Title
Modeling Region-Referenced Longitudinal Functional Electroencephalography Data

Author
Scheffler, Aaron Wolfe

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Modeling Region-Referenced Longitudinal Functional Electroencephalography Data

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Biostatistics

by

Aaron Wolfe Scheffler

2019

ABSTRACT OF THE DISSERTATION

Modeling Region-Referenced Longitudinal Functional Electroencephalography Data

by

Aaron Wolfe Scheffler
Doctor of Philosophy in Biostatistics
University of California, Los Angeles, 2019
Professor Damla Şentürk, Chair

Highly structured data collected in a variety of biomedical applications such as electroencephalography (EEG) are discrete samples of a smooth functional process observed across both temporal and spatial dimensions. EEG data is conceptualized as region-referenced longitudinal functional data in which the functional dimension captures local signal dynamics, the longitudinal dimension tracks changes over the course of an experiment, and the regional dimension indexes spatial information across electrodes on the scalp. This complex data structure exhibits intricate dependencies with rich information but its dimensionality and size produce significant obstacles for interpretation, estimation, and inference. Motivated by a series of EEG studies in children with autism spectrum disorder (ASD), a set of computationally efficient methods for these high-dimensional data structures are proposed that both maintain information along each dimension and yield interpretable components and inferences.

The first half of the work considers decompositions of the total variation. To begin, a multi-dimensional functional principal components analysis (MD-FPCA) is introduced which decomposes the total variation into subject- and electrode-level components and for each level employs a two-stage functional principal components decomposition sequentially across functional and longitudinal time. Next, a hybrid principal components analysis (HPCA) for region-referenced longitudinal functional EEG data is proposed which utilizes both vector and functional principal components analyses and does not collapse information along any of

the three dimensions of the data. The second half of the work shifts to modeling associations and introduces a covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM) for modeling scalar outcomes from region-referenced functional predictors. CARR-GFLM utilizes a tensor basis formed from one-dimensional discrete and continuous bases to estimate functional effects across a discrete regional domain while simultaneously adjusting for additional non-functional covariates, such as age. Proposed methods not only help identify neurodevelopmental differences between typically developing and ASD children but can also be used to study the heterogeneity within children with ASD. The performance of all proposed methods is studied via extensive simulations.

The dissertation of Aaron Wolfe Scheffler is approved.

Catherine Ann Sugar

Donatello Telesca

Shafali Spurling Jeste

Damla Şentürk, Committee Chair

University of California, Los Angeles

2019

This dissertation is dedicated to Marian Wolfe, Scott Scheffler, Robin Wolfe Scheffler, and Tess Armstrong.

Your support has meant everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The six years I spent in the Department of Biostatistics at the University of California, Los Angeles (UCLA) have been both challenging and productive. During this time, I relied on the support and cooperation of a large number of people and I would like to acknowledge their contributions here. Specific details pertaining to co-authorship of manuscripts on which this dissertation is based are deferred to Chapter 1.5. Financially, my doctoral work was supported both by the UCLA Graduate Division and a grant from the National Institute of General Medical Sciences [R01 GM111378-01A1].

First and foremost, my adviser and committee chair, Dr. Damla Şentürk, has been a tireless mentor and advocate on my behalf. More than anyone, she taught me how to engage in both good methodological research and productive collaborations. Equally important, she shared her experience and wisdom by preparing me for opportunities I could not anticipate or expect. Few people have had such a singular impact on my life and I am more than grateful for all of her efforts.

Throughout my PhD, my dissertation committee have been tremendously supportive as well. In addition to providing a necessary dose of levity at meetings, Dr. Donatello Telesca shared his keen insights many times when editing manuscripts or approaching methodological problems. Outside of my doctoral work, Dr. Telesca kindly included me in joint work with Dr. Joanne Weidhaas from the David Geffen School of Medicine, UCLA, to explore biomarkers of cancer toxicity. This experience was invaluable not only for the resulting exposure to new methods but also as an opportunity to engage in cross-disciplinary research. In terms of manuscript review and the fundamental scientific merit of my work, Dr. Catherine Ann Sugar was always the final arbiter. Her approach of not becoming overly concerned with the technicalities of the model but rather to focus on the underlying hypothesis (along with her piercing questions) has been formative in my own understanding of the role of biostatisticians as practitioners of good science. Finally, Dr. Shafali Spurling Jeste generously invited me into her lab to work with both her and her students, allowing me to draw on their expertise in electroencephalography. Without her openness to new statistical methodologies

and collaboration, the essential motivating data that underpin this dissertation would not have been available to me.

While Dr. Jeste granted me access to data produced in her laboratory, the data exchange and follow-up was conducted with two experts, Drs. Charlotte DiStefano and Abigail Dickinson. Dr. DiStefano coordinated collection and processing of much of the data found in this work and Dr. Dickinson conducted additional processing and helped me to understand the underlying neurological phenomena under study. I am extremely grateful to both of these talented researchers for their partnership and input.

A constant element in my time at UCLA has been the overwhelming support of the administrative staff. Our departmental student affairs officer, Roxy Naranjo, was instrumental in my numerous applications for funding and awards, ensuring I could work towards my degree without financial stress. Ivonne Nelson helped connect me with grant support and never tired of my repeated visits to her office to secure a replacement key when I had locked myself out of the office.

My peers at UCLA have also had a tremendous impact on my academic development. As an early PhD student, I co-authored my first paper "A multi-dimensional functional principal components analysis of EEG data" (which makes up the second chapter of this dissertation) with Dr. Kyle Hasenstab. In transitioning this project to me, Kyle demonstrated patience in working through code and explaining statistical formulations, as well as providing earnest and heartfelt advice in the subsequent years. During the middle course of my degree, I shared an office in CHS 21-253 (more of a closet) with Dr. Qian Li, one of the most brilliant students I have had the pleasure to work with. It always felt like a game of catchup with Qian (both on and off the basketball court) and I learned a tremendous amount about hard work and statistics from our interactions. After Qian departed, John Shamshoian literally filled his seat and we had countless conversations about functional data, often with John explaining some feature of Bayesian statistics that was elementary to him but mystical to me. We also attended seminars on campus from other departments, exploring topics from natural language processing to causal inference, making our time together both intellectual and social.

Lastly, I need to thank my family and Tess Armstrong who supported me both locally and from afar (in space and time). My girlfriend, Tess, has been a rock in my life for five out of my six years at UCLA (sorry we did not meet the first year!). Her intelligence and determination are inspiring to me. Not only that, I knew I could always count on her love and support and I am a better person for knowing her. My older brother, Robin Wolfe Scheffler, has been a role model and mentor for much of my life but this really got put to the test when I entered graduate school. Robin shared his own experiences in order to better mine and despite being a busy guy, he never hesitated to edit a fellowship application or offer encouragement. Finally, I would like to thank my amazing parents, Marian Wolfe and Scott Scheffler, for their complete devotion to my education at every stage of my life. Dad, your work ethic and devotion to supporting our family and my goals is moving. I am proud of you and thankful for everything you have done on my behalf. Mom, you have instilled so many values in me that I hold dear. I am grateful for no cell phones at the dinner table, your emphasis on education and graduate training, and making sure we shared time together growing up. You selflessly took attention away from your career in order to provide love and support to all those around you. I could not ask for a better set of parents and your collective efforts made this work possible.

VITA

2011	B.A. (Biochemistry), Columbia University
2011–2013	Researcher, Healthy Communities Institute, Berkeley, CA
2014–2016	Graduate Student Researcher, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles
2015	M.S. (Biostatistics), University of California, Los Angeles
2015–2019	Graduate Student Researcher, Department of Biostatistics, University of California, Los Angeles
2017–present	Statistical Consultant, MiraDx, Los Angeles, CA
2019	Carolbeth Korn Award for the outstanding graduate of the Fielding School of Public Health, University of California, Los Angeles

CHAPTER 1

Introduction

First demonstrated in 1929 by the German physiologist Hans Berger, electroencephalography (EEG) is the measurement and recording of electrical activity in the human brain by electrodes placed near the scalp [Luck, 2014]. While Berger's recordings were made subcutaneously using needles, contemporary EEG systems include high density electrode arrays containing between 64 and 256 electrodes placed on the scalp, each capturing local electrical potentials produced by synchronized excitatory or inhibitory post-synaptic activity in the cortex [Collura, 1993, Ombao, 2017]. The local potentials recorded at each electrode form EEG signals that capture both spatial and temporal changes in cortical neural synchrony on the millisecond time scale. Due to the excellent temporal resolution and relative ease of recording when compared to other imaging techniques, for instance functional magnetic resonance imaging (fMRI), practitioners from fields such as neurology, clinical psychology, and pharmacology have used EEG in a variety of experimental paradigms to study cognitive processes and disease states [Millsap and Maydeu-Olivares, 2010].

However, the structure of EEG signals presents many challenges from a statistical perspective that must be overcome in order to extract meaningful information. Chief among these challenges is the high dimensional structure of EEG data. EEG signals are observed uninterrupted for several minutes at multiple electrodes and the continuous signal is divided into shorter segments to facilitate analysis with electrodes, segments, and signals within segments forming regional, longitudinal, and functional dimensions, respectively (Figure 1.1). While we may assume that the functional repetitions evolve continuously within a single electrode over the course of an experiment, signal dynamics among regions may vary substantially due to localized nature of neuronal activity. Therefore, we consider the regional dimension

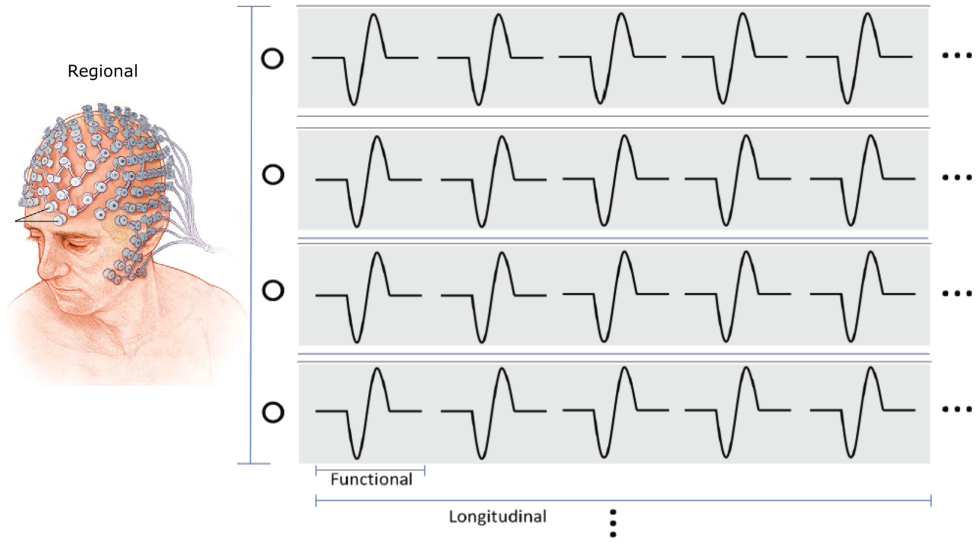


Figure 1.1: During a standard experimental paradigm, EEG signals are recorded at each electrode (regional dimension) and the continuous signal is divided into short segments (longitudinal dimension) with the signal within each segment encoding cortical neural activity (functional dimension).

to be discrete without the natural ordering provided by continuity. Thus, we conceptualize the EEG data as region-referenced longitudinal functional data in which the functional dimension captures local signal dynamics, the longitudinal dimension tracks changes in signal dynamics over the course of an experiment, and the regional dimension indexes spatial information across electrodes on the scalp. The resulting region-referenced longitudinal functional EEG data exhibits intricate dependencies rich with information that are difficult to capture using traditional statistical models. In order to apply standard statistical methods, the rich information contained in EEG data is often lost prior to modeling. EEG signals are collapsed across segments, resulting in an average signal that is further reduced into scalar features such as peak amplitude or frequency band power. Spatial information is similarly discarded by collapsing average signals or signal features across electrodes or restricting the analyses to a handful of electrodes contained in the high density electrode array. In this document, three functional data models are proposed that preserve the high dimensional structure of EEG data while simultaneously allowing for interpretation and analysis. The methods are demonstrated on EEG data collected in a series of studies that utilize event-related and

continuous paradigms to measure brain activity in young children with autism spectrum disorder (ASD). The remainder of this introduction discusses the basic framework of EEG experimental paradigms, defines the resulting EEG data structure, and presents functional data analysis as an alternative to the standard statistical analysis of EEG data to motivate the proposed methods.

1.1 Experimental paradigms for electroencephalography

Despite the wide variety of studies that employ EEG, experimental paradigms can essentially be classified into two categories, event-related paradigms and continuous paradigms. The distinguishing characteristic between these two categories is that event-related paradigms consider EEG signals time-locked to repeated presentation of stimuli, known as event-related potentials (ERPs), whereas continuous paradigms examine longer EEG signals without reference to a temporal event [Gross, 2014]. While both types of paradigms utilize EEG signals as a measure of brain activity, ERPs collected during event-related paradigms are analyzed in the time domain whereas EEG signals produced by continuous paradigms are typically analyzed in the frequency domain.

1.1.1 Event-related paradigms

The goal of event-related paradigms is to study the evolution of EEG signals in response to the presentation of time-locked stimuli. Subjects are repeatedly exposed to a stimulus, possibly under different conditions, and EEG signals are recorded at each electrode for a short time before and after exposure. Thus, each stimulus, or trial, generates an ERP waveform which is an instance of functional data. The experiments are made up of sequences of multiple trials, resulting in longitudinal functional data recorded at each electrode, adding a region dimension. The resulting ERP waveforms can be characterized by temporally ordered components called positivies (P) and negativities (N) that represent local maxima and minima of positive and negative electrical potential, respectively.

The characteristic components elicited vary among event-related paradigms based on the

stimuli presented. For example, in paradigms in which subjects are shown images, a component called the P3 peak is often observed 300 milliseconds after the stimulus and is thought to be associated with evaluation or categorization of stimuli [Daltrozzo and Conway, 2014]. The ERP components observed in a given experiment are typically studied by measuring their amplitude (size of component peak, measured in microvolts) or latency (time to component peak, measured in milliseconds). Because ERPs offer a way of measuring brain activity in response to stimuli, event-related paradigms have proven particularly useful in studying populations that are too young or developmentally delayed to provide overt behavioral responses [Luck, 2014]. Specifically, ERPs can act as covert measures of processing and have been highlighted as an instrumental method in the study of children with ASD [Jeste and Nelson, 2008].

1.1.2 Continuous paradigms

While event-related paradigms are concerned with the alteration of EEG signals in response to time-locked stimuli, continuous paradigms produce EEG signals recorded at each electrode over minutes or hours without reference to a specific temporal event. The resulting EEG signals are typically segmented into 1-3 second epochs and their analysis is focused on the spectral information in each epoch rather than changes in electrical potential over time. There exist a number of methods to transform epochs into the frequency domain but popular choices are the Fast Fourier Transform (FFT) and parametric autoregressive modeling [Ombao and Moon-Ho, 2006]. Thus, for each epoch there exists a corresponding spectral density which can be considered functional data. The collection of ordered epochs form longitudinal functional data recorded at each electrode, producing a similar data structure as found in event-related paradigms albeit with a different set of interpretations.

The exact understanding of the spectral information contained in each epoch depends on the presence or absence of stimulation during the course of the EEG recording [Gross, 2014]. One set of continuous paradigms take place during resting-state in which EEG signals are recorded on subjects not engaged in a task, typically with eyes closed or open, with the goal

of identifying functional connectivity among brain regions or baseline spectral characteristics. A second set of continuous paradigms collect EEG signals on subjects exposed to a task that lasts several minutes, such as reading or a sustained auditory stimulus. Spectral properties of the resulting task-based EEG are then correlated and contrasted across conditions and groups to identify differences or patterns. Similar to event-related paradigms, continuous paradigms have been heavily utilized as a method for studying neural function in children with ASD [Wang et al., 2013].

1.2 Standard statistical methods for EEG data

Despite the highly structured nature of EEG data, much of the information is discarded in order to allow for inference using standard statistical procedures, making the need for methods that model the dependency among EEG signals all the more urgent. Whether EEG is collected during an event-related or continuous paradigm, the observed signal can be considered a spatio-temporal stochastic process. Specifically, the EEG measured at each electrode is formed as a mixture of random signals corresponding in part to synchronized neuronal activity but also artifacts originating from a number of sources including facial muscles and nearby electromagnetic fields. Thus, artifact removal and correction is a key step in removing undesirable signals prior to statistical analysis and can impact the resulting data structure. A brief discussion of artifact rejection and correction will precede a summary of the loss of information that results from the application of standard statistical methods to EEG data.

1.2.1 Artifact rejection and correction

After pre-processing (i.e. segmentation, filtering, and referencing, see Cohen [2014] for extensive details), the rejection and correction of artifacts commonly takes place at the level of epochs for event-related and continuous paradigms, respectively. Note, for event-related paradigms, the formation of epochs for artifact rejection and correction precedes isolation of the ERP. Artifacts can be rejected by removing the entire portion of the contaminated signal

or corrected by an adjustment. Artifact rejection can be carried out by removing signals that exceed a pre-established amplitude threshold (e.g +/- 100 microvolts). On the other hand, artifacts may be corrected if the undesirable noise can be isolated and removed. One such method is Independent Components Analysis (ICA) which decomposes the EEG signal into independent components [Makeig et al., 1997]. Upon inspection, some components may exhibit voltage profiles that correspond to extracortical sources, such as muscular activity, and are removed when the signal is reconstructed. If a substantial number of artifacts are rejected rather than corrected, missingness can be introduced into the data structure which can produce challenges in statistical modeling.

1.2.2 Standard statistical method for the analysis of EEG data

Once artifacts are accounted for, EEG data is frequently collapsed across longitudinal, regional, and functional dimensions prior to statistical modeling due to a lack of appropriate methods, resulting in both a loss of information and interpretation. In event-related paradigms, the longitudinal dimension is collapsed by averaging ERPs across trials thus loosing potentially informative changes in ERPs over the course of the experiment. Further, regional information is often discarded by averaging ERPs across electrodes or restricting the analysis to a subset of predefined electrodes. Finally, functional information is summarized by features of the characteristic components such as latency and peak amplitude, which reduce the ERP from a curve into a single scalar value.

For continuous paradigms, the spectral information observed at each electrode and epoch is condensed in a similar manner. The observed spectral densities are averaged across epochs and summarized across electrodes to form single spectral density estimates for entire brain regions. The average spectral density is then integrated across clinically defined frequency bands of delta (0-4 Hz), theta (4-8 Hz), alpha (8-15 Hz), beta (15-32 Hz), and gamma (32-50 Hz), resulting in scalar summaries of spectral power for each band that do not account for the shape or trajectory of the spectral density.

Scalar summaries of ERPs and spectral information are then used as either outcomes or

predictors in generalized linear mixed models (GLMM), though it is often unclear whether Gaussian assumptions are reasonable even after the data is transformed (e.g. log transformed). If there are doubts about whether the observed data (or transformed data) are normally distributed, a number of robust non-parametric procedures have been applied including bootstrap analysis, permutation tests, Mann-Whitney U test, and Kruskal-Wallis test [Millsap and Maydeu-Olivares, 2010].

1.3 Functional data analysis

Rather than collapse the EEG data to accommodate traditional statistical methods, the observed signal (i.e. ERP or spectral density) may be considered discrete samples of an underlying functional process for which functional data analysis (FDA) offers a powerful methodological framework. FDA elucidates the underlying structure in this data by assuming that the basic unit of observation is a signal observed over some continuous domain [Ramsay and Silverman, 2010]. By modeling the entire functional process without collapsing information, functional data methods simultaneously maintain information and yield both interpretable components and inferences. While FDA grew from the analysis of one-dimensional curves, more and more applications produce functional data that are repeatedly observed across temporal or spatial domains (e.g. EEG recordings). The resulting data structure is rich in information but its dimensionality and size produce significant obstacles for interpretation, estimation, and inference. There is a critical need to expand existing FDA methods to accommodate this new generation of multi-dimensional functional data, motivating the methods proposed in this dissertation. A brief review of FDA follows, with emphasis on two work horses in the literature, functional principal components analysis and functional regression. These methods are detailed as they form the foundation for the functional models proposed in the following chapters.

1.3.1 Functional principal components analysis

Similar to principal components analysis (PCA) for multivariate data [Jolliffe, 2002], functional PCA (FPCA) is an instrumental method for dimension reduction for functional data [Wang et al., 2016]. Specifically, FPCA provides an orthogonal basis that explains more variation for a fixed dimension than any other basis, providing an optimal finite dimensional representation of an infinite dimensional functional process. Consider a sample of independent real-valued smooth functions, $X_i(t)$, $i = 1, \ldots, n$, defined on some continuous interval $t \in \mathcal{T}$. The mean and covariance functions, defined as $\mu(t) = E\{X_i(t)\}$ and $\Sigma(t, t') = \text{cov}\{X_i(t), X_i(t')\}$, provide functional information both across and within subjects, respectively. By Mercer's theorem, the covariance function can be expressed in terms of its spectral decomposition, $\Sigma\{t, t'\} = \sum_{k=1}^{\infty} \lambda_k \phi_k(t)\phi_k(t')$, where λ_k and $\phi_k(t)$ are ordered eigenvalues and eigenfunctions (functional principal components), respectively. The Karhunen-Loève (KL) expansion (FPCA expansion) [Karhunen, 1946, Loeve, 1946] of a random signal is given by,

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t),$$

where $\xi_{ik} = \int \{X_i(t) - \mu(t)\}\phi_k(t)dt$ are the functional principal component scores which capture the stochastic behavior of the process. The scores are uncorrelated and encapsulate all the stochastic information in the functional process with $E(\xi_{ik}) = 0$ and $\text{var}(\xi_{ik}) = \lambda_k$. If the KL expansion is truncated to include only K components, a finite dimensional approximation to the signal is obtained. In addition to providing an optimal basis, the functional principal components themselves provide useful interpretations by characterizing the directions of principal variation in the observed data.

Since the introduction of the KL expansion, a detailed literature has developed around the estimation of functional principal scores and components for both densely and sparsely observed functional data along a single dimension (see Wang et al. [2016] for a thorough review). In recent years, the literature surrounding FPCA has shifted to consider functional data with more complex dependency structures, including repeatedly measured functional

data [Crainiceanu et al., 2009, Di et al., 2009, Kundu et al., 2016, Morris and Carroll, 2006, Morris et al., 2003, Zipunnikov et al., 2014], longitudinally observed functional data [Greven et al., 2010, Chen and Müller, 2012, Park and Staicu, 2015, Chen et al., 2016], spatially correlated functional data [Baladandayuthapani et al., 2007, Giraldo et al., 2010, Zhou et al., 2010, Staicu et al., 2010, Liu et al., 2017], and multivariate functional data [Jaques and Preda, 2014, Chiou et al., 2014, Happ and Greven, 2018]. EEG and similar data structures present a unique set of challenges that are not addressed by previous FPCA methods and a major focus of this work is to develop computationally efficient and interpretable principal component decompositions to bridge this gap.

1.3.2 Functional regression

Whereas FPCA identifies the principal directions of variation in functional data to provide an optimal basis representation, functional regression methods model and characterize associations in functional data (for a comprehensive review, see Morris [2015] and Greven and Scheipl [2017]). Functional regression models can divided into three categories based on the role of the functional data, (1) scalar-on-function regression (SoFR), (2) function-on-scalar regression (FoSR), and (3) function-on-function regression (FoFR). While methods within these three categories share common attributes and assumptions, they also each carry a distinct set of challenges. For instance, FoFR and FoSR models often produce greater computational burden due to the size of design matrices, while SoFR must deal more directly with the 'small n, big p' problem given that the dimension of the functional predictor exceeds the number of observed scalar responses. In this dissertation, a SoFR model will be proposed and thus the remaining discussion of functional regression methods will focus on this class of models.

Consider the pair $\{y_i, X_i(t)\}$, where y_i is a scalar response and $X_i(t)$ is a functional predictor defined as in Section 1.3.1. The foundational functional regression model proposed by Ramsay and Dalzell [1991] assumed a linear relationship between predictor and outcome and was explicitly described for SoFR with a Gaussian response by Hastie and Mallows

[1993],

$$y_i = \beta_0 + \int X_i(t)\beta(t)dt + \epsilon_i,$$

where $\beta(t)$ is a regression function, β_0 is an intercept term, and $\epsilon_i \sim N(0, \sigma^2)$ is residual error. The effect of the functional predictor $X_i(t)$ is mapped by a regression function $\beta(t)$, the product of which are integrated to produce an effect on the scale of the outcome. Marx and Eilers [1999] extended this model to include exponential family responses and both models have been adapted to accommodate multilevel functional predictors and both non-parametric and non-linear frameworks (see Reiss et al. [2017] for more details). In the linear setting, estimation of the regression function $\beta(t)$ often takes place by projection onto a suitable basis (e.g. B-splines, Fourier basis, wavelets) and penalizing the resulting loadings to obtain desirable properties (e.g sparsity or smoothness). Just as was the case with FPCA, as the complexity of the functional predictors grow, challenges emerge in formulating and estimating a suitable functional regression model. While some methods exist to address multivariate (i.e. multiple functional signals defined on possibly different domains; Zhu et al. [2010], Gertheiss et al. [2013], Lian [2013]) and multi-dimensional (i.e. two- or higher-dimensional functional signals defined continuously on a single domain; Marx and Eilers [2005], Reiss and Ogden [2010], Goldsmith et al. [2014]) functional predictors, no existing methods accommodate the complex nature of EEG data and similar data structures that exhibit complex spatial and temporal correlations as detailed in the next section. A second major focus of this dissertation is to develop a SoFR method suited for the challenges of modeling region-referenced functional data such as the motivating EEG studies.

1.4 Overview of proposed methods

The collapse of longitudinal, regional, and functional dimensions is largely motivated by a lack of available statistical methods and software suited for the high-dimensional structure of EEG data. In order to preserve the high-dimensional structure of EEG data produced during event-related and continuous paradigms and allow for inference and interpretation

along the functional, longitudinal, and regional dimensions, three statistical methods are developed.

Chapter 2 develops multi-dimensional functional principal components analysis (MD-FPCA) to model longitudinal and spatial information in ERPs produced during an implicit learning paradigm in children with ASD. The decomposition is based on separation of the total variation into subject and subunit level variation which are further decomposed in a two-stage FPCA. In the context of EEG data, electrodes are grouped into scalp regions and within each scalp region electrodes are considered exchangeable. The proposed methodology is shown to be useful for modeling longitudinal trends in ERP functions while accounting for repetitions across electrodes within scalp regions, leading to novel insights into the learning patterns of children with ASD and their typically developing peers as well as comparisons between the two groups. Emphasis is placed on the longitudinal evolution of ERPs across experimental time due to the fact that changes in ERP morphology may provide insight into underlying cognitive processes [Daltrozzo and Conway, 2014].

Chapter 3 proposes a hybrid principal components analysis (HPCA) that provides a low-dimensional decomposition of EEG data via weak separability of the total covariance process. While MD-FPCA offers a flexible but intricate model of the total variance through a multi-level two-stage decomposition, HPCA offers a more parsimonious decomposition in terms of an empirical tensor basis based on one-dimensional eigenvectors and eigenfunctions obtained from the marginal covariances. The motivating example is a word segmentation paradigm in which TD children and children with ASD were exposed to a continuous speech stream. For each subject, continuous EEG signals recorded at each electrode were divided into one-second segments and projected into the frequency domain via Fast Fourier Transform. Following a spectral principal components analysis for dimension reduction, the resulting data consist of region-referenced principal power indexed regionally by scalp location, longitudinally by one-second segments, and functionally across frequencies. HPCA utilizes both vector and functional principal components analyses via a product of one dimensional eigenvectors and eigenfunctions obtained from the regional, functional, and longitudinal marginal covariances, to represent the observed data, providing a computationally feasible nonparametric

approach. A mixed effects framework is proposed to estimate the model components coupled with a bootstrap test for group level inference, both geared towards sparse data applications. Analysis of the data from the word segmentation paradigm leads to valuable insights about group-region differences among the TD and ASD children.

Chapter 4 shifts from covariance decompositions to introduce a covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM) that models clinical outcomes from region-referenced EEG data while adjusting for non-functional covariates. The proposed method is motivated by a resting state study in TD and ASD children evaluating an EEG biomarker called the peak alpha frequency (PAF), defined as the maximum of a prominent peak in the alpha frequency band. To retain the most information, oscillations in the spectral density within the entire alpha band rather than just the peak location are modeled as a functional predictor of diagnostic status (TD vs. ASD). CARR-GFLM utilizes a penalized tensor basis formed from discrete and continuous basis functions to estimate functional effects across a discrete regional domain while simultaneously adjusting for non-functional covariates. As case in point, alpha band oscillations exhibit developmental changes and thus chronological age must be accounted for when estimating effects. Application of CARR-GFLM to the resting state EEG data demonstrate that developmental trajectories in alpha spectral dynamics are associated with autism diagnostic status.

In summary, the three proposed methods maintain the high dimensional data structure observed in EEG data and allow for interpretation and analysis along each dimension. In Chapter 2, the proposed MD-FPCA decomposition based on the separation of the total variation into subject and subunit level variation models the longitudinal evolution of ERPs across experimental time and leads to novel insights into the learning patterns of children with ASD and their TD peers in an implicit learning paradigm. In Chapter 3, the HPCA decomposition provides a low dimensional approximation of region-referenced longitudinal functional EEG data via weak separability of the covariance process and a bootstrap procedure is proposed that allows for group-level comparisons of TD and ASD children in a task-based continuous paradigm. In Chapter 4, CARR-GFLM utilizes alpha oscillatory dynamics to model diagnostic status while accounting for developmental progression to illu-

minate developmental differences between TD and ASD children. Chapter 5 concludes with a discussion of the proposed methods and outlines future methodological developments that could follow this work.

1.5 Contributing publications

The contents of this dissertation are located in the publications and manuscripts detailed below. Authors with primary writing contributions are noted in italics under each reference.

Chapter 2: Multi-dimensional functional principal components analysis

> A. Scheffler †, K. Hasenstab †, D. Telesca, C. Sugar, S. Jeste, C. DiStefano, D. Şentürk (2017). A multi-dimensional functional principal components analysis of EEG data. *Biometrics* 73(3), 999-1009. (†Authors contributed equally)
> *Drafted by Aaron Scheffler, Kyle Hasenstab, and Damla Şentürk*

Chapter 3: Hybrid principal components analysis

> A. Scheffler, D. Telesca, Q. Li, C. Sugar, C. DiStefano, S. Jeste, D. Şentürk (2018). Hybrid principal components analysis for region-referenced longitudinal functional EEG data. *Biostatistics* (Epub ahead of print).
> *Drafted by Aaron Scheffler and Damla Şentürk*

Chapter 4: Covariate-adjusted region-referenced generalized functional linear model

> A. Scheffler, D. Telesca, Q. Li, C. Sugar, S. Jeste, A. Dickinson, C. DiStefano, D. Şentürk (2018). Covariate-adjusted region-referenced generalized functional linear model for EEG data. (under review)
> *Drafted by Aaron Scheffler and Damla Şentürk*

1.6 Software and implementation

The analyses carried out in this dissertation were performed on a 2.4 GHz 6-Core Intel Xeon processor operating MATLAB (v. 8.6.0.267246, The MathWorks Inc. [2015b]) and R (v. 3.5.1, R Core Team [2018]). Code for the proposed estimation and inference procedures are publicly available online at [https://github.com/aaron-scheffler], along with tutorials for step-by-step implementation of the proposed methodologies on simulated data.

CHAPTER 2

Multi-Dimensional Functional Principal Components Analysis

The electroencephalography (EEG) data created in event-related potential (ERP) experiments have a complex high-dimensional structure. Each stimulus presentation, or trial, generates an ERP waveform which is an instance of functional data. The experiments are made up of sequences of multiple trials, resulting in longitudinal functional data and moreover, responses are recorded at multiple electrodes on the scalp, adding an electrode dimension. Traditional EEG analyses involve multiple simplifications of this structure to increase the signal-to-noise ratio, effectively collapsing the functional and longitudinal components by identifying key features of the ERPs and averaging them across trials. Motivated by an implicit learning paradigm used in autism research in which the functional, longitudinal and electrode components all have critical interpretations, we propo gse a multi-dimensional functional principal components analysis (MD-FPCA) technique which does not collapse any of the dimensions of the ERP data. The proposed decomposition is based on separation of the total variation into subject and subunit level variation which are further decomposed in a two-stage functional principal components analysis. The proposed methodology is shown to be useful for modeling longitudinal trends in the ERP functions, leading to novel insights into the learning patterns of children with Autism Spectrum Disorder (ASD) and their typically developing peers as well as comparisons between the two groups. Finite sample properties of MD-FPCA are further studied via extensive simulations.

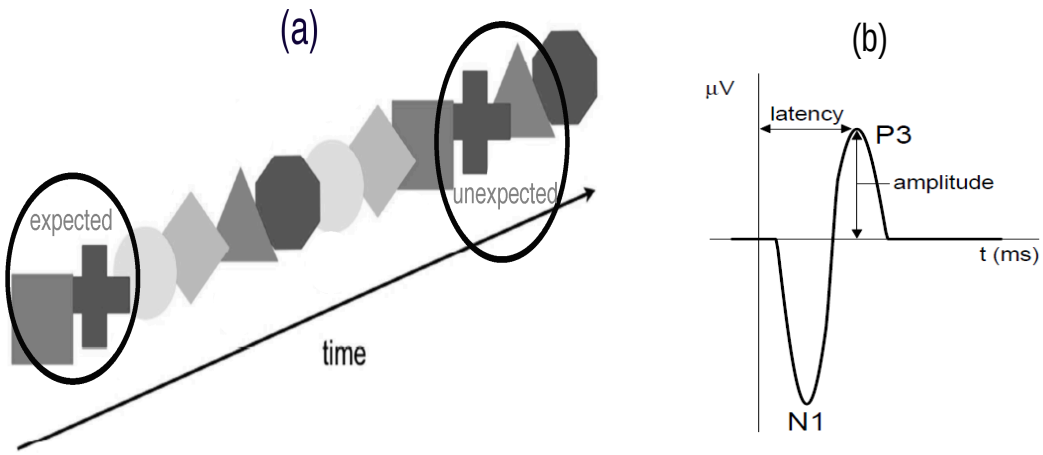


Figure 2.1: (a) The sequence of shape pairs in the implicit learning study. Transitions within a shape pair are labelled 'expected' (square and cross are a shape pair, so that the cross always follows the square); transitions between shape pairs are labelled 'unexpected'. (b) The ERP waveform containing the P3 and N1 phasic components from the implicit learning study.

2.1 Introduction

Electroencephalography (EEG) is a well-established noninvasive method for measuring spontaneous electrical activity across brain regions to identify neural function and cognitive states. Our motivating data is from a visual implicit learning study on young children with autism spectrum disorder (ASD) [Jeste et al., 2015]. The experiment involved event-related potentials (ERP) in which EEG signals were time locked to the presentation of a continuous sequence of colored shapes (visual stimuli) recorded in age-matched 2 to 5 year old typically developing and ASD children (Figure 2.1(a)). The six colored shapes, grouped into three shape pairs, were presented in random order. Transitions within a shape pair were labeled 'expected' since they could be learned (shape ordering within a pair was fixed) and transitions between shape pairs were labeled 'unexpected' since they could not be predicted. The goal of the study was to characterize implicit learning, defined as the detection of regular patterns in one's environment without a conscious awareness or intention to learn, by contrasting brain response to expected and unexpected transitions.

The data created in typical ERP studies as the one described above are rich and multi-dimensional. Each stimulus, corresponding to the presentation of a single shape, referred to as a trial, results in an ERP function with paradigm-specific phasic components. The P3 peak and N1 dip phasic components typically studied in this paradigm and thought to be related to cognitive processes and early category recognition are given in Figure 2.1(b) [Jeste et al., 2015]. Hence the experiment creates functional data (ERP curves) for each subject, collected longitudinally over trials (presentation of each shape) at multiple electrodes placed on the scalp. Due to the richness and multifaceted nature of the data along functional, longitudinal and electrode dimensions (repetitions over electrodes), typical practice involves multiple simplifications of the data before analysis. To increase the low signal-to-noise ratio (SNR) in raw ERP data, data are first collapsed in the longitudinal dimension, in which ERP functions observed over trials are averaged for each subject [Jeste et al., 2015, Gasser and Molinari, 1996]. In addition, aside from a few works on functional mixed effects modeling for the analysis of ERP data [Bugli and Lambert, 2006, Davidson, 2009], the functional dimension is typically summarized by the amplitude (magnitude of the peak or dip) or latency (time when the peak or dip occurs) of the phasic components of the averaged ERP functions. Hence, once both functional and longitudinal dimensions have been collapsed into one-dimensional data summaries, a repeated measures ANOVA can be used for analysis.

In this paper we propose, for the first time in the literature, a longitudinal principal components decomposition for the EEG data, which we refer to as multi-dimensional functional principal components analysis (MD-FPCA). MD-FPCA embodies all three dimensions (functional, longitudinal and electrode) of the ERP data, preserving the full complexity without stringent assumptions or data reduction. In order to increase the SNR without collapsing the longitudinal dimension over trials, we adopt MAP-ERP, a meta-preprocessing step based on a moving average of the ERP functions over trials in a sliding window [Hasenstab et al., 2015]. Capturing the longitudinal dimension is especially important in settings such as our motivating example, where patterns of learning correspond by definition to changes in ERP functions across trials. Previous studies in neuroscience and biomedical engineering have acknowledged that ERP function morphology may change over the course of a task. How-

ever, most prior work has focused on controlling for longitudinal trends [Gasser et al., 1983, Turetsky et al., 1989] rather than modeling them; the few works on modeling longitudinal trends have been limited to parametric forms [Rossi et al., 2007, De Silva et al., 2012]. We will build our proposed MD-FPCA on data produced by the novel meta-preprocessing step, MAP-ERP, capturing the continuum of longitudinal dynamics.

The literature on functional data analysis (FDA) [Ramsay and Silverman, 2010] has grown rapidly over the past two decades, with a considerable fraction of the work involving applications to longitudinal data [James et al., 2000, Müller, 2008, Şentürk and Müller, 2010]. More recently, there has been interest in analyzing multiple trajectories, with dependencies among the repeatedly measured functional data [Crainiceanu et al., 2009, Di et al., 2009, Kundu et al., 2016, Morris and Carroll, 2006, Morris et al., 2003, Zipunnikov et al., 2014]. Functional principal components decompositions for multilevel or longitudinal functional processes have been a major modeling theme in the FDA literature. Di et al. [2009] suggested decomposing sources of functional variation in an additive fashion via multilevel ANOVA, which we refer to as the ANOVA functional principal components decomposition (ANOVA-FPCA). Greven et al. [2010] proposed a decomposition based on a functional random intercept and slope to capture longitudinal variations, which we refer to as linear FPCA (LFPCA). Chen and Müller [2012] suggested a double decomposition (DFPCA) to capture potential nonlinear and nonparametric longitudinal trends within repeatedly observed functional data; parsimonious extensions of DFPCA have recently been proposed by Park and Staicu [2015] and Chen et al. [2016]. While ANOVA-FPCA models longitudinal repetitions as repeated measurements without a particular time ordering, similar to an ANOVA, LFPCA models longitudinal trends linearly, and DFPCA does not assume a parametric form. For spatially correlated functional data, Delicado et al. [2010] summarize limited works in three categories, analysis of geostatistical functional data [Baladandayuthapani et al., 2007, Giraldo et al., 2010, Zhou et al., 2010, Staicu et al., 2010, Liu et al., 2017] (mostly involving distance-based parametric correlation structures), point processes with associated functional data and functional areal data.

Our proposed MD-FPCA combines the flexible DFPCA modeling of longitudinal trends,

especially important for modeling learning trajectories in the motivating implicit learning experiment, with the decomposition of the total variation into subject and electrode level components as in ANOVA-FPCA, to embody all the dimensions of the multi-dimensional ERP data. MD-FPCA induces correlation between the electrode repetitions via random effects and utilizes multilevel random effects for extensions that involve data from multiple scalp regions. Following the initial ANOVA decomposition of the total variation, the proposed MD-FPCA involves a two-stage functional principal components decomposition of the subject and electrode level variations across functional and longitudinal time, leading to highly interpretable components contributing to the principal surfaces in a multiplicative fashion. Hence, even though multiple decompositions have been proposed for longitudinally observed or spatially correlated functional data in the literature, MD-FPCA is the first decomposition proposed for repeatedly measured longitudinal functional data which is tailored to model the specific features of the EEG data produced in ERP studies.

The remainder of the paper is organized as follows. Section 2.2 introduces the proposed MD-FPCA approach, compares it with other recently proposed functional principal components decompositions for longitudinally observed functional data, and outlines the extension of the methodology to analysis of data from multiple scalp regions. Section 2.3 provides insights gained from the implicit learning application, including comparisons of learning patterns in ASD and TD groups, summarized via the longitudinal trends in ERP functions across the experiment. We study the performance of the proposed decomposition in extensive simulation studies summarized in Section 2.4 and conclude with a brief discussion in Section 2.5.

2.2 Multi-dimensional functional principal components analysis (MD-FPCA)

2.2.1 The proposed MD-FPCA decomposition

Denote by $X_{ij}(t|s)$ a multilevel square integrable random function observed across continuous functional time t, $t \in \mathcal{T}$, at longitudinal time s, $s \in \mathcal{S}$, for subunit j, $j = 1, \ldots, J$, and subject i, $i = 1, \ldots, n$. In applications to the EEG data, data collected over electrodes represent the subunits within subjects, functional time is the time scale of the ERP function and longitudinal time corresponds to trials. The notation, $X_{ij}(t|s)$, for the multilevel longitudinal functional process is used to stress the two-stage nature of the Karhunen-Loève decompositions in MD-FPCA, where the first-stage expansions are conditional on a particular longitudinal time s, and second-stage decompositions describe the variations along longitudinal time s. The function $X_{ij}(t|s)$ is decomposed using a multilevel random effects model at each longitudinal time s,

$$X_{ij}(t|s) = \mu(t,s) + \eta_j(t,s) + Z_i(t|s) + W_{ij}(t|s) + \epsilon_{ij}(t|s), \tag{2.1}$$

where $\mu(t,s)$ and $\eta_j(t,s)$ are fixed functional effects that represent the overall mean function and subunit-specific shifts, respectively; $Z_i(t|s)$ and $W_{ij}(t|s)$ are the random subject- and subunit-specific deviations, respectively; and $\epsilon_{ij}(t|s)$ is measurement error with mean zero and variance σ_s^2. Denote the total variation of $X_{ij}(t|s)$ at a fixed longitudinal time s by $\Sigma_T(t,t'|s) = \text{cov}\{X_{ij}(t|s), X_{ij}(t'|s)\}$ and let $\widetilde{\Sigma}_T(t,t'|s) = \text{cov}\{X_{ij}(t|s), X_{ij}(t'|s)\} - \sigma_s^2 1_{\{t=t'\}}$ be the total variation without the measurement error with $1_{\{A\}}$ denoting the indicator function for event A. Assuming the subject and subunit-specific deviations, $Z_i(t|s)$ and $W_{ij}(t|s)$, are uncorrelated mean zero stochastic processes, (2.1) implies separation of the total variation $\widetilde{\Sigma}_T(t,t'|s)$ at each longitudinal time s into subject level $\Sigma^{(1)}(t,t'|s) = \text{cov}\{X_{ij}(t|s), X_{ij'}(t'|s)\} = \sum_k \lambda_k^{(1)}(s)\phi_k^{(1)}(t|s)\phi_k^{(1)}(t'|s)$ and electrode level $\Sigma^{(2)}(t,t'|s) = \widetilde{\Sigma}_T(t,t'|s) - \Sigma^{(1)}(t,t'|s) = \sum_p \lambda_p^{(2)}(s)\phi_p^{(2)}(t|s)\phi_p^{(2)}(t'|s)$ variation. Note that $\Sigma^{(1)}(t,t'|s)$ captures variation between electrodes within a subject and $\Sigma^{(2)}(t,t'|s)$ represents the remaining

second level variance; we refer to these quantities as subject and electrode level variations, respectively, to build intuition that this separation is analogous to an ANOVA decomposition. In this formulation, $\phi_k^{(1)}(t|s)$ and $\phi_p^{(2)}(t|s)$ are the first and second level eigenfunctions, describing modes of variation across functional time t at each longitudinal time s, and $\lambda_k^{(1)}(s)$ and $\lambda_p^{(2)}(s)$ are the first and second level eigenvalues.

Note that while the terminology 'level' is used to refer to the separation of the variation into the subject and electrode components, 'stage' is going to be used to refer to the two subsequent functional principal components decompositions applied to the deviation/variance defined at each level, first conditional on a particular longitudinal time s, and second describing variations along s. The first-stage Karhunen-Loève decompositions for $Z_i(t|s) = \sum_k \xi_{ik}(s)\phi_k^{(1)}(t|s)$ and $W_{ij}(t|s) = \sum_p \zeta_{ijp}(s)\phi_p^{(2)}(t|s)$ are carried out at each s, yielding

$$X_{ij}(t|s) = \mu(t,s) + \eta_j(t,s) + \sum_{k=1}^{\infty} \xi_{ik}(s)\phi_k^{(1)}(t|s) + \sum_{p=1}^{\infty} \zeta_{ijp}(s)\phi_p^{(2)}(t|s) + \epsilon_{ij}(t|s),$$

where $\xi_{ik}(s)$ and $\zeta_{ijp}(s)$ are the first and second level eigenscores, respectively, such that $\text{var}\{\xi_{ik}(s)\} = \lambda_k^{(1)}(s)$ and $\text{var}\{\zeta_{ijp}(s)\} = \lambda_p^{(2)}(s)$. In practice, the decompositions are truncated at only a small number of eigencomponents K and P (Appendix 2A). Note that the subunit repetitions within a subject, $W_{ij}(t|s)$ and hence $\zeta_{ijp}(s)$, are assumed to be independent across j, since the subunit dependency is modeled by the subject-specific random component $Z_i(t|s)$. Next, the second-stage Karhunen-Loève decompositions for the first and second level eigenscores, $\xi_{ik}(s) = \sum_{k'=1}^{\infty} \xi'_{ikk'}\psi_{kk'}^{(1)}(s)$, $\zeta_{ijp}(s) = \sum_{p'=1}^{\infty} \zeta'_{ijpp'}\psi_{pp'}^{(2)}(s)$, yield

$$X_{ij}(t|s) = \mu(t,s)+\eta_j(t,s)+\sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} \xi'_{ikk'}\psi_{kk'}^{(1)}(s)\phi_k^{(1)}(t|s)+\sum_{p=1}^{\infty}\sum_{p'=1}^{\infty} \zeta'_{ijpp'}\psi_{pp'}^{(2)}(s)\phi_p^{(2)}(t|s)+\epsilon_{ij}(t|s). \tag{2.2}$$

In (2.2), $\xi'_{ikk'}$ and $\zeta'_{ijpp'}$ are the eigenscores, $\lambda_{kk'}^{(1)} = \text{var}(\xi'_{ikk'})$ and $\lambda_{pp'}^{(2)} = \text{var}(\zeta'_{ijpp'})$ are the eigenvalues and $\psi_{kk'}^{(1)}(s)$ and $\psi_{pp'}^{(2)}(s)$ are the eigenfunctions describing the modes of variation across longitudinal time of the first-stage eigenscores.

We propose two decomposition summaries for MD-FPCA that are important in identify-

ing the contributions of different sources to the total variation in the analysis of the multilevel stochastic process $X_{ij}(t|s)$. While the first quantity, $\rho(s) = \{\sum_k \lambda_k^{(1)}(s)\}/\ \{\sum_k \lambda_k^{(1)}(s) + \sum_p \lambda_p^{(2)}(s)\}$, summarizes the proportion of variability explained by the subject level (first level) variation in the first-stage of MD-FPCA conditional on longitudinal time s, the second summary measure, $\rho = \{\sum_k \sum_{k'} \lambda_{kk'}^{(1)}\}/\{\sum_k \sum_{k'} \lambda_{kk'}^{(1)} + \sum_p \sum_{p'} \lambda_{pp'}^{(2)}\}$, captures the overall proportion of variability explained by the subject level variation in both stages of the MD-FPCA across longitudinal time s. The two summaries can be viewed as extensions of the intra-cluster correlation of the linear mixed effects framework to the decomposition of multilevel longitudinally observed functional processes. The intraclass correlations can also be interpreted as the average correlation between two subunits from the same subject, conditional on and across longitudinal time s, respectively. In applications to ERP data, repetitions over electrodes are considered subunits; within-subject correlations between these repetitions provide insight into the similarity of the trends across electrodes.

2.2.2 Comparison to other Karhunen-Loève decompositions

We briefly review three recently proposed principal components decompositions for functional processes, which are special cases of the proposed MD-FPCA, and highlight differences of MD-FPCA from a multilevel two-dimensional Karhunen-Loève decomposition. The three special cases are given without additive measurement error for simplicity. The ANOVA-FPCA of Di et al. [2009],

$$X_{ij}(t) = \mu(t) + \eta_j(t) + Z_i(t) + W_{ij}(t) = \mu(t,s) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(1)}(t) + \sum_{p=1}^{\infty} \zeta_{ijp}\phi_p^{(2)}(t), \qquad (2.3)$$

is a special case of MD-FPCA, where the decomposition does not have a longitudinal component. The repeatedly observed functional process $X_{ij}(t)$ is decomposed into an overall mean $\mu(t)$, a visit-specific mean $\eta_j(t)$, a subject-specific deviation from the visit-specific mean $Z_i(t)$ and a subject-visit-specific deviation $W_{ij}(t)$. Similar to MD-FPCA, $Z_i(t)$ and $W_{ij}(t)$, which are called the first and second level deviations, respectively, are expanded using Karhunen-Loève decompositions with first and second level eigenscores ξ_{ik} and ζ_{ijp} and eigenfunctions

$\phi_k^{(1)}(t)$ and $\phi_p^{(2)}(t)$, respectively. Note that repetitions of the functional process are modeled without a particular time ordering, similar to an ANOVA.

