Title
Functional Data Analysis Tools for the Analysis of High-Dimensional Brain Imaging Data

Author
Boland, Joanna Marie

Publication Date
2022

UNIVERSITY OF CALIFORNIA
Los Angeles

Functional Data Analysis Tools for the Analysis of High-Dimensional Brain Imaging Data

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

by

Joanna Marie Boland

2022

ABSTRACT OF THE DISSERTATION

Functional Data Analysis Tools for the
Analysis of High-Dimensional Brain Imaging Data

by

Joanna Marie Boland
Doctor of Philosophy in Biostatistics
University of California, Los Angeles, 2022
Professor Damla Şentürk, Chair

This dissertation develops methodology and presents applications of functional data analysis tools used in high-dimensional functional data settings. In particular, the tools detailed were intended for use when analyzing electroencephalography (EEG) measurements, which records spontaneous electrical activity in the brain at electrodes placed across the scalp, resulting in rich multidimensional functional data. EEG data is typically analyzed in either the time and/or frequency domains depending on the application: resting-state experiments are typically analyzed in the frequency domain, and task-related experiments are typically analyzed in the time domain. In the first chapter, we develop an algorithm for analyzing EEG data jointly in both the time and frequency and results in a method for analyzing high-dimensional EEG data that adds an additional level of specificity to our data application than is available in single-domain analysis alone. The second chapter of this dissertation showcases a Bayesian functional principal component analysis (BFPCA) model applied to a resting-state EEG experiment analyzed in the frequency domain. We develop a fully data-driven tool that relies on functional depth, a method to order a set of functional observations from the center outwards, to flexibly visualize uncertainty in the estimated posterior samples. The final chapter extends this visualization tool from use in BFPCA to Bayesian longitudinal FPCA (B-LFPCA) for analysis of longitudinal functional data, which is conceptualized as functional datum measured repeatedly over a set of longitudinal time points. We apply our flexible depth-based visualization tool in the higher-dimensional setting to an event-related EEG experiment analyzed in the time domain.

The dissertation of Joanna Marie Boland is approved.

Danh Nguyen

Catherine A. Sugar

Donatello Telesca

Damla Şentürk, Committee Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VITA

2013–2017	B.A. Statistics and Mathematical Economic Analysis, Rice University, Houston, Texas.
2017–2019	M.S. Biostatistics, University of California, Los Angeles (UCLA), Los Angeles, California.
2018	Graduate Student Researcher, Department of Human Genetics, UCLA, Los Angeles, California.
2018–present	Graduate Student Researcher, Department of Biostatistics, UCLA, Los Angeles, California.
2019–present	Ph.D. Student and Candidate, Department of Biostatistics, UCLA, Los Angeles, California.
2020–present	Graduate Teaching Assistant, Department of Biostatistics, UCLA, Los Angeles, California.
2022	Summer Associate, Health Economics and Outcomes Research, Analysis Group, Los Angeles, California.

PUBLICATIONS

Boland, J., Telesca, D., Sugar, C. A., Jeste, S., Goldbeck, C., and Şentürk, D. (2022) A study of longitudinal trends in time-frequency transformation of EEG data during a learning experiment. *Computational Statistics and Data Analysis*, 167: 107367.

Boland, J., Telesca, D., Sugar, C. A., Jeste, S., Dickinson, A., DiStefano C., and Şentürk, D. (2022) Central posterior envelopes for Bayesian functional principal component analysis. *Journal of Data Science* (In review).

Boland, J., Telesca, D., Sugar, C. A., Guindani, M., Jeste, S., Dickinson, A., DiStefano C., and Şentürk, D. (2022) Central posterior envelopes for Bayesian longitudinal functional principal component analysis. (In preparation).

CHAPTER 1

A study of longitudinal trends in time-frequency transformations of EEG data during a learning experiment

Abstract

EEG experiments yield high-dimensional event-related potential (ERP) data in response to repeatedly presented stimuli throughout the experiment. Changes in the high-dimensional ERP signal throughout the duration of an experiment (longitudinally) is the main quantity of interest in learning paradigms, where they represent the learning dynamics. Typical analysis, which can be performed in the time or the frequency domain, average the ERP waveform across all trials, leading to the loss of the potentially valuable longitudinal information in the data. Longitudinal time-frequency transformation of ERP (LTFT-ERP) is proposed to retain information from both the time and frequency domains, offering distinct but complementary information on the underlying cognitive processes evoked, while still retaining the longitudinal dynamics in the ERP waveforms. LTFT-ERP begins by time-frequency transformations of the ERP data, collected across subjects, electrodes, conditions and trials throughout the duration of the experiment, followed by a data driven multidimensional principal components analysis (PCA) approach for dimension reduction. Following projection of the data onto leading directions of variation in the time and frequency domains, longitudinal learning dynamics are modeled within a mixed effects modeling framework. Applications to a learning paradigm in autism depict distinct learning patterns throughout the experiment among children diagnosed with Autism Spectrum Disorder and their typically developing peers. LTFT-ERP time-frequency joint transformations are shown to bring an additional level of specificity to interpretations of the longitudinal learning patterns related to underlying cognitive processes, which is lacking in single domain analysis (in the time or the frequency domain only). Simulation studies show the efficacy of the proposed methodology.

1.1 Introduction

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder characterized by social interaction and communication impairments. Our motivating study was conducted at UCLA by our collaborator Dr. Shafali Jeste on implicit learning of children with ASD and their typically developing (TD) peers (Jeste et al. (2015)). Implicit learning is defined by the detection of irregularities in one's environment without a conscious awareness or intention to learn. Children two to five years old were shown a continuous stream of colored geometrical shapes on a computer screen (see Figure 1.1(a)). The shapes were presented in pairs and the children were expected to learn the order within the shape pairs as the experiment progressed. This constituted implicit learning, providing insights into core cognitive deficits and social behavior of the children in the two diagnostic groups (ASD vs. TD). As learning was expected to take place over the course of the experiment, in response to the repeated presentation of stimuli (colored geometrical shapes), referred to as trials, capturing longitudinal changes in electroencephalography (EEG) signals over trials was one of the main goals of the experiment.

EEG is a non-invasive and widely-available (low cost) brain imaging modality which records electrical activity in the brain. An event-related potential (ERP) is defined as the EEG waveform measured in response to presentation of each stimulus (e.g. a colored geometrical shape) in an EEG experiment. Analysis of EEG and ERP data date back to 1950's in a wide spectrum of biomedical applications including epilepsy, sleep disorders, multiple sclerosis, brain tumors, lesions, major affective disorder, schizophrenia, alcoholism, bipolar mood disorder, assessment of surgical outcomes, confirmation of brain death, and clinical trials for drug development (Gasser and Molinari (1996); Tierney et al. (2012)). Typical analysis of ERP data averages the ERP signal over all the trials of the experiment, enhancing the signal-to-noise ratio (SNR) (Gasser and Molinari (1996); Delorme and Makeig (2004); Tierney et al. (2012)). While this common technique is effective in increasing the SNR of the

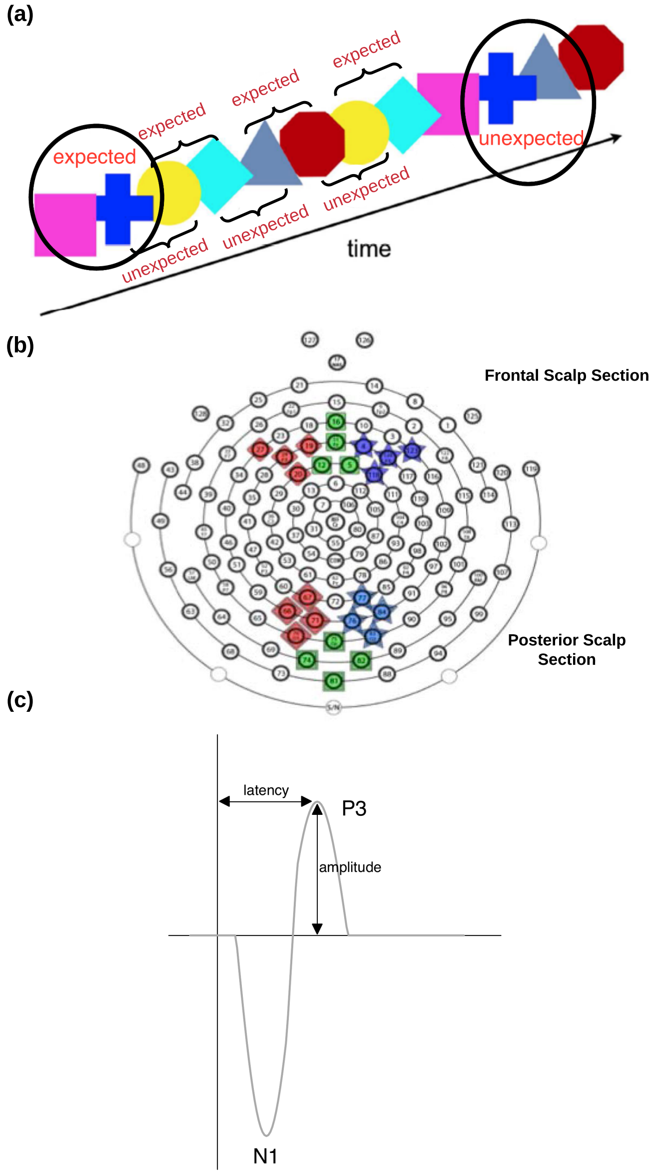


Figure 1.1: (a) Visualization of the implicit learning paradigm. The continuous stream of six-colored shapes are organized into three familiar pairs. The "expected" condition is defined as the transition between shapes within a shape pair, and the "unexpected" condition is defined as the transition between shape pairs. (b) The 24 electrodes of interest analyzed in the implicit learning paradigm in six total scalp regions (each containing four electrodes) within two scalp sections (frontal and posterior). (c) A depiction of the ERP phasic components P3 and N1 in the implicit learning paradigm.

ERP data, it collapses information gained during the course of the experiment. This longitudinal information is important, especially in learning experiments, where it characterizes the learning trends across the study participants, including speed of learning. Previous works have been proposed to study the longitudinal changes over the course of a learning experiment. Hasenstab et al. (2015) proposed the moving average preprocessed ERP (MAP-ERP) which averages ERPs over trials in a sliding-window to retain the inherent longitudinal information. Fiecas and Ombao (2016) proposed to study the longitudinal evolution of learning via the use of time-varying spectral densities in the frequency domain. Additional frequency domain approaches include Motta and Ombao (2012) using evolutionary factor analysis to study the multi-channel EEG dynamics across trials of a motor-visual task. For frequency domain analysis of EEG data from multiple subjects, see Krafty et al. (2011) and Krafty et al. (2017), where covariate effects on the power spectra of multiple time series are modeled in sleep studies.

The two previously proposed approaches of Hasenstab et al. (2015) and Ombao et al. (Fiecas and Ombao (2016); Ombao and Ho (2006)) for studying longitudinal trends in EEG experiments analyze ERP in different domains; in the time domain and the frequency domain, respectively. While the time domain analysis of ERP data concentrates on interpretations of the commonly studied ERP phasic components, such as the P3 as shown in Figure 1.1(c), the frequency domain analysis concentrates on interpretations of power from different frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-16 Hz), beta(16-32 Hz) and gamma (over 32 Hz). In our motivating implicit learning paradigm, N1 and P3 are the phasic components typically observed in the ERP waveforms. The N1 dip, with a short latency (time-delay), is thought to be related to early category recognition, while the P3 peak, with a long latency, is traditionally related to cognitive processes such as signal matching, decision making and memory updating (Bugli and Lambert (2006); Jeste et al. (2015)). In the frequency domain, frequencies in the delta and theta bands have been reported to contribute to a P3 phasic component where frequencies in the delta band are associated with evaluative cognitive processing, and in the theta band are associated with the orienting response to novel stimuli (Bernat et al. (2007); Harper et al. (2014)). Since both time and frequency domain analysis of ERP data carry different but complimentary information about the observed signal, we consider a time-frequency decomposition of the ERP data, targeting even richer information than is available in single domain analysis. We propose the longitudinal time-frequency transformation of ERP (LTFT-ERP) method, where a wavelet transformation is applied to the ERPs. LTFT-ERP, not only targets richer information in the signal through time-frequency transformations, it also allows modeling of longitudinal changes in the signal over trials throughout the learning experiment, adding an additional dimension for analysis (referred to as the longitudinal dimension).

Bernat et al. (2005, 2007) proposed time-frequency transformations (TFTs) of ERP data via wavelets and further included dimension reduction of the high dimensional time-frequency power surfaces through principal components analysis (PCA). To capture the longitudinal changes throughout the learning experiment, the resulting data from the proposed LTFT-ERP is even higher dimensional in our applications since the TFTs are repeated over multiple trials of the experiment. To adopt a data-driven approach to dimension reduction, similar to Bernat et al. (2005), we employ a multidimensional principal component analysis (MDPCA). We characterize vectorized TFT power surfaces as the functional dimension of the data and repetitions over trials as the longitudinal dimension. Under the simplifying assumption that the direction of variation in the functional dimension of the data stays the same for fixed slices along the longitudinal dimension, the eigenvectors in the functional dimension are obtained. Projections of the data onto the leading eigenvectors in the functional dimension allow us to study longitudinal changes in the resulting PCA scores via a mixed effects model. Finally, diagnostic-group level inference on longitudinal trends, representing learning dynamics, is derived via mixed effects modeling machinery.

This chapter is organized as follows. Section 1.2 outlines the proposed LTFT-ERP approach, including dimension reduction via MDPCA following the multidimensional TFT decompositions and modeling of the longitudinal trends via a mixed effects model. Simulation studies to study the efficacy of the LTFT-ERP in modeling of longitudinal trends in the MDPCA scores are outlined in Section 1.3, followed by applications to the implicit learning paradigm in Section 1.4. We conclude with a brief discussion given in Section 1.5.

1.2 The proposed longitudinal time-frequency transformation of ERP data (LTFT-ERP)

The proposed longitudinal time-frequency transformation (LTFT-ERP) starts with transformation of the ERP waveforms from each trial into the time-frequency (TFT) power surfaces using the wavelet transformation. Dimension reduction of the resulting trial specific time-frequency power surfaces is achieved by the data driven MDPCA in the third step. TFT power surfaces are first reshaped into two continuous dimensions (functional and longitudinal) before the application of MDPCA. The functional dimension represents wavelet power vectorized over ERP time and frequency within a trial and longitudinal dimension represents the repeatedly obtained TFT power surfaces over trials, providing the progression of the high-dimensional process throughout the experiment. Projection of the two-dimensional process onto the leading functional principal component vectors leads to the longitudinally estimated PCA scores which summarize changes in the signal over the course of the experiment. In a final step, longitudinal trends in the PCA scores are compared across participants in the two diagnostic groups (ASD vs. TD), leading to insights in speed and nature of learning. The main steps

Step 1: Multidimensional time-frequency transformation (MTFT-ERP) utilizing wavelets,

Step 2: Reshaping of TFT power surfaces into vectors,

Step 3: Dimension reduction via multidimensional PCA (MDPCA),

Step 4: Modeling of longitudinal trends in MDPCA scores via mixed effects modeling,

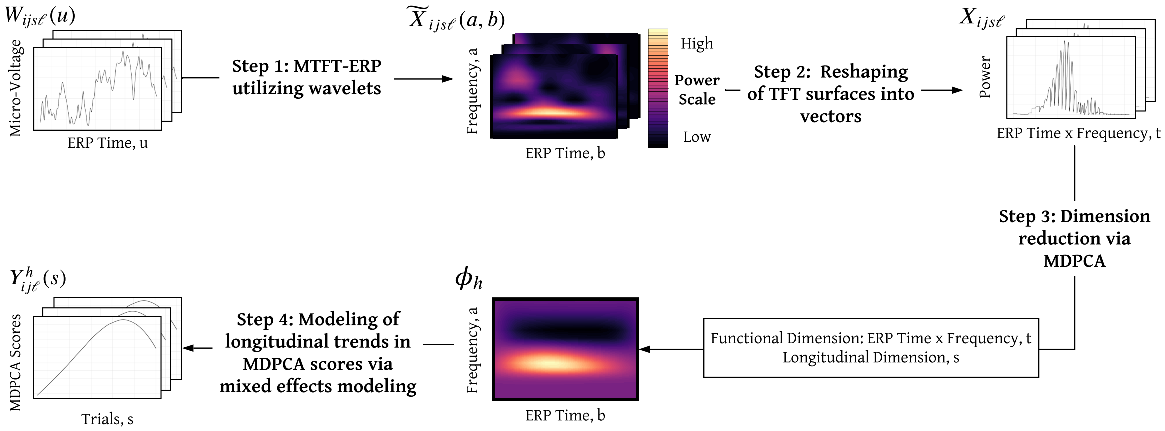


Figure 1.2: A flowchart of the LTFT-ERP algorithm. For each subject i, electrode j, on trial s and condition ℓ, Step 1 transforms the ERP waveform $\mathrm{W}_{ijs\ell}(u)$ into the TFT power surface $\widetilde{X}_{ijs\ell}(a,b)$ using the wavelet transformation. Step 2 reshapes the TFT power surface $\widetilde{X}_{ijs\ell}(a,b)$ into a vector $\mathbf{X}_{ijs\ell}$ in t where t denotes the functional dimension of ERP time × frequency. The longitudinal dimension is trials s, where Step 3 performs dimension reduction via MDPCA to target the h^{th} leading eigenvector ϕ_h in the functional dimension. The final step models the MDPCA scores as a function of trials (s) via mixed effects modeling.

of LTFT-ERP are depicted in Figure 1.2. We further expand on each step of the proposed LTFT-ERP algorithm in the following subsections.

1.2.1 MTFT-ERP utilizing wavelets

The ERPs waveforms are transformed into TFT power surfaces in the first step using the wavelet function which is a simple oscillating amplitude waveform that is localized in time. Let $\mathrm{W}_{ijs\ell}(u)$ denote the micro-voltage of the ERP for subject i, $i = 1, \ldots, \mathrm{N}$, from electrode j, $j = 1, \ldots, \mathrm{J}$, on trial s, $s \in \mathrm{S_i}$, in condition ℓ (expected/unexpected), $\ell \in \mathrm{L_{is}}$, observed at time u, $u = 1, \ldots, \mathrm{U}$, where N, J, U denote the total number of subjects, electrodes, time points within a trial, respectively, and $\mathrm{S_i}$ and $\mathrm{L_{is}}$ denote the sets of non-missing trials and conditions at trial s for subject i, respectively. The maximum number of conditions per trial, denoted by L, equals two in our application (expected vs. unexpected). The minimum and maximum possible number of trials per subject are denoted by s_{min} and S, respectively. In addition, subjects may be partitioned into multiple diagnostic groups (e.g. TD vs. ASD), but we omit additional subscripts denoting diagnostic groups for simplicity of notation.

For the wavelet transformation, a general form of the wavelet function is selected and is referred to as the mother wavelet. The mother wavelet is then systematically stretched and contracted in time with a set of "daughter wavelets" (Bernat et al. (2005)). The daughter wavelets are generated by the mother wavelet through scaling (frequency) and translation (time) parameters. The continuous and complex-valued Morlet wavelet,

$$\psi(u) = \pi^{-1/4} e^{i\omega u} e^{-u^2/2}, \tag{1.1}$$

commonly-used in the decomposition of ERP waveforms, is selected as the mother wavelet (Torrence and Compo (1998)). The "angular frequency", denoted by ω in (1.1), is set to 6, following previous literature (Farge (1992)). For a given scale/frequency $a > 0$ and translation/time parameter $b \in \mathbb{R}$, the resulting augmented daughter wavelet used in transformations is equal to $\psi\{(u-b)/a\}$. The wavelet transformation is then given as:

$$C_{ijs\ell}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \mathrm{W}_{ijs\ell}(u) \bar{\psi}\Big(\frac{u-b}{a}\Big) du,$$

where $C_{ijs\ell}(a,b)$ denotes the wavelet coefficients which are continuous and complex, and $\bar{\psi}\{(u-b)/a\}$ is the complex conjugate of the daughter wavelet. The power, denoted by $\widetilde{X}_{ijs\ell}(a,b)$, is calculated as the squared magnitude of the wavelet coefficients, i.e. $\widetilde{X}_{ijs\ell}(a,b) = |C_{ijs\ell}(a,b)|^2$.

For a selected grid of scale/frequency a_f, $f = 1, \ldots, F$, and translation/time parameters b_d, for $d = 1, \ldots, D$, the TFT power surfaces, $\widetilde{X}_{ijs\ell}$, $F \times D$ matrices with elements,

$\widetilde{X}_{ijs\ell}(a_f, b_d)$, are vectorized into

$$\begin{aligned}\mathbf{X}_{ijs\ell} &\equiv \mathrm{Vec}\big(\widetilde{X}^{\mathrm{T}}_{ijs\ell}\big) = \{\widetilde{X}_{ijs\ell}(a_1, b_1), \ldots, \widetilde{X}_{ijs\ell}(a_1, b_D), \ldots, \widetilde{X}_{ijs\ell}(a_F, b_1), \ldots, \widetilde{X}_{ijs\ell}(a_F, b_D)\}^{\mathrm{T}}, \\ &= \{X_{ijs\ell}(t_1), \ldots, X_{ijs\ell}(t_m)\}^{\mathrm{T}},\end{aligned}$$

where the indices $t_1, \ldots, t_m$ are interpreted as the "functional" dimension of the resulting process, encompassing both ERP time and frequency components ($m = F \times D$). The TFT power surfaces, vectorized in $\mathbf{X}_{ijs\ell}$, are observed repeatedly over trials s. Interpreting s as a coarse time scale, we refer to it as the "longitudinal" dimension. Hence we conceptualize the TFT dynamic of the original signal as a set of random quantities varying over both a functional and a longitudinal dimension. In the next section we outline how this random object is further reduced in dimensionality through data-driven MDPCA.

