Statistical Learning for the Spectral Analysis of

Time Series Data

by

Marie Tuft

BS, University of Utah, 2015

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Marie Tuft

It was defended on

July 23, 2020

and approved by

Robert T. Krafty, PhD, Associate Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Stewart J. Anderson, PhD, MA, Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Ada O. Youk, PhD, Associate Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Scott D. Rothenberger, PhD, Assistant Professor, Division of General Internal Medicine,

School of Medicine, University of Pittsburgh

Dissertation Director: Robert T. Krafty, PhD, Associate Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Statistical Learning for the Spectral Analysis of
Time Series Data

Marie Tuft, PhD

University of Pittsburgh, 2020

Abstract

Spectral analysis of biological processes poses a wide variety of complications. Statistical learning techniques in both the frequentist and Bayesian frameworks are required overcome the unique and varied challenges that exist in analyzing these data in a meaningful way. This dissertation presents new methodologies to address problems in multivariate stationary and univariate nonstationary time series analysis.

The first method is motivated by the analysis of heart rate variability time series. Since it is nonstationary, it poses a unique challenge: localized, accurate and interpretable descriptions of both frequency and time are required. By reframing this question in a reduced-rank regression setting, we propose a novel approach that produces a low-dimensional, empirical basis that is localized in bands of time and frequency. To estimate this frequency-time basis, we apply penalized reduced rank regression with singular value decomposition to the localized discrete Fourier transform. An adaptive sparse fused lasso penalty is applied to the left and right singular vectors, resulting in low-dimensional measures that are interpretable as localized bands in time and frequency. We then apply this method to interpret the power spectrum of HRV measured on a single person over the course of a night.

The second method considers the analysis of high dimensional resting-state electroencephalography recorded on a group of first-episode psychosis subjects compared to a group of healthy controls. This analysis poses two challenges. First, estimating the spectral density matrix in a high dimensional setting. And second, incorporating covariates into the estimate of the spectral density. To address these, we use a Bayesian factor model which decomposes the Fourier transform of the time series into a matrix of factors and vector of factor loadings. The factor model is then embedded into a mixture model with covariate dependent mixture weights. The method is then applied to examine differences in the power spectrum for first-episode psychosis subjects vs healthy controls.

Public health significance: As collection methods for time series data becomes ubiquitous in biomedical research, there is an increasing need for statistical methodology that is robust enough to handle the complicated and potentially high dimensionality of the data while retaining the flexibility needed to answer real world questions of interest.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

With the advancement of new technology for biomedical research and diagnosis comes a need for statistical techniques to find meaningful patterns from increasingly complex time series data. Often questions of interest have answers that are found by examining the frequency domain contents of these data, such as in examining how stress affects the autonomic nervous system by measuring oscillations in heart rate variability (HRV) (Hall et al., 2004) or using frequency band characteristics of high-density electroencephalography (EEG) to look for early signs of schizophrenia (Ferrarelli et al., 2019). The challenges that are posed in the attempt to estimate these frequency characteristics and interpret them are varied and complex. Therefore, a wide variety of statistical tools are required to meet the needs of researchers attempting to glean information to address modern problems such as these.

1.1 Frequency Domain Characteristics of a Time Series

Spectral analysis is the process used to examine the frequency characteristics of a signal. The main idea is that every time series is composed of sine and cosine waves oscillating at different frequencies. The power spectrum density of the process is akin to a frequency ANOVA, in that it shows which frequencies are present in the time series and how much influence they have over the signal. The Fourier transform is a powerful tool which allows us to calculate the spectral density. This transformation takes the autocorrelation sequence of a random process and decomposes it into its constituent frequencies (Shumway and Souffer, 2011).

Many current techniques used to estimate spectral densities are based upon the Fourier transformation. A comprehensive review of Fourier-based methodology can be found by reading the work by Brillinger (2002) or Shumway and Souffer (2011). Despite this large body of work devoted to estimating the power spectrum, there are still areas of advancement that need exploration.

1.2   Nonstationary Spectral Analysis

One of the assumptions when using the Fourier transformation to estimate the power spectrum is that the frequency characteristics of the time series do not vary over time. A time series that meets this assumption is called stationary. However, in biological applications this assumption is often violated. It is unreasonable, for example, to suppose that the heart beats at the same rate no matter what level of stress a human body is undergoing. Signals with features that change over time are called nonstationary.

Several existing techniques have been developed to handle nonstationary data. Most involve breaking the time series up into smaller pieces, with the underlying assumption that the frequency characteristics are evolving slowing over time (Adak, 1998). These smaller pieces are then considered piecewise stationary. The spectral density for a nonstationary process is represented as power varying across both frequency and time. Dahlhaus (1997) provides a good review of existing methodology to estimate this three dimensional surface.

1.3   High Dimensional Spectral Analysis

Nonstationarity is not the only complication that can arise when performing spectral analysis. Another area for consideration is multivariate spectral analysis. When analyzing multivariate signals researchers are not only interested in power for individual components, but also how they relate to one another. For instance, the electrical potentials gathered from one subject's EEG measurement not only contain frequency information about individual electrodes, but also in how different electrodes are correlated as a function of frequency. This relationship is called coherence.

While coherence provides important information about characteristics of the multivariate time series, the downside is that at every frequency the power spectrum is represented as a complex valued $P \times P$ Hermitian matrix, where P is the dimension of the time series. As the dimension of the time series increases, the number of parameters that must be estimated increases as well which can lead to ill conditioned matrices and unstable estimates. Wei (2019) provides an overview of challenges and methods to perform analysis on multivariate time series.

1.4   Outline of Work

This dissertation adds two new methods to the existing statistical toolbox in order to address problems that arise when analyzing nonstationary time series and high dimensional multivariate time series.

The first is detailed in Chapter 2 and is motivated by estimating the power spectrum of HRV time series in a way that provides insight into the workings of the autonomic nervous system (Malik et al., 1996). Because HRV is a nonstationary time series, it poses a specific challenge in that the frequency characteristics of its power spectrum can vary over time (Priestley, 1965). Furthermore, since time-varying power is estimated as a three dimensional surface, often clinicians use summarizing measures in their research, such as power within a band of frequencies. Our method hopes to provide an alternative by aiding in the interpretation of these structures by reframing the typical locally stationary Fourier estimate of the time varying spectrum in a penalized reduced rank regression setting. This allows for the power spectrum to be broken up into multiple unit-rank layers that are formed by multiplying an "importance" singular value, a left singular "time" vector, and a right singular "frequency" vector together. An adaptive sparse fused lasso penalty is imposed on these vectors that introduces sparsity and smoothness into the estimate. These layers can then be examined individually for patterns and the singular vectors provide a parsimonious representation of the time- and frequency-varying characteristics of the power spectrum.

The second method, outlined in Chapter 3 is geared toward incorporating covariates into estimation of the spectral density matrix of high dimensional multivariate time series. This type of methodology has many potential applications, but here we focus on the differences in 64 channel high density resting state EEG between healthy controls and first-break psychosis patients. Typical methodology involves averaging power across established frequency bands and using regression to test for differences in covariates. However we provide an analysis that would allow us to estimate the entire spectral density matrix conditional upon covariates. To do this, first we decrease the dimensionality of the problem by representing the likelihood of the Fourier transformation of the time series as a factor model. Then we model the covariate dependent spectral density matrix using a Bayesian mixture model. Mixture weights are represented using multinomial regression with a

Polya-Gamma data augmentation and provide flexible output of the covariate dependent estimate of the spectral density matrix.

Finally, in Chapter 4 a summary of the finding is given as well as suggestions for future work.

2.0   Empirical Localized Time-Frequency Analysis via Penalized Reduced Rank Regression

2.1   Introduction

Analysis of the frequency-domain characteristics of a nonstationary time series has many important applications in the biomedical field. For instance, examining the frequencies present in an EEG can give insight into functions of the brain (Valdés et al., 1992; Tang et al., 2013), the spectrum of oximetry data can lead to a diagnosis of obstructive sleep apnoea (Álvarez et al., 2009), and the power spectrum of heart rate variability (HRV) can give insight into functions of the autonomic nervous system (ANS) (Hall et al., 2004). However, despite the ever-increasing applications of nonstationary spectral analysis as a tool for understanding the human body, the inherent complexity of such power spectra presents unique challenges.

Since a nonstationary time series has features that change over time, the resulting power spectrum is a time-frequency surface in three dimensional space. This makes meaningful interpretation difficult. Adding to this challenge is the fact that typically only certain frequencies and times of the power spectrum contain useful information, but it is often unclear how to localize those frequencies and times from the surrounding white noise. Often clinical researchers resort to collapsing entire frequency bands of the power spectrum. For instance, to examine HRV a common summarizing measure is the ratio of the total power in the 0.05-0.15 Hz frequency band and the total power in the 0.15-0.40 Hz frequency band (Hall et al., 2004; Bonnet and Arand, 1997). While this provides a scaler measure of the power spectrum which can easily be used in basic statistical methods such as ANOVA, it also leads to inaccurate inference by propagating uncertainty and fails to capture the changing time-frequency dynamics of spectrum (Bruce et al., 2018).

There are several existing methods that deal with the estimation of the time-frequency surface. One class of these assumes that the time-varying processes underlying the time series are evolving slowly enough that one can partition the series into locally stationary time blocks. The power spectrum can then be estimated within each time block using traditional methods for stationary time series such as the periodogram Priestley (1965); Dahlhaus (1997); Adak (1998). Extensions

of this method minimize bias in the periodogram due to spectral leakage by using windowing functions or tapers Thomson (1982); Kim et al. (2018); Harris (1978). The estimates can be denoised by smoothing the periodogram directly (Shumway and Souffer, 2011) or by using splines (Guo et al., 2003). Typically important features of the power spectrum are localized using either frequency band estimation (Schröder and Ombao, 2019; Bruce et al., 2019) or temporal change point detection (Adak, 1998; Davis et al., 2006).

The SLEX model of Ombao et al. (2002) and its cousin Auto-SLEX (Ombao et al., 2001) use specialized basis functions that are smooth and localized in time and frequency simultaneously. However, choosing the best set of basis function for this method can be complicated and localization is dependent upon having multiple time series classes to discriminate between (Ho et al., 2008). Wavelet transform is another type of transformation which provides smooth, localized estimates of the power spectrum using a basis function called a mother wavelet that is scaled and shifted across windows of time (Daubechies, 1990). The main advantage of a wavelet transform over Fourier-based methods is that a wavelet uses longer time windows to calculate coefficients associated with lower frequencies and shorter time windows to calculate coefficients associated with higher frequencies. This addresses the resolution issues associated with the fixed time windows used in a Fourier transform (Zhang et al., 2003).For most biological applications, such as HRV, one expects nearby frequencies to behave similarly and that some frequency bands will be related through underlying dynamic processes. For example, a band of HRV frequencies could have similar characteristics during each period of REM throughout a night of sleep. In this scenario, the rich basis that allows wavelets to estimate localized time-frequency features makes isolating frequency bands in an interpretable manner challenging.

In this chapter we introduce a novel method that aims to provide a parsimonious, interpretable estimate of the time-varying power spectrum of nonstationary time series which we call SpeLLL (**Spe**ctra in **L**ow rank **L**ocalized **L**ayers). This method combines penalized reduced-rank regression with traditional Fourier techniques, leading to three unique properties. First, using singular value decomposition (SVD) the power spectrum can be broken up into mutually exclusive low-rank layers. These layers allow for related dynamics of the power spectrum to be examined individually. Because of this unique layering, uninteresting processes can be visually excluded from the spectrum if desired to allow for examination of the relationship between processes of interest. Second,

6

an adaptive sparse fused lasso penalty imposes a smooth and sparse structure onto each layer of the spectrum. The resulting estimate is less noisy and it is easy to locate related frequency bands and time blocks that hold dominate features of the power spectrum. Furthermore, the tuning parameters involved in the lasso give a large degree of control over the level of sparsity and smoothness in a given layer. And third, because each of these layers is unit-rank it can be expressed simply as the multiplication of two singular vectors (one associated with time and the other associated with frequency) and a singular value. This parsimonious representation allows for easy interpretation of the dominant time-varying or frequency-varying processes at play.

The rest of this chapter is outlined as follows. Section 2.2 gives a brief review of the underlying time series methods and outlines methodology to transform a time series from the time domain to the time-frequency domain. This common frequency domain technique is then reframed in the reduced rank regression setting in Section 2.3 where we introduce penalized regression and outline how to solve the resulting objective function. The results of a simulation study comparing SpeLLL to known methods are shown and elaborated on in Section 2.4. The method is then applied in Section 2.5 to examine characteristics of the HRV power spectrum for a chronically stressed subject. Finally, Section 2.6 offers an overview of the results and concluding remarks.

2.2   Time Varying Power Spectrum

2.2.1   Piecewise Locally Stationary Time Series

Nonstationary processes have characteristics that change over time. However, if those characteristics evolve slowly over time, then the process can be treated as locally stationary(Adak, 1998; Priestley, 1965).

Using the time-varying Cramér representation, a zero-mean locally stationary time series $X_{t,T}$ for $t = 1, ..., T$ can be represented as

$$X_{t,T} = \int_{-1/2}^{1/2} A(u, \omega) e^{2\pi i t \omega} \, dZ(\omega) \tag{2.1}$$

where $dZ(\omega)$ is an orthogonal increment, zero-mean process and $A(u, \omega)$ is a function of frequency ω and time u such that $A : [0, 1] \times [-1/2, 1/2] \rightarrow \mathbb{C}$ with $A(u, \omega) = A(u, -\omega)$. The time-dependent power spectrum of X_t at time u is given by $f(u, \omega) = |A(u, \omega)|^2$ (Adak, 1998).

We can break the time points u into B equal time blocks such that $B = T/t_b$ where t_b is the number of time points per block. Here, we assume that T is a multiple of B for the sake of convenience. Let $0 = u_0 < u_1 < ... < u_B = 1$ denote the endpoints of each of the B time blocks. Then for $i = 1, ...B$

$$X_{t,T} = \sum_{i=0}^{B} I(u) X_t^{(i)}$$

where $X_t^{(i)}$, $i = 1, ..., B$ are independent zero-mean stationary processes with spectra $f^{(i)}(\omega)$ and $I(u) = 1$ if $u \in (u_i, u_{i+1})$. Then it follows that the time-dependent spectrum of the piecewise stationary process at time u is given by

$$f(u, \omega) = \sum_{i=0}^{B} I(u) f^{(i)}(\omega) \tag{2.2}$$

2.2.2 Estimate of Spectrum

Within each of the $i = 1, ..., B$ time blocks we can estimate the local power spectrum $f^{(i)}(\omega)$ via the periodogram.

$$\mathbf{f}^{(i)}(\omega) = \frac{1}{t_B} \left| \sum_{t=1}^{T} I(t/T) X_t e^{-2\pi i \omega t} \right|^2 \tag{2.3}$$

Then, letting $Q = t_b/2$, $f^{(i)}(\omega)$ can be thought of as a $1 \times Q$ vector of power values at each of the Fourier frequencies used to calculate the periodogram for the ith time block. We then use each $f^{(i)}$ to build a matrix $\mathbf{Y}$ such that

$$\mathbf{Y}_{B \times Q} = \begin{bmatrix} \mathbf{f}^{(1)}(\omega) \\ \mathbf{f}^{(2)}(\omega) \\ \vdots \\ \mathbf{f}^{(B)}(\omega) \end{bmatrix} \tag{2.4}$$

Here $\mathbf{Y}$ represents a rough estimate of the time-varying power spectrum. However, this estimate of the power spectrum is based on the periodogram which is known to be asymptotically unbiased, but noisy and possessing a non-trivial bias in smaller sample sizes (Adak, 1998).

2.3 Penalized Reduced Rank Regression

To further refine our estimate the of time varying power spectrum $\mathbf{Y}$, we use penalized reduced rank regression.

The multivariate regression model is given as

$$\mathbf{Y} = \mathbf{XC} + \mathbf{E}$$

where $\mathbf{Y}$ is a $n \times Q$ matrix of responses, $\mathbf{X}$ is a $n \times B$ matrix of predictors, $\mathbf{C}$ is a $B \times Q$ coefficient matrix, and $\mathbf{E}$ is an $n \times Q$ matrix of error vectors assumed to be independently and identically distributed (Reinsel and Velu, 1998). In our specific setting, $\mathbf{Y}$ is the Fourier transformed estimate of the power spectrum that we seek to refine, and $\mathbf{C}$ is the final estimate of the power spectrum. Covariates related to each time bin can be included in the model via the columns of $\mathbf{X}$. Or if the typical estimate of the power spectrum as a time-frequency surface is desired, $\mathbf{X} = \mathbf{I}_B$.

Penalized reduced-rank regression uses two techniques to build upon classic multivariate regression. The first is restricting the rank r of the coefficient matrix $\mathbf{C}$ to be $r \leq \min(B, Q)$. This implies that there are $\min(B, Q) - r$ linear restrictions on $\mathbf{C}$ (Reinsel and Velu, 1998). This rank restriction allows for a parsimonious representation of the patterns found in the power spectrum. The second is penalizing the terms of the coefficient matrix to introduce smoothness and sparsity to estimate of the power spectrum $\mathbf{C}$. Forcing sparsity into the model allows us to remove contributions to the power spectrum from white noise while reducing the dimension of the time blocks and frequency bins. Adding smoothness allows us to take advantage of the fact that contributions to the spectrum from adjacent time blocks or frequency bins are likely to be similar. Our technique for introducing the penalties builds upon the methodology introduced in Chen et al. (2012).

Chen et al. (2012) uses singular value decomposition (SVD) to express matrix $\mathbf{C}$ as the sum of r rank one layers,

$$\mathbf{C} = \mathbf{UDV}^T = \sum_{k=1}^{r} d_k \mathbf{u}_k \mathbf{v}_k^T = \sum_{k=1}^{r} \mathbf{C}_k,$$

where $\mathbf{U}$ is a $B \times r$ matrix of left singular vectors, $\mathbf{V}$ is a $Q \times r$ matrix of right singular vectors, and $\mathbf{D}$ is a $r \times r$ diagonal matrix of ordered singular values such that $d_1 > d_2 > ... > d_r$. The kth rank one layer of $\mathbf{C}$ is then given by $\mathbf{C}_k = d_k \mathbf{u}_k \mathbf{v}_k^T$. We assume that all singular values are distinct.

Using SVD to decompose $\mathbf{C}$ in this manner allows for an appealing interpretation of the estimate of the power spectrum. It breaks the power spectrum up into r mutually exclusive layers which can be examined individually. Within each of the k layers, $\mathbf{u}_{k,(1\times B)}$ is associated with time effects on the power spectrum and $\mathbf{v}_{k,(1\times Q)}$ is associated with frequency effects. d_k gives the relative importance of the kth layer.

Where our method differs from Chen et al. (2012) is that an adaptive sparse fused lasso penalty is then imposed on the triplets $(d_k, \mathbf{u}_k, \mathbf{v}_k)$ for each of the $k = 1, .., r$ layers. The objective function is given by

$$\underset{(d,\mathbf{u},\mathbf{v})}{\mathrm{argmin}}\left\{\frac{1}{2}||\mathbf{Y} - \sum_{k=1}^{r} d_k\mathbf{u}_k\mathbf{v}_k^T||_F^2 + \sum_{k=1}^{r} Pe(\lambda_k^{(1)}, \lambda_k^{(2)}, \lambda_k^{(3)}, d_k, \mathbf{u}_k, \mathbf{v}_k)\right\} \qquad (2.5)$$

where $||\mathbf{u}_k||_2 = ||\mathbf{v}_k||_2 = 1$. Here $||.||_F$ indicates the the Frobenius norm, $\boldsymbol{\lambda}_k = (\lambda_k^{(1)}, \lambda_k^{(2)}, \lambda_k^{(3)})$ are the tuning parameters controlling the amount of sparsity and fusion in each layer, and $Pe(.)$ is the adaptive sparse fused lasso penalty given by

$$\begin{aligned} Pe(\lambda_k^{(1)}, \lambda_k^{(2)}, \lambda_k^{(3)}, d_k, \mathbf{u}_k, \mathbf{v}_k) &= \lambda_k^{(1)} \sum_{i=1}^{B}\sum_{j=1}^{Q} w_{ki}^{(u)} w_{kj}^{(v)} w_k^{(d)} |d_k u_{ki} v_{ku}| + \\ &\quad \lambda_k^{(2)} \sum_{i=2}^{B} m_{ki}^{(u)} |d_k u_{ki} - d_k u_{k,i-1}| + \\ &\quad \lambda_k^{(3)} \sum_{j=2}^{Q} m_{kj}^{(v)} |d_k v_{kj} - d_k v_{k,j-1}|. \end{aligned}$$

$\mathbf{w}^{(u)}, \mathbf{v}^{(u)}, w_{(d)}, \mathbf{m}^{(u)}$, and $\mathbf{m}^{(v)}$ are the adaptive weights which will be defined in subsection 2.3.1. Within this penalty, $\lambda_k^{(1)}$ controls the amount of sparsity by shrinking individual elements of $\mathbf{u}_k$ and $\mathbf{v}_k$ to zero. The corresponds with zeroing out a time block or frequency bin from layer k of the power spectrum. $\lambda_k^{(2)}$ and $\lambda_k^{(3)}$ control the smoothness of the estimate of the power spectrum by penalizing the distance between adjacent elements of $\mathbf{u}_k$ and $\mathbf{v}_k$, forcing similarity between neighboring frequencies and times in the spectral estimate.

2.3.1 Estimating a Single Layer

To estimate a single layer, we assume that $r = 1$ and therefore equation (2.5) reduces to minimizing the following objective function with respect to $(d, \mathbf{u}, \mathbf{v})$.

$$\frac{1}{2}||\mathbf{Y} - d\mathbf{X}\mathbf{u}\mathbf{v}^T||_F^2 = \lambda^{(1)} \sum_{i=1}^{B} \sum_{j=1}^{Q} w_i^{(u)} w_j^{(v)} w^{(d)} |du_i v_u| + \tag{2.6}$$

$$\lambda^{(2)} \sum_{i=2}^{B} m_i^{(u)} |du_i - du_{i-1}| + \lambda^{(3)} \sum_{j=2}^{Q} m_j^{(v)} |dv_j - dv_{j-1}|$$

where $d\mathbf{u}\mathbf{v}^T$ is the SVD of a single rank one layer of $\mathbf{C}$.

To create the weights we assume some consistent estimate of $\mathbf{C}$ is available, namely the ordinary least squares (OLS) estimate $\check{\mathbf{C}}$. The SVD of the OLS estimate is then given by $\check{d}\check{\mathbf{u}}\check{\mathbf{v}}^T$ and is used to calculate the weights such that

$$
\begin{aligned}
w^{(d)} &= |\check{d}|^{-\gamma}, \\
w_i^{(u)} &= |\check{u}_i|^{-\gamma}, \\
w_j^{(v)} &= |\check{v}_i|^{-\gamma}, \\
m_i^{(u)} &= |\check{u}_i - \check{u}_{i-1}|^{-\gamma} \text{ for } i = 2, ..., B, \\
m_j^{(v)} &= |\check{v}_j - \check{v}_{j-1}|^{-\gamma} \text{ for } j = 2, ..., Q.
\end{aligned}
$$

where γ is a non-negative parameter. Zou (2006) further discusses values for γ, but on the basis of our simulations we choose $\gamma = .25$.

