UNIVERSITY OF CALIFORNIA,
IRVINE

Visualization and Non–parametric Statistical Testing Methods for Multivariate Time Series

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

Duy Ngo

Dissertation Committee:
Associate Professor Babak Shahbaba, Chair
Professor Hernando Ombao
Associate Professor Yaming Yu
Associate Professor Zhaoxia Yu

2018

DEDICATION

To my family and friends,

for their endless encouragement and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor, Professor Hernando Ombao, for his valuable guidance and support throughout my PhD study. He is not only a great professor, but he is also my counselor in my life. Without his mentor this dissertation would not have been possible. I would like to thank Professor Babak Shahbaba, who has been giving encouragement and support during my academic research. I would also like to thank Professor Yaming Yu and Professor Zhaoxia Yu for sharing their valuable knowledge in statistics which have benefited my research. Lastly, I would like to thank my family and friends for all their love and encouragement.

CURRICULUM VITAE

Duy Ngo

EDUCATION

Doctor of Philosophy in Statistics **2018**
University of California, Irvine *Irvine, California*

Bachelor of Science in Mathematics **2013**
California State University, Fullerton *Fullerton, California*

RESEARCH EXPERIENCE

Graduate Research Assistant **2015–2017**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Instructor **2015, 2017**
University of California, Irvine *Irvine, California*

Teaching Assistant **2014–2018**
University of California, Irvine *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

An exploratory data analysis of electroencephalograms using the functional boxplots approach **2015**
Frontiers in Neuroscience

ABSTRACT OF THE DISSERTATION

Visualization and Non–parametric Statistical Testing Methods for Multivariate Time Series

By

Duy Ngo

Doctor of Philosophy in Statistics

University of California, Irvine, 2018

Associate Professor Babak Shahbaba, Chair

Many model-based methods have been developed over the last several decades for analysis of multivariate time series, such as electroencephalograms (EEG) in order to understand electrical neural data. In this dissertation, we propose to use the functional boxplot to analyze log periodograms of EEG time series data in the spectral domain. The functional boxplot approach produces a median curve – which is not equivalent to connecting medians obtained from frequency-specific boxplots. In addition, this approach identifies a functional median, summarizes variability and detects potential outliers. By extending functional boxplots analysis from one-dimensional curves to surfaces, surface boxplots are also used to explore the variation of the spectral power for the alpha ($8-12$ Hertz) and beta ($16-32$ Hertz) frequency bands across the brain cortical surface. By using rank-based nonparametric tests, we also investigate the stationarity of EEG traces across an exam acquired during resting-state by comparing the spectrum during the early vs. late phases of a single resting-state EEG exam.

Moreover, we present an exploratory data analysis tool for visualizing and testing the symmetric positive definite matrices (e.g., covariance, spectral and coherence matrices) in a multi–subject experimental setting. Our work is motivated by the clinician's interest to determine associations between brain functional connectivity (as measured by coherence) and patients' response to treatment. For each study participant, the geometric surface boxplot is

developed to characterize the distribution of coherence matrices through the median matrix and the 50% most central region of the data. The surface boxplot will also be used to detect the outlier coherence matrices as in the classical boxplot. To investigate the treatment effect, we develop a rank–based non–parametric approach to test for significant differences in coherence matrices between treatment and control groups. As an application, we demonstrate our proposed methods on coherence matrices, derived from an electroencephalograms, to determine potential associations treatment effect on patients with infantile spasms and hypsarrhythmia both before and after treatment.

Chapter 1

Introduction

1.1 Background

Multivariate time series are commonly produced in many different domains, from environment, through finance, to physical science. In some instances, electroencephalograms (EEGs) have been used for many decades to study the complex spatio–temporal dynamics of brain processes ([50]). Due to its excellent temporal resolution (sampling rates usually range from $100 - 1000$ Hz), EEGs can capture transient changes in brain activity, identify oscillatory behavior and study cross-dependence between EEG components. Since EEGs indirectly measure neuronal electrical activity, they can be used to infer the statistical properties of the underlying brain stochastic process. One such statistical property is the spectrum (or power spectrum) which decomposes the total variability in the EEG according to the contribution of oscillations at different frequencies. Most approaches to analyzing EEGs focus immediately on statistical modeling and spectral estimation. Here, we offer a systematic framework for exploring structures, patterns and features in the signal – prior to formal modeling. We explore the spectral properties only in a single channel using EEG traces from

several epochs.

One approach to estimating the spectrum using EEG traces is to fit a parametric time domain model, such as the autoregressive moving average (ARMA) model. Applications of parametric modeling of EEGs have a long history. See [4], [37] [38] and [41] among many others. When the spectrum of the EEG evolves over time (e.g., within an epoch), one could still use the ARMA model but allow the coefficients to vary over time. A key element in ARMA models is the order of the autoregressive (AR) and moving average (MA) components. These can be obtained objectively using an information-theoretic criterion such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Using these criteria, we obtain an optimal AR and MA order that jointly gives the best fit with the least complexity (as determined by the orders). BIC puts a heavier penalty for complexity compared to AIC and thus often gives a model with lower orders (lower complexity). From the parametric fit, we derive the estimates of the auto-correlation function and the spectrum. The theoretical background for parametric models are developed in [60], [66] and [8]).

One could also estimate the spectrum without resorting to a parametric model. Under this approach, the EEGs are considered to be superpositions of sines and cosines (Fourier waveforms) with different frequencies and random amplitudes. These random amplitudes (or coefficients) are computed using the fast Fourier transform (FFT). The squared magnitude of these amplitudes, often called the periodograms, are the data-analogues of the spectrum defined on discrete frequencies. The theoretical background on the frequency domain approach to time series is developed in [6] and [58]. This approach to analyzing EEGs continues to be popular in the cognitive and brain sciences. The following papers cover both methods and applications of spectral analysis to EEGs: [5], [47], [59] and [71], to name a few.

The common practice prior to spectral estimation is to pre-process EEGs, often to remove artifacts [48]. After artifact rejection and segmentation according to epochs, the spectrum

is estimated from each EEG trace. As noted, there is a lack a systematic framework for exploring structures, patterns and features in the signal – prior to formal modeling. Due to the complexity of EEG data, exploratory data analysis (EDA) plays an important role, especially when data are recorded from many epochs or trials during an experiment. For example, it is often expected that brain responses to the same stimulus ought to be relatively uniform, with minimal variation across epochs. In contrast, greater variability across epochs may be expected during neuroimaging studies that examine the brain in resting-state, as cognitive processes can vary within and across sessions for individual subjects and across subjects. An appropriate EDA methods can provide insights into features of EEG, including similarities and variability of the brain responses across epochs to facilitate the statistical model. In this dissertation, we propose to use the functional boxplot (FBP) method originally developed by [72] to address these questions.

Moreover, in multichannel EEG studies, there has been a growing interest in studying brain oscillatory activity between brain regions. A common measure of functional brain connectivity is coherence which is essentially the squared cross–correlation between oscillations at different frequency of a pair of time series, and it is one of the most common measures of such dynamics ([51, 67]). Coherence between all channels at a specific frequency (or frequency band) is represented by a matrix whose off–diagonal element is an estimate of functional interactions between two brain regions in each frequency. One of the properties of the coherence matrix is that it is symmetric positive definite (SPD), and thus we aim to develop methods that respects this important property. Visualizing and extracting useful information from a sample of coherence matrices is challenging especially in high dimensions because each coherence matrix must be considered as a whole information unit. There are recent advances on analyzing SPD matrices, such as [70, 56, 26, 80, 13] and [33], but not enough consideration has been given to visualization of such SPD matrices. [30] introduced a surface boxplot for visualization and exploratory analysis of samples of images, but this method does not take into account the non–Euclidean geometry of the space of SPD matrices.

To overcome this limitation, we propose a geometric surface boxplot for samples of coherence matrices. By using the geometric surface boxplot, we address two primary objectives. The *first* objective is to determine the median coherence matrix by ordering a sample of matrices based some depth measure. Here, the median matrix is that which is the deepest (in a sense that it is "covered" by the most number of pairs of matrices). The *second* objective is to detect outlying matrices again based on the ordering of the depth values of these matrices. These outliers are demonstrated by unusual connectivity between pairs of signals and hence would warrant further investigation. Subsequently, confirmed outliers, if necessary, can be removed in subsequent analyses. Thus our method pinpoints those epochs when these unusual coherence matrices were observed.

One can define a multivariate order statistic for sample of SPD matrices by ranking the data points according to their depth. Then a matrix–valued object has a high rank when it is near the center whereas point far from the center has lower rank. As studied in [43] and [44] for functional data, the test ranks of data points of the second sample are computed with respect to their depth in the first sample. Using this concept and notion of data depth, we develop two rank–based tests for two samples of matrix–valued objects under the homogeneity hypothesis, i.e., the two samples of matrix–valued objects come from the same distribution.

1.2 Contributions

The main contributions of this dissertation are as follows:

1. A functional boxplot to visualization and exploratory analysis for samples of curves (log periodograms of EEG time series data).

2. A geometric surface boxplot to visualization and exploratory analysis for samples of

SPD matrices (coherence matrices).

3. Non–parametric hypothesis testing procedures to test for difference between two populations or groups of curves and SPD matrices.

1.3 Outline

This dissertation is organized as follows. In Chapter 2, we present functional boxplot with an application on a motor skill acquisition study at the Neuro-rehabilitation laboratory at the University of California, Irvine (Principal Investigator: Steven C. Cramer). In Chapter 3, we introduce geometric surface boxplot with an application on infantile spasms dataset from Children Hospital of Orange County. Non–parametric hypothesis testing procedures to test for difference between two populations or groups of curves and SPD matrices are also presented in Chapter 2 and 3, respectively. My future works are discussed in the last Chapter 4.

Chapter 2

Functional Boxplot

2.1 Introduction

The methods presented in this chapter are motivated by a motor skill acquisition study at the Neuro-rehabilitation laboratory at the University of California, Irvine (Principal Investigator: Steven C. Cramer). In the previous study, EEG was recorded from 17 subjects both during resting-state prior to motor skill training and during motor skill training using dense-array EEG (256 electrodes) as shown in Figure 2.1. The resting-state EEG exam was three minutes, and during post-processing, was segmented into one-second non-overlapping epochs. As demonstrated in [79], the spectral features of the resting-state EEGs when combined with a partial least squares regression analysis, was predictive of an individual's subsequent ability to acquire a novel motor skill. These may be of clinical importance to the field of rehabilitation, as improved methods for stratifying patients may significantly improve response to treatment and assist allotment of limited resources. We present an exploratory spectral analysis (ESA) of resting-state EEG traces using functional boxplots for one subject. In spectral analysis, the spectrum is an important stochastic property of

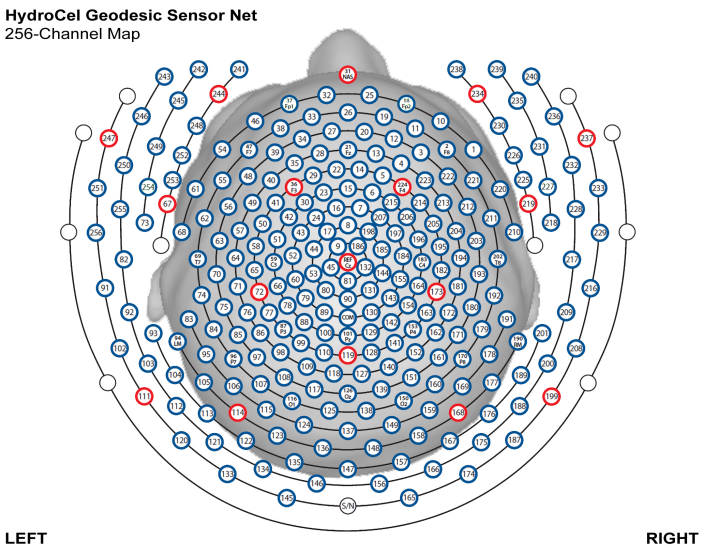


Figure 2.1: Map of channels on the scalp.

the signal. It indicates the amount (or proportion) of variance that is explained by each frequency bin. Thus the spectrum or the log spectrum of the EEG signal can be used to examine relative amounts of variability explained by slow (delta or theta) waves and fast (alpha or beta) waves. Throughout this analysis, we obtain a *sample* spectral curve by smoothing the log periodograms of each one-second EEG epoch, and treat it as one observation unit in the FBP. By using the FBP, we address three primary objectives. The *first* objective is to identify the median, i.e., the most characteristic spectral curve rather than the pointwise frequency-specific medians. In addition, outliers are demonstrated by their unusual sample log spectral curve, and can be caused by extra-brain artifacts, including eye blinks, eye movements and muscle movements in the EEG signal. Subsequently, confirmed outliers will be removed from subsequent analyses. The advantage of the FBP approach, over the usual pointwise boxplot method, is that it identifies epochs that have potential outlying spectral curves.

The *second* objective is to compare the median curves and the variability of the spectral curves from multiple phases of the resting state period. To test the stationarity of the EEG signal over the entire recording, we compare the spectral curves and the frequency-specific

spatial distribution of spectral power during the early phase (first 60 epochs) versus the late phase (last 60 epochs). Evidence against stationarity must be taken seriously since this would suggest an evolution of brain processes across the recording ([21]). Moreover, the FBP approach is able to provide some characterization of the variation of the sample log spectral curves across EEG recording. In experiments comparing more than one group (e.g., healthy controls vs patients with stroke), it would be also interesting to determine whether groups differ with respect to consistency (uniformity) of the EEG signal over time.

The *third* objective is to investigate the spatial variability of spectral power across the brain for a given frequency band using the surface boxplot, which is a generalization of the FBP. Using the surface boxplots approach, it is possible to identify cortical regions (or channels) that, relative to the other channels, exhibit a high proportion of beta power. The beta band is particular interest to neuroscientists, as changes in beta activity have a good association with motor function ([63] and [39]).

The remainder of the chapter is organized as follows. In Section 2.2, we present a comprehensive exploratory method which consists of the following: a review of the spectra in Section 2.2.1, a demonstration of automatic bandwidth selector for periodogram smoothing using the gamma generalized crossvalidation criterion in Section 2.2.2, some remarks on smoothing the periodogram in Section 2.2.3, a description of the functional boxplots in Section 2.2.4, a description of the surface boxplots in Section 2.2.5, and a demonstration of testing for differences in mean curves between families of curves in Section 2.2.6. In Section 2.2.3, we examine the finite sample performance of the proposed exploratory method. In Section 2.2.4, the resting-state EEG data are analyzed. Finally, in Section 2.2.5, conclusions and future work are discussed.

2.2 Method for Exploratory Spectral Analysis (ESA)

In this section, we review the methods that are needed for ESA of the EEG data. In Section 2.2.1, we first formally define the spectrum and then discuss a consistent estimator which is obtained by smoothing the periodogram using a bandwidth that is automatically selected by the gamma generalized cross-validation (Gamma-GCV) method described in Section 2.2.2. Next, we highlight two remarks on smoothing the periodogram in Section 2.2.3, then we present the functional boxplots method in Section 2.2.4 and surface boxplots method in Section 2.2.5. Finally, we present a rank sum test which tests for differences in median curves or surfaces between families of curves or surfaces in Section 2.2.6.

2.2.1 Spectrum

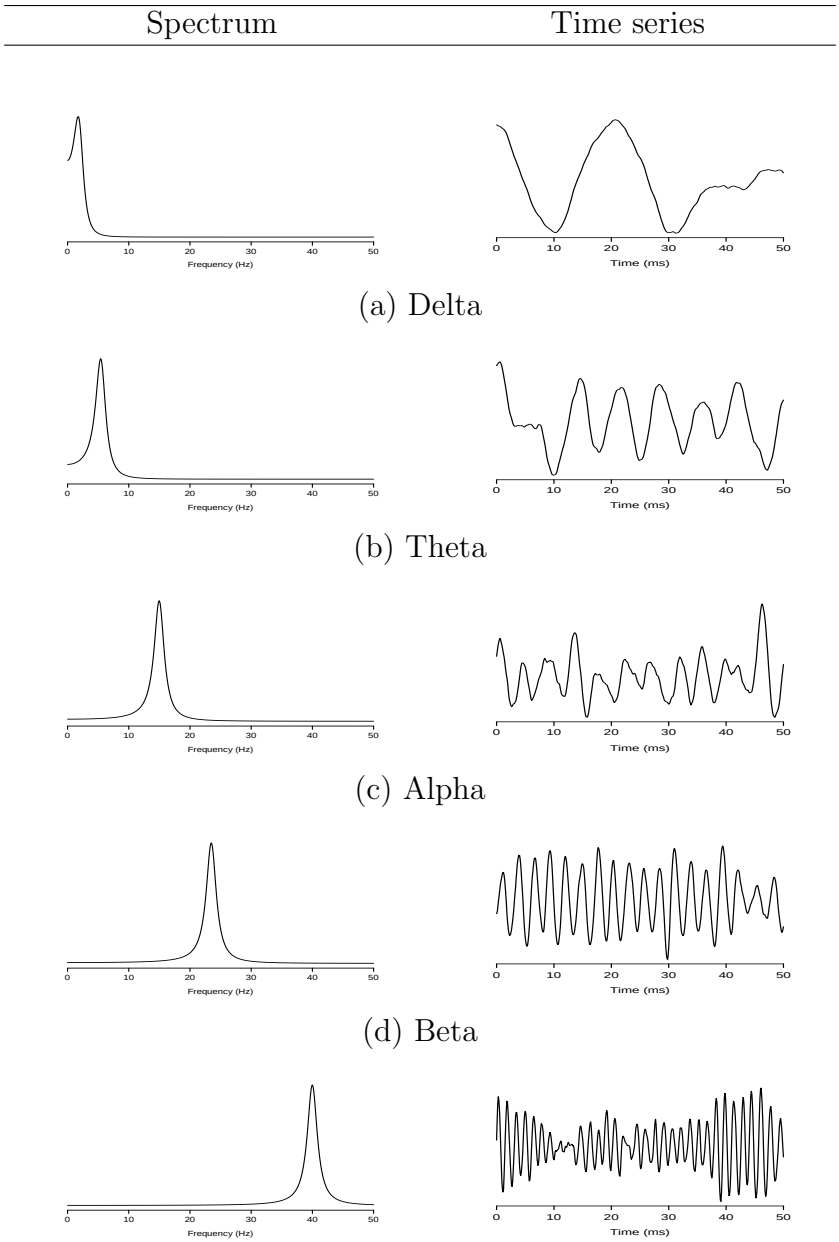
The spectrum of an EEG signal (which is assumed to be stationary) can give the amount of variance contributed by oscillatory components (from delta to beta band activity). Let $X(t),\ t = \ldots, -1, 0, 1, \ldots$ be a zero-mean stationary time series with covariance function $\gamma(\tau) = E\left(X(t)X(t+\tau)\right) (\tau = \ldots -1, 0, 1, \ldots)$ that is assumed to be absolutely summable, i.e., $\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| < \infty$. The spectrum, denoted $f(\omega)$, is defined to be

$$f(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\ e^{-i2\pi\omega\tau}, \quad \omega \in \left[-\frac{1}{2}, \frac{1}{2}\right].$$

The starting point for estimating $f(\omega)$ is the periodogram. Denote $I(\omega_k)$ to be the periodogram computed from a finite sample of the stationary process $X(0), X(2), \ldots, X(T-1)$ at frequency $\omega_k = k/T$ which is defined to be

$$I(\omega_k) = \frac{1}{T}\left|\sum_{t=0}^{T-1} X(t)\ e^{-i2\pi\omega_k t}\right|^2, \quad k = -[\![T/2]\!] - 1, \ldots, [\![T/2]\!],$$

where $[\![T/2]\!]$ is the quotient of $T/2$. To characterize the spectra of the EEG signals, we



(a) Delta

(b) Theta

(c) Alpha

(d) Beta

(e) Gamma

Figure 2.2: Left: the spectrum of second order auto-regressive processes AR(2) with power concentrated at the delta (0-4 Hertz), theta (4-8 Hertz), alpha (8-16 Hertz), beta (16-32 Hertz) and gamma (32-50 Hertz) bands. Right: realizations from each corresponding AR(2) process.

classify the oscillatory patterns of periodograms into four primary frequency bands: delta (0-4 Hz), theta (4-8 Hz), alpha (8-16 Hz), beta (16-32 Hz) and gamma (32-50 Hz) as shown

in Figure 2.2. Since each frequency band is defined by a range, we define $\widehat{S}(\Omega)$ to be the estimated spectral power at the Ω band:

$$\widehat{S}(\Omega) = \int_{\omega \in \Omega} I(\omega) d\omega.$$

It is well known that the periodogram $I(\omega_k)$ is an asymptotically unbiased estimator for $f(\omega_k)$, but it is inconsistent because its variance approaches a positive constant when $T \to \infty$. Therefore, to reduce the variance, we smoothed the periodogram. A number of nonparametric smoothing methods have been proposed including the kernel smoother ([42, 52]), wavelet ([28]), smoothing spline ([76, 55]), or local polynomial ([20]). For kernel smoothing, ([52]) developed an automatic span selector via the generalized crossvalidation criterion for generalized additive models based on the *deviance* which is discussed in Section 2.2.

2.2.2 Automatic Span Selector Using the Gamma Generalized Crossvalidation Method

From [6] (Theorem 5.2.6), $I(\omega_k)$ follows an asymptotic distribution

$$I(\omega_k) \sim \begin{cases} Gamma(1, f(\omega_k)) & k = 1, \ldots, T/2 - 1 \\ Gamma(\frac{1}{2}, 2f(\omega_k)) & k = 0, T/2, \end{cases}$$

where $I(\omega_0), \ldots, I(\omega_{T/2})$ are independent. As a caveat, we note here that the actual result requires that the number of frequencies is fixed and does not depend on T. However, in most applications, this is often ignored. This result can be equivalently stated as $I(\omega_k)/f(\omega_k) \sim \epsilon_k$ where $\epsilon_k \dot{\sim} \chi^2(1)$ when $k = 0$ or $T/2$ and $\epsilon_k \dot{\sim} \frac{1}{2}\chi^2(2)$ when $k = 1, \ldots, T/2 - 1$. As noted, we need to smooth the periodogram $I(\omega_k)$ to produce a consistent estimator for $f(\omega_k)$. Let

$\widehat{f}_p(\omega_k)$ be a smoothed periodogram estimator of $f(\omega_k)$ which we define to be

$$\widehat{f}_p(\omega_k) = \sum_{j=-p}^{p} W_{p,j} I(\omega_{k+j}) \quad k = 0, \ldots, T/2, \text{and } j = -p, \ldots, p$$

where $2p+1$ is the smoothing span and $W_{p,j}$ are nonnegative weights that satisfy the following conditions for any fixed p:

$$W_{p,j} = W_{p,-j} (j = 1, \ldots, p), \sum_{j=-p}^{p} W_{p,j} = 1.$$

Generally, the weights are chosen so that $W_{p,j}$ is a decreasing function of p, but [60] shows that the choice of the weights $W_{p,j}$ is of secondary importance to the value of the span or bandwidth. Thus, for simplicity, we use the boxcar smoother with weights defined by $W_{p,j} = 1/(2p+1)$ for all $j = -p, \ldots, p$. The gamma generalized crossvalidation method selects p to minimize the generalized crossvalidated deviance function

$$GCV(p) = \frac{M^{-1} \sum_{j=0}^{M-1} D(I(\omega_j), \widehat{f}_p(\omega_j))}{(1 - \text{tr}(H_p)/M)^2},$$

where $M = T/2+1$. The deviance $D(I(\omega_j), \widehat{f}_p(\omega_j))$ can be chosen as $q_j\{-\log(I(\omega_j)/\widehat{f}_p(\omega_j)) + (I(\omega_j) - \widehat{f}_p(\omega_j))/\widehat{f}_p(\omega_j)\}$ ([49]). Here, $q_j = 1 - 0.5\mathcal{I}\{j = 0, M-1\}$, and $\mathcal{I}$ is the indicator function. The H_p is the smoother matrix with smoothing parameter p, and the term $(1 - tr(H_p)/M)^2$ often referred to as the model degrees of freedom, can be expressed in terms of the weight at the center of the smoothing window: $(1 - W_{p,0})^2$. Then, the generalized crossvalidated deviance function can be written as

$$GCV(p) = M^{-1} \sum_{j=0}^{M-1} q_j \left\{ \frac{-\log(I(\omega_j)/\widehat{f}_p(\omega_j)) + (I(\omega_j) - \widehat{f}_p(\omega_j))/\widehat{f}_p(\omega_j)}{(1 - W_{p,0})^2} \right\}.$$

2.2.3 Remarks

For frequencies over 100 Hz, the periodogram values are almost negligible because the signals underwent low–pass filtering at 100 Hz. , so for simplicity, we will only show the spectrum over the frequency range of $0-100$ Hertz. In Figure 2.3, we show the location of channel 197 in right pre-motor region at the resting-state. Figure 2.4 gives an illustration of smoothing the periodograms for randomly selected epochs 3, 85 and 160 for a fixed channel 197. It can be seen that the power at these periodograms are dominated by low frequencies, and the values of smoothing span minimizing the generalized crossvalidated deviance function are

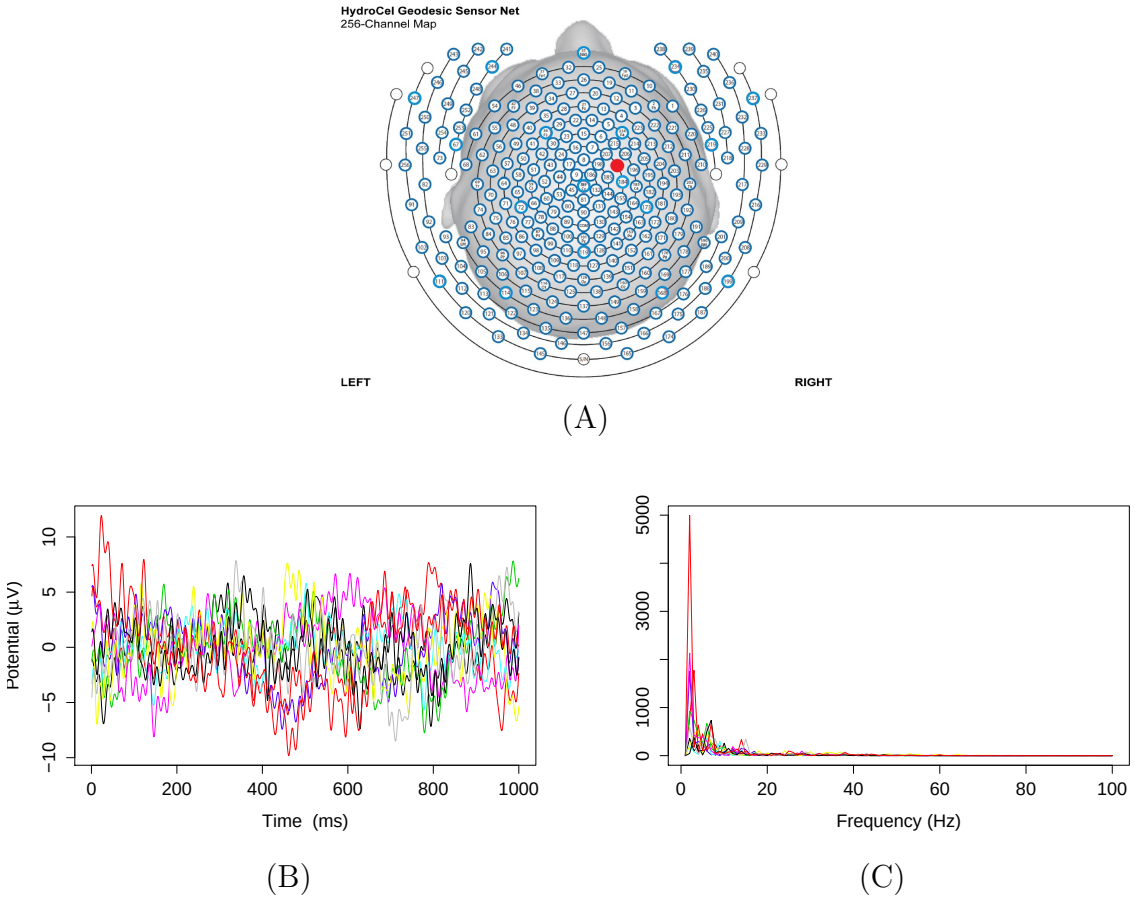
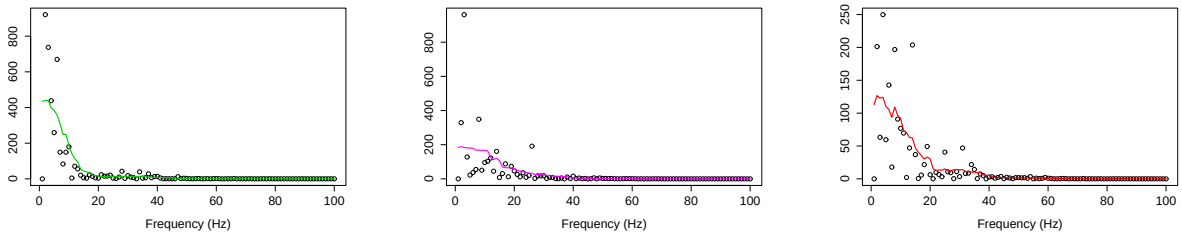


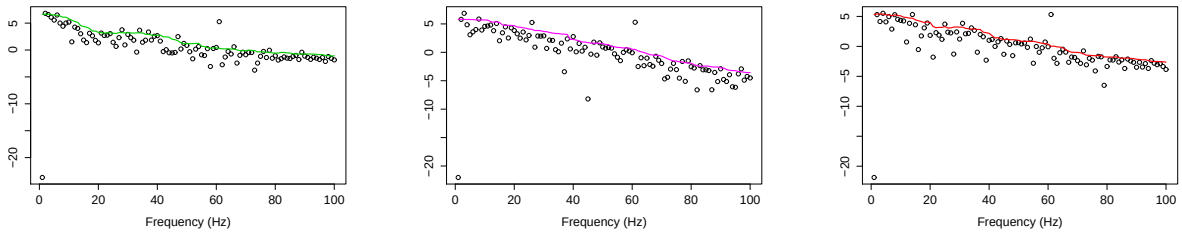
Figure 2.3: EEG time series and raw periodograms after filtering out frequency 60 HZ by averaging method of channel 197 (right pre-motor region) for the first 10 traces. (A) Red circle represents right pre-motor region (channel 197) on the scalp, (B) EEG time series of right pre-motor region (channel 197) and (C) Raw periodograms after filtering out frequency 60 HZ by averaging method.

about 3-5. Also, the smoothing lines reasonably approximate the periodograms and the small bandwidths preserve the peaks. Second, since the distribution of $I(\omega_k)$ is a multiple of the



(A) (B) (C)

Figure 2.4: Smoothing periodograms at randomly selected epochs: (A) Epoch 3, (B) Epoch 85 and (C) Epoch 160 of channel 197 (in the right pre-motor region) using the bandwidth that was automatically selected by the gamma generalized crossvalidation (gamma-GCV) method.



