

UCLA

UCLA Electronic Theses and Dissertations

Title

Analysis of High-Dimensional Functional Electroencephalography Data with Applications to Neurodevelopmental Disorders

Permalink

<https://escholarship.org/uc/item/16g1q0qd>

Author

O'Banion, Emilie

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Analysis of High-Dimensional Functional Electroencephalography Data with Applications
to Neurodevelopmental Disorders

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

by

Emilie Joyce O'Banion

2023

© Copyright by
Emilie Joyce O'Banion
2023

ABSTRACT OF THE DISSERTATION

Analysis of High-Dimensional Functional Electroencephalography Data with Applications
to Neurodevelopmental Disorders

by

Emilie Joyce O'Banion

Doctor of Philosophy in Biostatistics

University of California, Los Angeles, 2023

Professor Damla Şentürk, Chair

Due to the increasing availability of highly structured data, data analysis methods that are computationally efficient are necessary to study the dependencies between functional responses. The methods developed in this dissertation are motivated by a set of electroencephalography (EEG) studies in children with autism spectrum disorder (ASD).

The first chapter of this body of work focuses on finding a set of components across different conditions within multiple experimental tasks using principle ERP reduction (pERP-RED). These underlying components are referred to as principle ERPs (pERPs) and are found using principal component analysis and independent component analysis. The second chapter then provides a low-dimensional decomposition of the highly structured EEG data through multilevel hybrid principal component analysis (M-HPCA) by decomposing the total variation into between- and within-subjects variation using ideas from ANOVA and separating the variation further using vector and functional principal component analyses. Finally, the third chapter models associations between scalar outcomes, such as a diagnosis, with region-referenced longitudinal functional predictors, such as EEG data collected across multiple visits, by introducing region-referenced longitudinal functional generalized linear models (RRLF-GLM). Utilizing a tensor product of discrete and continuous bases, the regression coefficient is able to capture the relationship that changes to the EEG response

have on the outcome. Each of these three proposed methods can be used to understand the differences in brain responses between typically developing children and children with ASD.

The dissertation of Emilie Joyce O'Banion is approved.

Hilary Jeanne Aralis

Chad J. Hazlett

Donatello Telesca

Damla Şentürk, Committee Chair

University of California, Los Angeles

2023

This dissertation is dedicated to Timothy, Robert, and Renee O'Banion, who's support has made everything possible. I also cannot forget my mother, Christie Morlan, this time. This is the big one, mama.

Finally, I dedicate this dissertation to my daughter, Mollie O'Banion. May you always know that you can do anything and your mom will support you the whole way.

TABLE OF CONTENTS

List of Figures	x
List of Tables	xvii
Vita	xviii
1 Introduction	1
1.1 Electroencephalography	1
1.1.1 Event-related paradigms	2
1.1.2 Resting state paradigms	3
1.2 Functional data analysis	3
1.2.1 Functional principal components analysis	3
1.2.2 Scalar-on-function regression	4
1.3 Overview of proposed methods	5
1.4 Software and implementation	7
2 Principle ERP Reduction and Analysis: Estimating and using <i>principle ERP</i> waveforms underlying ERPs across tasks, subjects, and electrodes	8
2.1 Introduction	8
2.2 Methods	12
2.2.1 The pERP-RED Algorithm	12
2.2.2 The pERP-Space Analysis	18
2.3 Simulation	22
2.3.1 Simulation Design	22
2.3.2 Simulation Results	24

2.4	Applications	32
2.4.1	ASD Study	32
2.4.2	ADHD Study	38
2.5	Discussion	48
3	Multilevel hybrid principal components analysis for region-referenced multilevel functional EEG data	51
3.1	Introduction	51
3.2	Multilevel Hybrid Principal Components Analysis (M-HPCA)	55
3.2.1	M-HPCA decomposition	55
3.2.2	M-HPCA estimation algorithm	60
3.2.3	Mixed effects modeling and the computationally efficient MM algorithm	63
3.3	Applications	71
3.3.1	Audio odd-ball paradigm	71
3.3.2	Day-to-day test-retest reliability of PSD	77
3.4	Simulation Study	83
3.5	Discussion	85
4	Region-referenced longitudinal functional generalized linear model for EEG data	88
4.1	Introduction	88
4.2	Proposed region-referenced longitudinal functional GLM	92
4.2.1	Statistical framework and modeling	92
4.2.2	Estimation and inference	94
4.3	Data analysis	98

4.3.1	Data structure and methods	98
4.3.2	Data analysis results	99
4.4	Simulation	103
4.4.1	Data generation	103
4.4.2	Results	106
4.5	Discussion	107
5	Conclusion	112
A	pERPred	114
A.1	Supplementary data	114
A.2	ASD data collection and pre-processing	114
A.3	ADHD data collection and pre-processing	115
A.4	Online Application for Exploring Results	115
B	M-HPCA	119
B.1	Details of the M-HPCA estimation algorithm: Estimation of covariances and measurement error variance	119
B.2	MM algorithm details	120
B.2.1	Derivatives	120
B.2.2	Determinant	121
B.2.3	Likelihood Ratio Test for Weak Separability	122
B.3	Bootstrap tests for condition differentiation	123
B.3.1	Within-group condition differentiation	123
B.3.2	Between-group condition differentiation	124
B.4	Simulation design	125

B.5	Visual Assessment of Weak Separability in Data Applications	125
C	RR-GFLM	136
C.1	Simulation: Additional figures	136

LIST OF FIGURES

2.1 Illustration of the pERP-RED algorithm, described in text. First, at the within-subject level, correlation across electrodes is used to estimate a smaller number of “principle electrodes” for each subject, which we term “subject-region” scores. Next, these subject-region scores are combined together across subjects, then reduced to “principal subject-region” scores. These data are then reshaped to give separate columns for each principal subject-region score on each task/condition. The resulting waveforms are decomposed by ICA in order to arrive at the final pERPs. 14

2.2 (a) Boxplots of R_{pERP}^2 from regressing the true pERPs on the estimated pERPs in 100 simulation runs show that the pERP-RED algorithm is correctly identifying 5 as the true number of pERPs. (b) Boxplots of the R_{test}^2 obtained by regressing ERP records from the test set on the estimated pERPs, for both the high and low noise simulation cases. In both cases, the boxplots of R_{test}^2 identify 5 as the true number of pERPs as the value of R_{test}^2 levels off sharply at that value. . . . 25

2.3 (a) The true pERPs are regressed on the estimated pERPs from the high noise simulation. The larger the coefficient, the darker the color in the figure and the stronger the match between true and estimated pERPs. The first true pERP is loaded heavily onto the second estimated pERP. (b) The true pERPs are regressed on the estimated pERPs from the low noise simulation in this case. 26

2.4 The true (solid line) and estimated (dashed line) pERPs are matched based on the regression coefficients from regressing the true pERPs on the estimated pERPs in the low noise (top row) and high noise (bottom row) simulations. Recall that pERPs can be estimated only up to rotations (linear combinations) of those that generated the data. Accordingly, in some cases the true and estimated pair look almost identical, but in others the estimated pERPs appear as combinations of true pERPs. 27

2.5	The boxplots of R_{test}^2 as a function of the number of pERPs estimated across 100 simulations with 60 (first row) and 80 percent variation retained (second row) and low (first column), medium (second column) and high (third column) correlation between electrodes and tasks for $N = 25$ (red), 50 (green), and 100 (blue). The variability in prediction accuracy decreases with increasing sample size, as expected.	29
2.6	The boxplots of R_{test}^2 as a function of the number of pERPs estimated across 100 simulations with 60 (first row) and 80 percent variation retained (second row) and $N = 25$ (first column), 50 (second column), and 100 (third column) for low (red), medium (green) and high (blue) correlation between electrodes and tasks. With increasing correlation, we see increasing variability in prediction accuracy .	30
2.7	The boxplots of R_{test}^2 as a function of the number of pERPs estimated across 100 simulations with $N = 25$ (first row), 50 (second row), and 100 (third row) and low (first column), medium (second column) and high (third column) correlation between electrodes and tasks for 60 (red) and 80 (blue) percent variation retained. When the correlation is high, retaining more variation in PCA corresponds to retaining more noise, leading to worse prediction accuracy.	31
2.8	In both paradigms, the image appears for the full 2000ms. (a) In the audio paradigm, a spoken word is heard 1000ms after the onset of the image. (b) In the visual paradigm, the word image is seen 1000ms after the onset of the image.	34
2.9	(a) The observed average ERP waveform for the image condition, by group. No N1 is evident. (b) ERP reconstructed using only pERP2 , i.e. pERP 2 scaled by its contribution to the waveform (O1 plotted here). This component is detected despite the apparent lack of an N1 in the overall ERP because we also identify and model pERP 1, which captures spillover from the prior trial. The numbers shown are the mean weight with SE, and APSD in parentheses. The APSD reflects the subject variability and is much larger in the vASD group. Thus the vASD group has an average response very similar to that of the TD group, but there is a great deal more variability in the vASD group.	37

2.10	(a) ERP waveforms at F4. (b) A significant difference was found on pERP 9 between the mvASD and TD groups when contrasting match vs. mismatch. The numbers reported in the figure are the mean, with the SE and APSD in parentheses. This echoes a previously reported finding that the mismatch condition led to a deeper negativity at 700-800ms, thought to reflect semantic integration.	39
2.11	(a) In the SDRT task, a sequence of screens are presented. A fixation cross for 500ms, the cue dots for 2000ms, another fixation cross for 3000ms, and the probe dot for 3000ms. Within the 3000ms probe phase, the subject responds. (b) In the CPT task, the subject is presented with either an ‘X’ or ‘not X’ for 250ms, after which there is a blank screen for either 1000, 2000 or 4000ms before the next trial.	41
2.12	The N1/P2 as observed (solid black line) and as reconstructed from pERPs 4 (red line) and 5 (blue line) at CZ, at each of the five trials types. The loadings and standard errors for those loadings is shown in each figure. The corresponding t-statistics (mean divided by SE) range from 2.9 to 22 across these 10 cases, leading to firmly to rejecting the null hypothesis (of zero loading) in each.	43
2.13	(a) The observed ERP at Cz. The ‘X Correct’ and ‘Not X Incorrect’ tasks are used for this contrast because neither include a motor response. The large positivity in the ‘X Correct’ task is expected as it indicates the update/novelty of seeing a rare event. (b) The large positivity in the ‘X Correct’ task from 400-800ms can be decomposed into pERPs 7 (red), 8 (green), and 9 (blue). The numbers shown are the mean and SE of the coefficients.	45

3.1	(a) In the audio odd-ball paradigm, the image appears for the full 2000ms and a spoken word is heard 1000ms after the onset of the image. The functional observation to be considered is time-locked to the onset of the audio, creating an ERP that is from -100ms to 1000ms. (b) The ERP waveform displaying the characteristic components P200 and N400. (c) In the ABC-CT feasibility study, PSD are recorded on two visits, approximately a week apart. (d) The estimated mean relative PSD $\mu(t)$ as a function of frequency in Hz.	52
3.2	(a) The estimated overall mean function $\mu(t)$ in the audio odd-ball paradigm. (b) Positions of the 6 electrodes used in the analysis of the audio odd-ball paradigm highlighted. (c) Positions of the 18 electrodes used in the analysis of the ABC-CT feasibility study data highlighted.	72
3.3	The estimated leading participant-level eigenvectors $\nu_{d1}^{(1)}(r)$ (given in (a)) and condition-level eigenvector $\nu_{d1}^{(2)}$ (given in (b)) represent a contrast between electrodes from the anterior and posterior regions of the scalp for all groups. The estimated leading two eigenfunctions for the participant-level variation $\phi_{d\ell}^{(1)}$ (middle row) and condition-level variation $\phi_{dm}^{(2)}$ (bottom row) for each diagnostic group.	75
3.4	The estimated group, region and condition-specific mean functions $\eta_{aj}(r, t)$ in the audio odd-ball paradigm.	78
3.5	The estimated leading (a) participant-level eigenvector $\nu_{d1}^{(1)}(r)$ and (b) day-level eigenvector $\nu_{d1}^{(2)}(r)$. The estimated leading two eigenfunctions for the participant-level variation $\phi_{d\ell}^{(1)}$ (middle row) and day-level variation $\phi_{dm}^{(2)}$ (bottom row) for each diagnostic group.	81
4.1	(a) Group-specific mean alpha band log spectral density from the right parietal region. Darker lines correspond to an increase in age. (b) Positions of the 7 electrodes used in the analysis of the first year of life data, highlighted based on the region in which they are averaged.	89

4.2	Estimated mean function $\eta(r, s, t)$ for the (a) parietal left, (b) parietal, and (c) parietal right regions, where darker lines correspond to an increase in age. . . .	100
4.3	(a) The average mean-centered functional predictor for TD (grey) and ASD (black) children. (b) The estimated regression function $\beta(r, s, t)$. (c) The point-wise product of the mean-centered functional predictor and the estimated regression function, representing the average area under the curve for TD (grey) and ASD (black) children.	101
4.4	The predicted probabilities of ASD and their 95% confidence intervals for the study participants, where true group membership is denoted by grey for TD and black for ASD.	102
4.5	The RSE of the regression function across 200 Monte Carlo runs for the discrete (black) and continuous (grey) longitudinal dimension cases.	108
4.6	The coverage probabilities of the regression function across 200 Monte Carlo runs for the discrete (black) and continuous (grey) longitudinal dimension cases. . . .	109
4.7	The out-of-sample AUC across 200 Monte Carlo runs for the (a) RRLF-GLM and (b) m-GFLM models for the discrete (black) and continuous (grey) longitudinal dimension cases.	110
A.4.1	In the ASD Data Browser, there is a panel that displays the estimated pERPs as well as the plot of the R_{test}^2 used to determine the number of pERPs.	116
A.4.2	Each individual observed ERP can be reconstructed using the weights calculated using regression.	117
A.4.3	Headmaps can be created in order to visualize where the pERPs are being seen across the scalp. In this case, it can be seen that pERPs 1, 2, 6, and 7 are loaded onto heavily for the Image task.	118

B.5.1(a) The true and estimated overall mean functions $\mu(t)$ corresponding to the 10th, 50th and 90th percentile RSE values across groups based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15$ and high noise. (b, c) The true and estimated $\eta_{dj}(r, t)$ corresponding to the 10th, 50th and 90th percentile RSE values across groups $d = 1, 2$ and regions based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15$ and high noise. 128

B.5.2 The true and estimated first (first column) and second (second column) leading marginal eigenvectors for level 1 $v_{dk}^{(1)}(r)$ (first row) and level 2 $v_{dp}^{(2)}(r)$ (second row) and eigenfunctions $\phi_{dl}^{(1)}(t)$ (third row) and $\phi_{dm}^{(2)}(t)$ (fourth row) corresponding to the 10th, 50th and 90th percentile RSE values across groups based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15$ and high noise. 129

B.5.3(a, b) The estimated power of the proposed bootstrap test for condition differentiation based on 200 Monte Carlo runs. (c) Coverage and (d) length of the bootstrap CIs proposed for M-HPCA ICC based on 200 Monte Carlo runs. (e) The median run time for the proposed MM algorithm (green line) versus the popular R package `lme4` (red line) from the dense (solid line) and sparse (dashed line) simulation design at $n_d = 15$ and low noise, as a function of the total number of product components. 130

B.5.4(a, c, e) The estimated leading participant-level marginal eigenfunction $\phi_{d1}^{(1)}(t)$ (solid) compared to the leading participant-level eigenfunctions from the C4, CZ, and P3 regions. (b, d, f) The estimated second leading participant-level marginal eigenfunction $\phi_{d2}^{(1)}(t)$ (solid) compared to the second leading participant-level eigenfunctions from the C4, CZ, and P3 regions. 131

B.5.5(a, c, e) The estimated leading condition-level marginal eigenfunction $\phi_{d1}^{(2)}(t)$ (solid) compared to the leading condition-level eigenfunctions from the C4, CZ, and P3 regions. (b, d, f) The estimated second leading condition-level marginal eigenfunction $\phi_{d2}^{(2)}(t)$ (solid) compared to the second leading condition-level eigenfunctions from the C4, CZ, and P3 regions. 132

B.5.6(a, c, e) The estimated leading participant-level marginal eigenvectors $v_{d1}^{(1)}(r)$ (bottom row) compared to the leading participant-level eigenvectors obtained at fixed slices of $t = 0, 200, 400, 600, 800$. (b, d, f) The estimated leading condition-level marginal eigenvectors $v_{d1}^{(2)}(r)$ (bottom row) compared to the leading conditional-level eigenvectors obtained at fixed slices of $t = 0, 200, 400, 600, 800$	133
B.5.7The estimated leading participant-level (a, c) and second leading participant-level (b, d) marginal eigenfunctions $\phi_{d1}^{(1)}(t), \phi_{d2}^{(1)}(t)$ (solid) compared to the participant-level eigenfunctions from the Fz, C3, and Pz regions. The estimated leading day-level (e, g) and second leading day-level (f, h) marginal eigenfunctions $\phi_{d1}^{(2)}(t), \phi_{d2}^{(2)}(t)$ (solid) compared to the leading day-level eigenfunctions from the Fz, C3, and Pz regions.	134
B.5.8(a, c) The estimated leading participant-level marginal eigenvector $v_{d1}^{(1)}(r)$ (bottom row) compared to the leading participant-level eigenvectors obtained at fixed slices of $t = 4, 6, 9, 13, 30$. (b, d) The estimated leading day-level marginal eigenvector $v_{d1}^{(2)}(r)$ (bottom row) compared to the leading day-level eigenvectors obtained at fixed slices of $t = 4, 6, 9, 13, 30$	135
C.1.1The rank across 200 Monte Carlo runs for the discrete (black) and continuous (grey) longitudinal dimension cases.	137
C.1.2The squared error across 200 Monte Carlo runs for the discrete (black) and continuous (grey) longitudinal dimension cases.	138
C.1.3The estimated regression function with median RSE (dashed) and true regression function (solid) for (a) $N = 70$, (b) $N = 200$, (c) $N = 500$, and (d) $N = 1000$	139
C.1.4The estimated regression function with median RSE (first through fourth columns) and true regression function (last column) for all simulation settings.	140
C.1.5The out-of-sample accuracy across 200 Monte Carlo runs for the (a) RRLF-GLM and (b) m-GFLM models for the discrete (black) and continuous (grey) longitudinal dimension cases.	141

LIST OF TABLES

2.1	We consider 5 separate trial types/ tasks drawn from the two ASD experiments (Audio and Visual). From the two ADHD experiments (SDRT and CPT), we consider 8 separate trial types/ tasks.	34
2.2	The APSD captures variability across participants in the weight given to each pERP. We see that the ADHD group has a much higher APSD for many of the pERPs than the TD group.	47
3.1	The 50% (10%, 90%) percentiles of the relative squared errors, normalized mean squared errors based on 200 Monte Carlo runs from the simulation design at $n_d = 15, 50$ with low and high noise for the sparse set-up.	84
B.1	Percentiles 50% (10%, 90%) of the relative squared errors, normalized mean squared errors, and both total and marginal FVE across groups for model components based on 200 Monte Carlo runs from the simulation design at $n_d = 15, 50$ for the dense simulation.	126

VITA

- 2016 B.S. (Applied Mathematics and Statistics), California State Polytechnic
University, Pomona
- 2018 - present Graduate Student Researcher, Department of Biostatistics, University of
California, Los Angeles
- 2019 M.S. (Biostatistics), University of California, Los Angeles

CHAPTER 1

Introduction

The aim of this dissertation is to provide methods to study the heterogeneity in electroencephalography (EEG) data collected from children with neurodevelopmental disorders such as autism spectrum disorder (ASD) or attention deficit hyperactivity disorder (ADHD). For context, a brief review of essential topics is provided: electroencephalography, functional principal component analysis (FPCA), and functional regression. The review is followed by a summary of the main contributions: proposed methods, contributing publications, and available software.

1.1 Electroencephalography

An electroencephalogram (EEG) is a way to measure electrical activity in the brain using electrodes attached to the scalp and was first demonstrated in 1929 by Hans Berger (Luck, 2014). Over the decades, it has been shown that EEG can be used in scientific and clinical settings, but there are still a number of challenges in analyzing EEG data. The signal-to-noise ratio is often poor which leads researchers to find ways of amplifying the signal, which usually means collapsing an important dimension of the data. In addition to the signal-to-noise ratio, the observed EEG data is often a combination of many different sources of activity that make it difficult to isolate neuro-cognitive processes.

Statistical methods must also consider the high-dimensional structure of this data. EEG signals are recorded continuously throughout the experiment time at multiple electrodes across the scalp and the signal is split into segments. Therefore each observation of functional data has a regional and longitudinal dimension. The set of EEG signals are typically

collapsed by either averaging across regions, longitudinal timepoints, or both. Even further, researchers may reduce the signal into scalar features such as peak amplitude within a user-specified window in the functional domain. The methods developed as part of this dissertation circumvent the need to use scalar features and aim to utilize each of the dimensions to investigate the dominant modes of variation in the observed data and how they differ between diagnostic groups. The following describes the two paradigms of EEG experiments that will be used to demonstrate the methods developed in this dissertation.

1.1.1 Event-related paradigms

EEG recordings time-locked to an event of interest, such as a stimulus or participant response, produce an event-related potential (ERP) waveform. Experimental designs in this case utilize contrasting task conditions in order to investigate task-related changes to the ERP waveform. Each presentation of the stimulus or participant response generates an ERP that represents the functional observation. The contrasting task conditions create a multilevel dimension. Finally, the data is recorded at each electrode, creating a regional dimension. The goal of ERP experiments is to study the expected components, which are characterized by peaks or positivities (P) and valleys or negativities (N). These components are named for the direction (P or N) and the time at which they occur, e.g. the P300 is a positivity at 300 ms.

Experiments are designed to elicit a specific set of characteristic ERP components based on the cognitive processes they are thought to reflect (Luck, 2014). For example, a simple oddball paradigm is described in Luck (2014) in which participants were presented an ‘X’ on the screen 80% of the trials and an ‘O’ the other 20% of the trials. They were asked to press one button for ‘X’ and a different button for ‘O’. ERPs were averaged across trials within conditions and there was a set of components that were consistently seen in these averages. Namely, the P100 is connected to the initial visual response and the P300 is thought to represent context updating. The P300 has been shown to have a larger amplitude when the probability of the event is lower (unexpected). These types of experiments can be useful in studying the differences in brain responses between children with ASD and their typically

developing (TD) peers.

1.1.2 Resting state paradigms

Resting-state paradigms differ from event-related paradigms in that they are not time-locked to an event of interest. These experiments do not require an outward response and do not have a repeated stimulus. For example, EEG signals may be recorded over several minutes while participants gaze at a screen with floating bubbles while relaxed. The EEG signals are segmented into 1-3 second pieces and transformed into the frequency domain using Fast Fourier Transform (FFT). The resulting power spectral density is considered the until of observation and represents functional data. The goal of resting-state paradigms is to understand baseline spectral characteristics. In the studies considered within this dissertation, participants are children observed during their first year of life and the development of the spectral density over time is the focus since it can reveal early neurodevelopmental disruptions.

1.2 Functional data analysis

Each ERP or PSD can be considered samples from an underlying functional process and methodology from functional data analysis (FDA) has been widely developed, see Ramsay & Silverman (2005) for a full review. It is assumed that the units of observation are smooth curves $X_1(t), \dots, X_n(t)$ across a continuous domain \mathcal{T} . However, recently effort has been made in extending methods for one-dimensional functional data to accommodate functional data observed repeatedly for multiple regions or longitudinal timepoints, as described in the previous section.

1.2.1 Functional principal components analysis

Functional principal components analysis (FPCA) is a dimension reduction tool for functional data analogous to principal components analysis (PCA) for scalar multivariate data

(Wang et al., 2016). Similar to PCA, the goal of FPCA is to find the dominant modes of variation in the functional observations. The added bonus of FPCA is the ability to convert infinite-dimensional functional data into a finite set of principal component scores.

Let $X_i(t), i = 1, \dots, n$, denote the sample of smooth functions defined on the continuous interval $t \in \mathcal{T}$, then denote the mean function $\mu(t) = E\{X_i(t)\}$ and covariance function $\Sigma(t, t') = \text{cov}\{X_i(t), X_i(t')\}$. Mercer's theorem (Mercer, 1909) gives the spectral decomposition

$$\Sigma(t, t') = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(t'),$$

where $\lambda_k, k = 1, \dots, \infty$ denote the descending real-valued non-negative eigenvalues and the set of $\phi_k(t), k = 1, \dots, \infty$ denote the orthogonal eigenfunctions. Karhunen (1946) and Loeve (1946) independently introduced the Karhunen-Loève (KL) expansion

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

where $\xi_{ik} = \int_{\mathcal{T}} \{X_i(t) - \mu(t)\} \phi_k(t) dt$ denote the the functional principal component scores. The scores are uncorrelated across k , $E(\xi_{ik}) = 0$, and $\text{var}(\xi_{ik}) = \lambda_k$. Practically, the number of components K is truncated to a finite number and is chosen to balance bias (which shrinks with increasing K) and variance (which grows with increasing K).

As previously mentioned, FDA and FPCA literature has grown to accommodate functional data with complex dependency structures, e.g. multilevel functional data (Crainiceanu et al., 2009; Di et al., 2009; Kundu et al., 2016; Morris & Carroll, 2006; Morris et al., 2003), longitudinal functional data (Greven et al., 2010; Park & Staicu, 2015), and region-referenced longitudinal functional data (Scheffler et al., 2020).

1.2.2 Scalar-on-function regression

While there are three main categories of functional regression, based on the type of response and predictor: (1) scalar-on-function, (2) function-on-scalar, and (3) function-on-function.

The model proposed within this dissertation falls into the first category so the remainder of this section will focus solely on scalar-on-function regression, although Morris (2015) provides a comprehensive review of all three categories.

Let y_i denote the scalar response and $X_i(t)$ denote the functional predictor, $i = 1, \dots, n$. Scalar-on-function regression for Gaussian responses was first introduced in Hastie & Mallows (1993) with the model

$$y_i = \beta_0 + \int_{\mathcal{T}} X_i(t)\beta(t)dt + \epsilon_i,$$

where β_0 is an intercept term, $\beta(t)$ is the regression function, and ϵ_i is the mean-zero normally-distributed residual error with variance σ^2 . Marx & Eilers (1999) introduced the generalized functional linear model for non-Gaussian responses

$$g\{E(y_i)\} = \beta_0 + \int_{\mathcal{T}} X_i(t)\beta(t)dt + \epsilon_i,$$

where $g(\cdot)$ is an invertible link function. Methodological development has been made in the choice of basis functions and regularization approach. There have also been efforts made in addressing multivariate (Gertheiss et al., 2013) and multidimensional (Reiss & Ogden, 2010; Staicu et al., 2020) functional predictors, but none of them include the three dimensions in EEG data described above. However, this dissertation aims to develop a method that accommodates region-referenced longitudinal functional data in its entirety.

1.3 Overview of proposed methods

Three statistical methods have been developed in order to model the high-dimensional data produced by EEG experiments that avoid collapsing dimensions whenever possible. Along with these methods, computationally efficient software packages have been created to make analysis easily accessible. The methods that have been published are noted below.

Chapter 2 develops principle ERP reduction (pERP-RED) and analysis to amplify the ERP components found across different conditions within multiple experimental tasks. Standard

analysis of this ERP data would require finding the peak or mean amplitude within a user-specified window in which an ERP component is expected to peak. The proposed pERP-RED algorithm characterizes the entire set of contributing ERPs using a linear combination of underlying waveforms that span the entire ERP, called principle ERPs (pERPs). Using two PCA-based noise reduction steps and an ICA step for blind source separation, pERP-RED is able to accommodate multiple experimental tasks with contrasting conditions. The pERP-RED algorithm is accompanied by a set of analytical tools that allow researchers to examine their ERP experiments in terms of the pERPs found using pERP-RED. Rather than analyzing a pre-determined set of ERP components based on previous literature, pERP-RED facilitates the discovery of unexpected differences.

E. Campos †, C. Hazlett †, P. Tan, H. Truong, S. Loo, C. DiStefano, S. Jeste, D. Şentürk (2020). Principle ERP reduction and analysis: Estimating and using principle ERP waveforms underlying ERPs across tasks, subjects and electrodes. *NeuroImage* 212, 116630. († Authors contributed equally)
Drafted by Emilie Campos, Chad Hazlett, and Damla Şentürk

Chapter 3 develops multilevel hybrid principal component analysis (M-HPCA) to model region-referenced multilevel functional data with a low-dimensional decomposition by decomposing the total variation into between- and within-subjects variation and using weak separability to further separate the variation into regional and functional directions separately. The resulting model components are estimated in a mixed effects modeling framework via a computationally efficient minorization-maximization algorithm coupled with bootstrap. Inference is performed using a parametric bootstrap approach that facilitates group-level inference as well as within-group inference. M-HPCA is showcased with two studies of individuals with autism. While ERP responses to match vs mismatch conditions are compared in an audio odd-ball paradigm in the first study, short-term reliability of the PSD across visits is compared in the second.

E. Campos, A.W. Scheffler, D. Telesca, C. Sugar, C. DiStefano, S. Jeste, A.R. Levin, A. Naples, S.J. Webb, F. Shic, G. Dawson, S. Faja, J.C. McPartland, D. Şentürk (2022). Multilevel hybrid principal components analysis for region-referenced functional electroencephalography data. *Statistics in Medicine* 41(19),

3737-3757.

Drafted by Emilie Campos and Damla Şentürk

Chapter 4 introduces region-referenced longitudinal functional generalized linear model (RRLF-GLM) to relate a scalar outcome to a region-referenced longitudinal functional predictor, such as EEG data collected at multiple visits. Rather than collapsing the PSD by including the peak alpha frequency or mean power from the alpha band in a linear mixed effects model, the entire PSD morphology is used as a functional predictor of diagnostic status. RRLF-GLM uses a tensor product of discrete and continuous basis functions for the regression coefficient and is able to accommodate either a discrete or continuous longitudinal dimension. Since the alpha band dynamics are known to change over time, it was imperative to create a model that took into account the longitudinal changes that may differ between children with ASD and TD children.

1.4 Software and implementation

Code for the proposed estimation and inference procedures in each chapter are publicly available along with their tutorials for implementation at <https://github.com/emjcampos>.

CHAPTER 2

Principle ERP Reduction and Analysis: Estimating and using *principle ERP* waveforms underlying ERPs across tasks, subjects, and electrodes

2.1 Introduction

Electroencephalography (EEG) captures changes in voltage measured at the scalp, reflecting the firing of many similarly oriented neurons (Luck & Kappenman, 2011). When EEG recordings are time-locked to an event of interest, such as the onset of a trial or a participant’s response, event-related potential (ERP) waveforms are obtained. Through experimental designs utilizing contrasting task conditions, task-related changes to the ERP waveform are scrutinized to identify the “ERP components” that are thought to reflect particular brain processes. Subsequent studies – both in research and clinical settings – then seek to evoke and measure these ERP components in an effort to index the cognitive processes thought to generate them.

The goal of most ERP analyses, as described in the foundational text of Luck (2014), is “to provide an accurate measurement of the size and timing of the underlying ERP component with minimal distortion from noise and from other spatially and temporally overlapping components” (p285). Despite a number of approaches that seek to reveal underlying components of ERP waveforms as we do here, the standard approach for measuring ERP components is to *take the average or peak amplitude over an investigator-selected time window*, when the target ERP is *expected to peak*. This fails to remove the distortions from overlapping components, and the equating of components with peaks is deeply problematic.

Concretely, when an ERP component is defined and named, the functional process it reflects may be implied by virtue of the task and the contrast used to isolate it. However, in any subsequent study with different tasks and different participants, it can be problematic to attribute a peak that is observed at a similar time — or the potentiation/attenuation in it — to the same component, with the same functional meaning that was described in the previous studies. This is because a range of ongoing components may be differentially triggered by the different task conditions or participants. The overlapping activity of these unknown components can make the amplitude in a targeted interval higher or lower, and can push observed peaks/valleys in the waveform to occur either earlier or later in time. Moreover, even components typically associated with large peaks (such as the visual N1) may produce no observable peak as a result of such overlap.

These serious consequences of the overlapping nature of ERP components are widely recognized, including in the standard texts of Luck (2014) and Luck & Kappenman (2011). A less recognized problem with the common practice of taking an average or peak amplitude over a time window relates to the entire analytical tradition of targeting and reporting results for just one or several “components” from an ERP study. Such a selective approach to analyzing ERP experiments is understandable under the standard approach, where a reliance on selected intervals complicates efforts to characterize an entire waveform or to measure the magnitude of each component in a contributing set of waveforms, as we propose here. However focusing on so few features of the waveform limits what we can rigorously learn from our experiments, and slows the discovery of differences in components that may be of research or clinical value, all while amplifying selective-reporting concerns due to user-selection of these intervals.

In order to provide an alternative to the standard practice of using windowed peak or mean amplitude to define and measure ERP components, multiple approaches have attempted to decompose trial-wise or grand-averaged ERP waveforms into underlying sources which span the entire epoch. These approaches combine Principle Components Analysis (PCA) and Independent Component Analysis (ICA) techniques. Earlier works, beginning

with Makeig et al. (1995) used ICA to carry out subject-level decompositions of the signal. Such subject-level decompositions require additional steps, generally relying on clustering techniques, in order to connect components extracted from one subject with those of others. This post-estimation clustering requirement is avoided by “multi-subject” decomposition approaches for ERP. The multi-subject decomposition of Spencer et al. (2001), called spatiotemporal PCA, involves two PCA steps: a “spatial” one that reduces the electrode dimension through PCA on data concatenated across subjects, and a “temporal PCA” to reduce the dimension in the temporal domain. Other multi-subject approaches, recently reviewed by Huster & Raud (2018), combine both PCA and ICA approaches for noise reduction and source separation, respectively. Multi-level Group ICA (mlGICA) of Eichele et al. (2011) and temporal-concatenation group ICA (tcGICA) of Cong et al. (2013) both consider trial-wise ERP and conduct one or two PCA steps for dimension reduction at the electrode or subject levels, followed by a final ICA step for source separation.

To introduce our approach, we begin with the core assumption: the observed ERP for any subject in any task/experiment and condition/trial-type, and at any electrode, is approximated by a linear combination of underlying waveforms spanning the entire epoch, called *principle ERPs* (pERPs). We propose the pERP-reduction (pERP-RED) algorithm to estimate these pERPs. The pERP-RED algorithm, similar in spirit to other multi-subject approaches, begins with two PCA-based noise reduction steps. The first to construct “regions” similar to the “virtual electrodes” of Spencer et al. (2001), but separately for each subject; the second to reduce the subject-region-task specific ERP averages into a smaller set that explains most of the sample variation. This is followed by an ICA step for blind source separation. However, pERP-RED has important differences from prior proposals. First, we emphasize that pERP-RED is designed for reducing data not only across multiple subjects but also across multiple tasks. Though this is not explicitly prevented by other methods, it is the intended use-case only for pERP-RED. A wider range of tasks/experiments stimulates a wider range of neural processes, each with their own signatures that get mixed into the observed signal, which we seek to separate out and measure. The multi-task approach of

pERP-RED takes advantage of this rich form of data and estimates a set of pERPs with greater generalizability. Another useful benefit of the multi-experiment setting, in this early validation stage, is that it allows a test of whether different tasks, trial types or conditions are built up from shared components in the expected ways. Second, the electrode dimension reduction PCA steps of spatial PCA, mlGICA and tcGICA, all performed on matrices concatenated over subjects or trials, assume no missingness on electrodes, identical trial orderings, identical scalp topographies, or identical projections of components onto electrodes across subjects. The electrode reduction PCA step of pERP-RED, by comparison, avoids these assumptions by running dimension reduction separately for each subject in the first step. Subject-region-task specific ERP averages are only combined at the second PCA noise reduction step, which does not assume any homogeneity over the subject-specific components concatenated from the first step.

However, more important than these algorithmic differences, our approach is unique in arguing for a new way to analyze one’s data using these pERPs, and the set of tools we develop and make available to do so. Despite existing algorithms for estimating underlying components, the standard practice of using windowed peak or mean amplitudes continues to dominate in most published ERP studies. We argue that much more can and should be done to analyze one’s ERP waveforms using these underlying components than has been the case in any prior work. Our primary objective is to provide researchers with a set of analytical tools for examining their ERP experiments in terms of these components, which we call “pERP-space analysis”. That is, rather than attempting to measure a single component’s magnitude, coefficients describing the contribution of *all* pERPs in any given ERP waveform can be estimated and reported. These estimated pERP coefficients are directly interpretable as the magnitude of each pERP’s contribution to the observed waveform, and are equipped with standard errors. This allows the practitioner to analyze and report both the presence/absence and the strength of each pERP in each condition, group or scalp region, and to make inferences on condition or group contrasts. Additionally, we provide tools to track across-subject heterogeneity of the magnitude of each pERP’s contribution (e.g. het-

erogeneity in the pERPs, for anxious vs. non-anxious youth). By characterizing entire ERP waveforms in terms of these contributions (and their heterogeneity) rather than one or several time windows, this approach greatly facilitates the *discovery* of unexpected differences that may prove to be useful in research or clinical application, while better limiting user discretion (e.g. the choice of interval) and selective reporting (of one component/interval rather than others).

We outline the proposed pERP-RED algorithm and pERP-space analysis steps in the next section. Properties of the proposed approach are demonstrated using simulations (Section 2.3) and two separate real data examples (Section 4.3), each employing multiple tasks to derive a common set of pERPs followed by pERP-space analyses. Finally, to increase accessibility we provide an R package, `pERPred`. Investigators may keep their current pre-processing procedures and workflows up to and including the step of generating ERP waveforms at the subject level. These can then be imported into our software, allowing pERPs to be estimated and all the pERP-space analyses to be visualized through a graphical interface employing simple drop-down menus. We also make results available for the applied examples here through a publicly available browser-based implementation of that interface to encourage exploration and familiarization with the proposed pERP-space analyses.

2.2 Methods

In what follows we provide a detailed description of both the *pERP-RED* algorithm used to derive pERPs from an original dataset possibly spanning many participants, groups, and tasks, and then a set of tools that comprise the *pERP-space analysis* that we foresee being of value to a wide audience of investigators for the practical use of pERPs in research.

2.2.1 The pERP-RED Algorithm

The pERP-RED algorithm is motivated by the goal of estimating an underlying set of component waveforms, herein referred to as pERPs, such that the ERPs formed by time-locked

averages at any given electrode, participant, or trial type is approximately a weighted combination of these components. Our approach, depicted in Figure 2.1, involves (i) a series of data *concentrating* steps that turn a larger number of noisy waveform records (i.e. average ERPs for a given trial type, participant, electrode) into a smaller number of less noisy ones; and (ii) steps that generate maximally independent, “unmixed” components from these concentrated signals. Unlike some related approaches (such as spatiotemporal PCA, Spencer et al. (2001)) we do not conceptualize steps nor results as either “temporal” or “spatial” in the data reduction; rather we estimate the pERPs thought to underlie all ERPs (across all subjects, electrodes, and tasks) first, after which it is possible to both see how these load onto a given waveform and to see how these loadings are spatially distributed.

To describe these in detail, we first set notation. Let i denote subjects, $i = 1, \dots, N$; v tasks, $v = 1, \dots, V$; e electrodes, $e = 1, \dots, E_i$; t the number of time points, $t = 1, \dots, T$; and p the number of pERPs, $p = 1, \dots, P$. Note that each subject not only have electrodes in slightly different locations on the scalp, but may also have different numbers of electrodes owing to dropped channels after data cleaning steps. By beginning with the reduction of electrodes to “regions” within subject, this is unproblematic. The steps of the pERP-RED algorithm are then:¹

1. *Data initialization.* Data are first split, by subject into a training set and test set. Normalize each of the resulting ERPs to have unit variance. Proceed to next steps using only the training set.
2. *Electrode reduction.* The first data concentration or reduction step transforms electrode-level data within each subject to “region” level data within subjects, thereby reducing the electrode dimension. This is done by applying PCA to each of the N subject-specific matrices of size $(T \times V) \times E_i$, with data across electrodes in the columns and each tasks’ data concatenated in the rows. This produces N matrices of size $(T \times V) \times R_i$,

¹An example demonstrating the algorithm with the applications used here is implemented in R and can be viewed interactively at www.github.com/emjcampos/perpred. The pERPRED package for the R language allowing users to implement this on their own data will be made available upon publication.

Schematic of pERP-RED algorithm

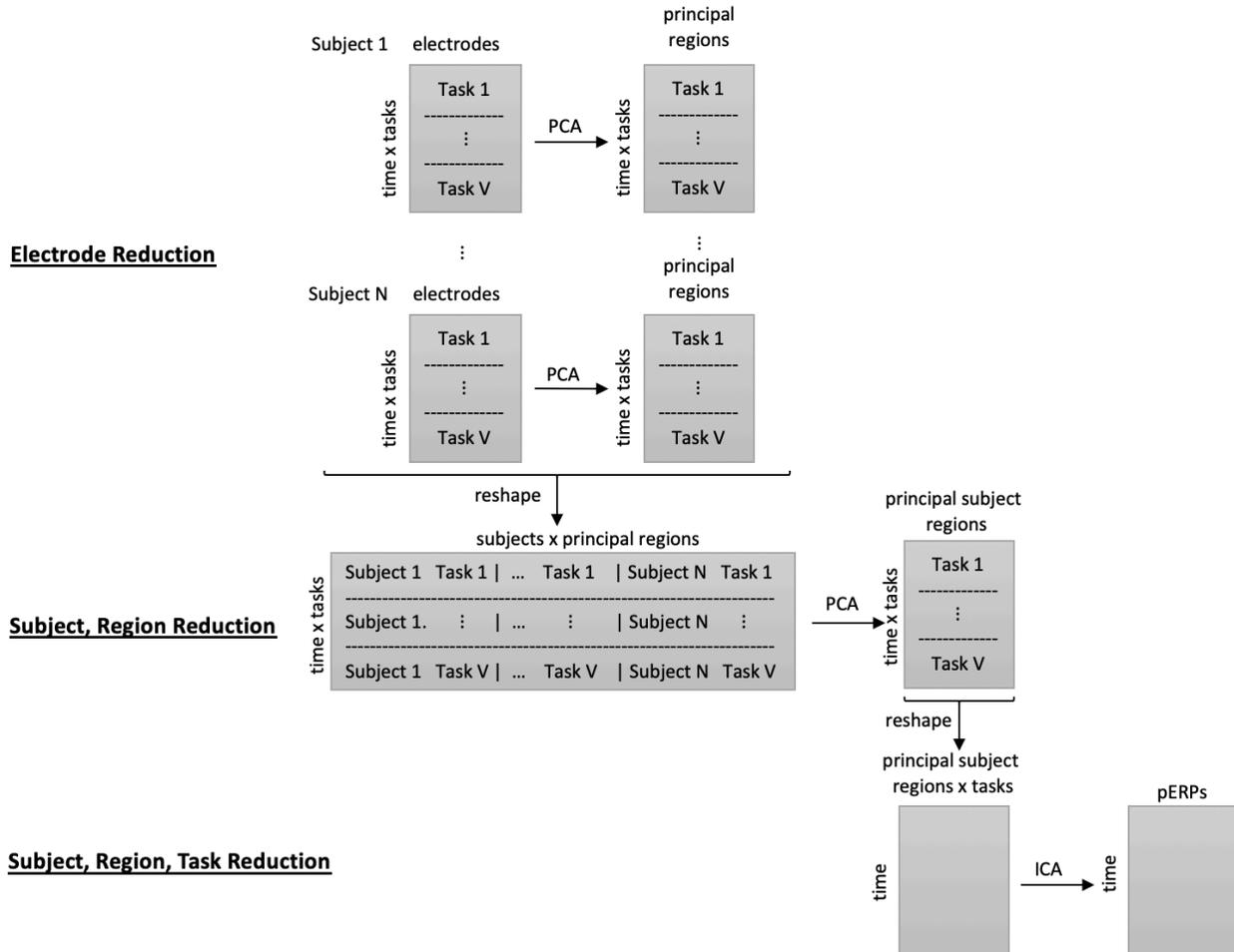


Figure 2.1: Illustration of the pERP-RED algorithm, described in text. First, at the within-subject level, correlation across electrodes is used to estimate a smaller number of “principle electrodes” for each subject, which we term “subject-region” scores. Next, these subject-region scores are combined together across subjects, then reduced to “principal subject-region” scores. These data are then reshaped to give separate columns for each principal subject-region score on each task/condition. The resulting waveforms are decomposed by ICA in order to arrive at the final pERPs.

where R_i is the number of principal regions that explains a pre-specified amount of variance. By default we set this threshold at 80%.

3. *Subject-region reduction.* The second data reduction step begins by reshaping the above data into a matrix of size $(T \times V) \times (\sum_{i=1}^N R_i)$ with all of the principal regions as the columns and the task and time dimensions concatenated in the rows. Each task-region ERP record is normalized to unit variance. Then PCA is used to generate N_R principal subject-regions.
4. *Source separation.* The data are then reshaped into a matrix of size $T \times (V \times N_R)$ with all task principal subject-regions as the columns and time as the rows. Fast ICA is then used to produce P principle ERPs, where P may be chosen by regressing the true signal (in the test set) onto the components and obtaining an R^2 value.

We make several remarks on these choices. First, in this paper, we use subject-level averages over trials of a given type as the “records” that first enter the algorithm above. This need not be the case: if for example one is interested in practice effects, then earlier trials and later trials may be averaged separately. Note that even having averaged trials together within participants, once the pERPs are estimated, it is still possible to go back and examine whether the amplitudes of pERPs vary from earlier trials to later trials (we further discuss the analysis of these amplitudes below). Hence one need not posit that “all the trials being averaged together are the same”, and variation in how pERPs contribute to different trials in the original, un-averaged data can still be studied. Statistically, the only requirement is that the data that enter the algorithm should include various weightings of the underlying components we aim to recover. Records need not (and certainly will not) be homogeneous in these components or mixtures of them. Put differently, any averaging that occurs prior to beginning the algorithm above should not “annihilate” or cancel out any components that would otherwise have been of interest.

Next, the purpose of splitting the data into a training and test set is to allow us to choose the appropriate dimensionality of the data in the final step, and to allow an unbiased evaluation of how well the principle ERPs explain the signal in a set of subjects not used to

derive those principle ERPs. We use 2/3 for training and 1/3 for testing but this choice is left to the user. Once the training-test set split has been used to decide upon the number of components, the components are re-estimated using all of the data.

At each data reduction step, we normalize the records. Equivalently, the covariance matrix underlying the PCA steps is actually a correlation matrix. In the “electrode reduction” step, the aim is to go from a large number of electrodes that may have highly correlated signals to a smaller set, each of which provides uncorrelated information relative to the others. That is, if we imagine a set of electrodes that all have a correlation of nearly one or negative one with each other, we would choose to collapse these into a single signal. We call the resulting principle components “regions”, though they need not and likely will not reflect coherent spatial regions. The subsequent “subject-region reduction” is analogous: if there are groups of participants that have “region” (combined electrode) signals that are highly correlated, this information can be combined with little loss.

In the final step, we use ICA rather than PCA because the aim is to take the “concentrated” data from the two reduction steps, and estimate components that are believed to be maximally independent. It thus resembles a blind source separation problem for which ICA is well suited, much as it is widely applied in source localization algorithms. This does not imply that the resulting components are functionally independent or meaningful sources.

Several choices remain user-driven at present. The first regards the proportion of variance each of the PCA steps must explain during the data concentration steps. Keeping a larger number of components implies that more of the data will survive into the final estimates, but at the cost of keeping more noise as well. It can also become computationally costly. Our supposition in choosing to keep enough components to cover 80% of the variation is that this should be sufficient to recover almost all of the true signal of value, whereas going beyond this is apt to introduce mostly noise.

The ICA step, too, requires choosing the number of components to estimate, P . In principle, $V \times N_R$ independent components can be kept. However there is no reason to expect that there are exactly $V \times N_R$ meaningful signals to separate. The risk of estimating too many

components is that over-decomposition may occur, splitting single meaningful components into separate estimated sources. Similarly, the risk of estimating too few components is that under-decomposition occurs, with separate sources remaining “mixed” together to alias as one (see e.g. Kairov et al. (2017)). In our case, over-decomposition is less harmful, because we will later characterize a given ERP waveform according to the linear combination of pERPs that compose it. The estimated pERPs can thus still “work together” to explain the average ERP for a particular condition, group or person. Thus, over-decomposition will simply lead to a wider set of pERPs being implicated and discussed as contributing to a given waveform or contrast.

Subject to this, however, the smallest set of components that will suffice is preferable for purposes of analysis and interpretation. We choose the number of final components P according to how well this set can explain actual ERP data, in a test set not used to estimate the pERPs. Denote the estimated pERPs as $\Phi = [\Phi_1, \Phi_2, \dots, \Phi_P]$. Let $j = 1, \dots, J$ be a counter running over each condition, subject, and electrode combination in the test set, with Y_j being the observed ERP record for the j^{th} condition/group/electrode. Without loss of generality, assume Y_j has been demeaned. Each of the J observed ERPs in the test set are then regressed individually on the matrix of the estimated pERPs, fitting the model

$$Y_j = \Phi\beta_j + \epsilon_j.$$

The variance in the test ERPs explained by the pERPs is then

$$R_{\text{test}}^2 = 1 - \frac{\|Y_j - \Phi\beta_j\|_{\mathcal{F}}^2}{\|Y_j\|_{\mathcal{F}}^2}$$

where $\|\cdot\|_{\mathcal{F}}^2$ denotes the Frobenius 2-norm, which corresponds effectively to a variance times N . We then choose the number of pERPs, P , such that raising P would result in relatively little additional gain in R_{test}^2 . Once P is chosen, all the data (in the training and test sets) are used to re-estimate the pERPs that will be used in the pERP-space analysis. We demonstrate this below.

2.2.2 The pERP-Space Analysis

The above procedure describes our proposal for how pERPs can be obtained. We now turn to how the estimated pERPs can be used in a complete analytical approach that avoids the problems inherent in the practice of using windowed peak (or mean) amplitude. The central concept is that any observed ERP waveform (for some condition over some group or individual at some electrode) can be recast as a vector of coefficients describing the magnitude of each pERP’s contribution to that ERP.