Note that the tuning parameters F and D utilized in the wavelet transformations determine the resolution of the time-frequency transformation. While larger values of F and D allow for a higher resolution decomposition in frequency and time, respectively, these values are bounded by considerations of the total number of available EEG (over subjects, electrodes, conditions and trials) for the MDPCA decompositions. More specific guidance is provided between the ratio of the number of TFT power surfaces included in estimation of the trial-specific covariances and $m = F \times D$, to ensure stability of the proposed PCA decompositions, in the next section.

1.2.2 Dimension reduction via MDPCA

For dimension reduction of the TFT power surfaces obtained in the previous section, we borrow ideas from marginal functional PCA (FPCA) (Park and Staicu (2015)). Marginal FPCA of longitudinally observed functional data relies on the assumption that the direction of variation, not the covariance itself, in the functional dimension of the data stays the same for fixed slices along the longitudinal dimension. This assumption, if assumed along both functional and longitudinal directions, called weak-separability (Chen and Müller (2012); Chen et al. (2017)), is weaker than the commonly assumed strong separability in high-dimensional data settings, which implies constant covariance along one dimension for fixed values of the other dimensions of the data. When applied to our two-dimensional TFT power surfaces, the assumption of constant direction of variation implies that the direction of variation along the functional dimension of ERP time and frequency (denoted by t) in the TFT power surfaces, stays the same across trials in the longitudinal dimension, denoted by s. This leads to the construction of a marginal covariance in the functional dimension, evaluated as an average of all functional covariances obtained at fixed trials, in estimation of the common functional directions of variation, captured by the functional eigenvectors.

For each trial s, we define the trial-specific functional covariance as $\Sigma_s := Cov(\mathbf{X}_{ijs\ell}) \in \mathbb{R}^{m\times m}$. Estimation of this quantity hinges on a moving window estimator, borrowing information across adjacent trials. Let A_s represent overlapping sets of trials of varying lengths for trials $s = s_{min}, \ldots, \mathrm{S}$, with the maximum number of trials within a set denoted by k,

$$\mathrm{A}_s = \begin{cases} [s_{min}, 2s - s_{min}], & s < \frac{k}{2} \\ [s - \frac{k}{2} + 1, s + \frac{k}{2}], & \frac{k}{2} \leqslant s \leqslant S - \frac{k}{2} \\ [2s - S, S], & s > S - \frac{k}{2}. \end{cases}$$

For a specific trial s, the estimator of Σ_s pools all vectors $\mathbf{X}_{ijs'\ell}$, s.t. $s' \in A_s$. This set of TFT power vectors is then centered by subtracting a mean vector $\overline{\mathbf{X}}_s \in \mathbb{R}^m$, obtained averaging across subjects i, electrodes j, conditions ℓ and trials in A_s. We denote the mean centered vectors as $\mathbf{X}^c_{ijs'\ell} := \mathbf{X}_{ijs'\ell} - \overline{\mathbf{X}}_s$. The ensuing estimator for the trial-specific covariance is defined as:

$$\hat{\Sigma}_s = \frac{1}{N_s - m} \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{s' \in A_s} \sum_{\ell \in L_{is'}} \mathbf{X}^c_{ijs'\ell} \mathbf{X}^{c\mathrm{T}}_{ijs'\ell}. \tag{1.2}$$

Note that in the formulation above, TFTs are merged across diagnostic groups, scalp sections and conditions, in targeting trial-specific covariances. This merge requires that the direction of variation (captured by the estimated eigenvectors of the subsequent marginal covariance) are similar across different grouping of TFTs (determined by diagnostic groups, conditions and scalp sections). Appendix A.1 Figures A.1-A.12 display the similarity across eigenvectors of marginal covariances targeted within the eight subgroups of subjects (two diagnostic groups, two conditions and two scalp sections) in our data application, justifying the merging of TFTs used in trial-specific covariances. Hence, the total number of TFT power vectors included in estimation of Σ_s is $N_s = J \times (\sum_i L_{is'}) \times |A_s|$, where $|A_s|$ denotes the number of elements in set A_s. The moving window parameters s_{min}, S and k, defining the number of trials included in each sliding window in A_s (ranging from 1 to k), are selected to maintain a 5:1 ratio between the number of TFT power surfaces included in estimation of Σ_s and $m = F \times D$. This procedure is recommended to ensure stability of the proposed PCA decompositions.

Given trial-specific covariances $\hat{\Sigma}_s$, an estimator, $\hat{\Sigma}$, of the functional marginal covariance, is targeted by a method of moments approach by averaging all trial-specific covariances at $s = s_{min}, \ldots,$ S. Note that when the number of trials gets large, trial-specific covariances from a smaller set of trials (than the entire trial set) may be averaged in targeting the functional marginal covariance for computational feasibility. In this setting, the invariant directions of functional variation across trials, are targeted by the estimated m by 1 eigenvectors, ϕ_h, $h = 1, \ldots, m$, of the m by m functional marginal covariance matrix,

$$\hat{\Sigma} = \sum_{h=1}^{m} \lambda_h \phi_h \phi_h^{\mathrm{T}},$$

with $\lambda_1 \geqslant \ldots \geqslant \lambda_m$ denoting the ordered eigenvalues (Greven et al. (2010); Di et al. (2009)). The decomposition is typically truncated to include the first H, $H < m$, leading eigencomponents in applications. H is selected using the elbow in the fraction of variance explained where all components are selected before a relative flattening in the plot. Figures 4 and 5 display the relative flattening in the fraction of variance explained plots in our data application. The mean centered TFT power vectors $\mathbf{X}_{ijs\ell} - \overline{\mathbf{X}}$, where $\overline{\mathbf{X}}$ denotes the overall mean TFT power vector (averaged over all conditions, subjects, electrodes, and trials), are then projected onto the H estimated leading functional eigenvectors, to obtain the longitudinal MDPCA scores:

$$Y_{ij\ell}^{h}(s) = < \mathbf{X}_{ijs\ell} - \overline{\mathbf{X}}, \phi_h >,$$

where $< \cdot, \cdot >$ denotes the inner product. We allude to the longitudinal modeling of the MDPCA scores that is outlined in the next section by representing the estimated MDPCA scores as functions of the longitudinal argument s, representing trials.

Formulation of the functional marginal covariance from trial-specific covariances utilizes the constant direction of functional variation assumption across trials in the longitudinal dimension. This assumption can be assessed via checking the degree of similarity between eigenvectors of trial-specific covariances across trials. Plots of the estimated six leading eigenvectors from trial-specific covariances evaluated at trials $20, 40$ and 60 are given in Appendix A.2 Figures A.13-A.18 in our data application. Plots display sufficient similarity across trials, signaling no violation of the assumption of constant variation across trials.

1.2.3 Modeling of longitudinal trends in MDPCA scores

We model the longitudinal trends in the MDPCA scores, $Y_{ij\ell}^{h}(s)$, corresponding to the leading H functional eigenvectors, across trials via a linear mixed effects model. Multilevel random effects at the subject, $i = 1, \ldots, \mathrm{N}$, and electrode region, $r = 1, \ldots, \mathrm{R}$, levels are utilized to model the dependency of the data within subjects and electrodes within a scalp region (3 scalp regions containing 4 electrodes are depicted in Figure 1.1(b) for our data application, within the frontal and posterior scalp sections) where spatial correlation between electrodes may exist. Additionally, we employ spline basis functions in modeling both the fixed and random effects to represent changes in the MDPCA scores longitudinally across trials.

Let $Y^{h}_{ij(r)\ell}(s)$ denote the hth leading MDPCA score for subject i, electrode j within region r, and condition ℓ, targeted as a function of the longitudinal index, trials (s). We model the MDPCA scores using fixed effects and a two-level random effects structure for subject and region. In our data application, the fixed effects parameters include an intercept, trial (represented by a natural cubic B-spline with four knots), group (ASD vs. TD), scalp section (frontal vs. posterior), condition (expected vs. unexpected), as well as higher order interactions (two- to four-way) between the main effects. This leads to a total of 40 fixed effects components. Let β^{h} denote the 40×1 column vector of fixed effects parameters, and let 5×1 vectors b^{h}_{i} and b^{h}_{ir} represent subject and region-level random effects, respectively. Further, let $q_{ijs\ell}$ denote the 1×40 row vector of the fixed effects matrix Q_{i}, corresponding to trial s, electrode j, and condition ℓ, and $q_{ij(r)s}$ be the 1×5 row vector of the random effects matrix $Q_{ij(r)}$ corresponding to trial s and electrode j in region r. We model $\mathrm{Y}^{h}_{ij(r)\ell}(s)$ by

$$\begin{aligned}
Y^{h}_{ij(r)\ell}(s) &= q_{ijs\ell}\beta^{h} + q_{ij(r)s}b^{h}_{i} + q_{ij(r)s}b^{h}_{ir} + \epsilon^{h}_{ij(r)s\ell}, \\
b^{h}_{i} &\sim MVN(0, D^{1h}_{5\times 5}), \\
b^{h}_{ir} &\sim MVN(0, D^{2h}_{5\times 5}), \\
\epsilon^{h}_{ij(r)s\ell} &\sim N(0, \sigma^{2}_{h}),
\end{aligned}$$

where D^{1h} and D^{2h} represent the random effects covariance matrices at the subject and region levels, respectively (Greven et al. (2010); Di et al. (2009)). In addition, $\epsilon^{h}_{ij(r)s\ell}$ represents the error term for the hth principal component with variance σ^{2}_{h}. This multi-level mixed effects model leads to the following covariance structure assuming the random level effects are independent of the error term

$$\begin{aligned}
\mathrm{Var}\{Y^h_{ij(r)\ell}(s)\} &= q_{ij(r)s}D^{1h}q'_{ij(r)s} + q_{ij(r)s}D^{2h}q'_{ij(r)s} + \sigma^2_h, \\
\mathrm{cov}\{Y^h_{ij(r)\ell}(s), Y^h_{ij(r)\ell}(s')\} &= q_{ij(r)s}D^{1h}q'_{ij(r)s'} + q_{ij(r)s}D^{2h}q'_{ij(r)s'}, \quad \forall i, j(r), s \neq s', \\
\mathrm{cov}\{Y^h_{ij(r)\ell}(s), Y^h_{ij'(r)\ell}(s)\} &= q_{ij(r)s}D^{1h}q'_{ij'(r)s} + q_{ij(r)s}D^{2h}q'_{ij'(r)s}, \quad \forall i, j(r) \neq j'(r), s,
\end{aligned}$$

for within region correlation and

$$\mathrm{cov}\{Y^h_{ij(r)\ell}(s), Y^h_{ij'(r')\ell}(s)\} = q_{ij(r)s}D^{1h}q'_{ij'(r')k}, \quad \forall i, j(r) \neq j'(r'), s,$$

for within subject across region correlation.

We use the same design matrix for both the subject and region random effects to reduce the complexity of the model, but these two matrices can be taken to be different. In addition, the model framework can be extended to accommodate additional covariates that can model the dependency structure of the data in further detail. The number of equispaced knots for the cubic B-splines is chosen using AIC criteria in order to obtain a sufficient degree of smoothness in the modeling of the MDPCA scores (Shi et al. (1996); Rice and Wu (2001)). The model parameters are estimated using restricted maximum likelihood (REML). See Appendix A.3 Figure A.19 for histograms of the modeled eigenscores, displaying relatively symmetric distributions in our data applications, signaling no violation of the assumed normality assumption.

1.3 Simulation studies

1.3.1 Simulation setup

We conduct simulations to study the efficacy of the LTFT-ERP algorithm in modeling the longitudinal trends of the MDPCA scores. For data generation, we utilize the mixed effects fits in modeling the leading six MDPCA scores from the delta frequency band in our data application, since the leading six eigencomponents explained a higher proportion (approximately 90%) of the variation in the data from the delta frequency band than the theta band. In assessing the efficacy of the LTFT-ERP algorithm, we utilize mean error (ME) in estimation of the TFT mean power vector and functional eigenvector, in addition to group, condition and scalp section-specific mean longitudinal trajectories of the MDPCA scores and utilize prediction error (PE) in prediction of subject and region-specific longitudinal trajectories of MDPCA scores. We consider multiple simulation scenarios at varying signal-to-noise ratios (SNR $= 0.4, 0.8$ and 1.6) and sample sizes ($N = 80, 160$ and 320), where results are reported based on 200 Monte Carlo runs.

The data fits in modeling the six leading MDPCA score trajectories and the estimated functional eigenfunctions from the delta frequency band are utilized to simulate the underlying TFT power vectors, denoted by $\mathbf{X}_{ijs\ell}^{signal}$. Additional error is added directly to the simulated TFT vectors, to avoid performing back wavelet transformations to target raw ERPs, which requires additional phase information. In order to simulate the six leading longitudinal MDPCA score trajectories, $Y_{ij(r)\ell}^{h}(s)$, h=1, $\ldots$,6, we first simulate subject and region-specific random effects, $b_i^h \sim MVN(0, D^{1h})$ and $b_{ir}^h \sim MVN(0, D^{2h})$, where the variance components, D^{1h} and D^{2h}, estimated in our data from the delta frequency band, are multiplied by 0.15 to reduce variability in the score trajectories to guarantee positivity of the generated TFT power vectors. Once the random effects are simulated, we generate MDPCA score trajectories via $Y_{ij(r)\ell}^{h}(s) = q_{ijs\ell}\beta^h + q_{ij(r)s}b_i^h + q_{ij(r)s}b_{ir}^h$, for $i = 1, \ldots,$ N subjects, $j = 1, \ldots, 24$ electrodes, $s = 5, \ldots, 80$ trials and $\ell \in \{1, 2\}$ (expected/unexpected) conditions, mimicking our data analysis. The design matrices for the fixed and random effects are utilized as described in Section 1.2.3, with half of the sample designated as belonging to the ASD sample and the other half to the TD sample and the fixed effects parameter vectors, β^h, are taken as estimated from the data fits from the delta frequency band.

The simulated true TFT power vectors are obtained via $\mathbf{X}_{ijs\ell}^{signal} = \overline{\mathbf{X}} + \sum_{h=1}^{6} Y_{ij(r)\ell}^{h}(s)\phi_h$, utilizing the six leading eigencomponents, where ϕ_h denotes the estimated functional eigenvectors and $\overline{\mathbf{X}}$ denotes the estimated TFT mean power vector in the delta frequency band. Additional error is added directly to the power vectors, to obtain the simulated TFT power vectors, $\mathbf{X}_{ijs\ell} = \mathbf{X}^{signal}_{ijs\ell} + \mathbf{X}^{noise}_{ijs\ell}$. The error, $\mathbf{X}^{noise}_{ijs\ell}$, is obtained as the TFT power vector (via a wavelet transformation, corresponding to the squared radius in the polar coordinates) of the Euclidean vector (X, Y), where coordinates X and Y are both generated from a $N(0, c^2)$ distribution. Since r^2/c will follow a χ^2_2 distribution, c is chosen to correspond to varying SNR ratios, SNR $= 0.4, 0.8$, and 1.6, where SNR is defined as the ratio of the standard deviation of $\mathbf{X}^{signal}_{ijs\ell}$ to the standard deviation of $\mathbf{X}^{noise}_{ijs\ell}$ (which equals $2c$). Finally, missingness is induced by randomly removing a fraction of the generated TFT power vectors by sampling with replacement from the missingess profiles of the subjects in our data set from the implicit learning paradigm.

LTFT-ERP algorithm is applied to the simulated TFT power vectors to first target the estimated TFT mean power vector, $\widehat{\overline{\overline{\mathbf{X}}}}$ (averaged over all conditions, subjects, electrodes, and trials), and the leading functional eigenvector, $\widehat{\phi}_1$. Projections onto the leading eigenvector, yield the estimated leading MDPCA scores, $\widehat{Y}^1_{ij(r)\ell}(s)$, where the group, condition and scalp section-specific true and estimated mean trajectories are denoted by $E[Y^1_{ij(r)\ell}(s)] = q_{ijs\ell}\beta^1$ and $E[\widehat{Y^1_{ij(r)\ell}(s)}] = q_{ijs\ell}\widehat{\beta}^1$, respectively, where β^1 is estimated from the implicit learning paradigm data fits in the delta frequency band and used in data generation and $\widehat{\beta}^1$ is estimated from the mixed effects modeling of the simulated MDPCA scores, $\widehat{Y}^1_{ij(r)\ell}(s)$. To mimic our data application mean trajectories over trials are considered in eight subgroups ($g = 1, \ldots, 8$) determined by two diagnostic groups (TD vs. ASD), two conditions (expected vs. unexpected) and two scalp sections (frontal vs. posterior). The true and estimated subject and region-specific predictions are denoted by $E[Y^1_{ij(r)\ell}(s)|b^1_i, b^1_{ir}] = q_{ijs\ell}\beta^1 + q_{ij(r)s}b^1_i + q_{ij(r)s}b^1_{ir}$ and $E[\widehat{Y^1_{ij(r)\ell}(s)}|b^1_i, b^1_{ir}] = q_{ijs\ell}\widehat{\beta}^1 + q_{ij(r)s}\widehat{b}^1_i + q_{ij(r)s}\widehat{b}^1_{ir}$, where b^1_i and b^1_{ir} are the true random effects values simulated from the above specified normal distributions in data generation and $\widehat{\beta}^1$, $\widehat{b}^1_i$, and $\widehat{b}^1_{ir}$ are estimated based on simulated data. Note that run times for a single simulation are approximately 20, 10, and 5 minutes at sample sizes $N = 320$, 160, and 80, respectively.

The mean errors (ME) in estimation of the TFT mean power vector and the leading functional eigenvector are defined by

$$\text{ME} = \frac{\sum_{p=1}^{m} \left| \overline{\mathbf{X}}(t_p) - \widehat{\overline{\mathbf{X}}}(t_p) \right|}{\sum_{p=1}^{m} \left| \overline{\mathbf{X}}(t_p) \right|}, \text{ and ME} = \frac{\sum_{p=1}^{m} \left| \phi_1(t_p) - \widehat{\phi}_1(t_p) \right|}{\sum_{p=1}^{m} \left| \phi_1(t_p) \right|},$$

respectively, for $m = F \times D = 1764$ (with $D = 63$ and $F = 28$), similar to our data application. The ME in estimation of MDPCA mean trajectories for subjects (in two diagnostic groups: TD vs. ASD), electrodes (in two scalp regions frontal vs. posterior) and conditions (expected vs. unexpected) from a total of eight unique subgroup trajectories ($\{i, j, \ell\} \in g = 1, \ldots, 8$), summed over trials $s = 5, \ldots 80$, similar to our data application, is defined as

$$\text{ME} = \frac{\sum_{g=1}^{8} \sum_{s=5}^{80} \left| E[Y^1_{ij(r)\ell}(s)] - E[\widehat{Y^1_{ij(r)\ell}}(s)] \right|}{\sum_{g=1}^{8} \sum_{s=5}^{80} \left| E[Y^1_{ij(r)\ell}(s)] \right|}.$$

The PE in prediction of the subject and region-specific MDPCA score trajectories is defined as

$$\text{PE} = \frac{\sum_{i,j,s,\ell} \left| E[Y^1_{ij(r)\ell}(s) | b_i, b_{ir}] - E[\widehat{Y^1_{ij(r)\ell}(s)} | b_i, b_{ir}] \right|}{\sum_{i,j,s,\ell} \left| E[Y^1_{ij(r)\ell}(s) | b_i, b_{ir}] \right|},$$

where $i = 1, \ldots, \text{N}$ denotes subjects, $j = 1, \ldots, 24$ denotes electrodes from six scalp regions, $s = 5, \ldots, 80$ denotes trials, and $\ell \in \{1, 2\}$ denotes the two conditions (expected/unexpected).

Table 1.1: Medians and (10th, 90th) percentiles of simulation performance metrics (ME and PE) from 200 Monte Carlo runs at varying SNRs (SNR $= 0.4, 0.8$, and 1.6) and sample sizes ($N = 80, 160$, and 320).

		Mean Power	Eigenvector	MDPCA Scores	
N	SNR	ME	ME	ME	PE
80	0.4	0.81460 (0.80966, 0.82040)	0.89373 (0.88256, 0.90372)	0.376 (0.324, 0.497)	0.564 (0.534, 0.621)
	0.8	0.40759 (0.40359, 0.41065)	0.69500 (0.65928, 0.72499)	0.293 (0.228, 0.415)	0.426 (0.380, 0.498)
	1.6	0.20421 (0.20144, 0.20652)	0.32873 (0.28398, 0.37164)	0.255 (0.184, 0.554)	0.272 (0.221, 0.728)
160	0.4	0.81453 (0.81069, 0.81836)	0.89293 (0.88269, 0.90094)	0.344 (0.309, 0.389)	0.561 (0.532, 0.590)
	0.8	0.40756 (0.40489, 0.41020)	0.69260 (0.67283, 0.71025)	0.245 (0.203, 0.345)	0.418 (0.388, 0.507)
	1.6	0.20394 (0.20169, 0.20564)	0.32145 (0.28761, 0.35842)	0.206 (0.146, 0.422)	0.265 (0.224, 0.510)
320	0.4	0.81462 (0.81215, 0.81717)	0.89206 (0.88639, 0.89865)	0.333 (0.300, 0.429)	0.566 (0.544, 0.670)
	0.8	0.40766 (0.40579, 0.40918)	0.68958 (0.67272, 0.70487)	0.236 (0.198, 0.356)	0.418 (0.394, 0.599)
	1.6	0.20382 (0.20255, 0.20522)	0.32116 (0.29401, 0.34458)	0.182 (0.140, 0.376)	0.270 (0.237, 0.573)

1.3.2 Simulation results

The medians, 10th and 90th percentiles of the ME in estimation of the TFT mean power vector and functional eigenvector, in addition to group, condition and scalp section-specific mean longitudinal trajectories of the MDPCA scores and of the PE in prediction of subject and region-specific longitudinal trajectories of MDPCA scores are given in Table 1.1 at three SNRs (SNR $= 0.4, 0.8$, and 1.6) and three sample sizes ($N = 80, 160$, and 320). Tables 1.2 and 1.3 report the ME and PE, respectively, of the MDPCA score trajectories separately within the eight subgroups, determined by the two diagnostic groups (TD vs. ASD), two conditions (expected vs. unexpected) and the two scalp sections (frontal vs. posterior). This provides a better understanding of the subgroup variability in these measures. In addition, Figure 1.3 displays the estimated fixed effects means of the MDPCA scores from the run with the median ME value for one of the eight subgroups, ASD expected posterior, at varying SNRs and sample sizes.