(A) (B) (C)

Figure 2.5: Log bias-corrected periodograms of epochs: (A) Epoch 3, (B) Epoch 85 and (C) Epoch 160 from Channel 197 (Right pre-motor region)

spectral density, its variance (which depends on $f(\omega_k)$) also changes across the frequencies ω_k. To stabilize the variance across frequencies and to standardize comparisons of median curves across two phases (early vs late phases of the resting-state EEG recording) we will use the log transformed periodograms. It is convenient then, that the variance of the log periodograms at each frequency is constant and takes the approximate value of $\frac{\pi^2}{6}$. Moreover, while the periodogram is approximately unbiased for the spectrum, the log periodogram is no longer (approximately) unbiased for the log spectrum due to Jensen's inequality. This is easily fixed by adding the Euler Mascheroni constant 0.57721 to log transformed periodograms to obtain

the log bias-corrected periodograms ([76]). Let $g(\omega_k)$ be the true log spectrum, then $Y_r(\omega_k)$, the log bias of the corrected periodogram at epoch r, is defined as

$$Y_r(\omega_k) = g(\omega_k) + 0.57721, \quad k = 0, 1, \ldots, T/2.$$

Figure 2.5 gives the log bias-corrected periodograms, $Y_r(\omega_k)$, corresponding to Figure 2.4. Throughout this chapter, we will apply the gamma crossvalidation method to obtain the optimal smoother of log bias-corrected periodograms.

2.2.4 Functional Boxplots

The functional boxplot is constructed in a similar manner to the classical (pointwise) boxplot. Each observation will be sorted based on decreasing values of some depth measure, and band depth is one notion. A curve is said to be "deeply situated" within a sample of curves if it is covered by many bands from pairs of curves. This idea is an extension of a pointwise boxplot where the median is also located "deep" in a sample because it is situated in the middle of the boxplot and hence covered by many pairs of points. Here, our observation units are curves (or real-valued functions) which are the log bias-corrected periodograms $Y_r(\omega_k)$, $k = 0, \ldots, T/2$ over many epochs r. The notion of a band depth was introduced in [46] through a graph-based approach to order all sample curves which we briefly describe. Suppose that a curve $Y(\omega_k)$ is the subset of the plane $G(Y(\omega_k)) = \{(\omega_k, Y(\omega_k)) : \omega_k \in \mathcal{A} = [0, T/2]\}$. The band in $\mathbb{R}^2$ can be delimited by a number J of curves, and this number is fixed as $J = 2$ in our study. Now, let Y_α, Y_β be two continuous functions, $L_k = \min(Y_\alpha(\omega_k), Y_\beta(\omega_k))$, and $U_k = \max(Y_\alpha(\omega_k), Y_\beta(\omega_k))$. Then the band delimited by Y_α, Y_β is

$$B(Y_\alpha, Y_\beta) = ((\omega_k, Y'(\omega_k)) : \omega_k \in \mathcal{A}, L_k \leq Y'(\omega_k) \leq U_k)\,.$$

Let $Y_1, \ldots, Y_n$ be n independent sample curves, then the band depth for a given curve $Y_i, i = 1, \ldots, n$ is defined as

$$BD(Y_i) = \binom{n}{2}^{-1} \sum_{\alpha=1,\ldots,n;\ \beta=1,\ldots,n} \mathcal{I}\{G(Y_i) \subseteq B(Y_\alpha, Y_\beta)\}$$

where $\mathcal{I}(\cdot)$ is the indicator function. When $J = 2$, there are $\binom{n}{2}$ possible bands delimited by two curves. The limit of the band depth BD is that it does not measure the proportion of curve inside the band. Thus, [46] also proposed a modified band depth method (MBD), which measures the proportion of a curve Y_i that is actually in a band:

$$MBD(Y_i) = \binom{n}{2}^{-1} \sum_{\alpha=1,\ldots,n;\ \beta=1,\ldots,n} \lambda\{A(Y_i; Y_\alpha, Y_\beta)\}$$

where $A(Y_i; Y_\alpha, Y_\beta) \equiv \{\omega_k \in \mathcal{A} : L_k \leq Y_i \leq U_k\}$, $\lambda(Y_i) = \lambda(A(Y_i; Y_\alpha, Y_\beta))/\lambda(\mathcal{A})$, and λ is a Lebesgue measure on $\mathcal{A}$. We notice that the MBD computation will be time-consuming when n is large, so we use an exact fast method from [75] to compute the MBD for the EEG data.

Based on the ranks of the depths of the curves, the functional boxplots can provide the descriptive statistics, such as the 50% central region, the median curve, and the maximum and minimum non-outlying curves. Moreover, the potential outliers can be detected by the 1.5 times inter-quartile range (IQR) empirical rule, which is commonly used for classical boxplots. The boundary region is defined as 1.5 times the height of the 50% central region. Any curves outside this region are considered potential outliers. In contrast with a constant factor 1.5 in classical boxplot, a factor 1.5 in functional boxplot can be modified due to potential spatio-temporal outliers. This is because the curves from different locations will be spatially correlated, and there can be dependence in time/frequency for each curve ([73]).

2.2.5 Surface Boxplots

Similar to functional boxplots, one can compute the data depth of all the observations, then order them according to decreasing depth values. Suppose that the observed sample surfaces, $z_1(s), \ldots, z_n(s), s \in \mathcal{S}$, where $\mathcal{S}$ is a region in $\mathbb{R}^2$. The information unit for such a dataset is the entire surface. To order sample surfaces, we need to generalize univariate order statistics to surfaces. To this end, we generalize the MBD with $J = 2$ to $\mathbb{R}^3$ through a volume. [30] define the sample modified volume depth (MVD) to be

$$MVD_n(z) = \binom{n}{2}^{-1} \sum_{1 \le i_1 \le i_2 \le n} \lambda_r A(z; z_{i_1}, z_{i_2}),$$

where $A(z; z_{i_1}, z_{i_2}) \equiv \mathbf{s} \in \mathcal{S} : \min_{r=i_1,i_2} z_r(\mathbf{s}) \le z(\mathbf{s}) \le \max_{r=i_1,i_2} z_r(\mathbf{s})$, and the measurement $\lambda_r(z) = \frac{\lambda(A(z;z_{i_1},z_{i_2}))}{\lambda(\mathcal{S})}$, if λ is the Lebesgue measure on $\mathbb{R}^3$. A sample median surface is a surface from the sample with the largest sample modified volume depth value, designed by $\arg\max_{z \in z_1,\ldots,z_n} MVD_n(z)$. If there are ties, the median will be the average of the surfaces maximizing the sample modified volume depth.

The first step for constructing surface boxplots is the surface ordering. Sample surfaces are ordered from the center outwards based on their MVD values, inducing the order $z_{[1]}, z_{[2]}, \ldots, z_{[n]}$. The sample α central region is naturally defined as the volume delimited by the α proportion ($0 < \alpha < 1$) of the deepest surfaces. In particular, the sample 50% central region is

$$C_{0.5} = \{(\mathbf{s}, z(\mathbf{s})) : \min_{r=1,\ldots,[n/2]} z_{[r]}(s) \le z(\mathbf{s}) \le \max_{r=1,\ldots,[n/2]} z_{[r]}(s)\},$$

where $[n/2]$ is the smallest integer not less than $n/2$. The border of the 50% central region is defined as the inner envelope representing the box in a surface boxplot. This is the surface analogue of the first and third quartiles of the classical boxplot. The median surface in the box is the one with the largest depth value. Because the ordering is from the center outwards,

the volume of the central region increases as α increases. Hence, the maximum envelope, or the outer envelope, is defined as the border of the maximum non-outlying central region. To determine this region, we propose to identify outlying surfaces by an empirical rule similar to the 1.5 times the 50% central region rule in a functional boxplot. The fences (or the upper and lower surface boundaries for flagging potential outliers) are obtained by inflating the inner envelope (as defined above) by 1.5 times the height of the 50% central region. Any surface crossing the fences are flagged as potential outliers. The factor 1.5 can be also adjusted as in the adjusted functional boxplots to take into account spatial autocorrelation and possible correlations between surfaces.

2.2.6 Testing for Differences in Median Between Families of Curves or Surfaces

To compare the median curves from two populations of curves, [46] proposed the rank sum test. Let $\tilde{\mu}_Y$ and $\tilde{\mu}_{Y'}$ be the median curves of two populations Y and Y' respectively. Define the null hypothesis to be

$$H_0 : \tilde{\mu}_Y = \tilde{\mu}_{Y'} \text{ for all } \mu.$$

Suppose that we observe two sets of curves, namely $\{y_1, \ldots, y_n\}$ and $\{y'_1, \ldots, y'_m\}$. Then define the reference sample to be $\{r_1, \ldots, r_k\}$ which is from one of the two observed sets with $k \geq \max(n, m)$. The position of a particular y_i for $i = 1, \ldots, n$, or y'_j for $j = 1, \ldots, m$ with respect to the reference sample r, is defined as

$$R(y_i) = \frac{1}{n} \sum_{l=1}^{n} \mathcal{I}\{MBD(z_l) \leq MBD(y_i)\},$$

$$R(y_j') = \frac{1}{m}\sum_{l=1}^{m} \mathcal{I}\{MBD(z_l) \leq MBD(y_j')\},$$

where MBD is the modified band depth defined in previous section, and $\mathcal{I}$ is the indicator. Then we can order the values $R(y_i)$ and $R(y_i')$ from the smallest to the largest, and their ranks are between 1 and $n+m$. The test statistics $T = \sum_{l=1}^{m} \operatorname{rank} R(y_j')$, then under the null hypothesis H_0, the distribution of T is the distribution of the sum of m numbers that are randomly chosen from $1, 2, \ldots, n+m$ ([74]).

2.2.7 Remarks on the applications of functional and surface boxplots

In this chapter, we use functional and surface boxplots to explore the structure of EEGs. However, these methods are general and can be applied to other types of data such as growth data and climate time series ([74]).

2.3 Simulation Study

The purpose of the simulation study is to examine the performance of the exploratory spectral methods under various experimental settings. In Section 2.3.1, we demonstrate the performance of the FBP on the smoothed log periodograms of a mixture of two first order autoregressive time series, denoted AR(1). In Section 2.3.2, we illustrate the rank sum test to compare the functional median from two families of curves.

2.3.1 Functional Boxplot Simulation Study

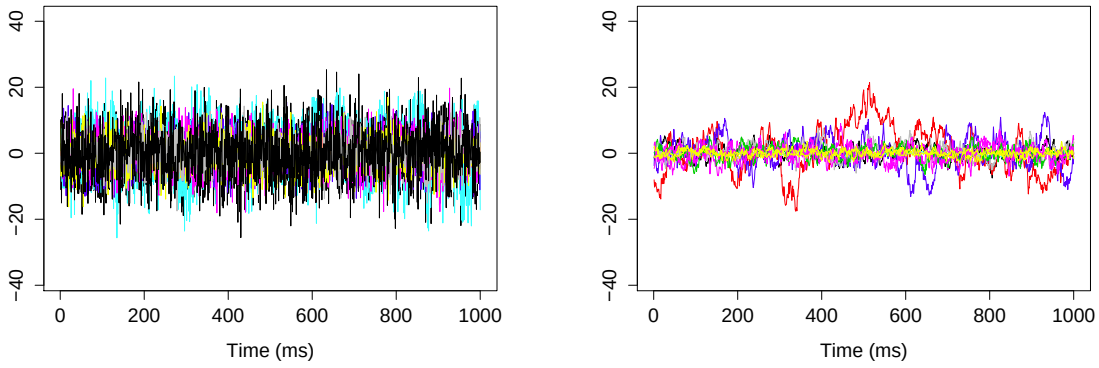
For the r^{th} epoch, let $U_{1r}(t)$ be an AR(1) process with its spectra dominated by high frequencies and $U_{2r}(t)$ be another AR(1) with its spectra mostly containing low frequencies. The AR(1) parameters are allowed to vary across epochs. Here, we set $t \in T = \{1, \ldots, 1000\}$. We define $X_r(t)$ to be the mixture of $U_{1r}(t)$ and $U_{2r}(t)$, such that

$$X_r(t) = a_{1r}U_{1r}(t) + a_{2r}U_{2r}(t)$$

where $r = 1, \ldots, 220$, a_{1r} and a_{2r} are weighted coefficients of $U_{1r}(t)$ and $U_{2r}(t)$ respectively. Then, the model for high and low frequency AR(1) processes are defined as

$$U_{\ell r}(t) = \phi_{\ell r}U_{\ell r}(t-1) + W_{rt}$$

where $\ell = 1, 2$ and $W(t)$ is white noise. In this setting, the high and low frequency AR(1)



(A) Group 1 (B) Group 2

Figure 2.6: Time series AR(1) for group 1 and group 2

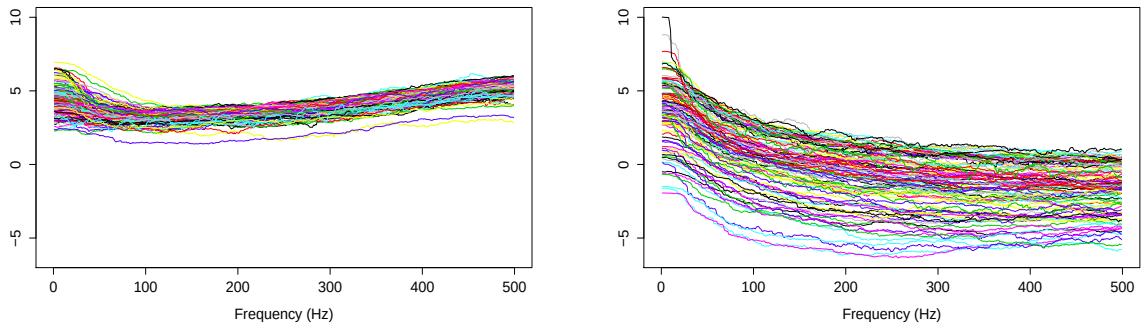
are distinguished by the value of $\phi_{\ell r}$. For example, for high frequency $U_{1r}(t)$, we set $\phi_{1r} = 0.9 + \xi_r$, where ξ_r are independent and identically distributed from $\mathcal{N}(0, 0.001)$. Similarly, for low frequency $U_{2r}(t)$, we set $\phi_{2r} = -0.5 + \eta_r$, and η_r are also independent and identically

distributed from $\mathcal{N}(0, 0.001)$. Here, we need the variance of ξ_r and η_r to be small so that it guarantees causality, i.e. $\xi_r \in (-1, 1)$ and $\eta_r \in (-1, 1)$. Next, we split the 220 subjects into two groups, such that the first group will include both high and low frequency series, $U_{1r}(t)$ and $U_{2r}(t)$, while the second group will only have the high frequency series $U_{1r}(t)$. To split $X_r(t)$ into two groups, we set the weight coefficients a_{1r} and a_{2r} as following

$$a_{1r} \sim \mathcal{N}(10, 1) \text{ for } r = 1, \ldots, 220$$

$$a_{2r} \sim \mathcal{N}(5, 1), \text{ for } r = 1, \ldots, 120, \text{ and } a_{2r} \sim \mathcal{N}(0, 0.001) \text{ for } r = 121, \ldots, 220.$$

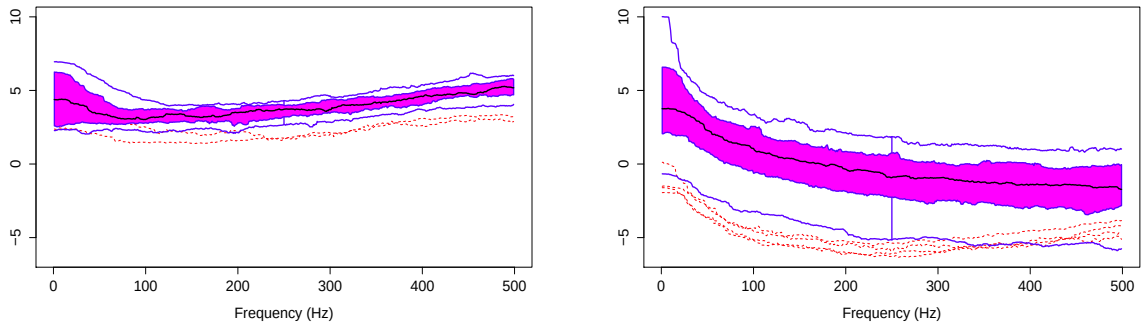
The two groups of $X_r(t)$ are shown in Figure 2.6. Using the gamma generalized crossvali-



(A) Group 1 (B) Group 2

Figure 2.7: Smoothed log bias-corrected periodograms for Group 1 and Group 2

dation method, Figure 2.7 displays the log bias-corrected periodograms for each group, and Figure 2.8 shows the corresponding FBPs. Note that group 1 is dominated by both high (right) and low (left) frequencies while group 2 includes only low frequencies. Thus, the functional median of group 1 should have two peaks, one each in high and low frequency ranges, while the functional median of group 2 has only one peak in the low frequency range. In Figure 2.8, the black curve is the median curve in the center of the functional boxplot. The two median curves from each group have clearly summarized the typical power distribution for each group. The blue curves in the center form the envelope of the 50% central



(A) Group 1 (B) Group 2

Figure 2.8: Functional boxplots of Group 1 and Group 2 with a black curve representing the median curve, the pink area denoting the 50% central region, the two inside blue curves indicating the envelopes of 50% central region, the two outside blue curves representing for two non-outlying extreme curves, and the red dashed curves illustrating the outlier candidates detected by 1.5 times the 50% central region rule.

region. The blue curves outside of the 50% central region are the non-outlying minimum and maximum curves. It is worth remarking that the envelope of group 1 is smaller than the envelope of group 2, and therefore, we demonstrate that group 2 has more dispersion than group 1. Moreover, the envelope of group 1 is in the middle of the non-outlying minimum and maximum curves, while the envelope of group 2 tends to move upwards. This indicates that group 2 shows more skewness than group 1. The red dashed curve in Figure 2.8 denotes the outliers. We see that the curves from group 1 that are dominated by high frequencies only are detected as outliers while the curves from group 2 that include both high and low frequencies are detected as outliers.

In order to illustrate the usefulness of the functional boxplot compared to the pointwise boxplot, we introduce a simulation study which randomly chooses 10 bias-corrected log periodograms among 160 total periodograms. We simulate an outlying curve by adding additional noise across the 0-100 Hz frequency range, and close to the center for the remaining frequencies. Figure 2.9-a shows the simulation data including the 10 random bias-corrected log periodograms (grey curves) and a simulated outlying curve (red curve). In Figure 2.9-b, the functional boxplot successfully detects the simulated outlying curve and other outliers.

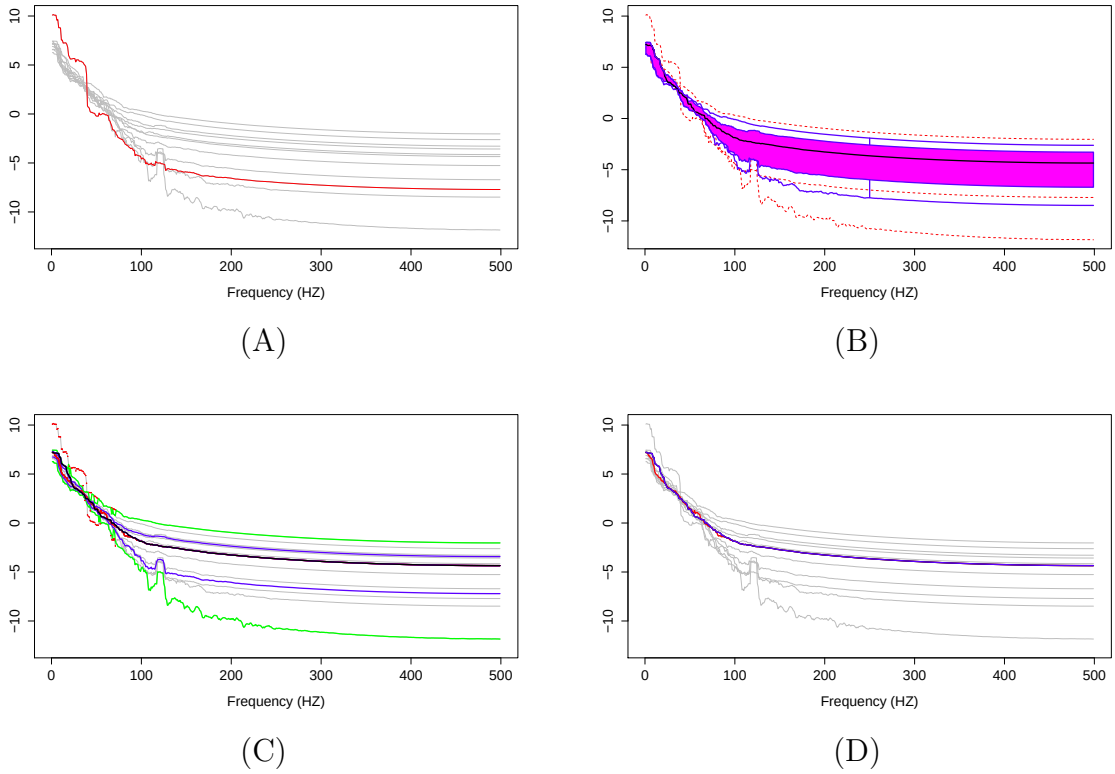


Figure 2.9: (A) Simulation data with grey curves representing sample curves, a red curve denoting the simulated outlying curve (B) Functional boxplots, (C) Pointwise boxplots with black curve representing a mean curve, blue curves for the envelope of the 50% central region, the green curves for the non-outlying minimum and maximum curves, and the red points for outliers, and (D) two median curves obtaining by functional boxplots method (blue) and pointwise boxplot method (red) are shown in the same plot.

However, Figure 2.9-c shows that the pointwise boxplot fails to detect the simulated outlying curve, and provides some disconnected outlying curves across frequencies. We also notice that the non-outlying maximum and minimum curves of pointwise boxplot are actually the outlying curves detected by functional boxplot. Figure 2.9-d compares the two median curves from these two methods, and by visual inspection, there is a slight difference between the two median curves at low frequencies. Thus, functional boxplot can be a non-parametric method to obtain the median curve and the variability around it for EEG data compared to pointwise boxplot.

2.3.2 Rank Sum Test Simulation Study

To investigate the performance of this nonparametric test, we simulated two sets of curves, which are defined as below:

$$Y_{\ell,r}(\omega_k) = f_\ell(\omega_k) + a_r g(\omega_k) + h_r(\omega_k),$$

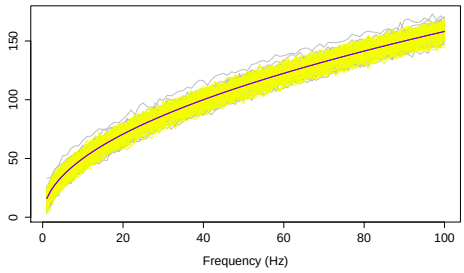
where $r = 50$, $\ell = 1, 2$, $g(\omega_k) = 1$ for all ω_k, and ω_k is defined as $\omega_k = k/100$, where $k = 1, \ldots, 100$. In the model, $f_1(\omega_k)$ and $f_2(\omega_k)$ are the mean functions; $a_r \overset{\text{iid}}{\sim} N(0, 5)$ and $h_r(\omega_k) \overset{\text{iid}}{\sim} N(0, 2)$ represent the variation between and within the curves, respectively. Let the function $f_1(\omega_k)$ be defined as

$$f_1(\omega_k) = 5 \cdot \sqrt{1000 \cdot \omega_k},$$

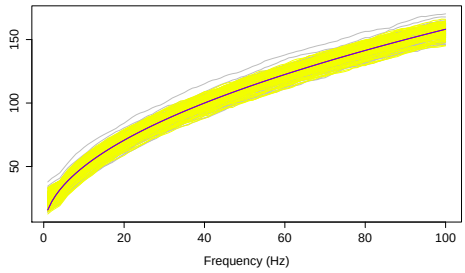
and consider three different cases:

1. The two means are identical, let $f_2(\omega_k) = f_1(\omega_k)$ for all ω_k.
2. There is a slight deviation between the two means; define $f_2(\omega_k) = 5 \cdot \sqrt{900 \cdot \omega_k}$.
3. There is an appreciable deviation between the two means; let $f_2(\omega_k) = f_1(\omega_k) + 2k/3$.

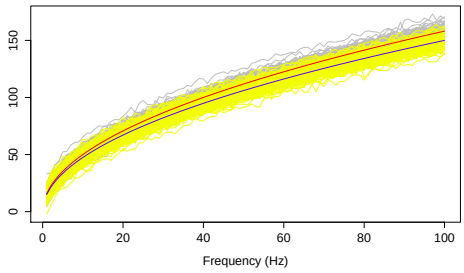
We applied the kernel average smoother with window size 7 to smooth each curve from these two families. Figure 2.10 illustrates the simulated curves (left panel) and the smoothed curves (right panel). In order to investigate the rank sum test performance in each case, we simulated two families of curves and obtain p-values of rank sum test; this procedure was repeated 1000 times. Let the type I error α be 5%, we report the percentage of time that the rank sum test rejects $H_o : f_1(\omega_k) = f_2(\omega_k)$ for all ω_k in Table 2.1. Overall, the rank sum test method performed well in each case. When the two families are identical, this method rejected the null hypothesis of equality only 44 times (4.4%) out of 1000 times, which is close



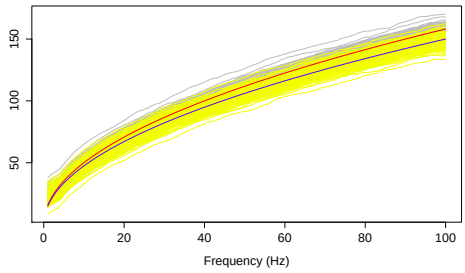
(A) Case 1



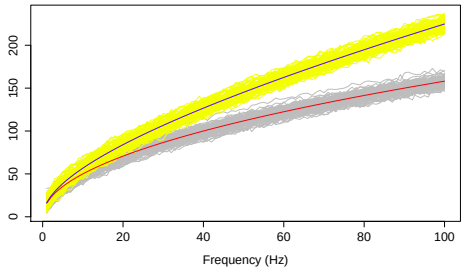
(B) Case 1: Smoothed



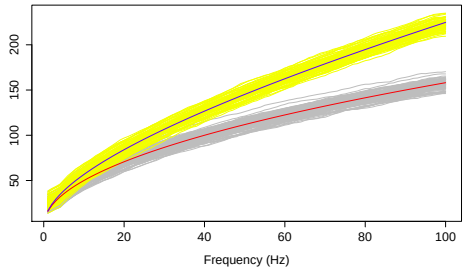
(A) Case 2



(B) Case 2: Smoothed



(A) Case 3



(B) Case 3: Smoothed

Figure 2.10: The two families of simulated curves, $Y_{1,r}$ and $Y_{2,r}$. The grey shaded area represents the first family, and the yellow shaded area is for the second family. The red and blue lines are the first and second mean functions, f_1 and f_2, respectively.

to the nominal α. When the two families are nearly identical, this method rejects 605 times (the power is 60.5%), and when the two families are completely different, the power is 100%. Thus this method demonstrates power and sensitivity to differences.

