Title
Modeling Time-varying Trends in ERP Data with Applications to an Implicit Learning Paradigm in Autism

Author
Hasenstab, Kyle Andrew

Publication Date
2015

University of California

Los Angeles

Modeling Time-varying Trends in ERP Data with Applications to an Implicit Learning Paradigm in Autism

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

by

Kyle Andrew Hasenstab

2015

ABSTRACT OF THE DISSERTATION

Modeling Time-varying Trends in ERP Data with Applications to an Implicit Learning Paradigm in Autism

by

Kyle Andrew Hasenstab

Doctor of Philosophy in Statistics
University of California, Los Angeles, 2015
Professor Damla Şentürk, Co-chair
Professor Catherine Ann Sugar, Co-chair

Event-related potential (ERP) studies are a subset of experimental frameworks within the field of electroencephalography (EEG) that focus on ERPs, the electrical potential outputted by a subject's brain when presented with an implicit task in the form of stimuli. Data comprise an ERP repetition observed for each stimulus across electrodes on the scalp, producing a complex data structure consisting of a functional, longitudinal and spatial dimension. In typical ERP studies, the dimension of data is reduced into a single measure for each subject by cross-sectionally averaging ERP across longitudinal and spatial repetitions. Features are then extracted from the averaged ERP and analyzed using simple statistical methods, ignoring additional information that may be found in the collapsed dimensions. In this dissertation, three types of methodology are proposed for preserving and analyzing the lost dimensions of ERP data. The first method, moving average processed ERP (MAP-ERP), is a two-step approach comprised of a meta-preprocessing step to preserve longitudinal information and a weighted mixed effects regression framework to allow modeling of the resulting meta-preprocessed data. The proposed robust functional clustering (RFC) algorithm identifies substructures in features of the longitudinal ERP processes while accounting for subject-level covariance heterogeneity induced by meta-preprocessing. Finally, the proposed multidimensional functional principal components analysis (MD-

FPCA) utilizes a two-stage procedure to summarize important characteristics across all three dimensions of the ERP data structure into an interpretable, low-dimensional form. Proposed methods are applied to a study on neural correlates of visual implicit learning in young children with autism spectrum disorder (ASD). Applications of the proposed methods reveal meaningful trends and substructures in the implicit learning processes of ASD children when compared to typically developing controls. Results indicate proposed methodology effectively preserves important information contained within the multiple dimensions of ERP.

The dissertation of Kyle Andrew Hasenstab is approved.

Donatello Telesca

Shafali Spurling Jeste

Frederic R Paik Schoenberg

Catherine Ann Sugar, Committee Co-chair

Damla Şentürk, Committee Co-chair

University of California, Los Angeles
2015

To my loving wife Susan...and Buzz

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Damla Senturk for rescuing me when I was academically homeless. You kept me on track and gave me valuable academic and professional guidance when I needed it most.

I would also like to thank my committee members and co-authors Dr. Shafali Jeste, Dr. Frederic Paik Schoenberg, Dr. Catherine Sugar and Dr. Donatello Telesca for providing me with valuable feedback on my research.

Thank you Aaron and Qian for your supportive input in our LF meetings.

To the Jeste Lab - Thank you for supplying us with the data and patiently guiding me through EEG preprocessing procedures. The data is rich with information and made for a wonderful project.

To Glenda Jones - Thank you for putting up with all of my administrative inquiries. My program and the entire department would not be able to function without you!

I am extremely grateful to my parents, whose sacrifices allowed me to pursue my academic and life goals.

Thank you to my family, the Hasenstabs (especially you Mark, Kim, Kristen, Garret and Branden) for supporting me through life's challenges.

Last but not least - I would like to express by deepest appreciation to Susan. Without your support and love, this would not have been possible.

This research was made possible by the graduate research fellowship grant DGE-1144087 (KH) from the National Science Foundation.

Thank you all so much!

Vita

2008–2009	Research Assistant, Department of Economics, UC Irvine.
2009	B.S. (Mathematics), University of California, Irvine.
2009	B.A. (Quantitative Economics), University of California, Irvine.
2008–2010	Math Standards Specialist, MIND Research Institute.
2011	Teaching Assistant, Department of Statistics, UCLA.
2011	Teaching Assistant, Explorations in Statistics Research Program, Columbia University.
2011	Reader, Department of Statistics, UCLA.
2011–2013	Graduate Student Researcher, Center for Embedded Networked Sensing, UCLA.
2012	M.S. (Statistics), UCLA.
2012–2015	National Science Foundation Graduate Research Fellowship Program (NSF-GRFP) Fellow, UCLA
2013–present	Graduate Student Researcher, Center for Autism Research and Treament, UCLA.
2015	Mathematical Statistician Intern, Center for Disclosure Avoidance Research, US Census Bureau

PUBLICATIONS AND PRESENTATIONS

Jeste, S. S., Kirkham, N., Şentürk, D., Hasenstab, K., Sugar, C., Kupelian, C., Baker, E., Sanders, A. J., Shimizu, C., Norona, A., Paparella, T., Freeman, S. F.N. and Johnson, S. P. (2014), Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Developmental Science* **18**, 90–105.

McEvoy, K., Hasenstab, K., Şentürk, D. Jeste, S. (2014) Physiologic artifacts in the resting state EEG of young children: Methodological considerations for noisy data. *Brain Imaging and Behavior* **9**, 104–114.

Hasenstab, K., Sugar, C., Telesca, D., Jeste, S., McEvoy, K. and Şentürk, D. (2015) Identifying longitudinal trends within EEG experiments. *Biometrics*,
doi: 10.1111/biom.12347.

Hasenstab, K., Sugar, C., Telesca, D., Jeste, S., and Şentürk, D. (2015) Robust Functional Clustering of Longitudinal ERP Trends. *Biostatistics*, In Revision.

Hasenstab, K., Scheffler, A., Telesca, D., Sugar, C., Jeste, S., and Şentürk, D. (2015) A Multi-Dimensional Functional Principal Components Analysis of EEG Data. In Preparation.

Invited oral presentation “Functional Analysis of ERP Trends Across an Implicit Learning Paradigm in Autism” at Centers for Disease Control and Prevention, Chronic Viral Diseases Branch (2015)

Invited oral presentation “Functional Analysis of ERP Trends Across an Implicit Learning Paradigm in Autism” at US Census Bureau, Associate Directorate of Research and Methodology (2015)

CHAPTER 1

Introduction

The continuous evolution in our ability to measure, record and process complex biomedical data has led to new opportunities as well as new challenges in the development of evidence based medical care and treatment. These rapid technological advances have resulted in a wealth of highly detailed patient level data and have provided the computational resources necessary to process this information. Through the use of modern computing, the development of new statistical methodology to analyze these data can potentially provide a more in depth understanding of the underlying mechanisms to certain diseases and disorders as well as epidemiological patterns. One such medium for recording data containing a highly complex structure that may benefit from these methodological advances is the electroencephalogram (EEG). EEG is a noninvasive method for measuring the communication between neurons of the brain (in the form of electrical impulses) over time. Since the 1950's, this approach has been utilized in the evaluation of numerous brain disorders, including epilepsy, sleep disorders, major affective disorder, schizophrenia, alcoholism and bipolar mood disorder [1, 2]. The data structure of EEG is quite complex, spanning the functional, longitudinal and spatial dimensions. EEG curves are functionally observed across a continous time domain and segmented into meaningful, functional repetitions that are spatially observed across several electrodes on the scalp. EEG functions may contain systematic changes in signal across repetitions, producing longitudinally observed functional data. Although EEG data contain a plethora of information on the cognitive processes of an individual, the majority of statistical methods used to make inference on these disorders do not take full advantage of the large amounts of information contained within the EEG. Standard techniques involve averaging EEG across the longitudinal and spatial dimensions to increase the signal-to-noise ratio of the observed curves and to simplify the data structure for analysis, contributing to the loss of potentially useful information on the evolution of EEG signal across electrodes over the course of an experiment. Single measures are then extracted from these averaged curves, ignoring the remaining information contained across the functional domain. In addition, data preprocessing unnecessarily suppresses segments of EEG that are thought to affect inference during subsequent analyses. This dissertation proposes a series of statistical methods to analyze and interpret complex EEG data while preserving information that is typically lost across multiple dimensions in standard analyses. Methods are discussed under two experimental frameworks of EEG analysis, resting-state EEG and event-related potentials. Resting-state EEG methods used to mitigate data loss during preprocessing and averaging are introduced in the context of two co-authored publications within this chapter. Three statistical methodologies for analyzing data from event-related potentials frameworks are proposed throughout the dissertation and are the primary focus of this research.

1.1 Resting-state EEG Experiments

Resting-state EEG studies record brain activity when a subject is not performing an explicit task, such as when subjects are sleeping, laying with their eyes closed or presented with a constant or neutral stimulus. The primary goal of these studies is to understand the brain's sensory, cognitive and motor functions within and across groups of interest by observing the cooperation of the brain's neuronal structures [3]. Although resting-state EEG is recorded in the temporal domain, data from these studies are typically analyzed in the frequency domain. A famous example of a resting-state EEG study is the Sleep Heart Health Study designed to investigate the impact of sleep disorders on health outcomes using polysomnography [4]. Other examples of resting-state EEG studies include those of Kim et al. (2013), Wang et al. (2013) and Lee et al. (2014) [5, 6, 7].

Data resulting from resting-state EEG experiments consist of functional curves observed over time collected across several electrodes on the scalp. Due to the noisy behavior and complex data structure of resting-state data, EEG must undergo a series of preprocessing steps in order to convert the data into an interpretable and manageable form prior to analysis. These include filtering (to smooth out noise outside a specified range of frequencies), bad channel replacement (a smoothing procedure performed when signals from an individual scalp electrode are compromised), voltage thresholding (to eliminate EEG with exceedingly high magnitudes), baseline correction (to standardize a subject's measurements to their own baseline period) and re-referencing (to standardize measurements to an average across all the electrodes). In addition, segments containing artifacts (eye blinks, saccades or muscle contractions) are usually removed due to their influence on inference during analysis. The data structure of the resulting preprocessed data is quite complex, consisting of a functional dimension (EEG curves observed over time) and a spatial dimension (EEG curves observed across several electrodes on the scalp). Segmentation of EEG into non-overlapping epochs adds a third additional repetition-specific dimension to the data. In order to reduce the repetition and spatial dimensions, EEG segments are averaged across repetitions and electrodes within regions of interest for each subject, resulting in a single EEG curve per region per subject. Averaged EEG curves are then represented as a function of frequency via Fast Fourier Transform. The frequency domain of the spectral curves is partitioned into bands where power levels extracted from these bands are analyzed using a variety of statistical methods.

1.1.1 Methods for Artifact Removal in Resting-State EEG

An important step in the preprocessing procedure described above is the removal of artifacts. Although visual inspection is an effective approach for removing trials affected by artifacts, removal of entire artifact segments may result in the loss of important EEG signal. Portions of EEG data recorded during an artifact occurrence may not be affected by a particular artifact and EEG signal may be salvagable. Effects of physiologic artifacts on resting state oscillations are investigated in McEvoy et al. (2014) where they develop methodological considerations for dealing with noisy data attributed to artifacts [8]. The goal of the study was to propose artifact rejection methods during preprocessing of resting state EEG data. Mean spectral power of artifact data (eye blinks, saccades and muscle contractions) were statistically compared to mean spectral power of artifact-free data across different power band frequencies (theta (4-7 hz), alpha (8-12 hz), beta (13-30 hz) and gamma (35-45 hz)) and regions of the scalp to identify regions and power bands most affected by specific artifact occurrences.

Statistical analysis was performed using the log transformed absolute and relative power estimates of the theta, alpha, beta and gamma bands. Eight log-power bands (4 bands × 2 power types) were modeled with a linear mixed effects model using artifact type, region, and all higher order interactions and predictors. All models contained a random intercept term to account for subject-specific heterogeneity and controlled for unequal averaging across each segment category. Mean contrasts between artifact-free power and artifact power were calculated for the eight power bands, three artifact categories (eye blinks, EMG, saccades) and nine regions, totaling 27 tests per model and 216 tests overall. Inference was adjusted for multiple testing in all 216 tests using the false discovery rate. The 216 contrasts of the eight linear mixed effects models indicated that mean absolute and relative power was affected differently by artifacts in separate regions. Results suggest the systematic removal of artifact segments depend on the power and region of interest and the nature of the visual stimulus. Specifics on the methodological considerations for artifact removal in resting state EEG studies are discussed thoroughly in McEvoy et al. (2014).

1.1.2 Spatial Dimension Reduction in Resting-State EEG

In resting-state EEG experiments, EEG from electrodes that are spatial neighbors usually exhibit similar behavior, therefore electrodes are typically grouped into regions of spatial proximity. Averaging data across electrodes within these regions results in the loss of information on the variability of EEG power. In addition, EEG is averaged across all relevant electrodes when differences between power across regions is not found to be significant using standard statistical tools. In a 2015 study, Jeste et al. (2015) proposed principal components analysis (PCA) for spatial dimension reduction to summarize EEG power across electrode regions into a set of scores accounting for the majority of EEG power variability [9]. They used PCA to reduce EEG power across nine regions of the scalp into a single component in order to investigate the association between resting state EEG oscillations and cognition in preschoolers with ASD. The sample for this study included 54 ASD and 43 TD children whose verbal IQ (VIQ) and non-verbal IQ (NVIQ) scores were recorded as measures of cognition and language. Absolute power from the theta, alpha and gamma bands were extracted from the FFTs of the EEG and log transformed for analysis.

The log transformed powers were analyzed using PCA separately across the theta, alpha and gamma bands in order to summarize the power data across nine regions of interest into comprehensive scores. VIQ and NVIQ were regressed on the PCA scores of the first component to test whether each power band in the ASD and TD groups significantly predicted cognition. VIQ and NVIQ have a censored distribution due to a lower bound in IQ measures. To cope with this distributional structure, Tobit models were also fit to account for IQ floor effects. Multiple linear and Tobit regressions found gamma band power to significantly predict VIQ in the TD group and theta band power to robustly predict both VIQ and NVIQ in the ASD group. The results of this study emphasized the plausibility of using EEG in investigating the heterogeneity of language cognition in ASD and suggested EEG band power may serve as a electrophysiological marker or predictor of cognitive function for young children with ASD. In addition, this study introduced Tobit regression as an additional tool to account for IQ censoring in phenotype analyses. Finally, this study shows PCA is a viable approach to spatial dimension reduction and EEG analysis under certain conditions.

1.2 Event-related Potentials (ERP) Experiments

Event-related potentials studies focus on the electrical potential of a subject when presented with a sequence of stimuli and provides insight into the evolution of brain activity across a learning task. EEG recorded in response to these stimuli are known as event-related potentials (ERP) and are the primary focus of this dissertation. Analysis of ERP data is quite challenging due to its complex structure and noisy behavior. ERP waveforms are recorded across multiple electrodes for each stimulus, producing data consisting of a spatial dimension (electrodes), a functional dimension (EEG recorded over time) and a longitudinal dimension (EEG recorded at each stimulus). Since EEG has high time resolution and low spatial resolution, primary analyses typically focus on the amplitude (microvolts) and latency (milliseconds) of temporal features (peaks) in the ERP waveforms for each subject (Luck, 2005). Electrical response to sensory stimuli is very small (on the order of a few microvolts) relative to the magnitude of spontaneous EEG, yielding a low signal-to-noise ratio in observed ERP, making the ERP features of interest unidentifiable. To cope with this well-known phenomenon, stimuli are applied repeatedly and the resulting ERP waveforms are averaged across the stimulus repetitions (longitudinal dimension) for each subject [1, 2, 10]. ERP also undergo additional preprocessing, similar to the steps described for resting-state EEG data. Subsequently, features are extracted from the preprocessed ERP, producing a single observation for each subject. Experiments involving multiple stimulus conditions (different sequences of stimuli) may produce multiple feature observations for each subject. In addition, ERP may be averaged within predefined regions of the scalp if region-specific effects are of particular interest. Following extraction of these features, the data are analyzed using simple statistical methods such as t-tests and repeated measures ANOVA.

Standard preprocessing procedures for the analysis of ERP result in information loss on all three data dimensions. As a result of collapsing the longitudinal dimension by averaging ERP across trials within subjects, methods fail to capture potentially informative changes in ERP signals that may emerge over the course of the experiment. Averaging ERP across electrodes results in the loss of spatial covariation at the electrode level. If spatial information is retained, information on within-region variability, which can provide insight into brain connectivity, is still lost due to within-region averaging. Furthermore, these methods selectively incorporate single measures of ERP features into their analyses, reducing the entire functional dimension of the observed

ERP into a single observation. In this dissertation, three types of methodology are developed for preserving and analyzing the lost dimensions of ERP data. Since ERP studies focus on the electrical response to a sequence of stimuli presented over time, the focus of this dissertation is on the longitudinal component of the data structure corresponding to the trials resulting from the repeated stimuli. The motivating application for the proposed methods is a co-authored study on neural correlates of visual implicit learning in young children with autism spectrum disorder (ASD). The goal of the study was to provide insights about pathways to core deficits in ASD through a better understanding of implicit learning, which is thought to play a critical role in social behavior [11]. Implicit learning, defined as the detection of regularities in one's environment without a conscious awareness or intention to learn [11, 12], is measured through differences in the ERP, specifically the ERP features, across multiple stimulus conditions. Although the study successfully identified differences in learning between the ASD group and typically developing (TD) controls as well as heterogeneity within the ASD group, analysis was performed exclusively in the cross-sectional domain, failing to capture the evolution of the TD and ASD implicit learning processes over the course of the experiment.

1.3 Summary of Methods Developed for ERP Analysis

In order to preserve the longitudinal information in the ERP data, a meta-preprocessing step based on applying a moving average across trials is proposed in Chapter 2. The proposed meta-preprocessing procedure strikes a balance between the need to average over trials to enhance the signal-to-noise ratio and 'over-averaging' to the point where the data for each subject are reduced to a single overall ERP and all longitudinal information is lost. This procedure is embedded in a weighted mixed effects regression framework to allow modeling of the resulting longitudinal data. Spatial observations across electrodes are incorporated into the model using a multilevel covariance structure to account for within-subject correlation. The proposed unified framework comprising the meta-processing and the weighted linear mixed effects modeling steps is referred to as MAP-ERP (Moving-Averaged-Processed ERP) throughout the dissertation. The proposed MAP-ERP procedure is able to effectively estimate the mean processes of ERP features over the course of the experiment as well as provide inference on group differences using ERP features defined at the electrode level.

MAP-ERP was designed to analyze ERP feature processes under the mean and covariance homogeneity assumption. Autism spectrum disorder contains a wide array of neurodevelopmental disorders with varying severity. The wide spectrum of autism suggests the existence of heterogeneous learning processes over time that are potentially observable through individuals' EEG. Identification of these subgroups is vital in understanding the diversity of learning within groups and identifying the differences and similarities in learning between groups. The method proposed in Chapter 3 is a robust functional clustering (RFC) algorithm applied to meta-preprocessed ERP data and is used to identify subgroups in the learning processes of ASD and TD children. Robustness in this context refers to the special care taken to incorporate information on covariance heterogeneity induced by meta-preprocessing into the RFC algorithm, setting it apart from prior functional clustering algorithms proposed in the literature. Application of RFC to the implicit learning paradigm identifies multiple clusters in the TD and ASD groups, supporting the ASD heterogeneity findings of Jeste et al. (2015) [11] and providing new insight into how these implicit learning subgroups evolve over time.

The purpose of MAP-ERP and RFC are to extract and analyze temporal information about specific features of the ERP waveforms while incorporating information on the multilevel spatial structure. Although these algorithms preserve the longitudinal and spatial dimensions of the data, analysis is focused on a single measure of the entire ERP curve, ignoring the remaining data on the functional curve and the additional information it provides. The proposed multidimensional functional principal components analysis (MD-FPCA) algorithm of Chapter 4 takes into account all three dimensions of the ERP data structure and summarizes its most important characteristics into an interpretable, low-dimensional form. Multilevel FPCA is applied to each longitudinal trial point across the functional ERP time domain and the scores from these decompositions are then modeled using a single-level FPCA across the trial domain. Resulting model parameters include the mean surfaces, eigensurfaces, eigenvalues and score distributions of the ERP processes. Due to the complexity of the MD-FPCA algorithm summarizing functional, longitudinal and spatial information, the major challenge of this “three-dimensional” approach is correctly interpreting its components. In addition to an in-depth description of the MD-FPCA algorithm, a short discussion on model interpretation in the context of the the implicit learning paradigm is provided.

MAP-ERP and RFC explore a longitudinal dimension that is typically lost in ERP studies. MD-FPCA incorporates the functional dimension of ERP into the longitudinal and spatial analysis. The methods proposed in this dissertation are able to provide insight into the learning processes of ASD and TD children and lead to interpretable group differences over time with respect to patterns of implicit learning. The dissertation explores these approaches in detail and is organized into the following chapters. Chapters 2 and 3 introduce the proposed MAP-ERP and RFC methods, respectively, as stand alone papers and can be found in Hasenstab et al. (2015) [13, 14]. A description of the MD-FPCA algorithm and the interpretation of its model components are located in Chapter 4. Discussion of the methods proposed and future directions for research are located in Chapter 5.

CHAPTER 2

Identifying Longitudinal Trends within EEG Experiments

This chapter proposes a method for preserving the longitudinal and spatial dimensions of ERP data by salvaging potentially important changes in the magnitude and form of ERP signals over the course of the experiment. This is a stand-alone paper taken from a first author publication in *Biometrics* [13]. We develop a meta-preprocessing step utilizing a moving average of ERP across sliding trial windows, to capture such longitudinal trends. We embed this procedure in a weighted linear mixed effects model to describe longitudinal trends in features such as ERP peak amplitude and latency across trials while adjusting for the inherent heteroskedasticity created at the meta-preprocessing step. The proposed unified framework, including the meta-processing and the weighted linear mixed effects modeling steps, is referred to as MAP-ERP (Moving-Averaged-Processed ERP). We perform simulation studies to assess the performance of MAP-ERP in reconstructing existing longitudinal trends and apply MAP-ERP to data from young children with autism spectrum disorder (ASD) and their typically developing counterparts to examine differences in patterns of implicit learning, providing novel insights about the mechanisms underlying social and/or cognitive deficits in this disorder.

2.1 Introduction

Both spontaneous electroencephalogram (EEG) signals and event-related potentials (ERP), which represent EEG recorded in response to stimuli, are noninvasive methods for measuring brain activity with very high time resolution. They have been in use since the 1950's in diverse biomedical settings including epilepsy, sleep disorders, multiple sclerosis, brain tumors, lesions, major affective disorder, schizophrenia, alcoholism, bipolar mood disorder, assessment of surgical outcomes, confirmation of brain death and clinical trials for drug development [1, 2, 15]. An ERP waveform (curve/signal/morphology) consists of characteristic components that span time. A typical ERP waveform from an implicit learning paradigm, containing the commonly studied P3 and N1 phasic components in this literature, is given in Figure 2.1 (a). Note that ERP waveforms can contain multiple phasic components, and the focus of analysis may be on different features in other applications. In our working example the N1 dip has a short latency (time-delay) and signifies early category recognition, while the P3 peak is task dependent due to its long latency and is traditionally related to cognitive processes such as signal matching, decision making and memory updating [11, 16, 17].

The analysis of ERP data is usually performed in the time domain. Differential brain response to sensory stimuli is very small (on the order of a few microvolts) as a fraction of spontaneous EEG, yielding a low signal-to-noise ratio (SNR) in ERP studies. To cope with this well-known phenomenon, stimuli are applied repeatedly and the resulting ERP waveforms are averaged across the trials for each subject [1, 2, 10]. Other common preprocessing steps include artifact detection (of irregularities in the signals caused by events such as blinks, saccades or muscle contractions), bad channel replacement (a smoothing procedure performed when signals from an individual scalp electrode are compromised), referencing (to standardize measurements to an average across all the electrodes) and baseline corrections (to standardize a subject's measurements to their own baseline period).

Since EEG has high time resolution and low spatial resolution (the space component corresponds to the ERP signals measured simultaneously at different electrodes placed on the skull), primary analyses typically focus on the amplitude (microvolts) and latency (milliseconds) of specific peaks (e.g., N1, P3) in the averaged ERP curves for each subject instead of spatial features [18, 19, 20, 21, 22]. As a result of averaging ERP across trials within subjects, the traditional methods fail to capture potentially

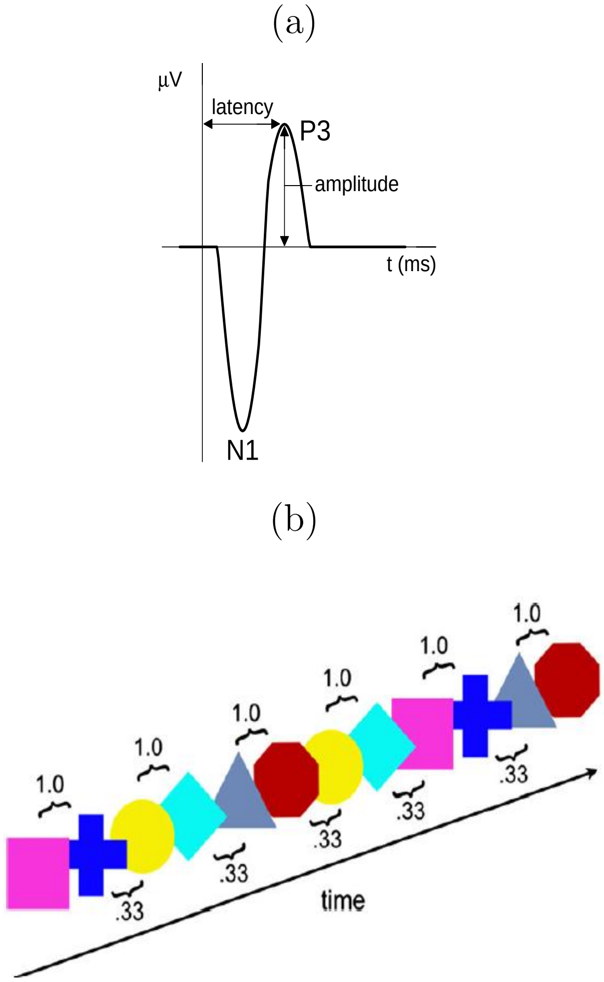


Figure 2.1: (a) A typical ERP waveform containing the P3 and N1 phasic components from the implicit learning paradigm. (b) Visualization of the implicit learning paradigm. The continuous stream of six colored shapes (pink squares, blue crosses, yellow circles, turquoise diamonds, gray triangles and red octagons) are organized into three familiar pairs. The ‘expected’ condition is defined as the transition from shape 1 to shape 2 in the familiar shape pair (with probability 1) and the ‘unexpected’ condition is defined as the transition from shape 2 to shape 1 of any shape pair (with probability 0.33).

informative changes in ERP signals that may emerge over the course of the experiment. In this paper we focus on modeling the longitudinal component of the data structure corresponding to the trials resulting from the repeated stimuli. We propose a meta-preprocessing step based on applying a moving average across trials to preserve the longitudinal information in ERP data. The proposed meta-preprocessing procedure strikes a balance between the need to average over trials to enhance the signal-to-noise ratio and 'over-averaging' to the point where the data for each subject are reduced to a single overall ERP and all longitudinal information is lost. This procedure is embedded in a weighted mixed effects regression framework to allow modeling of the resulting longitudinal data. The proposed unified framework comprising of the meta-processing and the weighted linear mixed effects modeling steps is referred to as MAP-ERP (Moving-Averaged-Processed ERP) throughout the manuscript.

Our working example is from a study on neural correlates of implicit learning in young children with autism spectrum disorder (ASD) [11]. ASD is a neurodevelopmental disorder defined by impairments in social behavior, communication and the presence of restricted interests. Implicit learning is characterized by detection of regularities in one's environment without a conscious awareness or intention to learn [11, 12, 23, 24]. The goal of the study is to provide insights about pathways to core deficits in ASD through a better understanding of implicit learning, which is thought to play a critical role in social behavior [11]. The study involved 2 to 5 year old typically developing (TD) and ASD children, exposed to a continuous stream of six colored shapes (pink squares, blue crosses, yellow circles, turquoise diamonds, gray triangles and red octagons; see Figure 2.1 (b)). The shapes were organized into three pairs such that the sequence within the pairs was always the same but the pairs themselves occurred in random order across the experiment. For example, a pink square was always followed by a blue cross (pair) but the symbol that followed a blue cross could be any of the three 'first' symbols of a pair. As a result, the transitions within a pair (pink square to blue cross) were 'expected' or could be 'learned' while the transitions between pairs were 'unexpected' or not predictable. Each transition from one shape to the next was considered a stimulus, resulting in an ERP waveform. Differences in the ERP signals (amplitude, shape, timing) between the expected and unexpected trials were thought to reflect the degree of implicit learning in young children while longitudinal changes in this contrast would indicate how the learning process evolves. The paradigm included 120 trials (repeated stimuli) for each of the two conditions, resulting in 240 ERP waveforms per child.

The goal of the original study was to look at differences in implicit learning between TD and ASD children in order to provide insights about mechanisms underlying social and/or cognitive deficits in this disorder. While the traditional analysis simply compares the average difference between expected and unexpected trials, determining whether the groups differ in terms of how the ERP signals change as the children learn the shape patterns over the course of the task is also important, e.g., for seeing whether implicit learning occurs at different speeds. Due to the low signal-to-noise ratio, it is also not feasible to analyze the original longitudinally collected ERP curves on a trial by trial basis. As an example, in Figure 2.2 (top left panel), an ERP waveform is plotted for one subject from a single trial in the right frontal region of the brain. Due to the low signal-to-noise ratio in a single trial, the N1 dip and P3 peak are not recognizable in their standard respective time intervals as defined in Jeste et al. (2015) [11]. Now consider an average of 30 ERP waveforms from adjacent trials, plotted in Figure 2.2 (top right panel). The N1 dip and P3 peak are easily recognizable due to the enhancement in the signal-to-noise ratio. Our proposed meta-preprocessing procedure relies on moving averages of ERP waveforms over trials to preserve trends in implicit learning, focusing on the longitudinal analysis of features such as the P3 amplitude. These features are then assessed via a weighted mixed effects model, which adjusts for inherent heteroskedasticity created at the meta-preprocessing step.

Previous studies in neuroscience and biomedical engineering have acknowledged that ERP morphology may change over the course of a task. However, most prior work has focused on controlling for longitudinal trends [25, 26, 27, 28, 29, 30] rather than viewing them as one of the central research questions. Moreover, most of the work on modeling longitudinal trends has been limited to parametric forms [31, 32, 33, 34, 35]. In contrast, our proposed meta-preprocessing step, utilizing the moving average idea

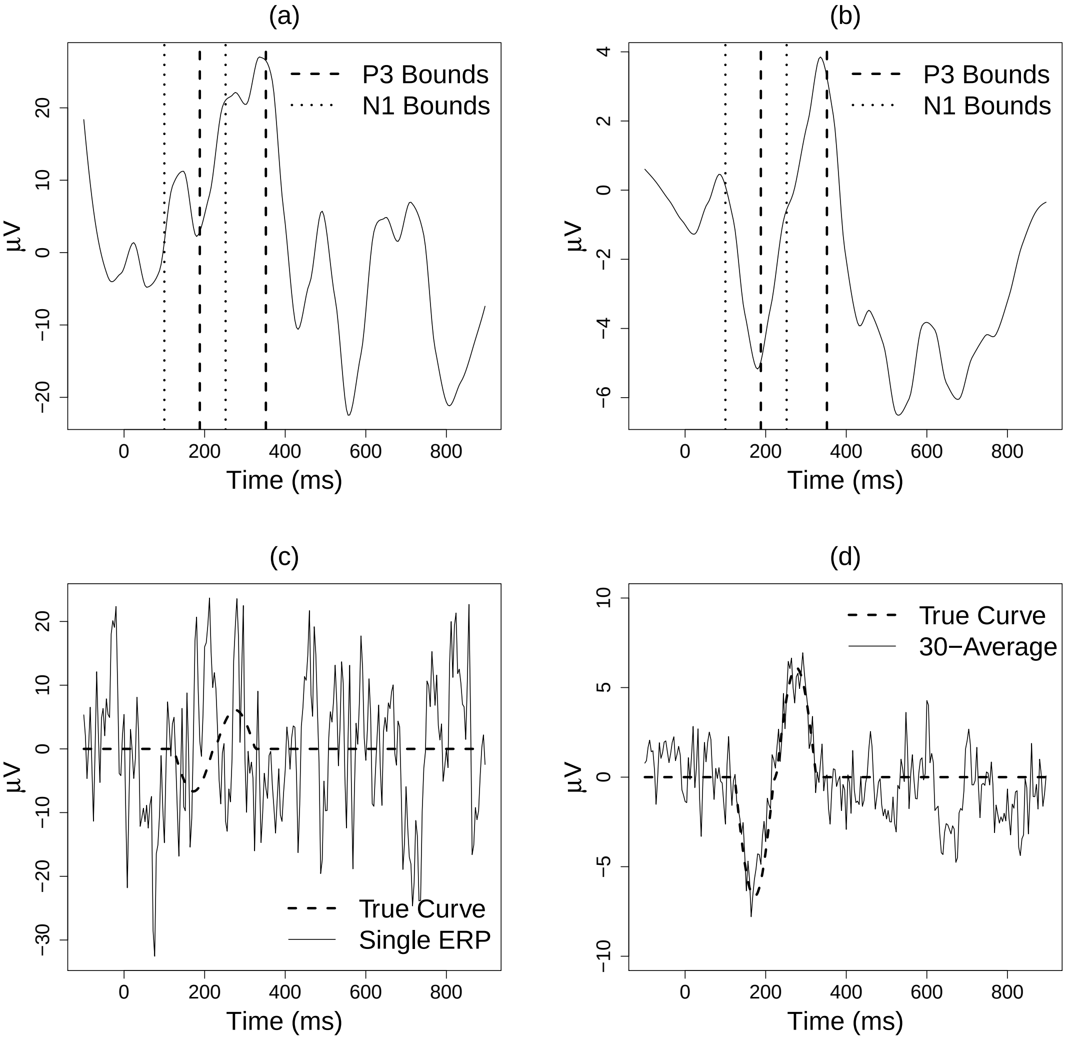


Figure 2.2: (a) ERP waveform for a single subject, trial, electrode and condition in the frontal region after preprocessing, (b) average of the first 30 consecutive ERP waveforms for a single subject, electrode and condition, (c) simulated ERP waveform for a single subject, trial and electrode, (d) average of 30 simulated ERP waveforms for a single subject and electrode. Vertical boundaries denoted by the dotted and dashed lines in (a) and (b) correspond to the search of the locations of the N1 and P3 components and are [100ms, 250ms] and [190ms, 350ms], respectively (Jeste et al., 2014).

described above, does not make any parametric assumptions. It is only after extracting the longitudinal features of the ERP morphology via a flexible nonparametric approach that we utilize a parametric model, the proposed weighted linear mixed effects regression, to describe changes in ERP features over the trials. Techniques such as sub-ensemble averaging and block averaging, which combine disjoint subsets of ERP waveforms across trials, require long recording sessions and still result in coarse-grained repeated measurements that do not take full advantage of the information on continuous longitudinal trends that is inherently available in the raw data [36, 37, 38]. These studies provide snapshots of the longitudinal information, while the proposed moving average captures the continuum of longitudinal dynamics.