Table 1.1 shows a clear and expected trend in the ME and PE with varying SNR values, where both metrics decrease with increasing SNR values. Note that this trend is present at all three sample sizes. For estimation of the MDPCA trajectories, ME values for the

Table 1.2: Medians and (10th, 90th) percentiles of ME from 200 Monte Carlo runs at varying SNRs (SNR $= 0.4, 0.8,$ and 1.6) and sample sizes ($N = 80, 160,$ and 320) for each of the eight subgroups determined by diagnostic group, condition and scalp section.

			ME of MDPCA Scores			
			Frontal		Posterior	
Group	N	SNR	Expected	Unexpected	Expected	Unexpected
ASD	80	0.4	0.547 (0.281, 0.882)	0.567 (0.377, 0.876)	0.348 (0.252, 0.452)	0.235 (0.171, 0.395)
		0.8	0.387 (0.232, 0.769)	0.402 (0.241, 0.721)	0.265 (0.169, 0.375)	0.167 (0.105, 0.319)
		1.6	0.384 (0.195, 0.853)	0.382 (0.214, 0.835)	0.242 (0.126, 0.513)	0.202 (0.107, 0.505)
	160	0.4	0.488 (0.272, 0.738)	0.507 (0.332, 0.734)	0.314 (0.253, 0.401)	0.219 (0.160, 0.279)
		0.8	0.331 (0.214, 0.602)	0.352 (0.234, 0.579)	0.224 (0.151, 0.312)	0.130 (0.077, 0.237)
		1.6	0.242 (0.133, 0.664)	0.245 (0.132, 0.636)	0.189 (0.115, 0.367)	0.171 (0.093, 0.349)
	320	0.4	0.463 (0.293, 0.728)	0.501 (0.362, 0.691)	0.311 (0.242, 0.389)	0.209 (0.156, 0.301)
		0.8	0.345 (0.205, 0.566)	0.328 (0.224, 0.506)	0.205 (0.145, 0.293)	0.120 (0.080, 0.229)
		1.6	0.230 (0.108, 0.480)	0.228 (0.122, 0.459)	0.166 (0.103, 0.322)	0.155 (0.077, 0.296)
TD	80	0.4	0.266 (0.178, 0.465)	0.284 (0.162, 0.523)	0.632 (0.484, 0.798)	0.679 (0.422, 0.989)
		0.8	0.202 (0.107, 0.340)	0.235 (0.127, 0.439)	0.431 (0.279, 0.649)	0.653 (0.400, 0.949)
		1.6	0.142 (0.070, 0.564)	0.174 (0.098, 0.571)	0.330 (0.191, 0.726)	0.595 (0.345, 1.137)
	160	0.4	0.238 (0.175, 0.336)	0.258 (0.171, 0.382)	0.580 (0.445, 0.682)	0.594 (0.413, 0.805)
		0.8	0.178 (0.112, 0.300)	0.206 (0.134, 0.351)	0.377 (0.265, 0.542)	0.551 (0.402, 0.776)
		1.6	0.115 (0.063, 0.349)	0.129 (0.075, 0.405)	0.289 (0.153, 0.546)	0.514 (0.285, 0.844)
	320	0.4	0.237 (0.188, 0.388)	0.267 (0.187, 0.420)	0.571 (0.464, 0.683)	0.561 (0.402, 0.729)
		0.8	0.165 (0.115, 0.336)	0.206 (0.134, 0.367)	0.356 (0.266, 0.512)	0.526 (0.400, 0.710)
		1.6	0.093 (0.053, 0.393)	0.115 (0.072, 0.414)	0.266 (0.171, 0.495)	0.466 (0.309, 0.753)

Table 1.3: Medians and (10th, 90th) percentiles of PE from 200 Monte Carlo runs at varying SNRs (SNR $= 0.4, 0.8$, and 1.6) and sample sizes ($N = 80, 160$, and 320) for each of the eight subgroups determined by diagnostic group, condition and scalp section.

			PE of MDPCA Scores			
			Frontal		Posterior	
Group	N	SNR	Expected	Unexpected	Expected	Unexpected
ASD	80	0.4	0.617 (0.564, 0.699)	0.625 (0.572, 0.709)	0.561 (0.503, 0.662)	0.519 (0.471, 0.618)
		0.8	0.464 (0.406, 0.541)	0.466 (0.407, 0.542)	0.418 (0.359, 0.517)	0.387 (0.327, 0.495)
		1.6	0.285 (0.228, 0.751)	0.290 (0.229, 0.751)	0.272 (0.204, 0.725)	0.257 (0.193, 0.682)
	160	0.4	0.607 (0.558, 0.664)	0.613 (0.570, 0.667)	0.551 (0.510, 0.602)	0.511 (0.472, 0.563)
		0.8	0.454 (0.416, 0.544)	0.456 (0.416, 0.539)	0.414 (0.374, 0.519)	0.381 (0.344, 0.486)
		1.6	0.283 (0.239, 0.540)	0.286 (0.239, 0.539)	0.258 (0.215, 0.557)	0.245 (0.199, 0.525)
	320	0.4	0.612 (0.580, 0.712)	0.619 (0.584, 0.711)	0.554 (0.528, 0.647)	0.516 (0.488, 0.606)
		0.8	0.455 (0.421, 0.653)	0.457 (0.426, 0.645)	0.410 (0.380, 0.606)	0.379 (0.352, 0.569)
		1.6	0.289 (0.254, 0.591)	0.291 (0.255, 0.584)	0.267 (0.225, 0.594)	0.256 (0.215, 0.560)
TD	80	0.4	0.505 (0.458, 0.584)	0.524 (0.477, 0.612)	0.615 (0.563, 0.715)	0.611 (0.556, 0.710)
		0.8	0.378 (0.326, 0.465)	0.398 (0.342, 0.476)	0.451 (0.390, 0.537)	0.466 (0.400, 0.550)
		1.6	0.235 (0.185, 0.667)	0.248 (0.198, 0.692)	0.298 (0.230, 0.734)	0.312 (0.238, 0.698)
	160	0.4	0.501 (0.458, 0.555)	0.522 (0.477, 0.569)	0.610 (0.570, 0.664)	0.608 (0.566, 0.653)
		0.8	0.375 (0.341, 0.442)	0.395 (0.358, 0.458)	0.453 (0.411, 0.559)	0.469 (0.421, 0.560)
		1.6	0.231 (0.196, 0.486)	0.244 (0.209, 0.504)	0.279 (0.234, 0.540)	0.296 (0.244, 0.558)
	320	0.4	0.505 (0.473, 0.630)	0.525 (0.492, 0.648)	0.614 (0.583, 0.722)	0.613 (0.575, 0.718)
		0.8	0.375 (0.349, 0.544)	0.394 (0.364, 0.554)	0.450 (0.417, 0.595)	0.466 (0.429, 0.597)
		1.6	0.234 (0.206, 0.533)	0.250 (0.219, 0.540)	0.289 (0.253, 0.563)	0.306 (0.264, 0.571)

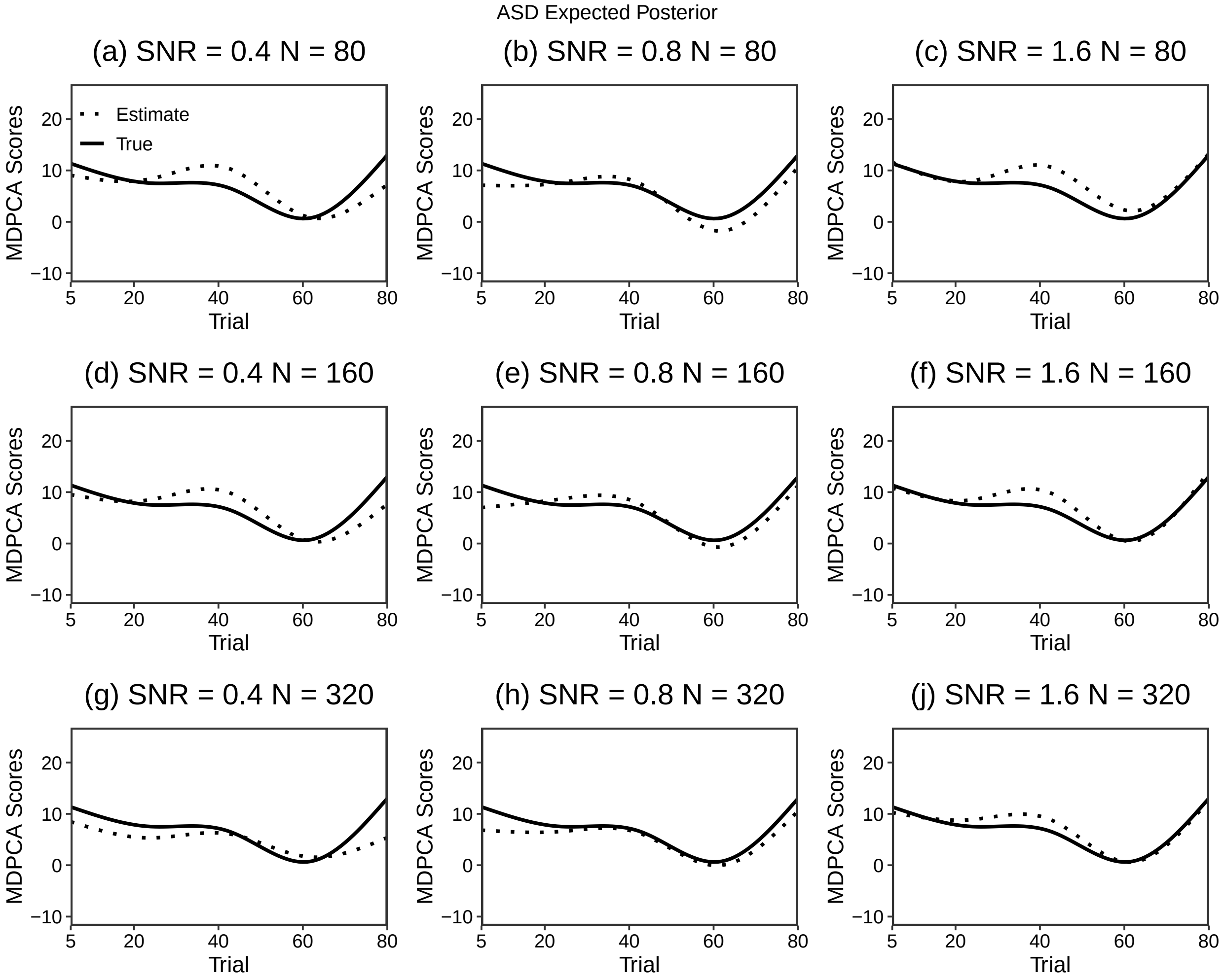


Figure 1.3: True (solid) and estimated (dotted) fixed effects mean trajectories from the run with the median ME value for the MDPCA scores of the subgroup ASD expected posterior, at varying SNRs (rows) and sample sizes (columns).

mean longitudinal trajectories are smaller than the PE values for subject and region-specific predictions, as expected, since it is harder to predict record-specific trajectories than recover group means in a mixed effects model. In addition, there is a subtle trend of decreasing ME with increasing sample size in estimation of the mean longitudinal trajectories, as expected. The ME values reported are largely stable across the three sample sizes for mean power and eigenvector estimation, perhaps due to the large total number of ERPs observed even for the lower sample size at $N = 80$, due to repetitions over electrodes and conditions. (The lower sample size at $N = 80$ is selected not only to mimic the sample size in our data application, but also to guarantee stable MDPCA decompositions.)

Tables 1.2 and 1.3 highlight the variability in the above outlined trends across the eight subgroups. More specifically while the decreasing trend with increasing SNR is pronounced for the PE values consistently across all eight subgroups, they are not consistently observed across all the subgroups in the ME values (the trends with increasing SNR are stronger for ME in some subgroups, such as both the TD and ASD groups in the unexpected condition from the frontal scalp section). This is again due to the comparatively higher PE values for the harder task of prediction compared to estimation of the group means, reflected through the ME metric. In addition, there are subtle differences in the trends observed across across scalp sections in two diagnostic groups. While ME values are larger in the ASD group in the frontal section (compared to the posterior section), they are larger in the TD group in the posterior section. Note that these differences are due to the different mixed effects coefficients used in data generation (based on fits to the original data). For example, the targeted contrasts are closer to zero in the ASD group in the frontal section (compared to posterior) and the TD group in the posterior section (compared to frontal), leading to larger ME values due to the standardization by a smaller integrand in the denominator. Figure 1.3 displays that the group mean estimates get closer to the true mean with increasing SNR and sample size for the MDPCA scores from the ASD group in the expected condition in the posterior scalp section. Results confirm that even at lower SNR values, where higher ME values are observed in the estimation of the mean power and eigenvector, group mean trajectories of the MDPCA scores are effectively recovered through the efficient mixed effects modeling proposed.

1.4 Data analysis

1.4.1 The implicit learning paradigm

The implicit learning paradigm included 27 children with Autism Spectrum Disorder (ASD) and 34 typically developing (TD) peers ranging in age from 2-5 years old. Children were presented with a continuous stream of six-colored shapes (pink squares, blue crosses, yellow circles, turquoise diamonds, gray triangles, and red octagons) which were grouped into three pairs. Presentation of each shape was a trial, resulting in an ERP waveform. The order within the paired shapes stayed the same throughout the experiment and hence could be learned as the experiment progressed, while the order across shape pairs was random. Therefore, the transition within a shape pair corresponded to the "expected" condition of the experiment and across shape pairs corresponded to the "unexpected" condition ($L = 2$, see Figure 1.1(a)). Implicit learning was detected through the difference in the ERP waveforms between the expected and unexpected conditions, and the rate of change in these differences reflected the evolution of implicit learning in the different diagnostic groups (ASD vs. TD) throughout the experiment.

There were 120 shapes presented in each condition, corresponding to the 120 trials of the experiment. EEG was recorded at $J = 24$ electrodes located in three regions (left, central, right) within two scalp sections (frontal and posterior) (see Figure 1.1(b)). In addition, EEG was sampled at 250 Hz, producing a total of $U = 250$ ERP time points per waveform spanning 1000ms. For further details on the preprocessing steps of the data, the readers are deferred to Hasenstab et al. (2015). In addition to the preprocessing steps, ERP data with no variability in ERP time have been removed from 3 ASD and 2 TD children (0.13% of the total ERP records have been removed before analysis).

1.4.2 Data analysis and interpretations

The implicit learning paradigm lead to two well-known ERP components: N1, related to early category recognition, and P3, related to cognitive processes such as signal matching, decision making and memory updating (Bugli and Lambert (2006); Jeste et al. (2015)). In our analysis, we concentrate on modeling power in the delta (.5 to 4 Hz) and theta (4 to 8 Hz) frequency bands, specifically in the way they contribute to the P3 phasic peak. Contributions of power from these two frequency bands to P3 in learning paradigms have been associated with cognitive processing (delta), and to orienting to a novel stimulus (theta) in previous works (Bernat et al. (2005, 2007); Harper et al. (2014)). A two-step filtering process was applied to the individual pre-processed ERP waveforms that were down-sampled at every fourth-time point (leading to a total of $D = 63$ translation parameters in the TFT decompositions). First, a 3rd order high-pass Butterworth filter at 1.25 Hz was applied to separate the true signal from the direct current (DC) shift observed in the data. The DC shift is a well-known phenomenon in which the electrode system measures additional voltage around 1 Hz alongside the EEG signal due to chemical reactions between the metallic surface of the electrode and the conductive gel applied to the scalp. Then, we applied high-pass and low-pass 3rd order Butterworth filters at 4 Hz to the previously filtered signal to separate the delta (.5 to 4 Hz) and theta (4 to 8 Hz) frequency bands.

Hasenstab et al. (2015) which only modeled P3 amplitudes during the paradigm showed that implicit learning (signaled by the differences in the P3 peak amplitudes between the expected and unexpected conditions) took place between trials 5 and 80, with maximal condition differentiation occurring around trial $s = 30$. Hence, we considered modeling longitudinal trends within this trial range, $s_{min} = 5$ to $S = 80$, with $k = 30$ maximal trials used in sliding windows A_s in estimation of trial-specific functional covariances. For computational efficiency, trial-specific functional covariances from every third trial were averaged in targeting the marginal functional covariance. These choices guaranteed enough total number of TFT power surfaces for stable PCA decomposition of the marginal functional covariance, where TFT decompositions involved a total of $m = D \times F = 63 \times 28 = 1764$ total parameters (total scale parameters in frequency were $F = 28$). A sensitivity analysis to the choice of the total number of trials in the sliding windows A_s was carried out on the final mixed effects modeling results of the derived MDPCA scores, where k values within the range $[20, 40]$ lead to similar results.

Before obtaining the leading eigenvectors in the functional domain, we assessed whether data from different diagnostic groups, conditions, scalp regions or frequency bands can be merged in estimation of the functional marginal covariances, as well as assessing the assumption of constant direction of functional variance across trials, leading to estimation of the marginal functional covariance. While the eigenvectors showed similar directions of variation within frequency bands across the diagnostic groups, conditions and scalp sections, they were sufficiently different between the two frequency bands considered, as expected due to initial filtering of ERPs (Appendix A.1 Figures A.1-A.12). This observation led us to consider two separate functional marginal covariances and eventually two separate longitudinal analysis of the MDPCA scores in the two frequency bands, which enabled the study of group differences across the two experimental conditions (expected vs. unexpected) within each frequency band. In addition, similarity of the directions of variation, captured by the eigenvectors of trial-specific functional covariances Σ_s for trials $s \in \{20, 40, 60\}$, signaled no violation of the constant direction of functional variation assumption for the data, leading to estimation of the functional marginal covariance within each frequency band (Appendix A.2 Figures A.13-A.18).

The estimated six leading eigenvectors ϕ_h, $h = 1, \ldots, 6$, of the functional marginal covariances explained approximately 60% and 90% of the total functional variation in the theta and delta frequency bands, respectively, where all six MDPCA scores associated with the six leading functional eigenvectors were modeled through the proposed mixed effects modeling framework. The mixed effects modeling for both the delta and theta frequency bands used four knots for the spline fits for the trial effect, selected by AIC. Multilevel random effects were considered at the subject and region levels, where the lme function in the `R` package nlme was used to fit the models.

The estimated six leading eigenvectors of the functional marginal covariance for the delta and theta frequency bands are displayed in Figures 1.4 and 1.5, respectively, where bounds for the P3 component, expected in the time window of [190, 350] ms, are depicted with vertical dashed horizontal lines. While the eigenvectors for the delta frequency band (Figure 1.4) span longer time intervals and do not isolate to the P3 expected time window that would allow interpretations for power contributions directly to the P3, the estimated first and third leading eigenvectors for the theta frequency band (Figure 1.5) highlight power contributions directly to the P3 component from lower theta (4-6 Hz) and higher theta (5-8 Hz) frequency intervals, respectively. Figures 1.4 and 1.5 also display contrasts for ((ASD expected - ASD unexpected) - (TD expected - TD unexpected)), based on the mixed effects modeling of the MDPCA scores, obtained by projecting data onto the estimated functional eigenvectors. More specifically, the estimated contrasts along with their the associated 95% bootstrap pointwise intervals (CIs), based on resampling from subjects with replacement, are given in solid black, while the 95% pointwise CIs based directly on the mixed effects modeling are shaded in gray. Note that while the mixed effects modeling based CIs only incorporate model based uncertainty, pointwise bootstrap CIs incorporate uncertainty from the entire LTFT-ERP algorithm including the MTFT and MDPCA decompositions, hence are wider as a result.

Also provided in Figures 1.4 and 1.5 are the 95% simultaneous bootstrap intervals (dashed black). Simultaneous CIs are based on a double bootstrap procedure, where the standard deviation of the estimated contrast is targeted by the first layer of bootstrap and the unknown constant $c_{0.95}$ multiplying the standard deviation is obtained from the second layer. The unknown constant $c_{0.95}$ is chosen as the 95th percentile of the distribution of the supremum of the normalized deviation of the bootstrap contrast estimate from the original contrast estimate in the data (normalized by the standard deviation of the bootstrap contrast targeted by the second layer of bootstrap) (See Dong et al. (2016) for additional details). Simultaneous confidence bands can help test for group differences in the longitudinal trends in the MDPCA scores (in the entire contrast as a function of trials) and provide insight into trials that contribute to the detected group differences. We base the interpretations below on significance of group differences in condition differentiation based on the simultaneous bootstrap CIs, which are the most conservative among the three sets of CIs provided.