Step 1: Individual scoring. We start by using pERPs to extract data about a condition within an individual. Here condition refers to trial types within a single experiment, such as match and mismatch conditions within the audio paradigm of Section 4.1. For individual i and condition c , at electrode e , let the observed ERP be denoted by the vector $Y_{i,c,e}$. Henceforth we suppress the electrode index, e , for simplicity. Let Φ be the matrix whose columns contain the pERPs. We first regress $Y_{i,c}$ on the pERPs to obtain coefficients,

$$\omega_{i,c} = (\Phi^T \Phi)^{-1} \Phi^T Y_{i,c}. \quad (2.1)$$

The vector $\omega_{i,c}$ can thus be thought of as a vector of coefficients, amplitudes, loadings, or weights that encode the ERP $Y_{i,c}$ in terms of the contributions of each pERP. As $\omega_{i,c}$ is a vector of size P by 1, with each element representing loadings onto each pERP, operations performed on such vectors below (such as subtraction or averaging) are performed element-wise. Note that this brings about a substantial dimension reduction: ERPs that contain 1000 data points are represented using only 10-15 dimensions in the empirical examples explored below while explaining 80-90% of the total variation.

Second, when investigators are interested in a between-condition contrast, we also extract that contrast at the individual level. That is, suppose we are interested in how condition c compares to condition c' , obtaining $\omega_{i,c-c'} = \omega_{i,c} - \omega_{i,c'}$. If done literally, this implies computing $(\Phi^T \Phi)^{-1} \Phi^T Y_{i,c} - (\Phi^T \Phi)^{-1} \Phi^T Y_{i,c'}$. However, note that this is numerically the

same as $(\Phi^T \Phi)^{-1} \Phi^T (Y_{i,c} - Y_{i,c'})$, which first differences the ERPs for condition c and c' within subject i before regressing them on Φ . The benefit of doing this differencing within-subject is that later statistical tests that assume independence across subjects but not across conditions within subject will remain valid.

Finally, to reduce redundancy in what follows, we use the notation $\omega_{i,c-c'}$ generically even where investigators may be interested only in comparing scores from condition c to a null value of zero, and not to another condition.

Step 2: Summary across individuals. The values of $\omega_{i,c-c'}$ can then be regarded as data representing individual i 's response on condition c compared to c' (or on c alone if c' is omitted). These are measurements that can be summarized by conventional statistics such as the mean and standard deviation, while the mean can then be characterized by its standard error. We construct these quantities next.

Over all individuals, the mean vector, $\bar{\omega}_{c-c'}$, of size P by 1, would be given simply by $\frac{1}{N} \sum_i \omega_{i,c-c'}$. We further consider group-wise means written as

$$\bar{\omega}(g)_{c-c'} = \frac{1}{N_g} \sum_{i \in G_g} \omega_{i,c-c'}$$

for group g , where G_g denotes the set of indices, $\{i\}$, for subjects falling in group g and N_g denotes the size of the group.

To describe variability in the loadings across individuals in group g on condition c or on the contrast of c and c' , we construct the standard deviation within the group. For clarity we label this the across-person standard deviation (APSD) for group g ,

$$APSD(g)_{c-c'} = \sqrt{\sum_{i \in G_g} \frac{(\omega_{i,c-c'} - \bar{\omega}(g)_{c-c'})^2}{N_g - 1}}.$$

Finally, to facilitate inferences about group-wise means or mean contrasts, we construct

the standard errors that reflect the variation in those means,

$$SE(g)_{c-c'} = \sqrt{\frac{\sum_{i \in G_g} \frac{(\omega_{i,c-c'} - \bar{\omega}(g)_{c-c'})^2}{N_g - 1}}{N_g}}.$$

Step 3: Description and inference. With these statistics, the investigator can perform a range of analyses. A first question is how much each pERP contributed to a given condition (c) or contrast ($c - c'$). This can be tested for any grouping g , by constructing

$$t(g)_{c-c'} = \frac{\bar{\omega}(g)_{c-c'}}{SE(g)_{c-c'}}.$$

Note that the tests would be carried out element-wise on the P elements of $t(g)_{c-c'}$, corresponding to contributions from the P pERPs. If we are interested in knowing which pERP contributed with a detectably non-zero weight to a given condition c , this would be done for a single condition c (i.e., “setting c' to zero”). If we would instead like to know if the difference in contribution of any pERP to condition c versus c' is distinguishable from 0, then this would employ both c and c' . Note that the proposed approach can be used to quickly “screen” ERPs to determine which pERPs are contributing detectably non-zero amounts to a given condition or contrast. Our suggested reporting style, as illustrated below, would automatically provide such t-statistics (and associated p-values) for all pERPs in a table. This avoids selective reporting and aids in discovering unexpected differences, while producing a fixed number of tests.

Second, we may be interested not only in which pERP contributes to a given condition or contrast for individuals in group g , but how this differs from what we see in group g' . We can similarly construct

$$t(g, g')_{c-c'} = \frac{\bar{\omega}(g)_{c-c'} - \bar{\omega}(g')_{c-c'}}{\sqrt{SE(g)_{c-c'}^2 + SE(g')_{c-c'}^2}}$$

to test against the null hypothesis that two groups/conditions have the same mean contri-

butions from each pERP. Note that the intended usage here is for non-overlapping groups g and g' , which we treat as independent samples. Such contrast statistics can also be reported on all pERPs in table form for transparency. Again we illustrate this usage below, and such tables are automatically generated in the software provided.

Third, investigators may be interested not only in how pERP loadings on a condition or contrast vary between groups, but in whether certain groups have higher/lower variability than others. They can thus compare $APSD(g)_{c-c'}$ across groups. Fourth, recalling that ω quantities are available at each electrode e , topographic head maps can be shown as well for conditions/contrasts such as $\bar{\omega}(g)_{c-c'}$ or for between-group differences in conditions or contrasts, $\bar{\omega}(g)_{c-c'} - \bar{\omega}(g')_{c-c'}$. These can be shown either for these magnitude estimates themselves, or on the corresponding t-statistics to give a sense of statistical scale. We show examples in our applications below and provide this functionality in the software.

Other uses of pERPs

We note several other potential use cases that are less analogous to traditional ERP analysis of components but serve as additional uses for $\omega_{i,c-c'}$ values once extracted. First, investigators may wish to compare $\omega_{i,c-c'}$ to behavior or clinical measures for person i . Second, pERPs provide a means of ERP “cleaning”: one can reconstruct ERPs using only the derived pERPs. This leaves out components of the signal that have been deemed to be noise. Such a cleaning approach may be especially useful when visualizing individual or single-trial ERPs.

Third, the pERP analysis approach could in principle be used for participant rejection as well. Specifically, individuals for whom the pERPs collectively explain less of their signal must be very noisy, suggesting problems in data collection, or at least pointing to their divergence from the rest of the sample. Fourth and finally, the pERP-space analysis can also be used for outlier detection: individuals with very unusual values of $\omega_{i,c-c'}$ could be examined.

2.3 Simulation

The proposed pERP-RED algorithm is studied extensively through simulations in which the sample size, correlation among electrodes and tasks, percent of variation used to retain PCA components and signal-to-noise-ratio are varied. In addition to assessing the efficacy of pERP-RED, the simulations also target providing readers guidance on the choice of tuning parameters and comparing pERP-RED to alternate algorithms. Depending on the amount of noise in the simulated data, the run time of pERP-RED ranges between 3 and 12 minutes (longer for more noise) at $N = 50$.

2.3.1 Simulation Design

The observed signal is assumed to be a linear combination of the pERPs. Let $Y_{i,v,e}(t)$ denote the ERP signal observed for subject i at task v and electrode e . Note that we employ here the term “tasks” to refer to trial types generally, possibly defined across experiments, whereas the term “condition” refers to trial types within an experiment. The data generation model used in the simulation is:

$$Y_{i,v,e}(t) = \sum_{p=1}^P k_{p,v,e} \phi_p^*(t) + \sum_{p=1}^P \xi_{p,i,v,e} \phi_p^*(t) + \sum_{\ell=1}^L \alpha_{\ell,i,v,e} \psi_{\ell}(t) + \zeta_{i,v,e}(t), \quad (2.2)$$

where $\phi_p^*(t)$, $\psi_{\ell}(t)$ and $\zeta_{i,v,e}(t)$ denote the “true” pERPs, Fourier basis functions and measurement error, respectively and $k_{p,v,e}$ denotes the task- and electrode-specific weights, and $\xi_{p,i,v,e}$ and $\alpha_{\ell,i,v,e}$ denote the subject-, task- and electrode-specific weights of the respective bases decompositions. The total number of tasks, denoted V , the total number of true pERPs, denoted P , the total number of time points, denoted by T , the total number of Fourier bases, denoted by L , and the total number of electrodes per subject, denoted by E , are set to equal 9, 5, 500, 7, and 40, respectively.

The first term in (2.2) reflects that ERPs observed at each task and electrode are composed of a weighted average of the true pERPs, denoted by $\phi_p^*(t)$. To simulate a single

$\phi_p^*(t)$, we draw a function from a reproducing kernel Hilbert space of smooth, relatively low frequency functions. The functions in this space are simply the superposition of rescaled Gaussian kernels “centered” around different time-points. Specifically, $\phi_p^*(t)$ is the summation of 20 rescaled kernel functions, $\phi_p^*(t) = \sum_{m=1}^{20} a_m \exp(-(t - t_m)^2/h_p)$, where the center t_m of each kernel is chosen uniformly from $[0, 1]$, and the kernel bandwidth h_p is set to $0.3(-0.125 + 0.375p)$. The coefficients a_m rescaling each kernel is drawn from a uniform $(0, 1)$ distribution. These simulated signals are then rotated by ICA to form a maximally independent set of bases. The coefficients, $k_{p,v,e}$, are drawn independently from a normal distribution with mean zero and variance $\sigma_k^2 = 0.25$.

The second term in (2.2) represents subject-specific deviations from task- and electrode-specific signal. The subject-specific weights $\xi_{p,i,v,e}$ are drawn from a matrix normal distribution (to create a dependence structure within tasks and across electrodes) with mean zero and covariance matrices $\Sigma_{p,v}$ and $\Sigma_{p,e}$ of dimension $V * V$ and $E * E$ (identically and independently drawn over subjects). The covariance matrices for the $P = 5$ total pERPs are created in the same way for the electrode and task dimensions, starting out with a matrix of 1 on the diagonal and the correlation ρ elsewhere. The value of ρ will be varied to assess the effects of tuning parameter choices outlined in Section 2.3.2. These covariance matrices are then multiplied by a factor of $0.1(P - p)$, for $p = 1, \dots, 5$, respectively, so that each true ERP component is represented differently.

The third term in (2.2) represents a noise component that is structured in time. The Fourier bases contain $\psi_0(t) = 1$, and pairs of sine and cosine functions $\psi_{2r-1}(t) = \sqrt{2} \sin(2\pi r t)$ and $\psi_{2r}(t) = \sqrt{2} \cos(2\pi r t)$ for $r = 1, \dots, (L - 1)/2$, with $L = 7$. The coefficients $\alpha_{\ell,i,v,e}$ are generated in the same way as $\xi_{c,i,v,e}$. Specifically, $\alpha_{\ell,i,v,e}$ are drawn from a matrix normal distribution with mean zero and covariance matrices $\Sigma_{\ell,v}$ and $\Sigma_{\ell,e}$ which equal 1 on the diagonal and ρ on the off diagonal terms. The covariance matrices are multiplied by a factor of $0.05(\ell + 1)$ for $\ell = 1, \dots, 7$.

The last term in (2.2) is the independent and identically distributed measurement error. The measurement error $\zeta_{i,v,e}(t)$ is generated (independently over i , v , and e) from a

normal distribution with mean zero and variance σ_{error}^2 . The variance ratio of the signal $\sum_{p=1}^P k_{p,v,e} \phi_p^*(t) + \sum_{p=1}^P \xi_{p,i,v,e} \phi_p^*(t)$ to noise $\sum_{\ell=1}^L \alpha_{\ell,i,v,e} \psi_{\ell}(t) + \zeta_{i,v,e}(t)$ is varied in different simulation set-ups to equal 0.6 and 1, referred to as the high and low noise cases, respectively.

2.3.2 Simulation Results

The estimation accuracy of pERP-RED is assessed using two measures: R_{test}^2 and R_{pERP}^2 . The first measure R_{test}^2 is calculated in the last step of the pERP-RED algorithm as given in (2.2.1) to assess the estimation accuracy of the predicted records in the test set. The second measure, denoted by R_{pERP}^2 , assesses the performance of pERP-RED in targeting the true pERPs, which can only be estimated in simulations. To obtain R_{pERP}^2 , the matrix of true pERPs $\Phi^* = [\phi_1^*, \dots, \phi_5^*]$ is regressed on the estimated pERPs $\Phi = [\phi_1, \dots, \phi_P]$ to determine the degree to which linear combinations of the estimated pERPs can account for (any linear combination of) the true pERPs used in the simulation. That is: how accurately did we recover the true pERPs, up to rotations? R_{pERP}^2 is then given by the proportion of variance in the true pERPs explained by the estimated ones

$$R_{\text{pERP}}^2 = 1 - \frac{\|\Phi^* - \widehat{\Phi}^*\|_{\mathcal{F}}^2}{\|\Phi^*\|_{\mathcal{F}}^2},$$

where $\widehat{\Phi}^*$ is the prediction of the true ERPs given by the estimated ones.²

Before summarizing results from multiple simulation settings corresponding to varying sample sizes, correlations among electrodes and tasks and percent of variation used to retain PCA components, we present results from a single simulation set-up with $N = 50$, $\rho = 0.5$ (medium correlation induced among of tasks and electrodes) and 80% of variation used to retain PCA components in pERP-RED, in more detail. Figures 2.2a and 2.2b show a steep increase in R_{pERP}^2 and R_{test}^2 up to 5 pERPs estimated for both the high and low noise cases. This is evidence that the procedure is estimating the bases of the pERPs

² R_{pERP}^2 can be computed more efficiently as follows. Let the singular value decomposition of Φ be given as $\Phi = UAV^T$, where $\widehat{\Phi}^* = \Phi(\Phi^T\Phi)^{-1}\Phi^T\Phi^* = UU^T\Phi^*$ and $R_{\text{pERP}}^2 = 1 - \|\Phi^* - UU^T\Phi^*\|_{\mathcal{F}}^2 / \|\Phi^*\|_{\mathcal{F}}^2$.

Regression measures used to assess estimation

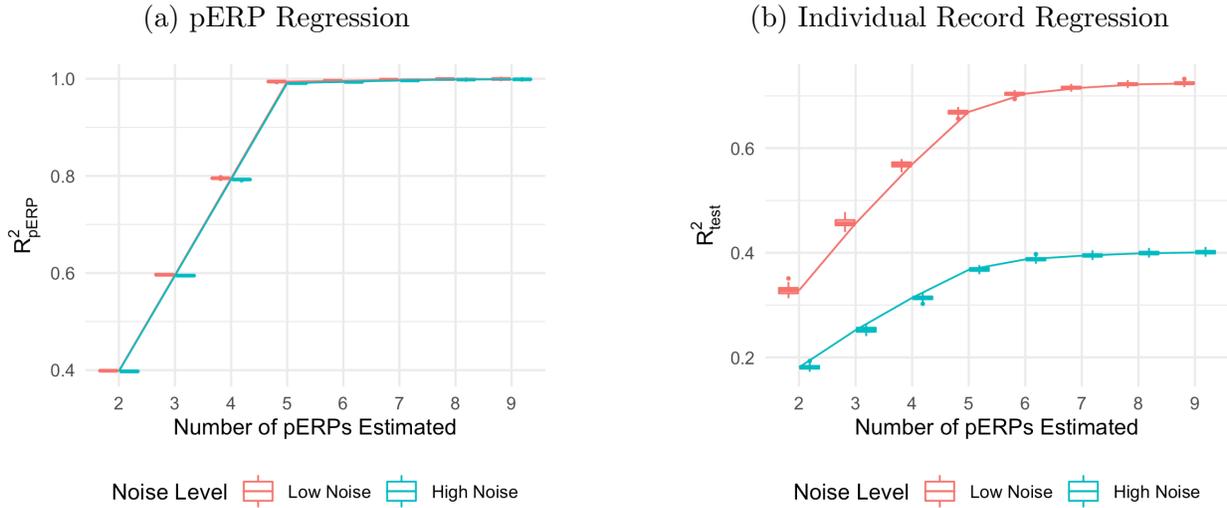


Figure 2.2: (a) Boxplots of R_{pERP}^2 from regressing the true pERPs on the estimated pERPs in 100 simulation runs show that the pERP-RED algorithm is correctly identifying 5 as the true number of pERPs. (b) Boxplots of the R_{test}^2 obtained by regressing ERP records from the test set on the estimated pERPs, for both the high and low noise simulation cases. In both cases, the boxplots of R_{test}^2 identify 5 as the true number of pERPs as the value of R_{test}^2 levels off sharply at that value.

correctly and identifying 5 as the true number of components. Figure 2.3 shows the regression coefficients and highlights the estimated and true component pairs. We see that under either the high noise or low noise case, each true pERP simulated was fitted by a combination of the estimated pERPs. At times there is a nearly one-to-one correspondence, but in general we see that each true pERP is a linear combination of estimated pERPs. Figure 2.4 displays each true pERP overlaying an estimated one which most heavily contributed to the estimation of the true pERP shown (using coefficients from Figure 2.3). This visually shows the nearly one-to-one relationship between some true and estimated pERPs, while in other cases linear combinations are required.

Regression coefficient heatmaps from regressing the true components on the pERPs

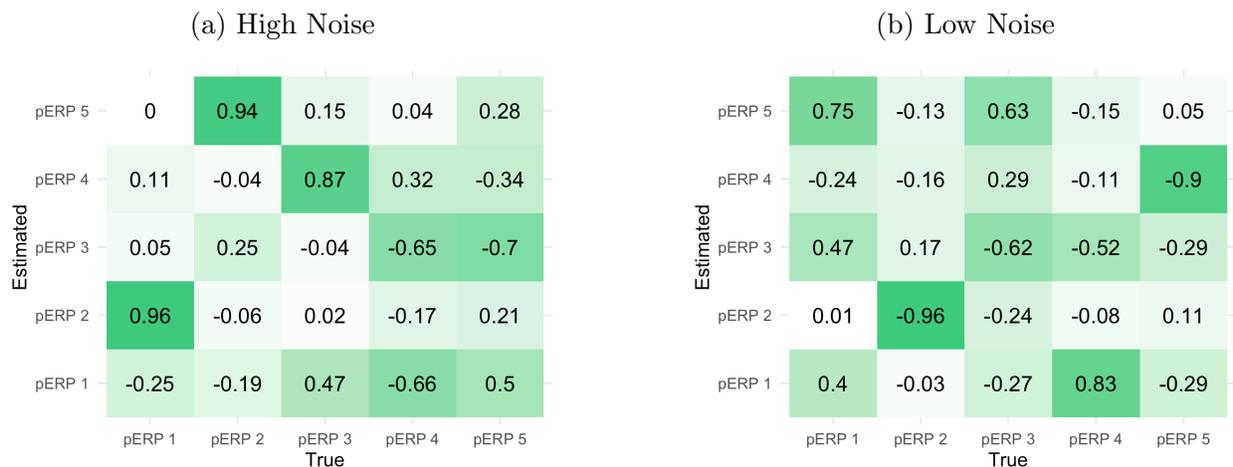


Figure 2.3: (a) The true pERPs are regressed on the estimated pERPs from the high noise simulation. The larger the coefficient, the darker the color in the figure and the stronger the match between true and estimated pERPs. The first true pERP is loaded heavily onto the second estimated pERP. (b) The true pERPs are regressed on the estimated pERPs from the low noise simulation in this case.

Effects of sample size, correlation level and percent of variation retained in PCA steps

For assessing variation in performance, we varied the sample size $N = 25, 50,$ and $100,$ correlation among electrodes and tasks $\rho = 0.1, 0.5,$ and 0.9 (used in generation of ξ and α), and the percent of variation used for retaining components in the PCA steps, for both the low and high noise cases. Trends from low and high noise cases are similar; hence we present results from the low noise case and defer results from the high noise case to the supplementary materials. While the effects are assessed both on R_{pERP}^2 and R_{test}^2 , variation in sample size, correlation ρ , amount of noise, and percent of variation chosen are found not to affect the algorithm’s ability to recover the true pERPs where R_{pERP}^2 displays the same pattern as in Figure 2.2a in all simulation cases. Hence we only report results on R_{test}^2 below.

While all three sample sizes considered provide sufficient information for recovery of the records (with steady R_{test}^2 levels across sample size), the variability in prediction accuracy decreases with increasing sample size, as expected, depicted in the narrower boxplots for larger

pERP Comparison

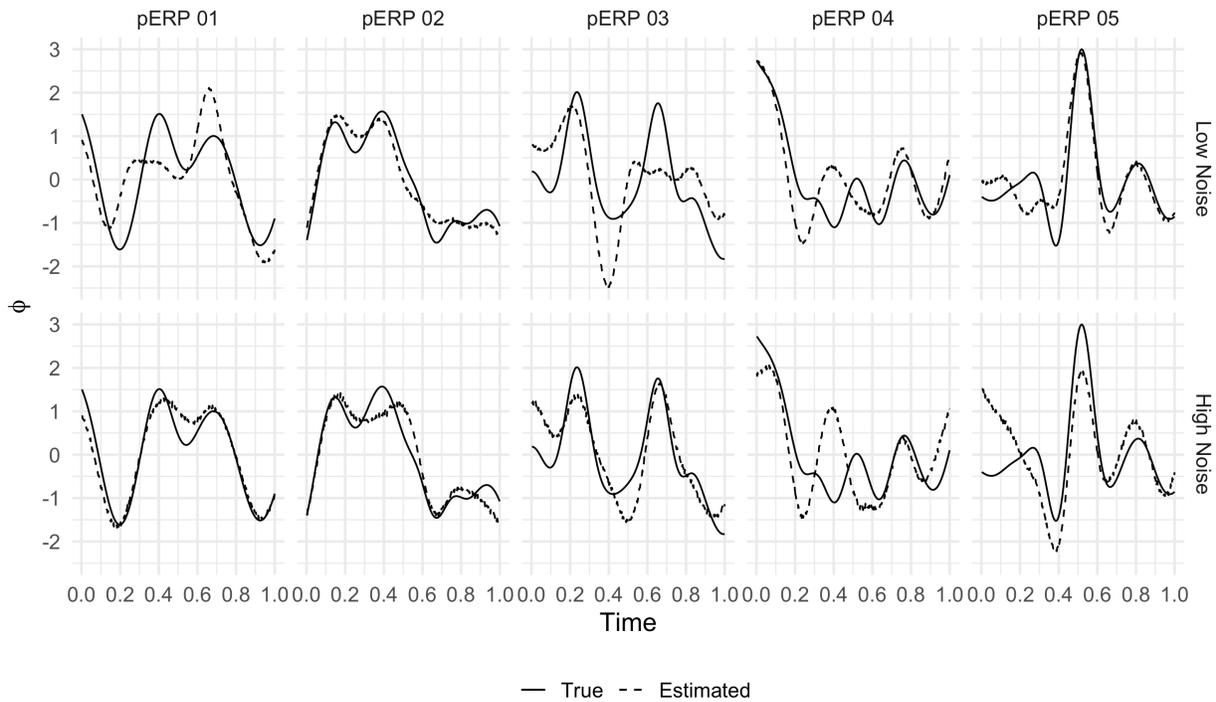


Figure 2.4: The true (solid line) and estimated (dashed line) pERPs are matched based on the regression coefficients from regressing the true pERPs on the estimated pERPs in the low noise (top row) and high noise (bottom row) simulations. Recall that pERPs can be estimated only up to rotations (linear combinations) of those that generated the data. Accordingly, in some cases the true and estimated pair look almost identical, but in others the estimated pERPs appear as combinations of true pERPs.

N in Figure 2.5. Since higher correlations across electrodes and tasks correspond to smaller effective total number of records, the prediction accuracy gets worse (with greater variability) as depicted in Figure 2.6. The percent of variation retained for PCA does not appear to affect the R_{test}^2 except for the high correlation case. When the correlation among electrodes and tasks is high, retaining more variation in PCA corresponds to retaining more noise, since repetitions over tasks and electrodes look more similar, leading to worse prediction accuracy (Figure 2.7).

Method Comparisons

The pERP-RED algorithm is compared to three different sets of basis expansions: Fourier, functional principal component analysis (FPCA), and a set derived using a single PCA and ICA similar to pERP-RED. Using Fourier bases, many more components are required to achieve the same predictive accuracy (see Supplementary Figures 4 through 7). This is expected, because these bases are fixed *ex ante* rather than being data-derived. The other three methods, all deriving their bases from the data, produce very similar prediction accuracy in terms of both R_{test}^2 and R_{pERP}^2 values; plots deferred to the supplementary materials). Additionally, a central feature of the pERP-RED approach is that the estimated bases are chosen not arbitrarily or *ex ante* but rather to describe the variation that was important to the original data structure. Consequently, as demonstrated in our applications, the non-arbitrary nature of these basis leads to substantive interpretations of the processes they may be detecting.

The difference between pERP-RED and FPCA is the ICA step employed by pERP-RED, which arises due to a conceptual difference in what information they seek to extract. FPCA chooses the bases that will explain the greatest variation in the data with the fewest components. By comparison, pERP-RED follows a similar logic in the initial reduction steps then employs an ICA in the final step because, as in other blind source-separation problems, maximizing the independence (not just orthogonality) of the components is a means to extract maximally “unmixed” underlying signals. These two approaches are expected to

Effects of Sample Size

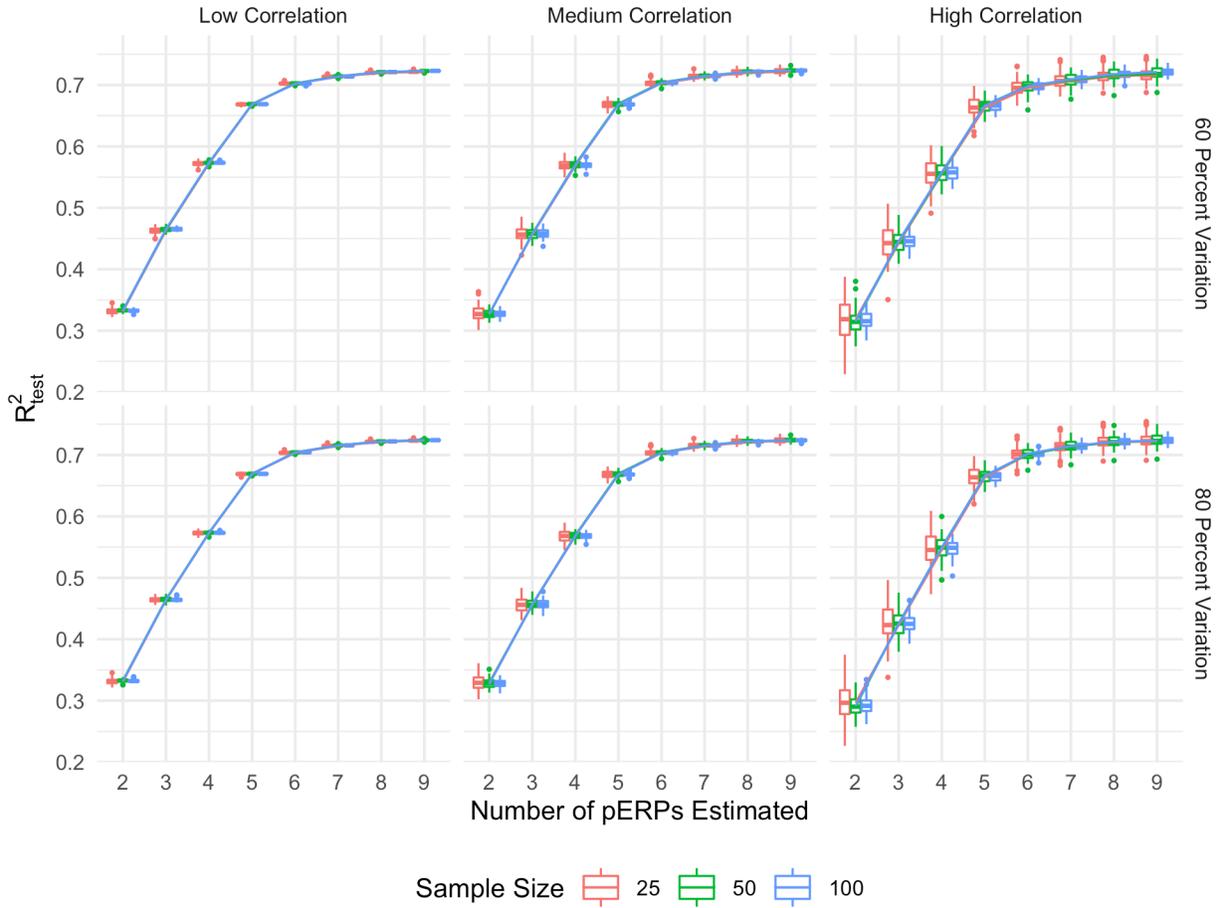


Figure 2.5: The boxplots of R^2_{test} as a function of the number of pERPs estimated across 100 simulations with 60 (first row) and 80 percent variation retained (second row) and low (first column), medium (second column) and high (third column) correlation between electrodes and tasks for $N = 25$ (red), 50 (green), and 100 (blue). The variability in prediction accuracy decreases with increasing sample size, as expected.

Effects of Correlation among Electrodes and Tasks

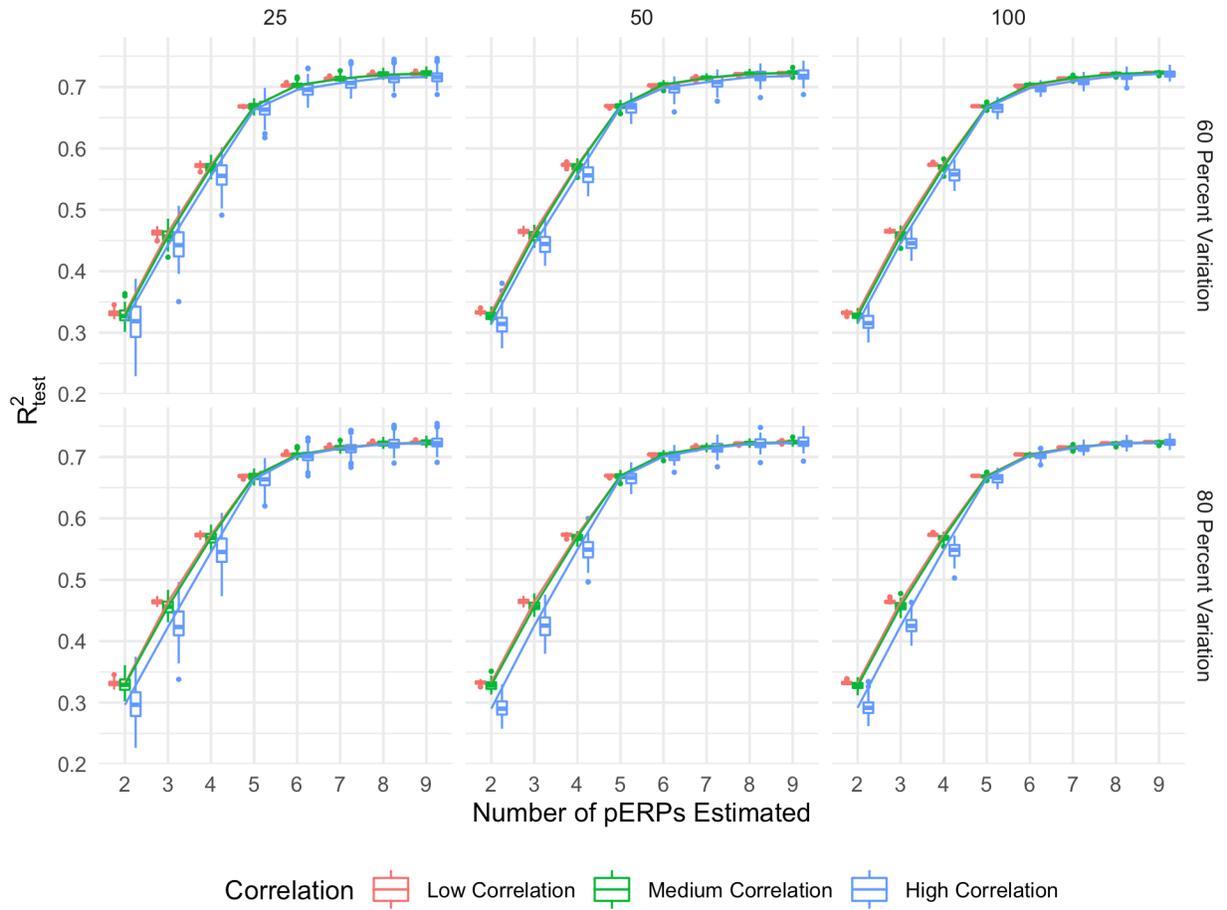


Figure 2.6: The boxplots of R^2_{test} as a function of the number of pERPs estimated across 100 simulations with 60 (first row) and 80 percent variation retained (second row) and $N = 25$ (first column), 50 (second column), and 100 (third column) for low (red), medium (green) and high (blue) correlation between electrodes and tasks. With increasing correlation, we see increasing variability in prediction accuracy .

Effects of Percent Variation Retained in PCA Steps

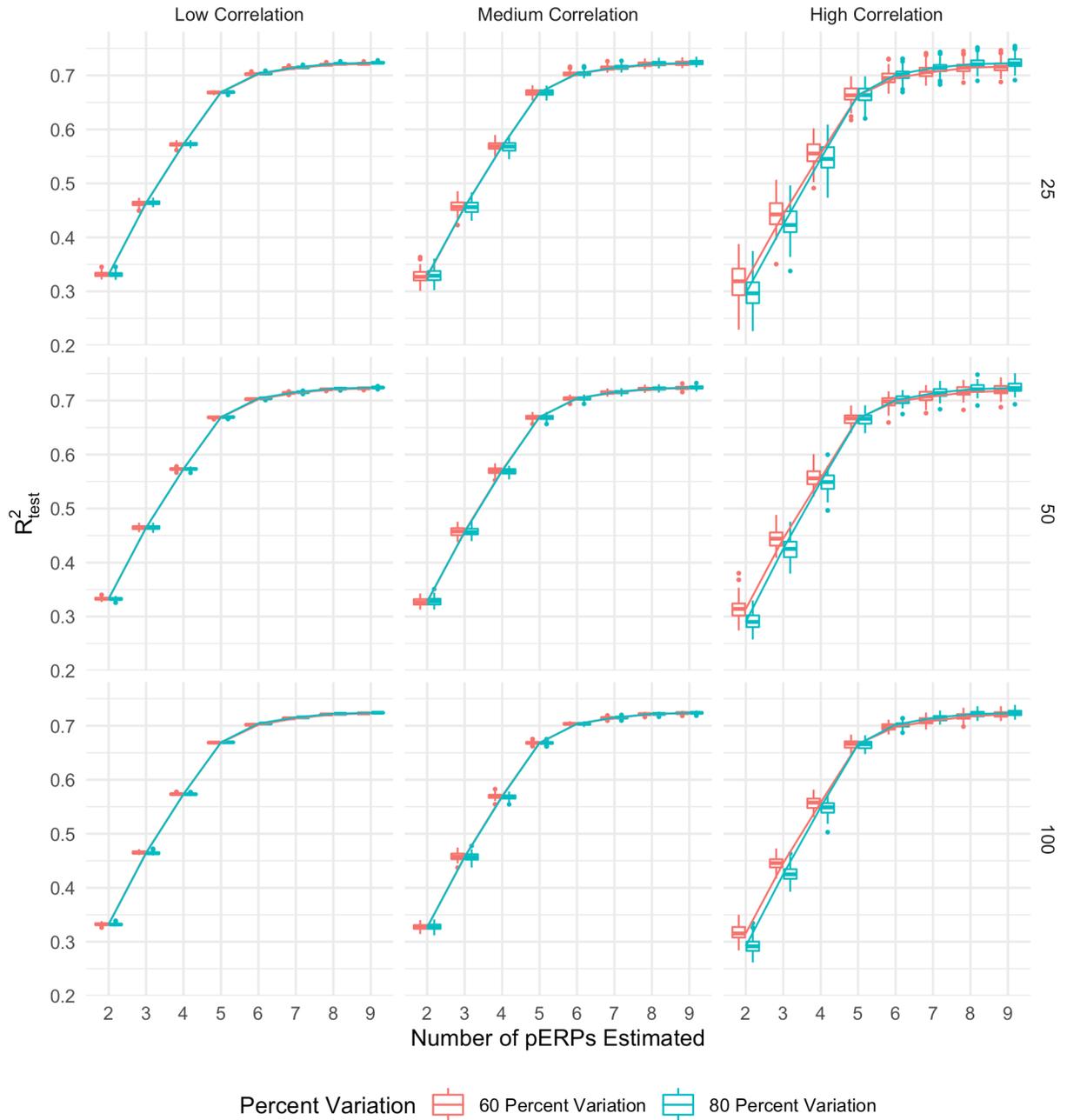


Figure 2.7: The boxplots of R^2_{test} as a function of the number of pERPs estimated across 100 simulations with $N = 25$ (first row), 50 (second row), and 100 (third row) and low (first column), medium (second column) and high (third column) correlation between electrodes and tasks for 60 (red) and 80 (blue) percent variation retained. When the correlation is high, retaining more variation in PCA corresponds to retaining more noise, leading to worse prediction accuracy.

explain similar amounts of variation in the data, but the latter adds greater value to the intended interpretation of their pERPs. As expected, our simulation shows very similar values for R_{pERP}^2 and R_{test}^2 in the two methods, with FPCA having a slightly lower R_{pERP}^2 (Supplementary Figures 4 and 5) and pERP-RED having slightly lower R_{test}^2 (Supplementary Figures 6 and 7).

As a final comparison, we consider a variation of pERP-RED we will call “single-PCA”, where there is only one data concentration (PCA) step and one ICA step. The single PCA would be conducted on the matrix formed by using all of the subjects and electrodes in the columns. We see in simulations (Supplementary Figures 4 through 7) that the pERP-RED and the “single-PCA” method perform similarly in terms of signal recovery in this setting on values of R_{pERP}^2 and R_{test}^2 . However the two have important differences in practice. First, the single-PCA method is limited in that the number of subjects times electrodes cannot exceed the number of time points times tasks because the PCA step would fail without additional constraints. The pERP-RED approach does not have this limitation, since the first PCA will reduce the number of electrodes before multiplying by the subjects. Second, by separately doing the electrode reduction within subject, pERP-RED allows subjects to vary arbitrarily in their electrode topography (and missingness), the projections of sources onto those electrodes, and the subject-specific components for their weightings. By contrast, the “single-PCA” reduction effectively assumes homogeneity in all these features across subjects. Lastly, pERP-RED allows the user to control the amount of variation used in each PCA step separately if they deem appropriate.

2.4 Applications

2.4.1 ASD Study

The first application of pERP-RED examines EEG data from a study of the neural mechanisms underlying language impairment in children with Autism Spectrum Disorder (ASD) (DiStefano *et al.*, 2019). This group has been difficult to study using traditional paradigms

that require following instructions or providing behavioral responses. Thirty-one children, aged 5-11 years old, who participated in the below described paradigms and provided sufficient high quality EEG data were considered: 14 typically developing (TD), 10 verbal ASD (vASD), and 7 minimally verbal ASD (mvASD). The three groups were age-matched. In the ASD participants, diagnoses had been made prior to enrollment, through clinical diagnosis by independent clinical psychologists, child psychiatrists, and/or developmental pediatricians. These diagnoses were confirmed by the research team using the Autism Diagnostic Observation Schedule (ADOS) and Social Communication Questionnaire (SCQ).

Two picture-word matching paradigms were used, one audio and one visual. The word stimuli included 60 basic nouns taken from the MacArthur-Bates Communicative Development Inventories-2nd edition. Examples of words included animals such as a bird or a dog and inanimate objects such as a doll or a bike. In the audio paradigm, a picture of the word would appear on a white background and the child would hear a spoken word that was either the same (*match* condition) or a word neither semantically nor phonologically matched the word image (*mismatch* condition). In both conditions, the picture image appeared for 2000ms, where the auditory stimulus was played after 1000ms after the picture image was shown (see Figure 2.8). Each word image pair appeared twice—once in the matched and once in the mismatched condition. No behavioral response was required. Trials were presented in four blocks of 30 trials each, totaling 6 minutes.

The visual paradigm used the same nouns and same number of trials, but in each an image of the word appeared on screen after 1000ms rather than an audio recording (see Figure 2.8). As in the auditory paradigm, the two conditions are “match” or “mismatch”. In both paradigms, the trials were video recorded in order to remove trials in which the participants were not looking at the screen. Data collection, cleaning, and pre-processing steps are described in the Appendix.

ASD Experiments

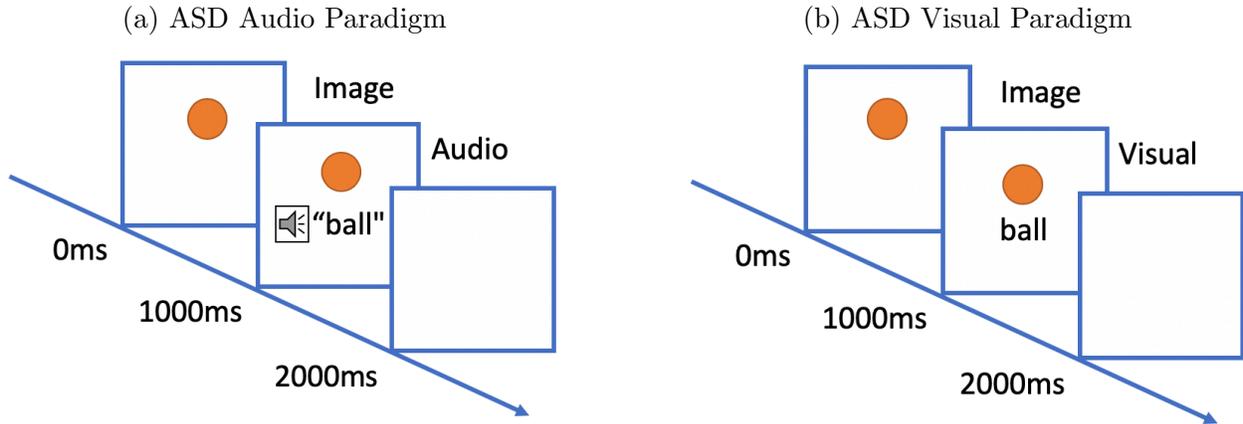


Figure 2.8: In both paradigms, the image appears for the full 2000ms. (a) In the audio paradigm, a spoken word is heard 1000ms after the onset of the image. (b) In the visual paradigm, the word image is seen 1000ms after the onset of the image.

ASD and ADHD Tasks and Trial Types

ASD		ADHD	
Image	Visual Match	SDRT Cue	CPT X Correct
Audio Match	Visual Mismatch	SDRT Probe	CPT X Incorrect
Audio Mismatch		SDRT Response Correct	CPT Not X Correct
		SDRT Maintenance	CPT Not X Incorrect

Table 2.1: We consider 5 separate trial types/ tasks drawn from the two ASD experiments (Audio and Visual). From the two ADHD experiments (SDRT and CPT), we consider 8 separate trial types/ tasks.

Results for ASD study

We chose to retain enough components to explain 90 percent of the total variation in the PCA steps of the pERP-RED algorithm. Ten pERPs achieve an R^2 value of 0.89 and are given in Figure A.4.1 ordered by their peak locations. We regress each demeaned record (subject and electrode ERP from each task), denoted by Y_j , on the set of estimated pERPs, Φ , to obtain the corresponding coefficients, ω_j . These coefficients can be used to reconstruct a less noisy version of each record as is shown in Appendix Figure A.4.2. To enhance transparency and encourage exploration of this approach, we provide a user-friendly data browser (ASD Data Exploration, https://perpred.shinyapps.io/asd_exploration/), where all analysis described here and others can be reproduced through a graphical interface. The user can thus become acquainted with the method before applying pERP-RED in their own lab. Screenshots from the application, providing graphics that mirror those used here, are provided in Appendix A.4.

To investigate the spatial distribution of pERP loadings, headmaps can be used to plot the estimated coefficients (averaged over all subjects) across the scalp as Figure A.4.3 shows for the *image* task. The coefficients of the first two pERPs are loaded heavily onto in the O1 and O2 electrodes. Notably, the overall ERP waveform from the image task (Figure 2.9, top) does not look as might be expected, showing no evidence of an N1 due to visual activation. However, our approach explains why the waveform does not include a negative deflection consistent with the N1 component, and recovers the ability to analyze the amplitude of a “hidden” component resembling the expected N1. Specifically, the time course of pERP 1 shows ongoing activity from the prior trial in the -100 to 0 ms interval. Such temporal overlap is to be expected given the fast rate of trials and that the ITI was fixed rather than jittered. However, the time course of pERP 2 is akin to the expected N1, peaking negatively around 90ms. The observed overall waveform, receiving contributions from both pERP 1 and 2, shows a relatively flat signal from -100 to 100ms. Figure 2.9 shows the contribution of pERP 2 to the overall waveform, in the TD and vASD groups. Loadings on pERP 2 are significant in all diagnostic groups, as would be expected for a visual task, and do not show significant

group differences. However, the variation in loadings across participants, indexed by the APSD, is 60% higher for the vASD group than for the TD group. Such higher variation in the vASD than TD group is generally to be expected. This is an important example of where the pERP-RED and pERP-space analysis is useful: Supposing the N1 indexed important activity we wish to measure for research or clinical purposes, it would have been ineffective to have measured it in this particular experiment using a peak or mean negativity around 100ms. Yet, a component resembling the N1 was elicited and can be indexed using the coefficient on pERP 2. This component shows similar average magnitude in the TD and vASD groups, but far greater variability in the latter.

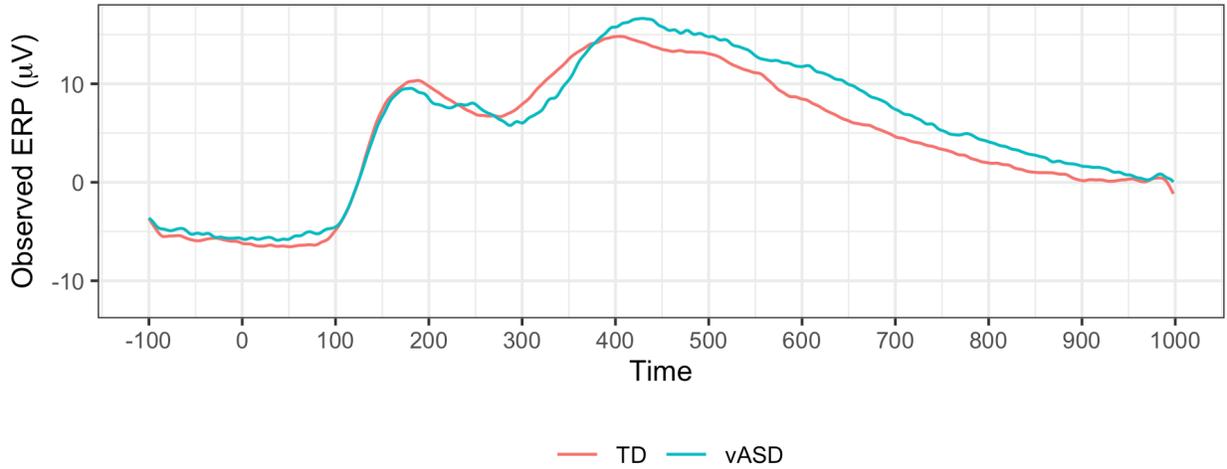
As noted, this approach also makes it easy to examine all pERPs for their loadings or significant differences in contrasts. While increasing the number of comparisons, this is done without user discretion of the sort allowed by user-driven selection of time intervals. For the image task, in addition to the expected N1 component that is recovered by this method, significant coefficients for pERPs 6 and 7 are found at O1 and O2, peaking positively around 450ms. This late positivity, present in all diagnostic groups, may be related to anticipation of the next stimulus. No other pERPs showed loadings that were statistically distinguishable from zero at these locations.

In the visual paradigm, the largest loadings are observed for pERP 7 across diagnostic groups found in the frontal and central regions (electrodes F3, F4, FZ, C3, CZ), corresponding to a negativity observed around 500ms. For loadings onto pERP 7, the mismatch condition was associated (though not significantly) with a stronger negativity than the match condition, as expected, in the TD group (possibly related to greater semantic processing).

For the auditory paradigm, a significant group difference emerges frontally at electrode F4, in pERP 9 when comparing activity between the TD and mvASD groups in the contrast of match vs. mismatch conditions (Figure 2.10). In the TD group, the mismatch condition leads to a deeper negativity than the match condition, seen around 700-800ms. This late negativity, expected to be larger in the mismatch condition, has been linked to semantic integration in the previous work of DiStefano et al. (2019).

ASD Image Contrast

(a) Observed ERP for image condition at O1, TD and vASD



(b) Contribution of pERP 2 for image condition, O1

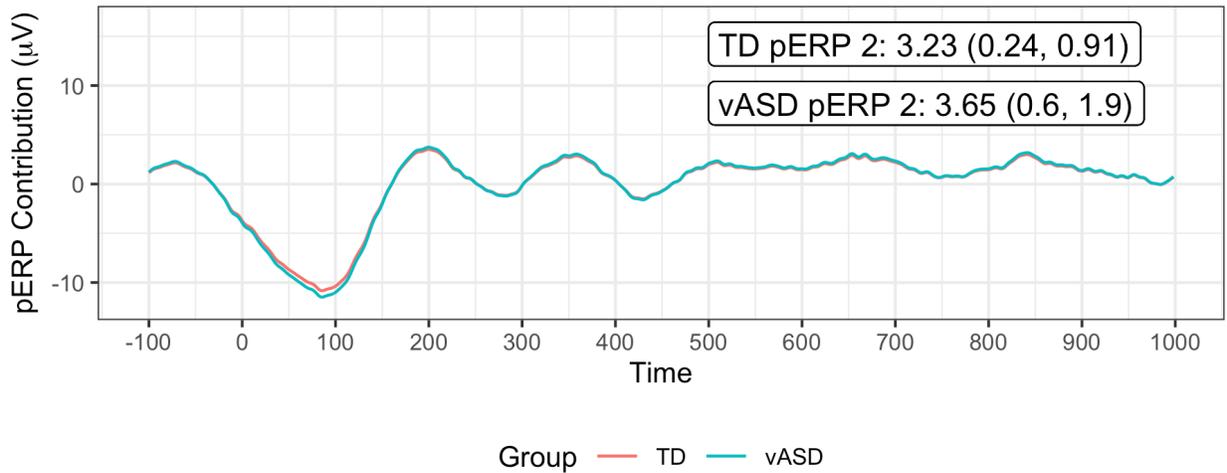


Figure 2.9: (a) The observed average ERP waveform for the image condition, by group. No N1 is evident. (b) ERP reconstructed using only pERP2, i.e. pERP 2 scaled by its contribution to the waveform (O1 plotted here). This component is detected despite the apparent lack of an N1 in the overall ERP because we also identify and model pERP 1, which captures spillover from the prior trial. The numbers shown are the mean weight with SE, and APSD in parentheses. The APSD reflects the subject variability and is much larger in the vASD group. Thus the vASD group has an average response very similar to that of the TD group, but there is a great deal more variability in the vASD group.

