016.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.

Assaf, M., Jagannathan, K., Calhoun, V. D., Miller, L., Stevens, M. C., Sahl, R., O'boyle, J. G., Schultz, R. T., and Pearlson, G. D. (2010). Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage*, 53(1):247–256.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2003). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press.

Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., and Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, 118(12):2765–2773.

Beckmann, C. F. and Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152.

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.

Bluhm, R. L., Miller, J., Lanius, R. A., Osuch, E. A., Boksman, K., Neufeld, R., Théberge, J., Schaefer, B., and Williamson, P. (2007). Spontaneous low-frequency fluctuations in the BOLD signal in schizophrenic patients: Anomalies in the default network. *Schizophrenia Bulletin*, 33(4):1004–1012.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley and Sons, Inc.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *The Annals of Statistics*, pages 1709–1722.

Britz, J., Van De Ville, D., and Michel, C. M. (2010). BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *Neuroimage*, 52(4):1162–1170.

Broyd, S. J., Demanuele, C., Debener, S., Helps, S. K., James, C. J., and Sonuga-Barke, E. J. (2009). Default-mode brain dysfunction in mental disorders: A systematic review. *Neuroscience & Biobehavioral Reviews*, 33(3):279–296.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, 1124(1):1–38.

Calhoun, V., Adali, T., and Pearlson, G. (2001a). Independent component analysis applied to fMRI data: a generative model for validating results. In *Neural Networks for Signal Processing XI, 2001. Proceedings of the 2001 IEEE Signal Processing Society Workshop*, pages 509–518. IEEE.

Calhoun, V. D. and Adali, T. (2006). Unmixing fMRI with independent component analysis. *Engineering in Medicine and Biology Magazine, IEEE*, 25(2):79–90.

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

Cardoso, J. (1997). Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing*, 4:112–114.

Chawla, M. (2011). PCA and ICA processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison. *Applied Soft Computing*, 11(2):2216–2226.

Chen, A. and Bickel, P. (2005). Consistent independent component analysis and prewhitening. *IEEE Transactions on Signal Processing*, 53(10):3625–3632.

Chen, A. C., Feng, W., Zhao, H., Yin, Y., and Wang, P. (2008). EEG default mode network in the human brain: Spectral regional field powers. *Neuroimage*, 41(2):561–574.

Chen, J.-L., Ros, T., and Gruzelier, J. H. (2013). Dynamic changes of ICA-derived EEG functional connectivity in the resting state. *Human Brain Mapping*, 34(4):852–868.

Comon, P. (1994). Independent component analysis– a new concept? *Signal Processing*, 36(3):287–314.

Comon, P. and Jutten, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press.

Congedo, M., John, R. E., De Ridder, D., and Prichep, L. (2010). Group independent component analysis of resting state EEG in large normative samples. *International Journal of Psychophysiology*, 78(2):89–99.

Cordes, D. and Nandy, R. R. (2006). Estimation of the intrinsic dimensionality of fMRI data. *Neuroimage*, 29(1):145–154.

Cover, T. and Thomas, J. (1991). *Elements of Information Theory*. Wiley.

Cressie, N. A. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, 3:159–180.

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., and Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global Environmental Change*, 18(4):598–606.

Cutter, S. L., Boruff, B. J., and Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*, 84(2):242–261.

Cutter, S. L. and Finch, C. (2008). Temporal and spatial changes in social vulnerability to natural hazards. *Proceedings of the National Academy of Sciences*, 105(7):2301–2306.

De Vos, M., De Lathauwer, L., and Van Huffel, S. (2011). Spatially constrained ICA algorithm with an application in EEG processing. *Signal Processing*, 91(8):1963–1972.

Delorme, A. and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Delorme, A., Makeig, S., Fabre-Thorpe, M., and Sejnowski, T. (2002). From single-trial EEG to brain area dynamics. *Neurocomputing*, 44:1057–1064.

Delorme, A., Makeig, S., and Sejnowski, T. (2001). Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. Paper presented at Internation Workshop on ICA (San Diego, CA).

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34:1443 – 1449.

Diener, E. and Suh, E. (1997). Measuring quality of life: Economic, social, and subjective indicators. *Social Indicators Research*, 40(1-2):189–216.

Duncan, O. D. (1974). Developing social indicators. *Proceedings of the National Academy of Sciences*, 71(12):5096–5102.