The objective function (2.6) is multiconvex. For a fixed value of $\mathbf{u}$, the minimization of $\tilde{\mathbf{v}} = d\mathbf{v}$ is given by the objective function

$$\frac{1}{2}||\mathbf{y} - \mathbf{X}^{(v)}\tilde{\mathbf{v}}||^2 + \lambda^{(1,v)} \sum_{j=1}^{q} w_j^* |\tilde{v}_j| + \lambda^{(3)} \sum_{j=2}^{q} m_j^* |\tilde{v}_j - \tilde{v}_{j-1}|, \tag{2.7}$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{X}^{(v)} = \mathbf{I}_q \otimes \mathbf{X}\mathbf{u}$, $\lambda^{(1,v)} = \lambda^{(1)} \sum_{i=1}^{B} w_i^{(u)} |u_i|$, $\mathbf{w}^* = w^{(d)}\mathbf{w}^{(v)}$, and $\mathbf{m}^* = w^{(d)}\mathbf{m}^{(v)}$. The symbol '$\otimes$' denotes the Kronecker product. For this minimization we ignore the

11

portion of equation (2.6) dealing with the fusion of $\mathbf{u}$ since it is constant with respect to $\tilde{\mathbf{v}}$. Rewriting in this fashion shows that (2.7) is simply an adaptive sparse fused lasso problem in $\tilde{\mathbf{v}}$.

For a fixed value of $\mathbf{v}$ the minimization with respect to $\tilde{\mathbf{u}} = d\mathbf{u}$ is given by the objective function

$$\frac{1}{2}||\mathbf{y} - \mathbf{X}^{(u)}\tilde{\mathbf{u}}||^2 + \lambda^{(1,u)}\sum_{i=1}^{p} w_i^\dagger|\tilde{u}_i| + \lambda^{(2)}\sum_{i=2}^{p} m_i^\dagger|\tilde{u}_i - \tilde{u}_{i-1}| \tag{2.8}$$

where $\mathbf{X}^{(u)} = \mathbf{v} \otimes \mathbf{X}$, $\lambda^{(1,u)} = \lambda^{(1)}\sum_{i=j}^{q} w_j^{(v)}|v_j|$, $\mathbf{w}^\dagger = w^{(d)}\mathbf{w}^{(u)}$, and $\mathbf{m}^\dagger = w^{(d)}\mathbf{m}^{(u)}$. Similarly as in (2.7), here we ignore the fusion of $\mathbf{v}$ since it is a constant with respect to $\tilde{\mathbf{u}}$ and notice that (2.8) is also an adaptive sparse fused lasso problem.

By breaking (2.6) up in this manner we can take advantage of its structure and solve using block coordinate descent. Namely, for fixed values of tuning parameters $\boldsymbol{\lambda}$ we can use the following steps:

(a) Choose a non-zero initial value of $\hat{\mathbf{u}}$, $\breve{\mathbf{u}}$ being a convenient choice.

(b) Fix $\mathbf{u} = \hat{\mathbf{u}}$. Minimize (2.7) to obtain $\tilde{\mathbf{v}}$. Set $\hat{d} = ||\tilde{\mathbf{v}}||_2$ and $\hat{\mathbf{v}} = \hat{d}^{-1}\tilde{\mathbf{v}}$

(c) Fix $\mathbf{v} = \hat{\mathbf{v}}$. Minimize (2.8) to obtain $\tilde{\mathbf{u}}$. Set $\hat{d} = ||\tilde{\mathbf{u}}||_2$ and $\hat{\mathbf{u}} = \hat{d}^{-1}\tilde{\mathbf{u}}$

(d) Iterate steps (b) and (c) until $\hat{\mathbf{C}} = \hat{d}\hat{\mathbf{u}}\hat{\mathbf{v}}^T$ converges. Convergence is achieved when $||\hat{\mathbf{C}}_l - \hat{\mathbf{C}}_{l-1}||_F/||\hat{\mathbf{C}}_{l-1}||_F < \epsilon$ for the lth iteration and for some tolerance level ϵ.

This algorithm also works when $\mathbf{u}$ is updated first instead of $\mathbf{v}$. Either way, for each step (b) and (c), equation (2.6) becomes an adaptive sparse fused lasso problem which can be solved using alternating direction method of multipliers (ADMM) within each step (Li et al., 2014; Boyd et al., 2010). See Appendix A.1 for details.

2.3.1.1   Parameter Tuning   There are three regularization parameters $\boldsymbol{\lambda} = (\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)})$ per layer. For smaller problems it is possible to use K-fold cross validation (CV) to choose $\boldsymbol{\lambda}$. However as the sampling rate of the underlying time series increases, CV becomes very computationally intense. Following the methodology of Chen et al. (2012), we use a Bayesian information criterion (BIC) as a more computationally efficient method of choosing $\boldsymbol{\lambda}$.

Let $(\hat{d}^{\boldsymbol{\lambda}}, \hat{\mathbf{u}}^{\boldsymbol{\lambda}}, \hat{\mathbf{v}}^{\boldsymbol{\lambda}})$ be the fitted value of $(d, \mathbf{u}, \mathbf{v})$ for the regularization parameter vector $\boldsymbol{\lambda}$. Then

$$\text{BIC}(\boldsymbol{\lambda}) = \log[\text{SSE}(\boldsymbol{\lambda})] + \frac{\log(Qn)}{Qn}\text{df}(\boldsymbol{\lambda}) \tag{2.9}$$

12

where $\text{SSE}(\boldsymbol{\lambda}) = ||\mathbf{Y} - \hat{d}^{\boldsymbol{\lambda}}\mathbf{X}\hat{\mathbf{u}}^{\boldsymbol{\lambda}}\hat{\mathbf{v}}^{\boldsymbol{\lambda}T}||_F^2$ is the sum of squared error. $\text{df}(\boldsymbol{\lambda})$ is the degrees of freedom of the model. The number of non-zero coefficients, given by

$$\text{df}(\boldsymbol{\lambda}) = \sum_{i=1}^{B} I(\hat{u}_i^{\lambda} \neq 0) + \sum_{j=1}^{Q} I(\hat{v}_i^{\lambda} \neq 0) - 1,$$

is an unbiased estimator of the degrees of freedom in a sparse fused lasso setting (Tibshirani et al., 2005). One degree of freedom is lost because there are two constraints, $||\mathbf{u}||_2 = 1$ and $||\mathbf{v}||_2 = 1$ and one additional free parameter d (Chen et al., 2012).

Since tuning three parameters can be computationally intensive, we used a two-stage grid search to determine the regularization path. In the first stage, a sparse grid is used to locate an approximate region of values, $\boldsymbol{\lambda}_k(\text{BIC}_1)$, which minimizes BIC. In the second stage, a finer grid within a neighborhood of $\boldsymbol{\lambda}_k(\text{BIC}_1)$ is used. This provides a speed advantage since for iteration t we are able to use the estimate $\hat{\mathbf{C}}$ from $\boldsymbol{\lambda}_{t-1}$ as the initial guess for $\boldsymbol{\lambda}_t$ provided $\boldsymbol{\lambda}_t > \boldsymbol{\lambda}_{t+1}$.

2.3.1.2 Spectral Bands Frequency bands in the power spectrum $\hat{\mathbf{C}}_k$ can be found where there are adjacent groups of zero and non-zero elements of $\hat{\mathbf{v}}$. Similarly, temporal break points can be found where there are neighboring zero and non-zero values of $\hat{\mathbf{u}}$.

2.3.2 Multiple Layers

To solve objective function (2.5) for $r > 1$ we follow the iterative exclusion extraction algorithm used in (Chen et al., 2012). First an initial estimate $\dot{\mathbf{C}}$ is found using ordinary least squares methodology and decomposed such that $\dot{\mathbf{C}} = \sum_{k=1}^{r} \dot{d}_k \dot{\mathbf{u}}_k \dot{\mathbf{v}}_k^T = \sum_{k=1}^{r} \dot{\mathbf{C}}_k$. Then we seek for a $\hat{\mathbf{C}}$ with a sparse, smooth SVD structure near $\dot{\mathbf{C}}$ by breaking up the problem into r individual rank one regressions on layer $\mathbf{Y}_k$. The steps are as follows:

(a) For each $k \in \{1, ..., r\}$

    (i) Construct the weights $\mathbf{w}^{(u)}, \mathbf{v}^{(u)}, w_{(d)}, \mathbf{m}^{(u)}$, and $\mathbf{m}^{(v)}$ based on the SVD of $\dot{\mathbf{C}}$.

    (ii) Create the layer $Y_k = Y - X(\dot{\mathbf{C}} - \dot{\mathbf{C}}_k)$

    (iii) Find $(\hat{d}_k, \hat{\mathbf{u}}_k, \hat{\mathbf{v}}_k)$ by performing penalized reduced rank regression of $\mathbf{Y}_k$ on $\mathbf{X}$. The tuning parameters $\boldsymbol{\lambda}_k$ can be found using CV or BIC.

(b) Set $\hat{\mathbf{C}} = \sum_{k=1}^{r} \hat{d}_k \hat{\mathbf{u}}_k \hat{\mathbf{v}}_k^T$

(c) Set $\dot{C} = \hat{C}$ and iterate over (a) and (b) until $||\dot{C} - \hat{C}||/||\dot{C}|| < \epsilon$ for some tolerance level ϵ.

Note that because of step (ii), the computations for step (iii) can be performed in parallel, speeding up the process.

2.3.2.1 Choosing r   Thus far we have assumed that the rank r of **C** is known. Chen and Chan (2016) has shown that for a choice of rank $\hat{r} > r$ all estimated layers $\hat{C}_k$ for $k \in \{r + 1, ..., \hat{r}\}$ become vanishingly small. Therefore, it is better to overestimate r than underestimate it.

In practice, one way to choose an initial guess for r is to examine the singular values of the initial guess $\breve{d}$. From here a cutoff may be obtained of singular values that are too small to produce a non-zero layer of $\hat{C}$.

2.4   Simulations

2.4.1   Simulation Scenarios

There are three specific properties of SpeLLL that we were interested in examining via simulation: accuracy in recovering the rank structure of the power spectrum, accuracy in localization in time and frequency, and interpretablity.

For accuracy in estimation of the spectral layers, we compare the spectral estimate from SpeLLL to the singular value decomposition of the estimate of the power spectrum recovered from the multitaper method (Thomson, 1982), which is the average of multiple estimates of the power spectrum generated from orthogonal tapers; a short time Fourier transformation (STFT) with a Hanning window (Harris, 1978), which breaks the time series up into stationary pieces using a windowing function; and a wavelet transform using the Morlet wavelet basis (Torrence and Compo, 1998; Daubechies, 1990). Because the wavelet transform provides an estimate of the power spectrum at each time point, unlike Fourier based methods which require time bins, we selected a similar number of time points to use in the comparison to adjust for the finer grid of the wavelet.
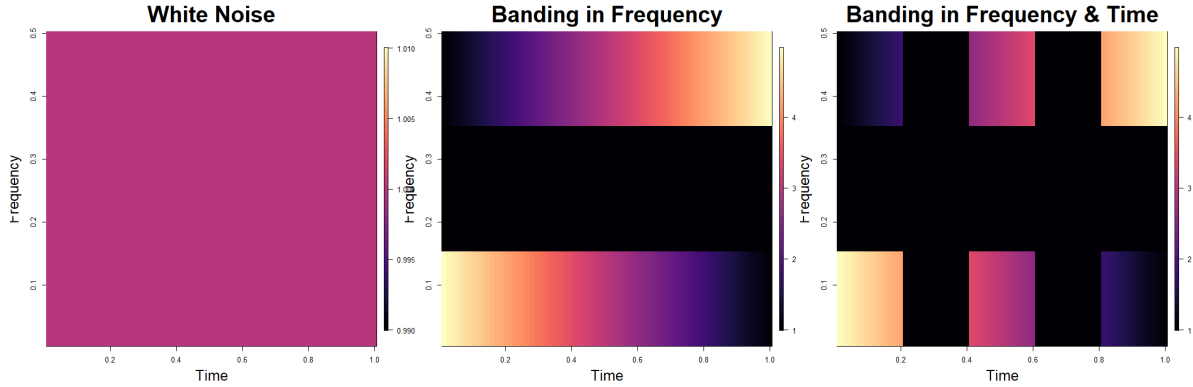
Figure 2.1: Target power spectrum for simulation scenarios.

White noise, banding in frequency, banding in frequency and time target power spectra for simulation scenarios to assess performance of SpeLLL.

To explore how accurate the localization properties of SpeLLL are, we will compare it to the EBA method (Bruce et al., 2019), which is a Baysian technique that searches for frequency bands in the power spectrum and PELT (Killick et al., 2012) which uses cost functions to look for temporal break points. For interpretability, we visually examine the results of method as compared to the estimates recovered by the multitaper, STFT and wavelet transform.

To evaluate the relative performance of the methods, we considered the three different simulation scenarios similar to the scenarios in (Bruce et al., 2019). The first is simply white noise (WN) with the following spectral density function

$$f_1(u, \omega) = 1 \text{ for } \omega \in (0, 0.05).$$

The second scenario involves banding in frequency (BF) and has the spectral density

$$f_2(u, \omega) = \begin{cases} 5 - 4u & \text{for } \omega \in (0, 0.15) \\ 1 & \text{for } \omega \in [0.15, 0.35) \\ 1 + 4u & \text{for } \omega \in [0.35, 0.5). \end{cases}$$

The third involves banding in both time and frequency (BFT) and has the spectral density

$$f_3(u, \omega) = \begin{cases} f_2(u, \omega) & \text{for } u \in [0, 0.2] \cup (0.4, 0.6] \cup (0.8, 1] \\ f_1(u, \omega) & \text{for } u \in (0.2, 0.4] \cup (0.6, 0.8] \end{cases}$$

15

The target spectrum for these scenarios are shown in Figure 2.1.

For each scenario we consider three different sampling densities $T = 500$, $T = 5000$, and $T = 15000$. The number of blocks for each of these sample sizes is $B = 25, 100,$ and 200 respectively.

2.4.2 Spectral Rank Structure Estimation

To measure the accuracy of the estimate of the power spectrum for each of the three methods compared we used scaled mean squared error calculated for each spectral layer, SMSE $= 100||\mathbf{C}_i - \hat{\mathbf{C}}_i||_F^2/BQ$ for $i = 1, ..., r$ where r is the layer of the spectral density. Because the comparison methods were not designed specifically to estimate the rank structure of the spectral density matrix, we only examined the first r layers of each where r is the true value of the number of layers of the spectral density. In practice, both the multitaper and STFT method contained a non-trivial amount of information in layers greater than r. Table 2.1 shows the SMSE for each method under each scenario.

In the WN scenario, SpeLLL has a lower SMSE than the STFT and wavelet estimate for all three sample sizes. This is probably due to the fact that neither the STFT or the wavelet have a means to remove white noise from the spectral estimate. It has a lower SMSE than multitaper for $T = 500$ and ties for the two larger sample sizes. The smoothing feature of the multitaper method seems to allow it to perform at a similar level to SpeLLL for this scenario.

In the BF scenario, SpeLLL recovers both layers of the rank structure of the power spectrum more accurately than the other methods for all sample sizes, except in the $T = 500$ scenario for the second layer. This layer is a little more complex in shape (see Figure 2.2) than the first layer and it appears wavelet is slightly better at accurately capturing this dynamic in the smallest sample size. But as T increases, SpeLLL's ability to smooth the estimate and introduce zeroes again allows it to be the most accurate method of the four.

In the BFT scenario, the wavelet is slightly more accurate for smaller sample sizes. But when $T = 15000$ SpeLLL again is the most accurate method of the three. This reinforces the trend seen in the BF scenario where as the sample size increases, the fusion and lasso penalty allow for SpeLLL to recover the true rank structure of the matrix more accurately.

16

Table 2.1: Comparison of SpeLLL to exisiting methods: rank structure.

Scenario	Layer	T	SpeLLL	Multitaper	STFT	Wavelet
WN	First	500	1.10 (0.72)	1.93 (0.30)	8.94 (0.55)	5.88 (0.23)
		5000	0.25 (0.16)	0.25 (0.03)	3.80 (0.05)	1.78 (0.04)
		15000	0.06 (0.01)	0.06 (0.01)	1.55 (0.01)	1.04 (0.02)
BF	First	500	3.31 (0.55)	7.90 (0.65)	7.46 (0.72)	2.72 (0.22)
		5000	0.64 (0.10)	1.57 (0.08)	4.03 (0.07)	0.85 (0.03)
		15000	0.29 (0.03)	0.79 (0.03)	1.64 (0.06)	0.50 (0.01)
	Second	500	2.21 (0.39)	4.63 (0.46)	5.66 (0.72)	1.20 (0.10)
		5000	0.30 (0.10)	0.86 (0.08)	1.02 (0.06)	0.31 (0.02)
		15000	0.11 (0.02)	0.36 (0.03)	0.40 (0.01)	0.18 (0.01)
BFT	First	500	3.29 (0.30)	6.38 (0.64)	8.28 (0.63)	2.12 (0.21)
		5000	0.79 (0.10)	1.20 (0.09)	4.00 (0.06)	0.68 (0.03)
		15000	0.29 (0.03)	0.59 (0.03)	1.63 (0.01)	0.40 (0.01)
	Second	500	2.04 (0.27)	4.18 (0.45)	5.27 (0.64)	1.08 (0.10)
		5000	0.38 (0.16)	0.77 (0.08)	0.91 (0.06)	0.28 (0.01)
		15000	0.11 (0.02)	0.32 (0.03)	0.36 (0.01)	0.16 (0.01)

Scaled MSE reported as mean (standard deviation) of 1000 simulation runs.

Table 2.2: Comparison of SpeLLL to existing methods: estimated number of bands in frequency/time.

Scenario	T	Time			Frequency		
		PELT	SpeLLL	Truth	SpeLLL	EBA	Truth
	500	1.00	1.00	1	1.00	1.00	1
WN	5000	1.00	1.00	1	1.00	1.00	1
	15000	1.00	1.00	1	1.00	1.00	1
	500	1.02	1.06	1	3.37	1.37	3
BF	5000	1.05	1.00	1	3.02	3.00	3
	15000	1.46	1.00	1	3.00	3.33	3
	500	1.08	7.31	5	2.50	1.18	3
BFT	5000	4.52	4.99	5	3.02	3.54	3
	15000	5.00	4.99	5	3.01	3.01	3

Simulation results comparing the number of estimated frequency bands or temporal break points to true value for 1000 simulation runs.

2.4.3 Frequency and Time Localization

To measure the accuracy in localizing bands of frequency and time we use two metrics. The first is simply the number of bands that exist in the power spectrum, denoted as b. For the white noise scenario, $b = 1$ in both the frequency and time dimension. For the banding in frequency scenario, $b = 3$ in the frequency dimension and $b = 1$ in the time dimension. For the banding in frequency and time scenario, $b = 3$ in the frequency dimension and $b = 5$ in time dimension.

The second metric is the Rand Index, which is typically used in a clustering setting to measure the accuracy of groupings of clustered points. In our setting, we can think of assigning time and frequency bins to banded clusters and use the Rand Index to assess the accuracy of the partition points for these clusters. The Rand Index ranges from 0 to 1, with values close to 1 indicating that the estimate of the partitions are in the correct place (Rand, 1971).

The results for estimating b are shown in Table 2.2. In time, SpeLLL generally converged much faster than PELT to the correct number of bands. In the banding in frequency scenario, PELT seemed to find a number of false bands which only got worse as sample size increased. In the banding in frequency and time scenario, SpeLLL was closer to the correct number of bands in the low sample size and converged more rapidly to the correct number as sample size increased. In frequency, SpeLLL also outperformed the EBA algorithm. This is especially true in low sample sizes. SpeLLL converged to the proper number of frequency bands more quickly than EBA.

The results for assessing the accuracy of the band placement via the Rand Index is shown in 2.3. In time, SpeLLL outperformed PELT for every sample size in the banding in frequency and banding in frequency/time scenarios. The issue with PELT finding false temporal breakpoints in the banding in frequency scenario affected the Rand Index in banding in frequency. In frequency, SpeLLL outperformed EBA for every sample size in the banding in frequency and banding in frequency/time scenarios.

Overall, SpeLLL is not only able to estimate the number of bands better than EBA or PELT, but it also places the location of the breakpoints for those bands much more accurately than either method. This is especially true in low sample sizes.

Table 2.3: Comparison of SpeLLL to existing methods: Rand Index for frequency/time bands.

| Scenario | T | Time | | Frequency | |
		PELT	SpeLLL	SpeLLL	EBA
	500	1.00	1.00	1.00	1.00
WN	5000	1.00	1.00	1.00	1.00
	15000	1.00	1.00	1.00	1.00
	500	1.00	0.99	0.83	0.42
BF	5000	0.99	1.00	0.97	0.80
	15000	0.90	1.00	0.99	0.85
	500	0.20	0.85	0.69	0.35
BFT	5000	0.77	0.99	0.96	0.80
	15000	0.95	0.99	0.98	0.87

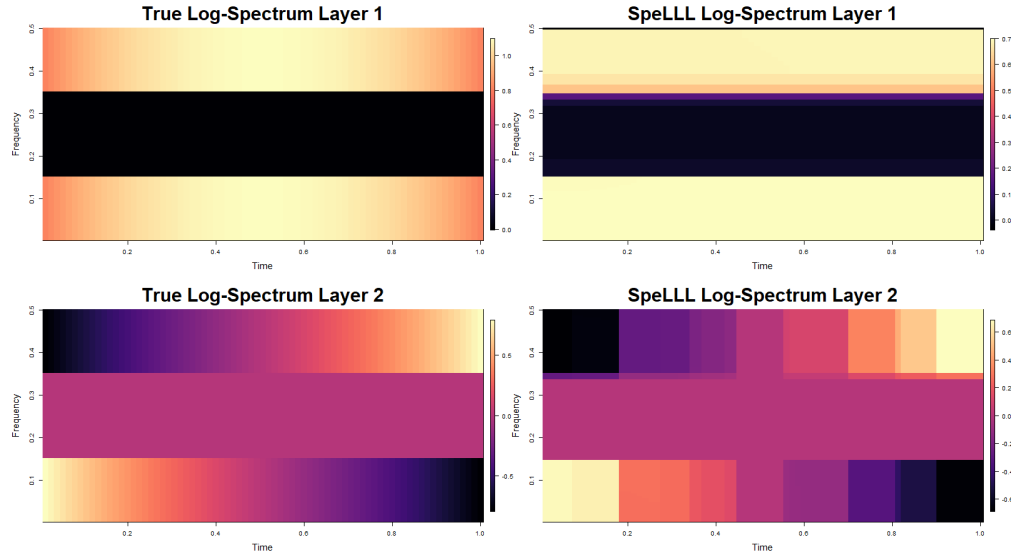Simulation results comparing the Rank Index between methods for 1000 simulations runs.

Figure 2.2: True and estimated rank structure for one run of banding in frequency.

True and estimated log power spectrum values from a single run of the banding in frequency scenario. The top row shows the results from the first layer recovered by SpeLLL compared to the first layer of the SVD of the true power spectrum. The bottom row compares the second layers.

2.4.4 Interpretablity

The one of the unique strengths of SpeLLL is the ability to break the power spectrum up into layers which are then processed individually. Figure 2.2 shows each of the two spectral layers for the banding in time scenario. We can clearly see that SpeLLL shrinks the white noise estimates to zero and forces adjacent time and frequency bins to be similar to one another. Simulating this scenario requires two steps. The first step involves running a high pass and a low pass filter over two time series. The second step adds a coefficient which makes the high pass time series more dominate later in time and the low pass time series more dominate early in time. Then the resulting time series are combined.