The prior work of DiStefano et al. (2019) study only the audio paradigm, and not the visual paradigm. Their main findings include a deeper negativity for mismatch than match trials, which they label as both an N4 and a later negativity, similar to the pERP 9 component picked up by the pERP-RED analysis. In relation to the N4, we see “N4-like” pERPs (5 and 6), peaking around 300 ms and 400 ms respectively. At Pz for example, we see the expected deeper negativity on these two pERPs for mismatch than for match trials, in each group. In the TD group, this difference in the amplitudes is large and statistically significant for pERP 5 ($t = 2.1$) though less so for pERP 6 ($t = 1.4$). In the vASD and mvASD, the differences appear similar but miss significance, ranging from $t = 1.4$ to $t = 1.8$.

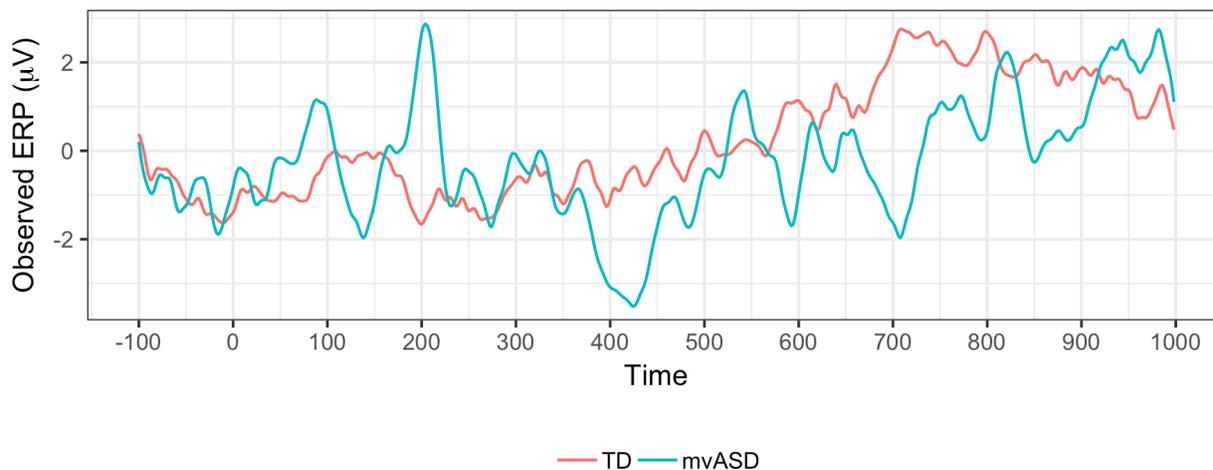
Relatedly, we also notice pERP 4, which peaks slightly earlier around 240 ms. This component also generates a significantly deeper negativity for mismatch than match in the TD group ($t = 2.1$) and in the mvASD group ($t = 3.0$), but interestingly shows no difference in the vASD group ($t = 0.1$). This is consistent with a finding in DiStefano et al. (2019) that perhaps surprisingly, mvASD participants appear more similar in many ways to TD than does the vASD group. Further, the vASD group may have far greater heterogeneity than the others: the APSD values for pERP 4, are on the order of 0.5-0.7 for the match and mismatch conditions in both the TD and mvASD groups, but are 1.2 (match) and 1.6 (mismatch) in the vASD group. That is, the standard error of coefficients across participants in the vASD group is 2-3 times larger than in TD or mvASD groups, again echoing the heightened heterogeneity of the vASD groups noted in DiStefano et al. (2019).

2.4.2 ADHD Study

We also test the methods on a dataset that is larger in terms of participants, diversity of tasks, and electrodes: a study of cognitive control in youth, aged 7-17 years old, with and without Attention Deficit and Hyperactivity Disorder (ADHD) with $N = 331$, of whom $N_{\text{ADHD}} = 242$ are in the ADHD group (clinicaltrials.gov ID: NCT00429273). Additional methodological details can be found in Lenartowicz et al. (2019). We examined data pooled from two sets of tasks: a spatial delayed response task (SDRT), and a continuous performance task (CPT)

ASD Sound match - mismatch contrast

(a) Sound predicted ERP mvASD vs TD



(b) pERP contribution to Sound contrast

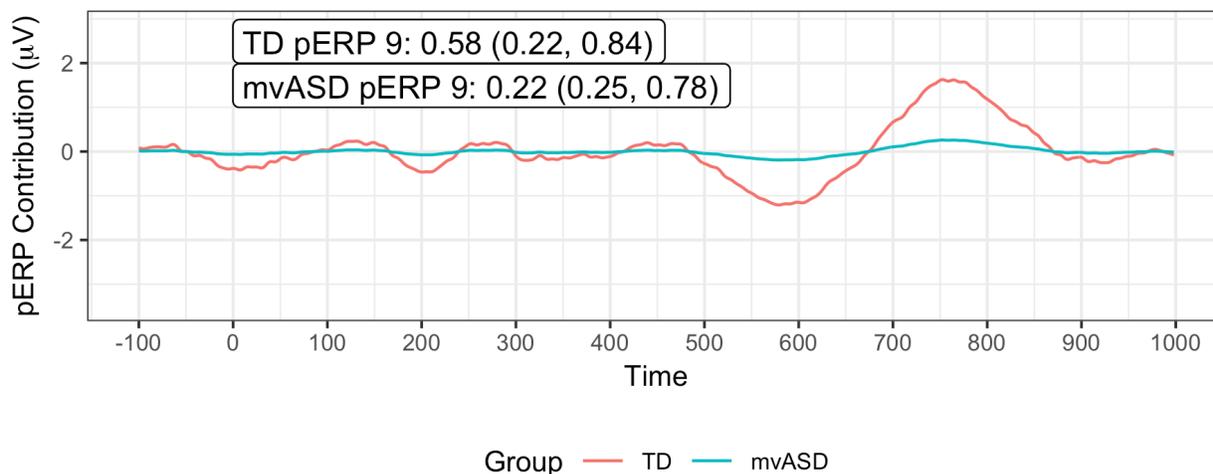


Figure 2.10: (a) ERP waveforms at F4. (b) A significant difference was found on pERP 9 between the mvASD and TD groups when contrasting match vs. mismatch. The numbers reported in the figure are the mean, with the SE and APSD in parentheses. This echoes a previously reported finding that the mismatch condition led to a deeper negativity at 700-800ms, thought to reflect semantic integration.

(see Figure 2.11). In the SDRT, participants are instructed to pay attention to the location of yellow dots on the screen against a black background. Each trial consists of a 500 ms fixation cross, followed by the yellow stimulus dots for 2000 ms (cue). Subjects are presented varying memory load of 1, 3, 5, or 7 dots at a time, in random order. Then a blank screen with a fixation cross appears for 3000ms (maintenance phase). Finally, a single probe green dot is presented for 3000ms (retrieval phase), during which the participant indicates whether the position of the probe matched the position of any of the yellow dots (pressing the left arrow to indicate a match, the right arrow to indicate a non-match). On half of the trials the green dot matched the location of a yellow dot and on half it did not. The trial ends when the participant responds, after which a fixed 3000 ms ITI (a blank screen) is displayed. If no response is made during the 3000 ms response interval, a message appears indicating that the subject did not respond and the next trial begins. Subjects are required to score above 50% accuracy on a series of practice rounds in order to move on to experimental blocks. There are 48 trials in each experimental block, and 2 blocks, with equal numbers of load sizes and equal numbers of match and no-match trials. Total experimental time is about 7 minutes per block, 14 minutes total. Including the training and practice trials the entire task requires roughly 17 minutes. ERPs are time-locked to cue, to probes, and to participant responses.

The CPT requires sustained attention (or vigilance) and response inhibition. Subjects are presented single letters (A, B, C, D, F, I, L, O, T, or X), one at a time in the center of the computer screen. The subject is instructed to press and release the spacebar as quickly as they can after viewing each letter – except when the letter is “X”, which indicates they should make no response. There are 360 continuous trials randomly presented with three different ISI lengths: 1000, 2000, or 4000 ms. Each letter is always presented for 250 ms, leaving either 750, 1750, or 3750 ms of remainder response time per trial, see Figure 2.11. Epochs are time-locked to stimulus onset. For present purposes, we categorize trials into just four types based on combinations of the cue (“X ” or “not X”) and response (“correct” or “incorrect”), Table 2.1. Data collection, cleaning, and pre-processing steps are described

ADHD Experiment Timelines

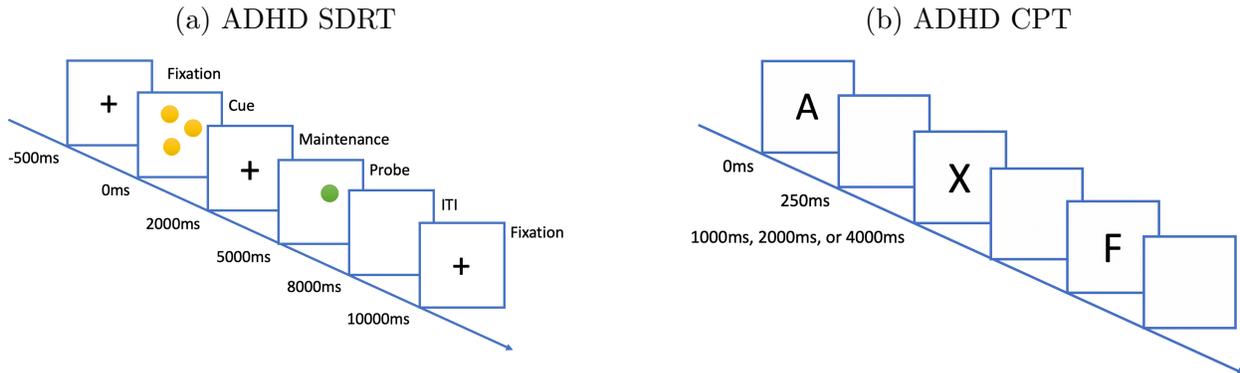


Figure 2.11: (a) In the SDRT task, a sequence of screens are presented. A fixation cross for 500ms, the cue dots for 2000ms, another fixation cross for 3000ms, and the probe dot for 3000ms. Within the 3000ms probe phase, the subject responds. (b) In the CPT task, the subject is presented with either an ‘X’ or ‘not X’ for 250ms, after which there is a blank screen for either 1000, 2000 or 4000ms before the next trial.

in the Appendix.

No other results from the CPT task of this dataset have been published as yet. For the SDRT, Lenartowicz *et al.* (2019) describes a number of results, focusing principally on spectral EEG analyses, and particularly alpha modulation during the encoding and maintenance stages as a predictor of impairment. This study did however index vigilance using the P2 event-related potential in response to the onset of the fixation, finding a nearly-significant smaller P2 in the ADHD group compared to the TD group at FCz.

Results for ADHD study

While the SDRT and CPT data are too rich to fully examine here, we consider a number of analyses that illustrate the pERP approach. The CPT data cannot be made public as yet, but all analyses of the SDRT described here can be replicated using our online data browser (https://perpred.shinyapps.io/adhd_exploration/). In comparison to Lenartowicz *et al.* (2019), at FCz we do not see significant evidence for a difference between the ADHD and TD groups on a P2-like component following the fixation’s onset. However, if the ADHD and TD groups differ in vigilance to task relevant stimuli, we might expect to see

this not only in the P2 to fixation, but in components such as the N1 and P2 following the encoding cue itself. Continuing to look at FCz, we do see pERPs reflecting an apparent P2 (pERP 5, a positive peak near 200 ms post encoding cue), and the N1 (pERP 4, a negative peak near 130 ms). Both show significantly non-zero loadings in the expected directions, for both ADHD and TD groups. While the differences in these average loadings (for ADHD vs. TD groups) do not reach statistical significance, the TD group shows a marginally larger loading on the N1-like component (pERP 4) with a t-statistic of 1.2 for the comparison. We consider the N1/P2 more extensively below.

There is a great deal more to report from this dataset, and we focus here on analyses that aid in evaluating the validity or usefulness of a pERP-based approach. First, as a validation exercise, we test the degree to which the algorithm identifies a N1/P2 complex in both the CPT and SDRT tasks. The N1/P2 complex has been widely identified in tasks involving visual stimuli and attention, such as the CPT and SDRT. If our approach is effective in constructing pERPs that reflect important or common sources of variation, having pooled both the CPT and SDRT to construct those pERPs, we might expect to see such components. Figure 2.12 shows the two pERPs noted above that appear to correspond to a N1 (pERP 4) and P2 (pERP 5), though the latter also contains a small negative deflection prior to 200ms. In the CPT experiment, taking Cz as an example, all four (visual) stimulus-locked trial types (X(No Go)-correct, X(No Go)-incorrect, notX(Go)-correct, notX(Go)-incorrect) show weights on pERPs 4 and 5 that are distinguishable from zero significantly, suggesting that the four trials types elicited pERPs 4 and 5 as would be expected. Furthermore, as shown in Figure 2.12, in each of the CPT trial types, the reconstructed signal using just these two components reproduce the expected form of a N1/P2 complex following the visual stimulus. The same holds for the all of the cue-locked trials of the SDRT experiment, and not for the response-locked or probe-locked trial types.

Next we turn to task specific expectations to further probe the validity of the approach. In the CPT task, trials with an “X (No Go)” are relatively rare, and are thus expected to produce an update or novelty signal typically associated with the P3. We thus expect

N1/P2 Complex in Five ADHD Trial Types

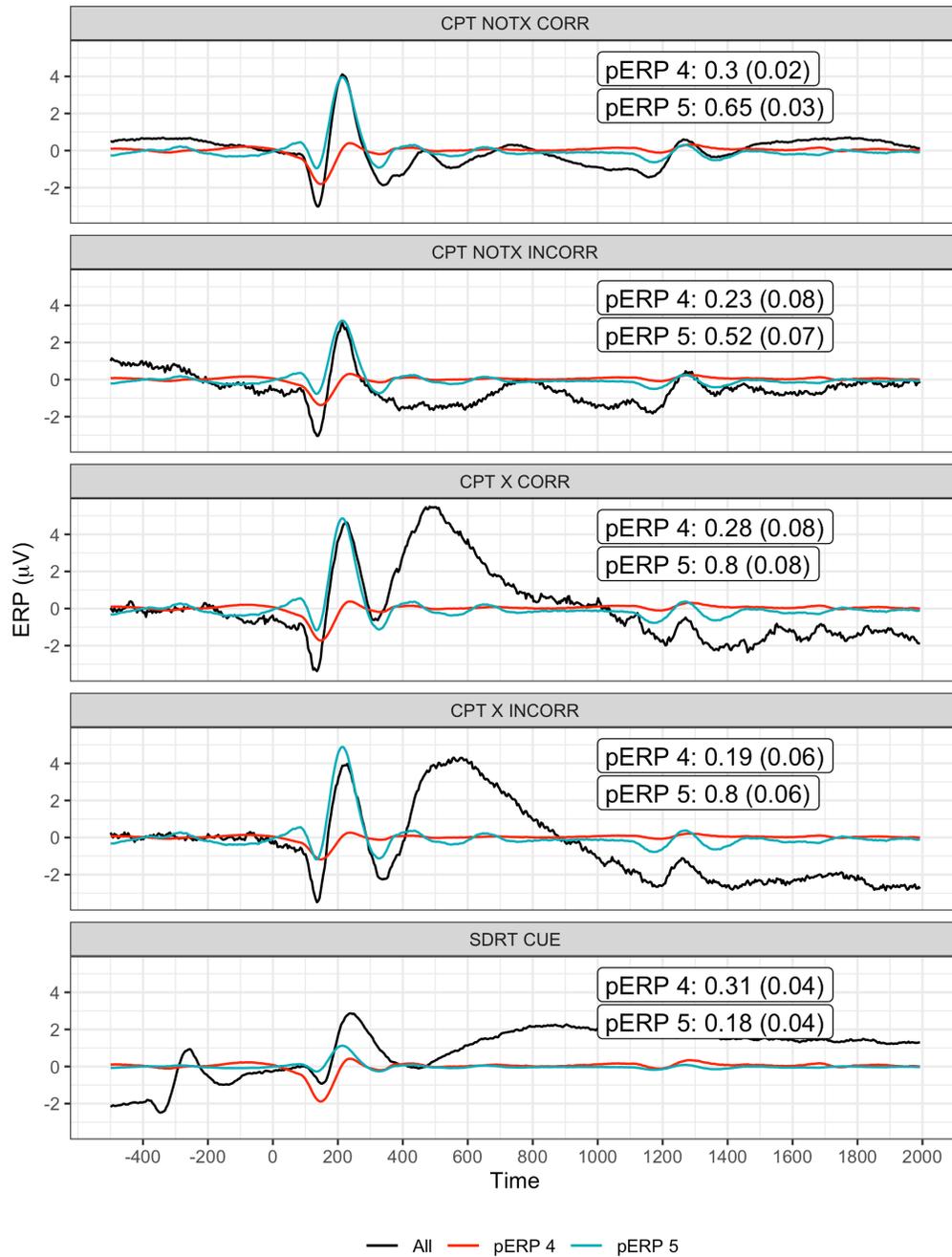


Figure 2.12: The N1/P2 as observed (solid black line) and as reconstructed from pERPs 4 (red line) and 5 (blue line) at CZ, at each of the five trials types. The loadings and standard errors for those loadings is shown in each figure. The corresponding t-statistics (mean divided by SE) range from 2.9 to 22 across these 10 cases, leading to firmly to rejecting the null hypothesis (of zero loading) in each.

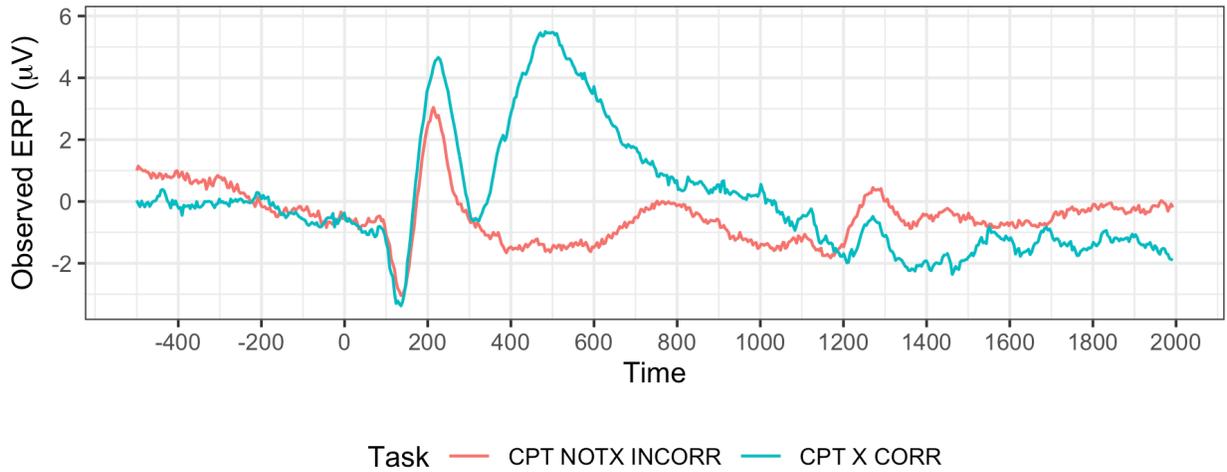
to see activity at 300-500ms in a contrast of “Not X (Go)-incorrect” versus “X (No Go)-, correct” in central and frontal areas. We contrast these two trial types because neither contain motor activity due to the response. Indeed, the observed ERP waveforms show a large positivity in the “X (No Go)-correct” trials, and not in the “Not X (Go)-incorrect” trials at sites such as Cz, see Figure 2.13 (top). This large difference between these two waveforms is quite broad, perhaps containing more than just a conventional P3, or averaging together positive deflections that peak at varying latencies. The estimated pERPs support such a hypothesis: We see that this contrast is explained not by a single broadly shaped pERP, but by pERPs 7, 8, and 9. These three pERPs each show a relatively sharp positivity, but at different latencies: 375ms for pERP 7, 475ms for pERP 8, and 625ms for pERP 9, see Figure 2.13 (bottom). In fact, all trial types with an “X” (i.e. No Go trials) regardless of response show heavy loadings on these three pERPs, whereas trial types without an “X” (and thus the majority stimulus type), regardless of response, show little weight on these pERPs, consistent with the expectation that pERPs 7, 8, and 9 relate to an updating or novelty or updating signal. As these three components have different latencies, they combine to form the slower peaking positivity seen in the overall waveform.

This highlights a potential use for such decompositions as compared to examining mean/peak amplitudes in the superposed ERP alone. Investigators could use the loadings on each of these three components, at the individual or group level to see if individual differences in clinical diagnosis or task performance correspond to different coefficients on these three novelty-related pERPs. In this case, the coefficients and their contrasts on these three pERPs are remarkably similar at least across the ADHD and non-ADHD groups. They also do not systematically vary as a function of participant age. Whether these individual, low-dimensional scores are more informative or predictive (of other behaviors, symptomology, or other traits) than a simple mean amplitude over some interval or a measure of latency remains an interesting question for the next phase of research with pERPs.

Turning to the SDRT task, we take the opportunity to illustrate another use of the pERP approach: characterizing heterogeneity. In addition to the standard errors that are used for

CPT X Correct vs Not X Incorrect

(a) CPT predicted ERP X vs Not X



(b) pERP contribution to the long positivity in the ‘CPT X Correct’ task

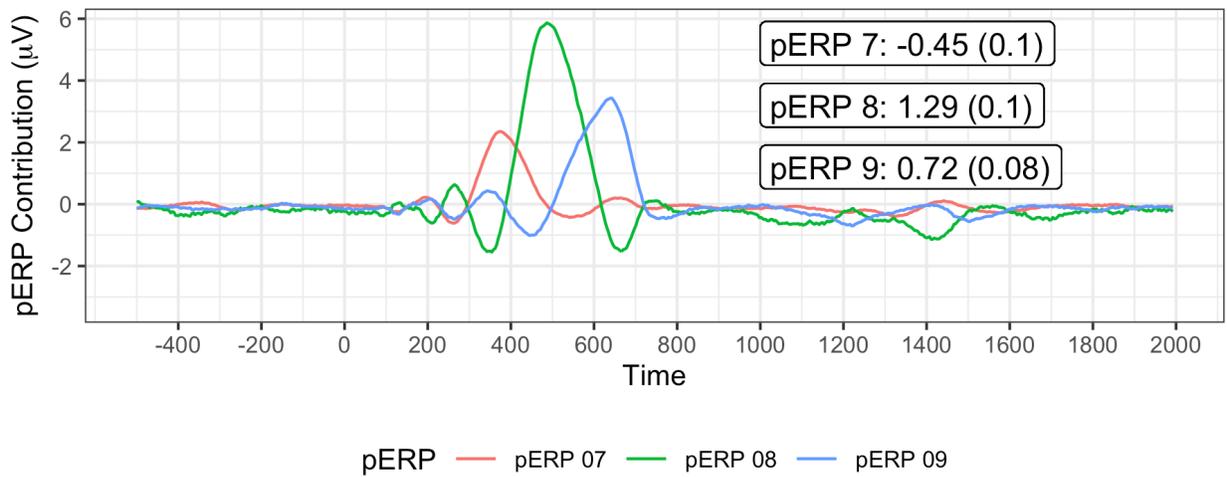


Figure 2.13: (a) The observed ERP at Cz. The ‘X Correct’ and ‘Not X Incorrect’ tasks are used for this contrast because neither include a motor response. The large positivity in the ‘X Correct’ task is expected as it indicates the update/novelty of seeing a rare event. (b) The large positivity in the ‘X Correct’ task from 400-800ms can be decomposed into pERPs 7 (red), 8 (green), and 9 (blue). The numbers shown are the mean and SE of the coefficients.

inference, we provide the “across participant standard deviations” (APSD), which characterize how the loadings on a given pERP vary from one participant to the next. Consider, for example, the waveform for the maintenance period in the SDRT. Because ADHD has implications for the ability to maintain attention and working memory resources on task, we may expect differences here. Table 2.2 shows loadings for all pERPs in the maintenance condition at electrode Cz, separately for typically developing (TD) participants and those categorized as either ADHD-inattentive or ADHD-combined (containing participants diagnosed with both inattention and hyperactivity). On the loadings, we see larger differences between the TD group and the ADHD-combined group, though these do not reach statistical significance³, while the TD and ADHD-inattentive group tend to be more similar. Returning to the question of heterogeneity, the TD and ADHD-inattentive group have very similar levels of heterogeneity in loadings, as indicated by the APSD value on each pERP. By contrast, the ADHD-combined group has higher APSD values on every pERP, indicating greater heterogeneity across the individuals that make up this group. Heterogeneity is especially large in the first two pERPs, and again for pERPs 10-13 onward. The early pERPs 1 and 2 appear to reflect activity before the cue has disappeared from the screen, while pERPs 10-13 contribute to a late ongoing positivity in the waveform — perhaps related to maintenance of task-relevant attention or working memory.

Using the pERP-space approach has several benefits over picking an interval, computing a mean or peak amplitude by subject, and compare the standard deviations across those measures by group. First, it seeks to mitigate the overlap problem that is inherent in using peak/mean amplitude, which sometimes makes peaks of important components difficult or impossible to locate, as in the N1 example in the ASD data above. Second, it avoids the requirement of picking the interval over which to do this, and thus reduces the risk of overmining the data to find an interval that works. Third, the pERP-space approach pinpoints variability in terms of the pERPs involved rather than a broad interval. By doing so, it suggests which pERPs show most variability and would thus be interesting to examine at

³On each of the first five pERPs, taken separately, the t-statistic for a difference in loadings between the TD and ADHD-combined group are approximately 1.6.

ADHD and TD Groups: Maintenance Condition (Cz)

pERP	Combined			Inattention			TD		
	Mean (SE)	t	APSD	Mean (SE)	t	APSD	Mean(SE)	t	APSD
pERP 01	-0.11 (0.27)	-0.41	2.94	0.32 (0.14)	2.30	1.50	0.41 (0.15)	2.67	1.44
pERP 02	0.18 (0.28)	0.65	3.01	-0.29 (0.15)	-1.91	1.68	-0.41 (0.18)	-2.28	1.69
pERP 03	0.1 (0.18)	0.58	1.91	-0.11 (0.13)	-0.86	1.41	-0.32 (0.15)	-2.10	1.44
pERP 04	0.02 (0.1)	0.23	1.07	0.09 (0.09)	1.06	0.94	0.22 (0.1)	2.25	0.90
pERP 05	0.21 (0.11)	1.96	1.15	0.1 (0.08)	1.17	0.90	-0.02 (0.09)	-0.20	0.87
pERP 06	0.13 (0.11)	1.22	1.17	0.07 (0.09)	0.81	0.94	0.08 (0.1)	0.86	0.90
pERP 07	0.28 (0.12)	2.30	1.29	0.35 (0.09)	3.80	1.02	0.45 (0.09)	4.89	0.86
pERP 08	-0.13 (0.12)	-1.03	1.34	-0.14 (0.09)	-1.66	0.94	-0.31 (0.1)	-3.11	0.93
pERP 09	0.07 (0.15)	0.51	1.56	-0.01 (0.07)	-0.12	0.79	-0.07 (0.09)	-0.80	0.80
pERP 10	-0.17 (0.17)	-1.02	1.83	0.04 (0.08)	0.54	0.88	-0.03 (0.08)	-0.43	0.74
pERP 11	-0.01 (0.11)	-0.11	1.18	-0.04 (0.06)	-0.69	0.70	-0.06 (0.07)	-0.82	0.68
pERP 12	0.09 (0.14)	0.64	1.55	-0.08 (0.06)	-1.29	0.65	-0.04 (0.06)	-0.73	0.56
pERP 13	-0.08 (0.18)	-0.45	1.97	0.09 (0.08)	1.07	0.91	0.13 (0.11)	1.18	1.07
pERP 14	-0.07 (0.15)	-0.48	1.56	0.1 (0.11)	0.90	1.25	0.11 (0.11)	1.02	1.01
pERP 15	-0.26 (0.19)	-1.37	2.01	0.16 (0.13)	1.23	1.42	0.14 (0.15)	0.96	1.38

Table 2.2: The APSD captures variability across participants in the weight given to each pERP. We see that the ADHD group has a much higher APSD for many of the pERPs than the TD group.

the individual (or even trial-by-trial) level to compare against behavior or diagnostic status.

2.5 Discussion

We describe a method for estimating an underlying set of components and then analyzing ERP waveforms in terms of these components. The fundamental idea is that any ERP waveform, from any electrode, subject, condition, and task, can be represented as a weighted combination of an underlying set of waveforms. The pERP-RED algorithm provides one reasoned approach to empirically estimate an underlying set of components (pERPs) from a corpus of ERP data, taken across participants and tasks. We then describe how investigators can use these pERPs to extract information from the ERP waveform. To ensure the accessibility of this approach, we provide software that enables users to import ERP waveforms, estimate pERPs, and conduct the pERP-space analyses described in this paper (the `pERPred` package in `R`).

This approach has many advantages relative to the standard practice of extracting the mean or peak amplitude in a user-specified window and presuming the result indexes the magnitude of a known ERP component of *ex ante* interest. First and foremost, it sidesteps the long-standing concern that differences observed at a particular interval may be due to a component other than that of interest. It similarly addresses concerns of spillover from the prior trial. As a result, this approach not only dissociates the measurement of components from the measurement of peaks, but can measure components that don't even visibly appear as a peak in the waveform due to overlap (as in the N1 in the ASD example above). It can also show that one broad slow-peaking structure may be explained as a combination of multiple underlying components, each of which can be measured on the group or individual level (as for the P3 in the ADHD example above). The ability to measure the contribution of every pERP contributing to the observed waveform, for each group or individual, makes the pERP approach particularly relevant to investigators seeking to discover biomarkers from group or individual level data. Relatedly, the pERP-space approach is useful for characterizing heterogeneity across individuals in a given group, assessing the variation of the relative

contributions of every pERP for every participant via the APSD score we provide.

An additional advantage relates to broader research practice in terms of selective reporting and discovery. Our approach employs multiple comparisons, but the number of hypotheses that will be tested (at least at a given electrode and for a given contrast) is simply the number of pERPs, and can be accounted for through any multiple comparison correction the user would like. Investigators can transparently report all significant and non-significant results for a given condition or contrast, characterizing (non) differences in any pERP, and thus over the entire time course. By contrast, in standard practice, the user chooses the time interval(s) and reports only differences there. Moreover, there is no clear way to exhaust the set of components that may or may not differ in a conventional analysis as can be done in a pERP-space analysis. Our approach facilitates discovery of differences in underlying components not previously of interest, yet avoids a user-driven search for such differences.

A number of areas for future exploration exist, in particular to further probe the validity or test the usefulness of this approach. First, it is natural to extract the weight on a given pERP of interest at the trial-wise level, in order to see how they might predict or relate to trial-wise behavior. This has parallels in recent work on high-dimensional modeling of EEG data. For example, Hasenstab et al. (2015) model longitudinal trends over trials within ERPs from a single learning task. Hasenstab et al. (2017); Scheffler et al. (2020); Fiecas & Ombao (2016); Ombao et al. (2018) consider high-dimensional modeling of EEG including the trial dimension using functional data and time series analysis approaches. While functional data analysis has seen rapid growth over the past two decades, with important recent work modeling multivariate hierarchical functional data Di et al. (2009); Shou et al. (2015); Happ & Greven (2018); Zhang et al. (2019), existing works have not considered data structures observed across multiple experiments, groups, subjects, electrodes, tasks and conditions, as we have done in this paper. We consider development of high-dimensional functional principal components decompositions suitable for the complex hierarchical structure considered in this paper as an important direction of future work. Second, a source localization approach

conducted on pERPs (as activated in a given task) would be a useful next step. Since its early days (e.g. Makeig et al. (1995)) ICA has been proposed as a means of isolating sources that are more likely to represent separate cortical generators and thus be well suited for a source localization step through later dipole modeling. Whereas ICA waveforms for source localization are computed separately on each subject, the fact the pERPs are computed globally on a multi-subject dataset provides new opportunities for seeing how a single pERP localizes, in different subjects, without having to use clustering or other approaches to link components in one subject to those in another. Specifically, the headmap constructed for a given pERP in a given condition can be used as input to a dipole modeling algorithm. It is then an empirical question as to whether a given pERPs activity in a given condition will localize to similar regions in separate subjects. This provides ample opportunity for falsification, since nothing in the algorithm constrains the spatial distribution of a pERP to be similar across participants.

Finally, we want to point to several limitations of pERP-RED. First, pERP-RED can only accommodate data in which the product of the total number of time points and the total number of tasks exceeds the number of all principal regions across all subjects (derived in the first PCA step of electrode reduction). This limitation is a by product of the second PCA step of subject, region reduction in pERP-RED. This condition was satisfied in both data applications. A second limitation is that the derived pERPs are not penalized in time to attain a desired degree of smoothness. Related to the functional data analysis approaches mentioned above that would consider the structure along the time dimension, an interesting direction of future work would be to also consider penalization in time such as localized PCA Chen & Lei (2015) which can set estimated basis functions to zero in time windows of little variation, further enhancing interpretation of the derived pERPs.

CHAPTER 3

Multilevel hybrid principal components analysis for region-referenced multilevel functional EEG data

3.1 Introduction

Electroencephalography (EEG) is a non-invasive and low-cost modality to analyze brain activity (Kumar & Bhuvanewari, 2012). EEG experiments can be categorized into two groups: (1) resting state experiments and (2) task-based ones or event-related potential designs. In the latter category, EEG is recorded during presentation of a sequence of stimuli, in which the brain response, called an event-related potential, is time-locked to the onset of each stimulus. While EEG data from stimulus-based experiments are often analyzed in the time domain via ERPs, EEG data from resting state are typically analyzed in the frequency domain via power spectral density (PSD). The generated data in the time (ERPs) or the frequency (PSDs) domain can be viewed as region-referenced functional objects, recorded across the scalp, representing the regional dimension. In addition, the data typically also have a multilevel structure in which high-dimensional observations are collected across varying experimental conditions within a single visit or repeatedly across multiple visits.

Our first motivating study was on lexical-semantic processing of youth with autism spectrum disorder (ASD) conducted by our collaborators at the UCLA Center for Autism Research and Treatment (DiStefano *et al.*, 2019). The goal of the study was to uncover neural mechanisms underlying language impairment in ASD. One quarter of children with ASD are minimally verbal and little is known about the underlying neural mechanisms. EEG, a modality that does not require a behavioral or verbal response, was a valuable tool in

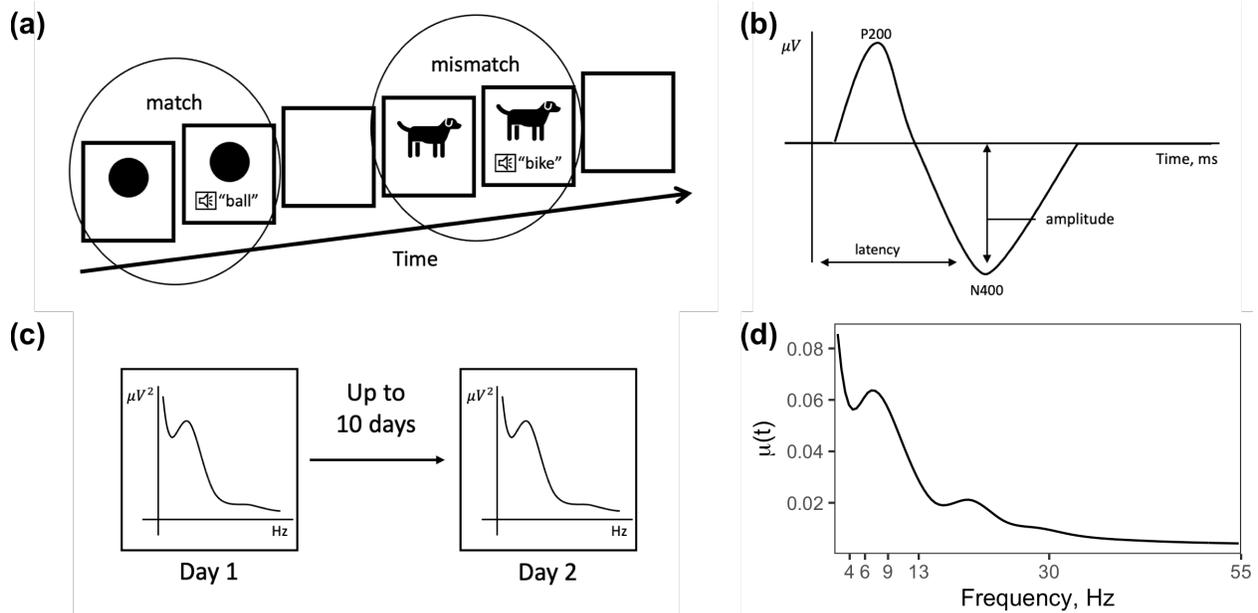


Figure 3.1: (a) In the audio odd-ball paradigm, the image appears for the full 2000ms and a spoken word is heard 1000ms after the onset of the image. The functional observation to be considered is time-locked to the onset of the audio, creating an ERP that is from -100ms to 1000ms. (b) The ERP waveform displaying the characteristic components P200 and N400. (c) In the ABC-CT feasibility study, PSD are recorded on two visits, approximately a week apart. (d) The estimated mean relative PSD $\mu(t)$ as a function of frequency in Hz.

reaching study goals. The children were presented with a sequence of images followed by an audio recording of the word that either matched or did not match the image, constituting the two conditions (match vs. mismatch) of the experiment (Figure 3.1a). ERP time-locked to each audio recording, measured across the scalp for both experimental conditions represents the multilevel region-referenced functional data in the time domain. Our second motivating study, which produced multilevel region-referenced functional data in the frequency domain was conducted by the Autism Biomarkers Consortium for Clinical Trials (ABC-CT) (McPartland *et al.*, 2020). EEG data were recorded during resting state, across two visits, which were approximately a week apart (Webb *et al.*, 2020). The resulting PSD collected across visits and scalp regions represents the multilevel region-referenced functional data in the frequency domain (Fig 3.1b).

Typical analysis of high-dimensional multilevel EEG data relies on collapsing the functional and regional dimensions. Instead of capturing the entire morphology of the ERP or

the PSD, the functional dimension is collapsed when characteristic phasic components of ERPs are targeted in the time domain or specific band power are targeted in the frequency domain. The collapse of the regional dimension corresponds to restricting analyses to a pre-determined scalp region. While such simplifications allow the use of common modeling techniques, simplifying analysis, they can fail to portray the more comprehensive differences in the entire ERP signal or the PSD across the scalp.

Functional data analysis (FDA) techniques (Ramsay & Silverman, 2005) can adequately handle trends in the entire ERP or PSD waveform. The literature on functional data analysis has grown rapidly over the past two decades, with a growing interest in analyzing high-dimensional functional data with dependencies among curves created by spatial or temporal proximity. Functional principal components analysis (FPCA) have been a major modeling theme in the FDA literature, with a focus on efficient and interpretable representations of functional variability in highly structured settings. For spatially-correlated functional observations Giraldo et al. (2010); Zhou et al. (2010) and Liu et al. (2017) utilized distance-based parametric correlation structures in functional modeling, while Baladandayuthapani et al. (2008) considered Bayesian regression splines and Staicu et al. (2010) built a nonparametric spatial process coupled with FPCA. For region-referenced EEG data, Scheffler et al. (2020) proposed Hybrid PCA (HPCA), combining vector and functional principal components analysis to decompose the high-dimensional signal along the regional and functional dimensions. HPCA relies on weak separability of the higher-dimensional covariance process (Chen et al., 2017; Lynch & Chen, 2018) and utilizes a product of one dimensional eigenvectors and eigenfunctions, obtained from the regional and functional marginal covariances, to represent the observed data in lower dimensions. Weak separability, which is less stringent than strong separability of the covariance processes in higher dimensions, assumes that the covariance can be represented by a weighted sum of separable covariance components and hence implies that the direction of variation (i.e. eigenvectors/eigenfunctions) along one dimension of the EEG data stays the same across fixed slices of the other dimension. For multilevel functional data, with an assumed exchangeable correlation structure among repetitions (experimental

conditions or visits), Di et al. (2009) proposed Multilevel Functional Principal Component Analysis (MFPCA) to decompose the total variation in univariate functional data into between- and within-subject variation. MFPCA has been extended for big data applications (Crainiceanu et al., 2009; Zipunnikov et al., 2011) and for sparse multilevel functional data (Di et al., 2014), however all extensions considered were limited to univariate functional observations. Note that recent works have also been proposed in the parallel literature of spectral analysis of time series that can model trends in the entire PSD without collapsing of dimensions. For spectral analysis of multiple dependent time series Krafty et al. (2011) proposed a mixed effects spectral representation of the data, while Macaro & Prado (2014) and Cadonna et al. (2019) took a bayesian nonparametric approach and Fiecas & Ombao (2016) modeled continuously evolving spectral densities across trials of an EEG learning experiment. For multi-subject applications, Hart et al. (2020) proposed modeling of PSD from resting state EEG of a sample of twin pairs enrolled in the Minnesota Twin Family Study.

We borrow ideas from HPCA (proposed for multidimensional functional data) and MFPCA (proposed for multilevel functional data) to decompose the variation in multilevel region-referenced EEG data in either the time or the frequency domain through the proposed Multilevel Hybrid Principal Components Analysis (M-HPCA). The multidimensionality refers to the EEG data collected across the scalp, constituting region-referenced functional data. The multilevel nature of the data is portrayed by our two motivating studies, outlined above, with repeatedly observed region-referenced EEG collected across experimental conditions in the audio odd-ball paradigm and across multiple visits in the ABC-CT feasibility study. Similar to MFPCA, M-HPCA begins by decomposing the total variation in the high-dimensional signal into between- and within-subject variation. In addition, relying on weak separability of the between- and within-subject variance (across the regional and functional dimensions), the higher-dimensional signal is represented using a basis set that are formed by products of one dimensional eigenvectors and eigenfunctions, obtained from the marginal regional and functional between- and within-covariances. Eigenscores, representing the mag-

nitude of variation in the subject-specific signal along the leading directions of variation, are estimated using mixed effects modeling.

A major challenge in estimation of the modeling components (including eigenscores and eigenvalues) via mixed effects machinery is the size of the data considered. Standard packages fail to scale up to the size of the high-dimensional multilevel region-referenced EEG data considered. To overcome this major challenge, we couple representation of the high-dimensional covariance matrices as weighted sums of lower dimensional building blocks, under the weak separability assumption, with an efficient minorization-maximization (MM) algorithm. The proposed solution achieves effective dimension reduction, leading to major computational savings, where estimation relies on inversion of much smaller building blocks, enabling applications to the high-dimensional EEG data. Relying on the achieved computational savings, inference is performed via bootstrap. The proposed M-HPCA decomposition is introduced in Section 3.2.1, followed by the proposed estimation algorithm, the mixed effects modeling and the MM algorithm outlined in Sections 3.2.2 and 3.2.3, respectively. Varied uses of M-HPCA are showcased with two data applications in Section 4.3. In the audio odd-ball paradigm, M-HPCA is used to study group differences in brain response to the two experimental conditions of the paradigm (match vs. mismatch). In the second application to the ABC-CT feasibility data, M-HPCA is used to assess short-term reliability of PSD (across study visits), through the proposed M-HPCA intraclass correlation (M-HPCA ICC) for multilevel region-referenced EEG data. Short-term reliability is an important step towards assessment of potential biomarkers for ASD. Extensive simulations, reported in Section 4.4, are used to assess the finite sample properties of the proposed approach, followed by a discussion section. To increase accessibility, an R package, `mhpca`, accompanies this article.

3.2 Multilevel Hybrid Principal Components Analysis (M-HPCA)

3.2.1 M-HPCA decomposition

We briefly review MFPCA and HPCA before we introduce the proposed multilevel hybrid principal components analysis (M-HPCA). For modeling multilevel functional data, MFPCA considered repeatedly observed functional data $Y_{ij}(t)$ on subjects $i = 1, \dots, n$, and repetitions $j = 1, \dots, J$. A two-way functional ANOVA model was proposed

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

where $\mu(t)$ denotes the overall mean function, $\eta_j(t)$ is the repetition-specific shift from the overall mean function, $Z_i(t)$ is the subject-specific deviation from the repetition-specific mean function and finally $W_{ij}(t)$ is the residual subject- and repetition-specific deviation. The correlation in the data is modeled through the two uncorrelated stochastic processes $Z_i(t)$ and $W_{ij}(t)$, leading to the decomposition of the total variation into between- ($\text{cov}\{Z_i(t), Z_i(t')\}$) and within-subject ($\text{cov}\{W_{ij}(t), W_{ij}(t')\}$) variation. This ANOVA decomposition implies an exchangeable correlation structure among the repetitions and allows for projection of the data onto a lower dimensional set of bases functions derived from FPCA of the between- and within-subject variation. It also allows for comparisons of the percent variation explained at each level of the data (subject vs. repetition).

In contrast, HPCA considered region-referenced functional data, where functional data is collected across the discrete dimension of region. Denoting the region-referenced functional data by $Y_i(r, t)$ for subject $i = 1, \dots, n$ and region $r = 1, \dots, R$, HPCA considered the representation

$$Y_i(r, t) = \mu(t) + \eta(r, t) + Z_i(r, t) + \epsilon_i(r, t),$$

where $\mu(t)$ denotes the overall mean function, $\eta(r, t)$ is the region-specific shift from the overall mean function and $Z_i(r, t)$ is the subject and region-specific random deviation. In the HPCA decomposition, the variation is modeled through the mean-zero stochastic process

$Z_i(r, t)$. Under the weak separability of the high-dimensional covariance of $Z_i(r, t)$, directions of variation (eigenvectors and eigenfunctions) are targeted along the regional and functional dimensions using PCA and FPCA, respectively. The data is then represented in lower dimensions through the product of the one-dimensional eigenvectors and eigenfunctions estimated. Hence in contrast to MFPCA, HPCA does not assume an exchangeable correlation structure among regions. Instead, it models region as a second discrete dimension added onto the continuous functional dimension and models variability along this additional dimension by PCA within the weak separability framework of the higher-dimensional covariance process. First decomposing the total variation into between and within covariance processes and assuming results in eigencomponents that are highly interpretable – giving rise to participant- and repetition-specific regional and variation separately.

M-HPCA combines different features from the two decompositions outlined above in modeling the multilevel region-referenced EEG data across multiple diagnostic groups. Let $Y_{dij}(r, t)$ denote the multilevel region-referenced functional observations for subject i , $i = 1, \dots, n_d$, from group d , $d = 1, \dots, D$, for repetition j , $j = 1, \dots, J$, in region r , $r = 1, \dots, R$ at time t , $t \in \mathcal{T}$. $Y_{dij}(r, t)$ is assumed to be square-integrable over the functional domain \mathcal{T} , which could represent time or frequency in the analysis of the ERP or PSD waveforms, respectively. While EEG data is typically available on a common set of regions across the scalp, data may be missing over repetitions, where each subject is observed at a different set of experimental conditions or visits. For ease of notation, we provide modeling details using the common set of repetitions $j = 1, \dots, J$ across subjects, however developments can easily be extended to accommodate missingness over j and are provided in more generality with subject-specific repetition sets in Section 3.2.2. The proposed M-HPCA decomposition is given as

$$Y_{dij}(r, t) = \mu(t) + \eta_{dj}(r, t) + Z_{di}(r, t) + W_{dij}(r, t) + \epsilon_{dij}(r, t),$$

where $\mu(t)$ denotes the overall mean function, $\eta_{dj}(r, t)$ is the group and repetition-specific shift from the overall mean function, $Z_{di}(r, t)$ and $W_{dij}(r, t)$ are the subject-specific and subject- and repetition-specific deviations, respectively, and $\epsilon_{dij}(r, t)$ is the independent measurement

error with mean zero and variance σ_d^2 .

Similar to the functional ANOVA model of MFPCA, the two stochastic processes $Z_{di}(r, t)$ and $W_{dij}(r, t)$ are assumed to be uncorrelated and will be referred to as the level 1 or between-subject, and level 2 or within-subject deviations. The proposed M-HPCA decomposes the total variation in the data, $K_{d, \text{Total}}\{(r, t), (r', t')\} \equiv \text{cov}\{Y_{dij}(r, t), Y_{dij}(r', t')\} = \text{cov}\{Z_{di}(r, t), Z_{di}(r', t')\} + \text{cov}\{W_{dij}(r, t), W_{dij}(r', t')\} + \sigma_d^2 \delta\{(r, t), (r', t')\}$ into between-, $K_{d, B}\{(r, t), (r', t')\} \equiv \text{cov}\{Y_{dij_1}(r, t), Y_{dij_2}(r', t')\} = \text{cov}\{Z_{di}(r, t), Z_{di}(r', t')\}$ and within-subject covariance, $K_{d, W}\{(r, t), (r', t')\} \equiv K_{d, \text{Total}}\{(r, t), (r', t')\} - K_{d, B}\{(r, t), (r', t')\} = \text{cov}\{W_{dij}(r, t), W_{dij}(r', t')\}$, where $\delta\{(r, t), (r', t')\}$ denotes the indicator function for $\{(r, t) = (r', t')\}$. The subscripts B and W are used to denote the between- and within-subject covariances, respectively.

Similar to HPCA, the proposed M-HPCA utilizes weak separability of the between- and within-subject covariances to decompose the signal in lower dimensions. Weak separability, which is less stringent than strong separability of the covariance processes in higher dimensions, assumes that the between- and within-subject covariances can be represented as a weighted sum of separable covariance components along the regional and functional dimensions. This assumption implies that the direction of variation (i.e. eigenvectors/eigenfunctions), not the covariance itself, along one dimension of the data (regional/functional), stays the same across fixed slices of the other dimension (functional/regional). Violations of the weak separability assumption on the between- and within-subject covariances is checked in data applications by visually assessing the similarity of the estimated eigenvectors and eigenfunctions across fixed values in the regional and functional domains (Section 4.3). In addition to visual assessments, we also check the weak separability assumption via a likelihood ratio test for the random effects correlation structure of the subject- and subject-region-specific eigenscores of the product eigencomponents proposed below in the mixed effects modeling framework of Section 3.2.3.