Dyrholm, M., Makeig, S., and Hansen, L. K. (2007). Model selection for convolutive ICA with an application to spatiotemporal analysis of EEG. *Neural Computation*, 19(4):934–955.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, pages 54–75.

Federal Emergency Management Agency (2017a). The National Flood Insurance Program. `https://www.fema.gov/national-flood-insurance-program`. Online; accessed 21-March-2017.

Federal Emergency Management Agency (2017b). The National Flood Insurance Program Data Definitions. `http://bsa.nfipstat.fema.gov/reports/data_definitions.html`. Online; accessed 10-April-2017.

Flexer, A., Bauer, H., Pripfl, J., and Dorffner, G. (2005). Using ICA for removal of ocular atrifacts in EEG recorded from blind subjects. *Neural Networks*, 18:998–1005.

Frank, R. M. and Frishkoff, G. A. (2007). Automated protocol for evaluation of electromagnetic component separation (APECS): Application of a framework for evaluating statistical methods of blink extraction from multichannel EEG. *Clinical neurophysiology*, 118(1):80–97.

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

Gaeta, M. and Lacoume, J. (1990). Source separation without prior knowledge: the maximum likelihood solution. *Proceedings of EUSIPCO*, pages 621–624.

Garrity, A. G., Pearlson, G. D., McKiernan, K., Lloyd, D., Kiehl, K. A., and Calhoun, V. D. (2007). Aberrant "default mode" functional connectivity in schizophrenia. *American Journal of Psychiatry*, 164(3):450–457.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

Ghosh, J. and Dunson, D. B. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, 18(2):306–320.

Giannakopoulos, X., Karhunen, J., and Oja, E. (1999). An experimental comparison of neural algorithms for independent component analysis and blind separation. *International Journal of Neural Systems*, 9(2):99–114.

Gini, C. (1912). Variabilità e mutabilità. *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi*, 1.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23):e215–e220.

Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648.

Greicius, M. D., Srivastava, G., Reiss, A. L., and Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 101(13):4637–4642.

Gusnard, D. A., Akbudak, E., Shulman, G. L., and Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(7):4259–4264.

Gusnard, D. A. and Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, 2(10):685–694.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

He, T., Clifford, G., and Tarassenko, L. (2006). Application of independent component analysis in removing artefacts from the electrocardiogram. *Neural Computing & Applications*, 15(2):105–116.

Hogan, J. W. and Tchernis, R. (2004). Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. *Journal of the American Statistical Association*, 99(466):314–324.

Huster, R. J., Debener, S., Eichele, T., and Herrmann, C. S. (2012). Methods for simultaneous EEG-fMRI: an introductory review. *Journal of Neuroscience*, 32(18):6053–6060.

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley.

Hyvarinen, A. and Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411–430.

Joyce, C. A., Gorodnitsky, I. F., and Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41(2):313–325.

Jung, T., Humphries, C., Lee, T., Makeig, S., McKeown, M., Iragui, V., and Sejnowski, T. (1998). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems*, 10:894–900.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(02):163–178.

Kachenoura, A., Albera, L., Senhadji, L., and Comon, P. (2008). ICA: A potential tool for BCI systems. *IEEE Signal Processing Magazine*, 25(1):57–68.

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

Klimesch, W., Doppelmayr, M., Roehm, D., Pöllhuber, D., and Stadler, W. (2000). Simultaneous desynchronization and synchronization of different alpha responses in the human electroencephalograph: a neglected paradox? *Neuroscience letters*, 284(1):97–100.

Kwok, R. K., Engel, L. S., Miller, A. K., Blair, A., Curry, M. D., Jackson II, W. B., Stewart, P. A., Stenzel, M. R., Birnbaum, L. S., et al. (2017). The GuLF STUDY: A prospective study of persons involved in the Deepwater Horizon oil spill response and clean-up. *Environmental Health Perspectives (Online)*, 125(4):570.

Lau, T., Gwin, J., and Ferris, D. (2012). How many electrodes are really needed for EEG-based mobile brain imaging? *Journal of Behavioral and Brain Science*, 2:387–393.

Laufs, H. (2010). Multimodal analysis of resting state cortical activity: What does EEG add to our knowledge of resting state BOLD networks? *Neuroimage*, 52(4):1171.

Ledesma, R. D. and Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2):1–11.