In contrast, the linear FPCA (LFPCA) of Greven et al. [2010] models longitudinal trends in a repeatedly observed functional process linearly. The double FPCA (DFPCA) decomposition of Chen and Müller [2012],

$$X_i(t|s) = \mu(t,s) + Z_i(t|s) = \mu(t,s) + \sum_{k=1}^{\infty} \xi_{ik}(s)\phi_k(t|s) = \mu(t,s) + \sum_{k=1}^{\infty}\sum_{p=1}^{\infty} \zeta_{ikp}\psi_{kp}(s)\phi_k(t|s), \tag{2.4}$$

does not assume a parametric form for the longitudinal time trend and thus can capture very flexible dynamics, similar to MD-FPCA. Note that the decomposition is for a longitudinally observed functional process that is not repeatedly observed, hence is a special case of MD-FPCA which considers the repeatedly observed longitudinal functional process $X_{ij}(t|s)$ with subject- and subunit-specific deviations. In (2.4), $X_i(t|s)$ is decomposed at a grid of longitudinal times s, yielding the mean function $\mu(t,s)$, eigenfunctions $\phi_k(t|s)$ and subject-specific random eigenscores $\xi_{ik}(s)$ at the first-stage. The eigenscores at each longitudinal time s are then decomposed further at the second-stage to yield eigenscores ζ_{ikp} and eigenfunctions $\psi_{kp}(s)$. Only decompositions LFPCA and DFPCA are suitable for situations in which interest centers on detecting changes in the ERP functions across longitudinal time. Moreover, in applications to the implicit learning paradigm, even LFPCA may be quite restrictive, since it requires these trends to be linear. Hence the proposed MD-FPCA combines the more flexible DFPCA to model longitudinal trends with ANOVA-FPCA to model repetitions over electrodes (the electrode dimension), enabling us to compare the nature of the implicit learning processes of children with ASD and their typically developing peers.

In comparing MD-FPCA to a multilevel two-dimensional Karhunen-Loève decomposition, note that the proposed MD-FPCA in (2.2) implies principal surfaces across both the functional and longitudinal domains,

$$X_{ij}(t|s) = \mu(t,s) + \eta_j(t,s) + \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} \xi'_{ikk'}\varphi^{(1)}_{kk'}(t,s) + \sum_{p=1}^{\infty}\sum_{p'=1}^{\infty} \zeta'_{ijpp'}\varphi^{(2)}_{pp'}(t,s) + \epsilon_{ij}(t|s), \tag{2.5}$$

such that $\varphi^{(1)}_{kk'}(t,s) = \phi^{(1)}_k(t|s)\psi^{(1)}_{kk'}(s)$ is a product of the first- and second-stage eigenfunctions, describing the variation conditional on longitudinal time s and along longitudinal time s, respectively; similarly $\varphi^{(2)}_{pp'}(t,s) = \psi^{(2)}_{pp'}(s)\phi^{(2)}_p(t|s)$. However, the principal surfaces $\varphi^{(1)}_{kk'}(t,s)$ and $\varphi^{(2)}_{pp'}(t,s)$ in (2.5) are not the eigenfunctions of the unconditional subject and subunit level covariance operators. In other words, the proposed MD-FPCA is distinct from a multilevel two-dimensional Karhunen-Loève decomposition with eigenscores θ_{ik} and ν_{ijp} and two-dimensional orthogonal eigenfunctions $\omega^{(1)}_k(t,s)$ and $\omega^{(2)}_p(t,s)$,

$$X_{ij}(t,s) = \mu(t,s) + \eta_j(t,s) + \sum_{k=1}^{\infty} \theta_{ik}\omega^{(1)}_k(t,s) + \sum_{p=1}^{\infty} \nu_{ijp}\omega^{(2)}_p(t,s) + \epsilon_{ij}(t,s).$$

For the multilevel two-dimensional Karhunen-Loève decomposition, the unconditional total covariance of $X_{ij}(t,s)$ minus measurement error, $\widetilde{\Sigma}_T(t,s,t',s') = \text{cov}\{X_{ij}(t,s), X_{ij}(t',s')\} - \sigma^2_s 1_{\{t=t',s=s'\}}$, would be decomposed into subject $\Sigma^{(1)}(t,s,t',s') = \text{cov}\{X_{ij}(t,s), X_{ij'}(t',s')\}$ and subunit level $\Sigma^{(2)}(t,s,t',s') = \widetilde{\Sigma}_T(t,s,t',s') - \Sigma^{(1)}(t,s,t',s')$ covariances. Then the covariation at both the subject and subunit levels would be expanded with two-dimensional functional principal component expansions, $\Sigma^{(1)}(t,s,t',s') = \sum_k \tau^{(1)}_k \ \omega^{(1)}_k(t,s)\omega^{(1)}_k(t',s')$, $\Sigma^{(2)}(t,s,t',s') = \sum_p \tau^{(2)}_p \omega^{(2)}_p(t,s)\omega^{(2)}_p(t',s')$ with eigenvalues $\tau^{(1)}_k$ and $\tau^{(2)}_p$ and eigenfunctions $\omega^{(1)}_k(t,s)$ and $\omega^{(2)}_p(t,s)$.

A major advantage of the proposed MD-FPCA is that while the multilevel two-dimensional Karhunen-Loève decomposition would require decomposition and hence smoothing of multiple four-dimensional covariance surfaces at the subject and subunit levels, the proposed MD-FPCA involves decompositions of only two-dimensional covariance surfaces due to the two-stage structure, leading to ease in implementation and savings in computational costs. Conditioning on longitudinal time s at the first-stage lowers the dimension of the covariance surface considered at both stages of the proposed algorithm. A second major advantage, as will be demonstrated in our application to the implicit learning study, is in interpretations of the decomposition components. The two-stage structure leads to additional decomposition components, such as the first- and second-stage eigenfunctions, which help with the interpretation of complex variation patterns in higher dimensions.

2.2.3 Extension to data from multiple scalp regions

In applications to the ERP data from the implicit learning paradigm, we consider four electrodes in the right frontal region of the scalp where maximal condition differentiation is detected. Motivated by the exchangeable correlation structure among electrodes within the same scalp region observed in the longitudinal functional ERP data, MD-FPCA models electrode repetitions by a random effect similar to an ANOVA approach. However MD-FPCA can be extended within the same ANOVA framework for analysis of data from multiple regions of interest on the scalp by an additional level of random effects at the scalp region level to account for differences between electrodes from different regions.

Denote by $X_{irj}(t|s)$ a multilevel square integrable random function observed across continuous functional time t, $t \in \mathcal{T}$, at longitudinal time s, $s \in \mathcal{S}$, for electrode j, $j = 1, \ldots, J$, within scalp region r, $r = 1, \ldots, R$, and subject i, $i = 1, \ldots, n$. Separating the total variation into variability at the subject, region and electrode levels at each longitudinal time s leads to

$$X_{irj}(t|s) = \mu(t,s) + \eta_r(t,s) + \alpha_{rj}(t,s) + Z_i(t|s) + W_{ir}(t|s) + U_{irj}(t|s) + \epsilon_{irj}(t|s), \qquad (2.6)$$

where $\mu(t,s)$, $\eta_r(t,s)$ and $\alpha_{rj}(t,s)$ are the overall mean function, region- and electrode-specific shifts, respectively; $Z_i(t|s)$, $W_{ir}(t|s)$ and $U_{irj}(t|s)$ are the random subject-, region- and electrode-specific deviations, respectively; and $\epsilon_{irj}(t|s)$ is measurement error with mean zero and variance σ_s^2. Denote the total variation of $X_{ij}(t|s)$ at a fixed longitudinal time s by $\Sigma_T(t,t'|s) = \text{cov}\{X_{irj}(t|s), X_{irj}(t'|s)\}$ and let $\widetilde{\Sigma}_T(t,t'|s) = \text{cov}\{X_{irj}(t|s), X_{irj}(t'|s)\} - \sigma_s^2 1_{\{t=t'\}}$. Decomposition (2.6) implies separation of the total variation $\widetilde{\Sigma}_T(t,t'|s)$ at each longitudinal time s into subject level $\Sigma^{(1)}(t,t'|s) = \text{cov}\{X_{irj}(t|s), X_{ir'j'}(t'|s)\} = \sum_k \lambda_k^{(1)}(s)\phi_k^{(1)}(t|s)$ $\phi_k^{(1)}(t'|s)$, region level $\Sigma^{(2)}(t,t'|s) = \text{cov}\{X_{irj}(t|s), X_{irj'}(t'|s)\} = \sum_p \lambda_p^{(2)}(s)\phi_p^{(2)}(t|s)\phi_p^{(2)}(t'|s)$ and electrode level $\Sigma^{(3)}(t,t'|s) = \widetilde{\Sigma}_T(t,t'|s) - \Sigma^{(1)}(t,t'|s) - \Sigma^{(2)}(t,t'|s) = \sum_v \lambda_v^{(3)}(s)\phi_v^{(3)}(t|s)$ $\phi_v^{(3)}(t'|s)$ variation. The second-stage Karhunen-Loève decompositions applied to the eigenscores $\xi_{ik}(s) = \sum_{k'=1}^{\infty} \xi'_{ikk'}\psi_{kk'}^{(1)}(s)$, $\zeta_{irp}(s) = \sum_{p'=1}^{\infty} \zeta'_{irpp'}\psi_{pp'}^{(2)}(s)$ and $\beta_{irjv}(s) = \sum_{v'=1}^{\infty} \beta'_{irjvv'}\psi_{vv'}^{(3)}(s)$ from the first-stage Karhunen-Loève decompositions, $Z_i(t|s) = \sum_k \xi_{ik}(s)\phi_k^{(1)}(t|s)$, $W_{ir}(t|s) =$

$\sum_p \zeta_{irp}(s)\ \phi_p^{(2)}(t|s)$ and $U_{irj}(t|s) = \sum_v \beta_{irjv}(s)\phi_v^{(3)}(t|s)$, yield the extended MD-FPCA decomposition

$$\begin{aligned} X_{ij}(t|s) &= \mu(t,s) + \eta_r(t,s) + \alpha_{rj}(t,s) + \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} \xi'_{ikk'}\psi_{kk'}^{(1)}(s)\phi_k^{(1)}(t|s) \\ &\quad + \sum_{p=1}^{\infty}\sum_{p'=1}^{\infty} \zeta'_{irpp'}\psi_{pp'}^{(2)}(s)\phi_p^{(2)}(t|s) + \sum_{v=1}^{\infty}\sum_{v'=1}^{\infty} \beta'_{irjvv'}\psi_{vv'}^{(3)}(s)\phi_v^{(3)}(t|s) + \epsilon_{ij}(t|s). \end{aligned}$$

Similar to MD-FPCA, $\xi'_{ikk'}$, $\zeta'_{irpp'}$ and $\beta'_{irjvv'}$ denote the second-stage eigenscores, $\lambda_k^{(1)}(s) = \text{var}\{\xi_{ik}(s)\}$, $\lambda_p^{(2)}(s) = \text{var}\{\zeta_{irp}(s)\}$ and $\lambda_v^{(3)}(s) = \text{var}\{\beta_{irjv}(s)\}$ are the first-stage eigenvalues, $\lambda_{kk'}^{(1)} = \text{var}(\xi'_{ikk'})$, $\lambda_{pp'}^{(2)} = \text{var}(\zeta'_{ijpp'})$, $\lambda_{vv'}^{(3)} = \text{var}(\beta'_{irjvv'})$ are the second-stage eigenvalues and $\psi_{kk'}^{(1)}(s)$, $\psi_{pp'}^{(2)}(s)$ and $\psi_{vv'}^{(3)}(s)$ are the second-stage eigenfunctions.

Note that the above outlined extension models correlations between electrodes within a scalp region and correlations between electrodes across different scalp regions as exchangeable. While in our experience, the first assumption is easier to verify for EEG data, the second assumption may be relaxed by modeling spatial correlations between electrodes across different scalp regions based on anatomical distances between the scalp regions. Distance-based correlations can be modeled by the addition of a spatial process to the expansion in (2.6) similar to the approach taken by [Staicu et al., 2010]. Authors model multilevel functional data where spatial correlations are modeled at the lowest level of the hierarchy via a random spatial process. For EEG applications, the random spatial process would be added at the region level rather than the lowest level of the hierarchy which is the electrode level. This extension requires further developments and is identified as a topic for future research. Finally note that both the mean and the random effects structures can include additional terms such as diagnostic group (e.g. ASD, TD) or condition (e.g. expected, unexpected) for incorporation of different experiment specific factors into MD-FPCA. For illustration of the methodology, we model condition difference trajectories directly in applications to the implicit learning paradigm in the next section, where the diagnostic groups are modeled separately since we expect the ASD and TD groups to be different in mean trends as well as covariation.

2.3 Application to the implicit learning study

2.3.1 Description of the data structure

In our motivating implicit learning study, EEG data are recorded on 37 ASD and 34 TD children for 120 trials per condition (expected and unexpected) for each subject at 128 electrodes. The EEG signals are sampled at 250Hz, producing 250 within-trial time points per waveform spanning 1000ms. The standard preprocessing steps of the data include artifact detection, bad channel replacement, referencing and baseline corrections. Next, the meta-preprocessing of Hasenstab et al. [2015] is applied to the data, following the preprocessing steps, to increase the SNR to a level at which P3 peak locations can be identified without collapsing the entire longitudinal dimension (via averaging the ERP curves over all trials). The meta-preprocessing step averages ERP functions separately for each subject, electrode and condition, in a moving window of 30 trials to identify the P3 peak location in the averaged ERPs (see Appendix 2B for more details). We capture the shape of the entire P3 peak from these averaged ERPs by examining a 140ms window around the P3 peak identified in the meta-preprocessing step (i.e. functional time domain around the P3 peak location is $t \in$[-70ms, 70ms] in 4ms increments). The length of the functional time domain of 140ms is determined by scientific practice and our own observation of the length of the entire P3 peak across trials. Note that the ERP curves in the described functional domain are already aligned across subjects, trials and electrodes, since we consider a symmetric window around each P3 peak. Since the interest lies in condition differentiation characterizing implicit learning, we focus on ERP difference functions obtained by subtracting the meta-preprocessed ERP corresponding to the unexpected condition from the expected condition. For illustration of the proposed methods, MD-FPCA is applied to meta-preprocessed difference ERPs from four electrodes in the right frontal region of the scalp, observed in trials $s \in [5, 60]$, where maximal condition differentiation is detected. Note that the longitudinal time domain starts at trial 5 to avoid boundary effects.

The proposed MD-FPCA algorithm is applied to the ASD and TD groups separately. Five subjects are removed as outliers prior to analysis. One of the removed subjects in the

ASD group did not have available data until trial 20 and the remaining four subjects (two in each diagnostic group) had ERP difference functions more than 2 standard deviations away from their respective group means for most of the observations across both functional and longitudinal time domains. In addition, a single electrode is omitted from two subjects in the TD group due to highly nonhomogeneous trends compared to the other electrodes. The bandwidths of the mean functions and covariance smooths are selected using GCV and visual inspection of the one- and two-dimensional smooths. The selected bandwidths for the two-dimensional smoothing of the overall and subunit mean functions and total and within covariances in the first-stage decompositions are (30ms, 30 trials), (30ms, 30 trials), (15ms, 15ms), (15ms, 15ms) in the ASD group and (30ms, 30 trials), (30ms, 30 trials), (5ms, 5ms), (5ms, 5ms) in the TD group. The selected bandwidths for the mean functions and covariances in the second-stage decompositions are (15 trials, 15 trials), (5 trials, 5 trials) in the ASD group and (10 trials, 10 trials), (15 trials, 15 trials) in the TD group.

2.3.2 Data analysis results

Overall mean surface estimates $\mu(t, s)$ of ERP difference trajectories for both diagnostic groups are given in Figure A2.5 (a-b). The ASD mean surface displays a trend of positive concave condition differentiation across trials that is uniform across ERP time. The mean surface peaks around trial 35 where there is a slight differential increase around the P3 peak location (indexed by functional time $t = 0$). In contrast to the ASD mean surface, the TD group exhibits a trend of negative differentiation across trials with much smaller magnitude, including a prominent dip of negative differentiation around trial 25. Since the mean surfaces represent condition differentiation, the opposing mean trends between diagnostic groups imply that children with ASD have higher EEG values in the expected condition while those in the TD group have higher values in the unexpected condition. This is consistent with our previous findings in Hasenstab et al. [2015] and Hasenstab et al. [2016]. Another difference between diagnostic groups is in the timing of maximal condition differentiation (trial 35 for ASD and trial 25 in TD). This implies that while both diagnostic groups differentiate between the conditions, implying implicit learning, the children in the TD group are learning

at higher speeds. Finally, while the entire P3 peak trajectory is increasing until trial 35 in the ASD group, there is a narrower window around the P3 peak that is minimized at the time of maximal condition differentiation in TD youth. The electrode specific means have similar patterns to the overall mean surfaces and are deferred to Figures A2.2 and A2.3.

Estimated leading subject level first-stage eigenfunctions $\phi_k^{(1)}(t|s)$ and second-stage eigenfunctions $\psi_{kk'}^{(1)}(s)$ are given in Figures 2.3 and 2.4 for the ASD and TD groups, respectively. Recall that while the eigenfunctions $\phi_k^{(1)}(t|s)$ display modes of variation in the functional dimension at a fixed longitudinal time s, the eigenfunctions $\psi_{kk'}^{(1)}(s)$ display modes of variation of the first-stage eigenscores in the longitudinal dimension. The products of these two quantities create subject level principal surfaces $\varphi_{kk'}^{(1)}(t,s)$ in (2.5) capturing the variation along both dimensions. Note that the model components $\phi_k^{(1)}(t|s)$ and $\psi_{kk'}^{(1)}(s)$ within themselves are quantities of interest and viewing them together provides an easily interpretable summary of the total variation conditional on and along longitudinal time.

In the ASD group, the uniform variation across ERP time in the leading component $\phi_1^{(1)}(t|s)$ (Figure 2.3 (a)), coupled with $\psi_{1k'}^{(1)}(s)$, $k' = 1, 2$ (Figure 2.3 (b)) displaying variation along trials, indicate that majority of the variation is in the longitudinal/trial dimension at intermediate and later trials, $k' = 1$ (solid, corresponding to 27% of total variation explained), and at the boundary trials, $k' = 2$ (dashed, corresponding to 18.4% of variation explained). The resulting product principal surfaces and surface intervals $\mu(t,s) \pm \sqrt{\lambda_{1k'}^{(1)}}\varphi_{1k'}^{(1)}(t,s)$ for $k' = 1$ and 2, leading to the same interpretations, are given in Figures A2.6-7, Figure A2.5 (c) and Figure A2.5 (a), respectively. The second component $\phi_2^{(1)}(t|s)$ (Figure 2.3 (c)) of the subject level variation conditional on longitudinal time, captures a uniformly concave mode of variation in ERP time maximized at the P3 peak location $t = 0$. Estimated $\psi_{21}^{(1)}(s)$ (Figure 2.3 (d)) (solid, 1.8%), capturing modes of variation in the trial direction, indicates that the variation around the P3 peak is maximized at trial 35, the trial of maximum positive condition differentiation in the overall mean surface in the ASD group. There is additional variation in the boundary and intermediate trials in component $\psi_{22}^{(1)}(s)$ (Figure 2.3 (d)) (dashed, 1%).

As in the flat contour of the ASD leading eigenfunction, the leading component $\phi_1^{(1)}(t|s)$

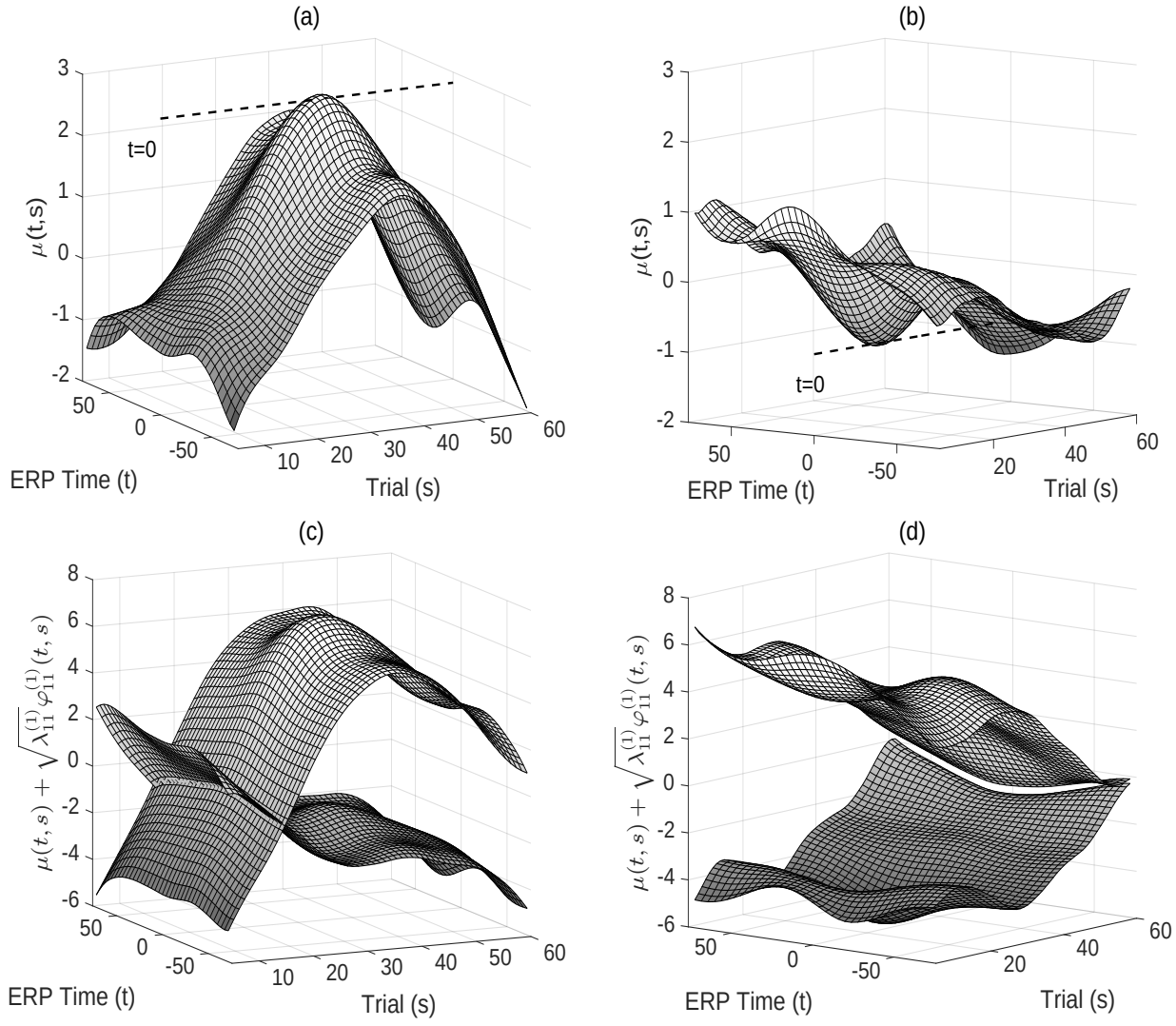


Figure 2.2: (a-b) Estimated mean surfaces, $\mu(t, s)$, for the ASD and TD groups, respectively. (c-d) Estimated surface intervals, $\mu(t, s) \pm \sqrt{\lambda_{11}^{(1)}}\varphi_{11}^{(1)}(t, s)$, for the ASD and TD groups, respectively.

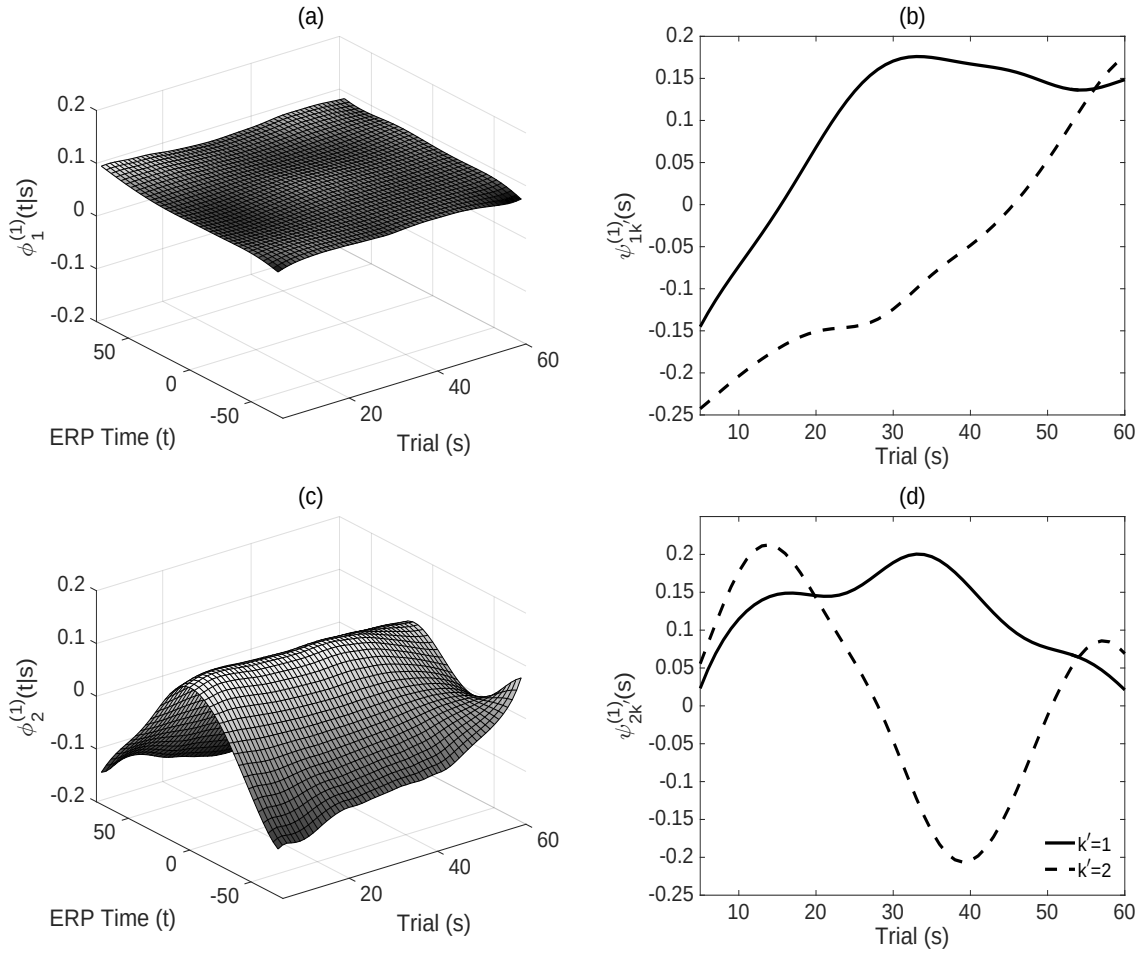


Figure 2.3: (a, c) Estimated leading subject level first-stage eigenfunctions, $\{\phi_k^{(1)}(t|s)\}$, $k = 1, 2$, respectively, for the ASD group. (b, d) Estimated leading subject level second-stage eigenfunctions, $\{\psi_{kk'}^{(1)}(s)\}$, $k, k' = 1, 2$.

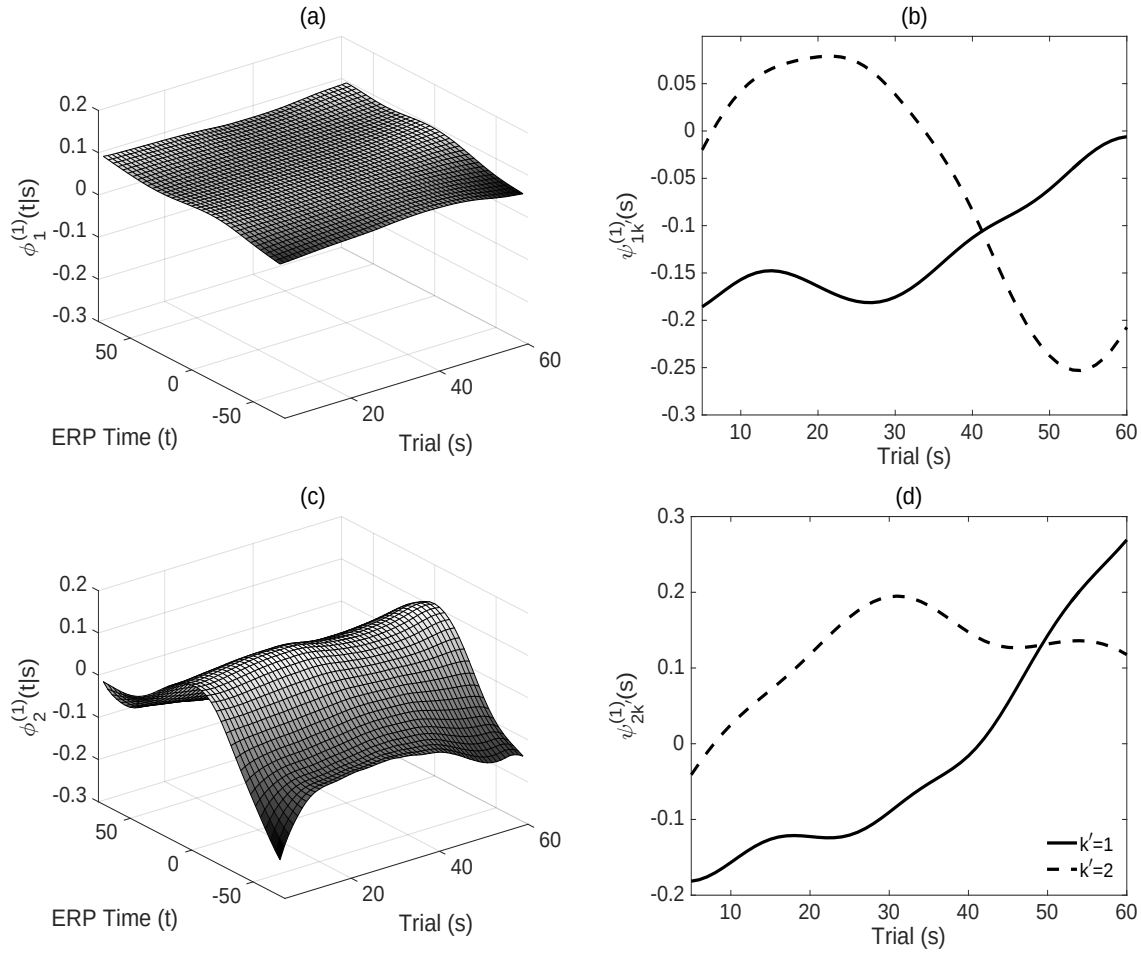


Figure 2.4: (a, c) Estimated leading subject level first-stage eigenfunctions, $\{\phi_k^{(1)}(t|s)\}$, $k = 1, 2$, respectively, for the TD group. (b, d) Estimated leading subject level second-stage eigenfunctions, $\{\psi_{kk'}^{(1)}(s)\}$, $k, k' = 1, 2$.

(Figure 2.4 (a)) for the TD group is also fairly flat with majority of the variation still in the longitudinal/trial dimension at the early and intermediate trials as captured by $\psi_{11}^{(1)}(s)$ (Figure 2.4 (b)) (solid, 36.9%) and later trials as captured by $\psi_{12}^{(1)}(s)$ (Figure 2.4 (b)) (dashed, 18.6%). (See Figure A2.5 (d) and Figure A2.5 (b) for surface intervals $\mu(t,s) \pm \sqrt{\lambda_{1k'}^{(1)}}\varphi_{1k'}^{(1)}(t,s)$, $k' = 1$ and 2, respectively.) The estimated $\phi_2^{(1)}(t|s)$ (Figure 2.4 (c)) captures leftover variation around the peak location and in the boundaries of the functional/ERP time domain, with variation in the trial direction maximized at boundary trials as reflected in $\psi_{21}^{(1)}(s)$ (Figure 2.4 (d)) (solid, 1.3%) and intermediate trials as reflected in $\psi_{22}^{(1)}(s)$ (Figure 2.4 (d)) (dashed, 0.9%). In summary, while the majority of the variation is in the longitudinal/trial dimension for both ASD and TD groups, most of the variation is observed at intermediate and later trials in the ASD group and at early and intermediate trials in the TD group. For interpretations on the electrode level variation and subject-specific eigenscores, see Appendix 2C.

The number of principal components at both stages of the MD-FPCA fit are selected to explain at least 90% of the variation. The breakdown of the total variation explained by each component of the subject and electrode level decompositions are given in Table 1 for the ASD and TD groups. While two components are selected in the first-stage decompositions uniformly across levels and diagnostic groups, three to four components are needed in the second-stage decompositions. The variability explained by the subject level variation in both stages of the MD-FPCA across longitudinal time (ρ) is estimated to be 62% and 72%, respectively, in the ASD and TD groups. This indicates that the longitudinal functional trajectories observed at the four electrodes in the right frontal region within a subject behave similarly, as expected, and the majority of the total variation is explained at the subject level for both diagnostic groups. Nevertheless, the similarity among electrodes seems to be larger in the TD group, although the difference between diagnostic groups is not found to be significant (90% percentile CI's for ρ for ASD and TD groups are $(0.53, 0.71)$ and $(0.67, 0.82)$, respectively, based on a bootstrap procedure on the meta-preprocessing and MD-FPCA using 200 data sets sampled with replacement from subjects). Figure A2.8 (a-b) display the estimated proportion of variability $\rho(s)$ explained at the subject level in the first-stage of MD-FPCA for the ASD and TD groups, respectively. The average estimated $\rho(s)$

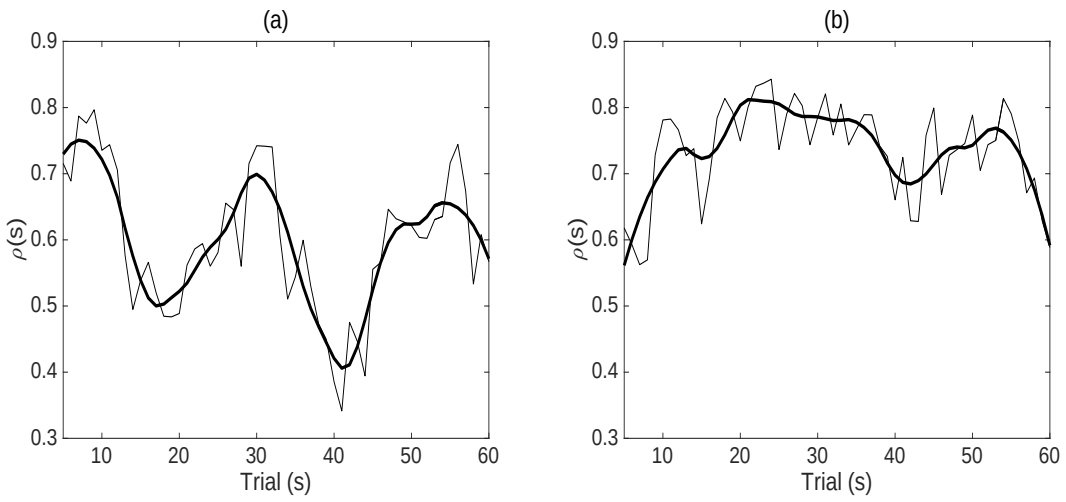


Figure 2.5: (a-b) Estimated proportion of variability explained at the subject level in the first-stage of MD-FPCA for the ASD and TD groups, respectively. The thin black line corresponds to the raw proportion of variability explained while the thick line corresponds to its smooth.

values again are higher in the TD group, where locations (along s) of maximum condition differentiation (trials 20 to 30) correspond to higher estimated $\rho(s)$ values in both groups. The similarity among electrodes within a subject seems to get stronger as the children start differentiating between the conditions, especially in the TD group.

Table 2.1: The number of principal components at both stages of the MD-FPCA fit selected to explain at least 90% of the variation.

Level 1					Level 2				
	ASD		TD			ASD		TD	
	$k=1$	$k=2$	$k=1$	$k=2$		$p=1$	$p=2$	$p=1$	$p=2$
$k'=1$	.270	.018	.369	.013	$p'=1$	.217	.006	.118	.007
$k'=2$	.184	.010	.186	.009	$p'=2$	.085	.004	.074	.005
$k'=3$	.133	.005	.141	.002	$p'=3$	.062	.002	.055	.003
$k'=4$	/	.004	/	/	$p'=4$	/	/	.014	.001

To conclude, we briefly highlight the additional insights gained by utilizing all the dimensions of the data in the analysis without collapsing across longitudinal, functional or electrode repetitions. Note that the analysis of the longitudinal functional condition differentiation trajectories averaged over the electrodes, collapsing the electrode dimension, can

be carried out by DFPCA of Chen and Müller [2012] which is a special case of MD-FPCA. Modeling the electrode dimension allowed us to study the electrode level variability, including comparisons to the variability at the subject level and the direction of the variation at the electrode level. More specifically, we learned that the majority of the variability was explained at the subject level in both groups (62% and 72% in ASD and TD, respectively). In addition, the proposed index $\rho(s)$ provided a more detailed depiction of the proportion of total variability at the subject level as a function of longitudinal time.

Similarly, the analysis of the data collapsed over either the functional or longitudinal dimensions can be carried out by ANOVA-FPCA of Di et al. [2009], another special case of MD-FPCA. Not collapsing the longitudinal dimension (enabled by the meta-preprocessing and MD-FPCA) revealed critical information in the application to the implicit learning study. We were able to characterize the entire learning process as well as its speed in addition to comparisons between groups. There is exploratory evidence that the TD group starts differentiating between the two conditions of the experiment earlier (trial 25) than the ASD group (trial 35). Modeling the P3 waveform instead of just the P3 peak amplitude (see our previous work [Hasenstab et al., 2015, 2016]) allowed us to compare the variation in the longitudinal and functional dimensions. Variations/changes over longitudinal time (trials) explain more of the total variation in the data than variation in the functional dimension. We also observed that longitudinal changes in the P3 waveform morphology were different between the two groups. While the entire P3 peak trajectory increased until trial 35 in the ASD group, the TD group showed condition differentiation in a much narrower functional time window around the P3 peak at the time of maximal condition differentiation.

2.4 Simulation

We study the finite sample properties of MD-FPCA through extensive simulations outlined in Appendix 2D. MD-FPCA recovers the true first- and second-stage model components for both small ($N = 30$) and moderate ($N = 100$) sample sizes, under varying SNRs (between 1 and 100) and sparsity levels in the longitudinal time domain, with up to 40% of data

missing at random longitudinal time points per subject. The median relative squared errors (RSEs) for all model components decrease with a denser design, increasing sample size and a higher SNR with the exception of the RSEs of the second-stage eigenfunctions which do not change with increasing SNR. This may be due to the fact that these quantities do not directly depend on data observed with measurement error.

2.5 Discussion

The proposed MD-FPCA has been presented under general settings without stringent assumptions on the separability of the longitudinal, functional and electrode covariances. Note that under the additional assumptions that modes of variability in the functional dimension stay the same across longitudinal times and electrode locations, or that modes of variability in the longitudinal dimension stay the same across functional times and electrode locations, more parsimonious versions of MD-FPCA can be derived using the marginal and product FPCA ideas of Park and Staicu [2015] and Chen et al. [2016]. These extensions would lead to a common set of eigenfunctions in functional time across longitudinal times and electrode locations and/or a common set of eigenfunctions in longitudinal time across functional time and electrode locations. Finally, while we focused on modeling the P3 peak curves in the current application, MD-FPCA can be extended to model the entire ERP waveform in the functional dimension. This extension would require warping of the ERP waveforms after meta-preprocessing according to data features (e.g. N1, P3) while simultaneously carrying out the multi-dimensional functional principal components decompositions.

Supporting materials

The proposed estimation algorithm, the meta-preprocessing step, additional data analysis interpretations, and the simulation studies are available in the appendices. MATLAB code, a sample simulated data set and a tutorial for implementing the MD-FPCA algorithm are also available on Github [https://github.com/aaron-scheffler/MD-FPCA].

CHAPTER 3

Hybrid Principal Components Analysis

Electroencephalography (EEG) data possess a complex structure that includes regional, functional, and longitudinal dimensions. Our motivating example is a word segmentation paradigm in which typically developing (TD) children and children with Autism Spectrum Disorder (ASD) were exposed to a continuous speech stream. For each subject, continuous EEG signals recorded at each electrode were divided into one-second segments and projected into the frequency domain via Fast Fourier Transform. Following a spectral principal components analysis, the resulting data consist of region-referenced principal power indexed regionally by scalp location, functionally across frequencies and longitudinally by one-second segments. Standard EEG power analyses often collapse information across the longitudinal and functional dimensions by averaging power across segments and concentrating on specific frequency bands. We propose a hybrid principal components analysis (HPCA) for region-referenced longitudinal functional EEG data which utilizes both vector and functional principal components analyses and does not collapse information along any of the three dimensions of the data. The proposed decomposition only assumes weak separability of the higher-dimensional covariance process and utilizes a product of one dimensional eigenvectors and eigenfunctions, obtained from the regional, functional, and longitudinal marginal covariances, to represent the observed data, providing a computationally feasible nonparametric approach. A mixed effects framework is proposed to estimate the model components coupled with a bootstrap test for group level inference, both geared towards sparse data applications. Analysis of the data from the word segmentation paradigm leads to valuable insights about group-region differences among the TD and verbal and minimally verbal children with ASD. Finite sample properties of the proposed estimation framework

and bootstrap inference procedure are further studied via extensive simulations.

3.1 Introduction

Approximately 30% of children with Autism Spectrum Disorder (ASD) never gain spoken language (referred to as 'minimally verbal') and the reasons are largely unknown [Tager-Flusberg and Kasari, 2013]. A major barrier in conducting research with minimally verbal children is the limited availability of appropriate assessment techniques. The recording of electroencephalography (EEG) signals during our motivating study, involving a word segmentation paradigm, gave researchers a unique opportunity to compare and contrast neurocognitive processes involved in language and communication development among verbal ASD (vASD), minimally verbal ASD (mvASD) and typically developing (TD) children, without relying on the children's ability to understand directions or provide an overt behavioral response. EEG is a popular non-invasive method for measuring voltage fluctuations across scalp regions in order to characterize neurocognitive processes and disorders. Children listened to a continuous speech stream which contained four 'made-up' words, each composed of three different phonemes or units of sound (Figures 3.1(a)-(b)). The four words were repeated 45 times in random order such that no word was used twice in a row and there was no time gap between words. The full experiment took 144 seconds. Children were expected to segment the speech stream, i.e. identify boundaries between words, by recognizing the differential patterns in the phonemes [Scott-Van Zeeland et al., 2010].

EEG studies, including both event-related and resting state paradigms, create high-dimensional data with regional, functional and longitudinal dimensions. Data from resting state paradigms are typically analyzed in the frequency domain, while event-related paradigms, where stimuli are applied repeatedly throughout the experiment, are analyzed either in the time or frequency domain. In our word segmentation paradigm, an event-related study, quantities considered of interest are in the frequency domain. Hence EEG signals, collected from an 128 electrode sensor net, were divided into one-second segments and projected into the frequency domain via Fast Fourier Transform (FFT). Given the fact

that EEG signals have low spatial resolution and that neighboring electrodes have similar power spectra, spectral principal components analysis (PCA) has been proposed to combine information from EEG signals recorded at electrodes within a scalp region [Ombao and Moon-Ho, 2006]. This pre-processing step produces region-referenced principal power, following a region-referenced longitudinal functional stochastic process. Specifically, the scalp locations represent the regional dimension, principal power obtained across frequencies represents the functional dimension and the one-second EEG segments represent the longitudinal dimension. Similarly, if the quantities of interest in an event-related paradigm are in the time domain, event-related potentials (ERP) time-locked to each stimulus (potentially combined over electrodes within a scalp region) would represent the functional dimension, and repetitions of the stimuli throughout the experiment would represent the longitudinal dimension. Note that all three dimensions of the observed data carry distinct interpretations and that longitudinal time (captured through segments across the experiment) may play an important role, especially in learning paradigms in which the focus is on changes over experimental time as learning evolves.

Standard analysis of high-dimensional EEG data involves collapsing information along multiple dimensions. The longitudinal dimension is collapsed when power spectra are averaged over segments or ERP curves are averaged over stimuli. Similarly, analysis of spectral power from specific frequency bands or specific ERP curve features corresponds to collapsing of the functional dimension, while averages over scalp regions collapse the regional dimension of the data We propose a hybrid principal components analysis (HPCA) for region-referenced longitudinal functional EEG data that does not collapse any of the three dimensions. We call the proposed decomposition hybrid, since it combines vector principal components analysis along the regional dimension (lacking a time order) and functional principal components analysis along the longitudinal and functional dimensions, providing an efficient nonparametric means of modeling high-dimensional EEG data. The HPCA decomposition involves a product of one-dimensional eigenvectors and eigenfunctions obtained from marginal covariances along the three dimensions of the data. A central assumption in this low dimensional, and hence computationally feasible, framework is the weak separability of the overall covari-

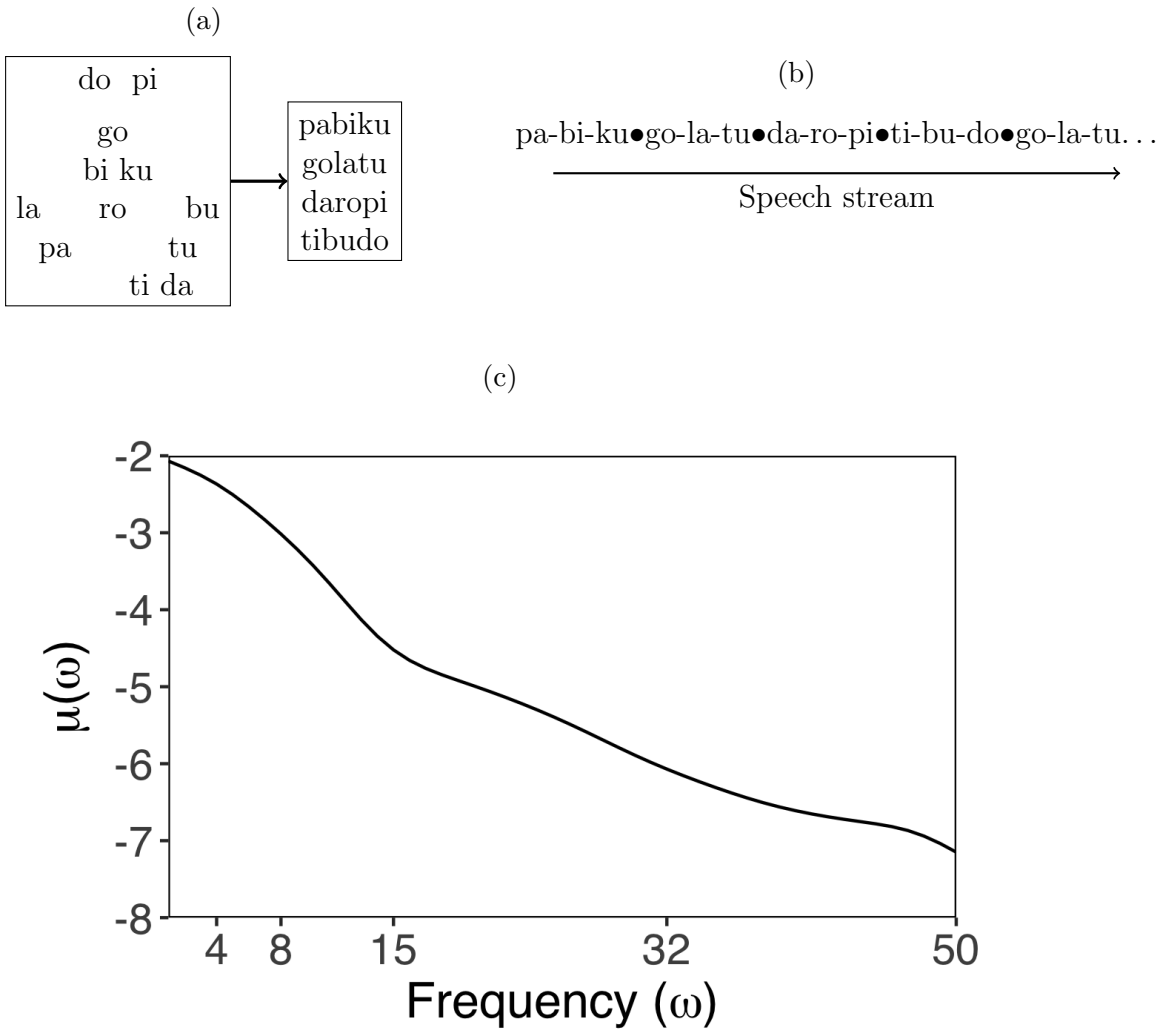


Figure 3.1: (a) Four 'made-up' words formed by concatenating three phonemes from a set of twelve phonemes without repetition in the word segmentation paradigm. (b) The artificial speech stream generated during the word segmentation paradigm. Breaks between phonemes are denoted by a dash and breaks between words are denoted by a dot. (c) The estimated mean log principal power $\mu(\omega)$ for subjects pooled across the TD, vASD, and mvASD groups.

ance process of the observed data. The concept of weak separability, recently proposed by Chen and Lynch [2017], refers to the idea that the covariance can be approximated by a weighted sum of separable covariance components and implies that the direction of variation (i.e. eigenvectors/eigenfunctions) along one of the three dimensions of the EEG data is the same across fixed slices of the other two dimensions. Note that this assumption is weaker than the commonly assumed strong separability of covariance surfaces in higher dimensions, which requires that the entire covariance structure, not only the directions of variation, is the same up to a constant across fixed slices of the other dimensions.

The literature on functional data analysis has proliferated over the past two decades, with methodological developments motivated by the complex dependency structures of repeatedly measured curves. Most of the recent developments on functional principal components analysis (FPCA) consider either longitudinally or spatially repeated functional data but not both. For longitudinally repeated functional data, Di et al. [2009] proposed multilevel ANOVA decompositions. Greven et al. [2010] extended their work to linear longitudinal decompositions, and Chen and Müller [2012], Park and Staicu [2015], Chen et al. [2016] and Hasenstab et al. [2017] considered more flexible nonlinear forms. For spatially repeated functional data, Staicu et al. [2010], Zhou et al. [2010] and Liu et al. [2017] considered parametric forms, while Huang et al. [2017] proposed a nonparametric decomposition. Of the proposed methods, only Hasenstab et al. [2017] decomposed both longitudinal and regional sources of functional variation in three dimensions via a multi-dimensional FPCA procedure (MD-FPCA). MD-FPCA, motivated by the analysis of the high-dimensional event-related ERP data in the time domain (through ERPs), treated scalp regions as exchangeable. The proposed HPCA method relaxes this assumption and involves a much simpler and computationally efficient decomposition via the weak separability of the covariance process. Product FPCA of Chen et al. [2016] also relies on weak separability and involves a product of one-dimensional eigenfunctions in the proposed decomposition; but their developments are obtained for two-dimensional functional data. HPCA extends product FPCA approach of Chen et al. [2016] to higher dimensions targeting region-referenced longitudinal functional EEG data and combining vector and functional principal components analysis. In addition,

while developments for product FPCA have only focused on densely measured longitudinally observed functional data, the estimation and inference procedures proposed for HPCA focus on sparse EEG data applications.