Utilizing weak separability, the common eigenfunctions and eigenvectors along each of the two dimensions (functional and regional) can be targeted using the marginal between- and

within-subject covariances. The marginal functional between- and within-subject covariances are defined as the sum of the between- and within-subject covariances across all regions,

$$\begin{aligned}\Sigma_{d,\mathcal{T},B}(t,t') &\equiv \sum_{r=1}^R K_{d,B}\{(r,t),(r,t')\} = \sum_{\ell=1}^{\infty} \tau_{d\ell,\mathcal{T}}^{(1)} \phi_{d\ell}^{(1)}(t) \phi_{d\ell}^{(1)}(t'), \text{ and} \\ \Sigma_{d,\mathcal{T},W}(t,t') &\equiv \sum_{r=1}^R K_{d,W}\{(r,t),(r,t')\} = \sum_{m=1}^{\infty} \tau_{dm,\mathcal{T}}^{(2)} \phi_{dm}^{(2)}(t) \phi_{dm}^{(2)}(t').\end{aligned}$$

In the above formulation $\phi_{d\ell}^{(1)}(t)$ and $\phi_{dm}^{(2)}(t)$ denote the level 1 and level 2 eigenfunctions and $\tau_{d\ell,\mathcal{T}}^{(1)}$, $\tau_{dm,\mathcal{T}}^{(2)}$ denote the respective eigenvalues. Superscripts (1) and (2) are used to differentiate quantities at the two levels of the data (subject vs. repetition, or between- vs within-subject). The marginal regional between- and within-subject covariance matrices are similarly defined as integrals of the between- and within-subject covariances across the functional domain,

$$\begin{aligned}(\Sigma_{d,\mathcal{R},B})_{r,r'} &\equiv \int_{\mathcal{T}} K_{d,B}\{(r,t),(r',t)\} dt = \sum_{k=1}^R \tau_{dk,\mathcal{R}}^{(1)} \mathbf{v}_{dk}^{(1)}(r) \mathbf{v}_{dk}^{(1)}(r'), \text{ and} \\ (\Sigma_{d,\mathcal{R},W})_{r,r'} &\equiv \int_{\mathcal{T}} K_{d,W}\{(r,t),(r',t)\} dt = \sum_{p=1}^R \tau_{dp,\mathcal{R}}^{(2)} \mathbf{v}_{dp}^{(2)}(r) \mathbf{v}_{dp}^{(2)}(r'),\end{aligned}$$

where $(A)_{r,r'}$ denotes the (r,r') -th element of matrix A , $\mathbf{v}_{dk}^{(1)}(r)$ and $\mathbf{v}_{dp}^{(2)}(r)$ denote level 1 and level 2 eigenvectors, and $\tau_{dk,\mathcal{R}}^{(1)}$, $\tau_{dp,\mathcal{R}}^{(2)}$ are the respective eigenvalues.

Next, the data is projected onto a basis set formed by the product of the one dimensional eigenvectors and eigenfunctions, obtained from the marginal regional and functional covariances,

$$\begin{aligned}Y_{dij}(r,t) &= \mu(t) + \eta_{dj}(r,t) + Z_{di}(r,t) + W_{dij}(r,t) + \epsilon_{dij}(r,t) \\ &= \mu(t) + \eta_{dj}(r,t) + \sum_{k=1}^R \sum_{\ell=1}^{\infty} \zeta_{di,k\ell} \mathbf{v}_{dk}^{(1)}(r) \phi_{d\ell}^{(1)}(t) \\ &\quad + \sum_{p=1}^R \sum_{m=1}^{\infty} \xi_{dij,pm} \mathbf{v}_{dp}^{(2)}(r) \phi_{dm}^{(2)}(t) + \epsilon_{dij}(r,t).\end{aligned}\tag{3.1}$$

In the above decomposition, $\zeta_{di,k\ell} = \langle Z_{di}(r, t), \mathbf{v}_{dk}^{(1)}(r)\phi_{d\ell}^{(1)}(t) \rangle = \sum_{r=1}^R \int_{\mathcal{T}} Z_{di}(r, t) \mathbf{v}_{dk}^{(1)}(r)\phi_{d\ell}^{(1)}(t) dt$ denotes the subject-specific scores defined through the projection of $Z_{di}(r, t)$ onto the level 1 product eigen components $\mathbf{v}_{dk}^{(1)}(r)\phi_{d\ell}^{(1)}(t)$, and $\xi_{dij,pm} = \langle W_{dij}(r, t), \mathbf{v}_{dp}^{(2)}(r)\phi_{dm}^{(2)}(t) \rangle = \sum_{r=1}^R \int_{\mathcal{T}} W_{dij}(r, t) \mathbf{v}_{dp}^{(2)}(r)\phi_{dm}^{(2)}(t) dt$ denotes the subject-repetition-specific scores defined through the projection of $W_{dij}(r, t)$ onto the level 2 product eigen components $\mathbf{v}_{dp}^{(2)}(r)\phi_{dm}^{(2)}(t)$. The subject- and subject-repetition-specific scores are uncorrelated within and across the two levels of the data. The M-HPCA decomposition leads to the following breakdown of the total covariance

$$\begin{aligned} K_{d,\text{Total}}\{(r, t), (r', t')\} &= \sum_{k=1}^R \sum_{\ell=1}^{\infty} \lambda_{d,k\ell}^{(1)} \mathbf{v}_{dk}^{(1)}(r)\phi_{d\ell}^{(1)}(t) \mathbf{v}_{dk}^{(1)}(r')\phi_{d\ell}^{(1)}(t') \\ &\quad + \sum_{p=1}^R \sum_{m=1}^{\infty} \lambda_{d,pm}^{(2)} \mathbf{v}_{dp}^{(2)}(r)\phi_{dm}^{(2)}(t) \mathbf{v}_{dp}^{(2)}(r')\phi_{dm}^{(2)}(t') + \sigma_d^2 \delta\{(r, t), (r', t')\} \\ &= K_{d,B}\{(r, t), (r', t')\} + K_{d,W}\{(r, t), (r', t')\} + \sigma_d^2 \delta\{(r, t), (r', t')\} \end{aligned}$$

where $\lambda_{d,k\ell}^{(1)} = \text{var}(\zeta_{di,k\ell})$, and $\lambda_{d,pm}^{(2)} = \text{var}(\xi_{dij,pm})$, $K_{d,B}\{(r, t), (r', t')\} = \sum_{k=1}^R \sum_{\ell=1}^{\infty} \lambda_{d,k\ell}^{(1)} \mathbf{v}_{dk}^{(1)}(r)\phi_{d\ell}^{(1)}(t) \mathbf{v}_{dk}^{(1)}(r')\phi_{d\ell}^{(1)}(t')$, and $K_{d,W}\{(r, t), (r', t')\} = \sum_{p=1}^R \sum_{m=1}^{\infty} \lambda_{d,pm}^{(2)} \mathbf{v}_{dp}^{(2)}(r)\phi_{dm}^{(2)}(t) \mathbf{v}_{dp}^{(2)}(r')\phi_{dm}^{(2)}(t')$.

The number of product eigen components included in (3.1) is truncated in practice at K , P , L , and M for the between- and within-subject regional and functional eigen components, respectively, using the fraction of variance explained (FVE). We set FVE to 99% in the marginal functional and regional covariance expansions, however reduce the number of resulting product components for inference on condition differentiation through the parametric bootstrap to avoid overfitting after estimation of the eigenscores and the variance components in a mixed effects framework, outlined in more detail in Section 3.2.3.

3.2.2 M-HPCA estimation algorithm

The proposed M-HPCA estimation algorithm is outlined in Algorithm 1. We provide additional details on the first three steps of the algorithm targeting means, covariances and the

product eigen components and defer discussions on details of Step 4, including the mixed effects modeling to target the eigenscores and the variance components and the associated computationally efficient MM algorithm to the next section.

In Step 1, the overall mean, $\mu(t)$, and group-region-repetition-specific means $\eta_{dj}(r, t)$ are targeted using a spline smoother. Due to possible imbalances in sample sizes across diagnostic groups, the smoothed mean functions within each group are averaged in Step 1(a) to target the overall mean function $\mu(t)$. The group-region-repetition-specific means are then targeted in Step 1(b) by smoothing the demeaned data within each group. Next, in Step 2, the raw total, $\widehat{K}_{d, \text{Total}}\{(r, t), (r', t')\}$, and between-subject covariances, $\widehat{K}_{d, B}\{(r, t), (r', t')\}$, are targeted using the empirical covariance estimates within the same subject and repetition, $\text{cov}\{Y_{dij}(r, t), Y_{dij}(r', t')\}$, and within the same subject across different repetitions, $\text{cov}\{Y_{dij_1}(r, t), Y_{dij_2}(r', t')\}$, respectively. The raw within-subject covariance $\widehat{K}_{d, W}\{(r, t), (r', t')\}$ is estimated as the difference between the total and within-subject covariances. The raw marginal functional and regional covariances are targeted by averaging the summed or integrated raw covariances over the regional and functional domains, respectively, followed by smoothing of the raw marginal functional covariances before performing FPCA. Bivariate penalized spline smoothers are utilized in smoothing of the raw marginal functional covariances, where the degree of smoothness is determined via REML. Before smoothing the marginal functional within-subject covariance, the diagonal elements of the raw covariance, which are prone to measurement error effects (similar to the diagonal of the raw total covariance), are removed (Yao *et al.*, 2005; Park & Staicu, 2015). Measurement error variance σ_d^2 is then targeted as the mean difference between the diagonal elements of the smoothed and unsmoothed marginal functional within-subject covariances. Since the diagonal elements of the marginal regional within-subject covariance are similarly effected by measurement error (due to measurement error effects on the diagonal of raw total covariance and within-subject covariance), the estimated σ_d^2 is subtracted from the diagonal entries of the raw marginal regional within-subject covariance. Further details on estimation of the means and covariances are provided in the Supplementary Material Appendix A.

Algorithm 1 M-HPCA Estimation Algorithm

Step 1: Estimation of Mean Functions

- a) Calculate $\widehat{\mu}_d(t)$ by applying a spline smoother to all observed data within group d , $\{t, Y_{dij}(r, t) : i = 1, \dots, n_d; r = 1, \dots, R; t \in \mathcal{T}, j = 1, \dots, J\}$. Obtain the overall mean by $\widehat{\mu}(t) = \{\sum_{d=1}^D \widehat{\mu}_d(t)\}/D$.
- b) Calculate $\widehat{\eta}_{dj}(r, t)$ by applying a spline smoother to all mean-centered data within group d and repetition j and region r , $\{t, Y_{dij}(r, t) - \widehat{\mu}(t) : i = 1, \dots, n_d; t \in \mathcal{T}\}$.

Step 2: Estimation of Covariances and Measurement Error Variance

- a) Raw Covariances: Calculate $\widehat{K}_{d, \text{Total}}\{(r, t), (r', t')\}$, $\widehat{K}_{d, B}\{(r, t), (r', t')\}$ and $\widehat{K}_{d, W}\{(r, t), (r', t')\} = \widehat{K}_{d, \text{Total}}\{(r, t), (r', t')\} - \widehat{K}_{d, B}\{(r, t), (r', t')\}$.
- b) Marginal Covariances: Calculate the raw marginal functional covariances $\widehat{\Sigma}_{d, \mathcal{T}, B}(t, t')$ and $\widehat{\Sigma}_{d, \mathcal{T}, W}(t, t')$ and raw marginal regional covariances $\widehat{\Sigma}_{d, \mathcal{R}, B}$ and $\widehat{\Sigma}_{d, \mathcal{R}, W}$.
- c) Smoothed Covariances:
 - i. Calculate $\widetilde{\Sigma}_{d, \mathcal{T}, B}(t, t')$ and $\widetilde{\Sigma}_{d, \mathcal{T}, W}(t, t')$ by applying a bivariate penalized spline smoother to $\widehat{\Sigma}_{d, \mathcal{T}, B}(t, t')$ and $\widehat{\Sigma}_{d, \mathcal{T}, W}(t, t')$, respectively.
 - ii. Calculate $\widehat{\sigma}_d^2$ as the mean difference between the diagonals of $\widetilde{\Sigma}_{d, \mathcal{T}, W}(t, t')$ and $\widehat{\Sigma}_{d, \mathcal{T}, W}(t, t')$.
 - iii. Calculate $\widetilde{\Sigma}_{d, \mathcal{R}, W}$ by removing the estimated measurement error variance $\widehat{\sigma}_d^2$ from the diagonal entries of the raw marginal regional covariance $\widehat{\Sigma}_{d, \mathcal{R}, W}$.

Step 3: Estimation of Marginal Eigen Components

- a) Employ FPCA on $\widetilde{\Sigma}_{d, \mathcal{T}, B}(t, t')$ and $\widetilde{\Sigma}_{d, \mathcal{T}, W}(t, t')$ to estimate the eigenvalue, eigenfunction pairs, $\{\tau_{d\ell, \mathcal{T}}^{(1)}, \phi_{d\ell}^{(1)}(t) : \ell = 1, \dots, L\}$ and $\{\tau_{dm, \mathcal{T}}^{(2)}, \phi_{dm}^{(2)}(t) : m = 1, \dots, M\}$, respectively.
- b) Employ PCA on $\widetilde{\Sigma}_{d, \mathcal{R}, B}$ and $\widetilde{\Sigma}_{d, \mathcal{R}, W}$ to estimate the eigenvalue, eigenvector pairs $\{\tau_{dk, \mathcal{R}}^{(1)}, v_{dk}^{(1)}(r) : k = 1, \dots, K\}$ and $\{\tau_{dp, \mathcal{R}}^{(2)}, v_{dp}^{(2)}(r) : p = 1, \dots, P\}$, respectively.

Step 4: Estimation of Variance Components and Eigenscores via Mixed Effects Modeling and the proposed MM Algorithm

- a) Calculate $\widehat{\zeta}_{dig}$, $\widehat{\xi}_{dijh}$, $\widehat{\lambda}_{dg}^{(1)}$, $\widehat{\lambda}_{dh}^{(2)}$ and $\widehat{\sigma}_d^2$ by fitting a linear mixed effects model using the proposed MM algorithm.
 - b) Select G' and H' such that $FVE_{d, B}(G'), FVE_{d, W}(H') > 0.8$, $d = 1, \dots, D$, for inference on condition differentiation, via parametric bootstrap.
-

In Step 3, FPCA and PCA of the marginal functional and regional covariances are employed to target the level 1, $\{\widehat{\tau}_{dl,\mathcal{T}}^{(1)}, \widehat{\phi}_{dl}^{(1)}(t)\}$, $\{\widehat{\tau}_{dk,\mathcal{R}}^{(1)}, \widehat{v}_{dk}^{(1)}(r)\}$, and level 2 eigen components, $\{\widehat{\tau}_{dm,\mathcal{T}}^{(2)}, \widehat{\phi}_{dm}^{(2)}(t)\}$, $\{\widehat{\tau}_{dp,\mathcal{R}}^{(2)}, \widehat{v}_{dp}^{(2)}(r)\}$. In order to target the product eigen components, $\widehat{v}_{dk}^{(1)}(r)\widehat{\phi}_{dl}^{(1)}(t)$ and $\widehat{v}_{dp}^{(2)}(r)\widehat{\phi}_{dm}^{(2)}(t)$ in (3.1), the number of marginal eigen components K , L , P , M are selected to explain at least 99% of the variation in the FPCA and PCA decompositions of the marginal functional and regional covariances. The estimated product eigen components are considered fixed in the below described mixed effects formulation (Step 4) targeting the subject- and subject-repetition-specific eigenscores, $\zeta_{di,k\ell}$ and $\xi_{dij,pm}$, their variances $\lambda_{d,k\ell}^{(1)}$ and $\lambda_{d,pm}^{(2)}$ and measurement error variance σ_d^2 .

3.2.3 Mixed effects modeling and the computationally efficient MM algorithm

The mixed effects model which considers the projection of the observed data onto the estimated level 1 and level 2 product eigen components denoted by $\varphi_{dg}^{(1)}(r, t) = v_{dk}^{(1)}(r)\phi_{dl}^{(1)}(t)$, and $\varphi_{dh}^{(2)}(r, t) = v_{dp}^{(2)}(r)\phi_{dm}^{(2)}(t)$, is given by

$$\begin{cases} Y_{dij}(r, t) = \mu(t) + \eta_{dj}(r, t) + \sum_{g=1}^G \zeta_{dig}\varphi_{dg}^{(1)}(r, t) + \sum_{h=1}^H \xi_{dijh}\varphi_{dh}^{(2)}(r, t) + \epsilon_{dij}(r, t) \\ \zeta_{dig} \sim N(0, \lambda_{dg}^{(1)}), \xi_{dijh} \sim N(0, \lambda_{dh}^{(2)}), \epsilon_{dij}(r, t) \sim N(0, \sigma_d^2), \end{cases} \quad (3.2)$$

where $g = (k-1) + K(\ell-1) + 1$ and $h = (p-1) + P(m-1) + 1$ denote the single indices for the $G = KL$ and $H = PM$ total product eigen components at level 1 and 2, respectively, and $\text{var}(\zeta_{dig}) = \lambda_{dg}^{(1)}$, $\text{var}(\xi_{dijh}) = \lambda_{dh}^{(2)}$. In the following developments, we will assume that subjects are observed at a subject-specific set of repetitions (with a total of J_i repetitions for subject i), to accommodate missing data over experimental conditions or visits.

We next introduce the relevant notation to represent the mixed effects model in (3.2) and the details on the proposed computationally efficient MM algorithm in vector form. Let \mathbf{Y}_{di} denote the $TRJ_i \times 1$ vector obtained by stacking $Y_{dij}(r, t) - \widehat{\mu}(t) - \widehat{\eta}_{dj}(r, t)$ over the functional grid ($t = 1, \dots, T$), regions, and repetitions for subject i , $i = 1, \dots, n_d$. Further let $\boldsymbol{\varphi}_{dg}^{(1)} = \{\varphi_{dg}^{(1)}(r_1, t_1), \dots, \varphi_{dg}^{(1)}(r_1, t_T), \dots, \varphi_{dg}^{(1)}(r_R, t_1), \dots, \varphi_{dg}^{(1)}(r_R, t_T)\}^T$ and $\boldsymbol{\varphi}_{dh}^{(2)} =$

$\{\varphi_{dh}^{(2)}(r_1, t_1), \dots, \varphi_{dh}^{(2)}(r_1, t_T), \dots, \varphi_{dh}^{(2)}(r_R, t_1), \dots, \varphi_{dh}^{(2)}(r_R, t_T)\}^T$ denote the g^{th} and h^{th} $TR \times 1$ vectorized level 1 and level 2 product eigen components, respectively. The vector product eigen components are then merged in the $TR \times G$ and $TR \times H$, level 1 and level 2 matrices, $\Phi_d^{(1)} = (\varphi_{d1}^{(1)}, \varphi_{d2}^{(1)}, \dots, \varphi_{dG}^{(1)})$ and $\Phi_d^{(2)} = (\varphi_{d1}^{(2)}, \varphi_{d2}^{(2)}, \dots, \varphi_{dH}^{(2)})$. Let $\Lambda_d^{(1)}$ and $\Lambda_d^{(2)}$ denote the $G \times G$ and $H \times H$ diagonal matrices containing the level 1 and level 2 variance components $\{\lambda_{d1}^{(1)}, \dots, \lambda_{dG}^{(1)}\}$ and $\{\lambda_{d1}^{(2)}, \dots, \lambda_{dH}^{(2)}\}$, on the diagonal. Finally let $\mathbf{1}_{J_i}$ and $\mathbf{1}_{J_i \times J_i}$ denote the $J_i \times 1$ column and $J_i \times J_i$ matrix of ones, respectively, and \mathbf{I}_{J_i} denote the $J_i \times J_i$ identity matrix, with \otimes denoting the Kronecker product.

The mixed effects model in (3.2) is given in vector form as

$$\begin{cases} \mathbf{Y}_{di} = \mathbf{Z}_{di}\boldsymbol{\zeta}_{di} + \mathbf{W}_{di}\boldsymbol{\xi}_{di} + \boldsymbol{\epsilon}_{di} \\ \boldsymbol{\zeta}_{di} \sim \text{MVN}\left(0, \Lambda_d^{(1)}\right), \boldsymbol{\xi}_{di} \sim \text{MVN}\left(0, \mathbf{I}_{J_i} \otimes \Lambda_d^{(2)}\right), \boldsymbol{\epsilon}_{di} \sim \text{MVN}\left(0, \sigma_d^2 \mathbf{I}_{TRJ_i}\right), \end{cases} \quad (3.3)$$

for $i = 1, \dots, n_d$, where $\mathbf{Z}_{di} = \mathbf{1}_{J_i} \otimes \Phi_d^{(1)}$ and $\mathbf{W}_{di} = \mathbf{I}_{J_i} \otimes \Phi_d^{(2)}$ are the $TRJ_i \times G$ and $TRJ_i \times HJ_i$ design matrices, $\boldsymbol{\zeta}_{di} = (\zeta_{di1}, \dots, \zeta_{diG})^T$ and $\boldsymbol{\xi}_{di} = (\xi_{dij_11}, \dots, \xi_{dij_1H}, \dots, \xi_{dij_{J_i}1}, \dots, \xi_{dij_{J_i}H})^T$ denote the $G \times 1$ and $HJ_i \times 1$ random effects vectors, and $\boldsymbol{\epsilon}_{di}$ denotes the $TRJ_i \times 1$ error vector. Estimates of the variance components, $\Lambda_d^{(1)}$, $\Lambda_d^{(2)}$, and σ_d^2 and subject-specific and subject-repetition-specific scores, $\boldsymbol{\zeta}_{di}$ and $\boldsymbol{\xi}_{di}$, are targeted via the proposed computationally efficient MM algorithm.

A major challenge in fitting of the mixed effects model in (3.3) is the size of the data. Even in data sets with small number of subjects, such as the ABC-CT feasibility study, with $n_d = 25$ subjects within diagnostic group (and $J = 2$ visits, $T = 108$ frequencies, $R = 18$ regions), the total number of random effects $G + HJ_i$ can get large ($G \approx 60$ and $H \approx 70$). Standard packages maximizing the log-likelihood directly using Newton's method (e.g. `lme4` via REML) prove to be not feasible and computationally intensive (for a comparison of run times for R packages `lme4` (?) and `mhpca` in simulation studies, see Section 4.4). To address this major challenge, we propose a novel MM algorithm, specifically tailored to M-HPCA. MM algorithms share the same advantages as expectation-maximization (EM) algorithms

such as simplicity and stability while avoiding large matrix manipulations, but MM generally converges faster than EM in the presence of more than two variance components (see Zhou et al. (2019) for more on stability and efficiency of MM algorithms). Instead of targeting to maximize the log-likelihood directly, the proposed MM algorithm maximizes a surrogate function, $f_d(\mathbf{\Lambda}_d^{(1)}, \mathbf{\Lambda}_d^{(2)}, \sigma_d^2 | \mathbf{\Lambda}_d^{(1)(c)}, \mathbf{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)})$ that minorizes the log-likelihood, where superscript (c) denotes the iteration index of the MM. The proposed algorithm iterates between estimation of the eigenscores, ζ_{di}, ξ_{di} , and the variance components, $\mathbf{\Lambda}_d^{(1)}, \mathbf{\Lambda}_d^{(2)}$, and σ_d^2 , where the convergence is determined by the computed log-likelihood at the values of the parameters selected to maximize the surrogate function. The proposed surrogate function is much easier to maximize with respect to the variance components, due to the additive structure of the surrogate where terms involve one variance component at a time. In addition, taking advantage of the weak separability assumption and the independence of random effects at different levels of the data, we achieve additional computational gains by representing the high-dimensional covariance matrices as weighted sums of lower dimensional building blocks. The proposed MM algorithm is outlined on page 69, where we outline details that are specifically tied to computational savings below, and defer other details to Supplementary Material Appendix B.

In order to derive the surrogate function that minorizes the log-likelihood $\ell_d(\mathbf{\Lambda}_d^{(1)}, \mathbf{\Lambda}_d^{(2)}, \sigma_d^2) = \sum_{i=1}^{n_d} \ell_{di}(\mathbf{\Lambda}_d^{(1)}, \mathbf{\Lambda}_d^{(2)}, \sigma_d^2)$, where $\ell_{di}(\mathbf{\Lambda}_d^{(1)}, \mathbf{\Lambda}_d^{(2)}, \sigma_d^2) = -(\log \det \mathbf{\Sigma}_{di})/2 - (\mathbf{Y}_{di}^T \mathbf{\Sigma}_{di}^{-1} \mathbf{Y}_{di})/2$, $i = 1, \dots, n_d$, and $\mathbf{\Sigma}_{di} = \mathbf{Z}_{di} \mathbf{\Lambda}_d^{(1)} \mathbf{Z}_{di}^T + \mathbf{W}_{di} (\mathbf{I}_{J_i} \otimes \mathbf{\Lambda}_d^{(2)}) \mathbf{W}_{di}^T + \sigma_d^2 \mathbf{I}_{TRJ_i}$, consider the following minorizations and majorizations (denoted by \preceq and \succeq). Using the convexity of the mapping $(\mathbf{X}, \mathbf{Y}) \mapsto \mathbf{X}^{-1} \mathbf{Y} \mathbf{X}^{-1}$ and the lemmas given in Zhou et al. (2019),

there is a majorization of $\Sigma_{di}^{(c)} \Sigma_{di}^{-1} \Sigma_{di}^{(c)}$

$$\begin{aligned}
\Sigma_{di}^{(c)} \Sigma_{di}^{-1} \Sigma_{di}^{(c)} &= \left\{ \mathbf{Z}_{di} \Lambda_d^{(1)(c)} \mathbf{Z}_{di}^T + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{(2)(c)} \right) \mathbf{W}_{di}^T + \sigma_d^{2(c)} \mathbf{I}_{TRJ_i} \right\} \\
&\quad \times \left\{ \mathbf{Z}_{di} \Lambda_d^{(1)} \mathbf{Z}_{di}^T + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{(2)} \right) \mathbf{W}_{di}^T + \sigma_d^2 \mathbf{I}_{TRJ_i} \right\}^{-1} \\
&\quad \times \left\{ \mathbf{Z}_{di} \Lambda_d^{(1)(c)} \mathbf{Z}_{di}^T + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{(2)(c)} \right) \mathbf{W}_{di}^T + \sigma_d^{2(c)} \mathbf{I}_{TRJ} \right\} \\
&\asymp \left(\mathbf{Z}_{di} \Lambda_d^{(1)(c)} \mathbf{Z}_{di}^T \right) \left(\mathbf{Z}_{di} \Lambda_d^{(1)} \mathbf{Z}_{di}^T \right)^+ \left(\mathbf{Z}_{di} \Lambda_d^{(1)(c)} \mathbf{Z}_{di}^T \right) \\
&\quad + \left\{ \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{(2)(c)} \right) \mathbf{W}_{di}^T \right\} \left\{ \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{(2)} \right) \mathbf{W}_{di}^T \right\}^+ \\
&\quad \times \left\{ \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{(2)(c)} \right) \mathbf{W}_{di}^T \right\} \\
&\quad + \frac{\sigma_d^{4(c)}}{\sigma_d^2} \mathbf{I}_{TRJ_i}
\end{aligned}$$

which is denoted by \mathbf{U} . This majorization implies $-\Sigma_{di}^{-1} \succeq -\Sigma_{di}^{-1(c)} \mathbf{U} \Sigma_{di}^{-1(c)}$. Additionally, using the convexity of the mapping $\mathbf{A} \mapsto -\log \det \mathbf{A}$, the minorization of $-\log \det \Sigma_{di}$ is given as

$$\begin{aligned}
-\log \det \Sigma_{di} &\geq -\log \det \Sigma_{di}^{(c)} - \text{tr} \left[\Sigma_{di}^{-1(c)} \left(\Sigma_{di} - \Sigma_{di}^{(c)} \right) \right] \\
&= -\log \det \Sigma_{di}^{(c)} - \text{tr} \left(\Sigma_{di}^{-1(c)} \Sigma_{di} \right) - \text{tr} \left(\mathbf{I}_{TRJ_i} \right).
\end{aligned}$$

Hence using the minorizations of Σ_{di}^{-1} and $-\log \det \Sigma_{di}$, the i th subject contribution to the log-likelihood is minorized by the surrogate function

$$\begin{aligned}
f_{di} \left(\Lambda_d^{(1)}, \Lambda_d^{(2)}, \sigma_d^2 \mid \Lambda_d^{(1)(c)}, \Lambda_d^{(2)(c)}, \sigma_d^{2(c)} \right) \\
&= -\frac{1}{2} \text{tr} \left(\Sigma_{di}^{-1(c)} \Sigma_{di} \right) - \frac{1}{2} \mathbf{Y}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{U} \Sigma_{di}^{-1(c)} \mathbf{Y}_{di} + q^{(c)} \\
&= -\frac{1}{2} \left[\text{tr} \left(\mathbf{Z}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{Z}_{di} \Lambda_d^{(1)(c)} \right) + \text{tr} \left\{ \mathbf{W}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{(2)(c)} \right) \right\} + \text{tr} \left(\sigma_d^2 \Sigma_{di}^{-1(c)} \right) \right. \\
&\quad \left. + \zeta_{di}^{(c)T} \Lambda_d^{-1(c)} \zeta_{di}^{(c)} + \xi_{di}^{(c)T} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{-1(2)(c)} \right) \xi_{di}^{(c)} + \frac{\sigma_d^{4(c)}}{\sigma_d^2} \mathbf{Y}_{di}^T \Sigma_{di}^{-2(c)} \mathbf{Y}_{di} \right] + q^{(c)},
\end{aligned}$$

where $q^{(c)}$ is a constant free of the variance components or scores. The overall minorization function across all individuals $f_d(\Lambda_d^{(1)}, \Lambda_d^{(2)}, \sigma_d^2 \mid \Lambda_d^{(1)(c)}, \Lambda_d^{(2)(c)}, \sigma_d^{2(c)}) =$

$\sum_{i=1}^{n_d} f_{di}(\boldsymbol{\Lambda}_d^{(1)}, \boldsymbol{\Lambda}_d^{(2)}, \sigma_d^2 | \boldsymbol{\Lambda}_d^{(1)(c)}, \boldsymbol{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)})$ can be maximized quickly and easily since terms in the minorization function involve variance components one at a time. More specifically the partial derivatives with respect to $\Lambda_{d,gg}^{(1)}$, $\Lambda_{d,hh}^{(2)}$ and σ_d^2 involve only the (first and fourth), (second and fifth) and (third and sixth) terms above, respectively. Derivatives used in deriving the variance components are provided in Supplementary Material Appendix B, where derived estimators are provided in Algorithm 2.

Algorithm 2 outlines the proposed MM algorithm which iterates between estimation of the eigenscores, ζ_{di} , ξ_{di} , and the variance components, $\boldsymbol{\Lambda}_d^{(1)}$, $\boldsymbol{\Lambda}_d^{(2)}$, and σ_d^2 . The algorithm begins with initialization of the variance components, denoted by $\boldsymbol{\Lambda}_d^{(1)(0)}$, $\boldsymbol{\Lambda}_d^{(2)(0)}$, and $\sigma_d^{2(0)}$. The initial value $\sigma_d^{2(0)}$ is set to $\widehat{\sigma}_d^2$ estimated in Step 2 (c)(ii) of the M-HPCA estimation procedure, as the mean difference between the diagonals of the smoothed and unsmoothed marginal functional within-subject covariance. The initializations $\boldsymbol{\Lambda}_d^{(1)(0)}$ and $\boldsymbol{\Lambda}_d^{(2)(0)}$, are set to identity matrices for simplicity which has worked well in practice. The estimates of the subject- and subject-repetition-specific scores are updated using the best linear unbiased predictors (BLUPs) in lines 7 and 8: $\zeta_{di}^{(c)} = \boldsymbol{\Lambda}_d^{(1)(c)} \mathbf{Z}_{di}^T \boldsymbol{\Sigma}_{di}^{-1(c)} \mathbf{Y}_{di}$ and $\xi_{di}^{(c)} = (\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)(c)}) \mathbf{W}_{di}^T \boldsymbol{\Sigma}_{di}^{-1(c)} \mathbf{Y}_{di}$. Next we review representation of the high-dimensional covariance matrices as weighted sums of lower dimensional building blocks and fast computation of log-likelihood at the estimated parameters evaluated for convergence, which all lead to further computational savings.

A major challenge in the algorithm is inverting the $TRJ_i \times TRJ_i$ covariance matrix, $\boldsymbol{\Sigma}_{di}$. We begin by borrowing tools from Crainiceanu et al. (2009) in simplifying $\boldsymbol{\Sigma}_{di}$ in terms of the $TR \times TR$ between and within covariances, denoted by $\mathbf{K}_{d,B} \equiv \boldsymbol{\Phi}_d^{(1)} \boldsymbol{\Lambda}_d^{(1)} \boldsymbol{\Phi}_d^{(1)T}$ and $\mathbf{K}_{d,W} \equiv \boldsymbol{\Phi}_d^{(2)} \boldsymbol{\Lambda}_d^{(2)} \boldsymbol{\Phi}_d^{(2)T}$,

$$\begin{aligned} \boldsymbol{\Sigma}_{di} &= \mathbf{Z}_{di} \boldsymbol{\Lambda}_d^{(1)} \mathbf{Z}_{di}^T + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right) \mathbf{W}_{di}^T + \sigma_d^2 \mathbf{I}_{TRJ_i} \\ &= \mathbf{1}_{J_i \times J_i} \otimes \boldsymbol{\Phi}_d^{(1)} \boldsymbol{\Lambda}_d^{(1)} \boldsymbol{\Phi}_d^{(1)T} + \mathbf{I}_{J_i \times J_i} \otimes \left(\sigma_d^2 \mathbf{I}_{TR} + \boldsymbol{\Phi}_d^{(2)} \boldsymbol{\Lambda}_d^{(2)} \boldsymbol{\Phi}_d^{(2)T} \right) \\ &= \mathbf{1}_{J_i \times J_i} \otimes \mathbf{K}_{d,B} + \mathbf{I}_{J_i \times J_i} \otimes \left(\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W} \right). \end{aligned}$$

Using the fact that the covariance is a sum of two Kronecker products, its inverse is targeted

by

$$\begin{aligned} \boldsymbol{\Sigma}_{di}^{-1} &= \mathbf{I}_{J_i} \otimes (\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W})^{-1} \\ &\quad - \mathbf{1}_{J_i \times J_i} \otimes \left\{ (\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W})^{-1} \mathbf{K}_{d,B} (\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W} + J_i \mathbf{K}_{d,B})^{-1} \right\}. \end{aligned}$$

Note that the above expression requires the inversion of only two $TR \times TR$ matrices: $(\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W})^{-1}$ and $(\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W} + J_i \mathbf{K}_{d,B})^{-1}$. Nevertheless, since the inversions are subject-specific and could still be computationally cumbersome, we achieve further simplifications in the two $TR \times TR$ matrices to be inverted using the Woodbury matrix identity: $(\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{U} (\mathbf{C}^{-1} + \mathbf{VA}^{-1} \mathbf{U})^{-1} \mathbf{VA}^{-1}$:

$$\begin{aligned} (\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W})^{-1} &= \left(\sigma_d^2 \mathbf{I}_{TR} + \boldsymbol{\Phi}_d^{(2)} \boldsymbol{\Lambda}_d^{(2)} \boldsymbol{\Phi}_d^{(2)\top} \right)^{-1} \\ &= \frac{1}{\sigma_d^2} \mathbf{I}_{TR} - \frac{1}{\sigma_d^4} \boldsymbol{\Phi}_d^{(2)} \left\{ \boldsymbol{\Lambda}_d^{-1(2)} + \frac{1}{\sigma_d^2} \boldsymbol{\Phi}_d^{(2)\top} \boldsymbol{\Phi}_d^{(2)} \right\}^{-1} \boldsymbol{\Phi}_d^{(2)\top}, \quad (3.4) \end{aligned}$$

$$\begin{aligned} (\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W} + J_i \mathbf{K}_{d,B})^{-1} &= \left(\mathbf{Q}^{-1} + J_i \boldsymbol{\Phi}_d^{(1)} \boldsymbol{\Lambda}_d^{(1)} \boldsymbol{\Phi}_d^{(1)\top} \right)^{-1} \\ &= \mathbf{Q} - J_i \mathbf{Q} \boldsymbol{\Phi}_d^{(1)} \left\{ \boldsymbol{\Lambda}_d^{-1(1)} + J_i \boldsymbol{\Phi}_d^{(1)\top} \mathbf{Q} \boldsymbol{\Phi}_d^{(1)} \right\}^{-1} \boldsymbol{\Phi}_d^{(1)\top} \mathbf{Q}, \quad (3.5) \end{aligned}$$

where $\mathbf{Q} \equiv (\sigma_d^2 \mathbf{I}_{TR} + \mathbf{K}_{d,W})^{-1}$. Note that expressions in (3.4) and (3.5), which only rely on the inversion of the $H \times H$ matrix $\boldsymbol{\Lambda}_d^{-1(2)} + \boldsymbol{\Phi}_d^{(2)\top} \boldsymbol{\Phi}_d^{(2)} / \sigma^2$ and the $G \times G$ matrix $\boldsymbol{\Lambda}_d^{-1(1)} + J_i \boldsymbol{\Phi}_d^{(1)\top} \mathbf{Q} \boldsymbol{\Phi}_d^{(1)}$, are used in inverting $\boldsymbol{\Sigma}_{di}$ in line 6 of the MM algorithm. In addition, due to weak separability, $\boldsymbol{\Lambda}_d^{(2)}$ and $\boldsymbol{\Lambda}_d^{(1)}$ are both diagonal matrices which are much easier to invert. Further simplifications are achieved for complete data (with no missing repetitions), where the covariance $\boldsymbol{\Sigma}_{di}$ is no longer subject-specific and hence can be inverted outside the subject-specific for loop (starting on line 5), leading additional computational savings. Algorithm run time comparisons between `lme4` and `mhpcr` in Section 4.4 include both incomplete and complete data set-ups.

In the final step of the proposed MM algorithm, the log-likelihood is evaluated at the

estimated model parameters for assessing convergence. Additional computational savings are targeted in evaluation of the log-likelihood, more specifically in evaluation of the determinant of Σ_{di} . Details are deferred to Supplementary Material Appendix B.2. Percent change in the log-likelihood is considered as the criterion of convergence in the proposed MM algorithm (line 17). In applications, δ is set to 0.001, corresponding to a 0.1% change in the log-likelihood, which has worked well in practice. Finally, weak separability is tested via a likelihood ratio test for the random effects correlation structure of the subject- and subject-region-specific eigenscores, $\zeta_{di,kl}$ and $\xi_{dij,pm}$, in the proposed mixed effects modeling framework. Weak separability corresponds to an independent/diagonal correlation structure on the eigenscores (Lynch & Chen, 2018). Hence we propose to test the independent correlation structure against the alternative of a compound symmetric correlation structure for the eigenscores. The compound-symmetric covariance has also been used to test for weak separability in Scheffler et al. (2020) and is preferred over a completely unstructured covariance in order to maximize the stability of the estimates targeted since the number of parameters in the unstructured covariance would increase rapidly with the number of product components. Details of the estimation procedure under the compound symmetric correlation structure of the scores are deferred to Supplementary Materials Appendix B.3. Results from the likelihood ratio test for weak separability are reported in data applications of Section 4.3, along with visual assessments of the similarity of the estimated eigenvectors and eigenfunctions across fixed values in the regional and functional domains.

One advantage of the proposed multilevel decomposition M-HPCA is the ability to compare the variability explained at each level of the data (subject vs. repetition) across the functional and regional dimensions. This comparison leads to a new concept of intraclass correlation for multilevel region-referenced functional data (M-HPCA ICC). M-HPCA ICC is defined as the proportion of the total variation in the multidimensional signal that is explained at the subject level, $\rho_{dW} = \sum_{g=1}^{\infty} \lambda_{dg}^{(1)} / (\sum_{g=1}^{\infty} \lambda_{dg}^{(1)} + \sum_{h=1}^{\infty} \lambda_{dh}^{(2)})$. A higher M-HPCA ICC signals that more of the total variation is explained by heterogeneity at the subject level, implying that repeatedly observed region-referenced functional data are similar within

Algorithm 2 MM algorithm

Input: $\mathbf{Y}_{di}, \mathbf{Z}_{di}, \mathbf{W}_{di}, i = 1, \dots, n_d$

Output: MLE $\hat{\zeta}_{di}, \hat{\xi}_{di}, \hat{\Lambda}_d^{(1)}, \hat{\Lambda}_d^{(2)}, \hat{\sigma}_d^2, i = 1, \dots, n_d$

- 1: Initialize $\Lambda_d^{(1)(0)}$ and $\Lambda_d^{(2)(0)}$ as identity matrices, and $\sigma_d^{2(0)}$ as $\hat{\sigma}_d^2$
 - 2: **repeat**
 - 3: $\mathbf{K}_{d,B}^{(c)} \leftarrow \Phi_d^{(1)} \Lambda_d^{(1)} \Phi_d^{(1)T}$
 - 4: $\mathbf{K}_{d,W}^{(c)} \leftarrow \Phi_d^{(2)} \Lambda_d^{(2)} \Phi_d^{(2)T}$
 - 5: **for** $i = 1, \dots, n_d$ **do**
 - 6: $\Sigma_{di}^{-1(c)} \leftarrow \mathbf{I}_{J_i} \otimes (\sigma_d^{2(c)} \mathbf{I}_{TR} + \mathbf{K}_{d,W}^{(c)})^{-1} - \mathbf{1}_{J_i \times J_i} \otimes \{(\sigma_d^{2(c)} \mathbf{I}_{TR} + \mathbf{K}_{d,W}^{(c)})^{-1} \mathbf{K}_{d,B}^{(c)} (J_i \mathbf{K}_{d,B}^{(c)} + \sigma_d^{2(c)} \mathbf{I}_{TR} + \mathbf{K}_{d,W}^{(c)})^{-1}\}^*$
 - 7: $\zeta_{di}^{(c)} \leftarrow \Lambda_d^{(1)(c)} \mathbf{Z}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{Y}_{di}$
 - 8: $\xi_{di}^{(c)} \leftarrow (\mathbf{I}_J \otimes \Lambda_d^{(2)(c)}) \mathbf{W}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{Y}_{di}$
 - 9: **end for**
 - 10: **for** $g = 1, \dots, G$ **do**
 - 11: $\Lambda_{d,gg}^{(1)(c+1)} \leftarrow \sqrt{\left\{ \sum_{i=1}^{n_d} \zeta_{i,g}^{2(c)} \right\} / \left\{ \sum_{i=1}^{n_d} \left(\mathbf{Z}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{Z}_{di} \right)_{gg} \right\}}$
 - 12: **end for**
 - 13: **for** $h = 1, \dots, H$ **do**
 - 14: $\Lambda_{d,hh}^{(2)(c+1)} \leftarrow \sqrt{\left\{ \sum_{i=1}^{n_d} \sum_{j=1}^{J_i} \xi_{i,h+(j-1)H}^{2(c)} \right\} / \left\{ \sum_{i=1}^{n_d} \sum_{j=1}^{J_i} \left(\mathbf{W}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{W}_{di} \right)_{(h+(j-1)H)(h+(j-1)H)} \right\}}$
 - 15: **end for**
 - 16: $\sigma_d^{2(c+1)} \leftarrow \sigma_d^{2(c)} \sqrt{\left\{ \sum_{i=1}^{n_d} \mathbf{Y}_{di} \Sigma_{di}^{-2(c)} \mathbf{Y}_{di} \right\} / \left\{ \sum_{i=1}^{n_d} \text{tr} \left(\Sigma_{di}^{-1(c)} \right) \right\}}$
 - 17: **until** $\left| \ell_d^{(c+1)}(\Lambda_d^{(1)}, \Lambda_d^{(2)}, \sigma_d^2) - \ell_d^{(c)}(\Lambda_d^{(1)}, \Lambda_d^{(2)}, \sigma_d^2) \right| < \delta \left| \ell_d^{(c)}(\Lambda_d^{(1)}, \Lambda_d^{(2)}, \sigma_d^2) \right|$
-

* Using the simplifications made in Equations (3.4) and (3.5).

subjects. Utilizing the G and H variance components estimated from the mixed effects modeling at the subject and subject-repetition levels, respectively, we target ρ_{dW} by

$$\hat{\rho}_{dW} = \frac{\sum_{g=1}^G \hat{\lambda}_{dg}^{(1)}}{\sum_{g=1}^G \hat{\lambda}_{dg}^{(1)} + \sum_{h=1}^H \hat{\lambda}_{dh}^{(2)}}.$$

In our second application to the ABC-CT feasibility data (presented in the next section), M-HPCA ICC is utilized to quantify the short-term reliability of the PSD across visits that are approximately a week apart. M-HPCA ICC, which is a single number summarizing the reliability in the high-dimensional EEG signal across frequencies and scalp regions, is a useful measure for researchers that want to assess reliability of PSD in general as a biomarker for ASD. Note that the proposed M-HPCA ICC is different from the functional ICC (fICC) proposed by Di et al. (2009) based on MFPCA, since fICC captures similarity across repetitions of functional data but does not consider a regional dimension. Hence M-HPCA ICC summarizes, in a single measure, the agreement not only across the functional dimension but also the regional one. We provide a percentile bootstrap confidence interval for M-HPCA ICC for inference, formed by resampling from subjects (see Section 3.3.2.2 for more details).

We also propose a parametric bootstrap based inference for condition differentiation within a group or across groups. In our applications to the audio odd-ball paradigm, the main interest is on group comparisons of the brain response to the two experimental conditions (match vs. mismatch). While a larger number of product components (G and H at the subject and repetition levels, respectively) are used in the initial mixed effects fitting, leading to a more accurate assessment of total variation and hence M-HPCA ICC, for the parametric bootstrap proposed on condition differentiation, a more parsimonious model is preferred for resampling. The leading product components G' and H' that explain at least 80% of variation at the subject- and subject-repetition levels are retained for drawing parametric bootstrap samples under the null, i.e. $FVE_{dB}(G') > 0.8$ and $FVE_{dW}(H') > 0.8$ for $d = 1, \dots, D$, where $FVE_{dB}(G') = (\sum_{g=1}^{G'} \hat{\lambda}_{dg}^{(1)}) / (\sum_{g=1}^G \hat{\lambda}_{dg}^{(1)})$ and $FVE_{dW}(H') = (\sum_{h=1}^{H'} \hat{\lambda}_{dh}^{(2)}) / (\sum_{h=1}^H \hat{\lambda}_{dh}^{(2)})$. Note that a different number of product components may be selected for each level and group. Once the number of product components needed is determined, the mixed effects

model is fit for a second time and the variance components estimated from the second fit are utilized for the parametric bootstrap (see Section 3.3.1.2 and Supplementary Material Appendix C for more details).

3.3 Applications

3.3.1 Audio odd-ball paradigm

3.3.1.1 Data structure and methods

The data for our first application was collected by our collaborators, at UCLA, on 31 children aged between 5 and 11 years old, to study neural mechanisms underlying language impairment in children with ASD (DiStefano et al., 2019). Some EEG paradigms don't require a behavioral response, which was instrumental in studying neural mechanisms among the sample of 14 typically developing (TD), 10 verbal ASD (vASD), and 7 minimally verbal ASD (mvASD) children. ASD diagnoses was made prior to enrollment and confirmed by the research team using the Autism Diagnostic Observation Schedule (ADOS) and Social Communication Questionnaire (SCQ). The paradigm was a picture-word matching paradigm (Fig 3.1a), where participants were presented a picture on a white background and then heard a spoken word that either matched (*match* condition) or did not match the picture (*mismatch* condition). The picture appeared for 2000ms and the auditory stimulus was played 1000ms after the picture was presented. The word stimuli were chosen from 60 basic nouns taken from the MacArthur-Bates Communicative Development Inventories-2nd edition. Each word appeared twice – once in the match condition and once in the mismatch condition. There were four blocks of 30 trials each, totaling 6 minutes (trials where participants were not looking at the screen were removed). Data collection, cleaning, and pre-processing steps are described further in DiStefano et al. (2019).

Previous analysis of the data by DiStefano et al. (2019) highlighted group differences in the ERP characteristic components N400 and a wide negativity from 600-900ms referred to

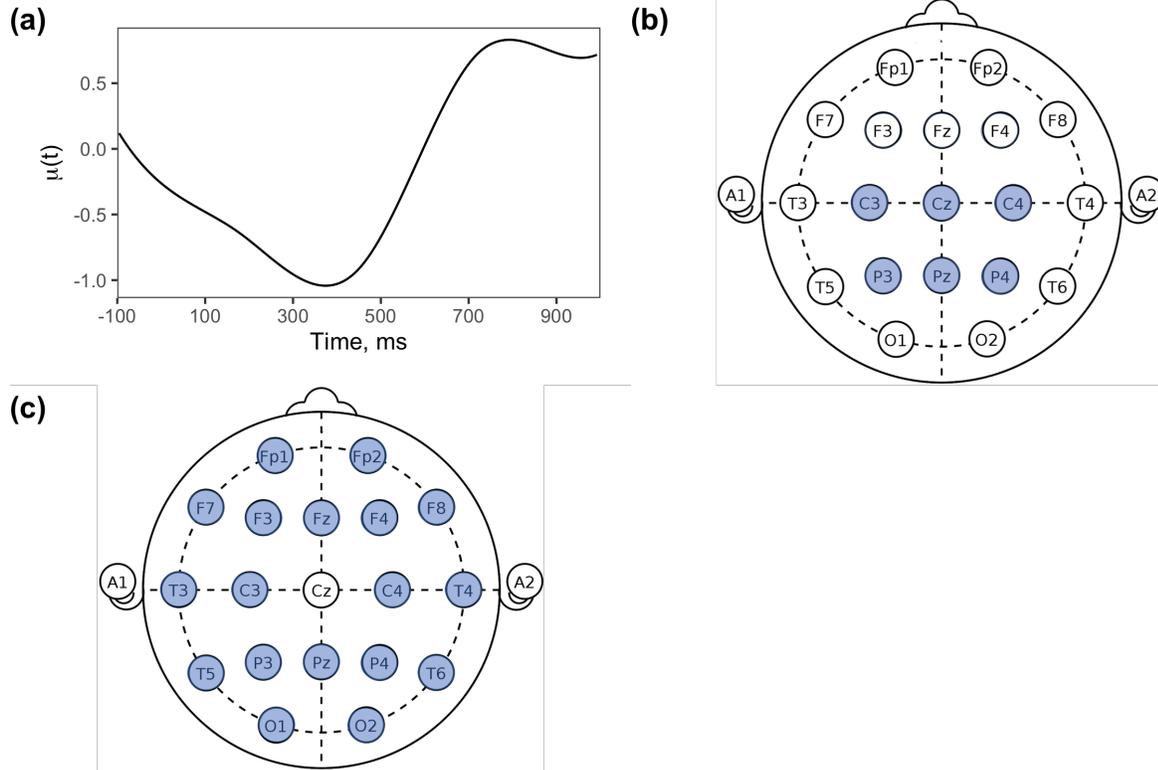


Figure 3.2: (a) The estimated overall mean function $\mu(t)$ in the audio odd-ball paradigm. (b) Positions of the 6 electrodes used in the analysis of the audio odd-ball paradigm highlighted. (c) Positions of the 18 electrodes used in the analysis of the ABC-CT feasibility study data highlighted.

as the late negative component (LNC). The N400 component is observed in the estimated overall mean function (Figure 3.2a). The N400 is related to lexical-semantic processing and LNC is related to cognitive integration following semantic processing (DiStefano *et al.*, 2019). DiStefano *et al.* (2019) reported a greater negativity in N400 in the mismatch condition compared to the match condition across all diagnostic groups and found a greater negativity in LNC in the mismatch condition compared to the match condition in region-specific comparisons.