Referring to Figure 2.2, we can see that the first layer shows two clearly defined bands of power, one at low frequency and one at high frequency. The bands of high and low power are analogous to the high and low pass filters used to generate this simulated time series. In layer 2 we see that within the frequency bands captured by layer, power is increasing over time for high frequencies and decreasing over time for low frequencies. This seems to correspond to adding the coefficient
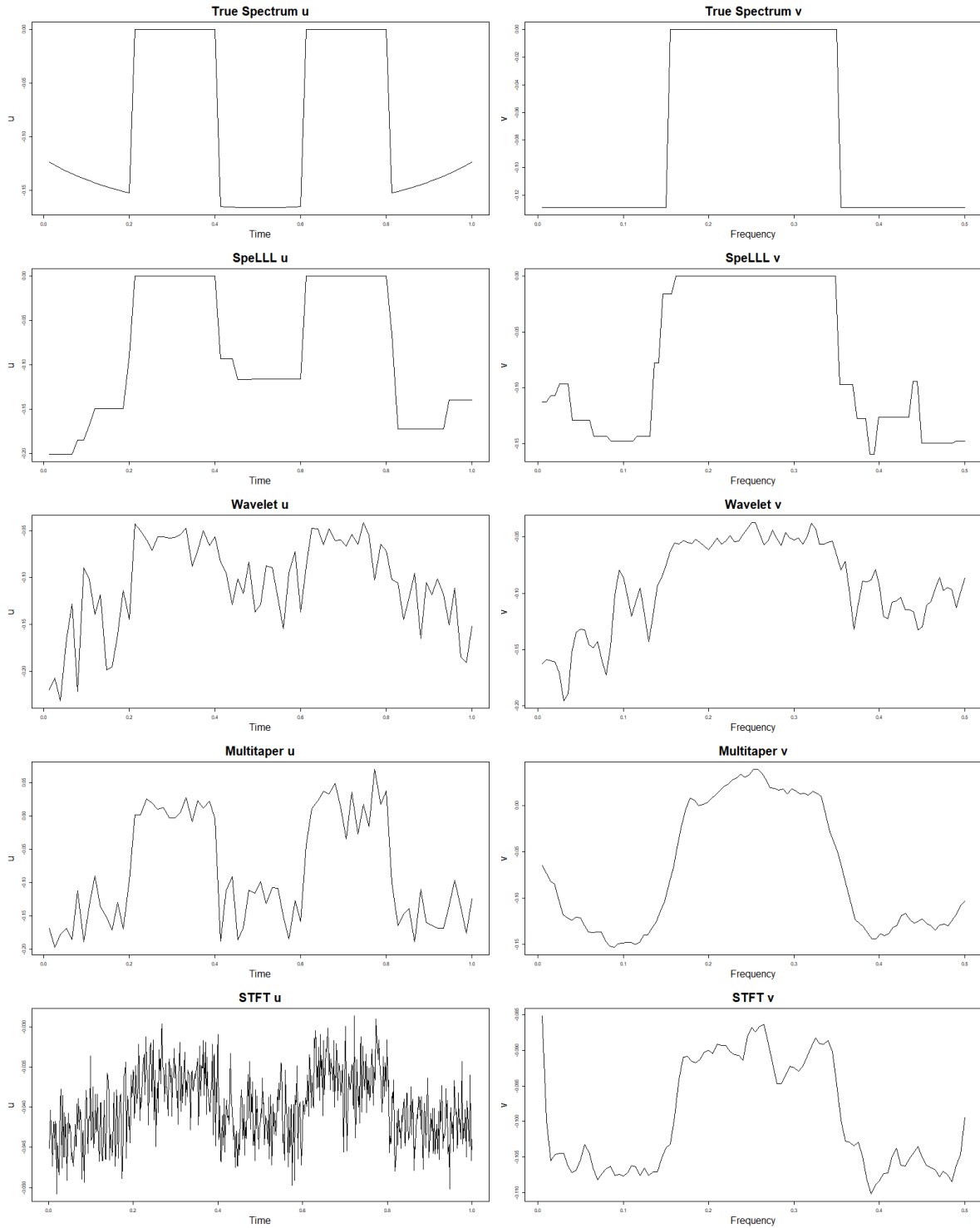
Figure 2.3: Visual comparison of $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ between methods.

The recovered $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors for the first layer of the BFT scenario with $T = 15000$. The first row is the true $\mathbf{u}$ and $\mathbf{v}$, and the rest are from the four spectral estimation techniques considered in this section as indicated by the plot titles.

22

to make the high pass more dominate later in time and the low pass dominant early in time. By breaking up the spectrum into layers and then smoothing and adding sparsity individually, SpeLLL is able to clearly capture the processes involved in simulating the time series for these scenarios.

Another unique property which allows SpeLLL to recover interpretable estimates of the spectral density is the smoothness and sparsity of the estimated $\mathbf{u}$ and $\mathbf{v}$ vectors. The $\mathbf{u}$ vector provides a summary of the temporal dynamics and the $\mathbf{v}$ vector provides a summary of the frequency dynamics of the power spectrum. These vectors can also be created by performing SVD on other methods, but the result is noisy and not sparse.

Figure 2.3 shows the $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors recovered for the first layer of the BFT scenario with $T = 15000$. The top row is the $\mathbf{u}$ and $\mathbf{v}$ from the true power spectrum. We see that the changes in time and frequency are generally smooth with an abrupt shift to zero during the time and frequency periods of white noise for this scenario. The rows below show the $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ recovered from each of the methods tested in this section. When visually comparing the four methods, it is easy to see that the fusion penalty introduced by SpeLLL allows the estimate of the $\mathbf{u}$ and $\mathbf{v}$ vectors to be much less noisy than the other methods. This makes it easier to spot patterns across time or frequency. The other noteworthy difference between SpeLLL and the other methods is that portions of the $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors are set completely to zero. This makes identifying periods of white noise in the power spectrum simple compared to the other methods where it is unclear whether the peaks in $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are white noise or just areas of low power.

Overall, the smoothness and sparsity of the $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors recovered by SpeLLL make them a powerful tool to be able to interpret time and frequency patterns in the power spectrum. This will be well illustrated in the next section as we apply SpeLLL to real world data.

2.5   Application: Heart Rate Variability

Stress can have a negative impact on the way that we sleep. It is relatively straightforward to examine how stress and worry delay the time it takes to fall asleep. However, it takes more sophisticated techniques to study how stress affects the nonrapid eye movement (NREM) and rapid eye movement (REM) stages of a night of sleep (Hall et al., 2004). Tracking how stress affects

the body as it cycles between NREM, REM, and wakefulness is critical to understanding the full picture of the relationship between sleep and stress. One method researchers are increasingly using to attempt to quantify stress is through tracking the function of the autonomic nervous system (ANS) by measuring heart rate variability (HRV) (Malik et al., 1996; Shaffer and Ginsberg, 2017; Bonnet and Arand, 1997).

HRV is the elapsed time between consecutive heart beats. Power in certain frequencies of the spectral density function of HRV is associated with arousal in different systems of the ANS. Power in the high frequency (HF) band ($0.15 - 0.4$ Hz) is associated with the parasympathetic, or rest/relax, response. Increased power in lower frequency (LF) band ($0.04 - 0.15$ Hz) accompanied by a decline of power in the HF band is associated with the sympathetic, or stress, response. And power in very low frequency (VLF) band ($< .04$ Hz) is associated with slow regulatory mechanisms of the body (Malik et al., 1996). In general, REM sleep typically shows increased activity in the sympathetic response and NREM sleep shows increased activity in the parasympathetic response (Hall et al., 2004).

To examine the sleep characteristics of an individual under stress, we analyze the HRV time series from a participant in the AgeWise Caregiver Study. Participants in this study were the primary caregiver for spouses with Alzheimer's, Parkinson's, or other advanced forms of dementia who reported disruptive sleep problems. For further information on inclusion/exclusion criteria for this study, refer to (Taylor et al., 2015).

2.5.1 Data Processing

The HRV of the participant randomly selected from the study pool was detrended, interpolated, and sampled at a rate of once per second over 8.6 hours of sleep. The resulting time series contained $T = 31680$ time points. Figure 2.4 shows a plot of the resulting HRV time series. The colored bar at the top represents which sleep stage (NREM, REM, or Awake) the participant was in at the given time point. These sleep stages were derived from an electroencephalogram by a trained sleep technician using established protocols.

This time series was then partitioned into thee minute intervals based on the suggestion by (Malik et al., 1996), leading to $B = 170$ time blocks. A Fourier transform was then applied to
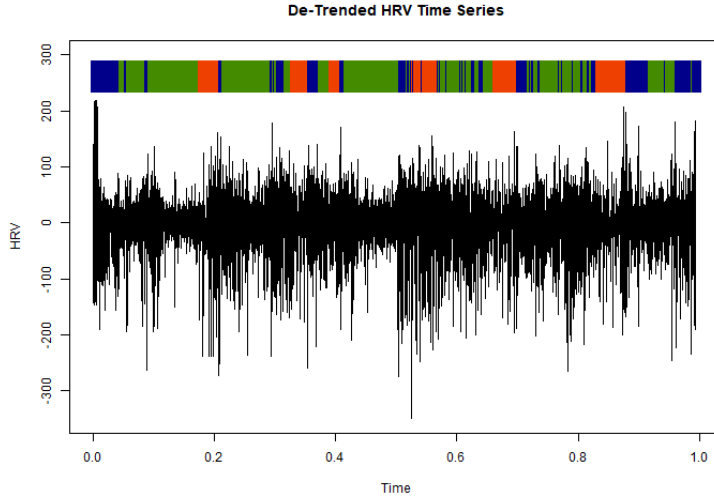
Figure 2.4: Processed HRV data.

The processed HRV time series measured over the course of a night for a single participant. The corresponding sleep stages as reported by a trained sleep technician are indicated in the colored bar at the top of the plot. Blue = awake; green = NREM; orange = REM.

each time block to calculate the periodogram for each interval. Each of the B transformations was then stacked to create a 170×90 matrix $\mathbf{Y}$. This matrix was then centered and logged.

Two different $\mathbf{X}$ matricies were used to analyze the data. For the first analysis we used $\mathbf{X} = \mathbf{I}$ to get a general estimate of the rank structure of the power spectrum. In the second analysis we incorporated the sleep stage information into the columns of $\mathbf{X}$. To do this, the time blocks were further partitioned into sleep stage epochs where one epoch consisted of all adjacent time blocks with the same sleep stage. For instance, the first b blocks of time where the participant was awake would be one epoch. Then if the participant entered a period of NREM, the next b blocks would be considered a second epoch and so on. All together there were 34 sleep stage epochs throughout the night. The columns of $\mathbf{X}$ then consisted of indicators such that

$$\mathbf{X} = \left[I(b_i = 1)', I(b_i = 2)', ..., I(b_i = 34)' \right] \text{ for } i = 1, ..., 170$$

where b_i is the sleep epoch number for each time bin. Then the columns of $\mathbf{X}$ were permuted so that the sleep stages were adjacent to each other. Namely, the first awake epoch was adjacent to the second awake epoch and so on. The graph for the fusion penalty was then modified so that the
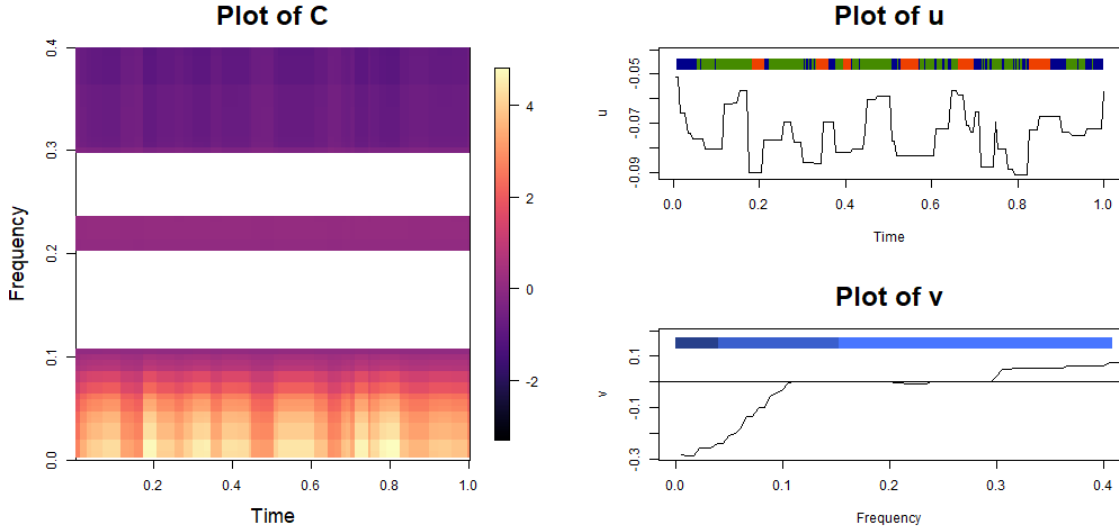
25

Figure 2.5: First estimated layer of HRV log power spectrum for $\mathbf{X} = \mathbf{I}$.

First layer of HRV log power spectrum (left) and its corresponding $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors (right). Frequency time points with zero value have been removed in the plot of $\hat{\mathbf{C}}$. The colored bar over the plot of $\hat{\mathbf{u}}$ indicates which sleep stage the participant was in for a given time bin. Blue=awake; green=NREM; orange=REM. The colored bar over the plot of $\hat{\mathbf{v}}$ indicates which of the traditionally used frequency bands a given frequency is in. Dark=VLF; medium=LF; light=HF.

last awake epoch would not be linked to the first NREM epoch and the last NREM epoch would not be linked to the first REM epoch.

2.5.2 Analysis for $\mathbf{X} = \mathbf{I}$

Ordinary least squares reduced-rank regression was performed on $\mathbf{Y}$ to get an initial estimate for the SpeLLL algorithm. The singular values of this initial estimate were examined and it was determined that the rank of $\mathbf{Y}$ was 3. The algorithm was then applied to decompose, smooth, and localize the HRV power spectrum.

Figure 2.5 shows the results for the first layer of the HRV power spectrum. The plot of $\hat{\mathbf{C}}_1$ shows three regions of nonzero power. The plot of $\hat{\mathbf{v}}_1$ indicates that most of this power occurs in the VLF frequency band and extends into the LF frequency band. This seems to indicate a slow, regulatory process that dominates the power spectrum throughout the night. Two additional frequency bands occur in the HF band of the power spectrum. The plot of $\hat{\mathbf{u}}_1$ shows no localization
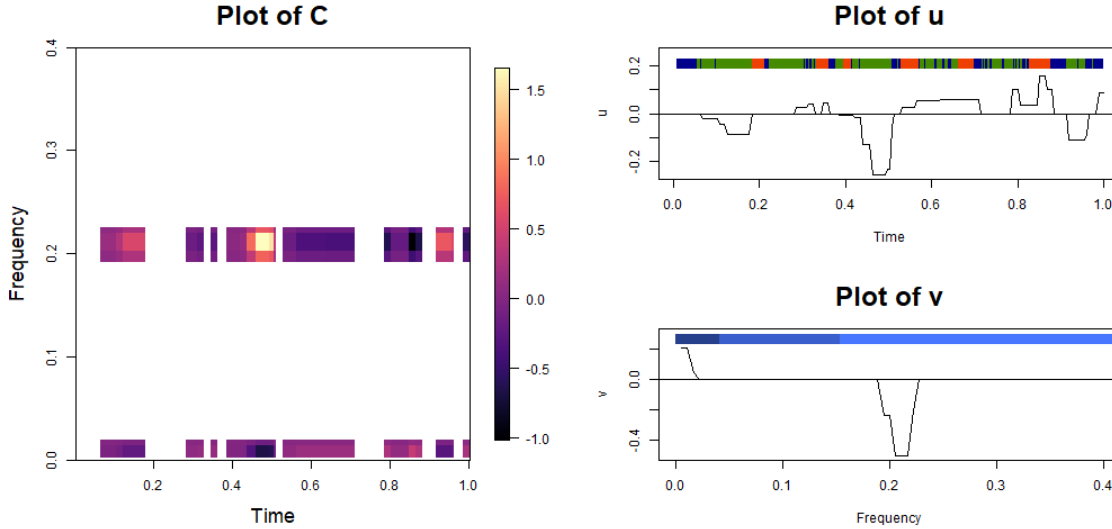
26

Figure 2.6: Second estimated layer of HRV log power spectrum for $\mathbf{X} = \mathbf{I}$.

Second layer of HRV log power spectrum (right) and its corresponding $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors (left). Frequency time points with zero value have been removed in the plot of $\hat{\mathbf{C}}$. Color bar for $\hat{\mathbf{u}}$: Blue=awake, green=NREM, orange=REM. Color bar for $\hat{\mathbf{v}}$: dark=VLF, medium=LF, light=HF.

in time (likely due to problems with resolving low frequency signals in time), but it does show drops in power occurring in periods of NREM sleep directly before REM sleep.

Figure 2.6 shows the results for the second layer of the HRV power spectrum. The plot of $\hat{\mathbf{C}}_2$ shows two regions of nonzero power, and the plot of $\hat{\mathbf{v}}_2$ indicates that these bands occur in the VLF and HF frequency range. For the HF band, the plot of $\hat{\mathbf{u}}_2$ shows three main peaks of power which correspond to periods of NREM sleep. Interestingly, these three periods are identifiable in Figure 2.4 as regions where the HRV time series is tighter about the zero axis. Since the parasympathetic nervous system is more likely to be activated during NREM sleep (Bonnet and Arand, 1997), this finding agrees with the current HRV literature.

Figure 2.7 shows the results for the third layer of the HRV power spectrum. The plot of $\hat{\mathbf{C}}_3$ shows two bands of power, one lower frequency band, which has a peak early and a peak the end of the night, and another which has a peak in the later half of the night. The plot of $\hat{\mathbf{v}}_3$ indicates that these two bands occur in the LF and HF frequency regions. Since the LF band includes some influences from the sympathetic nervous system as well as the parasympathetic nervous system, typically clinicians think of the sympathetic contribution to the HRV power spectrum in
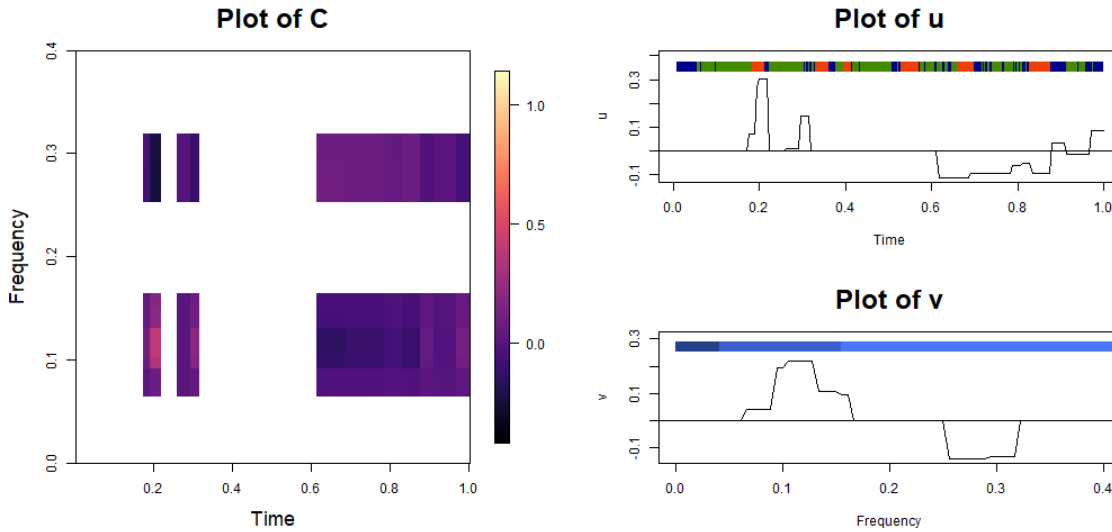
Figure 2.7: Third layer of HRV log power spectrum for $\mathbf{X} = \mathbf{I}$.

Third layer of HRV log power spectrum (left) and its corresponding $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors (right). Frequency time points with zero value have been removed in the plot of $\hat{\mathbf{C}}$. Color bar for $\hat{\mathbf{u}}$: Blue=awake, green=NREM, orange=REM. Color bar for $\hat{\mathbf{v}}$: dark=VLF, medium=LF, light=HF.

terms of a ratio of LF/HF contributions (Bonnet and Arand, 1997). This shows up in this layer as decreased regions of power in the HF band directly tied to increased regions of power in the LF band. Examining the plot of $\hat{\mathbf{u}}_3$ shows that the peaks of power in the LF band in the first half of the night correspond to the first episode of REM and the awake period before the second episode of REM. The peaks of power in the second half of the night occur in the HF band as the subject rapidly cycles between awake, NREM, and REM and then the power switches to the LF band right before the subject wakes up for the day.

The overall estimate of the HRV power spectrum is found by summing all three layers together. Figure 2.8 showcases one unique feature of SpeLLL. If certain layers are not of particular interest in examining features of the power spectrum, they can simply be left out of the summation as a means of filtering out unimportant information. Here layer one dominates the estimate of the power spectrum, but the VLF power band does not provide important information regarding how the participant's sleep varied over the course of the night. The plot of only layer two and three gives a clearer view of where peaks of power occurred throughout the night and at what frequency.
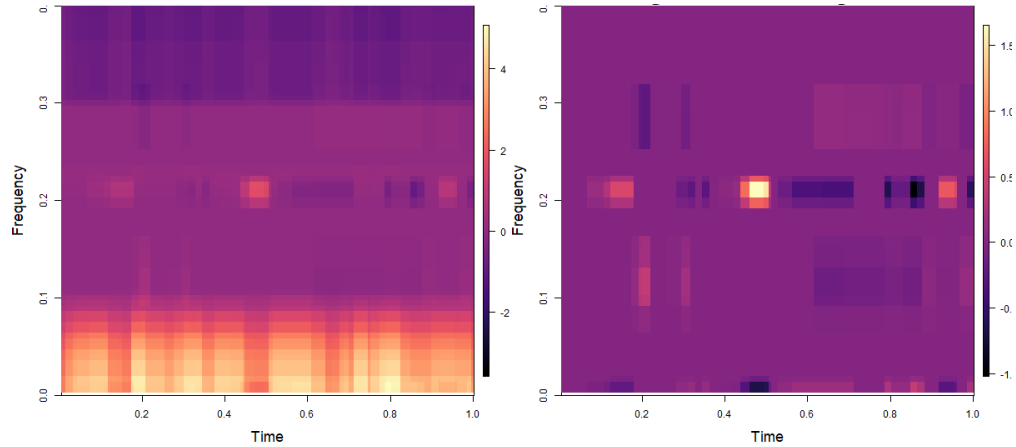
Figure 2.8: Final estimate of HRV log power spectrum for $\mathbf{X} = \mathbf{I}$.

Left: All three layers summed together for a final estimate of $\hat{\mathbf{C}}$. Right: The estimate of the spectrum without the first layer, namely $\hat{\mathbf{C}}_2 + \hat{\mathbf{C}}_3$. This illustrates the ability of SpeLLL to remove less pertinent information from the power spectrum if desired.

2.5.3 Analysis for Sleep Epoch X

Ordinary least squares regression was performed and it was determined that the rank of $\hat{\mathbf{C}}$ for this analysis is 2. Figure 2.9 shows the estimate of the power spectrum recovered from SpeLLL and Figure 2.10 plots the $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors for both layers. In both layers the contribution to the power spectrum from the epochs where the participant was awake were removed almost entirely.

The plots of $\hat{\mathbf{v}}$ for the first and second layer reflect patterns related to the LF/HF ratio that researchers typically connect with the sympathetic response of the ANS. In layer 1 this ratio would be large because of a concentration of power in the LF band and a smaller ridge of power in the HF band indicating a stress response of the body. This pattern is seen in the earlier NREM and REM epochs, perhaps indicating that the subject took longer to settle into restful sleep. It is also seen in the later NREM and REM epochs as the participant shifts into wakefulness.