The outline of the paper is as follows. Section 3.2 introduces spectral PCA as a pre-processing step with minimal loss of information that produces region-referenced longitudinal functional data. Section 3.3 introduces the HPCA decomposition, develops an innovative mixed effects framework for estimation of the model components, specifically geared towards sparse data applications, and outlines a bootstrap procedure for group-level inference. We highlight that the developments for sparse data applications are novel. Prediction of subject-specific scores based on sparse data have not yet been considered for decompositions based on weak separability of the covariance process, such as the product FPCA. The proposed mixed effects framework is also utilized to assess the weak separability assumption via the random effects correlation structure. Section 3.4 provides insights from the word segmentation paradigm, including inference on group-region differences in spectral dynamics among TD, vASD, and mvASD children. We assess the proposed decomposition and the associated bootstrap test with an extensive simulation study summarized in Section 3.5 and conclude with a discussion in Section 3.6.

3.2 Spectral PCA and the resulting region-referenced longitudinal functional EEG data

Given that EEG signals measured on neighboring electrodes are highly multi-collinear due to their spatial proximity, the analysis of EEG data collected from high density electrode arrays is often preceded by reduction of the electrode dimension to discard redundant information and facilitate interpretation. When analysis takes place in the frequency domain, dimensional reduction is often unsatisfactorily carried out by selecting spectra from a single electrode or averaging spectra within a scalp region. Alternatively, given that electrodes within a scalp region possess similar spectra, spectral PCA has been proposed to pool spectral information within a scalp region with minimal loss of information. Spectral PCA applications in the

analysis of time series data date back to Brillinger [1981], but we follow a more recent application to EEG data by Ombao and Moon-Ho [2006]. They utilize spectral PCA as an exploratory tool to consolidate power spectra in a scalp region by utilizing overlapping segments of the continuous multi-channel time-series recorded at multiple electrodes in a seizure study. In contrast, we perform spectral PCA on non-overlapping EEG segments as a pre-processing step to be followed by scalp-wide analysis.

We highlight the outline of spectral PCA procedure here and defer details to Appendix A of the Supplementary Materials. Fourier coefficients at a fixed frequency are obtained via FFT for EEG signals measured from electrodes within the same scalp region and collected in a region-specific periodogram matrix. Following smoothing of each term of the periodogram matrices over frequencies, principal power is defined as the normalized leading eigenvalue of the smoothed periodogram matrix, representing the common variation in the fixed frequency across the electrodes (relative to variation in other frequencies) in a given scalp region along the direction of the leading eigenvector. The interpretation of principal power is closely tied to the goal of spectral PCA in combining signals across electrodes within a given scalp region. The assumption that electrodes within a scalp region have similar spectral densities implies that the region-specific periodogram matrix at a particular frequency would be of low rank. Hence, extracting the largest eigenvalue would serve as a reasonable summary of the spectral dynamics within a brain region. While our analysis focuses on the largest eigenvalue as principal power, note that second and third eigenvalues can also be modeled similarly via HPCA, allowing further analysis of the spectral dynamics among brain regions. Spectral PCA being applied at each segment and region for each subject, yields region-referenced longitudinal functional data, i.e. principal power as a function of region r, frequency ω and segment s, denoted by $Y_{di}(r, \omega, s)$. If a given subject does not have valid data at a fixed segment then the principal power for that segment is considered missing. We model $Y_{di}(r, \omega, s)$ as a summary measure of the power dynamics across the scalp.

3.3 Hybrid principal components analysis (HPCA)

3.3.1 The HPCA decomposition

Let $Y_{di}(r, \omega, s)$ denote the log principal power, which comprises region-referenced longitudinal functional data observed for subject i, $i = 1, \ldots, n_d$, from group d, $d = 1, \ldots, D$, in region r, $r = 1, \ldots, R$, at frequency ω, $\omega \in \Omega$ and segment s, $s \in \mathcal{S}$. Here Ω and $\mathcal{S}$ represent the functional and longitudinal domains, respectively, and $Y_{di}(r, \omega, s)$ is assumed to be square-integrable. Even though subjects may not be observed at all segments $s \in \mathcal{S}$, we use a common index set in the formulations below for notational ease. Note that the smoothing-based estimation procedure proposed in the next section, will readily extend to subject-specific sparse longitudinal domains. Further let $Z_{di}(r, \omega, s) = Y_{di}(r, \omega, s) - \mu(\omega, s) - \eta_d(r, \omega, s) - \epsilon_{di}(r, \omega, s)$ denote a de-meaned and de-noised region-referenced stochastic process, where $\mu(\omega, s)$ and $\eta_d(r, \omega, s)$ denote the functional fixed effects that represent the overall mean function and group-region shifts, respectively, and $\epsilon_{di}(r, \omega, s)$ denotes the measurement error with mean zero and variance σ_d^2.

The proposed HPCA decomposition provides a lower dimensional approximation of a stochastic process defined over regional, functional and longitudinal dimensions in terms of an empirical orthonormal basis based on eigenvectors and eigenfunctions obtained from the marginal covariances in each dimension. A central assumption of HPCA is the weak separability of the overall three-dimensional covariance process, which implies that the direction of variation (i.e. eigenvectors/eigenfunctions) along any one of the three dimensions of the EEG data is the same across fixed slices of the other two dimensions. This assumption is less stringent than the strong separability commonly assumed in the analysis of spatio-temporal stochastic processes, which requires that the entire covariance process along one dimension, not only the direction of variation, is the same up to a constant across fixed slices of the other dimensions. Note that the eigenfunctions or eigenvectors being the same does not necessarily imply the same covariance surface at fixed slices of the other dimensions due to weighting through the eigenvalues. We refer readers to Chen and Lynch [2017] for a detailed comparison of weak versus strong separability and note that we propose two separate checks

for the weak separability assumption in Section 3.3.2 and Appendix D of the Supplementary Materials, through a test for the correlation structure of the random effects in the mixed effects modeling and through visualization of the data, respectively.

Under weak separability, the common eigenfunctions and eigenvectors along each of the three dimensions can be estimated using the marginal covariances. Let the functional and longitudinal marginal covariance surfaces be defined as

$$\begin{aligned}\Sigma_{d,\Omega}(\omega,\omega') &= \sum_r \int_{\mathcal{S}} \mathrm{cov}\{Z_{di}(r,\omega,s), Z_{di}(r,\omega',s)\}ds = \sum_{\ell=1}^{\infty} \tau_{d\ell,\Omega}\phi_{d\ell}(\omega)\phi_{d\ell}(\omega'),\\ \Sigma_{d,\mathcal{S}}(s,s') &= \sum_r \int_{\Omega} \mathrm{cov}\{Z_{di}(r,\omega,s), Z_{di}(r,\omega,s')\}d\omega = \sum_{m=1}^{\infty} \tau_{dm,\mathcal{S}}\psi_{dm}(s)\psi_{dm}(s'),\end{aligned}$$

and let $\Sigma_{d,\mathcal{R}}$ denote the regional marginal covariance matrix with (r,r')-th element equal to

$$(\Sigma_{d,\mathcal{R}})_{r,r'} = \int_{\mathcal{S}}\int_{\Omega} \mathrm{cov}\{Z_{di}(r,\omega,s), Z_{di}(r',\omega,s)\}d\omega ds = \sum_{k=1}^{R} \tau_{dk,\mathcal{R}}\mathrm{v}_{dk}(r)\mathrm{v}_{dk}(r'),$$

where $\phi_{d\ell}(\omega)$ and $\psi_{dm}(s)$ are the common eigenfunctions of the functional and longitudinal marginal covariance surfaces, respectively; $\mathrm{v}_{dk}(r)$ are the common eigenvectors for the regional marginal covariance matrix; and $\tau_{d\ell,\Omega}$, $\tau_{dm,\mathcal{S}}$ and $\tau_{dk,\mathcal{R}}$ are the respective eigenvalues. While we estimate the regional marginal covariance matrix nonparametrically, we note that parametric approaches have been quite popular for modeling spatial covariances. An important difference of the current EEG application from typical environmental applications is that in the latter spatial data may typically be observed only once over the location grid at a fixed time point, while we observe the region-specific longitudinal functional EEG data repeatedly over subjects. Parametric assumptions to interpolate information across regions are thus not necessarily needed in modeling the spatial dependence in our application and we use a nonparametric region marginal covariance matrix, mimicking the nonparametric marginal functional and longitudinal covariance surfaces.

Utilizing the eigenfunctions and eigenvectors of the marginal covariances, the HPCA

decomposition of $Y_{di}(r,\omega,s)$ is given as

$$\begin{aligned} Y_{di}(r,\omega,s) &= \mu(\omega,s)+\eta_d(r,\omega,s)+Z_{di}(r,\omega,s)+\epsilon_{di}(r,\omega,s) \\ &= \mu(\omega,s)+\eta_d(r,\omega,s)+\sum_{k=1}^{R}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\xi_{di,k\ell m}\mathrm{v}_{dk}(r)\phi_{d\ell}(\omega)\psi_{dm}(s)+\epsilon_{di}(r,\omega,s) \end{aligned} \quad (3.1)$$

In (3.1), the subject-specific scores $\xi_{di,k\ell m}$ are defined through the projection, $\langle Z_{di}(r,\omega,s), \mathrm{v}_{dk}(r)\phi_{d\ell}(\omega)\psi_{dm}(s)\rangle = \sum_{r=1}^{R}\int\int Z_{di}(r,\omega,s)\mathrm{v}_{dk}(r)\phi_{d\ell}(\omega)\psi_{dm}(s)d\omega ds$, of the de-meaned and de-noised stochastic process, $Z_{di}(r,\omega,s)$, onto the orthonormal bases $\mathrm{v}_{dk}(r)\phi_{d\ell}(\omega)\psi_{dm}(s)$ defined as the product of the one-dimensional eigenfunctions and eigenvectors of the marginal covariances. Note that the set of subject-specific scores $(\xi_{di,k\ell m})$ are uncorrelated over regions, frequencies and segments under weak separability. Hence, the proposed HPCA expansion also leads to a decomposition of the total covariance, $\Sigma_{d,T}\{(r,\omega,s),(r',\omega',s')\}$, of $Y_{di}(r,\omega,s)$, as follows,

$$\begin{aligned} &\Sigma_{d,T}\{(r,\omega,s),(r',\omega',s')\} = \mathrm{cov}\{Z_{di}(r,\omega,s),Z_{di}(r',\omega',s')\}+\sigma_d^2\delta\{(r,\omega,s),(r',\omega',s')\} \\ = &\sum_{k=1}^{R}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\tau_{d,k\ell m}\mathrm{v}_{dk}(r)\phi_{d\ell}(\omega)\psi_{dm}(s)\mathrm{v}_{dk}(r')\phi_{d\ell}(\omega')\psi_{dm}(s')+\sigma_d^2\delta\{(r,\omega,s),(r',\omega',s')\}, \end{aligned}$$

where $\tau_{d,k\ell m} = \mathrm{var}(\xi_{di,k\ell m})$ and $\delta\{(r,\omega,s),(r',\omega',s')\}$ denotes the indicator for $\{(r,\omega,s) = (r',\omega',s')\}$. Note that the total covariance is written as a weighted sum of separable regional, functional and longitudinal covariances. One way of assessing the weak separability assumption will be to examine the correlation structure of the subject-specific decomposition scores $\xi_{di,k\ell m}$ via the mixed effects modeling framework proposed in Section 3.3.2.

In practice, the HPCA decomposition is truncated to include only K, L, and M eigencomponents for the regional, functional and longitudinal marginal covariances in the expansion, respectively, with truncation based on the fraction of variance explained (FVE). A general guideline is to initially include marginal eigencomponents in the HPCA expansion that explain approximately 90% of variation in their respective marginal covariances. Some of these components may be eliminated after subject-specific scores and their associated variance components are estimated via the proposed mixed effects modeling framework of Section

3.3.2, which provide an overall estimate of FVE, not only for the separate marginal covariances, but for the covariance based on the entire data. Details on the selection of the number of eigencomponents are deferred to Section 3.3.2.

Note that the three-dimensional (3-D) HPCA introduced in (3.1) reduces to a two-dimensional (2-D) HPCA with the regional and functional dimensions when the longitudinal dimension may not be of interest or may not exhibit change. Given that the remaining indices and arguments are unchanged, let $Y_{di}(r,\omega)$ denote the region-referenced functional data with a weakly separable covariance process. Utilizing the eigenfunctions and eigenvectors of the marginal covariances, the 2-D HPCA decomposition of $Y_{di}(r,\omega)$ can be given as,

$$\begin{aligned} Y_{di}(r,\omega) &= \mu(\omega) + \eta_d(r,\omega) + Z_{di}(r,\omega) + \epsilon_{di}(r,\omega) \\ &= \mu(\omega) + \eta_d(r,\omega) + \sum_{k=1}^{R}\sum_{\ell=1}^{\infty} \xi_{di,k\ell} \mathrm{v}_{dk}(r)\phi_{d\ell}(\omega) + \epsilon_{di}(r,\omega), \end{aligned}$$

where model components and the decomposition of the total variance are defined as in the 3-D HPCA by omitting the longitudinal argument s. The functional dimension can similarly be collapsed leading to the 2-D HPCA with only the regional and longitudinal dimensions. The discussion will continue to center on the 3-D HPCA with the understanding that extensions to 2-D HPCA are available by omitting one of the continuous arguments.

Motivated by the high-dimensional EEG data, both 3-D and 2-D HPCA extend the product FPCA of Chen et al. [2016] for longitudinally observed functional data by the addition of a regional dimension. Moreover, HPCA involves a hybrid decomposition for the region-referenced longitudinal functional EEG data, combining vector and functional principal components analysis under the assumption of weak separability. Another important divergence from the product FPCA formulation is in estimation. Motivated by the longitudinally sparse EEG data, we next propose a novel mixed effects procedure framework for estimation of the model components, specifically geared towards sparse data applications (with low number of repetitions and irregular spacing in observations over the longitudinal dimension). The estimation and testing procedures proposed for the product FPCA largely depend on projection techniques which are applicable only to densely measured longitudinal

functional data [Chen et al., 2016, Chen and Lynch, 2017].

3.3.2 Estimation of model components

The section below outlines the estimation of all the model components, including functional fixed effects, marginal covariances and eigencomponents, a novel mixed effects framework for estimation of subject-specific decomposition scores and associated variance components, and a recommendation to select the number of eigencomponents included in the proposed HPCA. We begin by introducing the HPCA estimation algorithm.

Algorithm: *HPCA Estimation Procedure*

---

1. Estimation of Fixed Effects
   (a) Calculate $\hat{\mu}(\omega, s) = \sum_{d=1}^{D} \hat{\mu}_d(\omega, s)$ by applying a bivariate penalized spline smoother to all observed data $\{\omega, s, Y_{di}(r, \omega, s) : \ i = 1, \ldots, n_d; \ r = 1, \ldots, R; \ \omega \in \Omega; \ s \in \mathcal{S}\}$.
   (b) Calculate $\hat{\eta}_d(r, \omega, s)$ by applying a bivariate penalized spline smoother to all observed data $\{\omega, s, \widehat{Y}_{di}(r, \omega, s) - \hat{\mu}(\omega, s) : \ i = 1, \ldots, n_d; \ \omega \in \Omega; \ s \in \mathcal{S}\}$.
2. Estimation of Marginal Covariances and Measurement Error Variance
   (a) Calculate $\widetilde{\Sigma}_{d,\Omega}(\omega, \omega')$ and $\widetilde{\Sigma}_{d,\mathcal{S}}(s, s')$ by applying bivariate penalized spline smoothers to the pooled covariances, $\widehat{\Sigma}_{d,\Omega}(\omega, \omega')$ and $\widehat{\Sigma}_{d,\mathcal{S}}(s, s')$, respectively.
   (b) Calculate $\hat{\sigma}_d^2$ by averaging the measurement error variance estimates $\hat{\sigma}_{d,\Omega}^2$ and $\hat{\sigma}_{d,\mathcal{S}}^2$.
   (c) Calculate $\widetilde{\Sigma}_{d,\mathcal{R}}$ by removing the estimated measurement error variance $\hat{\sigma}_d^2$ from the diagonal entires of the pooled covariance $\widehat{\Sigma}_{d,\mathcal{R}}$.
3. Estimation of Marginal Eigencomponents
   (a) Employ FPCA on $\widetilde{\Sigma}_{d,\Omega}(\omega, \omega')$ and $\widetilde{\Sigma}_{d,\mathcal{S}}(s, s')$ to estimate the eigenvalue, eigenfunction pairs, $\{\tau_{d\ell,\Omega}, \phi_{d\ell}(\omega) : \ell = 1, \ldots, L\}$ and $\{\tau_{dm,\mathcal{S}}, \psi_{dm}(s) : m = 1, \ldots, M\}$, respectively.
   (b) Employ PCA on $\widetilde{\Sigma}_{d,\mathcal{R}}$ to estimate the eigenvalue, eigenvector pairs $\{\tau_{dk,\mathcal{R}}, \mathrm{v}_{dk}(r) : k = 1, \ldots K\}$.
4. Estimation of Variance Components and Subject-Specific Scores via Linear Mixed Effects Models
   (a) Calculate $\hat{\kappa}_{dg}$ and $\hat{\sigma}_d^2$ by fitting the proposed linear mixed effects model.
   (b) Calculate $\hat{\zeta}_{dig}$ as the BLUP $\hat{\zeta}_{dig} = E(\zeta_{dig} | \boldsymbol{Y}_{di})$.
   (c) Select G' such that $FVE_{dG'} > 0.8$ for $d = 1, \ldots, D$ and form predictions $\widehat{Y}_{di}(r, \omega, s)$.

We defer details on steps 1-3 to Appendix B of the Supplementary Materials in which we refer readers to previous works on well-established mean, covariance, and eigencomponent estimation. However, we briefly highlight two novel estimation procedures found in step 2 for the measurement error variance, σ_d^2, and regional marginal covariance, $\Sigma_{d,\mathcal{R}}$. While previous authors obtain estimates of the measurement error variance using smoothing techniques on the raw covariance from a single dimension [Yao et al., 2005, Park and Staicu, 2015], we adapt this method to high dimensional settings by pooling information across both the functional and longitudinal marginal covariances. We then use this pooled estimate to remove the measurement error variance from the diagonals of the raw regional marginal covariance, which as a matrix is not amenable to smoothing techniques. Thus, we are able to leverage information from both the functional and longitudinal dimensions to obtain a decontaminated estimate of the regional marginal covariance.

In step 4, we utilize the estimated functional fixed effects and marginal eigencomponents to propose a linear mixed effects framework for modeling sparsely observed region-referenced longitudinal functional EEG data. In addition to allowing estimation of subject-specific scores under the assumption of their joint normality with the measurement error, the proposed mixed effects framework also provides final estimates for the corresponding variance components and the measurement error variance. The variance components estimates associated with the subject-specific scores are utilized in selection of the number of eigencomponents included in the HPCA decomposition via estimation of the proportion of variance explained, as well as in the construction of a hypothesis testing procedure for group-level inference via the bootstrap. Finally, the proposed mixed effects framework provides an opportunity to check the weak separability assumption via examining the correlation structure of the random effects.

For ease of notation, we replace the triple index $k\ell m$ in HPCA truncated at K, L, and M with a single index $g = (k-1) + K(\ell - 1) + KL(m-1) + 1$,

$$Y_{di}(r, \omega, s) = \mu(\omega, s) + \eta_d(r, \omega, s) + \sum_{g=1}^{G} \zeta_{dig}\varphi_{dg}(r, \omega, s) + \epsilon_{di}(r, \omega, s),$$

where $\varphi_{dg}(r,\omega,s) = \mathrm{v}_{dk}(r)\phi_{d\ell}(\omega)\psi_{dm}(s)$, $\zeta_{dig} = \langle Z_{di}(r,\omega,s), \varphi_{dg}(r,\omega,s)\rangle$, $\kappa_{dg} = \mathrm{cov}(\zeta_{dig})$ and $G = KLM$. Denote by $\boldsymbol{Y}_{di}$ the vectorized form of $Y_{di}(r,\omega,s)$ over the subject-specific region, frequency and segment grid for subject i, $i = 1,\ldots,n_d$. In our EEG application, while the region and frequency grids are the same for all subjects, the segment grid is subject-specific due to data quality issues. Similar subject-specific vectorized forms for the functional fixed effects, $\mu(\omega,s)$ and $\eta_d(r,\omega,s)$, the region-referenced stochastic process $Z_{di}(r,\omega,s)$, the measurement error $\epsilon_{di}(r,\omega,s)$, and the multidimensional orthonormal basis $\varphi_{dg}(r,\omega,s)$ are denoted by $\boldsymbol{\mu}_i$, $\boldsymbol{\eta}_{di}$, $\boldsymbol{Z}_{di}$, $\boldsymbol{\epsilon}_{di}$ and $\boldsymbol{\varphi}_{dig}$, respectively. Note that the mean vectors $\boldsymbol{\mu}_i$, $\boldsymbol{\eta}_{di}$ are indexed by subject since they are defined over the subject-specific region, frequency and segment grids. Under the assumption that ζ_{dig} and $\boldsymbol{\epsilon}_{di}$ are jointly Gaussian, the proposed linear mixed effects model is given as

$$\boldsymbol{Y}_{di} \;=\; \boldsymbol{\mu}_i + \boldsymbol{\eta}_{di} + \boldsymbol{Z}_{di} + \boldsymbol{\epsilon}_{di} = \boldsymbol{\mu}_i + \boldsymbol{\eta}_{di} + \sum_{g=1}^{G} \zeta_{dig}\boldsymbol{\varphi}_{dig} + \boldsymbol{\epsilon}_{di}, \quad \text{for} \quad i = 1,\ldots,n_d. \tag{3.2}$$

The model can be fit separately in each group, $d = 1,\ldots,D$, with both $\boldsymbol{\mu}_i$ and $\boldsymbol{\eta}_{di}$ previously obtained by smoothing. The functional, longitudinal and regional dependencies in $\boldsymbol{Y}_{di}$ are induced through the subject-specific random effects ζ_{dig} in (3.2). Given estimates for $\boldsymbol{\mu}_i$, $\boldsymbol{\eta}_{di}$ and $\boldsymbol{\varphi}_{dig}$, estimates of the variance components, κ_{dg} and σ_d^2 are obtained using maximum likelihood.

Following Yao et al. [2005] in using conditional expectations to estimate subject-specific scores for sparse functional data, the ζ_{dig} are estimated using best linear unbiased prediction (BLUP), $\hat{\zeta}_{dig} = E(\zeta_{dig}|\boldsymbol{Y}_{di}) = \hat{\kappa}_{dg}\hat{\boldsymbol{\varphi}}_{dig}\widehat{\boldsymbol{\Sigma}}^{-1}_{\boldsymbol{Y}_{di}}(\boldsymbol{Y}_{di} - \hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\eta}}_{di})$, where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}_{di}} = \sum_g \hat{\kappa}_{dg}\hat{\boldsymbol{\varphi}}_{dig}\hat{\boldsymbol{\varphi}}'_{dig} + \hat{\sigma}_d^2\boldsymbol{I}_i$ with $\boldsymbol{I}_i$ denoting the identity matrix of the same dimension as the length of the vectorized response $\boldsymbol{Y}_{di}$. Compared to the projection-based estimator of the subject-specific random effects in Chen et al. [2016] and Chen and Lynch [2017], which is only applicable for densely measured longitudinal functional two-dimensional process observed without measurement error, the proposed approach via mixed effects modeling is specifically geared towards sparse region-referenced longitudinal functional EEG data observed with measurement error. It also allows for assessing the weak separability assumption via a likelihood ratio test for

the independence of the random effects (for details see Appendix D of the Supplementary Materials).

The subject-specific scores and variance components estimated via the proposed mixed effects model are used to obtain predicted subject-specific trajectories and to choose the number of eigencomponents included in the HPCA decomposition. Using subject-specific scores estimated from the mixed effects model, subject-specific trajectories can be predicted via $\widehat{Y}_{di}(r,\omega,s) = \hat{\mu}(r,\omega,s)+\hat{\eta}_d(r,\omega,s)+\sum_{g=1}^{G'}\hat{\zeta}_{dig}\hat{\varphi}_{dig}(r,\omega,s)$, where G' denotes a set of eigencomponents such that the total fraction of variance explained ($FVE_{dG'}$) is greater than 0.8 in all groups $d = 1,\ldots,D$. We recommend starting with a larger number $G = KLM$ of eigencomponents in the mixed effects modeling used for the estimation of $(\kappa_{dg} : g = 1,\ldots,G)$. In order to estimate the group-specific fraction of total variance explained via the G eigencomponents, we consider the quantity, $FVE_{dG} = \{n_d\sum_{g=1}^{G}\hat{\kappa}_{dg}\}/[\sum_{i=1}^{n_d}\{||Y_{di}(r,\omega,s) - \hat{\mu}(\omega,s) - \hat{\eta}_d(r,\omega,s)|| - R|\Omega||\mathcal{S}|\hat{\sigma}_d^2\}]$, where $||f(r,\omega,s)||^2 = \sum_{r=1}^{R}\int\int f(r,\omega,s)^2 d\omega ds$. Note that the above formulation utilizes variance components estimates $\hat{\kappa}_{dg}$ and $\hat{\sigma}_d^2$ obtained from the proposed mixed effects model and considers the ratio of the variance in the G eigencomponents to the total variation in the observed data $Y_{di}(r,\omega,s)$ without measurement error. The denominator of FVE_{dG} does not use variation in a large number of eigencomponents to estimate the total variation in the observed data due to computational costs in fitting the proposed mixed effects model, but instead uses the three-dimensional norm of the de-meaned data, similar to the approach by Chen et al. [2016]. As a result, a limitation of FVE_{dG} is that when measurement error variance is overestimated and scaled by a factor of $R|\Omega||\mathcal{S}|$, FVE_{dG} may exceed 1.

3.3.3 Group-level inference via bootstrap

To test the null hypothesis that all groups have equal means in the scalp region r, i.e. $H_0 : \eta_d(r,\omega,s) = \eta(r,\omega,s)$ for $d = 1,\ldots D$, we propose a parametric bootstrap procedure based on the HPCA decomposition. The proposed parametric bootstrap generates outcomes based on the estimated model components under the null hypothesis as $Y_{di}^b(r,\omega,s) = \hat{\mu}(\omega,s)+$

$\hat{\eta}(r,\omega,s) + Z^b_{di}(r,\omega,s) + \epsilon^b_{di}(r,\omega,s) = \hat{\mu}(\omega,s) + \hat{\eta}(r,\omega,s) + \sum_{g=1}^{G'} \zeta^b_{dig}\hat{\varphi}_{dig}(r,\omega,s) + \epsilon^b_{di}(r,\omega,s)$ in region r and as $Y^b_{di}(r,\omega,s) = \hat{\mu}(\omega,s) + \hat{\eta}_d(r,\omega,s) + Z^b_{di}(r,\omega,s) + \epsilon^b_{di}(r,\omega,s) = \hat{\mu}(\omega,s) + \hat{\eta}_d(r,\omega,s) + \sum_{g=1}^{G'} \zeta^b_{dig}\hat{\varphi}_{dig}(r,\omega,s) + \epsilon^b_{di}(r,\omega,s)$ in the other regions, where subject-specific scores and measurement error are sampled from $\zeta^b_{dig} \sim \mathcal{N}(0,\hat{\kappa}_{dg})$ and $\epsilon^b_{di}(r,\omega,s) \sim \mathcal{N}(0,\hat{\sigma}^2_d)$. The proposed test statistic $T_r = [\sum_{d=1}^{D} \int\int \{\hat{\eta}_d(r,\omega,s) - \hat{\eta}(r,\omega,s)\}^2 d\omega ds]^{1/2}$ is based on the norm of the sum of the departures of the estimated group-region shifts $\hat{\eta}_d(r,\omega,s)$ from the estimate of the common shift across groups, $\eta(r,\omega,s)$. The common region shift estimate $\hat{\eta}(r,\omega,s)$, under the null, is set to the point-wise average of the group-region shift estimates, $\hat{\eta}_d(r,\omega,s),\ d = 1,\ldots,D$. We utilize the proposed parametric bootstrap to estimate the distribution of the test statistic T_r. The proposed procedure can be extended to test for equal means from specific frequency bands (i.e. subsets of Ω). We defer steps of the bootstrap algorithm to Appendix C of the Supplemental Materials.

3.4 Application to the word segmentation data

3.4.1 Data structure

In our motivating word segmentation study, EEG data were recorded for 144 seconds using an 128 electrode HydroCel Geodesic Sensor Net for 9 TD, 13 vASD, and 19 mvASD children ranging between 4 to 12 years of age. The EEG data is divided into non-overlapping segments of 1.024 seconds, producing a maximum of 140 observable segments for each subject at each electrode. Descriptions on the pre-processing steps and the final study sample are deferred to Appendix D of the Supplementary Materials. We consider 11 regions made up of 4-7 electrodes; left and right for the temporal region (LT and RT) and left, right, and middle for the frontal, central, and posterior regions (LF, RF, MF, LC, RC, MC, LP, RP and MP, respectively). We employ the spectral PCA described in Section 3.2 and Appendix A as a pre-processing procedure to reduce the spectra within each brain region to its corresponding log transformed principal power. The functional domain ranges from 0 to 50 Hz, to include the clinically defined frequency bands of delta ($0-4$ Hz), theta ($4-8$ Hz), alpha ($8-15$ Hz), beta ($15-32$ Hz), and gamma ($32-50$ Hz). Even though HPCA captures power dynamics across

the total frequency domain, we note that the gamma band was of particular interest in the word segmentation study since a higher gamma power is associated with better performance in cognitive processes.

We employ a 3-D HPCA to model log principal power as a function of region, frequency and segment. Based on the 3-D HPCA decomposition, we observe minimal variability in the segment dimension in both the functional fixed effects (Figures S1(b) and S5) and leading marginal eigenfunctions (Figure S2(c)), accounting for more than 85% of the marginal segment variation in each group (Table S1). Collectively, these two observations suggest that log principal power dynamics do not substantially change in the segment dimension both within subjects and among groups. Therefore, we collapse the segment dimension by averaging log principal power across segments within regions and employ a 2-D HPCA decomposition to model the resulting average log principal power $Y_{di}(r, \omega)$ as a function of region and frequency. Thus, we utilize the 3-D HPCA decomposition to justify the collapse of the segment dimension allowing for a more interpretable analysis based on the 2-D HPCA decomposition. Finally, we illustrate the benefit of modeling the unreduced frequency dimension by integrating the average log principal power $Y_{di}(r, \omega)$ over clinically defined frequency bands and comparing separate linear mixed models (LMMs) of the resulting region-referenced log principal power bands with the 2-D and 3-D HPCA. In the LMMs, group-region dynamics are captured through group-region interactions while within-subject region variation is modeled using a subject-specific random intercept. LMMs were fit using `nlme` [Pinheiro et al., 2017]. For the 2-D and 3-D HPCA decompositions, the smoothing parameters for the functional fixed effects and marginal covariances were selected by GCV/REML.

3.4.2 Data analysis results

We present full results from the 2-D HPCA decomposition but defer details from the 3-D HPCA decomposition, including detailed checks of the weak separability assumption on the 2-D and 3-D covariance processes, to Appendix D of the Supplementary Materials. Our main focus is inference on group-region differences but we will briefly discuss the estimated model

Table 3.1: FVE of the marginal covariances for the selected eigencomponents in each diagnostic group in the 2-D HPCA decomposition. The number of eigencomponents are chosen to explain at least 90% FVE.

TD		vASD		mvASD	
$\mathcal{R}$	Ω	$\mathcal{R}$	Ω	$\mathcal{R}$	Ω
0.652	0.698	0.706	0.653	0.583	0.656
0.113	0.159	0.112	0.249	0.133	0.231
0.084	0.091	0.083	-	0.113	0.048
0.058	-	-	-	0.062	-
-	-	-	-	0.042	-

components from the 2-D HPCA decomposition. Table A3.1 displays the eigencomponents for the regional and functional marginal covariances that explain at least 90% marginal FVE in all three diagnostic groups. The leading four, three, and five regional marginal eigenvector and three, two, and three functional marginal eigenfunctions are collectively found to explain 0.998, 1.000, and 0.999 of the total FVE (FVE_{dG}) in the TD, vASD, and mvASD groups, respectively.

In the functional dimension, the first leading marginal eigenfunction $\phi_{d1}(\omega)$ (Figure 3.2(a)) displays increasing variation with increasing frequency for all diagnostic groups, with the peak observed in the beta and gamma bands (15-50 Hz). The second leading marginal eigenfunction $\phi_{d2}(\omega)$ (Figure 3.2(b)) displays peak variation mostly in the beta band (15-32 Hz). The first two eigenfunctions together explain at least 85% of the variation in the functional marginal covariance in all three diagnostic groups. In the regional dimension, the weights of the first leading marginal eigenvector $\mathrm{v}_{d1}(r)$ (Figure 3.2(c)) are uniform across scalp locations in all the diagnostic groups, implying equal variation, while the weights of the second leading marginal eigenvector $\mathrm{v}_{d2}(r)$ (Figure 3.2(d)) are highest for the LT and RP regions, and MF and RF regions for the TD and vASD groups, respectively. In the mvASD group, the leading components signal a contrast between LF and MP regions. The first two regional marginal eigenvectors together explain at least 70% of the variation in the regional marginal covariance in all three diagnostic groups.

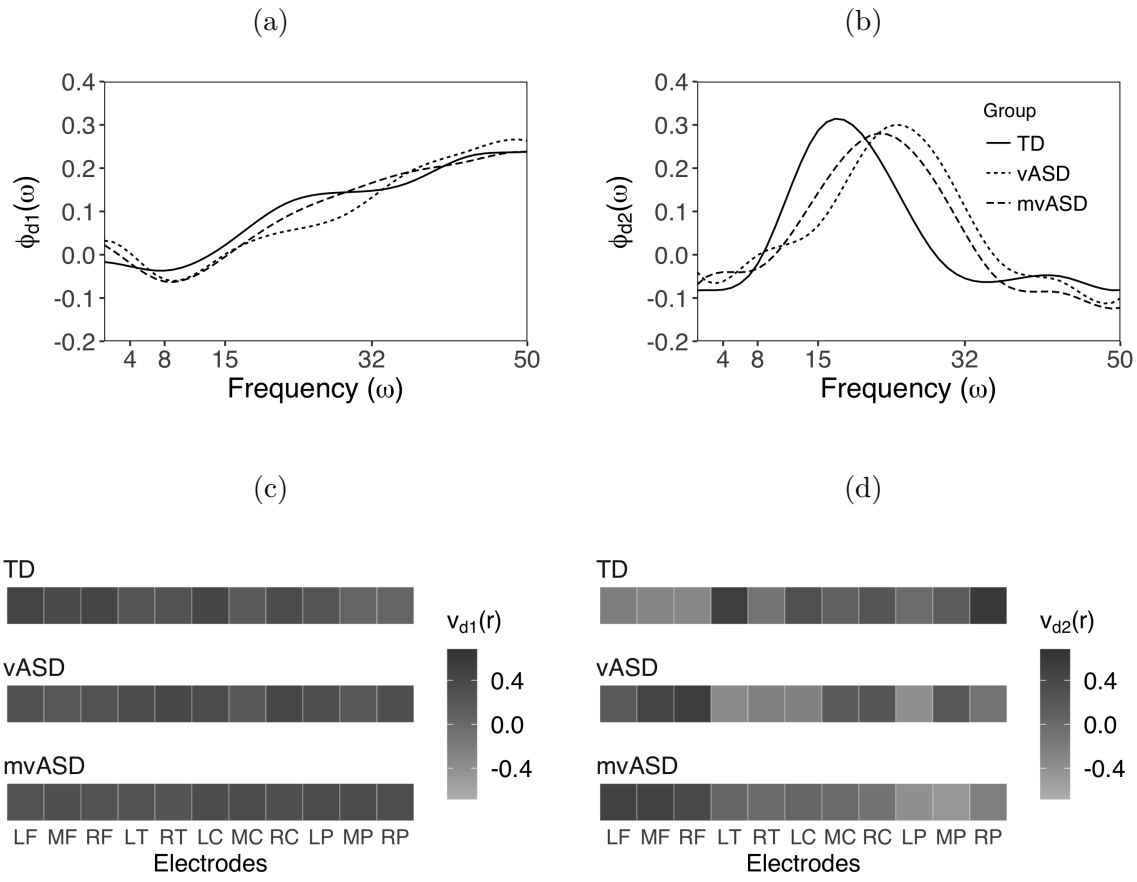


Figure 3.2: (a, b) Estimated first and second leading functional and longitudinal marginal eigenfunctions $\phi_{d1}(\omega)$ and $\phi_{d2}(s)$. (c, d) Estimated first and second leading regional marginal eigenvectors $\mathrm{v}_{d1}(r)$ and $\mathrm{v}_{d2}(r)$.

The estimated overall mean log principal power $\mu(\omega)$ curve, given in Figure 3.1(c), follows the well known trend of decreasing power with increasing frequency. In order to test for differences in the group-region means among the three diagnostic groups, we utilize the bootstrap test proposed in Section 3.3.3 originally for the 3-D HPCA decomposition, which can be extended to the 2-D HPCA decomposition via the test statistic $T_r = [\sum_{d=1}^{D} \int \{\hat{\eta}_d(r, \omega) - \hat{\eta}(r, \omega)\}^2 d\omega]^{1/2}$. For each scalp region r, we test the null hypothesis that the three diagnostic groups share a common mean which takes the form $H_0 : \eta_d(r, \omega, s) = \eta(r, \omega, s)$ and $H_0 : \eta_d(r, \omega) = \eta(r, \omega)$, $d = 1, 2, 3$, for 3-D and 2-D bootstrap procedures, respectively. The 2-D and 3-D bootstrap tests find significant differences among the group-region means for the three frontal regions: LR, RF, and MF (Figures 3.3 (a, c, e) and S5 (a-f), $p<.05$)

across the full frequency domain. The 3-D bootstrap procedure also identifies a significant difference among the group-region means for the total frequency domain in the LT, MC, and RP, although for the LT and RP regions this may be ascribed to edge effects in the segment dimension inflating the observed test statistic and for the MC region the 2-D bootstrap test is nearly significant (p=.05). While the 2-D and 3-D bootstrap tests provide insight into group-level dynamics for the full frequency domain, we also employ their band-specific extensions to enhance interpretation and enable comparisons with band-specific LMMs.

Table 3.2 displays the results of hypothesis tests for all scalp regions and frequency bands from the three separate models, the 2-D and 3-D HPCA decompositions and a set of band-specific LMMs. The greatest variation in group-region means from the 2-D HPCA decomposition are observed in LF and RF regions for the gamma band (Figure 3.3(a, c), p<.05), with the highest gamma principal power observed in the TD group, followed by the mvASD and vASD groups regions as evidenced by their relative difference from the group-region averages (Figure 3.3(b, d)). The mvASD group appears to have higher gamma principal power than the vASD group in the LF and LR regions, contradicting the expectation that the ordering of verbal impairment would be mirrored in group-region shifts in gamma activity which is thought to signal cognitive processes. One reason could be that the three diagnostic groups were not age-matched. The age distribution of vASD group had minimal overlap with those of the TD and mvASD groups and was over 20 months younger on average than the other diagnostic groups which may explain its lower gamma principal power. Further evidence of this age imbalance may be observed in LF and RF regions for the theta band in which the vASD group displays higher activity than the TD and mvASD groups across brain regions, consistent with the expected trend that theta activity is higher in younger children (Figures 3.3(a, c)). Finally, for each brain region the TD group followed by the vASD group have the highest alpha activity which is thought to be associated with relaxation, suggesting that mvASD children are not as relaxed as their verbally able peers.

The 2-D HPCA decomposition enhances the analysis of principal power by not only capturing the whole frequency domain but also by detecting significant differences among group-region means that are missed when the frequency domain is collapsed into specific

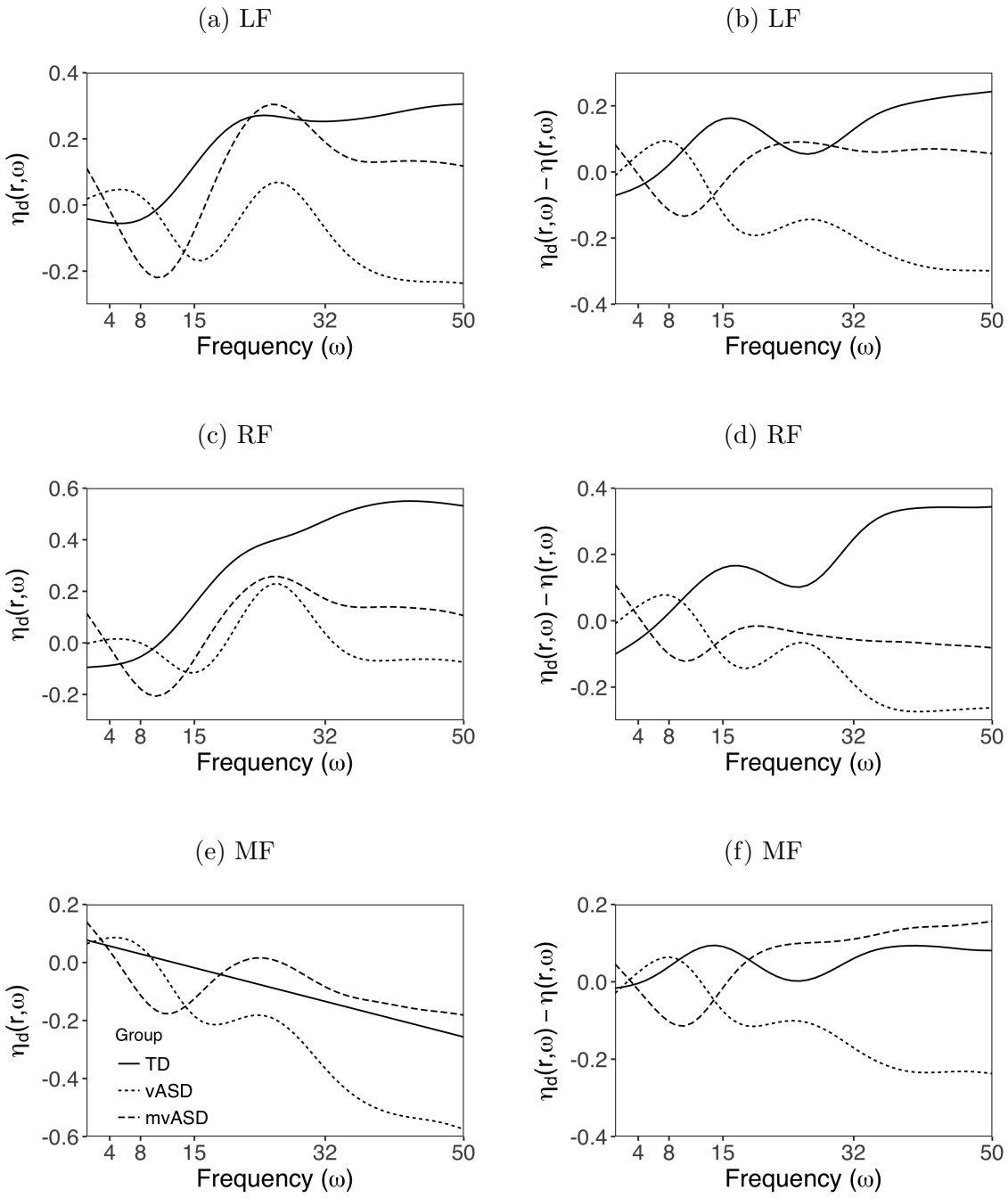


Figure 3.3: (a, c, e) The estimated group-region shifts $\eta_d(r, \omega)$ in the left, right and middle frontal regions in the TD, vASD, and mvASD groups. (b, d, f) The differences of the estimated group-region shifts $\eta_d(r, \omega)$ from group-region averages in the left, right and middle frontal regions in the TD, vASD, and mvASD groups. Note, the quantity $\eta_d(r, \omega) - \eta(r, \omega)$ forms the basis of the proposed bootstrap test statistic.

bands and modeled via band-specific LMMs. In the MF region for the gamma band, the TD and mvASD groups display higher principal power than the vASD group (Figure 3.3(e, f)) and the null hypothesis of a common group-region mean is rejected by the 2-D bootstrap test but not by the band-specific LMM. In addition, the 2-D bootstrap procedure finds significant differences in theta band dynamics among groups in all regions but the RC region while the LMM finds no significant differences among group-region means. By collapsing the frequency dimension prior to modeling, analysis methods such as the LMM cannot capture dynamics within frequency bands among groups (e.g. two signals crossing in a given interval) that may be modeled by maintaining the full frequency dimension.

3.5 Simulation

We studied the finite sample properties of the proposed HPCA and the bootstrap test for group-level inference via extensive simulations. While the results of the simulations are summarized here briefly, we defer details including data generation, discussion of the total and marginal FVEs, and further details on the bootstrap test to Appendix E of the Supplementary Materials. We conducted simulations for two sample sizes ($n_d = 15$ and 50), two signal-to-noise ratios (SNRs= 2.5 and 10) and two data sparsity levels (complete, partial in the longitudinal domain), for a total of eight settings. The lower sample size and sparsity levels were chosen to mimic the word segmentation data. To assess the performance of the proposed estimation algorithm in targeting the components of HPCA, we utilize normalized mean squared errors (MSE) and relative squared errors (RSE), based on the norms of the deviations of the estimate from the targeted quantities. In addition, we report the total and marginal fraction of variance explained along the regional, functional, and longitudinal dimensions, $FVE_{dK,\mathcal{R}}$, $FVE_{dL,\Omega}$, and $FVE_{dM,\mathcal{S}}$, based on the K, L, and M marginal eigencomponents included in the decomposition, respectively.

Figures A3.9 and S7 display the estimated model components based on 200 Monte Carlo runs from the sparse simulation set-up with $n_d = 15$ and high SNR. The estimated overall mean function and group-region shift with the median RSE values (Figures S7(b),(d)) track

the corresponding true surfaces (Figures 7(a),(c)). The estimated functional and longitudinal marginal eigenfunctions (Figure A3.9) are displayed from runs with RSE values at the 10th, 50th, and 90th percentiles, overlaid by their true quantities. Even with a small sample size, HPCA captures the periodicity and magnitude of the true components; patterns of estimated components from the dense case are similar and are deferred to Figures S8 and S9. Tables A3.2 and S2 display median, 10th, and 90th percentile RSE, normalized MSE values, and both total and marginal FVEs based on 200 Monte Carlo runs corresponding to the estimated HPCA components from all eight simulation settings. In general, the RSEs for all model components decrease with higher sample size and lower level of sparsity in the data. The fitted surfaces $Y_{di}(r, \omega, s)$ are the most susceptible to changes in SNR. The RSEs associated with the marginal eigencomponents are not sensitive to changes in SNR, suggesting that the estimation procedure successfully corrects for measurement error when obtaining the marginal covariances. For simulation set-ups with $n_d = 15$, the $90th$ percentile RSE for the marginal eigenvectors and eigenfunctions can exceed 1 but we note that 15 subjects is small for PCA and FPCA decompositions and that for $n_d = 50$ the comparable RSE values improve dramatically. The estimated level is on target and the power increases faster as one moves away from the null for the larger sample size, as expected (Figure S10).

3.6 Discussion

We proposed a hybrid principal components analysis technique (HPCA) which combines tools from vector and functional principal components analysis to decompose three-dimensional region-referenced longitudinal functional EEG data in a computationally efficient manner through the product of the one-dimensional eigenvectors and eigenfunctions of marginal covariances. Hence, the proposed estimation procedure scales up well to large datasets since estimation of the covariances and eigencomponents are performed within the marginal dimensions. To ease the computational burden in fitting the proposed mixed effects model for large data applications, the size of the grid chosen along each marginal dimension affecting the length of the design matrices can be controlled. Note also that the number of subjects

in most EEG studies are similar to those in our data application, hence HPCA would be applicable in most EEG paradigms.

The proposed estimation procedure centered around weak separability was developed to specifically handle realistic scenarios observed in EEG studies with potentially sparse data in the longitudinal dimension measured with noise. Note that similar ideas can be used to handle sparsity in the functional and regional dimensions as well. The HPCA decomposition paves the way for future work on regression analysis involving the high-dimensional EEG signals. A particular question of interest in autism research centers around relating behavioral outcomes to information within the EEG signals collected during an experiment. HPCA is a promising dimension reduction tool to enable regression modeling involving high-dimensional EEG signals.

Supporting materials

The reader is directed to the appendices for further details on the spectral PCA, HPCA estimation, bootstrap test, data analysis and simulation results. R programs and a tutorial for implementing HPCA can be found at Github [https://github.com/aaron-scheffler/HPCA].

Table 3.2: From left to right within each column grouping, results from the total spectral domain and band-specific hypothesis tests for all scalp regions for the 3-D bootstrap procedure, 2-D bootstrap procedure, and linear mixed models (LMMs), respectively. P-values less than .05 are displayed in bold.