3.3.1.2 Data analysis results

We consider a subset of 6 electrodes (C3, Cz, C4, P3, Pz, P4) in the regional dimension (Figure 2b) where weak separability was not rejected. Representative electrodes are chosen

by generating clusters of the original 128 electrodes in the (left, middle, and right) central (C3, Cz, C4) and parietal (P3, Pz, P4) regions. The functional dimension includes 137 equally spaced time points between -96ms and 992ms. In addition, the two experimental conditions (match vs. mismatch) constitute the multilevel structure. For visual assessment of weak separability, we compared the first two leading eigenfunctions along the functional dimension at the participant/subject and condition/repetition levels across fixed slices of the regional dimension (Figures B.5.4 and B.5.5) and the first leading eigenvector along the regional dimension at the participant and condition levels across fixed slices of the functional dimension (Figure B.5.6). Directions of variation along one dimension for fixed slices along the second dimension show sufficient similarity at both the participant and condition levels. In addition, the likelihood ratio test for assessing the weak separability assumption on the (between- and within-subject) covariances was not rejected with ($p = .20$, $p = .08$) in TD, ($p = .24$, $p = .4$) in vASD and ($p = .75$, $p = .38$) in mvASD groups. Recall that the number of eigencomponents is selected to explain 99% of total variation in eigendecompositions of the marginal functional and regional covariances. The MM algorithm has a total number of product components equal to the product of the number of marginal regional and functional eigencomponents retained, separately at the participant and condition levels. Once the variance components are estimated by the MM, enough product components are retained to explain 80% of the total variation at each level, in order to draw group-level inference on condition differentiation via parametric bootstrap. The total computing time of M-HPCA including the proposed MM algorithm is approximately 2 minutes on a 3.7 GHz 6-Core Intel Core i5 operating R 4.0.3 (R Core Team, 2021).

Even though a larger number of components are retained in the eigendecompositions of the marginal regional covariance (5 participant-level eigenvectors are retained for the TD and vASD groups, with 4 participant-level eigenvectors retained for the mvASD group, and 3 condition-level eigenvectors are retained for all groups), we interpret the leading eigenvector estimated at the participant and condition levels, explaining most of the variation. The leading participant- and condition-level eigenvectors ($\nu_{d1}^{(1)}(r)$, $\nu_{d1}^{(2)}(r)$), plotted in Figures 3.3a

and Figure 3.3b, respectively, account for (64.8%, 51.4%) of the total marginal variation in TD, (67.1%, 53.4%) in vASD and (47.4%, 53.1%) in mvASD. The estimated eigenvectors at both levels represent a contrast between electrodes from the anterior, central and posterior regions of the scalp for all the groups and hence connects to how dipoles produce ERP signal oriented in opposite directions.

For the functional dimension, we discuss interpretations of the leading two eigenfunctions at both the participant and condition levels, explaining the majority of the associated marginal functional variation. (The total number of eigencomponents retained for the mixed effects modeling via the proposed MM algorithm are (3, 5), (3, 5) and (4, 5) for TD, vASD and mvASD groups at the (participant, condition) levels, respectively.) The leading two eigenfunctions explain (64.8%, 21.8%), (67.1%, 18.1%) and (47.5%, 37%) of the total marginal variation for the TD, vASD and mvASD groups at the participant-level and (51.4%, 30.2%), (53.4%, 30.8%) and (53.1%, 31.2%) of the total marginal variation for the TD, vASD and mvASD groups at the condition-level. While the leading participant- and condition-level eigenfunctions in the TD and vASD groups both signal variability in N400 and LNC components (Figure 3.3c, e), the second leading eigenfunction at the participant-level highlights a contrast between the P200 and LNC components (Figure 3.3d) for the TD group and N400 and LNC for the vASD group and the second leading eigenfunction at the condition-level highlights a contrast between the N400 and LNC components (Figure 3.3f) for the TD group and P200 and LNC components for vASD group. The mvASD group portrays different directions of variation than the other two groups at both the participant and condition levels. At the participant-level, most of the variation is observed in the late LNC component in mvASD (Figure 3.3c), followed by additional variation in N400 (Figure 3.3d), while at the condition-level most of the variation is observed in the contrast between P200 and LNC components (Figure 3.3e) followed by additional variation around N400 and LNC (Figure 3.3f). Hence while most of the variation across subjects is attributable to the late LNC component followed by the N400 in the mvASD group, most of the variation across subjects is equally attributable to both the N400 and LNC components in the TD and vASD groups.

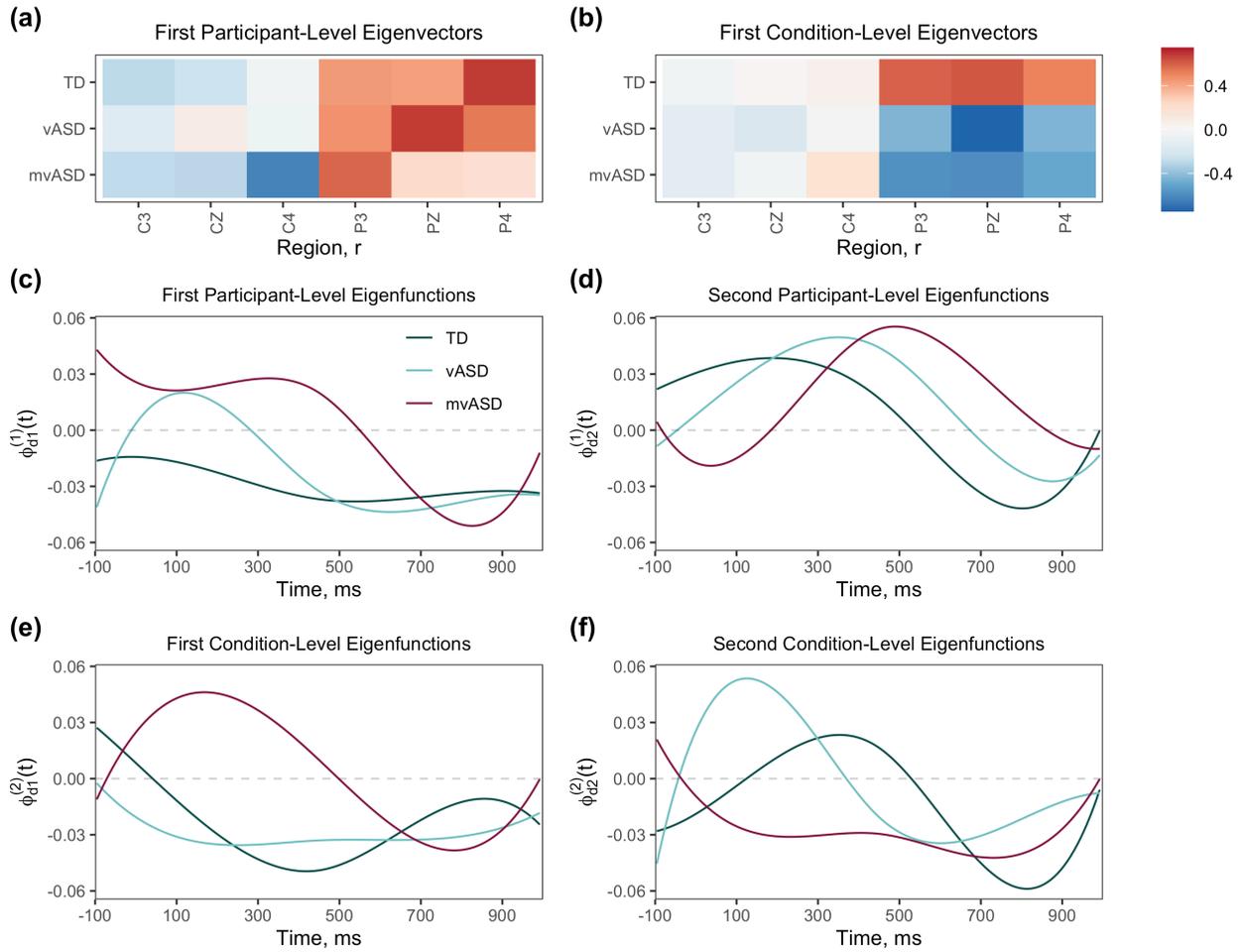


Figure 3.3: The estimated leading participant-level eigenvectors $\nu_{dl}^{(1)}(r)$ (given in (a)) and condition-level eigenvector $\nu_{dl}^{(2)}$ (given in (b)) represent a contrast between electrodes from the anterior and posterior regions of the scalp for all groups. The estimated leading two eigenfunctions for the participant-level variation $\phi_{dl}^{(1)}$ (middle row) and condition-level variation $\phi_{dl}^{(2)}$ (bottom row) for each diagnostic group.

For variation across conditions, most of the variation is in the contrast between P200 and LNC in mvASD, while TD and vASD groups show again roughly equal variability in both the N400 and LNC components. The increased variation observed in the LNC component at both levels for the mvASD group highlights the heterogeneity in this group in cognitive integration, following semantic processing.

In order to draw group-level inference via parametric bootstrap, 4, 6, and 5 participant-level product components are retained in the TD, vASD, and mvASD groups, respectively, while 9, 7, and 7 condition-level product components are retained in the TD, vASD, and mvASD groups, respectively, explaining at least 80% of the variation at each level. The first test is for the within group condition differentiation, i.e. $H_0 : \eta_{d_j}(r, t) = \eta_d(r, t)$ for $j = 1, 2$, where the two conditions share a common mean in region r . The proposed test statistic, denoted by $T_r = \left[\sum_{j=1}^2 \int \{ \hat{\eta}_{d_j}(r, t) - \hat{\eta}_d(r, t) \}^2 dt \right]^{1/2}$, with $\hat{\eta}_d(r, t)$ representing the pointwise average of $\hat{\eta}_{d_j}(r, t)$ for $j = 1, 2$, targets the distance between the group-repetition-region-specific means and their pointwise average. Parametric bootstrap samples are drawn under the null hypothesis using the estimated product components and generated scores and errors sampled from normal distributions with the variance components estimated by the fitted MM algorithm. Test statistics computed in the bootstrap samples estimate the null distribution of the test statistic T_r , leading to a p-value. A similar hypothesis test is also proposed for testing the condition differentiation across groups, i.e. $H_0 : \eta_{d_{1j_1}}(r, t) - \eta_{d_{1j_2}}(r, t) = \eta_{d_{2j_1}}(r, t) - \eta_{d_{2j_2}}(r, t)$ in region r , also via parametric bootstrap. Details on both tests are provided in Supplementary Materials Appendix C.

While significant group differences are not detected in condition differentiation, there are significant region-specific within group condition differentiations detected. While condition differentiation does not approach significance in the TD group, the estimated condition-specific mean functions from almost all the regions show a condition difference in the TD group, mostly due to the LNC component (see Figures 3.4a-b, where LNC is reflected as a positivity in the posterior region due to the dipole effect). LNC in the mismatch condition has a greater amplitude compared to the match condition (similar to findings of DiStefano et

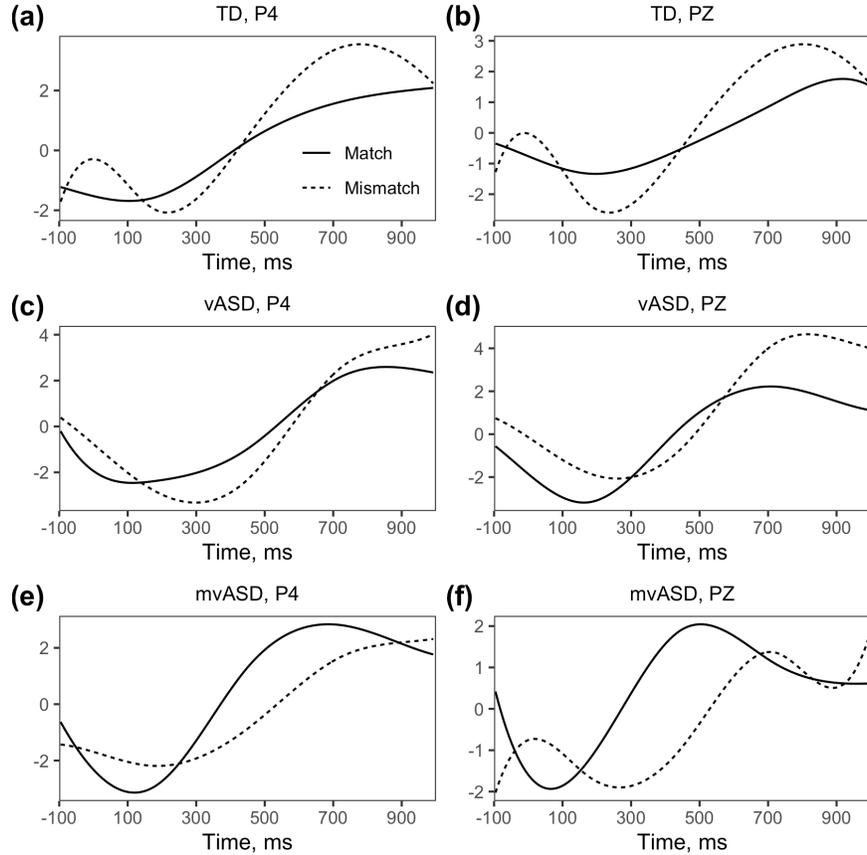


Figure 3.4: The estimated group, region and condition-specific mean functions $\eta_{dj}(r, t)$ in the audio odd-ball paradigm.

al. DiStefano *et al.* (2019)). The condition differentiation is strongest in the mvASD group at P4 ($p = .04$) and Pz ($p = .02$) mostly due to the N400 (see Figure 3.4e-f), with a greater amplitude in the match than the mismatch. Hence while the TD and vASD groups shows trending condition differentiation in the N400 and LNC components consistently across the regions, signaling semantic processing and cognitive integration, differentiation detected in mvASD is in the opposite direction (with greater amplitude in the match condition). Note that while previous findings of DiStefano *et al.* (2019) characterize condition differences in only the amplitude of the expected characteristic components N400 and LNC, the proposed M-HPCA compares the entire ERP waveform across conditions, allowing for more granular findings in the morphology of the group mean differences and directions of variation in the entire ERP waveform as plotted in Figures 3.3 and 3.4.

3.3.2 Day-to-day test-retest reliability of PSD

3.3.2.1 Data structure and methods

Our second application is a resting-state paradigm where the functional observations are the PSD in the frequency domain. The frequency domain ranges from 0 to 55 Hz, containing the frequency bands delta (0-4 Hz), theta (4-6 Hz), low alpha (6-9 Hz), high alpha (9-13 Hz), beta (13-30 Hz), and gamma (30-55 Hz). The main objective is to study the short-term reliability of the entire PSD waveform using EEG data from 25 TD and 22 youth diagnosed by ASD (aged 6 to 11 years old) in the feasibility study of ABC-CT, collected on two days, a median of 6 days apart (Levin et al., 2020). The participants were shown silent, chromatic digital videos as part of a resting EEG eyes open paradigm. On each day, six 30 second videos were shown in three blocks of 1 minute, for a total of 3 minutes. The PSD was calculated by segmenting the data into two-second intervals, and averaging the PSDs of artifact-free segments for each participant and electrode. The relative PSD was calculated by dividing the absolute power by total power, estimated by the integral of the PSD over the entire frequency domain. PSD from $R = 18$ electrodes, chosen from the 10-20 system, was analyzed to generate a robust signal decomposition (Figure 3.2c). For more details on the study, including data collection, cleaning, and pre-processing steps, please see Levin et al. (2020).

3.3.2.2 Data analysis results

The weak separability assumption on the between- and within-subject covariances is checked by comparing the first two leading eigenfunctions along the functional dimension at the participant/subject and day/repetition levels across fixed slices of the regional dimension (Figure B.5.7) and the first leading eigenvector along the regional dimension at the participant and day levels across fixed slices of the functional dimension (Figure B.5.8). Directions of variation along one dimension for fixed slices along the second dimension show sufficient similarity at both the participant and day levels, signaling no violation of the weak separability assump-

tion. In addition, the likelihood ratio test for assessing the weak separability assumption on the (between- and within-subject) covariances fails to reject the null hypothesis of the diagonal correlation structure with ($p = .50$, $p = .76$) in TD and ($p = .88$, $p = .14$) in ASD groups. As in our first data example, the number of eigencomponents is selected to explain 99% of total variation in eigendecompositions of the marginal functional and regional covariances. The variance components corresponding to product eigenscores are estimated by the proposed MM algorithm and bootstrap CIs (by resampling from subjects with replacement) are formed for M-HPCA ICC ρ_{dW} , quantifying the group-specific day-to-day reliability of the entire PSD. We highlight interpretations of the leading eigenvector of the marginal covariance in the regional dimension and leading two eigenfunctions of the marginal covariance in the functional dimension, explaining most of the variation at both the participant and day levels. The total computing time of M-HPCA including the proposed MM algorithm is approximately 5 minutes on a 3.7 GHz 6-Core Intel Core i5 operating R 4.0.3 (R Core Team, 2021).

The leading eigenvectors of the marginal regional covariance at both the participant and day levels are plotted in Figure 3.5. Eigenvectors from both levels signal constant variation in PSD across the entire scalp (explaining 82.4% and 80.8% of the marginal variation in TD and ASD at the participant-level, and 38.6% and 48.3% in TD and ASD at the day-level). Functional ICCs ($.742$, $.730$) in ASD and (0.635 , 0.685) in TD for Day 1 and Day 2), estimated by MFPCA at each visit to compare variability across subjects and regions, also signal good agreement of PSD across the scalp (i.e. more of the total variation is explained at the subject level than the regional level), consistent with the constant variation observed in the relatively flat marginal regional eigenvectors. In the functional dimension, the leading participant-level eigenfunctions show that most of the variation across subjects is observed in the alpha peak amplitude in both groups (explaining 58.2% and 47.2% of the participant-level marginal functional variance in TD and ASD) (Figure 3.5c). Note that the leading participant-level product eigencomponent (obtained as the product of the mostly constant leading marginal eigenvector from the regional dimension and the leading

marginal eigenfunction from the functional dimensions) can be interpreted as the alpha peak amplitude variation across subjects in PSDs averaged across the entire scalp and explain 28% and 20% of the total combined variation across both levels of the data as estimated by the proposed MM. The second leading participant-level eigenfunction signals variation in location of the dominant peak across subjects (explaining 24.6% and 24.7% of the participant-level marginal functional variance in TD and ASD) (Figure 3.5d). While the dominant peak lie in frequencies in low and high alpha bands in the TD group, the dominant peak is observed in the high alpha and beta bands in the ASD group. The leading second product eigencomponent which can be interpreted as the variation in the dominant peak location across subjects in PSDs averaged across the entire scalp, explain 16% and 15% of the total combined variation across both levels of the data. The components that explain the most amount of variation are reversed at the day-level, where most of the variation is explained by location of the dominant peak, followed by a second component related to variation in the amplitude of the dominant peak. More specifically, the leading day-level eigenfunctions signal the variation in the dominant peak location within alpha to beta bands in ASD, within theta to alpha in TD (explaining 53.3% and 48.3% of the marginal variation in TD and ASD) (Figure 3.5e), while the second leading day-level eigenfunctions signal peak alpha amplitude variation in both groups (explaining 31.2% and 30.2% of the marginal functional variation in TD and ASD) (Figure 3.5f). The associated day-level product eigencomponent which can be interpreted as the variation in the dominant peak location across subjects in PSDs averaged across the entire scalp, explain 5% and 3% of the total combined variation across both levels of the data.

Note that percentages of total variation explained by the product components reported in Figure 3.5 are much higher for participant-level product components compared to day-level product components, showing that the estimated M-HPCA ICC, calculated as the ratio of the variation explained at the participant-level to total variation will also be high. This signals good agreement in within subject day-to-day PSD, where most of the variation is due to differences across subjects, not differences across days. Consistent with this intuition the

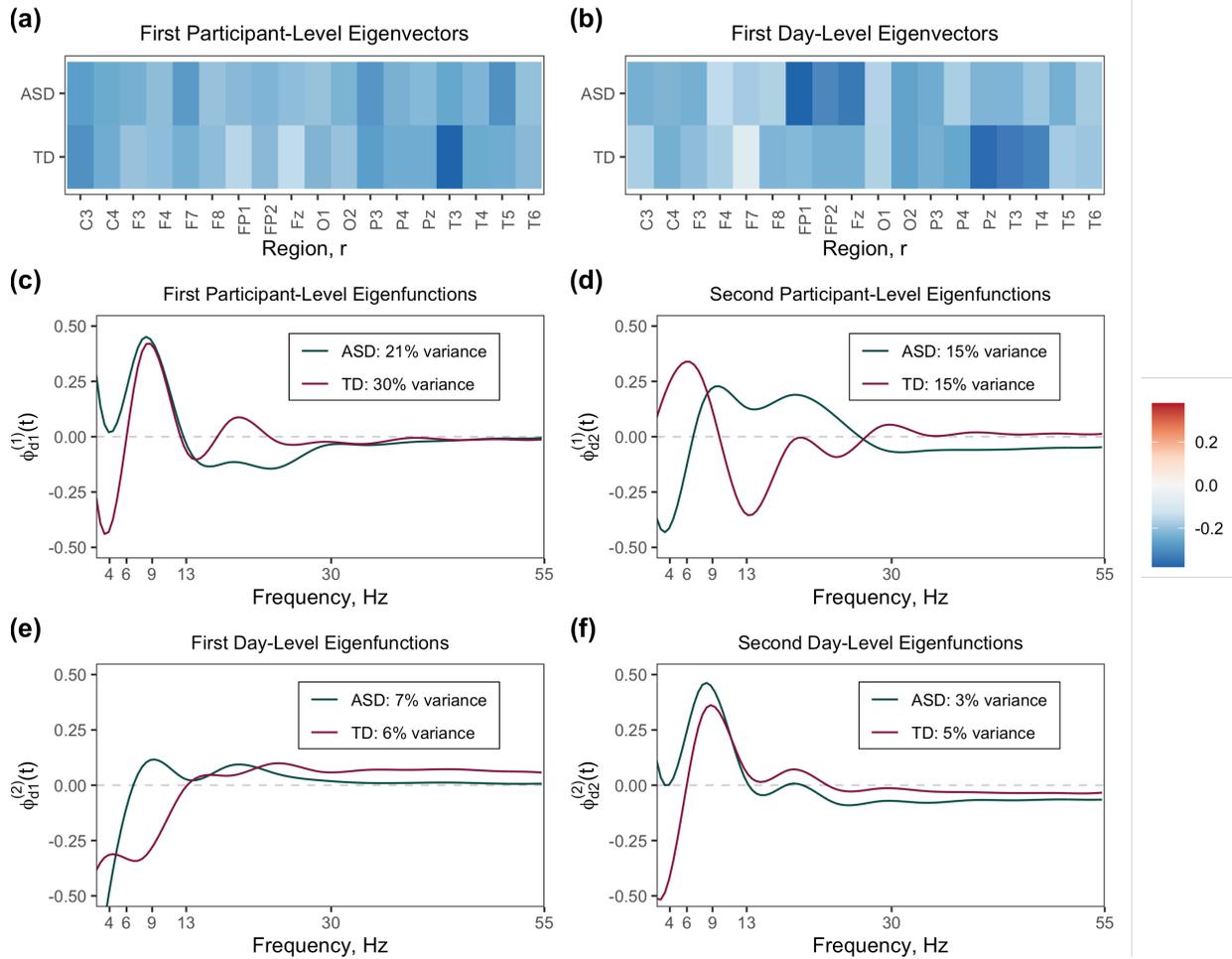


Figure 3.5: The estimated leading (a) participant-level eigenvector $\nu_{dl}^{(1)}(r)$ and (b) day-level eigenvector $\nu_{dl}^{(2)}(r)$. The estimated leading two eigenfunctions for the participant-level variation $\phi_{dl}^{(1)}$ (middle row) and day-level variation $\phi_{dm}^{(2)}$ (bottom row) for each diagnostic group.

M-HPCA ICC is estimated to be 0.673 [95% CI (0.626, 0.793)] for ASD and 0.656 for TD [95% CI (0.639, 0.776)]. Bootstrap percentile confidence intervals are formed based on 200 bootstrap data sets sampled with replacement. The reported M-HPCA ICC values are based on estimated variance components (via MM) for the 54 and 77 total product components at the participant-level for TD and ASD groups, respectively, and 98 total product components at the day-level for both groups.

Obtained results are consistent with findings of Levin et al. (2020) who reported ICCs for scalp averaged PSDs across days within subjects using MFPCA of Di et al. (2009). Authors also reported on alpha peak amplitude variations that explain most of the variation at the participant level, followed by variation in location in the dominant peak, resulting in high functional ICC estimates for both groups. The dominant direction of variation is reported as peak frequency (peak location), followed by peak amplitude, at the day-level, similar to our findings. While the analysis of Levin et al. (2020) accounts for the multilevel structure of the data (with participant and day levels), it ignores the regional dimension and variation across electrodes. By modeling data also along the regional dimension, M-HPCA provides independent support for the previous study of Levin et al. (2020) on the scalp-wide average of PSD. The leading eigenvectors of the marginal regional covariance are found to be quite flat, signaling relatively constant variation in PSDs across the scalp, which justifies the simplification in the data through collapsing of the regional dimension via the scalp average. The work of Levin et al. (2020) also reports that data at the peripheral electrodes (FP1, FP2, T3, T4, O1, O2) would be expected to have the most variation, which we also observed in the marginal eigenvectors plotted in Figure 3.5. A sensitivity analysis which estimated M-HPCA ICC without the peripheral electrodes yields even higher ICC values for both groups.

3.4 Simulation Study

The finite sample performance of the proposed M-HPCA, including the bootstrap inference proposed for M-HPCA ICC and condition differentiation, is assessed at a total of eight

simulation set-ups: two sample sizes ($n_d = 15$ and $n_d = 50$), two noise levels ($\sigma_d^2 = 0.25$ and $\sigma_d^2 = 1$), and two sparsity levels (dense and sparse; sparsity induced at the repetition-level). The data generation model with two repetitions, nine regions and two marginal between- and within-subject eigenfunctions and eigenvectors, totaling four product components at each level of the data is given in detail in Supplementary Materials Appendix B.4. In order to mimic missing repetitions in EEG data, sparsity is induced with 30% of subjects randomly missing one of the two repetitions. Marginal eigencomponents are chosen to explain at least 99% of the variation in the marginal covariance decompositions. Bandwidths used in smoothing of the means and functional covariances are selected by REML. In assessing the performance of the M-HPCA algorithm in targeting model components, we utilize relative squared error (RSE) which is given by $\|\widehat{f}(t) - f(t)\|^2/\|f(t)\|^2$, $\|\widehat{f}(r) - f(r)\|^2/\|f(r)\|^2$ and $\|\widehat{f}(r, t) - f(r, t)\|^2/\|f(r, t)\|^2$, for the one-dimensional functional, vector components and the two-dimensional components, respectively, where $\|f(r)\|^2 = \sum_{r=1}^R f^2(r)$, $\|f(t)\|^2 = \int f^2(t)dt$, and $\|f(r, t)\|^2 = \sum_{r=1}^R \int f^2(r, t)dt$. Estimation of the scalar components is assessed using normalized mean squared error (MSE), $MSE(\theta) = (\widehat{\theta} - \theta)^2/\theta^2$. Results are reported based on 200 Monte Carlo runs.

The estimated mean functions and marginal eigenvectors and eigenfunctions from the sparse simulation set-up with $n_d = 15$ and $\sigma_d^2 = 1$ (high noise) are displayed in Supplementary Material Figure B.5.1 and B.5.2, respectively. The estimates from Monte Carlo runs with the 10th, 50th, and 90th percentile RSE are shown overlaying the true functions. The figures show that the model components are targeted well even for the set-up with lower sample size, high noise, and sparsity in repetitions. The median, 10th, and 90th percentile RSE and normalized MSE values for model parameter estimates, including estimated means, eigenfunctions, eigenvectors, eigenvalues, error variance, response predictions and M-HPCA ICC are provided in Table 3.1 and Supplementary Materials Table B.1 for sparse and dense designs, respectively. Percentiles of estimates are reported combined over groups, repetitions, components and subjects. More specifically, while summary measures in estimation of $\mu(t)$ are reported over 200 Monte Carlo runs, those for σ_d^2 , $\eta_{dj}(r, t)$, $(v_{dk}^{(1)}(r), v_{dp}^{(2)}(r), \phi_{dl}^{(1)}(t))$,

Table 3.1: The 50% (10%, 90%) percentiles of the relative squared errors, normalized mean squared errors based on 200 Monte Carlo runs from the simulation design at $n_d = 15, 50$ with low and high noise for the **sparse** set-up.

	Low Noise		High Noise	
	$n_d = 15$	$n_d = 50$	$n_d = 15$	$n_d = 50$
$\mu(t)$.009 (.003, .029)	.003 (<.001, .008)	.009 (.003, .030)	.003 (<.001, .009)
$\eta_{dj}(r, t)$.036 (.017, .075)	.011 (.005, .022)	.045 (.024, .085)	.014 (.007, .026)
$Y_{dij}(r, t)$.181 (.104, .531)	.182 (.104, .533)	.445 (.299, .798)	.445 (.299, .799)
$\phi_{d1}^{(1)}(t)$.021 (.003, .123)	.005 (<.001, .025)	.024 (.004, .125)	.005 (<.001, .028)
$\phi_{d2}^{(1)}(t)$.098 (.034, .324)	.031 (.015, .079)	.119 (.044, .384)	.037 (.017, .094)
$\phi_{d1}^{(2)}(t)$.024 (.004, .140)	.007 (.001, .027)	.034 (.007, .163)	.010 (.002, .034)
$\phi_{d2}^{(2)}(t)$.088 (.020, .523)	.030 (.008, .159)	.139 (.042, .653)	.051 (.015, .200)
$\nu_{d1}^{(1)}(r)$.070 (.007, .569)	.022 (.002, .114)	.088 (.016, .626)	.027 (.005, .124)
$\nu_{d2}^{(1)}(r)$.141 (.034, .647)	.034 (.009, .133)	.219 (.065, .759)	.048 (.016, .162)
$\nu_{d1}^{(2)}(r)$.096 (.014, .672)	.027 (.005, .144)	.166 (.037, .875)	.048 (.015, .188)
$\nu_{d2}^{(2)}(r)$.206 (.041, .931)	.042 (.010, .173)	.517 (.159, 1.399)	.103 (.038, .335)
λ_{dg}	.074 (.003, .389)	.018 (<.001, .106)	.082 (.003, .468)	.019 (<.001, .120)
λ_{dh}	.057 (.002, .328)	.014 (<.001, .086)	.100 (.005, .533)	.018 (<.001, .116)
σ_d^2	.014 (.004, .033)	.016 (.004, .035)	<.001 (<.001, .002)	<.001 (<.001, .002)
ρ_{dW}	.006 (<.001, .039)	.001 (<.001, .009)	.007 (<.001, .048)	.002 (<.001, .013)

$\phi_{dm}^{(2)}(t)$, $(\lambda_{d,k\ell}, \lambda_{d,pm})$ and $Y_{dij}(r, t)$ are reported over $D \times 200$, $D \times J \times 200$, $D \times 2 \times 200$, $D \times 4 \times 200$ and $D \times n_d \times J \times 200$ runs, respectively, where $D = 2$, $J = 2$ and $K = P = L = M = 2$. The RSE and normalized MSE values decrease with increasing sample size, decreasing noise and denser designs, showing that model components are targeted effectively, except for the RSE for σ_d^2 , which gets smaller with increasing noise level. The RSE most sensitive to the noise level is the one for response predictions, since the model does not try to fit the noise. In addition, RSEs for the level 2 eigenvectors are slightly higher than level 1 eigenvectors, which could be due to the additional adjustment of the diagonals of the marginal regional within-subject covariance matrix (by subtraction of the estimated measurement error).

To assess the performance of the proposed bootstrap test for condition differentiation, we consider the null hypothesis $H_0 : \eta_{11}(1, t) = \eta_{12}(1, t)$, for equality of the group-region-

repetition-specific means for the two conditions ($j = 1, 2$) in group $d = 1$ and region $r = 1$. To generate data under growing degrees of departure from the null, the data generation model described in Supplementary Material Appendix B.4 is modified where $\eta_{1j}(r, t) = 0$ for $r \neq 1$ and $\eta_{1j}(1, t) = (-1)^j \Delta$ for $j = 1, 2$, with larger Δ representing larger deviations from the null. The estimated level and power of the test are assessed based on 200 Monte Carlo runs where p-values are based on 200 bootstrap samples. Supplementary Material Figure B.5.3 (a) and (b) display power with increasing Δ , where the power increases faster with increasing sample size. The level of the test is estimated at (0.055, 0.039), (0.075, 0.065), (0.115, 0.080) and (0.090, 0.070) for $n_d = (15, 50)$, in the (dense, low noise), (dense, high noise), (sparse, low noise) and (sparse, high noise) set-ups, respectively, all hovering around the target level 0.050. In addition, we assess coverage and length of the proposed bootstrap CI for M-HPCA ICC. Coverage values hover around the target value 0.95 across sample size, while the length of the CI decrease with increasing sample size, decreasing noise and denser designs, as expected (Supplementary Material Figure B.5.3 (c) and (d)).

Finally, Supplementary Material Figure B.5.3 (e) compares run times of the proposed MM algorithm with `lme4` in fitting the mixed effects model of Section 3.2.3, as a function of the total number of product components at each level of the data. The comparison highlights the computational savings achieved by the proposed MM algorithm and shows that it stays computationally feasible in large data applications with increasing number of product components. Computational time increases exponentially with increasing number of product components for `lme4`.

3.5 Discussion

The proposed M-HPCA models EEG data in its full complexity, including the functional and regional features as well as the repeated observations over experimental conditions or visits. EEG, both from the time and the frequency domains are targeted under the umbrella of multilevel region-referenced functional data. M-HPCA relies on dimension reduction techniques via modeling of the exchangeable correlation structure across repetitions using an

ANOVA-type decomposition of the total variation into between- and within-subject components. In addition, weak separability, FPCA and PCA are used to represent the functional and regional variability in lower dimensions. The proposed dimension reduction is coupled with a computationally efficient MM algorithm that is specifically designed to take advantage of the lower dimensional representation provided by M-HPCA. The computational savings proposed via representation of the high-dimensional covariances in terms of much-smaller building blocks within the proposed MM estimation framework enable the modeling of EEG in its full complexity, which cannot be achieved by standard packages.

Modeling the entire data without collapsing its dimensions enables characterization of the comprehensive features in the entire ERP/PSD morphology across the entire scalp. In addition, modeling of the multilevel structure leads to a new concept of intraclass correlation that summarizes agreement between repetitions of this high-dimensional signal. The study of condition differentiation within or across diagnostic groups is also achieved with the proposed multilevel modeling framework, in which the efficient implementation of M-HPCA via the proposed MM makes inference via parametric bootstrap feasible. The diverse applications of M-HPCA are showcased with two data examples, including data from a task-based paradigm analyzed in the time-domain via ERPs and data from resting state analyzed in the frequency domain via PSDs. While condition differentiation between the match vs. mismatch conditions is the object of inference in the audio odd-ball paradigm, short-term reliability of PSDs is quantified via M-HPCA ICC in the ABC-CT feasibility study. The lower dimensional representation of the EEG data can also be useful in incorporation of the EEG into regression models, where relationships between the estimated subject- or subject-repetition specific scores and behavioral outcomes can be studied.

The proposed estimation algorithm has been presented for data with missingness at the repetition level (subjects with missing visits/conditions), since this is a common missing pattern for EEG data where data are typically sampled on a dense functional grid. M-HPCA can be extended to cases where there is also missingness in the functional dimension. However such an extension would require the design matrices, \mathbf{Z}_{di} and \mathbf{W}_{dij} , and covariance ma-

trices, $\Phi_d^{(1)} \Lambda_d^{(1)} \Phi_d^{(1)\top}$ and $\Phi_d^{(1)} \Lambda_d^{(1)} \Phi_d^{(1)\top}$, associated with subject- and subject-region-specific eigenscores in the proposed mixed effects framework, to be computed at the subject and repetition level, due to varying matrix dimensions of complete data. This would lead to additional computational costs. However if the missingness is at random, then a presmoothing in the dense functional dimension may help impute possible missing values and avoid the additional computational costs.

CHAPTER 4

Region-referenced longitudinal functional generalized linear model for EEG data

4.1 Introduction

Electroencephalography (EEG) is a low cost and non-invasive modality to study the brain and its activity. EEG is recorded using electrodes arranged on the scalp and is analyzed in the frequency domain via power spectral density (PSD) for resting-state paradigms. Generated data can be viewed as region-referenced functional objects and when these signals are observed across multiple visits or trials, the data gains a longitudinal dimension that can either be considered discrete, in the case of a small number of visits, or continuous, in the case of a large number of trials. Children with autism spectrum disorder (ASD) are a group that can benefit from being studied using EEG resting-state paradigms since the focus is on internal brain responses, which can explain differences in observable behavioral responses. Heterogeneity in cognitive ability of children with ASD is not well understood.

Our motivating study was on oscillatory development across the first year of life conducted by our collaborators at the UCLA Center for Autism Research and Treatment (Dickinson *et al.*, 2022). The goal of the originating study was to characterize how the changes in oscillatory development differ between children with ASD and their typically developing (TD) peers. Of particular interest was the differences in maturation of the dominant peak in the alpha range (6-9 Hz), which can be seen in the group-averaged log PSD in the parietal right region in Figure 4.1(a). Data was collected at four visits during infancy (3, 6, 9, and 12 months) and ASD diagnosis was made at 36 months. At each visit, task-free spontaneous

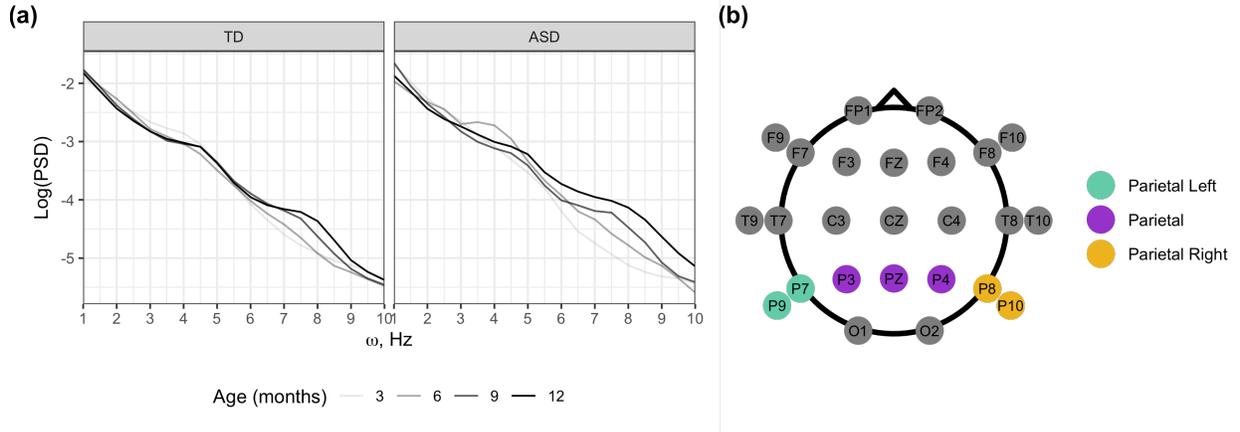


Figure 4.1: (a) Group-specific mean alpha band log spectral density from the right parietal region. Darker lines correspond to an increase in age. (b) Positions of the 7 electrodes used in the analysis of the first year of life data, highlighted based on the region in which they are averaged.

EEG was collected during a resting-state paradigm while the child was held in their caregiver’s lap and kept calm. EEG was recorded across electrodes, representing the regional dimension, and transformed into a PSD across the frequency domain, representing the functional dimension. The resulting PSDs form the region-referenced longitudinal functional data. This data structure is complicated as there is a continuous functional dimension, and discrete regional and longitudinal dimensions. We will use the first year of life data as a region-referenced longitudinal functional predictor of diagnostic status.

Functional data analysis methodology has grown rapidly over the past three decades in order to handle trends in the entire functional observation, especially high-dimensional functional data with dependencies among curves. Functional data that is correlated due to a longitudinal-based design, such as the repeated visits in our motivating example, has been studied under the umbrella of longitudinal functional data analysis. Longitudinal functional data analysis has become increasingly popular as seen in the works of Morris *et al.* (2003); Morris & Carroll (2006); Di *et al.* (2009); Greven *et al.* (2010); Hasenstab *et al.* (2017); Scheffler *et al.* (2020); Campos *et al.* (2022); Boland *et al.* (2022); Cui *et al.* (2022).

Functional regression methods, first introduced by Ramsay & Dalzell (1991) and formalized by Hastie & Mallows (1993); Marx & Eilers (1999) for Gaussian and non-Gaussian

responses, respectively, for relating functional predictors to scalar responses are termed scalar-on-function regression models (SoFR). Methods that deal with multivariate and multidimensional functional predictors will be of particular relevance as we try to relate a region-referenced longitudinal functional predictor to a scalar response. Reiss et al. (2017); Morris (2015) each provide an excellent review of these methods as well as others. One of the major areas of development in functional regression modeling is in the method of regularization used. Some models opt for basis functions chosen *a priori* such as the spline and wavelet approaches taken in Hastie & Mallows (1993); Marx & Eilers (1999); Crainiceanu & Goldsmith (2010); Goldsmith et al. (2011) which rely on either a quadratic penalty or Bayesian frameworks. Other models rely on data-driven basis functions such as those found using functional principal component analysis (FPCA) in Yao et al. (2005) or functional partial least squares in Reiss & Ogden (2007). Xiao et al. (2021) developed a linear model for generalized responses (i.e. scalar, binary, Poisson) with functional and non-functional covariates and recently, Sun & Wang (2022) proposed an adaptive group LASSO for scalar outcomes and functional predictors.

SoFR methods for longitudinal functional data have also been considered. Most commonly, methods stemming from functional principal component regression (FPCR) (Reiss & Ogden, 2007) have been developed which rely on decomposing the variability in the data using some form of functional principal components analysis (FPCA) then regressing the outcome on the eigenscores from that decomposition. Methods with a longitudinal scalar outcome as well as longitudinal functional predictor have been increasingly popular, as seen in the following literature. These methods can be categorized into three groups based on the form of the regression function: scalar, functional, or longitudinal and functional. Crainiceanu et al. (2009) utilized multilevel functional principal components analysis (MFPCA) to first decompose the variation in longitudinal functional data then used the scores from the eigencomponents in a scalar generalized multilevel mixed model (GLMM) for a single scalar outcome, similar to our motivating data application. Speed improvements for MFPCA were made for extremely large data in Zipunnikov et al. (2011) which does not

need to load all of the data at once and creatively targets the BLUP estimates using the singular value decomposition of the covariance. Additional improvements were made in Cui *et al.* (2022) by extending methods used in fast covariance estimation (FACE) (Xiao *et al.*, 2016). Longitudinal FPCR (LFPCR) (Gertheiss *et al.*, 2013) uses the eigenscores estimated using the longitudinal FPCA (LFPCA) (Greven *et al.*, 2010) decomposition of the functional predictor in a non-functional generalized mixed-model framework. Longitudinal penalized functional regression (LPFR) (Goldsmith *et al.*, 2012) performed FPCA across all functional observations along with a spline basis for the coefficient function in a mixed-model framework with random effects to account for the correlation between observations within subjects. This method does not account for the longitudinal structure of the functional predictors. Kundu *et al.* (2016) incorporates a more flexible regression function that can differ over longitudinal time. In this case, the regression function is assumed to vary as a user-specified function of longitudinal time, which in their application is a polynomial. A generalized ridge penalty is used to constrain the regression estimates. Staicu *et al.* (2020) created a modeling framework in which both the longitudinal and functional dimensions of the time-varying regression function are assumed to be smooth. FPCR methods require interpretation of the shape of the estimated eigenfunctions then the results of the model are interpreted based upon those interpretations. These methods still do not fully encompass all dimensions of our motivating data since they do not include a regional dimension.

Our proposed region-referenced longitudinal functional generalized linear model (RRLF-GLM) aims to provide regression coefficients directly by constructing the regression coefficient using a tensor product of discrete and continuous bases, as previously seen in Scheffler *et al.* (2019). Therefore the functional effects are across the regional and longitudinal domains and removes the need to collapse any of the dimensions using averaging. RRLF-GLM is able to accommodate either a discrete or continuous longitudinal dimension, which makes it suitable to a broad range of applications.

Existing SoFR methods cannot accommodate region-referenced longitudinal functional predictors, but we are able to compare prediction accuracy from averaging the functional

predictors from each region across visits and considering that functional signal as multivariate functional data. This will allow us to use multivariate GFLMs (m-GFLM) in simulations as well as the data application and will show the predictive benefit of including the longitudinal dimension in analysis.

The proposed RRLF-GFLM is introduced in Section 4.2. Use of the RRLF-GLM for longitudinal data is showcased in Section 4.3 with an application to the first year-of-life data. Extensive simulations, reported in Section 4.4, assess the finite sample properties of the proposed approach and a brief discussion is given in Section 4.5. An R package to implement RRLF-GLM for longitudinal data accompanies this article and is available at www.github.com/emjcampos/rRLF-glm.

4.2 Proposed region-referenced longitudinal functional GLM

4.2.1 Statistical framework and modeling

The proposed region-referenced longitudinal functional GLM borrows ideas from the recently proposed covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM) (Scheffler et al., 2019) in relating region-referenced longitudinal functional predictors for each subject to a scalar outcome, by decomposing the regression function as a tensor product of discrete and continuous bases. For $i = 1, \dots, n$ subjects, observe the data $\{y_i, X_i(r, s, t)\}$, where y_i is the scalar response, $X_i(r, s, t)$ is a region-referenced longitudinal functional predictor observed at region $r = 1, \dots, R$, and time $t, t \in \mathcal{T}$. The functional predictor $X_i(r, s, t)$ is assumed to be square integrable and smooth over the functional domain \mathcal{T} . When the longitudinal domain is discrete, denote visit $s = 1, \dots, S$, and when the longitudinal domain is continuous denote $s \in \mathcal{S}$. For simplicity, each of the domains is assumed to be observed on a regular grid. However, in data applications where any of these domains is sparse, multilevel hybrid principal components analysis (M-HPCA) by Campos et al. (2022) can be used to reconstruct the functional predictor when the longitudinal dimension is discrete or hybrid principal component analysis (HPCA) by Scheffler et al. (2020) when

the longitudinal dimension is continuous. The subsequent models for the discrete and continuous cases are similar in that the sum across discrete longitudinal time points is analogous to the integral across continuous longitudinal time points. Therefore, the proposed model for region-referenced longitudinal functional predictors is given by

$$\begin{aligned}
y_i &\sim \mathcal{F}(\mu_i, \boldsymbol{\vartheta}) \\
g(\mu_i) &= \begin{cases} \sum_{r=1}^R \sum_{s=1}^S \int_{\mathcal{T}} \{X_i(r, s, t) - \eta(r, s, t)\} \beta(r, s, t) dt & \text{discrete case,} \\ \sum_{r=1}^R \int_{\mathcal{S}} \int_{\mathcal{T}} \{X_i(r, s, t) - \eta(r, s, t)\} \beta(r, s, t) dt ds & \text{continuous case,} \end{cases} \\
&= \begin{cases} \sum_{r=1}^R \sum_{s=1}^S \int_{\mathcal{T}} X_i^c(r, s, t) \beta(r, s, t) dt & \text{discrete case,} \\ \sum_{r=1}^R \int_{\mathcal{S}} \int_{\mathcal{T}} X_i^c(r, s, t) \beta(r, s, t) dt ds & \text{continuous case,} \end{cases} \tag{4.1}
\end{aligned}$$

where \mathcal{F} is an exponential family distribution with condition expectation $\mu_i = E\{y_i | X_i(r, s, t), \boldsymbol{\vartheta}\}$, $\boldsymbol{\vartheta}$ vector of nuisance parameters and $g(\cdot)$ invertible link function. The region-referenced mean curves are given by $\eta(r, s, t) = E\{X_i(r, s, t)\}$ and $X_i^c(r, s, t) = X_i(r, s, t) - \eta(r, s, t)$ denotes the mean-centered subject-specific functional predictor. The regression function $\beta(r, s, t)$ is assumed to be smooth in the continuous domains over the discrete domains, i.e. when the longitudinal domain is discrete, then the $\beta(r, s, t)$ is smooth in the functional domain within a specified region r and visit s . For a fixed discrete s , regression function captures weights on functional predictor across functional domain and how this changes over the R regions and for a fixed r , regression function captures weights on functional predictor across functional domain and how this changes over the S visits. The proposed model reduces to the m-GFLM model when either $R = 1$ and S is discrete or when $S = 1$ and the standard GFLM for a functional predictor and scalar response given in Marx & Eilers (1999) when $R = S = 1$.