The paper is organized as follows. In Section 2.2, we outline a novel meta-preprocessing step for ERP experiments which preserves the longitudinal structure of the data while enhancing the signal-to-noise ratio. A weighted linear mixed effects modeling approach following the meta-preprocessing step of Section 2.2 is outlined in Section 2.3. In Section 2.4, we present a simulation to assess the performance of MAP-ERP, the proposed unified modeling framework. The analysis of the ERP data from the motivating implicit learning study on ASD is given in Section 2.5. We conclude with final remarks in Section 2.6.

2.2 Proposed meta-preprocessing step to preserve longitudinal information in ERP data

Our working example is the first to date to study implicit learning in young children with ASD via the use of EEG. Prior studies on implicit learning have typically focused exclusively on high-functioning children with ASD, as they can more readily follow directions and engage in paradigms requiring sustained attention [12, 39, 40, 41, 42]. Use of EEG, which does not require as much direct engagement, made it possible to include a younger group of children with ASD, with a focus on characterizing differences in implicit learning patterns that will distinguish them from typically developing peers. EEG data were recorded for 120 trials per condition (expected and unexpected) for each subject at 128 electrodes and were preprocessed using NetStation 4.4.5 software (Electrical Geodesics, Inc.). The standard preprocessing steps outlined in Web Appendix A produced trajectories of ERP waveforms for 37 ASD and 34 TD children to be fed into the proposed meta-preprocessing algorithm. The number of trials with usable data ranged from 10 to 120 per subject per condition. The EEG signals were sampled at 250Hz, producing 250 within-trial time points per waveform spanning 1000ms. To illustrate the methods, we focus on data from 12 electrodes in the frontal region which are described in Jeste et al. (2015) [11].

The proposed meta-preprocessing step is a moving average designed to increase the signal-to-noise ratio of ERP to a level at which curve attributes such as peak amplitudes and latency are identifiable, while preserving longitudinal information. Let $X_{ijk\ell}(t)$ represent the micro-voltage of the ERP of subject i from electrode j, on trial k in condition ℓ (expected/unexpected) observed at time t, $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k \in K_i$, $\ell \in L_{ik}$ and $t = 1, \ldots, T$, where N is the total number of subjects, J is the total number of electrodes, T is the total number of time points within a trial, K_i is the set of non-missing trials for subject i and L_{ik} is the set of non-missing conditions at trial k for subject i. The maximum number of conditions per subject per trial, denoted by L, is 2 in our application. The maximum possible number of trials per subject per condition, denoted by K, is 120. Further, let B_k represent overlapping sets of trials of varying lengths with the maximum number of trials within a set denoted by b,

$$B_k = \begin{cases} [1, 2k-1], & \text{if } k < \frac{b}{2}, \\ [k - b/2 + 1, k + b/2], & \text{if } \frac{b}{2} \le k \le K - \frac{b}{2}, \\ [2k - K, K], & \text{if } j > K - \frac{b}{2}. \end{cases}$$

Sets not on the boundary contain b elements and the number of elements shrink linearly towards 1 at the boundaries. The B_k are used as sliding trial windows in the moving average of ERP within the meta-preprocessing step. Since the goal is to extract continuous longitudinal trends within EEG experiments, including overlapping sets of trials helps to target the continuum of features across several ERPs. Alternative options include kernel smoothing; we chose to utilize a moving average in our applications to simplify the quantification of heterogeneity in these averages due to missingness. As will be outlined in Section 2.3, in the proposed weighted linear mixed effects model framework, longitudinal attributes captured from averaging over smaller numbers of ERP waveforms, as in the boundary sets or sets with a larger proportion of missing trials, will receive lower weights. This will help to mitigate boundary effects and place more weight on intervals with more trials, and hence more information.

In the algorithm introduced below, we let $\widetilde{X}_{ijk\ell}(t)$ represent the cross-sectional averages of ERPs within the sliding trial windows B_k and let $\widetilde{Y}_{ijk\ell}$, $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k \in M_i$, $\ell \in Q_{ik}$, represent longitudinal features (such as amplitude or latency of peaks) of ERPs captured from these cross-sectional averages, where M_i is the set of non-missing trials for subject i and Q_{ik} is the set of non-missing conditions for subject i at trial k after meta-preprocessing. Note that the notation $Y_{ijk\ell}$ is reserved to represent the true features of the underlying ERP signal to be used in subsequent sections and that different set notations, M_i and Q_{ik}, are used to index the observed trials for subject i and the observed conditions for subject i at trial k, respectively, to accommodate differences in the missingness structure which may be induced by the proposed meta-preprocessing step on the longitudinal features. The meta-preprocessing is applied separately for each subject, electrode and condition. The proposed algorithm can be summarized in the following steps.

1. For fixed i, j, k and ℓ, consider the set of ERP waveforms $\{X_{ijk\ell}(t) : k \in B_k\}$. If the set is empty, $\widetilde{Y}_{ijk\ell}$ will be missing and the algorithm proceeds to consider the set associated with trial $k+1$.

2. Calculate the cross-sectional mean curve, $\widetilde{X}_{ijk\ell}(t) = (1/c_{ijk\ell}) \sum_{k \in B_k} X_{ijk\ell}(t)$, of the subsetted waveforms, where $c_{ijk\ell}$ equals the number of trials in the set $\{X_{ijk\ell}(t) : k \in B_k\}$.

3. Smooth $\widetilde{X}_{ijk\ell}(t)$ over t to identify the locations of the ERP features of interest.

4. Use the locations from step 3 to define the longitudinal features $\widetilde{Y}_{ijk\ell}$ for $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k \in M_i$, $\ell \in Q_{ik}$. If the feature of interest is a peak

latency, then step 3 will be enough for the assignment of $\widetilde{Y}_{ijk\ell}$; if on the other hand it is a peak amplitude, then the locations from step 3 will be used for assignment of peak amplitudes. The assignment of peak amplitudes will utilize the values of the original cross-sectional mean waveforms, $\widetilde{X}_{ijk\ell}(t)$, rather than those from their smoothed versions in order to minimize bias.

5. Repeat steps 1 through 4 for $k = 1, \dots, K$.

The meta-preprocessing step extracts features from the dense ERP curves using a moving average across trials k, producing longitudinal features across trials for each subject, electrode and condition. Since the analysis after the meta-preprocessing is performed on the longitudinal time component (trials), the argument (t) is omitted from $\widetilde{Y}_{ijk\ell}$, denoting the longitudinal features. We use a loess smooth in the third step of the proposed algorithm where the bandwidth is selected via 10-fold cross-validation. To identify the feature locations on the smooth, which correspond to the locations of the P3 component in our application, we utilize a peak detection algorithm which identifies optima within the time interval $t \in [190\text{ms}, 350\text{ms}]$ [11]. If the peak location is on the boundary of the specified time interval, the time interval is gradually widened in the direction of the initial boundary peak until a new peak is identified that is not on the boundary [11]. If the interval is widened to twice the range of the initial interval and the peak still lies on the boundary, the peak observation is considered missing. While the locations of the features are identified on the smoothed data to minimize the effects of random noise, the features themselves, such as peak amplitudes, are assigned on the original cross-sectional averages to minimize bias. In our application, the size of the sliding trial window, b, is chosen to be 30, which corresponds to the minimum number of curves required to increase the signal-to-noise ratio to a level where the desired features are recognizable. However we also include a sensitivity analysis to show that the data analysis results are sufficiently robust to this choice. For cases where the focus of interest may be the entire ERP waveform instead of the particular features, automatic selection of b was proposed by Turetsky et al. (1989) [30], based on minimizing the estimated mean average square error of the smoothed data.

2.3 Analysis of meta-preprocessed ERP data via a weighted linear mixed effects model

Mixed effects regression is a powerful modeling tool which accounts for multi-level heterogeneity and complex temporal trends via the use of fixed and random effects. We propose a weighted linear mixed effects model to analyze the longitudinal features $\widetilde{Y}_{ijk\ell}$, $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k \in M_i$, $\ell \in Q_{ij}$, extracted by our meta-preprocessing algorithm. We reiterate that the longitudinal components of the proposed linear mixed effects model are the extracted features defined across trials k (not ERP time t). Our application focuses on ERP data from 12 electrodes in the front of the scalp, from left, middle and right regions. Each region contains four electrodes and electrodes within each region are spatial neighbors. Our goals in the application to the implicit learning paradigm are to model the dynamics of P3 amplitude across trials and to study differences between the TD and ASD groups. We utilize multi-level random effects at the subject and electrode region levels to model dependency of the data for a given subject in a particular electrode region where the spatial correlations are the strongest. We further utilize spline basis functions in modeling both fixed and random effects to describe the functional dependency of the P3 amplitudes across trials in a reduced dimensional space.

In addition, our proposal includes an adjustment to the standard linear mixed effects framework to account for the heteroskedasticity induced by averaging different numbers of trials at different time points during the meta-preprocessing step. More specifically, consider the plot of the P3 amplitudes (denoted by $\widetilde{Y}_{ijk\ell}$) produced by the proposed meta-preprocessing algorithm against values of $c_{ijk\ell}$, the number of ERP waveforms averaged in the B_k trial window in step 2 of the algorithm, displayed in Figure 2.3 (a). There is a sharp decrease in the variance of the amplitudes with increasing $c_{ijk\ell}$, suggesting that features extracted from averages in sliding trial windows with fewer ERP waveforms are less precise, as would be expected. In order to account for this systemic heteroskedasticity, separate variance components are allowed for different bins of $c_{ijk\ell}$ in the linear mixed effects formulation below. Note that similar weighting ideas have been considered previously to correct for heteroskedasticity in the context of averaging of ERP curves but only for cross-sectional data [43].

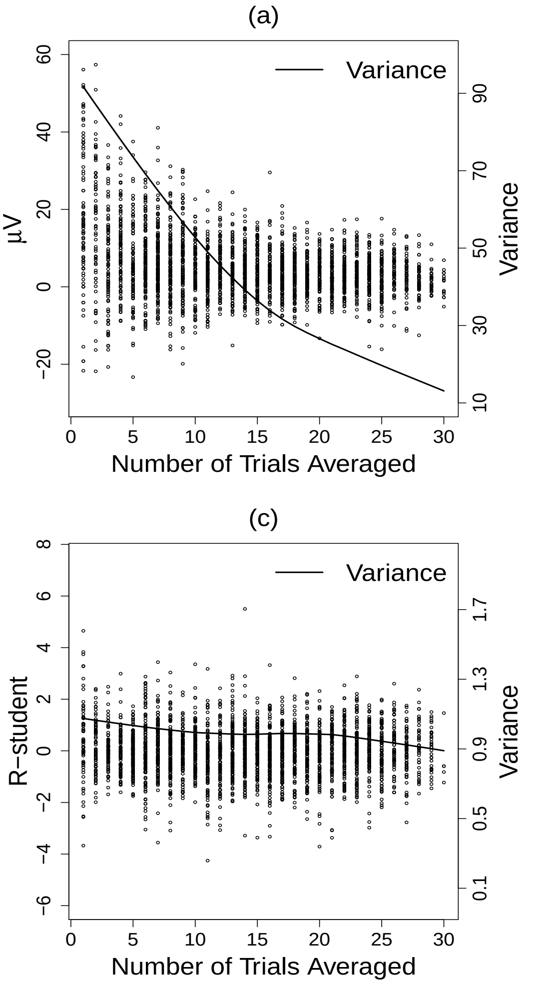


Figure 2.3: (a) P3 amplitudes as a function of number of trials averaged, $c_{ijk\ell}$, in the analysis of the ERP data from the implicit learning paradigm, (b) studentized residuals obtained from the mixed effects model without weights as a function of $c_{ijk\ell}$, (c) studentized residuals from the proposed weighted mixed effects model as a function of $c_{ijk\ell}$. Variance of the P3 amplitudes and the residuals are shown (solid line) corresponding to the y-axes given on the right hand side.

Let s index a partition (grouping/binning) of the range of $c_{ijk\ell} = 1, \ldots, b$. The choice of the total number of bins will be a trade-off between model parsimony and realistic representation of the heteroskedasticity in the longitudinal features. Let $\widetilde{Y}^{(s)}_{ij(r)k\ell}$ represent the longitudinal features of the meta-preprocessed ERP, from subject i, at electrode j within region r, with trial window B_k and condition ℓ, where the number, $c_{ijk\ell}$, of ERPs averaged in B_k falls into the sth partition. Note that the region index r

is added in this notation, since our proposed modeling will be addressing dependencies in the data within electrodes from the same part of the scalp. We model the $\widetilde{Y}^{(s)}_{ij(r)k\ell}$ using fixed effects and a two level random effects structure (subject and region). Fixed effects parameters include an intercept, trial (represented by a natural cubic B-spline with 5 knots), condition (expected vs. unexpected) and group (ASD vs. TD), along with all two-way and three-way interactions, totaling 24 fixed effects components. Let $\boldsymbol{w}_{ik\ell}$ be the 1×24 row vector corresponding to trial k and condition ℓ of the fixed effects matrix $\boldsymbol{W}_i$ and let $\boldsymbol{\beta}$ be the 24×1 column vector of fixed effects parameters. Further let $\boldsymbol{z}_{ij(r)k}$ represent the 1×6 row vector (including an intercept and trial, represented by a natural cubic B-spline with 5 knots) corresponding to trial k of the random effects matrix $\boldsymbol{Z}_{ij(r)}$ for electrode j in region r. Index ℓ for condition is not needed for $\boldsymbol{z}_{ij(r)k}$ in our application, since the random effects design matrix contains spline bases to model within subject correlations over trials. Vectors $\boldsymbol{b}_i$ and $\boldsymbol{b}_{ir}$ of dimension 6×1 represent subject and region level random effects. We model $\widetilde{Y}^{(s)}_{ij(r)k\ell}$ by

$$\begin{aligned}
&\widetilde{Y}^{(s)}_{ij(r)k\ell} = \boldsymbol{w}_{ik\ell}\boldsymbol{\beta} + \boldsymbol{z}_{ij(r)k}\boldsymbol{b}_i + \boldsymbol{z}_{ij(r)k}\boldsymbol{b}_{ir} + \varepsilon^{(s)}_{ij(r)k\ell}, \\
&\boldsymbol{b}_i \sim MVN(\boldsymbol{0}, \boldsymbol{D}^1_{6\times 6}), \quad \boldsymbol{b}_{ir} \sim MVN(\boldsymbol{0}, \boldsymbol{D}^2_{6\times 6}), \quad \varepsilon^{(s)}_{ij(r)k\ell} \sim N\left(0, \sigma^2_s\right),
\end{aligned} \tag{2.1}$$

where $\boldsymbol{D}^1$ and $\boldsymbol{D}^2$ represent the random effects covariance matrices at the subject and region levels, respectively, and $\varepsilon^{(s)}_{ij(r)k\ell}$ represents the error term for the sth partition with variance σ^2_s and uncorrelated over different subjects, regions, electrodes, trials and conditions. The subject and region level random effects are assumed to be independent of the error term, leading to the following covariance structure: $\mathrm{Var}(\widetilde{Y}^{(s)}_{ij(r)k\ell}) = \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^1\boldsymbol{z}'_{ij(r)k} + \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^2\boldsymbol{z}'_{ij(r)k} + \sigma^2_s$, $\mathrm{cov}(\widetilde{Y}^{(s)}_{ij(r)k\ell}, \widetilde{Y}^{(s)}_{ij(r)k'\ell}) = \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^1\boldsymbol{z}'_{ij(r)k'} + \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^2\boldsymbol{z}'_{ij(r)k'}$, $\forall i, j(r), k \neq k'$ and $\mathrm{cov}(\widetilde{Y}^{(s)}_{ij(r)k\ell}, \widetilde{Y}^{(s)}_{ij'(r)k\ell}) = \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^1\boldsymbol{z}'_{ij'(r)k} + \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^2\boldsymbol{z}'_{ij'(r)k}$, $\forall i, j(r) \neq j'(r), k$, for within region correlation and $\mathrm{cov}(\widetilde{Y}^{(s)}_{ij(r)k\ell}, \widetilde{Y}^{(s)}_{ij'(r')k\ell}) = \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^1\boldsymbol{z}'_{ij'(r')k}$, $\forall i, j(r) \neq j'(r'), k$, for within subject across region correlation.

We use a common design matrix $\boldsymbol{Z}_{ij(r)}$ for the subject and region random effects but they can be taken to be different in other applications. The degree of smoothness for the natural B-splines is determined through the selection of the number of knots which are typically equispaced. Automatic selection methods for the number of knots include AIC, BIC and cross-validation [44, 45]; we use AIC in our applications. Note that the mixed effects framework can easily be extended to adapt to the application at hand, for example by modeling further dependency structures or including baseline covariates. Restricted maximum likelihood (REML) is used to estimate model (2.1) due to its ability to produce unbiased estimates of the variance and covariance parameters. The proposed framework for addressing heteroskedasticity by using separate variance components assigns higher weights to the outcome values with lower variability (i.e. from trial windows with more ERP waveforms and therefore more information), and hence is referred to as a weighted model.

2.4 Simulation

We conduct simulations to study the performance of MAP-ERP, including both the meta-preprocessing and the weighted mixed effects modeling, in reconstructing longitudinal trends over trials in ERP studies. In particular, we include comparisons of our approach with a simple procedure that uses raw single-trial ERP to model longitudinal trends under various signal-to-noise ratios (SNR). Results are presented for four sample sizes, $N = 20, 40, 80$ and 160, which are typical for ERP studies, and are based on 200 Monte Carlo runs. Detailed explanations of the simulation setup along with definitions of mean error (ME) and prediction error (PE) used to evaluate proposed methodology are given in Web Appendix B.

The medians, 2.5th and 97.5th percentiles of the MEs and PEs for the two modeling procedures are given in Table 2.1 for varying SNRs and sample sizes. As expected, since the meta-processing step enhances the overall SNR, MAP-ERP leads to consistently smaller ME and PE compared with the single-trial approach across all SNR settings and sample sizes. This is also displayed in Figure 2.4, which gives the estimated fixed effects means and pointwise confidence intervals from the run with the median ME value for each approach for three SNR settings at $N = 40$. Asymptotic pointwise

Table 2.1: Median and (2.5th, 97.5th) percentiles of simulation performance metrics (ME and PE) for varying SNRs from 200 Monte Carlo runs with sample sizes $N = 20, 40, 80$ and 160.

		ME		PE	
N	SNR	Single	MAP-ERP	Single	MAP-ERP
20	0.4	.531 (.493, .563)	.040 (.020, .067)	.526 (.476, .572)	.164 (.141, .200)
	0.6	.225 (.190, .258)	.032 (.014, .069)	.222 (.194, .253)	.116 (.100, .137)
	0.8	.108 (.073, .148)	.039 (.013, .076)	.111 (.097, .129)	.094 (.082, .108)
40	0.4	.530 (.503, .558)	.033 (.019, .053)	.528 (.498, .559)	.164 (.144, .187)
	0.6	.224 (.198, .248)	.030 (.014, .058)	.221 (.205, .243)	.115 (.104, .129)
	0.8	.109 (.081, .135)	.037 (.013, .064)	.112 (.102, .122)	.093 (.085, .102)
80	0.4	.530 (.501, .553)	.029 (.019, .043)	.529 (.507, .551)	.164 (.150, .180)
	0.6	.222 (.203, .242)	.030 (.016, .049)	.223 (.210, .236)	.115 (.108, .125)
	0.8	.107 (.087, .127)	.037 (.021, .058)	.112 (.104, .119)	.093 (.087, .100)
160	0.4	.531 (.503, .548)	.026 (.019, .034)	.528 (.515, .545)	.163 (.152, .175)
	0.6	.224 (.203, .236)	.029 (.017, .045)	.222 (.214, .232)	.115 (.109, .122)
	0.8	.109 (.090, .122)	.037 (.025, .054)	.111 (.107, .117)	.093 (.089, .098)

confidence intervals for the estimated fixed effects means are formed using variance-covariance estimates of the model components. The true fixed effects mean trajectory lies within the 95% pointwise confidence intervals based on MAP-ERP and outside those based on the single-trial approach, further emphasizing the effectiveness of MAP-ERP at capturing longitudinal trends.

The single-trial modeling approach shows a clear improvement in both ME and PE with increasing SNR, but these summary metrics do not change with increasing sample size. For low SNRs, the single-trial approach primarily captures the noise component of the raw ERP, producing highly inflated and inaccurate amplitude estimates. As SNR increases (i.e. as noise in the raw ERP reduces), the amplitude estimates produced by the single-trial approach shrink towards the true values as depicted by the relatively sharp decreases in the two performance metrics. However the single trial approach still cannot handle the noise in the ERP trajectories effectively enough where the performance can show improvement within the considered range of sample sizes.

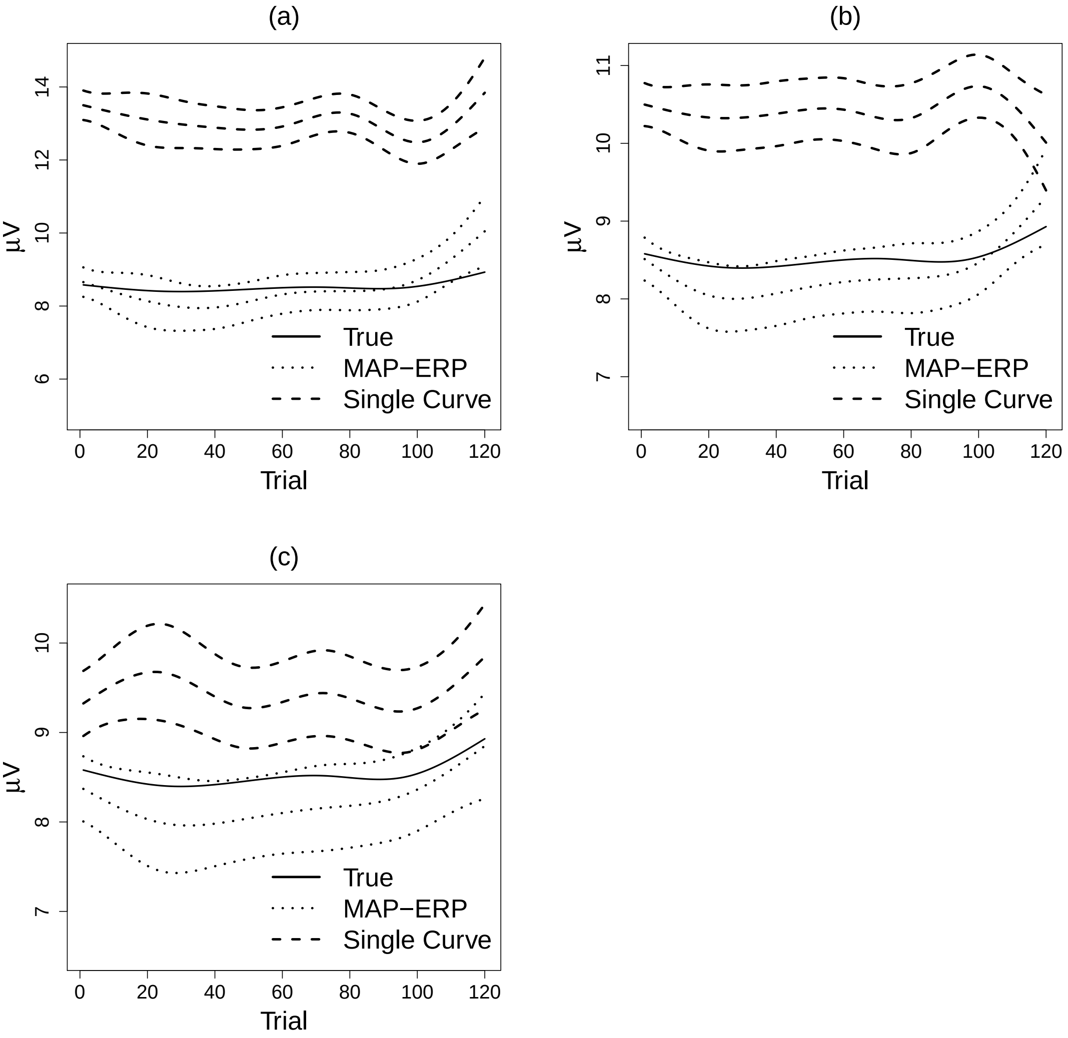


Figure 2.4: Estimated fixed effects mean trajectories along with 95% confidence intervals corresponding to the median of the ME in simulations based on the single curve (trial) approach and MAP-ERP for (a) SNR=0.4, (b) 0.6 and (c) 0.8.

The ME values for MAP-ERP are consistently low regardless of the varying SNR, showing the effectiveness of the proposed meta-preprocessing step in enhancing the signal, especially compared to the single-trial approach. Even though the proposed method is effective for a wide range of SNRs, there are still some subtle issues. Specifically, there are two opposing dynamics, noise and latency jitter, which are still observable in the estimated mean fixed effects trajectories. While noise tends to inflate the amplitude estimates, latency jitter (the fact that the timing of the P3 peak differs across trials and subjects) tends to dampen the amplitude estimates due to possible misalignment in averaging. As illustrated in Figure 2.4 (a) for SNR=0.4, noise and latency jitter effects roughly cancel each other and the estimated mean appears to have little visible bias. In contrast, the effects of latency jitter dominate for SNR$=0.6, 0.8$ and the estimated mean curve consistently lies below the true curve. The effect of sample size on ME of MAP-ERP is observable at SNR=0.4 where the noise effects cancel with the latency effects; in this case ME decreases with increasing sample size as expected. However, since increasing sample size does not particularly help with the effects of latency jitter, ME does not significantly improve with increasing sample size for SNR=0.6 or 0.8. We ran additional simulations without latency jitter to solidify the above interpretations on bias with varying SNR and sample size (Web Appendix C). Results reveal that ME decreases with increasing SNR and sample size as expected when latency effects are not present, dampening the mean amplitude estimates. Dampening effects of the latency jitter are discussed further in the Discussion Section.

Similarly, the PE for MAP-ERP improves with increasing SNR due to the reduced levels of noise, but is constant over sample sizes due to the flexible modeling of subject specific effects in the proposed mixed effects model via the inclusion of the spline terms. The proposed modeling allows for reconstruction of complex subject specific functional trends which leads to small PE values across sample size, especially compared to the single curve approach.

2.5 Analysis of the ERP data from the implicit learning paradigm

In order to compare the ASD and TD groups in the expected and unexpected conditions over the course of the task, the longitudinal P3 peak amplitudes $\widetilde{Y}^{(s)}_{ij(r)k\ell}$ from 12 electrodes were obtained using our proposed meta-processing procedure for $N = 71$ subjects with up to $K = 120$ trials obtained per condition and were then analyzed via the weighted linear mixed effects model described in Section 2.3. Five knots were used for the spline fits. This choice was made using a combination of AIC and subject matter expertise, and was found to reflect the complexity of the data well without undersmoothing. Six equal sized groups (i.e. $s = 1, \ldots, 6$) of $c_{ijk\ell} \in [1, 30]$ were used to model effects of heteroskedasticity. The random effects for the subject and region levels were allowed to have a unique variance for each random effect but were assumed to be independent. The number of averaged curves $c_{ijk\ell}$ was also included as a predictor in the mixed model but was not found to be significant and is hence omitted from the final analysis presented here. For comparison, we model P3 amplitudes using both weighted and unweighted linear mixed effects models, where the unweighted model does not allow for separate variance components with varying $c_{ijk\ell}$. Both models were fit using SAS PROC MIXED-REML.

The two model fits lead to largely similar fixed effects mean trajectories except at the boundaries where the unweighted fit displays more extreme values. This is not surprising since the reduction in $c_{ijk\ell}$ at the boundaries produces noisy cross-sectional averages and inflated amplitude estimates. The weighted model effectively stabilizes amplitude estimates by down-weighting these noisy observations. In addition, even though the difference is small in magnitude, the standard errors for the fixed effects estimates from the weighted model are consistently smaller than those from the unweighted model. In standard regression analysis, conventional standard errors bias up when observations with high leverage (far from predictor mean) are associated with smaller residual variance. Consistent with this observation, most of the spline basis predictor observations with high leverage are associated with smaller residual variance. The variances of the studentized residuals from the weighted model appear to be relatively constant across $c_{ijk\ell}$ compared to those from the unweighted model which have a strong downward trend in variance for increasing $c_{ijk\ell}$. This implies that the weighted model has adjusted for the heteroskedasticity in the data effectively.

Since the main interest of the original study was in comparing ASD to TD subjects, we present results in terms of the differences on condition differences in amplitudes (i.e. (ASD expected - unexpected) - (TD expected - unexpected)). In addition to pointwise confidence intervals formed based on estimates from the mixed effects model, we also form 90% bootstrap percentile confidence bands based on 200 bootstrap samples drawn from the original subjects' ERP trajectories with replacement. Bootstrap CI's account for the entire two-step procedure as well as variability in the sampling of the ASD and TD subjects. The contrast resulting from averaging 30 ERP in sliding trial windows in the meta-preprocessing step is displayed in Figure 2.5 (c). During the first 60 trials, the ASD group appears to have a larger amplitude difference than the TD group with a maximum difference between conditions around trial 30 for both groups. Group differences are found to be reliably significant based on the bootstrap bands between trials 20 and 50. Although both ASD and TD subjects appear to be able to differentiate between expected and unexpected conditions, implying implicit learning is taking place, and they do so at a similar speed, the pattern of discrimination seems to differ between the two groups. While the expected minus unexpected condition mean difference is positive for the ASD group, it is negative in TD children (plots not displayed). The absolute condition difference in the amplitudes remains smaller after trial 60 until the end of the experiment. One possible interpretation is that children may be less engaged in the task after this point in the trial. These results are consistent across different window sizes b =20, 30 and 40 for the proposed meta-preprocessing step (Figures 2.5 (b) through (d)). Hence, inferences based on longitudinal data produced by the meta-preprocessing algorithm appear fairly robust to moderate changes in window size in our application. We also display group differences based on analyzing single-trial data in Figure 2.5 (a). The bootstrap bands for this approach are very wide due to the low signal-to-noise ratios of the empirical ERP and the ampli-

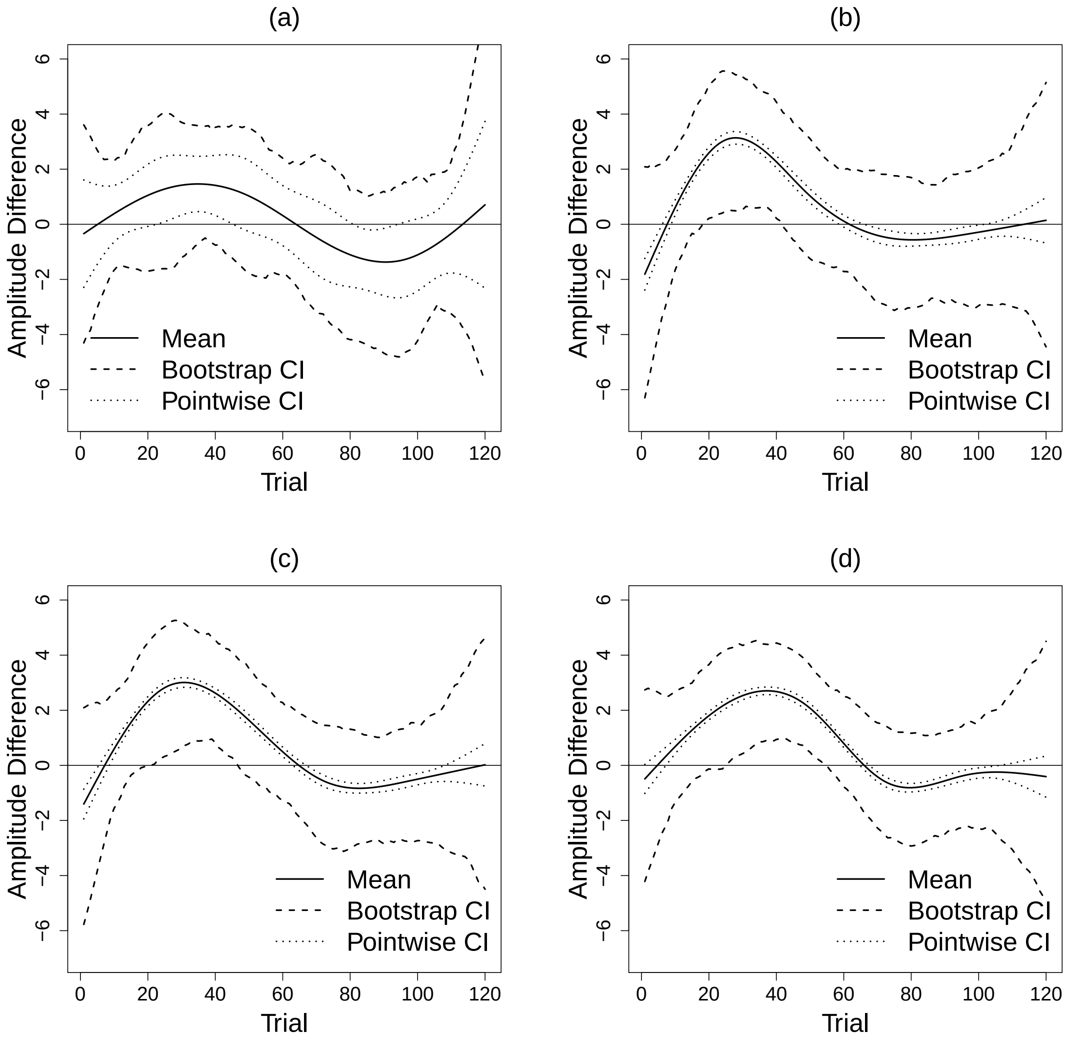


Figure 2.5: Estimated mean group and condition difference ((ASD expected - unexpected) - (TD expected - unexpected)) trajectories based on the (a) single-trial approach and MAP-ERP with window sizes (b) b=20, (c) b=30 and (d) b=40. 90% bootstrap bands from 200 runs are also provided (dashed lines).

tude differences between groups are no longer significant. In contrast MAP-ERP is able to identify the regions of significant group differences due to the increase in the signal-to-noise ratio.

We highlight that previously published results on the ERP data from this implicit learning paradigm completely ignored the longitudinal component of the data due to averaging over trials [11]. In addition, the single-trial approach is too noisy to depict any group differences across trials. In contrast, MAP-ERP approach provides novel insights on a new longitudinal dimension that is typically lost in analysis of ERP data, leading to interpretable group differences over trials with respect to patterns of implicit learning.