	First case	Second case	Third case
Percentage of time rejecting H_o	44	605	1000

Table 2.1: Rank Sum test study result.

2.4 Analysis of Resting-State EEGs Data

2.4.1 Data Description

In this chapter, we analyze EEG data from one participant in a resting-state EEG study approved by the Institutional Review Board of the University of California, Irvine. The over-arching aim of this study was to identify a pattern of EEG-derived coherence acquired during rest-state that could predict subsequent response to training on a novel motor skill. During EEG acquisition, subjects sat quietly with both feet flat on the floor, and were instructed to fixate their gaze to the center of a fixation cross. Each recording was 3 minutes in duration. While the original EEG recording included 256 channels, only 194 were used in subsequent analyses, as extra-brain artifacts, including cheek and neck muscle artifacts and heart rhythms, are more likely to contaminate EEG signals recorded from electrodes overlying cheek and neck regions. Following data acquisition, pre-processing steps included: 100 Hz low pass filter; EEG segmentation into 1-second consecutive, non-overlapping epochs; mean detrend; and EEG signal re-reference to mean signal across all 194 channels. In addition, a combination of visual inspection and Infomax Independent Component Analysis decomposition were used to remove extra-brain artifacts, including eye blinks, eye movements, muscle artifact, and heart rhythm artifacts. The final dataset consisted of 160 epochs, with each epoch lasting 1 second, and $T = 1000$ time points for each epoch.

The goals of the present analysis are as follows: In Section 2.4.2, we closely examined a representative channel in the pre-motor region (specifically channel 197 in this dataset). Since EEGs are not well localized in space (as opposed to local field potentials), conclusions

are constrained to the sensor space. However, electrical activity captured in channel 197 reflects activity roughly around the pre-motor area. Specifically, we estimated the (log) spectrum for each epoch to identify any frequency bin or frequency band that accounts for the majority of the power spectrum. Moreover, using the method of estimating the functional medians, we obtained an estimate of the median curve from the log periodogram curves obtained from several epochs. The median curve is interpreted as a "typical" (log) spectral profile across several epochs. Using this method, we also identified outlier curves which could also be interpreted as epochs with "unusual" EEG activity. In Section 2.4.3, we investigated the possibility of non-stationarity across the 3 minute resting-state EEG recording. Our specific goal was to compare the log spectrum during the early phase (first 60 epochs) of the recording with the log spectrum during the late phase (last 60 epochs) of the recording, and identify frequency bands that exhibit any differences between the early vs late phases. In Section 2.4.4, we studied the spatial variation of power, at each of the five frequency bands: delta, theta, alpha, beta and gamma, across all 194 channels, with the goal of identifying regions that exhibit relatively greater proportion of spectral power in each of the five frequency bands of interest. Finally, we compared the spatial variation for each of the five bands during the early vs late phases of the resting-state EEG recording.

2.4.2 Functional medians of the pre-motor log spectral curves

The log of the bias-corrected periodograms at the representative channel (channel 197) that approximately overlies cortex of the pre-motor region recorded for several traces and the functional boxplots are displayed in Figure 2.11-a. The functional median curve is represented by the black curve, which is located inside the 50% central region, shaded area. The two blue curves outside of the shaded area are the non-outlying maximum and minimum curves. Similar to a functional boxplot, we show in Figure 2.11-b the pointwise boxplot (per frequency point), where the black curve is the median obtained by connecting the medians

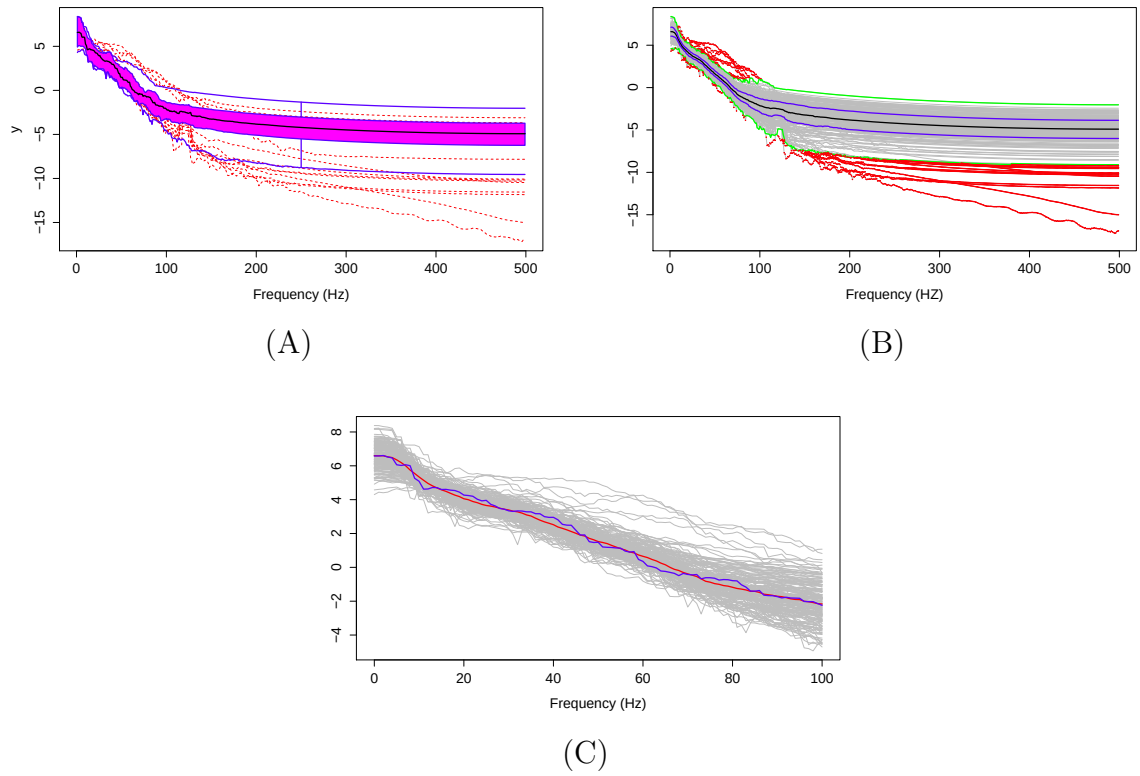


Figure 2.11: (A) The functional boxplots, (B) Pointwise boxplots of log bias-corrected periodograms, and (C) Two median curves obtaining by functional boxplot method (blue) and pointwise boxplot method (red) are shown in the same plot.

at each frequency point; the blue curves form the central region (50-th percentile region); the green curves are two non-outlying extreme curves. We compared these two median curves in Figure 2.11-c and noted a slight discrepancy between these median curves derived using a functional boxplot and the pointwise boxplot, with an emphasis on the low frequency range. The main difference between the functional median and the point-wise median curve is in the interpretation. The former is one of the curves from a recorded epoch, whereas the latter may not be an actual curve. Hence the latter cannot really be interpreted as a "typical" curve from a family of curves formed from several epochs. Moreover, the functional boxplots approach allows us to identify specific epochs that produce "unusual" or outlying log bias-corrected periodogram curves. Note that in the plots, the grey curves are the log bias-corrected periodograms of 160 epochs and the red curves are outliers. Figure 2.11-b

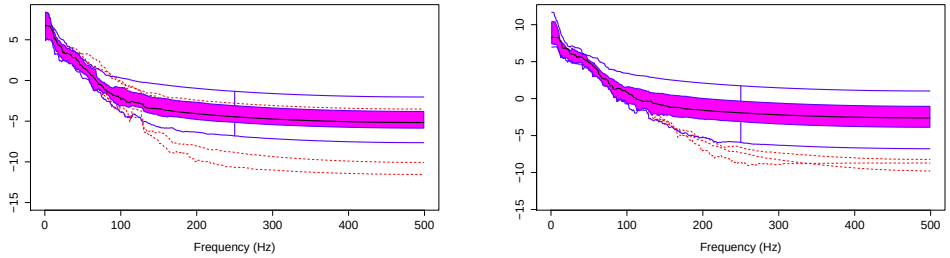
also shows that these outlying curves are discontinuous around the frequency bin centered at 100 Hz.

2.4.3 Testing for stationarity of EEG epochs across the entire resting-state

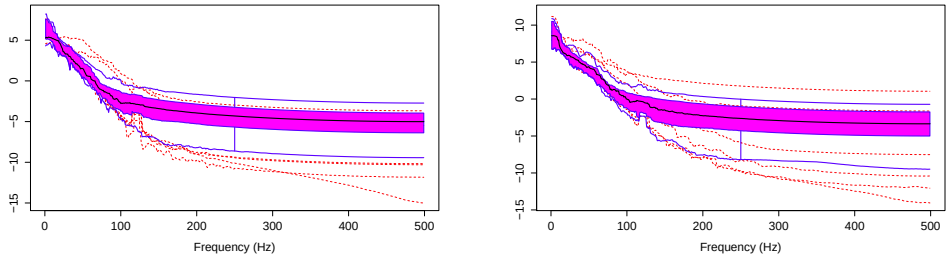
In the previous section, the functional boxplot provided descriptive statistics for the log bias-corrected periodograms of 160 epochs from the pre-motor region. Note that there were originally 180 epochs but 20 had to be removed from further analysis due to extra-brain artifact contamination. Our interest now is to test whether resting-state brain activity evolved across the 3 minute EEG recording. While there are many ways to characterize such an "evolution" of the underlying brain processes, here we will specifically look into changes on the log spectral curves for early vs late phases of the resting-state EEG recording. In this case, a change in the log spectral power in early vs late phases would indicate non-stationarity of the EEG signal across the resting-state recording.

The null hypothesis of stationarity here is that the true median curves of the early and last phrases are identical. We test this hypothesis using the rank sum test with the significance level set to 0.05. We defined the early phase to include the first 60 epochs (60 seconds) of the 3 minute recording and the late phase to include the last 60 epochs. In Figure 2.12, we display the functional boxplots and the other descriptive statistics for each phase. A visual inspection suggests that the median curves are only slightly different from each other for electrodes that approximately overlie the pre-motor region. More significant differences are noted for electrodes that approximately overlie the prefrontal region (see Figure 2.12-c). Moreover, the rank sum test failed to reject the null hypothesis, as the p-value is 0.56. Therefore, the two median curves are not significantly different and the hypothesis of stationarity in the pre-motor regions is not rejected. This is not entirely unexpected since the whole three-minute

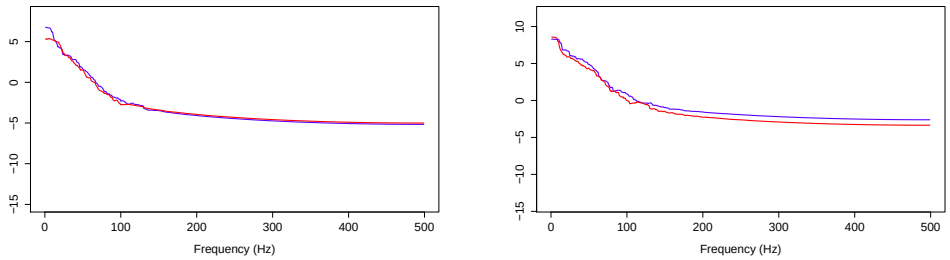
Pre-motor region	Left prefrontal region



(a)Functional boxplots of epochs 1-60



(b)Functional boxplots of epochs 101-160



(c)The median curves of the early (blue) and last (red) phrases.

Figure 2.12: Comparing median curves of the early and last phrases from pre-motor region and left pre-frontal region

recording was purely resting-state. There was no experimental stimulus and the time frame was short.

Next, we use the same testing procedure at this particular channel in the pre-motor region (channel 197) to test the same null hypothesis of non-evolution of the brain process at each of the other channels across the 3 minute EEG recording. Among the 194 total channels, 18 channels were identified that demonstrated a significant difference in median curves during

the early versus late phase at a significance level of 0.05. These channels are represented by colored circles in Figure 2.13. Of these 18, channel 29 (approximately overlying the supplementary motor area) has the lowest p-value at 10^{-4}. Since we repeat the same test for 194 channels, we used the Bonferroni correction so that the significance level for each test was set to $0.05/194 = 2 \times 10^{-4}$. Indeed, only channel 29 (anterior supplementary motor area) survived the stringent threshold after the Bonferroni correction.

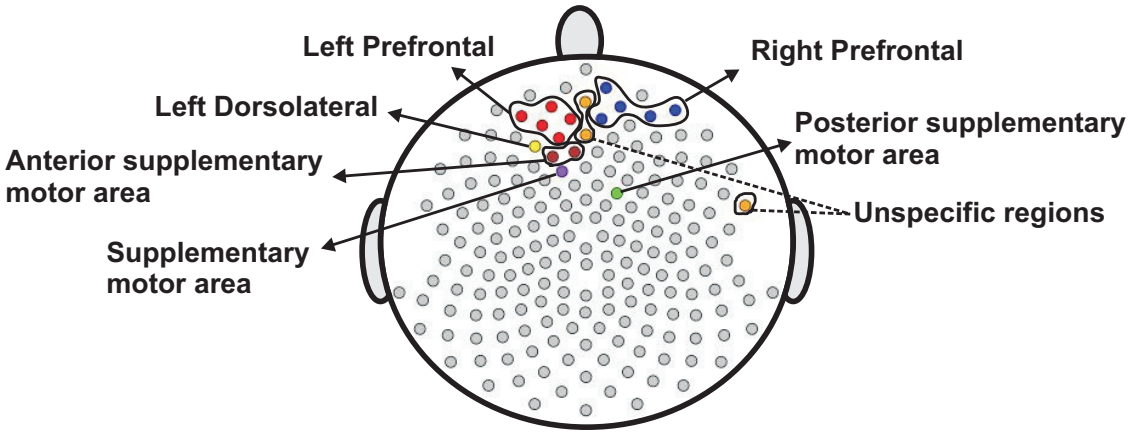
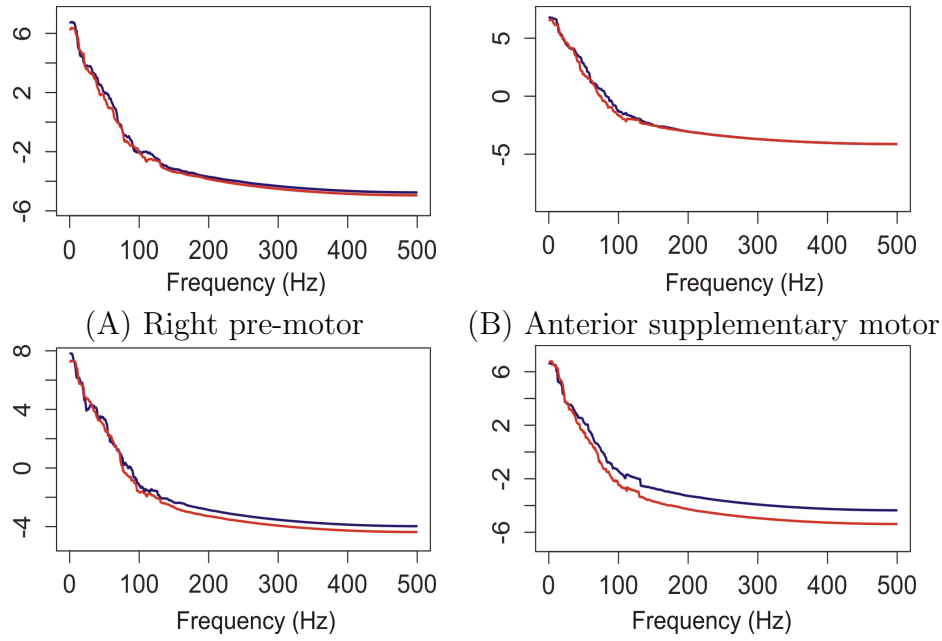


Figure 2.13: Color circles represent channels, which have significant difference between the median curve of first 60 epochs and the median curve of last 60 epochs at $\alpha = 0.05$. Grey circles represent channels which do not have significant difference between the median curve of first 60 epochs and the median curve of last 60 epochs at $\alpha = 0.05$.

The tests for temporal stationarity at each channel (local spatial tests) revealed several channels having a significant difference between the median curves of the early versus last phases of the EEG recording. As a next step, we studied stationarity in each of 19 predefined regions of the cortex. In this analysis, the representative EEG signal for each region was obtained by averaging the EEG signal-epochs over all channels within each region. The plots in Figure 2.14 suggest that the median curves for the early versus late phases of the EEG recording are similar for EEG signals recorded from channels that approximately overlie right pre-motor and anterior supplementary motor regions, but different in the right



(A) Right pre-motor (B) Anterior supplementary motor

(C) Right pre-frontal (D) Left parietal

Figure 2.14: Median curves of the early phase (first 60 epochs, in blue) and the late phase (last 60 epochs, in red) in the right pre-motor, anterior supplementary motor, right pre-frontal and left parietal regions

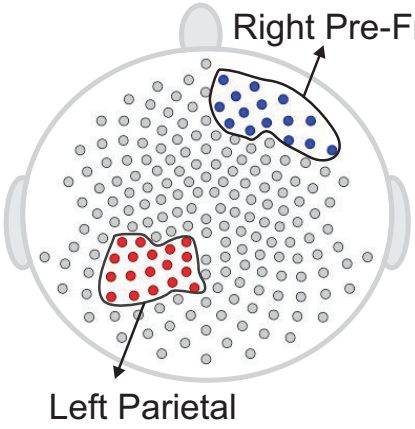


Figure 2.15: Testing for difference between the early and late phases of the resting-state for each region. The right pre-frontal regions (blue circles) and the left parietal regions (red circles) exhibit significant non-stationarity at level 0.05.

pre-frontal and left parietal regions. Indeed, we conclude from the rank sum test that there is significant difference between the early versus late phases in cluster of channels that approximately overlie the right pre-frontal (p-value = 0.01) and the parietal regions (p-value

= 0.029). We found that the right pre-frontal region is significantly non-stationary (i.e., early and late phases differ) at level 0.05 (see Figure 2.15). This result overlaps with the channel-specific tests, in which several of the channels identified to be non-stationary in the single channel tests are included in the predefined right pre-frontal region. In contrast, while the cluster of electrodes that overlie the left parietal region was found to be non-stationary in the region-by-region tests, none of the 18 channels that were identified to be non-stationary in the single channel tests are part of the left parietal cluster. Therefore, the additional averaging step across group of channels may improve signal-to-noise in this type of analysis. A similar phenomenon was also noted for predefined clusters of electrodes overlying at the left pre-frontal region.

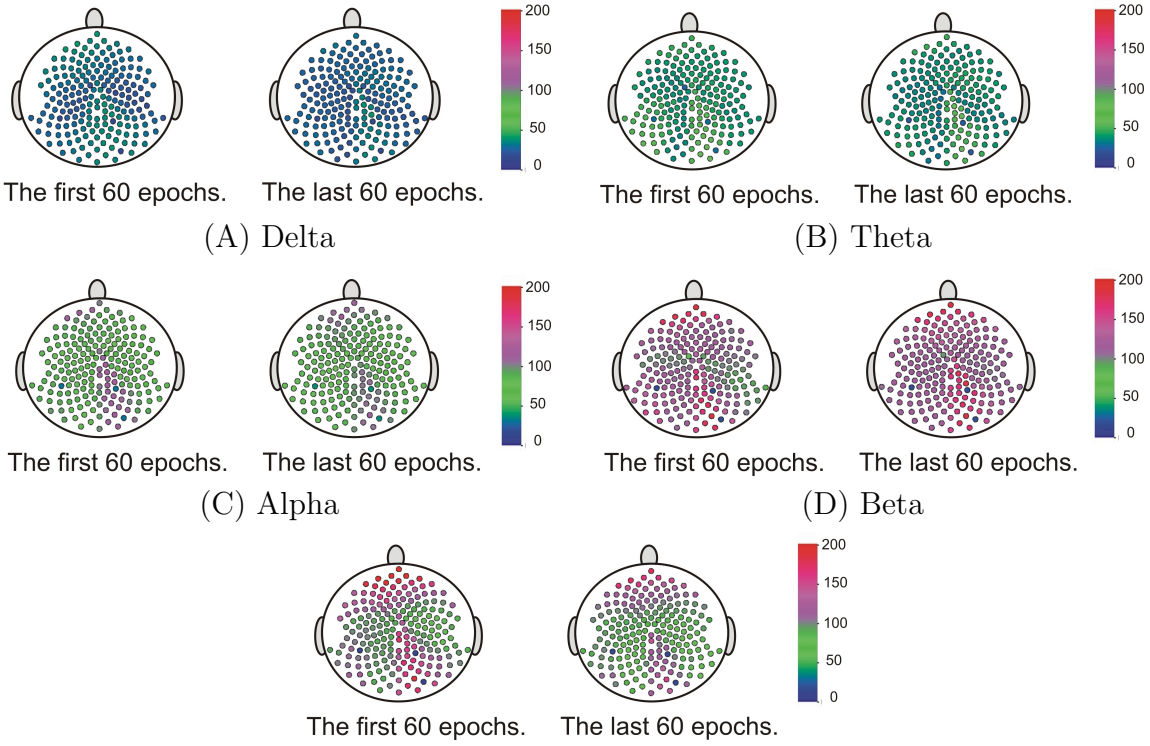
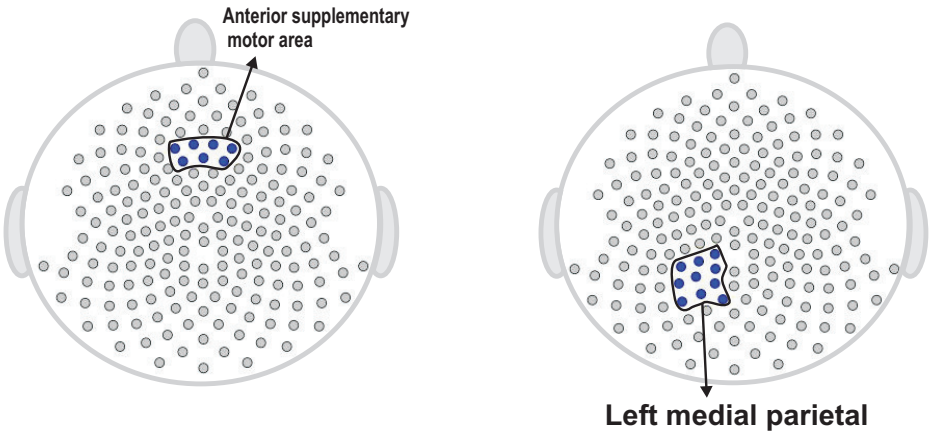


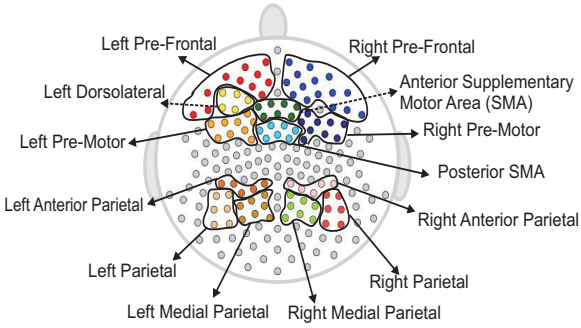
Figure 2.16: The median surfaces of five frequency bands.

2.4.4 The variation of spectral power at each frequency band across the entire cortex

Our goal here is to test whether the spectral power at each frequency band differed across the cortical surface. We first computed the estimate of the spectral power for each channel at each epoch. Starting with the delta band, for each epoch we construct a $2-D$ surface plot of the delta power across the entire cortical surface of 194 channels. These surfaces were then grouped according to the early and late phases of the resting-state. We then



(A) Theta (B) Beta



(C) Gamma

Figure 2.17: Rank sum test results for regions. Color circles in a bounded curve represents a region, which has a significant difference between the first 60 epochs and the last 60 epochs.

applied the surface boxplot method for each frequency band to obtain the median surfaces.

In Figure 2.16, we present the median surface for five frequency bands in the early and late phases. The color blue represents the low spectral power while red is for high power. In Figure 2.16, it is interesting that even during resting-state there is relatively high spectral power at the beta and gamma bands – which are both associated with higher cognitive processing ([19]).

The next step is to test for differences between the early and late phases of the EEG recording for each of the five frequency bands of interest. Using the rank sum test, the delta and alpha bands do not have significant difference between the early and late phases. However, theta, beta and gamma bands show significant differences. In Figure 2.17, the colored regions indicate significant differences between the first and last phases while the gray color regions indicate no significant differences between these two phases. For the theta band, the rank sum test rejected the null hypothesis at only one region which is the cluster of electrodes overlying anterior supplementary motor. For the beta band, the rank sum test identified differences at the left medial parietal region. For the gamma band, there were 13 regions (out of 19) with significant difference between the early and late phases. Since the gamma band is wider than other bands, an estimated spectrum powers' variation across channel in gamma band is expected to be smaller than the estimated spectrum powers' variation in other bands. In Section 4.3, we tested the stationarity for each region. Figure 2.15 shows two regions, namely, the right pre-fontal and left parietal, which are significantly non-stationary across all frequencies between the early and late phases. Figure 2.17 shows that the cluster of electrodes overlying the left parietal region exhibits significant non-stationarity in the beta and gamma bands while the cluster of electrodes overlying the right pre-fontal region is significantly non-stationary only in the gamma band.

2.5 Conclusion

This study has extended the use of the classical boxplot to FBP, which is a new visualization tool to analyze functional neuroimaging data, including EEG. The primary findings from the current study demonstrate the functional boxplot is useful for both characterizing the spectral distribution of both simulated and real EEG data and identifying potential outliers in a continuous EEG signal.

In the current implementation of the FBP, ranked sample curves are used to characterize the EEG spectrum by defining a 50% central region, a median curve, and maximum and minimum non-outlying curves. Thus, the shape, size, and length of the functional boxplot can be used to characterize the distribution of the dataset, including the skewedness and degree of variability of the EEG recording. Therefore, potential application of the FBP in this context includes comparing functional boxplots derived from EEG recordings before and after an experimental intervention (e.g., across a period of motor skill training), comparing mean functional boxplots derived from EEG recordings in healthy and diseased experimental groups, and comparing mean functional boxplots derived from EEG recordings during resting-state versus task.

An additional use of the FBP, as demonstrated by the current results, is to identify potential outliers of the EEG recording. Extra-brain artifacts, including eye blinks, eye movements, heart rhythms measured at pulse points downstream, and muscle movements can cause large deviations in the EEG signal, and represent a significant hurdle in EEG signal processing ([18]). As a method for identifying outliers in the EEG signal, the FBP could be used to rapidly identify periods of an EEG recording that show high likelihood for contamination by artifacts. In clinical applications, the continuous EEG recording has demonstrated promise as a method for monitoring neural function in patients who have compromised level of consciousness ([27]) or changes in neural function in patients undergoing neurosurgical

interventions ([17]). The use of FBP to identify outliers in the EEG recording represents a novel method for determining periods of the EEG recording that represent changes in consciousness in patients with a compromised level of consciousness, or for determining changes in neural function across neurosurgical intervention.

The current study also presents an application of the FBP to examine resting-state EEG data acquired from a single individual by comparing EEG signals acquired during early versus late phases of the 3 minute EEG recording. This result has important implications for resting-state studies of neural activity, as many neuroimaging studies that examine resting-state brain function assume resting-state neural activity to be static. However, recent studies that examine dynamic changes in resting-state neural activity suggest momentary change in cognitive processes can cause non-stationarity in resting-state function ([12] and [32]). In contrast, the current results show that the majority of channels demonstrate stationarity across the recording period, and provide support for the assumption that the average EEG signal is static across a 3 minute EEG recording. Combined with previous findings, the current results suggest that while momentary changes in cognitive processes result in non-stationary fluctuations of the time series, when averaged across a 60 second subset of the complete 3 minute EEG recording, the EEG signal is relatively static. This is supported by the current results that show channels which demonstrate non-stationarity of the EEG signal when comparing early and late phases of the recording include electrodes that overlie the right prefrontal region, which is associated with higher-order cognitive processes ([45]). Thus, the assumption of stationarity in resting-state functional neuroimaging studies may be more appropriate for non-cognitive networks, including the motor network. Regardless, further work is needed to determine the minimal time-frame in which EEG signal demonstrate stationarity.