The regression function is approximated by a linear combination of basis functions using

a tensor product of the discrete and continuous bases,

$$\beta(r, s, t) \approx \sum_{k_r=1}^{K_r} \sum_{k_s=1}^{K_s} \sum_{k_t=1}^{K_t} \phi_{k_r}(r) \phi_{k_s}(s) \phi_{k_t}(t) \theta_{k_r, k_s, k_t},$$

where the basis functions in t , denoted by $\phi_{k_t}(t)$, can be any set of continuous basis functions such as functional principal components or B -splines, and the basis functions in r denoted by $\phi_{k_r}(r)$ can be chosen as any set of discrete basis functions such that $\text{span}[\{\phi_{k_r}(r)\}_{k_r=1}^{K_r}] \subseteq \mathbb{R}^R$. The basis functions in s can be any set of discrete basis functions $\text{span}[\{\phi_{k_s}(s)\}_{k_s=1}^{K_s}] \subseteq \mathbb{R}^S$ or continuous basis functions, in the discrete and continuous longitudinal cases, respectively. One choice for the discrete basis functions are the columns of an identity matrix. The unknown coefficients are collected into θ_{k_r, k_s, k_t} to be estimated. The shape of this regression function is controlled by the marginal basis functions as well as the quadratic penalization of θ , which we follow from Scheffler et al. (2019) to be a Kronecker sum of marginal penalties along each dimension r , s , and t , i.e.

$$\text{pen}(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \boldsymbol{\theta}^T \mathbf{P}(\boldsymbol{\lambda}) \boldsymbol{\theta}$$

$$\mathbf{P}(\boldsymbol{\lambda}) = \lambda_r (\mathbf{I}_{K_a} \otimes \mathbf{P}_r \otimes \mathbf{I}_{K_t}) + \lambda_j (\mathbf{P}_j \otimes \mathbf{I}_{K_r} \otimes \mathbf{I}_{K_t}) + \lambda_t (\mathbf{I}_{K_r} \otimes \mathbf{I}_{K_t} \otimes \mathbf{P}_t), \quad (4.2)$$

where $\lambda = (\lambda_r, \lambda_j, \lambda_t)$ positive penalty parameters and \mathbf{P}_r , \mathbf{P}_j , and \mathbf{P}_t positive semi-definite penalty matrices. The regression function is assumed to be smooth along the continuous dimensions so a differencing penalty can be used for a B -spline basis. For the discrete dimensions, the regression function is not assumed to be smooth so the choice of penalty structure needs more care. When there is no prior knowledge, a ridge penalty can be used to promote smaller coefficients.

4.2.2 Estimation and inference

The region-referenced mean function $\eta(r, s, t)$ is first estimated within each region and visit (discrete case) or region alone (continuous case) based on data pooled across all subjects. In

the discrete case, smoothing is accomplished using penalized B -splines, and in the continuous case, it is done by projecting onto a tensor basis of penalized marginal B -splines in s and t . In both cases, estimation and smoothing parameter selection are performed using restricted maximum likelihood methods via the `gam()` function in the `mgcv` package in R. Due to possible imbalances in group sizes, the region-referenced mean function within each group can be averaged instead of pooling across all subjects. First, we introduce the model when the longitudinal dimension is discrete, then make the appropriate changes for a continuous longitudinal dimension. Let (t_1, \dots, t_H) denote the observed regular functional grid for the functional predictor, then the proposed RRLF-GLM in (4.1) can be written as

$$\begin{aligned}
& \sum_{r=1}^R \sum_{s=1}^S \int_{\mathcal{T}} X_i^c(r, s, t) \beta(r, st) dt \\
& \approx \sum_{r=1}^R \sum_{s=1}^S \int_{\mathcal{T}} X_i^c(r, s, t) \sum_{k_r=1}^{K_r} \sum_{k_s=1}^{K_s} \sum_{k_t=1}^{K_t} \phi_{k_r}(r) \phi_{k_s}(s) \phi_{k_t}(t) \theta_{k_r, k_s, k_t} dt \\
& \approx \sum_{r=1}^R \sum_{s=1}^S \sum_{h=1}^H \sum_{k_r=1}^{K_r} \sum_{k_s=1}^{K_s} \sum_{k_t=1}^{K_t} w_h X_i^c(r, s, t_h) \phi_{k_r}(r) \phi_{k_s}(s) \phi_{k_t}(t_h) \theta_{k_r, k_s, k_t} \\
& = \mathbf{X} \mathbf{W} (\mathbf{\Phi}_s \otimes \mathbf{\Phi}_r \otimes \mathbf{\Phi}_t) \boldsymbol{\theta}, \tag{4.3}
\end{aligned}$$

where \otimes denotes the standard Kronecker product. Additionally, in (4.3) $\mathbf{\Phi}_r = \{\phi_{k_r}(r)_{k_r=1, \dots, K_r}^{r=1, \dots, R}\}$, $\mathbf{\Phi}_s = \{\phi_{k_s}(s)_{k_s=1, \dots, K_s}^{s=1, \dots, S}\}$, and $\mathbf{\Phi}_t = \{\phi_{k_t}(t_h)_{k_t=1, \dots, K_t}^{h=1, \dots, H}\}$ denote $R \times K_r$, $S \times K_s$, and $H \times K_t$ matrices, whose columns are evaluations of the marginal basis function in r , s , and t , respectively. The predictor matrix \mathbf{X} is an $n \times RSH$ matrix with row i equal to the vectorized subject-specific functional predictor $X_i^c(r, s, t)$ and the matrix $\mathbf{W} = \text{diag}\{(w_{rsh})_{r=1, \dots, R; s=1, \dots, S}^{h=1, \dots, H}\}$ denotes the $RSH \times RSH$ diagonal matrix of weights that correspond to summing across the R regions and S visits in order to approximate the integral over the functional domain.

When the longitudinal dimension is continuous, let (s_1, \dots, s_L) denote the observed regular

grid of longitudinal points, then the proposed RRLF-GLM in (4.1) can be written as

$$\begin{aligned}
& \sum_{r=1}^R \int_{\mathcal{S}} \int_{\mathcal{T}} X_i^c(r, s, t) \beta(r, s, t) dt ds \\
& \approx \sum_{r=1}^R \int_{\mathcal{S}} \int_{\mathcal{T}} X_i^c(r, s, t) \sum_{k_r=1}^{K_r} \sum_{k_s=1}^{K_s} \sum_{k_t=1}^{K_t} \phi_{k_r}(r) \phi_{k_s}(s) \phi_{k_t}(t) \theta_{k_r, k_s, k_t} dt ds \\
& \approx \sum_{r=1}^R \sum_{l=1}^L \sum_{h=1}^H \sum_{k_r=1}^{K_r} \sum_{k_s=1}^{K_s} \sum_{k_t=1}^{K_t} w_l w_h X_i^c(r, s, t) \phi_{k_r}(r) \phi_{k_s}(s) \phi_{k_t}(t) \theta_{k_r, k_s, k_t} \\
& = \mathbf{X} \mathbf{W} (\mathbf{\Phi}_s \otimes \mathbf{\Phi}_r \otimes \mathbf{\Phi}_t) \boldsymbol{\theta}, \tag{4.4}
\end{aligned}$$

where $\mathbf{\Phi}_r = \{\phi_{k_r}(r)_{k_r=1, \dots, K_r}^{r=1, \dots, R}\}$, $\mathbf{\Phi}_s = \{\phi_{k_s}(s)_{k_s=1, \dots, K_s}^{l=1, \dots, L}\}$, and $\mathbf{\Phi}_t = \{\phi_{k_t}(t)_{k_t=1, \dots, K_t}^{h=1, \dots, H}\}$ denote $R \times K_r$, $L \times K_s$, and $H \times K_t$ matrices, whose columns are evaluations of the marginal basis function in r , s , and t , respectively. The predictor matrix \mathbf{X} is an $n \times RLH$ matrix with row i equal to the vectorized subject-specific functional predictor $X_i^c(r, s, t)$ and the matrix $\mathbf{W} = \text{diag}\{(w_{rlh})_{r=1, \dots, R; l=1, \dots, L}^{h=1, \dots, H}\}$ denotes the $RLH \times RLH$ diagonal matrix of weights that correspond to summing across the R regions in order to approximate the integral over the functional and longitudinal domains.

The following estimation procedure is similar for the discrete and continuous cases, where the appropriately sized matrices/vectors are used in each step. Following Scheffler et al. (2019), the coefficient vector $\boldsymbol{\theta}$ is estimated using penalized least squares, where the penalized log-likelihood is given by

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

where $\mathbf{u} = g^{-1}\{\mathbf{X} \mathbf{W} (\mathbf{\Phi}_j \otimes \mathbf{\Phi}_r \otimes \mathbf{\Phi}_t) \boldsymbol{\theta}\}$ and $\ell(\mathbf{u}, \boldsymbol{\vartheta} | \mathbf{y}) = \sum_{i=1}^n \ell(\mu_i, \boldsymbol{\vartheta} | y_i)$ denotes the log-likelihood function for the response. The penalized likelihood is optimized using a Laplace approximation of REML criterion then optimizes the approximated likelihood using Newton-Raphson updates, which is accomplished using the `gam()` function from the `mgcv` package in R (Wood, 2017, 2011).

The projection of the regression function $\beta(r, s, t)$ onto a tensor basis reduced the di-

mension somewhat, but the number of basis functions ($K_r K_s K_t$) can still be quite large and exceed the number of subjects, which makes estimation unstable so an additional reduction must be made. The second dimension reduction is made by retaining the right singular vectors of the design matrix $\mathbf{D} = \mathbf{XW}(\Phi_s \otimes \Phi_r \otimes \Phi_t)\boldsymbol{\theta}$, for the discrete and continuous longitudinal cases (Scheffler et al., 2019; Reiss & Ogden, 2010, 2007). Now, the penalized log-likelihood is minimized using the response mean function $\mathbf{u} = g^{-1}(\mathbf{D}\mathbf{V}_q\tilde{\boldsymbol{\theta}})$, where \mathbf{V}_q is the matrix containing the q leading columns from the SVD of the design matrix $= \mathbf{U}\mathbf{E}\mathbf{V}^T$ and $\tilde{\boldsymbol{\theta}}$ denotes the coefficient vector for the reduced design matrix $\mathbf{D}\mathbf{V}_q$. In practice, the number of components q is chosen such that $\sum_{s=1}^q \kappa_s^2 / \sum_{s=1}^{K_r K_s K_t} \kappa_s^2 > .95$ where κ_s are the ordered singular values from the SVD of \mathbf{D} . This corresponds to the minimum number of components needed to explain 95% of the total variation in the design matrix D . The penalty structure is updated with the transformation $\text{pen}(\tilde{\boldsymbol{\theta}}|\boldsymbol{\lambda}) = \tilde{\boldsymbol{\theta}}^T \mathbf{V}_q^T \mathbf{P}(\boldsymbol{\lambda}) \mathbf{V}_q \tilde{\boldsymbol{\theta}}$.

Bayesian point-wise confidence intervals are estimated rather than the naive point-wise confidence intervals since the naive methods are based on biased estimates of the coefficient vector $\tilde{\boldsymbol{\theta}}$ (Scheffler et al., 2019; Wood, 2006).

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

where $\boldsymbol{\gamma}$ denotes an estimate of $\tilde{\boldsymbol{\theta}}$, \mathbf{Z} denotes the $n \times n$ matrix with entries $Z_{ii} = \{g'(\mu_i)^2 V(\mu_i)\}^{-1}$ with μ_i the i th entry of \mathbf{u} , $\tilde{\mathbf{V}}$ is the variance function and $V(\mu_i)\sigma^2$ is the variance of y_i , and σ^2 denotes the scale parameter defined by \mathcal{F} . The covariance of $\tilde{\boldsymbol{\theta}}|\mathbf{y}$ can be adjusted by \mathbf{V}_q in order to get the desired posterior distribution of $\boldsymbol{\theta}|\mathbf{y}$. These Bayesian point-wise confidence intervals provide better coverage despite the biased parameter estimates. The finite sample properties are studied using simulations in Section 4.4 and interpretation is demonstrated in Section 4.3.

4.3 Data analysis

4.3.1 Data structure and methods

The data from our first application was collected by our collaborators at UCLA, on 47 children to study how deviations in early oscillatory development may differentiate those with neurodevelopmental disorders such as autism spectrum disorder (ASD). Spontaneous EEG data were collected at 4 visits across the first year of life (3, 6, 9, and 12 months) and clinical outcomes were assessed at 36 months of age. Thirty-seven children were categorized as typically developing (TD) and 10 with ASD, all of which attended all four visits. Data collection, cleaning, and pre-processing steps are described further in Dickinson *et al.* (2022). The functional grid includes $H = 19$ points, equally spaced from 1-10 Hz. The spectral density is normalized to unit area by dividing by the total integral in order to allow comparisons between electrodes and subjects. The relative PSD is log transformed. The three regions of interest are created by averaging nearby electrodes at the back of the scalp (parietal left: P7 and P9; parietal: P3, PZ, and P4; parietal right: P9 and P10) and are displayed in Figure 4.1(b).

Previous analysis of this data by Dickinson *et al.* (2022) conducted FPCA across all observations and used generalized linear mixed models to model the longitudinal trends of the eigenscores from each of the three leading eigenfunctions. Main effects of diagnostic group (TD, ASD), scalp region (frontal, central, occipital), and age, as well as the two- and three-way interactions along with a subject-level random intercept were included in the model. The estimated eigenfunctions corresponded to known distinct developmental trends. The leading eigenfunction accounted for 71% of the functional variation in the data and corresponded to a power increase in the alpha range (6-9 Hz), meaning higher scores on this eigenfunction indicate stronger increases in alpha power. It was found that there was a significant positive main effect of age for the first eigenscore, indicating that there was an increase in alpha power over time for all groups. The interaction for ASD and age was also found to be significant and positive, implying a quicker increase in alpha power over the first

year of life in the ASD group compared to the TD group.

The visit-region-referenced mean functions $\eta(r, s, t)$ are obtained using a basis of penalized B -splines with 10 degrees of freedom and second degree difference penalties. The regression function $\beta(r, s, t)$ is estimated by projecting onto a tensor product of basis functions along the region, visit, and frequency domains, where the marginal bases $\Phi_r = [\mathbf{1}_R, \mathbf{I}_R]$ and $\Phi_s = [\mathbf{1}_S, \mathbf{I}_S]$ with $K_r = 4$ and $K_s = 5$ and Φ_t is formed using evaluations of cubic B -splines with $K_t = 6$ degrees of freedom. A second-order difference penalty is used for \mathbf{P}_t to guarantee smoothness in the functional domain and a ridge style penalty is used for both the visit and regional domains $\mathbf{P}_s = [\mathbf{0}_S, \mathbf{I}_S]$ and $\mathbf{P}_r = [\mathbf{0}_R, \mathbf{I}_R]$ since we do not have any prior knowledge on the dependency of visits or regions. The first entry in each of the ridge-style penalty matrices corresponds to the first basis vector and is left unpenalized to absorb the effect across all visits or regions, respectively. The number of columns in the design matrix \mathbf{D} using all of these basis functions is $K_r K_s K_t = 120$. The rank of the reduced design matrix is 23, accounting for 99% of the total variation. The total computing time is approximately .2 seconds on a 3.7 GHz 6-Core Intel Core i5 operating R 4.1.1 and using the `gam()` function from the `mgcv` package (version 1.8-36). The penalty parameters chosen using REML are estimated to be $\lambda_r = 0.00027$, $\lambda_s = 0.00001$, and $\lambda_t = 4.51015$.

4.3.2 Data analysis results

The visit-region-referenced mean functions $\eta(r, s, t)$ are given in Figure 4.2 and show the known decrease in power with an increase in frequency as well as the dominant peak in the 6-9 Hz range that emerges as the children age. The visit-region-referenced mean functions are estimated across subjects rather than within groups since it is assumed that the functional predictors will retain group differences. The mean-centered $X_i^c(r, s, t)$ is targeted using these mean functions and used in modeling.

Figure 4.3 shows the estimated regression function for the parietal right (PR) region, which is highlighted due to its high contribution to the log-odds of ASD diagnosis $1/n \sum_{i=1}^n \sum_{s=1}^S \int_{\Omega} |X_i^c(r, s, t) \hat{\beta}(r, s, t)| d\omega$. The results from this region are representative of

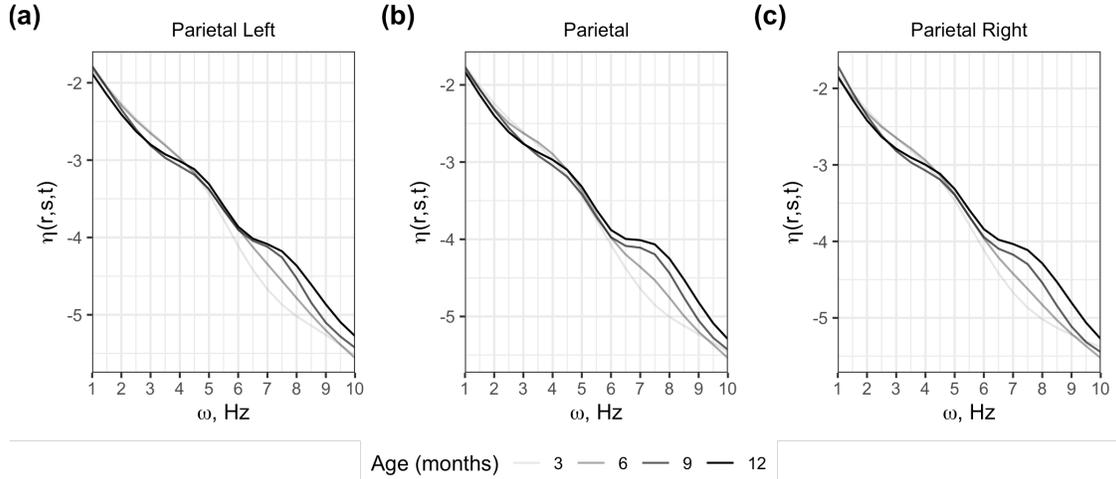


Figure 4.2: Estimated mean function $\eta(r, s, t)$ for the (a) parietal left, (b) parietal, and (c) parietal right regions, where darker lines correspond to an increase in age.

the other two regions. From Figure 4.3(a), which displays the average mean-centered functional predictor for the TD and ASD children in the parietal right region, it is obvious there are differences between the groups as the children age. On average, at 3 months the largest differences are in the dominant alpha peak between 6-9 Hz, with TD children having an alpha peak with a higher amplitude than the ASD children. At 6 months, this relationship reverses and by 12 months, the largest difference is towards the end of the alpha range.

The estimated regression functions for the parietal right region $\hat{\beta}(\text{PR}, s, t)$ are given in Figure 4.3(b) and put mostly positive weight on the spectral density for the entire frequency domain at 6 months and beyond. The Bayesian point-wise confidence intervals do not include zero at 6 months for 2-3 Hz and at 12 months at 1.5-2.5 Hz, but they are wide due to the small sample size. Figure 4.3(c) displays the point-wise product of the mean-centered functional predictor and the corresponding regression function, where the area under the curve corresponds to the linear effect on the log odds of an ASD diagnosis. At 6 months, the average TD child has a lower peak alpha amplitude and the positive weighting of the regression coefficient results in a predicted probability of ASD that is less than 0.5. The average child with ASD has a higher peak alpha amplitude by 12 months and the positive weighting from the regression coefficient results in a predicted probability of ASD that is more than 0.5.

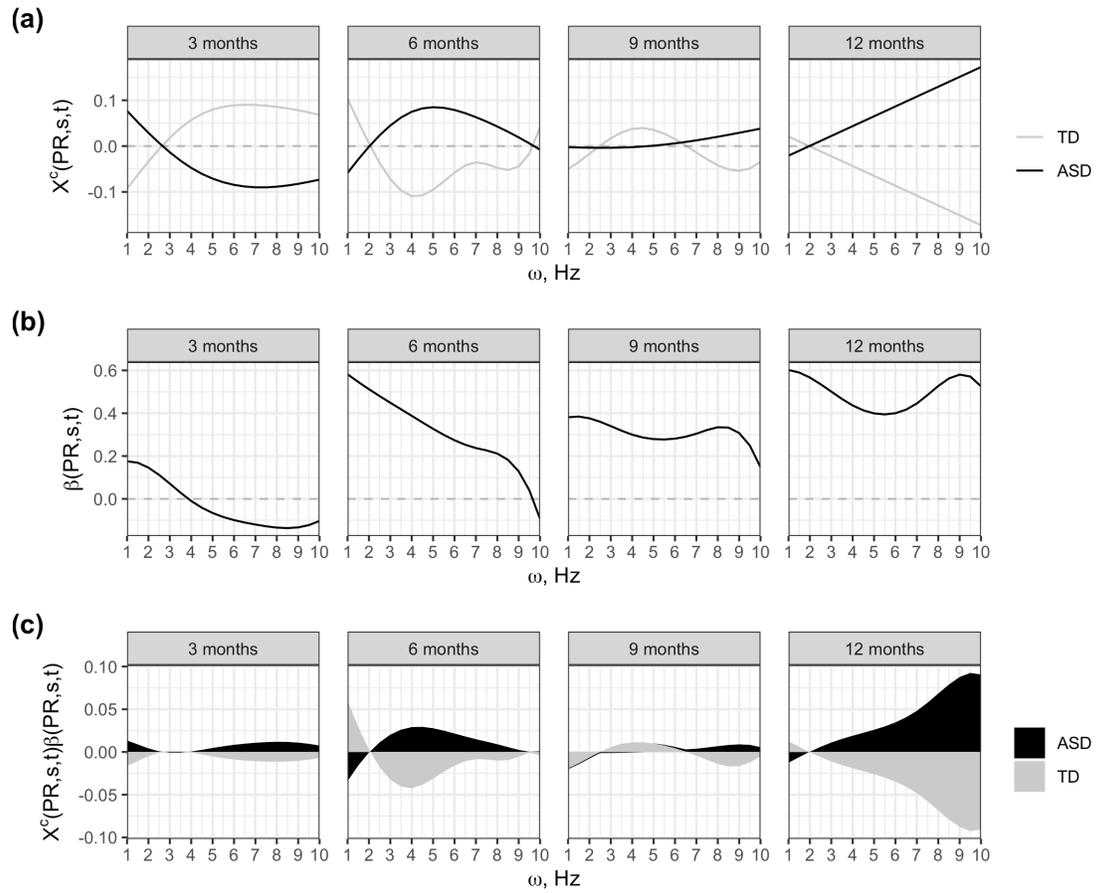


Figure 4.3: (a) The average mean-centered functional predictor for TD (grey) and ASD (black) children. (b) The estimated regression function $\beta(r, s, t)$. (c) The point-wise product of the mean-centered functional predictor and the estimated regression function, representing the average area under the curve for TD (grey) and ASD (black) children.

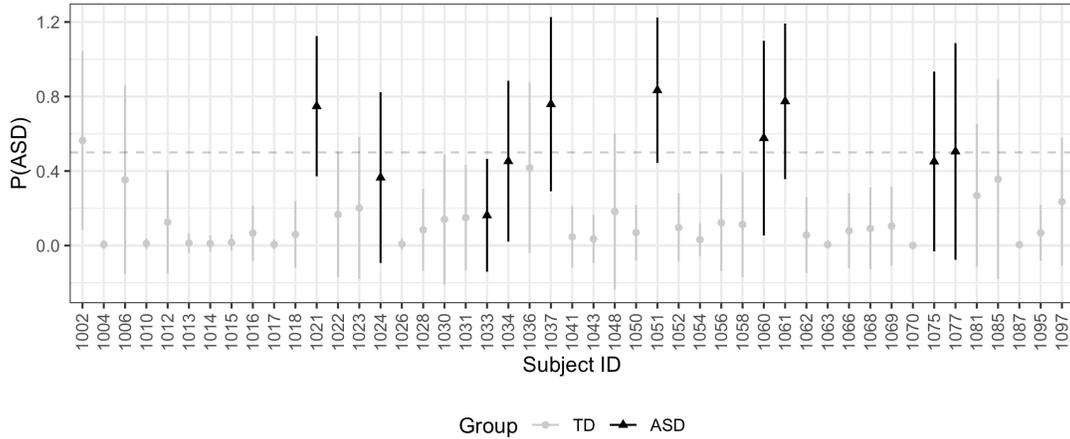


Figure 4.4: The predicted probabilities of ASD and their 95% confidence intervals for the study participants, where true group membership is denoted by grey for TD and black for ASD.

The 95% confidence intervals for the predicted probabilities of ASD diagnosis are targeted on the logit scale based on the posterior distribution of $DV_q \tilde{\theta} | \mathbf{y}$ and transformed to the probability scale (see Figure 4.4). The small sample size of the ASD group results in wide confidence intervals for the ASD group. Using a threshold of 0.5, the sensitivity (sens = $P\{\widehat{ASD}|ASD\}$), specificity (spec = $P\{\widehat{TD}|TD\}$), and area under the receiver operating characteristic curve (AUC), using leave-one-out cross validation (LOOCV) are sens = .44; spec = .90; and AUC = .78. In comparison, when the data is averaged across visits and mGFLM is used, sens = .44; spec = .90; and AUC = .75. Results from averaging are similar to the results from the uncollapsed data. The estimated regression function is moderately similar for 6 months and beyond, so averaging across the longitudinal dimension does not result in a loss of too much information in this application. We will show in simulations that RRLF-GLM outperforms mGFLM when the regression function is known to change over longitudinal time.

4.4 Simulation

The finite sample performance of the proposed RRLF-GLM for longitudinal data is assessed in several simulation set-ups, in which sample size, the type of longitudinal domain (discrete vs. continuous), and correlation among regions and visits are varied.

4.4.1 Data generation

The functional grid is chosen as 50 equidistant points in $[0, 1]$ and the total number of regions $R = 3$. The longitudinal grid is chosen as 4 or 20 equidistant points in $[0, 1]$, when the longitudinal dimension is discrete or continuous, respectively. The mean function $\eta(r, s, t)$ is set to zero for simplicity. The regression function $\beta(r, s, t)$ varies across regions and along the functional and longitudinal domains

$$\beta(r, s, t) = \begin{cases} -2 \cos(t\pi/6 + \pi s), & r = 1, \\ 2 \sin(t\pi/6 + \pi s), & r = 2, \\ 2 \sin(2t\pi/6 + \pi s) & r = 3. \end{cases}$$

The subject-specific functional predictors are defined as $X_i(r, s, t) = \sum_{k=1}^5 \xi_{ikrs} \psi_k(t)$, where $\psi_k(t)$, $k = 1, \dots, 5$ are a common set of basis functions across the regions and longitudinal visits, and in this case are chosen to be cubic B -splines. The matrix of subject-specific scores are generated from

$$\boldsymbol{\xi}_{ik} = \begin{bmatrix} \xi_{ik11} & \dots & \xi_{ik1S} \\ \xi_{ik21} & \dots & \xi_{ik2S} \\ \vdots & \ddots & \vdots \\ \xi_{ikR1} & \dots & \xi_{ikRS} \end{bmatrix} \sim \text{Matrix Normal}(\mathbf{0}, \Sigma_r, \Sigma_s),$$

where Σ_r and Σ_s are the $r \times r$ and $S \times S$ (or $H \times H$ when the longitudinal dimension is continuous) matrices that control the regional and longitudinal dependency of the functional

predictors, respectively. The regional dependency is chosen to have a compound symmetric structure where the diagonal entries are set to one and the off-diagonal entries are set to ρ_r . When the longitudinal dimension is discrete, the dependency among visits is also chosen to have a compound symmetric structure with off-diagonal entries set to ρ_s . When the longitudinal dimension is continuous, dependency is chosen to be a first-order autoregressive structure with homogenous variances, i.e.

$$\Sigma_j = \begin{bmatrix} 1 & \rho_s & \rho_s^2 & \dots & \rho_s^H \\ \rho_s & 1 & \rho_s & \dots & \rho_s^{H-1} \\ \rho_s^2 & \rho_s & 1 & \dots & \rho_s^{H-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_s^H & & & & 1 \end{bmatrix}$$

The ρ_r and ρ_s values are used as tuning parameters for the level of dependency across regions and longitudinal visits. The binary scalar outcomes are then simulated from $y_i \sim \text{Bernoulli}(\mu_i), i = 1, \dots, n$, where the subject-specific probabilities are targeted using $\text{logit}(\mu_i) = \sum_{r=1}^R \sum_{s=1}^S \int_{\mathcal{T}} X_i^c(r, s, t) dt$ when the longitudinal dimension is discrete and using $\text{logit}(\mu_i) = \sum_{r=1}^R \int_{\mathcal{S}} \int_{\mathcal{T}} X_i^c(r, s, t) dt ds$ when it is continuous.

The finite sample performance of the proposed RRLF-GLM is assessed across 200 Monte Carlo runs in a total of 72 simulation settings: four sample sizes ($n = 70, 200, 500, 1000$), discrete and continuous longitudinal dimension, and three levels of regional and longitudinal dependency ($\rho_r = 0.0, 0.2, 0.4$ and $\rho_s = 0.0, 0.2, 0.4$). The relative squared error is used to assess the regression function estimates and is targeted as $\text{RSE}(r, s, t) = \|\hat{\beta}(r, s, t) - \beta(r, s, t)\|^2 / \|\beta(r, s, t)\|^2$, where

$$\|\beta(r, s, t)\|^2 = \begin{cases} \sum_{r=1}^R \sum_{s=1}^S \int_{\mathcal{T}} \beta^2(r, s, t) dt & \text{discrete case, and} \\ \sum_{r=1}^R \int_{\mathcal{S}} \int_{\mathcal{T}} \beta^2(r, s, t) dt ds & \text{continuous case.} \end{cases}$$

The coverage probability of the Bayesian point-wise confidence intervals is assessed using the proportion of times the true regression function lies within the estimated confidence interval over the 200 Monte Carlo runs. In each Monte Carlo run, an additional 200 subjects are simulated to be used as a validation set for assessing prediction accuracy. Prediction accuracy is assessed using the AUC in the validation set and a common number of observations supports comparisons across simulation settings.

The region-referenced mean functions $\eta(r, s, t)$ are targeted by pooling data across all subjects and smoothing using penalized splines over the functional domain when the longitudinal domain is discrete and bivariate penalized splines over the functional and longitudinal domains when the the longitudinal domain is continuous, with 5 degrees of freedom along the functional and longitudinal domains. Second-order difference penalties are used for continuous domains and a ridge penalty is used for discrete domains. The marginal basis matrices for continuous domains are formed as evaluations of the cubic B -splines with 5 degrees of freedom and discrete domains are formed as a column of ones appended to the front of an identity matrix, i.e. for the regional dimension $\Phi_r = [\mathbf{1}_3, \mathbf{I}_3]$. Therefore the model in all simulation settings has $K_r \times K_s \times K_t = 4 \times 5 \times 5 = 100$ coefficients. SVD is used to reduce the dimension of the design matrix so that 95% of the total variation is retained and across simulation settings, the rank of the design matrix increases with sample size, decreases with increases in ρ_r and ρ_s . When $\rho_r = \rho_s = 0.0$, the median rank is 25, 29, 31, 31 (discrete); 25, 31, 33, 33 (continuous) for $n = 70, 200, 500, 1000$, respectively (see Figure C.1.1). Median smoothing parameters for the discrete case $(\lambda_r, \lambda_s, \lambda_t)$: $N = 70$ (0.000002, 0.025, 0.101), $N = 200$ (0.0000035, 0.032, 0.141), $N = 500$ (0.000008, 0.032, 0.018), $N = 1000$ (0.00002, 0.032, 0.212) and the continuous case: $N = 70$ (0.0000006, 0.037, 0.087), $N = 200$ (0.0000005, 0.040, 0.105), $N = 500$ (0.0000005, 0.030, 0.187), $N = 1000$ (0.000008, 0.028, 0.248). There is no trend found with dependency among regions or longitudinal repetitions. The amount of time required based on 100 iterations for the Monte Carlo runs increases with sample size and is lower when the longitudinal dimension is discrete. Finally, the predictive performance of RRLF-GLM is compared to m-GFLM using functional predictors averaged

across the longitudinal domain.

4.4.2 Results

The mean function $\eta(r, s, t)$ is set to zero for simplicity which means the relative squared error cannot be calculated, but the squared error is calculated as $SE(r, s, t) = \|\hat{\eta}(r, s, t)\|^2$, where $\|\eta(r, s, t)\|^2 = \sum_{r=1}^R \sum_{s=1}^S \eta^2(r, s, t)dt$ or $\|\eta(r, s, t)\|^2 = \sum_{r=1}^R \int_{\mathcal{S}} \int_{\mathcal{T}} \eta^2(r, s, t)dt ds$, when the longitudinal dimension is discrete or continuous, respectively, and is displayed in Figure C.1.2 for each simulation setting. As sample size increases, the error decreases as expected. Also of note, the mean function $\hat{\eta}(r, s, t)$ is targeted more easily when the longitudinal dimension is continuous. However, as correlation increases, the error also increases slightly.

Figures C.1.3 and C.1.4 display the estimated regression function with median RSE, which get closer to the true function as the sample size increases. Figure 4.5 displays the RSE values for the regression function $\beta(r, s, t)$ and clearly decrease as the sample size increases in both the (a) discrete and (b) continuous cases, as expected. When the dependency between visits ρ_s is low and the longitudinal dimension is continuous, the RSE is lower. Increasing dependency between functional predictors, either regional or longitudinal, means that they share more information and can increase the RSE in the same way that multicollinearity can affect multivariate regression. Figure 4.6 displays the coverage probabilities for each simulation run and approaches the nominal level of 95% as the sample size increases. Coverage decreases as the correlation between regions and visits increases. The number of points considered is large (discrete: $3 \times 4 \times 50 = 600$, continuous: $3 \times 30 \times 50 = 4500$), so the outlying points have been jittered for presentation.

The out-of-sample accuracy displayed in Figure C.1.5 is higher for the continuous longitudinal case and increases as the sample size increases. The out-of-sample AUC displayed in Figure 4.7 is also higher for the continuous longitudinal case and increases as the sample size increases. When holding the regional dependency constant, the AUC increases as the longitudinal dependency increases. When compared to mGFLM, the out-of-sample AUC using RRLF-GLM for longitudinal EEG data is higher in every simulation setting. The largest

gains are made when the longitudinal dimension is continuous and when regional and/or longitudinal correlation is low.

4.5 Discussion

The proposed RRLF-GLM is a flexible model that can estimate functional effects or bivariate longitudinal functional effects across a non-smooth regional domain, where the regression function is projected onto a Kronecker product of discrete and continuous basis functions. RRLF-GLM relies on a penalty structure and dimension reduction techniques that benefit from the projection's Kronecker structure. The number of shrinkage parameters is kept small, but allows each dimension of the data to be smoothed appropriately. Unlike previously proposed regression methods for longitudinal functional data, RRLF-GLM does not rely on interpretation of FPCA components.

RRLF-GLM is used to model associations between diagnostic status and power spectral density over the first year of life in children with ASD and TD children. The PSD is known to change rapidly over the first year of life and deviations from typical development can help improve knowledge of neurodevelopmental conditions such as ASD. We find that differences across the scalp in region-referenced longitudinal PSD between children with ASD and TD children are reasonable predictors of diagnostic status and further investigation of these differences may lead to earlier diagnosis and intervention.

This model has been demonstrated using a developmental EEG study, but the methodology can be useful in many other scientific applications in which a multivariate functional predictor is observed longitudinally and the scalar response of interest is observed once. The longitudinal dimension could potentially be trial-level data, as in the case of experiments where learning is known to occur. If it is assumed that the functional predictors change over time and that change differs between diagnostic groups, then modeling the longitudinal trend is advantageous. However, trial-level data typically has a very low signal-to-noise ratio. This could be mitigated using a sliding window average, which has been proposed by

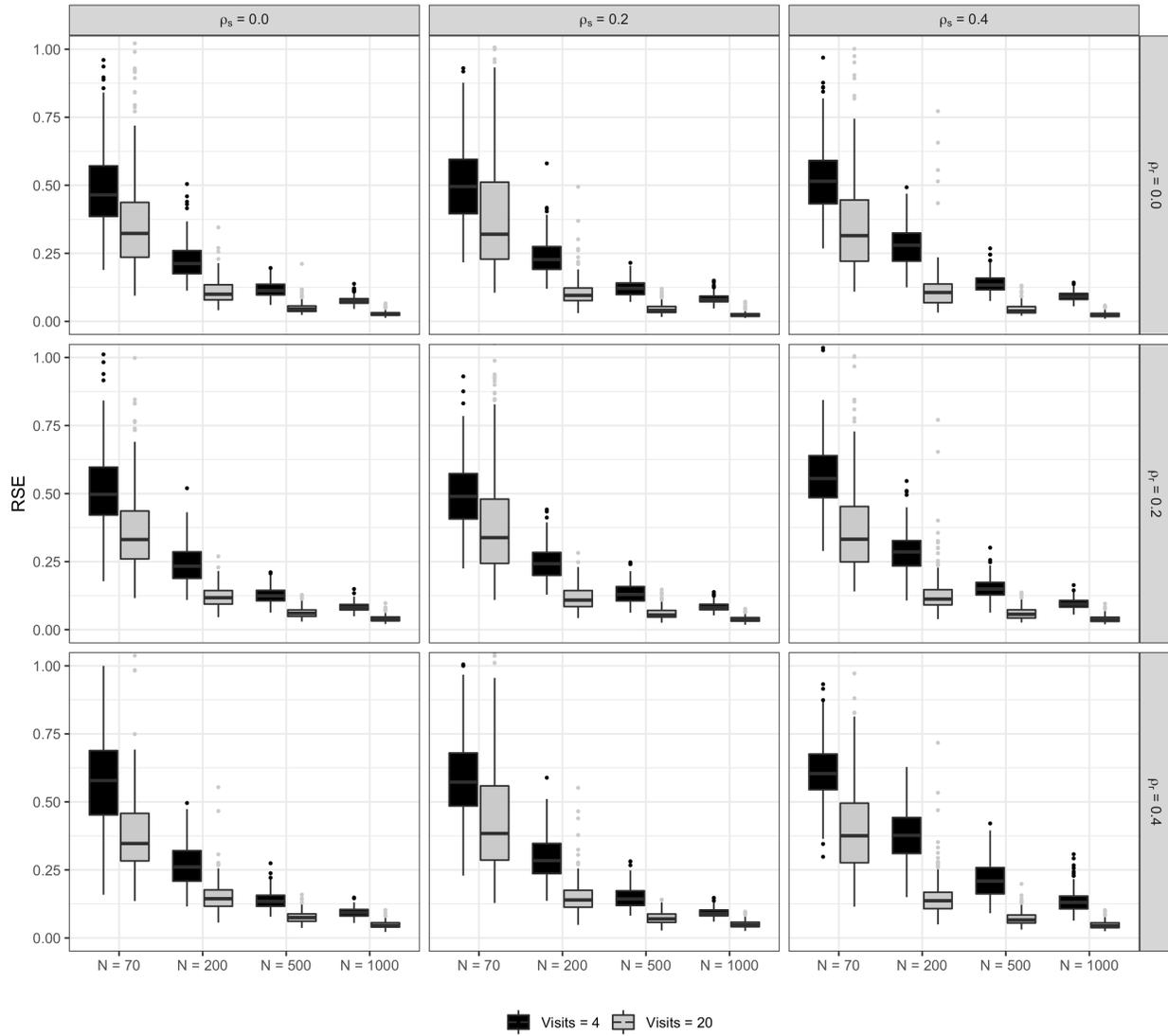


Figure 4.5: The RSE of the regression function across 200 Monte Carlo runs for the discrete (black) and continuous (grey) longitudinal dimension cases.

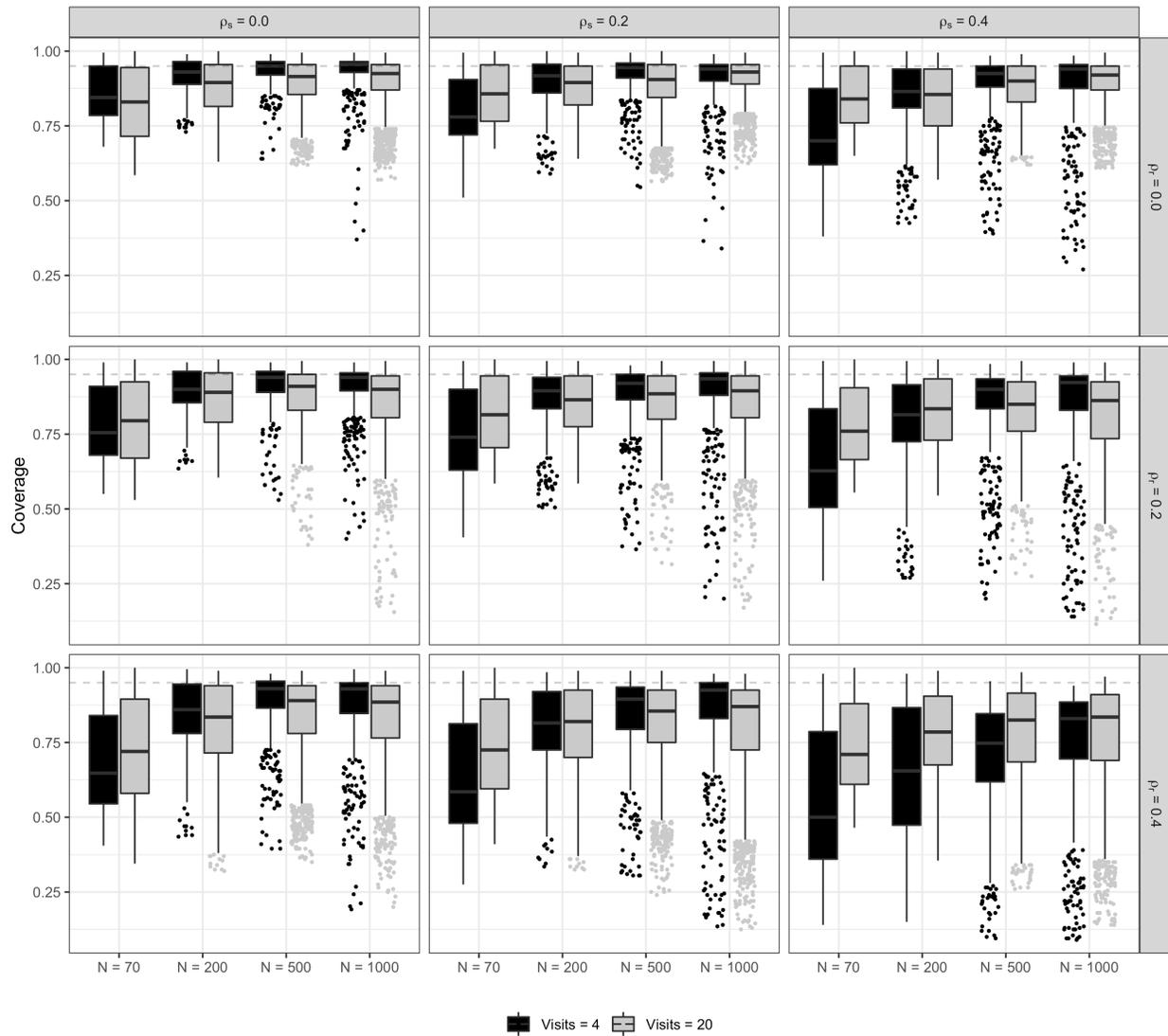


Figure 4.6: The coverage probabilities of the regression function across 200 Monte Carlo runs for the discrete (black) and continuous (grey) longitudinal dimension cases.

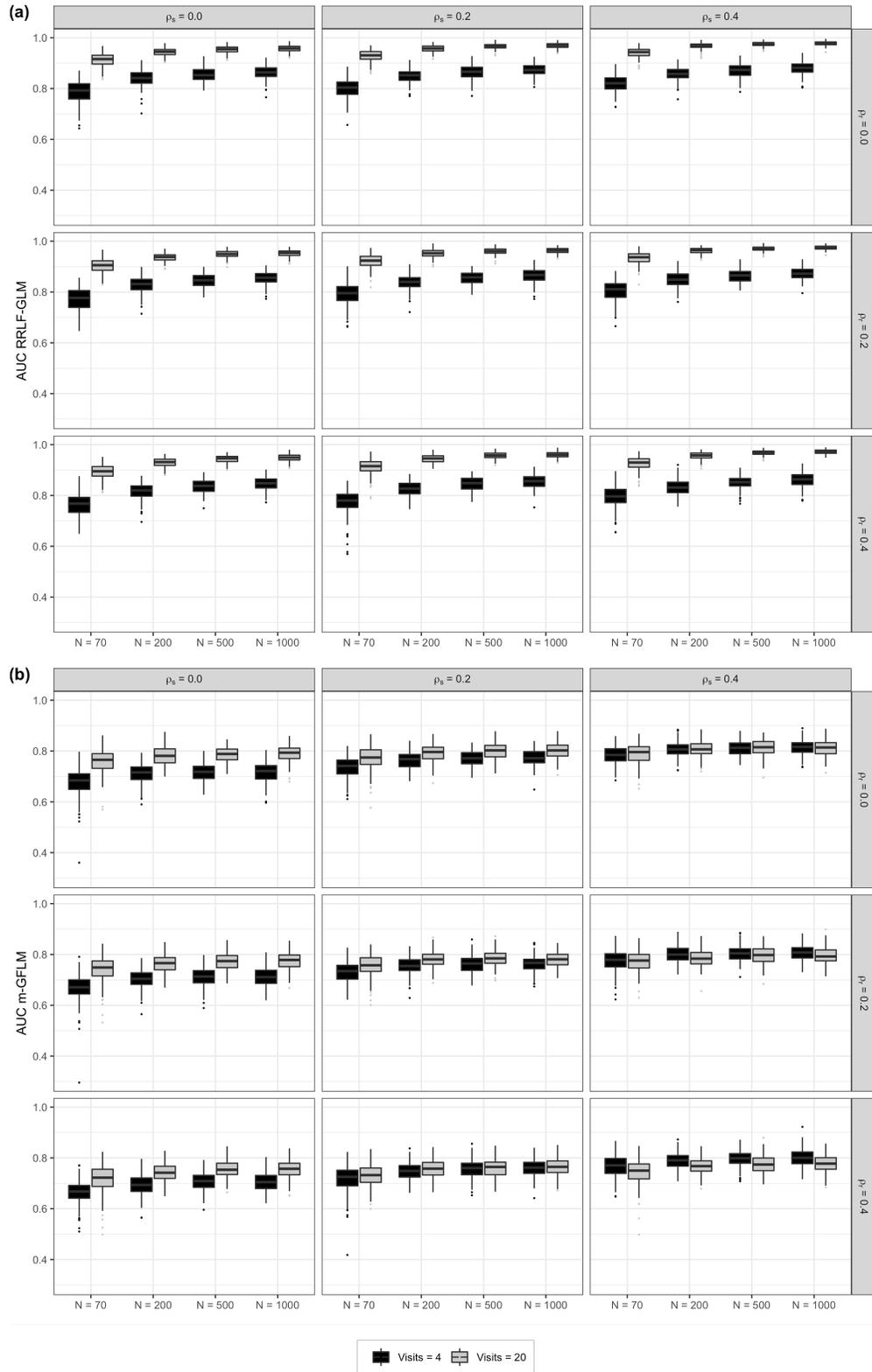


Figure 4.7: The out-of-sample AUC across 200 Monte Carlo runs for the (a) RRLF-GLM and (b) m-GFLM models for the discrete (black) and continuous (grey) longitudinal dimension cases.

Hasenstab et al. (2015). RRLF-GLM can also be applied to other EEG applications such as those that use experiment time and are analyzed in the time domain.

CHAPTER 5

Conclusion

The EEG applications provided within this dissertation highlight how methods for highly structured data are needed for description and inference, without collapsing any of the dimensions. However, with such large data, computational challenges need to be addressed. Therefore the methods proposed here have cleverly used dimension reduction tools and computational tricks to model EEG data without collapse.

In Chapter 2, principle ERP reduction (pERP-RED) analysis provides interpretable ERP components found across different conditions within multiple experimental tasks rather than relying on peak or mean amplitude within a user-specified window. The entire set of contributing ERPs can be translated into a linear combination of underlying waveforms that span the entire ERP. In the two applications, components found during traditional analysis were confirmed using pERPs and additional insights were made that were not found previously. In the ASD study, it was found that the largest loadings for the visual paradigm were on a pERP that was interpreted as relating to semantic processing in all groups. In the ADHD study, the pERP-RED algorithm was able to pick up the known N1/P2 complex related to visual stimuli and attention. Both of these analyses showed that the pERP-space analysis approach mitigated the overlap issue that is inherent in peak or mean amplitude analysis.

In Chapter 3, multilevel hybrid principal components analysis (M-HPCA) introduced a decomposition for multilevel region-referenced functional data. The total variation was decomposed the total variation into between- and within-subjects variance in a familiar ANOVA way, then utilized both vector and functional principal components analysis to break those covariance blocks down further. The model components were estimated using a mixed

effects modeling framework in conjunction with a computationally efficient minorization-maximization algorithm developed specifically for this covariance structure. Applications of the M-HPCA decomposition were highlighted in two different data examples, each with their own set of analysis goals. The audio-oddball paradigm found significant condition differentiation in the N400 and LNC components consistently across regions for both the TD and vASD groups, signaling semantic processing and cognitive integration, which agreed with previous findings. The day-to-day test-retest reliability of PSD study showcased the proposed measure of variability explained at each level of the data, called the M-HPCA ICC. It was found that there was a reasonable amount of agreement between between days within the same subject for both diagnostic groups.

In Chapter 4, a region-referenced longitudinal functional generalized linear model (RRLF-GLM) was introduced in order to relate clinical outcomes to the EEG data collected across multiple visits. The functional effects were estimated jointly across the discrete and continuous domains projecting the regression function onto a tensor product of discrete and continuous bases and using a penalty structure that relied on a sum of kronecker products along each dimension. RRLF-GLM was applied to the first year-of-life data where the dynamics of the observed region-referenced longitudinal functional data are known to change over time, but it was unknown how those changes differ between diagnostic groups. It was found that the model was able to distinguish between the two diagnostic groups using all dimensions of the region-referenced longitudinal functional data.

Appendix A

pERPred

A.1 Supplementary data

Supplementary data including additional figures referred to in Section 2.3 can be found online with this article.

A.2 ASD data collection and pre-processing

EEG was recorded at 500 Hz using a 128-channel HydroCel Geodesic Sensor Net. Prior to EEG data cleaning, the electrodes were interpolated and reduced to 25 in EEGlab by spherical interpolation (Perrin *et al.*, 1989) using the function *eeg_interp*. EEG was digitally filtered using a 0.3 to 30 Hz bandpass filter, segmented into 1000 ms epochs starting at 100 ms before the stimulus (picture, audio or word) onset, and baseline corrected using mean voltage during the 100 ms pre-stimulus baseline period. An automatic artifact detection tool was used to reject electrodes with amplitude difference (max-min) greater than 150mV, usually due to excessive electrode movement, or movement of the cap. Following this automatic artifact detection, each trial was visually inspected to remove any remaining channels that contained EMG, eye-blink, or eye-movement artifacts from further analysis. Trials with more than 15% bad channels were rejected. Only subjects with more than 10 artifact-free trials per condition were accepted for further analysis. For pERP-RED analysis, ERPs averaged over trials, time-locked to the picture onset, the auditory stimulus, and the word image for both match and mismatch conditions are considered producing 5 trial types: image, sound match, sound mismatch, text match, and text mismatch. Epochs considered for pERP-RED

were restricted to -100 to 1000ms.