2.6 Discussion

We have proposed a meta-preprocessing procedure for ERP studies that enhances the signal-to-noise ratio while still retaining longitudinal trends across trials. Longitudinal features may be an important focus in various ERP studies, such as the implicit learning paradigm analyzed in this manuscript where speed of acquisition contributes significantly to the characterization of implicit learning in young children with ASD compared to typically developing controls. A recent work focused on longitudinal trends in a local field potentials study [46]. While the proposal focuses on repeated ERP signal recorded in response to multiple applications of a stimuli, the proposed methodology is applicable more broadly to studies involving repetitions of a systematic signal observed with noise, such as heart beat, breath cycle or eye blinks. Following the meta-preprocessing step, we also proposed a weighted linear mixed effects model that is shown to describe longitudinal trends in ERP features effectively in simulation studies.

The proposed method is a two-step approach comprising of a meta-preprocessing step followed by a weighted linear mixed effects model. One reason for using a two-step approach was to connect to the rich literature on the analysis of ERP data, which focuses on modeling particular features of ERP curves (e.g., amplitudes, latency) that are readily interpretable scientifically. Our meta-preprocessing step enables researchers to focus on these readily interpretable features, while preserving longitudinal trends over trials in contrast to the common practice in this literature. Another advantage of the two-step approach is that the initial novel meta-preprocessing step is left modular to be flexibly coupled with an array of secondary analyses options for the extracted features. One limitation of the current approach based on moving averages is that misalignment of the ERP features can potentially lead to slight underestimation of the true peak amplitudes. This is a challenging issue, since alignment would de facto require that features be identified before averaging, which is impractical due to noise levels in the individual raw ERP. Even though we were able to draw combined inference for the two steps of MAP-ERP via bootstrap confidence intervals in our application, we identify the development of formal inference procedures as a direction for further research.

While the current manuscript focuses on analyzing particular features of ERP such as peak amplitudes or latencies, which is a common practice in the EEG literature, we note that recent studies have proposed using functional data analysis for analysis of the ERP curves in their entirety [16, 47]. The proposed meta-preprocessing algorithm creates longitudinal functional data, i.e. repeated ERP waveforms over the trials. Analysis of such data using functional techniques is a very interesting open problem.

Chapter 2 Appendices

Appendix A: Preliminary preprocessing steps of the EEG data from the implicit learning paradigm

Our working example on visual implicit learning included EEG recorded for 120 trials per condition (expected and unexpected) for each subject using 128 electrodes and were preprocessed using NetStation 4.4.5 software (Electrical Geodesics, Inc.). The raw profiles were first run through a 0.3-50 Hz bandpass filter to smooth out noise below and above this range of frequencies. The resulting EEG profiles were then segmented by condition into event-locked trials of 1000ms total duration with a pre-stimulus baseline period of 100ms used for baseline correction. Using an automated artifact detection tool, trials with more than 15% of electrodes having amplitude differences larger than 150μV (max-min $>$ 150 microvolts) were classified as bad trials and were omitted from further analysis. ERP waveforms were then visually inspected to remove any remaining trials which contained eye artifacts such as blinks, muscle contractions or saccades. Trials with fewer than 15% bad electrodes underwent bad channel replacement, a data smoothing procedure using a spherical spline interpolation algorithm [48], while subjects with more than 15% bad electrodes were discarded. Subjects with fewer than

10 good trials per condition were excluded from subsequent analyses. Finally, the EEG data were re-referenced and baseline corrected.

Appendix B: Simulation setup

In our simulations we utilize EEGLAB [10], a software package in Matlab, to simulate the true underlying ERP signals and background ERP noise, the two primary components needed to mimic realistic ERP waveforms at varying signal-to-noise ratios. Our application to implicit learning focuses on longitudinal trends in the amplitude of the P3 peak, but the literature suggests that ERP waveforms from the frontal electrodes also contain a prominent N1 dip [11]. We begin by generating peak and dip amplitudes for the P3 and N1, respectively, for a true underlying ERP signal under a single group and condition (hence the omission of subscript ℓ) using two separate linear mixed effects models

$$
\begin{aligned}
&Y_{ij(r)k} = \boldsymbol{w}_{ik}\boldsymbol{\beta} + \boldsymbol{z}_{ij(r)k}\boldsymbol{b}_i + \boldsymbol{z}_{ij(r)k}\boldsymbol{b}_{ir} + \varepsilon_{ij(r)k},\\
&\boldsymbol{b}_i \sim MVN(\boldsymbol{0}, \boldsymbol{D}^1_{6\times 6}), \quad \boldsymbol{b}_{ir} \sim MVN(\boldsymbol{0}, \boldsymbol{D}^2_{6\times 6}), \quad \varepsilon_{ij(r)k} \sim N\left(0, \sigma^2\right),
\end{aligned}
$$

for $i = 1, \ldots, N$, $j = 1, \ldots, 12$ and $k = 1, \ldots, 120$. The response $Y_{ij(r)k}$ represents the true amplitude for the P3 peak and N1 dip which is taken to have the functional form used in the weighted linear mixed modeling procedure used for the analysis of our motivating data set. To make the simulation as realistic as possible, the parameters are chosen to mimic the pattern of effects seen in the ASD group under the unexpected condition. Since our goal is to assess accuracy in modeling longitudinal trends, rather than group differences, the simulation uses an intercept term and a natural cubic B-spline basis with equidistant knot placement for the trial number, included as both a fixed and random effect, but no group or condition indicators. The columns of the 120×6 fixed and random design matrices $\boldsymbol{W}$ and $\boldsymbol{Z}_{j(r)}$ correspond to the intercept and five B-spline basis functions. True fixed effects parameters for the P3 and N1 amplitudes are $\boldsymbol{\beta} = (8.580, -0.128, -0.013, -0.136, -0.031, 0.481)$ and $\boldsymbol{\beta} = (-8.654, -0.048, 0.327, 0.079, -0.173, -0.555)$, respectively. These values correspond to a linear transformation ($g(x) = x/10+8$ for P3, $g(x) = x/10-8$ for N1) of the original fitted values in our real-world data to guarantee positive P3 and negative N1 amplitudes on an appropriate scale. To obtain an appropriate signal-to-noise ratio, the fitted variance components were similarly divided by 100 to generate the simulation values. The error variances $\sigma^2 = 0.094$ and 0.072 for the P3 and N1 amplitudes, respectively, are chosen to mimic the fitted variance estimates obtained for the $s = 6$th partition of c_{ijk} under the unexpected condition in their respective weighted linear mixed effects models for the ASD data. In an attempt to keep the total number of variance components low in the model, the random effects covariance matrices, $\boldsymbol{D}^1_{6\times 6}$ and $\boldsymbol{D}^2_{6\times 6}$, have a simple covariance structure which assumes a unique variance for each random effect where random effects are assumed to be independent. The variances of the random effects are chosen to be $(0.389, 0.537, 0.498, 0.760, 2.327, 1.672)$ and $(0.061, 0.139, 0.172, 0.123, 0.387, 0.246)$ for $\boldsymbol{D}^1$ and $\boldsymbol{D}^2$, respectively, for P3 and $(0.466, 0.595, 0.738, 0.497, 2.542, 1.347)$ and $(0.051, 0.132, 0.127,\ 0.100, 0.294, 0.193)$ for $\boldsymbol{D}^1$ and $\boldsymbol{D}^2$, respectively, for N1, similar to our data. The random effects $\boldsymbol{b}_i, \boldsymbol{b}_{ir}$ and the residual error $\varepsilon_{ij(r)k}$ are simulated independently. Note that the simple covariance structure assumed for the random effects does not imply that a simple covariance structure is assumed on the response. The proposed structure still accommodates complex dependency structures over trials in data within subjects and electrode regions via the design matrices containing spline basis functions.

After the P3 and N1 amplitudes are simulated via the above linear mixed models, the true ERP signals, denoted $X_{ijk}^{signal}(t)$, are simulated using the `peak` function in EEGLAB, creating true ERP trajectories with a positive (P3) and negative (N1) phasic sinusoidal peak [10, 49]. Frequencies for the peak functions are sampled from a normal distribution with a mean of 5Hz and standard deviation of 0.1Hz. Latencies of the P3 and N1 peaks are sampled from normal distributions with means of 270ms and 175ms and a standard deviation of 10ms, respectively. The `noise` function in EEGLAB is then used to simulate background EEG, denoted $X_{ijk}^{noise}(t)$, using an inverse Fourier transformation with randomized phase [10, 49]. The noise term consists of the sum of a series of phase-randomized sinusoids, each with amplitudes varying by frequency based on the power spectrum of empirical EEG data.

Next, we obtain the empirical ERP waveforms by summing the true ERP signals and the background noise, $X_{ijk}(t) = X_{ijk}^{signal}(t) + c * X_{ijk}^{noise}(t)$ [49]. The noise component is multiplied by the constant c in this sum, to vary the SNR, which is defined to be the ratio of the standard deviations for the signal over the standard deviation of the noise component, as in Quiroga and van Luijtelaar (2003) [50]. The standard deviation of the signal is calculated in the nonzero region of the ERP waveform containing the P3 and N1 peaks; the standard deviation of the noise component is calculated over the entire range of t. Finally, we induce realistic missingness in the data, randomly removing a fraction of the simulated ERP for each subject by sampling with replacement from the missingness profiles in our implicit learning data set. This ensures that the smoothed data generated by the meta-processing step in the simulations will have an appropriate pattern of heteroskedasticity.

The performance characteristics of both MAP-ERP and the single-trial analysis are evaluated in terms of how accurately they (i) reconstruct the mean longitudinal P3 amplitude trajectories and (ii) predict the subject-specific amplitudes. The amplitude estimates for the single-trial analysis are obtained by applying the peak detection algorithm on loess smooths of all trials for each subject, condition and electrode without cross-sectional averaging and by using the standard (unweighted) linear mixed effects model. Mixed models for both MAP-ERP and the single-trial case are estimated using restricted maximum likelihood. The number of knots was determined by a preliminary simulation study. The most commonly selected number of knots, by AIC, across the MAP-ERP and single-trial preliminary runs were fixed for the real simulation at 6 knots for SNR=$0.4, 0.6$ and 5 knots for SNR=0.8 at $N = 20$ and 40; 6 knots for SNR=0.4 and 5 knots for SNR=0.6 and 0.8 at $N = 80$ and 160. Selections correspond to lower number of knots at higher SNR values as expected [44, 45]. Let $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{b}}_{ir}$ denote the 6×1 (or 7×1, corresponding to 5 or 6 knots, respectively) estimated fixed effects parameters, subject level random effects, and region level random effects from the proposed weighted linear mixed effects model, respectively. Further, let $\boldsymbol{W}$ and $\boldsymbol{Z}_{j(r)}$ represent the fixed and random effects matrices of dimension 120×6 (or 120×7), where the columns represent the intercept and five (or 6) B-spline basis functions defined over 120 trials before the missing data is induced. We denote the (120×1) true and estimated mean fixed effects trajectories by $E(\boldsymbol{Y}) = \boldsymbol{W}\boldsymbol{\beta}$ and $\widehat{E(\boldsymbol{Y})} = \boldsymbol{W}\hat{\boldsymbol{\beta}}$, respectively. The 120×1 true and estimated subject-specific amplitude trajectories are denoted by $E(\boldsymbol{Y}|\boldsymbol{b}_i, \boldsymbol{b}_{ir}) = \boldsymbol{W}\boldsymbol{\beta} + \boldsymbol{Z}_{j(r)}\boldsymbol{b}_i + \boldsymbol{Z}_{j(r)}\boldsymbol{b}_{ir}$ and $E(\widehat{\boldsymbol{Y}|\boldsymbol{b}_i, \boldsymbol{b}}_{ir}) = \boldsymbol{W}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_{j(r)}\hat{\boldsymbol{b}}_i + \boldsymbol{Z}_{j(r)}\hat{\boldsymbol{b}}_{ir}$ for $i = 1, \ldots, N$. We evaluate MAP-ERP via mean error (ME) and prediction error (PE), defined as

$$\text{ME} = \frac{\sum_{k=1}^{120} \left| [E(\boldsymbol{Y})]_k - [\widehat{E(\boldsymbol{Y})}]_k \right|}{\sum_{k=1}^{120} |[E(\boldsymbol{Y})]_k|}$$

and

$$\text{PE} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{12} \sum_{k=1}^{120} \left| [E(\boldsymbol{Y}|\boldsymbol{b}_i, \boldsymbol{b}_{ir})]_k - [E(\widehat{\boldsymbol{Y}|\boldsymbol{b}_i, \boldsymbol{b}}_{ir})]_k \right|}{\sum_{i=1}^{N} \sum_{j=1}^{12} \sum_{k=1}^{120} |[E(\boldsymbol{Y}|\boldsymbol{b}_i, \boldsymbol{b}_{ir})]_k|},$$

where $[\cdot]_k$ is used to denote the kth element of the vectors $E(\boldsymbol{Y})$, $\widehat{E(\boldsymbol{Y})}$, $E(\boldsymbol{Y}|\boldsymbol{b}_i, \boldsymbol{b}_{ir})$ and $E(\widehat{\boldsymbol{Y}|\boldsymbol{b}_i, \boldsymbol{b}}_{ir})$. Proposed measures are normalized by the sum of the absolute deviation of the true mean trajectory and the true subject-specific amplitudes, respectively.

Appendix C: MAP-ERP simulation without latency jitter

We ran additional simulations without latency variation, i.e. without variation in the location of the peaks. The medians, 2.5th and 97.5th percentiles of the MEs and PEs for MAP-ERP are given in Web Table 2.2. In addition, Web Figure 2.6 displays the estimated fixed effects means and pointwise confidence intervals from the run with the median ME value for varying sample size and SNR settings. Similar to the results from simulations with latency jitter, PE of MAP-ERP consistently decreases with increasing SNR across varying sample sizes. Unlike the previous simulations with latency jitter, where noise and latency effects roughly cancel each other at SNR=0.4, the estimated mean trajectory lies above the true one consistently across all sample sizes at SNR=0.4 without latency jitter, since noise tends to inflate the amplitude estimates. This is due to the inflation of the amplitude estimates at low SNR, in the meta-preprocessing step which does not change with sample size. Therefore ME does not decrease with increasing sample size. This is similar to the SNR=0.6 and 0.8 cases with latency jitter, where ME does not decrease with sample size due to the dominating effects of latency jitter dampening the estimated mean response trajectories. However as SNR increases to 0.6 and 0.8 without latency jitter, the estimated mean trajectory gets closer to the true one, and ME decreases with increasing sample size. Hence overall results confirm that ME decreases with increasing SNR and sample size as expected when latency effects are not present, dampening the mean amplitude estimates.

Table 2.2: Median and (2.5th, 97.5th) percentiles of MAP-ERP simulation performance metrics (ME and PE) without latency jitter for varying SNRs from 200 Monte Carlo runs with sample sizes $N = 20, 40, 80$ and 160.

N	SNR	ME	PE
	0.4	.080 (.040, .144)	.199 (.169, .247)
20	0.6	.035 (.017, .085)	.132 (.114, .155)
	0.8	.025 (.011, .063)	.101 (.087, .118)
	0.4	.080 (.047, .116)	.198 (.174, .225)
40	0.6	.031 (.016, .061)	.132 (.119, .148)
	0.8	.019 (.007, .042)	.101 (.091, .112)
	0.4	.079 (.051, .110)	.199 (.182, .221)
80	0.6	.027 (.014, .051)	.133 (.123, .145)
	0.8	.015 (.006, .035)	.102 (.095, .110)
	0.4	.079 (.056, .099)	.199 (.187, .210)
160	0.6	.028 (.013, .044)	.132 (.125, .140)
	0.8	.014 (.006, .028)	.101 (.096, .106)

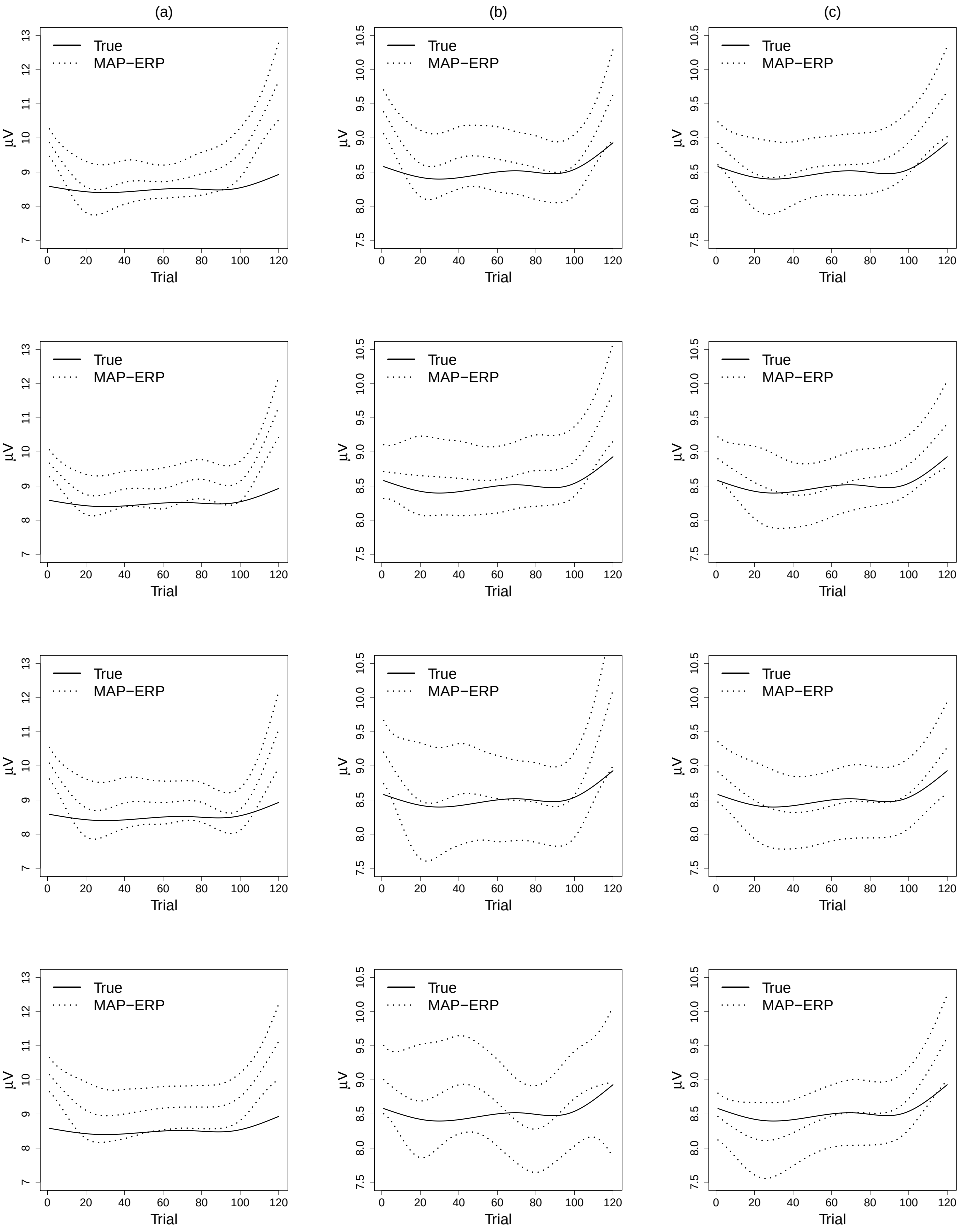


Figure 2.6: Estimated fixed effects mean trajectories along with 95% confidence intervals corresponding to the median of the ME in simulations without latency jitter based on MAP-ERP. Columns correspond to different signal-to-noise ratios (a) SNR=0.4, (b) 0.6 and (c) 0.8. Rows from top to bottom correspond to sample sizes N=20, 40, 80 and 160, respectively.

CHAPTER 3

Robust Functional Clustering of ERP Trends Across an Implicit Learning Paradigm in Autism

This chapter proposes a robust functional clustering (RFC) algorithm to identify subgroups within electroencephalography (EEG) data. This is a stand-alone paper taken from a first author publication in *Biostatistics* that is currently in revision [14]. The proposed RFC is an iterative algorithm based on functional principal component analysis (FPCA), where cluster membership is updated via predictions of the functional trajectories obtained through a nonparametric random effects model. We consider functional data resulting from event-related potential (ERP) waveforms representing EEG time-locked to stimuli over the course of an implicit learning experiment, after applying the previously proposed meta-preprocessing step. The resulting functional ERP components (peak amplitudes and latencies) following meta-preprocessing inherently exhibit covariance heterogeneity due to low data quality over some stimuli inducing the averaging of different numbers of waveforms in sliding windows of the meta-preprocessing step. The proposed RFC algorithm incorporates this known covariance heterogeneity into the clustering algorithm, improving cluster quality, as illustrated in the data application and extensive simulation studies. ASD is a heterogeneous syndrome and identifying subgroups within ASD children is of interest for understanding the diverse nature of this complex disorder. Applications to the implicit learning paradigm identify subgroups within ASD and TD children with diverse learning patterns over the course of the experiment, which may inform clinical stratification of ASD.

3.1 Introduction

Electroencephalography (EEG) is a noninvasive method for measuring spontaneous electrical activity across brain regions over time. As a method to identify neural function and cognitive states, it has been studied in diverse biomedical settings including epilepsy, sleep disorders, multiple sclerosis, brain tumors and lesions, major affective disorder, schizophrenia, alcoholism and bipolar disorder ([1, 2]). Here we consider an application to a study of visual implicit learning in young children with Autism Spectrum Disorders (ASD) [11]. ASD has a highly heterogeneous presentation, making it difficult to tease apart underlying mechanistic pathways to core deficits. The goal of this paper is to provide insights into those pathways through a better understanding of implicit learning, defined as the detection of regular patterns in one's environment without a conscious awareness to learn. Age-matched 2 to 5 year old typically developing (TD) and ASD children were presented with a continuous sequence of six colored shapes organized into three shape pairs (Figure 3.1(a)). Shapes within pairs appeared in the same order but the pairs themselves occurred in random order. Transitions within a shape pair were labeled 'expected' since they could be learned and transitions between shape pairs were 'unexpected' since they could not be predicted.

EEG signals, time locked to visual stimuli (e.g. presentation of colored shapes), result in event-related potential (ERP) waveforms containing the P3 and N1 phasic components shown in Figure 3.1(b). While the focus is on the P3 and N1 components in this particular paradigm, other phasic components may be studied in different applications. The P3 peak of the ERP waveform is thought to be related to cognitive processes such as signal matching, decision making and memory updating while the N1 dip represents early category recognition [11, 16]. Implicit learning is assessed through differences in the amplitude (size of the peak) and latency (time when the peak occurs) of the ERP components between the expected and unexpected conditions.

It is natural to seek inference about potential differences in ERP variation between TD and ASD groups in the implicit learning paradigm [11]. However, ASD is a heterogeneous syndrome characterized by impairments in social communication and the

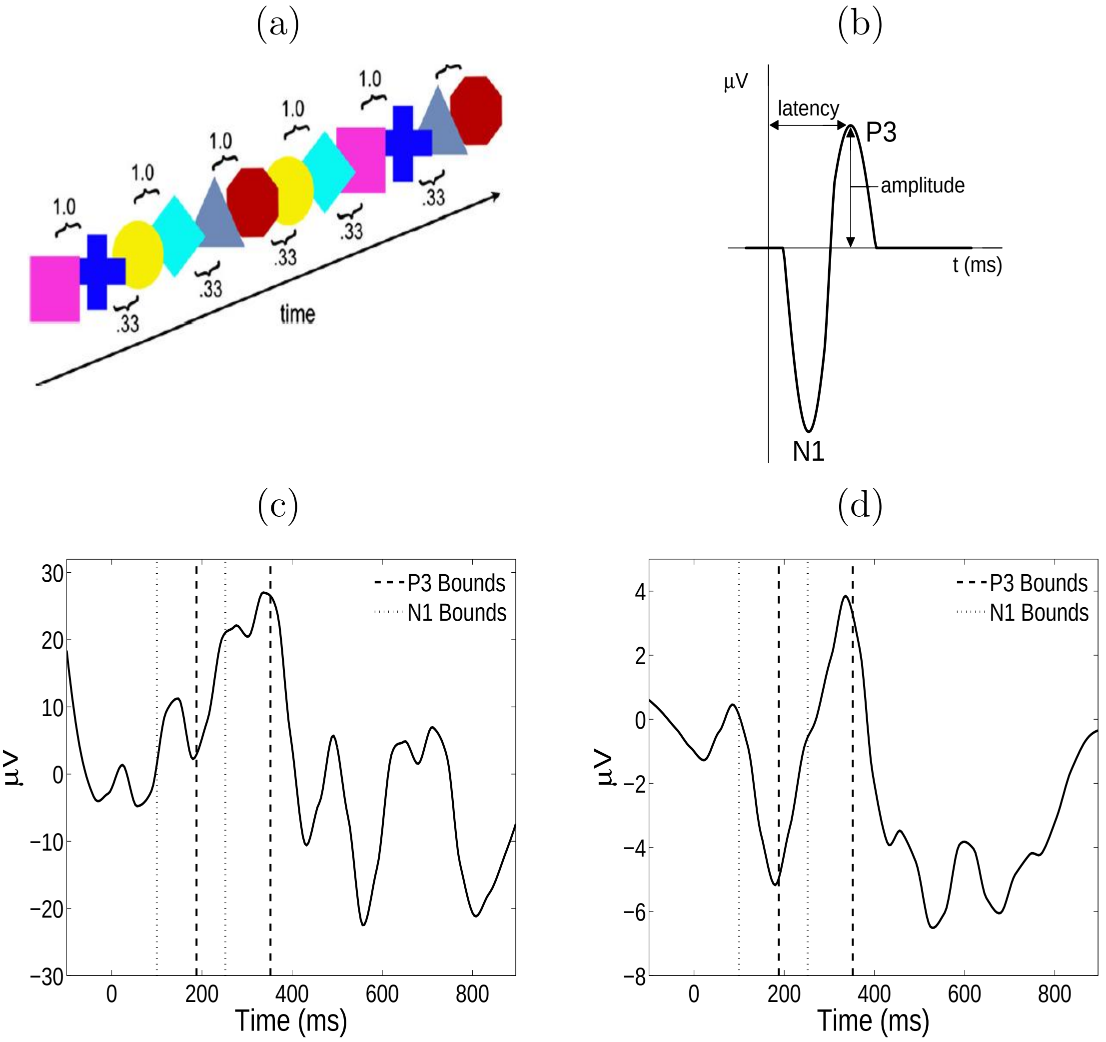


Figure 3.1: (a) The sequence of shape pairs in the implicit learning paradigm. (b) The ERP waveform containing the P3 and N1 phasic components from the implicit learning study. (c) ERP waveform from a single subject, condition, electrode and trial in the right frontal region of the scalp after preprocessing. (d) The average of the first 30 consecutive ERP waveforms for the same subject, electrode and condition.

presence of restricted interests and repetitive behaviors [51]. Hence, in addition to contrasting learning patterns of TD and ASD groups, identifying subgroups within ASD children with distinct learning patterns is also of interest for understanding the diverse nature of this complex disorder. We therefore propose a robust functional clustering (RFC) algorithm to more finely grain learning patterns within TD and ASD children. A specific feature of our proposal is to make maximal use of the existing structural information in ERPs induced by data quality to improve clustering accuracy even in small samples.

Typical analysis of ERP data focus on summaries of key components, such as peak amplitude and latency. Specifically, to increase the low signal-to-noise ratio in raw ERP data, the waveforms resulting from repeated stimuli (referred to as trials) are averaged for each subject so that the ERP components are identifiable. Hasenstab et al. proposed a meta-preprocessing step for the analysis of ERP data, based on a moving average, which increases the signal-to-noise ratio of the observed ERPs while preserving changes in ERPs across trials [13]. Meta-preprocessing retains valuable longitudinal information which is lost by the common practice of averaging ERP trajectories across all trials. Capturing these trends is especially important in settings such as our motivating example, where patterns of learning correspond by definition to changes of ERP features across trials. We will build our proposed clustering algorithm on the functional data produced by this novel meta-preprocessing step, consisting of ERP components (e.g. P3 peak amplitude) obtained over trials (representing time) of the experiment for each subject and condition (expected/unexpected).

An important issue with the meta-preprocessed functional ERP components is covariance heterogeneity. Trials resulting in low data quality are common in EEG studies, especially in experiments involving young children. A standard approach is to remove these trials in the data cleaning steps. The meta-preprocessing recovers information in these low data quality trials by smoothing over them, leading to different numbers of waveforms being averaged in the sliding windows, and hence to covariance heterogeneity in the functional data. We propose a robust functional clustering algorithm that accounts for this known source of covariance heterogeneity in the trajectories, setting it apart from previously proposed functional clustering algorithms.

Clustering or classification of functional data typically involves either regularization or filtering. Regularization involves discretization of the time interval followed by the application of standard multivariate clustering or classification methods. Because the resulting data are high-dimensional and highly correlated, a regularization constraint is typically applied to the covariance structure in model-based methods [52, 53, 54, 55, 56, 57, 58, 59]. Other model-based approaches include proposals by Luan and Li (2003) [60] and James and Sugar (2003) [61] where special attention is given to sparse and irregularly spaced functional data. Filtering methods work by projecting each curve onto a finite-dimensional set of basis functions, such as B-splines or functional principal components (FPCA), and then applying standard clustering or classification algorithms to the resulting basis coefficients [61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73]. Serban and Jiang extended filtering-based functional approaches to hierarchical data structures using multilevel FPCA in the context of hard and soft clustering [74]. Multilevel functional data consisting of expression profiles from genes of immune cells were considered.

Chiou and Li introduced another filtering method, k-centers functional clustering (FC), based on FPCA to identify homogeneous clusters within the sample sharing a cluster-specific mean function and a cluster-specific covariance surface [75, 76]. We build on this approach to incorporate the known covariance heterogeneity in the meta-preprocessed functional data induced by the data quality issues into the proposed clustering algorithm. Our proposal assumes a common covariance structure within the TD and ASD groups if data with acceptable quality were collected at all the trials. First we identify induced covariance subsets within each diagnostic group with similar low quality data patterns over time (trials of the ERP experiment). Fixing the covariance subsets, trajectories are clustered according to the estimated mean trends where covariance subset and cluster memberships are allowed to differ. Mean trajectory and covariance surface estimates are updated iteratively in a nonparametric fashion, where cluster memberships are updated in a reclassification step based on a nonparametric random effects model. The proposed RFC increases cluster accuracy by incorporating the known covariance heterogeneity into the functional clustering algorithm as illustrated in our data analysis and simulation studies. We further extend the proposed RFC for multilevel functional data to be applicable to the meta-preprocessed ERP components obtained from multiple electrodes on the scalp.

The remainder of the paper is organized as follows. Section 3.2 describes the cleaning and meta-preprocessing of ERP data in detail. Section 3.3 provides background on FPCA and introduces the proposed RFC algorithm for single and multilevel functional data. Section 3.4 applies the proposed RFC to the autism study and compares the results to those obtained from alternative algorithms including FC of Chiou and Li (2007) [75]. We study the performance of the proposed algorithm in extensive simulations summarized in Section 3.5 and conclude with a brief discussion (Section 3.6).

3.2 Description of data cleaning and meta-preprocessing steps and the resulting multilevel functional data

In the motivating study of implicit learning, EEG data were recorded for 120 trials per condition (expected/unexpected) for each of the 34 TD and 37 ASD children at 128 electrodes. The EEG signals were sampled at 250Hz, producing 250 within-trial time points per waveform, spanning 1000ms. Standard preprocessing steps included artifact detection (of physiological noise that could obscure the signal such as blinks, saccades or muscle contractions), bad channel replacement (a smoothing procedure performed when signals from an individual scalp electrode are compromised), referencing (to standardize measurements to an average of one across all the electrodes) and baseline corrections (to standardize a subject's measurements to their own baseline period). Despite the standard preprocessing steps, the ERP data had a small signal-to-noise ratio (SNR), making it difficult to identify components, such as peak amplitudes and latency, on trial specific ERPs. Figure 3.1(c) displays a single ERP waveform for one subject from a single trial recorded in the right frontal region of the scalp. The P3 peak and N1 dip are unrecognizable due to the low SNR. The meta-preprocessing step of Hasenstab et al. utilizes a moving average of ERPs across sliding trial windows to increase the SNR in order to identify ERP components [13]. Figure 3.1(d) displays an average of 30 ERP waveforms from adjacent trials, where the P3 peak and N1 dip are easily recognized due to the increased SNR. Components of interest such as peak amplitudes are extracted from these averaged ERP waveforms. For a detailed outline of the meta-preprocessing step, readers are referred to Hasenstab et al. (2015) [13].

In addition to identifying the magnitudes of the key ERP components over trials, the meta-preprocessing provides information on the variance of the extracted components. EEG experiments involving young children tend to have larger amounts of trials with low quality data due to head movements or lack of cooperation. Hence, the number of ERPs averaged in the sliding windows of the meta-preprocessing step may not be the same. This difference introduces a known form of covariance heterogeneity in the ERP components. Components extracted from ERPs averaged over a smaller number of trials will have larger variance. For illustration of the methods, we consider P3 peak amplitude trajectories between the 5th and 60th trials of the experiment, where implicit learning is thought to be maximal, and analyze data from the four electrodes in the right frontal region of the scalp. The first four trials are excluded for stability of the results. The focus will be on the P3 amplitude difference trajectories between expected and unexpected conditions, resulting in multilevel functional data (electrodes nested in subjects) on P3 amplitude differences and the number of averaged ERPs. We first cluster the multilevel functional data on the number of averaged ERPs to determine the induced covariance subsets, which are assumed to be known in the proposed RFC algorithm outlined below. More details on this initial multilevel functional clustering to determine the covariance subsets are deferred to Section 3.4.1.

3.3 Robust functional clustering (RFC)

Our work builds on the FC algorithm proposed by Chiou and Li (2007) [75]. The original formulation of this clustering procedure assumes identical mean and covariance cluster membership. In the analysis of ERP data, the assumption of within-cluster covariance homogeneity may not be warranted. More precisely, we aim to make use of the covariance subset information, induced in the meta-preprocessing step due to data quality issues. We consider n_v covariance subsets, but we do not require subset membership to necessarily overlap with cluster membership in the proposed RFC which clusters trajectories according to mean trends. Even though the covariance subsets are assumed known, the covariance surfaces cannot be estimated with unknown cluster membership of the functional trajectories. Hence the proposed algorithm involves a nonparametric iterative mean and covariance update to estimate cluster structures. These cluster structures are used for updating cluster membership of each functional trajectory via predictions based on a nonparametric random effects model of the truncated Karhunen-Loève expansions. We introduce basic principles in Section 3.3.1 and outline the proposed algorithm in Section 3.3.2. Extensions to multilevel functional data are given in Section 3.3.3.