Additional future work is focused on developing a new method for computing confidence bands for the median curve. This method needs to consider the data as a whole. One possible

approach is a re-sampling method, in which the notion of band depth is used to construct a 95% confidence band. A potential limitation of the re-sampling method is that there is the potential for multiple curves demonstrating ties with respect to band depth, thus affecting the resultant confidence band. One of the assumptions of the current smoothed periodogram method is that the log bias-corrected periodogram is an unbiased estimator of spectrum. Future work will provide further investigation of this assumption as the current method includes several levels of periodogram manipulation, including smoothing with the gamma generalized crossvalidation, log transformation, and correction by adding Euler Mascheroni constant. In conclusion, the current study presents a novel implementation of the functional boxplot and demonstrates promise as a method for exploratory analysis of complex, high-dimensional neuroimaging datasets, including EEG data.

Chapter 3

Geometric Surface Boxplot

3.1 Introduction

Infantile spasms (IS), also known as West syndrome, was one of the first epilepsy syndromes described in children less than one year old ([78]). One ongoing challenge is to understand and evaluate the underlying neuroanatomic and electrophysiologic mechanisms that cause infantile spasms. Diagnosing infantile spasms is often based on an abnormal electroencephalographic (EEG) pattern called hypsarrhythmia as shown in Figure 3.1. It is typically described as high amplitude and chaotic, and it is associated with multifocal, independent epileptiform discharges ([29, 35]). Since hypsarrhythmia does not occur in other epilepsy syndromes, a unique functional neural network is widely believed to underlie hypsarrhythmia. Previous work by [9] found that abnormally increased EEG coherence in long–distance connections is associated with hypsarrhythmia. [35, 36]; and [54] also found abnormal brain connectivity patterns as well as widely varying EEG patterns in IS patients with hypsarrhythmia.

Motivated by these several findings, our goal in this paper is to develop visualization tech-

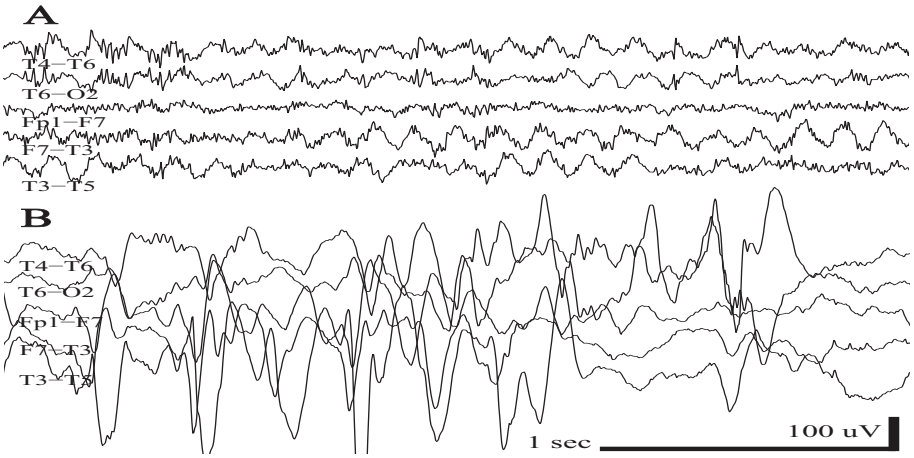


Figure 3.1: Example EEG traces from (A) a control subject containing no epileptic activity; (B) Awake EEG from an infantile spasms patient with hypsarrhythmia.

niques for connectivity that may assist in identifying a differentiable pattern among patients who have infantile spasms versus those who do not have spasms. Moreover, these visualization techniques are an important part of data exploration and modeling. Even though previous research suggests that hypsarrhythmia reflects aberrant brain connectivity, its relationship to treatment response has not been thoroughly evaluated. Thus, in this paper, we develop visualization techniques and novel statistical methods to analyze brain physiological data (in particular, multi–channel EEGs) to better understand the associations between changes in between-channel connectivity during the seizure-free epochs and the response to treatment.

In multichannel EEG studies, a common measure of functional brain connectivity is coherence which is essentially the squared cross–correlation between oscillations at different frequency of a pair of time series, and it is one of the most common measures of such dynamics ([50, 51, 67, 23, 22]). Coherence between all channels at a specific frequency (or frequency band) is represented by a matrix whose off–diagonal element is an estimate of functional interactions between two brain regions in each frequency. One of the properties of the coherence matrix is that it is symmetric positive definite (SPD), and thus we aim to develop methods that respects this important property. Visualizing and extracting useful

information from a sample of coherence matrices is challenging especially in high dimensions because each coherence matrix must be considered as a whole information unit. There are recent advances on analyzing SPD matrices, such as [70, 56, 26, 80, 13] and [33], but not enough consideration has been given to visualization of such SPD matrices. [30] introduced a surface boxplot for visualization and exploratory analysis of samples of images, but this method does not take into account the non–Euclidean geometry of the space of SPD matrices. The reasoning is that the SPD matrices do not live in a vector space, therefore, commonly used Hilbert space based methods for functional data, such as the band depth are not applicable.

To overcome this limitation, we propose a transformation approach, mapping the SPD matrices to a Hilbert space of functions, and constructed a geometric surface boxplot (GSBox) in the transformed space for samples of coherence matrices. By using the GSBox, we address two primary objectives. The *first* objective is to determine the median coherence matrix by ordering a sample of matrices based some depth measure. Here, the median matrix is that which is the deepest (in a sense that it is "covered" by the most number of pairs of matrices). The *second* objective is to detect outlying matrices again based on the ordering of the depth values of these matrices. These outliers are demonstrated by unusual connectivity between pairs of signals and hence would warrant further investigation. Subsequently, confirmed outliers, if necessary, can be removed in subsequent analyses. Thus our method pinpoints those epochs when these unusual coherence matrices were observed. In addition, the EEG signals are very prone to artifacts, so detecting outliers is very helpful for application to EEG.

One can define a multivariate order statistic for sample of SPD matrices by ranking the data points according to their depth. Then a matrix–valued object has a high rank when it is near the center whereas point far from the center has lower rank. As studied in [43] and [44] for functional data, the test ranks of data points of the second sample are computed with respect to their depth in the first sample. Using this concept and notion of data

depth, we develop two rank–based tests for two samples of matrix–valued objects under the homogeneity hypothesis, i.e., the two samples of matrix–valued objects come from the same distribution.

As previously noted, this work is motivated by our clinical collaborators' interest to determine if changes in between–channel connectivity (as measured by coherence) are associated with the patients' response to treatment. In particular, we investigate the difference between pre– and post–treatment brain functional connectivity in an awake resting state for two groups of patients: responders, whose hypsarrhythmia and spasms were both completely resolved; and non–responders, patients that had persistent spasms or hypsarrhythmia or both after treatment. Further, we compared the post–treatment connectivity of the completely responders with the control subjects. Similar to the univariate case, when two samples are paired or are repeated measurements (e.g., pre– and post–treatment from the same subject), we propose a signed rank test based on data depth to assess whether two samples have the same distribution. We also develop a rank sum test for two independent samples (e.g., complete–responders vs. control).

The main contributions of this paper are the following: (1) a GSBox to visualization and exploratory analysis for samples of SPD matrices; and (2) non–parametric hypothesis testing procedures to test for difference between two populations or groups of SPD matrices. The rest of the paper is organized as follows. In Section 2, we review the concept of coherence for multivariate stationary signals. We present the space of SPD matrices and the GSBox in Section 3. The non–parametric tests for two samples of SPD matrices are presented in Section 4. We demonstrate our proposed methodology for simulation study in Section 5. The EEG data analysis are reported in Section 6, and conclusions in Section 7.

3.2 Background: Multitaper Coherence Analysis for Stationary Signals

In this section, we first introduce some background material on spectral and coherence estimation for (second–order) stationary univariate and multivariate time series signals in the frequency domain, and then we present multitaper coherence for stationary signals. Let $X(t)$ be a zero mean univariate stationary time series with the autocovariance function is defined as $\gamma(\tau) = E\left(X(t)X(t+\tau)\right) (\tau = \ldots -1, 0, 1, \ldots)$ that is assumed to be absolutely summable, i.e., $\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| < \infty$. Then the spectrum of $X(t)$ at frequency $\omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ is given by $f(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\ e^{-i2\pi\omega\tau}$, and the inverse transform of the spectral density is the autocovariance function which has the representation $\gamma(\tau) = \int_{-1/2}^{1/2} e^{i2\pi\omega\tau} f(\omega)\mathrm{d}\omega$. When $\tau = 0$, $\gamma(0) = \mathrm{Var}\left(X(t)\right) = \int_{-1/2}^{1/2} f(\omega)\mathrm{d}\omega$ which expresses the total variance of $X(t)$ as the integrated spectral density. In our IS study, we are interested in the behavior of functional connectivity at different frequency bands, such as delta (0–4 Hz), theta (4–8 Hz), alpha (8–16Hz), beta (16–32 Hz), and gamma (32–50 Hz). These frequency bands are already established in the clinical literature (see [11, 64, 77, 10] for a review of normal and pathological oscillations).

For a finite sample of the stationary process $X(1), X(2), \ldots, X(T)$ of length T, a non–parametric estimate of $f(\omega)$ is the periodogram $I(\omega_k)$ which is defined to be $I(\omega_k) = \frac{1}{T}\left|\sum_{t=0}^{T-1} X(t)\ e^{-i2\pi\omega_k t}\right|^2, \quad k = 1, \ldots, [\![T/2 - 1]\!]$, where $[\![A]\!]$ is the integer part of A. Mean–squared consistent estimator can be derived by smoothing periodograms. A number of nonparametric smoothing periodogram methods have been proposed including the kernel smoother ([42, 52]), wavelet ([28]), smoothing spline ([76, 55]), or local polynomial ([20]).

Now we extend the spectrum background for univariate case to the multivariate. Let $\mathbf{X}(t) = [X_1(t), \ldots, X_P(t)]^T$ be a zero–mean P–dimensional multivariate stationary signal with a $P \times P$ autocovariance function $\Psi(\tau) = \mathrm{Cov}\left(\mathbf{X}(t), \mathbf{X}(t-\tau)\right)$ for $\tau = \ldots -1, 0, 1, \ldots$. If

$\sum_{\tau=-\infty}^{\infty} ||\Psi(\tau)|| < \infty$, the spectral density matrix of $\mathbf{X}(t)$ is $\mathbf{f}(\omega) = \sum_{\tau=-\infty}^{\infty} \Psi(\tau) e^{-i2\pi\omega\tau}$, where $\omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]$, and the inverse transform is $\Psi(\tau) = \int_{-1/2}^{1/2} \mathbf{f}(\omega) e^{i2\pi\omega\tau} \mathrm{d}\omega$, where $\tau = \ldots -1, 0, 1, \ldots$. For each frequency $\omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]$, $\mathbf{f}(\omega)$ is a non–negative definite Hermitian matrix, whose diagonal elements $f_{pp}(\omega)$ are the autospectra and the off-diagonal element $f_{pq}(\omega)$ are the cross spectra. Throughout this paper, we assume that $\mathbf{f}(\omega)$ is nonsingular for all $\omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]$. An important application of spectral analysis of multivariate time series is to study the linear dependence between oscillatory activity in the different components. In particular, coherence, at frequency ω, between the p–th component time series $X_p(t)$ and q–th component $X_q(t)$ is defined as $\rho_{pq}(\omega) = \frac{|f_{pq}(\omega)|^2}{f_{pp}(\omega) f_{qq}(\omega)}$, whose form is similar to that of squared correlation. The magnitude of coherence lies between 0 and 1 so that a value that is close to 1 indicates a strong linear association between $X_p(t)$ and $X_q(s)$ at frequency ω. Thus, one could interpret this as two time series with ω-oscillatory activity that are highly correlated. Moreover, if the coherence is zero at all frequencies, the time series $X_p(t)$ and $X_q(s)$ are uncorrelated ([51]).

To estimate the coherence $\rho_{pq}(\omega)$, we first estimate the autospectra and cross spectrum via the periodogram then smooth them to obtain a consistent estimator. For kernel smoothing of the periodogram matrix, a common bandwidth is used for all components in order to guarantee that the smoothed periodogram matrix (spectral matrix estimate) is positive definite ([6]). The weakness of this approach is that it assumes uniform smoothness across all elements of the spectral density matrix. To overcome this limitation, [15] proposed to estimate the multivariate spectrum by smoothing the Cholesky decomposition of a multitaper spectral estimator which is defined as follows.

Suppose that we have M independent tapering functions $\ell_m(t)$ where $m = 1, \ldots, M$, the $P-$dimensional tapered Fourier coefficient vector of a zero–mean $P-$dimensional multivariate stationary signal $\mathbf{X}(t)$ at frequency $\omega_k = k/T$ for $(k = 1, \ldots, [\![T/2 - 1]\!])$ is defined

as

$$\mathbf{D}_m(\omega_k) = \sum_{t=1}^{T} e^{-2\pi i \omega_k t}\, \ell_m(t)\, \mathbf{X}(t).$$

A particular choice of taper $\ell_m(t)$ is a sinusoidal taper, which is proposed by [62], for $m = 0, \ldots, M$ where

$$\ell_m(t) = \left(\frac{2}{T+1}\right)^{1/2} \sin\left(\frac{\pi t(m+1)}{T+1}\right), \quad \sum_{t=1}^{T} \left(\ell_m(t)\right)^2 = 1, \quad \sum_{t=1}^{T} \ell_m(t)\ell_{m'}(t) = 0 \quad (m \neq m').$$

Then the $P \times P$ multitaper periodogram at frequency ω_k is

$$\mathbf{I}(\omega_k) = \frac{1}{M} \sum_{m=1}^{M} \mathbf{D}_m(\omega_k)\mathbf{D}_m(\omega_k)^*,$$

where $(^*)$ denotes the conjugate transpose. Let $\mathbf{C}(\omega_k)$ be the $P \times P$ lower triangular matrix, which is the complex Cholesky decomposition of $\mathbf{I}(\omega_k)$, such that $\mathbf{C}(\omega_k)\mathbf{C}(\omega_k)^* = \mathbf{I}(\omega_k)$ where elements $c_{jj'}(\omega_k) = 0$ for $j' > j$. Note that $\mathbf{C}(\omega_k)$ is unique when the diagonal elements $c_{jj}(\omega_k)$ are strictly positive. To obtain the unique $\mathbf{C}(\omega_k)$, we constrain the diagonal elements $c_{jj}(\omega_k)$ to be positive.

Remark. The choice of tapers $\ell_m(t)$ do not significantly affect the final estimate of the spectral matrix $\mathbf{f}(\omega_k)$ since the multitaper method provides an initial smoothing periodogram $\mathbf{I}(\omega_k)$, and indeed we smooth the Cholesky decomposition $\mathbf{C}(\omega_k)$.

We use the kernel smoothing method for smoothing separately each Cholesky element, i.e., the diagonal elements $c_{jj}(\omega_k)$, the real part of $c_{jj'}(\omega_k)$, $\mathrm{Re}(c_{jj'}(\omega_k))$, and imagine part of $c_{jj'}(\omega_k)$, $\mathrm{Im}(c_{jj'}(\omega_k))$, for $j > j'$ and $j, j' = 1, \ldots, P$. Denote $c(\omega_k)$ to be a generic element which is either $c_{jj}(\omega_k)$, $\mathrm{Re}(c_{jj'}(\omega_k))$ or $\mathrm{Im}(c_{jj'}(\omega_k))$, and $\widehat{c}(\omega_k)$ to be a smoothed estimator

of $c(\omega_k)$ which we define to be

$$\widehat{c}(\omega_k) = \sum_{q=-\tau}^{\tau} W_{\tau,q} c(\omega_{k+q}) \quad k = 0, \ldots, [\![T/2 - 1]\!], \text{ and } q = -\tau, \ldots, \tau$$

where $2\tau+1$ is the smoothing span and $W_{\tau,q}$ are nonnegative weights that satisfy the following conditions for any fixed τ:

$$W_{\tau,q} = W_{\tau,-q} \quad (q = 1, \ldots, \tau), \qquad \sum_{q=-\tau}^{\tau} W_{\tau,q} = 1.$$

[60] shows that the choice of the weights $W_{\tau,q}$ is of secondary importance to the value of the span or bandwidth, so for simplicity, we use the boxcar smoother with weights defined by $W_{\tau,q} = 1/(2\tau + 1)$ for all $j = -\tau, \ldots, \tau$. Following [42], we select τ which minimizes the generalized crossvalidated function defined as

$$GCV(\tau) = \frac{\mathcal{T}^{-1} \sum_{j=0}^{\mathcal{T}-1} \left(c(\omega_j) - \widehat{c}_\tau(\omega_j)\right)^2}{(1 - \mathrm{tr}(H_\tau)/\mathcal{T})^2},$$

where $\mathcal{T} = T/2 + 1$ and $\mathcal{I}$ is the indicator function. The H_τ is the smoother matrix with smoothing parameter τ, and the term $(1 - tr(H_\tau)/\mathcal{T})^2$ often referred to as the model degrees of freedom, can be expressed in terms of the weight at the center of the smoothing window $(1 - W_{\tau,0})^2$.

After obtaining the smoothed $\widehat{\mathbf{C}}(\omega_k)$, by Theorem 1 ([15]), we can reconstruct the multivariate spectral $\widetilde{\mathbf{C}}(\omega_k)$ from $\widehat{\mathbf{C}}(\omega_k)$, and the spectral estimator $\widehat{\mathbf{f}}(\omega_k) = \widetilde{\mathbf{C}}(\omega_k)\widetilde{\mathbf{C}}(\omega_k)^*$. Now we can obtain the estimate of coherence between two series $X_j(t)$ and $X_{j'}(t)$ at frequency ω_k such as

$$\widehat{\rho}_{jj'}(\omega_k) = \frac{\left|\widehat{f}_{jj'}(\omega_k)\right|^2}{\widehat{f}_{jj}(\omega_k)\widehat{f}_{j'j'}(\omega_k)},$$

where $\widehat{f}_{..}(\omega_k)$ is an individual component of $\widehat{\mathbf{f}}(\omega_k)$. Moreover, our study consider the coherence between any pair of signals in five primary frequency bands that are mostly used in the literature, namely, delta (0–4 Hz), theta (4–8 Hz), alpha (8–16Hz), beta (16–32 Hz), and gamma (32–50 Hz). Since each frequency band is defined on a range of frequencies, we define the estimated coherence power at the Ω band as

$$\widehat{\rho}_{jj'}(\Omega) = \frac{\left|\int_{\omega\in\Omega} \widehat{f}_{jj'}(\omega)d\omega\right|^2}{\int_{\omega\in\Omega} \widehat{f}_{jj}(\omega)d\omega \int_{\omega\in\Omega} \widehat{f}_{j'j'}(\omega)d\omega}.$$

For a $P-$dimensional multivariate stationary signal $\mathbf{X}(t)$, let $\Gamma(\Omega)$ be the $P \times P$ coherence matrix whose elements are $\rho_{jj'}(\Omega)$ estimated by $\widehat{\rho}_{jj'}(\Omega)$ for $j, j' = 1, \ldots, P$. Note that $\Gamma(\Omega)$ is a symmetric (real–valued) positive definite (SPD) matrix.

In our IS dataset, the multivariate signals are often recorded in many epochs, so we denote $\Gamma_r(\Omega)$ as the coherence matrix at epoch r for $r = 1, \ldots, R$. Our goal is to characterize variations of the sample of coherence matrices, such as obtaining the median, the first and third quantile, and the non–outlying maximum and minimum coherence matrices. In order to obtain the descriptive statistics in univariate setting, we compute the distance of all observations, then order them according to their distance values. Therefore, in the next section, we present some backgrounds on the space of SPD matrices, and Log–Euclidean metric to compute the distance of all observed SPD matrices. Our proposed GSBox is constructed by using the Log–Euclidean metric.

3.3 Geometric Surface Boxplot (GSBox)

When a sample of SPD matrices, i.e., sample coherence matrices are available, we wish to compute descriptive statistics for this sample. Using the Euclidean metric is problematic since SPD matrices do not lie in a vector space ([57] and [2]), so the ordinary arithmetic

operations '+' or '−' may not be well–defined. For example, the subtraction of two SPD matrices may not be SPD. Moreover, [2] found that the Euclidean averaging of SPD matrices can lead to a swelling effect meaning that the determinant of the average can be greater than the determinant of each matrix. In addition, [25] remarked that interpolation of SPD matrices is possible, but extrapolation can produce indefinite matrices.

To overcome these problems, we transform the data by applying the matrix logarithm (one–to–one correspondence with the original SPD matrices), and work on the matrix logarithms of the SPD matrices. The matrix logarithm is defined in Appendix A.1. Since the Log-transformed space has a vector space structure (and is a complete metric space), we can apply our usual statistical methods on the transformed matrices. Essentially, what we are doing is working in the space of SPD matrices equipped with the Log–Euclidean metric instead of the usual Euclidean metric. There are other metrics available in the literature, such as the affine-invariant ([57]), Cholesky or Procrustes metrics. However, working with the Log–Euclidean metric is conceptually simple as we can apply usual Euclidean methods on the Log–transformed matrices. In this section, we first introduce Log–Euclidean metric in Section 3.1. In Section 3.2, we present our proposed GSBox to visualize the sample of coherence matrices.

3.3.1 The Log–Euclidean Metric

The Log–Euclidean metric was proposed by [2], and the geodesic distance between two SPD matrices based on Log–Euclidean metric is called Log–Euclidean distance. In order to define the Log–Euclidean metric, we first introduce some notations. Let $\mathcal{M}(P)$ be the inner product space of $P \times P$ dimensional real matrices equipped with the Euclidean (i.e., Frobenius) inner product $\langle \Gamma, \Gamma' \rangle_F = \mathrm{tr}\left(\Gamma^T \Gamma'\right)$ for $\Gamma, \Gamma' \in \mathcal{M}(P)$. Let $GL(P) \in \mathcal{M}(P)$ be a general linear group containing only nonsingular matrices. Then, the exponential of a

matrix Γ in $GL(P)$ is defined as $\exp(\Gamma) = \sum_{j=0}^{\infty} \frac{\Gamma^j}{j!}$. The matrix logarithm $\log(\Gamma)$ is defined as the inverse of the exponential, and the calculation of $\log(\Gamma)$ is provided in Appendix A.1. For a general nonsingular matrix, the existence and uniqueness of the logarithm may not be valid, but it is well defined for a SPD matrix. Now we denote $\mathcal{S}(P) = \{\Gamma \in \mathcal{M}(P), \Gamma^T = \Gamma\}$ as the subset of $\mathcal{M}(P)$ consisting of all $P \times P$ symmetric matrices, and denote $\mathcal{P}(P) = \{\Gamma \in \mathcal{S}(P), x^T \Gamma x > 0 \text{ for all } x \in \mathbb{R}^n \setminus \{0\}\}$ as the subset of $\mathcal{S}(P)$ consisting of all symmetric positive-definite matrices. As showed in [57] and [2], $\mathcal{P}(P)$ is a convex cone, i.e., if $\Gamma, \Gamma' \in \mathcal{P}(P)$, then $\Gamma + j\Gamma' \in \mathcal{P}(P)$ for any $j > 0$. For an example, we consider $\Gamma \in \mathcal{P}(2)$ such as $\Gamma = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ where $ac - b^2 > 0$ and $a, c > 0$. Since the constraints can be plotted as a cone in $\mathbb{R}^3$, the SPD matrix Γ can be viewed as a point in $\mathbb{R}^3$ and strictly lie inside this cone.

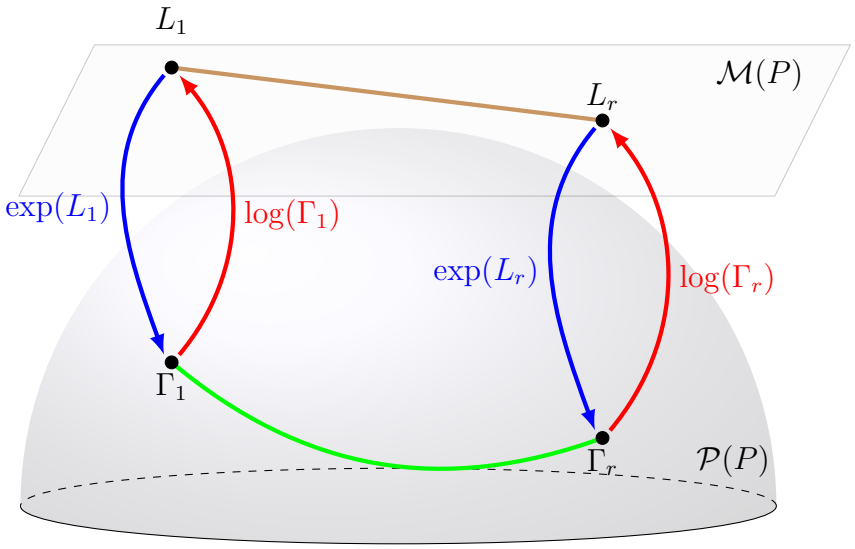


Figure 3.2: An illustration of the Log–Euclidean distance. The red line represents for log transformation while the blue line is for exponential transformation. The brown line connecting L_1 and L_r is the Euclidean distance, and the green line connecting Γ_1 and Γ_r is the Riemannian distance. Notice that these two distances are not equal to each other.

The Log–Euclidean metric essentially corresponds to a Euclidean metric in the domain of SPD matrix logarithms, so compared to other metrics, the Log–Euclidean metric requires lower computational cost. Hence, we employ the Log–Euclidean distance to measure distance between two SPD matrices, and the Log–Euclidean distance is defined as $\text{dist}(\Gamma_1, \Gamma_2) =$

$\|\log(\Gamma_1) - \log(\Gamma_2)\|_F$. Note that the Log–Euclidean metric does not have affine–invariance properties as the Riemannian metric. As discussed in [2], obtained results under the Log–Euclidean and affine–invariant metric are usually quite similar. For instance, there is no swelling effect when interpolating SPD matrices, and the Log–Euclidean metric satisfies some invariance properties, in particular the Log–Euclidean distance between SPD matrices is invariant under congruence transformations by permutation or rotation matrices, which is useful in practice. Next we define the logarithmic and exponential maps associated with the Log–Euclidean metric in the following theorem which will applied to simulated data in Section 5.

Theorem 3.1. *Suppose that* $\Gamma_1, \Gamma_2 \in \mathcal{P}(P)$*, the logarithmic and exponential maps associated with the Log–Euclidean metric can be expressed as* $Log_{\Gamma_1}(\Gamma_2) = \log(\Gamma_2) - \log(\Gamma_1)$*, and* $Exp_{\Gamma_1}(\Gamma_2) = \exp(\log(\Gamma_1) + \log(\Gamma_2))$*. Then* $Exp_{\Gamma_1}\left(Log_{\Gamma_1}(\Gamma_2)\right) = \Gamma_2$*.*

We prove Theorem 3.1 in Appendix A.2. Under the Log–Euclidean distance, the distance between two SPD matrices on the SPD manifold is computed by transforming SPD matrices onto the space $\mathcal{M}(P)$ and obtaining the Frobenius norm. In addition, for the sample of SPD matrices, it is necessary to discuss properties of unbiasedness. Suppose that Γ_r for $r = 1, \ldots, R$ are identically and independently distributed from some distribution with mean μ and variance Σ. Following [68]'s Definition 1 and 2, let M be an estimator of μ, and we define the unbiasness of M under Log–Euclidean distance in Definition 3.1 below.

Definition 3.1. *Let* M *be an estimator of* $\mu \in \mathcal{P}(P)$*, and* $\mathbf{0}_P$ *be the* $P \times P$ *matrix whose elements are all 0's. The bias of* M *relative to* μ *corresponding to the Log–Euclidean distance is defined as* $b(M, \mu) = E\left[Log_\mu(M)\right] = E\left[\log(M)\right] - \log(\mu)$*. An estimator* M *is said to be unbiased if and only if* $b(M, \mu) = \mathbf{0}_P$*.*

By using Log–Euclidean distance, we develop the GSBox approach to compute the median SPD matrix which is robust to outlying matrices. The advantage of our approach is that

it identifies outliers and provide the descriptive statistics, such as the 50% central regions, the maximum and minimum non-outlying surfaces. Here we will consider a SPD matrix as a surface and now proceed to introduce GSBox in the next section.