Building on the interpretations from the first leading eigenvector for the theta band that relatively isolate contributions to P3 from lower theta frequencies, inference on longitudinal trends of MDPCA scores from simultaneous bootstrap CIs identify significant group differences for the ((ASD expected - ASD unexpected) - (TD expected - TD unexpected)) contrast in the posterior scalp section ((Figure 1.5(b)). More specifically, for the leading MDPCA scores, signaling power from lower theta frequencies, the group condition differentiation contrast is significant in the posterior scalp section, where the simultaneous bootstrap CIs do not contain zero at trials 48 through 54. Note that positive scores on PC1 signal higher theta power contributions to P3, when multiplied by a positive leading eigenvector. Figure 1.6 (b) and (d) further display the fitted leading score trajectories in the two conditions for the two diagnostic groups, respectively. The condition differentiation is larger in the ASD group (higher theta power contributions to P3 in the unexpected condition) in the posterior section trials 48-54 in PC1 (where simultaneous bootstrap CI for the group condition differentiation contrast does not contain zero). Hence the significant group condition differentiation is tied to the larger condition difference in this range in the ASD group. Condition differentiation in theta power contributions to P3 in the posterior scalp sections are typically associated with sensory and visual processing where the significant group condition contrast can be interpreted as the ASD group displays signs of visual processing in the later parts of the experiment.

Note that in the frontal scalp section, the TD group displays larger condition differentiation with higher theta power contribution to P3 in the unexpected condition (compared to the expected condition) for later parts of the experiment, after trial 60 (Figure 1.6(a)), where condition differentiation typically signals orienting to the novel stimulus when observed in the frontal scalp section. However while the pointwise CIs do not contain zero at trials 63 and 64 for the group condition differentiation in the frontal scalp section, the simultaneous bootstrap CIs do contain zero and do not signal significant group differences in the front scalp section (Figure 1.5(a)).

1.4.3 Comparison of results to analysis from the time and frequency domains

The LTFT-ERP analysis, not only allows us to study dynamic trends related to learning throughout the experiment, it also allows for added interpretation from the associated time-frequency transformations, where power contributions to specific ERP components from different frequency bands carry different interpretations related to underlying cognitive processes. To highlight the specific added interpretations gained from the LTFT-ERP with the time-frequency joint transformations, we review in this section results from the time domain only and frequency domain only analysis.

Note that the prior longitudinal modeling of the implicit learning paradigm by Hasenstab et al. (2015) is carried out strictly in the ERP time domain, where differences in the the longitudinal trends of the amplitude of the P3 component across trials in the frontal scalp region is depicted via a mixed effects model. Hasenstab et al. (2015) detected condition differentiation in both diagnostic groups around trial 30 of the experiment with different directions of differentiation. More specifically, while the ASD group exhibited higher P3 amplitude in the expected condition, the TD group exhibited higher P3 amplitude in the unexpected condition (results based on pointwise bootstrap CIs). In a parallel analysis of longitudinal trends in the frequency domain, we found that relative delta and theta power stay relatively constant across trials and do not display differences across conditions for both diagnostic groups.

While the time domain only analysis of Hasenstab et al. (2015) was able to connect to learning trends across the two diagnostic groups based on condition differentiation (in the P3 amplitudes), the analysis could not connect results to specific underlying cognitive processes. Note that in the time-frequency joint analysis, we are able to utilize time-frequency specific information along with scalp section, to connect results to specific processes. More specifically theta power contributions to P3 in the posterior scalp section are typically associated with sensory and visual processing, while theta power contributions to P3 in the frontal scalp section are typically associated with orienting to novel stimulus (Bernat et al. (2007)). Hence the time-frequency joint analysis brings an additional level of specificity to interpretations which is lacking in time domain only or frequency domain only analysis.

1.5 Discussion

We have proposed LTFT-ERP to model the longitudinal trends in the ERP signal over trials/duration of an EEG experiment. Longitudinal changes during the course of an experiment may be the main interest in some studies exemplified by the implicit learning paradigm, where longitudinal trends capture speed and nature of learning among TD and ASD children. LTFT-ERP utilizes time-frequency transformations to retain valuable information from both the time and frequency domains, enhancing interpretability of the findings. In addition, under the assumption of constant direction of variation in the functional domain, the algorithm borrows information across trials (i.e. the longitudinal dimension) in identifying the leading eigenvectors in the functional domain (carrying both time and frequency information) through the proposed MDPCA. This is a major stabilization tool for the proposed algorithm as it enhances the signal and enables further modeling of the longitudinal trends in the longitudinal MDPCA scores, obtained through projections of the signal onto the leading functional eigenvectors. Application to the implicit learning paradigm uncovers distinct learning patterns throughout the experiment among children diagnosed with ASD and their typically developing peers.

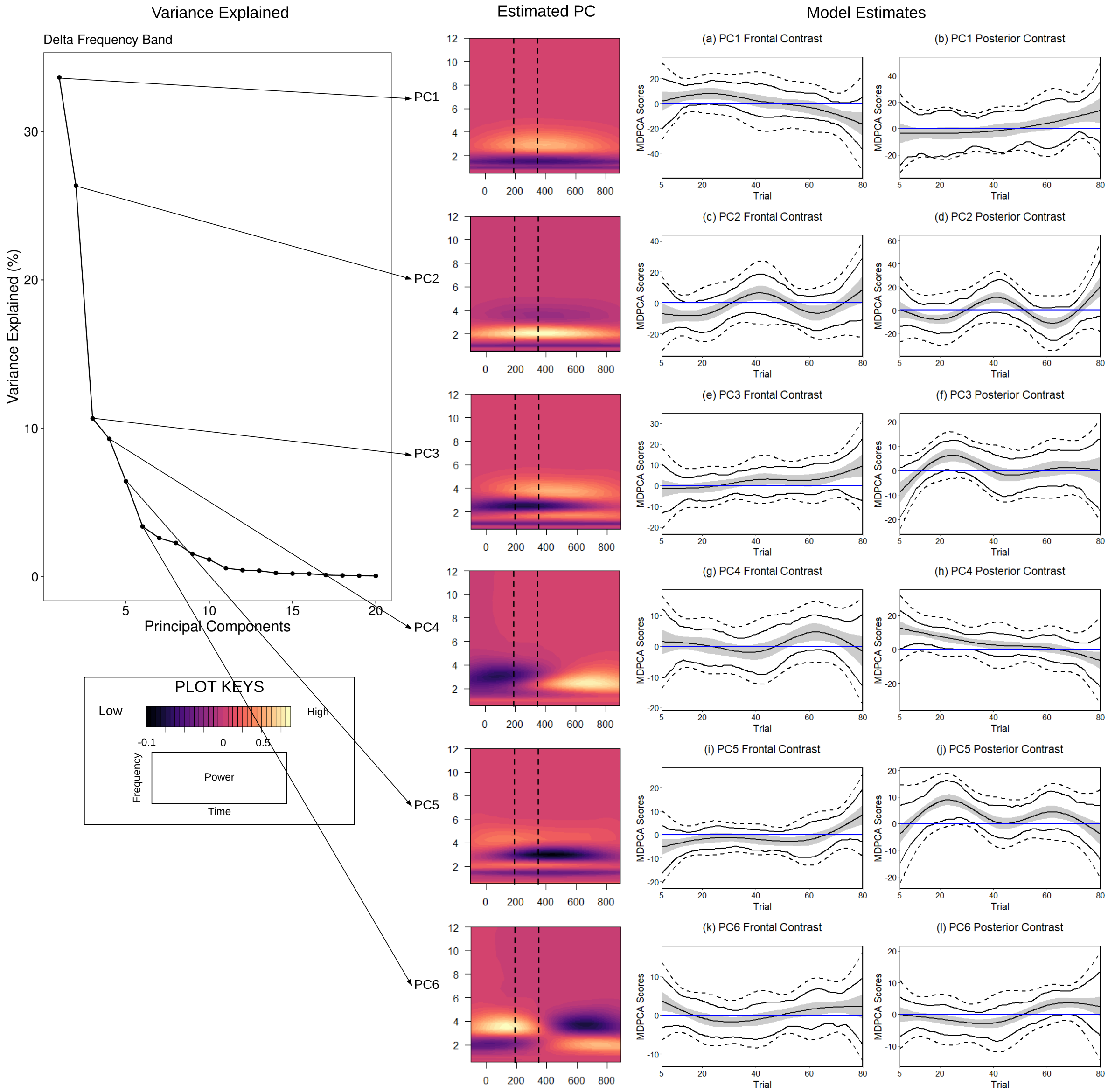


Figure 1.4: Summary of results from the proposed LTFT-ERP algorithm from the delta frequency band. The six estimated leading functional eigenvectors and their corresponding percent of variance explained are depicted. Contrasts for ((ASD expected - ASD unexpected) - (TD expected - TD unexpected)), based on the mixed effects modeling of the MDPCA scores, are also depicted. The contrasts and the associated 95% pointwise and simultaneous bootstrap intervals, based on resampling from subjects with replacement, are given in solid black and dashed black, respectively, while the 95% pointwise confidence intervals based on the mixed effects modeling are shaded in gray. A blue horizontal line at zero is included for ease of interpretation.

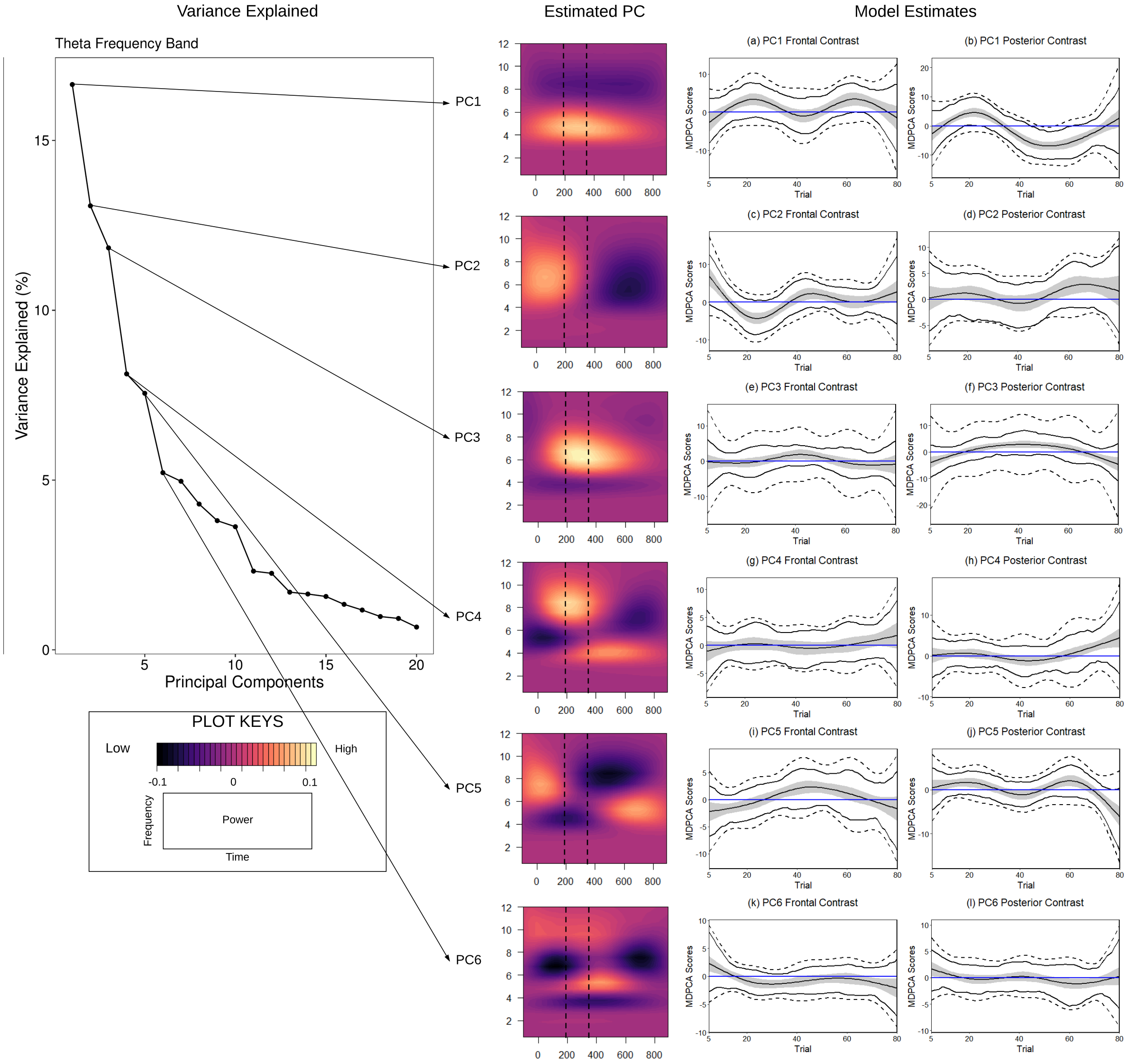


Figure 1.5: Summary of results from the proposed LTFT-ERP algorithm from the theta frequency band. The six estimated leading functional eigenvectors and their corresponding percent of variance explained are depicted. Contrasts for ((ASD expected - ASD unexpected) - (TD expected - TD unexpected)), based on the mixed effects modeling of the MDPCA scores, are also depicted. The contrasts and the associated 95% pointwise and simultaneous bootstrap intervals, based on resampling from subjects with replacement, are given in solid black and dashed black, respectively, while the 95% pointwise confidence intervals based on the mixed effects modeling are shaded in gray. A blue horizontal line at zero is included for ease of interpretation.

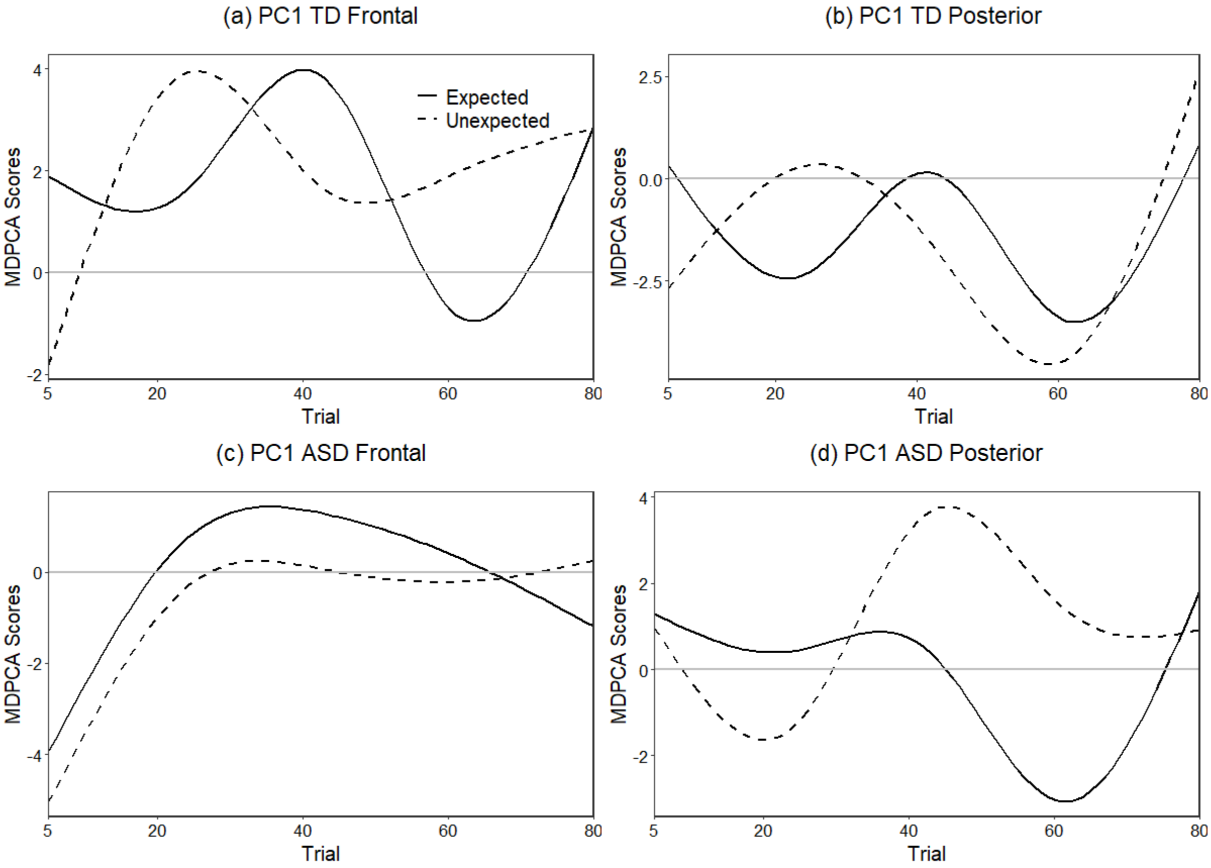


Figure 1.6: Estimated mean trajectories of the MDPCA scores for the leading first eigencomponent in the theta frequency band for each group, scalp section, and condition (expected solid and unexpected dashed). A grey horizontal line at zero is included for ease of interpretation.

CHAPTER 2

Central posterior envelopes for Bayesian functional principal component analysis

Abstract

Bayesian methods provide direct inference in functional data analysis applications without reliance on bootstrap techniques. A major tool in functional data applications is the functional principal component analysis which decomposes the data around a common mean function and identifies leading directions of variation. Bayesian functional principal components analysis (BFPCA) provides uncertainty quantification on the estimated functional model components via the posterior samples obtained. We propose central posterior envelopes (CPEs) for BFPCA based on functional depth as a descriptive visualization tool to summarize variation in the posterior samples of the estimated functional model components, contributing to uncertainty quantification in BFPCA. The proposed BFPCA relies on a latent factor model and targets model parameters within a mixed effects modeling framework using modified multiplicative gamma process shrinkage priors on the variance components. Functional depth provides a center-outward order to a sample of functions. We utilize modified band depth and modified volume depth for ordering of a sample of functions and surfaces, respectively, to derive at CPEs of the mean and eigenfunctions within the BFPCA framework. The proposed CPEs are showcased in extensive simulations. Finally, the proposed CPEs are applied to the analysis of a sample of power spectral densities (PSD) from resting state electroencephalography (EEG) where they lead to novel insights on diagnostic group differences among children diagnosed with autism spectrum disorder and their

typically developing peers across age.

2.1 Introduction

The literature on functional data analysis (FDA) has seen rapid growth in the past two decades in the analysis of data where the basic unit of measurement is a high-dimensional object such as a curve, surface or an image (Ramsay and Silverman (2005)). The wide spectrum of application areas include neuroscience, engineering, medicine, economics and geosciences. A major tool for dimension reduction is the functional principal component analysis (FPCA) for modeling functional variability in the data in lower dimensions (Wang et al. (2016); Yao et al. (2012); Cardot (2007)). Recent literature on FPCA models complex dependencies among the functional observations that are observed in close proximity with respect to time or space (Chen and Müller (2012); Greven et al. (2010); Crainiceanu et al. (2009); Di et al. (2009); Hasenstab et al. (2017); Scheffler et al. (2020); Campos et al. (2022); Zipunnikov et al. (2011); Baladandayuthapani et al. (2008); Staicu et al. (2010)). Bayesian FPCA (BFPCA) offers uncertainty quantification on the functional model components, including the mean and eigenfunctions, via credible intervals, without the need for bootstrap. Developments are typically based on expansion of the functional observation or the functional model components on a set of basis functions, followed by dimension reduction. Suarez and Ghosal (2017) expanded functional model components on a basis set and modeled the covariance function via an approximate spectral decomposition, while Montagna et al. (2012) proposed a Bayesian latent factor regression model (BLFRM) that expands each functional observation as a linear combination of a high-dimensional basis set and placed a latent factor model on the basis coefficients. Effective basis selection is achieved in the latter approach via the multiplicative gamma process shrinkage (MGPS) prior of Bhattacharya and Dunson (2011) placed on the factor loadings. Traditional data summaries for uncertainty in the Bayesian setting rely on parametric assumptions or the use of pointwise quantiles. Parametric credible intervals employed in Crainiceanu et al. (2007) assume approximate posterior normality to form pointwise or simultaneous credible intervals through estimation of pointwise variation in the posterior sample. Krivobokova et al. (2010) proposed quantile credible intervals using the estimated posterior pointwise quantiles, and extended developments to simultaneous quantile credible intervals by rescaling of the bounds of the pointwise credible intervals by a common factor.

The parametric and quantile credible intervals, used to describe uncertainty of the functional posterior estimates, have some drawbacks. Parametric credible intervals are by-design symmetric around the pointwise mean and are dependent on distributional assumptions that may not hold when modeling assumptions are violated. Quantile credible intervals allow for asymmetry using a rank-based approach, however they rely on pointwise ranks while approximating the ordering of a posterior sample of functions. To address these drawbacks, we propose central posterior envelopes (CPEs), which are not credible intervals, but are descriptive visualization tools to summarize the variation in the posterior summaries of the functional model components, contributing to uncertainty quantification in BFPCA. The proposed CPEs do not have to be symmetric around the estimated mean and are based on functional depth in ordering of a posterior sample of curves, rather than pointwise quantiles. Additionally, CPEs are formed using envelopes delineated by subsets of the posterior functional sample, and hence are fully data-driven, summarizing the variation in the posterior sample without any parametric assumptions.