3.3.1 Functional model

A common modeling strategy assumes the observed functional trajectory for subject i, $y_i(t)$, to be a realization of a stochastic process $Y_i(t)$ where $t \in \mathcal{T} = [0, T]$. The random function $Y_i(t)$ has mean $\mu(t) = E\{Y_i(t)\}$ and covariance $\mathrm{cov}\{Y_i(s), Y_i(t)\} = \Sigma(s,t) + \sigma^2 I(s = t)$, leading to the Karhunen-Loève (K-L) expansion, $Y_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t) + \epsilon_i(t)$, where $\phi_k(t)$ are the eigenfunctions of Σ with corresponding eigenvalues λ_k and $\epsilon_i(t)$ is measurement error with mean zero and variance σ^2. Eigenfunctions are orthonormal and are assumed in non-increasing order ($\lambda_1 \geq \lambda_2 \ldots$) such that their sum is finite $\sum_{k=1}^{\infty} \lambda_k < \infty$. The scores $\xi_{ik} = \int \{Y_i(t) - \mu(t)\}\phi_k(t)\mathrm{d}t$ are the projections of the $Y_i(t) - \mu(t)$ in the direction of the k^{th} eigenfunction $\phi_k(t)$ and are uncorrelated with $E(\xi_{ik}) = 0$ and $Var(\xi_{ik}) = \lambda_k$.

To allow for granularity in the foregoing model, we allow subclusters within each of the overarching clinical groups. Specifically, we assume that $Y_i(t)$ is sampled from a mixture of stochastic processes, with cluster membership indexed by $c \in \{1, 2, \ldots, C\}$. To account for covariance heterogeneity, we allow for covariance subsetting, indexed by $v \in \{1, 2, \ldots, n_v\}$, which may be different from cluster membership. Conditioning on cluster membership c and covariance subset v, means and covariances of the subprocesses are given as $E\{Y_i\} = \mu^{(c)}(t)$, $\mathrm{cov}\{Y_i(s), Y_i(t)\} = \Sigma^{(v)}(s,t) + \sigma^{2(v)} I(s = t)$. The measurement error $\epsilon_i^{(v)}(t)$ has mean zero and variance $\sigma^{2(v)}$ for covariance subset v. We note that for full generality, measurement error can be allowed to change across covariance subsets but in practice one will often contain it to be the same across v. The mean functions of the subprocesses are only associated with the cluster membership c while the covariance terms are associated with both clusters and covariance subsets since the mean function of cluster c is used in the estimation of both $\Sigma^{(v)}(s,t)$ and $\sigma^{2(v)}$. The covariance terms are not indexed by c for ease of notation. It is assumed that each subprocess has a K-L expansion with corresponding mean function $\mu^{(c)}(t)$ and eigenvalues $\lambda_k^{(v)}$ and corresponding eigenfunctions $\phi_k^{(v)}(t)$ such that $\Sigma^{(v)}(s,t) = \sum_k \lambda_k^{(v)} \phi_k^{(v)}(s) \phi_k^{(v)}(t)$, $s, t \in \mathcal{T}$ and $\xi_{ik}^{(c,v)} = \int \{Y_i(t) - \mu^{(c)}(t)\} \phi_k^{(v)}(t) \mathrm{d}t$.

The updating of the cluster membership in the proposed RFC will utilize functional predictions based on the nonparametric random effects model,

$$Y_i^{(c,v)}(t) = \mu^{(c)}(t) + \sum_{k=1}^{K_v} \xi_{ik}^{(c,v)} \phi_k^{(v)}(t) + \epsilon_i^{(v)}(t). \tag{3.1}$$

Methods for selecting the number of components K_v in (3.1) include cross-validation [77], Akaike's Information Criterion [78], and percentage of variance explained [76, 79, 80]. We found that choosing components to explain 90% of the variation works well in our applications. For a trajectory $Y_i(t)$ from covariance subset v, $Y_i^{(c,v)}(t)$ in (3.1) will be the truncated K-L expansion and hence will be a good approximation of $Y_i(t)$ if $Y_i(t)$ actually belongs to cluster c, but may match poorly if the current cluster assignment is incorrect. Hence, the cluster membership updating will compare an observed curve for subject i, $y_i(t_{ip})$ $(p = 1, \ldots, T_i)$, from covariance subset v to its estimated predictions $\widehat{y}_i^{(c,v)}(t_{ip})$ via (3.1) from each of the $c = 1, \ldots, C$ clusters and assign cluster membership according to the criterion

$$c^*(y_i) = \underset{c \in \{1,\ldots,C\}}{\arg\min} \left[\sum_{p=1}^{T_i} \left\{ y_i(t_{ip}) - \widehat{y}_i^{(c,v)}(t_{ip}) \right\}^2 \right]^{1/2}. \tag{3.2}$$

3.3.2 Single-level RFC

The mean functional trajectory is obtained by local linear smoothing; smooth covariance surfaces are estimated via a two dimensional local least squares algorithm. To eliminate the effects of measurement error, the diagonal elements of the raw covariance matrix are removed before the two-dimensional smoothing step. A nonparametric functional principal component analysis step will be employed on the smooth estimate of the covariance surface by a standard discretization procedure to estimate the eigenfunctions and eigenvalues. In order to guarantee that the covariance matrix is non-negative definite, any eigenvalues with negative estimates and their corresponding eigenfunctions are removed from the functional principal component decompositions of the covariances. For explicit expressions of the estimated mean, eigenfunctions and eigenvalues, denoted by $\widehat{\mu}(t)$, $\widehat{\phi}_k(t)$ and $\widehat{\lambda}_k$, respectively, as well as the covariance surfaces, readers are referred to Şentürk and Nguyen (2011) [81] and Şentürk et al. (2013) [82]. Bandwidths for the smooths may be selected using cross-validatio or generalized cross-validation (GCV) [77]. For a computationally efficient bandwidth choice in the proposed smoothing procedures, we adopt the generalized cross-validation algorithm of Liu and Muller (2008) [83].

For dense functional data, subject-specific scores can be estimated using projection, $\hat{\xi}_{ik} = \int \{y_i(t) - \widehat{\mu}(t)\} \widehat{\phi}_k(t)\mathrm{d}t$; for sparse data applications, the best linear unbiased prediction (BLUP) of the scores has been proposed by Yao et al. (2005) [78]. Score estimates obtained through BLUP for both single-level and multilevel functional data are deferred to the online Supplementary Material. The estimates of the mean, eigenfunctions, eigenvalues and subject-specific scores for trajectories in covariance subset v and cluster c, denoted by $\widehat{\mu}^{(c)}(t)$, $\widehat{\phi}_k^{(v)}(t)$, $\widehat{\lambda}_k^{(v)}$ and $\hat{\xi}_{ik}^{(c,v)}$, respectively, are obtained similarly based on trajectories in the specific subset and cluster in the iterative steps of the RFC algorithm as outlined below.

An initial clustering is performed using the k-means algorithm applied to the estimated subject-specific scores obtained via a FPCA model fitted to the entire sample. Given a fixed number of clusters C, the k-means clustering is performed on the leading K scores where K is selected via percentage of variance explained. We denote the cluster assignment for subject i in iteration r of the algorithm by $c_i^{(r)}$ and note that covariance subset assignments do not change throughout the algorithm; only the clusters are updated in each iteration. Given the initial clustering results, we estimate predictions for the ith subject's trajectory from all clusters. Specifically, in order to arrive at the predictions, we first obtain estimates of the mean function $\widehat{\mu}_{(-i)}^{(c)}(t)$ based on all trajectories assigned to cluster c while leaving out the i^{th} subject's trajectory. Note that only the mean functions of the cluster containing subject i need to be re-estimated. Next we obtain the mean centered trajectories of all subjects in covariance subset v, which contains the ith subject trajectory. The eigenfunctions of covariance subset v are estimated based on mean centered trajectories of the subjects in the subset, leaving out the mean centered trajectory of the ith subject. Note that eigenfunction estimates, denoted by $\widehat{\phi}_{k(-i)}^{(v)}(t)$, need to be estimated for only one covariance subset, the subset that contains the ith subject's trajectory. The last component for the predictions are the scores, which can be estimated based on the leave-one-out mean and eigenfunction estimates for $c = 1, \ldots, C$, based on the estimated projection $\hat{\xi}_{ik}^{(c,v)} = \int \{y_i(t) - \widehat{\mu}_{(-i)}^{(c)}(t)\} \widehat{\phi}_{k(-i)}^{(v)}(t) \mathrm{d}t$ for dense functional data. Using all estimated components, we arrive at the predicted trajectory for subject i from cluster $c = 1, \ldots, C$ as

$$\widehat{y}_i^{(c,v)}(t_{ip}) = \widehat{\mu}_{(-i)}^{(c)}(t_{ip}) + \sum_{k=1}^{K_v} \widehat{\xi}_{ik}^{(c,v)} \widehat{\phi}_{k(-i)}^{(v)}(t_{ip}).$$

The ith subject's trajectory is then reclassified into cluster $c_i^{(r+1)}$ such that $c_i^{(r+1)} = \arg\min_{\{c=1,\ldots,C\}} [\sum_{p=1}^{T_i} \{y_i(t_{ip}) - \widehat{y}_i^{(c,v)}(t_{ip})\}^2]^{1/2}$. Note that the sum above is taken over all observation time points for subject i and that different weighting schemes can be implemented if observation times in certain intervals are thought to be more informative than others in determining cluster membership.

The above steps are performed for all subject trajectories $i = 1, \ldots, n$, resulting in an updated set of cluster assignments $\{c_i^{(r+1)}, i = 1, \ldots, n\}$ for the $(r + 1)$st iteration of the algorithm. The iterative process continues until none of the cluster memberships change. The cluster and covariance subset components are re-estimated at each iteration of the RFC algorithm; for sufficiently large samples sizes, one may ignore the leave-one-curve out procedure when calculating predictions $\widehat{y}_i^{(c,v)}(t_{ip})$ in order to significantly reduce computational time, assuming negligible bias. As with other clustering algorithms, RFC requires the number of clusters (and covariance subsets) to be known *a priori*. In our applications we set both the number of clusters and covariance subsets to two due to limitations in sample size (there are $n = 32$ and 34 children in the TD and ASD groups after removal of outliers). Readers are referred to Li and Chiou (2011) [84] for an extensive discussion on methods for selecting the number of clusters in the context of functional data.

Chiou and Li (2007) [75] discuss identifiability conditions for their FC algorithm and show that the cluster eigenspaces cannot be subsets of each other, there cannot be two identical cluster mean functions and that if a cluster mean function belongs to its own cluster's eigenspace, then another mean function cannot belong to that same eigenspace. Note that unlike our proposed RFC, the FC algorithm of Chiou and Li assumes all curves within a cluster have the same covariance and uses both mean and covariance differences to identify clusters. In contrast, the proposed RFC clusters functional trajectories only based on differences in mean trends, since cluster and covariance subset memberships do not necessarily overlap. Hence identifiability conditions for the proposed RFC include that 1) the cluster mean functions cannot be the same and that 2) the cluster mean functions cannot lie in any covariance subset eigenspace. Note that the identifiability of the covariance subsets (via the assumption that eigenspaces cannot be subsets of each other) is no longer needed for RFC, since the covariance subsets are assumed to be known *a priori*. However, while cluster mean functions lying in their own eigenspaces is not a problem for FC, where FC cluster and covariance subset memberships overlap, it poses an identifiability issue for RFC, where memberships do not necessarily overlap. Since the first identifiability condition is standard, we examine only the second condition through simulation studies (Section 3.5).

3.3.3 Multilevel RFC

In this section we extend the RFC algorithm to multilevel functional data. Multilevel functional data refers to functional data collected in a hierarchy of units such as subject-specific ERP feature trajectories observed at multiple electrodes (subunits) on the scalp. Let $Y_{ij}(t_{ijp})$ denote a functional response observed for subject i, on subunit j at time point t_{ijp}, $p = 1, \ldots, T_{ij}$. Total functional variation in $Y_{ij}(t)$, $t \in \mathcal{T}$, can be decomposed via functional analysis of variance (FANOVA) such that

$$Y_{ij}(t) = \mu(t) + \eta_j(t) + Z_i(t) + W_{ij}(t) + \epsilon_{ij}(t),$$

where $\mu(t)$ and $\eta_j(t)$ are fixed functional effects that represent the overall mean function and subunit (e.g. electrode-specific) shifts, respectively, $Z_i(t)$ and $W_{ij}(t)$ are the subject- and subunit-specific deviations, respectively, and $\epsilon_{ij}(t)$ is measurement error with mean zero and variance σ^2 [79, 80]. The deviations $Z_i(t)$ and $W_{ij}(t)$ are assumed to be uncorrelated mean zero stochastic processes. As with the K-L decompositions for the single-level functional data, decomposition across both levels of variation results in $Y_{ij}(t) = \mu(t) + \eta_j(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(1)}(t) + \sum_{\ell=1}^{\infty} \zeta_{ij\ell}\phi_\ell^{(2)}(t) + \epsilon_{ij}(t)$ where $\phi_k^{(1)}(t)$ and $\phi_\ell^{(2)}(t)$ are level 1 and level 2 eigenfunctions, and ξ_{ik} and ζ_{ijl} are subject-specific scores with mean zero and variance $\lambda_k^{(1)}$ and $\lambda_\ell^{(2)}$, respectively [79, 80]. Note that $\phi_k^{(1)}(t)$ and $\phi_\ell^{(2)}(t)$ may not be mutually orthogonal. The above formulation models the dependency between subunit-specific trajectories within a subject, while still allowing covariance surfaces within subunits to be different from covariance surfaces across subunits. In this application we consider multilevel functional data from four electrodes in the right frontal region of the scalp. However we note that more complex FANOVA models can be developed with additional layers in the multilevel structure (e.g. electrodes nested within multiple brain regions).

Similar to the single-level case, we assume that $Y_{ij}(t)$ is sampled from a mixture of subprocesses with cluster means and induced covariance subsets. Motivated by similar low data quality patterns, and hence similar functional trajectories of number of averaged ERPs from sliding windows across electrodes, covariances at both levels of the multilevel deviations are grouped in the same covariance subset. In addition, since mean functions within subjects across electrodes are quite similar in our application, cluster membership of different electrode trajectories within a subject are assumed to be the same, where clusters are determined more based on the shapes of the overall means $\mu(t)$. Hence, conditional on cluster c and covariance subset v, the means and covariances of the subprocesses are given by

$$E\{Y_{ij}(t)\} = \mu^{(c)}(t) + \eta_j^{(c)}(t), \quad \operatorname{cov}\{Y_{ij}(s), Y_{ij'}(t)\} = \Sigma_B^{(v)}(s,t), \quad j \neq j', \tag{3.3}$$

$$\operatorname{cov}\{Y_{ij}(s), Y_{ij}(t)\} = \Sigma_T^{(v)}(s,t) = \Sigma_B^{(v)}(s,t) + \Sigma_W^{(v)}(s,t) + \sigma^{2(v)} I(s=t), \tag{3.4}$$

where $\Sigma_T^{(v)}(s,t)$ is the overall covariance function, $\Sigma_B^{(v)}(s,t)$ and $\Sigma_W^{(v)}(s,t)$ are the between and within subunit covariance functions, respectively, and $\sigma^{2(v)} I(s=t)$ represents the error variance on the diagonal of $\Sigma_T^{(v)}(s,t)$. The between and within structures may be decomposed further using their eigenvalues and eigenfunctions such that

$$\Sigma_B^{(v)}(s,t) = \sum_k \lambda_k^{(1,v)} \phi_k^{(1,v)}(s) \phi_k^{(1,v)}(t)$$
$$\Sigma_W^{(v)}(s,t) = \sum_\ell \lambda_\ell^{(2,v)} \phi_\ell^{(2,v)}(s) \phi_\ell^{(2,v)}(t).$$

The multilevel RFC adopts the same structure as the single-level algorithm. Since we are interested in clustering individuals largely by subject-level (level 1) differences and level 2 is mainly used to incorporate dependencies within subjects across electrodes, initial clustering is performed by applying k-means to the estimated level 1 scores (ξ_{ik}) based on the entire sample. Note that if the initial clusters are far from the true clustering, this could adversely affect cluster quality and the RFC may converge to a local optimum. Hence, robustness to initial clustering results should be studied in applications. Cluster membership updates still utilize functional predictions based on the estimated nonparametric truncated multilevel random effects model

$$\widehat{y}_{ij}^{(c,v)}(t_{ijp}) = \widehat{\mu}_{(-i)}^{(c)}(t_{ijp}) + \widehat{\eta}_{j(-i)}^{(c)}(t_{ijp}) + \sum_{k=1}^{K_v} \widehat{\xi}_{ik}^{(c,v)} \widehat{\phi}_{k(-i)}^{(1,v)}(t_{ijp}) + \sum_{\ell=1}^{L_v} \widehat{\zeta}_{ijl}^{(c,v)} \widehat{\phi}_{\ell(-i)}^{(2,v)}(t_{ijp}), \tag{3.5}$$

where the $(-i)$ notation denotes that the observed multilevel functional data for subject i has been left out while obtaining the estimates. In (3.5), K_v and L_v refer to the number of eigen components selected for the between and within levels, respectively. First and second level eigenfunctions in (3.5) are estimated based on subunit trajectories within subjects in covariance subset v, centered by subtraction of their overall mean $\widehat{\mu}^{(c)}_{(-i)}(t)$ and the subunit-specific mean deviation $\widehat{\eta}^{(c)}_{j(-i)}(t)$ for subjects in cluster c. First and second level subject-specific scores are estimated as projections similar to the single-level case for dense functional data [79] and via BLUP estimates for sparse cases [80], where explicit forms are provided in the Appendix. Given a set of observed curves $y_{ij}(t_{ijp})$ and predicted curves $\widehat{y}^{(c,v)}_{ij}(t_{ijp})$ for subject i, cluster membership is assigned such that

$$c_i^{(r+1)} = \underset{c\in\{1,\ldots,C\}}{\arg\min} \left(\sum_{j}^{J} \left[\sum_{p=1}^{T_{ij}} \left\{ y_{ij}(t_{ijp}) - \widehat{y}^{(c,v)}_{ij}(t_{ijp}) \right\}^2 \right]^{1/2} \right).$$

In contrast to the single-level case, the above criterion also sums over J subunits within a subject.

3.4 Application to the implicit learning study

We utilize the proposed multilevel RFC algorithm to cluster P3 amplitude difference trajectories within ASD and TD groups. Following the data cleaning and meta-preprocessing steps, differences in amplitudes are computed for each trial between expected and unexpected conditions; trials which do not have valid data for both conditions are considered missing. To determine the covariance subsets, the number of ERPs (from sliding windows in the moving average) are further averaged across conditions, where the number of ERPs are observed to be quite similar for the two conditions. Five subjects are removed as outliers prior to analysis. Two of the removed subjects (one in each diagnostic group) did not have observed data until trial 20 of the experiment, and the remaining three subjects had amplitude differences more than 2 standard deviations away from their respective group means for most of the trials.

3.4.1 Determination of the covariance subsets

We begin by clustering the multilevel functional trajectories of the number of averaged ERPs to determine the covariance subsetting. Covariance subsets are determined by a k-means clustering of the level 1 scores in the multilevel FPCA decompositions. Due to small sample sizes in both the TD and ASD groups (32 and 34 children in TD and ASD groups, respectively), we explore two clusters and two covariance subsets via RFC. The number of averaged ERPs from all 4 electrodes are plotted in Figures 3.2(a)-(b) for the two covariance subsets identified within the TD and ASD groups. Lower numbers of averaged ERPs correspond to higher variance. The numbers of averaged ERPs increase to their maximum value of 30 around trial 20 in the first covariance subset. The separation between covariance subsets is larger in the ASD group with respect to shapes and magnitudes of the trajectories due to lower numbers of averaged ERPs, suggesting stronger covariance heterogeneity. The second covariance subset within ASD has consistently low numbers of averaged ERPs across the first 60 trials. In contrast, the trajectories in the second covariance subset within TD are more similar in shape to those in the first covariance subset but with smaller magnitudes. These observations are consistent with the estimated covariance subset eigenfunctions (Figures 3.2(c)-(d)) obtained after the estimation of the cluster means via RFC. The estimated leading eigenfunction for the second covariance subset within TD shows that much of the variability in the trajectories is observed at later trials, where the number of averaged ERPs decrease. Nevertheless, the estimated leading eigenfunctions differ more in the ASD group than the TD group. The major differences are in the earlier trials, where the second covariance subset within the ASD subjects has lower numbers of averaged ERPs.

3.4.2 New scientific insights gained from the RFC results

Estimated cluster means and 90% bootstrap bands obtained from the RFC algorithm are shown in Figures 3.3(a)-(b) for the TD and ASD groups, respectively. Bandwidths for the mean and covariance smooths are selected using generalized cross-validation and visual assessment to maximize cluster quality, where selected bandwidths are 5 and 10

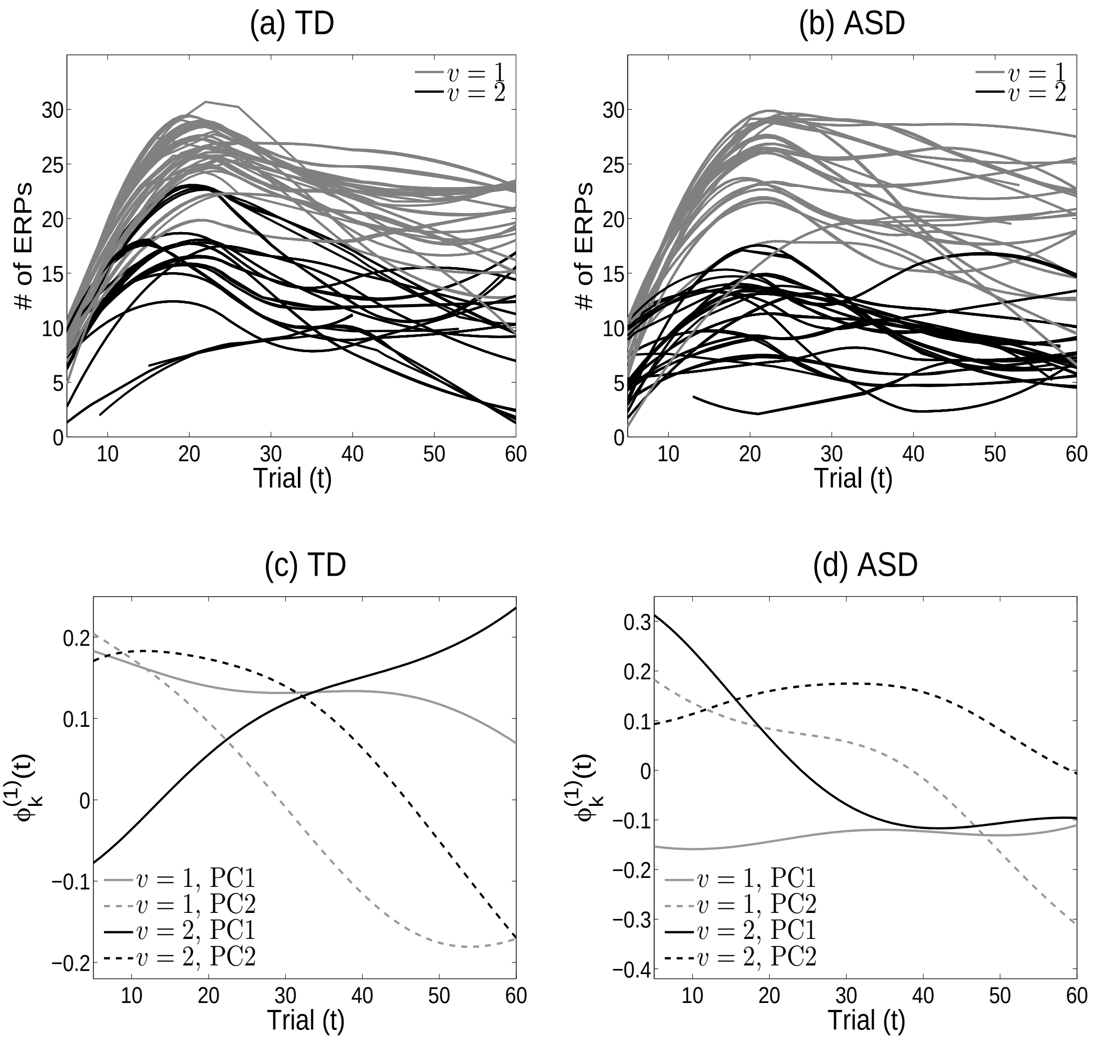


Figure 3.2: The number of averaged ERP trajectories from the two covariance subsets for the TD (a) and ASD (b) children. Estimated eigenfunctions for the two covariance subsets are given in plots (c) and (d) where the gray and black trajectories correspond to the covariance subset index and the solid and dashed lines represent the first and second principal components, respectively.

for mean and covariance smoothing, respectively. The percentile confidence bands are based on 200 bootstrap samples chosen with replacement from TD and ASD subject-specific ERP data. The data cleaning and meta-preprocessing steps are applied to the resampled ERP data followed by covariance subsetting and RFC clustering. Hence, in addition to assessing the variability in the proposed RFC algorithm, the bootstrap procedure also includes variability associated with the meta-preprocessing of the data and sampling variation within the TD and ASD groups. While resulting confidence intervals are wide, given the small sample sizes of our application, we note that the shapes of the cluster mean trajectories are fairly preserved in the bootstrap bands. Bootstrap clusters are mapped to the cluster means of the original sample such that the distance between them is minimized. The percentage of times a subject is assigned to their original mean cluster over the 200 bootstrap runs is averaged across all subjects to be used as a measure of RFC cluster consistency. Despite the small sample size of the groups, RFC clustering is found to be fairly consistent, its subjects in the bootstrap sample being assigned their original clusters 76% and 77% of the time for the TD and ASD groups, respectively. In addition to the plots of the estimated cluster means, Figures 3.3(c)-(d) display the electrode-specific cluster means which are quite similar within clusters, implying small within subject between-electrode variation. Hence, we further display smoothed amplitude difference trajectories across electrodes within the identified clusters in the top rows of Figures 3.4 and 3.5 for the TD and ASD groups, respectively.

While interpreting the results, we review previous findings in the analysis of the data from the implicit learning study and highlight new insights offered by the proposed RFC. Hasenstab et al. were able to model longitudinal changes in the ERP components via meta-preprocessing and used a weighted linear mixed effects model (LMM) to analyze the meta-preprocessed data to study mean trends of P3 amplitude differences between conditions in the TD and ASD groups [13]. Figure 3.3(e) shows the mean condition differentiation (expected - unexpected condition) across trials for the TD and ASD groups displayed with 90% bootstrap confidence bands. Both groups seem to differentiate between the two conditions at similar speeds (peak differentiation around

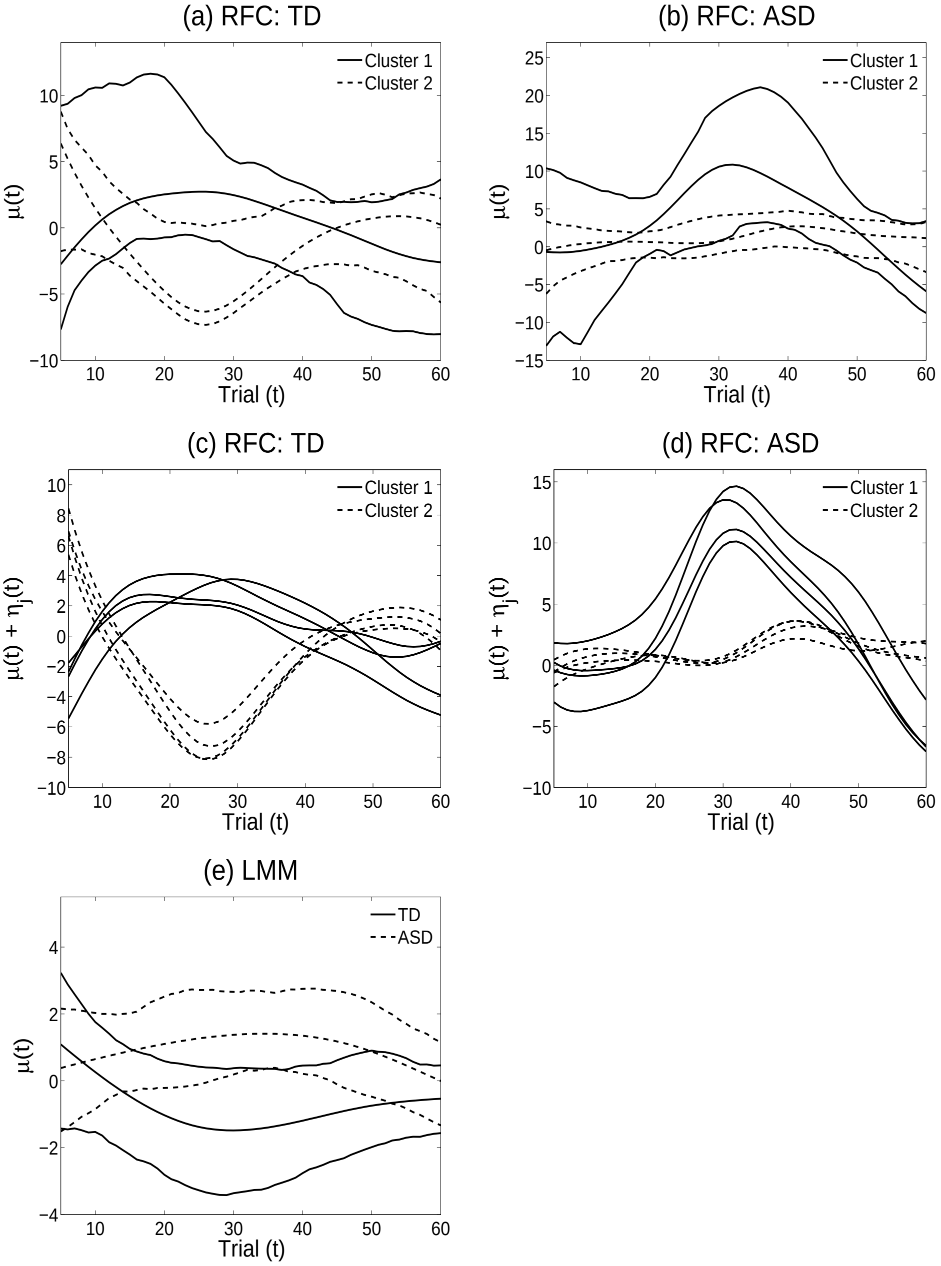


Figure 3.3: The estimated cluster mean functions obtained from RFC for the TD (a) and ASD (b) groups along with 90% bootstrap confidence bands. The estimated electrode-specific cluster mean functions are also plotted for the TD (c) and ASD (d) groups. The estimated mean functions from the linear mixed effects model of Hasenstab et al. (2015) overlaying 90% bootstrap bands are given in (e).

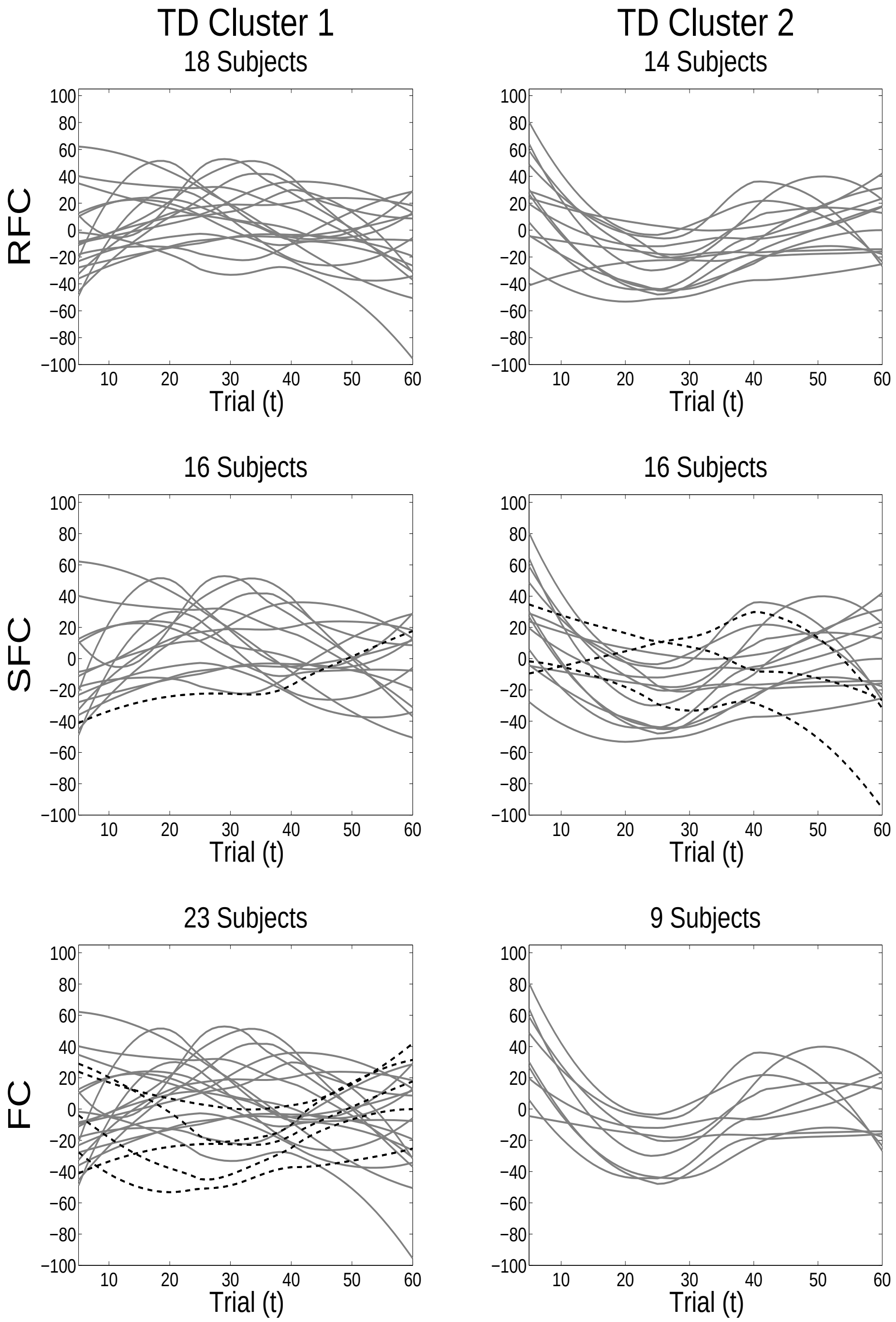


Figure 3.4: The smoothed P3 amplitude difference trajectories across electrodes for each algorithm (row) and cluster (column) within the TD group. The trajectories in SFC and FC with different clustering assignment than the proposed RFC are given dashed.