	3-D HPCA / 2-D HPCA / LMM																	
Region	Delta (0-4 Hz)			Theta (4-8 Hz)			Alpha (8-15 Hz)			Beta (15-32 Hz)			Gamma (32-50 Hz)			Total (0-50 Hz)		
LT	**.005**	**.000**	**.001**	**.000**	**.000**	.174	**.000**	**.000**	**.000**	.080	.215	.413	**.025**	.145	.217	**.015**	.095	-
RT	**.030**	**.000**	**.001**	**.000**	**.015**	.442	**.005**	**.000**	**.001**	.210	.410	.777	.240	.465	.517	.125	.350	-
LF	.235	.075	.435	**.000**	**.000**	**.014**	.080	**.005**	**.026**	**.035**	.065	.073	**.000**	**.030**	**.038**	**.005**	**.020**	-
RF	.125	**.005**	.074	**.000**	**.005**	.062	.090	**.000**	**.034**	**.040**	.105	.179	**.000**	**.000**	**.013**	**.000**	**.000**	-
MF	.615	.550	.679	**.010**	**.000**	.198	.085	**.000**	**.033**	.170	.125	.218	**.030**	**.030**	.088	**.025**	**.035**	-
LC	**.025**	**.020**	**.005**	**.010**	**.000**	.163	**.020**	**.005**	**.002**	.130	.160	.387	.185	.275	.262	.110	.175	-
MC	.085	**.045**	.272	**.000**	**.000**	.232	**.000**	**.000**	**.029**	.075	.150	.532	**.040**	.105	.362	**.025**	.050	-
RC	**.000**	**.035**	**.001**	**.010**	.125	.624	**.000**	**.000**	**.003**	.260	.370	.748	.130	.345	.396	.100	.320	-
LP	**.045**	**.015**	**.004**	**.005**	**.045**	.450	**.000**	**.000**	**.002**	.120	.205	.542	.310	.795	.822	.110	.350	-
MP	**.025**	**.010**	**.040**	**.000**	**.015**	.253	**.000**	**.000**	**.015**	.115	.180	.490	.310	.545	.646	.130	.265	-
RP	**.035**	.130	**.029**	**.000**	**.010**	.162	**.000**	**.000**	**.002**	.135	.260	.644	.065	.220	.386	**.020**	.155	-

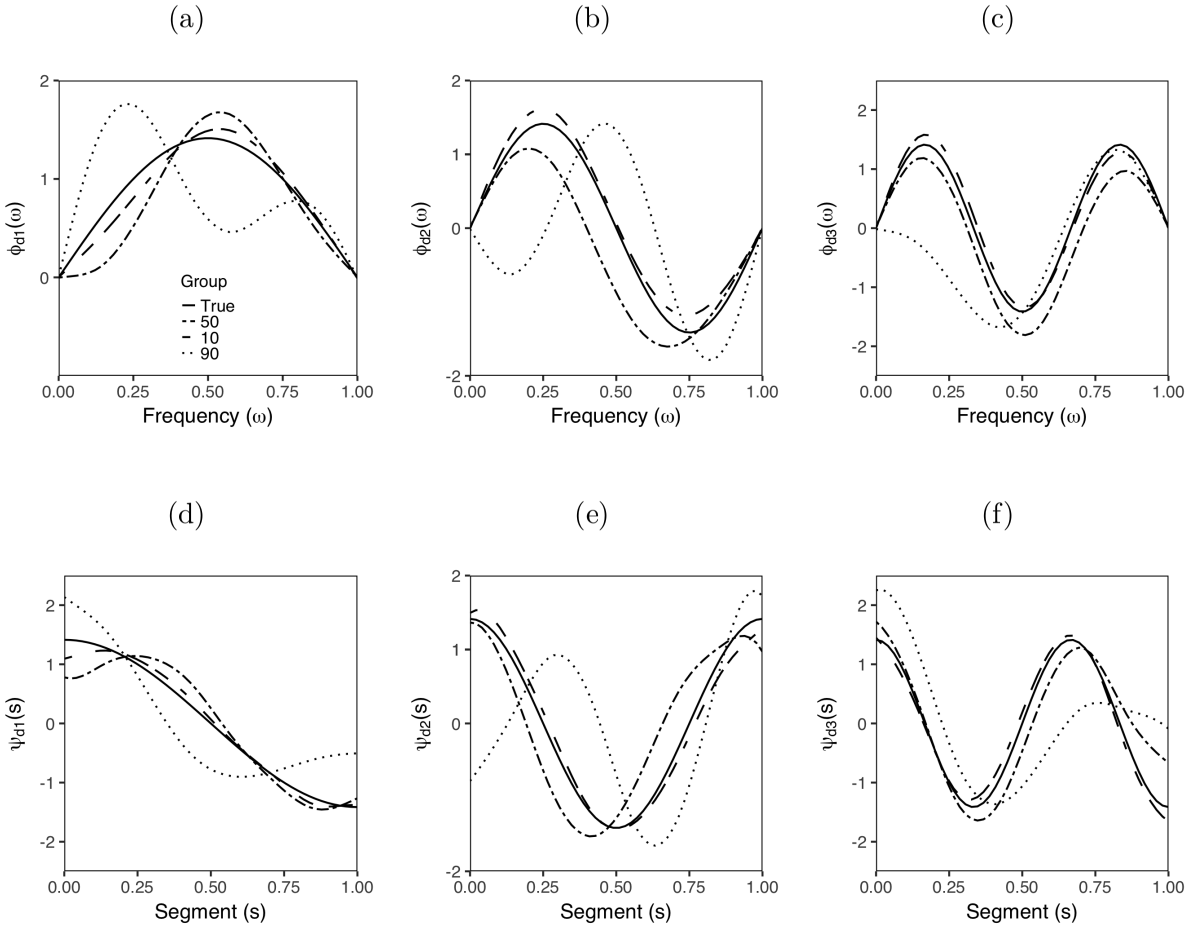


Figure 3.4: The true and estimated functional (first row) and longitudinal (second row) marginal eigenfunctions corresponding to the 10th, 50th, and 90th percentile RSE values across groups based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15$ and high SNR.

Table 3.3: Percentiles 50% (10%, 90%) of the relative squared errors (RSE), normalized mean squared errors (MSE) , and both total and marginal fraction of variance explained (FVE) across groups for model components based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15, 50$ for low and high SNR. Due to their small magnitude, MSE values are scaled by a factor of 10^3 for presentation.

	Low SNR				High SNR			
	$n_d = 15$		$n_d = 50$		$n_d = 15$		$n_d = 50$	
$\mu(\omega, s)$	0.000	(0.000, 0.000)	0.000	(0.000, 0.000)	0.000	(0.000, 0.000)	0.000	(0.000, 0.000)
$\eta_d(r, \omega, s)$	0.004	(0.003, 0.006)	0.001	(0.001, 0.002)	0.004	(0.003, 0.006)	0.001	(0.001, 0.002)
$Y_{di}(r, \omega, s)$	0.149	(0.072, 0.155)	0.077	(0.072, 0.155)	0.016	(0.015, 0.035)	0.016	(0.015, 0.035)
$\mathrm{v}_{d1}(r)$	0.085	(0.019, 0.422)	0.025	(0.005, 0.095)	0.075	(0.012, 0.348)	0.023	(0.003, 0.082)
$\mathrm{v}_{d2}(r)$	0.259	(0.063, 1.166)	0.076	(0.017, 0.350)	0.236	(0.039, 1.118)	0.057	(0.011, 0.250)
$\mathrm{v}_{d3}(r)$	0.198	(0.042, 1.025)	0.059	(0.012, 0.332)	0.159	(0.024, 0.954)	0.039	(0.005, 0.189)
$\phi_{d1}(\omega)$	0.080	(0.015, 0.376)	0.022	(0.003, 0.078)	0.074	(0.009, 0.351)	0.021	(0.004, 0.083)
$\phi_{d2}(\omega)$	0.267	(0.033, 1.109)	0.057	(0.007, 0.232)	0.231	(0.029, 0.966)	0.058	(0.008, 0.243)
$\phi_{d3}(\omega)$	0.143	(0.023, 1.073)	0.038	(0.006, 0.196)	0.137	(0.018, 0.857)	0.037	(0.004, 0.211)
$\psi_{d1}(s)$	0.092	(0.011, 0.414)	0.026	(0.004, 0.088)	0.078	(0.012, 0.330)	0.027	(0.004, 0.087)
$\psi_{d2}(s)$	0.223	(0.025, 1.130)	0.068	(0.011, 0.325)	0.212	(0.023, 1.044)	0.060	(0.009, 0.247)
$\psi_{d3}(s)$	0.143	(0.022, 0.871)	0.059	(0.012, 0.297)	0.141	(0.021, 0.899)	0.051	(0.008, 0.221)
$FVE_{dK,\mathcal{R}}$	0.985	(0.974, 0.993)	0.989	(0.983, 0.994)	0.995	(0.992, 0.998)	0.997	(0.995, 0.998)
$FVE_{dL,\Omega}$	0.974	(0.962, 0.986)	0.980	(0.972, 0.988)	0.990	(0.984, 0.994)	0.993	(0.990, 0.996)
$FVE_{dM,\mathcal{S}}$	0.938	(0.916, 0.957)	0.963	(0.948, 0.974)	0.986	(0.979, 0.992)	0.992	(0.988, 0.995)
$FVE_{dG'}$	1.009	(0.981, 1.038)	1.011	(0.994, 1.030)	1.007	(0.995, 1.019)	1.010	(1.003, 1.016)
$\tau_{d,klm}$	0.106	(0.004, 0.769)	0.031	(0.001, 0.232)	0.101	(0.003, 0.686)	0.032	(0.001, 0.226)
σ_d^2	0.025	(0.001, 0.131)	0.005	(0.000, 0.034)	0.001	(0.000, 0.005)	0.000	(0.000, 0.002)

CHAPTER 4

Covariate-Adjusted Region-Referenced Generalized Functional Linear Model

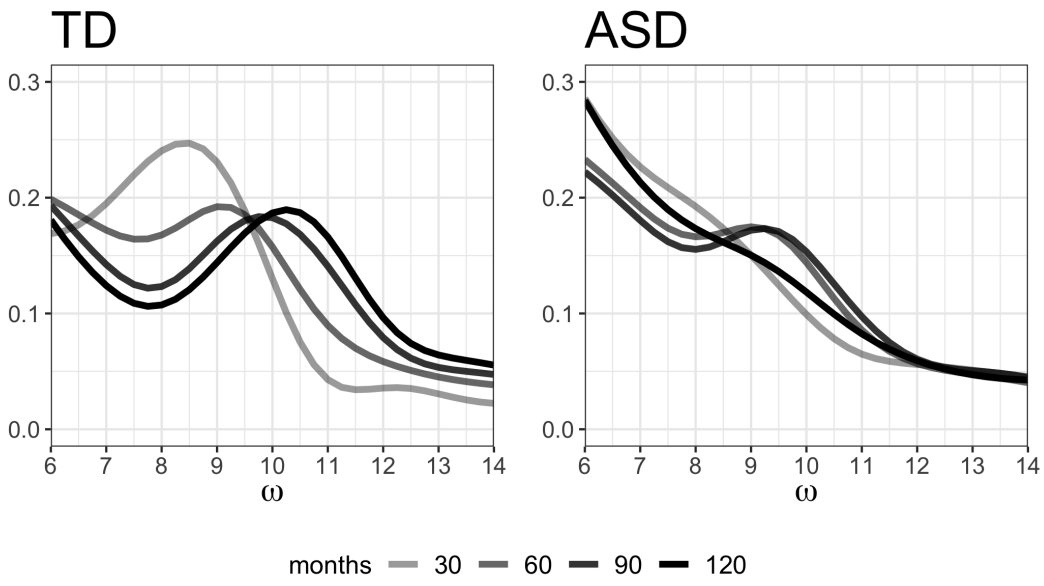
Electroencephalography (EEG) studies produce region-referenced functional data in the form of EEG signals recorded across electrodes on the scalp. It is of clinical interest to relate the highly structured EEG data to scalar outcomes such as diagnostic status. In our motivating study, resting state EEG is collected on both typically developing (TD) children and children with Autism Spectrum Disorder (ASD) aged two to twelve years old. The peak alpha frequency (PAF), defined as the location of a prominent peak in the alpha frequency band of the spectral density, is an important biomarker linked to neurodevelopment and is known to shift from lower to higher frequencies as children age. To retain the most amount of information from the data, we consider the oscillations in the spectral density within the alpha band, rather than just the peak location, as a functional predictor of diagnostic status (TD vs. ASD), adjusted for chronological age. A covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM) is proposed for modeling scalar outcomes from region-referenced functional predictors, which utilizes a tensor basis formed from one-dimensional discrete and continuous bases to estimate functional effects across a discrete regional domain while simultaneously adjusting for additional non-functional covariates, such as age. The proposed methodology provides novel insights into differences in neural development of TD and ASD children. The efficacy of the proposed methodology is investigated through extensive simulation studies.

4.1 Introduction

Children with Autism Spectrum Disorder (ASD) display a wide range of cognitive ability compared to their typically developing (TD) peers, yet the neural processes underlying this variability are not well understood [Dickinson et al., 2018]. In our motivating study, resting-state electroencephalograms (EEG) were recorded on both ASD and TD children aged two to twelve years old, allowing researchers to compare and contrast neural processes between the two diagnostic groups over a wide developmental range. Of particular interest was the location of a single prominent peak in the spectral density located within the alpha frequency band (6-14 Hz) called the peak alpha frequency (PAF). PAF has been shown to index neural development in TD children, where it shifts from lower to higher frequencies as children grow older [Miskovic et al., 2015, Valdas-Hernandez et al., 2010] .Recent research suggests that this chronological shift (from lower to higher frequencies) in PAF is delayed or possibly absent in children with ASD [Dickinson et al., 2018, Edgar et al., 2015] . This phenomena can be seen in our motivating data where slices of the group-specific bivariate mean surface of the spectral density (across age and frequency) at ages 30, 60, 90 and 120 months from the T8 electrode are plotted in Figure 4.1(a). The PAF, resembled by the location of the 'humps' in the spectral density, is more pronounced and displays a greater shift with age in the TD children compared to their peers diagnosed with ASD.

While the PAF location is well defined in sample averages, estimating a subject-electrode specific PAF presents many challenges, including the variability in estimation of the spectral densities and the potential for multiple local maxima [Corcoran et al., 2018]. In addition, identifying a single PAF inherently collapses information in the data across the alpha frequency band into a single number. To retain the most information from the data, we consider the spectral density across the alpha band as a functional observation and model associations between alpha band spectral dynamics and diagnostic status. In our motivating study, EEG signals are observed uninterrupted for several minutes across a high density electrode array and the continuous signal is divided into two-second segments before Fast Fourier Transform (FFT) to guarantee stationarity. The spectral density is then averaged across segments to

(a)



(b)

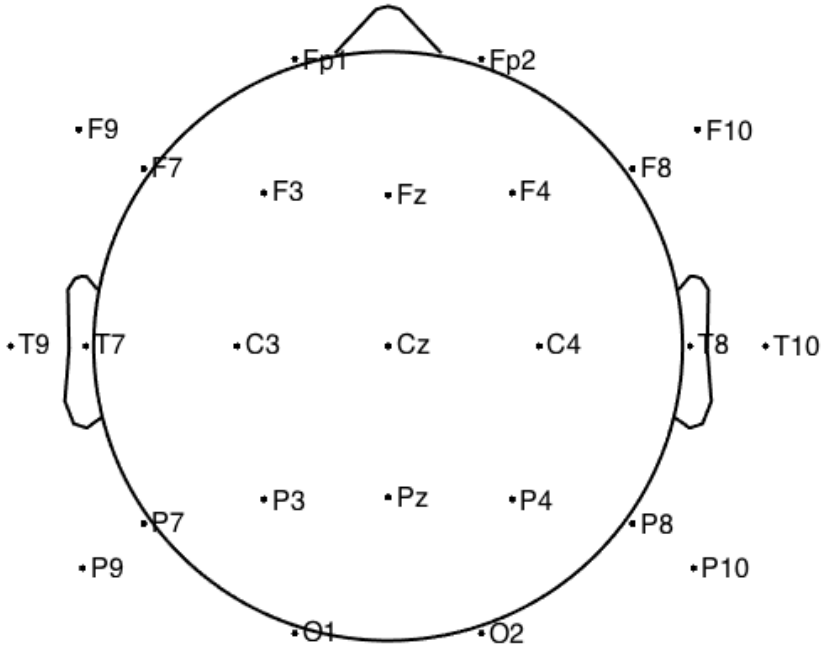


Figure 4.1: (a) Slices of the group-specific bivariate mean alpha band spectral density (across age and frequency (6-14 Hz)) at ages 30, 60, 90 and 120 months from the T8 electrode. Darker lines correspond to older children. (b) A schematic diagram of the 10-20 system 25 electrode montage.

increase the signal-to-noise ratio. The resulting spectral densities obtained across electrodes form the region-referenced functional data, with the spectral densities and the electrodes referred to as the functional and regional dimensions of the data. In order to model the association between diagnostic status and the high-dimensional EEG data, two methodological obstacles must be addressed. First, EEG signals recorded at each electrode result in a region-referenced functional predictor for which an appropriate functional regression model does not exist. Second, the relationship between the alpha band spectral dynamics and diagnostic status is expected to change with age and thus the potential regression model must allow for covariate-adjustments when estimating functional effects. To address both issues, we propose the covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM) that jointly estimates covariate-adjusted functional effects at each region by first projecting the regression function onto a tensor basis and then performing dimension reduction to produce a well-posed problem.

Since the introduction of the functional linear model (FLM) by Ramsay and Dalzell [1991], functional regression methods have been formalized into three categories based on the role of the functional data object: (1) scalar-on-function, (2) function-on-scalar, and (3) function-on-function regression models [Ramsay and Silverman, 2010, Morris, 2015]. Given that our goal is to relate a region-referenced functional predictor (region-referenced EEG spectral densities) to a scalar response (ASD diagnostic status), we restrict our discussion to relevant scalar-on-function regression methods (SoFR) particularly with respect to multivariate (i.e. multiple functional signals defined on possibly different domains) and multi-dimensional (i.e. two- or higher-dimensional functional signals defined continuously on a single domain) functional predictors. Hastie and Mallows [1993] were the first to formally define a FLM for a Gaussian response and Marx and EilersMarx and Eilers [1999] broadened this foundational model to include exponential family responses by proposing a generalized FLM (GFLM). Both models have been extended to accommodate multilevel functional predictors and adapted to non-parametric and non-linear frameworks [Reiss et al., 2017]. Considerable methodological development has focused on appropriate regularization strategies for settings in which multivariate or multi-dimensional functional predictors are

observed, where estimation is often performed via projection of the corresponding regression function(s) onto smooth basis functions with regularization imposed via the basis coefficients. While regularization for multivariate functional predictors is enforced within each distinct functional domain [Zhu et al., 2010, Gertheiss et al., 2013, Lian, 2013], regularization for multi-dimensional functional predictors is enforced by assuming continuity across the functional domain [Marx and Eilers, 2005, Reiss and Ogden, 2010, Goldsmith et al., 2014]. Specific to EEG and local field potentials, recent works by Gao et al. [2016] and Gao et al. [2018] develop methods for vector-valued electrical potentials recorded from multiple electrodes but these models focus on capturing longitudinal dynamics and cluster structures over the course of a recording session, respectively, rather than modeling associations with a scalar response.

Our proposed CARR-GFLM makes two important contributions to the existing literature. First, to our knowledge no SoFR method accommodates region-referenced functional predictors, i.e. correlated functional data observed over a non-smooth regional domain. To address this challenge, we consider a tensor basis that is a mixture of discrete and continuous basis functions. A corresponding penalty structure is developed to ensure smoothness of the regression function within each region along the functional dimension, with joint penalization across the regional domain. Second, we allow for the region-referenced regression function to vary across a continuous covariate, in our application, age. In the setting of a one-dimensional functional predictor, Wu et al. [2010] proposed a varying-regression functional linear regression model where regression effects not only vary across functional time but also across a scalar covariate. Authors estimate regression effects by targeting the functional covariance processes conditional on specific values of the scalar covariate via kernel smoothing methods. This estimation approach does not scale up well for higher dimensional functional data (e.g. region-referenced EEG spectral densities) given the reliance on computationally intensive kernel methods. Different from the approach in Wu et al. [2010], we add age as an argument to the tensor basis considered in estimation. The tensor basis is formed as a kronecker product of a marginal bases in the functional, regional and covariate domains leading to greater computational efficiency. The resulting number of tensor basis

functions may exceed the number of subjects in some applications; hence we further consider the singular value decomposition (SVD) of the design matrix as in Reiss and Ogden [2007] and Reiss and Ogden [2007] to ensure the problem is well-posed.

Note that existing SoFR methods do not provide an adequate covariate-adjusted modeling framework for region-referenced functional predictors. Given that region-referenced functional predictors are observed over a discrete regional domain, methods for multi-dimensional functional predictors which assume continuity of the regression function across each dimension cannot be used in the analysis of region-referenced functional predictors. Considering the functional signal from each region as multivariate functional data and applying existing multivariate GFLM (m-GFLM) techniques would require either a global regularization parameter across all dimensions or a separate regularization parameter for each region, both of which are undesirable due to the possibility of under fitting or over fitting the data, respectively. In addition to less than desirable regularization, existing multivariate methods do not allow for covariate adjustments when modeling the regression effects, as these adjustments have only been proposed in the literature for a single functional predictor byWu et al. [2010]. We show the favorable predictive performance of the proposed CARR-GFLM in comparison to the existing simpler approaches of m-GFLM, ignoring covariate effects, and a multivariate GFLM with a linear interaction term between the covariate and the functional predictor (m-GFLMi) in simulation studies and data applications.

The paper is organized as follows. Section 4.2 introduces the proposed model and develops estimation and inferential procedures. Section 4.3 discusses application of the proposed method to resting state EEG data from our motivating study, focusing on inference and interpretation of the estimated regression coefficients. Section 4.4 assesses performance of the proposed methodology via a simulation study. We conclude with a brief discussion in Section 4.5.

4.2 The proposed covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM)

4.2.1 Statistical framework and modeling

Suppose for $i = 1, \dots n$ subjects, we observe the data $\{y_i, X_i(a_i, r, \omega), a_i\}$, where y_i is a scalar response, $X_i(a_i, r, \omega)$ is a region-referenced functional predictor observed at region r, $r = 1, \dots, R$, frequency ω, $\omega \in \Omega$ and non-functional scalar covariate $a_i \in \mathcal{A} \subset \mathbb{R}$. While Ω and $\mathcal{A}$ are both continuous domains, they represent a functional and a non-functional covariate domain, respectively. The predictor $X_i(a_i, r, \omega)$ is assumed square-integrable and smooth over the functional domain Ω. Given our motivating data, we assume a_i is a scalar covariate though it could constitute a real valued vector of continuous covariates. For notational convenience, a regular grid for observations is assumed in the regional and functional dimensions, however, note that for sparse data applications in either the regional or functional dimension or both, the hybrid principal components analysis (HPCA) of Scheffler et al. [2018] can be used to reconstruct the full functional predictor. Throughout the remainder of the paper, scalars will be represented by lower case letters (b), vectors by lower case bold letters ($\boldsymbol{b}$), and matrices by upper case bold letters ($\boldsymbol{B}$).

First consider the region-referenced GFLM allowing for a region-referenced functional predictor,

$$y_i \sim \mathcal{F}(\mu_i, \boldsymbol{\vartheta})$$
$$g(\mu_i) = \sum_{r=1}^{R} \int \{X_i(r, \omega) - \eta(r, \omega)\}\beta(r, \omega)d\omega = \sum_{r=1}^{R} \int X_i^c(r, \omega)\beta(r, \omega)d\omega \qquad (4.1)$$

where $\mathcal{F}$ is an exponential family distribution with conditional expectation $\mu_i = E\{y_i | X_i(r, \omega), \boldsymbol{\vartheta}\}$, $\boldsymbol{\vartheta}$ denoting the vector of nuisance parameters and $g(\cdot)$ denoting an invertible link function. The region-referenced mean curve for all subjects is denoted by $\eta(r, \omega) = E\{X_i(r, \omega)\}$ and the mean centered subject-specific functional predictor which captures subject-level deviations from the region-referenced mean curve is denoted by $X_i^c(r, \omega) = X_i(r, \omega) - \eta(r, \omega)$. The

region-referenced regression function $\beta(r,\omega)$ models the linear association between $g(\mu_i)$ and $X_i^c(r,\omega)$, where $\beta(r,\omega)$ is not assumed to be smooth across the R regions. Note that the region-referenced GFLM in (4.1) is different from a multivariate GFLM with R separate functional predictors (possibly evaluated over different functional domains) in that the R functional predictors considered for (4.1) all represent spectral densities evaluated over the same domain, hence modeled in the next section with a single tensor basis and a combined smoothing parameter. Fixing $R = 1$ yields a standard GFLM for a functional predictor and scalar response as described in Marx and Eilers [1999].

Next, consider the proposed CARR-GFLM where the regression relationship changes as a function of a non-functional covariate. In our motivating data for example, the association between a subject's diagnostic status and alpha band spectral dynamics depends on chronological age. Thus, the proposed model for region-referenced functional predictors is given by

$$g(\mu_i) = \sum_{r=1}^{R} \int \big\{X_i(a_i,r,\omega) - \eta(a_i,r,\omega)\big\}\beta(a_i,r,\omega)d\omega = \sum_{r=1}^{R} \int X_i^c(a_i,r,\omega)\beta(a_i,r,\omega)d\omega \tag{4.2}$$

where $\mu_i = E\{y_i|X_i(a_i,r,\omega), a_i, \boldsymbol{\vartheta}\}$, $\eta(a,r,\omega) = E\{X_i(a_i,r,\omega)|a_i\}$, $X_i^c(a_i,r,\omega)$ and $\beta(a_i,r,\omega)$ denote the conditional expectation, the region-referenced mean surface, the mean-centered subject-specific functional predictor and the regression function that now all depend on the covariate a, respectively. The regression function $\beta(a,r,\omega)$ in a specific region r is assumed to be smooth in both the functional domain Ω and the covariate domain $\mathcal{A}$, allowing borrowing of information across the range of the covariate values observed in the sample in estimation. For a fixed a, the regression function $\beta(a,r,\omega)$ captures the different weights placed on the functional predictor across the frequency domain and how these relations change over the R regions. Changes over a add to this interpretation by depicting how this regression relation can vary over the different values of the covariate a. The proposed model reduces to the varying-coefficient functional linear model of Wu et al. [2010] for $R = 1$.

The regression function $\beta(a,r,\omega)$ is approximated by a linear combination of basis func-

tions that are formed as a tensor product of discrete and continuous marginal basis functions in a, r, and ω,

$$\beta(a, r, \omega) \approx \sum_{k_a=1}^{K_a} \sum_{k_r=1}^{K_r} \sum_{k_\omega=1}^{K_\omega} \phi_{k_a}(a)\phi_{k_r}(r)\phi_{k_\omega}(\omega)\theta_{k_a,k_r,k_\omega}$$

where the basis functions in a and ω, denoted by $\phi_{k_a}(a)$ and $\phi_{k_\omega}(\omega)$, respectively, can be chosen to be any set of continuous basis functions appropriately combined with a quadratic penalty, such as functional principal components or B-splines. The basis functions in r, denoted by $\phi_{k_r}(r)$, is a set of discrete basis functions such that $\text{span}\big[\{\phi_{k_r}(r)\}_{k_r=1}^{K_r}\big] \subseteq \mathbb{R}^R$ (e.g. columns of an identity matrix). The unknown coefficients of the projection, denoted by $\theta_{k_a,k_r,k_\omega}$ are collected into the vector $\boldsymbol{\theta}$ and are estimated. Note that the total number (K_a, K_r and K_ω) of basis functions considered in each dimension, is chosen to be sufficiently large to capture the regression function behavior before penalization. The shape of the resulting regression function can be controlled by both the choice of the marginal basis functions and the quadratic penalization of $\boldsymbol{\theta}$. We follow Wood [2006a] to construct a general penalty structure which is formed by a kronecker sum of marginal penalties along each dimension a, r and ω,

$$\text{pen}(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\lambda})\boldsymbol{\theta},$$
$$\boldsymbol{P}(\boldsymbol{\lambda}) = \lambda_a(\boldsymbol{P}_a \otimes \boldsymbol{I}_{K_r} \otimes \boldsymbol{I}_{K_\omega}) + \lambda_r(\boldsymbol{I}_{K_a} \otimes \boldsymbol{P}_r \otimes \boldsymbol{I}_{K_\omega}) + \lambda_\omega(\boldsymbol{I}_{K_a} \otimes \boldsymbol{I}_{K_r} \otimes \boldsymbol{P}_\omega),$$

where $\boldsymbol{\lambda} = (\lambda_a, \lambda_r, \lambda_\omega)$ denotes a vector of positive penalty parameters and $\boldsymbol{P}_a$, $\boldsymbol{P}_r$, and $\boldsymbol{P}_\omega$ denote the positive semi-definite penalty matrices, that control the degree of smoothness or shrinkage along each marginal dimension. For the dimensions along which the regression function is expected to be smooth (i.e. the functional and covariate dimensions), a differencing penalty can be used for a B-spline basis to penalize rapid change in the coefficients. In case of the regional dimension, the regression function is not assumed to be smooth and the choice of penalty structure requires more deliberation. In situations where there is no a priori knowledge of the dependency of the functional effects among regions, a ridge penalty

would promote smaller coefficients without imposing a prior dependency structure. If prior knowledge is available, one way to induce dependency across regions would be through a Gaussian Markov Random Field prior which have been applied in spatial analysis and generalized additive models [Besag, 1974, Wood, 2017]. The choice of the penalty structure plays an important role in inference through the posterior distribution of the coefficient vector $\boldsymbol{\theta}$ and should be considered carefully.

To our knowledge, this is the first application of a tensor basis formed from a mixture of discrete and continuous basis functions for estimating functional regression effects. Note that the proposed penalty structure with one penalty parameter for each marginal dimension, strikes a balance between under smoothing and increased computational burden with a larger number of penalty parameters and over smoothing with a smaller number of penalty parameters. For example, penalizing the regression function in each region separately would lead to R separate pairs of penalty parameters $(\lambda_{ra}, \lambda_{r\omega})$, $r = 1, \ldots, R$, increasing the number of penalty parameters and hence the computational burden significantly. Alternatively, λ_r, can be set to zero, effectively controlling smoothness across Ω and $\mathcal{A}$ at each region with just two parameters, possibly leading to over smoothing.

4.2.2 Estimation and inference

The proposed CARR-GFLM model is fit using the general additive model (GAM) framework of Wood [2017] for which there exists both stable optimization routines and theory for inference via confidence intervals. The region-referenced mean surface $\eta(a, r, \omega)$ is estimated prior to model fitting separately for each region based on pooled data across all subjects via smoothing achieved by projection onto a tensor basis of penalized marginal B-splines in a and ω. Estimation and smoothing parameter selection are carried out by restricted maximum likelihood (REML) methods. Let $(\omega_1, \ldots, \omega_H)$ denote the regular functional grid where the region-referenced functional predictor is observed. The proposed CARR-GFLM

in (4.2) can be written in matrix notation as,

$$\sum_{r=1}^{R}\int X_i^c(a_i,r,\omega)\beta(a_i,r,\omega)d\omega \approx \sum_{r=1}^{R}\int X_i^c(a_i,r,\omega)\sum_{k_a=1}^{K_a}\sum_{k_r=1}^{K_r}\sum_{k_\omega=1}^{K_\omega}\phi_{k_a}(a_i)\phi_{k_r}(r)\phi_{k_\omega}(\omega)\theta_{k_a,k_r,k_\omega}d\omega$$

$$\begin{aligned}&\approx \sum_{r=1}^{R}\sum_{h=1}^{H}\sum_{k_a=1}^{K_a}\sum_{k_r=1}^{K_r}\sum_{k_\omega=1}^{K_\omega} w_h X_i^c(a_i,r,\omega_h)\phi_{k_a}(a_i)\phi_{k_r}(r)\phi_{k_\omega}(\omega_h)\theta_{k_a,k_r,k_\omega}\\ &= \quad \big[\boldsymbol{\Phi}_a \otimes_r \big\{\boldsymbol{X}\ \boldsymbol{W}\ \big(\boldsymbol{\Phi}_r \otimes \boldsymbol{\Phi}_\omega\big)\big\}\big]\boldsymbol{\theta} \qquad (4.3)\end{aligned}$$

where $\otimes$ and $\otimes_r$ denote the standard and the row kronecker products, respectively. For two matrices with the same number of rows, the row kronecker product $\otimes_r$ forms a new matrix by taking the kronecker product of the rows of each matrix. In (4.3), $\boldsymbol{\Phi}_a = \big\{\phi_{k_a}(a_i)_{k_a=1,\ldots,K_a}^{i=1,\ldots,n}\big\}$ denotes an $n \times K_a$ matrix with ith row containing evaluations of the marginal basis functions at a_i; and $\boldsymbol{\Phi}_r = \big\{\phi_{k_r}(r)_{k_r=1,\ldots,K_r}^{r=1,\ldots,R}\big\}$ and $\boldsymbol{\Phi}_\omega = \big\{\phi_{k_\omega}(\omega_h)_{k_\omega=1,\ldots,K_\omega}^{h=1,\ldots,H}\big\}$ denote $R \times K_r$ and $H \times K_\omega$ matrices, respectively, whose columns contain evaluations of the marginal basis functions in r and ω. The predictor matrix $\boldsymbol{X}$ is a $n \times RH$ matrix containing the vectorized subject-specific functional predictor $X_i^c(a_i,r,\omega)$ in its ith row. Finally, $\boldsymbol{W} = \text{diag}\big\{(w_{rh})_{r=1,\ldots,R}^{h=1,\ldots,H}\big\}$ denotes the $RH \times RH$ diagonal matrix of weights that correspondingly sum across the R regions and is used to approximate the integral in ω. More information on defining appropriate marginal basis function can be found in Wood [2017].

The coefficient vector $\boldsymbol{\theta}$ is estimated by penalized least squares, with the penalized log-likelihood given by

$$\ell_p(\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\vartheta}|\boldsymbol{y}) = \ell(\boldsymbol{u},\boldsymbol{\vartheta}|\boldsymbol{y}) - \frac{1}{2}\text{pen}(\boldsymbol{\theta}|\boldsymbol{\lambda})$$

where $\boldsymbol{u} = g^{-1}\big(\big[\boldsymbol{\Phi}_a \otimes_r \big\{\boldsymbol{X}\boldsymbol{W}\ \big(\boldsymbol{\Phi}_r \otimes \boldsymbol{\Phi}_\omega\big)\big\}\big]\boldsymbol{\theta}\big)$ and $\ell(\boldsymbol{u},\boldsymbol{\vartheta}|\boldsymbol{y}) = \sum_{i=1}^{n}\ell(\mu_i,\boldsymbol{\vartheta}|y_i)$ denotes the log-likelihood function for the response distribution $\mathcal{F}(\mu_i,\boldsymbol{\vartheta})$. The penalized log-likelihood is maximized using REML rather than generalized cross-validation (GCV) due to the superior performance of REML reported in numerical studies. Reiss and Todd Ogden [2009] The

penalized likelihood can be maximized in a number of different ways. One popular approach is to treat $\boldsymbol{\lambda}$ as a precision parameter in a generalized linear mixed model [Wood, 2006a, Goldsmith et al., 2011]. However, this often produces covariance structures that are difficult to implement in standard software; a challenge that can be circumvented by suitable transformations of the marginal basis functions that divide the coefficient vector into sets of fixed effects and independent and identically distributed Gaussian random effects. This solution may still not be desirable since it introduces additional penalty parameters that may be hard to interpret [Wood et al., 2013]. We opt to maximize the penalized likelihood with the `gam()` function in the R package `mgcv` which finds an approximate REML criterion via Laplace approximation and optimizes the approximated likelihood using Newton-Raphson updates. The procedure iterates between estimating the $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ using standard penalized regression methods [Wood, 2011, 2017].

By projecting the regression function $\beta(a, r, \omega)$ onto a tensor basis, we perform an initial dimension reduction step. However, for SoF regression with multi-dimensional predictors, the number of basis functions may still greatly exceed the number of subjects, suggesting that the dimension of the basis is too large to be estimated well and further dimension reduction may be needed. Rather than restricting the number of basis functions in the tensor product, we perform a second dimension reduction step by only retaining the leading right singular vectors of the design matrix $\boldsymbol{D} = \boldsymbol{\Phi}_a \otimes_r \left\{ \boldsymbol{X}\boldsymbol{W} \left(\boldsymbol{\Phi}_r \otimes \boldsymbol{\Phi}_\omega \right) \right\}$ as in Reiss and Ogden [2007] and Reiss and Ogden [2010]. Therefore, we minimize the penalized log-likelihood based on the response mean function,

$$\boldsymbol{u} = g^{-1}\big(\boldsymbol{D}\boldsymbol{V}_q\tilde{\boldsymbol{\theta}}\big), \tag{4.4}$$

where $\boldsymbol{V}_q$ denotes the matrix containing the q leading columns from the singular value decomposition $\boldsymbol{U}\boldsymbol{E}\boldsymbol{V}^{\mathrm{T}}$ of the design matrix $\boldsymbol{D}$ and $\tilde{\boldsymbol{\theta}}$ denotes the coefficient vector for the reduced dimensional design matrix $\boldsymbol{D}\boldsymbol{V}_q$. In applications, we retain the minimum number of components q that explain 95% of the total variation in the design matrix $\boldsymbol{D}$, i.e. q is the minimum number of components that satisfy $\sum_{s=1}^{q} \kappa_s^2 / \sum_{s=1}^{K_a K_r K_\omega} \kappa_s^2 > .95$, where κ_s are the

ordered singular values from the SVD of $\boldsymbol{D}$. The penalty structure can easily be updated to accommodate the SVD of the design matrix, $\text{pen}(\tilde{\boldsymbol{\theta}}|\boldsymbol{\lambda}) = \tilde{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{V}_q^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\lambda})\boldsymbol{V}_q\tilde{\boldsymbol{\theta}}$. This dimension reduction serves two purposes, (1) it ensures that there is a unique solution that maximizes the penalized log-likelihood and (2) it allows for the use of the inferential machinery of `mgcv` which requires that the number of coefficients is less than or equal to the number of subjects. As emphasized earlier, this is a common issue with multi-dimensional predictors in SoF regression where the dimension of the scalar response may be much lower than the dimension of the functional predictors.

Due to the quadratic penalty, estimates of the coefficient vector $\tilde{\boldsymbol{\theta}}$ (associated with the the reduced dimensional design matrix $\boldsymbol{D}\boldsymbol{V}_q$) are biased and thus naive point-wise confidence intervals calculated based on the covariance matrix of the estimated coefficient vector can produce poor coverage. Therefore, we adopt the Bayesian point-wise confidence intervals described in Wood [2006b] which are based on the large sample limit of the posterior distribution $\tilde{\boldsymbol{\theta}}|\boldsymbol{y}$, which can be obtained by a default option in `gam()`. More specifically,

$$\tilde{\boldsymbol{\theta}}|\boldsymbol{y} \sim \mathcal{N}\left[\boldsymbol{\gamma}, \left\{\boldsymbol{V}_q^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{D}\boldsymbol{V}_{\boldsymbol{q}} + \boldsymbol{V}_q^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\lambda})\boldsymbol{V}_q\right\}^{-1}\right]$$

where $\boldsymbol{\gamma}$ denotes the estimate of $\tilde{\boldsymbol{\theta}}$, $\boldsymbol{Z}$ denotes an $n \times n$ matrix containing entries $Z_{ii} = \left\{g'(\mu_i)^2 V(\mu_i)\right\}^{-1}$ with μ_i equal to the ith entry of $\boldsymbol{u}$ in (4.4), V denotes the variance function such that $V(\mu_i)\sigma^2$ is equal to the variance of y_i and σ^2 is a scale parameter defined by $\mathcal{F}$. The covariance of $\tilde{\boldsymbol{\theta}}|\boldsymbol{y}$ can be adjusted by $\boldsymbol{V}_q$ in order to recover the posterior distribution of $\boldsymbol{\theta}|\boldsymbol{y}$. The Bayesian point-wise confidence intervals have been reported to lead to better coverage than those based on the covariance of the parameter estimates. We study the finite sample properties of the proposed methodology including coverage of the proposed Bayesian point-wise confidence intervals in the simulations of Section 4.4.

4.3 Data Analysis

4.3.1 Data structure and methods

In our motivating study, EEG data was sampled at 500 Hz for two minutes using a 128-channel HydroCel Geodesic Sensor Net on 58 ASD and 39 TD children 25 to 144 months old (groups were age matched). EEG recordings were made under an 'eyes-open' paradigm in which bubbles were presented on a screen in a sound-attenuated room [Dickinson et al., 2018]. Four electrodes near the eyes were removed prior to recording, in order to improve the comfort of the subjects. To facilitate independent component analysis for identification of artifacts, the data was interpolated to the international 10-20 system 25 electrode montage ($R = 25$) by spherical interpolation as described in Perrin et al. [1989] and implemented in the function `eeg_interp` from `EEGLAB` [Delorme and Makeig, 2004] (Figure 4.1 (b)). Following interpolation and ICA, the signal was reconstructed without components attributed to non-neural sources of signal, such as electromyogram (EMG) or non-stereotyped artifacts, and re-referenced to the average of all electrodes. For spectral analysis, the first 38 seconds of artifact free EEG data was used for each subject, where 38 seconds was the minimum amount of artifact free data available on all the subjects. Spectral density estimates were obtained by Welch's method [Welch, 1967]. The 38 second EEG signal was divided into two second Hanning windows with 50% overlap and transformed into the frequency domain via FFT. For each electrode, the spectral densities at each overlapping segment were averaged, resulting in electrode-specific estimates of the spectral density.

Since the interest is on the location of the dominant alpha peak and the general shape of the alpha band spectral dynamics more than alpha band power, the alpha ($\Omega = (6\text{ Hz}, 14\text{ Hz})$) spectral density normalized to a unit area (through division by its integral) is considered as the region-referenced functional predictor to facilitate comparisons across electrodes and subjects. As a result of the sampling scheme, the grid along the functional domain has a frequency resolution of .25 Hz and thus includes $H = 33$ points. Smooths of the region-referenced mean surface $\eta(a, r, \omega)$ are obtained as described in Section 4.2.2 using a tensor basis of penalized cubic B-splines (with 15 and 4 degrees of freedom in the frequency and

age domains, respectively) and second degree difference penalties along each dimension. In order to avoid bias in estimation of the region-referenced mean surfaces due to the observed imbalance in sample size between diagnostic groups, we re-weight the data such that the two diagnostic groups contribute equally to the region-referenced mean surface smoothing. The mean centered subject-specific functional predictors, $X_i^c(a_i, r, \omega) = X_i(a_i, r, \omega) - \eta(a_i, r, \omega)$, are obtained by subtracting age conditional slices of $\eta(a, \omega, r)$ from the observed subject-specific alpha spectral densities.

The regression function $\beta(a, r, \omega)$ is estimated by projection onto a tensor basis formed as the product of basis functions along the age, region/electrode and frequency dimensions. The marginal basis matrices $\boldsymbol{\Phi}_a$ and $\boldsymbol{\Phi}_\omega$ are formed as evaluations of cubic B-splines with $K_a = 5$ and $K_\omega = 10$ degrees of freedom, respectively. The regional basis matrix is equal to $\boldsymbol{\Phi}_r = \left[\mathbf{1}_R, \boldsymbol{I}_R\right]$ with $K_r = 26$, where $\mathbf{1}_R$ is an $R \times 1$ vector of 1's. Second order marginal difference penalties are utilized for both $\boldsymbol{P}_a$ and $\boldsymbol{P}_\omega$ to ensure smoothness over the functional and age domains. Given that we do not have any prior knowledge regarding the dependency of alpha spectral effects across electrodes, we employ a ridge style penalty across the regional dimension, $\boldsymbol{P}_r = \left[\mathbf{0}_R, \boldsymbol{I}_R\right]$, where $\mathbf{0}_R$ is an $R \times 1$ vector of 0's. The zero entry along the diagonal of $\boldsymbol{P}_r$ corresponds to the basis vector $\mathbf{1}_R$ which is left unpenalized to absorb the common effect across all electrodes. The remaining regional basis vectors are penalized and loadings on them represent the electrode-specific deviations from the overall effect across the scalp. The number of columns in the resulting design matrix $\boldsymbol{D}$ is $K_a K_r K_\omega = 1300$.

We carry out comparisons between the proposed CARR-GFLM and the existing approaches of m-GFLM and m-GFLMi. For the m-GFLM, the region-referenced alpha spectral densities are treated as multivariate functional data where the functional effect at each region is estimated by projection onto a basis of cubic B-splines with ten degrees of freedom. The m-GFLMi includes a main effect of age as well as a linear interaction term between age and the region-referenced alpha spectral densities for which an additional set of functional effects are estimated. For each functional effect, regularization is enforced using a separate smoothing parameter with a second degree difference penalty. Similar to CARR-GFLM, the number of basis functions for m-GFLM and m-GFLMi is too large to be estimated well

and the portion of the design matrix encoding functional effects is reduced by SVD with appropriate adjustments made to the penalty structure. The number of columns (i.e. rank) of the SVD reduced design matrices for the CARR-GFLM, m-GFLM, and m-GFLMi models is 66, 57, and 50, respectively, accounting for approximately 99% of the total variation. The threshold for the proportion of variation explained is slightly higher in the data analysis than suggested in Section 4.2.2 in order to ensure model convergence. Models are fit using the `gam()` function from `mgcv` (version 1.8-24) on a 2.4 GHz 6-Core Intel Xeon processor operating R (version 3.5.1) with a mean computation time of 26.6 seconds based on ten runs. Penalty parameters are selected via REML and estimated to be $\lambda_a = 0.0063$, $\lambda_r = 0.0042$, and $\lambda_\omega = 0.0078$.

4.3.2 Data analysis results

Slices of the region-referenced mean surface $\eta(a, r, \omega)$ representing the electrode-specific mean alpha spectral density at ages 30, 60, 90, and 120 months for the T8 and T10 electrodes (right temporal) are given in Figure 4.2. Across subjects and electrodes, the PAF increases with increasing chronological age. Since functional predictors are supposed to retain group differences to predict diagnostic status, region-referenced mean surfaces are estimated across subjects (rather than within diagnostic groups) and subtracted from the observed alpha spectral densities to obtain the mean-centered functional predictors $X_i^c(a_i, r, \omega)$ used in modeling. It is expected that, on average, subjects within each diagnostic group will deviate from the region-referenced mean surfaces in a distinct manner, allowing for characterization of patterns in the alpha spectral density that are predictive of ASD diagnosis.

The results from fitting the CARR-GFLM model to our motivating data for the T8 electrode are shown in Figure 4.3 and are representative in shape and sign of other electrodes across the scalp. In order to visualize information across chronological age, results are shown as cross sections at 30, 60, 90, and 120 months. The T8 electrode is highlighted because among all electrodes it produces the highest average contribution to the log-odds of ASD diagnosis, $1/n \sum_{i=1}^{n} \int_\Omega |X_i^c(a_i, r, \omega)\hat{\beta}(a_i, r, \omega)| d\omega$, which can be interpreted as a measure of

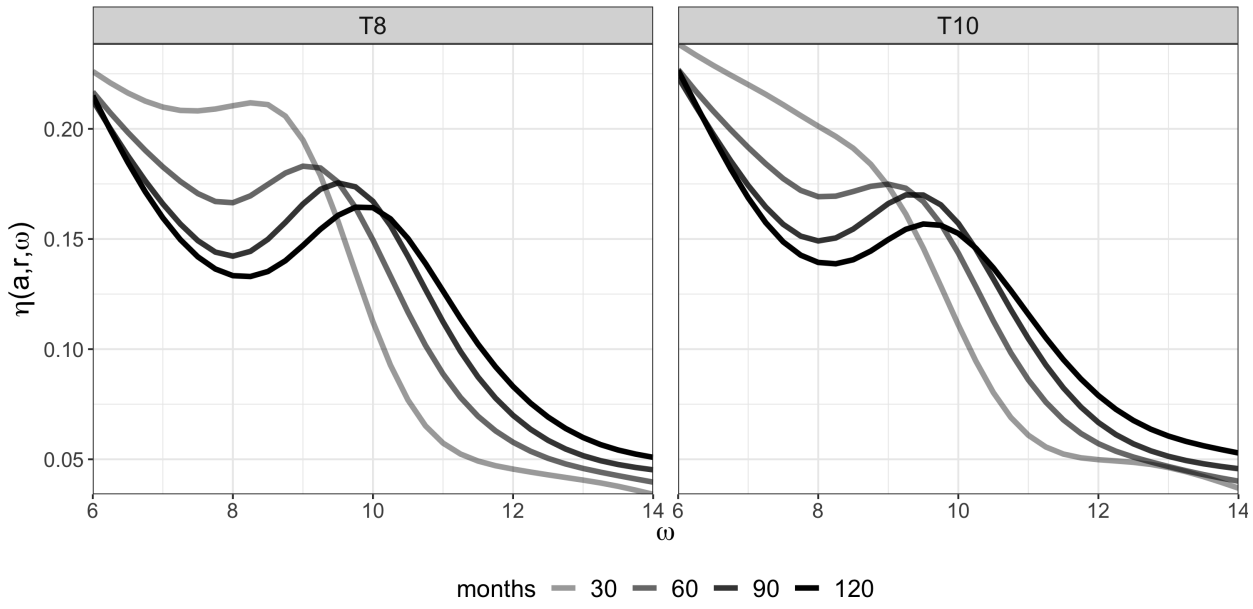


Figure 4.2: Slices of the region-referenced mean surface $\eta(a, r, \omega)$ representing the electrode-specific mean alpha spectral density at ages 30, 60, 90, and 120 months for the T8 and T10 electrodes. Darker lines correspond to older children.

the absolute effect of a given electrode across all subjects. Referring to Figure 4.1(b), the three electrodes with the highest average contribution to the log-odds of ASD diagnosis, the T8, T10, and F8 electrodes, are located in the right temporal and frontal region of the scalp, suggesting differences in these brain regions have the strongest effect on whether a subject is predicted to have an ASD diagnosis. The average mean-centered functional predictors for the TD and ASD children displayed in in Figure 4.3(top row) provide insights into group differences. On average at 30 months old, TD children display higher alpha power between 8 to 10 Hz. This changes over the course of development and by 120 months ASD children display higher alpha power between 6 to 10 Hz and TD children show higher power between 10 to 12 Hz.

The slices of the estimated regression function $\widehat{\beta}(a, r, \omega)$ for the T8 electrode are plotted in Figure 4.3(middle row). The estimated regression function puts mostly negative weight on the spectral density for frequencies between 8 to 10 Hz and positive weight for frequencies between 10 to 14 Hz for subjects aged 30 to 60 months. After 60 months the shape of the regression function begins to flip, with positive weight on frequencies between 6 and 10 Hz

and negative weight for those between 10 and 14 Hz. The Bayesian point-wise confidence intervals are wide including zero, due to the small sample size within diagnostic groups; however they exclude zero at 11.25 Hz for children between 124 to 134 months (not shown). The point-wise product of the average mean-centered functional predictors for each group and the regression function are displayed in Figure 4.3(bottom row) where the shading represents the area under the curve for each group which ultimately encapsulates the linear effect on the log odds of ASD diagnosis. Before 60 months, the average TD child has a PAF between 8 to 10 Hz and hence due to the negative weighting by the regression coefficient, this results in a predicted probability of ASD diagnosis that is less than .5. Similarly, since the average TD child of age older than 60 months have PAF between 10 to 14 Hz, a negative weight in that domain results in a predicted probability of ASD diagnosis that is again less than .5. Similar descriptions can be formed for the ASD group and on average the areas under the curve produced by the point-wise products formed in Figure 4.3(bottom row) are in accordance with the true diagnostic status.