In layer 2, the plot of $\hat{\mathbf{v}}_2$ shows a different LF/HF ratio. Here we see almost no contribution to the power spectrum from the LF and a prominent ridge of power from the HF. This is consistent with a parasympathetic response. This pattern occurs mostly in the middle sleep epochs for the NREM stage as the subject settles in to restful sleep and is reversed in the final epoch of REM
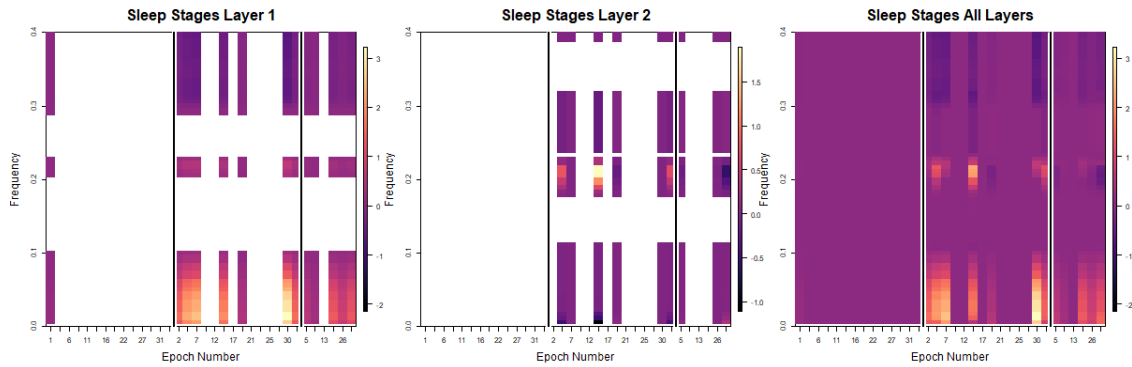
29

Figure 2.9: Estimated value of HRV log power spectrum for sleep epoch **X**.

The estimated value of $\hat{\mathbf{C}}$ for each layer and overall. Time frequency points which contribute zero to the estimate are removed from the plot for the first and second layers. The black lines within each layer plot denotes the boundaries between sleep stage contributions to the estimate: left = Awake, middle = NREM, right = REM.
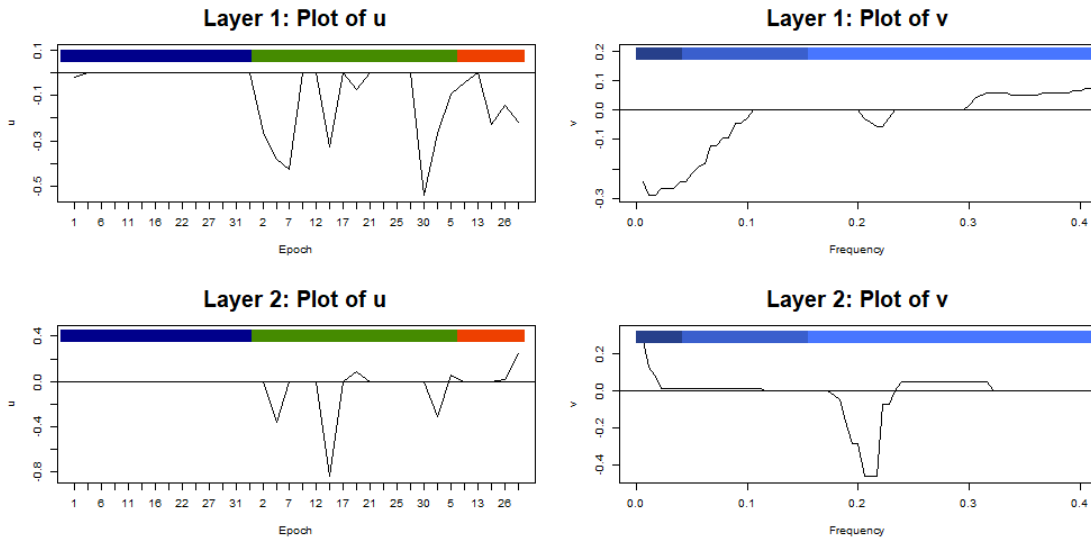

Figure 2.10: Estimated left and right singular vectors for sleep epoch **X**.

Estimate of $\hat{\mathbf{v}}$ and $\hat{\mathbf{u}}$ for first and second layer. Color bar for **u**: Blue=awake, green=NREM, orange=REM. Color bar for **v**: dark=VLF, medium=LF, light=HF.

sleep. This reversal could be due to an active dream or the sympathetic nervous system ramping up in preparation for the morning.

Overall this analysis reflects similar dynamics to those seen when $\mathbf{X} = \mathbf{I}$, but summarized in a more parsimonious way.

2.6 Discussion

The SpeLLL algorithm reframes a traditional method to handle nonstationary time series data as a penalized reduced-rank regression problem. The unique advantages of moving to this framework are threefold. First, it uses SVD to break the estimate of the power spectrum up into multiple layers which can be examined individually. This brings out patterns in the data which may get lost when the power spectrum is considered as a whole. In the first HRV analysis, dynamics captured in the third layer were difficult to pick up on when viewing the power spectrum overall. However when examined by itself, the increased power in the LF band for the first period of REM was quite clear. Second, it imposes smoothness and sparsity to each layer of the spectral estimate which greatly improves the interpretability of the output. This was shown in the HRV data as localized bands in frequency and time that could be easily classified according to sleep stage or ANS frequency band. Third, it provides a parsimonious estimate of the power spectrum, since each spectral layer consists of only two vectors and a scaler. Plotting the $\mathbf{u}$ and $\mathbf{v}$ for the HRV data alongside meaningful frequency and time categorizations allowed for easy examination of frequency and time domain characteristics.

Additionally, this model allows for the incorporation of time or frequency specific covariates. In the HRV analysis we used sleep stage epochs to include within the model further information of interest. This provided us with interpretable LF/HF characteristics of the HRV power spectrum that were less clear without using this information. Another potential source of covariate information would be known frequency bands. One could transpose the matrix $\mathbf{Y}$ and create an $\mathbf{X}$ matrix which indicates which frequency band a particular frequency bin landed in.

A more complicated extension would be to take advantage of the parsimonious output of SpeLLL and extend it to the multiple time series case via penalized tensor principal components

31

analysis. This would allow researchers to estimate a time-varying basis vector $\mathbf{u}$ and a frequency-varying basis vector $\mathbf{v}$ for a larger number of participants. A third vector would then be added to the mix which could give information on how the time/frequency basis vectors varied based on subject level covariates.

3.0 Weighted Factor Model for High Dimensional Stationary Time Series

3.1 Introduction

As time series data collection becomes easier, cheaper, and more prevalent, the amount of research being done involving multiple processes measured over time has increased. Many researchers are interested in finding patterns that exist within and across the power spectra of these processes as well as how these patterns relate to covariates of interest. However, the analysis becomes more challenging with increasing dimensionality of the time series.

For instance, the motivating application of this chapter involves the analysis of 64-channel high density resting-state electronencephalography (EEG) measured on first-episode psychosis patients (FEP) vs healthy controls (HC). Resting state EEG measures brain activity during wakefulness while the subject is in a relaxed state. Because it is free of the confounding effects of level of attention and motivation that exist in EEG measured while the subject is performing a task, the frequency domain characteristics of this type of EEG can give insight into intrinsic differences in brain function between HC and FEP (Khanna et al., 2015). These differences could potentially serve as a clinical biomarker for schizophrenia. However because the dimensionality of the time series information gathered by 64-channel EEG on each individual is so large, existing methodology to analyze this type of data in relation to subject-level covariates is sparse.

The difficulties faced when performing this type of estimation are two-fold. First the high dimensionality of the problem requires that the number of parameters estimated at each frequency is on the order of P^2, where P is the dimension of the time series. This poses serious computational challenges due to the unstable condition number of large spectral density matrices. Existing methodology for analyzing high dimensional stationary time series includes shrinkage estimators (Ledoit and Wolf, 2004; Bohm and von Sachs, 2009; Fiecas et al., 2010; Fiecas and Ombao, 2011; Fiecas and von Sachs, 2014; Schneider-Luftman and Walden, 2016), which involves shrinking the estimate of the spectral density to a target matrix; thresholding estimators (Sun et al., 2018), which are extensions of recent developments in the field of covariance matrix estimators (Bickel and Levina, 2008; Cai and Zhou, 2011) and are derived from thresholding averaged periodiograms; and

factor models, which involve decomposing the time series into a vector of factors and a loading matrix. Factor models have been extensively studied in the stationary time domain–see Ensor (2013) for a review paper. However, there are few examples of factor models in the frequency domain (Stoffer, 1999; Macaro and Prado, 2014; Li et al., 2019). Furthermore, none of these methods provide for the use of covariates in the estimation procedure or allow estimation of a common power spectrum for multiple multivariate time series.

In the multiple univariate time series setting, Cadonna et al. (2019) and Bertolacci et al. (2019) use a Bayesian mixture model to estimate power spectra dependent on covariates. AdaptSpecX uses a Gibbs sampler to estimate the power spectrum of nonstationary multiple time series data. White it allows for a large number of time series, it does not provide an adaptation to estimate the coherence necessary for multiple multivariate spectral analysis. On the other hand, Cadonna et al. (2019) does provide some discussion on how to adapt their method to accommodate the challenges of analyzing multivariate time series data, but admit that in practice their method would only be able to handle at most a 2-3 dimensional time series.

In this chapter we introduce a novel method for the estimation of covariate dependent high dimensional stationary power spectra. Our approach uses an extension of the Whittle likelihood based factor model representation of the power spectrum from Li et al. (2019) that allows us to estimate a common spectral density for N independent multivariate time series. The use of a factor model paradigm allows us mitigate the challenges inherent in the high dimensional setting by representing the spectral density matrix using a $Q \times P$ loading matrix where $Q < P$. This effectively reduces the number of parameters to be estimated. Then we embed the resulting estimation procedure into a Bayesian mixture model framework (Gelman et al., 2014). The power spectrum is therefore decomposed into several basis functions multiplied by covariate dependent weights modeled as a multinomial logistic regression. This allows us to estimate a spectral density matrix conditional on any combination of subject-level covariates.

The rest of the chapter is outlined as follows. Section 3.2 gives an overview of the methodology for this chapter and provides an outline of the sampling scheme used. A simulation study is then performed in Section 3.3 to assess the performance of the model in known conditions. Then in Section 3.4 the 64 channel EEG data is analyzed. And the chapter is concluded in Section 3.5 with an overview of results.

3.2 Methodology

The machinery for this model consists of two main parts. In section 3.2.1 we review the factor model introduced by Li et al. (2019) to estimate the power spectrum from a single stationary high dimensional multivariate time series. Then in section 3.2.2 we modify this framework to estimate a common power spectrum from N multivariate time series. We extend this further in section 3.2.3 by embedding the factor model into a Bayesian mixture model framework which will allow us to introduce covariates into the estimation procedure.

3.2.1 Likelihood for Single Multivariate Time Series

Let $\mathbf{X_t} = (X_{1,t}, ..., X_{P,t})'$ for $t = 1, ..., T$ be a single P dimensional vector-valued time series. $\mathbf{X_t}$ has a factor model Cramér representation given by

$$\mathbf{X}_t = \int_0^1 \Lambda(\omega) \exp(2\pi i \omega t) d\mathbf{Z}(\omega) + \boldsymbol{\epsilon}, \tag{3.1}$$

that decomposes $\mathbf{X}_t$ into a set of Q common factors and factor loadings. Here $d\mathbf{Z}(\omega)$ is an Q dimensional mean-zero orthogonal incremental process with unit variance, $\boldsymbol{\epsilon}$ is a P dimensional white noise term, and $\Lambda(\omega)$ is a $P \times Q$ loading matrix at frequency $\omega \in \mathbb{R}$.

The power spectrum is then given by

$$f(\omega) = \Lambda(\omega)\Lambda(\omega)^* + \Sigma_\epsilon \tag{3.2}$$

where Λ^* is the conjugate transpose of Λ and Σ is a $P \times P$ covariance matrix of $\boldsymbol{\epsilon}$. The spectral matrix $f(\omega)$ then becomes a complex-valued Hermitian matrix of dimension $P \times P$ at every frequency ω (Li et al., 2019; Ensor, 2013; Macaro and Prado, 2014).

The advantage of this structure is it allows for an infinite number of factors with quickly decaying loadings. This is accomplished by using a tensor product penalized spline and multiplicative gamma process shrinkage prior which is robust against the choice of number of factors. By choosing $Q < P$ the curse of dimensionality is mitigated somewhat allowing for estimation of high dimensional spectra.

To map $\mathbf{X}_t$ into the frequency domain, the discrete Fourier transform at each frequency k is used such that

$$\mathbf{Y}_k = T^{-1/2} \sum_{t=1}^{T} \mathbf{X}_t \exp(-2\pi i \omega_k j), \text{ for } k = 1, ..., K,$$

where $\omega_k = k/T$ are the Fourier frequencies and $K = \lfloor (T-1)/2 \rfloor$. By some regularity conditions, each $\mathbf{Y}_k$ are independently distributed complex normal such that $\mathbf{Y}_k \sim CN(\mathbf{0}, f(\omega_k))$ (Brillinger, 2002).

Following Li et al. (2019), we let $\mathbf{D}_k = \int_0^{\omega_k} \mathbf{Z}(\omega)\, d\omega$, $\mathbf{E_k} = T^{-1/2} \sum_{t=1}^{T} \epsilon_t \exp(-2\pi i \omega_k t)$, and $\Lambda_k = \Lambda(\omega_k)$. This leads us to

$$\mathbf{Y}_k \approx \Lambda_k \mathbf{D}_k + \mathbf{E}_k,$$

with $\mathbf{D}_k \sim CN(\mathbf{0}, I_Q)$ and $\mathbf{E}_k \sim CN(\mathbf{0}, \Sigma_\epsilon)$. Then the conditional Whittle likelihood for this series is given by

$$L(\mathbf{Y}|\Lambda, \mathbf{D}, \Sigma) \approx \prod_{k=1}^{K} \prod_{p=1}^{P} \left[\sigma_{\epsilon,p}^{-1} \exp\left(\sigma_{\epsilon,p}^{-1} |Y_{kp} - \Lambda_k^{(p)} \mathbf{D}_k|^2 \right) \right] \tag{3.3}$$

where $\mathbf{Y}$ is the discrete Fourier transforms of the time series such that Y_{kp} is the pth element of $\mathbf{Y}_k$. Λ and $\mathbf{D}$ denote collections of loadings and factors respectively. And $\sigma_{\epsilon,p}^2$ is the pth diagonal element of Σ_ϵ. We refer the interested reader to (Li et al., 2019) for the prior distributions of Λ and $\sigma_{\epsilon,p}$ for this model.

3.2.2 Likelihood for Multiple Multivariate Time Series

Now we wish to extend this model to include $j = 1, ..., N$ independent multivariate time series. Let $\mathbf{X}_{jt} = (X_{j,1,t}, ..., X_{j,P,t})$ for $t = 1, ..., T$ and $j = 1, ..., N$ be the P dimensional multivariate time series for the jth subject at time t. Similar to equation 3.1 above, this time series has a factor model representation given by

$$\mathbf{X}_{jt} = \int_0^1 \Lambda(\omega) \exp(2\pi i \omega t) d\mathbf{Z}_j(\omega) + \boldsymbol{\epsilon}, \tag{3.4}$$

Here we assume that all N time series share a loading matrix $\Lambda(\omega)$, with individual factors and independent identically distributed errors. Following the same methodology as above and Fourier transforming the $\mathbf{X}_{jt}$'s into the frequency domain gives

$$\mathbf{Y}_{jk} \approx \Lambda_k \mathbf{D}_{jk} + \mathbf{E}_k$$

where $\mathbf{D}_{jk}$, $\mathbf{E}_k$, and Λ_k have the same distributions as in Section 3.2.1. It then follows that the conditional Whittle likelihood for $\mathbf{Y}$ is

$$L(\mathbf{Y}|\Lambda, \mathbf{D}, \Sigma) \approx \prod_{j=1}^{N} \prod_{k=1}^{K} \prod_{p=1}^{P} \left[\sigma_{\epsilon,p}^{-1} \exp\left(\sigma_{\epsilon,p}^{-2} |Y_{jkp} - \Lambda_{kp} \mathbf{D}_{jk}|^2 \right) \right] \tag{3.5}$$

Here the notation is the same as equation 3.3, but includes an extra subscript j to account for more than one independent time multivariate series.

The prior that is chosen for $\sigma_{\epsilon,p}$ is an independent Half-t(ν, G_ϵ) where the hyperparameters are known constants (Gelman, 2006). Often this distribution is modeled using a mixture of inverse gamma distributions $IG(a,b)$, with densities $p(x) \propto x^{-(a+1)} \exp(-b/x)$, $x > 0$ (Wand et al., 2012). Here we use $p(\sigma_{\epsilon,p}^2|g_{\epsilon,p}) \sim IG(\nu/2, \nu/g_{\epsilon,p})$ and $p(g_{\epsilon,p}) \sim IG(1/2, 1/G_{\epsilon,p}^2)$.

The prior for Λ allows for the loadings to decay for large P, thus most of the information exists in the first few factors. The real and imaginary parts of the entires are modeled using tensor products of Bayesian penalized splines (Krafty et al., 2017; Li et al., 2019) and multiplicative gamma process shrinkage priors (Bhattacharya and Dunson, 2011). Here the first S Demmler-Reinsch basis functions are used such that

$$Re(\Lambda_{pq}(\omega)) = \alpha_{pq0} + \sum_{s=1}^{S-1} \alpha_{pqs}\sqrt{2}\cos(2\pi s\omega) \tag{3.6}$$

$$Im(\Lambda_{pq}(\omega)) = \sum_{s=1}^{S} \beta_{pqs}\sqrt{2}\sin(2\pi s\omega). \tag{3.7}$$

Krafty and Collinge (2013) shows that these basis functions improve performance by restricting the functions of the real parts of the matrix to be even and the imaginary parts to be odd. Conditioned on a smoothing parameter τ^2, independent $N(0, [2\pi s]^{-1}\tau^2)$ priors are placed on the coefficients. To concentrate the information in the first q factors, we follow the methodology of Li et al. (2019) and use a prior that is a tensor product of the penalize spline prior and a gamma process shrinkage prior. The prior distributions for the spline coefficients then become

$$\begin{aligned}
\alpha_{pq0} &\sim N\left(0, \psi_{q(re)}^{-1}\right), \alpha_{pqs} \sim N\left(0, \psi_{q(re)}^{-1}\tau_{pq(re)}^2(2\pi s)^{-1}\right) \text{ for } s = 1, ..., S-1 \\
\beta_{pqs} &\sim N\left(0, \psi_{q(im)}^{-1}\tau_{pq(im)}^2(2\pi s)^{-1}\right) \text{ for } s = 1, ..., S
\end{aligned}$$

And the prior distributions for the smoothing and shrinkage parameters are

$$\tau_{pq(.)} \quad \sim \quad \text{Half-}t(\nu, G_\tau^2)$$

$$\phi_{1(.)} \quad \sim \quad G(a_1, 1) \text{ and } \phi_{q(.)} \sim G(a_2, 1) \text{ for } q \geq 2$$

where a_1 and a_2 are fixed constants, $(.)$ denotes (re) or (im), and $\psi_{q(.)} = \prod_{h=1}^{q} \phi_{h(.)}$.

The choice of these priors allows for two computational advantages. First the shrinkage parameters are increasing for $a_2 > 1$, allowing for an infinite number of factors with decaying loadings. This provides a good representation of the multivariate spectrum and reduces sensitivity to the number of factors chosen. Second we only require $S < K$ basis functions for the Bayesian smoothing spline, where K would be full rank. This allows us to simplify the computation without introducing error into the fit (Krafty et al., 2017; Li and Krafty, 2019).

3.2.3   Introducing Covariates via a Bayesian Mixture Model

Suppose now that $\mathbf{X}_j$ for $j = 1, ..., N$ has an M dimensional vector of associated covariates $\mathbf{v}_j = (v_1, ..., v_M)'$. We incorporate $\mathbf{v}_j$ into the estimate of the spectral density matrix via a finite Bayesian mixture model with covariate-dependent mixture weights. The overall likelihood for $\mathbf{Y}_j$ is given by

$$L(\mathbf{Y}_j|\mathbf{v}_j, \Lambda, \Sigma_\epsilon, \mathbf{D}_j) \propto \sum_{h=1}^{H} \pi_h(\mathbf{v}_j) L_h(\mathbf{Y}_j|\Lambda^{(h)}, \Sigma_\epsilon^{(h)}, \mathbf{D}_j^{(h)}) \tag{3.8}$$

where $h = 1, ..., H$ denotes the mixture component, $\Lambda^{(h)}, \Sigma_\epsilon^{(h)}, \mathbf{D}^{(h)}$ are the parameters associated with with hth component, and $0 \leq \pi_h(\mathbf{v}_j) \leq 1$ and $\sum_{h=1}^{H} \pi_h(\mathbf{v}_j) = 1$ (Bertolacci et al., 2019; Gelman et al., 2014).

To aid in the estimation of this model, latent variables $z_j \in \{1, ..., H\}$ are introduced such that

$$(\mathbf{Y}_j|z_j = h) \sim L_h(\mathbf{Y}_j|\Lambda^{(h)}, \Sigma_\epsilon^{(h)}, \mathbf{D}^{(h)}),$$

where L_h is the conditional Whittle likelihood given in Equation 3.5. This formulation allows us to estimate the parameters for component h using the $\mathbf{Y}_j$'s that are assigned to the hth component by z_j using the methodology previously recorded in Section 3.2.2.

There exist many methods to model the mixture weight $\pi_h(\mathbf{v})$ such as the logit stick-breaking prior of Rigon and Durante (2017) used by Bertolacci et al. (2019) in their spectral density mixture model. Cadonna et al. (2019) model frequency weights using logistic regression. Here the mixture weights are modeled using a multinomial logistic regression. Component membership $z_{j'} = (z_{j'1}, ..., z_{j'H})'$ where $z_{j'h} = I(z_j = h)$ is regressed on covariates $\mathbf{v}_j$ to estimate an $m \times H$ coefficient matrix γ. Then

$$\pi_h(\mathbf{v}_j) = \frac{\exp(\mathbf{v}_j \gamma_h)}{\sum_{h=1}^{H} \exp(\mathbf{v}_j \gamma_h)} \tag{3.9}$$

where γ_H has been constrained to 0 to maintain identifiability.

Classically a Metropolis-Hastings algorithm has been used to estimate γ for regression models with binomial likelihoods. However, more recently Polson et al. (2013) showed that γ can be sampled using a Gibbs sampler formulation by introducing a data-augmentation scheme which incorporates a Polya-Gamma distributed latent variable. A multivariate normal prior is placed on γ_h so that $\gamma_h \sim MVN(\mathbf{m}_{0h}, \mathbf{V}_{0h})$ and γ_h is sampled conditional on η_{ih} where η_{ih} is distributed as a Polya-Gamma. For more details see Appendix section A.2.

3.2.4 Sampling Scheme

Let $\mathbf{z} = (z_1, ..., z_N)$, $\boldsymbol{\gamma} = \gamma_1, ..., \gamma_H$, $\Theta_h = \{\Lambda^{(h)}, \Sigma_\epsilon^{(h)}, \mathbf{D}_j^{(h)}\}$, $\Theta = \{\Theta_1, ..., \Theta_h\}$. Initialize all parameters. Then the sampling scheme is as follows:

1. Draw $(\Theta_h | z, \mathbf{Y})$ for $h = 1, ..., H$ using the methodology in section A.2.1.
2. Draw $(z | \boldsymbol{\gamma}, \Theta, \mathbf{Y})$ for $j = 1, ..., N$ using the methodology in A.2.3.
3. $(\gamma_h | z)$ for $h = 1, ..., H$ using the Polya-Gamma augmentation mentioned earlier in section.