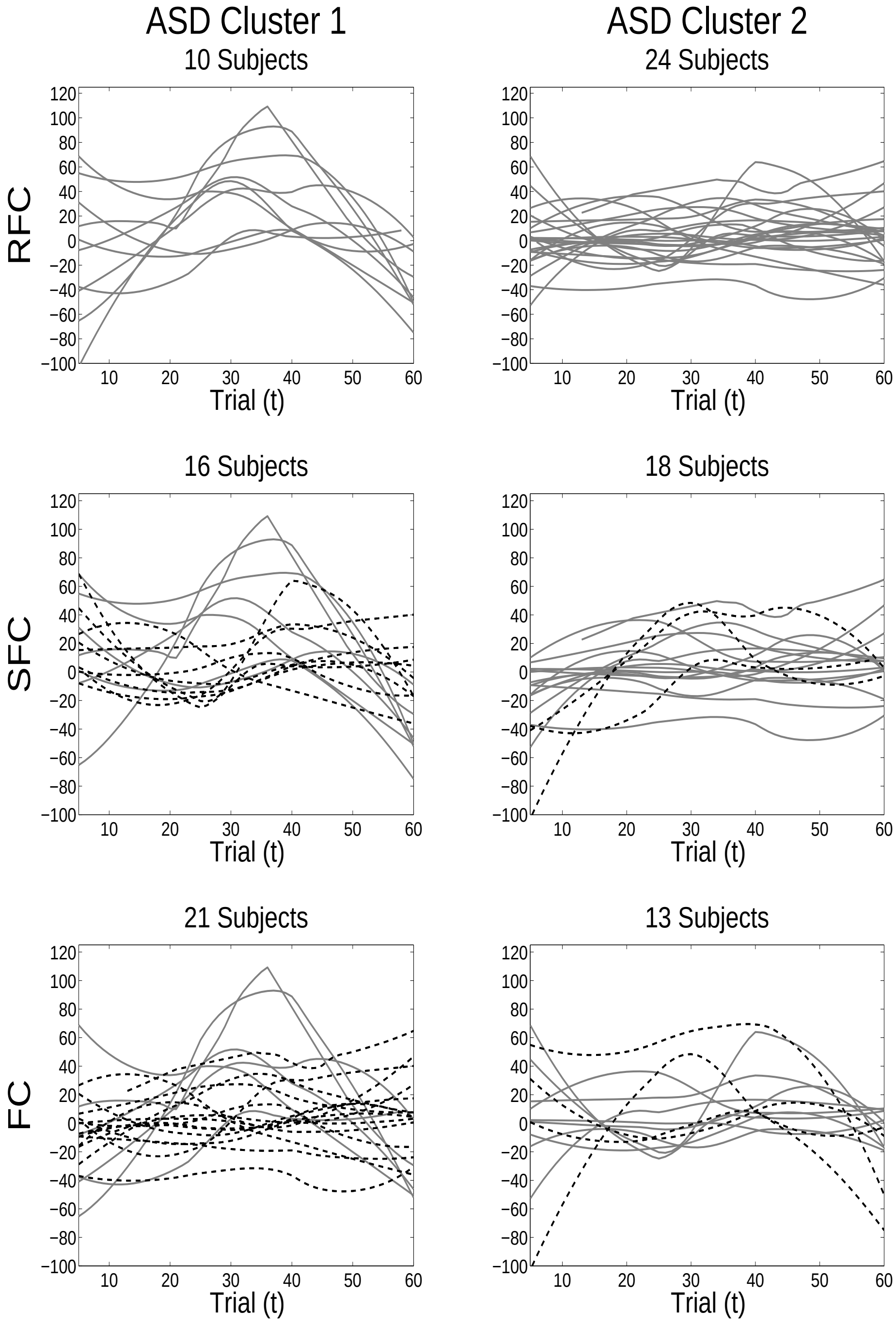


Figure 3.5: The smoothed P3 amplitude difference trajectories across electrodes for each algorithm (row) and cluster (column) within the ASD group. The trajectories in SFC and FC with different clustering assignment than the proposed RFC are given dashed.

trial 30), implying implicit learning is taking place. Furthermore, the differentiation is in opposite directions; while the ASD group has a positive mean condition difference, the TD group has a negative mean. Even though only the positive mean of the ASD group was found significant between trials 30 and 40 based on the bootstrap confidence bands, group differences are found to be reliably significant between trials 20 and 50 (not displayed) implying different patterns of discrimination between the two groups.

While the LMM modeling of the meta-preprocessed data provides novel insights on a new longitudinal dimension that is typically lost in analysis of ERP data, leading to interpretable group differences over trials with respect to patterns of implicit learning, this approach centers around only the mean trends and does not provide insights into the possible diversity of learning patterns within groups. Figures 3.3(a)-(b) show the mean condition differentiation of two diverse clusters of children identified within the TD and ASD groups, overlaid with their respective 90% bootstrap confidence bands. The TD group contains two clusters with roughly equal numbers of children showing condition differentiation in opposing positive and negative directions, while the ASD group comprises of a subgroup of children (n=24) with a flat mean condition differentiation and another subgroup (n=10) with a positive mean differentiation pattern. While the average pattern over the two subgroups within the TD and ASD groups are consistent with the findings of Hasenstab et al. (2015) [13], with a negative overall mean differentiation pattern for TD and a positive overall mean pattern for ASD, they identify diverse subgroups within each diagnostic group, implying that not all TD and ASD children display opposing trends of condition differentiation. In fact, most children in the ASD group are in the cluster with a flat differentiation pattern indicating little or no implicit learning, while others differentiate positively between the conditions, similar to roughly half of the TD children. RFC analysis shows that the negative differentiation pattern of half of the TD children is not shared by children with ASD. These findings provide novel insights into the diversity of implicit learning patterns within each group, while also enable comparisons across groups.

3.4.3 Comparison of clustering results

The RFC clustering results are further compared to clusters obtained via a simpler version of the algorithm that assumes a single covariance subset (referred to as the single subset functional clustering (SFC)) and a multilevel extension of the FC algorithm of Chiou and Li (2007) [75]. Smoothed amplitude difference trajectories across electrodes from all clustering algorithms are also displayed in Figures 3.4 and 3.5. For the TD group, SFC yields similar clustering results to RFC with a few differences in cluster assignments and an equal subject split across clusters. In contrast, the SFC results are quite different than RFC for the ASD group, allocating several of the subjects from the cluster with the flat mean to the cluster with the positive mean. This is consistent with the prior observations where trajectories of the number of averaged ERPs and estimated eigenfunctions confirm that the ASD group displays higher covariance heterogeneity than the TD group, which makes the single covariance subset assumption of SFC harder to justify. The clustering results from FC are different from those obtained from RFC in both the TD and ASD groups. For the TD group, FC assigns subjects from the cluster with the negative mean to the cluster with the positive mean. For the ASD group, FC assigns almost half of the subjects from the cluster with the flat mean to the cluster with the positive mean. The FC algorithm clusters subject trajectories according to both the mean and covariance trends. Hence due to the covariance heterogeneity in the TD and ASD groups, FC is unable to robustly identify clusters according to differences in the mean trends.

We also compare the three clustering algorithms according to a multilevel extension of the Davies-Bouldin Index (DBI) [85]. The DBI is an internal cluster evaluation metric that assesses within cluster variation and between cluster separation, where smaller DBI values correspond to better clustering results and greater cluster separation. Define $S_c = (1/N_c J) \sum_{i=1}^{N_c} \sum_{j=1}^{J} [\sum_{p=1}^{T_{ij}} \{y_{ij}^{(c,v)}(t_{ijp}) - \mu^{(c)}(t_{ijp}) - \eta_j^{(c)}(t_{ijp})\}^2]^{1/2}$ to be a measure of within cluster variation for mean cluster c where N_c represents the number of subjects in cluster c. Further define $M_{c,c'} = \sum_{j=1}^{J} (\sum_{p=1}^{T_{ij}} [\{\mu^{(c)}(t_{ijp}) + \eta_j^{(c)}(t_{ijp})\} - \{\mu^{(c')}(t_{ijp}) + \eta_j^{(c')}(t_{ijp})\}]^2)^{1/2}$ to be a measure of separation between two clusters c and c'. The DBI is then defined to be $(1/n_c) \sum_{c=1}^{n_c} \max_{c \neq c'} \{(S_c + S_{c'})/M_{c,c'}\},\ c' = 1, \ldots, C,$

where the definition reduces to $(S_c + S_{c'})/M_{c,c'}$ when $C = 2$. The DBI values for (RFC, SFC, FC) are (2.16, 2.44, 2.53) in the TD group and (2.65, 3.47, 3.90) in the ASD group, respectively. Hence RFC achieves a better cluster separation over the other two algorithms and differences in the DBI values are greater in the ASD group. Performance of the three clustering algorithms are further compared via simulations (Section 3.5).

3.5 Simulation studies

We study the performance of the proposed RFC compared to FC and SFC, and study the performance of the algorithm under the second non-identifiability condition outlined in Section 3.3.2 that the cluster mean functions lie in the same or different covariance subset eigenspaces. We summarize the findings in this section and defer the simulation details to the online Supplementary Material. We conducted simulations under five scenarios. The first two cases correspond to the second non-identifiability condition with cluster mean functions lying in the eigenspace of the same covariance subset (case 1) and different covariance subsets (case 2). The cluster and covariance subset memberships are not assumed to be identical. The last three simulation scenarios correspond to the assumptions of RFC, SFC and FC, respectively: that the cluster and covariance subset memberships are not identical (case 3); that there is a single covariance subset for the entire sample (case 4); and cluster and covariance subset membership are set to be the same (case 5). All three algorithms perform poorly in the first two simulation cases of non-identifiability conditions, since the cluster means lying in the eigenspace of the covariance subsets is also a non-identifiable case for SFC and FC with non-overlapping cluster and covariance subset memberships. RFC outperforms SFC and FC in the third simulation case, improving cluster quality by incorporating the known covariance heterogeneity into the clustering of the mean trends. When the covariance groups are highly similar (simulation case 4), all three algorithms perform equally well as expected. In case (5), RFC is almost as effective in finding clusters as FC, where cluster and covariance subset membership overlap. As in case (3), SFC is unable to recover clusters under the multiple covariance subsets of case (5).

3.6 Discussion

We proposed a robust functional clustering algorithm to identify subgroups within ERP data. The proposed algorithm is tailored to account for known covariance heterogeneity in the sample and is applicable to multilevel functional data. The proposed clustering utilizes the longitudinal patterns over the course of the ERP experiment, which are typically lost in the standard practice of averaging ERPs over all trials and leads to novel insights into the diversity of implicit learning patterns within and across ASD and TD children. Note that the proposed RFC algorithm relies on consistent estimation of the cluster and covariance subset components such as the mean functions, covariance surfaces, eigenfunctions and eigenscores. Even though asymptotic consistency has been established for the proposed estimators (with increasing sample size), finite sample performance of these estimators may affect the performance of the RFC. Another issue is the consistency of the cluster and covariance subset components based on observations from estimated clusters. Almost sure convergence of cluster means for the classical k-means clustering algorithm was established by Pollard (1981) [86] for multivariate data. Chiou and Li (2007) [75] point out that owing to the complexity of convergence and slower convergence rates for estimating cluster means and covariance subset eigenfunctions in functional data, consistency results for the FC need development of further technical results. Similarly, consistency of the RFC requires further research.

Chapter 3 Appendices

Appendix A: Description of single level and multilevel BLUP score calculations

In this section we provide explicit formulas for the BLUP eigenscore estimators for single and multilevel sparse functional data. The single-level FPCA model can be considered a linear mixed effects model with fixed effects mean $\mu^{(c)}(t)$ and random effects $\xi_{ik}^{(c,v)}$. Random effects model matrices are the eigenfunctions corresponding to the random effects scores. Under this framework, scores may be estimated using their best linear unbiased predictions based on the conditional expectation of the scores given the data. Let $y_{ip} = y_i(t_{ip})$, $p = 1, \ldots, T_i$ denote an observation on the i^{th} subject at time t_{ip} and let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT_i})^T$ denote the $T_i \times 1$ observation vector across all time points. Further let $\hat{\boldsymbol{\mu}}_i^{(c)} = \{\hat{\mu}^{(c)}(t_{i1}), \ldots, \hat{\mu}^{(c)}(t_{iT_i})\}^T$, $\widehat{\boldsymbol{\Phi}}_i^{(v)}$ denote the $T_i \times K_v$ matrix of estimated eigenfunctions whose k^{th} column is the vector $\{\hat{\phi}_k^{(v)}(t_{i1}), \ldots, \hat{\phi}_k^{(v)}(t_{iT_i})\}^T$ and let $\widehat{\boldsymbol{\Lambda}}^{(v)}$ denote the $K_v \times K_v$ diagonal matrix $\widehat{\boldsymbol{\Lambda}}^{(v)} = \text{diag}(\hat{\lambda}_1^{(v)}, \ldots, \hat{\lambda}_{K_v}^{(v)})$. Then the $K_v \times 1$ estimator vector for the single-level functional principal component scores are given by

$$\hat{\boldsymbol{\xi}}_i^{(c,v)} = \widehat{\boldsymbol{\Lambda}}^{(v)} \widehat{\boldsymbol{\Phi}}_i^{(v)T} \left(\widehat{\boldsymbol{\Sigma}}_i^{(v)} + \hat{\sigma^2}^{(v)} I_{T_i \times T_i}\right)^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i^{(c)}),$$

where $\widehat{\boldsymbol{\Sigma}}_i^{(v)} = \widehat{\boldsymbol{\Phi}}_i^{(v)} \widehat{\boldsymbol{\Lambda}}^{(v)} \widehat{\boldsymbol{\Phi}}_i^{(v)T}$ and $I_{T_i \times T_i}$ denotes the $T_i \times T_i$ identity matrix. When estimating the BLUP eigenscore estimates in the leave-one-out procedure of the RFC algorithm, $\hat{\boldsymbol{\mu}}_i^{(c)}$, $\widehat{\boldsymbol{\Phi}}_i^{(v)}$ and $\widehat{\Lambda}^{(v)}$ are replaced with $\hat{\boldsymbol{\mu}}_{(-i)}^{(c)}$, $\widehat{\boldsymbol{\Phi}}_{(-i)}^{(v)}$ and $\widehat{\Lambda}_{(-i)}^{(v)}$.

For the multilevel extension, let $y_{ijp} = y_{ij}(t_{ijp})$, $p = 1, \ldots, T_{ij}$ denote an observation on the i^{th} subject and the j^{th} subunit at time t_{ijp} and let $\boldsymbol{y}_i = \{y_{i11}, \ldots, y_{i1T_{i1}}, \ldots, y_{iJ1}, \ldots, y_{iJT_{iJ}}\}^T$ denote the be a $(\sum_j T_{ij}) \times 1$ observation vector across all time points. Further let $\hat{\boldsymbol{\mu}}_i^{(c)} = \{\hat{\mu}^{(c)}(t_{i11}), \ldots, \hat{\mu}^{(c)}(t_{i1T_{i1}}), \ldots, \hat{\mu}^{(c)}(t_{iJ1}), \ldots, \hat{\mu}^{(c)}(t_{iJT_{iJ}})\}^T$, $\hat{\boldsymbol{\eta}}_j^{(c)} = \{\hat{\eta}_1^{(c)}(t_{i11}), \ldots, \hat{\eta}_1^{(c)}(t_{i1T_{i1}}), \ldots, \hat{\eta}_J^{(c)}(t_{iJ1}), \ldots, \hat{\eta}_J^{(c)}(t_{iJT_{iJ}})\}^T$, $\widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)}$ denote the $T_{ij} \times K_v$ matrix of estimated level 1 eigenfunctions with k^{th} column equal to $\{\hat{\phi}_k^{(1,v)}(t_{ij1}), \ldots, \hat{\phi}_k^{(1,v)}(t_{ijT_{ij}})\}^T$. Similarly let $\widehat{\boldsymbol{\Phi}}_{ij}^{(2,v)}$ denote the $T_{ij} \times L_v$ matrix of estimated level 2 eigenfunctions with ℓ^{th} column equal to $\{\hat{\phi}_\ell^{(2,v)}(t_{ij1}), \ldots, \hat{\phi}_\ell^{(2,v)}(t_{ijT_{ij}})\}^T$. The $K_v \times 1$ level 1 and $L_v J \times 1$ level 2 estimator vectors are then given by

$$\begin{aligned}
\hat{\boldsymbol{\xi}}_i^{(c,v)} &= \widehat{\boldsymbol{D}}_{i,\xi}^{(v)} \left(\widehat{\boldsymbol{\Sigma}}_{T,i}^{(v)}\right)^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i^{(c)} - \hat{\boldsymbol{\eta}}_i^{(c)}), \\
\hat{\boldsymbol{\zeta}}_i^{(c,v)} &= \widehat{\boldsymbol{D}}_{i,\zeta}^{(v)} \left(\widehat{\boldsymbol{\Sigma}}_{T,i}^{(v)}\right)^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i^{(c)} - \hat{\boldsymbol{\eta}}_i^{(c)}),
\end{aligned}$$

where

$$\begin{aligned}\widehat{\boldsymbol{D}}_{i,\xi}^{(v)} &= \{\widehat{\boldsymbol{\Lambda}}^{(1,v)}\widehat{\boldsymbol{\Phi}}_{i1}^{(1,v)^T}, \ldots, \widehat{\boldsymbol{\Lambda}}^{(1,v)}\widehat{\boldsymbol{\Phi}}_{iJ}^{(1,v)^T}\},\\ \widehat{\boldsymbol{D}}_{i,\zeta}^{(v)} &= \{\widehat{\boldsymbol{\Lambda}}^{(2,v)}\widehat{\boldsymbol{\Phi}}_{i1}^{(2,v)^T}, \ldots, \widehat{\boldsymbol{\Lambda}}^{(2,v)}\widehat{\boldsymbol{\Phi}}_{iJ}^{(2,v)^T}\},\end{aligned}$$

and $\widehat{\boldsymbol{\Sigma}}_{T,i}^{(v)}$ denotes the estimated $(\sum_j T_{ij}) \times (\sum_j T_{ij})$ block covariance matrix with the (j, j') block equal to $\widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)}\widehat{\boldsymbol{\Lambda}}^{(1,v)}\widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)^T} + \widehat{\boldsymbol{\Phi}}_{ij}^{(2,v)}\widehat{\boldsymbol{\Lambda}}^{(2,v)}\widehat{\boldsymbol{\Phi}}_{ij}^{(2,v)^T} + \hat{\sigma}^{2(v)} I_{T_{ij}\times T_{ij}}$ when $j = j'$ and $\widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)}\widehat{\boldsymbol{\Lambda}}^{(1,v)}\widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)^T}$ when $j \neq j'$.

Appendix B: Simulation studies

The goals of the simulations are to study the performance of the proposed RFC compared to FC and SFC, and to study the performance of the algorithm under the second non-identifiability condition outlined in Section 3.3.2 that the cluster mean functions lie in the same or different covariance subset eigenspaces. As in the data applications, we consider two clusters and covariance subsets with a total sample size of 35 subjects and $J = 4$ subunits. Response trajectories $y_{ij}^{(c,v)}(t)$ for cluster c and covariance group v are generated using the random effects model

$$y_{ij}^{(c,v)}(t) = \mu^{(c)}(t) + \sum_{k=1}^{K_v} \xi_{ik}^{(c,v)}\phi_k^{(1,v)}(t) + \sum_{\ell=1}^{L_v} \zeta_{ijl}^{(c,v)}\phi_\ell^{(2,v)}(t) + \epsilon_{ij}^{(v)}(t),$$

where subunit-specific mean functions $\eta_j^{(c)}(t)$ are taken to be zero for simplicity. The number of eigen components are taken to be two at both levels ($K_v = L_v = 2$) where eigenscores $\xi_{ik}^{(c,v)}$ and $\zeta_{ijl}^{(c,v)}$ are independently sampled from $N(0, \lambda_k^{(1,v)})$ and $N(0, \lambda_\ell^{(2,v)})$, respectively. The measurement error $\epsilon_{ij}^{(v)}(t)$ is sampled from $N(0, \sigma^{2(v)})$ and observation times t are regular and equally spaced in the interval $\mathcal{T} = [0, 1]$ with $T_{ij} = 40$. The eigenvalues and error variance are selected as $\lambda^{(1,1)} = (500, 200)/d$, $\lambda_1^{(2,1)} = (250, 100)/d$, $\lambda^{(1,2)} = (700, 300)/d$, $\lambda^{(2,2)} = (500, 250)/d$, $\sigma^{2(1)} = 10/d$, $\sigma^{2(2)} = 50/d$ with the divisor d chosen specifically for the particular simulation case considered. The means and eigenfunctions used are defined as follows: $\mu^{(1)}(t) = 2\exp\{-10(t - 0.3)^2\} + \exp\{-1(t - 0.75)^2\} - 1.5$, $\mu^{(2)}(t) = -3\exp\{-10(t - 0.35)^2\} - \exp\{(t - 0.5)^2\} + 2$,

$\phi_1^{(1,1)}(t) = \sqrt{2}\sin(\pi t)$, $\phi_2^{(1,1)}(t) = \sqrt{2}\cos(\pi t)$, $\phi_1^{(2,1)}(t) = \mathbf{1}t$, $\phi_2^{(2,1)}(t) = \sqrt{3}(2t-1)$, $\phi_1^{(1,2)}(t) = \sqrt{2}\sin(2\pi t)$, $\phi_2^{(1,2)}(t) = \sqrt{2}\cos(2\pi t)$, $\phi_1^{(2,2)}(t) = \sqrt{5}(6t^2-6t+1)$, $\phi_2^{(2,2)}(t) = \sqrt{7}(20t^3-30t^2+12t-1)$.

We study the performance of the RFC under five simulation scenarios. The first two cases correspond to the second non-identifiability condition with cluster mean functions lying in the eigenspace of the same covariance subset (case 1) and different covariance subsets (case 2). The cluster and covariance subset memberships are not assumed to be identical. We set $\mu^{(1)}(t) = 1.2\phi_1^{(1,2)}(t)$ and $\mu^{(2)}(t) = -\mu^{(1)}(t)$ for case 1 and $\mu^{(1)}(t) = 1.2\phi_1^{(1,1)}(t)$ and $\mu^{(2)}(t) = 1.2\phi_1^{(1,2)}(t)$ for case 2. The last three simulation scenarios correspond to the assumptions of RFC, SFC and FC, respectively: that the cluster and covariance subset memberships are not identical (case 3); that there is a single covariance subset for the entire sample (case 4); and cluster and covariance subset membership are set to be the same (case 5). In the first three cases, cluster and covariance subset memberships are sampled independently with equal probability. For case 4, we utilize the model components from $v = 2$ for the common covariance subset and set the covariance scale d to 700 (d is taken to be 2000 for the other four cases which leads to comparable SNR in the entire sample across simulation scenarios).

Performance of the clustering algorithms is evaluated using two common measures of cluster quality, the correct classification rate (CR) and the Rand index (RI). The correct classification rate measures the proportion of subjects correctly classified in reference to the true external clustering. CR is defined as the maximum proportion of correctly classified subjects among all cluster label correspondences between the current cluster output and the true external cluster reference. The second measure, RI, quantifies the similarity between two clustering results by accounting for pairings of subjects assigned to the same cluster or different clusters [87]. Specifically, let f and g represent two different clusterings and let f_i and g_i represent the cluster memberships of the i^{th} subject in these two clusterings, respectively. The RI is defined as

$$RI(f,g) = \sum_{i<k}\{I(f_i = f_k, g_i = g_k) + I(f_i \neq f_k, g_i \neq g_k)\}/\binom{n}{2}, \quad i,k = 1,\ldots,n,$$

where I denotes the indicator function and the index counts the subject pairings (e.g. for subjects i and k) that are assigned to the same or different clusters in both clusterings f and g. Larger CR and RI values correspond to higher cluster quality.

The means and percentiles of the CR and RI values from 200 Monte Carlo runs for the RFC, SFC and FC algorithms under the five simulation cases are given in Table 3.1. We also plot the medians of the cluster mean estimates across the 200 Monte Carlo runs along with the true cluster means in Figure 3.6 for different algorithms and simulation cases. The bandwidths of 0.05 and 0.15 for the mean and covariance smooths are selected using generalized cross-validation in a preliminary simulation study and are set at the selected values for the 200 Monte Carlo runs. All three algorithms perform poorly in the first two simulation cases of non-identifiability conditions (Figures 3.6(a)-(b)), since the cluster means lying in the eigenspace of the covariance subsets is also a non-identifiable case for SFC and FC with non-overlapping cluster and covariance subset memberships. RFC outperforms SFC and FC in the third simulation case (Figure 3.6(c)), improving cluster quality by incorporating the known covariance heterogeneity into the clustering of the mean trends. When the covariance groups are highly similar (simulation case 4), all three algorithms perform equally well as expected (Figure 3.6(d)). In case (5), RFC is almost as effective in finding clusters as FC, where cluster and covariance subset membership overlap. As in case (3), SFC is unable to recover clusters under the multiple covariance subsets of case (5) (Figure 3.6(e)). Due to the small sample size, all three algorithms perform poorly occasionally, as reflected in the lower CR and RI values for the 5th percentile.

Table 3.1: The correct classification rate (CR) and the Rand index (RI) means, medians and (5th, 95th) percentiles for the RFC, SFC and FC algorithms over 200 Monte Carlo runs. While the first two simulation cases correspond to the non-identifiability conditions, the last three correspond to the assumptions of RFC, SFC and FC algorithms, respectively.

		RFC		SFC		FC	
		Mean	Percentile	Mean	Percentile	Mean	Percentile
Case 1	CR	0.79	0.80 (0.54, 0.97)	0.84	0.86 (0.59, 1.00)	0.59	0.56 (0.51, 0.84)
	RI	0.69	0.67 (0.49, 0.94)	0.75	0.75 (0.50, 1.00)	0.52	0.49 (0.49, 0.73)
Case 2	CR	0.78	0.80 (0.51, 1.00)	0.62	0.60 (0.51, 0.89)	0.58	0.57 (0.51, 0.69)
	RI	0.71	0.67 (0.49, 1.00)	0.54	0.51 (0.49, 0.79)	0.51	0.50 (0.49, 0.56)
Case 3	CR	0.90	1.00 (0.54, 1.00)	0.76	0.74 (0.51, 1.00)	0.58	0.57 (0.51, 0.81)
	RI	0.87	1.00 (0.49, 1.00)	0.67	0.61 (0.49, 1.00)	0.52	0.50 (0.49, 0.69)
Case 4	CR	0.93	1.00 (0.56, 1.00)	0.96	1.00 (0.57, 1.00)	0.97	1.00 (0.69, 1.00)
	RI	0.90	1.00 (0.49, 1.00)	0.95	1.00 (0.50, 1.00)	0.96	1.00 (0.56, 1.00)
Case 5	CR	0.98	1.00 (0.86, 1.00)	0.77	0.77 (0.54, 0.97)	0.99	1.00 (0.97, 1.00)
	RI	0.97	1.00 (0.75, 1.00)	0.66	0.64 (0.49, 0.94)	0.99	1.00 (0.94, 1.00)

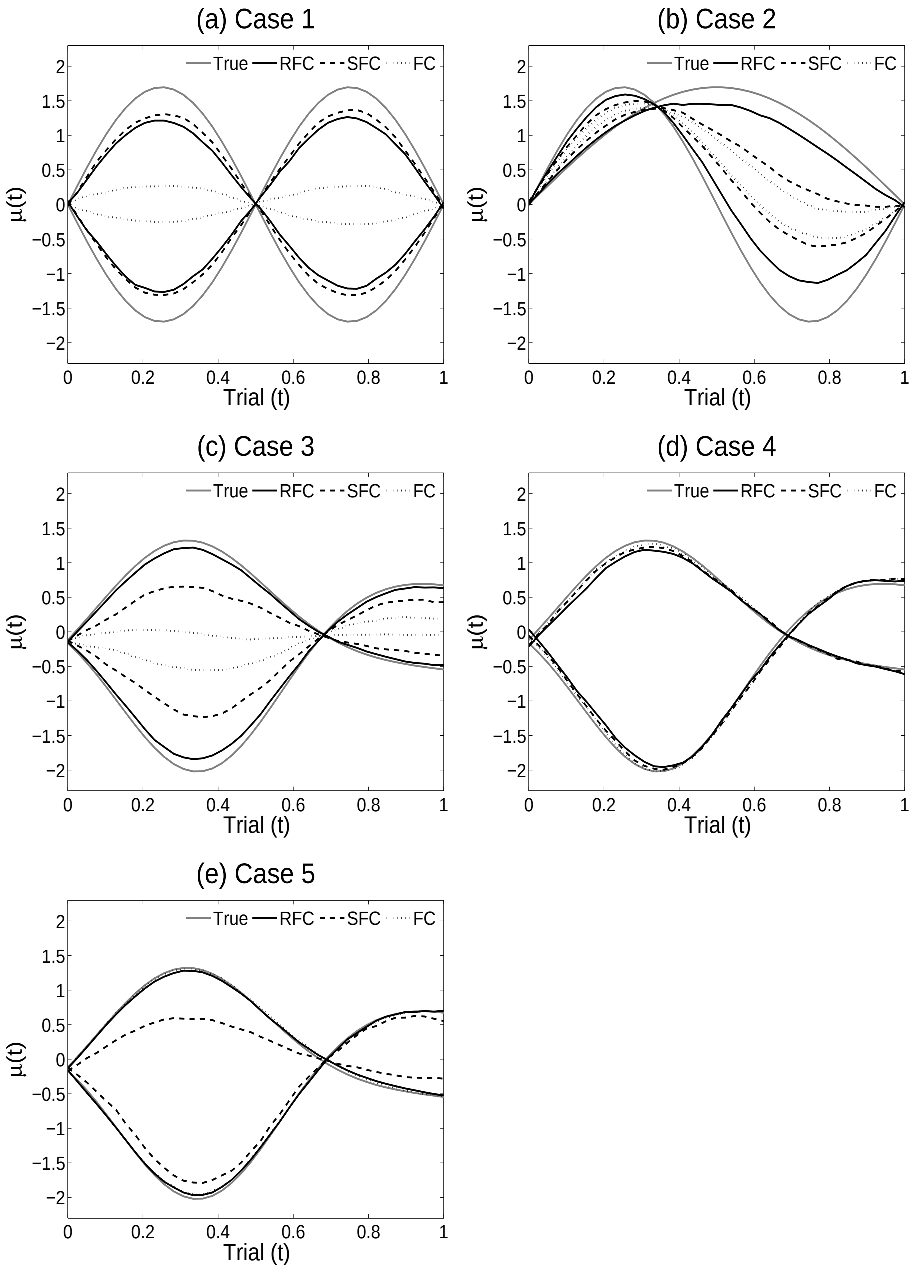


Figure 3.6: The estimated cluster mean functions for the RFC, SFC and FC algorithms over 200 Monte Carlo runs overlaying the true cluster mean functions.

CHAPTER 4

A Multi-Dimensional Functional Principal Components Analysis of EEG Data

Thus far, the proposed MAP-ERP and RFC methods have concentrated on the longitudinal and spatial dimensions of the ERP data. However, these methods only focus on a single measure of the ERP, the P3 magnitude, failing to incorporate information on the functional behavior of the observed ERP. This chapter proposes multidimensional functional principal components analysis (MD-FPCA), a two-stage approach for incorporating information from all three dimensions of the ERP. In addition to the longitudinal dimension of trials and spatial dimension of electrodes, as in Chapters 2 and 3, ERP measures on the entire P3 peak are included in the analysis. The proposed method is generalized to structures containing any repeatedly observed multilevel functional data. This chapter is a work in progress and will eventually be submitted for publication.

4.1 Introduction

The ubiquity of personalized mobile devices and advances in biomedical technologies have recently provided the scientific community with a plethora of unique and challenging data structures containing densely and continuously recorded data observed spatially across several hierarchical geographies. These advances have resulted in the rapid development of functional data analysis methods to temporally and spatially analyze trends hidden within these data. The majority of functional methods in practice are implemented on longitudinal data structures containing information on individual outcomes observed across time or other functional measures [88]. Yao et al. (2005)

[78] proposed PACE, a functional principal components analysis (FPCA) analysis approach for analyzing functional data in the presence of data sparsity. Although these methods are useful for examining complex functional trends, they are inappropriate for data structures containing correlated functional curves. Presently, much of the functional data collected contains a multilevel structure, where multiple functional observations are collected on the same individual. Di et al. (2009, 2014) [79, 80] proposed multilevel FPCA, a method designed to analyze between- and within-repetition variability within multilevel functional data under dense and sparse designs. Other methods for modeling dependent functional data with a hierarchical structure have been proposed by Morris et al. (2003), Morris and Carroll (2006) and Crainiceanu et al. (2009) [89, 90, 91]. Greven et al. (2010) [92] introduced a method for modeling longitudinally observed functional data, similar to the ERP data structure observed at each electrode. In addition, Muller and Chen (2012) [93] proposed double FPCA, a two-stage approach for analyzing entire functional curves observed longitudinally across another functional metric. We propose to extend the double FPCA method of Muller and Chen (2012) [93] to multilevel functional data using the multilevel FPCA decompositions of Di et al. (2009, 2014) [79, 80] in order to summarize and analyze information across the functional, longitudinal and spatial dimensions of event-related potentials (ERP) data.

Appropriate data for the proposed MD-FPCA method will contain a multilevel component by which data is observed at discrete, not necessarily smooth, units. The spatial electrode distribution of the ERP data is an example of this multilevel component. Note that we do not consider changes in electrical potential across the entire electrode netting to be smooth due to the low spatial resolution of ERP data. Furthermore, data is observed across dual temporal dimensions at each of these spatial units producing both functional and longitudinal measurements. The distinction between the terms ‘functional’ and ‘longitudinal’ are made based on study design and scientific meaning. Functions that are observed repeatedly across another temporal axis create longitudinal observations of each functional curve (e.g., entire ERP curves observed across several stimuli). The functional and longitudinal components of the repeated multilevel functional data are assumed to evolve smoothly across their respective axes at each spatial site. Spatial effects are considered discrete repetitions of each functional, longitudinal set and are included as random effects, as in previous chapters.