Functional depth has been proposed to generalize order statistics to functional data, providing a center-outward order to a sample of functions (López-Pintado and Romo (2009); Sun et al. (2012)) and have been extended in a variety of FDA applications including construction of the median or a trimmed mean function, functional boxplots (Sun and Genton (2012)), surface boxplots (Genton et al. (2014)), outlier detection via the outliergram (Arribas-Gil and Romo (2014)), robust rank, permutation and location tests for distributional and dispersion differences in two-sample functional data groups (López-Pintado and Wrobel (2017); López-Pintado and Qian (2020)). López-Pintado and Romo (2009) introduced the notion of modified band depth (MBD), extending the definition of band depth based on a graph-based approach, by measuring the proportion of time that a curve lies in the band delimited by a subsample of curves. Sun et al. (2012) derived a computationally efficient algorithm for calculating MBD that can rank millions of curves in seconds. Genton et al. (2014) further extended MBD to higher dimensional functional data through modified volume depth (MVD). We utilize both MBD and MVD to propose functional depth based CPEs for the mean function and eigenfunctions in BFPCA. MBD-CPEs are formed by ranking the posterior estimates and forming an envelope of a subset of the posterior estimates with the largest depth values. MVD-CPEs for eigenfunctions are formed via ranking of the posterior covariance surfaces.

The chapter is organized as follows. The BFPCA model considered is introduced in Section 2.2 along with an outline of the traditional posterior summaries for the BFPCA model components. The considered BFPCA is a simplification of previous formulations in literature, where a Bayesian estimation for model components is followed by derivation of the estimated mean and eigenfunctions through singular value decomposition of the estimated covariance surfaces. The proposed model uses a latent factor model to represent the functional observations where the mean function and latent factors are further expanded on a basis set. A normal-inverse gamma prior is placed on the coefficients of the mean function, and a modified multiplicative gamma process shrinkage (MMGPS) prior is placed on the factor loadings to induce sparsity in basis selection similar to BLFRM. A computationally efficient estimation procedure is proposed for the considered BFPCA via fully conjugate priors that leads to implementation through a Gibbs sampler. The proposed CPEs based on functional depth are outlined in Section 2.3, followed by simulation studies to showcase their finite sample performance in the presence of different types of functional outliers (Section 2.4). Section 2.5 outlines application of CPEs to analysis of a sample of power spectral densities (PSD) from resting state electroencephalography (EEG). Novel insights are provided on diagnostic group differences in the evolution of PSD across age among children diagnosed with autism spectrum disorder and their typically developing peers. A brief discussion is included in Section2.6.

2.2 Bayesian functional principal component analysis

2.2.1 Model specification

Let $Y_i(t) = f_i(t) + \epsilon_i(t)$ denote the observed noisy response measurements for subject i, $i = 1, \ldots, n$, represented as a sum of a smooth underlying function $f_i(t)$, and measurement error $\epsilon_i(t)$. The measurement error $\epsilon_i(t)$ is assumed to be i.i.d. with mean zero and variance σ_ϵ^2. The smooth function $f_i(t)$ is assumed to be square integrable with mean $\mu(t)$ and covariance $C(s,t) = \text{Cov}\{f_i(s), f_i(t)\} = \sum_{k=1}^{\infty} \rho_k \psi_k(s) \psi_k(t)$, where $\rho_1 \geqslant \rho_2 \geqslant \ldots$ denote the ordered nonnegative eigenvalues, and $\psi_1(t), \psi_2(t), \ldots$ denote the corresponding eigenfunctions. While the eigenfunctions describe direction of leading modes of variation in the functional data, the eigenvalues quantify the amount of variation explained by the different modes of variation. The Karhunen-Loève (KL) expansion of $f_i(t)$ is then given by $f_i(t) \approx \mu(t) + \sum_{k=1}^{\infty} \xi_{ik} \psi_k(t)$, where $\xi_{ik} = \int_t \{f_i(t) - \mu(t)\} \psi_k(t) dt$ denotes the kth subject-specific FPCA score with mean zero and variance ρ_k. In practice, the expansion is truncated to include K eigencomponents, $f_i(t) \approx \mu(t) + \sum_{k=1}^{K} \xi_{ik} \psi_k(t)$, based on fraction of variance explained (FVE), where the covariance is approximated by $C(s,t) \approx \sum_{k=1}^{K} \rho_k \psi_k(s) \psi_k(t)$. A total of $K = 2$ or 3 eigencomponents are typically retained in most applications (explaining more than 80% of the total variation), achieving effective dimension reduction.

The BFPCA model considered is based on a latent factor model constructed for $f_i(t) = \mu(t) + \sum_{\ell=1}^{L} \eta_{i\ell} \phi_\ell(t)$, where $\phi_\ell(t)$, $\ell = 1, \ldots, L$, denote the L latent components and $\eta_{i\ell} \sim N(0,1)$, $\ell = 1, \ldots, L$, denote the corresponding uncorrelated subject-specific scores. Next, the mean function and latent components are expanded on a set of R B-spline basis functions $(b_1(t), \ldots, b_R(t))$, $\mu(t) = \sum_{r=1}^{R} \beta_r b_r(t)$ and $\phi_\ell(t) = \sum_{r=1}^{R} \lambda_{r\ell} b_r(t)$, where β_r and $\lambda_{r\ell}$ denote the mean coefficients and factor loadings, respectively. This leads to the following expansion of $f_i(t) = \sum_{r=1}^{R} [\beta_r + \sum_{\ell=1}^{L} \eta_{i\ell} \lambda_{r\ell}] b_r(t)$, which in vector form can be written as the mixed effects model

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{f}_i + \boldsymbol{\epsilon}_i = B(\boldsymbol{\beta} + \Lambda \boldsymbol{\eta}_i) + \boldsymbol{\epsilon}_i, \\ \boldsymbol{\eta}_i &\sim N_L(\mathbf{0}_L, I_L), \quad \boldsymbol{\epsilon}_i \sim N_T(\mathbf{0}_T, \sigma_\epsilon^2 I_T), \quad i = 1, \ldots, n, \end{aligned} \tag{2.1}$$

where $\boldsymbol{Y}_i = \{Y_i(t_1), \ldots, Y_i(t_T)\}^\top$ denotes the response observed at a total of T time points, $\boldsymbol{f}_i = \{f_i(t_1), \ldots, f_i(t_T)\}^\top$ denotes the $T \times 1$ vector of underlying smooth functions and $\boldsymbol{\epsilon}_i = \{\epsilon_i(t_1), \ldots, \epsilon_i(t_T)\}^\top$ denotes the $T \times 1$ vector of measurement error. Furthermore, in (2.1), $B = (\mathbf{b}_1, \ldots, \mathbf{b}_R)$ denotes the $T \times R$ matrix of B-spline basis functions with $\mathbf{b}_r = \{b_r(t_1), \ldots, b_r(t_T)\}^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_R)^\top$ denotes the $R \times 1$ vector of mean coefficients, $\Lambda = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_L)$ denotes the $R \times L$ factor loading matrix with $\boldsymbol{\lambda}_\ell = (\lambda_{i\ell}, \ldots, \lambda_{R\ell})^\top$ and $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iL})^\top$ denotes the $L \times 1$ vector of subject-specific scores. Finally, $\mathbf{0}_L$ and $\mathbf{0}_T$ are used to denote the $L \times 1$ and $T \times 1$ vectors of zeros, respectively, and I_L and I_T are used to denote the $L \times L$ and $T \times T$ identity matrices.

Note that in the BFPCA formulation in (2.1), the total number of latent components L are typically larger than K, the number of eigencomponents retained in the FPCA expansion. Based on empirical studies, Shamshoian et al. (2022) report that the estimation of the mean function and the covariance surface is robust to different choices of R, the total number of basis functions used in the expansion, provided L is large. Following guidance from Shamshoian et al. (2022), R is selected to be a fraction of the total number of time points T, $R = \lfloor T/2 \rfloor$, to provide adequate smoothing of $f_i(t)$, and L is selected to be a fraction of R, $L = max(6, \lfloor R/4 \rfloor)$. Fully conditional conjugate priors on the variance components are selected to achieve efficient posterior estimation in BFPCA. More specifically, using Gaussian priors for the mean coefficients $\boldsymbol{\beta}$ and factor loadings $\boldsymbol{\lambda}_\ell$, noninformative prior for the error variance σ_ϵ^2 (proportional to a constant c), gamma prior for the variance of the mean coefficients σ_β^2 and a modified multiplicative gamma process shrinkage (MMGPS) prior for the variance components of the factor loading matrix $\sigma^2_{\lambda_{r\ell}}$,

$$\boldsymbol{\beta} \sim N_R\left(\mathbf{0}_R, \frac{1}{\sigma^2_\beta}\Omega^{-1}\right), \quad \sigma^2_\beta \sim \text{Gamma}\left(\frac{a_\beta}{2}, \frac{a_\beta}{2}\right), \quad \frac{1}{\sigma^2_\epsilon} \propto c \tag{2.2}$$

$$\boldsymbol{\lambda}_\ell \sim N_R\left(\mathbf{0}_R, \Sigma_{\lambda_\ell}\right), \quad \Sigma_{\lambda_\ell} = \text{diag}(\sigma^2_{\lambda_{1\ell}}, \ldots, \sigma^2_{\lambda_{R\ell}}), \quad \sigma^2_{\lambda_{r\ell}} = \varphi^{-1}_{r\ell}\tau^{-1}_\ell, \quad \varphi_{r\ell} \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

$$\tau_\ell = \prod_{h=1}^{\ell} \delta_h, \quad \delta_1 \sim \text{Gamma}(a_1, 1), \quad \delta_h \sim \text{Gamma}(a_2, 1) I(\delta_h > 1), \quad h \geqslant 2.$$

we target the posterior distributions in model (2.1) using a Gibbs sampler (posterior distributions and details on choice of hyperparameters, a_β, ν, a_1, a_2, are deferred to Appendix B.1). In (2.2), $\mathbf{0}_R$ denotes an $R \times 1$ vector of zeros, Ω is a positive-definite $R \times R$ penalty matrix, Σ_{λ_ℓ} is a $R \times R$ diagonal matrix comprised of the variance components, $\sigma^2_{\lambda_{1\ell}}, \ldots, \sigma^2_{\lambda_{R\ell}}$, for the the ℓth factor loading $\boldsymbol{\lambda}_\ell$, and $I(\cdot)$ denotes the indicator function. The variance components of the factor loading matrix, denoted by $\sigma^2_{\lambda_{r\ell}}$, are given a MMGPS prior adapted from Bhattacharya and Dunson (2011), where $\varphi_{r\ell}$ and τ_ℓ denote the element-wise and column-wise precisions, respectively. The column-wise precision τ_ℓ is the cumulative product of gamma distributed δ_h for $h = 1, \ldots, \ell$. The truncation of δ_h to be larger than one when $h \geqslant 2$ guarantees that τ_ℓ increases with ℓ, forcing columns of Λ, $\boldsymbol{\lambda}_\ell$, to get stochastically smaller as ℓ increases. Note that this mimics the estimation of eigenfunctions in FPCA with ordered (decreasing) eigenvalues and results in effective basis selection (Montagna et al. (2012)).

2.2.2 Traditional posterior summaries for BFPCA components

Posterior estimates of the mean coefficient vector $\boldsymbol{\beta}^{(m)} = (\beta^{(m)}_1, \ldots, \beta^{(m)}_R)^\top$ and factor loading matrix $\Lambda^{(m)} = (\boldsymbol{\lambda}^{(m)}_1, \ldots, \boldsymbol{\lambda}^{(m)}_L)$, with $\boldsymbol{\lambda}^{(m)}_\ell = (\lambda^{(m)}_{1\ell}, \ldots, \lambda^{(m)}_{R\ell})^\top$, where the superscript m, $m = 1, \ldots, M$, is used to index the posterior estimates obtained from the MCMC sampler after burn-in and thinning, leads to the posterior estimates of the mean function, $\mu^{(m)}(t) = \sum_{r=1}^R \beta^{(m)}_r b_r(t)$, and the covariance $C^{(m)}(s,t) = \sum_{r=1}^R \sum_{r'=1}^R \sum_{\ell=1}^L \lambda^{(m)}_{r\ell} \lambda^{(m)}_{r'\ell} b_r(s) b_{r'}(t)$. In an attempt to recover the additional interpretations offered by the lower dimensional representation of FPCA (where eigenfunctions describe the leading modes of variation in the functional data), we consider the singular value decomposition of $C^{(m)}(s,t) \approx \sum_{k=1}^{K} \rho_k^{(m)} \psi_k^{(m)}(s)\psi_k^{(m)}(t)$, targeting the posterior estimates of the eigenfunctions, $\psi_k^{(m)}(t)$, and eigenvalues, $\rho_k^{(m)}$. The total number of eigencomponents retained, K, is chosen by the mean or the median FVE calculated across the posterior samples. Since the sign of the eigenfunctions are not identifiable, we implement an additional alignment step in obtaining the posterior estimates of the eigenfunctions (see Appendix B.2 for further details).

The mean estimate is obtained by averaging the posterior estimates, $\widehat{\mu}(t) = (1/M)\sum_{m=1}^{M} \mu^{(m)}(t)$. The eigenfunctions and eigenvalues can be targeted in two ways. The first approach is to average the posterior estimates, $\widehat{\psi}_k(t) = (1/M)\sum_{m=1}^{M} \psi_k^{(m)}(t)$, $\widehat{\rho}_k = (1/M)\sum_{m=1}^{M} \rho_k^{(m)}$, $k = 1, \ldots, K$, similar to the mean estimate. An alternative approach is to first target the mean of the posterior covariances, $\widetilde{C}(s,t) = (1/M)\sum_{m=1}^{M} C^{(m)}(s,t)$, followed by SVD of $\widetilde{C}(s,t) \approx \sum_{k=1}^{K} \widetilde{\rho}_k \widetilde{\psi}_k(s)\widetilde{\psi}_k(t)$, leading to the eigenfunction and eigenvalue estimates obtained via covariance estimation, denoted by $\widetilde{\psi}_k(t)$ and $\widetilde{\rho}_k$, respectively. While we evaluate the finite sample performance of both point estimates for the eigenfunctions and eigenvalues via simulations, we center the traditional credible intervals for these quantities around the posterior average estimates.

A main advantage of BFPCA is the readily available inference provided for the FPCA components based on the posterior sample. While credible intervals can be constructed for the scalar components (including eigenvalues or FVE) using the standard deviation or percentiles obtained from the posterior sample, we center our discussion mainly on inference for the functional components, which is the focus of the proposed functional depth based approach. For the functional components of FPCA, i.e. the mean and eigenfunctions, traditional posterior summaries include pointwise and simultaneous, parametric and quantile credible intervals. In the formulations below, parametric and quantile credible intervals will be denoted by capital ‘P’ and ‘Q’, respectively, for ease of notation, while pointwise and simultaneous credible intervals will be distinguished by the superscripts ‘p’ and ‘s’. While the parametric intervals are based on variance of the estimates in the posterior sample, the quantile intervals leverage pointwise quantiles obtained from the posterior sample. Let $g(t)$ denote either the mean or eigenfunction of interest, observed at time points t_j, $j = 1, \ldots, T$, and let $\widehat{g}(t_j)$ and $\widehat{Var}\{g(t_j)\}$ denote the sample mean and variance of $g(t_j)$ calculated from the M MCMC samples $g^{(m)}(t)$. The $(1-\alpha)100\%$ pointwise parametric credible interval for $g(t)$ is given by $[\widehat{g}(t_j) \pm z_\alpha \sqrt{\widehat{Var}\{g(t_j)\}}; j = 1, \ldots, T]$ (denoted as $P^p_{1-\alpha}\{g(t)\}$), where $z_\alpha = \Phi^{-1}(1-\alpha/2)$ with Φ denoting the CDF of the standard normal distribution. For defining the simultaneous parametric credible intervals, let c_α denote the $(1-\alpha)$ sample quantile of $\max_{j=1,\ldots,T} \mid \{g^{(m)}(t_j) - \widehat{g}(t_j)\}/\sqrt{\widehat{Var}\{g(t_j)\}} \mid$ over the M posterior samples. Then the $(1-\alpha)100\%$ simultaneous parametric credible interval for $g(t)$ is given by $[\widehat{g}(t_j) \pm c_\alpha \sqrt{\widehat{Var}\{g(t_j)\}}; j = 1, \ldots, T]$ (denoted as $P^s_{1-\alpha}\{g(t)\}$) (Crainiceanu et al. (2007)).

For the pointwise and simultaneous quantile credible intervals, let $g_{\alpha/2}(t_j)$ and $g_{1-\alpha/2}(t_j)$ denote the pointwise $\alpha/2$ and $(1-\alpha/2)$ sample quantiles of $g^{(m)}(t)$, $m = 1, \ldots, M$ at $t = t_j$, respectively. Then $(1-\alpha)100\%$ pointwise quantile credible interval for $g(t)$ is given by $[\{g_{\alpha/2}(t_j), g_{1-\alpha/2}(t_j)\}; j = 1, \ldots, T]$ (denoted as $Q^p_{1-\alpha}\{g(t)\}$). The $(1-\alpha)100\%$ simultaneous quantile credible interval for $g(t)$ is given by $[\widehat{g}(t_j) + q\{g_{\alpha/2}(t_j) - \widehat{g}(t_j)\}, \widehat{g}(t_j) + q\{g_{1-\alpha/2}(t_j) - \widehat{g}(t_j)\}; j = 1, \ldots, T]$ (denoted as $Q^s_{1-\alpha}\{g(t)\}$), where q is a common factor that rescales the upper and lower bounds until $(1-\alpha)100\%$ of the posterior estimates are contained inside the credible interval (Krivobokova et al. (2010)). Note that the factor q is common across all time points t_j in the above formulation and that both pointwise and simultaneous quantile credible intervals rely on pointwise quantiles while trying to quantify uncertainty in estimation of a functional component. The proposed central posterior envelopes based on functional depth consider ranking of the entire functional estimates in the posterior sample, rather than relying on pointwise quantiles. The code for implementation of the quantile credible intervals can be found in the R package *acid*. The notations used for the traditional posterior summaries outlined above, as well as the functional depth based intervals proposed in the next section, are summarized for the readers reference in Table 2.1.

2.3 Proposed functional depth based CPEs for BFPCA components

The traditional posterior summaries outlined in Section 2.2.2 have potential pitfalls when describing posterior distributions of functional data. The parametric credible intervals rely on distributional assumptions, which may be violated under deviation from modeling assumptions. In addition, the parametric credible intervals are symmetric around the pointwise mean, which is restrictive when capturing potential asymmetry in posterior distributions in the presence of highly variable and/or skewed posterior samples. Although the quantile credible intervals are data-driven and more flexible with potential asymmetry, their construction relies on pointwise quantiles. In particular, instead of treating posterior samples as functional data, the quantile credible intervals are estimated by ranking of the posterior samples at each time point. Furthermore, scaling the pointwise quantiles by a common factor q across all time points in construction of the simultaneous quantile credible intervals may be too restrictive in modeling the spread in the functional posterior samples. In order to circumvent these issues, the use of functional depth is proposed to rank the functional posterior estimates to obtain fully data-driven CPEs that capture uncertainty in a flexible way.

Functional depth is a measure that provides a center-outward ordering of a sample of functional observations. In particular, functional depth ranks a sample of functions from the 'deepest' curve with the highest functional depth value, defined as the median curve, to the most outlying curve with the lowest functional depth value. Although there are a number of functional depth measures that have been proposed in the literature (see Zuo and Serfling (2000) and Gijbels and Nagy (2017)), modified band depth based on a graph-based approach has been quite popular in applications López-Pintado and Romo (2009). Let $g^{(1)}(t), \ldots, g^{(M)}(t)$, $t \in \mathcal{I}$, denote a sample of M functional posterior estimates, either for the mean or eigenfunctions, defined on a compact interval $\mathcal{I}$, where $\mathcal{I} \in \mathbb{R}$. The band in $\mathbb{R}^2$

delineated by a subset of u, $2 \leqslant u \leqslant M$, posterior estimates, $g^{(m_1)}(t), \ldots, g^{(m_u)}(t)$, drawn from the full posterior sample $\{g^{(1)}(t), \ldots, g^{(M)}(t)\}$ is given by

$$\mathcal{B}\left\{g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\right\} = \left[\{t, g(t)\} : t \in \mathcal{I}, \min_{v=m_1,\ldots,m_u} g^{(v)}(t) \leqslant g(t) \leqslant \max_{v=m_1,\ldots,m_u} g^{(v)}(t)\right].$$

Appendix Figures B.1(a) and (b) represent two bands $\mathcal{B}\left\{g^{(1)}(t), g^{(2)}(t)\right\}$ and $\mathcal{B}\left\{g^{(3)}(t), g^{(4)}(t)\right\}$, delimited by two curves, where in the first figure the entire graph of $g^{(m)}(t)$ and in the second a proportion of it is included in the band. Band depth of López-Pintado and Romo (2009) considers the proportion of bands $\mathcal{B}\left\{g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\right\}$ determined by u different curves $g^{(m_1)}(t), \ldots, g^{(m_u)}(t)$ containing the graph of $g(t)$. Modified band depth (MBD) extends band depth, such that rather than the proportion of bands that contain the entire graph of $g(t)$, MBD considers the proportion of time that the graph of $g(t)$ lies inside the bands. More specifically, let $\mathcal{A}_u\{g^{(m)}(t); g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\} = \{t \in \mathcal{I} : \min_{v=m_1,\ldots,m_u} g^{(v)}(t) \leqslant g^{(m)}(t) \leqslant \max_{v=m_1,\ldots,m_u} g^{(v)}(t)\}$ denote the set in the interval $\mathcal{I}$ where the function $g^{(m)}(t)$ lies inside the band $\mathcal{B}\{g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\}$. Further let $\mathcal{A}^*_u\{g^{(m)}(t); g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\} = \theta[\mathcal{A}_u\{g^{(m)}(t); g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\}]/\theta(\mathcal{I})$ denote the proportion of time that the curve $g^{(m)}(t)$ lies inside the band $\mathcal{B}\{g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\}$, where $\theta(\cdot)$ denotes the Lebesgue measure on $\mathcal{I}$. If $2 \leqslant U \leqslant M$ denotes a fixed total number of curves used to delineate a band, then the modified band depth (MBD) for the curve $g^{(m)}(t)$ in $g^{(1)}(t), \ldots, g^{(M)}(t)$ given U is

$$MBD_{M,U}\left\{g^{(m)}(t)\right\} = \sum_{u=2}^{U}\left[\binom{M}{2}^{-1} \sum_{1 \leqslant m_1 < m_2 < \ldots < m_u \leqslant M} \mathcal{A}^*_u\left\{g^{(m)}(t); g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\right\}\right].$$

MBD ranks each curve in the sample $g^{(1)}(t), \ldots, g^{(M)}(t)$ as the sum of $\mathcal{A}^*_u\{g^{(m)}(t); g^{(m_1)}(t), \ldots, g^{(m_u)}(t)\}$ over all possible combinations of bands delineated by $2 \leqslant U \leqslant M$ total curves that can be formed. Functions where a higher proportion of the curve lies in a higher number of bands get a higher MBD value, representing curves that are closer to the center of the sample. Those curves with lower MBD values have a lower proportion of the curve lying in a lower proportion of bands, representing outlying observations, resulting in an effective ranking of the sample. We follow common practice and set $U = 2$ in applications for computational efficiency (and drop U from the MBD notation), where bands delineated by all combinations of two curves are considered in the MBD definition (i.e. $MBD_M\{g^{(m)}(t)\} \equiv MBD_{M,2}\{g^{(m)}(t)\}$).