Lee, S. (2011). *Independent Component Analysis in Spectral Domain*. PhD thesis, University of North Carolina at Chapel Hill.

Lee, S., Shen, H., Truong, Y., Lewis, M., and Huang, X. (2011). Independent component analysis involving autocorrelated sources with an application to functional magnetic resonance imaging. *Journal of the American Statistical Association*, 106(495):1009–1024.

Li, H. and Maddala, G. (1996). Bootstrapping time series models. *Econometric Reviews*, 15(2):115–158.

Liu, X., Wall, M. M., and Hodges, J. S. (2005). Generalized spatial structural equation models. *Biostatistics*, 6(4):539–557.

Lopes, H. F., Salazar, E., and Gamerman, D. (2008). Spatial dynamic factor analysis. *Bayesian Analysis*, 3(4):759–792.

Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, 14(1):41–68.

Lusted, H. S. and Knapp, R. B. (1996). Controlling computers with neural signals. *Scientific American*, 275(4):82–87.

Makeig, S., Bell, A. J., Jung, T.-P., and Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, pages 145–151.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T.-P., Kindermann, S. S., Bell, A. J., and Sejnowski, T. J. (1997). Analysis of fMRI data by blind separation into independent spatial components. Technical report, DTIC Document.

McKeown, M. J., Sejnowski, T. J., et al. (1998). Independent component analysis of fMRI data: Examining the assumptions. *Human Brain Mapping*, 6(5-6):368–372.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R., Koppenhaver, A. M., Greischar, L. L., and Davidson, R. J. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *Neuroimage*, 49(3):2416–2432.

Meinecke, F., Ziehe, A., Kawanabe, M., and Müller, K. (2002). A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Transactions on Biomedical Engineering*, 49(12):1514–1525.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Mezzetti, M. (2012). Bayesian factor analysis for spatially correlated data: Application to cancer incidence data in Scotland. *Statistical Methods and Applications*, 21(1):49–74.

Mugglin, A. S. and Carlin, B. P. (1998). Hierarchical modeling in geographic information systems: Population interpolation over incompatible zones. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 111–130.

Musso, F., Brinkmeyer, J., Mobascher, A., Warbrick, T., and Winterer, G. (2010). Spontaneous brain activity and EEG microstates: A novel EEG/fMRI analysis approach to explore resting-state networks. *Neuroimage*, 52(4):1149–1161.

Nethery, R. C., Warren, J. L., Herring, A. H., Moore, K. A., Evenson, K. R., and Diez-Roux, A. V. (2015). A common spatial factor analysis model for measured neighborhood-level characteristics: The Multi-Ethnic Study of Atherosclerosis. *Health & Place*, 36:35–46.

Onton, J., Westerfield, M., Townsend, J., and Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Behavioral Reviews*, 30:808 – 822.

Oxfam America Inc (2009). Exposed: Social vulnerability and climate change in the US southeast. `https://policy-practice.oxfamamerica.org/static/oa3/files/Exposed-Social-Vulnerability-and-Climate-Change-in-the-US-Southeast.pdf`. Online; accessed 09-November-2016.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes, 3rd Edition*. McGraw-Hill.

Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(2):559.

Perdikaris, J. (2014). *Physical Security and Environmental Protection*. CRC Press.

Petersen, K., Hansen, L. K., Kolenda, T., Rostrup, E., and Strother, S. (2000). On the independent components of functional neuroimages. In *Third International Conference on Independent Component Analysis and Blind Source Separation*, pages 615–620.

Pham, D. and Garat, P. (1997). Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Transactions on Signal Processing*, 45(7):1712–1725.

Pham, D., Garrat, P., and Jutten, C. (1992). Separation of a mixture of independent sources through a maximum likelihood approach. *Proceedings of EUSIPCO*, pages 771–774.

Pollen, D. A. and Trachtenberg, M. C. (1972). Some problems of occipital alpha block in man. *Brain Research*, 41(2):303–314.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2):676–682.

Rowe, D. B. (1998). *Correlated Bayesian Factor Analysis*. PhD thesis, Citeseer.

Särelä, J. and Vigário, R. (2003). Overlearning in marginal distribution-based ICA: Analysis and solutions. *Journal of Machine Learning Research*, 4(Dec):1447–1469.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043.