A.3 ADHD data collection and pre-processing

Recording and pre-processing details are similar to those reported in Lenartowicz et al. (2019). All recordings were collected with a 40 channel Electrocap with electrodes positioned as in the 10/20 system. Impedences were reduced to 10 K Ω before recording on MANSCAN hardware and software. Continuous EEG was collected at 256 Hz, referenced by linked mastoids. Pre-processing of data was done using EEGLAB software (v.11.03.b, Delorme & Makeig (2004)). Data were high-pass filtered at 0.1 Hz. Noisy electrodes (determined by visual inspection) were removed from further analysis. In each subject, epochs with movement or muscle artifacts were identified and removed if signal power in that epoch exceeded the 85th percentile for $> 60\%$ of the channels. The resulting continuous data were then epoched and averaged per subject to arrive at the ERPs used in the main text. Epochs considered for pERP-RED were restricted to -500 to 2000ms.

A.4 Online Application for Exploring Results

ASD Data Exploration, pERP panel

ASD Data Exploration

Experiments

pERPs

Individual Reconstruction

Headmaps

Condition Differences

Our Principle ERP Reduction (pERP-RED) algorithm uses a series of data concentration and source separation steps to find a set of bases functions we call principle ERPs (pERPs). These bases are the set of underlying component waveforms such that the ERPs formed by time-locked averages at any given electrode, participant, or trial type is approximately a weighted combination of these components.

We use a measure similar to the typical R^2 in order to choose the number of pERPs estimated. Based on the R^2 value found here, we chose to estimate 10 pERPs.

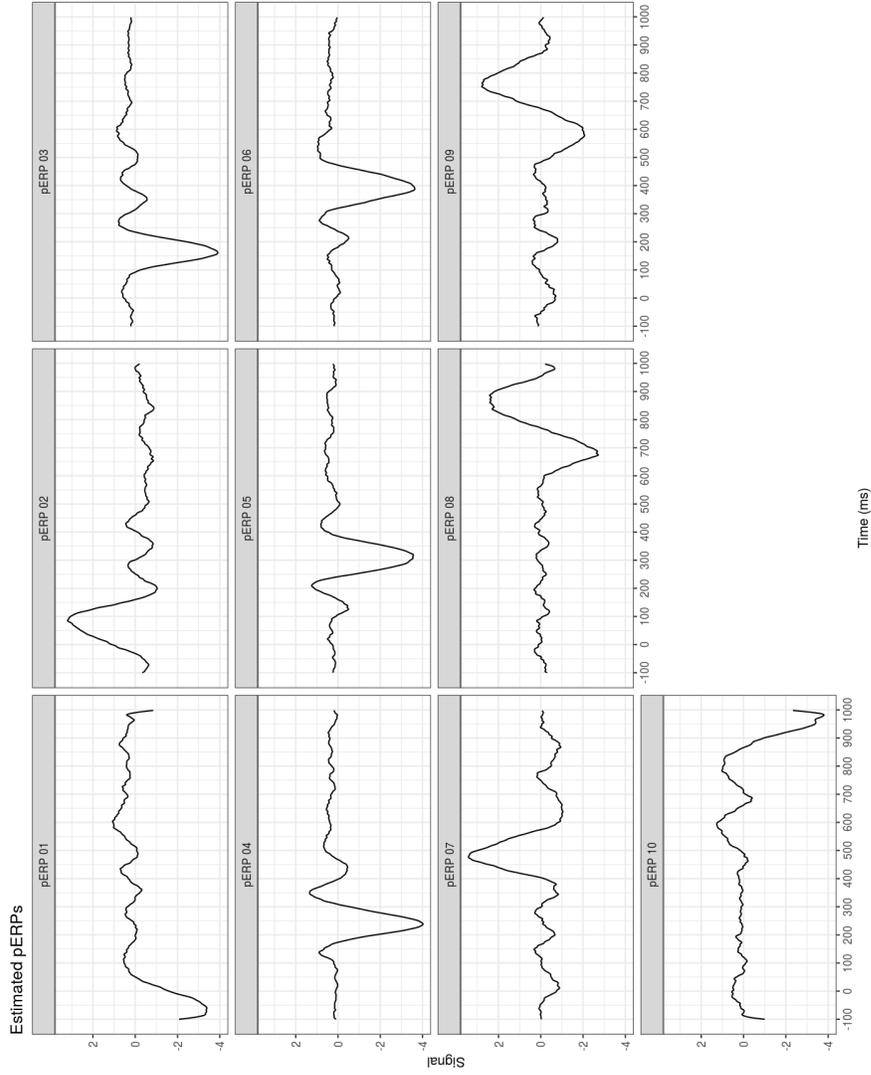
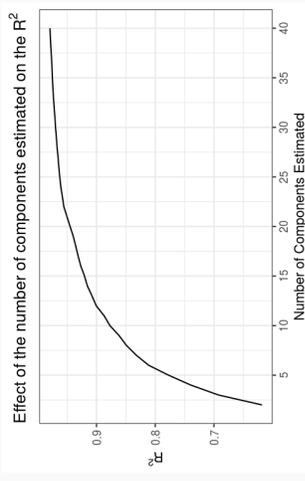


Figure A.4.1: In the ASD Data Browser, there is a panel that displays the estimated pERPs as well as the plot of the R_{test}^2 used to determine the number of pERPs.

ASD Data Exploration, Individual Reconstruction panel

ASD Data Exploration

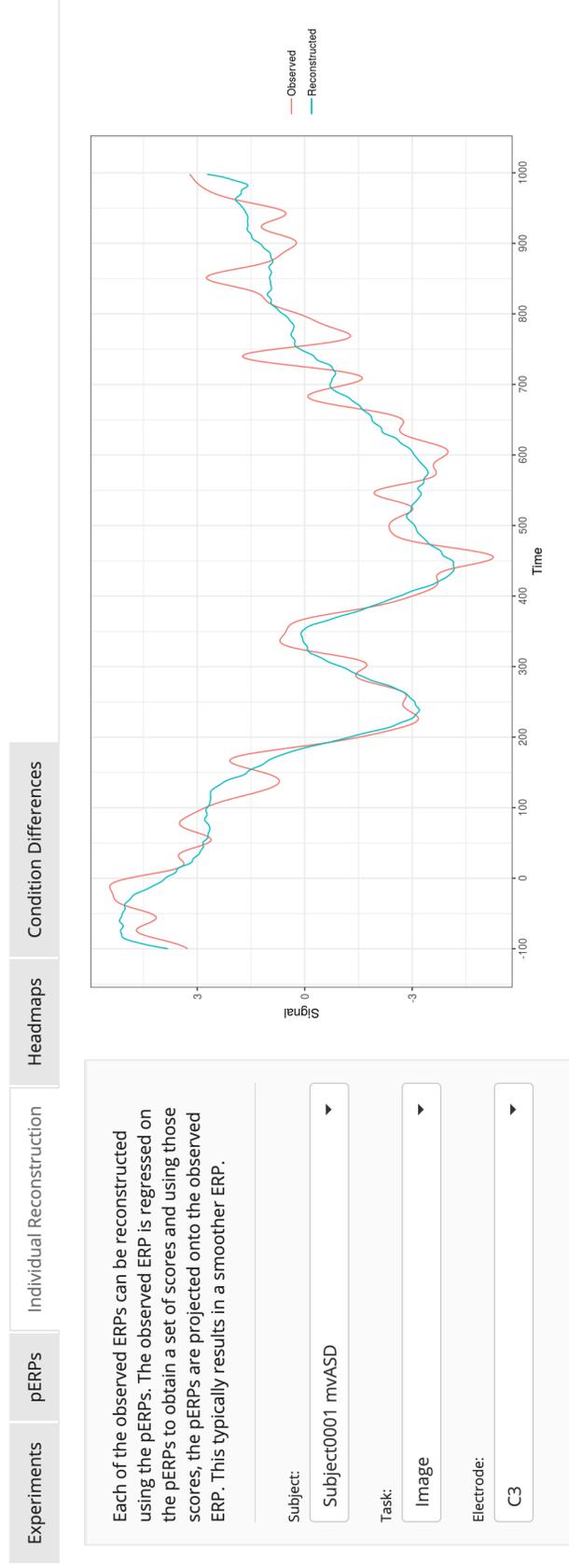


Figure A.4.2: Each individual observed ERP can be reconstructed using the weights calculated using regression.

ASD Data Exploration

Experiments
pERPs
Individual Reconstruction
Headmaps
Condition Differences

Task:

Standardized

Here, the scores from each task/condition were averaged across all subjects and plotted for each electrode to visualize where activity occurs on the scalp.

Select 'Standardized' if you would like to see the average scores divided by their standard deviations. In the 'Difference' headmaps, this would indicate a significant difference.

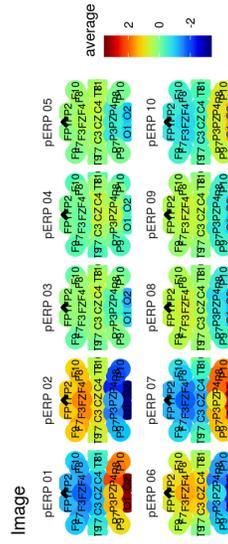


Figure A.4.3: Headmaps can be created in order to visualize where the pERPs are being seen across the scalp. In this case, it can be seen that pERPs 1, 2, 6, and 7 are loaded onto heavily for the Image task.

Appendix B

M-HPCA

B.1 Details of the M-HPCA estimation algorithm: Estimation of covariances and measurement error variance

- (a) Using the estimated mean functions, center the data $\hat{Y}_{dij}^*(r, t) = Y_{dij}(r, t) - \hat{\mu}(t) - \hat{\eta}_{dj}(r, t)$ and estimate the total covariance $\hat{K}_{d,Tot}\{(r, t), (r', t')\} = \frac{1}{Jn_d} \sum_{j=1}^J \sum_{i=1}^{n_d} \hat{Y}_{dij}^*(r, t) \hat{Y}_{dij}^*(r', t')$, the between-subject covariance $\hat{K}_{d,B}\{(r, t), (r', t')\} = \frac{1}{n_d} \sum_{i=1}^{n_d} \hat{Y}_{dij_1}^*(r, t) \hat{Y}_{dij_2}^*(r', t')$, and the within-subject covariance $\hat{K}_{d,W}\{(r, t), (r', t')\} = \hat{K}_{d,Tot}\{(r, t), (r', t')\} - \hat{K}_{d,B}\{(r, t), (r', t')\}$.
- (b) Estimate the raw marginal functional between- and within-subject covariances as the between- and within-subject covariances averaged over regions $\hat{\Sigma}_{d,\mathcal{T},B}(t, t') = \frac{1}{R} \sum_{r=1}^R \hat{K}_{d,B}\{(r, t), (r, t')\}$, $\hat{\Sigma}_{d,\mathcal{T},W}(t, t') = \frac{1}{R} \sum_{r=1}^R \hat{K}_{d,W}\{(r, t), (r, t')\}$. Estimate the raw marginal regional between- and within-subject covariance matrices as the between- and within-subject covariance averaged over time ($\hat{\Sigma}_{d,\mathcal{R},B}_{r,r'} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \hat{K}_{d,B}\{(r, t), (r', t)\}$, $(\hat{\Sigma}_{d,\mathcal{R},W})_{r,r'} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \hat{K}_{d,W}\{(r, t), (r', t)\}$, where $|\mathcal{T}|$ is the Lebesgue measure of \mathcal{T}).
- (c) Smooth the raw marginal functional between- and within-subject covariances estimated in step (b) using bivariate penalized splines with smoothing parameters selected through restricted maximum likelihood (REML) as proposed by Goldsmith et al.?. The diagonal elements are excluded in the smoothing of $\hat{\Sigma}_{d,\mathcal{T},W}(t, t')$ to avoid effects of measurement error.

Next, the measurement error variance σ_d^2 is estimated as the mean difference between the smoothed and unsmoothed diagonal elements of $\hat{\Sigma}_{d,\mathcal{T},W}(t, t')$. We use the estimated

measurement error variance to adjust the diagonal elements of the marginal regional within-subject covariance matrix. More specifically $(\tilde{\Sigma}_{d,\mathcal{R},W})_{r,r} = (\hat{\Sigma}_{d,\mathcal{R},W})_{r,r} - \hat{\sigma}_d^2$.

B.2 MM algorithm details

B.2.1 Derivatives

We provide here the partial derivatives of the proposed surrogate function with respect to the three variance components

$$\begin{aligned} \frac{\partial}{\partial \sigma_d^2} \sum_{i=1}^{n_d} f_{di} \left(\Lambda_d^{(1)}, \Lambda_d^{(2)}, \sigma_d^2 \mid \Lambda_d^{(1)(c)}, \Lambda_d^{(2)(c)}, \sigma_d^{2(c)} \right) &= \sum_{i=1}^{n_d} \sigma_d^2 \text{tr} \left(\Sigma_{di}^{-1(c)} \right) - \frac{\sigma_d^{4(c)}}{(\sigma_d^2)^2} \mathbf{Y}_{di}^T \Sigma_{di}^{-2(c)} \mathbf{Y}_{di}, \\ \frac{\partial}{\partial \Lambda_{d,gg}^{(1)}} \sum_{i=1}^{n_d} f_{di} \left(\Lambda_d^{(1)}, \Lambda_d^{(2)}, \sigma_d^2 \mid \Lambda_d^{(1)(c)}, \Lambda_d^{(2)(c)}, \sigma_d^{2(c)} \right) \\ &= \sum_{i=1}^{n_d} \text{vec} \left(\mathbf{Z}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{Z}_{di} - \Lambda_d^{-1(1)} \boldsymbol{\zeta}_{di}^{(c)} \boldsymbol{\zeta}_{di}^{(c)T} \Lambda_d^{-1(1)} \right)^T \frac{\partial}{\partial \Lambda_{d,gg}^{(1)}} \text{vec} \Lambda_d^{(1)}, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \Lambda_{d,hh}^{(2)}} \sum_{i=1}^{n_d} f_{di} \left(\Lambda_d^{(1)}, \Lambda_d^{(2)}, \sigma_d^2 \mid \Lambda_d^{(1)(c)}, \Lambda_d^{(2)(c)}, \sigma_d^{2(c)} \right) \\ = \sum_{i=1}^{n_d} \text{vec} \left(\mathbf{W}_{di}^T \Sigma_{di}^{-1(c)} \mathbf{W}_{di} - \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{-1(2)} \right) \boldsymbol{\xi}_{di}^{(c)} \boldsymbol{\xi}_{di}^{(c)T} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{-1(2)} \right) \right)^T \\ \frac{\partial}{\partial \Lambda_{d,hh}^{(2)}} \text{vec} \left(\mathbf{I}_{J_i} \otimes \Lambda_d^{(2)} \right), \end{aligned}$$

leading to the variance components estimators provided in Algorithm 1 of the main paper.

B.2.2 Determinant

Further computational savings are targeted in the last step of the algorithm in estimation of the log-likelihood. The log-likelihood contains the term $\log \det \boldsymbol{\Sigma}_{di}$ where

$$\begin{aligned}
& \det \boldsymbol{\Sigma}_{di} \\
&= \det \left\{ \sigma_d^2 \mathbf{I}_{TRJ_i} + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \mathbf{W}_{di}^T \right. \\
&\quad \left. + \mathbf{Z}_{di} \left(\boldsymbol{\Lambda}_d^{(1)} \right)^{1/2} \left(\boldsymbol{\Lambda}_d^{(1)} \right)^{1/2} \mathbf{Z}_{di}^T \right\} \\
&= \det \left\{ \sigma_d^2 \mathbf{I}_{TRJ_i} + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \mathbf{W}_{di}^T \right\} \\
&\quad \times \det \left\{ \mathbf{I}_G + \left(\boldsymbol{\Lambda}_d^{(1)} \right)^{1/2} \mathbf{Z}_{di}^T \left(\sigma_d^2 \mathbf{I}_{TRJ_i} + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \mathbf{W}_{di}^T \right)^{-1} \times \mathbf{Z}_{di} \left(\boldsymbol{\Lambda}_d^{(1)} \right)^{1/2} \right\}
\end{aligned} \tag{B.1}$$

using the identity $\det(\mathbf{A} + \mathbf{U}\mathbf{V}^T) = \det(\mathbf{A}) \det(\mathbf{I} + \mathbf{V}^T \mathbf{A}^{-1} \mathbf{U})$, and letting $\mathbf{A} = \sigma_d^2 \mathbf{I}_{TRJ_i} + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \mathbf{W}_{di}^T$, $\mathbf{U} = \mathbf{Z}_{di} \left(\boldsymbol{\Lambda}_d^{(1)} \right)^{1/2}$, and $\mathbf{V}^T = \left(\boldsymbol{\Lambda}_d^{(1)} \right)^{1/2} \mathbf{Z}_{di}^T$. Further simplifications are achieved via

$$\begin{aligned}
\det \boldsymbol{\Sigma}_{di} &= (\sigma^2)^{TRJ_i} \det \left\{ \mathbf{I}_{HJ_i} + \frac{1}{\sigma_d^2} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \mathbf{W}_{di}^T \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \right\} \times \\
&\quad \det \left(\mathbf{I}_G + \left(\boldsymbol{\Lambda}_d^{(1)} \right)^{1/2} \mathbf{Z}_{di}^T \left[\frac{1}{\sigma_d^2} \mathbf{I}_{TRJ_i} - \frac{1}{\sigma_d^2} \mathbf{W}_{di} \left\{ \sigma_d^2 \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{-1(2)} \right) + \mathbf{W}_{di}^T \mathbf{W}_{di} \right\}^{-1} \mathbf{W}_{di}^T \right] \times \right. \\
&\quad \left. \mathbf{Z}_{di} \left(\boldsymbol{\Lambda}_d^{(1)} \right)^{1/2} \right),
\end{aligned} \tag{B.2}$$

where the first determinant in (B.1) is further simplified by the determinant identity given above, letting $\mathbf{A} = \sigma_d^2 \mathbf{I}_{TRJ_i}$, $\mathbf{U} = \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2}$, and $\mathbf{V}^T = \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \mathbf{W}_{di}^T$. The Woodbury matrix identity $(\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{U} (\mathbf{C}^{-1} + \mathbf{VA}^{-1} \mathbf{U})^{-1} \mathbf{VA}^{-1}$ is used to simplify the inverse $\left(\sigma_d^2 \mathbf{I}_{TRJ_i} + \mathbf{W}_{di} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \left(\mathbf{I}_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right)^{1/2} \mathbf{W}_{di}^T \right)^{-1}$ in the second determinant of (B.1). Note that the simplified form in (B.2) yields computational savings where determinants are computed from lower dimensional matrices of size $HJ_i \times HJ_i$ and $G \times G$.

B.2.3 Likelihood Ratio Test for Weak Separability

Weak separability is tested via a likelihood ratio test for the random effects correlation structure of the subject- and subject-region-specific eigenscores, $\zeta_{di,kl}$ and $\xi_{dij,pm}$, in the proposed mixed effects modeling framework. Hence we propose to test the independent correlation structure against the alternative of a compound symmetric correlation structure for the eigenscores. Under the compound symmetric correlation structure of the random effects, the mixed effects model has the same form as in (3) where the matrices $\mathbf{\Lambda}_d^{(1)}$ and $\mathbf{\Lambda}_d^{(2)}$ now have $\lambda_g^{(1)}$ or $\lambda_h^{(2)}$ on the diagonal and $\rho^{(1)}$ or $\rho^{(2)}$ on the off-diagonal, respectively. The minorizing function and its derivatives with respect to the diagonal terms of the covariances remain the same as fitting the model under the weak separability assumption, where the partial derivatives with respect to σ_d^2 , $\Lambda_{d,gg}^{(1)}$, $\Lambda_{d,hh}^{(2)}$ are as given in Appendix B.2.1. The partial derivatives with respect to $\rho_d^{(1)}$ and $\rho_d^{(2)}$ are

$$\begin{aligned} & \frac{\partial}{\partial \rho^{(1)}} \sum_{i=1}^{n_d} f_{di} \left(\mathbf{\Lambda}_d^{(1)}, \mathbf{\Lambda}_d^{(2)}, \sigma_d^2 \mid \mathbf{\Lambda}_d^{(1)(c)}, \mathbf{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)} \right) \\ &= \sum_{i=1}^{n_d} \text{vec} \left(\mathbf{Z}_{di}^T \mathbf{\Sigma}_{di}^{-1(c)} \mathbf{Z}_{di} - \mathbf{\Lambda}_d^{-1(1)} \boldsymbol{\zeta}_{di}^{(c)} \boldsymbol{\zeta}_{di}^{(c)T} \mathbf{\Lambda}_d^{-1(1)} \right)^T \frac{\partial}{\partial \rho^{(1)}} \text{vec} \mathbf{\Lambda}_d^{(1)}, \end{aligned}$$

and

$$\begin{aligned} & \frac{\partial}{\partial \rho^{(2)}} \sum_{i=1}^{n_d} f_{di} \left(\mathbf{\Lambda}_d^{(1)}, \mathbf{\Lambda}_d^{(2)}, \sigma_d^2 \mid \mathbf{\Lambda}_d^{(1)(c)}, \mathbf{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)} \right) \\ &= \sum_{i=1}^{n_d} \text{vec} \left(\mathbf{W}_{di}^T \mathbf{\Sigma}_{di}^{-1(c)} \mathbf{W}_{di} - \left(\mathbf{I}_{J_i} \otimes \mathbf{\Lambda}_d^{-1(2)} \right) \boldsymbol{\xi}_{di}^{(c)} \boldsymbol{\xi}_{di}^{(c)T} \left(\mathbf{I}_{J_i} \otimes \mathbf{\Lambda}_d^{-1(2)} \right) \right)^T \times \\ & \quad \frac{\partial}{\partial \rho^{(2)}} \text{vec} \left(\mathbf{I}_{J_i} \otimes \mathbf{\Lambda}_d^{(2)} \right). \end{aligned}$$

Derivatives are set to zero and solved simultaneously using a Broyden update with the `nleqslv` package in R. The likelihood ratio test compares the log-likelihood under the null hypothesis of weak separability (ℓ_0) and the log-likelihood under the alternative hypothesis

of a compound symmetric correlation structure (ℓ_1) via the statistic is $-2(\ell_0 - \ell_1)$ which is distributed as $\chi(1)$ under the null.

B.3 Bootstrap tests for condition differentiation

B.3.1 Within-group condition differentiation

We first consider the null hypothesis that all conditions/repetitions have equal means within group d in the scalp region r , i.e. $H_0 : \eta_{dj_1}(r, t) = \eta_{dj_2}(r, t) = \dots = \eta_{dj_J}(r, t)$ or equivalently $H_0 : \eta_{dj}(r, t) = \eta_d(r, t)$ for $j = 1, \dots, J$. The proposed parametric bootstrap algorithm is outlined in Algorithm (3). Define $\hat{\eta}_d(r, t)$ as the point-wise average of $\hat{\eta}_{dj}(r, t)$ for $j = 1, \dots, J$. For each bootstrap sample b , the response is simulated under the null hypothesis based on the estimated model components as $Y_{dij}^b(r, t) = \hat{\mu}(t) + \hat{\eta}_d(r, t) + Z_{di}^b(r, t) + W_{dij}^b(r, t) + \epsilon_{dij}^b(r, t) = \hat{\mu}(t) + \hat{\eta}_d(r, t) + \sum_{g=1}^{G'} \zeta_{dig}^b \hat{\varphi}_{dig}^{(1)}(r, t) + \sum_{h=1}^{H'} \xi_{dijh}^b \hat{\varphi}_{dijh}^{(2)}(r, t) + \epsilon_{dij}^b(r, t)$ in region r and as $Y_{dij}^b(r', t) = \hat{\mu}(t) + \hat{\eta}_{dj}(r', t) + Z_{di}^b(r', t) + W_{dij}^b(r', t) + \epsilon_{dij}^b(r', t) = \hat{\mu}(t) + \hat{\eta}_{dj}(r', t) + \sum_{g=1}^{G'} \zeta_{dig}^b \hat{\varphi}_{dig}^{(1)}(r', t) + \sum_{h=1}^{H'} \xi_{dijh}^b \hat{\varphi}_{dijh}^{(2)}(r', t) + \epsilon_{dij}^b(r', t)$, in the other regions $r' \neq r$, where subject-specific scores, subject/repetition-specific scores, and measurement error are sampled from $\zeta_{dig}^b \sim N(0, \hat{\lambda}_{dg}^{(1)})$, $\xi_{dijh}^b \sim N(0, \hat{\lambda}_{dh}^{(2)})$ and $\epsilon_{dij}^b(r, t) \sim N(0, \hat{\sigma}_d^2)$, respectively. The proposed test statistic is given by $T_r = \left[\sum_{j=1}^J \int \{ \hat{\eta}_{dj}(r, t) - \hat{\eta}_d(r, t) \}^2 dt \right]^{1/2}$. Computation of the test statistic in the bootstrap samples provides the null distribution against which the test statistics T_r is compared, leading to a p-value.

Algorithm 3 Within-group condition differentiation

Input: For a fixed region, $r \in \{1, \dots, R\}$

Output: p-value

- 1: Generate B parametric bootstrap samples with sample sizes in each group identical to the observed data
 - 2: **for** $b = 1, \dots, B$ **do**
 - 3: $T_r^b = \sqrt{\sum_{j=1}^J \int \{ \hat{\eta}_{dj}^b(r, t) - \hat{\eta}_d^b(r, t) \}^2 dt}$
 - 4: **end for**
 - 5: Use $(1/B) \sum_{b=1}^B I(T_r^b > T_r)$ to estimate the p-value where $I(\cdot)$ denotes the indicator function and T_r is the test statistic from the original sample.
-

B.3.2 Between-group condition differentiation

The second test considered assesses whether the condition differentiation is the same across groups, i.e. $H_0 : \eta_{d_1j_1}(r, t) - \eta_{d_1j_2}(r, t) = \eta_{d_2j_1}(r, t) - \eta_{d_2j_2}(r, t)$. The proposed parametric bootstrap algorithm is provided in Algorithm (4). Let $\bar{\eta}_{j^-}(r, t) = \frac{1}{2}(\hat{\eta}_{d_1j^-}(r, t) + \hat{\eta}_{d_2j^-}(r, t))$ denote the averaged group-specific condition differences $\hat{\eta}_{dj^-}(r, t) = \hat{\eta}_{d_1j^-}(r, t) - \hat{\eta}_{d_2j^-}(r, t)$. For each bootstrap sample b , the response is simulated under the null hypothesis based on the estimated model components as $Y_{d_1j_1}^b(r, t) = \hat{\mu}(t) + \hat{\eta}_d(r, t) + \frac{1}{2}\bar{\eta}_{j^-}(r, t) + Z_{d_1}^b(r, t) + W_{d_1j_1}^b(r, t) + \epsilon_{d_1j_1}^b(r, t)$ for repetition 1 and $Y_{d_1j_2}^b(r, t) = \hat{\mu}(t) + \hat{\eta}_d(r, t) - \frac{1}{2}\bar{\eta}_{j^-}(r, t) + Z_{d_1}^b(r, t) + W_{d_1j_2}^b(r, t) + \epsilon_{d_1j_2}^b(r, t)$ for repetition 2 in region r and as $Y_{d_1j}^b(r', t) = \hat{\mu}(t) + \hat{\eta}_{dj}(r', t) + \sum_{g=1}^{G'} \zeta_{dig}^b \hat{\varphi}_{dig}^{(1)}(r', t) + \sum_{h=1}^{H'} \xi_{dijh}^b \hat{\varphi}_{dijh}^{(2)}(r', t) + \epsilon_{d_1j}^b(r', t)$ in the other regions $r' \neq r$. The proposed test statistic is given by $T_r = [\sum_{d=1}^2 \int \{\hat{\eta}_{dj^-}(r, t) - \bar{\eta}_{j^-}(r, t)\}^2 dt]^{1/2}$. The p-value is computed by comparing T_r to its distribution under the null, targeted by using the test statistics computed in the bootstrap samples.

Algorithm 4 Between-group condition differentiation

Input: For a fixed region, $r \in \{1, \dots, R\}$

Output: p-value

- 1: Generate B parametric bootstrap samples with sample sizes in each group identical to the observed data
 - 2: **for** $b = 1, \dots, B$ **do**
 - 3: $T_r^b = \sqrt{\sum_{d=1}^2 \int \{\hat{\eta}_{dj^-}^b(r, t) - \bar{\eta}_{j^-}^b(r, t)\}^2 dt}$
 - 4: **end for**
 - 5: Use $(1/B) \sum_{b=1}^B I(T_r^b > T_r)$ to estimate the p-value where $I(\cdot)$ denotes the indicator function and T_r is the test statistic from the original sample.
-

B.4 Simulation design

The response is generated from the model

$$\begin{aligned}
Y_{dij}(r, t) = & \mu(t) + \eta_{dj}(r, t) + \sum_{k=1}^2 \sum_{\ell=1}^2 \zeta_{di, k\ell} v_{dk}^{(1)}(r) \phi_{d\ell}^{(1)}(t) + \sum_{p=1}^2 \sum_{m=1}^2 \xi_{dij, pm} v_{dp}^{(2)}(r) \phi_{dm}^{(2)}(t) \\
& + \epsilon_{dij}(r, t),
\end{aligned} \tag{B.3}$$

with $K = L = P = M = 2$ for $J = 2$ repetitions, $R = 9$ regions, $D = 2$ groups and $i = 1, \dots, n_d$ subjects within each group. The functional grid is set to 50 equidistant points in $[0, 1]$. A total of eight simulation set-ups are considered: two sample sizes ($n_d = 15$ and 50), two noise levels ($\sigma_d^2 = 0.25$ and 1), and two sparsity levels (dense and sparse). The measurement error $\epsilon_{dij}(r, t)$ is simulated independently for subjects, repetitions, regions and time points, from a mean zero Gaussian distribution with variance σ_d^2 . The overall and group-region-repetition-specific mean functions are set to $\mu(t) = 5(t - 0.5)(t - 0.75)$, $\eta_{dj}(r, t) = 2(-1)^j \{t - (d - 1)\}^3$ for $d = 1, 2$ and $j = 1, 2$. The r -th element of the level 1 and 2 regional marginal eigenvectors equal $(v_{dk}^{(1)})_r = 1/2 \sin\{k\pi(r - 1)/8\}$, $(v_{dp}^{(2)})_r = 1/\sqrt{5} \cos\{p\pi(r - 1)/8\}$, respectively, for $r = 1, \dots, 9$, $k = 1, 2$ and $p = 1, 2$. The level 1 and 2 functional marginal eigenfunctions equal $\phi_{d\ell}^{(1)}(t) = \sqrt{2} \sin(\ell\pi t)$, $\phi_{dm}^{(2)}(t) = \sqrt{2} \cos(m\pi t)$, respectively, for $\ell = 1, 2$ and $m = 1, 2$. The subject-specific ($\zeta_{di, k\ell}$) and subject-repetition-specific scores ($\xi_{dij, pm}$) are simulated independently from a mean zero Gaussian distribution with variances $\tau_{d, k\ell} = 0.5^{(k-1)+K*(\ell-1)}$ and $\tau_{d, pm} = 0.5 \times 0.5^{(p-1)+P*(m-1)}$, respectively, for $k = 1, 2$, $\ell = 1, 2$, $p = 1, 2$ and $m = 1, 2$.

B.5 Visual Assessment of Weak Separability in Data Applications

The assumption of weak separability on the between- and within-subject covariance processes is assessed visually by comparing the first two leading eigenfunctions along the functional dimension at the participant/subject and condition/day/repetition levels across fixed slices of the regional dimension and the first leading eigenvector along the regional dimension at

Table B.1: Percentiles 50% (10%, 90%) of the relative squared errors, normalized mean squared errors, and both total and marginal FVE across groups for model components based on 200 Monte Carlo runs from the simulation design at $n_d = 15, 50$ for the **dense** simulation.

	Low Noise		High Noise	
	$n_d = 15$	$n_d = 50$	$n_d = 15$	$n_d = 50$
$\mu(t)$.004 (<.001, .018)	.001 (<.001, .006)	.005 (<.001, .017)	.002 (<.001, .006)
$\eta_{dj}(r, t)$.027 (.012, .061)	.008 (.004, .016)	.034 (.019, .065)	.010 (.006, .019)
$Y_{dij}(r, t)$.180 (.104, .533)	.182 (.104, .535)	.446 (.299, .801)	.446 (.299, .800)
$\phi_{d1}^{(1)}(t)$.014 (.002, .075)	.003 (<.001, .018)	.017 (.003, .076)	.004 (<.001, .020)
$\phi_{d2}^{(1)}(t)$.074 (.025, .192)	.024 (.013, .054)	.088 (.030, .220)	.028 (.015, .063)
$\phi_{d1}^{(2)}(t)$.011 (.001, .057)	.003 (<.001, .016)	.018 (.004, .073)	.005 (.001, .023)
$\phi_{d2}^{(2)}(t)$.020 (.006, .066)	.006 (.002, .018)	.063 (.020, .167)	.020 (.005, .044)
$\nu_{d1}^{(1)}(r)$.058 (.006, .416)	.014 (.001, .090)	.070 (.011, .443)	.017 (.003, .096)
$\nu_{d2}^{(1)}(r)$.091 (.022, .440)	.025 (.005, .095)	.140 (.050, .502)	.033 (.010, .112)
$\nu_{d1}^{(2)}(r)$.054 (.008, .383)	.015 (.002, .100)	.106 (.029, .583)	.031 (.010, .133)
$\nu_{d2}^{(2)}(r)$.074 (.019, .394)	.019 (.004, .102)	.328 (.109, .884)	.070 (.025, .174)
λ_{dg}	.059 (.003, .318)	.018 (<.001, .103)	.062 (.003, .366)	.018 (<.001, .113)
λ_{dh}	.035 (.001, .193)	.011 (<.001, .069)	.059 (.002, .325)	.014 (<.001, .092)
σ_d^2	.015 (.004, .031)	.016 (.004, .038)	<.001 (<.001, .002)	<.001 (<.001, .002)
ρ_{dW}	.004 (<.001, .023)	.001 (<.001, .007)	.005 (<.001, .031)	.001 (<.001, .007)

the participant and condition/day levels across fixed slices of the functional dimension in the audio odd-ball paradigm (Figures B.5.4, B.5.5, B.5.6) and the ABC-CT feasibility study (Figures B.5.7, B.5.8).

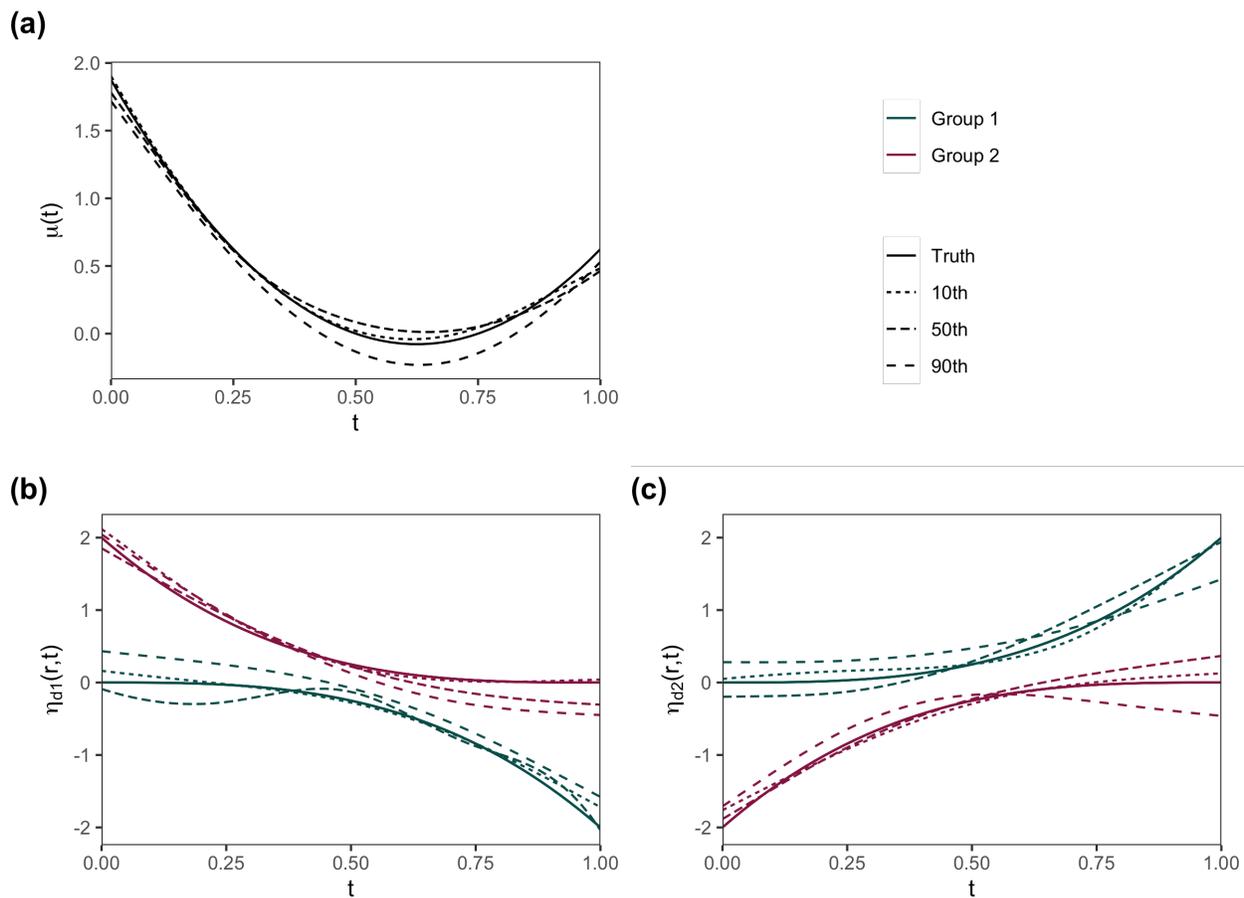


Figure B.5.1: (a) The true and estimated overall mean functions $\mu(t)$ corresponding to the 10th, 50th and 90th percentile RSE values across groups based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15$ and high noise. (b, c) The true and estimated $\eta_{dj}(r, t)$ corresponding to the 10th, 50th and 90th percentile RSE values across groups $d = 1, 2$ and regions based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15$ and high noise.

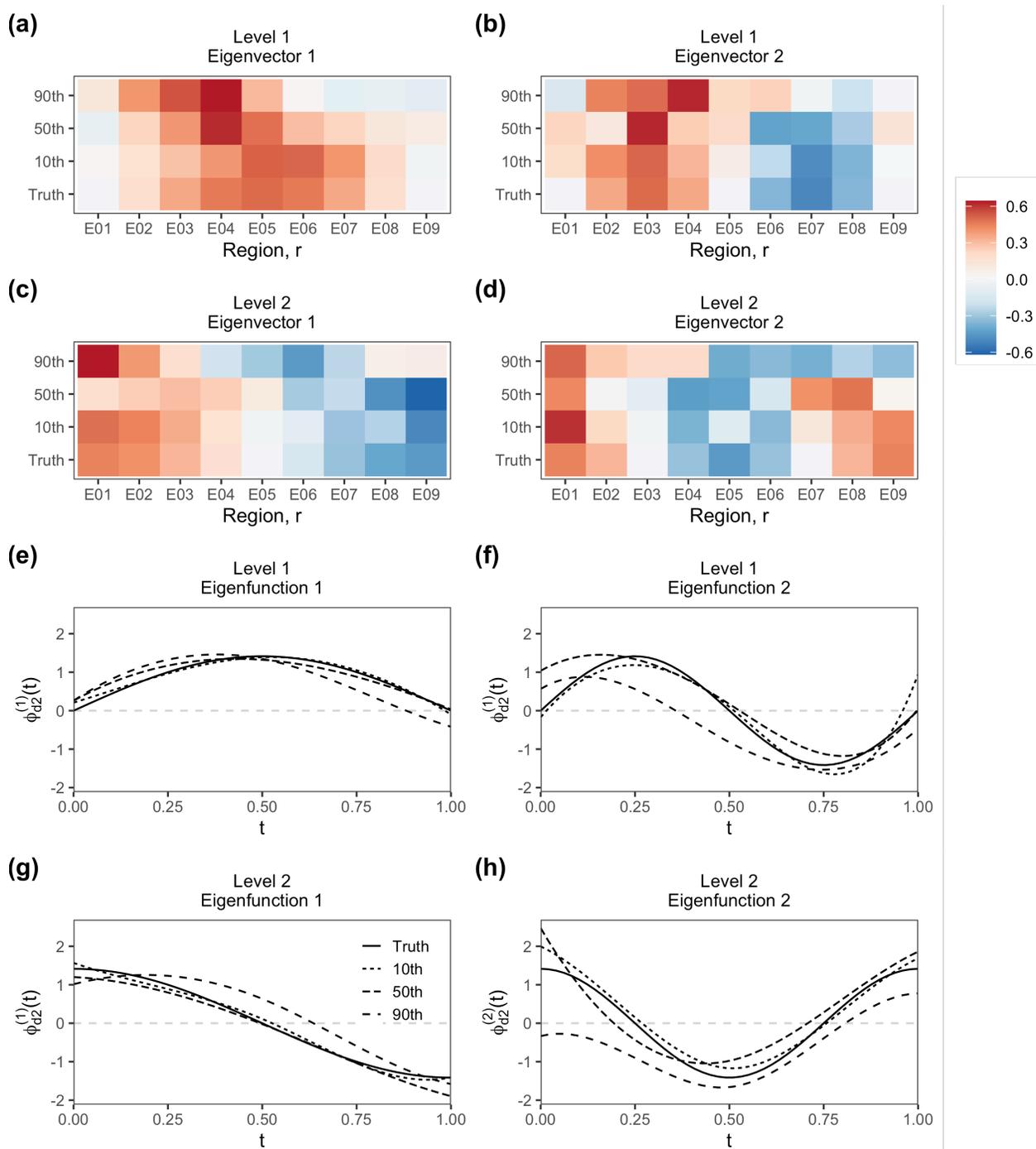


Figure B.5.2: The true and estimated first (first column) and second (second column) leading marginal eigenvectors for level 1 $v_{dk}^{(1)}(r)$ (first row) and level 2 $v_{dp}^{(2)}(r)$ (second row) and eigenfunctions $\phi_{dl}^{(1)}(t)$ (third row) and $\phi_{dm}^{(2)}(t)$ (fourth row) corresponding to the 10th, 50th and 90th percentile RSE values across groups based on 200 Monte Carlo runs from the sparse simulation design at $n_d = 15$ and high noise.

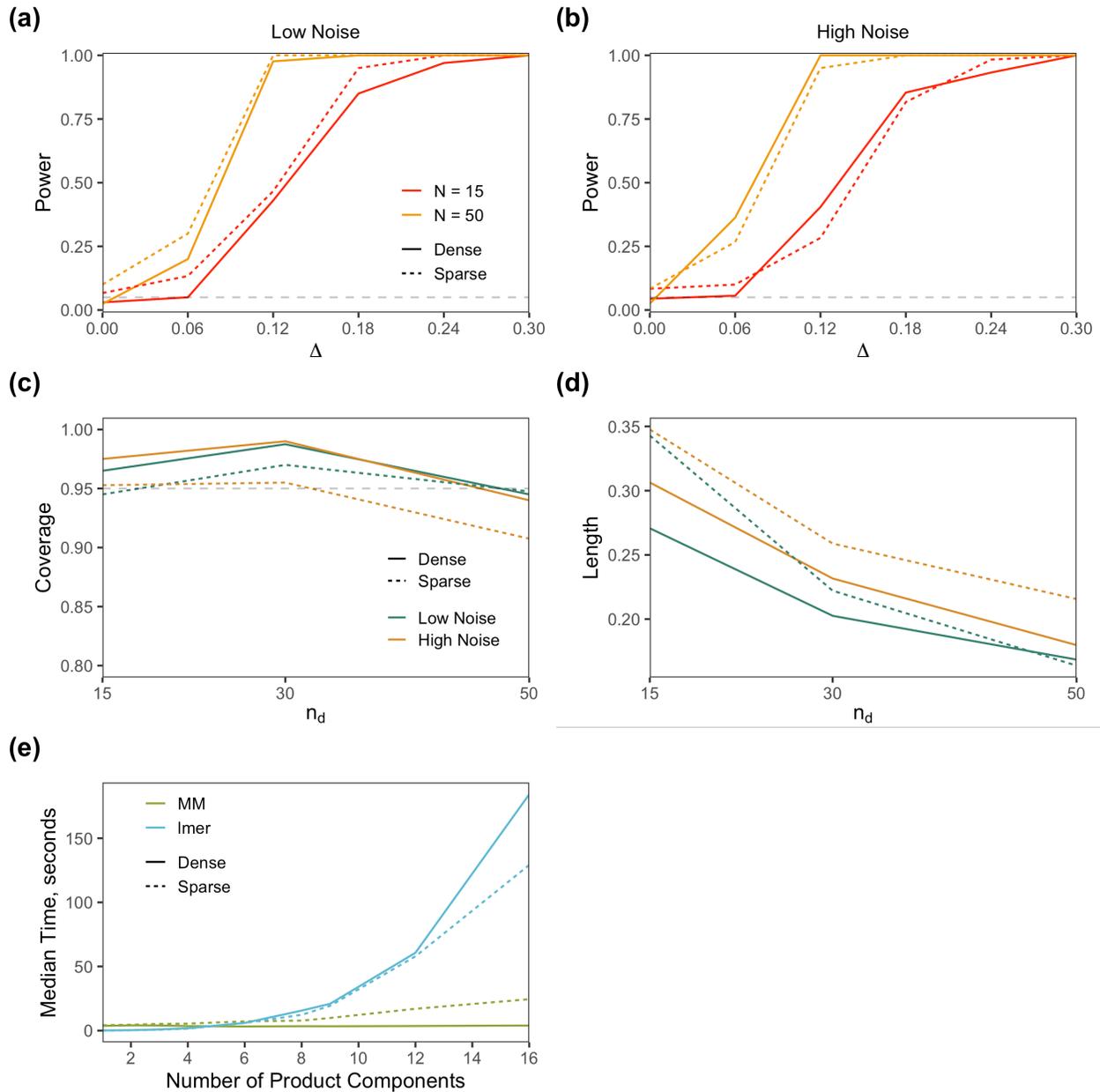


Figure B.5.3: (a, b) The estimated power of the proposed bootstrap test for condition differentiation based on 200 Monte Carlo runs. (c) Coverage and (d) length of the bootstrap CIs proposed for M-HPCA ICC based on 200 Monte Carlo runs. (e) The median run time for the proposed MM algorithm (green line) versus the popular R package `lme4` (red line) from the dense (solid line) and sparse (dashed line) simulation design at $n_d = 15$ and low noise, as a function of the total number of product components.

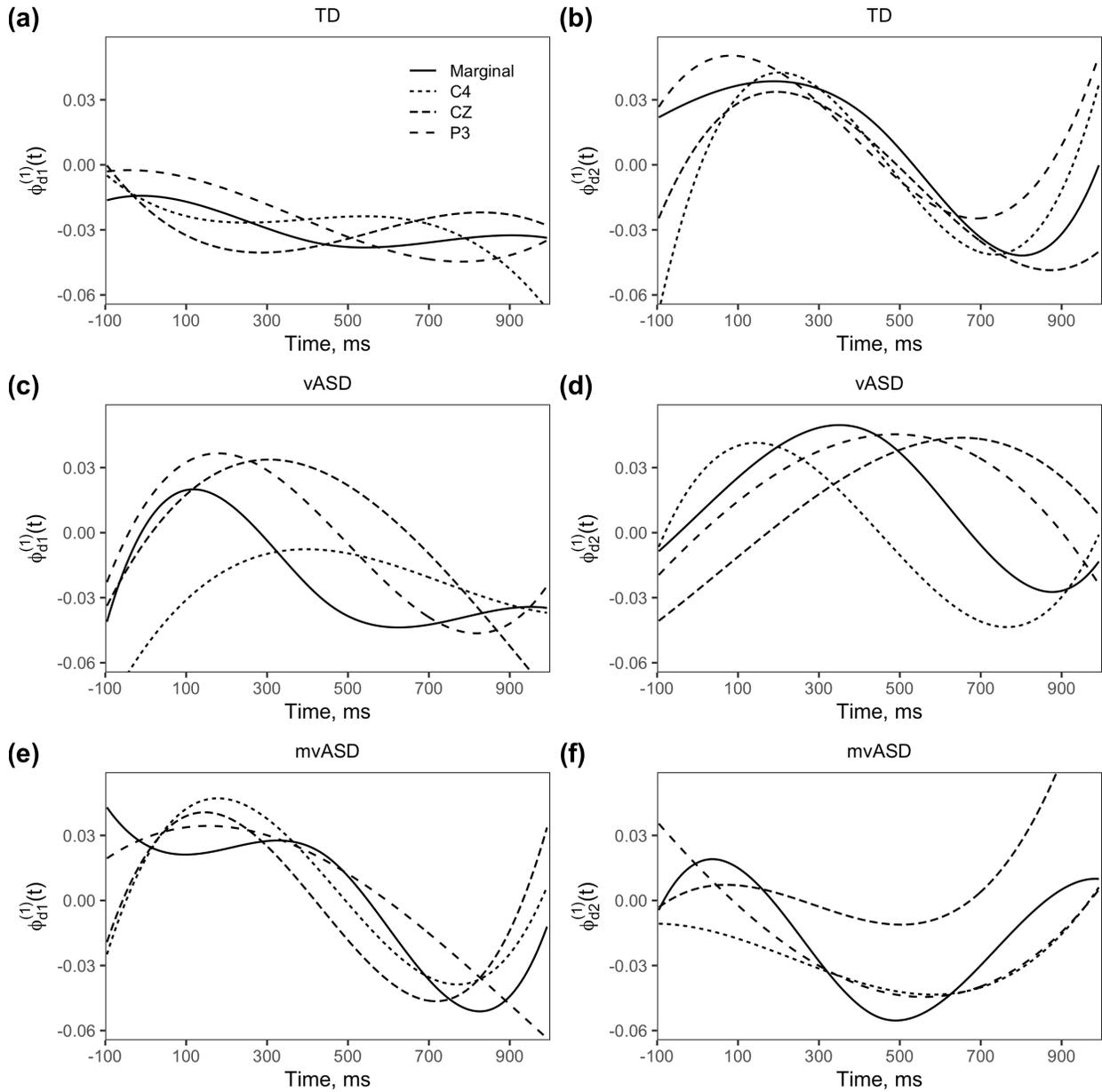


Figure B.5.4: (a, c, e) The estimated leading participant-level marginal eigenfunction $\phi_{d1}^{(1)}(t)$ (solid) compared to the leading participant-level eigenfunctions from the C4, CZ, and P3 regions. (b, d, f) The estimated second leading participant-level marginal eigenfunction $\phi_{d2}^{(1)}(t)$ (solid) compared to the second leading participant-level eigenfunctions from the C4, CZ, and P3 regions.