3.3.2 Geometric Surface Boxplot

In this section, we develop a GSBox for a sample of coherence matrices by using Log–Euclidean metric. The GSBox is constructed in a similar manner to the classical boxplot. Each observation will be sorted based on decreasing values of some depth measure (in this paper, we will consider the band depth). For ease of teminology, we call each coherence matrix as a *surface*, and in our study, let $\Gamma_r(\Omega)$, where $r = 1, \ldots, R$, be the observed surface at frequency band Ω and rth epoch. Then at frequency band Ω, a surface is said to be "deeply situated" within a sample of R surfaces if it is covered by most bands from pairs of surfaces. This idea is an extension of a classical boxplot where the median is also located "deep" in a sample because it is situated in the middle of the boxplot and hence covered by many pairs of points.

The first step to construct the GSBox is to compute the data depth of all the observations, then order them according to decreasing depth values. Suppose that the observed sample surfaces, such as $\Gamma_1(\Omega), \ldots, \Gamma_R(\Omega)$ from a region $\mathcal{S}$ in $\mathbb{R}^2$, and the information unit for such dataset is the entire surface. To compute the data depth of each surface, [30] defined the modified volume depth (MVD). Because the SPD matrices do not live in a vector space, therefore, commonly used Hilbert space based methods for functional data, such as the band depth and volume depth are not applicable. To solve this problem, we propose a transformation approach, mapping the SPD matrices to a Hilbert space of functions, and constructed a surface boxplot in the transformed space. Moreover, we employ the Log–Euclidean metric to compute the geodesic distance between two SPD matrices. Fig 3.2

illustrates the Log–Euclidean distance between Γ_1 and Γ_r which is measured by Euclidean metric between L_1 and L_r in the SPD matrix logarithmic domain. Therefore, we define a new sample geometric modified volume depth (GMVD) by applying MVD to the matrix logarithms as following

$$GMVD_R(\Gamma(\Omega)) = \binom{R}{2}^{-1} \sum_{1\le i_1\le i_2\le R} \lambda_{\mathcal{S}}\left(A\left(\log\left(\Gamma(\Omega)\right); \log\left(\Gamma_{i_1}(\Omega)\right), \log\left(\Gamma_{i_2}(\Omega)\right)\right)\right),$$

where $A\left(\log\left(\Gamma(\Omega)\right); \log\left(\Gamma_{i_1}(\Omega)\right), \log\left(\Gamma_{i_2}(\Omega)\right)\right) \equiv \min_{r=i_1,i_2} \log\left(\Gamma_r(\Omega)\right) \le \log\left(\Gamma(\Omega)\right) \le \max_{r=i_1,i_2} \log\left(\Gamma_r(\Omega)\right)$ and $\lambda_{\mathcal{S}}(\log\left(\Gamma(\Omega)\right)) = \frac{\lambda\left(A\left(\log(\Gamma(\Omega)); \log\left(\Gamma_{i_1}(\Omega)\right), \log\left(\Gamma_{i_2}(\Omega)\right)\right)\right)}{\lambda(\mathcal{S})}$, where λ is the Lebesgue measure on $\mathbb{R}^3$. A sample median surface is a surface from the sample with the largest sample geometric modified volume depth value, designed by $\arg\max_{\Gamma\in\Gamma_1,\ldots,\Gamma_R} GMVD_R(\Gamma(\Omega))$. If there are ties, the median will be the average of the surfaces with the largest sample GMVD.

Next based on the sample surfaces' GMVD values, the sample surfaces are ordered from the center outwards based on their GMVD values, inducing the order $\Gamma_{[1]}(\Omega) > \Gamma_{[2]}(\Omega) > \ldots > \Gamma_{[R]}(\Omega)$. The sample α central region is naturally defined as the volume delimited by the α proportion ($0 < \alpha < 1$) of the deepest surfaces. In particular, the sample 50% central region is

$$C_{0.5} = \{\Omega : \min_{r=1,\ldots,[R/2]} \Gamma_{[r]}(\Omega) \le \Gamma(\Omega) \le \max_{r=1,\ldots,[R/2]} \Gamma_{[r]}(\Omega)\},$$

where $[R/2]$ is the smallest integer not less than $R/2$. The border of the 50% central region is defined as the inner envelope representing the box in a surface boxplot. This is the surface analogue of the first and third quartiles of the classical boxplot. The median surface in the box is the one with the largest depth value. Because the ordering is from the center outwards, the volume of the central region increases as α increases. Hence, the maximum envelope, or the outer envelope, is defined as the border of the maximum non-outlying central region. To

determine this region, we propose to identify outlying surfaces by an empirical rule similar to the 1.5 times the 50% central region rule in a functional boxplot. The fences (or the upper and lower surface boundaries for flagging potential outliers) are obtained by inflating the inner envelope (as defined above) by 1.5 times the height of the 50% central region. Any surface crossing the fences are flagged as potential outliers. The factor 1.5 can be also adjusted as in the adjusted functional boxplots to take into account spatial autocorrelation and possible correlations between surfaces.

Remark on the applications of GSBox. In this paper, we will use the GSBox to explore the structure of brain functional connectivity from EEG data (with coherence as the metric for connectivity). However, these methods are general and can be applied to other types of data, which contain many SPD matrices (such as covariance matrices, partial coherence matrices, partial correlation matrices). For example, in medical imaging, SPD matrices (covariance tensors) provide a powerful framework for modeling the variability of the brain structure as shown in [24].

3.4 Non–Parametric Tests for Two Samples of SPD Matrices

The primary objective of our study is to investigate the change of coherence matrices of IS patients in response to treatment. In particular, for different frequency bands, we computed the coherence matrices for pre– and post–treatment of each patient and each control subject. We wish to test the hypotheses of no difference between the distributions of coherence matrices of the pre–treatment and post–treatment; and the distributions of matrices for the post–treatment patients and the controls. Since we have multiple subjects and to avoid multiple testing, we used the GSBox to obtain the median coherence matrix over all epochs for

each subject. Here the median coherence matrix is the most "deeply situated" within a sample of coherence matrices and robust to outlying matrices. Note that the median coherence matrix can be directly computed without using depth notion, but this median matrix is not one of observed coherence matrices. Then we apply rank–based non–parametric approach to decide whether the two samples, which includes median coherence matrices of each subject, come from the sample population. When the two samples of coherence matrices are paired, i.e. repeated measurements of the same subjects under pre–treatment and post–treatment, we introduce the signed rank test in Section 4.1. When the two samples of coherence matrices are independent, i.e. post–treatment versus control subjects, we present the rank sum test in Section 4.2.

3.4.1 Wilcoxon Signed Rank Test

Here we present the Wilcoxon signed rank test for paired coherence matrices. Suppose that $\Gamma_1^{(1)}, \ldots, \Gamma_n^{(1)} \overset{\text{iid}}{\sim} \mathcal{P}^{(1)}(P)$ and $\Gamma_1^{(2)}, \ldots, \Gamma_n^{(2)} \overset{\text{iid}}{\sim} \mathcal{P}^{(2)}(P)$ are paired sample SPD matrices, i.e. $\Gamma_j^{(1)}$ and $\Gamma_j^{(2)}$ $(j = 1, \ldots, n)$ are repeated measurements of the j–th subject under two different experimental conditions. Our null hypothesis is $H_o : \mathcal{P}^{(1)}(P) = \mathcal{P}^{(2)}(P)$, which is equivalent to the hypotheses that both samples come from the same population. To conduct this test, we shall use the test statistic given by

$$W_{1;n} = \sum_{j=1}^{n} \operatorname{sgn}\left(\operatorname{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})\right) \operatorname{rank}\left(\left|\operatorname{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})\right|\right) \tag{3.1}$$

where $\operatorname{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)}) = \operatorname{Tr}\left[\log\left\{\left(\Gamma_j^{(1)}\right)^{-1}\Gamma_j^{(2)}\right\}\right]$, so when two SPD matrices $\Gamma_j^{(1)}, \Gamma_j^{(2)}$ are identical, i.e. $\Gamma_j^{(1)} = \Gamma_j^{(2)}$, the distance between them $\operatorname{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})$ is 0. Moreover, it is noticed that $\operatorname{rank}\left(\left|\operatorname{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})\right|\right)$ is the usual rank of $\left|\operatorname{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})\right| \in \mathbb{R}^+$ in the set $\cup_{j=1}^{n}\left|\operatorname{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})\right|$. Here it is implicitly assumed that $\operatorname{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)}) \neq 0$ for each j, otherwise we exclude the zero difference scores and reduce the sample size n. We now

present some properties of test statistic $W_{1;n}$ in the following lemmas.

Lemma 3.4.1. *Under the null H_0, for each $j = 1, \ldots, n$, the distribution of diff$(\Gamma_j^{(1)}, \Gamma_j^{(2)})$ is symmetric around zero.*

The proof of Lemma 3.4.1 is provided in Appendix A.3. Since $\mathrm{E}\left(\mathrm{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})\right) = 0$ under the null hypothesis, the null distribution of $W_{1;n}$ follows in the same way as for the univariate Wilcoxon signed rank test, with expectation zero and variance $n(n+1)(2n+1)/6$ (see [3], Section 4.6).

Lemma 3.4.2. *The test statistic $W_{1;n}$ is invariant under congruence transformation of the observations, i.e. $\left(a\Gamma_1^{(1)}a^*, \ldots, a\Gamma_n^{(1)}a^*\right)$ for any $a \in GL(P)$.*

We provide the proof in Appendix A.4. Note that the test statistic only depends on the difference scores for each j, so the invariance of the test statistic follows as well.

Remark on Wilcoxon signed rank test. We have previously mentioned that the Log–Euclidean metric does not have full affine–invariance properties as the Riemannian metric. However, by equation (3) in Appendix A.3, the test statistics of Wilcoxon signed rank test is the same for both Log–Euclidean and Riemannian metric. If $\mathrm{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})$ is simply defined based on the Log–Euclidean metric (without using trace of a matrix), its distribution is still symmetric and around zero under the H_0 but will not hold the invariance property.

3.4.2 Wilcoxon Rank Sum Test

In the previous section, we developed the Wilcoxon signed rank test for two correlated coherence matrices populations. Here, we will develop the non–parametric test for independent samples of SPD matrices based on the Wilcoxon rank sum test. In this section, we propose rank sum test method, which is generalized to rank sum test for two populations of curves

proposed by [46]. By using GMVD, we can extend the idea of rank in the 2–D dimensional curve to higher dimensions, i.e. SPD matrices. Using the same notations as previous section, let $\Gamma_1^{(1)}, \ldots, \Gamma_n^{(1)} \overset{\text{iid}}{\sim} \mathcal{P}^{(1)}(P)$ and $\Gamma_1^{(2)}, \ldots, \Gamma_m^{(2)} \overset{\text{iid}}{\sim} \mathcal{P}^{(2)}(P)$, where $\Gamma_i^{(1)}$ and $\Gamma_j^{(2)}$ are independent for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Define the null hypothesis to be $H_0 : \mathcal{P}_1(p) = \mathcal{P}_2(P)$. Suppose that we have the reference sample $Z_1, \ldots, Z_k$ which is from one of the two observed samples, with $k \geq \max(n, m)$. The position of a particular $\Gamma_i^{(1)}$ for $i = 1, \ldots, n$, or $\Gamma_j^{(2)}$ for $j = 1, \ldots, m$ with respect to the reference sample Z, is defined as

$$V(\Gamma_i^{(1)}) = \frac{1}{n} \sum_{l=1}^{n} \mathcal{I}\{GMVD(Z_l) \leq GMVD(\Gamma_i^{(1)})\},$$

$$V(\Gamma_j^{(2)}) = \frac{1}{m} \sum_{l=1}^{m} \mathcal{I}\{GMVD(Z_l) \leq GMVD(\Gamma_j^{(2)})\},$$

where $GMVD$ is the modified volume depth defined in previous section, and $\mathcal{I}$ is the indicator. Then we can order the values $V(\Gamma_i^{(1)})$ and $V(\Gamma_j^{(2)})$ from the smallest to the largest, and their ranks are between 1 and $n + m$. The test statistic for the test is

$$W_{2;n,m} = \sum_{l=1}^{m} \operatorname{rank} V(\Gamma_j^{(2)}), \tag{3.2}$$

then under the null hypothesis H_0, the distribution of $W_{2;n,m}$ is the distribution of the sum of m numbers that are randomly chosen from $1, 2, \ldots, n + m$ without replacement ([74]). We reject the null hypothesis when $W_{2;n,m}$ is smaller than the critical value.

3.5 Simulation Study

In this section, we report results from two simulation studies with various experimental settings. The first simulation study illustrates the GSBox in Section 5.1, and we demonstrate the signed rank test (Equation 3.1) and rank sum test (Equation 3.2) in Section 5.2.

3.5.1 Geometric Surface Boxplot Simulation Study

We set $T = 200$ which is the same as the number of time points in IS EEGs dataset (see Section 6.1–6.3), the number of electrodes is $P = 15$, and $R = 100$ which is smaller number of epochs to illustrate the performance of our GSBox. First, we obtain the median coherence matrix of a control patient, named N001, at delta band $\widehat{\Gamma}_{\text{med, N001}}(\Omega_\delta)$ with its corresponding estimated spectrum density matrix $\widehat{f}_{\text{med, N001}}(\Omega_\delta)$ and Cholesky decomposition $\widehat{C}_{\text{med, N001}}(\Omega_\delta)$ for the first 100 epochs. Note that we demonstrate our proposed method at only delta band for illustration purpose. Then we generate the simulation data in two different settings:

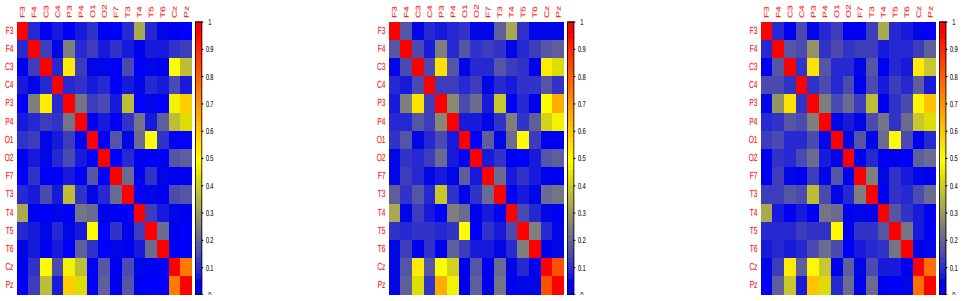
Setting 1: In this setting, we generate a sample of coherence matrices without including outlying matrices so that we can compare the median obtain from GSBox with the $\widehat{\Gamma}_{\text{med, N001}}(\Omega_\delta)$. For each epoch $r = 1, \ldots, R$, we generate $P-$ dimensional multivariate time series $\mathbf{X}_{(r)}(t) = \int_{-0.5}^{0.5} A_{(r)}(\omega) \exp^{i2\pi\omega t} dZ_{(r)}(\omega)$ by using Cramer representation, where

$$\mathrm{E}(dZ_{(r)}(\omega) dZ_{(r)}(\omega')^*) = \begin{cases} 0, & \omega \neq \omega' \\ I d\omega, & \omega = \omega', \end{cases}$$

and I is $P \times P$ identity matrix. The transfer function matrix $A_{(r)}(\omega)$ at epoch r is defined as $A_{(r)}(\Omega_\delta) = \widehat{C}_{\text{med, N001}}(\Omega_\delta) + B_{(r)}$, where $B_{(r)}$ is a lower triangle matrix whose diagonal elements are 0 and off-diagonal elements are normally distributed $\mathrm{N}(0, 0.01)$. For $\omega \notin \Omega_\delta$, we set $A_{(r)}(\omega) = \widehat{C}_{\text{med, N001}}(\omega)$. Under this setting, only the real part of cross-spectrum slowly changed over epoch r at delta band while the auto-spectrum remained the same. Thus, the coherence surfaces are also slowly changing over epochs.

Setting 2: With outlying surfaces. We used the same setting in previous simulation dataset, and we added two outlying coherence matrices from a control patient N001 at epochs 101 and 102 of simulation data in Setting 1.

For each simulation dataset, we compared our proposed method GSBox with the classical surface boxplot (SB) proposed by [30]. Figure 3.3 showed that the median coherence matrices obtained from both methods are similar to each other and to $\widehat{\Gamma}_{\text{med, N001}}(\Omega_\delta)$ of normal subject N001, and no outlying surface is detected for the Data 1.



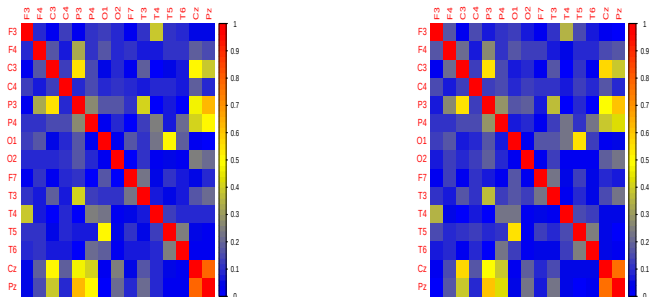
(A) (B) (C)

Figure 3.3: (A) The median coherence matrix $\widehat{\Gamma}_{\text{med, N001}}(\Omega_\delta)$ at delta band Ω_δ obtained from normal subject N001 for the first 100 epochs. (B) The median coherence matrix using the GSBox for the first simulation. (C) The median coherence matrix using SB.

For the second simulation data, Figure 3.4 shows the median coherence matrices obtaining by our method (panel A) and [30]'s method (panel B). The median coherence matrices obtained from both methods are similar as shown in Figure 3.3(A). For detecting outlying coherence matrices, both methods correctly detect two outlying matrices at epochs 101 and 102. However, SB incorrectly detected other coherence matrix at epoch 53 as shown in Figure 3.5(C). This demonstrates that our proposed method can be used as an exploratory tool to visualize and explore the structure of the sample of coherence surfaces.

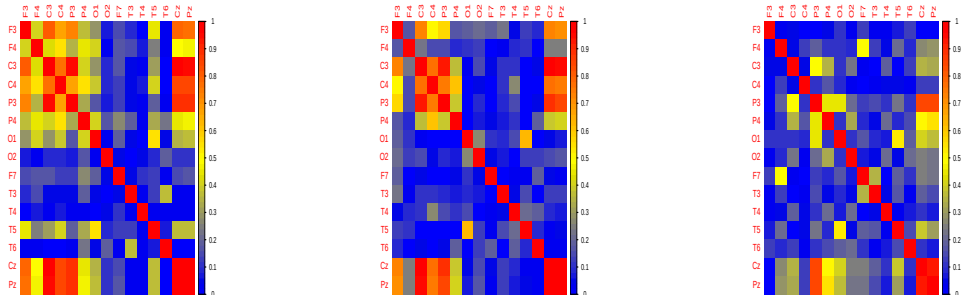
3.5.2 Non–parametric Rank test Simulation Study

In this section, we compare the signed rank test and rank sum test in the following setting. Suppose that $\Gamma_1^{(x)}, \ldots, \Gamma_n^{(x)} \overset{\text{iid}}{\sim} \mathcal{P}^{(x)}(P)$ and $\Gamma_1^{(y)}, \ldots, \Gamma_n^{(y)} \overset{\text{iid}}{\sim} \mathcal{P}^{(y)}(P)$. We wish to test the null hypothesis $H_o : \mathcal{P}^{(x)}(P) = \mathcal{P}^{(y)}(P)$. First we present how each $\Gamma_j^{(x)}$ and $\Gamma_j^{(y)}$ are generated then we compare the performance of the signed rank test and rank sum test under different



(A) (B)

Figure 3.4: (A) The median coherence matrix using GSBox for the second simulation. (B) The median coherence matrix using SB.



(A) Epoch 101 (B) Epoch 102 (C) Epoch 53

Figure 3.5: The outlying coherence matrices. Our proposed method and SB detected outlying coherence matrices (A) and (B). However, Epoch 53 (C) was incorrectly detected by SB as an outlying matrix.

settings.

Furthermore, suppose that μ_x and μ_y are means of two population $\mathcal{P}^{(x)}(P)$ and $\mathcal{P}^{(y)}(P)$ respectively, (Z_{jx}, Z_{jy}) and (R_{jx}, R_{jy}) are independent random variables which are identically distributed to some distributions $(\mathbf{0}_P, \Sigma_X)$ and $(\mathbf{0}_P, \Sigma_R)$, respectively. Let $\Gamma_j^{(x)} = \mathrm{Exp}_{\mu_{jx}}(R_{jx})$, where $\mu_{jx} = \mathrm{Exp}_{\mu_x}(Z_{jx})$ as showed in Figured 3.6. The population mean μ_x is on the space $\mathcal{P}(P)$, and is log transformed to L_x on $\mathcal{M}(P)$. Then we add two independent random variables Z_{jx} and R_{jx} to L_x by using Log–Euclidean metric, and the exponential transformation of the outcome back to $\mathcal{P}(P)$ is Γ_j^x. This setting can be viewed as a traditional linear mixed model in the way that each positive definite matrix $\Gamma_j^{(x)}$ has the same group–specific mean μ_x, the random effect Z_{jx} for the jth subject which represents the varia-

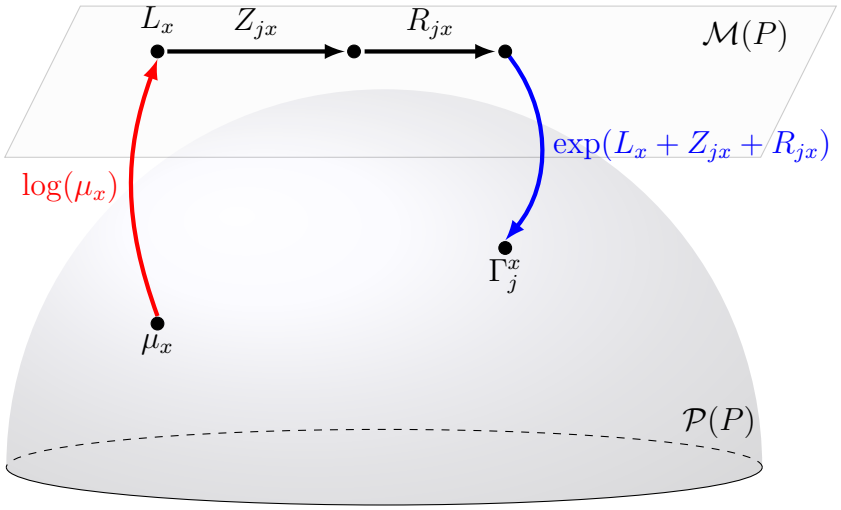


Figure 3.6: Illustration of generating SPD matrix Γ_j^x from a population mean μ_x. The population mean μ_x is on the space $\mathcal{P}(P)$, and the Log–transformed L_x on the vector space $T\mathcal{M}(P)$. Then we add two independent random variables Z_{jx} and R_{jx} to L_x by using Log–Euclidean metric, and the transformation of the outcome back to $\mathcal{P}(P)$ is Γ_j^x.

tion between subjects, and another random term R_{jx} taking into account of variation within each subject. From Theorem 3.1 and using Log–Euclidean distance, $\Gamma_j^{(x)}$ and $\Gamma_j^{(y)}$ can be expressed as

$$\begin{aligned}
\Gamma_j^{(x)} &= \exp\left(\log \mu_x + Z_{jx} + R_{jx}\right), \\
\Gamma_j^{(y)} &= \exp\left(\log \mu_y + Z_{jy} + R_{jy}\right), \qquad j = 1, \ldots, n.
\end{aligned}$$

Following Definition 3.1, we can verify that the bias $b(\Gamma_j^{(x)}, \mu_x) = \mathbf{0}_P$ since

$$\begin{aligned}
\mathrm{E}\left(\mathrm{Log}_{\mu_x} \Gamma_j^{(x)}\right) &= \mathrm{E}\left(\mathrm{Log}(\Gamma_j^{(x)}) - \mathrm{Log}(\mu_x)\right) \\
&= \mathrm{E}\left(\mathrm{Log}(\mu_x) + Z_{jx} + R_{jx} - \mathrm{Log}(\mu_x)\right) \\
&= \mathrm{E}\left(Z_{jx} + R_{jx}\right) \\
&= \mathbf{0}_P,
\end{aligned}$$

thus it implies that $\Gamma_j^{(x)}$ is an unbiased estimator of μ_x. The unbiasedness proof is similar for $\Gamma_j^{(y)}$. Next we demonstrate the power of signed rank test and rank sum test in different

scenarios.

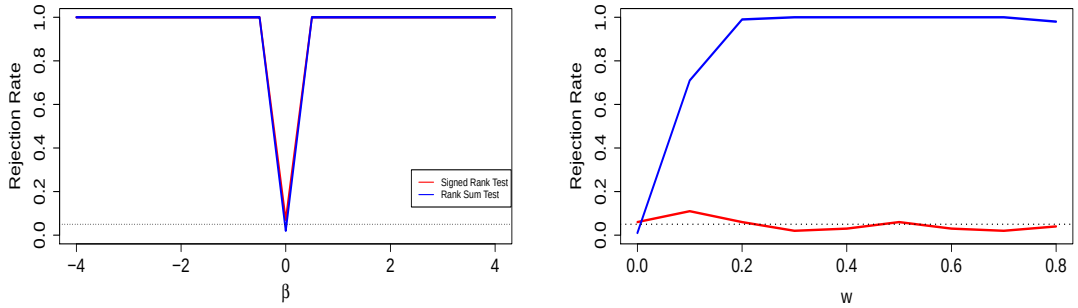
Setting 3 : Location shift. Let $\mu_x = \mu_y$, $Z_{jx} = \mathbf{0}_P$ and $Z_{jy} = \beta Z_{jv}$ where $Z_{jv} \overset{iid}{\sim} (\mathbf{0}_P, \Sigma_V)$. Under this setting, the sample mean μ_y move away from μ_x through the direction of Z_{jv} with the real–value weight β. Note that the two populations are independent, thus the power of rank sum test should larger than the power of signed rank test.

Setting 4 : Dependence. This simulated data illustrates the signed rank test perform better than rank sum test under dependence assumption between two populations. Let $\mu_x = \mu_y$, $Z_{jy} = (1 - \mathrm{w})Z_{jv} + \mathrm{w}Z_{jx}$ for $\mathrm{w} \in [-1, 1]$. Note that R_{jx} and R_{jy} are always independent. However, the dependence between Z_{jx} and Z_{jy} are related to w. When $\mathrm{w} = 0$, both two samples are independent, and when $|\mathrm{w}| < 1$ for $\mathrm{w} \neq 0$, two samples are more dependent when w is increasing. When $\mathrm{w} = 1$, the two samples are distributed from similar distribution.

Setting 5 : Similar to the setting in Data 4, but we set $\mu_x \neq \mu_y$. In particular, we observe the rejection rate of the two rank tests when the Log–Euclidean distances between μ_x and μ_y are $\mathrm{dist}(\mu_x, \mu_y) \approx 0.01$, $\mathrm{dist}(\mu_x, \mu_y) \approx 0.03$, and $\mathrm{dist}(\mu_x, \mu_y) \approx 0.1$.

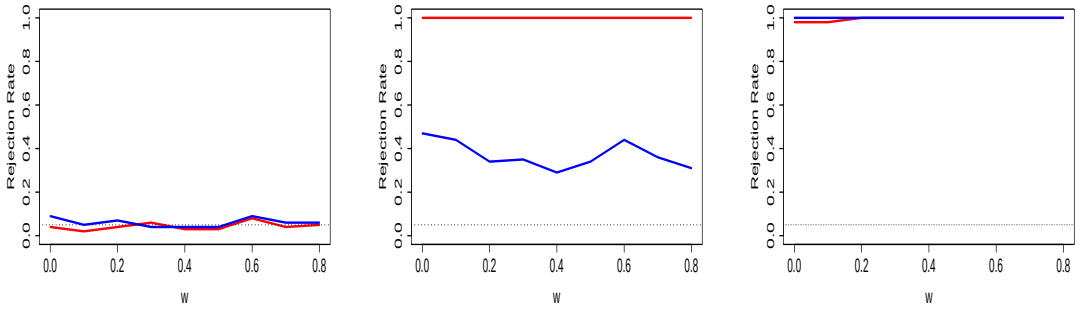
To generate the simulation data 3–5, we set the number of channels $P = 15$, the sample size $n = 100$, the location scale $\beta \in [-4, 4]$, the dependence weight $\mathrm{w} = \{0, 0.1, 0.2, \ldots, 0.8\}$. Moreover, the random component $Z_{i\cdot} = z_i^T V$, where $z_i \sim \mathrm{N}(\vec{0}_{p\times 1}, D)$, D is a $P \times P$ identity diagonal matrix, and $V = \sum_{k=1}^{P} e_k$. Here, e_k are the basis elements of an orthonormal basis of the space of Hermitian matrices. Similarly, the error terms $R_{i\cdot} = r_i^T V$, where $r_i \sim \mathrm{N}(\vec{0}_{p\times 1}, D)$. For each value of β and w, we run 500 replications. Finally, we use the estimated median coherence matrix $\widehat{\Gamma}_{\text{med, N001}}(\Omega_\delta)$ of normal subject N001 for the population means $\mu_{\cdot\cdot}$.