To get a better sense of the predictive performance of the CARR-GFLM in our data across electrodes stratified by developmental age, we look at the predicted probabilities of ASD diagnosis and their associated 95% confidence intervals for the study subjects in Figure 4.4(top row). The 95% confidence intervals are calculated on the logit scale based on the posterior distribution of $\boldsymbol{DV}_q\tilde{\boldsymbol{\theta}}|\boldsymbol{y}$ (see Section 4.2.2) and then transformed onto the probability scale. If subjects are classified based on a threshold of .5, then at younger and older ages CARR-GFLM discriminates between the two groups well. However, between 50 and 75 months the model has some trouble distinguishing the two diagnostic groups. This is likely because the differences in the alpha spectrum between the two groups are minimal at the median study age, suggesting the greatest group differences in alpha spectral dynamics occur at younger and older ages. In order to contrast the predictive performance of the CARR-GFLM with the existing models of m-GFLM and m-GFLMi, we compare performance measures including sensitivity (sens $= P\{\widehat{ASD}|ASD\}$), specificity (spec $= P\{\widehat{TD}|TD\}$), and the area under the receiver operating characteristic curve (AUC). Predicted probabilities are estimated using leave-one-out cross validation (LOOCV) in which each subject is iteratively withheld

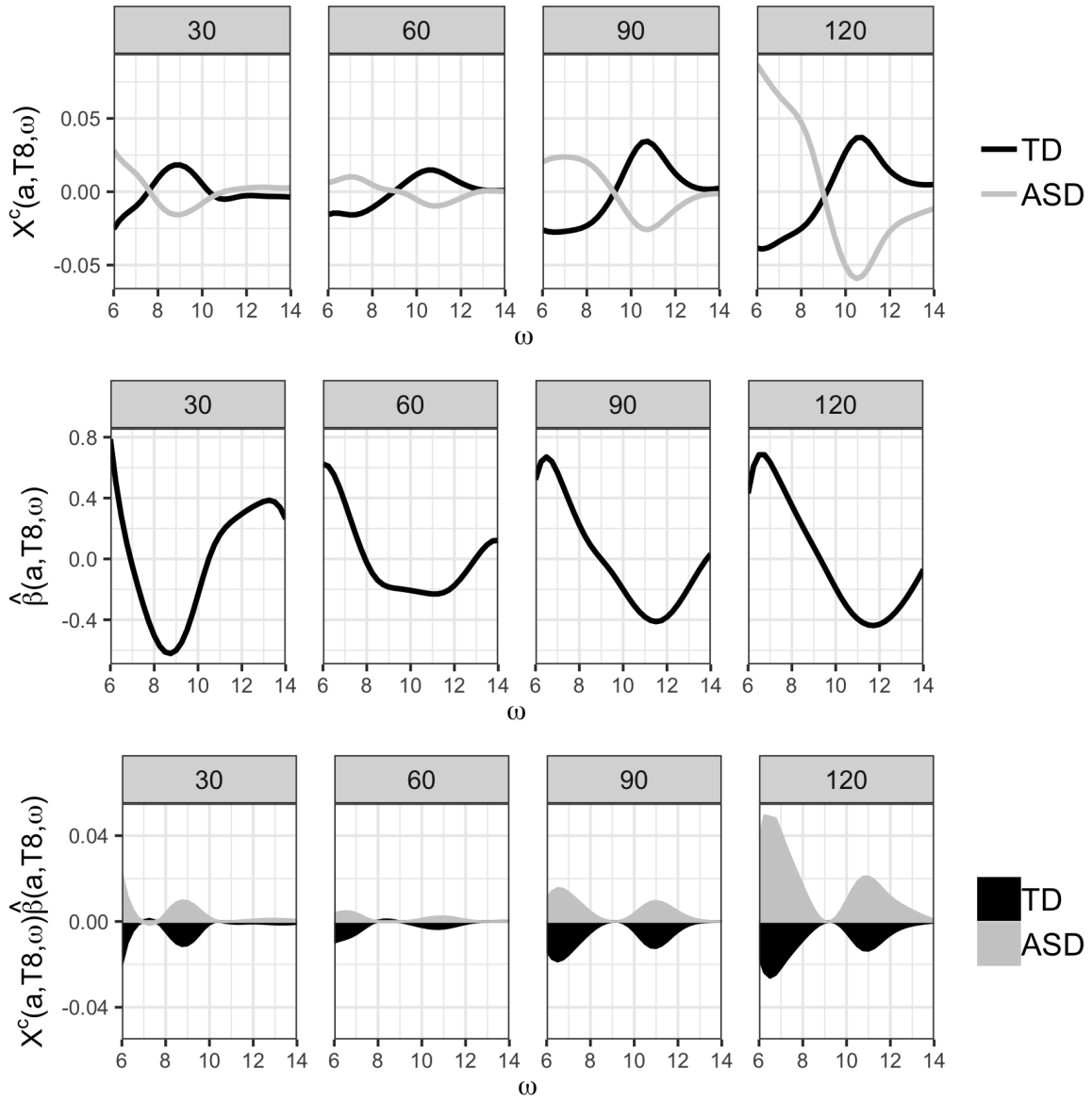


Figure 4.3: The results from fitting the CARR-GFLM model to the resting state EEG data for the T8 electrode. Results are presented with increasing age from 30 to 120 months organized by column. (top row) The average mean-centered functional predictor for (black) TD children and (grey) ASD children. (middle row) Cross sections of the estimated regression function. (bottom row) The point-wise product of the top two rows where the shading represents the average area under the curve for the (black) TD children and (grey) ASD children.

from the model data, models are fit, and then probabilities are predicted for the withheld subject. The CARR-GFLM (sens = .602; spec = .663; AUC = .635) outperforms both m-GFLM (sens = .527; spec = .588; AUC = .553) and m-GFLMi (sens = .822; spec = .363; AUC = .593) in terms of both balance between sensitivity and specificity and AUC.

It is clear from the superior performance of the CARR-GFLM and m-GFLMi models compared to m-GFLM that including age improves prediction of ASD diagnosis. To determine why CARR-GFLM outperforms m-GFLMi despite the fact that both models account for age, we consider the estimated regression functions from the T8 electrode for both models. For m-GFLMi (Figure 4.4(bottom row)), the association between the alpha spectral densities and ASD diagnosis is modeled as a function of two terms, the main effect, $X_i^c(T8, \omega)$, and the interaction term, $a_i X_i^c(T8, \omega)$. The estimated regression function for the main effect is flat compared to the estimated regression function of the interaction term which assumes a relatively linear decreasing trend with a positive weight on the alpha spectral density between 6 to 9 Hz and a negative weight between 9 to 14 Hz. These effects get stronger as children age. On the other hand, CARR-GFLM (Figure 4.3(middle row)), allows the regression function to vary in a non-linear manner across age, with the sign and shape of the regression function shifting across development. The benefit of the greater flexibility of CARR-GFLM over m-GFLMi is visible in the predicted probabilities for each model. While both models struggle to differentiate between the two groups between 50 and 75 months, CARR-GFLM (Figure 4.4(top row)) is able to discriminate between the two diagnostic groups at younger ages between 25 and 50 months while m-GFLMi (Figure 4.4 (middle row)) shows a clear bias towards a diagnosis of ASD, likely induced by the more rigid linear modeling structure.

4.4 Simulation

We assess the finite sample performance of the proposed methodology across a range of simulation settings. The data generating process for the simulation is described in Section 4.4.1 and simulation results are deferred to Section 4.4.2.

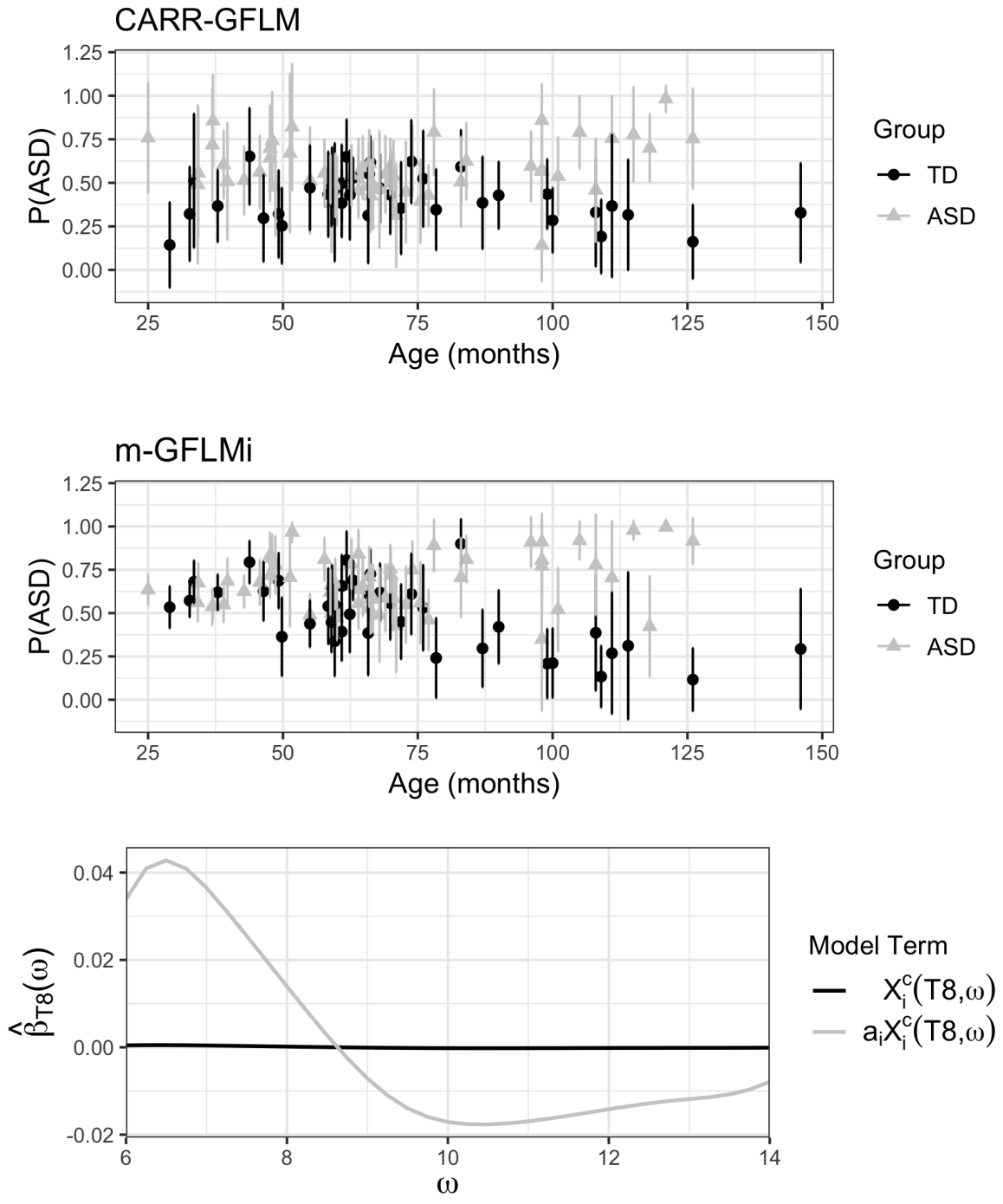


Figure 4.4: (top and middle row) The predicted probabilities of ASD diagnosis and their associated 95% confidence intervals for the study subjects, with true group membership denoted by black for TD and grey for ASD for the CARR-GFLM (top) and m-GFLMi (middle) models. (bottom row) The estimated regression functions for the main effects and interaction term from the T8 electrode for the m-GFLMi model.

4.4.1 Data generation

Binary scalar outcomes are simulated from $y_i \sim \text{Bernoulli}(\mu_i)$, $i = 1, \ldots, n$, where the subject-specific probabilities are formed on the log odds scale through the linear model $\text{logit}(\mu_i) = \sum_{r=1}^{R} \int X_i^c(a_i, r, \omega)\beta(a_i, r, \omega)d\omega$. The functional and covariate grids are chosen as 50 and 30 equidistant points in $[0, 1]$, respectively, with data generated for a total of $R = 15$ regions. For each subject, the observed covariate values are simulated from a discrete uniform distribution defined on the covariate grid in $[0, 1]$. The regression function $\beta(a, r, \omega)$ is constructed to vary both across regions and along the functional and covariate domains,

$$\beta(a, r, \omega) = \begin{cases} (-1)^r 2\cos(\omega r\pi/6 + \pi a), & r = 1, \ldots, 8 \\ (-1)^r 2\sin\big\{\omega(r-8)\pi/6 + \pi a\big\}, & r = 9, \ldots, 15. \end{cases}$$

The subject-specific functional predictors are formed by $X_i(a_i, r, \omega) = \sum_{k=1}^{5} \xi_{ikr}\psi_k(\omega)$, with a common set of basis functions $\psi_k(\omega)$, $k = 1, \ldots, 5$, across the 15 regions, where $\psi_k(\omega)$ are cubic B-splines and the vector of subject-specific scores is generated from $\xi_{ik} = (\xi_{ik1}, \ldots, \xi_{ikR})^{\mathrm{T}} \sim \mathcal{MVN}(\mathbf{0}, \Sigma_k)$. The $r \times r$ covariance matrix Σ_k controls the regional dependency of the functional predictors by inducing correlations between a subject's scores for the kth spline basis across the R regions. The covariance matrix Σ_k is chosen to have a compound symmetric structure with diagonal entries equal to one and off diagonal entries equal to ρ, a tuning parameter for the level of dependency across regions. For simplicity, we set $\eta(a_i, r, \omega)$ to zero.

We perform 500 Monte Carlo runs across nine simulation settings: three sample sizes ($n = 200, 500, 1000$) and three levels of regional dependency ($\rho = 0.0, 0.1, 0.3$). We use the relative squared error $\text{RSE}(a, r, \omega) = ||\hat{f}(a, r, \omega) - f(a, r, \omega)||^2 / ||f(a, r, \omega)||^2$ to assess the regression function estimates, where $||f(a, r, \omega)||^2 = \sum_{r=1}^{R} \int\int f^2(a, r, \omega)dad\omega$. The coverage probability of the Bayesian point-wise confidence intervals as a function of a, r and ω, is assessed by recording the proportion of times the regression function estimates lie within

the confidence interval over the 500 Monte Carlo runs. For each Monte Carlo run at a fixed sample size n, we generate $n + 200$ samples, where first n is used for estimation and the additional 200 samples are reserved as a validation set for assessing prediction accuracy. Prediction accuracy is assessed by the AUC in the validation sets. The use of validation sets with a common number of observations (200 samples) allows for comparisons of the AUC from different simulation settings.

The region-referenced mean curves are estimated by pooling data across subjects and performing bivariate penalized smoothing over the functional and covariate domains with a tensor basis of penalized cubic B-splines (with 10 and 5 degrees of freedom in the functional and covariate domain, respectively). In addition, second order difference penalties are used along the functional and the covariate dimension, similar to the data analysis. The marginal basis matrices $\boldsymbol{\Phi}_a$ and $\boldsymbol{\Phi}_\omega$ are formed as evaluations of the cubic B-splines with $K_a = 5$ and $K_\omega = 5$ degrees of freedom, respectively, and the regional basis matrix is equal to $\boldsymbol{\Phi}_r = \left[\mathbf{1}_{15}, \boldsymbol{I}_{15}\right]$ with $K_r = 16$. First order marginal difference penalties are utilized for both $\boldsymbol{P}_a$ and $\boldsymbol{P}_\omega$ to ensure smoothness over the functional, and the covariate domains. For the regional domain, a ridge style penalty $\boldsymbol{P}_r = \left[\mathbf{0}_{15}, \boldsymbol{I}_{15}\right]$ is employed as in the data analysis. The model for all simulation settings has $K_a \times K_r \times K_\Omega = 400$ coefficients and for each Monte Carlo run SVD is used to reduce the dimension of the design matrix such that the resulting columns account for at least 95% of the total variation. Across simulation settings, the rank of the SVD reduced design matrix increases with sample size and decreases as a function of ρ, with the median rank at $\rho = 0.0$ equal to 93, 128, 145, for $n = 200, 500$, and 1000, respectively. Moving from $\rho = 0.0$ to $\rho = 0.3$, the median rank decreases by approximately 20 at each sample size. The median penalty parameters across 500 Monte Carlo runs for $n = 500; \rho = 0.0$ are $\lambda_a = .0033$, $\lambda_r = .0036$, and $\lambda_\omega = .0043$, with penalty parameters displaying no median trend across sample size or dependency structure but slightly more variation among Monte Carlo runs at $n = 200$ compared to larger sample sizes. The average computation times based on ten iterations for Monte Carlo runs at $n = 200, 500, 1000$ are 7.6, 15.4, and 29.1 seconds, respectively. Predictive performance of CARR-GFLM is compared to the existing methods of m-GFLM and m-GFLMi with functional effects estimated by

projection onto a basis of cubic B-splines with five degrees of freedom and second degree differencing penalties.

4.4.2 Results

Figure 4.5 displays the results from 500 Monte Carlo runs under each simulation setting, with RSE values for the regression function $\beta(a, r, \omega)$ (top row) and coverage probabilities for the Bayesian point-wise confidence intervals (bottom row). As expected, RSE values decrease as sample size increases with the median RSE reduced by approximately a factor of 3 moving from $n = 200$ to $n = 1000$ at each level of regional dependency. For a fixed sample size, an increase in regional dependency produces a modest but consistent increase in median RSE (RSE is increased by 7.2% with ρ changing from 0.0 to 0.3 at $n = 200$). With increasing ρ, functional predictors at each region share more information and the estimated regression function may lose precision much akin to when a multivariate regression experiences multicollinearity. The true and estimated regression function $\beta(a, r, \omega)$ at three regions from the Monte Carlo run with the median RSE (0.327) for $n = 500$ and $\rho = 0.1$ is shown in Figure 4.6. Despite the non-negligible RSE, the shape, periodicity, and magnitude of the regression function is well preserved, suggesting that the accumulation of estimation error is evenly distributed across the regression functions from each region rather than being concentrated within a few regions. Note that $n = 200$ is a small sample size for functional regression settings, especially for binary functional regression. This explains the relatively high median RSE values observed for $n = 200$ (ranging between 0.57 and 0.65 for varying values of ρ). The coverage probabilities for each simulation setting approaches the nominal level of 95% as sample size increases. For sample sizes $n = 500$ and 1000, the median coverage observed is consistently larger than .83. Since the confidence intervals considered are point-wise, they are not expected to hit the nominal level uniformly over all a, r and ω. However for $n = 200$, coverage decreases significantly with increasing ρ. This may be due to the fact that the rank of the resulting design matrix after SVD is the smallest at $n = 200$ and $\rho = .3$, leading to narrower confidence intervals. Note that because coverage probabilities are calculated at each (a, r, ω), the number of points considered $15 * 30 * 50 = 22,500$ is large

and thus outliers have been jittered horizontally to improve presentation.

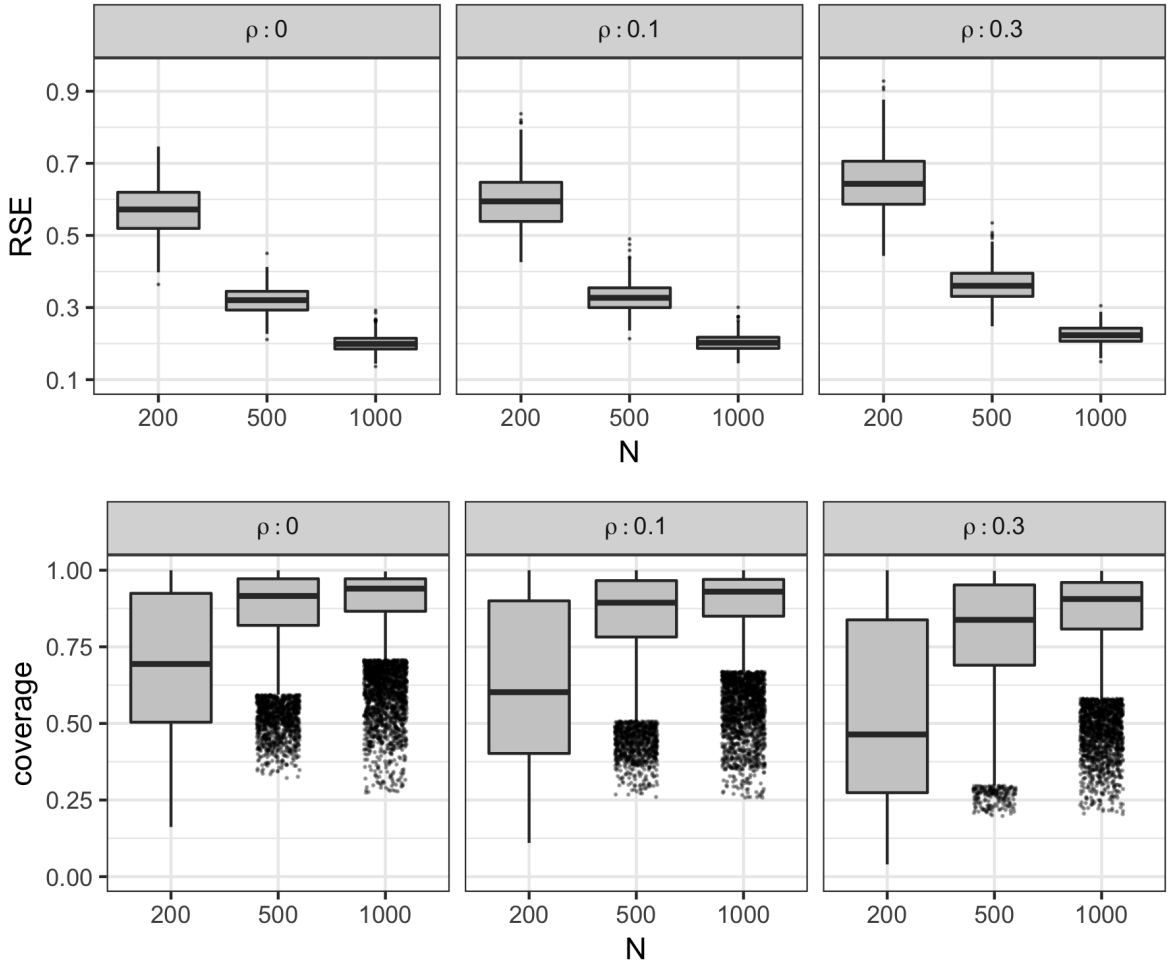


Figure 4.5: The simulation results from 500 Monte Carlo runs under each simulation setting ($\rho = 0, 0.1, 0.3$ in columns and $n = 200, 500, 1000$ in columns within panels). RSE values for the regression function $\beta(a, r, \omega)$ (top row) and the coverage probability for the Bayesian point-wise confidence intervals for a nominal level of 95% ((bottom row) are provided. Outliers are jittered horizontally to improve presentation.

Figure 4.7 compares the AUC from CARR-GFLM (top row) with m-GFLM (middle row) and m-GFLMi (bottom row) over 500 Monte Carlo runs under each simulation setting. For all models, the median AUC for the validation sets improves with increasing sample size and decreases with increasing regional dependency, though the differences observed for regional dependency are small. At each simulation setting, a descending trend is observed for median AUCs as one moves from CARR-GFLM to m-GFLMi to m-GFLM. While the median AUC

for CARR-GFLM is greater than .80 for sample sizes greater than $n = 200$, m-GFLM and m-GFLMi fail to exceed a median AUC of .75 for any simulation setting, suggesting that incorporating flexible covariate-adjustments is essential for predictive performance even at large sample sizes. The overall good AUC for CARR-GFLM suggests that despite the high model complexity, the regularization induced by the quadratic penalty and SVD avoids overfitting and allows for generalization of the fitted model to newly observed data.

4.5 Discussion

We propose a covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM) that estimates functional effects across a non-smooth regional domain while simultaneously adjusting for observed covariates. The proposed estimation procedure projects the regression function onto a tensor basis formed from a kronecker product of one-dimensional discrete and continuous basis functions. The tensor structure allows for construction of a flexible penalty structure that induces regularization along each dimension while at the same time controlling the number of shrinkage parameters. Even for a three-dimensional regression function, the number of elements in the tensor basis will often greatly exceed the number of observed subjects and thus SVD is utilized to reduce the dimension of the design matrix allowing the proposed model to be fit in standard software. The model can be generalized to accommodate a vector of covariates by introduction of additional marginal bases in the kronecker product.

The proposed method is used to model associations between diagnostic status and alpha band spectral dynamics in ASD and TD children across a broad developmental range. The challenge in estimating a single PAF at each electrode is a voided by considering the full alpha spectral density, where the information on the developmental stage of the child is integrated into the model by adjusting for chronological age. Thus, based on EEG data alone, we find that differences across the scalp in alpha band spectral dynamics between ASD and TD children at similar ages can predict diagnostic status reasonably well. This finding suggests that developmental differences in the alpha band spectral density may provide a promising

point of further investigation into the underlying neural differences between ASD and TD children. Performance of the CARR-GFLM model is compared to existing methods in both the data analysis and the simulation study and is found to provide superior prediction and inference. While the proposed model is motivated by a developmental EEG study, the methodology can be considered in applications involving other brain imaging modalities with a regionally-referenced functional predictor and an additional set of covariates.

Supporting materials

The code for the proposed estimation and inference procedures are made publicly available online on the Github page [https://github.com/aaron-scheffler/CARRGFLM], along with a tutorial for step-by-step implementation of the proposed methodology.

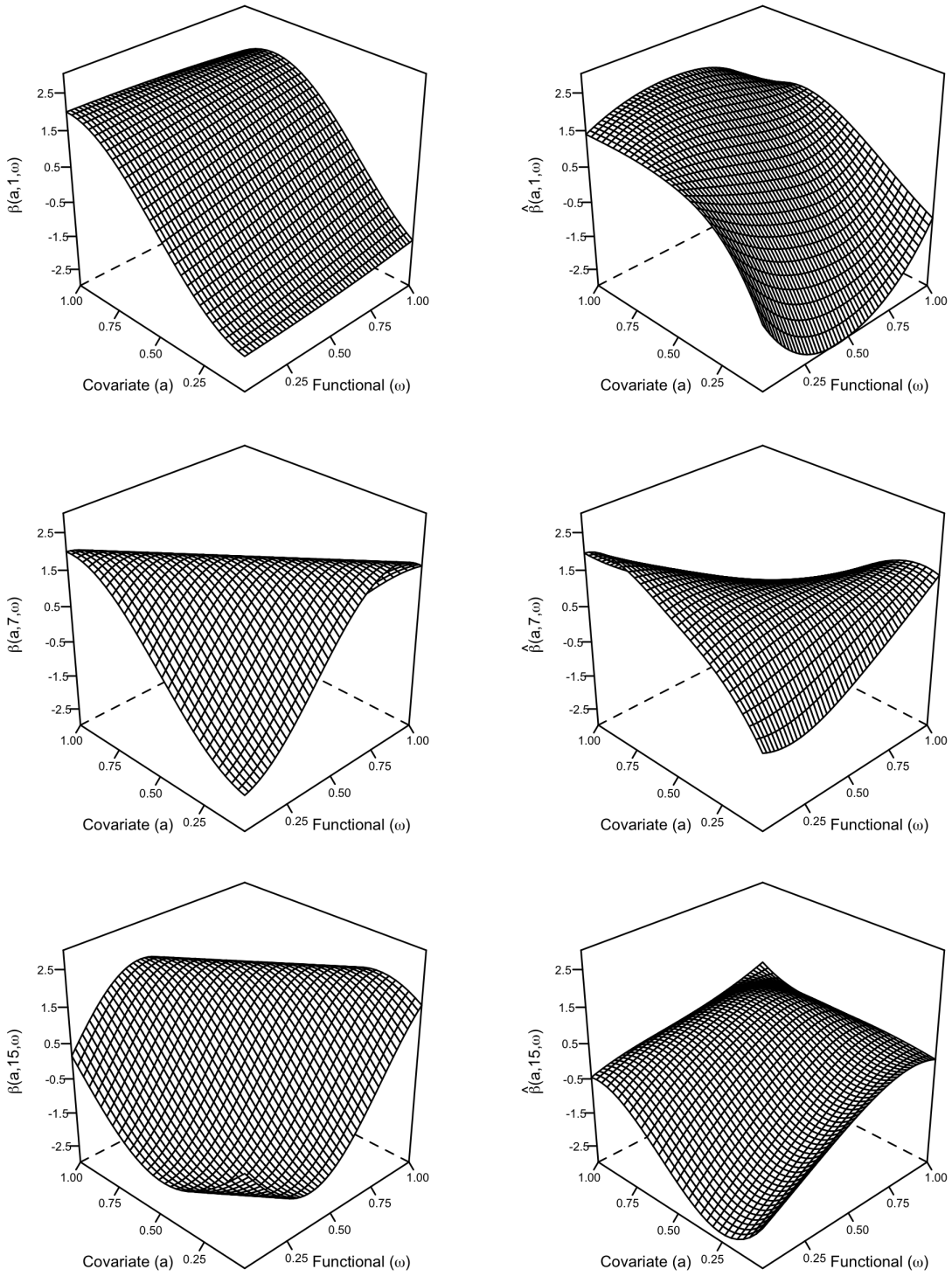


Figure 4.6: The true (left column) and estimated (right column) regression function $\beta(r, a, \omega)$ for regions $r = 1, 7, 15$ (descending rows) from the Monte Carlo run with the median RSE (0.327) under the simulation design, $n = 500$; $\rho = 0.1$.

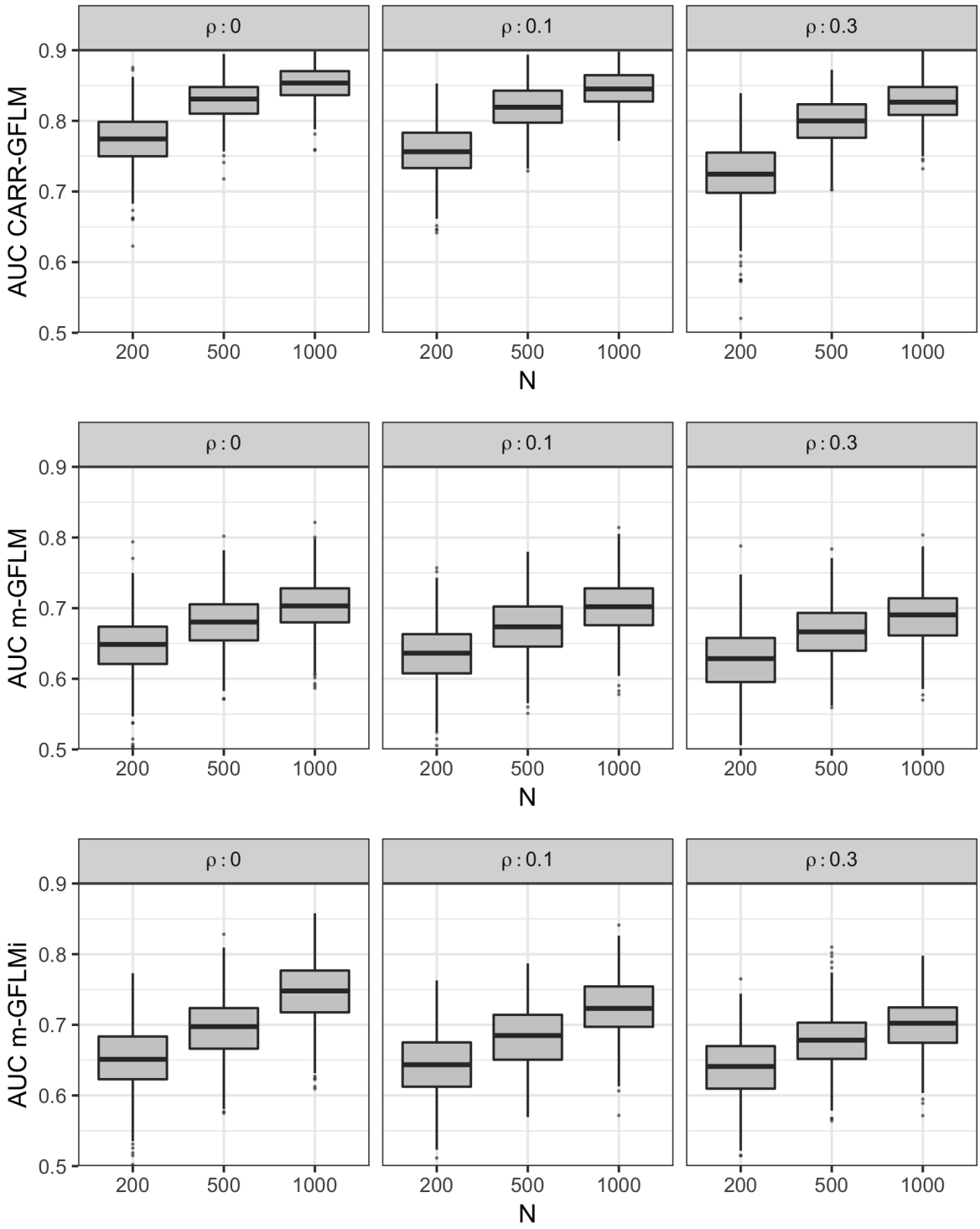


Figure 4.7: The AUC for the validation data sets from 500 Monte Carlo runs under each simulation setting ($\rho = 0, 0.1, 0.3$ in columns and $n = 200, 500, 1000$ in columns within panels) for the CARR-GFLM (top row), mGFLM (middle row), and m-GFLMi (bottom row).

CHAPTER 5

Conclusion

This concluding chapter provides a summary in Section 5.1 of the proposed methods and insights gained by their application to the motivating EEG studies in children with ASD. The challenges posed by region-referenced longitudinal functional data, and other multi-dimensional functional data structures, provide several avenues for future work which are discussed in Section 5.2.

5.1 Concluding summary

As is the case in many biomedical applications, EEG produces highly structured data that exhibits intricate dependencies with rich information but the dimensionality and size of the data can produce significant obstacles for interpretation, estimation, and inference. In studies that utilize EEG to measure neural activity, the resulting region-referenced longitudinal functional data is often collapsed along all three dimensions to facilitate analysis by standard statistical methods which produces a loss of valuable information. This doctoral work proposed three functional data methods methods that both maintain information along each dimension and yield interpretable components and inferences. Methodological development centers on extending two fundamental tools of FDA, functional principal components analysis and functional regression, from the setting of one-dimensional curves to region-referenced longitudinal functional signals. In doing so, several theoretical and computational challenges are addressed, namely developing flexible but low-dimensional representations of these complex data structures and their corresponding effects and applying appropriate regularization along each dimension of the data. The proposed methods are motivated by three exper-

imental paradigms that each utilize EEG as covert measure of brain activity in children with ASD and their TD peers, producing a common region-referenced longitudinal function data structure. By allowing the EEG data to be modeled without dimensional collapse, the proposed methods provide novel insights of the neural features underlying autism across a wide developmental spectrum.

In Chapter 2, the proposed multi-dimensional functional principal components analysis (MD-FPCA) decomposition provides a flexible and nonparametric decomposition based on separation of the total variation into subject and subunit level variation which are further decomposed in a two-stage FPCA across both longitudinal and functional dimensions. The MD-FPCA decomposition is the first proposed for repeatedly measured longitudinal functional data and characterizes functional variation both longitudinally over the course of the implicit learning paradigm and spatially over the scalp through an exchangeable correlation structure. In the implicit learning data, characterizing longitudinal trends in ERP waveforms over the course of the experiment is considered essential to track visual learning. Based on the MD-FPCA decomposition, TD children are found to achieve earlier visual condition differentiation than their ASD peers with substantial variation in condition differentiation among electrodes and throughout the course of the experiment. While the MD-FPCA decomposition is extremely flexible in its treatment of functional dynamics across the experiment (i.e. a separate FPCA is performed at each longitudinal grid point), the spatial correlation structure is somewhat restrictive due to the assumption of exchangeability of electrodes within a scalp region. This produces two potential limitations, (1) the flexibility of the longitudinal functional covariance results in added computational burden and (2) the exchangeability assumption requires justification that is not always scientifically available.

In Chapter 3, hybrid principal components analysis (HPCA) introduced an alternative decomposition for region-referenced longitudinal functional data that provides superior balance between parsimony and flexibility. Under the assumption of weak separability, the total covariance is decomposed via a tensor product of marginal eigenvectors and eigenfunctions obtained from the marginal regional, longitudinal, and functional covariances. By operating on the marginal covariances along each dimension, computation is alleviated while simul-

taneously allowing for a more unrestricted form for the regional dependency structure (i.e. similar to an empirical rather than a compound symmetric covariance matrix). A mixed effects modeling framework is used to estimate subject loadings onto the tensor basis, allowing for sparsity introduced during artifact correction in EEG studies. The mixed effects model allows for a parametric bootstrap procedure to perform group-level inference. Application of the HPCA decomposition to the word segmentation paradigm data finds the strongest differences in spectral dynamics among the TD, vASD, and mVASD groups were located in the left and right frontal regions of the scalp, particularly among high frequency oscillations.

In Chapter 4, attention shifts from decompositions of the total variation to the modeling of clinical outcomes from region-referenced EEG data while adjusting for non-functional covariates using a covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM). The motivating EEG data comes from a developmental study that produces two analytical challenges that must be overcome, (1) functional effects must be jointly fit across a discrete regional domain, and (2) the resulting regression function needs to be adjusted for covariate-effects analogous to a varying-coefficient model. Both challenges are addressed simultaneously by projecting the regression function onto a tensor basis of discrete and continuous bases in the regional, functional, and covariate dimensions. A key features in estimation is defining the penalty structure for the basis coefficients as a kronecker sum of penalties along each dimension. The resulting system is rank deficient and singular value decomposition is applied to to the design matrix and penalty structure to create a well-posed problem. Applying CARR-GFLM to the motivating data finds that PAF tends to increase with age across the scalp in TD children but not in ASD children and this developmental difference can be used to predict diagnostic status. This finding suggests that developmental differences in the alpha band spectral density may provide a promising point of further investigation into the underlying neural differences between ASD and TD children.

The proposed methods provide a set of inferential and analytical tools for region-referenced longitudinal functional data that preserve information along each dimension while simultaneously modeling functional associations and covariation. The finite sample properties of each estimation procedure are assessed via extensive simulation studies. Although the proposed

methods are motivated and applied to experimental EEG data, the models are also applicable to general settings in which longitudinal functional data are observed across discrete regions, such as fMRI brain imaging studies or streaming activity data from multiple sensors (i.e. gait monitoring).

5.2 Future work and directions

Region-referenced longitudinal functional data is highly complex and thus the opportunities for future work are bountiful both from a theoretical and a computational perspective. Several direct extensions of the proposed work are available. Similar to the covariate-adjustment in CARR-GFLM, introducing covariate-adjustments to high-dimensional covariance decompositions would allow for characterization of changes in dependency across clinically informative quantities (e.g. age). Another interesting project would be to weave together FPCA and functional regression by incorporating information from previously developed covariance decompositions to form adaptive penalty structures for use in multi-dimensional functional regression models [James and Silverman, 2005]. In addition, motivated by recent work in functional graphical models [Qiao et al., 2019] and dynamic connectivity [Lan et al., 2017, Li et al., 2019], modeling connectivity and graph structures of high-dimensional functional data would prove an interesting methodological problem to address as well as informative in many biomedical settings.

Another direction for future work is to draw connections between the literature on functional regression and a set of flexible optimization methods called the majorize-minimization (MM) algorithm [Lange, 2016]. The MM algorithm is a method for obtaining parameter estimates by bounding a cost function either above or below by a simpler convex function and then iteratively minimizing or maximizing the bounding function. By reducing the complexity of parameter estimation to a problem of convex optimization, the MM algorithm allows for stable fitting of complex models with the added numerical benefit of separating parameters during optimization. Although the MM algorithm has been applied to a variety of statistical problems, the connection has not yet been made with functional regression

models. By introducing the MM algorithm to the functional data literature, new approaches to variable selection and smoothing will be possible through fast optimization of a broad array of likelihood functions.

Appendix 2A: Estimation algorithm

In this section we provide an outline of the proposed estimation procedures and refer the reader to Şentürk and Nguyen (2011), Şentürk et al. (2013), Chen and Müller (2012) and Di et al. (2014) for further details. We also present explicit algorithm steps at the end of this section. Denote the observations collected on the i^{th} subject at subunit j, longitudinal time s_{ijq} and functional time $t_{ijq\ell}$ by $\{X_{ij}(t_{ijq\ell}|s_{ijq}),\ i = 1, \ldots, n;\ j = 1, \ldots, J;\ q \in Q_{ij};\ \ell \in L_{ijq}\}$, where the sets of observed longitudinal and functional times (Q_{ij} and L_{ijq}, respectively) are allowed to be different across subjects, subunits and longitudinal times. The global mean surface $\mu(t, s)$ is estimated by a two-dimensional smoother applied to all observed data $\{s_{ijq}, t_{ijq\ell}, X_{ij}(t_{ijq\ell}|s_{ijq}),\ i = 1, \ldots, n;\ j = 1, \ldots, J;\ q \in Q_{ij};\ \ell \in L_{ijq}\}$. The subunit-specific means $\eta_j(t, s)$ are similarly estimated by smoothing across all mean-centered observation pairs $\{s_{ijq}, t_{ijq\ell},\ X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}),\ i = 1, \ldots, n;\ q \in Q_{ij};\ \ell \in L_{ijq}\}$.

Let $\{s_q,\ q = 1, \ldots, Q\}$ be the unique set of longitudinal times among $\{s_{ijq},\ i = 1, \ldots, n;\ j = 1, \ldots, J;\ q \in Q_{ij}\}$. The two-stage estimation procedure involves the functional principal components decompositions of the subject and subunit covariance surfaces over functional time at a fixed longitudinal time s_q in the first-stage. The subject covariance surface $\Sigma^{(1)}(t, t'|s_q)$ is estimated by a two-dimensional smoothing of the products $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}\{X_{ij'}(t_{ij'q\ell'}|s_{ij'q}) - \hat{\mu}(t_{ij'q\ell'}, s_{ij'q}) - \hat{\eta}_{j'}(t_{ij'q\ell'}, s_{ij'q})\}$ over functional times $\{t_{ijq\ell},\ t_{ij'q\ell'}\}$ for all subjects and subunits from different electrodes, observed at $s_{ijq} = s_q$, $\ell \in L_{ijq}$, $\ell' \in L_{ij'q}$ and $j \neq j'$. The subunit covariance surface is estimated by the difference between the total and subject covariances, $\widehat{\Sigma}^{(2)}(t, t'|s_q) = \widehat{\widehat{\Sigma}}_T(t, t'|s_q) - \widehat{\Sigma}^{(1)}(t, t'|s_q)$, where the total covariance surface is obtained by the two-dimensional smoothing of $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}\{X_{ij}(t_{ijq\ell'}|s_{ijq}) - \hat{\mu}(t_{ijq\ell'}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell'}, s_{ijq})\}$ over functional times $\{t_{ijq\ell},\ t_{ijq\ell'}\}$ for all subjects and subunits from the same electrode, observed at $s_{ijq} = s_q$ and $\ell \neq \ell' \in L_{ijq}$. Note that the diagonal ($\ell = \ell'$) is excluded in the smoothing of the two-dimensional total covariance surface to eliminate the effects of mea-

surement error. The measurement error variance $\sigma^2_{s_q}$ can be estimated by smoothing the difference between the left out diagonal and the diagonal of the estimated total covariance surface, $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}^2 - \widehat{\widehat{\Sigma}}_T(t_\ell, t_\ell|s_q)$, over $\{t_{ijq\ell}\}$ for all subjects and subunits observed at $s_{ijq} = s_q$. The bandwidths used in the two-dimensional smoothing of the mean and covariance surfaces can be selected using cross-validation or generalized cross-validation (GCV).

After obtaining the covariance surface estimates at each longitudinal time s_q, a nonparametric functional principal component decomposition is employed on the smooth estimates of the covariance surfaces, $\widehat{\Sigma}^{(1)}(t, t'|s_q)$ and $\widehat{\Sigma}^{(2)}(t, t'|s_q)$, by a standard discretization procedure to estimate the subject and subunit level (first and second level) eigenvalues and eigenfunctions, $\lambda^{(1)}_k(s_q), \phi^{(1)}_k(t|s_q)$ and $\lambda^{(2)}_p(s_q), \phi^{(2)}_p(t|s_q)$, respectively. In order to guarantee the non-negative definiteness of the covariance matrix, the negative eigenvalue estimates and the corresponding eigenfunctions are removed from the functional principal component decomposition of the covariances. In addition, to maintain the smoothness of $\hat{\phi}^{(1)}_k(\cdot|s)$ and $\hat{\phi}^{(2)}_p(\cdot|s)$ over s, we determine the signs of consecutive eigenfunctions as follows. Let $\hat{\phi}^{(1)}_k(\cdot|s_q)$ be the first level eigenfunction estimate at the qth longitudinal time s_q. The sign of the eigenfunction, $\hat{\phi}^{(1)}_k(\cdot|s_q)$, is determined such that the L^2 distance to the previous eigenfunctions $\{\hat{\phi}^{(1)}_k(\cdot|s_{q-1}), \ldots, \hat{\phi}^{(1)}_k(\cdot|s_{q-q'})\}$ is minimized. We found comparisons up to $q' = 5$ previous eigenfunctions to yield good results in our implementations. This is recursively performed for the entire domain of s for both the first and second level eigenfunctions across k and p, respectively.

The decompositions given in (2) and (5) are truncated to include only the components containing the largest modes of variation. Specifically, the number of first-stage subject level (K) and subunit level (P) principal components are selected using percentage of variation explained. Let K_{s_q} and P_{s_q} be the smallest number of components satisfying the criteria $\{\sum_{k=1}^{K_{s_q}} \hat{\lambda}^{(1)}_k(s_q)\}/\{\sum_{k=1}^{M_1} \hat{\lambda}^{(1)}_k(s_q)\} > 0.9$ and $\{\sum_{p=1}^{P_{s_q}} \hat{\lambda}^{(2)}_p(s_q)\}/\{\sum_{p=1}^{M_2} \hat{\lambda}^{(2)}_k(s_q)\} > 0.9$ for first and second levels, respectively, where M_1 and M_2 are large. We select $K \equiv \max_q(K_{s_q})$ and $P \equiv \max_q(P_{s_q})$ to ensure the same number of principal components at each s_q for subsequent modeling of the first-stage eigenscores. The number of first (K'_k) and second level components

(P_p') of the second-stage decomposition are selected similarly to explain at least 90% of the total variation. Other methods for selecting the number of principal components include cross-validation (Rice and Wu, 2001) and Akaike's Information Criterion (Yao et al., 2005).

The first-stage subject and subunit level eigenscores, $\xi_{ik}(s_q)$ and $\zeta_{ijp}(s_q)$, are estimated using their best linear unbiased predictors (BLUP) as described in Di et al. (2014). Next, second-stage functional principal components decompositions are applied to the longitudinally observed first-stage eigenscores separately for $k = 1, \ldots, K$ and $p = 1, \ldots, P$. This involves the two-dimensional smoothing of the raw covariances $\{\hat{\xi}_{ik}(s_q)\hat{\xi}_{ik}(s_{q'})\}$ over $\{s_q, s_{q'}\}$, $q, q' = 1, \ldots, Q$ and $\{\hat{\zeta}_{ijp}(s_q)\hat{\zeta}_{ijp'}(s_{q'})\}$ over $\{s_q, s_{q'}\}$, $q, q' = 1, \ldots, Q$; $j, j' = 1, \ldots, J$. As with the first-stage decompositions, the smooth covariance surface estimates are discretized to estimate the second-stage subject and subunit level eigenvalues and eigenfunctions $\{\lambda_{kk'}, \psi^{(1)}_{kk'}(s)\}$ and $\{\lambda_{pp'}, \psi^{(2)}_{pp'}(s)\}$, respectively, for $k = 1, \ldots, K$; $k' = 1, \ldots, K_k'$; $p = 1, \ldots, P$; $p' = 1, \ldots, P_p'$. The second-stage eigenscores are also estimated using BLUP, yielding $\hat{\xi}'_{ikk'}$, $\hat{\zeta}'_{ijpp'}$. Estimated model components from both stages yield predictions of the subject and subunit trajectories over functional and longitudinal time,

$$\widehat{X}_{ij}(t|s) = \hat{\mu}(t,s) + \hat{\eta}_j(t,s) + \sum_{k=1}^{K}\sum_{k'=1}^{K_k'} \hat{\xi}'_{ikk'}\hat{\psi}^{(1)}_{kk'}(s)\hat{\phi}^{(1)}_{k}(t|s) + \sum_{p=1}^{P}\sum_{p'=1}^{P_p'} \hat{\zeta}'_{ijpp'}\hat{\psi}^{(2)}_{pp'}(s)\hat{\phi}^{(2)}_{p}(t|s).$$

The MD-FPCA estimation algorithm implemented in the MATLAB function `MultiLevelFuncLong.m` is outlined below. The custom functions developed specifically for MD-FPCA are listed alongside the relevant steps. For more detailed comments, please see the MATLAB code posted on Github. The runtime of a single Monte Carlo run for the sparsely observed data described in Appendix 2D with a sample size of $N = 30$ is 15 minutes and 32 seconds on a 2.4 GHz 6-Core Intel Xeon processor operating MATLAB R2015B. The runtime for the data analysis are 10 minutes and 17 seconds and 11 minutes and 31 seconds for the ASD and TD groups, respectively.