More specifics of the posterior distributions can be found in A.2.4.

3.2.4.1 Label Swapping
One issue to make note of in any mixture model setting is label swapping (Gelman et al., 2014). In the analysis for Section 3.3 and Section 3.4, we are interested in the total estimate of the power spectrum not in examining individual components directly. However, if one wanted to perform inference using individual mixture components an order criteria could be imposed as fourth step in the sampling scheme which would order the components by Frobenius norm of $\mathbf{\Lambda}^{(h)}$.

3.3 Simulations

3.3.1 Simulation Scenario

To assess the performance of this method we assumed that N multivariate time series were randomly sampled from two groups with basis functions $x_t^{(1)}$ and $x_t^{(2)}$ where

$$x_t^{(1)} = \epsilon_t + \Phi_{11}\epsilon_{t-1} + \Phi_{12}\epsilon_{t-2}$$

$$x_t^{(2)} = \epsilon_t + \Phi_{21}\epsilon_{t-1} + \Phi_{22}\epsilon_{t-2}.$$

The matrices Φ are $P \times P$ block diagonal such that $\Phi_{11} = I_{P/3} \otimes \Phi_{11}^\dagger$, $\Phi_{12} = I_{P/3} \otimes \Phi_{12}^\dagger$, $\Phi_{21} = I_{P/3} \otimes \Phi_{21}^\dagger$, and $\Phi_{22} = I_{P/3} \otimes \Phi_{22}^\dagger$. Here $\otimes$ indicates the kronecker product, $I_{P/3}$ is a $P/3 \times P/3$ identity matrix, and

$$\Phi_{11}^\dagger = \begin{bmatrix} 0.7 & 0 & 0 \\ 0.2 & -0.6 & 0 \\ 0.1 & 0.2 & 0.6 \end{bmatrix}, \Phi_{21}^\dagger = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.2 & 0.6 & 0 \\ -0.1 & -0.2 & -0.6 \end{bmatrix}$$

$$\Phi_{12}^\dagger = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \Phi_{22}^\dagger = \begin{bmatrix} -0.3 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The noise term $\epsilon_t \sim N_P(\mathbf{0}, \Omega)$ where $\Omega = I_{P/3} \otimes \Omega^\dagger$. $\Omega^\dagger$ denotes a 3×3 matrix with 1s on the diagonal and 0.5s on the off diagonal. Then the time series X_{jt} for $j = 1, ..., N$ are simulated as

$$X_{jt} = \begin{cases} 0.4x_t^{(1)} + 0.6x_t^{(2)} & \text{if } v_j = 0 \\ 0.6x_t^{(1)} + 0.4x_t^{(2)} & \text{if } v_j = 1 \end{cases}$$

where $v_j \sim BER(.5)$ indicates group membership and is the covariate of interest in the simulation.

The true power spectrum $f_j(\omega)$ for $j = 1, ..., N$ is then given by

$$f_j(\omega) = \begin{cases} 0.4^2 f^{(1)}(\omega) + 0.6^2 f^{(2)}(\omega) & \text{if } v_j = 0 \\ 0.6^2 f^{(1)}(\omega) + 0.4^2 f^{(2)}(\omega) & \text{if } v_j = 1 \end{cases}$$

where $f^{(b)}(\omega)$ for $b = 1, 2$ is the power spectrum for the bth basis function, and

$$f^{(b)}(\omega) = \Phi^{(b)}(\omega)\Omega\Phi^{(b)}(\omega)^*$$

with $\Phi(\omega)^{(b)} = I + \Phi_{b1}\exp(-2\pi i\omega) + \Phi_{b2}\exp(-4\pi i\omega)$.

3.3.2 Performance Metrics

We seek to quantify predication accuracy and performance in estimating association of covariate v with differences in the power spectrum.

Predication accuracy is measured using mean squared error (PMSE) obtained by averaging the MSE for each frequency across all levels of covariates. In this case since the possible values of the covariates are $v = 1$ and $v = 0$ the formula is given by

$$\text{PMSE} = \frac{\sum_{i=1}^{2} \sum_{k=1}^{K} ||\hat{f}_i(\omega_k) - f_i(\omega_k)||_F^2}{2K}$$

where $||.||_F$ is the Frobenius norm and $\hat{f}_i(\omega_k)$ is the estimated spectral density for $v = i$ at Fourier frequency k.

To assess whether the method recovers the true effect of covariate v we use the MSE of the difference between $v = 1$ and $v = 0$.

$$\text{EMSE} = \frac{\sum_{k=1}^{K} ||\widehat{\Delta f}(\omega_k) - \Delta f(\omega_k)||_F^2}{K}$$

where $\widehat{\Delta f}(\omega_k)$ is the difference in spectral density for $v = 0$ and $v = 1$ at frequency k, namely

$$\widehat{\Delta f}(\omega_k) = \hat{f}_{v=1}(\omega_k) - \hat{f}_{v=0}(\omega_k)$$

3.3.3 Comparison

There are no existing methods which provide a covariate dependent estimate of the entire power spectrum for multivariate time series. However, we seek to compare our method to an ad hoc procedure a clinician could feasibly use to provide a covariate based spectral density estimate.

First the multivariate power spectrum is estimated individually on each time series using a periodogram smoothed with a Daniell kernel. Then the matrix log of the $P \times P$ spectral density matrix at each frequency for each subject is calculated. Each diagonal and off diagonal element from the power spectrum from each time series is gathered into an $N \times k$ matrix. Then multivariate regression is performed and the estimated spectral density for the ith jth component is calculated as

$$\hat{\mathbf{Y}}_{ij} = \mathbf{V}\boldsymbol{\beta}$$

41

Table 3.1: Simulation results comparing mixture model to regression model.

	Mixture Model		Regression	
	PMSE	EMSE	PMSE	EMSE
n=40, p=9, t=500	0.18 (0.02)	0.22 (0.04)	0.54 (0.02)	0.26 (0.01)
n=80, p=9, t=500	0.15 (0.02)	0.16 (0.03)	0.52 (0.01)	0.20 (0.01)
n=40, p=18, t=500	0.63 (0.35)	0.78 (0.97)	2.22 (0.04)	1.06 (0.03)
n=40, p=9, t=1000	0.17 (0.01)	0.22 (0.03)	0.53 (0.01)	0.23 (0.01)

Results for simulations. PMSE = predicted mean squared error. EMSE = effect mean squared error. Reported as mean (sd) over 100 simulated data sets.

where $\hat{\mathbf{Y}}_{ij}$ is the estimated spectral density, $\boldsymbol{\beta}$ is the coefficient matrix obtained by regressing $\mathbf{Y}$ on $\mathbf{V}$. Then the spectrum is averaged across all $v = 0$ and $v = 1$ to create the ith jth component of the estimated spectrum of $v = 1$ and $v = 0$. Then the final step involved piecing the resulting spectral densities back together and taking the matrix exponential of each $P \times P$ matrix at each frequency.

The performance of this method is compared to the performance of the Bayesian mixture model on 100 datasets generated from the process described in 3.3.1. For these simulations we set the number of smoothing splines $S = 10$, the number of possible mixture components $H = 10$, the number of factors $Q = \lceil P/2 \rceil$.

3.3.4 Simulation Results

The PMSE and EMSE generated by the simulations using the Bayesian mixture model and the regression comparison method are reported in Table 3.1.

Our method outperforms the regression in every setting, both in estimating the value of the spectral density matrix and also in estimating the difference between the estimate of $v = 0$ and $v = 1$. This is especially true in the case when $P = 18$ as the prediction error for the regression is

nearly four times higher than that of the mixture model. This reflects the difficulties of estimating high dimensional spectral matrices. In the other settings the PMSE is a little over twice that of the mixture model. The value of EMSE is closer when comparing the two methods, but the mixture model again proves to be more accurate at estimating the true distance between the spectral matrices for $v = 0$ and $v = 1$.

Focusing specifically on the mixture model, the PMSE is generally lower than EMSE suggesting that the method is slightly better at estimating predicted power spectra than difference in two power spectra. However, the numbers are still comparable suggesting that its a robust method in both settings. Comparing to the $n = 40, p = 9, t = 500$ setting, we see that the method performs better when n or t increases since more information is being added to the model. However, the accuracy decreases when the dimension doubles, which again reflects the difficulties of high dimensional problems.

Figure 3.1 plots the estimated diagonal and off diagonal spectral density components for $v = 0$ and $v = 1$ against the predicted values for one data set in the $n = 80, p = 9, t = 500$ setting. In components where the spectral matrix was less similar between $v = 0$ and $v = 1$, such as the imaginary parts of the cross spectra, the recovered values more tightly overlay the truth. In components where the spectral matrix is more similar the method still shows good separation between the different covariate values.

3.4   Application: High Density EEG

Previous research has shown that there are intrinsic differences in the activity of healthy and schizophrenic brains. Because resting state EEG is measured when the brain is not focusing on a specific task, it is often utilized to quantify these differences (Khanna et al., 2015). Typically power in traditionally defined frequency bands is assessed. Delta waves (< 5 Hz) are typically associated with deep relaxation and sleep. Theta waves (5-7 Hz) tend to be most active during daydreaming and sleep. Alpha waves (7-15 Hz) are usually the most activated during a restful activity while awake. And beta waves (15-30 Hz) are associated with active thinking and concentration.
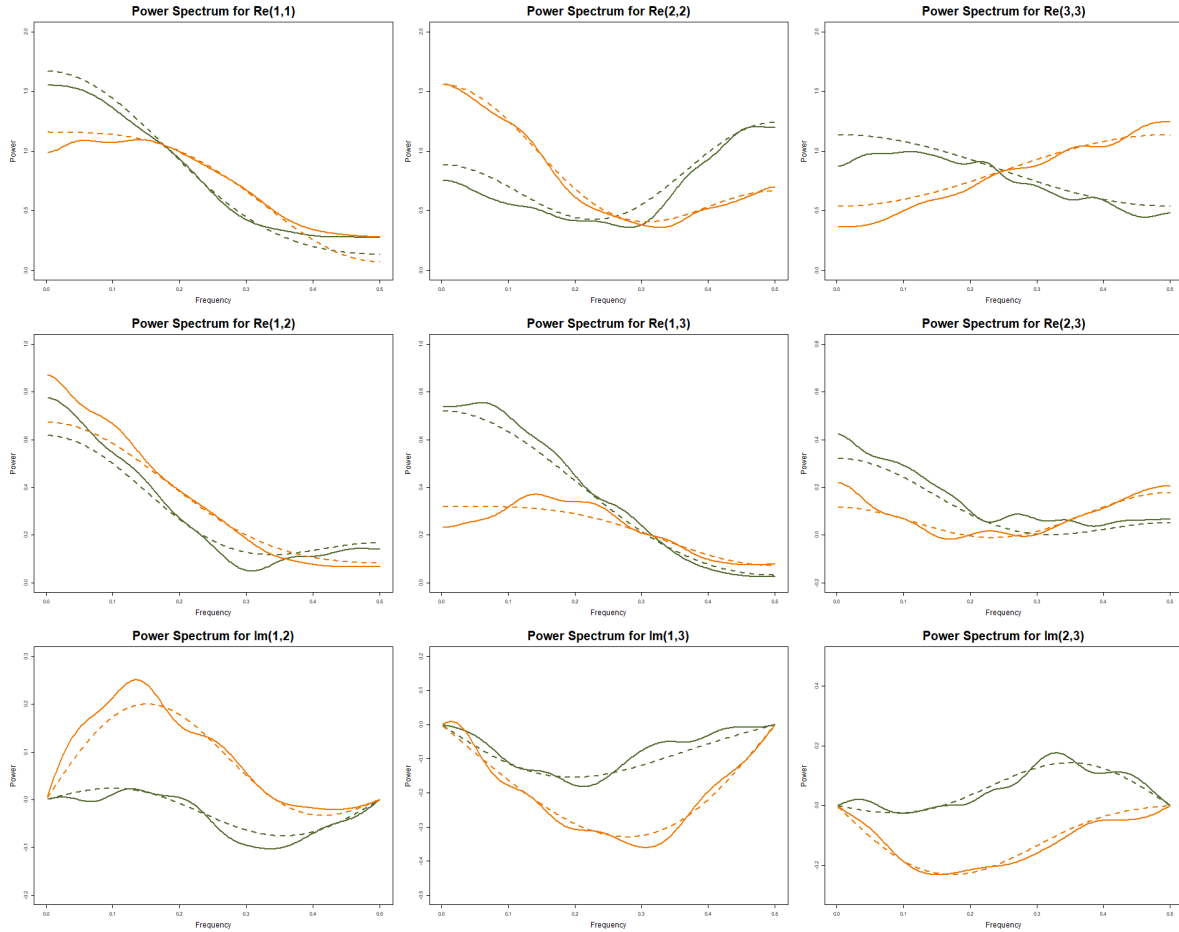
Figure 3.1: True and recovered values from one run of simulation.

Estimated power spectrum for first three components from one simulated data set from the $n = 80, p = 9, t = 500$ scenario. Orange is estimate for $v = 1$ and green is for $v = 0$. The solid line is the recovered estimate and the dashed line is the true power spectrum. Top row: Power spectrum for first three components, $f_{1,1}(\omega)$, $f_{2,2}(\omega)$, and $f_{3,3}(\omega)$. Middle row: Real parts of the cross spectra $f_{1,2}(\omega)$, $f_{1,3}(\omega)$, and $f_{2,3}(\omega)$. Bottom row: Imaginary parts of the cross spectra $f_{1,2}(\omega)$, $f_{1,3}(\omega)$, and $f_{2,3}(\omega)$.

44

Research comparing schizophrenic subjects to healthy controls has shown a variety of results, likely due to small sample sizes. However, a recent review paper by Newson and Thiagarajan (2019) reports that the most consistent findings are lower power in the alpha frequency band for schizophrenic patients and higher power in the delta and theta bands. This pattern of brain activity is indicative of an inappropriate arousal state and leads to a reduced ability to focus on relevant information. Interestingly some studies found that this trend held for the frontal part of the brain, but was flipped for the posterior (Begic et al., 2019; Kim et al., 2015).

Another important aspect of EEG data analysis is looking at brain connectivity. Typically this is measured using coherence between electrodes. In review papers focused on coherence between electrode signals, researchers found that subjects with schizophrenia have increased coherence compared with controls (French and Beumont, 1984). However, there was some conflicting results regarding frontal coherence in the delta and alpha frequency bands. Some studies reported increased coherence while other studies reported decreased coherence (Leocani and Comi, 1999)

We use data gathered on first episode psychosis (FEP) and health control (HC) subjects to demonstrate the effectiveness of the proposed method in quantifying the differences in the spectral density between FEP and HC.

3.4.1   Study Data Processing

The total sample size of this data set was $N = 43$, with 17 of these being HC and 26 being FEP. The FEP subjects were recruited from the emergency department of the Western Psychiatric Hospital of the University of Pittsburgh Medical Center while the HC were gathered from the community. All participants were ages $18 - 40$ and had no illness or developmental disorder that could potentially impact the central nervous system. EEG data was recorded using a 64-channel high-density electroencephalography net using the 10-10 system while the subjects were told to sit in a chair and relax. Signals were initially sampled at a rate of 250 Hz.

The data was then processed using MATLAB. Signals were down-sampled to 64 Hz and filtered using a 1 Hz high-pass filter and a 58 Hz low-pass filter to isolate frequencies of interest. Segments with large artifacts such as muscle activity or movement were removed from the recording by a trained EEG data manager. Data was then referenced to the average of the channels, and

more subtle artifacts such as ocular movement and cardiac signals were removed via independent component analysis in EEGLAB (Delorme and Makeig, 2004).

3.4.2 Analysis Overview

The Bayesian mixture model method was fit to the data set using $H = 10$ components, $S = 10$ basis functions, and $Q = 12$ factors. It was run for a total of $10,000$ iterations with a burn in of $2,500$ iterations. In this section we specifically focus on four electrodes: FC5, which is on the frontal left side of the brain, FC6, which is on the frontal right side of the brain; CP5 which is on the posterior left side of the brain; and CP6, which is on the posterior right side of the brain. We specifically choose electrodes from different hemispheres and different regions to see whether we could recover the patterns in the spectral and coherence features of the electrodes reported in the literature.

To assess spectral patterns in the electrodes, we plot the difference in power between HC and FEP as well as 95% credible intervals. This allows us to get an idea of frequency of the spectra which might be different between cases and controls. Additionally, we report frequency band collapsed values

$$\int_a^b f_{i,i}(\omega)\,d\omega$$

where a and b are the upper and lower range of the frequency band of interest and i indicates either FC5, FC6, CP5, or CP6. This quantity and a 95% credible interval is reported for both HC and FEP.

To assess coherence between electrodes, we plot the differences in pairwise squared coherences given by

$$\frac{|f_{ij}(\omega)|^2}{f_{ii}(\omega)f_{jj}(\omega)}$$

for $i \neq j$ where i, j indicates either FC5, FC6, CP5, or CP6. We also report frequency band collapsed values for squared coherences and the 95% credible interval for FEP and HC as

$$\frac{|\int_a^b f_{ij}(\omega)\,d\omega|^2}{\int_a^b f_{ii}(\omega)\,d\omega \int_a^b f_{jj}(\omega)\,d\omega}$$

where again a and b are the upper and lower range of the frequency band of interest.
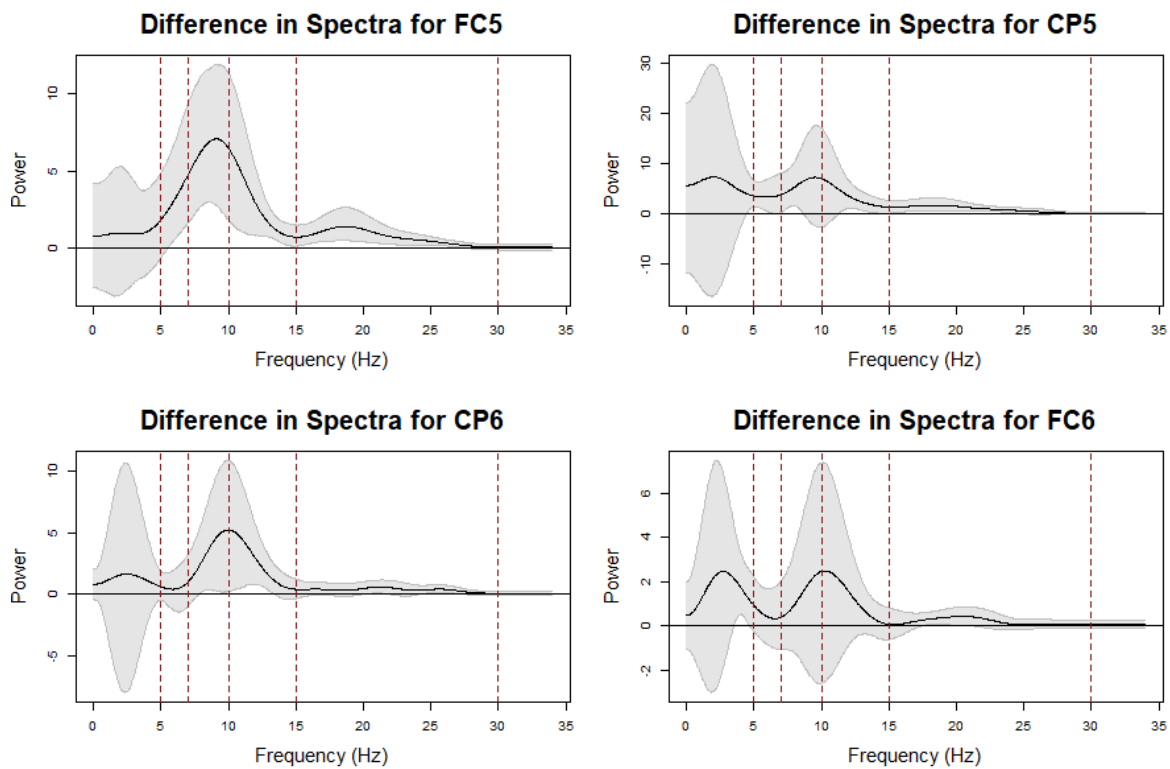
Figure 3.2: Power spectrum: estimated differences.

Plot of estimated difference in power spectrum for four selected electrodes as well as the 95% credible interval. Frequency bands are denoted as dotted red lines, delta, theta, low alpha, high alpha, beta.

3.4.3 Results

Plots of the mixture components along with their respective mixture weights for FEP and HC can be found in Appendix A.3.1.

3.4.3.1 Spectral Analysis of Individual Electrodes Figure 3.2 shows plots of the trajectory and corresponding 95% credible intervals of the difference in the spectrum for HC and FEP across frequency for the four electrodes of interest. Regions where the difference is greater than zero mean that the power is higher for HC and regions where the difference is less than zero mean that the power is higher for FEP. Frequency bands are denoted by dashed lines. The FC5 electrode showed the most difference between HC and FEP, with significantly decreased power for FEP in the theta, low/high alpha, and beta bands. CP5 and CP6 also showed decreased power for FEP in portions of the low and high alpha band. FC6 and CP5 showed decreased power for FEP in the higher frequencies of the delta band.

While we were able to replicate the finding of decreased alpha power for schizophrenic patients, we did not see the trend of increased delta and theta power. This could be due to several of the studies in the review paper relying on subjects taking antipsychotic medication, which impacts the generation of delta and theta band activity (Bochkarev et al., 2017). We also did not see a difference in the pattern of the frontal and posterior spectral density, with both regions exhibiting either lower power across frequency or no difference. Plots of all the spectral density for all electrodes can be found in Appendix A.3.2.

Table 3.2 shows the frequency collapsed measures for HC and FEP and the associated 95% credible intervals. Interestingly, when examining the frequency collapsed metrics the only significant differences appeared in the beta band of FC5 and CP5. This is likely a result of examining covariate effect in Figure 3.2 vs prediction based on the in Table 3.2. Since frequency collapsed measures are most often what is compared in the literature, this makes for an interesting dichotomy which illustrates a strength of using a mixture model for this type of comparison- the mixture model allows for an examination of frequencies of difference across the entire trajectory of the power spectrum without relying on measures which summarize the frequency band.

Table 3.2: Power spectrum: frequency band collapsed measures.

		Delta	Theta	Low Alpha	High Alpha	Beta
FC5	HC	21.2 (19.3–22.6)	13.7 (11.1–17.4)	34.7 (29.4–41.4)	21.4 (19.2–24.1)	3.6 (3.3–4.1)
	FEP	20.2 (17.7–22.0)	10.5 (9.2–12.1)	28.3 (22.9–34.7)	18.6 (15.6–21.9)	2.9 (2.6–3.1)
CP5	HC	65.6 (53.2–75.3)	22.9 (20.9–26.0)	52.2 (43.6–62.0)	28.2 (25.9–31.1)	4.8 (4.4–5.4)
	FEP	59.8 (70.0–48.1)	19.5 (16.8–21.7)	46.3 (36.1–57.5)	24.8 (20.9–28.8)	3.9 (3.5–4.3)
CP6	HC	30.4 (25.6–34.7)	14.3 (13.7–15.0)	33.2 (30.7–36.4)	21.0 (20.2–22.0)	4.0 (3.8–4.1)
	FEP	29.2 (24.9–33.3)	13.7 (12.0–15.5)	29.8 (25.2–34.2)	18.5 (16.0–20.5)	3.7 (3.3–3.9)
FC6	HC	23.4 (21.1–25.4)	11.7 (10.8–13.5)	26.4 (21.6–35.6)	17.9 (15.5–21.6)	3.1 (3.1–3.2)
	FEP	21.8 (19.1–24.2)	11.3 (9.8–13.1)	25.0 (19.2–34.1)	16.6 (13.7–20.4)	2.9 (2.7–3.1)

Absolute power in collapsed frequency bands for power spectra for four selected electrodes. Reported as estimate and 95% credible interval.