Figure 3.7(A) illustrates that when the distribution of two populations are the same, i.e. $\beta = 0$, the rank sum test and the signed rank test have similar rejection rate around significance



(A) Simulation setting 3 (B) Simulation setting 4

Figure 3.7: Simulation results: (A) location shift β in Setting 3, and (B) dependence w between two samples in Setting 4. Rejection rates of signed rank test (red line) and rank sum test (green line) for two samples of 20×20 dimensional matrices. There are 500 times for each values of β. The horizontal dotted line indicates the level $\alpha = 0.05$



(A) $\text{dist}(\mu_x, \mu_y) \approx 0.01$ (B) $\text{dist}(\mu_x, \mu_y) \approx 0.03$ (C) $\text{dist}(\mu_x, \mu_y) \approx 0.1$

Figure 3.8: Setting 5 result. Rejection rates of signed rank test (red line) and rank sum test (blue line) for two populations of 15×15 dimensional matrices. There are 500 times for each values of w. The horizontal dotted line indicates an significance level $\alpha = 0.05$.

level $\alpha = 0.05$. However, when β increases and two populations are independent, the power of the rank sum test is larger than the power of signed rank test. In addition, Figure 3.7(B) shows that when two populations are independent, i.e. $w = 0$, the rank sum test and the signed rank test have similar rejection rate around significance level $\alpha = 0.05$. However, when the dependence between two populations w increases, the the rank sum test tends to reject the H_0 more than the signed rank test.

In Figure 3.8, we investigate the reject rate of two rank tests for two populations whose

means are different. When the Log–Euclidean distance between two distribution populations' means is small, even though both test have low rejection rate, the rejection rate of signed rank test is greater than rank sum test's rate when w increases. When the distance between two distribution populations' means is moderate, the rejection rate of signed rank test is increasing as the dependence w increases, whereas the rejection rate of rank sum test is higher than signed rank test when w = 0, i.e. independent populations, and lower than signed rank test when w increases. This indicates that when two populations are highly correlated, the signed rank test is more appropriate than the rank sum test. In addition, when the distance between two populations' means is large, both tests have highly comparable rejection rate. Through the simulation study, we shows that our GSBox can visualize the distribution of sample of SPD matrices, and perform better than SB. Moreover, we demonstrate the power of rank sum test and signed rank test for two samples of SPD matrices. It shows that the signed rank test perform better than the rank sum test in the case that two samples are dependent. Conversely, the rank sum test is an appropriate testing method when two samples are independent in terms of type I error rate and power.

3.6 Visualization and Testing of Coherences of EEGs in the IS Study.

3.6.1 Experimental Design

This is a case–control study conducted by our co–author Dr. Shrey at the Children's Hospital of Orange County (CHOC). The IS patients were retrospectively identified with new-onset infantile spasms between August 2011 and December 2014 who exhibited hypsarrhythmia and immediately received initial treatment of infantile spasms with ACTH and/or vigabatrin. The video–EEG was recorded for each patients at diagnosis and after at least 10 days of

treatment. Here, twenty one controls were retrospectively identified as the patients without any known neurological diagnoses or EEG abnormalities. Control subjects were selected such that the group's overall distribution of ages was similar to the IS cohort (similar median value and IQR).

3.6.2 EEG Acquisition

Relevant clinical and demographic data were abstracted from the medical records. For each patient, digital scalp EEG recordings were retrospectively collected. All studies were acquired using the Nihon Kohden EEG acquisition system, with electrodes placed according to the international 10-20 standard as shown in Figure 3.9. All but four studies were originally

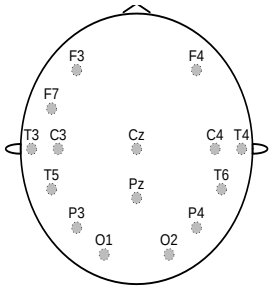


Figure 3.9: Map of channels on the scalp. There are total 19 cortical scalp electrodes for each subject. Electrode placement followed the 10-20 international recording system. M1 and M2 were referential electrodes applied to the left and right mastoid processes, respectively. FPz and Oz were not utilized in these recordings.

recorded with 200Hz sampling rate; the remaining four studies were originally recorded at 500 Hz and downsampled to 200Hz. For the IS patients, two separate epochs during wakefulness, each lasting twenty minutes or longer, were extracted; the first was isolated from the study performed at the time of the infantile spasms diagnosis (prior to treatment) and the second was from the subsequent follow-up EEG. The selection of EEG data segments

was performed without reviewer knowledge of treatment status or outcome, as each specific EEG study was de-identified and assigned a code prior to epoch abstraction. Epochs of awake EEG data were similarly extracted from the control subjects' recordings. Approval for this study was obtained from the Institutional Review Board of the Children's Hospital Orange County, and the requirement for informed consent was waived.

While hypsarrhythmia can sometimes be more prominent during sleep, all pre–treatment IS subjects in this study exhibited hypsarrhythmia during wakefulness. We analyzed awake EEG data for two methodological reasons. First, wakefulness is easily recorded in nearly every routine clinical EEG performed on infants, whereas sleep cannot be consistently captured without long epochs and relies upon extended EEG monitoring. This is particularly true of IS patients, who sleep less than healthy infants of the same age [34]. Second, the features of hypsarrhythmic EEG vary with stage of sleep [40], necessitating accurate sleep staging prior to quantitative analysis. Moreover, IS and hypsarrhythmia are associated with altered structure and progression of sleep stages, with notably diminished/absent REM sleep [34] and sleep spindles in stage 2 sleep [1]. Furthermore, these characteristics change during the course of treatment, making it infeasible to employ standard sleep staging procedures.

3.6.3 EEG Preprocessing

EEG preprocessing was performed on Matlab (R2017b, MathWorks, Inc.). The continuous EEG signal was re–referenced to the common average [14], low-pass filtered (3rd order Butterworth) at 0.5–55 Hz, removed 60–Hz noise, and segmented into non-overlapping one second epochs. Artifacts, which were caused by eye blinks, muscle activity, movement, and poor electrode contact and marked by a board-certified pediatric epileptologist, were removed from analysis. Moreover, EEG channels with constant artifact were also excluded dataset (one channel for control group and four channels for spasms group), which has been

published in [69].

We used the GSBox to visualization and exploratory analysis the samples of coherence matrices for all subjects, including control subjects, IS patients in which both spasms resolved and IS patients in which spasms did not resolve after treatment. We present the analysis in two stages. First, in Section 6.4, we closely examined the sample of coherence matrices of control subjects at five frequency bands used in clinical studies. By using our proposed GSBox method, we obtained subject–specific median coherence matrices (a large number of epochs $R > 1000$). The median coherence matrix is interpreted as the representative functional connectivity for each subject. In addition, we also identified outlying coherence matrices which could also be interpreted as epochs with abnormal EEG connectivity. Similarly, we continue investigating the sample of coherence matrices from IS patient before and after treatment. In Section 6.5, we used the median coherence matrices of each subject to compare the pre– and post–treatment coherence matrices for IS patients, and we also tested the differences between post–treatment from IS patients with the baseline from normal infants.

3.6.4 Visualizing Functional Connectivities

We illustrated the GSBox of samples of coherence matrices for a control infant N001 in Figure 3.10. It showed the first quartile coherence matrix (left), the median matrix (center) and the third quartile matrix (right) at gamma band. Similar to a classical boxplot, the median matrix is located between the first and third quartiles, which form the 50% central region. The median coherence matrix is one of the matrices from a recorded epoch, so it can be interpreted as the representative coherence matrix from a sample of coherence matrices formed from several epochs. Moreover, our GSBox identified specific epochs that produce abnormal or outlying coherence matrices which is illustrated in Figure 3.11. Compared to the median coherence matrix in Figure 3.10, the outlying coherence matrices exhibited stronger

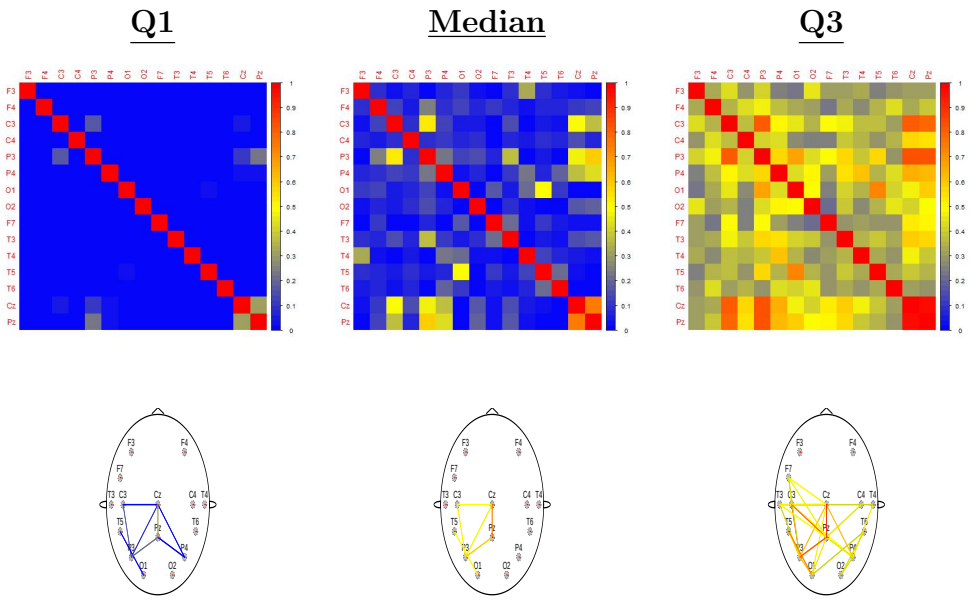


Figure 3.10: GSBox for control subject. We illustrate the 50% central region of sample of coherence matrices at gamma band. The top panel showed the coherence matrices with the functional connectivity on the scalp shown in the bottom by using 0.4 threshold for the median, Q1 and Q3 matrices. The left matrix is the first quartile coherence matrix, middle is the median matrix, and right is the third quartile matrix.

functional connectivities in many pairs of channels such as (C3, P3), (C3, Fz), (F3, Cz) and (P3, Cz). It showed that great variability across epochs may be expected in neuroimaging studies because cognitive processes can vary for individual subjects. Therefore, our proposed GSBox was able to provide insights into features of EEG such as similarities and variability of the brain functional connectivity across epochs to facilitate the statistical model. Moreover, Figure 3.12 showed the median coherence matrices in other frequency bands. A number of functional connectivities were greater in delta, theta and alpha bands than in beta and gamma bands. It means that for an awake control subject, he/she had more slow brain oscillation (delta, theta and alpha) activity than beta and gamma oscillations.

Our interest now is to investigate the sample of coherence matrices from two IS patients with hypsarrhythmia. Figure 3.13 illustrated the median coherence matrices of pre– and post–treatment for IS patient in which hypsarrhythmia was resolved. Compared between

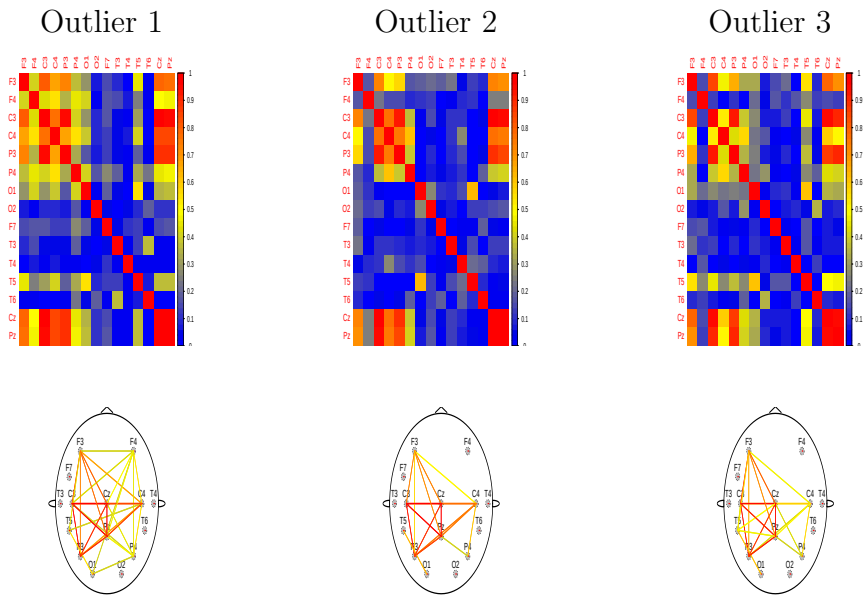
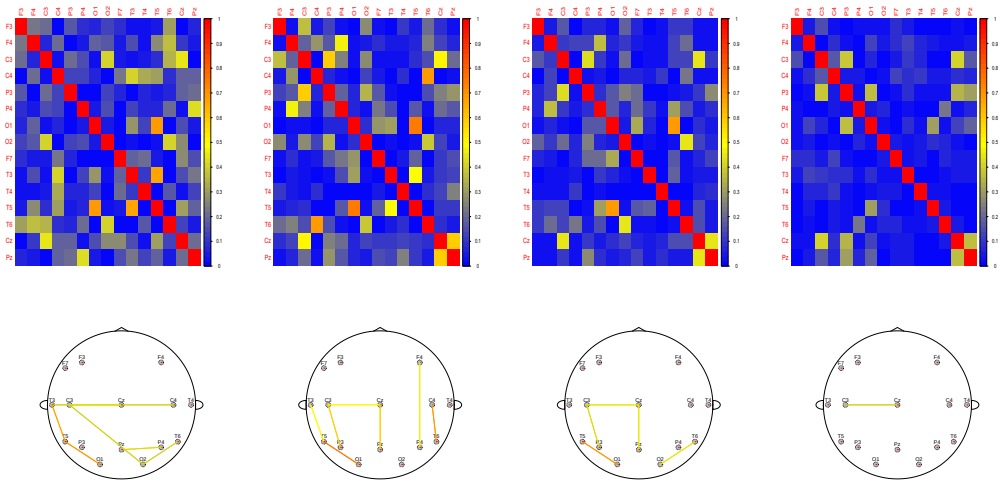


Figure 3.11: Several outlying coherence matrices for control subject at gamma band. We applied 0.4 threshold for the functional connectivity on the scalp shown in the bottom panel.



(A) Delta (B) Theta (C) Alpha (D) Beta

Figure 3.12: Median coherence matrices of control subject at delta, theta, alpha and beta bands. We applied 0.4 threshold for the functional connectivity on the scalp.

pre– and post–treatment, the method suggests that the functional connectivity in the theta and alpha bands is weaker at post–treatment than at pre–treatment. Furthermore, we notice that on average, the the strength and pattern of connectivity of post–treatment are similar

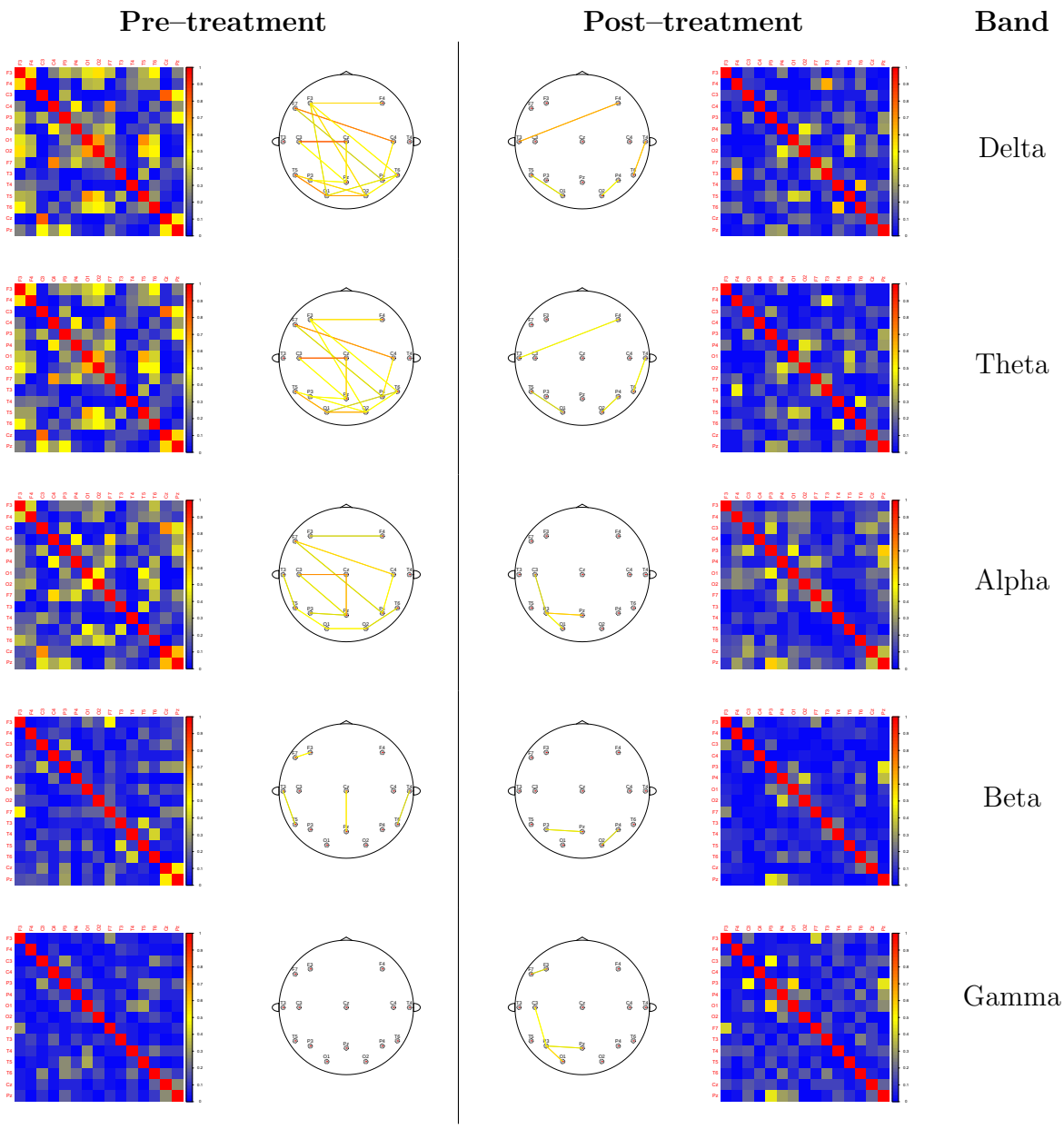


Figure 3.13: A comparison of the median coherence matrices between pre– and post–treatment for an IS patient whose spasms was resolved after treatment. The corresponding functional connectivity with 0.4 threshold on the scalp were shown next to each median coherence matrix.

to the control subject at each frequency band. For IS patient whose hypsarrhythmia did not resolve (a non–responder), the median coherence matrices for the pre– and post–treatment are give in Figure 3.14. We found that the strength and number of functional connectivities in delta, theta, beta and gamma bands became stronger after treatment. It also showed that

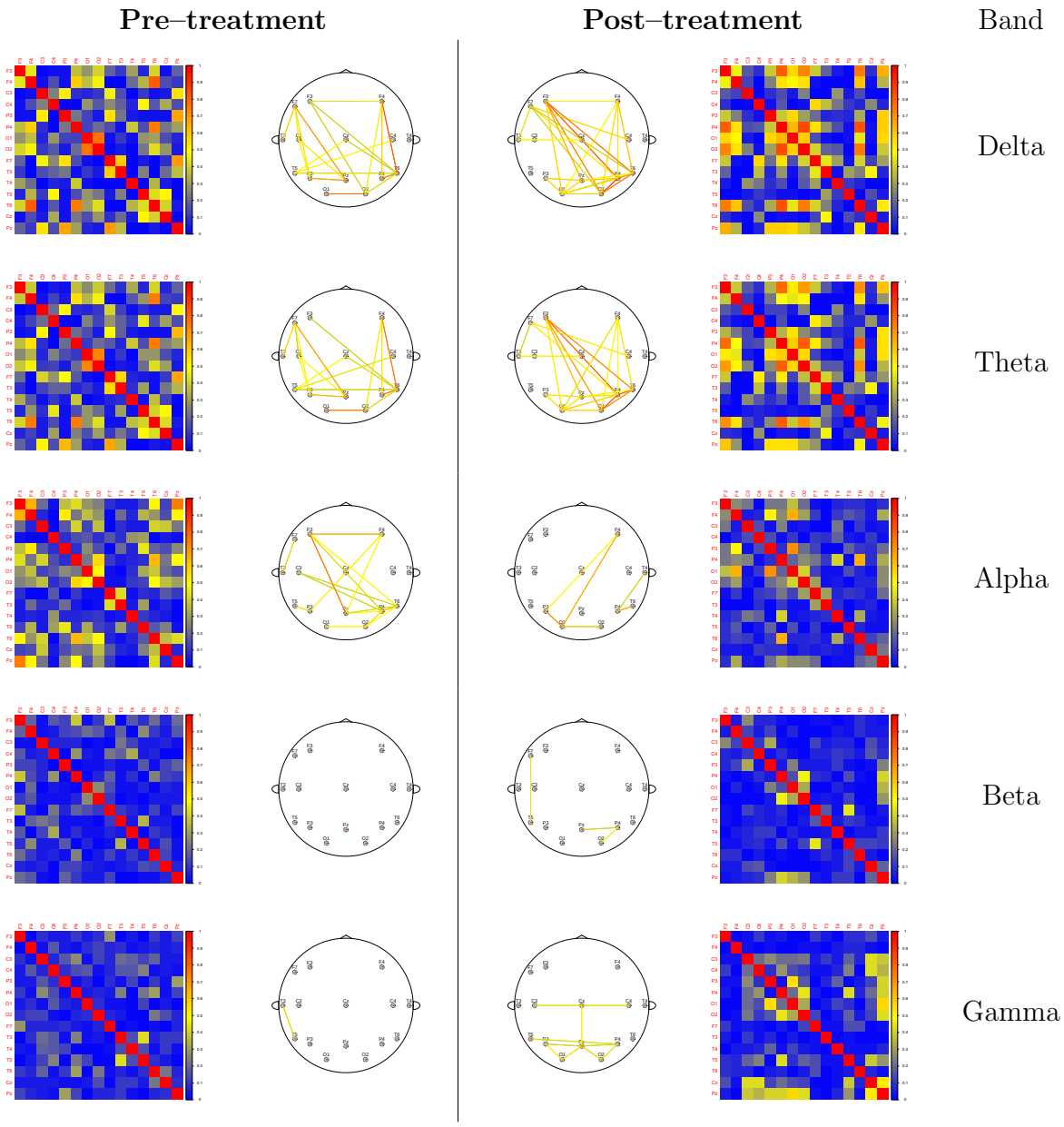


Figure 3.14: A comparison of the median coherence matrices between pre– and post–treatment for an IS patient whose spasms was not resolved. The corresponding functional connectivity with 0.4 threshold on the scalp were shown next to each median coherence matrix.

the median coherence matrices at delta, theta and gamma bands are different with median coherence matrices of control subject. Here we demonstrated our GSBox for three different subjects to obtain descriptive statistics for the sample of coherence matrices across epochs. Next, we tested the hypothesis for difference between the populations of coherence matrices

before and after treatment for all IS patients, and between the population of coherence matrices of IS patients after treatment with control subjects in the next section.

3.6.5 Testing for Association Between Treatment Response and Functional Connectivity

To investigate the change of functional connectivities in response to treatment, subjects are grouped according to their response to treatment: responders (Resp = 11), non–responders (NResp = 10), and control subjects (Normal = 21). Notice that non–responders post–treatment have persistent spasms, with or without hypsarrhythmia. We first applied the GSBox for each subject to obtain the median coherence matrix, which is represented the most centering or deepest matrix among sample of coherence matrices according their depth. Then we employed the signed rank test to test the difference between two dependent samples of coherence matrices for all subjects.

	Delta	Theta	Alpha	Beta	Gamma
Pre v.s. Post (NResp)	0.106 (33)	0.28 (48)	0.625 (16)	0.431 (23)	0.92 (4)
Pre v.s. Post (Resp)	0.2 (48)	**0.042** (50)	**0.007** (61)	0.174 (31)	0.464 (24)

Table 3.1: Signed rank test result of testing difference between pre–treatment coherence matrices with post–treatment coherence matrices of resolved and non–resolved IS patients. We reported p–values with corresponding test statistics inside parenthesis and the significance level 0.05.

Table 3.1 shows the result of signed rank test for two samples of coherence matrices pre– and post–treatment. We found that there is no significant difference in non–responders for all frequency bands. However, among responders, functional connectivity at post–treatment is significantly different from pre–treatment at the theta and alpha bands. Our conclusion is that the non-responders are not significantly different after treatment, but the responders have different connectivity patterns in the theta and alpha bands. This implies that theta

and alpha are the most informative frequency bands for functional networks associated with IS.

	Delta	Theta	Alpha	Beta	Gamma
Post (NResp) v.s. Normal	**3×10^{-4}** (-3.5)	**0.001** (-3.1)	**0.004** (-2.8)	**0.007** (-2.7)	0.53 (-0.6)
Post (Resp) v.s. Normal	**0.002** (-3.1)	**0.03** (-2.1)	0.059 (-0.1)	**10^{-4}** (-3.88)	**0.002** (-3.1)

Table 3.2: Rank sum test result of testing difference between post–treatment coherence matrices of resolved and non–resolved IS patients with coherence matrices of control subjects (Normal). We reported p–values with corresponding test statistics inside parenthesis and the significance level 0.05.

To investigate whether the post–treatment functional connectivities of IS patients are similar to control subjects' functional connectivities, Table 3.2 summarizes the result of rank sum test for two independent samples. The results show that the non-responders are significantly different than the control subjects in all except the gamma band. The results in the gamma band are not particularly relevant, as scalp–EEG gamma activity is not currently employed during standard clinical EEG analysis. Additionally, the gamma amplitude is very low, and it overlaps with the frequencies associated with muscle activity (artifact). Furthermore, Table 3.2 also shows that the responders are different than the controls in every band except alpha. It is noticed that theta has the second largest p–value. The results found here are consistent with Table 3.1 in the sense that Table 3.1 implies theta and alpha are most informative, and in Table 3.2 we see that, in responders, the functional networks in the theta and alpha bands look the most like the control subjects.

Remark. We did not take into account of the effects of volume conduction. The reasoning was that volume conduction should affect each subject approximately equally, so it should not preclude us from comparing the different subject groups. However, the spurious connections due to volume conduction happen so frequently that they can mask more subtle differences between groups. Note also that volume conduction affects the coherence measure but not

the GSBox and rank–based non–parametric methods that are the main focus of this paper.

3.7 Discussion

This study introduces a novel approach to visualization and exploratory analysis for the samples of SPD matrices (coherence matrices) which are observed from multiple subjects under different experimental settings. A new GSBox is introduced based on the surface boxplot. A rank sum test is developed for testing the differences between two families of matrices. The method were motivated by an IS with hypsarrhythmia study whose goal is to relate persistent hypsarrhythmia following treatment with adverse neuropsychological outcomes that have not been well understood. For the implementation of our proposed method, ranked sample coherence matrices are used to characterize the functional connectivity by defining a 50% central region, a median coherence matrix and also detecting outlying coherence matrices. Thus, it can be used to characterize the distribution of the dataset, including the skewness and degree of variability of the sample of matrices over EEG recording.

As illustrating in this study, the potential application of our proposed method is to comparing the sample of SPD matrices (e.g., coherence matrices, covariance matrices and diffusion tensor) in healthy and diseased experimental groups, or under different experimental settings. Moreover, an extension of GSBox is to measure the likelihood that a given matrix belongs to a distribution of matrices. For instance, we could develop GSBox to measure the distance between the post–treatment coherence matrix (for a single subject) with the median coherence matrix of the control population. This would provide a measure of the amount of improvement in an individual subject.

We conclude this section by discussing several limitations and plans for future work. First, our study characterized functional connectivity via coherence between each pair of signals,

but coherence does not provide a measure of the direct linear relationship between two components. The alternative approaches are partial coherence ([23]) Second, our proposed method does not compute the confidence surfaces for the median SPD matrix. One possible approach is a re–sampling method, in which the notion of band depth can be used to construct a 95% confidence surface. A potential limitation of the re–sampling scheme is that there is possible for multiple surfaces with tied depth, thus affecting the confidence surface.

3.8 Acknowledgments

This study was funded in part by an Institute of Clinical and Translational Sciences UC Irvine-Childrens Hospital of Orange County Collaborative Grant and a Childrens Hospital of Orange County Pediatric Subspecialty Faculty Tithe Grant.