MD-FPCA Algorithm [`MultiLevelFuncLong.m`]

---

1. Estimate $\mu(t,s)$ with a two-dimensional smoother applied to all observed data

$\{s_{ijq}, t_{ijq\ell}, X_{ij}(t_{ijq\ell}, s_{ijq}),\ i = 1, \ldots, n;\ j = 1, \ldots, J;\ q \in Q_{ij};\ \ell \in L_{ijq}\}$. [`MeanSmooth2D.m`]

2. Estimate $\eta_j(t, s)$ with a two-dimensional smoother applied to all mean-centered observation pairs $\{s_{ijq}, t_{ijq\ell},\ X_{ij}(t_{ijq\ell}, s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}),\ i = 1, \ldots, n;\ q \in Q_{ij};\ \ell \in L_{ijq}\}$. [`MeanSmooth2D.m`]

3. First-stage Karhunen-Loève decomposition:
   (a) For each longitudinal time s_q, $q = 1, \ldots, Q$, perform a multilevel FPCA for all data $\{t_{ijq\ell}, X_{ij}(t_{ijq\ell}|s_{ijq}),\ i = 1, \ldots, n;\ j = 1, \ldots, J;\quad \ell \in L_{ijq}\}$ observed at $s_{ijq} = s_q$. [`MultilevelFPCA.m`]
      i. Estimate $\Sigma^{(1)}(t, t'|s_q)$ with a two-dimensional smoother applied to the products $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}\{X_{ij'}(t_{ij'q\ell'}|s_{ij'q}) - \hat{\mu}(t_{ij'q\ell'}, s_{ij'q}) - \hat{\eta}_{j'}(t_{ij'q\ell'}, s_{ij'q})\}$ over functional times $\{t_{ijq\ell},\ t_{ij'q\ell'}\}$ for all subjects and subunits from different electrodes, observed at $\ell \in L_{ijq}$, $\ell' \in L_{ij'q}$ and $j \neq j'$. [`CovarianceSmooth.m`]
      ii. Estimate $\widetilde{\Sigma}_T(t, t'|s_q)$ with a two-dimensional smoother applied to the products $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}\{X_{ij}(t_{ijq\ell'}|s_{ijq}) - \hat{\mu}(t_{ijq\ell'}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell'}, s_{ijq})\}$ over functional times $\{t_{ijq\ell},\ t_{ijq\ell'}\}$ for all subjects and subunits from the same electrode, observed at $\ell \neq \ell' \in L_{ijq}$. [`CovarianceSmooth.m`]
      iii. Estimate $\Sigma^{(2)}(t, t'|s_q) = \widetilde{\Sigma}_T(t, t'|s_q) - \Sigma^{(1)}(t, t'|s_q)$ by taking the difference of $\widehat{\widetilde{\Sigma}}_T(t, t'|s_q) - \widehat{\Sigma}^{(1)}(t, t'|s_q)$.
      iv. Estimate $\sigma^2_{s_q} = \{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}^2 - \widetilde{\Sigma}_T(t_{ijq\ell}, t_{ijq\ell'}|s_q)$, observed at $\ell = \ell' \in L_{ijq}$.
   (b) Employ a FPCA on $\widehat{\Sigma}^{(1)}(t, t'|s_q)$ to estimate $\{\lambda^{(1)}_k(s_q), \phi^{(1)}_k(t|s_q)\}$.
   (c) Employ a FPCA on $\widehat{\Sigma}^{(2)}(t, t'|s_q)$ to estimate $\{\lambda^{(2)}_p(s_q), \phi^{(2)}_p(t|s_q)\}$.
   (d) For each longitudinal time s_q, $q = 1, \ldots, Q$, let K_{s_q} and P_{s_q} be the smallest number of components satisfying the criteria $\{\sum_{k=1}^{K_{s_q}} \hat{\lambda}^{(1)}_k(s_q)\}/\{\sum_{k=1}^{M_1} \hat{\lambda}^{(1)}_k(s_q)\} > 0.9$ and $\{\sum_{p=1}^{P_{s_q}} \hat{\lambda}^{(2)}_p(s_q)\}/\{\sum_{p=1}^{M_2} \hat{\lambda}^{(2)}_k(s_q)\} > 0.9$, where M_1, M_2 are large.
   (e) Set $K = \max_q(K_{s_q})$ and $P = \max_q(P_{s_q})$.
   (f) For each longitudinal time s_q, $q = 2, \ldots, Q$, determine the sign of $\hat{\phi}^{(1)}_k(\cdot|s_q)$ by minimizing the L^2 distance to the previous eigenfunction $\hat{\phi}^{(1)}_k(\cdot|s_{q-1})$. Repeat the procedure for $\hat{\phi}^{(2)}_p(\cdot|s_q)$.
   (g) Estimate $\xi_{ik}(s_q)$ and $\zeta_{ijp}(s_q)$ using their multilevel best linear unbiased predictors (BLUP). [`ComputeScores.m`]

4. Second-stage Karhunen-Loève decomposition:
   (a) Perform a two-dimensional smooth of the raw covariances $\{\hat{\xi}_{ik}(s_q)\hat{\xi}_{ik}(s_{q'})\}$ over $\{s_q, s_{q'}\}$, $q, q' = 1, \ldots, Q$, and $\{\hat{\zeta}_{ijp}(s_q)\hat{\zeta}_{ij'p}(s_{q'})\}$ over $\{s_q, s_{q'}\}$, $q = 1, \ldots, Q; j, j' = 1, \ldots, J$.
   (b) Employ a FPCA on the smooth covariance surface obtained in 4 (a) to estimate $\{\lambda_{kk'}, \psi^{(1)}_{kk'}(s)\}$.

(c) Employ a FPCA on the smooth covariance surface obtained in 4 (a) to estimate $\{\lambda_{pp'}, \psi^{(2)}_{pp'}(s)\}$.

(d) Let K'_k and P'_p be the smallest number of components satisfying the criteria $\{\sum_{k'=1}^{K'_k} \hat{\lambda}^{(1)}_{kk'}\}/\{\sum_{k'=1}^{M_1} \hat{\lambda}^{(1)}_{kk'}\} > 0.9$ and $\{\sum_{p'=1}^{P'_p} \hat{\lambda}^{(2)}_{pp'}\}/\{\sum_{p'=1}^{M_2} \hat{\lambda}^{(2)}_{pp'}\} > 0.9$, where M_1, M_2 are large.

(e) Estimate $\xi'_{ikk'}$ and $\zeta'_{ijpp'}$ using their BLUP.

Appendix 2B: The meta-preprocessing

Due to the low SNR of the ERP, components such as the P3 peak are hard to identify on trial specific ERPs. The meta-preprocessing of Hasenstab et al. (2015), utilizing a moving average of ERPs across sliding trial windows, increases the SNR without collapsing the longitudinal dimension (via the typical practice of averaging across all ERP trials) to extract longitudinal information from ERPs. For illustration of the increase in the SNR and identification of the P3 and N1 components, consider a single ERP waveform for one subject from a single trial recorded in the right frontal region of the scalp and an average of 30 ERP waveforms from adjacent trials plotted in Figures A2.1 (a) and (b), respectively. While the components are unrecognizable in Figure A2.1 (a) due to low SNR, P3 peak and N1 dip are easily recognized in the meta-preprocessed ERP in Figure A2.1 (b). Note that a limitation inherent to the problem at hand is that components such as P3 and N1 cannot be aligned across trials prior to the averaging in the meta-preprocessing, since alignment would de facto require that features be identified before averaging, which is impractical due to noise levels in the individual raw ERP.

Appendix 2C: Additional data analysis results: electrode-level variation and subject-specific eigenscores

The leading first-stage electrode level eigenfunctions $\phi_1^{(2)}(t|s)$ for the ASD and TD groups (shown in Figure A2.4 (a) and (c), respectively) both have relatively flat contours, indicating that the majority of the electrode level variation is also along the longitudinal/trial dimension. While second-stage eigenfunctions indicate variability at intermediate trials (Figure A2.4 (b)) (solid, 21.7%) and boundary trials (Figure A2.4 (b)) (dashed, 8.5%) in the ASD group, most of the variability in the TD group seems to be quite uniform across trials with slightly larger variance at intermediate and later trials (Figure A2.4 (d)) (solid, 11.8%), followed by variation in early trials (Figure A2.4 (d)) (dashed, 7.4%).

The subject-specific eigenscores are estimated to predict subject-specific trajectories and to study subgroups within diagnostic groups of ASD and TD. The median (10th and 90th) percentile relative squared error, as defined in Appendix 2D, for the predicted surfaces $X_{ij}(t|s)$ are .394 (.171, .811) and .350 (.141, .650) for the ASD and TD groups, respectively. These are reasonable values for the available data with a low SNR and a small sample size.

ASD is a highly heterogeneous disorder. In order to study subgroups with similar learning patterns within the diagnostic groups, we plot the two leading subject level eigenscores from the second-stage of MD-FPCA in Figures A2.9 (a) and (b) for the ASD and TD groups (the subject level eigenscores from the first-stage decompositions are displayed in Figure A2.8). In addition, the smoothed subject-specific amplitude difference trajectories (averaged over the four electrodes) across trials at the peak location ($t = 0$) partitioned by the median of the leading scores ξ'_{i11} are given in Figures A2.9 (c) and (d) for ASD and TD groups, respectively. The variation in the leading eigenscore in the ASD group represents the largest component of the total variation in the ASD group that is observed in the intermediate and later trials. While no major clusters seem to emerge in either diagnostic group, there seems to be a small cluster around zero in the ASD group. Specifically the small cluster around zero observed in Figure A2.9 (a), falling in the below median partition, are plotted in black in Figure A2.9 (c) with minimal variation along trials, signaling little or no implicit learning. The

second group plotted in gray shows a positive condition differentiation at intermediate trials, implying implicit learning is taking place. In contrast, the variation in the leading eigenscore in the TD group represents variation at early and intermediate trials, corresponding to two groups in TD with positive (plotted in gray) and negative (plotted in black) condition differentiation. These subgroups with distinct patterns of condition differentiation at the P3 peak location within the two diagnostic groups are similar to insights gained in our previous works (Hasenstab et al. (2016)) based on a clustering of subjects according to longitudinal trends over trials, ignoring the functional dimension.

Appendix 2D: Simulation

We study the finite sample properties of MD-FPCA through simulations. We generate data from the model

$$X_{ij}(t|s) = \mu(t,s) + \eta_j(t,s) + \sum_{k=1}^{2}\sum_{k'=1}^{2} \xi'_{ikk'}\psi^{(1)}_{kk'}(s)\phi^{(1)}_{k}(t|s) + \sum_{p=1}^{2}\sum_{p'=1}^{2} \zeta'_{ijpp'}\psi^{(2)}_{pp'}(s)\phi^{(2)}_{p}(t|s) + \epsilon_{ij}(t|s),$$

containing components from both stages of the MD-FPCA decomposition for $i = 1, \ldots, n$, $j = 1, \ldots, 4$. The grids for the longitudinal and functional time points, s and t, are chosen as 50 equidistant points in $[0, 1]$. The overall mean function is $\mu(t,s) = 10\sqrt{1-(t-.5)^2-(s-.5)^2}$ with a positive concave pattern similar to the mean function of the ASD group, and the subunit-specific shifts $\eta_j(t,s)$ are set to zero for $j = 1, \ldots, 4$, for simplicity. The first- and second-stage eigenfunctions at both the subject and subunit levels are defined as $\phi^{(1)}_1(t|s) = \sqrt{2}\cos\{\pi(t-s)\}$, $\phi^{(1)}_2(t|s) = \sqrt{2}\cos\{3\pi(t-s)\}$, $\phi^{(2)}_1(t|s) = \sqrt{2}\sin\{\pi(t-s)\}$, $\phi^{(2)}_2(t|s) = \sqrt{2}\sin\{3\pi(t-s)\}$, $\psi^{(1)}_{11}(s) = \sqrt{2}\sin(2\pi s)$, $\psi^{(1)}_{12}(s) = \sqrt{2}\cos(2\pi s)$, $\psi^{(1)}_{21}(s) = \sqrt{2}\cos(4\pi s)$, $\psi^{(1)}_{22}(s) = \sqrt{2}\sin(4\pi s)$, $\psi^{(2)}_{11}(s) = \sqrt{2}\sin(2\pi s)$, $\psi^{(2)}_{12}(s) = \sqrt{2}\cos(2\pi s)$, $\psi^{(2)}_{21}(s) = \sqrt{2}\cos(4\pi s)$, and $\psi^{(2)}_{22}(s) = \sqrt{2}\sin(4\pi s)$. The second-stage eigenscores at the subject and subunit levels, ξ_{i11}, ξ_{i12}, ξ_{i21}, ξ_{i22}, ζ_{ij11}, ζ_{ij12}, ζ_{ij21}, and ζ_{ij22} are simulated from mean zero Gaussian distributions with variances $\lambda^{(1)}_{11} = 3$, $\lambda^{(1)}_{12} = 2$, $\lambda^{(1)}_{21} = 1.5$, $\lambda^{(1)}_{22} = .75$, $\lambda^{(2)}_{11} = 3$, $\lambda^{(2)}_{12} = 2$, $\lambda^{(2)}_{21} = 1.5$, $\lambda^{(2)}_{22} = .75$, respectively, corresponding to the leading first-stage eigenscores $\xi_{i1}(s)$, $\xi_{i2}(s)$, $\zeta_{ij1}(s)$, and $\zeta_{ij2}(s)$ with variance functions $\text{var}\{\xi_{i1}(s)\} = \lambda^{(1)}_1(s) = \sum_{k'=1}^{2}\lambda^{(1)}_{1k'}\psi^{(1)}_{1k'}(s)$, $\text{var}\{\xi_{i2}(s)\} = \lambda^{(1)}_2(s) = \sum_{k'=1}^{2}\lambda^{(1)}_{2k'}\psi^{(1)}_{2k'}(s)$, $\text{var}\{\zeta_{ij1}(s)\} = \lambda^{(2)}_1(s) = \sum_{p'=1}^{2}\lambda^{(2)}_{1p'}\psi^{(2)}_{1p'}(s)$, and $\text{var}\{\zeta_{ij2}(s)\} = \lambda^{(2)}_2(s) = \sum_{p'=1}^{2}\lambda^{(2)}_{2p'}\psi^{(2)}_{2p'}(s)$. The measurement error $\epsilon_{ij}(t|s)$ is simulated independently over longitudinal and functional time from a mean zero Gaussian distribution with variance $\sigma^2(s) = \{7 + \cos(2\pi s)\}/c$, where c is used to vary the SNRs of the simulated data.

We run simulations under multiple scenarios with varying SNRs, sample sizes, and sparsity levels in the longitudinal time domain. A range of measurement error variance was obtained by varying the c values from 0.5 to 50, corresponding to SNRs of roughly 1 and

100, respectively. Web Tables 1 and 2 report results from the boundary cases of $c = 0.5$ and 50. In addition, results are reported for two sample sizes $N = 30$ (similar to the implicit learning data) and $N = 100$, as well as two sparsity levels at the longitudinal dimension. As in our motivating implicit learning example, while ERP data are recorded on a dense grid of functional time points, there may be some missing values on the longitudinal time grid due to trials with low data quality. Hence for the sparse design case, we consider 40% missing data at random longitudinal time points per subject. As with our data analysis, the number of principal components for both stages of the MD-FPCA decomposition are selected to explain at least 90% of the variation. The smoothing bandwidths are selected in a preliminary simulation study using GCV and are fixed for the full simulation at (.2, .2) for first-stage overall and subunit means for both sample sizes, at (.1, .1) for the total and subunit covariance surfaces for $N = 30$, at (.02, .02) for the total and subunit covariance surfaces for $N = 100$, and at (.05, .05) for the second-stage mean and covariance surfaces for both sample sizes. We utilize relative squared error $\text{RSE}(s,t) = \{||f(t,s) - \hat{f}(t,s)||^2\}/||f(t,s)||^2$ to assess estimates of the one and two dimensional decomposition components in MD-FPCA where $||f(s)|| = \int f(s)^2 ds$ and $||f(t,s)||^2 = \int\int f(t,s)^2 dsdt$, respectively. For assessment of estimated scalar model components, we utilize mean squared error (MSE).

Figures A2.10-A2.14 display estimated model components from 200 Monte Carlo runs for a dense design with $N = 100$ and $c = .5$ (low SNR). The estimated first-stage subject and subunit level eigenfunctions with the median RSE value (Figures A2.11-A2.12 (b) and (d)) track the true quantities (Figures A2.11-A2.12 (a) and (c)) closely, similar to the overall mean estimates (Figures A2.10 (a) and (b)). The estimated first-stage subject and subunit eigenscore variance functions, $\lambda_1^{(1)}(s)$, $\lambda_2^{(1)}(s)$, $\lambda_1^{(2)}(s)$, and $\lambda_2^{(2)}(s)$ (Figures A2.11-A2.12 (e)-(f)), second-stage eigenfunctions (Figures A2.13-A2.14 (a)-(d)) and the estimated proportion of variability explained at the subject-level in the first-stage of MD-FPCA, $\rho(s)$, (Figure A2.13 (c)), are given from runs with RSE values at the 10th, 50th and 90th percentiles, overlaying the true quantities. The estimates for the second leading first-stage eigenscore variance functions, the second-stage eigenfunctions, and $\rho(s)$ track the periodicity, magnitude, and shape of the true functions. The estimated leading first-stage eigenscore variance functions capture

the periodicity and shape but tend to overestimate the magnitude of the true functions.

Web Tables 1 and 2 display the median, 10th and 90th percentiles of the RSE and MSE for estimated MD-FPCA model components over the eight simulation designs. As expected, the RSEs for the mean surface, $\mu(t,s)$, first-stage eigenfunctions, $\phi_1^{(1)}(t|s)$, $\phi_2^{(1)}(t|s)$, $\phi_1^{(2)}(t|s)$, and $\phi_2^{(2)}(t|s)$, the first-stage eigenscore variance, $\lambda_1^{(1)}(s)$, $\lambda_2^{(1)}(s)$, $\lambda_1^{(2)}(s)$, and $\lambda_2^{(2)}(s)$, and the proportion of variability displayed at the subject-level, $\rho(s)$, decrease under a denser design, with increasing sample size and SNR. The fitted surfaces for $X_{ij}(t|s)$ seem to be the most sensitive to increasing SNR. The RSE for the low SNR design corresponds to only 10% of the area under the true subject specific surface and improves substantially moving to the high SNR design. The RSEs for second-stage eigenfunctions, $\psi_{kk'}^{(1)}(s)$ and $\psi_{pp'}^{(2)}(s)$, decrease with increasing sample size under a denser design, but do not change with increasing SNR. This may be due to the fact that these quantities are estimated in the second-stage of the algorithm, based on estimated quantities in the first-stage and do not depend directly on data observed with measurement error. In addition, note that N = 100, especially N = 30, are small sample sizes for functional principal components decompositions and estimates especially for the second-stage quantities get better with increasing sample size (see the drop in the 90th percentile for the RSEs of the second-stage eigenfunctions with sample size). Finally, the median MSE for ρ decreases under a denser design, with increasing sample size and SNR.

Table A2.1: Percentiles 50% (10%, 90%) of the relative squared error (RSE) for model components based on 200 Monte Carlo runs from six different simulation designs at $c = .5$ (low SNR), $N = 30, 100$ and sparse and dense longitudinal observations. For ρ, percentiles of the mean squared error (MSE) are reported.

	Dense				Sparse			
	N=30		N=100		N=30		N=100	
Low SNR								
$\mu(t,s)$	.002	(.001, .004)	.001	(.000, .001)	.002	(.001, .004)	.001	(.000, .001)
$X_{ij}(t\|s)$	.134	(.130, .138)	.131	(.129, .134)	.140	(.136, .145)	.134	(.131, .138)
ρ(s)	.013	(.005, .032)	.003	(.001, .008)	.023	(.014, .038)	.006	(.003, .010)
ρ	.003	(.000, .011)	.001	(.000, .004)	.003	(.000, .014)	.001	(.000, .004)
$\phi_1^{(1)}(t\|s)$	.039	(.019, .087)	.028	(.012, .088)	.075	(.042, .148)	.052	(.027, .129)
$\phi_2^{(1)}(t\|s)$	.143	(.087, .334)	.044	(.024, .103)	.259	(.155, .487)	.079	(.055, .153)
$\phi_1^{(2)}(t\|s)$	.017	(.010, .031)	.013	(.008, .029)	.030	(.018, .050)	.025	(.015, .050)
$\phi_2^{(2)}(t\|s)$	.068	(.053, .122)	.022	(.016, .039)	.121	(.081, .192)	.041	(.031, .063)
$\psi_{11'}^{(1)}(s)$	.082	(.002, 1.006)	.039	(.002, .270)	.118	(.012, 1.010)	.046	(.003, .264)
$\psi_{12'}^{(1)}(s)$	.091	(.003, 1.038)	.043	(.002, .291)	.134	(.016, 1.052)	.058	(.006, .283)
$\psi_{21'}^{(1)}(s)$	.037	(.008, .362)	.013	(.005, .053)	.056	(.017, .524)	.020	(.007, .073)
$\psi_{22'}^{(1)}(s)$	.039	(.007, .273)	.007	(.001, .043)	.064	(.017, .462)	.016	(.005, .058)
$\psi_{11'}^{(2)}(s)$	.042	(.002, .253)	.014	(.001, .082)	.036	(.004, .258)	.010	(.001, .059)
$\psi_{12'}^{(2)}(s)$	.046	(.003, .271)	.015	(.001, .089)	.039	(.005, .278)	.012	(.002, .064)
$\psi_{21'}^{(2)}(s)$	.014	(.005, .052)	.007	(.004, .017)	.019	(.007, .061)	.007	(.004, .019)
$\psi_{22'}^{(2)}(s)$	.012	(.003, .047)	.003	(.001, .012)	.021	(.008, .056)	.005	(.002, .014)
$\lambda_1^{(1)}(s)$	.051	(.019, .132)	.015	(.004, .046)	.094	(.049, .175)	.031	(.017, .063)
$\lambda_2^{(1)}(s)$	.224	(.118, .370)	.018	(.006, .044)	.245	(.152, .345)	.032	(.018, .061)
$\lambda_1^{(2)}(s)$	.024	(.010, .067)	.006	(.003, .015)	.040	(.024, .075)	.013	(.008, .024)
$\lambda_2^{(2)}(s)$	.234	(.161, .325)	.009	(.004, .020)	.242	(.174, .320)	.014	(.009, .022)

Table A2.2: Percentiles 50% (10%, 90%) of the relative squared error (RSE) for model components based on 200 Monte Carlo runs from six different simulation designs at $c = 100$ (high SNR), $N = 30, 100$ and sparse and dense longitudinal observations. For ρ, percentiles of the mean squared error (MSE) are reported.

	Dense				Sparse			
	N=30		N=100		N=30		N=100	
High SNR								
$\mu(t,s)$	.002	(.001, .004)	.001	(.000, .001)	.002	(.001, .005)	.001	(.000, .001)
$X_{ij}(t\|s)$	.006	(.005, .008)	.003	(.003, .006)	.011	(.009, .015)	.006	(.005, .009)
ρ(s)	.013	(.005, .033)	.003	(.001, .009)	.022	(.014, .040)	.006	(.004, .011)
ρ	.001	(.000, .007)	.000	(.000, .002)	.001	(.000, .006)	.000	(.000, .002)
$\phi_1^{(1)}(t\|s)$	.031	(.017, .081)	.022	(.009, .083)	.070	(.040, .128)	.044	(.022, .119)
$\phi_2^{(1)}(t\|s)$	.120	(.067, .375)	.033	(.016, .093)	.221	(.134, .411)	.060	(.034, .128)
$\phi_1^{(2)}(t\|s)$	.013	(.006, .026)	.009	(.003, .026)	.020	(.012, .038)	.016	(.008, .038)
$\phi_2^{(2)}(t\|s)$	.054	(.037, .095)	.012	(.006, .028)	.081	(.054, .153)	.021	(.012, .040)
$\psi_{11'}^{(1)}(s)$	.106	(.004, .937)	.043	(.001, .244)	.101	(.009, .931)	.034	(.003, .359)
$\psi_{12'}^{(1)}(s)$	.121	(.005, .972)	.048	(.002, .263)	.118	(.015, .979)	.040	(.006, .381)
$\psi_{21'}^{(1)}(s)$	.041	(.008, .618)	.015	(.005, .052)	.057	(.014, .492)	.022	(.007, .067)
$\psi_{22'}^{(1)}(s)$	.039	(.004, .509)	.009	(.001, .043)	.076	(.019, .453)	.015	(.005, .061)
$\psi_{11'}^{(2)}(s)$	.027	(.001, .257)	.011	(.000, .064)	.044	(.003, .306)	.008	(.001, .053)
$\psi_{12'}^{(2)}(s)$	.029	(.001, .277)	.012	(.001, .070)	.049	(.005, .328)	.011	(.002, .057)
$\psi_{21'}^{(2)}(s)$	.014	(.007, .040)	.007	(.005, .014)	.020	(.008, .052)	.007	(.004, .015)
$\psi_{22'}^{(2)}(s)$	.008	(.001, .032)	.002	(.001, .009)	.015	(.004, .050)	.003	(.001, .011)
$\lambda_1^{(1)}(s)$	.047	(.018, .136)	.015	(.004, .048)	.090	(.050, .167)	.030	(.017, .069)
$\lambda_2^{(1)}(s)$	.245	(.109, .374)	.015	(.004, .047)	.253	(.149, .380)	.028	(.017, .052)
$\lambda_1^{(2)}(s)$	.023	(.007, .056)	.006	(.002, .017)	.035	(.021, .064)	.011	(.007, .023)
$\lambda_2^{(2)}(s)$	.246	(.169, .325)	.007	(.002, .018)	.246	(.174, .334)	.011	(.007, .022)

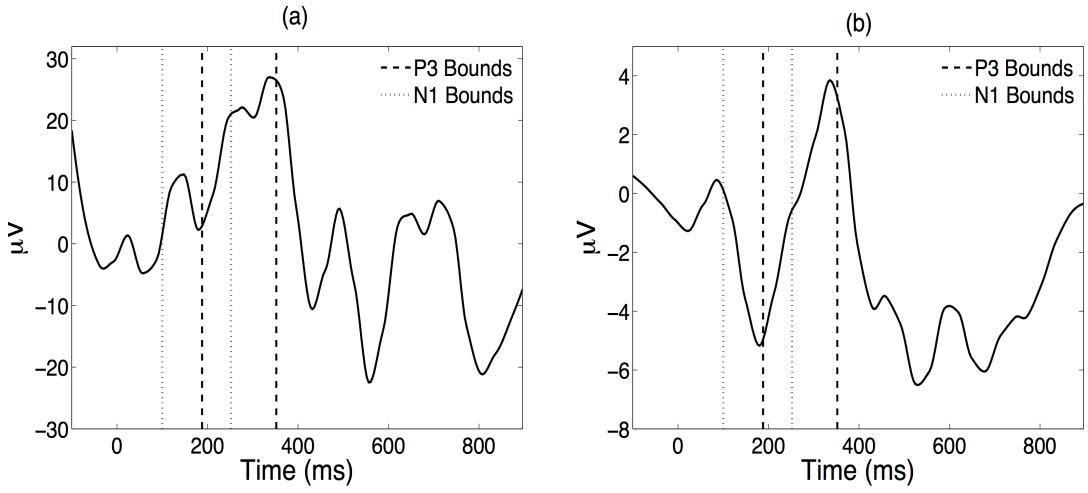


Figure A2.1: (a) ERP waveform from a single subject, condition, electrode and trial in the right frontal region of the scalp. (b) The average of the first 30 consecutive ERP waveforms for the same subject, electrode and condition after preprocessing.

Appendix 3A: Spectral PCA

We consider 11 scalp regions (left and right temporal (LT, RT) and left, right, and middle, frontal, central, and posterior regions (LF, RF, MF, LC, RC, MC, LP, RP, MP), each containing 4 to 7 electrodes, identified by our collaborators to be of interest in the word segmentation paradigm. Let $X_{dirjs}(t)$ denote the locally stationary mean zero time series observed on subject i, $i = 1, \ldots, n_d$, from group d, $d = 1, \ldots, D$, at electrode j, $j = 1, \ldots, J_r$, within region r, $r = 1, \ldots, R$, for segment s, $s = 1, \ldots, S$, at a sampling rate of U across discretized time t, $t = 0, \pm 1, \ldots, \pm U/2$. For each observation $X_{dirjs}(t)$, FFT is performed to obtain the set of Fourier coefficients $a_{dirjs}(\omega) = U^{-1/2} \sum_{t=-U/2+1}^{U/2} X_{dirjs}(t) e^{-2\pi i \omega t}$ across frequency $\omega = \{u + (U/2 - 1)\}/U$, $u = -U/2 + 1, \ldots, U/2$. Using the Fourier coefficients obtained, the raw $J_r \times J_r$ periodogram matrix $\mathbf{I}_{dirs}(\omega)$ is computed, where the (j, j')-th entry is equal to $a_{dirjs}(\omega)\bar{a}_{dirj's}(\omega)$, with $\bar{a}_{dirj's}(\omega)$ denoting the conjugate transpose of $a_{dirj's}(\omega)$. Next, kernel smoothing is applied to each (j, j')-th entry in the $U/2 + 1$ periodogram matrices across frequency ω, using a modified Daniell kernel to obtain a consistent estimate of the spectral density, denoted by $\tilde{\mathbf{I}}_{dirs}(\omega)$. To guarantee that the resulting smoothed periodogram matrix is non-negative definite, generalized cross validation (GCV) is used to

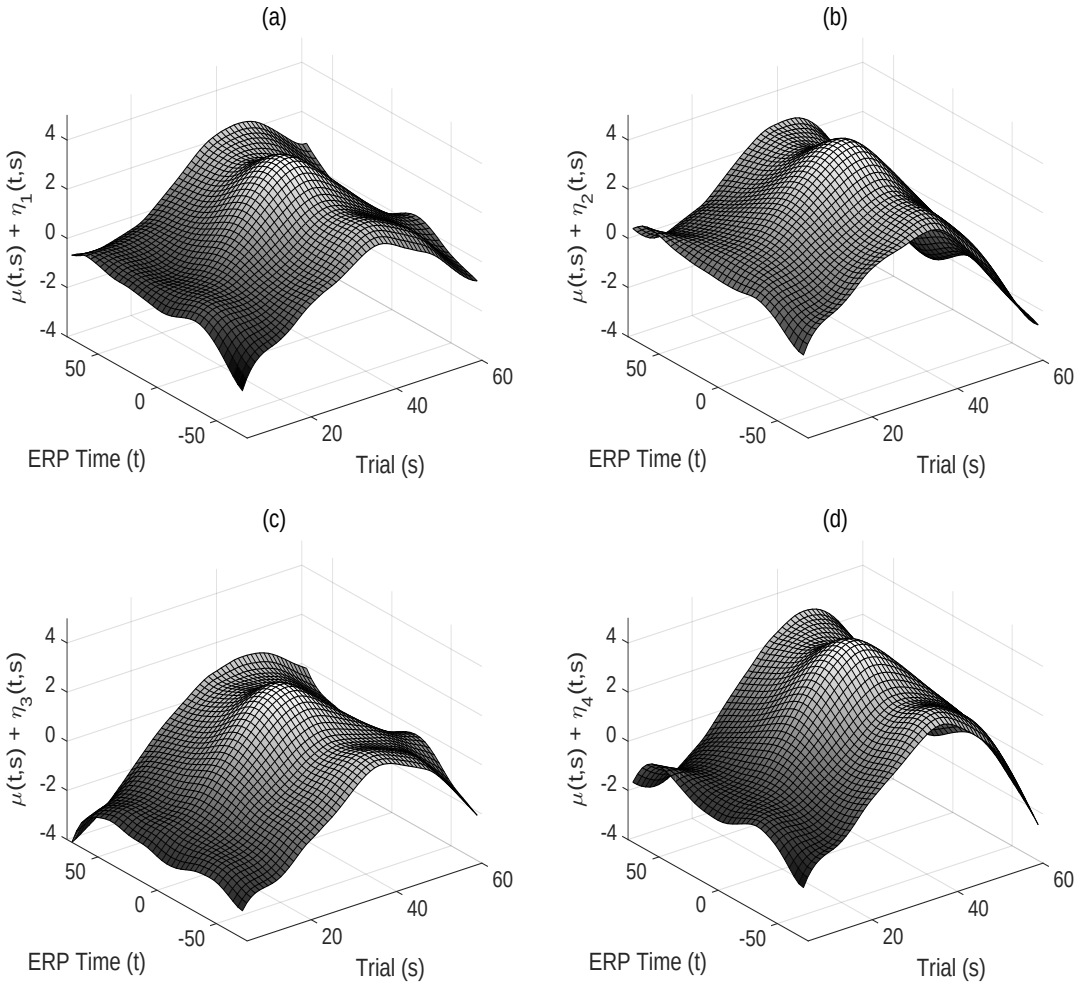


Figure A2.2: Estimated electrode-specific mean surfaces for the ASD group.

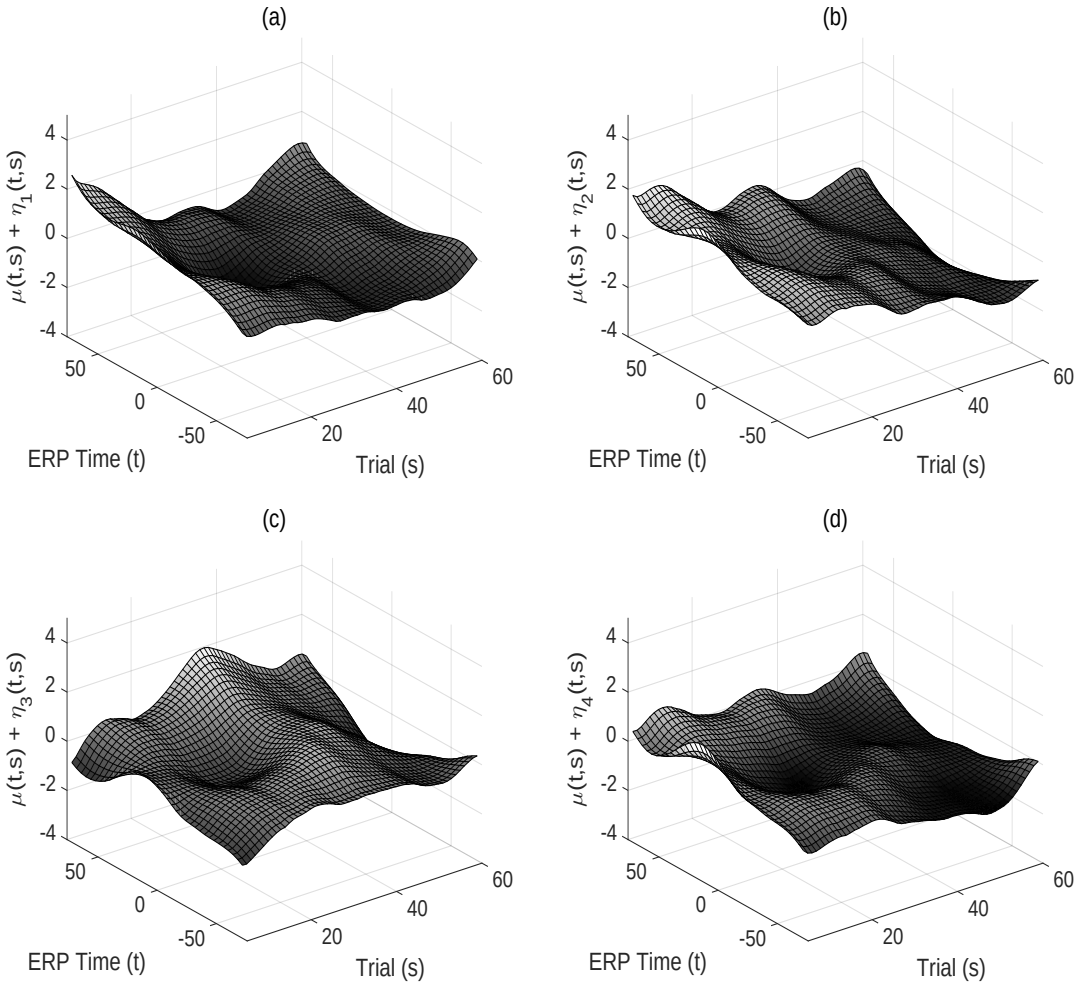


Figure A2.3: Estimated electrode-specific mean surfaces for the TD group.

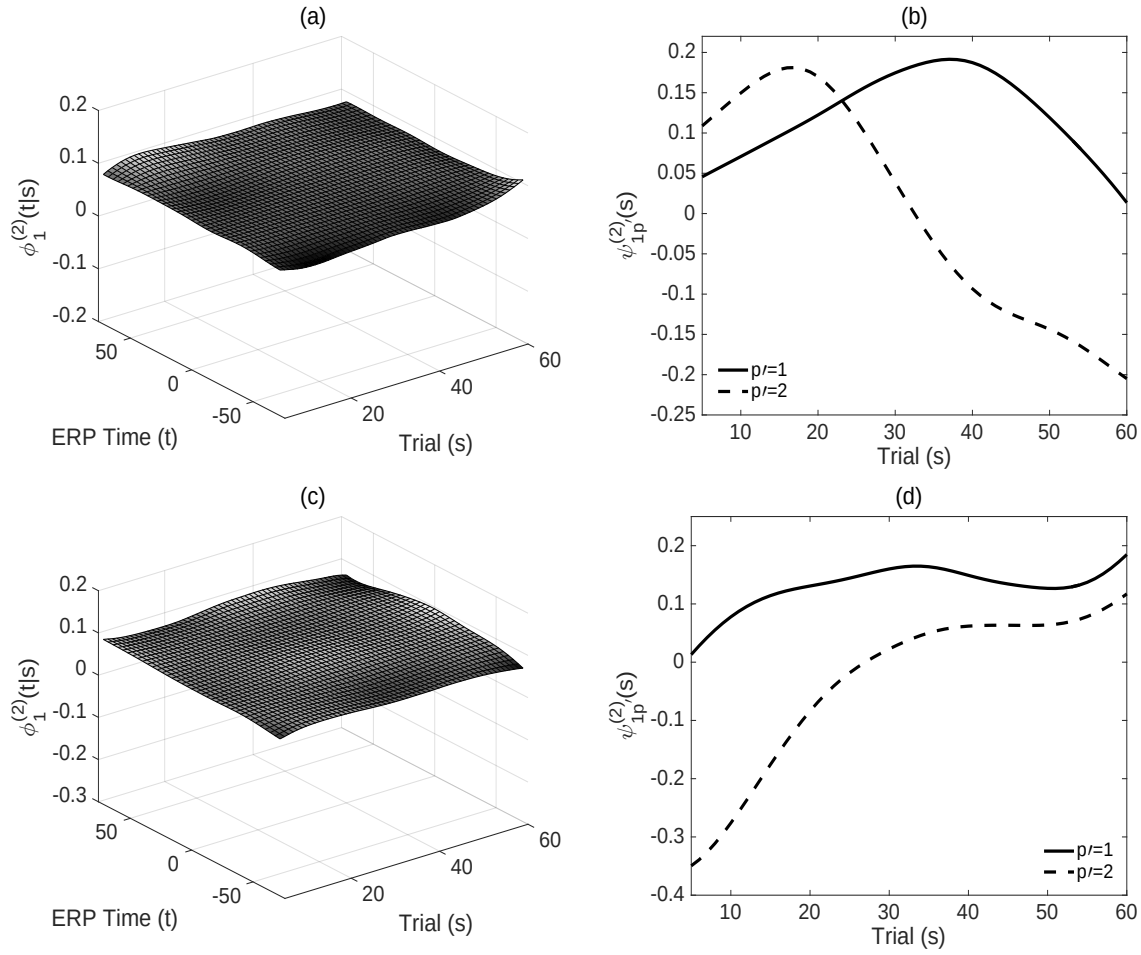


Figure A2.4: (a, c) Estimated leading electrode level first-stage eigenfunctions, $\{\phi_1^{(2)}(t|s)\}$ and (b, d) estimated leading electrode level second-stage eigenfunctions, $\{\psi_{1p'}^{(2)}(s)\}$, $p' = 1, 2$, for the ASD and TD groups, respectively.

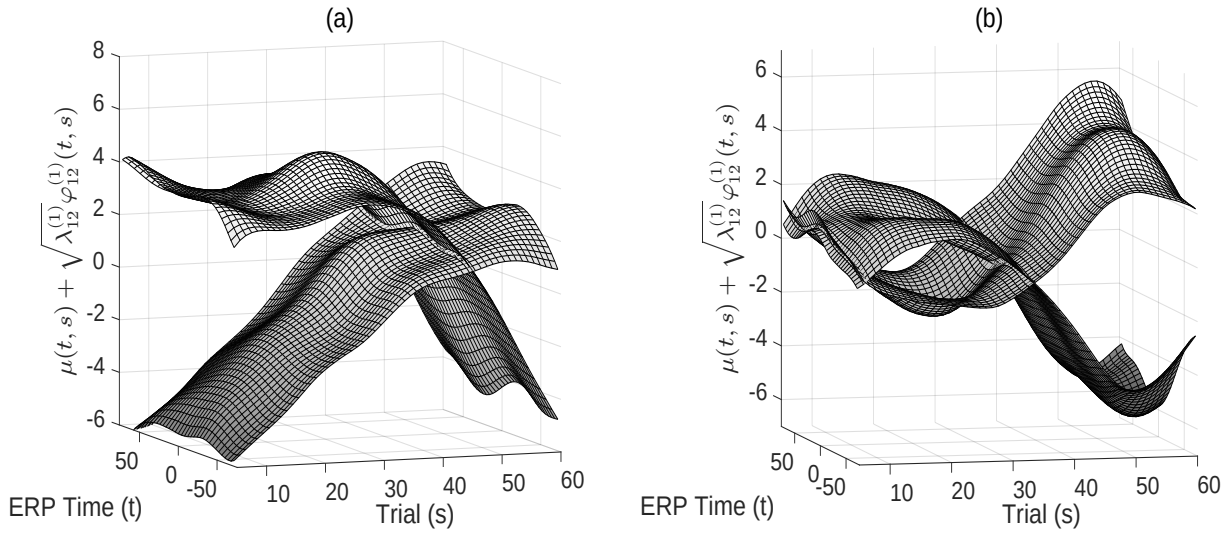


Figure A2.5: (a-b) Estimated surface intervals, $\mu(t,s) \pm \sqrt{\lambda_{12}^{(1)}}\varphi_{12}^{(1)}(t,s)$, for the ASD and TD groups, respectively.

select a common bandwidth for the smoothing of all the terms in the raw periodogram matrices. For each frequency ω, a principal components decomposition is performed on the smoothed periodogram matrix $\tilde{\mathbf{I}}_{dirs}(\omega)$ and the leading eigenvalue, denoted by $\lambda_{dirs1}(\omega)$, is extracted. We normalize the leading eigenvalue $\lambda_{dirs1}(\omega)$ across the frequency domain such that $\int \lambda_{dirs1}(\omega)d\omega = 1$, defined as principal power that is comparable across brain regions, similar to the relative power commonly calculated in EEG spectral analysis. While the leading eigenvector summarizes the contribution of the specific electrodes to total variation in a given frequency within a region, the principal power, as the normalized leading eigenvalue, represents the common variation in that frequency across the electrodes (relative to variation in other frequencies) in a given scalp region along the direction of the leading eigenvector.

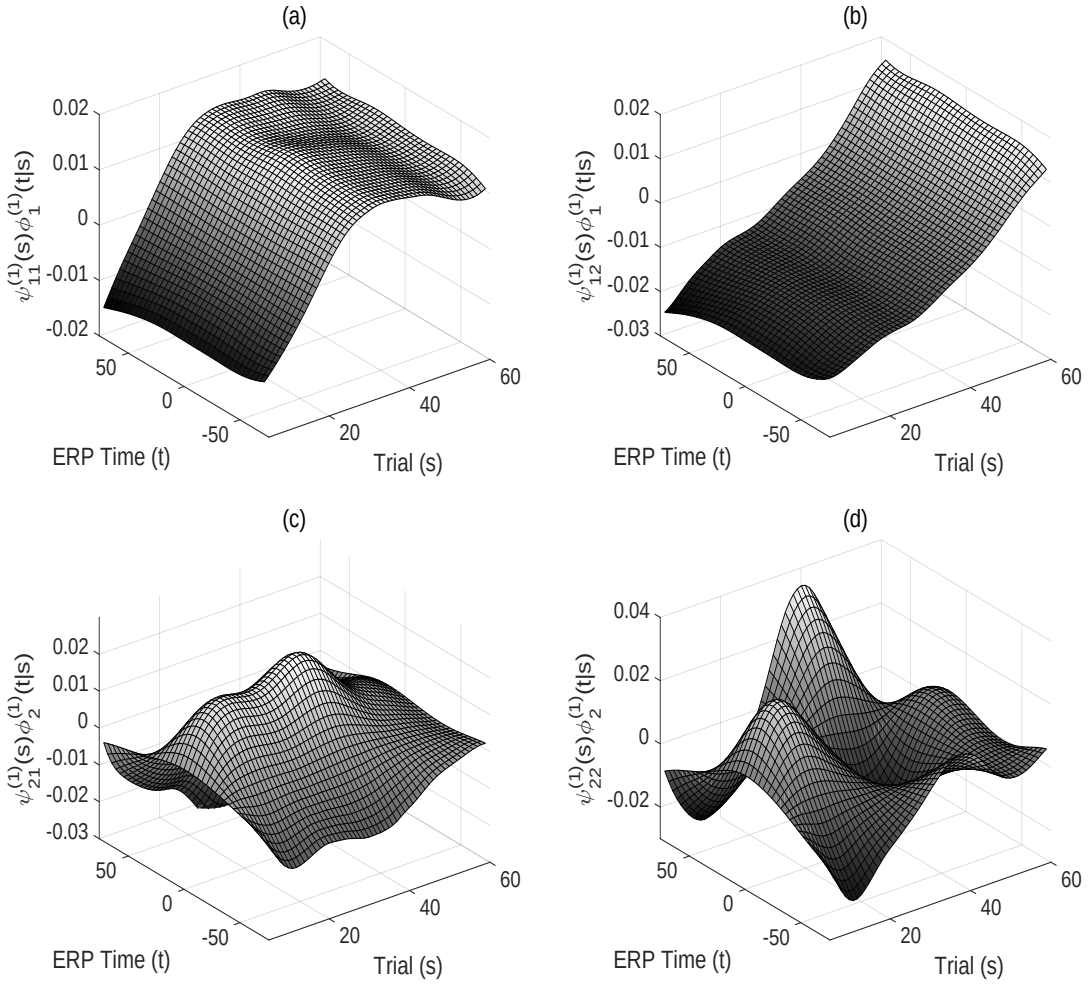


Figure A2.6: Estimated subject level principal surfaces $\varphi^{(1)}_{kk'}(t,s) = \psi^{(1)}_{kk'}(s)\phi^{(1)}_{k}(t|s)$ for (a) $k = 1, k' = 1$, (b) $k = 1, k' = 2$, (c) $k = 2, k' = 1$ and (d) $k = 2, k' = 2$ for the ASD group.

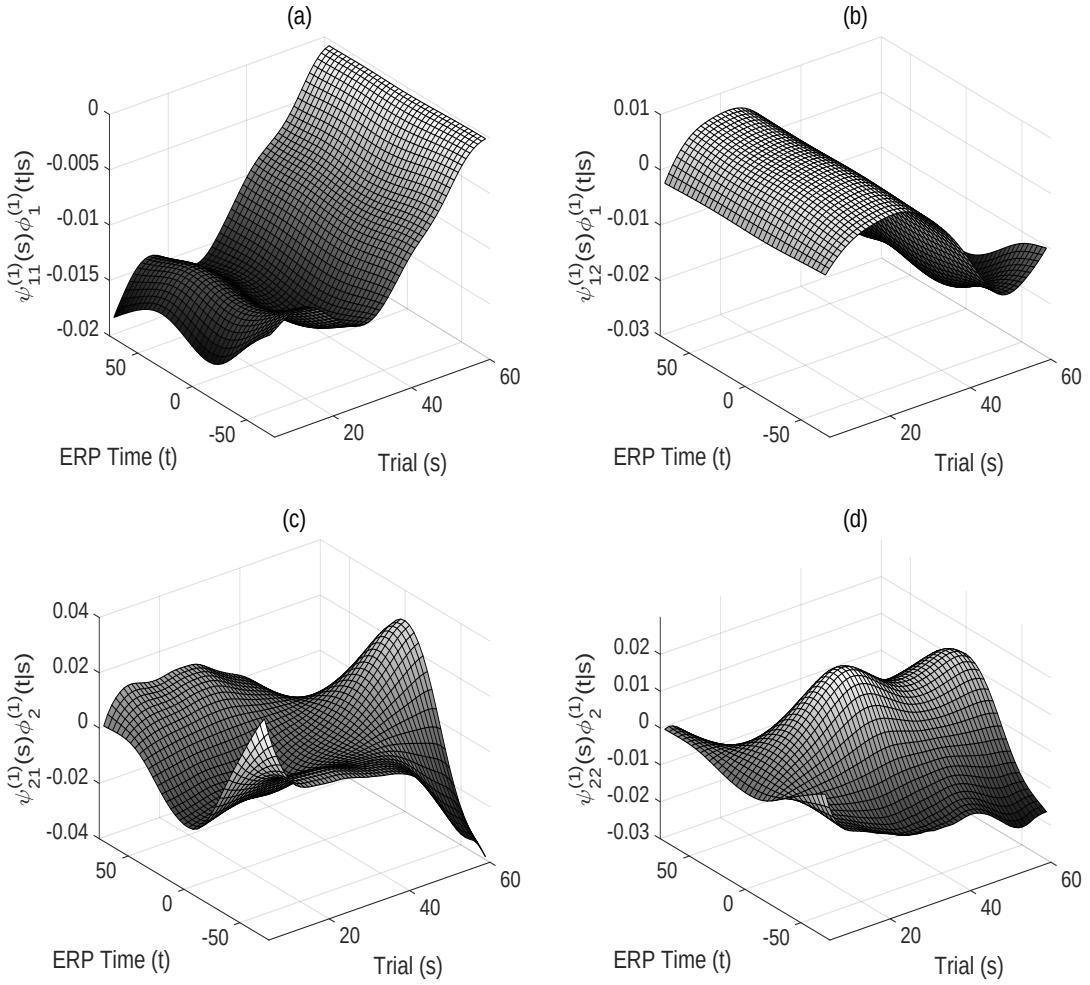


Figure A2.7: Estimated subject level principal surfaces $\varphi^{(1)}_{kk'}(t,s) = \psi^{(1)}_{kk'}(s)\phi^{(1)}_{k}(t|s)$ for (a) $k = 1, k' = 1$, (b) $k = 1, k' = 2$, (c) $k = 2, k' = 1$ and (d) $k = 2, k' = 2$ for the TD group.

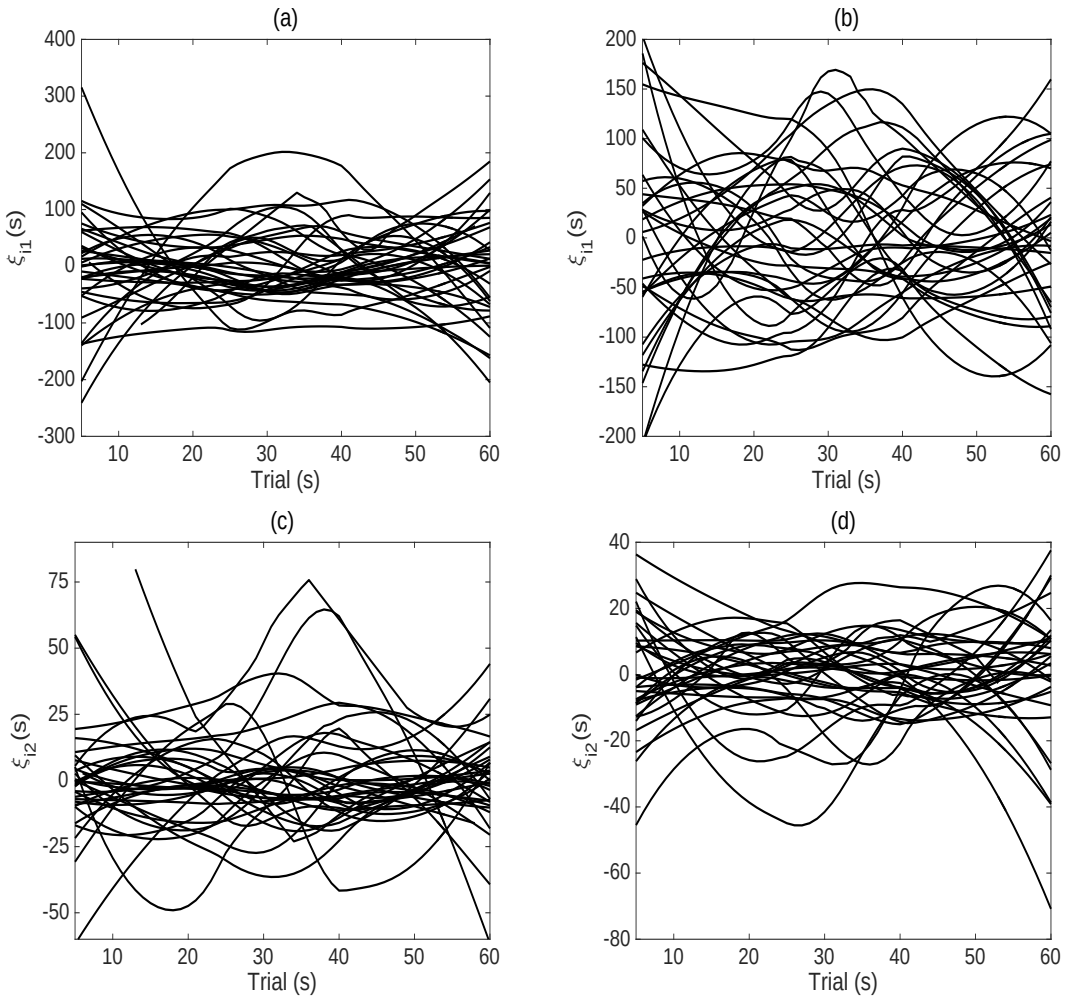


Figure A2.8: Estimated subject-specific eigenscores from the first-stage decompositions for the ASD ((a, c)) and TD ((b, d)) groups.