3.4.3.2 Coherence Between Electrodes Figure 3.3 illustrates plots of the coherence between the electrodes of interest. Similar to Figure 3.2, regions below the x axis indicate higher coherence in FEP while regions above the x axis indicate higher coherence in HC. Dashed lines denote boundaries of traditionally defined frequency bands.

In general, interhemispheric coherence was higher across all frequency bands for FEP patients except between the two frontal electrodes. These showed no difference except for a small range of frequencies in the beta band. We also found decreased coherence within the left hemisphere for FEP patients, and no difference in coherence within the right hemisphere. This finding is consistent with the idea that schizophrenia is a disorder of connectivity within the brain (Lynall et al., 2010) as well as some of the findings within French and Beumont (1984). A notable deviation from the patterns discussed in French and Beumont (1984) is the coherence between electrodes within the left hemisphere. Here FEP had decreased power compared to HC. This poses some questions regarding inter vs intra hemispheric coherence as well as left vs right side of brain in FEP and HC.
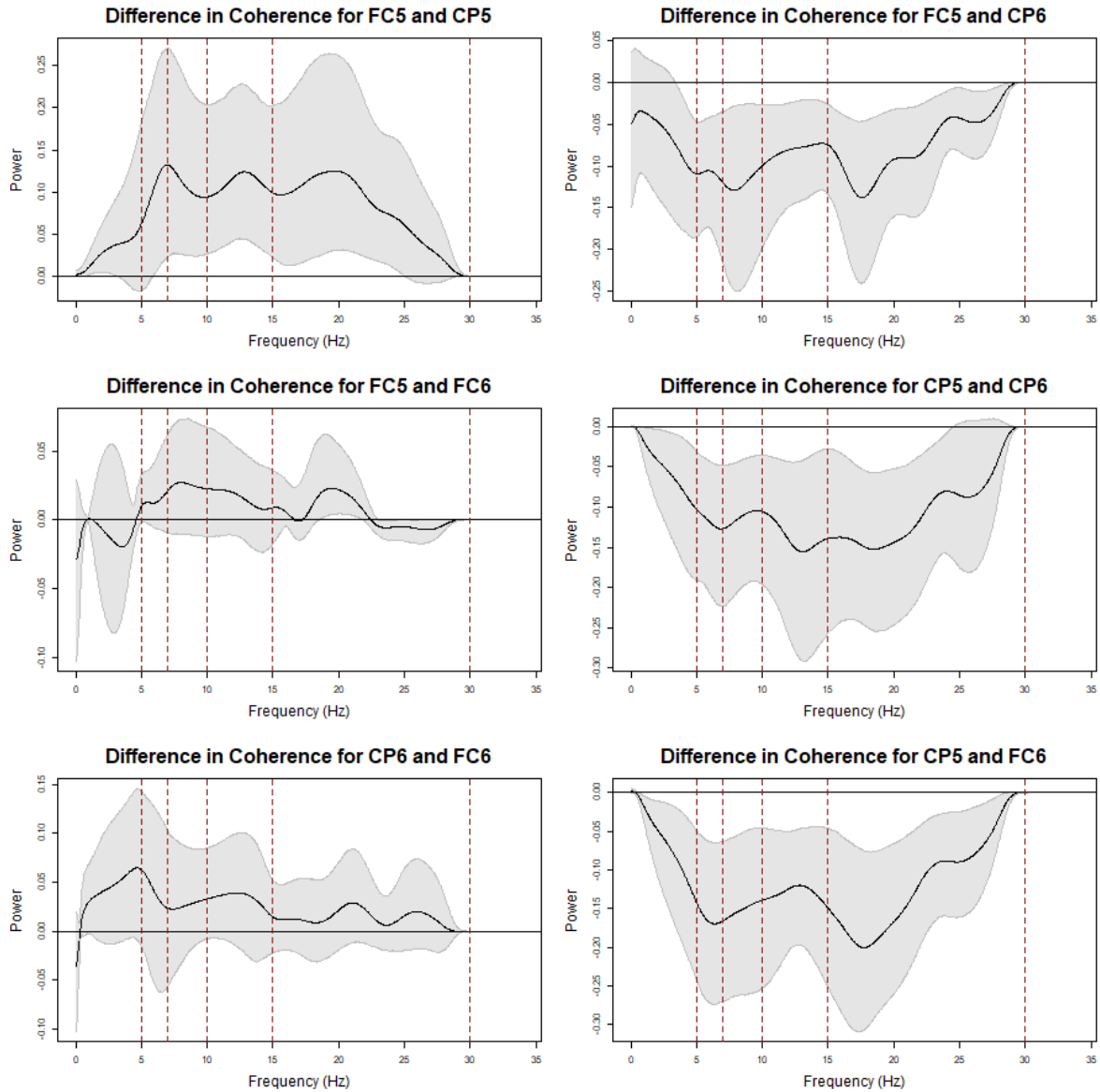
Figure 3.3: Coherence: estimated differences.

Plot of estimated difference in coherence between four selected electrodes as well as the 95% credible interval. Frequency bands are denoted as dotted red lines, delta, theta, low alpha, high alpha, beta.

Table 3.3: Coherence: frequency band collapsed measures.

		Delta	Theta	Low Alpha	High Alpha	Beta
FC5-CP5	HC	0.11 (0.08–0.17)	0.23 (0.14–0.36)	0.22 (0.12–0.36)	0.27 (0.20–0.37)	0.19 (0.13–0.28)
	FEP	0.08 (0.07–0.10)	0.13 (0.09–0.17)	0.12 (0.07–0.17)	0.16 (0.12–0.19)	0.11 (0.11–0.12)
FC5-CP6	HC	0.24 (0.16–0.30)	0.16 (0.08–0.23)	0.25 (0.14–0.35)	0.15 (0.08–0.20)	0.12 (0.07–0.17)
	FEP	0.30 (0.26–0.34)	0.27 (0.24–0.30)	0.37 (0.32–0.40)	0.23 (0.20–0.26)	0.19 (0.18–0.21)
FC5-FC6	HC	0.04 (0.02–0.07)	0.02 (0.00–0.05)	0.07 (0.03–0.13)	0.11 (0.07–0.15)	0.01 (0.00–0.02)
	FEP	0.05 (0.03–0.07)	0.01 (0.00–0.03)	0.05 (0.01–0.10)	0.09 (0.06–0.13)	0.00 (0.00–0.01)
CP5-CP6	HC	0.15 (0.09–0.20)	0.14 (0.07–0.21)	0.15 (0.08–0.21)	0.23 (0.11–0.32)	0.13 (0.06–0.20)
	FEP	0.20 (0.17–0.23)	0.26 (0.23–0.31)	0.26 (0.22–0.32)	0.36 (0.34–0.39)	0.23 (0.21–0.27)
CP6-FC6	HC	0.38 (0.34–0.42)	0.28 (0.21–0.35)	0.21 (0.16–0.27)	0.24 (0.20–0.29)	0.19 (0.17–0.23)
	FEP	0.34 (0.30–0.38)	0.24 (0.19–0.29)	0.18 (0.15–0.22)	0.21 (0.17–0.24)	0.18 (0.17–0.20)
CP5-FC6	HC	0.14 (0.08–0.20)	0.12 (0.03–0.21)	0.13 (0.03–0.24)	0.11 (0.03–0.18)	0.10 (0.03–0.17)
	FEP	0.20 (0.17–0.23)	0.28 (0.24–0.33)	0.28 (0.23–0.34)	0.24 (0.20–0.28)	0.21 (0.19–0.24)

Absolute power in collapsed frequency bands for coherences between four selected electrodes. Reported as estimate and 95% credible interval.

Table 3.3 shows the power in collapsed frequency bands for FEP and HC. In the delta frequency bands, no significant difference was found between the two groups for any coherence measure. In the theta band FEP had a higher interhemisphere coherence, with the electrodes affected being FC5-CP6 and CP5-CP6. In the low alpha band, FEP had higher coherence in the CP5-CP6 electrode pair. In the high alpha frequency band, FEP had lower coherence in FC5-CP5 and higher coherence in the FC5-CP6, CP5-CP6, and CP5-FC6 pairings. And in the beta band, FEP had lower coherence in FC5-CP5 and higher coherence in FC5-CP6 and CP5-CP6. This reflects similar, but less sensitive findings to what was reported for Figure 3.3 and has a similar pattern to the findings for the spectra from individual electrodes where the absolute power metrics are less sensitive to differences than examining the entire trajectory.

3.5 Discussion

In this chapter we developed a method to examine the characteristics of high dimensional spectral matrices dependent upon covariates of interest. This novel approach tackles the problem of high dimensionality by using a factor model to lower the number of parameters that need to be estimated, coupled with a prior that concentrates information into the first factors. This allows for stable estimates of the spectral density matrix as well as increased computational efficiency. Covariates are then introduced as weights in a Bayesian mixture model framework. This simplifies the problem into the estimation of several spectral basis functions which contribute to the overall estimate of the power spectrum through the covariate dependent weights.

As far as we know, this model is the first to deal with introducing covariates to the spectral estimate in a high dimensional multivariate setting. Previous methods have either focused on estimating high dimensional densities without covariates using such methodology as shrinkage, thresholding, or factor models (Schneider-Luftman and Walden, 2016; Sun et al., 2018; Ensor, 2013); or have provided covariate dependent estimates in the multiple time series or low dimensional multivariate setting (Cadonna et al., 2019; Bertolacci et al., 2019).

The main application of interest we were targeting was high density EEG. The advantages of this model over traditional methods to analyze EEG data are threefold. First, the method easily

outputs a smooth estimate of the spectral density for the entire trajectory across frequency for any combination of covariates a clinician might be interested in. This allows for easy identification of frequency dependent features. Second, traditional methodology involves using frequency band collapsed measures coupled with regression that requires the use of correction for multiple comparisons. Because our method exists in the Bayesian framework and we use 95% credible intervals, there is no need make such an adjustment. And third, our method is flexible enough to compare any function of the spectral density that might be of interest. This includes frequency band collapsed measures, differences in spectral densities, ect.

However, this method does not come without some drawbacks. First, implementing this method is computationally intense, requiring a long run time and large amount of storage. This can be alleviated somewhat by recoding the algorithm in packages such as `Rcpp` (Eddelbuettel and François, 2011) and implementing smart coding techniques to reduce storage requirements. And second, we rely on 95% credible intervals to conduct hypothesis testing which does not allow for an overall global test of covariate association with power spectrum. This also could be overcome by adding a step in the Gibbs sampler that would allow us to jump between a model with the covariate of interest and a model without and output the probability of one model over the other.

Future work would involve implementing both of these things, as well as looking at possible ways to extend the model. For instance, a hierarchical model could be implemented to model the mixture weights which would allow for increased subject-level variability. Other models of the mixture weights could also be explored, such a logit or probit stick breaking process. And finally, another natural extension would be to modify the model to accommodate nonstationary multivariate time series.

4.0 Discussion

In this dissertation we introduce novel methodologies to perform spectral analysis under a variety of different conditions. Specifically we were interested in creating an interpretable method for estimating the rank structure of a univariate nonstationary time series and also in deriving a method for covariate-dependent estimation of high dimensional multivariate spectral density matrices. Each of these new tools has wide variety of potential applications.

The SpeLLL algorithm differs from previous offerings of nonstationary spectral analysis by focusing specifically on recovering multiple low rank representations of the power spectrum. Penalized regression techniques introduce smoothness and sparsity into these estimates allowing for clearer interpretation of time and frequency features. Smoothed plots of the $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ vectors are especially helpful in visualizing these features, and also allow for accurate placement of temporal break points and frequency bands. As an added feature, information about either the time or frequency bins of the power spectrum can be incorporated into a design matrix $\mathbf{X}$. In the HRV application, we used sleep stage epochs to neatly summarize how the ANS functions throughout a night of sleep. However, it would also be possible to transpose the $\mathbf{Y}$ matrix and incorporate information about frequency bands in the design matrix. This is especially useful to clinicians who are are specifically interested in power at frequency bands pre-defined in the literature.

SpeLLL can be useful in any situation exploring nonstationary signals. Our application specifically involved tracking ECG of HRV throughout a night of sleep. But it could also be applied to other biological processes such as examining the frequency characteristics of an EEG electrode over time (Lim and Chia, 2015), or in analyzing data related to tremors in neurological diseases (de Lima et al., 200), or in extracting features related to pulse oximetery and a diagnosis of sleep apnea (Álvarez et al., 2009).

Further extensions of the SpeLLL algorithm could make it more flexible and more applicable to a wider variety of signals. For instance, a natural extension would be to extend the algorithm to handle multiple time series by using penalized tensor principal components analysis. Covariates could be incorporated by ordering the $\mathbf{Y}$ matrices from each subject by covariate level. Another area of future work would be to explore other types of penalties. A first step would be to incor-

porate separate lasso penalties on the $\mathbf{u}$ and $\mathbf{v}$ vectors. Another interesting addition would be to incorporate a group penalty to the regression (Simon et al., 2013). Adding a group penalty to $\mathbf{v}$ would be a natural way to incorporate information about pre-specified frequency bands and could be used in tandem with time-specific information added to the $\mathbf{X}$ matrix.

The SpeLLL algorithm has been coded in R, but further work needs to be done to submit it as a package to the Comprehensive R Archive Network (CRAN). The ADMM algorithm portion of it was coded in `Rcpp` (Eddelbuettel and François, 2011), but there are other functions that could be translated into C++ to facilitate faster convergence.

In the stationary high dimensional setting, our Bayesian mixture model is the only existing methodology which incorporates covariates into estimates of the spectral density matrix. Other methods focus on multiple time series (Cadonna et al., 2019; Bertolacci et al., 2019) which does not allow for an estimate of the coherence between two time series within the same subject. This is particularly important in EEG analysis where coherence is thought of as a measure of brain connectivity. By using penalized Bayesian splines the resulting estimates are smoothed across frequency, and the prior placed on the factor matrix allows us to reduce the dimension of the parameter space at each frequency to $P \times Q$ where $Q < P$. This improves the stability of the matrix and reduces computational complexity. Covariates are incorporated as weights in the mixture model. Each component of the mixture model can be thought of as a basis function and, provided the label swapping issue is dealt with, can also provide information about the power spectrum.

This method was specifically created to handle high density EEG signals, but could also be incorporated in any multivariate setting. For instance, it could be applied to look for covariate relationships with gait measurements in x, y, and z space (Henryk et al., 2019) or in looking at covariate-dependent frequency characteristics of infectious disease surveillance in many different areas at once (Pedeli and Karlis, 2019).

Future work in this area would include developing a robust method which would allow for a statistical test of the covariate of interest. A natural candidate is a Bayes factor. However this would require running the model twice and also becomes computationally unstable for longer series as the likelihood for $\mathbf{Y}$ involves multiplying PQN elements together. One way to deal with this would be to incorporate an indicator at each step to track which likelihood (covariate dependent or not) has a higher probability at each step. Another potential extension of this model is into the

nonstationary multivariate time series domain. This could be done by breaking the time series up into smaller segments and running the model on each segment, similar to Li et al. (2019).

The code for this method was written in R and incorporates sampling the coefficients of the $\boldsymbol{\Lambda}$ matrix in parallel to speed up computation. However, in order for the code to be usable by researchers further work needs to be done to decrease computation time. The accuracy of the likelihood for z_j could also be increased by using the `Rmpfr` package (Mächler, 2013) which would allow for arbitrary precision when calculating the likelihood of $\mathbf{Y}_j$ for each component.

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

<center>**Additional Methodological Details**</center>

A.1   ADMM for Adaptive Sparse Fused Lasso

Here we provide additional details for using the alternating direction method of multipliers (ADMM) algorithm (Boyd et al., 2010) to solve equations 2.7 and 2.8.

A.1.1   Theory

A general adaptive sparse fused lasso problem is given by minimizing

$$\frac{1}{2}||\mathbf{y} - \mathbf{Xb}||^2 + \lambda^{(1)} \sum_{i=1}^{n} w_i |b_j| + \lambda^{(2)} \sum_{i=2}^{n} m_i |b_i - b_{i-1}|. \tag{.1}$$

with respect to $\mathbf{b}$, where $\mathbf{y}$ is a vector of responses, $\mathbf{X}$ is a design matrix, and $\mathbf{w}$ and $\mathbf{m}$ are data driven weights. Here $\lambda^{(1)}$ and $\lambda^{(2)}$ are the tuning parameters for the lasso and fusion penalties respectively.

The minimization in Equation .1 can be rewritten as

$$\frac{1}{2}||\mathbf{y} - \mathbf{Xb}||^2 + \sum_{i=1}^{2n-2} \lambda_i^w |a_j| \text{ subject to } \mathbf{Db} - \mathbf{a} = 0. \tag{.2}$$

where $\boldsymbol{\lambda}^{(w)} = (\lambda_1 \mathbf{w}, \lambda_2 \mathbf{m})'$ and $\mathbf{D} = [\mathbf{I}_n | \mathbf{R}]'$, with $\mathbf{R}$ containing the graph for the fusion portion of the penalty such that

$$\mathbf{R} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \dots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix}.$$

<center>57</center>

Then the updating equations for the ADMM algorithm for the kth step with stepsize ρ are given by

$$
\begin{aligned}
\mathbf{b}^{(k)} &= (\mathbf{X}'\mathbf{X} + \rho\mathbf{D}'\mathbf{D})^{-1}(\mathbf{X}'\mathbf{y} + \rho\mathbf{D}'(\mathbf{a}^{(k-1)} - \mathbf{z}^{(k-1)})), \\
\mathbf{a}^{(k)} &= S_{\boldsymbol{\lambda}^w/\rho}(\mathbf{D}\mathbf{b}^{(k)} - \mathbf{z}^{(k-1)}). \\
\mathbf{z}^{(k)} &= z^{(k-1)} + \mathbf{D}\mathbf{b}^{(k)} - \mathbf{a}^{(k)}
\end{aligned}
$$

Here $S_{\boldsymbol{\lambda}^w/\rho}(\mathbf{D}\mathbf{b}^{(k)} - z^{(k-1)})$ is the soft-thresholding operator given by $S_j(k) = \max(0, k - j) - \max(0, -k - j)$ and is applied element-wise to the vectors $\boldsymbol{\lambda}^w/\rho$ and $\mathbf{D}\mathbf{b}^{(k)} - \mathbf{z}^{(k-1)}$.

The algorithm converges when $||s_1|| < \epsilon_1$ and $||s_2|| < \epsilon_2$ where

$$
\begin{aligned}
s_1 &= \rho\mathbf{D}'(\mathbf{a}^{(k)} - \mathbf{a}^{(k-1)}) \\
s_2 &= (\mathbf{D}\mathbf{b}^{(k)} - \mathbf{a}^{(k)})
\end{aligned}
$$

and ϵ_1, ϵ_2 are pre-specified error tolerances.

A.1.2   Rcpp Code

The ADMM algorithm was implemented using the `Rcpp` package (Eddelbuettel and François, 2011). By translating the R code to C++ the run time of the SpeLLL algorithm was greatly reduced.

```
#include <RcppArmadillo.h>
#include <soft_thresh.cpp>
#include <setZero.cpp>

using namespace arma;
// [[Rcpp::depends(RcppArmadillo)]]

// [[Rcpp::export]]
arma::mat admmFAST(arma::mat x, arma::vec y, float lam1,
float lam2, float rho,
arma::vec initial, arma::mat D,
arma::vec m, arma::vec w,
float dtol, float ptol, float z_thresh, int maxiter)
{

arma::vec lambda = join_cols(lam1*w,lam2*m);
arma::vec alpha = D * initial;
int t = alpha.n_rows;
arma::vec z = arma::zeros<arma::vec>(t);
arma::mat beta_half = inv((x.t()*x) + (rho * D.t()*D));
arma::vec beta = arma::zeros<arma::vec>(initial.n_rows);

for (int i = 0; i < maxiter; ++i){

arma::vec alphak=alpha;

beta = beta_half * ((x.t()*y) + (rho*D.t() * (alpha-z)));
```

```
arma::vec temp1 = (D * beta)+z;
arma::vec temp2 = lambda/rho;

alpha = sfThresh(temp1,temp2);

z = z + (D*beta) - alpha;

arma::vec d = rho * D.t() * (alpha-alphak);
arma::vec p = (D*beta) - alpha;

beta = setZero(beta, z_thresh);

if (norm(d,2) < dtol && norm(p,2) < ptol){
break;
}
}
```

A.2 Further Details of Sampling Scheme

A.2.1 Drawing Θ_h

Following the sampling scheme given in (Li et al., 2019), but modified to draw Θ_h for $\mathbf{Y}_j$ $j = 1, ..., N$,

1. The factor matrix $\mathbf{D}_{j,k}$ for $k = 1, ..., K$ and $j = 1, ..., N$ is sampled from the conditional posterior distribution

$$p(\mathbf{D}_{j,k}|\cdots) \sim CN(\mu_{D_kj}, \boldsymbol{\Sigma}_{D_kj}) \tag{.3}$$

where $\boldsymbol{\mu}_{D_kj} = \boldsymbol{\Sigma}_{D_kj}\Lambda_k^*\boldsymbol{\Sigma}_{D_kj}^{-1}\mathbf{Y}_{kj}$ and $\boldsymbol{\Sigma}_{D_kj} = (\Lambda_k^*\boldsymbol{\Sigma}_\epsilon^{-1}\Lambda_k + I_Q)^{-1}$.