The proposed MD-FPCA method is a two-stage approach that decomposes total variation into variation between and within functional repetitions across each level of the multilevel structure. The first stage of the two-stage procedure models the multilevel functional component of the data at each longitudinal repetition using multilevel FPCA. The resulting principal components scores are then observed as a function of longitudinal time and modeled using single-level FPCA. Given a set of ERP curves, $X_{ij}(t|s)$, from the i^{th} subject observed over functional time t (ERP time) at longitudinal time s (stimulus repetition) and electrode j, decompositions are performed using a two-stage Karhunen-Loève expansion,

$$\begin{aligned} X_{ij}(t|s) &= \mu(t|s) + \eta_j(t|s) + \sum_{k=1}^{\infty} \xi_{ik}(s)\phi_k^{(1)}(t|s) + \sum_{p=1}^{\infty} \zeta_{ijp}(s)\phi_p^{(2)}(t|s) && (4.1)\\ &= \mu(t|s) + \eta_j(t|s) + \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} \xi'_{ikk'}\psi_{kk'}^{(1)}(s)\phi_k^{(1)}(t|s) + \sum_{p=1}^{\infty}\sum_{p'=1}^{\infty} \zeta'_{ijpp'}\psi_{pp'}^{(2)}(s)\phi_p^{(2)}(t|s) && (4.2)\\ &= \mu(t|s) + \eta_j(t|s) + \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} \xi'_{ikk'}\varphi_{kk'}^{(1)}(t,s) + \sum_{p=1}^{\infty}\sum_{p'=1}^{\infty} \zeta'_{ijpp'}\varphi_{pp'}^{(2)}(t,s). && (4.3) \end{aligned}$$

Model terms $\mu(t|s)$ and $\eta_j(t|s)$ are the overall and subunit-specific mean surfaces, respectively. Terms $\phi_k^{(1)}(t|s)$ and $\phi_p^{(2)}(t|s)$ are the k^{th} and p^{th} eigenfunctions of the level 1 and level 2 decompositions, respectively, resulting from the multilevel FPCA at each longitudinal time s across functional time t with corresponding eigenscores $\xi_{ik}(s)$ and $\zeta_{ijp}(s)$. Model terms $\psi_{kk'}^{(1)}(s)$ and $\psi_{pp'}^{(2)}(s)$ are the eigenfunctions from performing FPCA on the stage one functional eigenscores. The product of the stage one and stage two eigenfunctions representing the modes of variation between and within functional repetitions produces eigensurfaces $\varphi_{kk'}^{(1)}(t,s) = \psi_{kk'}^{(1)}(s)\phi_k^{(1)}(t|s)$ and $\varphi_{pp'}^{(2)}(t,s) = \psi_{pp'}^{(2)}(s)\phi_p^{(2)}(t|s)$ of the overall repeated multilevel functional process. Stage two scores $\xi'_{ikk'}$ and $\zeta'_{ijpp'}$ are interpreted as projections of the data onto the complete eigenspace spanning both temporal dimensions across each level of the multilevel structure. These scores may be utilized for exploratory purposes such as clustering and regression. Note that the axes of t and s are defined as temporal metrics, however, other measures can be used in place of time provided that functions defined along these axes are intrinsically smooth.

The use of a two-stage approach has major advantages, analogous to the advantages discussed in Muller and Chen (2012) [93]. The proposed MD-FPCA method provides a low dimensional representation of complex data spanning three dimensions with computational efficiency. Estimates of the overall eigensurface are produced using several two-dimensional smooths across both temporal dimensions. In contrast, estimates of the overall eigensurface may also be obtained by doing a formal multilevel FPCA decomposition across both temporal dimensions in a single stage using a four-dimensional smooth. However, this would be computationally expensive, processing the high dimensional data structure much slower than the two-stage approach. Moreover, the four-dimensional smooth requires large amounts of data for each functional-longitudinal time pair, thus falling victim to the curse of dimensionality. Sparsity in the data may create instability in the estimates if this approach is selected. In addition, the two-stage decomposition of the total variation of the repeated multilevel functional data into the variation within and between functional repetitions provides another vehicle for interpretation. This multiplicative decomposition is important when the functional and longitudinal domains have inherently different meanings and interpretations as in the ERP from the implicit learning paradigm.

The main goal of this chapter is to propose a method that will take advantage of the information provided in the functional, longitudinal and spatial dimensions of ERP data. The MD-FPCA approach is applied to the ERP data of the implicit learning paradigm and is shown to supply additional information on the behavior of ERP components throughout the course of the experiment. This chapter is organized into the following sections. Section 4.2 describes the proposed MD-FPCA method in detail, including estimation and interpretation of its components. Section 4.3 applies the proposed MD-FPCA approach to the implicit learning paradigm and gives a thorough discussion of results. Simulations are conducted to assess the performance of the proposed MD-FPCA in estimating model components in Section 4.4. This is followed by a discussion of future directions in Section 4.5.

4.2 Multidimensional Functional Principal Components Analysis (MD-FPCA)

The MD-FPCA model is an extension of the double FPCA model of Muller and Chen (2012) [93] to multilevel functional data using multilevel FPCA approaches proposed by Di et al. (2009, 2014) [79, 80] described in Chapter 3. The multilevel extension allows for the incorporation of hierarchical units of functional data, such as the spatial dimension of the ERP from the implicit learning paradigm, into the modeling of repeated functional data. The MD-FPCA model and the estimation of its model components are introduced in Sections 4.2.1 and 4.2.2, respectively. In addition, interpretation of the MD-FPCA model components are included in Section 4.2.3.

4.2.1 MD-FPCA Model

Consider a squared integrable random function $X_{ij}(t|s)$ observed across continuous time t at longitudinal time s, subunit j and subject i, $t \in \mathcal{T}$, $s \in \mathcal{S}$, $j = 1, \ldots, J$, $i = 1, \ldots, n$ with covariance function $\Sigma_T(t, t'|s)$. Similar to Chapter 3, function $X_{ij}(t|s)$ may be decomposed using a multilevel random effects model at each longitudinal time s,

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

where $\mu(t|s)$ are fixed functional effects that represent the overall mean function and subunit-specific shifts, respectively, $Z_i(t|s)$ and $W_{ij}(t)$ are the random subject- and subunit-specific deviations, respectively, and $\epsilon_{ij}(t|s)$ is measurement error with mean zero and variance σ^2 [79, 80]. Assuming the subunit-specific deviations $Z_i(t|s)$ and $W_{ij}(t|s)$ are uncorrelated mean zero stochastic processes, a multilevel FPCA decomposition using Karhunen-Loève expansions at each level of covariation may be performed at each longitudinal time s,

$$X_{ij}(t|s) = \mu(t|s) + \eta_j(t|s) + \sum_{k=1}^{\infty} \xi_{ik}(s)\phi_k^{(1)}(t|s) + \sum_{p=1}^{\infty} \zeta_{ijp}(s)\phi_p^{(2)}(t|s) + \epsilon_{ij}(t|s), \tag{4.4}$$

where $\xi_{ik}(s)$ and $\zeta_{ijp}(s)$ are the level 1 and level 2 eigenscores, respectively, and $\phi_k^{(1)}(t|s)$ and $\phi_p^{(2)}(t|s)$ are the level 1 and level 2 eigenfunctions, respectively. The level 1 and level 2 scores of the first-stage decomposition are considered mean zero stochastic processes defined across longitudinal time s with $\text{var}\{\xi_{ik}(s)\} = \lambda_k^{(1)}(s)$ and $\text{var}\{\zeta_{ijp}(s)\} = \lambda_p^{(2)}(s)$. Subsequently, a second set of Karhunen-Loève expansions are performed on the scores such that

$$\xi_{ik}(s) = \sum_{k'=1}^{\infty} \xi'_{ikk'}\psi_{kk'}^{(1)}(s) \quad \text{and} \quad \zeta_{ijp}(s) = \sum_{p'=1}^{\infty} \zeta'_{ijpp'}\psi_{pp'}^{(2)}(s), \tag{4.5}$$

where $\xi'_{ikk'}$ and $\zeta'_{ijpp'}$ are the eigenscores and $\psi_{kk'}^{(1)}(s)$ and $\psi_{pp'}^{(2)}(s)$ are the eigenfunctions from the level 1 and level 2 second-stage FPCA, respectively, with corresponding eigenvalues $\lambda_{kk'}^{(1)}$ and $\lambda_{pp'}^{(2)}$. Note that a multilevel expansion may be used in place of the single-level expansion for the second stage electrode level decomposition on $\zeta_{ijp}(s)$ when assuming dependence across electrodes j within subject (for details, see Section 4.2.2). Combining equations 4.4 and 4.5 produces the proposed MD-FPCA model,

$$\begin{aligned}
X_{ij}(t|s) &= \mu(t|s) + \eta_j(t|s) + \sum_{k=1}^{\infty} \xi_{ik}(s)\phi_k^{(1)}(t|s) + \sum_{p=1}^{\infty} \zeta_{ijp}(s)\phi_p^{(2)}(t|s) + \epsilon_{ij}(t|s) \\
&= \mu(t|s) + \eta_j(t|s) + \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} \xi'_{ikk'}\psi_{kk'}^{(1)}(s)\phi_k^{(1)}(t|s) + \sum_{p=1}^{\infty}\sum_{p'=1}^{\infty} \zeta'_{ijpp'}\psi_{pp'}^{(2)}(s)\phi_p^{(2)}(t|s) + \epsilon_{ij}(t|s), \\
&= \mu(t|s) + \eta_j(t|s) + \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} \xi'_{ikk'}\varphi_{kk'}^{(1)}(t,s) + \sum_{p=1}^{\infty}\sum_{p'=1}^{\infty} \zeta'_{ijpp'}\varphi_{pp'}^{(2)}(t,s) + \epsilon_{ij}(t|s),
\end{aligned}$$

where $\varphi_{kk'}^{(1)}(t,s) = \psi_{kk'}^{(1)}(s)\phi_k^{(1)}(t|s)$ and $\varphi_{pp'}^{(2)}(t,s) = \psi_{pp'}^{(2)}(s)\phi_p^{(2)}(t|s)$ represent the overall eigensurfaces for the MD-FPCA decomposition. The decomposition may be truncated to only include the components containing the largest modes of variation,

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

where K and P are the number of level 1 and level 2 components from the first decomposition stage, respectively, and K'_k and L'_p are the number of level 1 and level 2 components of the second decomposition stage, respectively.

The proposed MD-FPCA model decomposes the total variation of repeated multilevel functional data into several components of covariation based on the two-stage procedure and the two-level mixed effects FPCA framework. Total covariation at each longitudinal time s is decomposed into within subunit and between subunit covariation based on the mixed effects framework. This is represented by the sum of the between $\Sigma_B(t, t'|s)$ and within $\Sigma_W(t, t'|s)$ covariance functions

$$\begin{aligned}
\Sigma_T(t, t'|s) &= \Sigma_B(t, t'|s) + \Sigma_W(t, t'|s) + \sigma^2, \\
\Sigma_B(t, t'|s) &= \sum_k \lambda_k^{(1)}(s)\phi_k^{(1)}(t|s)\phi_k^{(1)}(t'|s), \\
\Sigma_W(t, t'|s) &= \sum_p \lambda_p^{(2)}(s)\phi_p^{(1)}(t|s)\phi_p^{(1)}(t'|s).
\end{aligned}$$

Furthermore, MD-FPCA decomposes total variation into a multiplicative relationship of the variation within and between longitudinally observed functions. Accounting for functional variation across both the functional and longitudinal domains, total variation of the process $X_{ij}(t|s)$ can provide a covariance function estimate $\mathrm{cov}\{X_{ij}(t, s), X_{ij}(t', s')\}$ of the two-dimensional Karuhunen-Loève expansion so that

$$\begin{aligned}
\Sigma_T(t, s, t', s') &= \Sigma_B(t, s, t', s') + \Sigma_W(t, s, t', s') + \sigma^2, \\
\Sigma_B(t, s, t', s') &= \sum_{k,k'} \lambda_{kk'}^{(1)}\varphi_{kk'}^{(1)}(t, s)\varphi_{kk'}^{(1)}(t', s'), \\
\Sigma_W(t, s, t', s') &= \sum_{p,p'} \lambda_{pp'}^{(2)}\varphi_{pp'}^{(2)}(t, s)\varphi_{pp'}^{(2)}(t', s'),
\end{aligned}$$

where $\Sigma_T(t, s, t', s')$, $\Sigma_B(t, s, t', s')$ and $\Sigma_W(t, s, t', s')$ are estimates of the total, between and within covariance functions of the two-dimensional expansion, respectively, under the multiplicative assumption. Components of these decompositions of covariation provide insight into behavior of functional variation across both functional time t and longitudinal time s.

4.2.2 Estimation of Model Components

Estimation of the proposed MD-FPCA model is presented under the assumption that observations may be sparse in both the functional and longitudinal time domains.The two-stage estimation procedure involves performing MD-FPCA at each longitudinal time s on the repeated multilevel functional data and subsequently performing FPCA on the functional scores resulting from the stage one decomposition. Assume observations for the i^{th} subject $\{X_{ij}(t_{ijq\ell}|s_{ijq}), 1 \leq i \leq n, 1 \leq j \leq J_i, 1 \leq q \leq Q_{ij}, 1 \leq \ell \leq L_{ijq}\}$ are recorded across functional times $t_{ijq\ell}$ and longitudinal times s_{ijq} where j indexes the subunits, q is the index for the longitudinal time s_{ijq} of the i^{th} subject at subunit j and ℓ is the index for the functional time $t_{ijq\ell}$ of the q^{th} repeated function of the j^{th} subunit for the i^{th} subject. The global mean surface $\mu(t|s)$ is estimated by smoothing across all observation pairs $\{t_{ijq\ell}$, $X_{ij}(t_{ijq\ell}|s_{ijq})\}$. This may be done using a one-dimensional smoothing across t for each s or two-dimensional smoothing across both t and s to ensure smoothness in both dimensions. Smoothing in our application is performed across both dimensions. Since the time points are regular in our application, smoothing is performed on the matrix consisting of averages across subjects and subunits for each time pair (t, s). Subunit-specific deviations $\eta_j(t|s)$ are similarly estimated by smoothing across all mean-centered observation pairs $\{t_{ijq\ell},\ X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}|s_{ijq})\}$ for $j = 1, \ldots, J$. Bandwidths for the smooths may be selected using cross-validation, generalized cross-validation (GCV) or the geometric mean [77]. In the application to the implicit learning paradigm, bandwidths are selected based on visual inspection and interpretation.

Covariance functions are estimated using the multilevel FPCA framework of Di et al. (2014) [80] at each longitudinal time s. For fixed longitudinal time s, the between covariance function $\Sigma_B(t, t'|s)$ is estimated by performing a bivariate smooth on $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}|s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}|s_{ijq})\}\{X_{ij'}(t_{ij'q\ell'}|s_{ij'q}) - \hat{\mu}(t_{ij'q\ell'}|s_{ij'q}) - \hat{\eta}_{j'}(t_{ij'q\ell'}|s_{ij'q})\}$ against functional time pairs $\{t_{ijq\ell},\ t_{ij'q\ell'}\}$ for all i, j, j', s, ℓ, ℓ' where $j \neq j'$ (products are taken between different subunits j and j'). The total covariance function $\Sigma_T(t, t'|s)$ is estimated by performing a bivariate smooth on $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}|s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}|s_{ijq})\}$ $\{X_{ij}(t_{ijq\ell'}|s_{ijq}) - \hat{\mu}(t_{ijq\ell'}|s_{ijq}) - \hat{\eta}_j(t_{ijq\ell'}|s_{ij'q})\}$ against functional time pairs $\{t_{ijq\ell},\, t_{ijq\ell'}\}$ for all i, j, s, ℓ, ℓ' (products are within the same subunits j). The within covariance function is estimated by $\widehat{\Sigma}_W(t, t'|s) = \widehat{\Sigma}_T(t, t'|s) - \widehat{\Sigma}_B(t, t'|s)$.

Once covariance function estimates are obtained at each longitudinal time s, perform PCA on covariance functions $\widehat{\Sigma}_B(t, t'|s)$ and $\widehat{\Sigma}_W(t, t'|s)$ for each s. This produces eigenvalue-eigenfunction pairs $\{\hat{\lambda}_k^{(1)}(s), \hat{\phi}_k^{(1)}(t|s)\}$ and $\{\hat{\lambda}_p^{(2)}(s), \hat{\phi}_p^{(2)}(t|s)\}$ for the between and within covariance functions, respectively. Components corresponding to negative eigenvalues are omitted to maintain positive definite covariance functions. Measurement error variance $\sigma^2(s)$ is estimated by smoothing $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}|s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}|s_{ijq})\}^2 - \widehat{\Sigma}_T(t_{ijq\ell}, t_{ijq\ell}|s_{ijq})$ against functional times $t_{ijq\ell}$ for all i, j, q at each longitudinal time s. The number of components may be selected using percentage of variation explained with criteria

$$\max_s\{\sum_{k=1}^{K} \hat{\lambda}_k^{(1)}(s)\}/\{\sum_{k=1}^{M_1} \hat{\lambda}_k^{(1)}(s)\} > 0.9,$$
$$\max_s\{\sum_{P=1}^{P} \hat{\lambda}_p^{(2)}(s)\}/\{\sum_{p=1}^{M_2} \hat{\lambda}_k^{(2)}(s)\} > 0.9,$$

for levels one and two, respectively, where M_1 and M_2 are large [93]. This approach ensures that decompositions at each longitudinal time s contain the same number of principal components for subsequent modeling of the stage one scores. Other methods for selecting the number of components that may be extended to the MD-FPCA case include cross-validation [77] and Akaike's Information Criterion [78].

Based on the resulting principal components, the issue of crossing eigenvalues must be addressed. Let $\hat{\phi}_k^{(1)}(\cdot|s_0)$ be the level 1 eigenfunction estimate at the left endpoint s_0 of the longitudinal time domain. The sign of the next eigenfunction across s, $\hat{\phi}_k^{(1)}(\cdot|s_1)$, is determined such that the L^2 distance to the previous eigenfunction $\hat{\phi}_k^{(1)}(\cdot|s_0)$ is minimized. This is recursively performed for the entire domain of s for both the level 1 and level 2 eigenfunctions across k and p, respectively, and ensures smoothness of the eigenfunctions across s. Stage one scores $\hat{\xi}_{ik}(s)$ and $\hat{\zeta}_{ijp}(s)$ are then calculated from the ordered eigenfunctions using their best linear unbiased predictors (BLUP) described in Chapter 3.

After obtaining the stage one eigenscores functionally observed across s, perform FPCA across longitudinal time s on $\hat{\xi}_{ik}(s)$ for each k. This results in estimates for $\hat{\xi}_{ikk'}$, $\hat{\psi}^{(1)}_{kk'}(s)$. In addition, perform FPCA on $\hat{\zeta}_{ijp}(s)$ for each p, which results in estimates $\hat{\zeta}_{ijpp'}$, $\hat{\psi}^{(2)}_{pp'}(s)$. Note that the single-level FPCA on $\hat{\zeta}_{ijp}(s)$ assumes independence across electrodes within subject. The first stage in the MFPCA procedure decomposes the multilevel process into subject level and electrode level variation, specifically in the functional time direction. Although the majority of subject level information across both the functional and longitudinal dimensions is contained within the $\hat{\xi}_{ik}(s)$ trajectories, residual subject level effects in the longitudinal direction may remain in the $\hat{\zeta}_{ijp}(s)$ trajectories. These effects were shown to be negligible in the data analysis, where the majority of variability in the $\hat{\zeta}_{ijp}(s)$ processes were observed at the electrode level. However, note that the multilevel FPCA used to estimate components in the first stage may also be used to further decompose $\hat{\zeta}_{ijp}(s)$ into another set of subject and electrode level components when observing substantial subject-level variability in the $\hat{\zeta}_{ijp}(s)$ trajectories.

Resulting estimates from the two-stage MD-FPCA include model components $\hat{\mu}(t|s)$, $\hat{\eta}_j(t|s)$, $\hat{\xi}'_{ikk'}$, $\hat{\psi}^{(1)}_{kk'}(s)$, $\hat{\phi}^{(1)}_k(t|s)$, $\hat{\zeta}'_{ijpp'}$, $\hat{\psi}^{(2)}_{pp'}(s)$, $\hat{\phi}^{(2)}_p(t|s)$. In the presence of sparsity, the MD-FPCA approach may be used to predict trajectories across each dimension of the data. Predictions for this model may be performed using

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

4.2.3 Interpretation and Analysis of Model Components

Estimated components of the MD-FPCA model provide a wealth of information on the magnitude and covariation of repeated multilevel functional data across the several dimensions. Mean surfaces $\mu(t|s)$ and $\eta_j(t|s)$ provide insight into the evolution of the response across both functional and longitudinal time. Stage one eigenfunctions $\phi^{(1)}_k(t|s), \phi^{(2)}_p(t|s)$ describe modes of variation across functional time t at each longitudinal time s. The dimension of each repeated function is reduced into a set of principal components whose scores $\xi_{ik}(s), \zeta_{ijp}(s)$ are observable across longitudinal time s. Level 1 scores characterize the overall functional behavior across subunits of a given individual, in contrast to the level 2 scores, which characterize behavior between subunits.

Stage two eigenfunctions $\psi^{(1)}_{kk'}(s), \psi^{(2)}_{pp'}(s)$ describe the modes of variation of the stage one eigenscores. The corresponding scores $\xi'_{ikk'}, \zeta'_{ijpp'}$ are useful for characterizing the behavior of an individual's data both in the functional and longitudinal dimensions. The principal surfaces of the two-stage decomposition $\varphi^{(1)}_{kk'}(t, s), \varphi^{(2)}_{kk'}(t, s)$ provide information on the modes of variation across both time dimensions. Large principal surface values indicate high variability at a specific functional time t and longitudinal time s.

When analyzing multilevel data, an important measure is the proportion of variability explained by the level one component of the data. This may be calculated by

$$\rho = \frac{\sum_{k,k'} \lambda^{(1)}_{kk'}}{\sum_{k,k'} \lambda^{(1)}_{kk'} + \sum_{p,p'} \lambda^{(2)}_{pp'}}$$

and can be considered a longitudinal, functional extension to the intra-cluster correlation of the linear mixed effects model framework. The intraclass correlation extension may also be interpreted as the average correlation between two surfaces from the same subject. In addition, adding and subtracting a multiple of the principal surface to the mean surface $\{\mu(t|s) \pm \sqrt{\lambda^{(1)}_{kk'}} \varphi^{(1)}_{kk'}(t, s)\}$ provides a visualization of the behavior of individual surfaces scoring high or low in specific components. All of these measures are useful for summarizing the overall trends and dominant modes of variation in the repeated multilevel functional data.

4.3 Application to the Implicit Learning Study

4.3.1 Description of Data Structure

Chapters 2 and 3 have throughly examined the longitudinal dimension of the ERP data represented by the stimulus repetitions while accounting for the spatial distribution of electrodes. However, these analyses were performed using a single magnitude of the P3 peak feature on the entire ERP curve. We consider the functional dimension of the ERP in this application defined by the span of the entire P3 peak. As done previously, ERP were preprocessed according to the procedures described in Chapter 1.3. ERP were then meta-preprocessed using a moving average window of 30 to identify the P3 peak for each subject, electrode and condition across the first 60 trials, where maximal condition differentiation was observed to occur. Functional observations on the ERP curve within 70ms to the left and right of the P3 peak were extracted for each subject, electrode and condition. The width of the peak (140ms) was determined based on inspection of the ERP averaged across several trials. Analysis of implicit learning is performed on the expected - unexpected condition difference of the entire P3 peak observed over the course of the experiment. P3 peaks under the unexpected condition are subtracted from P3 peaks under the expected condition within subject, electrode and trial to produce P3 peak differentiation measures. Peak differentiation time (ERP time) is considered functional time t and trial time (ERP repetitions) is considered longitudinal time s. Note that all curves are aligned so that the location of the P3 peak is at the center of the functional time domain (-70, 70).

4.3.2 MD-FPCA Results

The proposed MD-FPCA algorithm is applied to the repeated multilevel P3 difference trajectories separately for the TD and ASD groups. Six subjects are removed as outliers prior to analysis. Two of the removed subjects (one in the TD and one in the ASD group) did not have observed data until trial 20 of the experiment, and the remaining four subjects (two in the TD and two in the ASD group) had peak condition differences more than 2 standard deviations away from their respective group means for most of the observations across either the functional or longitudinal time domains. In addition, a single electrode was omitted from the trajectories of two subjects in the TD group due to erratic and heterogeneous behavior. Since the emphasis of the study is on the functional and longitudinal dimensions of the ERP while only accounting for electrode effects, the analysis of the repeated functional ERP of the TD and ASD groups focuses on the results from the level one terms of the MD-FPCA models. Bandwidths of the mean and covariance smooths are selected based on visual inspection to produce optimal interpretation. Bandwidths for the stage one TD model fit were selected to be 30ms for the mean smooth and 15ms for the covariance smooth across a functional time domain of [-70ms, 70ms] incremented by 4ms. Stage two bandwidths were selected to be 10 trials and 5 trials for the mean and covariance smooths, respectively, across the longitudinal trial domain of [5 trials, 60 trials]. Note that the longitudinal time domain starts at trial 5 to avoid noisy endpoint effects produced by meta-preprocessing as in the previous chapter. Bandwidths for the ASD model fit are selected to be 30ms and 15ms for the mean and covariance smooths of the stage one decompositions, respectively, and 15 trials and 5 trials for the mean and covariance smooths of the stage two decompositions, respectively.

The overall mean surface estimates $\mu(t|s)$ of the P3 peak condition difference for the TD and ASD MD-FPCA fits are shown in the left and right panels of Figure 4.1, respectively. Note that the overall mean surfaces represent condition differentiation and not the actual P3 peak, which is only observed in the functional time domain. The ASD mean surface displays a trend of positive concave condition differentiation across trials that is uniform across peak time. The mean surface peaks around trial 35 where there is a slight differential increase at the P3 peak location ($t = 0$). 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 in negative differentiation around trial 25. Changes in condition differentiation across trials are not uniform in peak time, particularly at intermediate trials where there is a negative dip at $t = 0$ associated with positive values at the endpoints of the peak time domain. The positive and negative differentiation of the ASD and TD groups, respectively, are consistent with the results from Chapter 2. The smaller magnitude of the TD group mean surface is due to the omission of a largely negative outlier. The subunit-specific means of the ASD group (Figure 4.3) indicate consistency across all four electrodes. Although maintaining a consistent shape in condition differentiation, the TD subunit-specific means vary in magnitude at points of negative differentiation prior to trial 30, with strongest condition differentiation in the two electrodes shown in Figures 4.2 (a) and (d), respectively.

The number of principal components of the MD-FPCA fits are selected to explain 90% of the variation in both stages of the MD-FPCA. Table 4.1 breaks down the percentage of variance explained for each selected component across the subject and electrode levels of the MD-FPCA models for the TD and ASD groups. Two components explained at least 90% of the variation in the data in the stage one decompositions for both levels and groups. A larger number of components were required to exceed the 90% threshold in the stage two decompositions. The majority of variation is explained by the leading components of the level one decompositions. Differences in the variation explained across the subject and electrode levels is slightly smaller for the ASD group due to higher variability across electrodes within subject. In total, the subject level components explain approximately 72% and 62% of the total variation for the TD and ASD groups, respectively.

Table 4.1: Number of components selected in the stage one and stage two decompositions of the proposed MD-FPCA model using percentage of variation explained $> 90\%$. Indexes k and p indicate level one and level two components, respectively, and the prime indicates the stage two decomposition. The majority of variation is explained by the first component of the first stage decomposition at the subject level.

Level 1

	TD		ASD	
	$k = 1$	$k = 2$	$k = 1$	$k = 2$
$k' = 1$	.369	.013	.270	.018
$k' = 2$	.186	.009	.184	.010
$k' = 3$	.141	.002	.133	.005
$k' = 4$	/	/	/	.004

Level 2

	TD		ASD	
	$p = 1$	$p = 2$	$p = 1$	$p = 2$
$p' = 1$	.118	.007	.217	.006
$p' = 2$	.074	.005	.085	.004
$p' = 3$	.055	.003	.062	.002
$p' = 4$	.014	.001	/	/

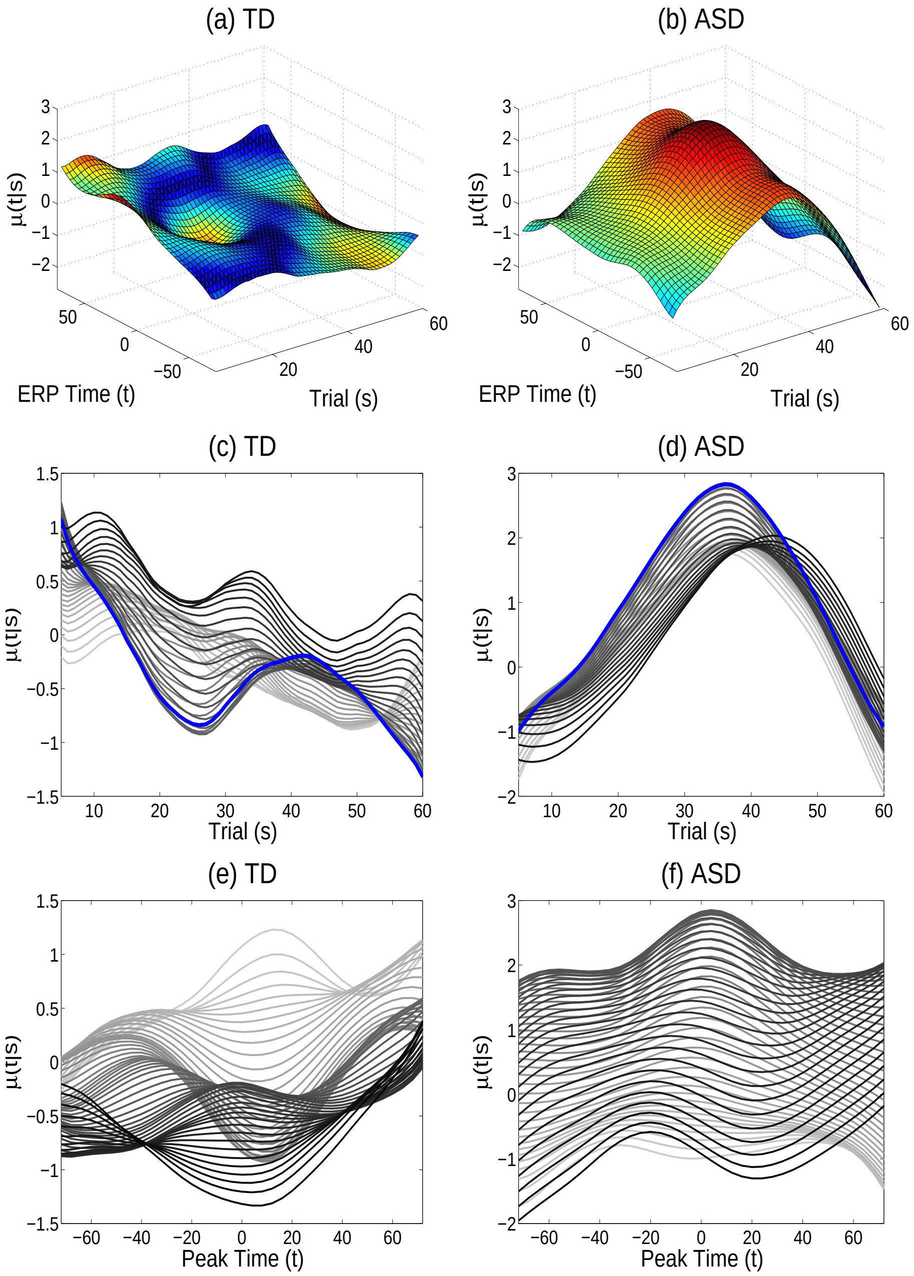


Figure 4.1: Multiple representations of mean surfaces $\mu(t|s)$ for the TD (left column) and ASD (right column) groups. The top panels are the complete mean surfaces, the middle panels are the mean surfaces across peak time represented in the trial domain and the bottom panels are the mean surfaces across trial time represented in the peak time domain. Light gray corresponds to earlier times on the time axis not shown and dark gray corresponds to later times. For example, light gray in (c) correponds to early peak times. The blue lines indicate the trajectory across trials at peak time $t = 0$. TD and ASD groups have opposing condition differentiation.

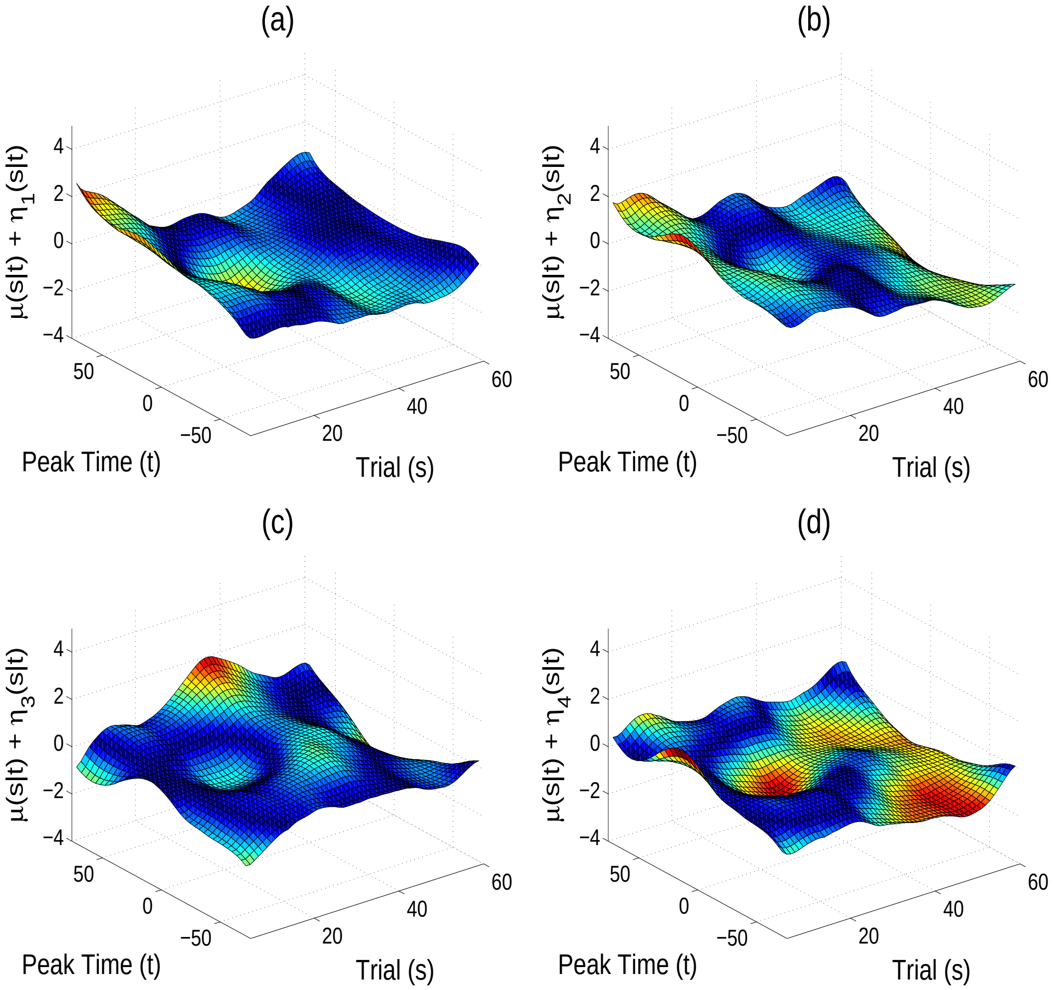


Figure 4.2: TD mean surfaces for each electrode: (a) $\mu(t|s)+\eta_1(t|s)$, (b) $\mu(t|s)+\eta_2(t|s)$, (c) $\mu(t|s)+\eta_3(t|s)$ and (d) $\mu(t|s)+\eta_4(t|s)$. Mean surfaces maintain the same shape as the overall mean but vary in magnitude.