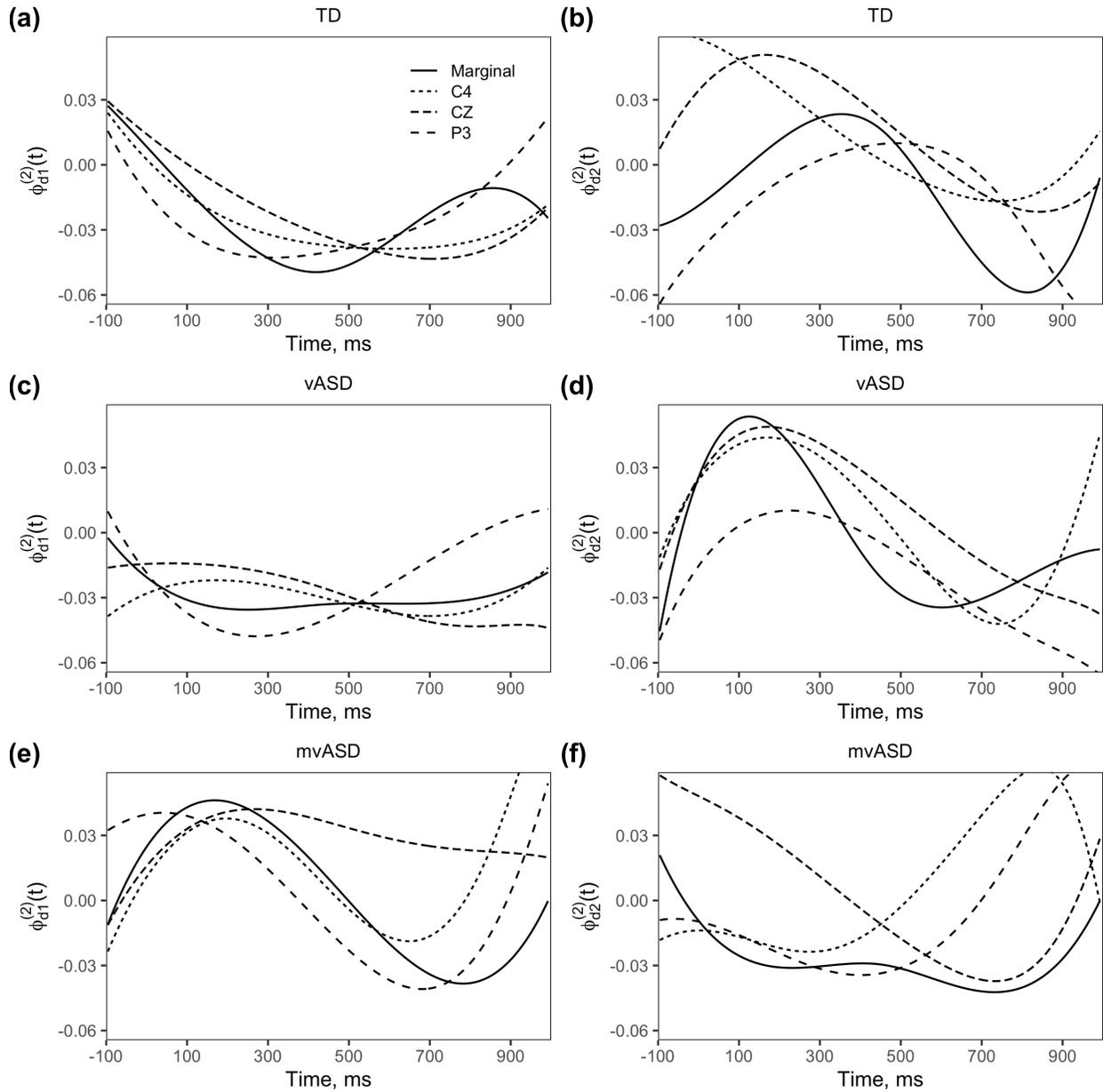


Figure B.5.5: (a, c, e) The estimated leading condition-level marginal eigenfunction $\phi_{d1}^{(2)}(t)$ (solid) compared to the leading condition-level eigenfunctions from the C4, CZ, and P3 regions. (b, d, f) The estimated second leading condition-level marginal eigenfunction $\phi_{d2}^{(2)}(t)$ (solid) compared to the second leading condition-level eigenfunctions from the C4, CZ, and P3 regions.

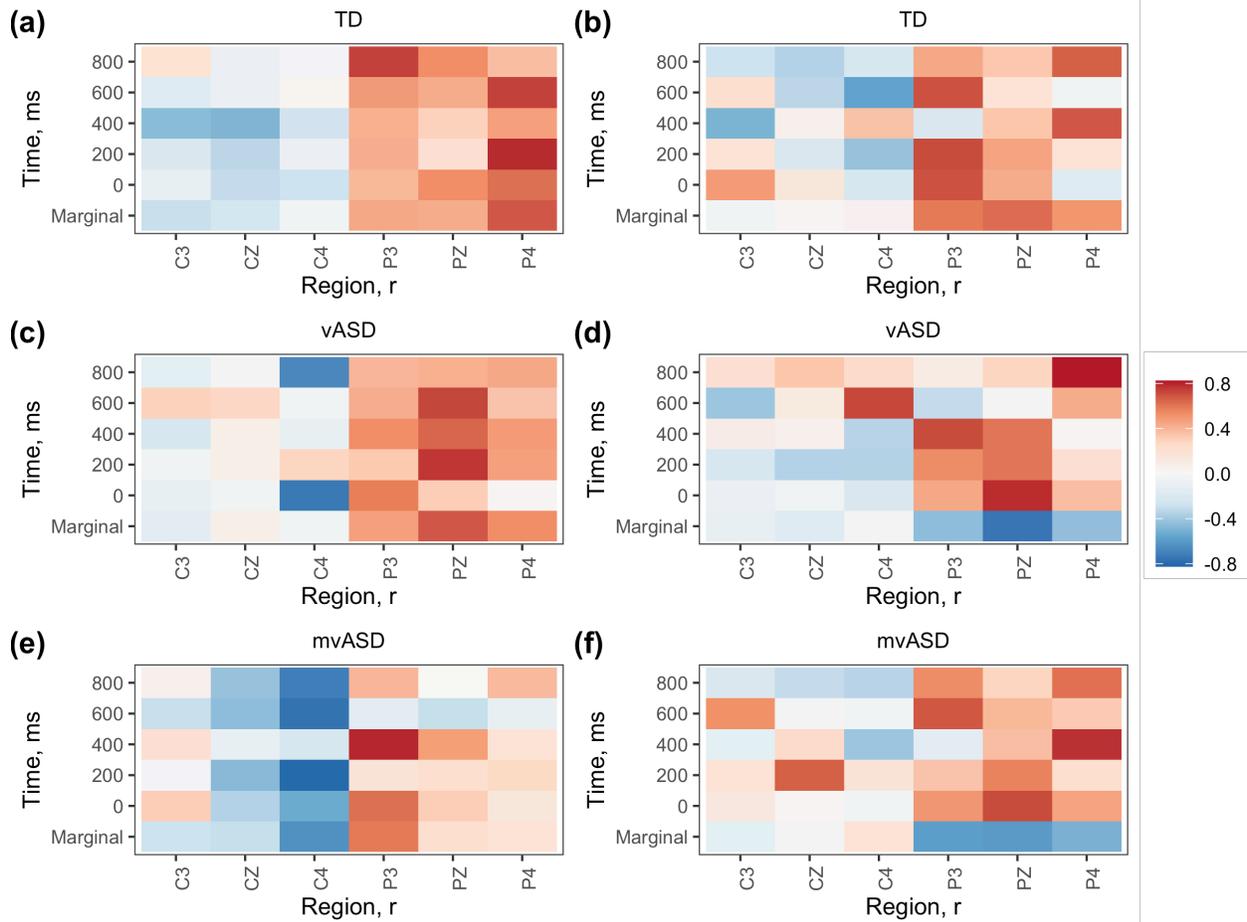


Figure B.5.6: (a, c, e) The estimated leading participant-level marginal eigenvectors $v_{d1}^{(1)}(r)$ (bottom row) compared to the leading participant-level eigenvectors obtained at fixed slices of $t = 0, 200, 400, 600, 800$. (b, d, f) The estimated leading condition-level marginal eigenvectors $v_{d1}^{(2)}(r)$ (bottom row) compared to the leading conditional-level eigenvectors obtained at fixed slices of $t = 0, 200, 400, 600, 800$.

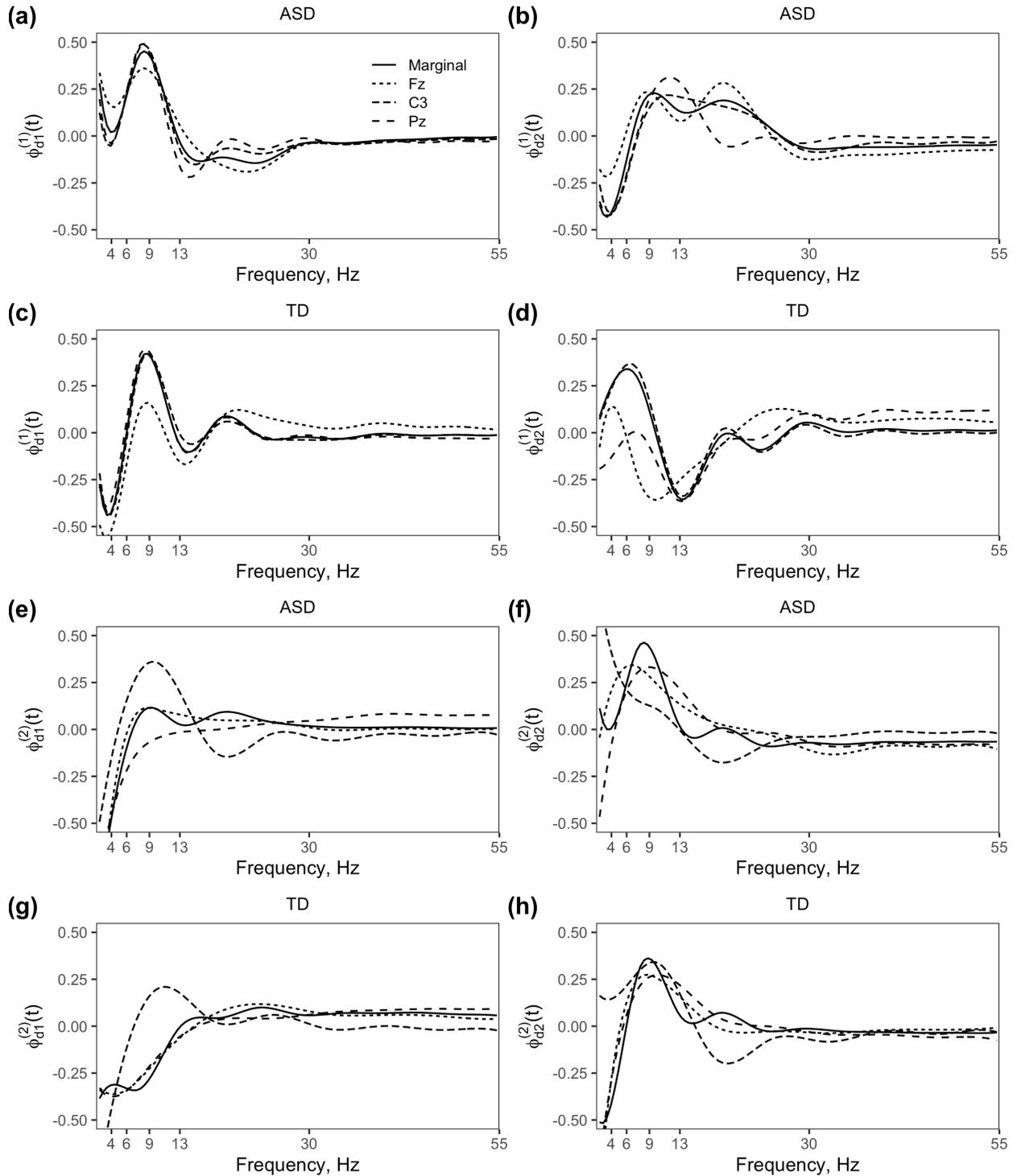


Figure B.5.7: The estimated leading participant-level (a, c) and second leading participant-level (b, d) marginal eigenfunctions $\phi_{d1}^{(1)}(t)$, $\phi_{d2}^{(1)}(t)$ (solid) compared to the participant-level eigenfunctions from the Fz, C3, and Pz regions. The estimated leading day-level (e, g) and second leading day-level (f, h) marginal eigenfunctions $\phi_{d1}^{(2)}(t)$, $\phi_{d2}^{(2)}(t)$ (solid) compared to the leading day-level eigenfunctions from the Fz, C3, and Pz regions.

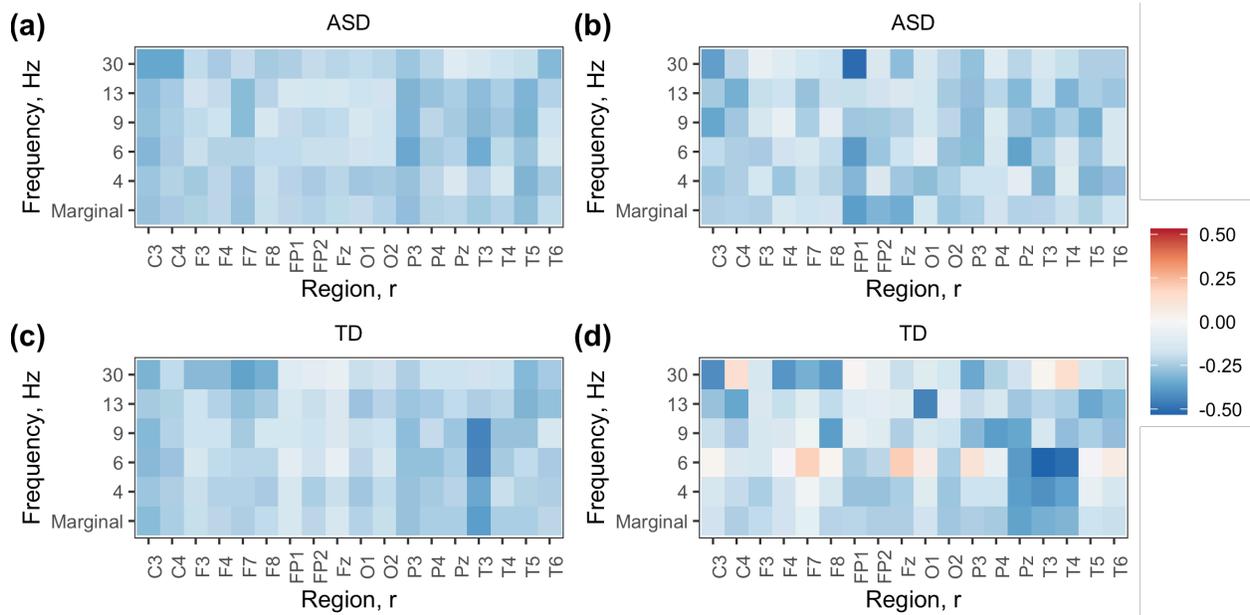


Figure B.5.8: (a, c) The estimated leading participant-level marginal eigenvector $v_{d1}^{(1)}(r)$ (bottom row) compared to the leading participant-level eigenvectors obtained at fixed slices of $t = 4, 6, 9, 13, 30$. (b, d) The estimated leading day-level marginal eigenvector $v_{d1}^{(2)}(r)$ (bottom row) compared to the leading day-level eigenvectors obtained at fixed slices of $t = 4, 6, 9, 13, 30$.

Appendix C

RR-GFLM

C.1 Simulation: Additional figures

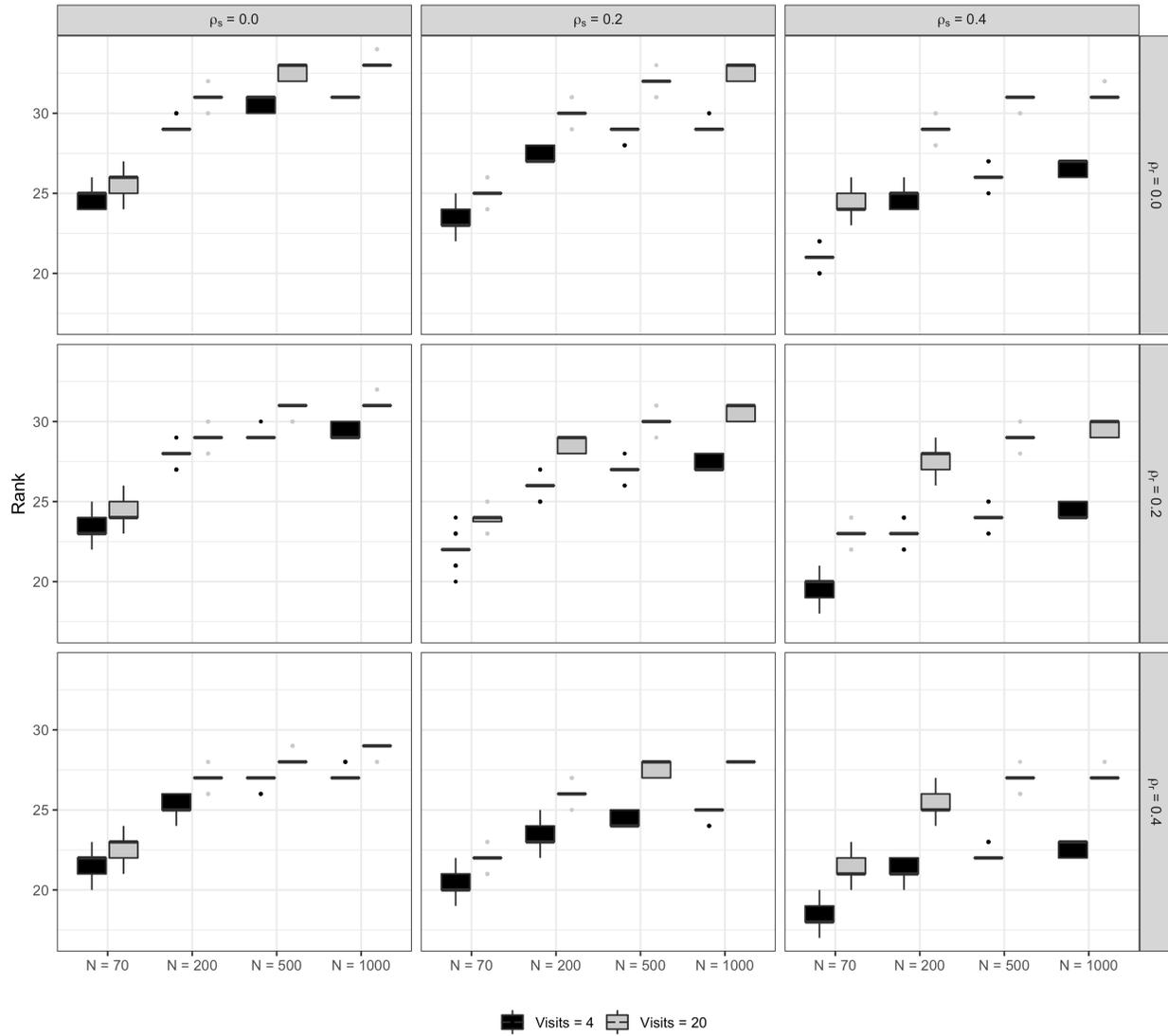


Figure C.1.1: The rank across 200 Monte Carlo runs for the discrete (black) and continuous (grey) longitudinal dimension cases.

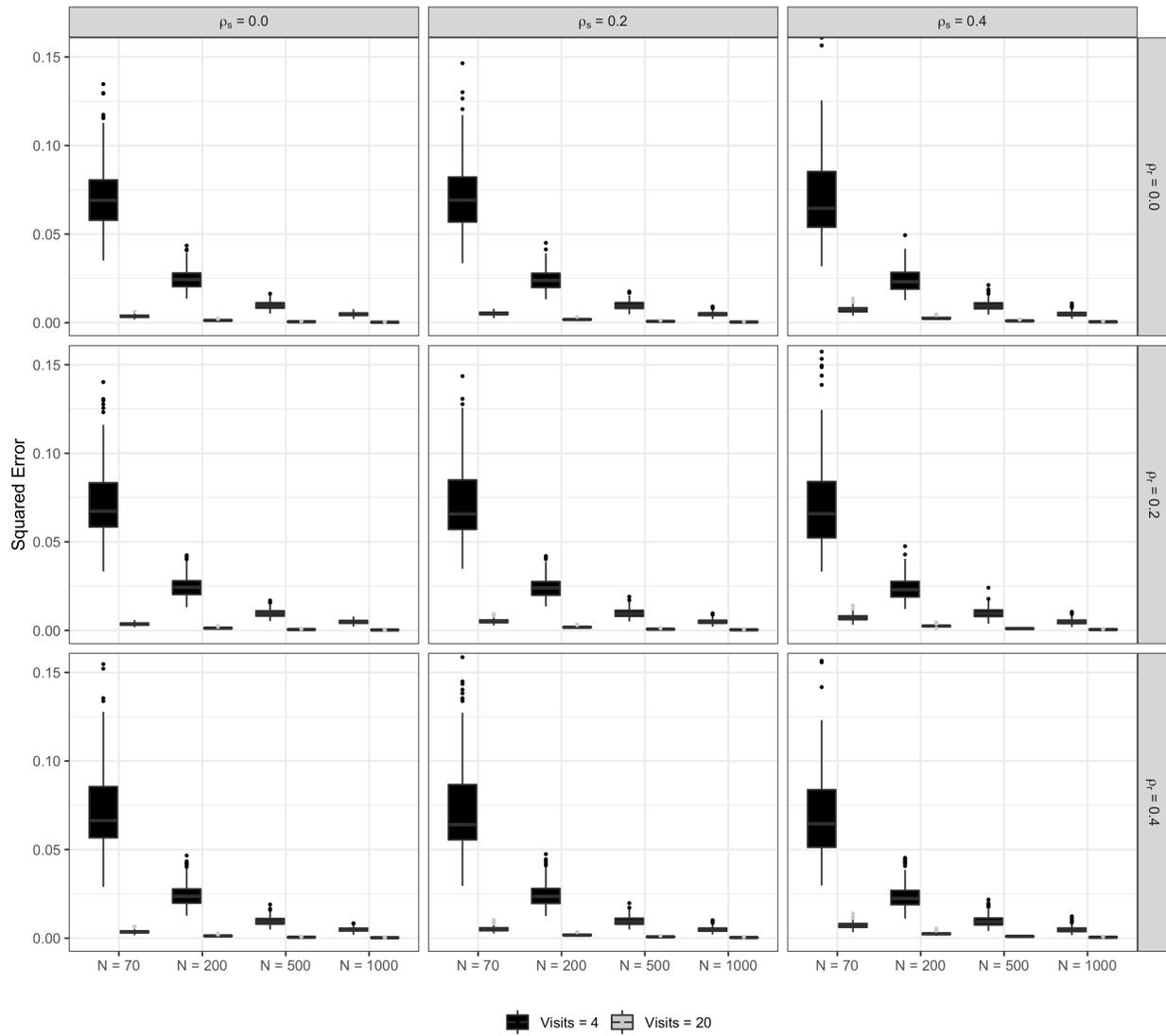


Figure C.1.2: The squared error across 200 Monte Carlo runs for the discrete (black) and continuous (grey) longitudinal dimension cases.

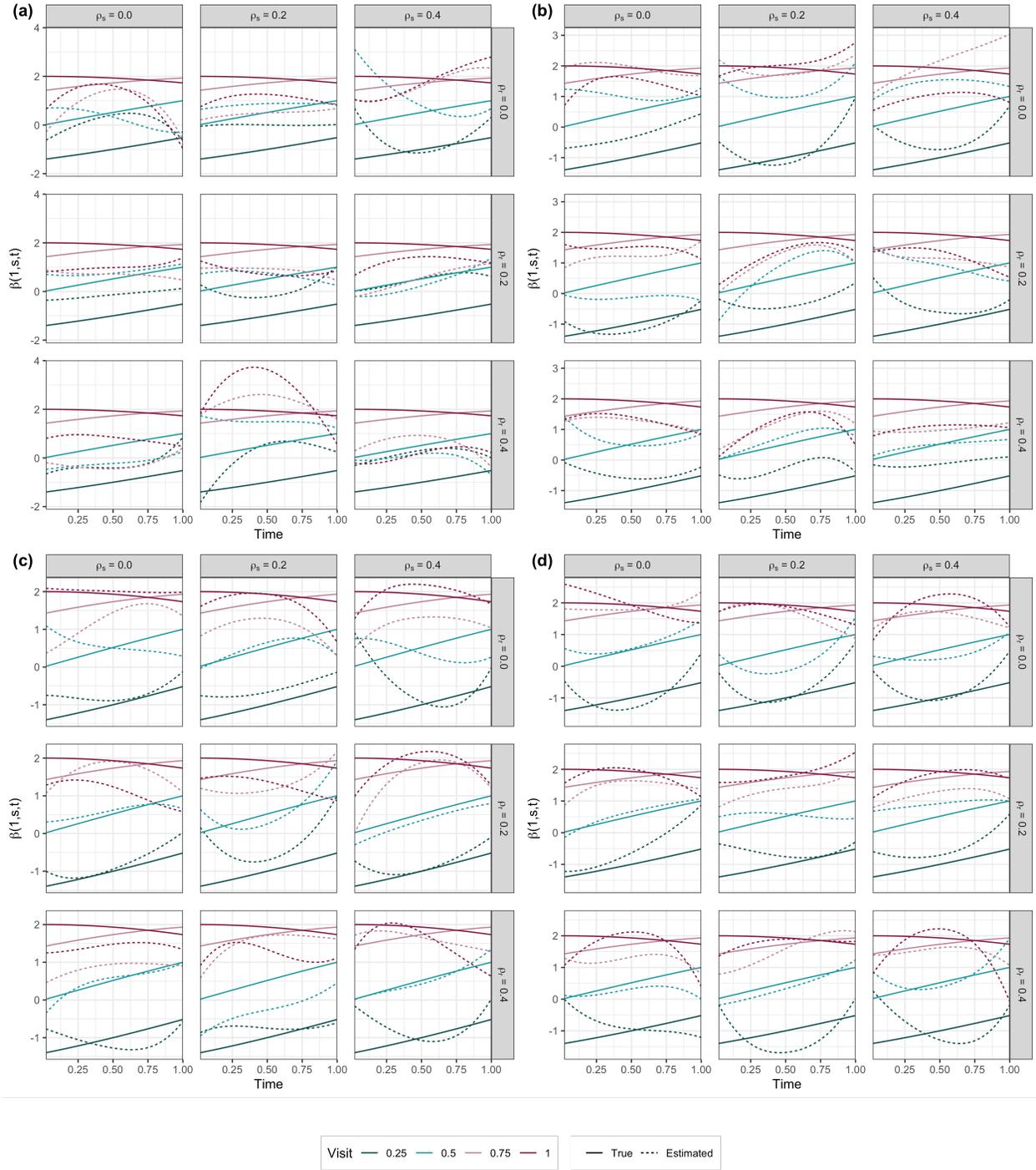


Figure C.1.3: The estimated regression function with median RSE (dashed) and true regression function (solid) for (a) $N = 70$, (b) $N = 200$, (c) $N = 500$, and (d) $N = 1000$.

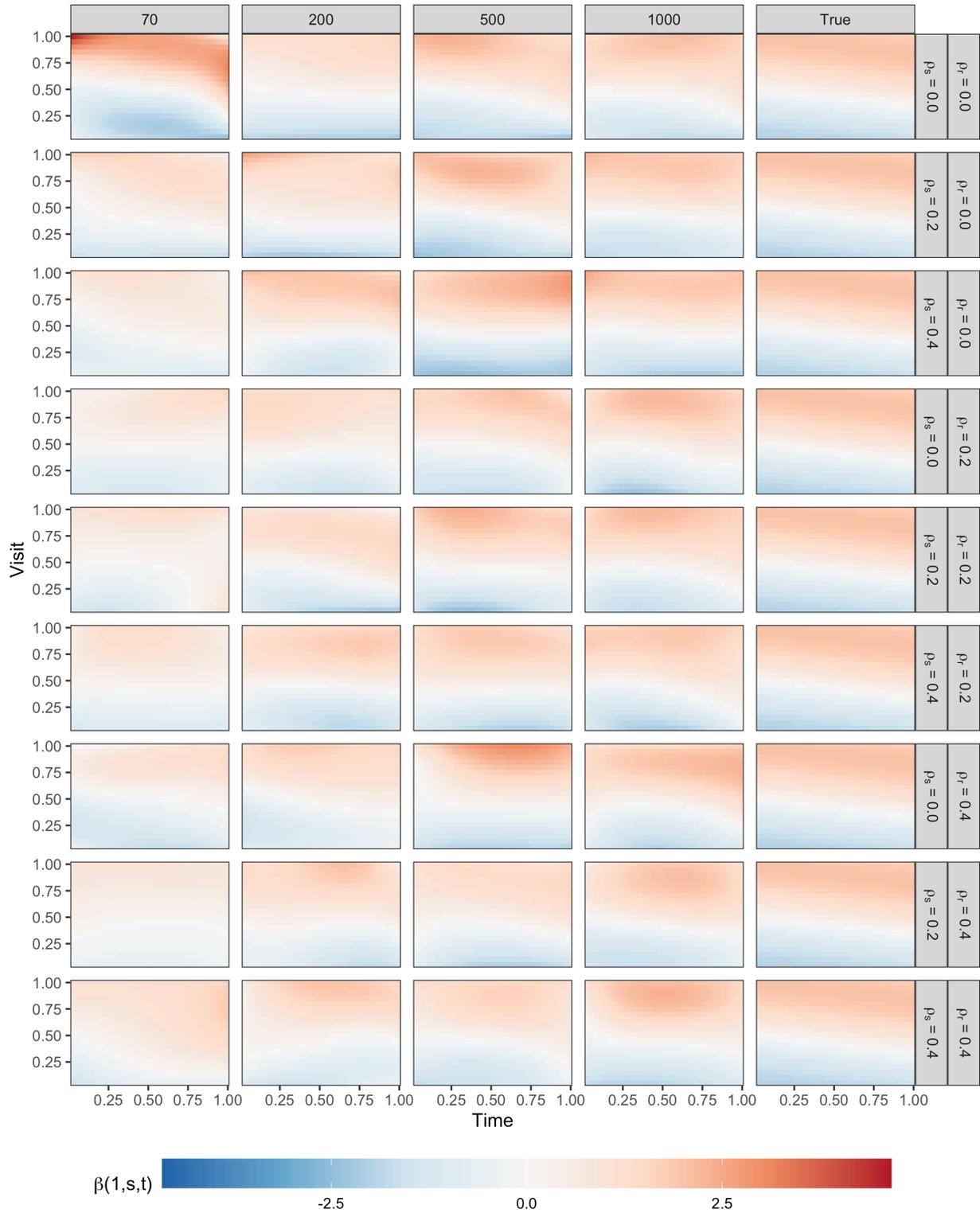


Figure C.1.4: The estimated regression function with median RSE (first through fourth columns) and true regression function (last column) for all simulation settings.

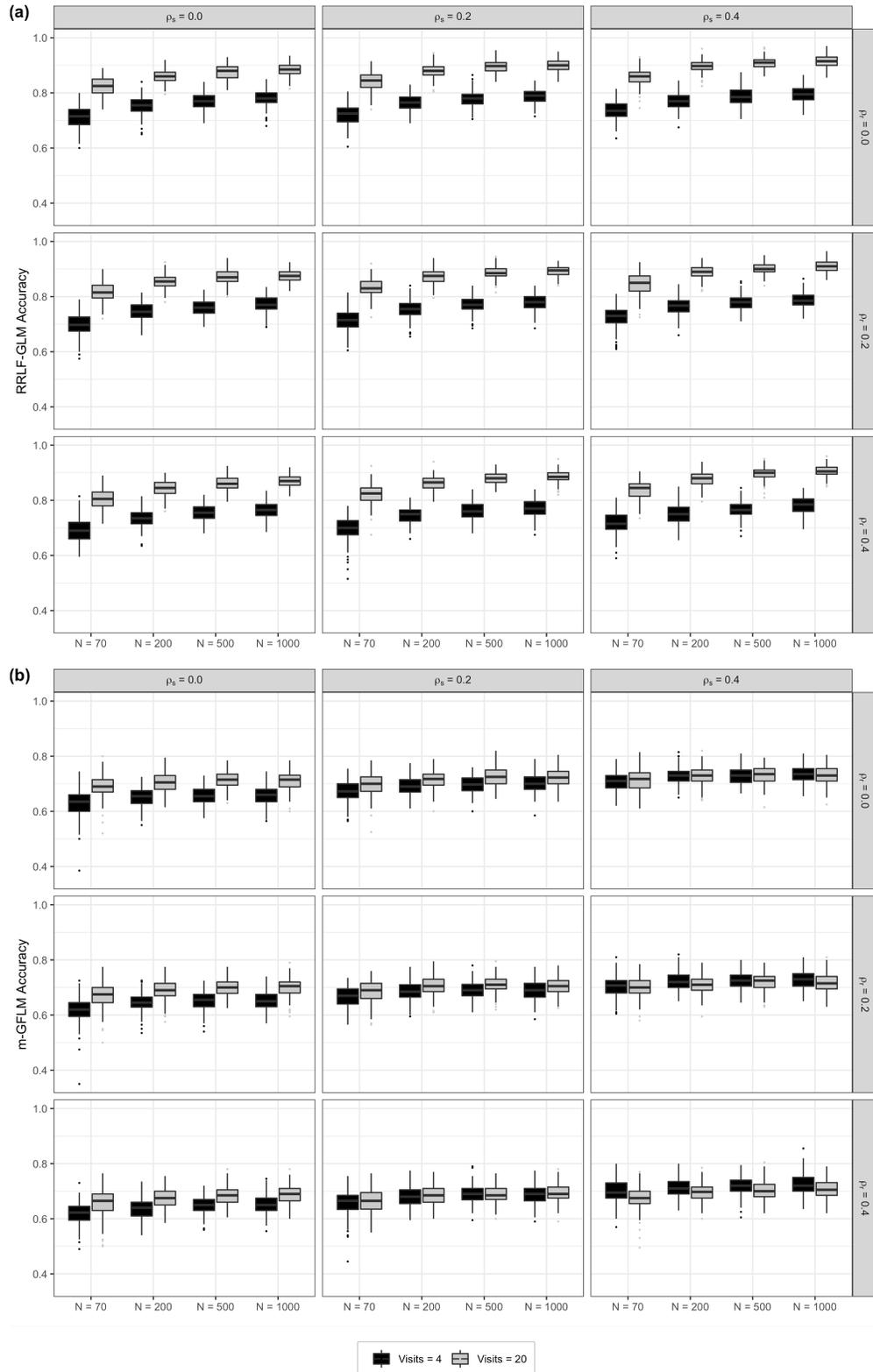


Figure C.15: The out-of-sample accuracy across 200 Monte Carlo runs for the (a) RRLF-GLM and (b) m-GFLM models for the discrete (black) and continuous (grey) longitudinal dimension cases.

BIBLIOGRAPHY

- Baladandayuthapani, Veerabhadran, Mallick, Bani K., Young Hong, Mee, Lupton, Joanne R., Turner, Nancy D., & Carroll, Raymond J. 2008. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics*, 64(1), 64–73.
- Boland, Joanna, Telesca, Donatello, Sugar, Catherine, Jeste, Shafali, Goldbeck, Cameron, & Şentürk, Damla. 2022. A study of longitudinal trends in time-frequency transformations of EEG data during a learning experiment. *Computational Statistics and Data Analysis*, 167, 107367.
- Cadonna, Annalisa, Kottas, Athanasios, & Prado, Raquel. 2019. Bayesian Spectral Modeling for Multiple Time Series. *Journal of the American Statistical Association*, 114(528), 1838–1853.
- Campos, Emilie, Scheffler, Aaron W., Telesca, Donatello, Sugar, Catherine, DiStefano, Charlotte, Jeste, Shafali, Levin, April R., Naples, Adam, Webb, Sara J., Shic, Frederick, Dawson, Geraldine, Faja, Susan, McPartland, James C., & Şentürk, Damla. 2022. Multilevel hybrid principal components analysis for regionreferenced functional electroencephalography data. *Statistics in Medicine*, 41(19), 3737–3757.
- Chen, Kehui, & Lei, Jing. 2015. Localized Functional Principal Component Analysis. *Journal of the American Statistical Association*, 100(511), 1266–1275.
- Chen, Kehui, Delicado, Pedro, & Müller, Hans-Georg. 2017. Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1), 177–196.
- Cong, Fengyu, He, Zhaoshui, Hämmäläinen, Jarmo, Leppänen, Paavo H.T., Lyytinen, Heikki, Cichocki, Andrzej, & Ristaniemi, Tapani. 2013. Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection. *Journal of Neuroscience Methods*, 212, 165–172.

- Crainiceanu, Ciprian M., & Goldsmith, A. Jeffrey. 2010. Bayesian functional data analysis using WinBUGS. Journal of Statistical Software, 32, 1–33.
- Crainiceanu, Ciprian M, Staicu, Ana-Maria, & Di, Chong-Zhi. 2009. Generalized Multilevel Functional Regression. Journal of the American Statistical Association, 104(488), 1550–1561.
- Cui, Erjia, Li, Ruonan, Crainiceanu, Ciprian M., & Xiao, Luo. 2022. Fast Multilevel Functional Principal Component Analysis. Journal of Computational and Graphical Statistics, aug, 1–33.
- Delorme, Arnaud, & Makeig, Scott. 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134, 9–21.
- Di, Chong Zhi, Crainiceanu, Ciprian M., Caffo, Brian S., & Punjabi, Naresh M. 2009. Multilevel functional principal component analysis. Annals of Applied Statistics, 3(1), 458–488.
- Di, Chongzhi, Crainiceanu, Ciprian M., & Jank, Wolfgang S. 2014. Multilevel sparse functional principal component analysis. Stat, 3(1), 126–143.
- Dickinson, Abigail, Campos, Emilie, Şentürk, Damla, & Jeste, Shafali. 2022. Atypical oscillatory development during infancy in ASD.
- DiStefano, Charlotte, Şentürk, Damla, & Jeste, Shafali Spurling. 2019. ERP evidence of semantic processing in children with ASD. Developmental Cognitive Neuroscience, 36, 100640.
- Eichele, Tom, Rachakonda, Srinivas, Brakedal, Brage, Eikeland, Rune, & Calhoun, Vince D. 2011. EEGIFT: Group Independent Component Analysis for Event-Related EEG Data. Computational Intelligence and Neuroscience, 2011, 129365.

- Fiecas, Mark, & Ombao, Hernando. 2016. Modeling the Evolution of Dynamic Brain Processes During an Associative Learning Experiment. Journal of the American Statistical Association, 111(516), 1440–1453.
- Gertheiss, Jan, Maity, Arnab, & Staicu, Ana-Maria. 2013. Variable selection in generalized functional linear models. Stat, 2(1), 86–101.
- Giraldo, R., Delicado, P., & Mateu, J. 2010. Continuous time-varying kriging for spatial prediction of functional data: An environmental application. Journal of Agricultural, Biological, and Environmental Statistics, 15(1), 66–82.
- Goldsmith, Jeff, Bobb, Jennifer, Crainiceanu, Ciprian M., Caffo, Brian, & Reich, Daniel. 2011. Penalized Functional Regression. Journal of Computational and Graphical Statistics, 20(4), 830–851.
- Goldsmith, Jeff, Crainiceanu, Ciprian M, Caffo, Brian, & Reich, Daniel. 2012. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. Tech. rept.
- Greven, Sonja, Crainiceanu, Ciprian, Caffo, Brian, & Reich, Daniel. 2010. Longitudinal functional principal component analysis. Electronic Journal of Statistics, 4(jan), 1022–1054.
- Happ, Clara, & Greven, Sonja. 2018. Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. Journal of the American Statistical Association, 113(522), 649–659.
- Hart, Brian, Guindani, Michele, Malone, Stephen, & Fiecas, Mark. 2020. A nonparametric Bayesian model for estimating spectral densities of resting-state EEG twin data. Biometrics, 1–11.
- Hasenstab, Kyle, Sugar, Catherine A., Telesca, Donatello, Mcevoy, Kevin, Jeste, Shafali, & Şentürk, Damla. 2015. Identifying longitudinal trends within EEG experiments. Biometrics, 71(4), 1090–1100.

- Hasenstab, Kyle, Scheffler, Aaron, Telesca, Donatello, Sugar, Catherine A., Jeste, Shafali, DiStefano, Charlotte, & Şentürk, Damla. 2017. A multi-dimensional functional principal components analysis of EEG data. Biometrics, 73(3), 999–1009.
- Hastie, Trevor, & Mallows, Colin. 1993. A Statistical View of Some Chemometrics Regression Tools: Discussion. Technometrics, 35(2), 140–143.
- Huster, René J., & Raud, Liisa. 2018. A Tutorial Review on Multi-subject Decomposition of EEG. Brain Topography, 31, 3–16.
- Kairov, Ulykbek, Cantini, Laura, Greco, Alessandro, Molkenov, Askhat, Czerwinska, Urszula, Barillot, Emmanuel, & Zinovyev, Andrei. 2017. Determining the optimal number of independent components for reproducible transcriptomic data analysis. BMC Genomics, 18, 1–13.
- Karhunen, K. 1946. Zur Spektraltheorie stochastischer Prozesse. *Annales Academiae scientiarum Fennicae. Series A. 1, Mathematica-physica.*
- Krafty, Robert T., Hall, Martica, & Guo, Wensheng. 2011. Functional mixed effects spectral analysis. Biometrika, 98(3), 583–598.
- Kumar, J. Satheesh, & Bhuvaneshwari, P. 2012. Analysis of Electroencephalography (EEG) Signals and Its Categorization A Study. Procedia Engineering, 38, 2525–2536.
- Kundu, Madan G., Harezlak, Jaroslaw, & Randolph, Timothy W. 2016. Longitudinal functional models with structured penalties. Statistical Modelling, 16(2), 114–139.
- Lenartowicz, Agatha, Truong, Holly, Salgari, Giulia C., Bilder, Robert M., McGough, James, McCracken, James T., & Loo, Sandra K. 2019. Alpha modulation during working memory encoding predicts neurocognitive impairment in ADHD. Journal of Child Psychology and Psychiatry and Allied Disciplines, 60(8), 917–926.
- Levin, April R., Naples, Adam J., Scheffler, Aaron Wolfe, Webb, Sara J., Shic, Frederick, Sugar, Catherine A., Murias, Michael, Bernier, Raphael A., Chawarska, Katarzyna, Daw-

- son, Geraldine, Faja, Susan, Jeste, Shafali, Nelson, Charles A., McPartland, James C., & entürk, Damla. 2020. Day-to-Day Test-Retest Reliability of EEG Profiles in Children With Autism Spectrum Disorder and Typical Development. Frontiers in Integrative Neuroscience, 14(April), 1–12.
- Liu, Chong, Ray, Surajit, & Hooker, Giles. 2017. Functional principal component analysis of spatially correlated data. Statistics and Computing, 27, 1639–1654.
- Loeve, Michel. 1946. Fonctions aléatoires du second ordre. Vol. 84. La Revue Scientifique.
- Luck, Steven J. 2014. An Introduction to the Event-Related Potential Technique. 2 edn. MIT Press.
- Luck, Steven J., & Kappenman, Emily S. 2011. The Oxford Handbook of Event-Related Potential Components. Oxford University Press.
- Lynch, Brian, & Chen, Kehui. 2018. A test of weak separability for multi-way functional data, with application to brain connectivity studies. arXiv:1703.10210.
- Macaro, Christian, & Prado, Raquel. 2014. Spectral Decompositions of Multiple Time Series: A Bayesian Non-parametric Approach. Psychometrika, 79(1), 105–129.
- Makeig, Scott, Ca, San Diego, Bell, Anthony J, & Sejnowski, Terrence J. 1995. Independent Component Analysis of Electroencephalographic Data. Pages 145–151 of: Proceedings of the 8th International Conference on Neural Information Processing Systems. Denver, Colorado: MIT Press.
- Marx, Brian D., & Eilers, Paul H.C. 1999. Generalized linear regression on sampled signals and curves: A p-spline approach. Technometrics, 41(1), 1–13.
- McPartland, James C., Bernier, Raphael A., Jeste, Shafali S., Dawson, Geraldine, Nelson, Charles A., Chawarska, Katarzyna, Earl, Rachel, Faja, Susan, Johnson, Scott P., Sikich, Linmarie, Brandt, Cynthia A., Dziura, James D., Rozenblit, Leon, Hellemann, Gerhard, Levin, April R., Murias, Michael, Naples, Adam J., Platt, Michael L., Sabatos-DeVito,

- Maura, Shic, Frederick, Şentürk, Damla, Sugar, Catherine A., & Webb, Sara J. 2020. The Autism Biomarkers Consortium for Clinical Trials (ABC-CT): Scientific Context, Study Design, and Progress Toward Biomarker Qualification. Frontiers in Integrative Neuroscience, 14(April), 1–7.
- Mercer, James. 1909. Functions of Positive and Negative Type, and their Connection the Theory of Integral Equations. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 209, 415–446.
- Morris, Jeffrey S. 2015. Functional Regression. Annual Review of Statistics and Its Application, 2(1), 321–359.
- Morris, Jeffrey S., & Carroll, Raymond J. 2006. Wavelet-based functional mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(2), 179–199.
- Morris, Jeffrey S., Vannucci, Marina, Brown, Philip J., & Carroll, Raymond J. 2003. Wavelet-Based Nonparametric Modeling of Hierarchical Functions in Colon Carcinogenesis. Journal of the American Statistical Association, 98(463), 573–583.
- Ombao, Hernando, Fiecas, Mark, Ting, Chee Ming, & Low, Yin Fen. 2018. Statistical models for brain signals with properties that evolve across trials. NeuroImage, 180, 609–618.
- Park, So Young, & Staicu, Ana-Maria. 2015. Longitudinal functional data analysis. Stat, 4(1), 212–226.
- Perrin, F, Pernier, J, Bertrand, O, & Echallier, J.F. 1989. Spherical splines for scalp potential and current density mapping. Electroencephography and clinical Neurophysiology, 72, 184–187.
- R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

- Ramsay, J. O., & Dalzell, C.J. 1991. Some Tools for Functional Data Analysis. Journal of the Royal Statistical Society: Series B (Methodological), 53, 539–572.
- Ramsay, J. O., & Silverman, B. W. 2005. Functional Data Analysis. Springer.
- Reiss, Philip T., & Ogden, R. Todd. 2007. Functional Principal Component Regression and Functional Partial Least Squares. Journal of the American Statistical Association, 102(9), 984–996.
- Reiss, Philip T., & Ogden, R. Todd. 2010. Functional Generalized Linear Models with Images as Predictors. Biometrics, 66(1), 61–69.
- Reiss, Philip T, Goldsmith, Jeff, Shang, Han Lin, & Ogden, R Todd. 2017. Methods for Scalar-on-Function Regression. International Statistical Review, 85(8), 228–249.
- Scheffler, Aaron, Telesca, Donatello, Li, Qian, Sugar, Catherine A., Distefano, Charlotte, Jeste, Shafali, & Şentürk, Damla. 2020. Hybrid principal components analysis for region-referenced longitudinal functional EEG data. Biostatistics, 21(1), 139–157.
- Scheffler, Aaron W., Telesca, Donatello, Sugar, Catherine A., Jeste, Shafali, Dickinson, Abigail, DiStefano, Charlotte, & Şentürk, Damla. 2019. Covariate-adjusted region-referenced generalized functional linear model for EEG data. Statistics in Medicine, 38(30), 5587–5602.
- Shou, Haochang, Zipunnikov, Vadim, Crainiceanu, Ciprian M., & Greven, Sonja. 2015. Structured functional principal component analysis. Biometrics, 71(1), 247–257.
- Spencer, Kevin M., Dien, Joseph, & Donchin, Emanuel. 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. Psychophysiology, 38(2), 343–358.
- Staicu, Ana Maria, Crainiceanu, Ciprian M., & Carroll, Raymond J. 2010. Fast methods for spatially correlated multilevel functional data. Biostatistics, 11(2), 177–194.

- Staicu, Ana Maria, Islam, Md Nazmul, Dumitru, Raluca, & van Heugten, Eric. 2020. Longitudinal dynamic functional regression. Journal of the Royal Statistical Society. Series C: Applied Statistics, 69(1), 25–46.
- Sun, Yifan, & Wang, Qihua. 2022. An adaptive group LASSO approach for domain selection in functional generalized linear models. Journal of Statistical Planning and Inference, 219, 13–32.
- Wang, Jane-Ling, Chiou, Jeng-Min, & Mller, Hans-Georg. 2016. Review of functional data analysis. Annual Review of Statistics and Its Application, 3, 257–295.
- Webb, Sara Jane, Shic, Frederick, Murias, Michael, Sugar, Catherine A., Naples, Adam J., Barney, Erin, Borland, Heather, Hellemann, Gerhard, Johnson, Scott, Kim, Minah, Levin, April R., Sabatos-DeVito, Maura, Santhosh, Megha, Şentürk, Damla, Dziura, James, Bernier, Raphael A., Chawarska, Katarzyna, Dawson, Geraldine, Faja, Susan, Jeste, Shafali, & McPartland, James. 2020. Biomarker Acquisition and Quality Control for Multi-Site Studies: The Autism Biomarkers Consortium for Clinical Trials. Frontiers in Integrative Neuroscience, 13(February), 1–15.
- Wood, Simon N. 2006. On confidence intervals for generalized additive models based on penalized regression splines. Australian and New Zealand Journal of Statistics, 48(4), 445–464.
- Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 73(1), 3–36.
- Wood, Simon N. 2017. Generalized Additive Models: An Introduction with R. 2nd edn. London, UK: Chapman and Hall/CRC Press.
- Xiao, Luo, Zipunnikov, Vadim, Ruppert, David, & Crainiceanu, Ciprian. 2016. Fast covariance estimation for high-dimensional functional data. Statistics and Computing, 26(1-2), 409–421.