Schalk Lab (2017). BCI2000. `http://www.schalklab.org/research/bci2000`. Online; accessed 23-April-2017.

Schoffelen, J.-M. and Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Human Brain Mapping*, 30(6):1857–1865.

Schomer, D. L. and Da Silva, F. L. (2012). *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins.

Shumway, R. and Stoffer, D. (2011). *Time Series Analysis and Its Applications, 3rd Edition*. Springer.

Simpier, T. T., Swann, D., Emmer, R., Simpier, S., and Schneider, M. (2010). Coastal resilience index: A community self-assessment. MASGP-08-014. `http://masgc.org/assets/uploads/publications/662/coastal_community_resilience_index.pdf`. Online; accessed 09-November-2016.

Smith, D. M. (1973). *The geography of social well-being in the United States: An introduction to territorial social indicators*. McGraw-Hill.

Smith, T. W. (1981). Social indicators. *Journal of Social History*, 14(4):739–747.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stakhovych, S., Bijmolt, T. H. A., and Wedel, M. (2012). Spatial dependence and heterogeneity in bayesian factor analysis: A cross-national investigation of Schwartz values. *Multivariate Behavioral Research*, 47(6):803–839.

Stine, R. (1987). Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association*, 82(400):1072–1078.

Taylor, C. L. and Hudson, M. C. (1970). World Handbook of Political and Social Indicators II. Section 1. Cross-National Aggregate Data. Technical report, Michigan University Ann Arbor Department of Political Science.

The MathWorks Inc. (2015). *MATLAB Release 2015b*. The MathWorks Inc., Natick, Massachusetts, United States.

Tian, L., Jiang, T., Wang, Y., Zang, Y., He, Y., Liang, M., Sui, M., Cao, Q., Hu, S., Peng, M., et al. (2006). Altered resting-state functional connectivity patterns of anterior cingulate cortex in adolescents with attention deficit hyperactivity disorder. *Neuroscience Letters*, 400(1):39–43.

US Bureau of the Census (1994). Geographic areas reference manual. `https://www.census.gov/geo/reference/garm.html`. Online; accessed 09-November-2016.

US Bureau of the Census (2013). American Community Survey Information Guide. `https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf`. Online; accessed 18-November-2016.

US Bureau of the Census (2016). American Community Survey Estimates, 2008-2012. Prepared by Social Explorer. `http://www.socialexplorer.com/`. Online; accessed 29-July-2016.

US Federal Bureau of Investigation (2010). Uniform Crime Reporting Data Online. `www.ucrdatatool.gov`. Online; accessed 09-November-2016.

US Federal Bureau of Investigation (2016). Uniform Crime Report Data. Prepared by Social Explorer. `http://www.socialexplorer.com/`. Online; accessed 11-October-2016.

Varoquaux, G., Sadaghiani, S., Pinel, P., Kleinschmidt, A., Poline, J.-B., and Thirion, B. (2010). A group model for stable multi-subject ICA on fMRI datasets. *Neuroimage*, 51(1):288–299.

Vigário, R. (1997). Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, 103:395–404.

Vigário, R., Särelä, J., Jousmiki, V., Hämäläinen, M., and Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering*, 47(5):589–593.

Wang, F. and Wall, M. M. (2003). Generalized common spatial factor model. *Biostatistics*, 4(4):569–582.

Whittle, P. (1952). Some results in time series analysis. *Scandinavian Actuarial Journal*, 1952(1-2):48–60.

Winter, S., Sawada, H., and Makino, S. (2003). Geometrical understanding of the PCA subspace method for overdetermined blind source separation. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 2, pages II–769. IEEE.

Wisbeck, J. O., Barros, A. K., Yy, A. K. B., and Ojeda, R. G. (1998). Application of ICA in the separation of breathing artifacts in ECG signals.

Xu, N., Gao, X., Hong, B., Miao, X., Gao, S., and Yang, F. (2004). BCI competition 2003-data set IIb: Enhancing P300 wave detection using ICA-based subspace projections for BCI applications. *IEEE Transactions on Biomedical Engineering*, 51(6):1067–1072.

Yuan, H., Zotev, V., Phillips, R., Drevets, W. C., and Bodurka, J. (2012). Spatiotemporal dynamics of the brain at restexploring EEG microstates as electrophysiological signatures of BOLD resting state networks. *Neuroimage*, 60(4):2062–2072.