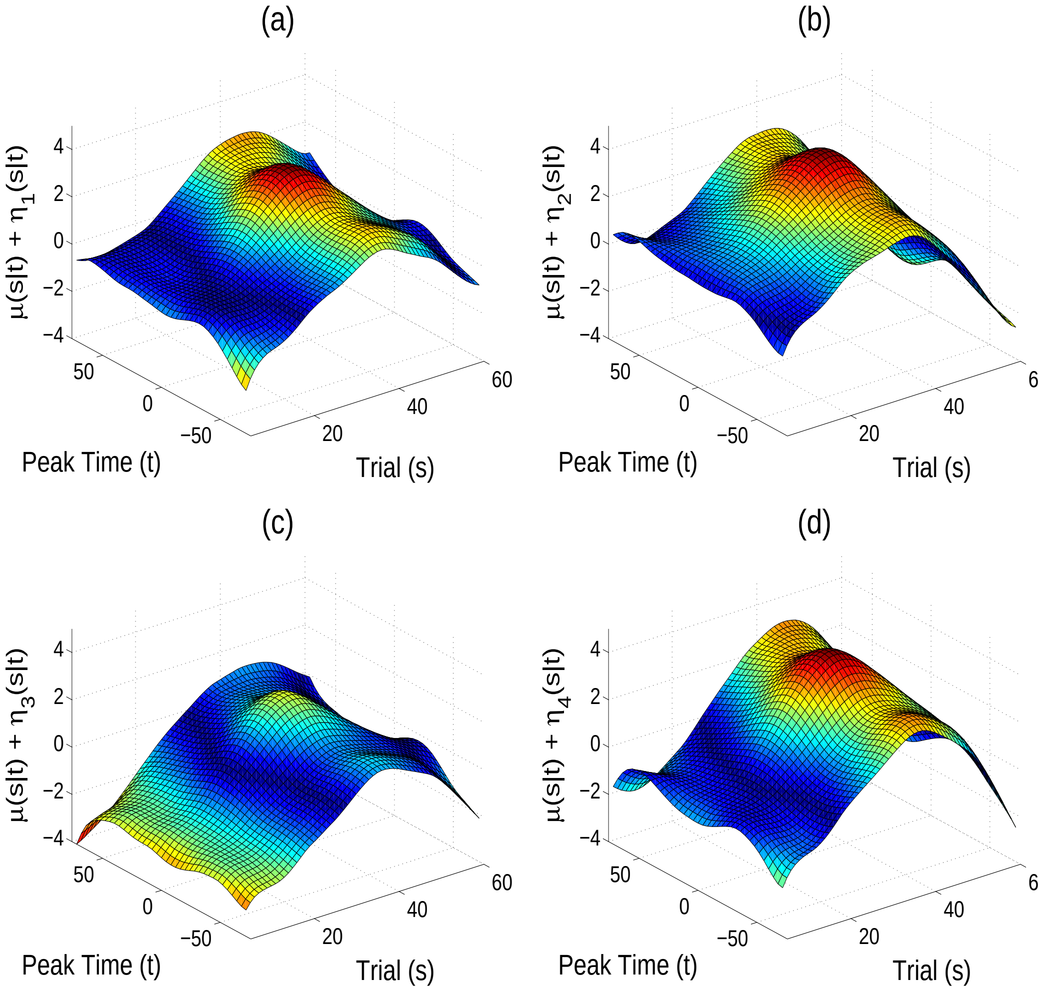
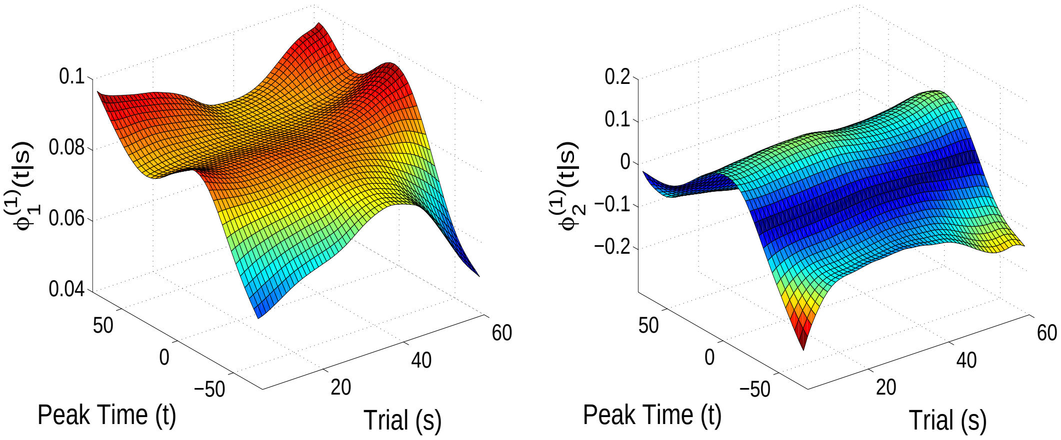


Figure 4.3: ASD mean surfaces for each electrode: (a) $\mu(t|s) + \eta_1(t|s)$, (b) $\mu(t|s) + \eta_2(t|s)$, (c) $\mu(t|s)+\eta_3(t|s)$ and (d) $\mu(t|s)+\eta_4(t|s)$. Mean surfaces are fairly homogeneous and appear similar across all electrodes.

Stage one level one eigenfunctions $\phi_k^{(1)}(t|s)$ and stage two level one eigenfunctions $\psi_{kk'}^{(1)}(s)$, displaying modes of variation in the functional and longitudinal directions, respectively, are shown in Figures 4.4 and 4.5, respectively, for the TD and ASD groups. Uniform variation across ERP time in the ASD leading component $\phi_1^{(1)}(t|s)$ indicates little or no variability around the peak location $t = 0$ with majority of ASD variation in the trial direction at intermediate and later trials for $k = 1, k' = 1$ (27%) and at the boundary trials for $k = 1, k' = 2$ (18.4%). Eigensurface $\phi_2^{(1)}(t|s)$ captures modes of variation around the peak $t = 0$ with additional modes of variation in the trial direction around trial 35 for $k = 2, k' = 1$ (1.8%), the location of the maximum positive condition differentiation in the overall mean surface for the ASD group. There is additional variation in the boundary trials surrounding the intermediate trials for component $k = 2, k' = 2$, explaining only 1% of total variability.

In contrast to the flat contour of the ASD leading eigensurface, the leading component $\phi_1^{(1)}(t|s)$ for the TD group captures modes of variation around the peak location $t = 0$ with modes of variation in the early and intermediate trials for $k = 1, k' = 1$ (36.9%) and later trials for $k = 1, k' = 2$ (18.6%). Note that the leading component $k = 1, k' = 1$ contains a slight increase in negative weight at trial 25, the location of the peak negative condition differentiation in the TD overall mean surface. Eigenfunctions $\phi_2^{(1)}(t|s)$ capture leftover variation around the peak location and variation in the boundaries of the ERP time domain, with modes of variation in the trial direction at boundary trials for $k = 2, k' = 1$ (1.3%) and intermediate trials $k = 2, k' = 2$ (0.9%). These results are consistent with overall eigensurfaces $\varphi_{kk'}^{(1)}(t, s), \varphi_{pp'}^{(2)}(t, s)$ for the TD and ASD groups shown in Figures 4.11 and 4.12, respectively. Modes of variation in the stage one and stage two components at the subject level are supported by the subject-specific trajectories across electrodes shown in Figure 4.8 and the stage one level one score trajectories $\xi_{i1}(s)$ shown in Figure 4.9. Variability in subject trajectories is fairly uniform across ERP time for the ASD group and there appears to be additional variability around the peak location $t = 0$ in the TD group. In addition, score trajectories in Figure 4.9 display the modes of variation at intermediate and boundary trials observed in the stage two level one eigenfunctions.

(a) TD (b) TD



(c) ASD (d) ASD

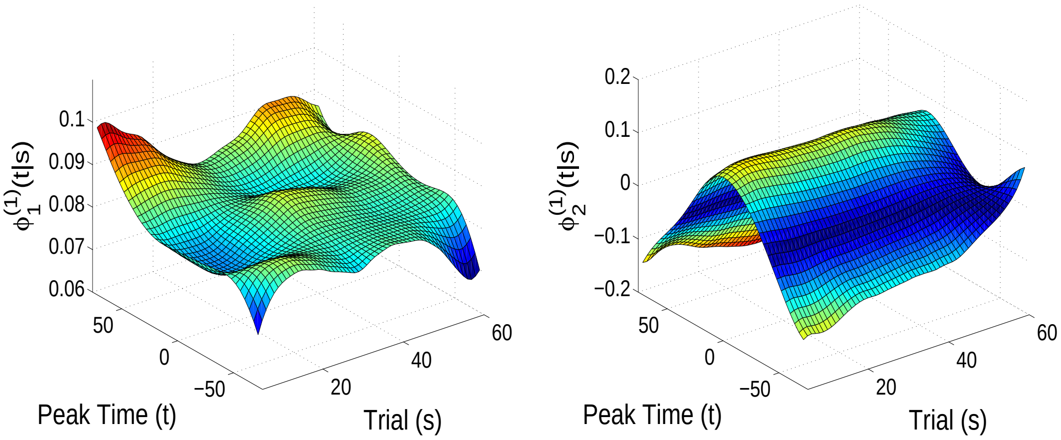


Figure 4.4: Subject-level eigenfunctions $\{\phi_k^{(1)}(t|s)\}$ for components $k = 1$ (left column) and $k = 2$ (right column) from the stage one decomposition across t for each trial s for the TD (top row) and ASD (bottom) row groups.

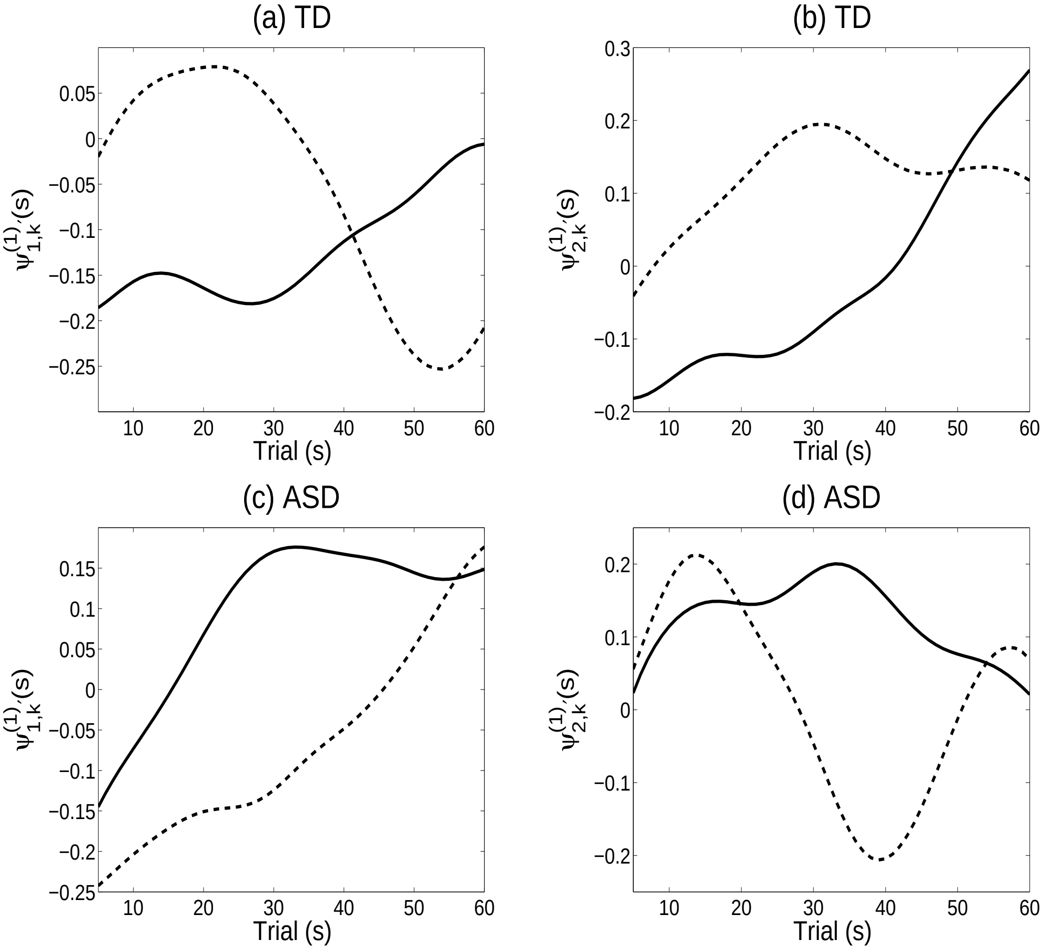


Figure 4.5: Subject-level eigenfunctions $\psi^{(1)}_{k,k'}(s)$ for $k = 1$ (left column) and $k = 2$ (right column) of the stage two decompositions for the TD group (top row) and ASD group (bottom row). Solid lines indicate $k' = 1$ and dashed lines indicate $k' = 2$.

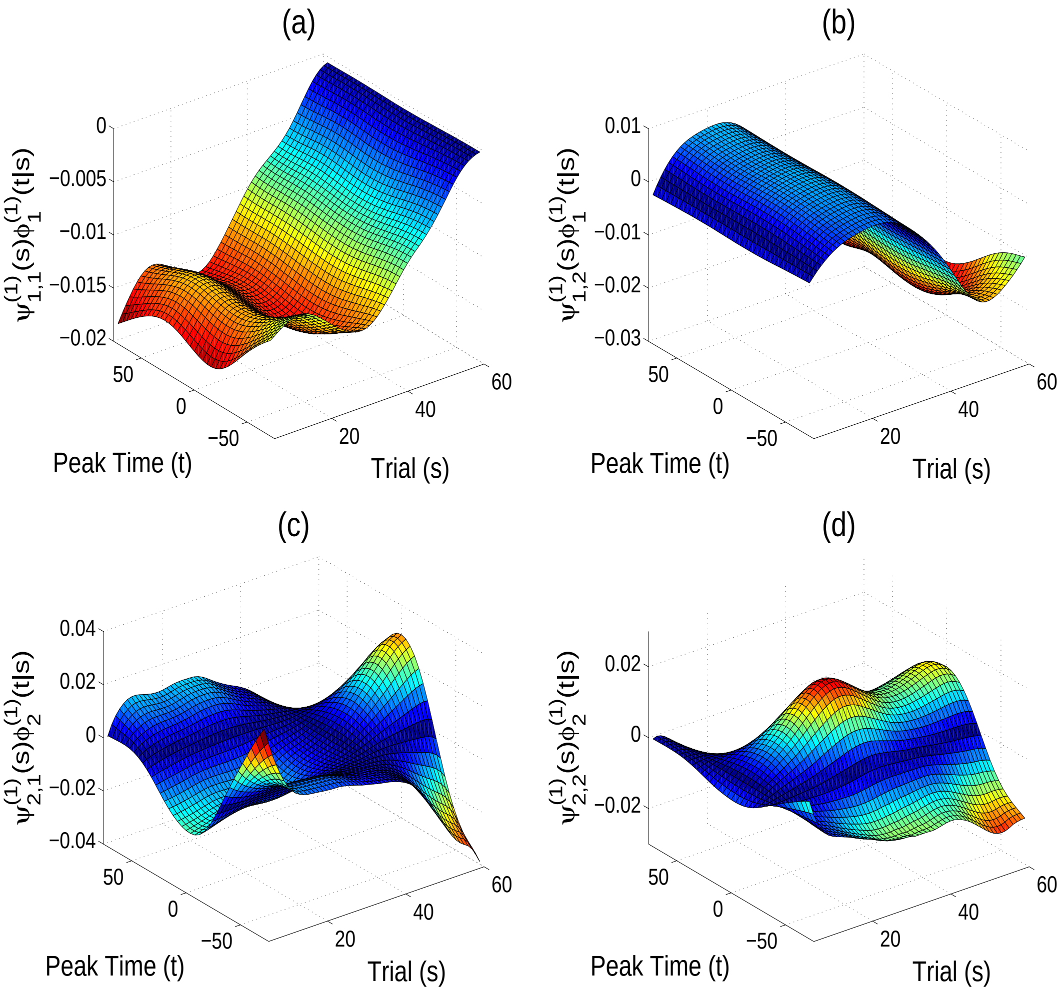


Figure 4.6: TD subject-level eigensurfaces $\varphi^{(1)}_{k,k'}(t,s) = \psi^{(1)}_{k,k'}(s)\phi^{(1)}_{k}(t|s)$ for (a) $k = 1, k' = 1$, (b) $k = 1, k' = 2$, (c) $k = 2, k' = 1$ and (d) $k = 2, k' = 2$. The overall eigensurfaces are dominated by the stage two eigensurface $\psi^{(1)}_{k,k'}(s)$ across the trial domain.

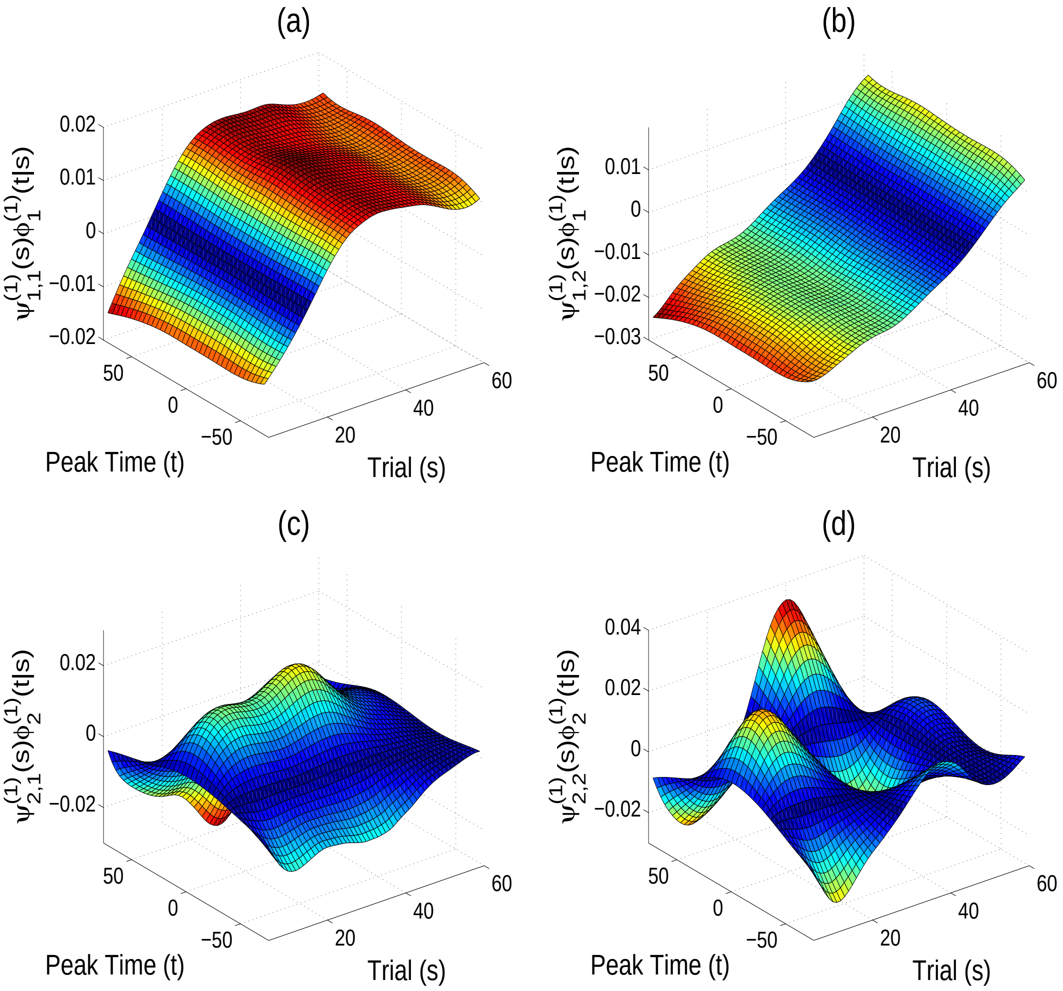


Figure 4.7: ASD subject-level eigensurfaces $\varphi^{(1)}_{k,k'}(t,s) = \psi^{(1)}_{k,k'}(s)\phi^{(1)}_k(t|s)$ for (a) $k = 1, k' = 1$, (b) $k = 1, k' = 2$, (c) $k = 2, k' = 1$ and (d) $k = 2, k' = 2$. The overall eigensurfaces are dominated by the stage two eigensurface $\psi^{(1)}_{k,k'}(s)$ across the trial domain.

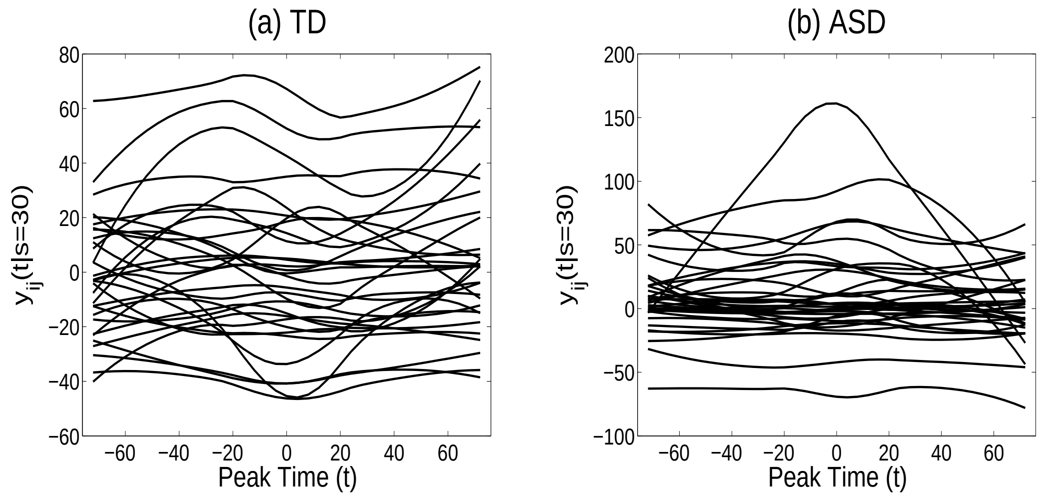


Figure 4.8: Smoothed peak difference trajectories across electrodes along the peak time domain at trial 30 for the (a) TD and (b) ASD groups. Both sets of trajectories appear quite flat across peak time with relatively constant variability across subjects, explaining the flat eigensurfaces of the stage one decomposition.

Leading scores for the principal eigensurfaces ($\xi'_{i1k'}$) are plotted in Figure 4.10 for the TD and ASD groups. Scores appear randomly scattered around zero with no visible clusters for the TD group, but there is a small cluster around zero for the ASD group. The remaining scores for component $k = 2$ (not shown) behave similarly. In order to further understand the behavior of subject-specific surfaces in relation to the magnitudes of the stage two scores, multiples of the principal surfaces are added to and subtracted from the mean surfaces as described in Section 4.2.3. TD intervals (Figure 4.11) suggest stage two leading scores with opposing signs produce surfaces with positive and negative condition differentiation. Intervals for the ASD group contain a surface with a large positive peak and a relatively flat surface for stage two scores with opposing signs. However, the flat surface increases for positive ξ'_{i12} during later trials. The shapes of these surface intervals suggest the existence of clusters in condition differentiation, primarily in the trial direction. Further evidence is shown in the mean surfaces of subjects with $\xi'_{i11} > 0$ and $\xi'_{i11} \leq 0$ (Figure 4.13). The two-dimensional mean surface representation of this score partition on the TD and ASD groups shows potential clustering in the means of condition differentiation. Additional methods used to incorporate more information from the MD-FPCA scores to identify formal subgroups in the ERP should be explored in the future.

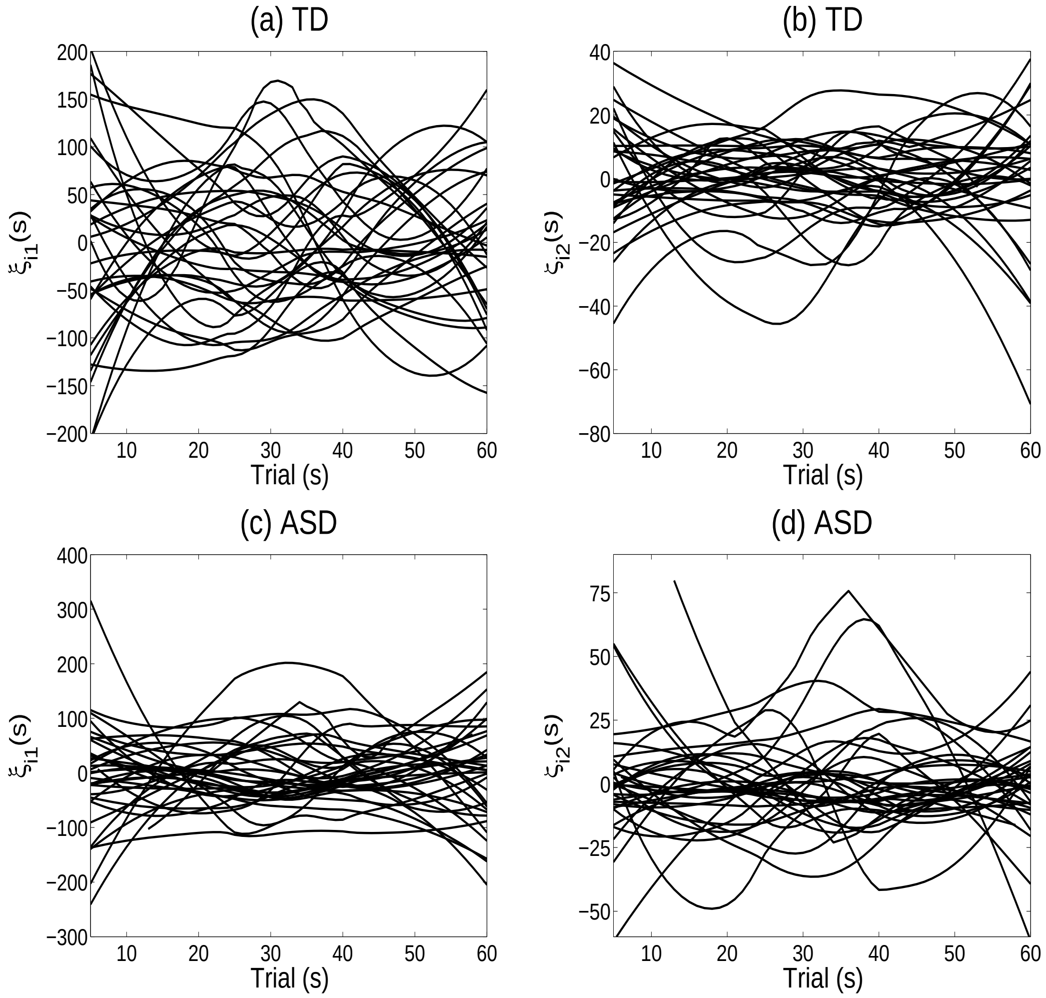


Figure 4.9: Scores $\xi_{ik}(s)$ for $k = 1$ (left column) and $k = 2$ (right column) from the stage one decompositions for the TD (top row) and ASD (bottom row) groups. Score trajectories of the leading component appear similar to the trajectories clustered by RFC.

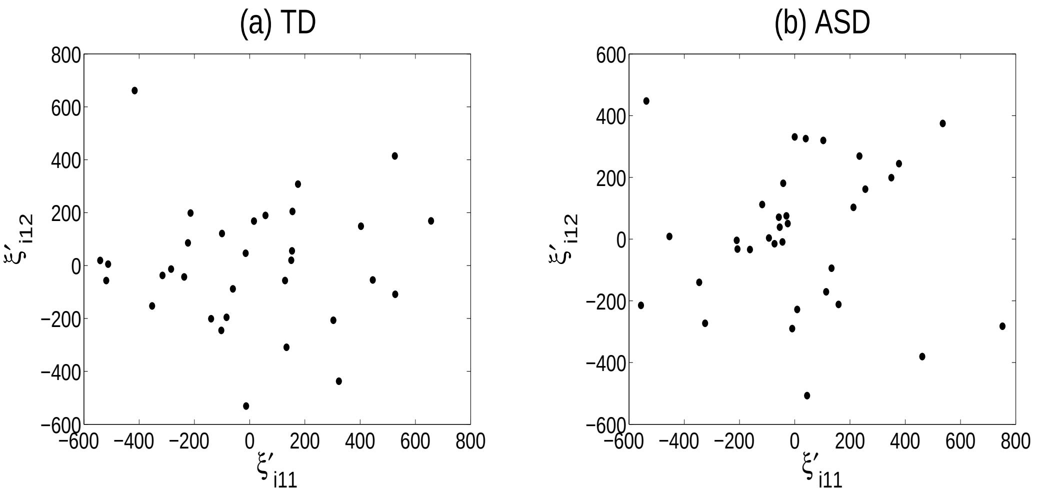


Figure 4.10: Subject level scores ξ'_{i12} vs ξ'_{i11} of the stage two decomposition for the (a) TD and (b) ASD groups. Scores appear randomly scattered across both axes with no visible clustering.

Until now, we have thoroughly discussed the results of the functional and longitudinal dimensions of the model. Although the spatial dimension is included as a nuisance component in the MD-FPCA to account for spatial covariation between electrodes, metrics can be derived from MD-FPCA terms to provide insight into the connectivity of the electrodes. Figure 4.14 shows the smoothed functional analogue of the intraclass correlations (Di et al., 2009) for the stage one multilevel FPCAs at each longitudinal time s for the TD and ASD groups. Both plots indicate the majority of total variability in the stage one decompositions is explained by subject level differences, especially for intermediate trials, implying high connectivity between the electrodes in the right frontal region relative to subject-level variability. Average correlations between peak differentiation trajectories from the same subject across peak time exceed 0.8 within the TD group. Maximum correlations for the ASD group occur during early and intermediate trials with correlations exceeding 0.7. The TD correlations are consistently larger than the correlations for the ASD group, but there appears to be larger subject level variability during early trials for the ASD group. In addition, there are increases in the intraclass correlation around the locations (along s) of maximum differentiation between the clusters found by the RFC algorithm in both groups. However note that these increases may also be attributed to some other unknown effect unrelated to the aforementioned clusters. The functional longitudinal extension to the intraclass

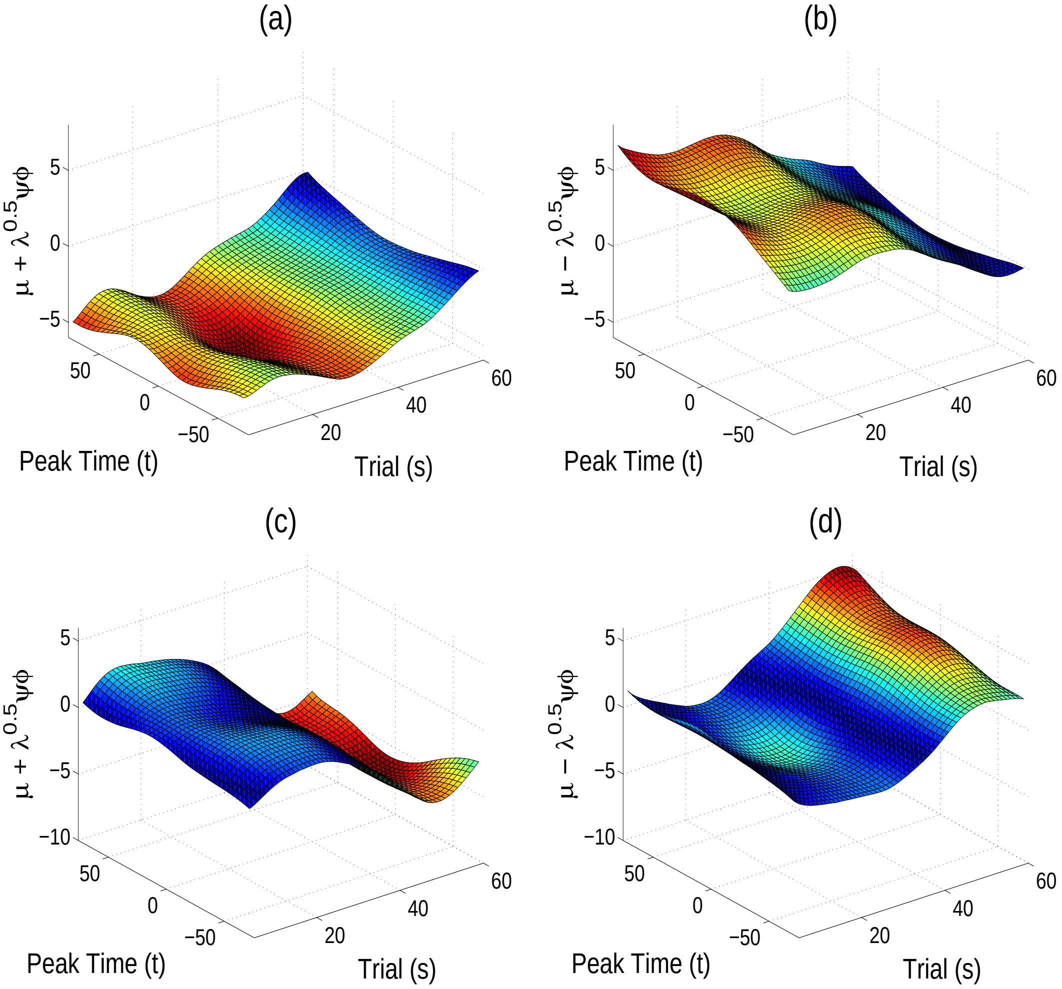


Figure 4.11: TD interval surfaces $\mu(t|s)\pm\sqrt{\lambda^{(1)}_{kk'}}\varphi^{(1)}_{kk'}(t,s)$ for $k=1, k'=1$ (top row) and $k=1, k'=2$ (bottom row). The left column corresponds to the added interval and the right column corresponds to the subtracted interval. According to the intervals, large positive and negative scores correspond to positive or negative peak differentiation.