MBD of the posterior estimates are utilized to obtain point estimates (i.e. MBD median) and functional band depth central posterior envelopes (i.e. MBD-CPEs for the BFPCA functional components ($\mu(t)$ and $\psi_k(t)$). While the MBD median equals the functional median of the posterior sample, the $(1-\alpha)100\%$ MBD-CPE is formed by the band delineated by the $1-\alpha$ deepest posterior estimates in the sample. The MBD median and $(1-\alpha)100\%$ MBD-CPE formed for the functional BFPCA component $g(t)$, denoted by $\widehat{m}\{g(t)\}$ and $D_{1-\alpha}\{g(t)\}$, respectively, are targeted via Algorithm 1. *Note that similar to functional depth based central envelopes proposed for functional data, CPEs are not credible intervals, but they rather are descriptive visualization tools that help summarize the variation in the posterior sample. When plotted for a grid of α values, CPEs display the central envelopes allowing for visualization of the most central regions of the functional posterior distributions.*

```
**Algorithm 1: MBD Median and MBD-CPE for $g(t) \equiv \mu(t)$ or $g(t) \equiv \psi_k(t)$**
Step 1: Calculate the MBD of the posterior samples: $MBD_M\{g^{(1)}(t)\}, \ldots, MBD_M\{g^{(M)}(t)\}$.
Step 2: Order the MBD values of the posterior sample from the smallest to the largest and
    denote the corresponding ordered samples as $g^{[1]}(t), \ldots, g^{[M]}(t)$.
Step 3: Calculate the $(1-\alpha)100\%$ MBD-CPE as
    $D_{1-\alpha}\{g(t)\} = B\left\{g^{[\lfloor \alpha M \rfloor+1]}(t), \ldots, g^{[M]}(t)\right\}$,
    and the MBD median as $\widehat{m}\{g(t)\} = g^{[M]}(t)$.
```

An alternative way for obtaining point estimates and CPEs for the eigenfunctions $\psi_k(t)$, $k = 1, \ldots, K$, is to rank the posterior covariance surfaces using modified volume depth (MVD). MBD has been extended to surface data as MVD to provide a way to rank two-dimensional functional data (Sun et al. (2012); Genton et al. (2014)) (for extensions to higher-dimensional functional data, see López-Pintado and Wrobel (2017)). Let $C^{(1)}(s,t), \ldots, C^{(M)}(s,t)$,

$(s,t) \in \mathcal{S}$, denote a sample of M posterior covariance estimates, defined on $\mathcal{S} \in \mathbb{R}^2$. Further let $\mathcal{A}_u\{C^{(m)}(s,t); C^{(m_1)}(s,t), \ldots, C^{(m_u)}(s,t)\} = \{(s,t) \in \mathcal{S} : \min_{v=m_1,\ldots,m_u} C^{(v)}(s,t) \leqslant C^{(m)}(s,t) \leqslant \max_{v=m_1,\ldots,m_u} C^{(v)}(s,t)\}$ denote the region in $\mathcal{S}$ where the covariance $C^{(m)}(s,t)$ lies inside the simplex delineated by the covariances $C^{(m_1)}(s,t), \ldots, C^{(m_u)}(s,t)$. In addition, $\mathcal{A}_u^*\{C^{(m)}(s,t);$
$C^{(m_1)}(s,t), \ldots, C^{(m_u)}(s,t)\} = \vartheta[\mathcal{A}_u\{C^{(m)}(s,t); C^{(m_1)}(s,t), \ldots, C^{(m_u)}(s,t)\}]/\vartheta(\mathcal{S})$ is used to denote the proportion of time that the covariance $C^{(m)}(s,t)$ lies inside the simplex formed by the covariances $C^{(m_1)}(s,t), \ldots, C^{(m_u)}(s,t)$, where $\vartheta(\cdot)$ denotes the Lebesgue measure extended to $\mathbb{R}^2$. Considering $U = 2$ total covariances to delineate a simplex (similar to the definition of MBD), MVD for the covariance $C^{(m)}(s,t)$ in $C^{(1)}(s,t), \ldots, C^{(M)}(s,t)$ is given as

$$MVD_M\left\{C^{(m)}(s,t)\right\} = \binom{M}{2}^{-1} \sum_{1 \leqslant m_1 < m_2 \leqslant M} \mathcal{A}_2^*\left\{C^{(m)}(s,t); C^{(m_1)}(s,t), C^{(m_2)}(s,t)\right\}.$$

MVD of the posterior covariance estimates are utilized to obtain point estimates (i.e. MVD median) and functional volume depth CPEs (i.e. MVD-CPEs) for $\psi_k(t)$, $k = 1, \ldots, K$. The MVD median and $(1-\alpha)100\%$ MVD-CPE formed for $\psi_k(t)$, denoted by $\widetilde{m}\{\psi_k(t)\}$ and $D^{\star}_{1-\alpha}\{\psi_k(t)\}$, respectively, are targeted via Algorithm 2. The algorithm starts with ranking the posterior covariances using MVD and obtaining their corresponding eigenfunctions. The MVD median for $\psi_k(t)$ equals the functional median of the posterior eigenfunctions, while the $(1-\alpha)100\%$ MBD-CPE for $\psi_k(t)$ is formed by the band delineated by the $1-\alpha$ deepest eigenfunction estimates in the sample, where the eigenfunction estimates are ordered according to MVD of their corresponding posterior covariances.

Functional depth based medians proposed above ($\widehat{m}\{\mu(t)\}$, $\widehat{m}\{\psi_k(t)\}$ and $\widetilde{m}\{\psi_k(t)\}$) estimate the central tendency in the posterior distributions of the mean and eigenfunctions and provide a realistic estimate equal to one of the observed functional posterior samples rather than relying on pointwise averages as is done in most of the traditional posterior summaries. In addition, the proposed functional depth based CPEs (denoted by $D_{1-\alpha}\{g(t)\}$ and

Algorithm 2: MVD Median and MVD-CPE for $\psi_k(t)$

```
Step 1: Calculate the MVD of the posterior covariances:
    $$MVD_M\{C^{(1)}(s,t)\}, \ldots, MVD_M\{C^{(M)}(s,t)\}.$$
Step 2: Order the MVD values of the posterior sample from smallest to largest and denote
    the corresponding covariances as $C^{[1]}(s,t), \ldots, C^{[M]}(s,t)$.
Step 3: The SVD of the ordered covariances lead to their corresponding ordered eigenfunctions
    $\psi_k^{[1]}(t), \ldots, \psi_k^{[M]}(t)$.
Step 4: Calculate the $(1-\alpha)100\%$ MVD-CPE as
    $$D^{\star}_{1-\alpha}\{\psi_k(t)\} = B\left\{\psi_k^{[\lfloor \alpha M \rfloor + 1]}(t), \ldots, \psi_k^{[M]}(t)\right\},$$
    and the MVD median as $\widetilde{m}\{\psi_k(t)\} = \psi_k^{[M]}(t)$.
```

$D^{\star}_{1-\alpha}\{\psi_k(t)\}$, corresponding to MBD and MVD based summaries, respectively), are formed from a band delineated from the $1-\alpha$ deepest subset of the functional posterior estimates. Hence, the proposed methods use functional depth measures to construct fully data-driven summaries that capture the uncertainty in the posterior estimates in a flexible way. Notations used for point estimates and credible intervals both in the traditional summaries and in the proposals are summarized in Table 2.1.

2.4 Simulation Studies

We consider five simulation scenarios to display the use of CPEs in describing the variation in the posterior samples in the presence of functional outliers. We also study the finite sample properties of the traditional summaries from Section 2.2.2 under different simulation scenarios. More specifically five simulation scenarios are proposed: Case 1 – no outliers, Case 2 – magnitude outlier, Case 3 – amount of variation outlier, Case 4 – time-shifted eigenfunction outlier, Case 5 – higher-frequency eigenfunction outlier. Magnitude outliers are generated by adding a constant deviation (with a random sign) to $\mu(t)$ for $t \geqslant T_i$, where $T_i \sim \text{Unif}[0,1]$. This adds constant variation to randomly selected portions of the unit time

Table 2.1: Notation used for point estimates for the BFPCA components (mean function $\mu(t)$, eigenfunction $\psi_k(t)$, eigenvalue ρ_k and covariance $C(s,t)$) and credible intervals for $\mu(t)$ and $\psi_k(t)$ obtained from traditional summaries as well as the proposed CPEs.

	Traditional Summaries				Proposed Summaries			
BFPCA Components	Point Estimates		Credible Intervals		Point Estimates		Central Posterior Envelopes (CPEs)	
$\mu(t)$	$\widehat{\mu}(t)$	Mean estimate	$P^p_{1-\alpha}\{\mu(t)\}$ $P^s_{1-\alpha}\{\mu(t)\}$ $Q^p_{1-\alpha}\{\mu(t)\}$ $Q^s_{1-\alpha}\{\mu(t)\}$	Pointwise parametric credible interval Simultaneous parametric credible interval Pointwise quantile credible interval Simultaneous quantile credible interval	$\widehat{m}\{\mu(t)\}$	MBD median	$D_{1-\alpha}\{\mu(t)\}$	MBD-CPE
$\psi_k(t)$	$\widehat{\psi}_k(t)$ $\widetilde{\psi}_k(t)$	Eigenfunction estimate Eigenfunction estimate via covariance estimation	$P^p_{1-\alpha}\{\psi_k(t)\}$ $P^s_{1-\alpha}\{\psi_k(t)\}$ $Q^p_{1-\alpha}\{\psi_k(t)\}$ $Q^s_{1-\alpha}\{\psi_k(t)\}$	Pointwise parametric Simultaneous parametric Pointwise quantile Simultaneous quantile	$\widehat{m}\{\psi_k(t)\}$ $\widetilde{m}\{\psi_k(t)\}$	MBD median MVD median	$D_{1-\alpha}\{\psi_k(t)\}$ $D^{\star}_{1-\alpha}\{\psi_k(t)\}$	MBD-CPE MVD-CPE
ρ_k	$\widehat{\rho}_k$ $\widetilde{\rho}_k$	Eigenvalue estimate Eigenvalue estimate via covariance estimation						
$C(s,t)$	$\widetilde{C}(s,t)$	Covariance estimate						

domain in $[0, 1]$, where more of the variation is added to the latter part of the domain. Amount of variation outliers are generated using larger eigenvalues ρ_k, which result in added variation throughout the unit interval. Finally, eigenfunction outliers are generated from time-shifted (Case 4) or higher frequency (Case 5) eigenfunctions, where additional variation is added along the direction of the time-shifted or higher frequency eigenfunctions used. Results are reported for outliers generated as $q = 10$ and 20% of the sample ($n = 50$), where details of data generation under the five simulation scenarios are deferred to Appendix B.3.

The BFPCA model is fitted using $R = 20$ B-spline basis functions with equidistant knots in $[0, 1]$ and $L = 6$ latent factors, for functional data observed at a uniform grid of 40 time points. Results are reported based on a total of 200 Monte Carlo runs with $25,000$ MCMC iterations ($5,000$ for burn-in and thinning at every 5th iteration), and $M = 4,000$ posterior estimates for each Monte Carlo run. Finite sample performance of point estimates of the functional model components (i.e. mean and eigenfunctions) and scalar model components (i.e. eigenvalues) are assessed via the standardized integrated mean squared error (IMSE), $IMSE_{\widehat{g}(t)} = [\int_t \{\widehat{g}(t) - g(t)\}^2 dt] / \int_t g^2(t) dt$, and the standardized mean squared error (MSE), $MSE_{\widehat{\rho_k}} = (\widehat{\rho}_k - \rho_k)^2/\rho_k^2$, respectively. The mean IMSE and MSE values from 200 Monte Carlo runs for the five simulation scenarios are summarized in Table 2.2. The traditional and proposed point estimates for the mean function and the leading two eigenfunctions from the Monte Carlo run with the median IMSE are given in Appendix Figure B.2, Figure 2.1 and Figure 2.2, respectively.

Appendix Figure B.3 and Figures 2.3 and 2.4 display CPEs from α cutoffs ranging from 0.05 to 0.95 for the mean function and the leading two eigenfunctions, respectively, from a single Monte Carlo run overlaying $M = 4,000$ posterior estimates (given in gray) for the five simulation scenarios (for outlier percentage equal to $q = 20\%$ in Figures 2.3 and 2.4). CPEs from increasing α cutoffs help visualize regions with the most central functional posterior observations in the sample. Note that while MBD-CPEs for the eigenfunctions are nested in each other for increasing α values, MVD-CPEs are not necessarily nested since they are based

Table 2.2: The mean standardized integrated mean squared error (IMSE) and standardized mean squared error (MSE) for both the traditional and functional depth-based point estimates from the 200 Monte Carlo runs. The five simulation cases correspond to: Case 1 – no outliers, Case 2 – magnitude outlier, Case 3 – amount of variation outlier, Case 4 – time-shifted eigenfunction outlier, Case 5 – higher-frequency eigenfunction outlier with $q = \{10, 20\}\%$ outliers.

		$q = 10\%$				$q = 20\%$			
Point Estimate	Case 1	Case 2	Case 3	Case 4	Case 5	Case 2	Case 3	Case 4	Case 5
	IMSE	**IMSE**				**IMSE**			
$\widehat{\mu}(t)$	0.0033	0.0077	0.0042	0.0040	0.0039	0.0117	0.0047	0.0037	0.0042
$\widehat{m}\{\mu(t)\}$	0.0036	0.0080	0.0045	0.0042	0.0042	0.0122	0.0048	0.0040	0.0045
$\widehat{\psi}_1(t)$	0.0303	0.4723	0.0441	0.0549	0.0320	0.8975	0.0693	0.0894	0.0395
$\widetilde{\psi}_1(t)$	0.0308	0.5104	0.0445	0.0568	0.0320	0.9375	0.0713	0.0903	0.0398
$\widehat{m}\{\psi_1(t)\}$	0.0344	0.5223	0.0481	0.0603	0.0358	0.9525	0.0768	0.0956	0.0442
$\widetilde{m}\{\psi_1(t)\}$	0.0367	0.4955	0.0481	0.0606	0.0400	0.9303	0.0723	0.0972	0.0518
$\widehat{\psi}_2(t)$	0.0482	1.1887	0.0565	0.0670	0.1187	1.3740	0.0790	0.1017	0.5130
$\widetilde{\psi}_2(t)$	0.0477	1.3365	0.0554	0.0672	0.1290	1.4922	0.0786	0.0971	0.5951
$\widehat{m}\{\psi_2(t)\}$	0.0559	1.3476	0.0644	0.0746	0.1424	1.4891	0.0869	0.1101	0.6284
$\widetilde{m}\{\psi_2(t)\}$	0.0676	1.3488	0.0682	0.0812	0.1364	1.4796	0.0867	0.1106	0.6197
	MSE	**MSE**				**MSE**			
$\widehat{\rho}_1$	0.0303	0.4723	0.0441	0.0549	0.0320	0.8975	0.0693	0.0894	0.0395
$\widetilde{\rho}_1$	0.0308	0.5104	0.0445	0.0568	0.0320	0.9375	0.0713	0.0903	0.0398
$\widehat{\rho}_2$	0.0482	1.1887	0.0565	0.0670	0.1187	1.3740	0.0790	0.1017	0.5130
$\widetilde{\rho}_2$	0.0477	1.3365	0.0554	0.0672	0.1290	1.4922	0.0786	0.0971	0.5951

on the functional depth rankings of the posterior covariance surfaces, rather than posterior eigenfunctions. This is also the reason why MVD-CPEs are typically wider than MBD-CPEs, incorporating variation from the entire covariance process, rather than only eigenfunction-specific variation. More specifically, the first simulation scenario of no outliers shows that the traditional point estimates ($\widehat{\psi}_k(t)$ and $\widetilde{\psi}_k(t)$) perform quite well (yielding small IMSE and MSE values) compared to their depth based counter parts ($\widehat{m}\{\psi_k(t)\}$ and $\widetilde{m}\{\psi_k(t)\}$) in the absence of outliers (Table 2.2). The magnitude outliers in the second simulation scenario add a constant deviation from the mean function over a random portion of the time domain ($t \in [T_i, 1]$) with a random sign. Due to the random sign of the constant deviation, rather than biasing the mean function, they increase the variation in the mean function estimation (IMSE values for mean estimation are higher for Case 2 than other simulation scenarios in Table 2.2). The addition of the constant deviation also leads to identifiability issues in eigenfunction estimation, where CPEs of the first eigenfunction portray the constant variation, especially in the second half of the unit time domain. MBD-CPEs up to $\alpha = 0.95$ and MVD-CPEs up to $\alpha = 0.50$ are constant in the second half of the unit interval (Figure 2.3). CPEs of the second eigenfunction capture variation along the first eigenfunction as the second major direction of variation (following the constant variation in the second half of the unit time interval) (Figure 2.4). This leads to higher IMSE values in eigenfunction estimation for Case 2 compared to other simulation cases. Note that the traditional point estimate $\widehat{\psi}_k(t)$ has the smallest IMSE values for Case 2, possibly due to the cancellation of the effects of the constant deviation with a random sign while averaging over the posterior samples.

Case 3 generates outliers with larger eigenvalues, which increase the variation along the eigenfunctions. Due to the shapes of the two eigenfunctions considered this adds variation across the entire unit interval, as is detected through the CPEs in Figures 2.3 and 2.4. While this additional variation does not bias the point estimates (except for ρ_k), the IMSE in estimation of the eigenfunctions are slightly higher than Case 1 with no outliers. Finally, Cases 4 and 5 consider direction of variation outliers where the functional data is generated under altered eigenfunctions: in Case 4 eigenfunctions are shifted in time, in Case 5 the frequency of the eigenfunctions is increased. Both cases lead to added variation in estimation of the eigenfunctions (consistent with higher IMSE values for targeting $\psi_k(t)$ compared to Case 1 in Table 2.2). More specifically, while the time shift in both eigenfunctions adds variation to eigenfunction estimation throughout the unit interval in Case 4, the added variation due to outliers with higher frequency eigenfunctions is apparent especially in Figures 2.4 (i) and (j) with higher frequency posterior estimates of $\psi_2(t)$ captured in the $\alpha = 0.95$ MBD-CPEs and $\alpha = 0.25$ MVD-CPEs.