Chapter 4

Other time–frequency and spike train models

In this chapter, I present other time–frequency and spike train models that were explored as part of my dissertation. These will be developed into two papers in the near future. An extension of my work in Chapter 2 is to investigate dependence between oscillatory activity in a pair of brain regions, and I propose a time frequency vector autoregressive model in Section 4.1. Another future work is modeling stochastic spiking neural networks in Section 4.2.

4.1 A time frequency vector autoregressive model (TF-VAR)

4.1.1 Motivation

In time series analysis, there is a need for rigorous approaches to measure the temporal dependence structure between time series. The autocorrelation function (ACF), which measures the linear dependencies in time series, has been widely used for exploring and testing dependence between time series. More precisely, the ACF measures correlation at all available time lags between each pair of components of the multivariate time series, and thus the ACF equals 0 when there is no linearly dependence between observed time series, and ACF equals to maximum absolute value 1 if observations are perfectly linearly related. In neuroscience, brain signals, such as electroencephalograms (EEGs), functional fMRI and local field potential (LFP), are used to study brain activity. Of key interest are the oscillations at different frequency bands. While correlation (and partial correlation) measure dependence between a pair of time series, these do not precisely reflect how oscillatory activity on one region may be associated with oscillatory activity in another.

Another popular measures of dependence between time series is coherence and partial coherence, which is essentially the cross–correlation and partial cross–correlation respectively, between oscillations at different frequency bands of each time series. An advantage of partial coherence is that it provides a measure of the direct linear relationship between two signals by removing the linear effects of other observed signals. Figure 4.1 demonstrates an advantage of partial coherence over correlation. In this figure, $F_1(t)$ and $F_2(t)$ are the latent sources which are low and high frequency AR(1) time series respectively, M is an indicator matrix, and $N(t)$ is noise. Then $X_1(t)$ is a mixture of low frequency component $F_1(t)$ and high frequency component $F_2(t)$, while $X_2(t)$ and $X_3(t)$ contains only low frequency com-

ponent $F_1(t)$ and high frequency component $F_2(t)$ respectively. Clearly, $X_2(t)$ and $X_3(t)$ are indirectly linked through $X_1(t)$, so correlation between $X_2(t)$ and $X_3(t)$ is non-zero, but partial coherence between $X_1(t)$ and $X_2(t)$ is non-zero at low frequency, and it is zero at high frequency. Here, we address two limitations of partial coherence. First, it captures dependence between time series oscillating only at the same frequency bands, and thus it does not take into account the interactions between oscillations at different frequency bands. Second, partial coherence measures the time-series dependence at the same time block, so the disadvantage is that it does not take into account potential non-stationarity (e.g., evolution of the dependence structure).

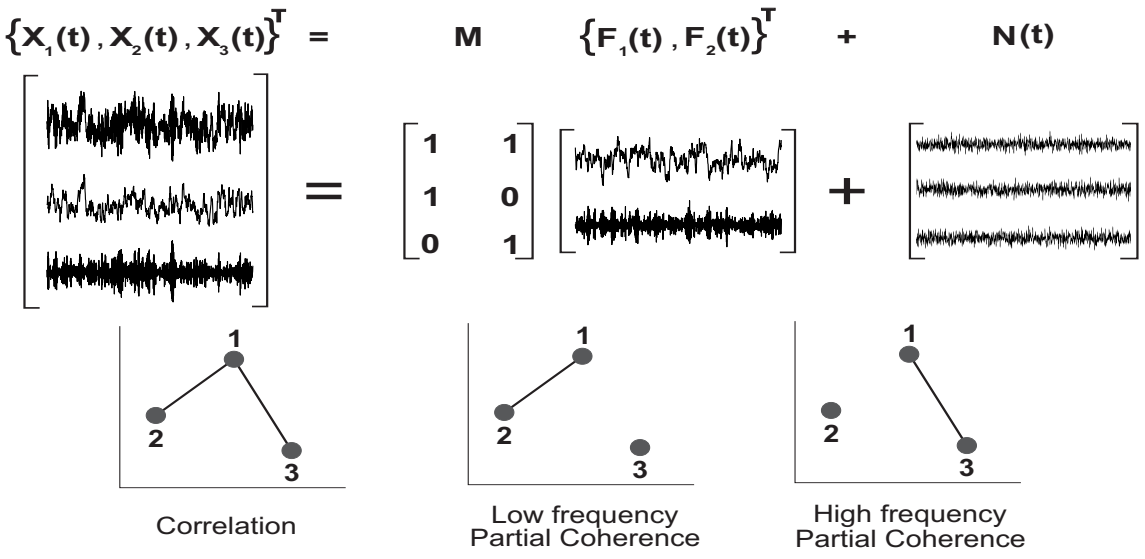


Figure 4.1: A comparison of the dependence measures. Left: $X_2(t)$ and $X_3(t)$ are indirectly dependent via $X_1(t)$. Middle: $X_1(t)$ and $X_2(t)$ are directly dependent via the low frequency component $F_1(t)$. Right: $X_1(t)$ and $X_3(t)$ are directly dependent via the high frequency component $F_2(t)$.

To overcome these limitations we propose a time frequency vector autoregressive model (TF-VAR). The proposed TFVAR model is different from the classical VAR model because it directly examines lead-lag dependence between the oscillatory features rather than the raw time series themselves. Consider a multivariate time series with components $X_1(t)$ and $X_2(t)$. Our proposed model will be utilized to address whether or not oscillatory activity

as measured by the periodograms for $\{X_1(t-1), X_2(t-1)\}$ at previous time $t-1$ leads to increased or decreased particular frequency band activity in time series $X_1(t)$ at the current time t. The methods presented is motivated by a motor skill acquisition study at the Neurorehabilitation laboratory at the University of California, Irvine (Principal Investigator: Dr. Steven C. Cramer). In the previous study, EEG was recorded from 17 subjects both during resting-state prior to motor skill training and during motor skill training using dense-array EEG (256 electrodes).

Our goals are (1) to develop a dimension reduction method to extract the best "representative" signals for multivariate time series ; (2) investigate dependence between oscillatory activity in a pair of brain regions. To achieve our goals, first, we introduce the spectral principal component analysis (Spec–PCA) to obtain the univariate signal from EEG signals in the same brain region. Second, we develop the time-frequency vector autoregressive model (TF–VAR) to examine dependence between brain wave oscillations in a pair of brain regions during the motor-learning state. In the preliminary studies, we shall look into dependence between the pre-frontal and supplementary motor regions of a single subject during the motor-learning experiment. First, we apply the spectral PCA to compute a univariate signal which summarizes multivariate time series in each regions in Section 4.1.2. Then we define the spectrum of the first component time series obtained from spectral PCA in Section 4.1.3, and we introduce a TF–VAR model in Section 4.1.4. Finally, we report preliminary analysis results in Section 4.1.5 and discuss our future work in Section 4.1.6.

4.1.2 Computing Signal Summary to Multivariate Time Series by Using Spectral Domain Principal Component Analysis

Since each brain region includes a large number of EEG signals, we apply the spectral principal component analysis to extract the best "representative" patterns for each brain region.

The principal components in frequency domain was introduced in [7] (Chapter 9). Here, we develop the idea from [67]. Let $\mathbf{X}(t)$ be a p–dimensional stationary zero–mean multivariate time series, and $U(t)$ be the univariate time series. Suppose that we are interested in representing $\mathbf{X}(t)$ by a summary signal $U(t)$ defined as

$$U(t) = \sum_{j=-\infty}^{\infty} a(t-j) * \mathbf{X}(j). \tag{4.1}$$

Indeed, $U(t)$ in Equation 4.1 is a process obtained from $\mathbf{X}(t)$ by linear filter with a $p \times 1$ vector filter $a(j)$ such that the coefficients are absolutely summable, i.e., $\sum_{j=-\infty}^{\infty} |a(j)| < \infty$. To reconstruct $\mathbf{X}(t)$ from $U(t)$, we can form a process $\widehat{\mathbf{X}}(t)$ as an estimate of $\mathbf{X}(t)$ such as

$$\widehat{\mathbf{X}}(t) = \sum_{j=-\infty}^{\infty} b(t-j)U(j), \tag{4.2}$$

where $b(j)$ is an absolutely summable $p \times 1$ filter. In this paper, we choose $a(j)$ and $b(j)$ so that the mean square error of reconstruction between $\widehat{\mathbf{X}}(t)$ and $\mathbf{X}(t)$ is minimized, i.e.

$$E\left\{(\mathbf{X}(t) - \widehat{\mathbf{X}}(t))^*(\mathbf{X}(t) - \widehat{\mathbf{X}}(t))\right\} \tag{4.3}$$

is minimized. Let $\Sigma(h)$ be the auto–covariance of $\mathbf{X}(t)$, and $S(\omega)$ is the spectral density matrix whose element $S_{pp}(\omega)$ is the auto–spectrum of the p-th component of $\mathbf{X}(t)$ and $S_{pq}(\omega)$ is the cross-spectrum between the p-th and q-th components. Furthermore, let $a(\omega)$ be the Fourier transform of $a(j)$, such that

$$a(\omega) = \sum_{j=-\infty}^{\infty} a(j) \exp(-2\pi i j \omega) \tag{4.4}$$

According to [7] (Theorem 9.3.1), the solution is obtained in the spectral domain with $a(\omega)$ to be the eigenvector $e_1(\omega)$ associated with the largest eigenvalue values $\lambda_1(\omega)$, and $b(\omega)$ to

be $\overline{a(\omega)}$. Thus, the filter values $a(j)$ can be computed such as

$$a(j) = \int_{-1/2}^{1/2} e_1(\omega) \exp(2\pi i j \omega) d\omega, \tag{4.5}$$

and the first principal component $U(t)$ can be formed by Equation 4.1.

4.1.3 Spectrum of the First Summary Time Series $U(t)$

The spectrum of an univariate time series $U(t)$ (which is assumed to be stationary) can give the amount of variance contributed by oscillatory components (from delta to beta band activity). Let $U(t)$, $t = \ldots, -1, 0, 1, \ldots$ be a zero-mean stationary time series with auto–covariance function $\gamma(\tau) = E\left(U(t)U(t+\tau)\right) (\tau = \ldots -1, 0, 1, \ldots)$ that is assumed to be absolutely summable, i.e., $\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| < \infty$. The spectrum of $U(t)$, denoted $f(\omega)$, is defined to be

$$f(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\ e^{-i2\pi\omega\tau}, \quad \omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]. \tag{4.6}$$

Moreover, the spectrum of $U(t)$ is actually the largest eigenvalue of the spectral matrix so that $f(\omega) = \lambda_1(\omega)$. To characterize the spectra of the EEG signals through the first component $U(t)$, which is the representative time series, we classify the oscillatory patterns of spectrum into four primary frequency bands: delta (0-4 Hz), theta (4-8 Hz), alpha (8-16 Hz), beta (16-32 Hz) and gamma (32-50 Hz) as shown in Figure 2.2. The plots on the left are the true spectra at each of the frequency band, and the plots on the right are corresponding observed time series. The starting point for estimating $f(\omega)$ is the periodogram of the first component time series $U(t)$. Denote $I(\omega_k)$ to be the periodogram computed from a finite sample of the stationary process $U(0), U(1), \ldots, U(T-1)$ at frequency $\omega_k = k/T$ which is

defined to be

$$I(\omega_k) = \frac{1}{T}\left|\sum_{t=0}^{T-1} U(t)\ e^{-i2\pi\omega_k t}\right|^2, \quad k = -[\![T/2]\!] - 1, \ldots, [\![T/2]\!], \tag{4.7}$$

where $[\![T/2]\!]$ is the quotient of $T/2$. Note that

$$I(\omega_k) \sim \begin{cases} \frac{f(\omega_k)}{2}\mathcal{X}_2^2 & k = 1, \ldots, T/2 - 1 \\ f(\omega_k)\mathcal{X}_1^2 & k = 0, T/2 \end{cases} \tag{4.8}$$

where $\mathcal{X}_\nu^2$ is a chi–square random variable with ν degree of freedom. Since the distribution of $I(\omega_k)$ is a multiple of the spectral density, its variance (which depends on $f(\omega_k)$) also changes across the frequencies ω_k. To stabilize the variance across frequencies, we will use the log transformed periodograms. It is convenient then, that the variance of the log periodograms at each frequency is constant and takes the approximate value of $\frac{\pi^2}{6}$. Moreover, while the periodogram is approximately unbiased for the spectrum, the log periodogram is no longer (approximately) unbiased for the log spectrum due to Jensen's inequality. This is easily fixed by adding the Euler Mascheroni constant 0.57721 to log transformed periodograms to obtain the log bias-corrected periodograms ([76]). Let $g(\omega_k)$ be the true log spectrum, then $Y^{(r)}(\omega_k)$, the bias corrected log periodogram at each trial r, is defined as

$$Y^{(r)}(\omega_k) = g(\omega_k) + 0.57721, \quad k = 0, 1, \ldots, T/2. \tag{4.9}$$

Since each frequency band is defined by a range, we define $Q_\ell^{(r)}(\Omega)$ to be the spectral power of the Ω band in region ℓ at trial r to be

$$Q_\ell^{(r)}(\Omega) = \int_{\omega\in\Omega} \log\ f_\ell^{(r)}(\omega) d\omega. \tag{4.10}$$

where $f_\ell^{(r)}$ is the spectrum of the representative signal of region ℓ at trial r. The data

analogue for $Q_\ell^{(r)}(\Omega)$ is the band-averaged bias-corrected log periodograms defined as

$$P_\ell^{(r)}(\Omega) = \text{ave}_{\ \omega_k \in \Omega} \left\{ Y_\ell^{(r)}(\omega_k) \right\}. \tag{4.11}$$

4.1.4 Time–Frequency Vector Auto–Regression (TF–VAR) Model

Follow the Equation 4.1, let $U_1^{(r)}(t)$ and $U_2^{(r)}(t)$ be the summary signals from two brain regions with observed signals $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$, respectively, recorded during trial r. Let $P_1^{(r)}(\Omega)$ and $P_2^{(r)}(\Omega)$ be band-averaged bias-corrected log periodograms of $U_1^{(r)}(t)$ and $U_2^{(r)}(t)$ at frequency band Ω respectively. In particular, $P_1^{(r)}(\delta)$ is the band-averaged bias-corrected log periodograms at the delta band of the first component $U_1^{(r)}(t)$ at trial r. To illustrate our notation we consider the TF–VAR model for $P_1^{(r)}(\delta)$ at delta band and single trial r to be

$$\begin{aligned} P_1^{(r)}(\delta) = \xi_0 &+ \xi_{11}(\delta,\delta)P_1^{(r-1)}(\delta) + \xi_{11}(\delta,\theta)P_1^{(r-1)}(\theta) + \xi_{11}(\delta,\alpha)P_1^{(r-1)}(\alpha) \\ &+ \xi_{11}(\delta,\beta)P_1^{(r-1)}(\beta) + \xi_{11}(\delta,\gamma)P_1^{(r-1)}(\gamma) + \xi_{12}(\delta,\delta)P_2^{(r-1)}(\delta) + \xi_{12}(\delta,\theta)P_2^{(r-1)}(\theta) \\ &+ \xi_{12}(\delta,\alpha)P_2^{(r-1)}(\alpha) + \xi_{12}(\delta,\beta)P_2^{(r-1)}(\beta) + \xi_{12}(\delta,\gamma)P_2^{(r-1)}(\gamma) + \epsilon_1^{(r)}(\delta) \end{aligned} \tag{4.12}$$

where $\xi_{ij}(\Omega,\Omega')$ is coefficient of $P_j^{(r-1)}(\Omega')$ at frequency Ω', and we assume that $\epsilon_1^{(r)}(\delta) \sim \mathcal{N}(0,\sigma^2)$. In the next step, we will consider the situation where the ϵ's are correlated across trials. Figure 4.2 illustrates our proposed model.

Remarks Here we give interpretations of some coefficients in the model. The coeffiient $\xi_{ij}(\Omega,\Omega')$ measures the effect on the Ω activity in region i at the current trial r by the Ω' activity in region j from the previous trial $r-1$.

For multiple trials such as $r = 1,\ldots,R$, the TF–VAR model in 4.12 can be generalized such

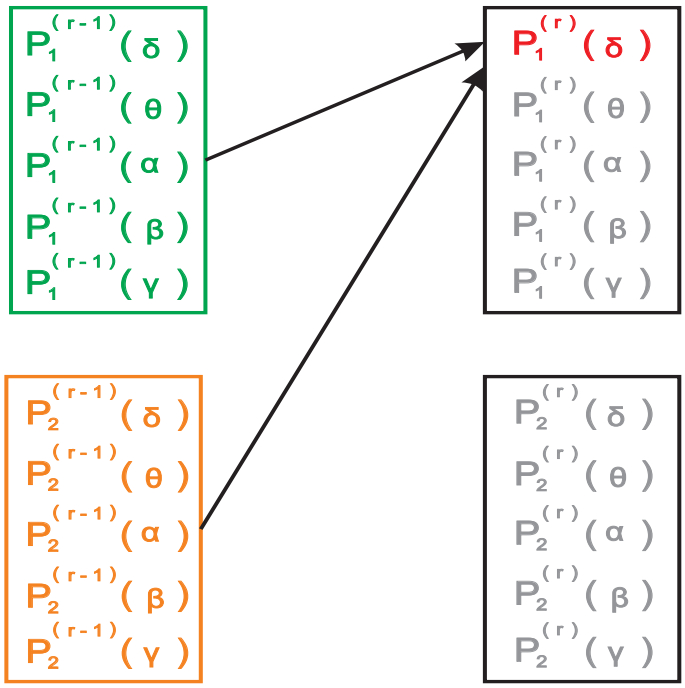


Figure 4.2: TF-VAR model illustration to study the association between the delta spectral power on region 1 during trial r with the spectral power at the different bands at the previous trial $r-1$

as

$$\mathbf{P}_1(\delta) = \mathbf{M}\boldsymbol{\xi} + \mathcal{E} \tag{4.13}$$

where $\mathbf{P}_1(\delta) = \left(P_1^{(1)}(\delta), P_1^{(2)}(\delta), \ldots, P_1^{(R)}(\delta)\right)^T$, $\mathbf{M}$ is a $(R-1)\times 11$ matrix defined as

$$\begin{bmatrix} 1 & P_1^{(0)}(\delta) & P_1^{(0)}(\theta) & \ldots & P_1^{(0)}(\gamma) & P_2^{(0)}(\delta) & P_2^{(0)}(\theta) & \ldots & P_2^{(0)}(\gamma) \\ 1 & P_1^{(2)}(\delta) & P_1^{(2)}(\theta) & \ldots & P_1^{(2)}(\gamma) & P_2^{(2)}(\delta) & P_2^{(1)}(\theta) & \ldots & P_2^{(1)}(\gamma) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & P_1^{(R-1)}(\delta) & P_1^{(R-1)}(\theta) & \ldots & P_1^{(R-1)}(\gamma) & P_2^{(R-1)}(\delta) & P_2^{(R-1)}(\theta) & \ldots & P_2^{(R-1)}(\gamma) \end{bmatrix}$$

and the coefficient 11×1 vector $\boldsymbol{\xi} = (\xi_0,\ \xi_{11}(\delta,\delta),\ \xi_{11}(\delta,\theta),\ \ldots\ ,\ \xi_{12}(\delta,\beta),\ \xi_{11}(\delta,\gamma))^T$. Moreover, we assume that $\mathcal{E} = \left(\epsilon_1^{(2)}, \ldots, \epsilon_1^{(R)}\right)^T$, where $\epsilon_1^{(j)} \sim \mathcal{N}(0, \sigma_j^2)$ for $j = 2, \ldots, R$. We denote $\widehat{\Phi}$ as the least square estimation of Φ, and $\widehat{\Phi} = \left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T\mathbf{P}_1(\delta)$.

4.1.5 Preliminary Analysis

In this section, we analyze EEG data from one participant in the *first* testing-state (Test 1) EEG study approved by the Institutional Review Board of the University of California, Irvine. The over-arching aim of this study was to identify a pattern of EEG-derived coherence acquired during rest-state that could predict subsequent response to training on a novel motor skill. During EEG acquisition, subjects sat quietly with both feet flat on the floor, and were instructed to fixate their gaze to the center of a fixation cross (Figure 4.3B). Each recording was 3 minutes in duration. While the original EEG recording included 256 channels, only 194 were used in subsequent analyses, as extra-brain artifacts, including cheek and neck muscle artifacts and heart rhythms, are more likely to contaminate EEG signals recorded from electrodes overlying cheek and neck regions. Following data acquisition, pre-processing steps included: 100 Hz low pass filter; EEG segmentation into 1-second consecutive, non-overlapping trials; mean detrend; and EEG signal re-reference to mean signal across all 194 channels. In addition, a combination of visual inspection and Infomax Independent Component Analysis decomposition were used to remove extra-brain artifacts, including eye blinks, eye movements, muscle artifact, and heart rhythm artifacts. The final dataset consisted of 160 trials (resting state) and 73 trials (motor-learning state), with each trial lasting 1 second, and $T = 1000$ time points for each trial.

The goals of the present analysis are as follows: First, we closely examine the dependence of delta oscillations in supplementary motor area (SMA) for single trial r, denoted $P_1^{(r)}(\delta)$, with other frequency bands oscillations in the same region $(P_1^{(r-1)}(\delta), \ldots, P_1^{(r-1)}(\gamma))$ and pre–frontal region $(P_2^{(r-1)}(\delta), \ldots, P_2^{(r-1)}(\gamma))$ at previous trial $r-1$. Second, we extend the analysis for multiple trials, and we report the results of other frequency bands.

Here, we investigate the dependence of brain oscillations in SMA and left pre–frontal regions. The location of two brain regions are shown in Figure 4.4. We illustrate the workflow

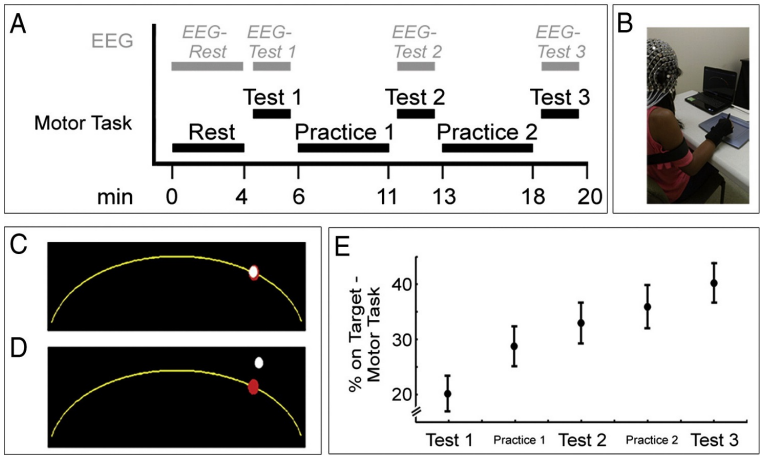


Figure 4.3: Motor learning experiment. Source: [79]

to obtain the bias-corrected log periodogram of the first component $U(t)$ across trials in Figure 4.5 From Equation 4.13, we define $\mathbf{P}_1(\delta)$ as a band-averaged bias-corrected log pe-

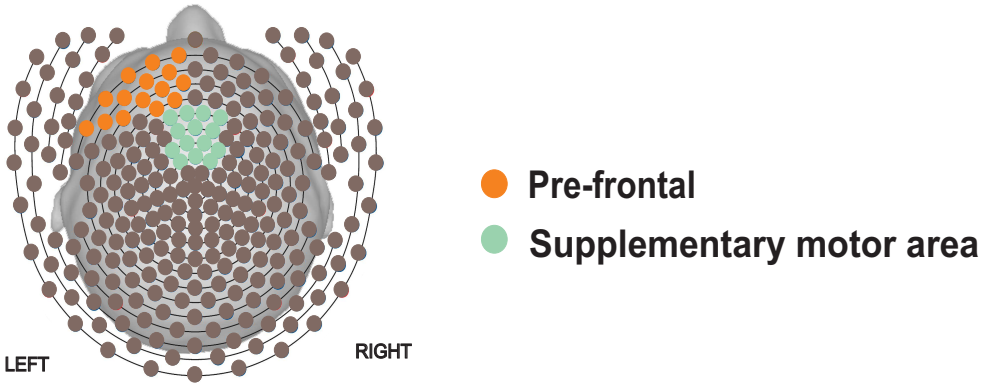


Figure 4.4: Location of SMA and left pre–frontal regions on the skull

riodogram the first component $U_1(t)$ of SMA region at delta band. In the matrix $\mathbf{M}$, each element $P_i^{(r-1)}(\Omega)$, for $r = 1, \ldots, R$, is a band-averaged bias-corrected log periodogram of the first component $U_i(t)$ of i region at Ω band from the previous trial $r-1$, where $i = 1, 2$ represents for SMA and left pre-frontal regions respectively.

The estimates of coefficient matrix $\boldsymbol{\xi}$ are summarized in Table 4.1. We found that the beta oscillations of the first component signal of left pre–frontal region at the previous trial are

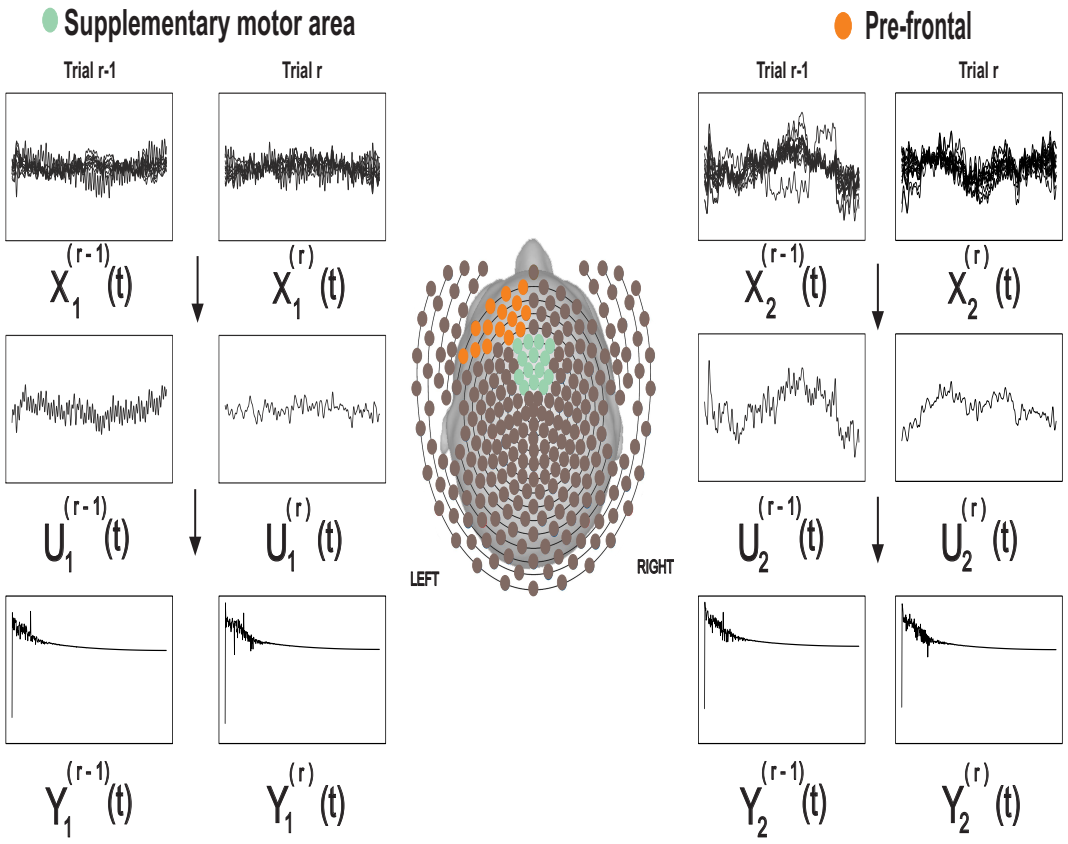
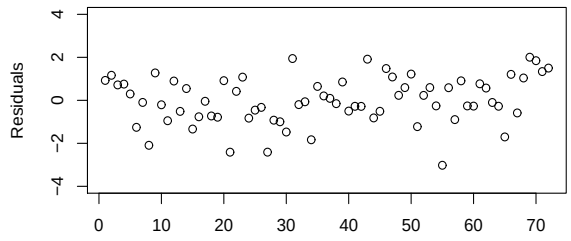


Figure 4.5: Illustration of workflow to obtain the first bias-corrected log periodogram of the first component across trials.