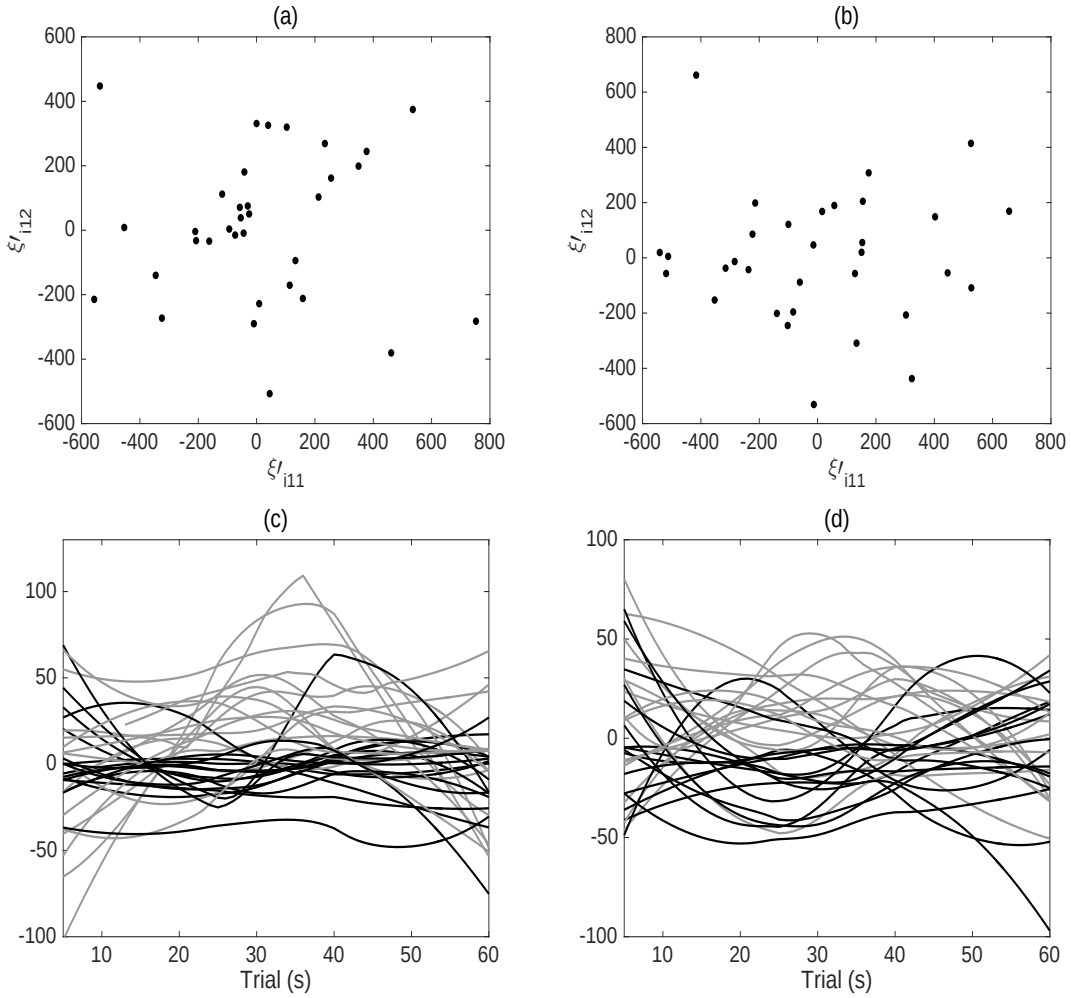


Figure A2.9: (a-b) The two leading subject level eigenscores from the second-stage decompositions for the ASD and TD groups, respectively. (c-d) The smoothed subject-specific amplitude difference trajectories at P3 peak location $t = 0$, partitioned by the median of the leading scores ξ'_{i11} for the ASD and TD groups, respectively.

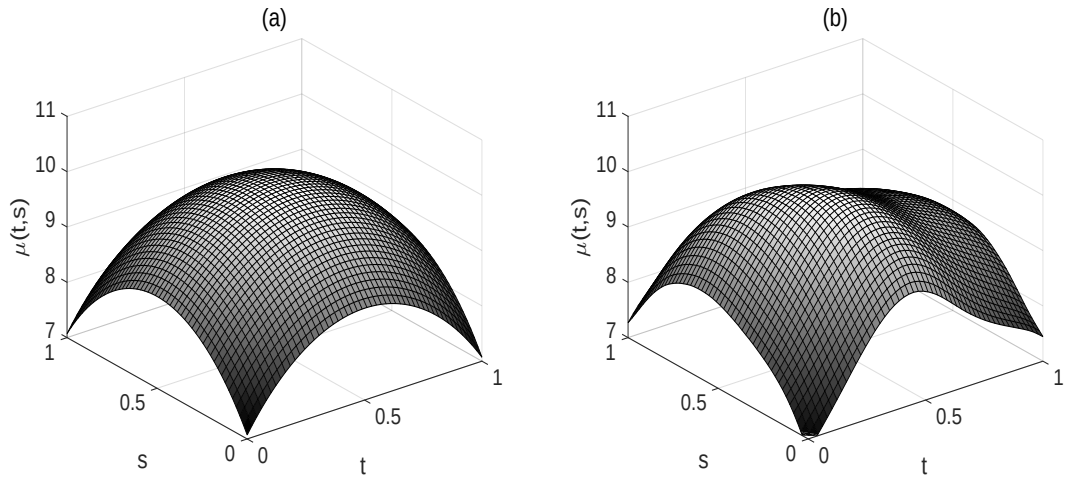


Figure A2.10: The true (a) and estimated (b) mean surfaces, $\mu(t, s)$, based on 200 Monte Carlo runs from the dense design scenario at $N = 100$ and low SNR. The estimated function corresponds to the Monte Carlo run with RSE value at the 50th percentile.

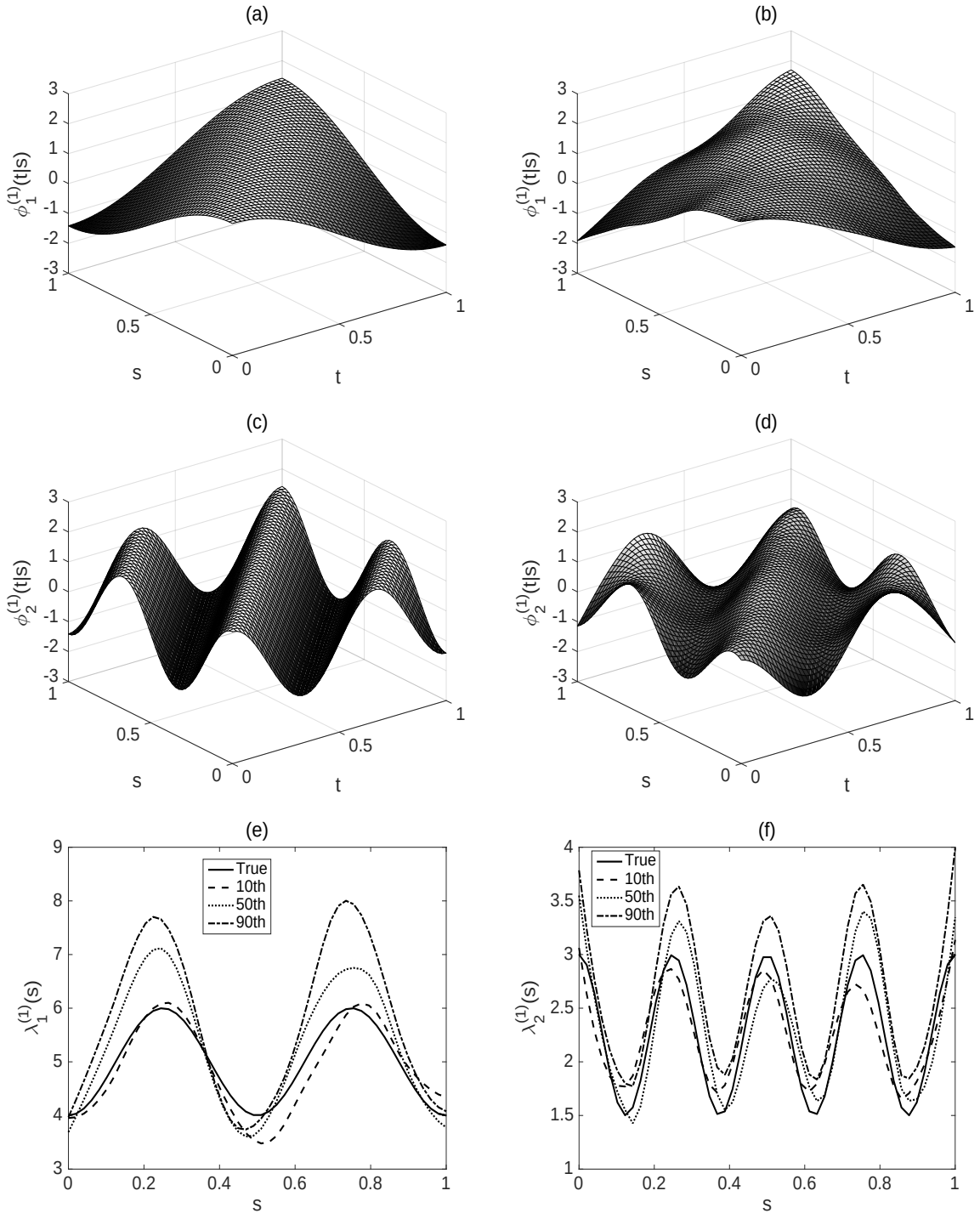


Figure A2.11: The true and estimated model components based on 200 Monte Carlo runs from the dense design scenario at $N = 100$ and low SNR. The true ((a, c)) and estimated ((b, d)) first-stage subject level eigenfunctions are displayed in the first two rows. The estimated functions correspond to the Monte Carlo run with RSE value at the 50th percentile. The true and estimated first-stage subject level eigenscore variance functions are given in (e) and (f) from runs with RSE values at the 10th, 50th and 90th percentiles.

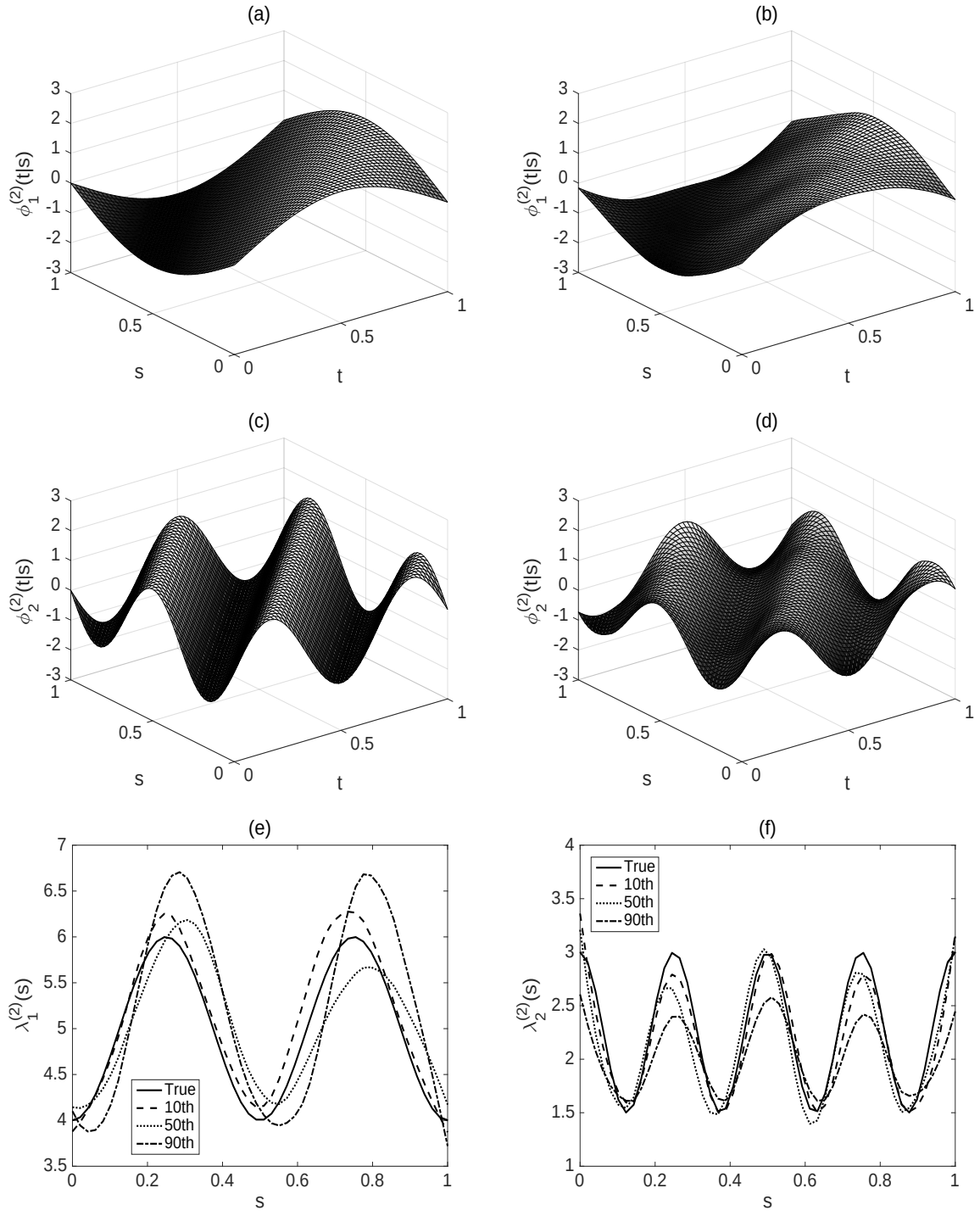


Figure A2.12: The true and estimated model components based on 200 Monte Carlo runs from the dense design scenario at $N = 100$ and low SNR. The true ((a, c)) and estimated ((b, d)) first-stage subunit level eigenfunctions are displayed in the first two rows. The estimated functions correspond to the Monte Carlo run with RSE value at the 50th percentile. The true and estimated first-stage subunit level eigenscore variance functions are given in (e) and (f) from runs with RSE values at the 10th, 50th and 90th percentiles.

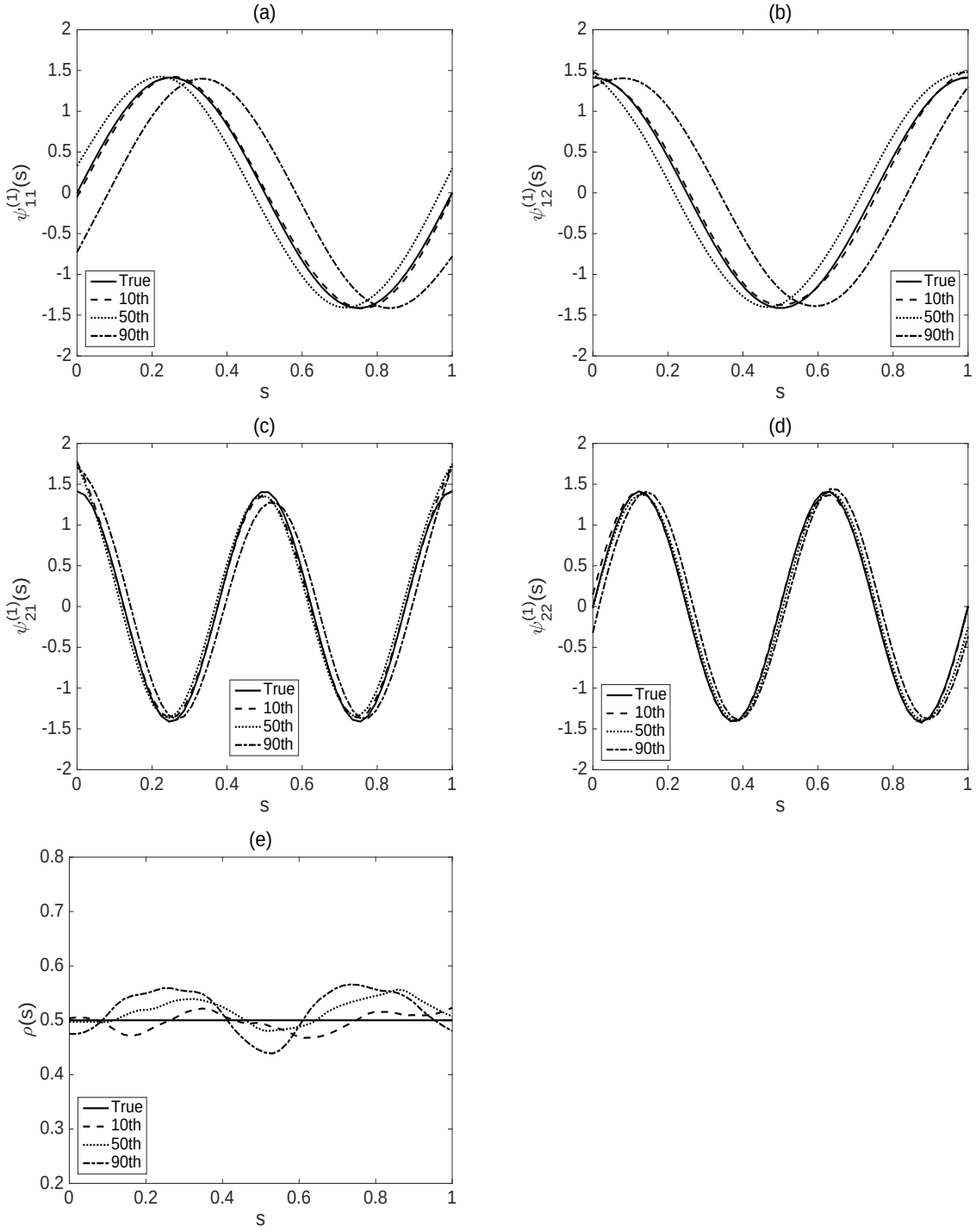


Figure A2.13: The true and estimated model components based on 200 Monte Carlo runs from the dense design scenario at $N = 100$ and low SNR. Estimated model components are given from runs with RSE values at the 10th, 50th and 90th percentiles. Displayed are the ((a-d)) second-stage subject level eigenfunctions and (e) the proportion of variability explained at the subject level, $\rho(s)$, in the first-stage decompositions.

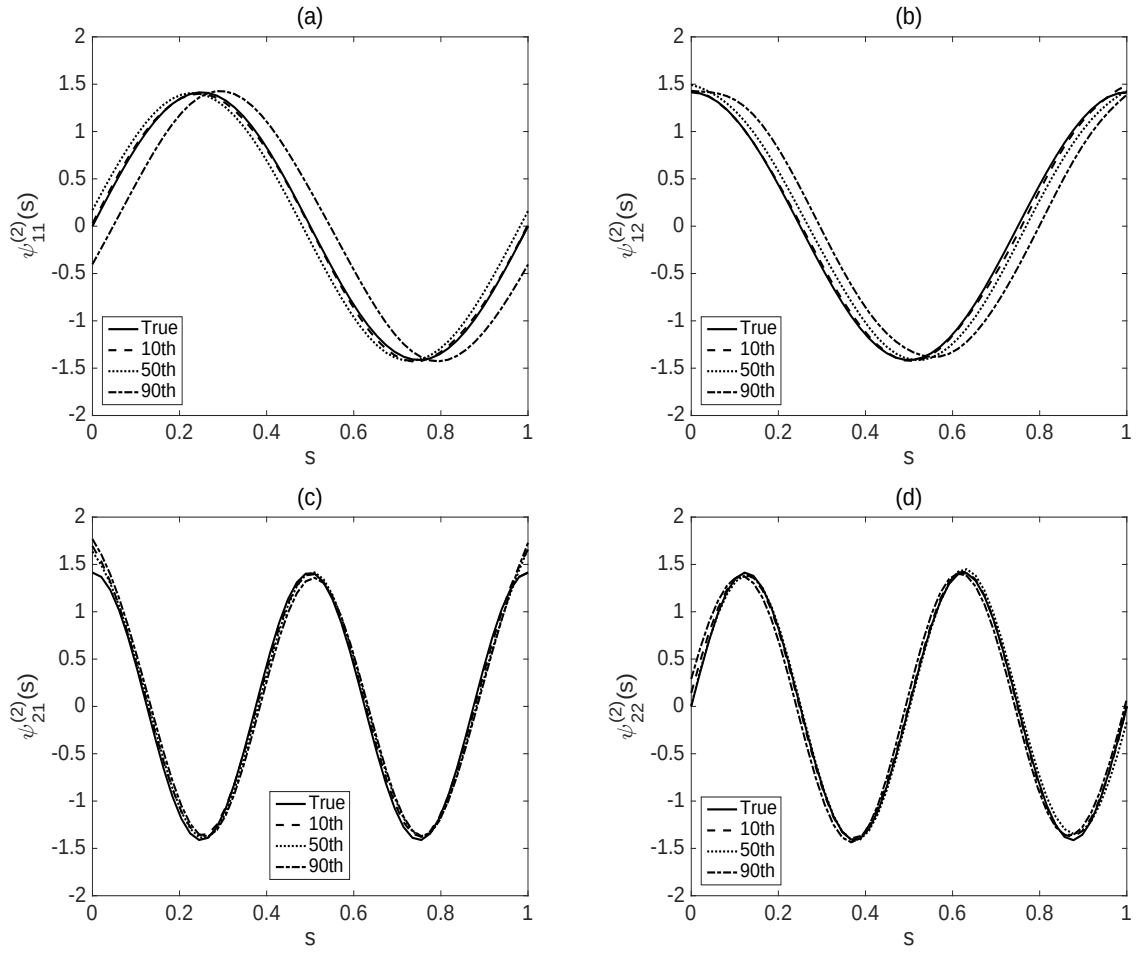


Figure A2.14: The true and estimated model components based on 200 Monte Carlo runs from the dense design scenario at $N = 100$ and low SNR. Estimated model components are given from runs with RSE values at the 10th, 50th and 90th percentiles. Displayed ((a-d)) are the second-stage subunit level eigenfunctions.

Appendix 3B: Estimation of model components

Presented below are details of the estimation algorithm deferred from Section 3.2 due to their utilization of well-established mean, covariance, and eigencomponent estimation procedures.

1. Estimation of the Fixed Effects:

(a-b) The overall mean and group-region shifts, $\mu(\omega, s)$ and $\eta_d(r, \omega, s)$, are estimated using bivariate penalized splines via the sandwich smoother of Xiao et al. [2013]. Given possible imbalances in group sample sizes, the estimated overall mean function is taken as a point-wise average of the smoothed mean functions within each group.

2. Estimation of the Marginal Covariances and Measurement Error Variance:

(a) After the functional fixed effects are estimated, let $\widehat{Y}^c_{di}(r, \omega, s) = Y_{di}(r, \omega, s) - \hat{\mu}(\omega, s) - \hat{\eta}_d(r, \omega, s)$ denote the centered data. We estimate the functional marginal covariance surface by first obtaining the pooled sample covariance, $\widehat{\Sigma}_{d,\Omega}(\omega, \omega') = \{\sum_{i=1}^{n_d} \sum_{r=1}^{R} \sum_{s \in \mathcal{S}} \widehat{Y}^c_{di}(r, \omega, s) \widehat{Y}^c_{di}(r, \omega', s)\} / (n_d R |\mathcal{S}|)$, where $|\mathcal{S}|$ denotes the total number of grid points observed in $\mathcal{S}$. Note that the sample covariance is pooled to accelerate computation. Given that the diagonals of the pooled sample covariance $\widehat{\Sigma}_{d,\Omega}(\omega, \omega')$ are inflated by the measurement error variance σ^2_d, the estimated functional marginal covariance surface $\widetilde{\Sigma}_{d,\Omega}(\omega, \omega')$ is obtained by smoothing the off-diagonal elements of the pooled sample covariance $\widehat{\Sigma}_{d,\Omega}(\omega, \omega')$ over frequencies (ω, ω') using bivariate penalized splines with smoothing parameters selected by restricted maximum likelihood (REML) as proposed by Goldsmith et al. [2012]. The estimated longitudinal marginal covariance surface, $\widetilde{\Sigma}_{d,\mathcal{S}}(s, s')$, is obtained similarly. If missingness in the data is not symmetric along either the longitudinal or functional dimensions, the proposed smoothing procedure can be adapted by incorporating weights to adjust for the number of observations at each grid point in the pooled marginal sample covariance.

(b) The measurement error variance σ^2_d is estimated by the difference of two smooths, the one-dimensional smooth fitted only to diagonal elements of the pooled marginal covariance and the diagonal elements of $\widetilde{\Sigma}_{d,\Omega}(\omega, \omega')$. The resulting measurement error variance estimate is denoted by $\hat{\sigma}^2_{d,\Omega}$. The estimated measurement error variance, $\hat{\sigma}^2_{d,\mathcal{S}}$, based on the longitudinal

covariance is obtained similarly. The two measurement error variance estimates, $\hat{\sigma}^2_{d,\Omega}$ and $\hat{\sigma}^2_{d,\mathcal{S}}$, are averaged to derive an initial estimate of the measurement error variance $\hat{\sigma}^2_d$, utilized below in estimation of the regional marginal covariance matrix.

(c) Unlike the marginal covariance surfaces, it is not possible to adjust for measurement error by smoothing over non-diagonal elements when estimating the regional marginal covariance matrix. Hence we utilize our initial estimate, $\hat{\sigma}^2_d$, of the measurement error variance to de-noise the pooled regional marginal covariance matrix. More specifically, the estimated regional marginal covariance matrix, $\widetilde{\Sigma}_{d,\mathcal{R}}$, is obtained by adjusting the diagonal elements of the pooled sample covariance, $(\widehat{\Sigma}_{d,\mathcal{R}})_{r,r'} = \{\sum_{i=1}^{n_d} \sum_{s\in\mathcal{S}} \sum_{\omega\in\Omega} \widehat{Y}^c_{di}(r,\omega,s)\widehat{Y}^c_{di}(r',\omega,s)\}/(n_d|\Omega||\mathcal{S}|)$, where $(\widetilde{\Sigma}_{d,\mathcal{R}})_{r,r'} = (\widehat{\Sigma}_{d,\mathcal{R}})_{r,r'}$ for $r \neq r'$ and $(\widetilde{\Sigma}_{d,\mathcal{R}})_{r,r'} = (\widehat{\Sigma}_{d,\mathcal{R}})_{r,r'} - \hat{\sigma}^2_d$ for $r = r'$.

3. Estimation of the Marginal Eigencomponents:

(a-b) The estimated marginal eigenvalue, eigenvector/function pairs $\{\hat{\tau}_{dk,\mathcal{R}}, \hat{\mathrm{v}}_{dk}(r)\}$, $\{\hat{\tau}_{d\ell,\Omega}, \hat{\phi}_{d\ell}(\omega)\}$, and $\{\hat{\tau}_{dm,\mathcal{S}}, \hat{\psi}_{dm}(s)\}$ are obtained by eigendecompositions of the marginal covariance matrix and surfaces, respectively [Park and Staicu, 2015, Chen et al., 2016]. Unlike previous works, the marginal eigendecompositions for HPCA utilize both functional and vector PCA. Initial values for the number of eigencomponents, K, L, and M, included in the truncated HPCA are chosen to explain at least 90% of the variation in the respective marginal covariance decompositions.

The runtime of a single Monte Carlo run for the longitudinally dense data described in Appendix E with a sample size of $n_d = 15$ is 9 minutes and 26 seconds on a 2.4 GHz 6-Core Intel Xeon processor operating R 3.3.2 [R Core Team, 2018].

Appendix 3C: Algorithm for the bootstrap test

Presented below is the algorithm for the bootstrap test proposed in Section 3.3.

Algorithm: *Bootstrap Test*:

For a fixed region, $r \in \{1, \ldots, R\}$, perform the following algorithm:

1. Generate B parametric bootstrap samples with sample sizes in each group identical to the observed data.
2. For the bth parametric bootstrap sample, calculate the test statistic

$$T_r^b = \sqrt{\sum_{d=1}^{D} \int \int \{\hat{\eta}_d^b(r, \omega, s) - \hat{\eta}^b(r, \omega, s)\}^2 d\omega ds},$$

   where $\hat{\eta}_d^b(r, \omega, s)$ and $\hat{\eta}^b(r, \omega, s)$ are both estimated based on the bth bootstrap sample.
3. Use $(1/B) \sum_{b=1}^{B} \mathbb{I}(T_r^b > T_r)$ to estimate the p-value where $\mathbb{I}(\cdot)$ denotes the indicator function and T_r is the test statistic from the original sample.

Appendix 3D: Application to the word segmentation data

D.1: Data Structure and Methods

Prior to FFT, the linear trend is removed and the EEG signal is normalized at each segment and electrode to facilitate comparison of the resulting spectral densities. Standard EEG pre-processing steps include bad electrode replacement, automatic segment rejection (QEEG (4.0) Auto Artifact Detection, Netstation 4.4.5), manual segment rejection due to muscle activity, saccades or eye movements, and re-referencing to the average scalp signal. Since the segment rejection varies from person to person, the support of the longitudinal dimension is subject-specific, but remains the same across electrodes within a subject. The mvASD children had the lowest number of good segments after the pre-processing steps (as few as 12), leading to data sparsity in the longitudinal dimension (Figure A3.1(a)). We exclude the first 4 segments for which only one subject in the TD group has valid data and, so the range of the longitudinal domain is $\mathcal{S} = \{5 : 140\}$. One mvASD subject was removed prior to analysis due to having valid data in less than ten segments.

D.2: Data Analysis Results

In order to asses the weak separability assumption on the covariance process, we compare the first and second leading eigenvectors and eigenfunctions along each of the three dimensions of the observed data across fixed slices of the other two dimensions for all three diagnostic groups of TD, vASD and mvASD children (Figures A3.2 and A3.3, second columns). For ease of interpretation, we also display leading eigenvectors and eigenfunctions of the regional, functional and longitudinal marginal covariances, depicting modes of marginal variation along the three dimensions (Figures A3.2 and A3.3, first columns). Table A3.1 displays the FVE by the eigencomponents for the regional, functional and longitudinal marginal covariances that explain at least 90% FVE in all three diagnostic groups. The leading five, four, and five regional marginal eigenvectors, five functional marginal eigenfunctions, and the two longitudinal marginal eigenfunction are found to explain $FVE_{dG} = 0.934,\ 0.944,$

and 0.925 of the total variation in the TD, vASD, and mvASD groups, respectively.

In the functional dimension, the first leading marginal eigenfunction $\phi_{d1}(\omega)$ (Figure A3.2(a)) displays increasing variation with increasing frequency for all diagnostic groups, with the peak observed in the beta and gamma bands (15-50 Hz). The second leading marginal eigenfunction $\phi_{d2}(\omega)$ (Figure A3.3(a)) displays peak variation mostly in the beta band (15-32 Hz). The first two eigenfunctions together explain at least 70% of the variation in the functional marginal covariance in all three diagnostic groups. The eigencomponents estimated at fixed slices of the other two dimensions have similar shapes within diagnostic groups and are also similar to the eigencomponents of the functional marginal covariance. Figures A3.2 and A3.3(b) illustrate these effects using slices corresponding to the LT region at segments $s = 40, 70, 100$.

In the longitudinal dimension, the first leading marginal eigenfunction $\psi_{d1}(s)$ (Figure A3.2(c)) displays constant variation across segments and explains more than 85% of the variation in the longitudinal marginal covariance for all diagnostic groups. The second leading longitudinal marginal eigenfunction $\psi_{d2}(s)$ (Figure A3.3(c)) oscillates around zero explaining less than 7% of the longitudinal marginal covariance for all diagnostic groups. As illustrated in the LT region at fixed frequencies $\omega = 30, 40, 50$, the estimated eigencomponents have similar shapes within diagnostic groups and with the eigencomponents of the longitudinal marginal covariance (Figures A3.2 and A3.3(d)).

In the regional dimension, the weights of the first leading marginal eigenvector $\mathrm{v}_{d1}(r)$ (Figure A3.2(e)) are uniform across scalp locations in all the diagnostic groups, implying equal variation, while the weights of the second leading marginal eigenvector $\mathrm{v}_{d2}(r)$ (Figure A3.3(e)) are highest for the MP and RP regions, and MF and RF regions for the TD and vASD groups, respectively. In the mvASD group, the leading components signal a contrast between LT and MC regions. The first two regional marginal eigenvectors together explain at least 70% of the variation in the regional marginal covariance. The leading eigenvectors obtained at fixed slices of $(\omega, s) = (30, 40), (30, 70), (30, 100)$ put reasonably similar weights on regions within diagnostic groups, at least for the first leading eigenvector (Figures A3.2 and A3.3(f)). However, the estimates may not be very reliable, based on sample sizes as low

as seven subjects per group, due to conditioning on the sparse longitudinal dimension. In addition to the visual inspection of the shapes of the eigenvectors and eigenfunctions along each dimension at fixed slices of the other two dimensions, we also performed a likelihood ratio test on the correlation structure of the subject-specific scores to check the weak separability assumption. Specifically, we test for an independence structure versus an alternative of a heterogenous compound-symmetric dependence structure, with potentially different variances on the diagonal. We chose the compound symmetric dependence as the alternative rather than unstructured to maximize the stability of the estimates given the number of covariance parameters required. The likelihood ratio test fails to reject the null hypothesis of the independence structure, corresponding to the weak separability assumption (p= $0.72, 0.79, 0.62$ for the TD, vASD and mvASD groups respectively), consistent with the results of our visual inspection. Weak separability of the covariance process is maintained after collapsing the longitudinal dimension given that the marginal longitudinal eigenfunctions are simply integrated out.

While there is sufficient evidence to justify the assumption of weak separability in the observed data, the question remains whether the even more restrictive property of strong separability can be assumed. Recall that strong separability requires that the entire covariance structure, not only the directions of variation along each dimension, is the same up to a constant across fixed slices of the other dimensions. The estimated functional covariance surfaces, $\Sigma_{d,\Omega}(\omega, \omega')$, are displayed for TD, vASD, and mvASD groups in descending rows for the RT region for fixed $s = 100$ (Figure A3.4 (a, c, e)) and RF region for fixed $s = 70$ (Figure A3.4 (b,d,f)). Visual inspection of the covariance surfaces within each group shows that the covariance structure in the functional dimension is not the same up to a constant across fixed slices of the regional and longitudinal dimensions, suggesting that the assumption of strong separability is too stringent for the observed data.

Appendix 3E: Simulation

We studied the finite sample properties of the proposed HPCA and the bootstrap test for group-level inference via extensive simulations. We generated data from the model

$$Y_{di}(r,\omega,s) = \mu(\omega,s) + \eta_d(r,\omega,s) + \sum_{k=1}^{3}\sum_{\ell=1}^{3}\sum_{m=1}^{3} \xi_{di,k\ell m}\mathrm{v}_{dk}(r)\phi_{d\ell}(\omega)\psi_{dm}(s) + \epsilon_{di}(r,\omega,s), \quad \text{(A.1)}$$

with $K = L = M = 3$ for $d = 1, 2$, and $i = 1, \ldots, n_d$. The functional and longitudinal grids were chosen as 50 equidistant points in $[0, 1]$, where data were generated for a total of $R = 9$ regions. The overall mean function was $\mu(\omega, s) = 5\sqrt{1 - (\omega - .5)^2 - (s - .5)^2}$ and the group-region shifts were $\eta_d(r, \omega, s) = 3(-1)^d\omega^2$, for $d = 1, 2$ and $r = 1, \ldots, R = 9$. The rth element of the regional marginal eigenvectors were $(\mathrm{v}_{dk})_r = 1/2\sin[\{k\pi(r-1)\}/8]$, for $d = 1, 2$, $r = 1, \ldots, 9$ and $k = 1, 2, 3$; the functional and longitudinal marginal eigenfunctions were $\phi_{d\ell}(\omega) = \sqrt{2}\sin(\ell\pi\omega)$, $d = 1, 2$, $\ell = 1, 2, 3$, and $\psi_{dm}(s) = \sqrt{2}\cos(m\pi s)$, $d = 1, 2$, $m = 1, 2, 3$, respectively. The subject-specific scores $\xi_{di,k\ell m}$ were simulated independently from mean zero Gaussian distributions with variances $\tau_{d,k\ell m} = \sqrt{.01/(k\ell m)}$, for k, ℓ and m ranging from 1 to 3. The measurement error $\epsilon_{di}(r, \omega, s)$ was simulated independently over region, frequency, and segment from a mean zero Gaussian distribution with variance $\sigma_d^2 = 5/c$, where c was used as a tuning parameter to vary the signal-to-noise ratio (SNR) of the simulated data.

We conducted simulations for two sample sizes, two SNRs and two data sparsity levels, for a total of eight settings. We considered SNR ratios of approximately 2.5 and 10, corresponding to $c = 2$ and 10, respectively, in the generation of the error variance. We simulated data at sample sizes $n_d = 15$ and 50 where the lower sample size mimics the within group sample size of the word segmentation data. In addition to the dense case with every subject observed at the full segment domain $\mathcal{S}$, we considered sparsity in the longitudinal dimension similar to the word segmentation data where subjects were observed at a subject-specific longitudinal domain, denoted by $\mathcal{S}_{di}$. To induce sparsity, we generated the number of segments observed for each subject, denoted by $|\mathcal{S}_{di}|$, from a uniform distribution on $[30, 50]$. Once the number of segments were generated for each subject, the segments

observed for subject i were selected randomly from the 50 grid points in $[0,1]$ constituting the subject-specific longitudinal domain $\mathcal{S}_{di}$. In all settings, marginal components were selected to explain at least 90% of the variation in the respective marginal covariance decompositions. Smoothing bandwidths were selected by GCV and REML for the functional fixed effects and marginal covariances, respectively. To assess the performance of the proposed estimation algorithm in targeting the functional and vector components of HPCA, we utilizes relative squared errors $\text{RSE}(s) = ||\hat{f}(s) - f(s)||^2/||f(s)||^2$, $\text{RSE}(\omega, s) = ||\hat{f}(\omega, s) - f(\omega, s)||^2/||f((\omega, s)||^2$, $\text{RSE}(r, \omega, s) = ||\hat{f}(r, \omega, s) - f(r, \omega, s)||^2/||f((r, \omega, s)||^2$ and $\text{RSE}(r) = ||\hat{f}(r) - f(r)||^2/||f(r)||^2$, for one- to three-dimensional functional components and the vector components, respectively, where $||f(s)||^2 = \int f^2(s)ds$, $||f(\omega, s)||^2 = \int\int f^2(\omega, s)d\omega ds$, $||f(r, \omega, s)||^2 = \sum_{r=1}^{R} \int\int f^2(\omega, s)d\omega ds$ and $||f(r)||^2 = \sum_{r=1}^{R} f^2(r)$. For assessment of the scalar components, we utilize normalized mean squared errors, $\text{MSE}= (\hat{\theta} - \theta)^2/\theta^2$, for an estimate $\hat{\theta}$ of a parameter θ. In addition, we report the total FVE, $FVE_{dG'}$, and the marginal FVEs along the regional, functional, and longitudinal dimensions, $FVE_{dK,\mathcal{R}}$, $FVE_{dL,\Omega}$, and $FVE_{dM,\mathcal{S}}$, based on the K, L, and M marginal eigencomponents included in the decomposition, respectively.

Figures A3.8 and A3.9 display the estimated model components based on 200 Monte Carlo runs from the dense simulation set-up with $n_d = 15$ and $c = 10$ (high SNR). The estimated overall mean function and group-region shift with the median RSE values (Figure A3.8(b),(d)) track the corresponding true surfaces (Figure A3.8(a),(c)). The estimated functional and longitudinal marginal eigenfunctions (Figure A3.9) are displayed from runs with RSE values at the 10th, 50th, and 90th percentiles, overlaid by their true quantities. Even with a small sample size, HPCA captures the periodicity and magnitude of the true components. Tables 3 and A3.2 display median, 10th, and 90th percentile RSE and normalized MSE values based on 200 Monte Carlo runs corresponding to the estimated HPCA components from all eight simulation settings. Note that since normalized measures of RSE and MSE were used, we report percentiles for the same quantity over combined Monte Carlo runs across groups and potentially subjects. More specifically, while performance measures for $\eta_d(r, w, s)$, $v_{dk}(r)$, $\phi_{d\ell}(w)$, $\psi_{dm}(s)$, σ_d^2, $FVE_{dG'}$, $FVE_{dK,\mathcal{R}}$, $FVE_{dL,\Omega}$, and $FVE_{dM,\mathcal{S}}$ are

reported over $D \times 200$ Monte Carlo runs, measures for $Y_{di}(r, w, s)$ and $\tau_{d,klm}$ are reported over $D \times n_d \times 200$ and $D \times K \times L \times M \times 200$ Monte Carlo runs, respectively.

Across simulation designs, the marginal FVEs almost always approach 1 due to the relatively small number of eigencomponents utilized in the data generation procedure. Despite this, trends in the marginal FVEs in all three dimensions are still detectable. The marginal FVE increases or stays the same with increasing sample size and decreasing sparsity due to improved estimation of the marginal eigencomponents. $FVE_{dG'}$ is similarly high across simulation designs, which is not surprising given the value of the marginal FVEs. However, as noted in Section 3.2, estimation of $FVE_{dG'}$ relies on estimates of the variance components and the three-dimensional norm of the demeaned observed data, which may explain why $FVE_{dG'}$ exceeds 1 in some instances.

To study the performance of the proposed bootstrap test for group-level inference, we considered testing the equality of the group-region mean shifts within the fifth region (with maximal variation in the leading marginal eigenvector), $H_0 : \eta_d(5, \omega, s) = \eta(5, \omega, s)$. Data were generated from model (A.1) with $\eta_d(r, \omega, s) = 0$ for $r \neq 5$ and $\eta_d(5, \omega, s) = (-1)^d \Delta$ for $r = 5$, where Δ was a tuning parameter used to vary the difference between the group-region shift. We conducted simulations under the sparse data design with $n_d = 15, 30, 50$ and $c = 2$ (low SNR) at varying levels of Δ, $\Delta = 0, .02, .04, .06, .08, .10, .12$, with $\Delta = 0$ corresponding to no difference in region shifts and $\Delta = .12$ to the largest group-region shift difference. We assessed the level and power of the proposed test based on 200 Monte Carlo runs with the p-values estimated from 200 bootstrap samples within each Monte Carlo run (Figure A3.10). The null hypothesis was rejected if the estimated p-value was less than or equal to .05. The level of the test is estimated at .085, .070, and .035 for $n_d = 15$, $n_d = 30$, and $n_d = 50$, respectively. The power is found to increase faster with increasing Δ for the larger sample sizes as expected, with a power of 1 reached at Δ equal to .12, .10, and .08 for $n_d = 15$, $n_d = 30$, and $n_d = 50$, respectively, suggesting that even at low SNR and small group sample sizes the proposed bootstrap procedure is sensitive to departures from the null hypothesis.

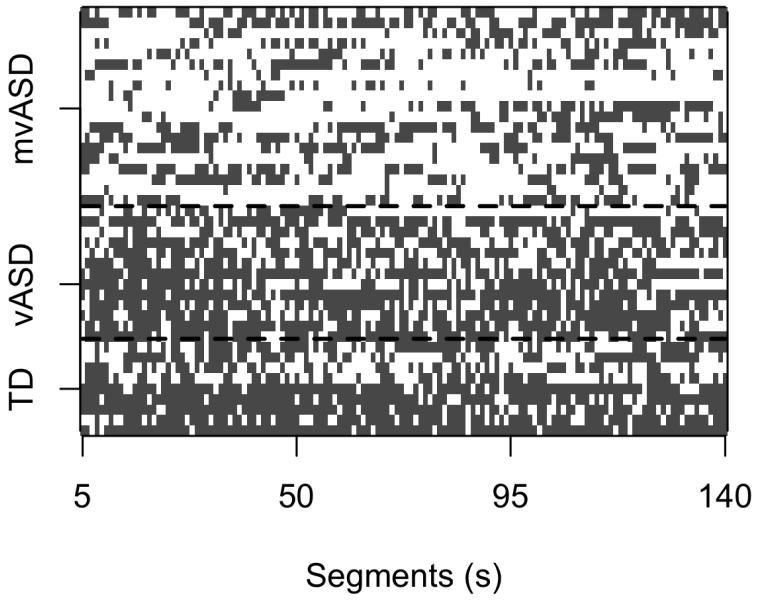
Table A3.1: FVE of the marginal covariances for the selected eigencomponents in each diagnostic group in the three dimensional HPCA decomposition. The number of eigencomponents are chosen to explain at least 90% FVE.

TD			vASD			mvASD		
$\mathcal{R}$	Ω	$\mathcal{S}$	$\mathcal{R}$	Ω	$\mathcal{S}$	$\mathcal{R}$	Ω	$\mathcal{S}$
0.590	0.595	0.890	0.613	0.575	0.892	0.571	0.579	0.857
0.114	0.139	0.049	0.122	0.181	0.060	0.136	0.174	0.063
0.099	0.080	-	0.101	0.085	-	0.111	0.083	-
0.078	0.065	-	0.067	0.049	-	0.077	0.052	-
0.051	0.042	-	-	0.042	-	0.046	0.040	-

Table A3.2: Percentiles 50% (10%, 90%) of the relative squared errors (RSE), normalized mean squared errors (MSE) , and both total and marginal fraction of variance explained (FVE) across groups for model components based on 200 Monte Carlo runs from the dense simulation design at $n_d = 15, 50$ for low and high SNR. Due to their small magnitude, MSE values are scaled by a factor of 10^3 for presentation.

	Low SNR				High SNR			
	$n_d = 15$		$n_d = 50$		$n_d = 15$		$n_d = 50$	
$\mu(\omega, s)$	0.000	(0.000, 0.000)	0.000	(0.000, 0.000)	0.000	(0.000, 0.000)	0.000	(0.000, 0.000)
$\eta_d(r, \omega, s)$	0.004	(0.003, 0.005)	0.001	(0.001, 0.002)	0.004	(0.002, 0.005)	0.001	(0.001, 0.002)
$Y_{di}(r, \omega, s)$	0.075	(0.072, 0.155)	0.076	(0.072, 0.155)	0.016	(0.015, 0.035)	0.016	(0.015, 0.035)
$\mathrm{v}_{d1}(r)$	0.073	(0.014, 0.373)	0.023	(0.005, 0.079)	0.063	(0.008, 0.266)	0.020	(0.002, 0.075)
$\mathrm{v}_{d2}(r)$	0.258	(0.038, 1.197)	0.070	(0.013, 0.321)	0.212	(0.026, 1.178)	0.057	(0.007, 0.233)
$\mathrm{v}_{d3}(r)$	0.176	(0.036, 1.216)	0.059	(0.012, 0.301)	0.148	(0.016, 1.109)	0.043	(0.004, 0.213)
$\phi_{d1}(\omega)$	0.062	(0.012, 0.295)	0.017	(0.003, 0.083)	0.066	(0.012, 0.286)	0.018	(0.003, 0.065)
$\phi_{d2}(\omega)$	0.194	(0.026, 0.874)	0.058	(0.008, 0.267)	0.202	(0.027, 0.761)	0.051	(0.006, 0.214)
$\phi_{d3}(\omega)$	0.134	(0.018, 0.790)	0.035	(0.005, 0.228)	0.121	(0.014, 0.682)	0.032	(0.004, 0.190)
$\psi_{d1}(s)$	0.071	(0.011, 0.292)	0.021	(0.003, 0.076)	0.075	(0.010, 0.274)	0.019	(0.004, 0.071)
$\psi_{d2}(s)$	0.218	(0.021, 0.858)	0.058	(0.009, 0.302)	0.219	(0.027, 1.049)	0.051	(0.007, 0.245)
$\psi_{d3}(s)$	0.128	(0.022, 0.735)	0.055	(0.013, 0.290)	0.141	(0.018, 0.904)	0.048	(0.010, 0.238)
$FVE_{dK,\mathcal{R}}$	0.985	(0.973, 0.993)	0.990	(0.983, 0.995)	0.995	(0.993, 0.997)	0.997	(0.995, 0.998)
$FVE_{dL,\Omega}$	0.978	(0.966, 0.987)	0.983	(0.976, 0.989)	0.992	(0.989, 0.994)	0.995	(0.993, 0.997)
$FVE_{dM,\mathcal{S}}$	0.939	(0.917, 0.959)	0.965	(0.953, 0.976)	0.986	(0.981, 0.991)	0.992	(0.990, 0.995)
$FVE_{dG'}$	0.996	(0.986, 1.007)	0.997	(0.993, 1.003)	0.999	(0.997, 1.001)	1.000	(0.999, 1.001)
$\tau_{d,klm}$	0.106	(0.004, 0.709)	0.031	(0.001, 0.233)	0.100	(0.004, 0.704)	0.029	(0.001, 0.226)
σ_d^2	0.003	(0.000, 0.016)	0.001	(0.000, 0.005)	0.003	(0.000, 0.020)	0.001	(0.000, 0.005)

(a)



(b)

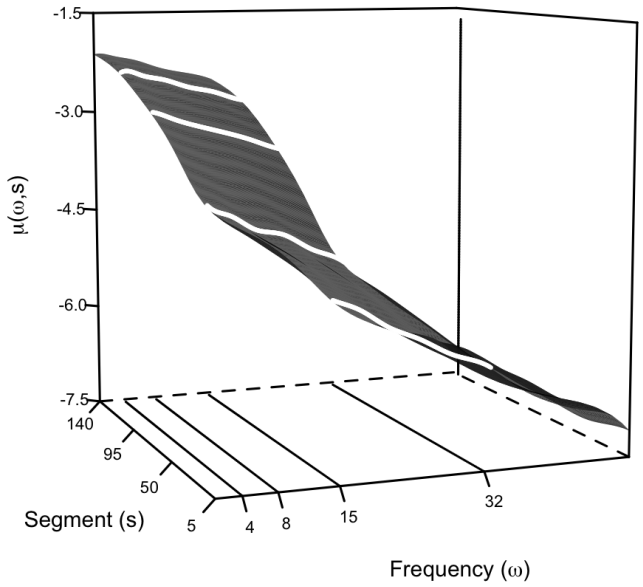


Figure A3.1: (a) Longitudinal sparsity plot: observed segments for each subject are shown in grey. Diagnostic groups are separated by black lines. (b) The estimated mean log principal power $\mu(\omega, s)$ for subjects pooled across the TD, vASD, and mvASD groups. The black vertical lines on the frequency axis separate the five frequency bands and the white lines show the boundaries projected onto the surface.