We also assess how well the traditional credible intervals reflect variation in the posterior sample under the five simulation scenarios. Appendix Figures B.4, B.5 and B.6 display the 95% parametric and quantile credible intervals and 95% CPEs for reference, for the mean function and the leading two eigenfunctions, respectively, from a single Monte Carlo run overlaying $M = 4,000$ posterior estimates for the five simulation scenarios (for outlier percentage equal to $q = 20\%$). The added variation in the posterior estimates of especially the eigenfunctions really help portray the shortcomings of the symmetry restrictions and constant multipliers used in enlarging of the simultaneous parametric and quantile credible intervals, respectively. Appendix Figures B.5 and B.6 (d), (g), (j) and (m) show that the symmetry restriction (around the pointwise mean) of the parametric credible intervals force the credible intervals to be too wide in one bound and too narrow in the other, especially when the variation in the posterior sample is not symmetric. While the quantile credible intervals perform better relative to their parametric counter parts, they also include regions that do not represent posterior sample variation, perhaps due to the restrictive enlargement of the credible intervals by a constant multiplier that is kept the same over the entire time domain. The latter point is best observed in Table 2.3, reporting two metrics: negative area ratio (NAR) and area ratio (AR). NAR and AR capture the ratio of the area of the credible interval or CPE that lies outside and inside of posterior sample (proportional to the area of the posterior sample), respectively. Hence, while higher AR values (AR $\in [0, 1]$)

correspond to a more realistic portrayal of the variation in the posterior sample estimates in the presence of functional outliers, NAR values greater than zero can signal problems with the credible intervals (i.e. via inclusion of regions that do not represent the posterior sample), possibly due to symmetry constraints. Similar to previous sections, let $g^{(1)}(t), \ldots, g^{(M)}(t)$ denote the M MCMC samples, where $g(t) \equiv \mu(t)$ or $\psi_k(t)$. In addition, denote the pointwise lower and upper bounds of the posterior sample by $g_{\min}(t_j) = \min_{m=1,\ldots,M}\{g^{(m)}(t_j)\}$ and $g_{\max}(t_j) = \max_{m=1,\ldots,M}\{g^{(m)}(t_j)\}$, respectively, at a total of T grid points t_j, $j = 1, \ldots, T$. Finally denote the pointwise lower and upper bounds of the discretized $(1-\alpha)100\%$ credible interval $H_{1-\alpha}\{g(t)\}$ by $H_{1-\alpha}\{g(t_j)\} = [L(t_j), U(t_j)]$ for t_j, $j = 1, \ldots, T$. Then NAR and AR are given as follows

$$
\begin{aligned}
NAR[H_{1-\alpha}\{g(t)\}] = & \frac{\sum_{j=1}^{T}[\{U(t_j) - g_{\max}(t_j)\}I\{U(t_j) > g_{\max}(t_j)\}]}{\sum_{j=1}^{T}\{g_{\max}(t_j) - g_{\min}(t_j)\}} \\
& + \frac{\sum_{j=1}^{T}[\{g_{\min}(t_j) - L(t_j)\}I\{g_{\min}(t_j) > L(t_j)\}]}{\sum_{j=1}^{T}\{g_{\max}(t_j) - g_{\min}(t_j)\}}
\end{aligned}
$$

$$
AR[H_{1-\alpha}\{g(t)\}] = \frac{\sum_{j=1}^{T}[\min\{U(t_j), g_{\max}(t_j)\} - \max\{L(t_j), g_{\min}(t_j)\}]}{\sum_{j=1}^{T}\{g_{\max}(t_j) - g_{\min}(t_j)\}}.
$$

For the proposed CPEs and pointwise quantile credible intervals, NAR always equals zero, since these summaries are based on pointwise or functional ordering of the data and therefore have to lie within the posterior sample. However the simultaneous quantile credible intervals may have NAR values larger than zero, since the bounds of the pointwise quantile intervals are rescaled by a common factor that is constant across t_j. The mean NAR and AR values based on the 200 Monte Carlo runs across the five simulation scenarios are summarized in Table 2.3. In summary, simultaneous credible intervals lead to higher AR values than their pointwise counterparts (parametric and quantile), as expected, and have nonzero NAR values except for pointwise quantile credible intervals. This also confirms that they can cover regions not representing posterior sample variation as is observed from the figures. The CPEs have AR values equal to or larger than all traditional credible intervals and have

NAR equal to zero by definition. Hence, CPEs provide a more flexible representation of the shape and spread of the posterior sample in the presence of functional outliers (or added variability in estimation due to violation of modeling assumptions) and can contribute to the visualization of the functional posterior sample in applications.

2.5 Data Application

We use CPEs in the analysis of EEG power spectral densities obtained on a sample of 58 children with autism spectrum disorder (ASD) and 39 of their typically developed (TD) peers at resting state (Dickinson et al. (2018)). The goal of the study was to characterize the shift in the peak alpha frequency (PAF), a neurological biomarker defined as the location of a single prominent peak in the alpha frequency band (6-14Hz) of the spectral density, across development. It was of particular interest to compare ASD and TD groups in their evolution of the alpha peak across chronological age. PAF has been shown to shift from lower to higher frequencies as children grow older in the TD group (Somsen et al. (1997); Stroganova et al. (1999); Dustman et al. (1999); Chiang et al. (2011); Cragg et al. (2011); Miskovic et al. (2015)), however previous research has suggested that this chronological shift in the location of the PAF may be delayed or absent in children with ASD (Edgar et al. (2015)). In our motivating study, electroencephalogram (EEG) data was sampled at 500Hz for 2 minutes using a 128-channel HydroCel Geodesic Sensor Net during an "eyes-open" resting-state paradigm in which bubbles were presented on a screen in a sound-attenuated room. The participants in the two diagnostic groups were age-matched with ages ranging from 25 to 146 months old with a median age of 66 and 65.8 months in the TD and ASD groups, respectively. To more generally capture the shape of the PSD in the alpha frequency band and to avoid the challenges involved in identifying a unique PAF for each subject, we consider scalp-averaged relative PSD from the alpha frequency band as our sample of functional data observed over $T = 33$ equidistant frequencies within 6-14Hz. The BFPCA

Table 2.3: The mean AR and NAR values for traditional and functional-depth based 95% credible intervals and CPEs over 200 Monte Carlo runs for all simulation cases with $q = \{10, 20\}\%$ outliers. $P^p_{1-\alpha}\{g(t)\}$, $P^s_{1-\alpha}\{g(t)\}$, $Q^p_{1-\alpha}\{g(t)\}$, $Q^s_{1-\alpha}\{g(t)\}$, $D_{1-\alpha}\{g(t)\}$ and $D^\star_{1-\alpha}\{g(t)\}$ denote the pointwise parametric and simultaneous credible intervals, quantile pointwise and simultaneous credible intervals, and MBD and MVD-CPEs, respectively.

			AR						NAR		
$g(t)$	q	Case	$P^p_{1-\alpha}\{g(t)\}$	$P^s_{1-\alpha}\{g(t)\}$	$Q^p_{1-\alpha}\{g(t)\}$	$Q^s_{1-\alpha}\{g(t)\}$	$D_{1-\alpha}\{g(t)\}$	$D^\star_{1-\alpha}\{g(t)\}$	$P^p_{1-\alpha}\{g(t)\}$	$P^s_{1-\alpha}\{g(t)\}$	$Q^s_{1-\alpha}\{g(t)\}$
$\mu(t)$	-	Case 1	0.522	0.795	0.523	0.796	0.870	-	0.000	0.000	0.000
	10%	Case 2	0.524	0.788	0.525	0.790	0.907	-	0.000	0.000	0.000
		Case 3	0.524	0.794	0.525	0.796	0.867	-	0.000	0.000	0.000
		Case 4	0.524	0.798	0.525	0.799	0.875	-	0.000	0.000	0.000
		Case 5	0.523	0.802	0.524	0.803	0.889	-	0.000	0.000	0.000
	20%	Case 2	0.526	0.782	0.527	0.784	0.910	-	0.000	0.000	0.000
		Case 3	0.524	0.790	0.525	0.791	0.859	-	0.000	0.000	0.000
		Case 4	0.522	0.792	0.523	0.793	0.868	-	0.000	0.000	0.000
		Case 5	0.523	0.807	0.524	0.808	0.904	-	0.000	0.000	0.000
$\psi_1(t)$	-	Case 1	0.431	0.668	0.432	0.669	0.688	0.838	0.000	0.008	0.000
	10%	Case 2	0.469	0.690	0.477	0.700	0.736	0.928	0.006	0.065	0.007
		Case 3	0.428	0.644	0.429	0.650	0.656	0.836	0.001	0.019	0.000
		Case 4	0.428	0.634	0.430	0.641	0.642	0.859	0.002	0.034	0.001
		Case 5	0.436	0.679	0.437	0.680	0.721	0.874	0.000	0.011	0.000
	20%	Case 2	0.426	0.639	0.430	0.645	0.666	0.878	0.002	0.032	0.003
		Case 3	0.444	0.645	0.448	0.654	0.648	0.846	0.005	0.047	0.001
		Case 4	0.481	0.684	0.490	0.697	0.691	0.900	0.008	0.086	0.002
		Case 5	0.430	0.675	0.432	0.676	0.727	0.904	0.000	0.014	0.001
$\psi_2(t)$	-	Case 1	0.483	0.757	0.483	0.760	0.815	0.902	0.000	0.007	0.001
	10%	Case 2	0.548	0.780	0.555	0.787	0.841	0.972	0.008	0.101	0.009
		Case 3	0.472	0.722	0.472	0.729	0.761	0.888	0.001	0.017	0.001
		Case 4	0.470	0.706	0.472	0.716	0.733	0.899	0.001	0.031	0.002
		Case 5	0.487	0.746	0.488	0.750	0.803	0.955	0.001	0.038	0.003
	20%	Case 2	0.507	0.736	0.513	0.743	0.792	0.965	0.005	0.071	0.005
		Case 3	0.482	0.711	0.485	0.723	0.731	0.887	0.004	0.044	0.003
		Case 4	0.514	0.734	0.521	0.751	0.745	0.923	0.007	0.082	0.006
		Case 5	0.589	0.842	0.583	0.844	0.915	0.988	0.004	0.111	0.010

model was estimated using $R = 16$ B-spline basis functions and $L = 6$ latent components, leading to $M = 4000$ posterior samples for each model component. Further information on pre-processing of the EEG data and the experiment are deferred to Appendix B.4.

Figures 2.5 (a) and (b) display a sample of the data obtained on 10 subjects from the TD and ASD diagnostic groups, respectively. Note that the data is quite noisy, where PAF is distinctly visible in only a subset of the subjects, where there is considerable variation in PAF and the amplitude of the alpha peak. Our goal is to summarize the mean and variation trends in the data using the BFPCA model and to characterize the variability in the posterior estimates of the model components using CPEs in both diagnostic groups. Also given in Figure 2.5 are the MBD median and 95% MBD-CPEs for the mean function estimated in both diagnostic groups (Figure 2.5 (c)) and in age-based subgroups within the TD and ASD samples (Figures 2.5 (d) and (e), respectively). The age-based subgroups were obtained within each diagnostic group by using the group-specific median age. While the PAF on the estimated overall means are similar across the two diagnostic groups (at 9 and 9.5Hz in TD and ASD, respectively), the trend in PAF across age-based subgroups is quite different within the TD and the ASD samples. While a clear developmental shift is observed in the TD sample (with PAF at 8.75 and 10.25Hz for young and old TD groups, respectively), the PAF in the ASD sample is quite similar across the two age-groups (observed around 9.5Hz). These results are consistent with previous literature and findings from our own work (Scheffler et al. (2019) and Scheffler et al. (2022)).

The CPEs from α cutoffs ranging from 0.05 to 0.95 for the leading two eigenfunctions for both diagnostic groups, overlaying $M = 4,000$ posterior estimates in gray, are given in Figure 2.6. The two leading eigencomponents explain more than 60% of the total variation in both groups (median and (2.5th, 97.5th) percentiles of FVE at 66.5% (59.5%, 73.4%) and 63.0% (57.8%, 68.7%) in the TD and ASD groups, respectively). The CPEs of the third and fourth leading eigenfunctions, explaining 14.9% (10.7%, 20.2%) and 10.2% (7.3%, 12.8%) FVE in the TD group and 15.6% (12.5%, 19.1%) and 11.7% (8.9%, 14.6%) FVE in the ASD group, respectively, are deferred to Appendix Figure B.8. While the leading eigenfunction mostly captures variation in PAF (location of the prominent alpha peak) especially in the TD group, the second leading eigenfunction captures variation in the magnitude of the alpha peak (Figures 2.6 (a), (b), (e) and (f)). The PAF varies between 8.75-10.25Hz in the TD group and over a slightly larger range between 7-9.75Hz in the ASD group (Figures 2.6 (a), (b), (c) and (d)). Note that posterior estimates of the leading eigenfunction are more variable in the TD group than the ASD group, where CPE-MVD envelopes are wider at lower alpha levels up to 0.35, similar to the observations from our simulation study. The variation difference between groups is perhaps due to the stronger developmental shift in PAF from lower to higher frequencies as children grow older in the TD group. The leading eigenfunction signals combination of variation in PAF and some variation in the magnitude of the alpha peak in the ASD group. The second leading eigenfunction capturing variation in the amplitude of the prominent alpha peak, signals variation around 9.25Hz in the TD group and at 7.25Hz and 9.25Hz in the ASD group (Figure 2.6 (e), (f), (g) and (h)). In the second leading eigenfunction, signaling variation in the amplitude of the prominent alpha peak, there is more variation in the ASD group compared to the TD group. For the readers reference CPEs for the third and fourth leading eigenfunctions, capturing remaining variation in the concavity of the PSD in the alpha frequency band and 95% parametric and quantile credible intervals along with 95% CPEs are deferred to Appendix Figures B.7, B.8 and B.9. Consistent with results from the simulation section, simultaneous parametric credible intervals include regions that do not represent variation in the posterior sample.

2.6 Discussion

We propose a descriptive tool to visualize the variation in the posterior sample of the functional model components of BFPCA. The BFPCA modeling considered relies on a latent factor model and MMGPS priors on the variance components, leading to an easy to implement estimation framework and a direct way for inference on the model components. We recover the highly interpretable mean and eigenfunctions following Bayesian estimation and propose functional depth based summaries for these quantities. The novel summaries proposed are shown to lead to a data-driven approach in portraying the variability of the functional model components. Traditional summaries rely on distributional assumptions or suboptimal symmetry constraints, and fail to treat the posterior sample as functional data. In contrast, the proposed summaries are based on ranking of the posterior sample for the functional components using functional depth. Two functional depth based summaries are considered, one based on direct ranking of the entire posterior functional sample and another through ranking of the associated covariance surfaces. Both approaches have been shown to lead to flexible modeling of the variation in the posterior sample, where the second leads to wider CPEs as expected, incorporating variation from the entire covariance process, rather than a single eigenfunction.

Extensions of the proposed methodology to higher dimensional functional data would be of interest, especially in EEG applications. EEG data is collected across the scalp, creating spatially indexed functional data. In addition, data are collected across multiple trials time locked to presentation of a sequence of stimuli in stimulus-based experiments and across time in resting state paradigms. When changes across experimental time are of interest, these repetitions can be viewed as an additional dimension of the observed data (i.e. as longitudinally observed functional data) and be part of analysis rather than collapsed via averaging. FPCA modeling has been considered for high-dimensional functional data, especially in EEG data applications involving a spatial or a longitudinal dimension (Shamshoian et al. (2022); Li et al. (2020); Campos et al. (2022); Scheffler et al. (2019, 2020); Hasenstab et al. (2017)). Developments rely on simplifying assumptions on the higher dimensional covariance via strong or weak separability. Extension of functional depth based central posterior envelopes to BFPCA for higher-dimensional functional data is an interesting direction for future research.

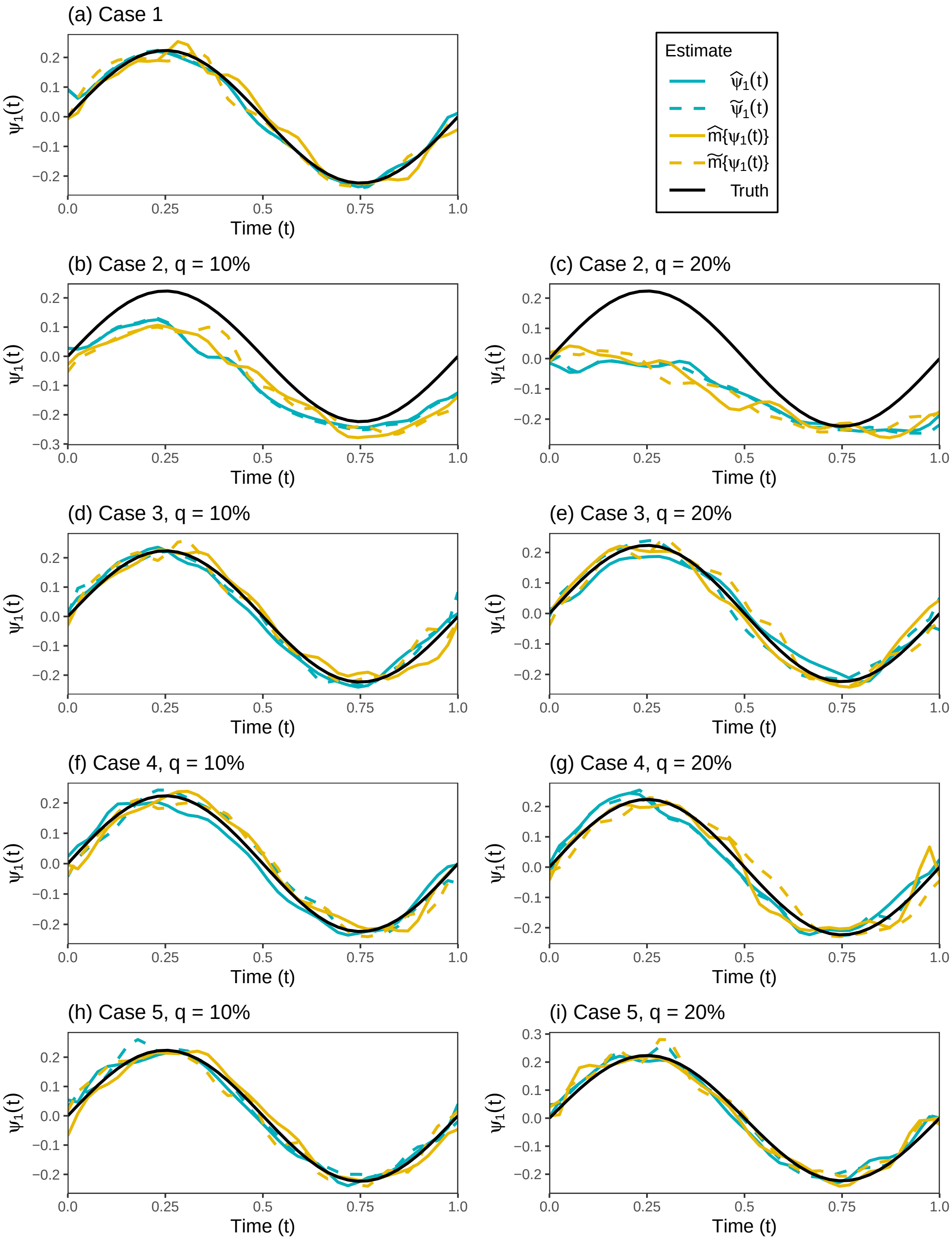


Figure 2.1: Point estimates of $\psi_1(t)$ for each simulation case with $q = 10$ and $q = 20\%$ outliers from runs with 50th percentile IMSE values. Eigenfunction estimates, eigenfunctions estimates via covariance estimation, MBD median and MVD median are given in solid blue, dashed blue, solid yellow and dashed yellow, overlaying the true function given in solid black.

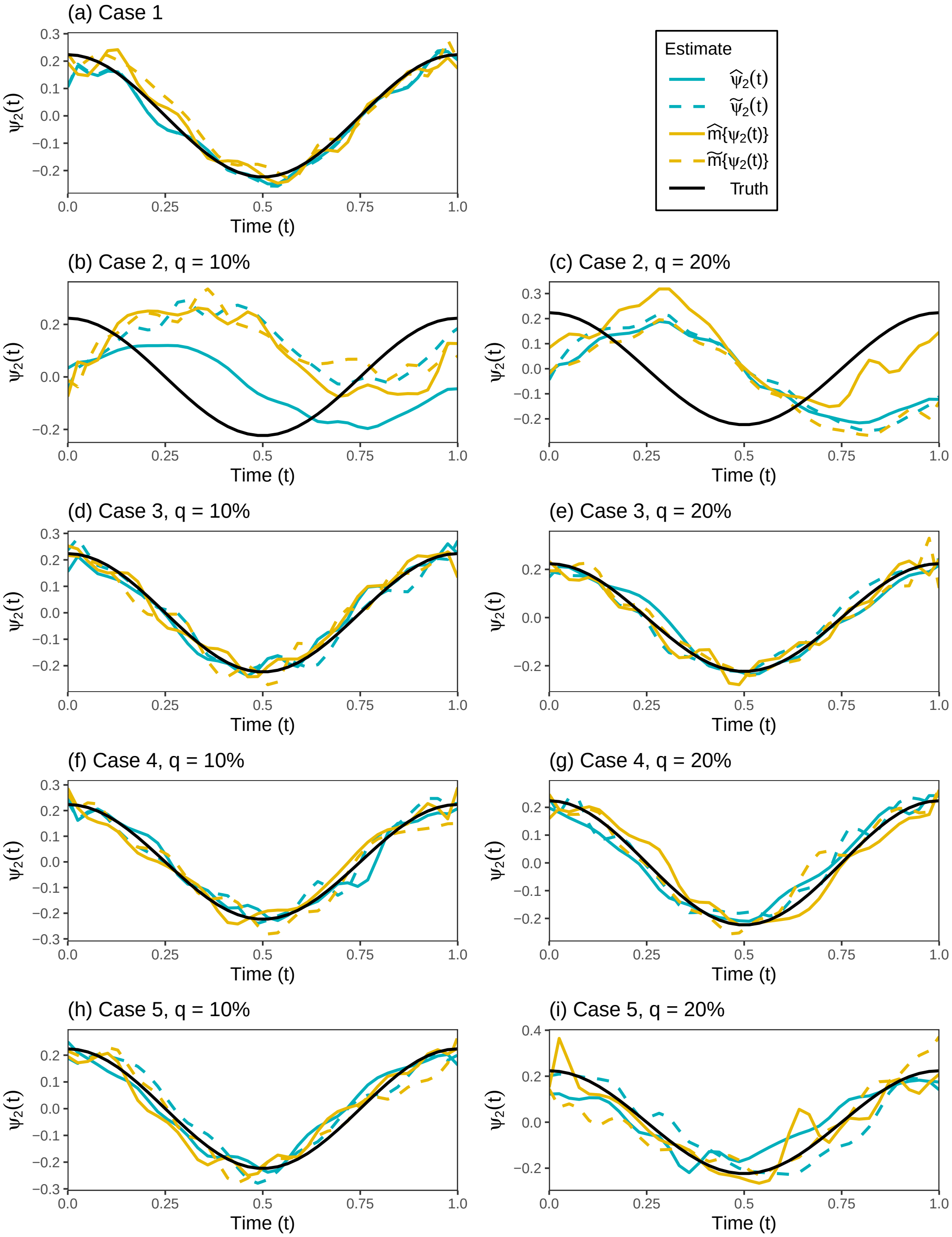


Figure 2.2: Point estimates of $\psi_2(t)$ for each simulation case with $q = 10$ and $q = 20\%$ outliers from runs with 50th percentile IMSE values. Eigenfunction estimates, eigenfunctions estimates via covariance estimation, MBD median and MVD median are given in solid blue, dashed blue, solid yellow and dashed yellow, overlaying the true function given in solid black.