2. Define $\Lambda_k^{(pq)}$ as the pqth element of Λ_k. Then all K frequencies of $\Lambda^{(pq)}$ are modeled as $\Lambda^{(pq)} = (\Lambda_1^{(pq)}, ..., \Lambda_K^{(pq)})' = W^{(re)}\alpha_{pq} + W^{(im)}i\beta_{pq}$ where $\boldsymbol{\alpha}_{pq} = (\alpha_{pq0}, ..., \alpha_{pqS-1})'$ and $\boldsymbol{\beta}_{pq} = (\beta_{1pq}, ..., \beta_{Spq})'$ are the coefficient vectors of the basis functions defined in equation **??** and $W^{(\cdot)}$ are the design matrices of the basis functions. The posterior distributions of $\boldsymbol{\alpha}_{pq}$ and $\boldsymbol{\beta}_{pq}$ for $p = 1, ..., P$ and $q = 1, ..., Q$ are given by

$$p(\boldsymbol{\alpha}_{pq}|\cdots) \sim CN(\mu_{\alpha_{pq}}, \boldsymbol{\Sigma}_{\alpha_{pq}}) \tag{.4}$$

$$p(\boldsymbol{\beta}_{pq}|\cdots) \sim CN(\mu_{\beta_{pq}}, \boldsymbol{\Sigma}_{\beta_{pq}}). \tag{.5}$$

where the mean and covariances of $\boldsymbol{\alpha}_{pq}$ and $\boldsymbol{\beta}_{pq}$ are

$$\Sigma_{\alpha_{pq}} = \left(\Omega^{(re)-1} + 2\sigma_{\epsilon,p}^{-2}\sum_{j=1}^N\sum_{k=1}^K |D_{jk}^{(q)}|^2 w_k^{(re)'} w_k^{(re)}\right)^{-1} \tag{.6}$$

$$\boldsymbol{\mu}_{\alpha_{pq}} = 2\sigma_{\epsilon,p}^{-2}\Sigma_{\alpha_{pq}}\left(\sum_{k=1}^K\sum_{j=1}^N Re\left[Y_{jkp}^* D_{jk}^{(q)} - \left(\sum_{h\neq q}\Lambda_k^{(ph)}D_{jk}^{(h)}\right)^* D_{jk}^{(q)*}\right]w_k^{(re)'}\right) \tag{.7}$$

$$\Sigma_{\beta_{pq}} = \left(\Omega^{(re)-1} + 2\sigma_{\epsilon,p}^{-2}\sum_{j=1}^N\sum_{k=1}^K |D_{jk}^{(q)}|^2 w_k^{(im)'} w_k^{(im)}\right)^{-1} \tag{.8}$$

$$\boldsymbol{\mu}_{\beta_{pq}} = 2\sigma_{\epsilon,p}^{-2}\Sigma_{\beta_{pq}}\left(\sum_{k=1}^K\sum_{j=1}^N Im\left[Y_{jkp}^* D_{jk}^{(q)} - \left(\sum_{h\neq q}\Lambda_k^{(ph)}D_{jk}^{(h)}\right)^* D_{jk}^{(q)*}\right]w_k^{(im)'}\right). \tag{.9}$$

$$\tag{.10}$$

Then if $w_k^{(.)}$ is the kth row of $W^{(.)}$,

$$
\begin{aligned}
\Omega_{pq}^{(re)} &= \mathrm{diag}\left[\psi_{q(re)}^{-1}, (2\pi\psi_{q(re)})^{-1}\tau_{pq(re)}^2, ..., (2\pi(S-1)\psi_{q(re)})^{-1}\tau_{pq(re)}^2\right] \\
\Omega_{pq}^{(im)} &= \mathrm{diag}\left[(2\pi\psi_{q(im)})^{-1}\tau_{pq(im)}^2, ..., (2\pi S\psi_{q(im)})^{-1}\tau_{pq(im)}^2\right].
\end{aligned}
$$

3. The smoothing parameters $\tau_{pq(.)}^2$ for $p = 1, ..., P$, $q = 1, ..., Q$ are sampled from the distributions

$$
\begin{aligned}
p(\tau_{pq(re)}^2|\cdots) &\sim IG\left(\frac{(S+\nu-1)}{2}, \frac{\tilde{\alpha}_{pq}'\tilde{\alpha}_{pq}}{2\psi_q^{-1}} + \frac{\nu}{g_{pq(re)}}\right) \\
p(\tau_{pq(im)}^2|\cdots) &\sim IG\left(\frac{(S+\nu)}{2}, \frac{\beta_{pq}'\beta_{pq}}{2\psi_q^{-1}} + \frac{\nu}{g_{pq(im)}}\right)
\end{aligned}
$$

where $\tilde{\alpha}_{pq} = (\alpha_{pq1}, ..., \alpha_{pq(s-1)})$. Then the hyperparameters $g_{pq(.)}$ are sampled from

$$
p(g_{pq(.)}|\cdots) \sim IG\left(\frac{\nu+1}{2}, \frac{\nu}{\tau_{pq(.)}^2} + \frac{1}{G_\tau^2}\right)
$$

4. The shrinkage parameters $\psi_{h(.)} = \prod_{h=1}^q \phi_{h(.)}$ are sampled from

$$
\begin{aligned}
p(\phi_{1(re)}|\cdots) &\sim Ga\left(a_1 + \frac{PQS}{2}, \sum_{q=1}^Q \Psi_{q(re)}^{(1)} \sum_{p=1}^P \frac{\tilde{\alpha}_{pq}'\tilde{\alpha}_{pq}}{2\tau_{pq(re)}^2} + \alpha_{pq0}^2\right) \\
p(\phi_{1(im)}|\cdots) &\sim Ga\left(a_1 + \frac{PQS}{2}, \sum_{q=1}^Q \Psi_{q(im)}^{(1)} \sum_{p=1}^P \frac{\beta_{pq}'\beta_{pq}}{2\tau_{pq(im)}^2}\right)
\end{aligned}
$$

for $h = 1$, and for $h \geq 2$

$$
\begin{aligned}
p(\phi_{h(re)}|\cdots) &\sim Ga\left(a_2 + \frac{PS(Q-h+1)}{2}, \sum_{q=1}^Q \Psi_{q(re)}^{(h)} \sum_{p=1}^P \frac{\tilde{\alpha}_{pq}'\tilde{\alpha}_{pq}}{2\tau_{pq(re)}^2} + \alpha_{pq0}^2\right) \\
p(\phi_{h(im)}|\cdots) &\sim Ga\left(a_2 + \frac{PS(Q-h+1)}{2}, \sum_{q=1}^Q \Psi_{q(im)}^{(h)} \sum_{p=1}^P \frac{\beta_{pq}'\beta_{pq}}{2\tau_{pq(im)}^2}\right)
\end{aligned}
$$

where $\Psi_{q(.)}^{(h)} = \prod_{t=1, t\neq h}^q \phi_{h(.)}$ for $h = 1, ..., Q$.

5. The error variances $\sigma_{\epsilon,p}$ for $p = 1, ..., P$ are sampled as

$$p(\sigma_{\epsilon,p}|\cdots) \sim IG\left(\frac{nK+\nu}{2}, \sum_{j=1}^{n}\sum_{k=1}^{K}|Y_{jkp} - \Lambda_k^{(p)}\mathbf{D}_{jk}|^2 + \frac{\nu}{g_{\epsilon,p}}\right)$$

where the conditional distribution of the hyperparameter is given by

$$p(g_{\epsilon,p}|\cdots) \sim IG\left((\nu+2)/2, \nu/\sigma_{\epsilon,p}^2 + 1/G_\epsilon^2\right)$$

A.2.2 R Code to Draw Θ

An abbreviated version of the implementation in R of Section A.2.1 used to draw Θ for one iteration of the Gibbs sampler. The full version uses an 'if else' statement which implements a slightly different version of the function when only a single $\mathbf{Y}_k$ is allocated to a component.

```
#Fxn for stationary factor model.
#
#INPUTS:
#Y = Fourier of multiple multivariate TS
#input = object with parameters and model objects
#S = number of smoothing splines
#Q = number of factors
#v,G_e, G_tau, a1, a2 = hyperparameters
#
#OUTPUTS:
#Lambda = matrix of factors
#D = matrix of loadings
#sigep, g_p = variance and latent variable
#Tau_i, Tau_r, g_i, g_r = Smoothing parameters and latent variable
#rho_i, rho_r = Shrinkage parameters
#

#If n = 1 a slightly different code is used since Y is no longer an array.

stat.fact.draw.user <- function(Y, input,
S=10, Q=12, v=2, G_e=10, G_tau=10, a1=5,a2=1){

Lambda <- input$Lambda
sigma <- input$sigep
Tau_i <- input$Tau_i
Tau_r <- input$Tau_r
rho_i <- input$rho_i
rho_r <- input$rho_r
g_i <- input$g_i
g_r <- input$g_r
g_p <- input$g_p
Yi <- Y
dms <- dim(Yi)
p <- dms[2]
k <- dms[1]
n <- dms[3]

freq <- 1:k/(2*k)

if (length(dms) == 2){
n <- 1
```

```
}

#Preallocate arrays
dist <- array(NA,c(k,p,n))
alphaalpha0 <- alphaalpha <- betabeta <- matrix(1,nrow=p,ncol=Q)
D_k <- array(NA,c(k,Q,n))
subq <- musum <- matrix(NA, nrow = k, ncol = n)
Lambda_temp <- array(0,c(p,Q,k))

#Create basis functions
w <- lin_basis_func(S,freq)

#--Draw Lambda
sigepinv <- diag(1/sigma^2)

for (i in 1:k){

  tempsig <- (cjt(Lambda[,,i]) %*% sigepinv %*% Lambda[,,i]) + diag(Q)
  sigma_d <- chol.qr.inv(tempsig)
  temp <- sigma_d$final %*% cjt(Lambda[,,i]) %*% sigepinv

 for (l in 1:n){

   mu_dk <- temp  %*% (Yi[i,,l])
   D <- mvnfast::rmvn(1,mu=c(Re(mu_dk),Im(mu_dk)),sigma=.5*t(sigma_d$invss))
   D_k[i,,l] <- D[1:Q] + 1i*(D[(Q+1):(2*Q)])

 }
}

psi_i <- cumprod(rho_i)
psi_r <- cumprod(rho_r)

for (m in 1:Q){
 for (j in 1:p){

  for (l in 1:n){
    subq[,l] <- rowSums(D_k[,,l]*t(Lambda[j,,])) - D_k[,m,l]*Lambda[j,m,]
    musum[,l] <- -D_k[,m,l]*Conj(subq[,l]) + Conj(Yi[,j,l])*D_k[,m,l]
  }


  musum2 <- apply(musum,1,sum)
  Areal = 2*t(w$xx_r) %*% Re(musum2)/sigma[j]^2
  Aimag = -2*t(w$xx_i) %*% Im(musum2)/sigma[j]^2

  Dabs <- apply(D_k[,m,],1,function(a) sum(abs(a)^2))

  B1 = (2*t(w$xx_r*Dabs) %*% w$xx_r)/sigma[j]^2
  B2 = diag(c(psi_r[m],psi_r[m]*rep(1,S-1)/Tau_r[j,m]))
  sigma_alpha <- solve(B1+B2)

  mu_alpha <- sigma_alpha %*% Areal

  B3 = (2*t(w$xx_i*Dabs) %*% w$xx_i)/sigma[j]^2
  B4 = diag(c(psi_i[m]*rep(1,S)/Tau_i[j,m]))
  sigma_beta <- solve(B3+B4)

  mu_beta <- sigma_beta %*% Aimag

  alpha = mvnfast::rmvn(1, mu = mu_alpha, sigma = sigma_alpha)
  beta = mvnfast::rmvn(1, mu = mu_beta, sigma = sigma_alpha)

  templambda <-  w$xx_r %*% t(alpha) + 1i * w$xx_i %*% t(beta)

  Lambda_temp[j,m,] <- templambda

  alphaalpha[j,m] <- t(alpha[-1]) %*% (alpha[-1]) #Stored for shrinkage and smoothing
```

```r
  betabeta[j,m] <- beta %*% t(beta)
  alphaalpha0[j,m] <- alpha[1]^2
 }
}

Lambda <- Lambda_temp


#--Draw Tau

tr <- alphaalpha %*% diag(psi_r)/2 + v/g_r
ti <- betabeta %*% diag(psi_i)/2 + v/g_i

Tau_r <- apply(tr, c(1,2), function(a) 1/rgamma(1,shape = (S+v-1)/2, rate = a))
Tau_i <- apply(ti, c(1,2), function(a) 1/rgamma(1,shape = (S+v)/2, rate = a))

g_r <- apply(Tau_r, c(1,2), function(a) 1/rgamma(1, shape = (v+1)/2, rate = v/a + 1/G_tau^2))
g_i <- apply(Tau_i, c(1,2), function(a) 1/rgamma(1, shape = (v+1)/2, rate = v/a + 1/G_tau^2))


#--Draw Rho

ra <- colSums(alphaalpha/(Tau_r)) + colSums(alphaalpha0)
rb <- colSums(betabeta/(Tau_i))

rho_r[1] <- rgamma(1, shape = a1 + (p*Q*S)/2, rate = 1 + .5* (1/rho_r[1])* psi_r %*% ra)
rho_i[1] <- rgamma(1, shape = a1 + (p*Q*S)/2, rate = 1 + .5* (1/rho_i[1])*psi_i %*% rb)

for (m in 2:Q){
  rho_r[m] <- rgamma(1, shape = a2 + (p*(Q-m+1)*S)/2,
        rate = 1 + .5*(1/rho_r[m])*psi_r[m:Q] %*% ra[m:Q])
  rho_i[m] <- rgamma(1, shape = a2 + (p*(Q-m+1)*S)/2,
        rate = 1 + .5*(1/rho_i[m])*psi_i[m:Q] %*% rb[m:Q])
}

#--Draw Sigma

for (i in 1:k){
 for (l in 1:n){

  dist[i,,l] <- abs(Yi[i,,l] - Lambda[,,i] %*% D_k[i,,l])^2

 }
}

dist2 <- apply(dist,2,sum)

srate<- dist2 + v/(g_p)
sigma <- sapply(srate, function(a) 1/rgamma(1,shape = n*k+v, rate = srate))

grate <- v/(sigma) + (1/G_e^2)
g_p <- sapply(grate, function(a) 1/rgamma(1, shape = (v+1)/2, rate = grate))

sigma <- sqrt(sigma)

}

return(list(Lambda = Lambda, D = D_k, sigep = sigma, Tau_i=Tau_i,
Tau_r=Tau_r, rho_i=rho_i, rho_r=rho_r, g_i=g_i, g_r=g_r, g_p=g_p))
}
```

A.2.3   Drawing z_i

The latent indicators for mixture component membership have the conditional probability mass function of

$$p(z_j = h | \cdots) = \frac{\pi_h(\mathbf{v}_j) L(Y_k^{(h)} | \Theta^{(h)})}{\sum_{\ell=1}^{H} \pi_\ell(\mathbf{v}_j) L(Y_k^{(\ell)} | \Theta^{(\ell)})}$$

with $\ell = 1, ..., H$. Computationally, this ratio can be unstable given that $L(Y_k^{(\ell)} | \Theta^{(\ell)})$ is the product of $p \times k$ elements. When implemented in R, a modified version was used which divides the top and bottom by $\pi_h(\mathbf{v}_j) L(Y_k^{(h)} | \Theta^{(h)})$ to improve stability in computation.

A.2.4   Drawing $\pi_h(\mathbf{v}_j)$

The weight for each component are drawn from a multinomial regression model that has been augmented using a Polya-Gamma distribution (**?**). The update for each component h has two steps. Let $C_{jh} = \log \sum_{\ell \neq h} v_j' \gamma_\ell$ and $\kappa_h = I(z_j = h) - 0.5$. First draw γ_h from

$$p(\gamma_h | \Delta_h) \sim N_m(\boldsymbol{\mu}_{\gamma h}, \Sigma_{\gamma h})$$

where m is the dimension of the covariate space, $\Sigma_h^{-1} = \mathbf{V}' \Delta_h \mathbf{V} + I_m$, and $\boldsymbol{\mu}_{\gamma h} = \Sigma_{\gamma h} \left(\mathbf{V}'(\kappa_h + \Delta_h c_h) \right)$ Then update hyperparameter $\Delta_h = \text{diag}(\{\delta_{jh}\})$ using

$$p(\delta_{jh} | \gamma_h) \sim PG(1, v_j' \gamma_h - C_{jh})$$

Table A1: Estimated value of mixture weights.

	HC	FEP
Component 1	0.18 (0.05, 0.38)	0.00 (0.00, 0.04)
Component 2	0.06 (0.00, 0.20)	0.07 (0.01, 0.20)
Component 3	0.77 (0.55, 0.92)	0.69 (0.50, 0.85)
Component 4	0.00 (0.00, 0.00)	0.11 (0.02, 0.25)
Component 5	0.00 (0.00, 0.00)	0.12 (0.03, 0.26)

Estimate value of mixture weights and corresponding 95% credible interval for HC and FEP for each component of the mixture model for the first 1000 iterations.

A.3   Additional Details for Analysis of EEG Data

A.3.1   Exploring Mixture Components

The section explores the mixture components recovered from the first 1000 iterations of the sampler after burn in. The equation of the power spectrum conditioned on covariate vector $\mathbf{v}$ is given by

$$\hat{f}(\omega|\mathbf{v}) = \sum_{h=1}^{H} \pi_h(\mathbf{v})\left[\Lambda(\omega)^{(h)}\Lambda(\omega)^{(h)*} + \Sigma_e^{(h)}\right]$$

for $h = 1, ..., H$. In our analysis of the EEG data, the number of mixture components recovered by the model was $H = 5$ and label swapping was handled using the method discussed in 3.2.4.1.

Table A1 shows the estimated value of mixture weights $\pi_h(\mathbf{v})$ and corresponding 95% credible intervals for the HC and FEP covariate patterns. Components 2 and 3 contributed to the estimate of the power spectrum for both HC and FEP, with component 3 contributing the most overall. Component 1 only added to the power spectrum for HC while components 4 and 5 only added to the estimate for FEP.

Figure A1, Figure A2, and Figure A3 plot the posterior estimate of the mixture components for the power spectra, real part of the cross spectra, and imaginary part of the cross spectra for the electrodes of interest in Section 3.4.

A.3.2 Plots of Spectral Density for All Electrodes

Figure A4 of this section shows plots of the difference between the diagonal elements of the spectral density between HC and FEP, namely $f_{i,i}(\omega|x=0) - f_{i,i}(\omega|x=1)$ for $i = 1, ..., 64$. Positive values indicate areas where the density for HC had more power than FEP and vice versa for the area with negative values. Frequency bands are denoted by vertical lines.
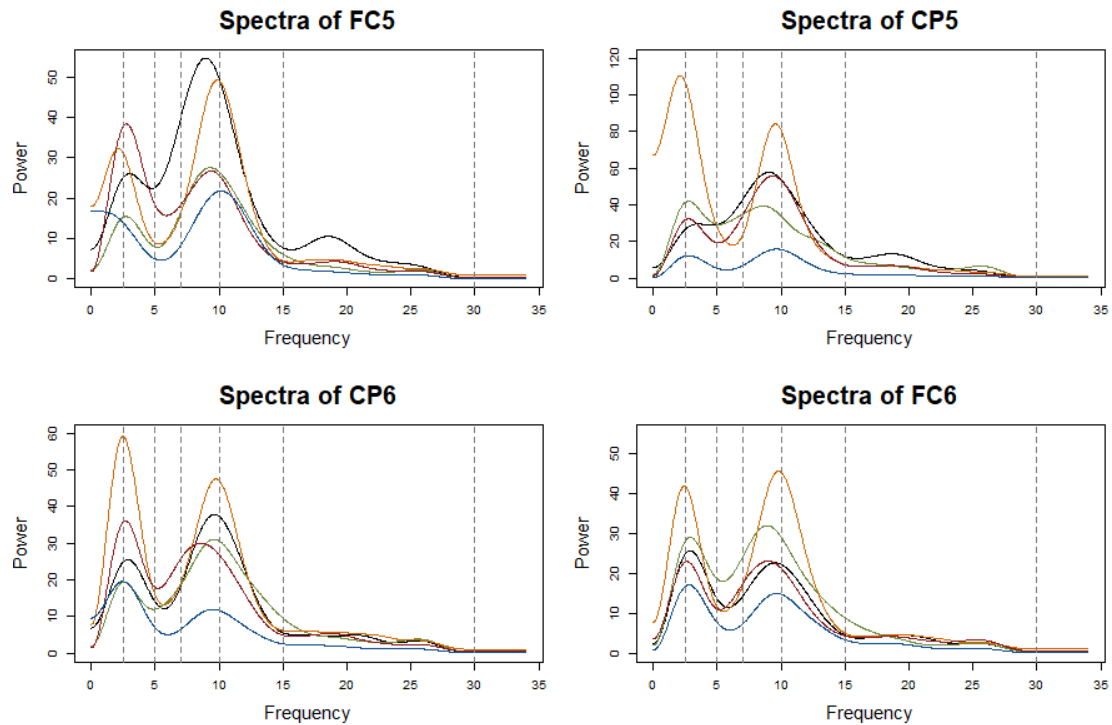
Figure A1: Components of mixture model: power spectrum.

Graph of the components of the mixture model for the spectrum of FC5, CP5, CP6, FC6 electrodes. Component 1 = black; component 2 = green; component 3 = orange; component 4 = red; component 5 = blue. Dashed lines mark traditional boundaries for the delta, theta, low alpha, high alpha, and beta frequency bands.
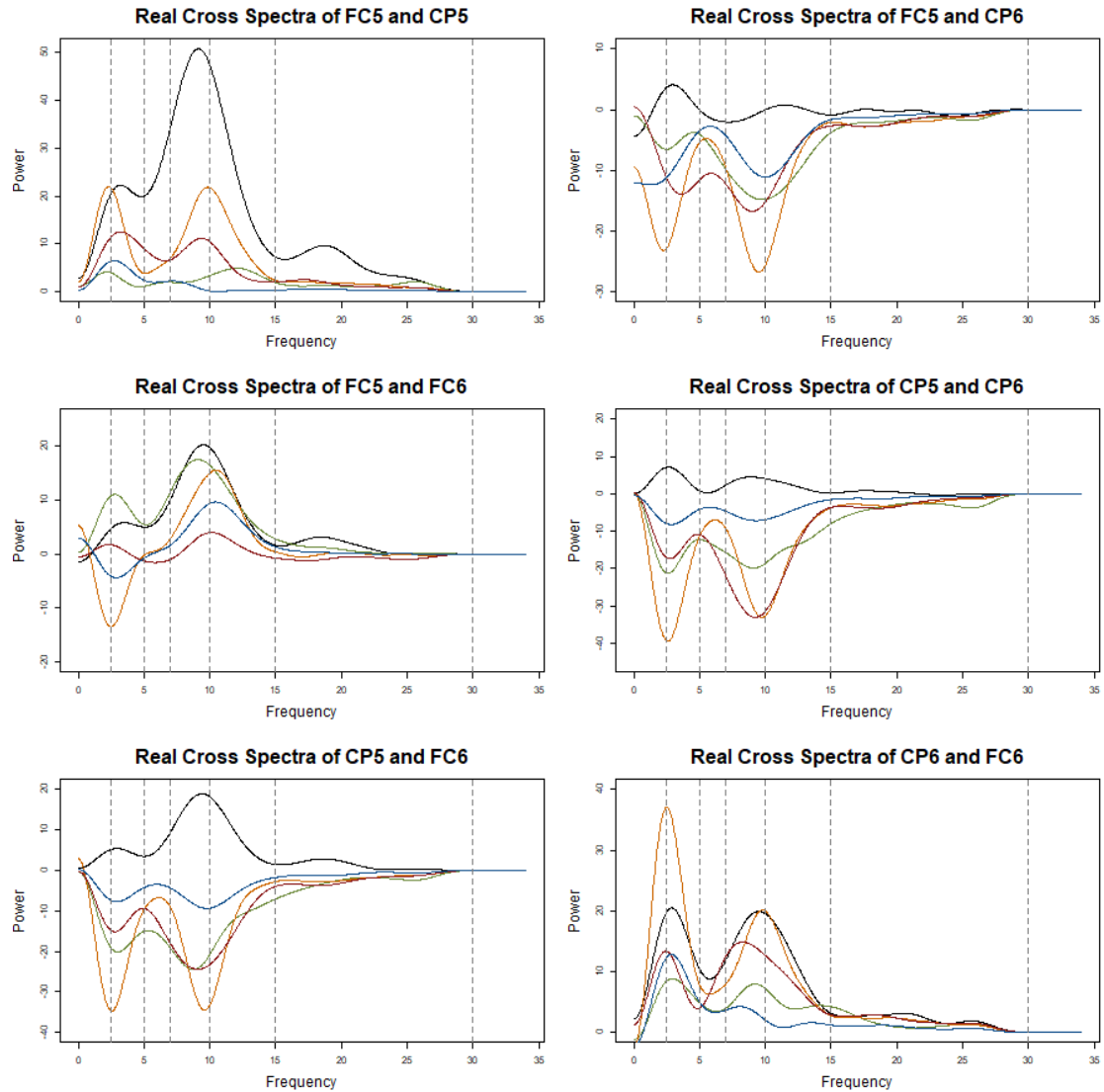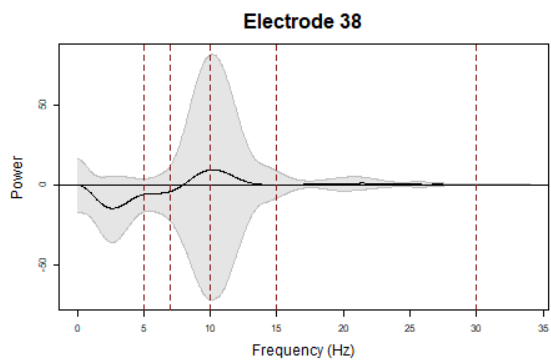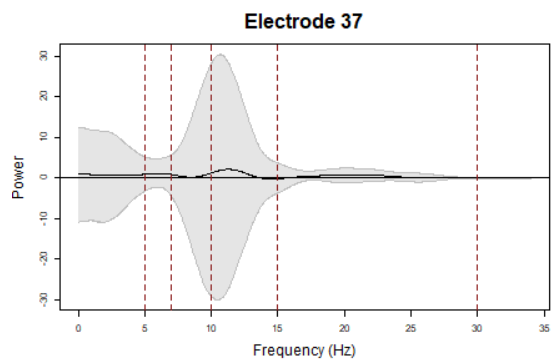
Figure A2: Components of mixture model: real cross spectra.