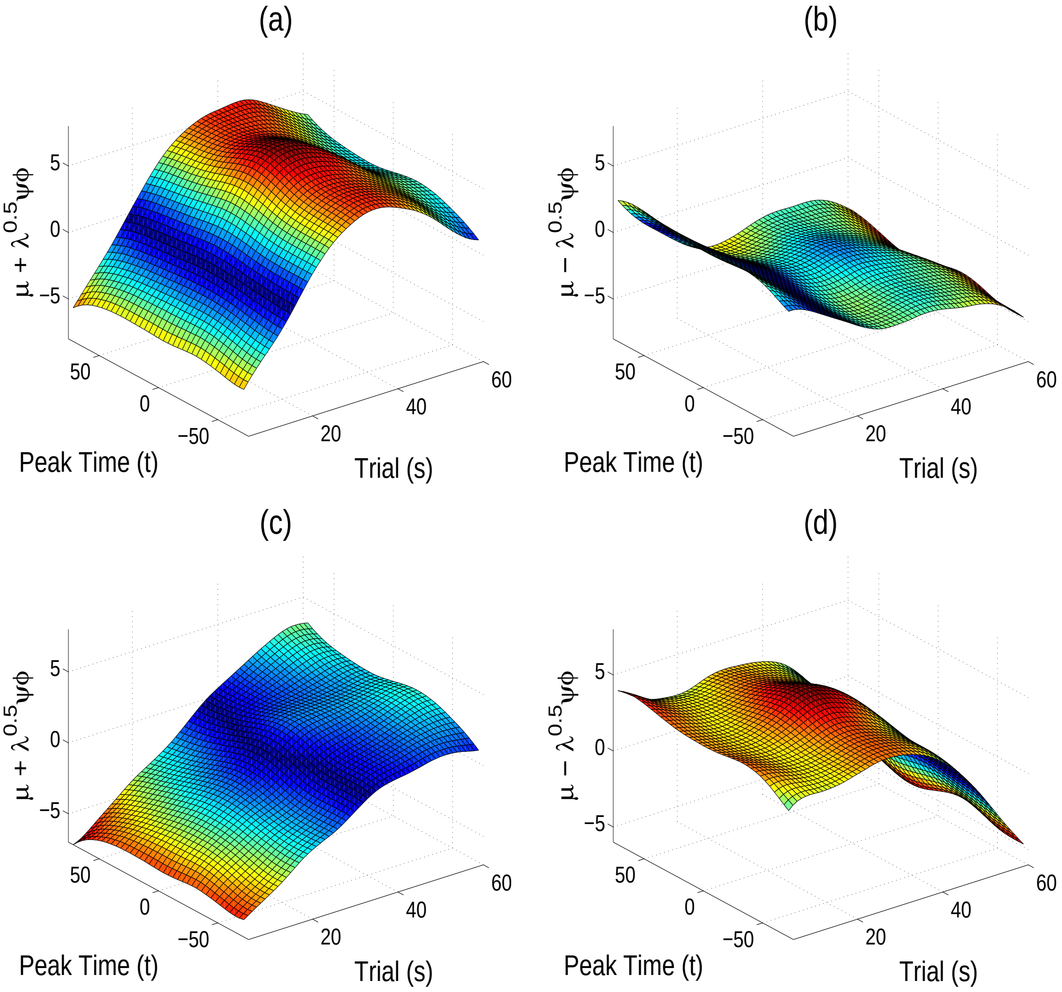


Figure 4.12: ASD interval surfaces $\mu(t|s) \pm \sqrt{\lambda^{(1)}_{kk'}}\varphi^{(1)}_{kk'}(t,s)$ for $k = 1, k' = 1$ (top row) and $k = 1, k' = 2$ (bottom row). The left column corresponds to the added interval and the right column corresponds to the subtracted interval. For $k = 1, k' = 1$, large positive scores correspond to a positive peak in differentiation across trials and large negative scores correspond to surfaces with little or no condition differentiation.

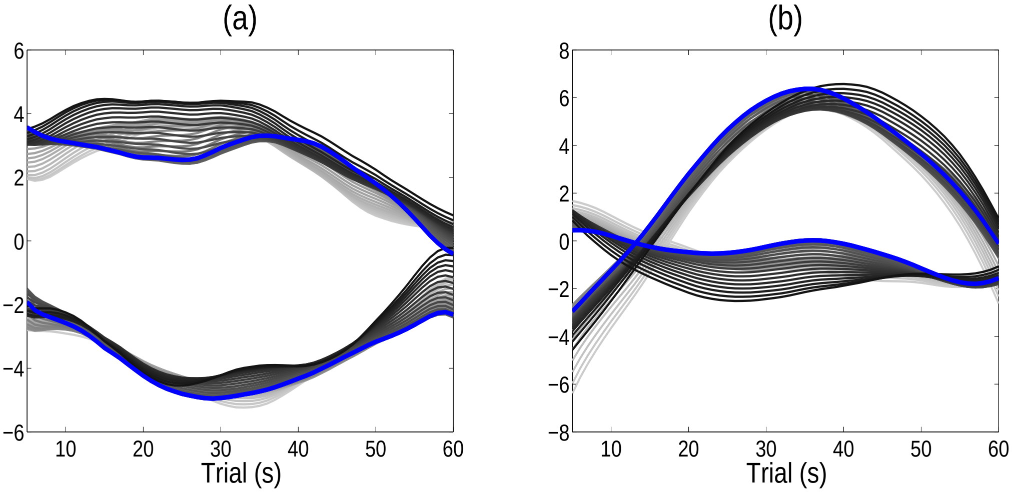


Figure 4.13: Mean surfaces of subjects with leading component scores $\xi'_{i11} > 0$ and $\xi'_{i11} \leq 0$ for each peak time plotted against the trial domain for the (a) TD and (b) ASD groups. A simple split between the leading scores shows evidence of clustering information contained within the leading components. Formal clustering procedures should be developed to further investigate methods for identifying these substructures.

correlation introduced in Section 4.2.3 was calculated for the TD group and the ASD group, indicating an average correlation between within-subject surfaces of 0.75 and 0.64 for the TD and ASD groups, respectively. A boostrapping procedure on the entire meta-preprocessing and MD-FPCA stages was performed using 200 runs to place distributions on these two statistics, producing the following 5% and 95% quantiles: TD (0.67, 0.82) and ASD (0.53, 0.71). Although the result is not significant, the proportion of total variability explained by the subject level across both the functional and longitudinal domains is larger for the TD group, similar to the multilevel correlations across s. However, the larger magnitude may be due to the larger separation of cluster substructures in the TD group, creating more variability at the subject level. Additional analyses within cluster substructures may be performed using MD-FPCA to verify this claim. Note the above analysis focuses primarily on the subject level components since our interest is in determining subject level differences in the condition differences of the entire ERP peak. Analysis of the level two components of the proposed MD-FPCA method is reserved as a future focus of analysis and interpretation when systematic variability in the subunits is of primary interest or is the dominant influence on the multilevel estimates.

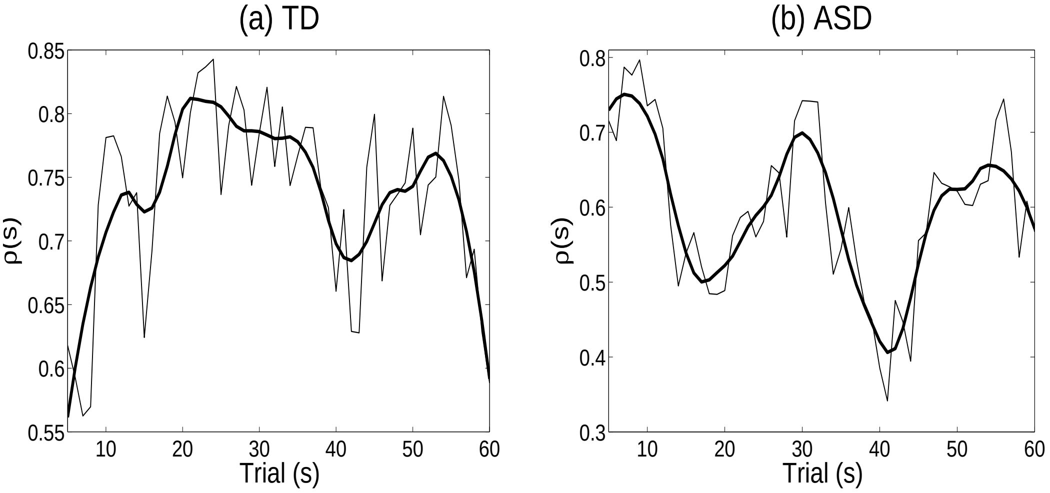


Figure 4.14: Stage one intraclass correlations $\rho(s)$ across trials for the (a) TD and (b) ASD groups. Both plots indicate that the majority of total variation is attributed to the subject-level variation in the ERP.

4.4 Simulation Study

Simulations are conducted to assess the performance of the proposed MD-FPCA in estimating the overall mean surfaces for each subunit and the overall eigensurfaces at each level of the multilevel structure across varying sample sizes. Data is generated using the originally estimated MD-FPCA model components in the data analysis for the TD group. The overall mean surface and the electrode-specific mean deviations from the TD application are used as true mean components for surfaces $\mu(t|s)$ and $\eta_j(t|s), j = 1, \ldots, 4$, respectively. The product of the stage one and stage two eigenfunctions at each level $\{\psi^{(1)}_{kk'}(s)\phi^{(1)}_{k}(t|s), \psi^{(2)}_{pp'}(s)\phi^{(2)}_{p}(t|s)\}$ from the TD analysis are used as the true overall eigensurfaces $\{\varphi^{(1)}_{kk'}(t, s), \varphi^{(2)}_{pp'}(t, s)\}$ using the first two components at each level $k, k' = 1, 2$ and $p, p' = 1, 2$. Corresponding level one scores $\xi'_{ikk'}$ and level two scores $\zeta'_{ijpp'}$ are sampled independently from the normal distribution with zero mean and variances {100000, 50000, 10000, 5000} for $\{\xi'_{i11}, \xi'_{i12}, \xi'_{i21}, \xi'_{i22}\}$ and variances {30000, 15000, 10000, 5000} for $\{\zeta'_{ij11}, \zeta'_{ij11}, \zeta'_{ij21}, \zeta'_{ij22}\}$, in order to achieve a signal-to-noise ratio similar to the observed surfaces of the TD group. Error variance is allowed to change across longitudinal time s and is selected from the error variance estimates from the first stage of the MD-FPCA TD application. Data is generated across the ERP time and trial domains of the working example with $J = 4$ subunits across sample sizes $n = 25, 50, 100$.

Simulated data are fit using the proposed MD-FPCA across 100 Monte Carlo runs. Accuracy of the model estimates are evaluated using the relative squared error of a functional longitudinal surface $f(t|s)$ and its estimate $\widehat{f}(t|s)$,

$$RSE = \frac{||f(t|s) - \widehat{f}(t|s)||^2}{||f(t|s)||^2},$$

where $||f(t|s)|| = \int_t \int_s f(t|s)^2 dsdt$ [93]. RSE is calculated for surface estimates $\{\widehat{\mu}(t|s) + \widehat{\eta}_j(t|s), \widehat{\varphi}^{(1)}_{kk'}(t,s), \widehat{\varphi}^{(2)}_{pp'}(t,s), \widehat{X}_{ij}(t|s)\}$. Reported RSE metrics for the predictions $\widehat{X}_{ij}(t|s)$ are averaged across subunits j, then across subjects i. Metrics for $\widehat{\mu}(t|s) + \widehat{\eta}_j(t|s)$ are averaged across subunits j. In addition, RSE metrics are calculated for predictions from the extended MD-FPCA denoted by $\widehat{X}^*_{ij}(t|s)$, which utilizes multilevel FPCA in the expansion of $\zeta_{ijp}(s)$, in order to examine the effect of the stage two within subject electrode independence assumption on model predictions.

The RSE metrics shown in Table 4.2 exhibit a decreasing trend with increasing sample size across the different surfaces as expected. RSE for the mean surfaces is somewhat large for $n = 25$ due to the small magnitude of the TD overall mean surface and the large signal-to-noise ratio of the data, however, estimates maintain the contour of the mean surface across simulation runs. Relative standard error of the eigensurface estimates is fairly small across samples but increases with decreases in component variation (components explaining less variability have larger RSE). Low RSE in the eigensurface estimates suggests the multiplicative decomposition at the subject and electrode levels accurately estimate the true overall eigensurfaces. Furthermore, MD-FPCA predictions for models containing the single-level or multilevel expansions of the $\zeta_{ijp}(s)$ trajectories are similar across sample sizes, suggesting the validity of the within subject electrode independence assumption on $\zeta_{ijp}(s)$.

4.5 Discussion

The proposed MD-FPCA approach is an effective method for analyzing and summarizing structurally complex data spanning several dimensions in low dimensional form.

Table 4.2: Percentiles {50%(10%,90%)} of relative squared error metrics for the surface estimates $\{\widehat{\mu}(t|s) + \widehat{\eta}_j(t|s), \widehat{\varphi}^{(1)}_{kk'}(t,s), \widehat{\varphi}^{(2)}_{pp'}(t,s), \widehat{X}_{ij}(t|s), \widehat{X}^*_{ij}(t|s)\}$. RSE exhibits a decreasing trend with increasing sample size across all estimates. Two-stage eigensurface estimates accurately estimate overall eigensurfaces at each level. Models $X_{ij}(t|s)$ and $X^*_{ij}(t|s)$ produce similar predictions.

	$n = 25$	$n = 50$	$n = 100$
$\mu + \eta_j$	1.842 (0.916, 3.547)	0.958 (0.351, 1.958)	0.461 (0.174, 1.165)
$\varphi^{(1)}_{11}$	0.046 (0.007, 0.300)	0.022 (0.005, 0.122)	0.009 (0.003, 0.053)
$\varphi^{(1)}_{12}$	0.054 (0.018, 0.300)	0.032 (0.008, 0.125)	0.014 (0.007, 0.056)
$\varphi^{(1)}_{21}$	0.244 (0.079, 0.725)	0.146 (0.061, 0.405)	0.108 (0.053, 0.233)
$\varphi^{(1)}_{22}$	0.348 (0.128, 0.758)	0.195 (0.091, 0.465)	0.154 (0.083, 0.280)
$\varphi^{(2)}_{11}$	0.039 (0.011, 1.909)	0.027 (0.006, 1.870)	0.009 (0.004, 1.071)
$\varphi^{(2)}_{12}$	0.074 (0.019, 1.905)	0.045 (0.017, 1.899)	0.027 (0.008, 1.055)
$\varphi^{(2)}_{21}$	0.264 (0.131, 2.267)	0.195 (0.108, 2.270)	0.164 (0.092, 1.743)
$\varphi^{(2)}_{22}$	0.211 (0.091, 2.197)	0.146 (0.066, 2.079)	0.120 (0.058, 2.078)
X_{ij}	0.150 (0.129, 0.185)	0.149 (0.130, 0.178)	0.148 (0.133, 0.160)
X^*_{ij}	0.148 (0.124, 0.188)	0.147 (0.125, 0.180)	0.144 (0.130, 0.157)

Application to the ERP from the implicit learning paradigm identified mean trends in condition differentiation of the entire P3 peak that is consistent with results from previous studies. MD-FPCA terms describing the within and between functional variations at the subject level supported the existence of the substructures of the TD and ASD groups identified by RFC in Chapter 3. MD-FPCA also provided insight into the dynamic structure of the functional peak difference over peak time across the longitudinal dimension of trials for both TD and ASD groups. Finally, the modeling of spatial covariation within and between ERP repetitions helped to describe the connectivity of electrodes relative to the variation encountered at the subject level.

Future studies should extend the consistency results and asymptotic convergence rates for the estimated model components of the double FPCA approach to the proposed MD-FPCA model. In addition, several extensions of the MD-FPCA method may be explored to improve upon its framework or to focus on different aspects of the ERP data. The analysis of the ERP in Section 4.3.2 did not include an in-depth analysis of the spatial dimension of ERP. Analysis of the level two components of the MD-FPCA model is reserved as a future focus when systematic variability in the subunits is the dominant influence on the multilevel estimates. Under the smoothness assumption, MD-FPCA may be altered to consider electrodes as an additional smooth dimension if the focus is on the spatial evolution of ERP signal in response to a sequence of stimuli. However, this would require that data be collected from a large electrode netting in order to provide sufficient continuity of the spatial information. This is feasible under these conditions due to the denseness of EEG observations across the functional domain. Further research on the clustering of repeated multilevel functional data should also be performed to identify subgroups in these data structures and to validate the evidence of the clusters visually identified in the level one components of the MD-FPCA results.

CHAPTER 5

Discussion and Future Work

This dissertation proposes three statistical methods (MAP-ERP, RFC, MD-FPCA) for analyzing and interpreting ERP across the functional, longitudinal and spatial dimensions that are typically collapsed or ignored in standard analyses. The meta-preprocessing step of the proposed MAP-ERP approach in Chapter 2 increases the signal-to-noise ratio of observed ERP in order to extract the magnitude and latency of ERP features observed across trial repetitions and electrodes, effectively preserving the longitudinal and spatial dimensions of the ERP. The proposed weighted linear mixed effects model allows for the visualization and analysis of the longitudinal feature processes across trials while incorporating feature information at the electrode level and correcting for heteroskedasticity induced by meta-preprocessing. MAP-ERP applied to the implicit learning paradigm found significant differences in implicit learning processes between the ASD subjects and TD controls and provided graphical evidence of their implicit learning trajectories. The RFC algorithm of Chapter 3 successfully identifies subgroups of feature trajectories while correcting for covariance heterogeneity induced by meta-preprocessing. RFC results from the implicit learning paradigm identified differences in substructures between the ASD and TD groups and provided a deeper understanding of the results observed in the MAP-ERP application. MD-FPCA incorporates an additional dimension, functional ERP time, into the analysis of ERP data. The MD-FPCA approach decomposes ERP across all three data dimensions into interpretable, low-dimensional modes of variation. Application to the implicit learning paradigm revealed the majority of variation in the trial direction, possibly due to the clusters present in the longitudinal dimension. However, decompositions in the functional direction provide valuable insight into the stability of the functional P3 peak condition difference based on the flat contour of the principal eigensurfaces across ERP time.

As evidenced by the wealth of results from the implementation of the proposed algorithms, utilizing ERP data across the three data dimensions can provide an in-depth understanding of the underlying mechanisms to certain diseases and disorders. Preservation of the functional dimension of ERP can offer additional measures of cognition in individuals with brain disorders. Classification of these ERP feature morphologies can provide neurological biomarkers that assist in their diagnoses. The preservation of longitudinal features of ERP is necessary when changes in electrical potential occur with dynamic sequences of stimuli over the course of an experiment. This is apparent in the implicit learning paradigm analyzed in Chapter 2, where speed of acquisition was found to contribute significantly to the characterization of implicit learning in young children with ASD compared to typically developing controls. Finally, incorporation of the spatial dimension can provide insight into within-subject electrode connectivity relative to between-subject covariation as discussed in Chapter 4. Data exploration of ERP across these three dimensions provides a complete picture of the cognitive processes of individuals with brain disorders; their combined analysis using the methods proposed provides a means of classification and analysis of these neural patterns by multiple attributes.

The entire focus of this dissertation has been on analyzing and interpreting information from ERP data using the proposed methods. Although meta-preprocessing has been specifically developed for the analysis of ERP data, the proposed methodology is applicable to studies involving repetitions of a systematic signal contaminated with noise, such as heart beat or breath cycle. The proposed weighted mixed effects model and the RFC algorithm may then be applied to the meta-preprocessed data if there are specific patterns of missingness producing heteroskedasticity or covariance heterogeneity. Standard modeling and clustering methods are feasible alternatives when these effects are not present. Unlike the MAP-ERP and RFC methods that have been specifically developed for the analysis of ERP data, the MD-FPCA method can be applied to any data containing repeated multilevel functional observations across some longitudinal domain.

Continuing research will involve several methodological extensions for taking advantage of the detailed information contained within the complex structure of EEG. MD-FPCA may be used to extend RFC to repeated multilevel functional data, where a major challenge is the interpretation of the clustered components. Clustering the scores of the second step of the MD-FPCA algorithm may also be investigated. RFC may be extended to model-based functional clustering algorithms using expectation-maximization, where covariance heterogeneity is accounted for using predefined random effects covariance parameterizations. In addition, future research may focus on the spatial distribution of ERP features rather than considering spatial effects as nuisance factors as in the proposed methods. The longitudinal component of the two-step FPCA algorithms can be replaced with the spatial coordinates of the EEG electrodes, providing spatial visualizations of the ERP features and summarizations of spatial modes of variation. Finally, future studies should consider larger sample sizes to allow for stable inferences to be made on group differences using the proposed MAP-ERP, RFC and MD-FPCA methods. Results from these studies, along with the results from these methodological extensions, offer a comprehensive summary of the most important and informative features of these multidimensional data structures.

References

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

[2] Tierney, A. L., Durnam, L. G., Farley, V. V., Flusberg, H. T. and Nelson, C. A. (2012). Developmental Trajectories of Resting EEG Power: An endophenotype of autism spectrum disorder. *PLoS One* **7**(6), e39127.

[3] Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L. and Corbetta M. (2007). Electrophysiological signatures of resting state networks in the human brain. *PNAS* **104**(32), 13170–13175.

[4] Quan, S. F., Howard, B. V., Iber, C., Kiley, J. P., Nieto, F. J., O'Connor, G. T., Rapoport, D. M., Redline, S., Robbins, J., Samet, J. M. and Wahl, P. W. (1997). The Sleep Heart Health Study: design, rationale, and methods. *Sleep* **20**(12), 1077–1085.

[5] Kim, D. J., Bolbecker, A. R., Howell, J., Rass, O., Sporns, O., Hetrick, W. P., Breier, A. and O'Donnell, B. F. (2013) Disturbed resting state EEG synchronization in bipolar disorder: A graph-theoretic analysis. *NeuroImage: Clinical* **2**, 414–423.

[6] Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y. and Sweeney, J. A. (2013) Resting state EEG abnormalities in autism spectrum disorders. *Journal of Neurodevelopmental Disorders* **5**(1), 24.

[7] Lee, S. H., Yoon, S., Kim, J. I., Jin, S. H. and Chung, C. K. (2014) Functional connectivity of resting state EEG and symptom severity in patients with post-traumatic stress disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **51**, 51–57.

[8] McEvoy, K., Hasenstab, K., Şentürk, D. and Jeste, S. (2014) Physiologic artifacts in the resting state EEG of young children: Methodological considerations for noisy data. *Biological Psychiatry*, In print.

[9] Jeste, S. S., DiStefano, C., Frohlich, J., Sanders, A., McEvoy, K., Hasenstab, K. and Şentürk, D. (2015) Resting state EEG oscillations predict cognition in preschoolers with ASD. In Preparation.

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

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

[12] Travers, B. G., Klinger, M. R., Mussey, J. L. and Klinger, L. G. (2010). Motor-linked implicit learning in persons with autism spectrum disorders. *Autism Research* **3**(2), 68–77.

[13] Hasenstab, K., Sugar, C., Telesca, D., Jeste, S., McEvoy, K. and Şentürk, D. (2015). Identifying longitudinal trends within EEG experiments. *Biometrics*, in press.

[14] Hasenstab, K., Sugar, C., Telesca, D., Jeste, S. and Şentürk, D. (2015) Robust functional clustering of longitudinal ERP trends. *Statistics in Medicine*, Submitted.

[15] Jeste, S. S. and Nelson, C. A. (2009). Event related potentials in the understanding of autism spectrum disorders: an analytical review. *Journal of Autism and Developmental Disorders* **39**(3), 495–510.

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

[17] Saavedra, C. and Bougrain, L. (2012). Processing Stages of Visual Stimuli and Event-Related Potentials. *The NeuroComp/KEOpS'2012 workshop* **2**, 1–5.

[18] Friedman, D. and Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research ad Technique* **51**(1), 6–28.

[19] van Veen, V. and Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology and Behavior* **77**(4-5), 477–482.

[20] Puce, A., Syngeniotis, A., Thompson, J. C., Abbott, D. F., Wheaton, K. J. and Castiello, U. (2003). The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. *NeuroImage* **19**(3), 861–869.

[21] Cabeza, R., Nyberg, L. and Park, D. (2005). *Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging.* New York: Oxford University Press.

[22] Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge: The MIT Press.

[23] Cohen, N. J. and Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science* **210**(4466), 207–210.

[24] Reber, P. J. and Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning and Memory* **1**(4), 217–229.

[25] Coppola, R., Tabor, R. and Buchsbaum, M. S. (1978). Signal to noise ratio and response variability measurements in single trial evoked potentials. *Electroencephalography and Clinical Neurophysiology* **44**(2), 214–222.

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

[27] Woestenburg, J. C., Verbaten, M. N. and Slangen, J. L. (1983). Stimulus information and habituation of the visual event related potential and the skin conductance reaction under task-relevance conditions. *Biological Psychology* **16**(3-4), 225–240.

[28] Mocks, J., Gasser, T. and Tuan, P. D. (1984). Variability of single visual evoked potentials evaluated by two new statistical tests. *Electroencephalography and Clinical Neurophysiology* **57**(6), 571–580.

[29] Mocks, J., Gasser, T., Pham, D. T. and Kohler, W. (1987). Trial-to-trial variability of single potentials: methodological concepts and results. *International Journal of Neuroscience* **33**(1-2), 25–32.

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

[31] Woestenburg, J. C., Verbaten, M. N., van Hees, H. H. and Slangen, J. L. (1983). Single trial ERP estimation in the frequency domain using orthogonal polynomial trend analysis (OPTA): estimation of individual habituation. *Biological Psychology* **17**(2-3), 173–191.

[32] Krieger, D., Berger, T. W. and Sclabassi, R. J. (1992). Instantaneous Characterization of Time-Varying Nonlinear Systems. *Transactions on Biomedical Engineering* **39**(4), 420–424.

[33] Krieger, D. and Sclabassi, R. J. (1994). Time-varying evoked potentials. *Journal of Medical Engineering and Technology* **18**(3), 96–100.

[34] Rossi, L., Bianchi, A. M., Merzagora, A., Gaggiani, A., Cerutti, S. and Bracchi, F. (2007). Single trial somatosensory evoked potential extraction with ARX filtering for a combined spinal cord intraoperative neuromonitoring technique. *Biomedical Engineering Online* **6**(2), 1–12.

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

[36] Verleger, R., Gasser, T. and Mocks, J. (1985). Short term changes of event related potentials during concept learning. *Biological Psychology* **20**(1), 1–16.

[37] Bruin, N. M. W. J., Ellenbroek, B. A., van Schaijk, W. J., Cools, A., Coenen, A. M. L. and van Luijtelaar, E. L. J. M. (2001). Sensory gating of auditory evoked potentials in rats: effects of repetitive stimulation and the interstimulus interval. *Biological Psychology* **55**(3), 195–213.

[38] Quiroga, R. Q. and van Luijtelaar, E. L. J. M. (2002). Habituation and sensitization in rat auditory evoked potentials: a single-trial analysis with wavelet denoising. *International Journal of Psychophysiology* **43**(2), 141–153.

[39] Mostofsky, S. H., Goldberg, M. C., Landa, R. J. and Denckla, M. B. (2000). Evidence for a deficit in procedural learning in children and adolescents with autism: implications for cerebellar contribution. *Journal of the International Neuropsychological Society* **6**(7), 752–759.

[40] Barnes, K. A., Howard, J. H., Howard, D. V., Gilotty, L., Kenworthy, L., Gaillard, W. D. and Vaidaya, C. J. (2008). Intact implicit learning of spatial context and temporal sequences in childhood autism spectrum disorder. *Neuropsychology* **22**(5), 563–570.

[41] Gidley, L. J. C. and Mostofsky, S. H. (2008). Evidence that the pattern of visuomotor sequence learning is altered in children with autism. *Autism Research* **1**(6), 341–353.

[42] Brown, J., Aczel, B., Jimenez, L., Kaufman, S. B. and Grant, K. P. (2010). Intact implicit learning in autism spectrum conditions. *Quarterly Journal of Experimental Psychology* **63**(9), 1789–1812.

[43] Stahl, D., Parise, E., Hoehl, S. and Striano T. (2010). Eye contact and emotional face processing in 6-month-old infants: Advanced statistical methods applied to event-related potentials. *Brain and Development* **32**(4), 305–317.

[44] Shi, M., Weiss, R. E. and Taylor, J. M. G. (1996). An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome using Flexible Random Curves. *Applied Statistics* **45**(2), 151–163.

[45] Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**(1), 253–259.

[46] Fiecas, M. and Ombao, H. (2015). Modeling the evolution of dynamic brain processes during an associative learning experiment. *Journal of the American Statistical Association,* submitted.

[47] Davidson, D. J. (2009). Functional mixed-effects models for electrophysiological responses. *Neurophysiology* **41**(1), 79–87.

[48] Srinivasan, R., Nunez, P. L., Tucker, D. M., Silberstein, R. B. and Cadusch, P. J. (1996). Spatial sampling and filtering of EEG with spline-Laplacians to estimate cortical potentials. *Brain Topography* **8**(4), 355–366.

[49] Yeung, N., Bogacz, R., Holroyd, C. B., Nieuwenhuis, S. and Cohen, J. D. (2007). Theta phase-resetting and the error-related . *Psychophysiology* **44**(1), 39–49.

[50] Quiroga, R. Q. and Garcia, H. (2003). Single trial event-related potentials with wavelet denoising. *Clinical Neurophysiology* **114**(2), 376–390.

[51] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders.* (5th ed.) Washington, DC.

[52] Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E. and Ruzzo, W. L. (2001). Model-based clustering and data transformations for gene expression data. *Bioinformatics* **17**, 977–987.

[53] Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**, 611–631.

[54] Li, J. (2005). Clustering based on multilayer mixture model. *Journal of Computational and Graphical Statistics* **14**, 547–568.

[55] Ma, P., Castillo-Davis, C. I., Zhong, W. and Liu, J. S. (2006). A data-driven clustering method for time course gene expression data. *Nucleic Acids Research* **34**, 1261–1269.

[56] Ma, P. and Zhong, W. (2008). Penalized clustering of large-scale functional data with multiple covariates. *Journal of the American Statistical Association* **103**, 625–636.

[57] Bouveyron, C. and Jacques, J. (2011). Model-based clustering of time series in group-specific functional subspaces. *Advances in Data Analysis and Classification* **5**, 281–300.

[58] Samé, A., Chamroukhi, F., Govaert, G. and Aknin, P. (2011). Model-based clustering and segmentation of time series with changes in regime. *Advances in Data Analysis and Classification* **5**, 301–321.

[59] Jacques, J. and Preda, C. (2013). Funclust: A curves clustering method using functional random variables density approximation. *Neurocomputing* **112**, 164–171.

[60] Luan, Y. and Li, H. (2003). Clustering of time-course gene expression data using a mixed-effects model with B-splines. *Bioinformatics* **19**, 474–482.

[61] James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association* **98**, 397–408.

[62] Yeung, K. Y. and Ruzzo, W. L. (2001). Principal component analysis for clustering gene expression data. *Bioinformatics* **17**, 763–774.

[63] Tarpey, T. and Kinateder, K. K. J. (2003). Clustering functional data. *Journal of Classification* **20**, 93–114.

[64] Abraham, C., Cornillon, P. A., Matzner-Løber, E. and Molinari, N. (2003). Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics* **30**, 581–595.

[65] García-Escudero, L. A. and Gordaliza, A. (2005). A proposal for robust curve clustering. *Journal of Classification* **22**, 185–201.

[66] Serban, N. and Wasserman, L. (2005). CATS: clustering after transformation and smoothing. *Journal of the American Statistical Association* **100**, 990–999.

[67] Ray, S. and Mallick, B. (2006). Functional clustering by Bayesian wavelet methods. *Journal of the Royal Statistical Society* **68**, 305–332.

[68] Kayano, M., Dozono, K. and Konishi, S. (2010). Functional cluster analysis via orthonormalized Gaussian basis expansions and its application. *Journal of Classification* **27**, 211–230.

[69] Delaigle, A., Hall, P. and Bathia, N. (2012). Componentwise classification and clustering of functional data. *Biometrika* **99**, 299–313.

[70] Gattone, S. A. and Rocci, R. (2012). Clustering curves on a reduced subspace. *Journal of Computational and Graphical Statistics* **21**, 361–379.

[71] Yamamoto, M. (2012). Clustering of functional data in a low-dimensional subspace. *Advances in Data Analysis and Classification* **6**, 219–247.

[72] Suarez, A. and Ghosal, S. (2013). Bayesian clustering of functional data using local features. *Bayesian Analysis* **1**, 1–15.

[73] Jacques, J. and Preda, C. (2014). Functional data clustering: a survey. *Advances in Data Analysis and Classification* **8**, 231–255.

[74] Serban, N. and Jiang, H. (2012). Multilevel functional clustering analysis. *Biometrics* **68**, 805–814.

[75] Chiou, J. and Li, P. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society* **69**, 679–699.

[76] Chiou, J. and Li, P. (2008). Correlation-based functional clustering via subspace projection. *Journal of the American Statistical Association* **103**, 1684–1692.

[77] Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society* **53**, 233–243.

[78] Yao, F., Müller, H. and Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

[79] Di, C., Crainiceanu, C. M., Caffo, B. S. and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics* **3**, 458–488.

[80] Di, C., Crainiceanu, C. M. and Jank, W. S. (2014). Multilevel sparse functional principal component analysis. *Stat* **3**, 126–143.

[81] Şentürk, D. and Nguyen, D. V. (2011). Varying coefficient models for sparse noise-contaminated longitudinal data. *Statistica Sinica* **21**, 1831–1856.

[82] Şentürk, D., Dalrymple, L. S., Mohammed, S. M., Kaysen, G. A. and Nguyen, D. V. (2013). Modeling time varying effects with generalized and unsynchronized longitudinal data. *Statistics in Medicine* **32**, 2971–2987.

[83] Liu, B. and Muller, H. G. (2008). *Functional Data Analysis for Sparse Auction Data.* Wiley: New York, 269–290.

[84] Li, P. and Chiou, J. (2011). Identifying cluster number for subspace projected functional data clustering. *Computational Statistics and Data Analysis* **55**, 2090–2103.

[85] Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**, 224–227.

[86] Pollard, D. (1981). Strong consistency of k-means clustering. *The Annals of Statistics* **9**, 135–140.

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

[88] Ramsay, J. O. (2005). *Functional Data Analysis*, Springer Series in Statistics, New York: Springer, 2nd ed.

[89] Morris, J. S., Vannucci, M., Brown, P. J. and Carroll, R. J. (2003). Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *Journal of the American Statistical Association* **98**, 573–597.

[90] Morris, J. S. and Carroll, R. J. (2006). Wavelet-based Functional Mixed Models. *Journal of the Royal Statistical Society* **68**, 179–199.

[91] Crainiceanu, C. M., Staicu, A.-M. and Di, C.-Z. (2009). Generalized Multilevel Functional Regression. *Journal of the American Statistical Association* **104**, 1550–1561.

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

[93] Muller, H. G. and Chen, K. (2012). Modeling Repeated Longitudinal Observations. *Journal of the American Statistical Association* **107**, 1599–1609.