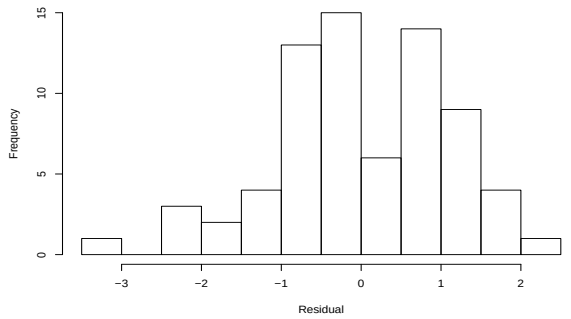
Coefficients	Estimate	Std	t statistics	p–values
ξ_0	-3.18	1.72	-1.84	**0.07**
$\xi_{11}(\delta,\delta)$	0.09	0.138	0.67	0.51
$\xi_{11}(\delta,\theta)$	0.31	0.21	1.49	0.14
$\xi_{11}(\delta,\alpha)$	50.20	0.28	0.72	0.47
$\xi_{11}(\delta,\beta)$	-0.52	0.35	-1.49	0.14
$\xi_{11}(\delta,\gamma)$	0.12	0.32	0.39	0.69
$\xi_{12}(\delta,\delta)$	0.009	0.13	0.07	0.94
$\xi_{12}(\delta,\theta)$	-0.2	0.22	-0.91	0.37
$\xi_{12}(\delta,\alpha)$	-0.12	0.34	-0.35	0.73
$\xi_{12}(\delta,\beta)$	0.69	0.31	2.25	**0.03**
$\xi_{12}(\delta,\gamma)$	-0.41	0.25	-1.63	0.11

Table 4.1: Estimates of coefficient matrix Φ

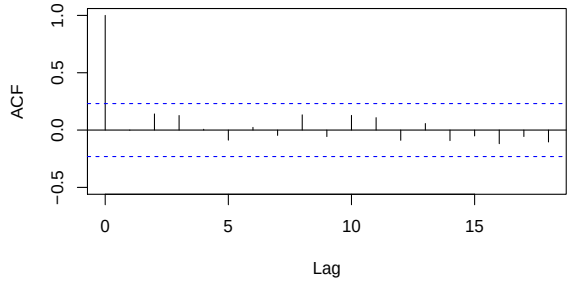
significantly (p–value= 0.03) related to delta oscillations of the first component signal of a SMA at the current trial. The interpretation of $\xi_{12}(\delta,\beta)$ is when the power spectrum of



(A)



(B)



(C)

Figure 4.6: (A) The residual plot across trials, (B) A histogram plot of residuals (C) The residual ACF plot.

beta band in left pre–frontal region increase 1 unit at previous trial, on average, the power spectrum of delta band in SMA will increase 0.69 unit at current trial where other power spectra are fixed. The residual and ACF plots are shown in Figure 4.6. The residual plot shows that linear regression is an appropriate approach for this data, and from the ACF plot, the $\epsilon_i^l(\delta)$ and $\epsilon_i^k(\delta)$ are uncorrelated for $l, k = 1, \ldots, R-1$.

Next, we investigate the dependence of SMA and left pre–frontal regions at different frequency bands. The analysis is summarized in Figure 4.7. It is interesting that the cross–dependence between the delta activity in SMA at current trial and the beta activity in pre–frontal region at a pass trial. Similarly, there is cross–dependence between the beta activity in SMA at current trial and the gamma activity in pre–frontal region at a pass trial.

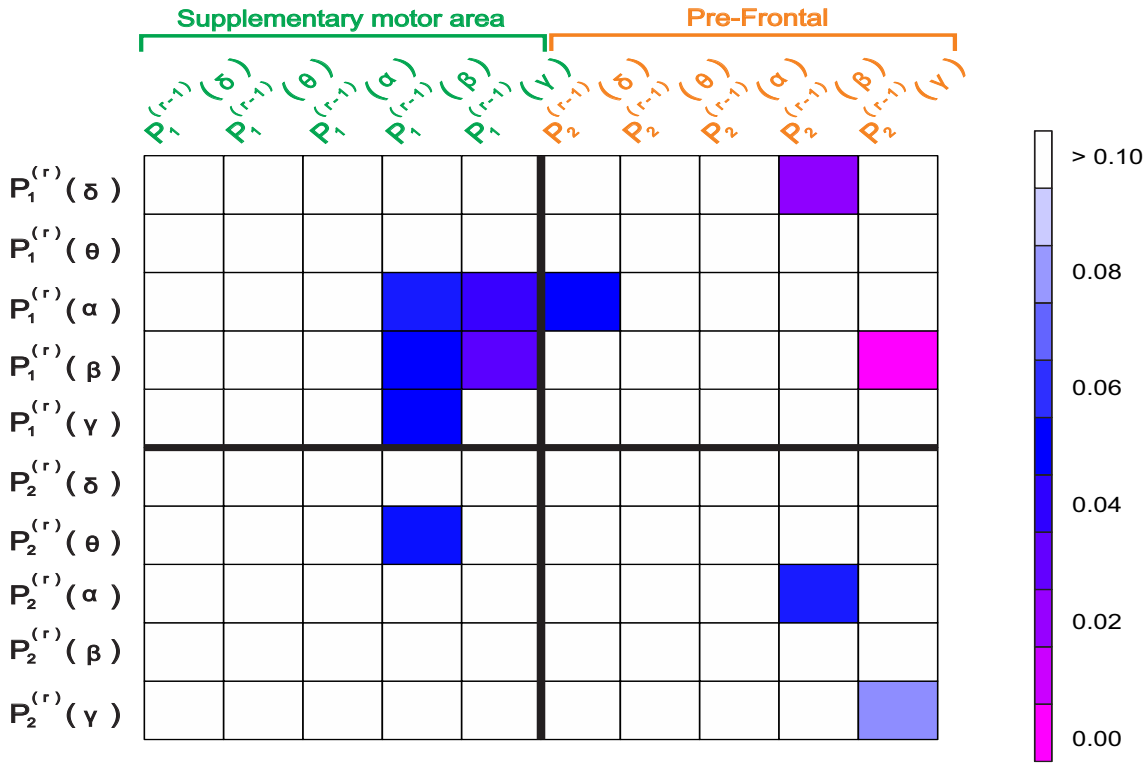


Figure 4.7: The p–values of coefficients. Each cell represents the p-value of coefficient $\xi_{ij}(\Omega.\Omega')$ corresponding to the row response variable $P_i^{(r)}(\Omega)$ and the column explanatory variable $P_j^{(r-1)}(\Omega')$

4.1.6 Future Work

We proposed a TF–VAR to examine the dependence between two brain regions at different frequency bands during the testing state of experiment. Our future works are

- Apply the proposed model to *specific* brain regions of interest and compare the analysis results of different experiment conditions.
- Find ways to simplify the analysis (reduce the number of parameters and increase power) by combining frequency bands. For example, instead of using five different bands, we combine δ and θ bands to form $(\delta\theta)$ band and similar to $(\alpha\beta\gamma)$ band.
- Incorporate subject-specific random effects to model variation across subjects.
- Investigate using procedures for correcting for multiple testing (e.g., Bonferroni, False Discovery Rate).

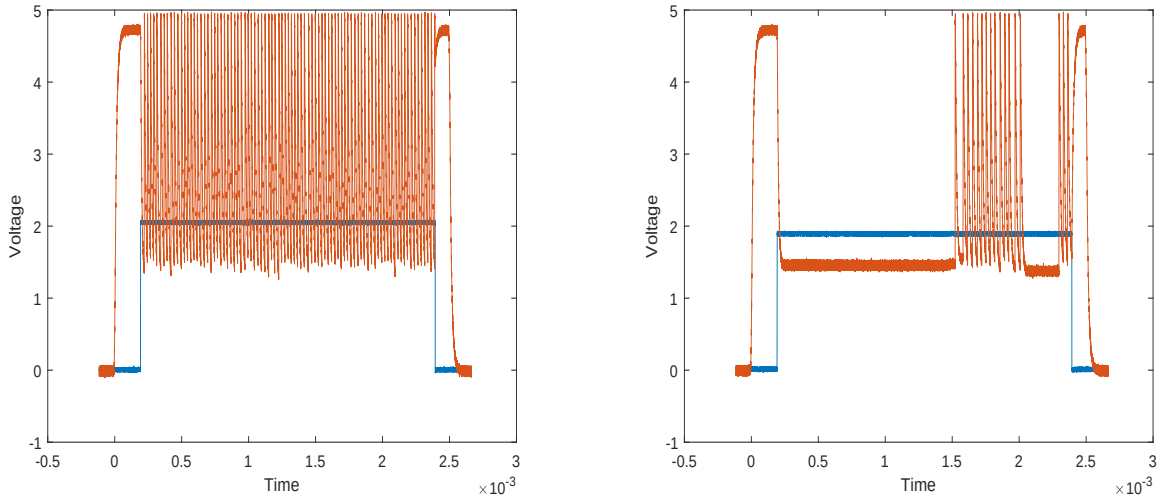
4.2 Modeling Stochastic Spiking Neural Networks

In brain–inspired (neuromorphic) computing, one of the challenges is to develop models that are compliant with the spatial and temporal constraints of the brain and whose parameters are easily interpretable. In an ongoing project led by Neuromorphic Machine Intelligence Lab led by Professor Emre Neftci (Department of Cognitive Sciences, University of California, Irvine), we focus on modeling the stochastic spiking which is produced by an insulator–metal–transition (IMT) device based on electrically induced phase–transition in series with a tunable resistance. The stochastic spiking in an IMT neuron are explained by threshold fluctuations and thermal noise. The neuron's membrane potential below firing threshold can be described by an Ornstein–Uhlenbeck process ([53]), where spikes can occur at anytime with a probability density which depends on the momentary distance between the membrane

potential and the threshold. When this probability density is assumed to be the exponential, the spikes can be formulated from a smooth underlying time-dependent function, i.e. firing rate. Point processes has been often used to describe and analyze spike trains ([31]). A common limitation of point process is that a spike train commonly assumed following a Poisson process, which is the simplest form of point process. Some alternative methods are to use inhomogeneous Poisson process or Bayesian nonparametric methods.

Another approach is using Gaussian process method for neural decoding, proposed by [65].It is often discretized the time to obtain at most one spike with each time interval. If a spike is active, a response will be recorded as 1; otherwise, it will be 0. Since the response variables are only 0 or 1, the marginal probability of firing within an interval can be modeled by the logistic function. For the firing rate, [65] used a continuous latent variable with Gaussian process prior. The advantages of Gaussian processes (GPs) are that GPs are powerful nonparametric distributions over continuous functions, and the covariance function can be modified to model the time-varying and history-dependent for each neuron's firing rate. However, a drawback of GPs is its expensively computation for large datasets due to an inversion of an $N \times N$ covariance matrix, where N is the number of data points. Another limitation of GPs is that the covariance function is assumed to be stationary over the experiment. In our case, the waveform of an IMT spiking neuron is not stationary for various input of the threshold fluctuations and thermal noise as show in Figure 4.8, where a high voltage produced more spikes than lower one. Hence, GPs can limit the modeling flexibility.

To overcome GPs' limitations, an alternative approach is to use infinite mixtures of GP experts proposed by [61], where the number of components of GP experts is automatically determined from dataset by using a Dirichlet process prior on the mixture component. As showed in [61], the mixture of GPs resolved the first limitation of a single GPs by inversing multiple smaller matrices instead of a large covariance matrix. In addition, mixture of GPs



(A) (B)

Figure 4.8: Experimental waveforms (red line) of an IMT spiking neuron for different input threshold fluctuations (blue line) at (A) 2.06V and (B) 1.9V.

composes multiple GPs which can overcome the second limitation of a single GPs. However, a drawback of this method is that it employs Monte Carlo Markov Chain (MCMC) sampling methods, which can be slow to converge. In addition, it is also difficult to diagnose the convergence of MCMC. Another approach is deep GPs (DGPs) proposed by [16] which is multi–layer neural networks, where each layer is modelled as a GP. As demonstrated in [16], DGPs is able to handle data with multi–modality (e.g., non-stationarity). Thus, my future work is to explore DGPs method on IMT spiking neurons and compare DGPs with other methods such as infinite mixtures of GP experts and neural networks.

Bibliography

[1] S. A. Altunel A, Altunel EO. The utility of the hypsarrhythmia paroxysm index and sleep spindles in eeg for predicting cognitive outcomes in a case series of infantile spasms. *Journal of Neurology & Neurophysiology*, 6(5):1–5, 2015.

[2] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications*, 29(1):328–347, 2007.

[3] V. Bagdonavicius, J. Kruopis, and M. S. Nikulin. *Nonparametric tests for complete data*. John Wiley & Sons, 2013.

[4] T. Bohlin. Comparison of two methods of modeling stationary eeg signals. *IBM Journal of Research and Development*, 17(3):194–205, 1973.

[5] S. L. Bressler and W. J. Freeman. Frequency analysis of olfactory system eeg in cat, rabbit, and rat. *Electroencephalography and clinical neurophysiology*, 50(1):19–24, 1980.

[6] D. R. Brillinger. *Time series: data analysis and theory*, volume 36. SIAM, 1981.

[7] D. R. Brillinger. *Time series: data analysis and theory*, volume 36. SIAM, 2001.

[8] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer Science & Business Media, 2009.

[9] S. A. Burroughs, R. P. Morse, S. H. Mott, and G. L. Holmes. Brain connectivity in west syndrome. *Seizure*, 23(7):576–579, 2014.

[10] G. Buzsaki. *Rhythms of the Brain*. Oxford University Press, 2006.

[11] G. Buzsáki and A. Draguhn. Neuronal oscillations in cortical networks. *science*, 304(5679):1926–1929, 2004.

[12] C. Chang and G. H. Glover. Time–frequency dynamics of resting-state brain connectivity measured with fmri. *Neuroimage*, 50(1):81–98, 2010.

[13] J. Chau and R. von Sachs. Positive-definite multivariate spectral estimation: a geometric wavelet approach. *arXiv preprint arXiv:1701.03314*, 2017.

[14] C. J. Chu, M. A. Kramer, J. Pathmanathan, M. T. Bianchi, M. B. Westover, L. Wizon, and S. S. Cash. Emergence of stable functional networks in long-term human electroencephalography. *Journal of Neuroscience*, 32(8):2703–2713, 2012.

[15] M. Dai and W. Guo. Multivariate spectral analysis using cholesky decomposition. *Biometrika*, 91(3):629–643, 2004.

[16] A. Damianou and N. Lawrence. Deep gaussian processes. In *Artificial Intelligence and Statistics*, pages 207–215, 2013.

[17] C. C. de Vos, S. M. van Maarseveen, P. J. Brouwers, and M. J. van Putten. Continuous eeg monitoring during thrombolysis in acute hemispheric stroke patients using the brain symmetry index. *Journal of Clinical Neurophysiology*, 25(2):77–82, 2008.

[18] A. Delorme, T. Sejnowski, and S. Makeig. Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis. *Neuroimage*, 34(4):1443–1449, 2007.

[19] A. K. Engel and P. Fries. Beta-band oscillations-signalling the status quo? *Current opinion in neurobiology*, 20(2):156–165, 2010.

[20] J. Fan and E. Kreutzberger. Automatic local smoothing for spectral density estimation. *Scandinavian Journal of Statistics*, 25(2):359–369, 1998.

[21] M. Fiecas and H. Ombao. Modeling the evolution of dynamic brain process during a learning association experiment. *Journal of the American Statistical Association*, (Under revision), 2015.

[22] M. Fiecas and H. Ombao. Modeling the evolution of dynamic brain processes during an associative learning experiment. *Journal of the American Statistical Association*, 111(516):1440–1453, 2016.

[23] M. Fiecas, H. Ombao, C. Linkletter, W. Thompson, and J. Sanes. Functional connectivity: Shrinkage estimation and randomization test. *NeuroImage*, 49(4):3005–3014, 2010.

[24] P. Fillard, V. Arsigny, X. Pennec, K. M. Hayashi, P. M. Thompson, and N. Ayache. Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *Neuroimage*, 34(2):639–650, 2007.

[25] P. T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, pages 87–98. Springer, 2004.

[26] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi. The geometric median on riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1):S143–S152, 2009.

[27] B. Fyntanidou, V. Grosomanidis, Z. Aidoni, G. Thoma, M. Giakoumis, E. Kiurzieva, and C. Skourtis. Bispectral index scale variations in patients diagnosed with brain death. In *Transplantation proceedings*, volume 44, pages 2702–2705. Elsevier, 2012.

[28] H.-Y. Gao. Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis*, 18(3):231–251, 1997.

[29] H. Gastaut. Clinical and electroencephalographical classification of epileptic seizures. *Epilepsia*, 11(1):102–112, 1970.

[30] M. G. Genton, C. Johnson, K. Potter, G. Stenchikov, and Y. Sun. Surface boxplots. *Stat*, 3(1):1–11, 2014.

[31] S. Grün, M. Diesmann, and A. Aertsen. Unitary events in multiple single-neuron spiking activity: I. detection and significance. *Neural Computation*, 14(1):43–80, 2002.

[32] E. C. Hansen, D. Battaglia, A. Spiegler, G. Deco, and V. K. Jirsa. Functional connectivity dynamics: Modeling the switching behavior of the resting state. *NeuroImage*, 105:525–535, 2015.

[33] A. Holbrook, S. Lan, A. Vandenberg-Rodes, and B. Shahbaba. Geodesic lagrangian monte carlo over the space of positive definite matrices: with application to bayesian spectral density estimation. *Journal of Statistical Computation and Simulation*, 88(5):982–1002, 2018.

[34] R. A. Hrachovy, J. D. Frost, and P. Kellaway. Sleep characteristics in infantile spasms. *Neurology*, 31(6):688–688, 1981.

[35] R. A. Hrachovy, J. D. Frost, and P. Kellaway. Hypsarrhythmia: variations on the theme. *Epilepsia*, 25(3):317–325, 1984.

[36] R. A. Hrachovy and J. D. Frost Jr. Infantile epileptic encephalopathy with hypsarrhythmia (infantile spasms/west syndrome). *Journal of clinical neurophysiology*, 20(6):408–425, 2003.

[37] A. Isaksson, A. Wennberg, and L. H. Zetterberg. Computer analysis of eeg signals with parametric models. *Proceedings of the IEEE*, 69(4):451–461, 1981.

[38] S. Jain and G. Deshpande. Parametric modeling of brain signals. In *Biotechnology and Bioinformatics, 2004. Proceedings. Technology for Life: North Carolina Symposium on*, pages 85–91. IEEE, 2004.

[39] R. A. Joundi, N. Jenkinson, J.-S. Brittain, T. Z. Aziz, and P. Brown. Driving oscillatory activity in the human cortex enhances motor performance. *Current Biology*, 22(5):403–407, 2012.

[40] P. Kellaway. Sleep and epilepsy. *Epilepsia*, 26(s1), 1985.

[41] A. D. Krystal, R. Prado, and M. West. New methods of time series analysis of non-stationary eeg data: eigenstructure decompositions of time varying autoregressions. *Clinical Neurophysiology*, 110(12):2197–2206, 1999.

[42] T. C. Lee. A simple span selector for periodogram smoothing. *Biometrika*, 84(4):965–969, 1997.

[43] R. Y. Liu and K. Singh. Ordering directional data: concepts of data depth on circles and spheres. *The Annals of Statistics*, pages 1468–1484, 1992.

[44] R. Y. Liu and K. Singh. A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421):252–260, 1993.

[45] S. F. Logue and T. J. Gould. The neural and genetic basis of executive function: attention, cognitive flexibility, and response inhibition. *Pharmacology Biochemistry and Behavior*, 123:45–54, 2014.

[46] S. López-Pintado and J. Romo. On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734, 2009.

[47] S. Makeig. Auditory event-related dynamics of the eeg spectrum and effects of exposure to tones. *Electroencephalography and clinical neurophysiology*, 86(4):283–293, 1993.

[48] S. Makeig, A. J. Bell, T.-P. Jung, T. J. Sejnowski, et al. Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, pages 145–151, 1996.

[49] P. McCullagh and J. A. Nelder. Generalized linear models. pages 33–4, 1989.

[50] P. L. Nunez and R. Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford university press, 2006.

[51] H. Ombao and S. Van Bellegem. Evolutionary coherence of nonstationary signals. *IEEE Transactions on Signal Processing*, 56(6):2259–2266, 2008.

[52] H. C. Ombao, J. A. Raz, R. L. Strawderman, and R. Von Sachs. A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*, 88(4):1186–1192, 2001.

[53] A. Parihar, M. Jerry, S. Datta, and A. Raychowdhury. Stochastic imt (insulator-metal-transition) neurons: An interplay of thermal and threshold noise at bifurcation. *Frontiers in neuroscience*, 12:210, 2018.

[54] P. Pavone, P. Striano, R. Falsaperla, L. Pavone, and M. Ruggieri. Infantile spasms syndrome, west syndrome and related phenotypes: what we know in 2013. *Brain and Development*, 36(9):739–751, 2014.

[55] Y. Pawitan and F. O'sullivan. Nonparametric spectral density estimation using penalized whittle likelihood. *Journal of the American Statistical Association*, 89(426):600–610, 1994.

[56] X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127, 2006.

[57] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[58] D. B. Percival and A. T. Walden. *Spectral analysis for physical applications*. Cambridge University Press, 1993.

[59] G. Pfurtscheller and A. Aranibar. Evaluation of event-related desynchronization (erd) preceding and following voluntary self-paced movement. *Electroencephalography and clinical neurophysiology*, 46(2):138–146, 1979.

[60] M. B. Priestley. Spectral analysis and time series. 1981.

[61] C. E. Rasmussen and Z. Ghahramani. Infinite mixtures of gaussian process experts. In *Advances in neural information processing systems*, pages 881–888, 2002.

[62] K. S. Riedel and A. Sidorenko. Minimum bias multiple taper spectral estimation. *IEEE Transactions on Signal Processing*, 43(1):188–195, 1995.

[63] A. K. Roopun, S. J. Middleton, M. O. Cunningham, F. E. LeBeau, A. Bibbig, M. A. Whittington, and R. D. Traub. A beta2-frequency (20–30 hz) oscillation in nonsynaptic networks of somatosensory cortex. *Proceedings of the National Academy of Sciences*, 103(42):15646–15650, 2006.

[64] A. Schnitzler and J. Gross. Normal and pathological oscillatory communication in the brain. *Nature reviews neuroscience*, 6(4):285, 2005.

[65] B. Shahbaba, B. Zhou, S. Lan, H. Ombao, D. Moorman, and S. Behseta. A semiparametric bayesian model for detecting synchrony among multiple neurons. *Neural computation*, 26(9):2025–2051, 2014.

[66] D. S. Shumway and D. S. Stoffer. *Time series analysis and its applications*, volume 3. Springer New York, 2000.

[67] R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.

[68] C. Smith. Subspace, and intrinsic cramer-rao bounds, may 2005. *IEEE Transactions on Signal Processing*, 53(5):1610–1630, 2005.

[69] R. J. Smith, A. Sugijoto, N. Rismanchi, S. A. Hussain, D. W. Shrey, and B. A. Lopour. Long-range temporal correlations reflect treatment response in the electroencephalogram of patients with infantile spasms. *Brain topography*, 30(6):810–821, 2017.

[70] S. T. Smith. Intrinsic cramér-rao bounds and subspace estimation accuracy. In *Sensor Array and Multichannel Signal Processing Workshop. 2000. Proceedings of the 2000 IEEE*, pages 489–493. IEEE, 2000.

[71] R. Srinivasan and S. Deng. Multivariate spectral analysis of eeg: Power, coherence and second-order blind identification. *The Biomedical Engineering Handbook*, 4, 2011.

[72] Y. Sun and M. G. Genton. Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2), 2011.

[73] Y. Sun and M. G. Genton. Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics*, 23(1):54–64, 2012.

[74] Y. Sun and M. G. Genton. Functional median polish. *Journal of agricultural, biological, and environmental statistics*, 17(3):354–376, 2012.

[75] Y. Sun, M. G. Genton, and D. W. Nychka. Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? *Stat*, 1(1):68–74, 2012.

[76] G. Wahba. Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, 75(369):122–132, 1980.

[77] L. M. Ward. Synchronous neural oscillations and cognitive processes. *Trends in cognitive sciences*, 7(12):553–559, 2003.

[78] W. J. West. On a peculiar form of infantile convulsions. *The Lancet*, 35(911):724–725, 1841.

[79] J. Wu, R. Srinivasan, A. Kaur, and S. C. Cramer. Resting-state cortical connectivity predicts motor skill acquisition. *NeuroImage*, 91:84–90, 2014.

[80] Y. Yuan, H. Zhu, W. Lin, and J. Marron. Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(4):697–719, 2012.

Appendix A

Appendix Title

A.1 The matrix logarithm $\log(\Gamma)$

For a P–dimensional SPD matrix Γ, let $\Lambda = \text{diag}\,(\lambda_1, \ldots, \lambda_P)$ be a $P \times P$ diagonal matrix with λ_j's as the eigenvalues of Γ, and $U = (\vec{u}_1, \ldots, \vec{u}_P)$ where $\vec{u_j}$'s is corresponding orthonormal eigenvector for λ_j. A matrix logarithm $\log(\Gamma)$ can be calculated in three steps. First, we obtain the spectral decomposition of Γ, i.e. $\Gamma = U^T \Lambda U$. Second, we take natural logarithm for each element of Λ to get $\tilde{\Lambda} = \text{diag}\,(\log(\lambda_1), \ldots, \log(\lambda_P))$. Finally, the matrix logarithm $\log(\Gamma)$ can be computed as $\log(\Gamma) = U^T \tilde{\Lambda} U$. Note that the matrix exponential of Γ can be obtain by using the scalar exponential each eigenvector λ_j's; therefore, this is an one–to–one transformation for a SPD matrix.

A.2 Proof of Theorem 3.1

We want to show that $\text{Exp}_{\Gamma_1}\left(\text{Log}_{\Gamma_1}(\Gamma_2)\right) = \Gamma_2$. By using the Log and Exp definitions in Theorem 3.1, we have

$$\begin{aligned}
\text{Exp}_{\Gamma_1}\left(\text{Log}_{\Gamma_1}(\Gamma_2)\right) &= \text{Exp}_{\Gamma_1}\left(\log(\Gamma_2) - \log(\Gamma_1)\right) \\
&= \exp(\log(\Gamma_1) + \log(\Gamma_2) - \log(\Gamma_1)) \\
&= \Gamma_2.
\end{aligned}$$

A.3 Proof of Lemma 3.4.1

Under the null $H_o : \mathcal{P}^{(1)}(P) = \mathcal{P}^{(2)}(P)$, we have

$$\begin{aligned}
\text{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)}) = \text{Tr}\left[\log\left\{\left(\Gamma_j^{(1)}\right)^{-1}\Gamma_j^{(2)}\right\}\right] &\stackrel{d}{=} \text{Tr}\left[\log\left\{\left(\Gamma_j^{(2)}\right)^{-1}\Gamma_j^{(1)}\right\}\right] \quad \text{(Under } H_0\text{)} \\
&\stackrel{d}{=} \text{Tr}\left[\log(\Gamma_j^{(1)}) - \log(\Gamma_j^{(2)})\right] \quad (1) \\
&\stackrel{d}{=} -\text{Tr}\left[\log\left\{\left(\Gamma_j^{(1)}\right)^{-1}\Gamma_j^{(2)}\right\}\right] \quad (2),
\end{aligned}$$

where (1) and (2) hold because the identities $\text{Tr}(\log(AB)) = \text{Tr}(\log(A)) + \text{Tr}(\log(B))$ and $\text{Tr}(\log(A^t)) = t\text{Tr}(\log(A))$ for any $A, B \in \mathcal{P}(P)$ and $t \in \mathbb{R}$. It implies that $\text{E}\left(\text{diff}(\Gamma_j^{(1)}, \Gamma_j^{(2)})\right) = 0$ under the null hypothesis.

A.4 Proof of Lemma 3.4.2

For any $a \in GL(P)$, we have

$$\begin{aligned}
\mathrm{Tr}\left[\log\left\{\left(a\Gamma_j^{(1)}a^*\right)^{-1}\left(a\Gamma_j^{(2)}a^*\right)\right\}\right] &= \mathrm{Tr}\left[\log\left(a\Gamma_j^{(2)}a^*\right)\right] - \mathrm{Tr}\left[\log\left(a\Gamma_j^{(1)}a^*\right)\right] \quad (3)\\
&= \mathrm{Tr}\left[\log\left(\Gamma_j^{(2)}\right)\right] - \mathrm{Tr}\left[\log\left(\Gamma_j^{(1)}\right)\right]\\
&\quad + \mathrm{Tr}\left[\log\left(a^*a\right)\right] - \mathrm{Tr}\left[\log\left(a^*a\right)\right] \quad (4)\\
&= \mathrm{Tr}\left[\log\left\{\left(\Gamma_j^{(1)}\right)^{-1}\Gamma_j^{(2)}\right\}\right] \quad (5)
\end{aligned}$$