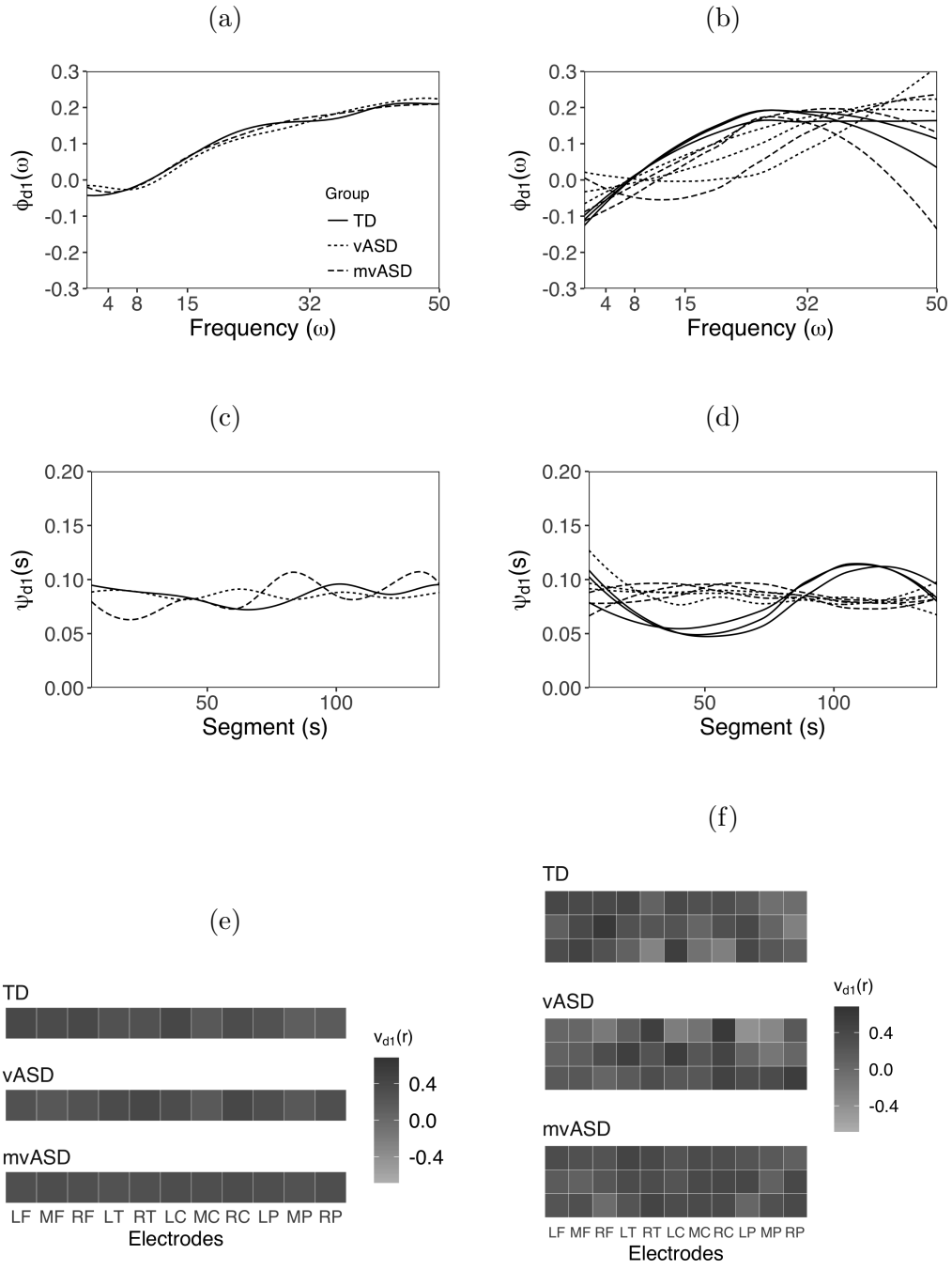


Figure A3.2: (a, c) Estimated first leading functional and longitudinal marginal eigenfunctions $\phi_{d1}(\omega)$ and $\psi_{d1}(s)$. (b, d) Estimated first leading functional and longitudinal eigenfunctions $\phi_{d1}(\omega)$ and $\psi_{d1}(s)$ in the LT region calculated at fixed values of $s = 40, 70, 100$ and $\omega = 30, 40, 50$, respectively. (e) Estimated first leading regional marginal eigenvectors $\mathrm{v}_{d1}(r)$. (f) Estimated first leading regional eigenvectors $\mathrm{v}_{d1}(r)$ obtained at fixed slices of $(\omega, s) = (30, 40), (30, 70), (30, 100)$ (top to bottom, respectively).

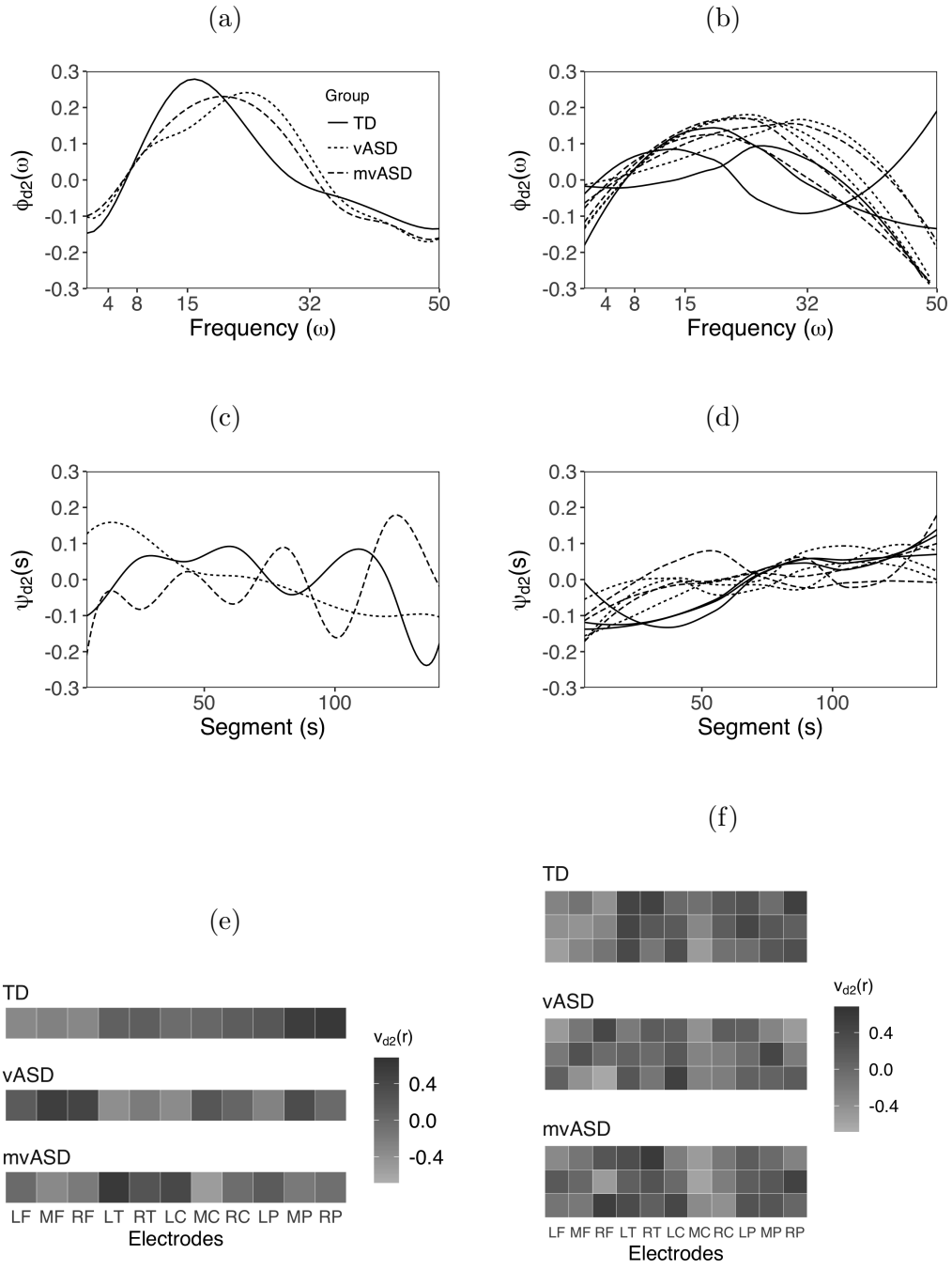


Figure A3.3: (a, c) Estimated second leading functional and longitudinal marginal eigenfunctions $\phi_{d2}(\omega)$ and $\psi_{d2}(s)$. (b, d) Estimated second functional and longitudinal eigenfunctions $\phi_{d2}(\omega)$ and $\psi_{d2}(s)$ in the LT region calculated at fixed values of $s = 40, 70, 100$ and $\omega = 30, 40, 50$, respectively. (e) Estimated second leading regional marginal eigenvectors $\mathrm{v}_{d2}(r)$. (f) Estimated second leading regional eigenvectors $\mathrm{v}_{d2}(r)$ obtained at fixed slices of $(\omega, s) = (30, 40), (30, 70), (30, 100)$ (top to bottom, respectively).

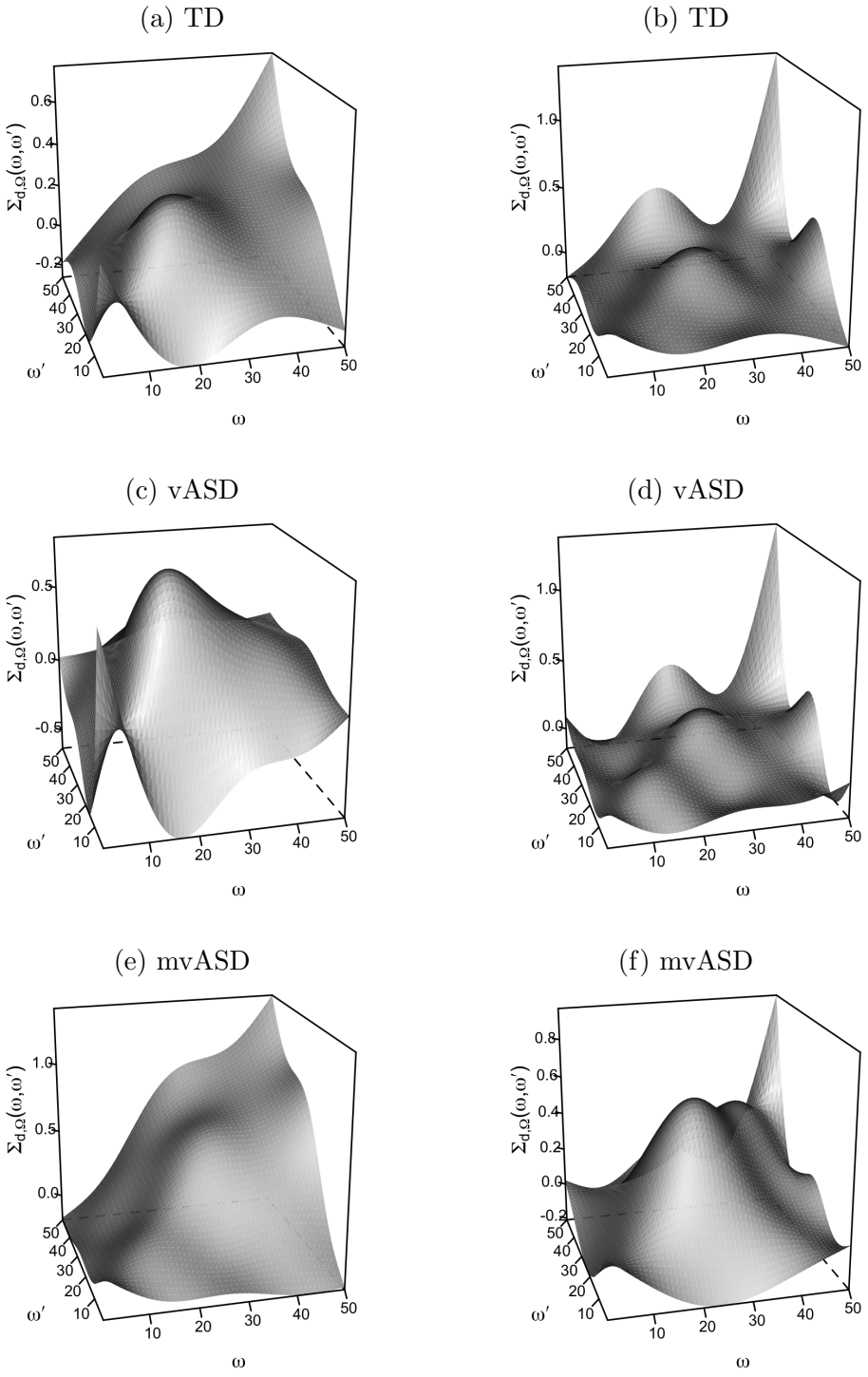


Figure A3.4: Estimated functional covariance surfaces, $\Sigma_{d,\Omega}(\omega, \omega')$, in the RT region for fixed $s = 100$ (left column) and RF region for fixed $s = 70$ (right column). Covariance surfaces are calculated within diagnostic groups, with the TD, vASD, and mvASD groups displayed in descending rows.

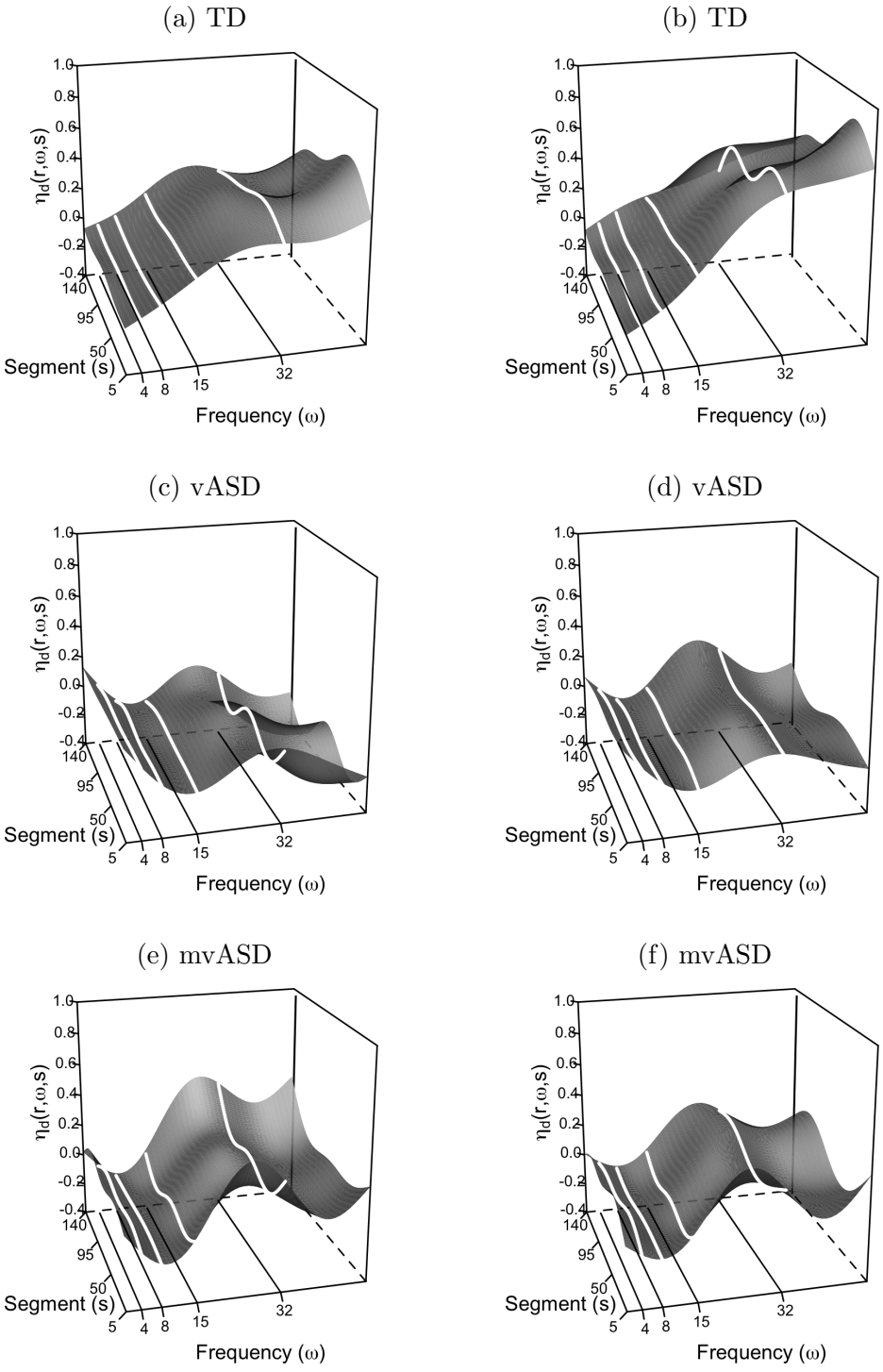


Figure A3.5: The estimated group-region shifts $\eta_d(r, \omega, s)$ in the LF (first column) and RF (second column) regions for the TD, vASD, and mvASD groups in descending rows, respectively. The black vertical lines on the frequency axis separate the five frequency bands and the white lines show the boundaries projected onto the surface.

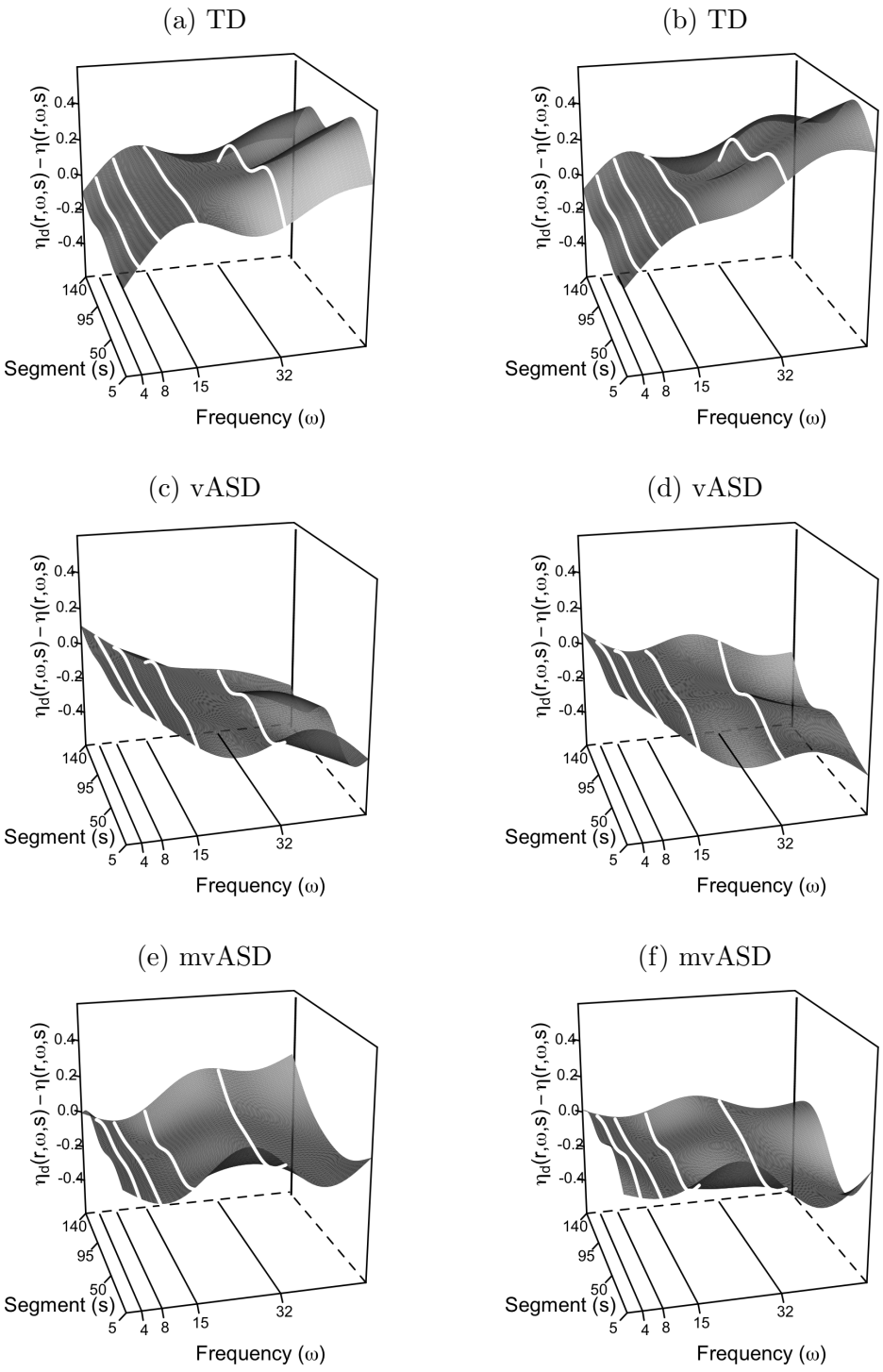


Figure A3.6: The differences of the estimated group-region shifts $\eta_d(r, \omega, s)$ from group-region averages in the LF (first column) and RF (second column) regions for the TD, vASD, and mvASD groups in descending rows, respectively. The black vertical lines on the frequency axis separate the five frequency bands and the white lines show the boundaries projected onto the surface.

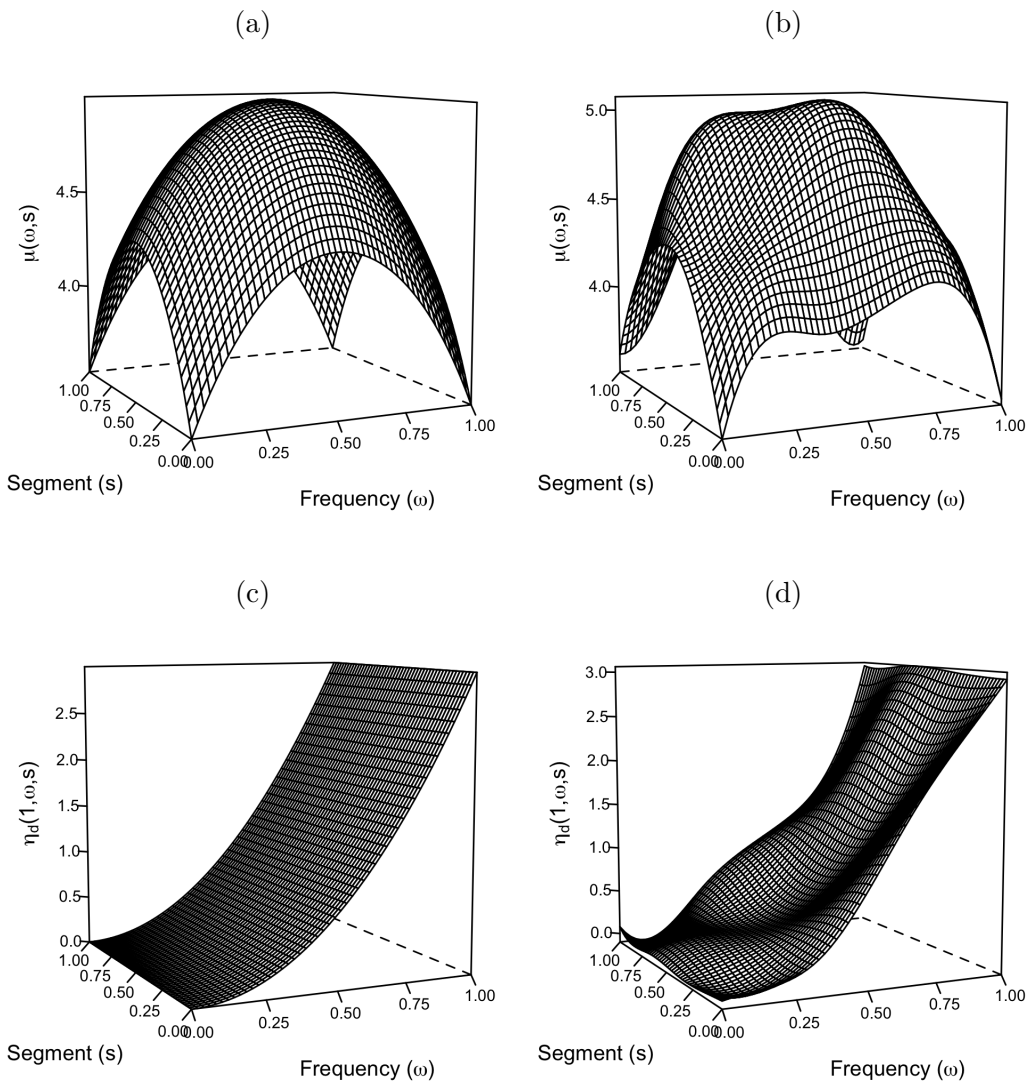


Figure A3.7: The true (a, c) and estimated (b, d) overall mean function $\mu(\omega, s)$ and group-specific shift $\eta_d(r, \omega, s)$, respectively, based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15$ and high SNR. The estimated quantities correspond to the Monte Carlo run with RSE value at the 50th percentile across groups.

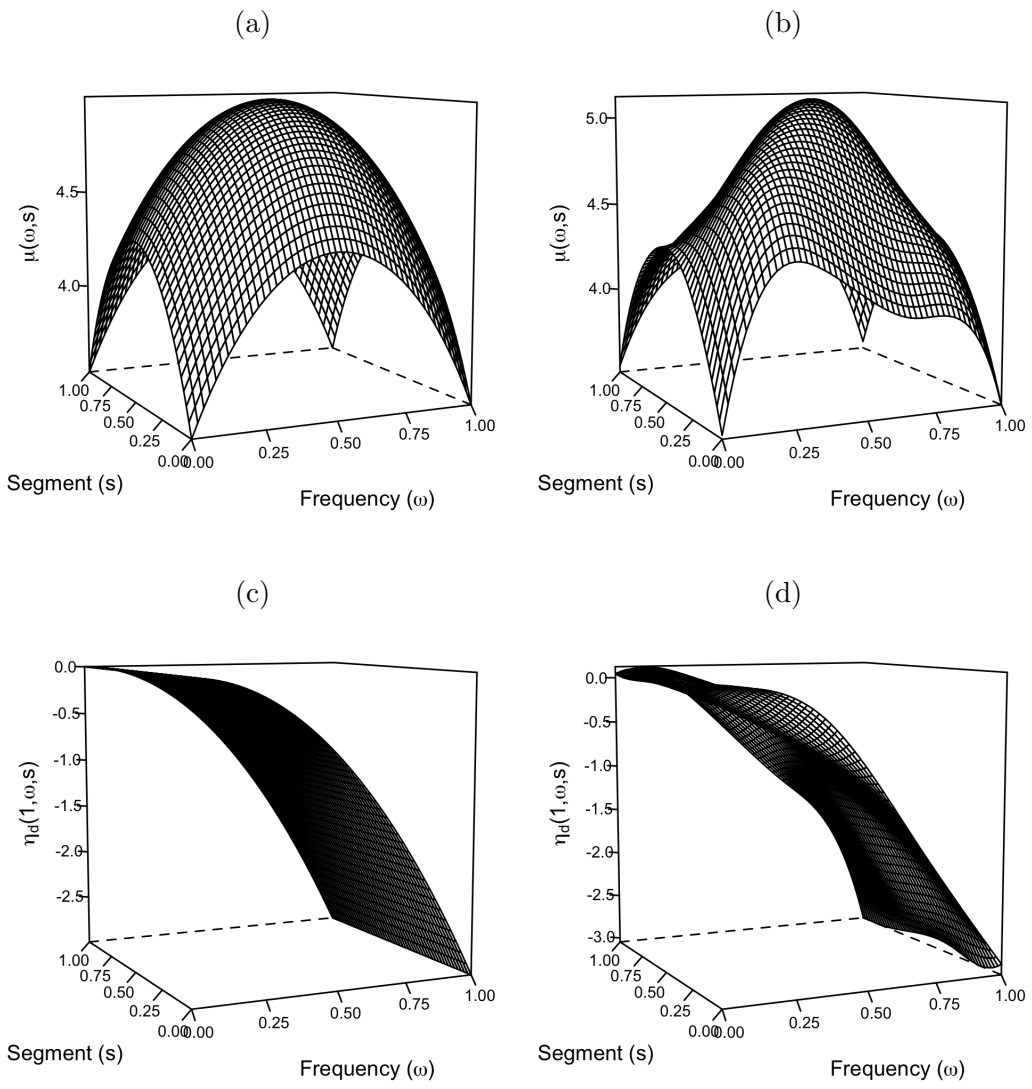


Figure A3.8: The true (a, c) and estimated (b, d) overall mean function $\mu(\omega, s)$ and group-region shift $\eta_d(r, \omega, s)$, respectively, based on 200 Monte Carlo runs from the dense simulation design at $n_d = 15$ and high SNR. The estimated quantities correspond to the Monte Carlo run with RSE value at the 50th percentile across groups.

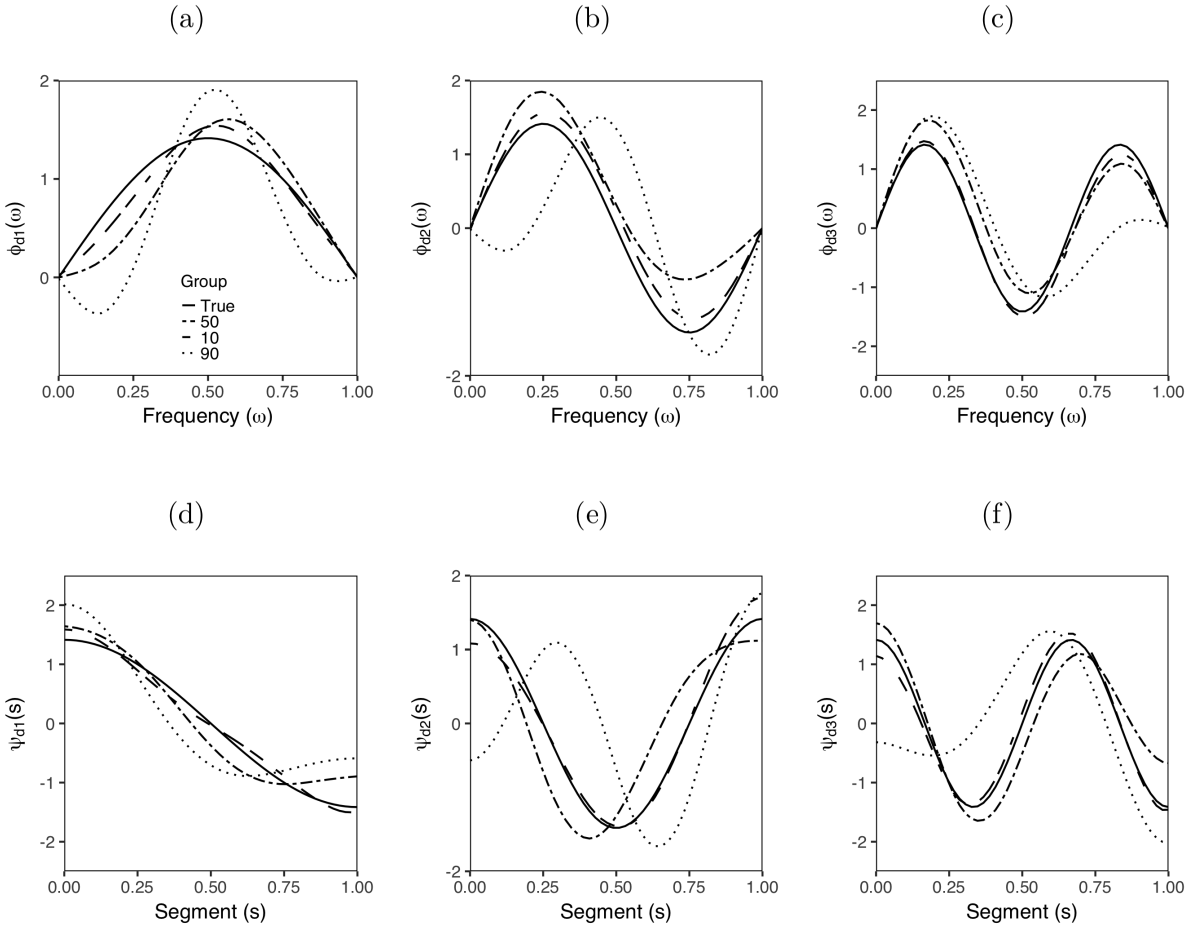


Figure A3.9: The true and estimated functional (first row) and longitudinal (second row) marginal eigenfunctions corresponding to the 10th, 50th, and 90th percentile RSE values across groups based on 200 Monte Carlo runs from the dense simulation design at $n_d = 15$ and high SNR.

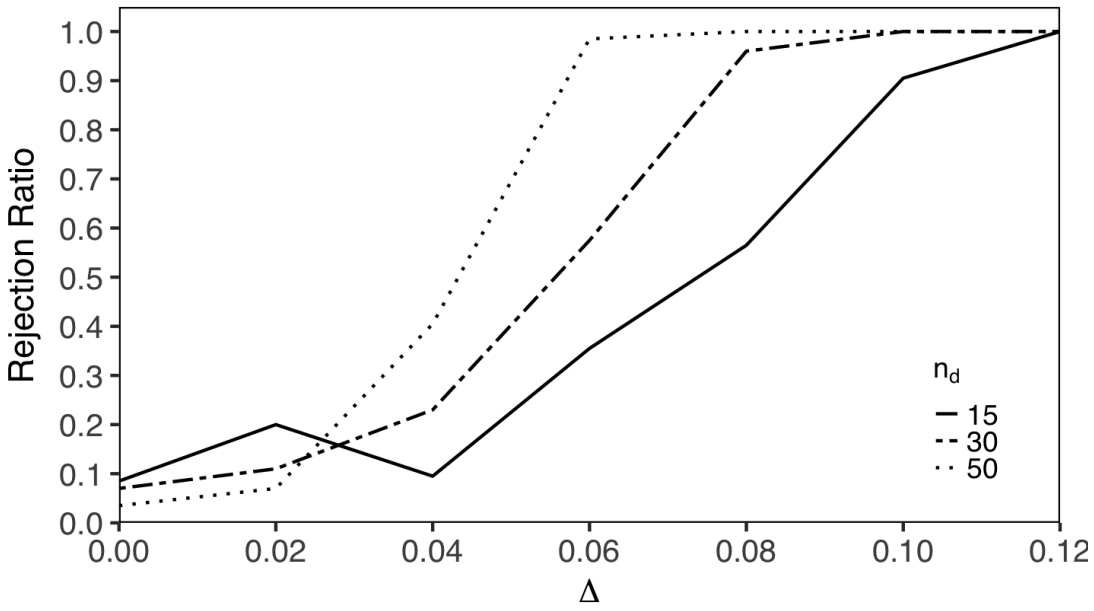


Figure A3.10: The level and power of the proposed parametric bootstrap procedure based on 200 Monte Carlo runs with the p-values estimated from 200 bootstrap samples within each Monte Carlo run.

Bibliography

V. Baladandayuthapani, B. Mallick, M. Hong, J. Lupton, N. Turner, and R. Carroll. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics*, 64(1):64–73, 2007.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.

D. Brillinger. *Time Series: Data Analysis an Theory*. Holden-Day Series in Time Series Analysis. Holden-Day, 1981.

C. Bugli and P. Lambert. Functional anova with random functional effects: an application to event-related potentials modelling for electroencephalograms analysis. *Statistics in Medicine*, 25(21):3718–3739, 2006.

K. Chen and B. Lynch. Weak separability for two-way functional data: concept and test. *arXiv:1703.10210*, 2017.

K. Chen and H.-G. Müller. Modeling repeated functional observations. *Journal of the American Statistical Association*, 107(500):1599–1609, 2012.

K. Chen, P. Delicado, and H.-G. Müller. Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 79(1):177–196, 2016.

J.-M. Chiou, Y.-T. Chen, and Y.-F. Yang. Multivariate functional principal components analysis: a normalization approach. *Statistica Sinica*, 24(4):1571–1596, 2014.

M. X. Cohen. *Analyzing neural time series data: theory and practice*. The MIT Press, 2014.

T. Collura. History and evolution of electroencephalographic instruments and techniques. *Journal of Clinical Neurophysiology*, 10(4):476–504, 1993.

A. W. Corcoran, P. M. Alday, M. Schlesewsky, and I. Bornkessel-Schlesewsky. Toward a reliable, automated method of individual alpha frequency (iaf) quantification. *Psychophysiology*, 55(7):e13064, 2018.

C. Crainiceanu, A. Staicu, and C. Di. Generalized multilevel functional regression. *Journal of the American Statistical Association*, 104(488):1550–1561, 2009.

D. Şentürk and H. Müller. Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association*, 105(491):1256–1264, 2010.

J. Daltrozzo and C. Conway. Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Frontiers in Human Neuroscience*, 8:437, 2014.

D. J. Davidson. Functional mixed-effect models for electrophysiological responses. *Neurophysiology*, 41(1):71–79, Feb 2009.

A. C. De Silva, N. C. Sinclair, and D. T. J. Liley. Limitations in the rapid extraction of evoked potentials using parametric modeling. *IEEE Transactions on Biomedical Engineering*, 59 (5):1462–1471, May 2012.

P. Delicado, R. Giraldo, C. Comas, and J. Mateu. Statistics for spatial functional data: some recent contributions. *Environmetrics*, 21(3-4):224–239, 2010.

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

C.-Z. Di, C. Crainiceanu, B. Caffo, and N. Punjabi. Multilevel functional principal component analysis. *The Annals of Applied Statistics*, 3(1):458–488, 2009.

A. Dickinson, C. DiStefano, D. Senturk, and S. S. Jeste. Peak alpha frequency is a neural marker of cognitive function across the autism spectrum. *European Journal of Neuroscience*, 47(6):643–651, 2018.

J. C. Edgar, K. Heiken, Y.-H. Chen, J. D. Herrington, V. Chow, S. Liu, L. Bloy, M. Huang, J. Pandey, K. M. Cannon, S. Qasmieh, S. E. Levy, R. T. Schultz, and T. P. L. Roberts. Resting-state alpha in autism spectrum disorder and alpha associations with thalamic volume. *Journal of Autism and Developmental Disorders*, 45(3):795–804, 2015.

X. Gao, W. Shen, B. Shahbaba, N. Fortin, and H. Ombao. Evolutionary State-Space Model and Its Application to Time-Frequency Analysis of Local Field Potentials. *arXiv*, art. arXiv:1610.07271, 2016.

X. Gao, W. Shen, and H. Ombao. Regularized matrix data clustering and its application to image analysis. *arXiv*, art. arXiv:1808.01749, 2018.

T. Gasser and L. Molinari. The analysis of the eeg. *Statistical Methods in Medical Research*, 5(1):67–99, 1996.

T. Gasser, J. Mocks, and R. Verleger. Selavco: A method to deal with trial-to-trial variability of evoked potentials. *Electroencephalography and Clinical Neurophysiology*, 55(6):717–723, 1983.

J. Gertheiss, A. Maity, and A.-M. Staicu. Variable selection in generalized functional linear models. *Stat*, 2(1):86–101, 2013.

R. Giraldo, P. Delicado, and J. Mateu. Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, 18(3):411–426, 2010.

J. Goldsmith, J. Bobb, C. Crainiceanu, C. Caffo, and D. Reich. Penalized functional regression. *Journal of Computational and Graphical Statistics*, 20(4):830–851, 2011.

J. Goldsmith, S. Greven, and C. Crainiceanu. Corrected confidence bands for functional data using principal components. *Biometrics*, 69(1):41–51, 2012.

J. Goldsmith, L. Huang, and C. M. Crainiceanu. Smooth scalar-on-image regression via spatial bayesian variable selection. *Journal of Computational and Graphical Statistics*, 23 (1):46–64, 2014.

S. Greven and F. Scheipl. A general framework for functional regression modelling. *Statistical Modelling*, 17(1-2):1–35, 2017.

S. Greven, C. Crainiceanu, B. Caffo, and D. Reich. Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, 4:1022–1054, 2010.

J. Gross. Analytical methods and experimental approaches for electrophysiological studies of brain oscillations. *Journal of Neuroscience Methods*, 228:57–66, 2014.

C. Happ and S. Greven. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659, 2018.

K. Hasenstab, C. Sugar, D. Telesca, K. McEvoy, S. Jeste, and D. Şentürk. Identifying longitudinal trends within eeg experiments. *Biometrics*, 71(4):1090–1100, 2015.

K. Hasenstab, C. Sugar, D. Telesca, S. Jeste, and D. Şentürk. Robust functional clustering of erp data with application to a study of implicit learning in autism. *Biostatistics*, 17(3): 484–498, 2016.

K. Hasenstab, A. Scheffler, D. Telesca, C. Sugar, S. Jeste, C. DiStefano, and D. Şentürk. A multi-dimensional functional principal components analysis of eeg data. *Biometrics*, 2017.

T. Hastie and C. Mallows. [a statistical view of some chemometrics regression tools]: Discussion. *Technometrics*, 35(2):140–143, 1993.

L. Huang, P. Reiss, L. Xiao, V. Zipunnikov, M. Lindquist, and C. Crainiceanu. Two-way principal component analysis for matrix-variate data, with an application to functional magnetic resonance imaging data. *Biostatistics*, 18(2):214–229, 2017.

G. James, T. Hastie, and C. Sugar. Principal component models for sparse functional data. *Biometrika*, 87(3):587–602, 2000.

G. M. James and B. W. Silverman. Functional adaptive model estimation. *Journal of the American Statistical Association*, 100(470):565–576, 2005.

J. Jaques and C. Preda. Model-based clustering for multivariate functional data. *Computational Statistics Data Analysis*, 71(C):92–106, 2014.

S. Jeste, N. Kirkham, D. Şentürk, K. Hasenstab, C. Sugar, C. Kupelian, E. Baker, A. Sanders, C. Shimizu, T. Norona, A.and Paparella, S. Freeman, and S. Johnson. Electrophysiological evidence of heterogeneity in visual statistical learning in young children with asd. *Developmental Science*, 18(1):90–105, 2015.

S. S. Jeste and C. A. Nelson. Event related potentials in the understanding of autism spectrum disorders: An analytical review. *Journal of Autism and Developmental Disorders*, 39(3):495, 2008.

I. T. A. Jolliffe. *Principal component analysis*. Springer, 2002.

K. Karhunen. Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae, AI*, 37:1–37, 1946.

M. Kundu, J. Harezlak, and T. Randolph. Longitudinal functional models with structured penalties. *Statistical Modelling*, 16(2):114–139, 2016.

S. Lan, A. Holbrook, G. A. Elias, N. J. Fortin, H. Ombao, and B. Shahbaba. Flexible Bayesian Dynamic Modeling of Correlation and Covariance Matrices. *arXiv*, 2017.

K. Lange. *MM optimization algorithms*. Society for Industrial and Applied Mathematics, 2016.

Q. Li, D. Senturk, C. A. Sugar, S. Jeste, C. DiStefano, J. Frohlich, and D. Telesca. Inferring brain signals synchronicity from a sample of eeg readings. *Journal of the American Statistical Association*, 0(0):1–18, 2019.

H. Lian. Shrinkage estimation and selection for multiple functional regression. *Statistica Sinica*, 23(1):51–74, 2013.

C. Liu, S. Ray, and G. Hooker. Functional principal component analysis of spatially correlated data. *Statistics and Computing*, 27(6):1639–1654, 2017.

M. Loeve. Fonctions aléatoires à décomposition orthogonale exponentielle. *La Revue Scientifique*, 84:159–162, 1946.

S. Luck. *An Introduction to the Event-Related Potential Technique*. The MIT Press, 2014.

S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, 94(20):10979–10984, 1997.

B. D. Marx and P. H. Eilers. Multidimensional penalized signal regression. *Technometrics*, 47(1):13–22, 2005.

B. D. Marx and P. H. C. Eilers. Generalized linear regression on sampled signals and curves: A p-spline approach. *Technometrics*, 41(1):1–13, 1999.

R. Millsap and A. Maydeu-Olivares. *SAGE Handbook of Quantitative Methods in Psychology*. SAGE, 2010.

V. Miskovic, X. Ma, C. Chou, M. Fan, M. Owens, H. Sayama, and B. Brandon E. Gibb. Developmental changes in spontaneous electrocortical activity and network organization from early to late childhood. *NeuroImage*, 118(Supplement C):237–247, 2015.

J. Morris. Functional regression. *Annual Review of Statistics and Its Application*, 2(1): 321–359, 2015.

J. Morris, M. Vannucci, P. Brown, and R. Carroll. Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *Journal of the American Statistical Association*, 98(463):573–583, 2003.

J. S. Morris and R. J. Carroll. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 68(2):179–199, 2006.

H. G. Müller. Functional modeling of longitudinal data. In G. Verbeke, editor, *Longitudinal Data Analysis*, chapter 10, pages 223–252. Chapman and Hall/CRC, New York, 2008.

H. Ombao. *Handbook of neuroimaging data analysis*. CRC Press, 2017.

H. Ombao and R. Moon-Ho. Time-dependent frequency domain principal components analysis of multichannel non-stationary signals. *Computational Statistics and Data Analysis*, 50(9):2339–2360, 2006.

S. Park and A.-M. Staicu. Longitudinal functional data analysis. *Stat*, 4(1):212–226, 2015.

F. Perrin, J. Pernier, O. Bertrand, and J. Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, 72 (2):184–187, 1989.

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2017. R package version 3.1-131.

X. Qiao, S. Guo, and G. M. James. Functional graphical models. *Journal of the American Statistical Association*, 114(525):211–222, 2019.

R Core Team. R: A language and environment for statistical computing, 2018.

J. O. Ramsay and C. J. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):539–572, 1991.

J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer, 2010.

P. Reiss, J. Goldsmith, H. Shang, and R. Ogden. Methods for scalar-on-function regression. *International Statistical Review*, 85(2):228–249, 2017.

P. T. Reiss and R. T. Ogden. Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, 102(479):984–996, 2007.

P. T. Reiss and R. T. Ogden. Functional generalized linear models with images as predictors. *Biometrics*, 66(1):61–69, 2010.

P. T. Reiss and R. Todd Ogden. Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 71(2): 505–523, 2009.

L. Rossi, A. Bianchi, A. Merzagora, A. Gaggiani, S. Cerutti, and F. Bracchi. Single trial somatosensory evoked potential extraction with arx filtering for a combined spinal cord intraoperative neuromonitoring technique. *BioMedical Engineering OnLine*, 6(2):1–2, Jan 2007.

A. Scheffler, D. Telesca, Q. Li, D. Şentürk, C. A. Sugar, C. Distefano, and S. Jeste. Hybrid principal components analysis for region-referenced longitudinal functional eeg data. *Biostatistics*, 2018.

A. Scott-Van Zeeland, K. Mcnealy, A. Wang, M. Sigman, S. Bookheimer, and M. Dapretto. No neural evidence of statistical learning during exposure to artificial languages in children with autism spectrum disorders. *Biological Psychiatry*, 68(4):345–351, 2010.

A.-M. Staicu, C. M. Crainiceanu, and R. J. Carroll. Fast methods for spatially correlated multilevel functional data. *Biostatistics*, 11(2):177–194, 2010.

H. Tager-Flusberg and C. Kasari. Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, 6(6):468–478, 2013.

The MathWorks Inc. Matlab, 2015b.

B. Turetsky, J. Raz, and G. Fein. Estimation of trial-to-trial variation in evoked potential signals by smoothing across trials. *Psychophysiology*, 26(6):700–712, 1989.

P. Valdas-Hernandez, A. Ojeda-Gonzalez, E. Martinez-Montes, T. Lage-Castellanos, A.and Virues-Alba, L. Valdas-Urrutia, and P. Valdes-Sosa. White matter architecture rather than cortical surface area correlates with the eeg alpha rhythm. *NeuroImage*, 49(3):2328–2339, 2010.

J. Wang, J. Barstein, L. Ethridge, M. Mosconi, Y. Takarae, and J. Sweeney. Resting state eeg abnormalities in autism spectrum disorders. *Journal of Neurodevelopmental Disorders*, 5(1):24, 2013.

J.-L. Wang, J.-M. Chiou, and H. Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295, 2016.

P. D. Welch. The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15:70–73, 1967.

S. Wood. *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, 2017.

S. N. Wood. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036, 2006a.

S. N. Wood. On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics*, 48(4):445–464, 2006b.

S. N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 73(1):3–36, 2011.

S. N. Wood, F. Scheipl, and J. J. Faraway. Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, 23(3):341–360, May 2013.

Y. Wu, J. Fan, and H.-G. Müller. Varying-coefficient functional linear regression. *Bernoulli*, 16(3):730–758, 08 2010.

L. Xiao, Y. Li, and D. Ruppert. Fast bivariate p-splines: the sandwich smoother. *Journal of the Royal Statistical Society. Series B (Methodological)*, 75(3):577–599, 2013.

F. Yao, H.-G. Müller, and J.-L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

L. Zhou, J. Huang, J. Martinez, A. Maity, V. Baladandayuthapani, and R. Carroll. Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association*, 105(489):390–400, 2010.

H. Zhu, M. Vannucci, and D. D. Cox. A bayesian hierarchical model for classification with selection of functional predictors. *Biometrics*, 66(2):463–473, 2010.

V. Zipunnikov, S. Greven, H. Shou, B. Caffo, D. Reich, and C. Crainiceanu. Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *The Annals of Applied Statistics*, 8(4):2175–2202, 2014.