Graph of the real components of the mixture model for the cross spectrum of FC5, CP5, CP6, FC6 electrodes. Component 1 = black; component 2 = green; component 3 = orange; component 4 = red; component 5 = blue. Dashed lines mark traditional boundaries for the delta, theta, low alpha, high alpha, and beta frequency bands.
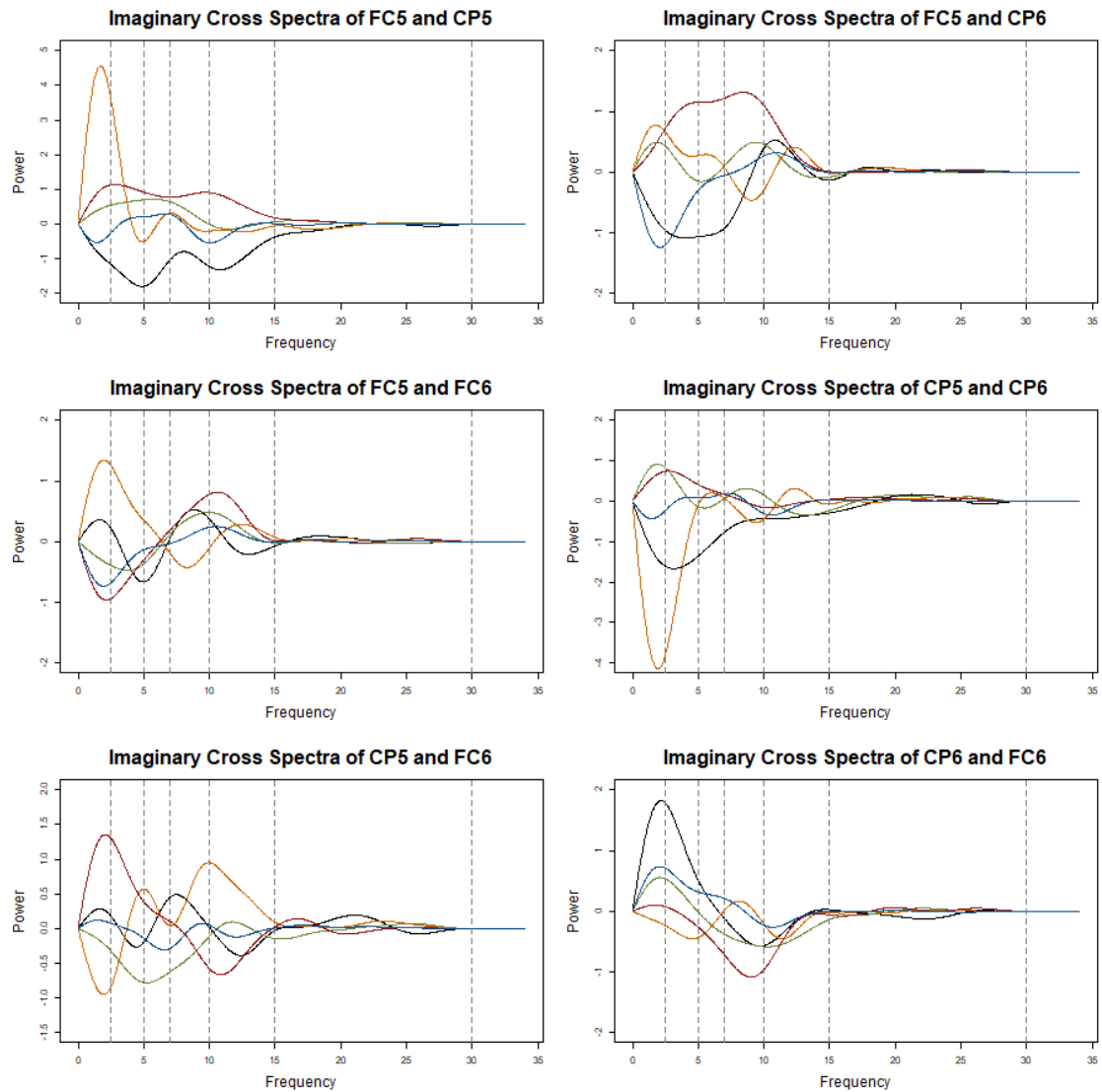
Figure A3: Components of mixture model: imaginary cross spectra.

Graph of the imaginary components of the mixture model for the cross spectrum of FC5, CP5, CP6, FC6 electrodes. Component 1 = black; component 2 = green; component 3 = orange; component 4 = red; component 5 = blue. Dashed lines mark traditional boundaries for the delta, theta, low alpha, high alpha, and beta frequency bands.

Electrode 1

Electrode 2

Electrode 3

Electrode 4

Electrode 5

Electrode 6

Electrode 7

Electrode 8

Electrode 9

Electrode 10

Electrode 11

Electrode 12

Electrode 13

Electrode 14

Electrode 15

Electrode 16

Electrode 25

Electrode 26

Electrode 27

Electrode 28

Electrode 29

Electrode 30

Electrode 31

Electrode 32

Electrode 33

Electrode 34

Electrode 35

Electrode 36

Electrode 37

Electrode 38

Electrode 39

Electrode 40

Electrode 41

Electrode 42

Electrode 43

Electrode 44

Electrode 45

Electrode 46

Electrode 47

Electrode 48

Electrode 49

Electrode 50

Electrode 51

Electrode 52

Electrode 53

Electrode 54

Electrode 55

Electrode 56

Figure A4: Plots of the difference in estimated power and 95% credible intervals for each electrode.

Bibliography

S. Adak. Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, 93:1488–1501, 1998.

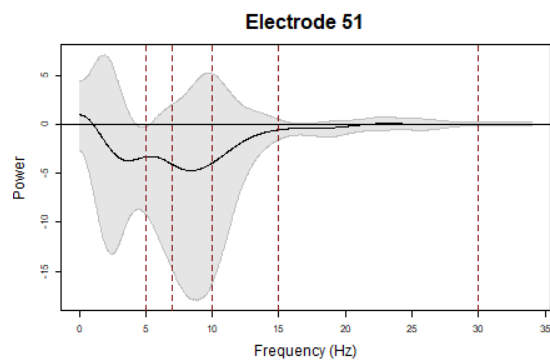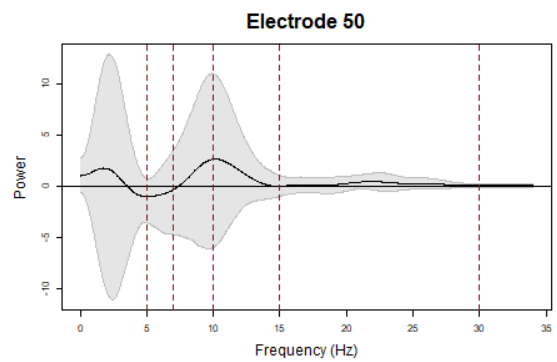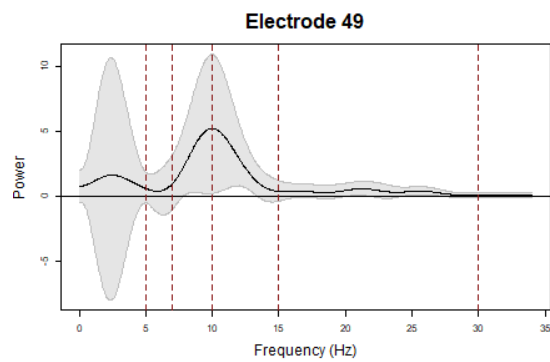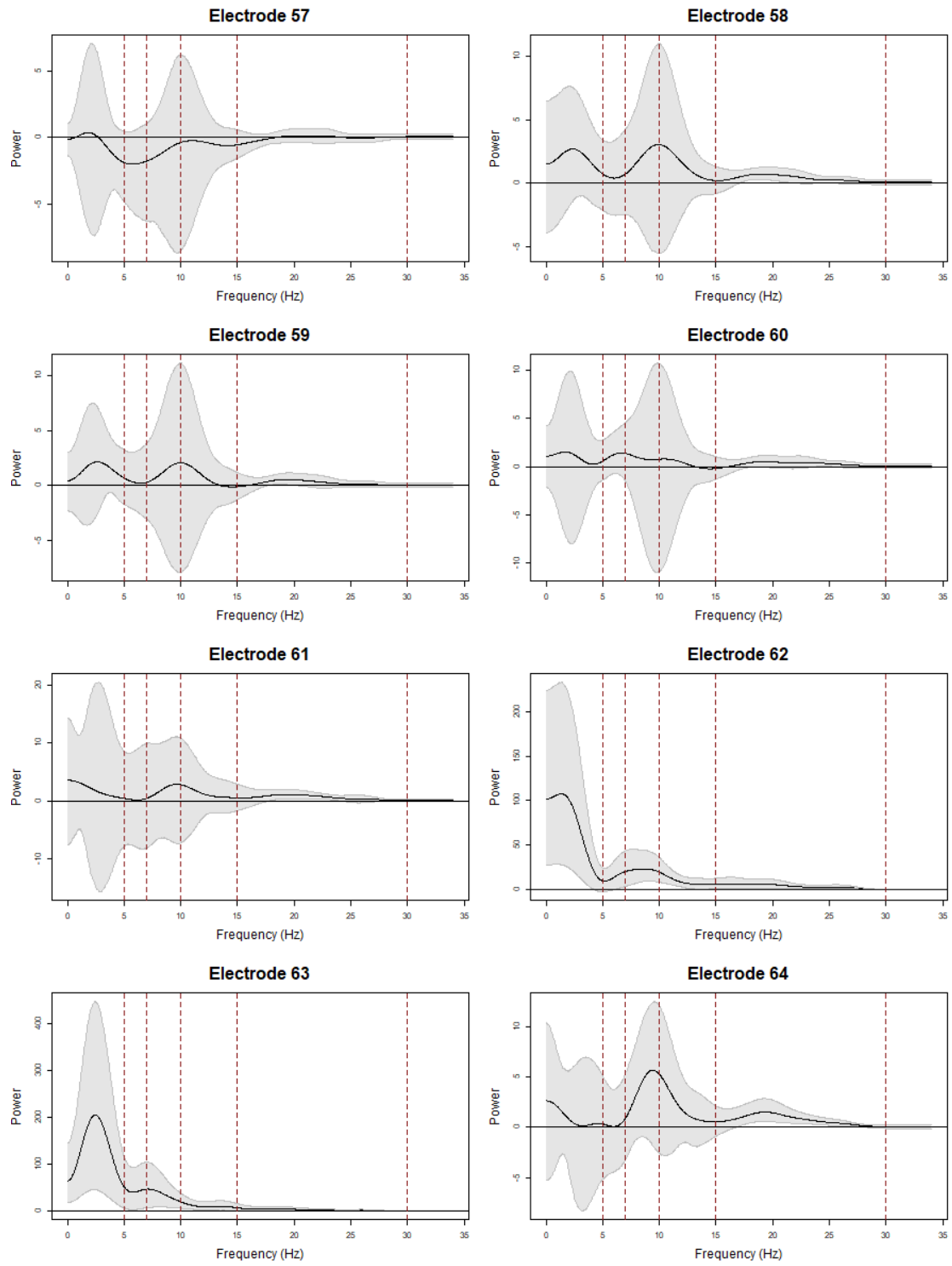D. Álvarez, R. Hornero, J. Víctor Marcos, F. Del Campo, and M. López. Spectral analysis of electroencephalogram and oximetric signals in obstructive sleep apnea diagnosis. *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, pages 400–403, 2009.

D. Begic, L. Hotujac, and N. Jokic-Begic. Quantitative eeg in positive and negative schizophrenia. *Acta Psychiatrica Scandinavica*, 101:307–311, 2019.

M. Bertolacci, O. Rosen, E. Cripps, and S. Cripps. Adaptspec-x: Covariate dependent spectral modeling of multiple nonstationary time series. *arXiv:1908.06622v1 [stat.ME]*, 2019.

A. Bhattacharya and D. B. Dunson. Sparse bayesian infinite factor models. *Biometrika*, 98:291–306, 2011.

P.J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, pages 2577–2604, 2008.

V. K. Bochkarev, A. V Kirenskaya, S. V Solnceva, and A. A. Tkachenko. Specificity of spatial organization of evoked eeg rhythms in patients with paranoid schizophrenia. *Zh Nevrol Psikhiatr Im S S Korsakova*, 117:29–35, 2017.

H. Bohm and R. von Sachs. Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis*, 100:913–935, 2009.

M. H. Bonnet and D. L. Arand. Heart rate variability: Sleep stage, time of night, and arousal influences. *Electroencephalography and Clinical Neurophysiology*, 102(5):390–396, 1997.

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2010.

D. R. Brillinger. *Time Series: Data Analysis and Theory*. SIAM, 2002.

S. A. Bruce, M. H. Hall, D. J. Buysse, and R. T. Krafty. Conditional apative bayesian spectral analysis of nonstationary biomedical time series. *Biometrics*, 74(1):260–269, 2018.

S. A Bruce, C. Y. Tang, M. H. Hall, and R. T. Krafty. Empirical frequency band analysis of nonstationary time series. *Journal of the American Statistical Association*, 0(0):1–31, 2019.

A. Cadonna, A. Kottas, and R. Prado. Bayesian spectral modeling for multiple time series. *Journal of the American Statistical Association*, 114:1838–1853, 2019.

T. Cai and H. H. Zhou. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106:6472–684, 2011.

K. Chen and K. Chan. A note on rank reduction in sparse multivariate regression. *Jounrnal of Statistical Theory and Practice*, 10(1):100–120, 2016.

K. Chen, K. Chan, and N. C. Stenseth. Reduced rank stochastic regression with a sparse singular value decomposition. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 74(2):203–221, 2012.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1): 1–37, 1997.

I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *Fundamental Papers in Wavelet Theory*, 5:442–486, 1990.

R. A. Davis, T. C.M. Lee, and G. A. Rodriguez-Yam. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239, 2006.

E. R. de Lima, A. O. Andrade, J. L. Pons, K. Kyberd, and S. J. Nasuto. Empirical mode decomposition: a novel technique for the study of tremor time series. *Medical and Biological Engineering and Computing*, 44:569–582, 200.

A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics. *Journal of Neuroscience Methods*, 134:9–21, 2004.

D. Eddelbuettel and R. François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011. doi: 10.18637/jss.v040.i08. URL http://www.jstatsoft.org/v40/i08/.

K. B. Ensor. Time series factor models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(2):97–104, 2013.

F. Ferrarelli, R. E. Kaskie, B. Graziano, C. C. Reis, and A. G. Casali. Abnormalities in the evoked frontal oscillatory activity of first-episode psychosis: A tms/eeg study. *Schizophrenia Research*, 206:436–439, 2019.

M. Fiecas and H. Ombao. The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics*, 5:1102–1125, 2011.

M. Fiecas and R. von Sachs. Data-driven shrinkage of the spectral density of a high-dimensional time series. *Electronic Journal of Statistics*, 8:2975–3003, 2014.

M. Fiecas, H Ombao, C. Linkletter, and J. Sanes. Functional connectivity: shrinkage estimation and randomization test. *NeuroImage*, 5:1102–1125, 2010.

C. C. French and J. G. Beumont. A critical review of eeg coherence studies of hemisphere function. *International Journal of Psychophysiology*, pages 241–254, 1984.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:1–19, 2006.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. Taylor and Francis Group, 3 edition, 2014. ISBN 978-1-4398-4095-5.

W. Guo, H. C. Ombao, and R. von Sachs. Smoothing spline anova for time-dependent spectral analysis. *Journal of the American Statistical Association*, 98(463):643–652, 2003.

M. Hall, R. Vasko, D. Buysse, H. Ombao, Q. Chen, J. D. Cashmere, D. Kupfer, and J. F. Thayer. Acute stress affects heart rate variability during sleep. *Psychosomatic Medicine*, 66(1):56–62, 2004.

F. J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

J. Henryk, A. Switonski, A. Michalczuk, P. Grabiec, M. Pawlyta, and K. Wojciechowski. Assessment of local dynamic stability in gait based on univariate and multivariate time series. *Computational and Mathematical Methods in Medicine*, 2019, 2019.

M. R. Ho, H. Ombao, J. C. Edgar, J. M. Cañive, and G. A. Miller. Time-frequency discrimniate analysis of meg signals. *NeuroImage*, 40(1):174–186, 2008.

A. Khanna, A. Pascual-Leone, C.M. Michel, and F. Farzan. Microstates in resting-state eeg: current status and future directions. *Neuroscience and Biobehavioral Reviews*, 49:105–113, 2015.

R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

J. W. Kim, Y. S. Lee, D. H. Han, K. J. Min, J. Lee, and K Lee. Diagnostic utility of quantitative eeg in un-medicated schizophrenia. *Neuroscience Letters*, 589:126–131, 2015.

S. E. Kim, M. K. Behr, D. Ba, and E. N. Brown. State-space multitaper time-frequency analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 115(1):E5–E14, 2018. doi: 10.1073/pnas.1702877115.

R. T. Krafty and W. O. Collinge. Penalized multivariate whittle likelihood for power spectrum estimation. *Biometrika*, 100:447–458, 2013.

R. T. Krafty, O. Rosen, D. S. Stouffer, D. J. Buysse, and M. H. Hall. Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology. *Journal of the American Statistical Association*, 112:1405–1416, 2017.

O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matriciess. *Journal of Multivariate Analysis*, 88:365–411, 2004.

L. Leocani and G. Comi. Eeg coherence in pathological conditions. *Journal of Clinical Neurophysiology*, 16:548, 1999.

X. Li, L. Mo, X. Yuan, and J. Zhang. Linearized alternating direction method of multipliers for sparse group and fused lasso models. *Computational Statistics and Data Analysis*, 79:203–221, 2014.

Z. Li and R. T. Krafty. Adaptive bayesian time-frequency analysis of multivariate time series. *Journal of the American Statistical Association*, 114:453–465, 2019.

Z. Li, O. Rosen, F. Ferrarelli, and R. T. Krafty. Adaptive bayesian spectral analysis of high-dimensional nonstationary time series. *arXiv:1910.12126v1 [stat.ME]*, 2019.

C. K. A. Lim and W. C. Chia. Analysis of single-electrode eeg rhythms using matlab to elicit correlation with cognitive stress. *International Journal of Computer Theory and Engineering*, 7, 2015.

M. Lynall, D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Muller, and E. Bullmore. Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience*, 30:9477–9487, 2010.

C. Macaro and R. Prado. Spectral decompositions of multiple time series: A bayesian non-parametric approach. *Psychometrika*, 79(1):105–129, 2014.

M. Mächler. Arbitrarily accurate computation with r: The rmpfr package. 2013.

M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz. Introduction to ecg time series variability analysis: A simple overview. *ECG Time Series Variability Analysis: Engineering and Medicine*, pages 1–12, 1996.

J. J. Newson and T. C. Thiagarajan. Eeg frequency bands in psychiatric disorders: a review of resting state studies. *Frontiers in Human Neuroscience*, 12:521, 2019.

H. Ombao, J. Raz, R. Von Sachs, and W. Guo. The slex model of a non-stationary random process, 2002.

H. C. Ombao, J. A. Raz, R. von Sachs, and B. A. Malow. Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association*, 96(454):543–560, 2001.

X. Pedeli and D. Karlis. An integer-valued time series model for multivariate surveillance. *Statistics in Medicine*, 39, 2019.

N. G. Polson, J. G. Scott, and J. Windle. Bayesian inference for logistic models using polya-gamma latent variables. *Journal of the American Statistical Association*, 108:1339–1349, 2013.

M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society*, 27(2):204–237, 1965.

W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

G. C. Reinsel and R. P. Velu. *Multivariate reduced-rank regression: theory and applications*. Springer, New York, 1 edition, 1998. ISBN 0387986014.

T. Rigon and D. Durante. Tractable bayesian density regression via logit stick-breaking priors. *Journal of Statistical Planning and Inference*, 211:131–142, 2017.

D. Schneider-Luftman and A.T. Walden. Partial coherence estimation via spectral matrix shrinkage under quadratic loss. *IEEE Transactions on Signal Processing*, 64:5767–5777, 2016.

A. L. Schröder and H. Ombao. Fresped: Frequency-specific change-point detection in epileptic seizure multi-channel eeg data. *Journal of the American Statistical Association*, 114(525):115–128, 2019.

F. Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5(September):1–17, 2017.

R. H. Shumway and D. S. Souffer. *Time series analysis and its applications*. Springer, New York, 3 edition, 2011. ISBN 9780387781884. doi: 10.1016/j.peva.2007.06.006.

N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. A sparse-group lasso. *Journal of Computational and Graphical Statistics*, 22, 2013.

D. Stoffer. Detecting common signals in multiple time series using the spectral envelope. *Journal of the American Statistical Association*, 94:1341–1356, 1999.

Y. Sun, Y. Li, and S. Kuceyeski, A.and Basu. Large spectral density matrix estimation by thresholding. *arXiv: 0902.0885 [stat.ME]*, 2018.

X. Tang, L. Xia, W. Liu, Y. Peng, D. Chen, T. Gao, and Y. Zeng. Analysis of frequency domain of eeg signals in clinical location of epileptic focus. *Clinical EEG and Neuroscience*, 44(1):25–30, 2013.

B. J. Taylor, L. A. Irish, L. M. Martire, G. J. Siegle, R. T. Krafty, R. Schulz, and M. H. Hall. Avoidant coping and poor sleep efficiency in dementia caregivers. *Psychosomatic Medicine*, 77 (9):1050–1057, 2015.

D. J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9): 1055–1096, 1982.

R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(1): 91–108, 2005.

C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79:61–78, 1998.

P. Valdés, J. Bosch, R. Grave, J. Hernandez, J. Riera, R. Pascual, and R. Biscay. Frequency domain models of the eeg. *Brain Topography*, 4(4):309–319, 1992.

M. Wand, J. Omergerod, S. Padoan, and R. Frühworth. Mean field variational bayes for elaborate distribution. *Bayesian Analysis*, 7:847–900, 2012.

W. W. S. Wei. *Multivariate Time Series Regression Models*. John Wiley and Sons, 2019.

Y. Zhang, Z. Guo, W. Wang, S. He, T. Lee, and M. Loew. A comparison of the wavelet and short-time fourier transforms for doppler spectral analysis. *Medical Engineering and Physics*, 25(7): 547–557, 2003.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.