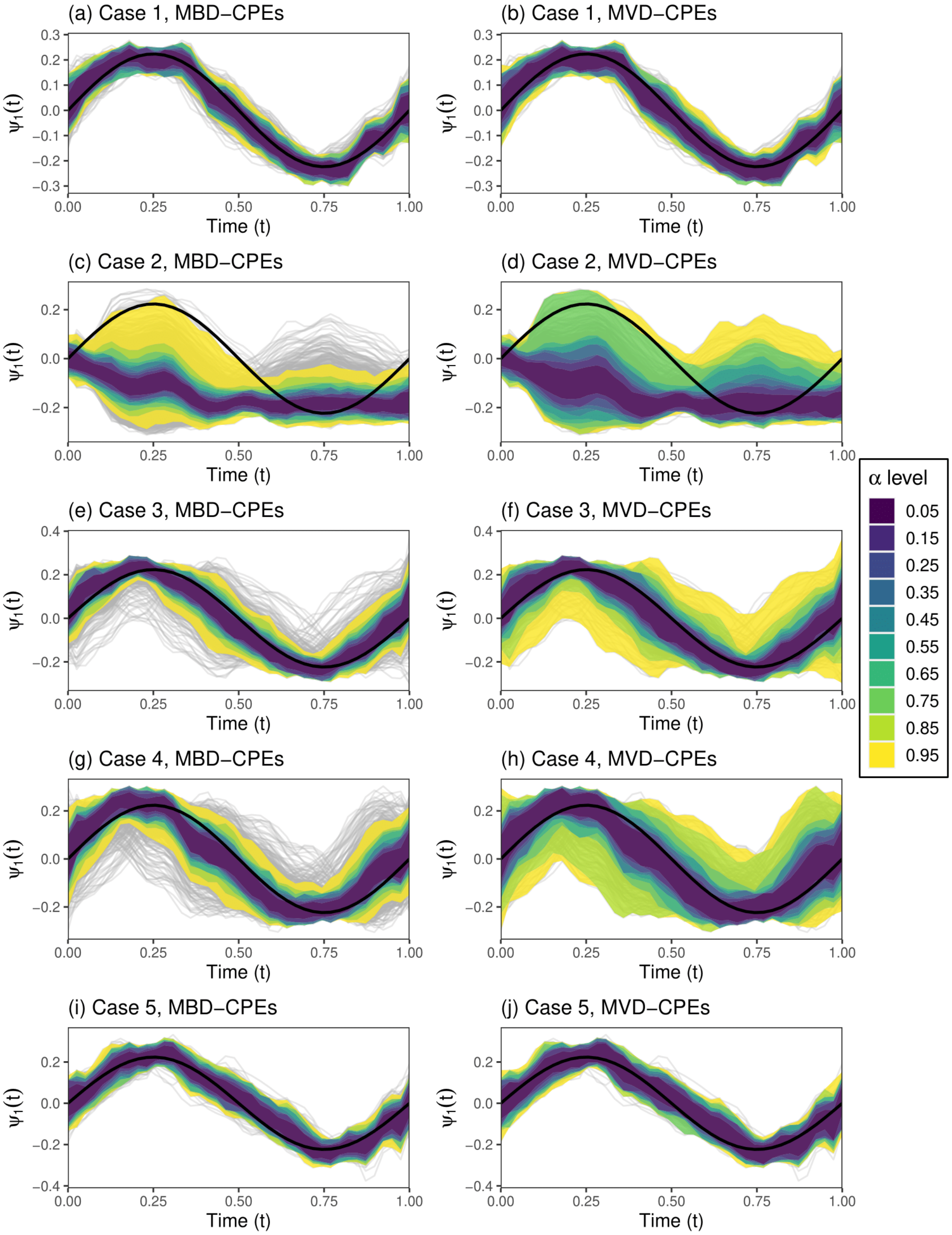


Figure 2.3: CPE contours of $\psi_1(t)$ for each simulation case with $q = 20\%$ outliers. The light grey solid lines, overlaying the true function in solid black, represent the sample of $M = 4000$ posterior estimates. The left and right hand columns display the MBD and MVD-CPEs, denoted by $D_{1-\alpha}\{\psi_1(t)\}$ and $D^{\star}_{1-\alpha}\{\psi_1(t)\}$, respectively, at a grid of α levels marked by varying contour colors.

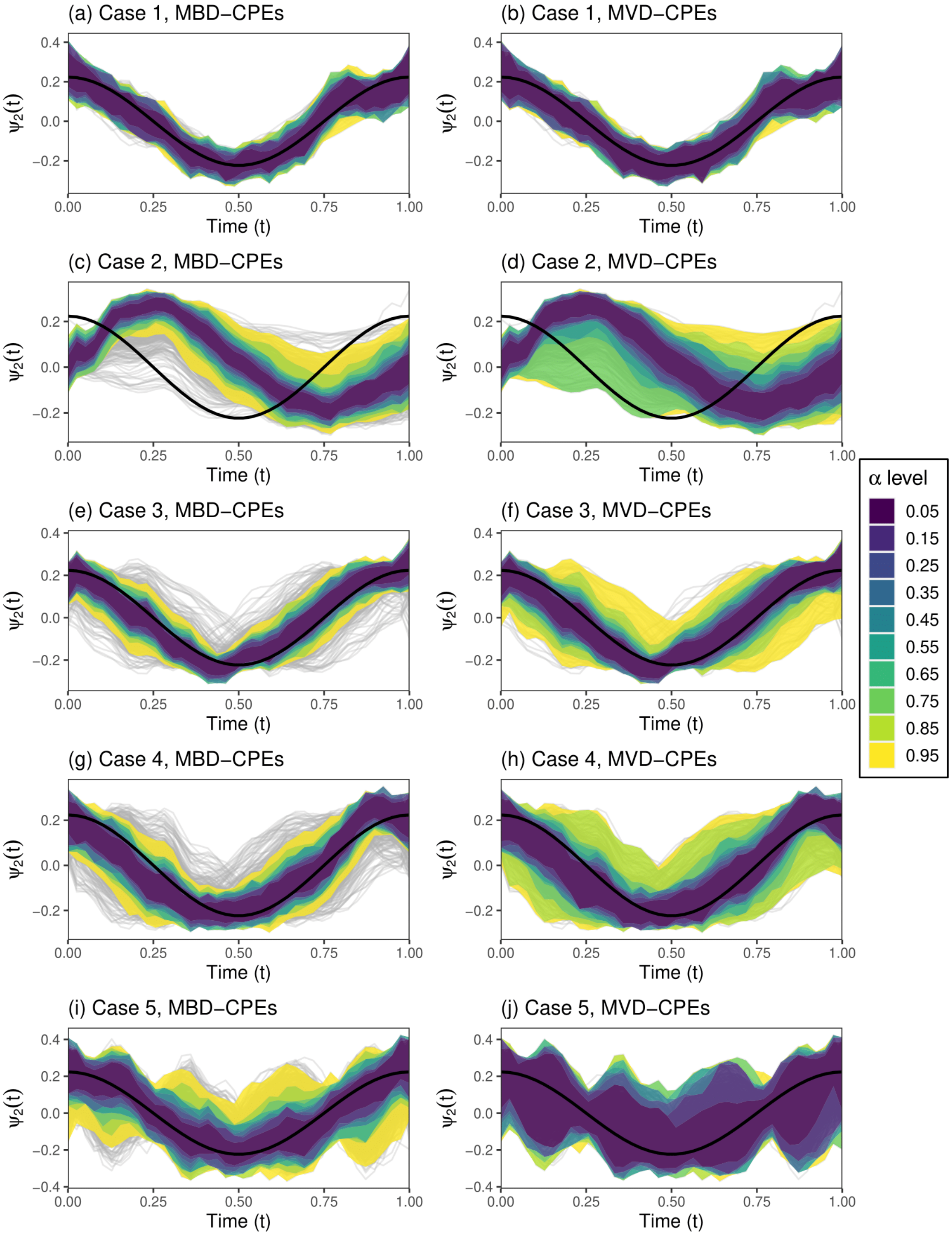


Figure 2.4: CPE contours of $\psi_2(t)$ for each simulation case with $q = 20\%$ outliers. The light grey solid lines, overlaying the true function in solid black, represent the sample of $M = 4000$ posterior estimates. The left and right hand columns display the MBD and MVD-CPEs, denoted by $D_{1-\alpha}\{\psi_1(t)\}$ and $D^{\star}_{1-\alpha}\{\psi_1(t)\}$, respectively, at a grid of α levels marked by varying contour colors.

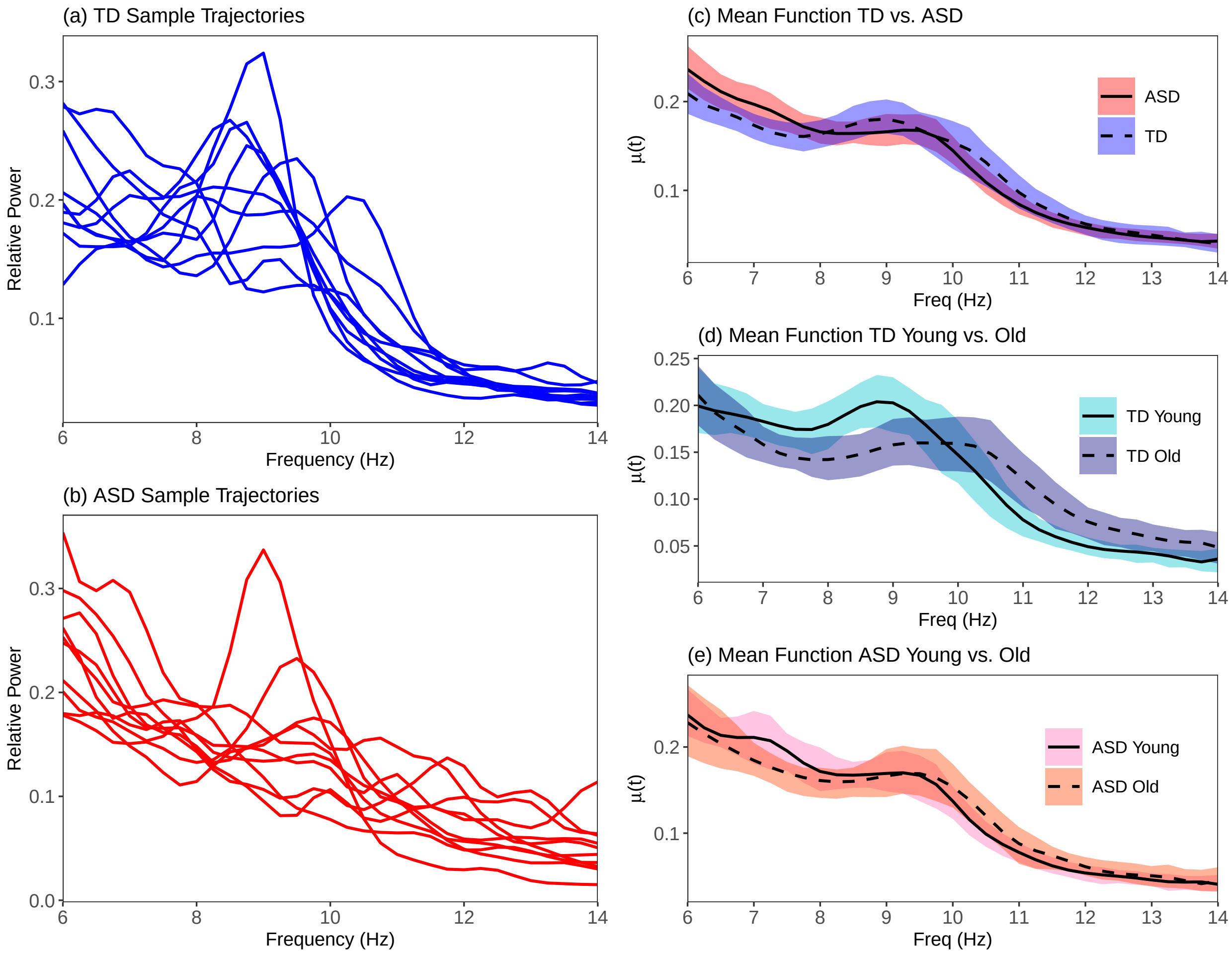


Figure 2.5: The left-hand columns display a sample of relative PSD obtained on 10 subjects from the TD (a) and ASD (b) diagnostic groups. The right-hand columns display the estimated MBD medians (solid or dashed black lines) and 95% MBD-CPEs (colored area) for the mean function in the TD and ASD groups (c), TD young and old groups (d), and ASD young and old groups (e).

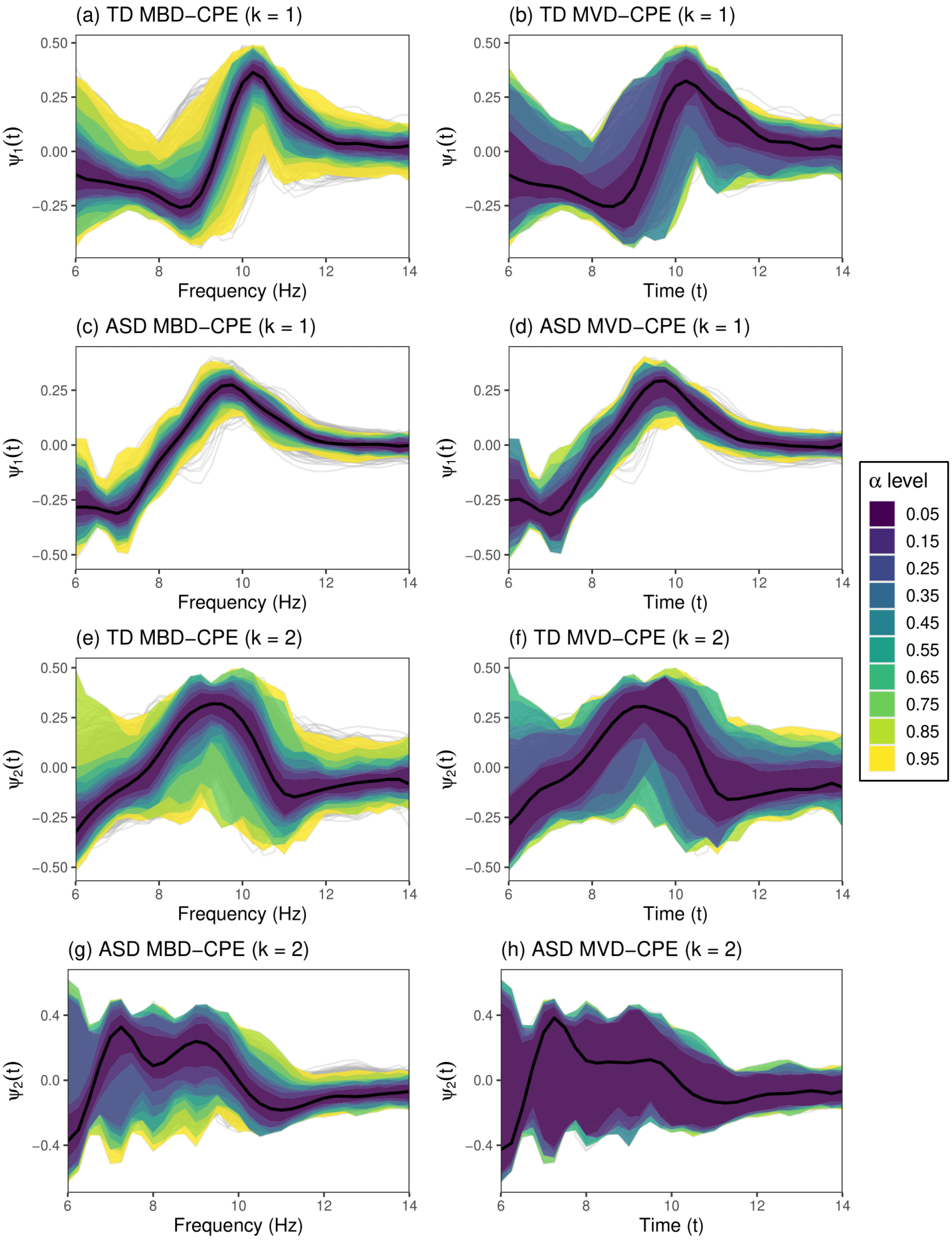


Figure 2.6: CPE contours of the two leading eigenfunctions for both ASD and TD groups in our data application, overlaying the posterior estimates given in gray. The left and right hand columns display the MBD and MVD-CPEs, denoted by $D_{1-\alpha}\{\psi_1(t)\}$ and $D^{\star}_{1-\alpha}\{\psi_1(t)\}$, respectively, at a grid of α levels marked by varying contour colors. The estimated MBD and MVD median are given in solid black in the right and left columns, respectively.

CHAPTER 3

Central posterior envelopes for Bayesian longitudinal functional principal component analysis

Abstract

Longitudinal functional data emerge in situations where functional observations for a subject are repeatedly measured over a set of time points. Bayesian longitudinal functional principal component analysis (B-LFPCA) decomposes this data into highly interpretable low-dimensional features under the assumption weak separability of the longitudinal and functional dimensions. Weak separability implies that the direction of variation along one of the dimensions stays constant across fixed slices of the other dimension, and vice versa, motivating the construction of the marginal longitudinal and functional covariances. The tensor product of the resulting marginal directions of variation along the longitudinal and functional dimensions provides a factorization of the signal into its longitudinal and functional components. The Bayesian modeling approach leads to readily available inference and data exploration via the posterior samples, and we propose central posterior envelopes (CPEs) to capture uncertainty in the low-dimensional features and mean function. CPEs are fully data-driven descriptive visualization tool that display the most-central regions of a posterior sample at specified α-level percentile contours obtained via functional depth. Functional depth is a measure that provides a center-outward ordering of a sample of curves and extends order statistics to functional data. The efficacy of the proposed CPEs under violations of weak separability and in the presence of outliers is demonstrated in extensive simulations. The B-LFPCA model and proposed CPEs are further showcased in an analysis

of longitudinally observed event-related potentials (ERPs) during an implicit learning experiment among children with autism spectrum disorder (ASD) and typically developed (TD) controls. Application of the CPEs leads to novel insights in diagnostic group differences not observed using traditional methods of uncertainty quantification in the Bayesian framework.

3.1 Introduction

Multidimensional functional data in the form of longitudinal functional data consist of high-dimensional functional observations such as curves, surfaces, or images, repeatedly measured over a set of longitudinally observed time points. A typical set of longitudinal functional data for subjects, $i = 1, \ldots, n$, is comprised of observations $Y_i(s, t)$ obtained at longitudinal time s (e.g. hospital visit), where each time point resulted in a functional datum with argument t (e.g. electrocardiogram measurement). Due to the close proximity of longitudinal time points, the correlation structure between functional observations for a given subject i is typically non-trivial and may be an important focus for a study. In electroencephalography (EEG) experiments, spontaneous electrical activity is recorded at multiple electrodes placed across the scalp and results in rich multidimensional functional data. Our motivating implicit learning study in children with autism spectrum disorder (ASD) (Jeste et al. (2015)) involved measuring EEG in response to visual stimulus, resulting in an event-related potential (ERP), over multiple trials in which the visual stimuli were repeatedly presented. Therefore, our study resulted in functional ERP data collected longitudinally over trials, and since learning was expected to occur over the duration of the experiment, appropriate modeling of the longitudinal trends was of particular interest.

Over the past two decades, an abundance of literature on functional data analysis (FDA) has been published (Ramsay and Silverman (2005)), and both frequentist and Bayesian approaches for modeling longitudinal functional data are a well-developed area of FDA. These methods often rely on dimension reduction techniques for capturing variability in a low-dimensional and easily interpretable form. A common tool employed for this is functional principal component analysis (FPCA) which decomposes the covariance structure of a signal into its primary directions of variations, commonly referred to as eigenfunctions. Previous work has utilized multidimensional or multilevel FPCA often in combination with mixed effects modeling to capture hierarchical dependencies (Chen and Müller (2012); Di et al. (2009); Wang et al. (2016); Greven et al. (2010); Yao et al. (2012); Suarez and Ghosal (2017); Montagna et al. (2012)) and has been developed for brain imaging studies such as our motivating study described previously (Hasenstab et al. (2017); Campos et al. (2022); Scheffler et al. (2020); Shamshoian et al. (2022)). Important generalizations have been made in multidimensional FPCA via the assumption of weak separability, in which one assumes that the leading directions of variation of the longitudinal (functional) dimension stay constant across fixed slices of the functional (longitudinal) dimension. This property motivates the construction of the marginal longitudinal and functional covariances, obtained by integrating the full covariance kernel of both dimensions over the functional and longitudinal dimensions, respectively (Chen et al. (2017); Park and Staicu (2015); Lynch and Chen (2018)). Projection of the mean-centered longitudinal functional data onto the tensor product of the resulting marginal longitudinal and functional directions of variation leads to product scores that are uncorrelated over the longitudinal and functional dimensions. Thus, an appealing decomposition of the multidimensional signal into the product scores and highly interpretable one-dimensional directions of variation along the longitudinal and functional dimensions (separately) is achieved under the assumption of weak separability. In this chapter, Bayesian longitudinal functional principal component analysis (B-LFPCA) under the assumption of weak separability is implemented following the data-adaptive Bayesian modeling approach for longitudinal functional data that was proposed in Shamshoian et al. (2022). An advantage of using a Bayesian modeling framework is uncertainty quantification via summaries of posterior samples obtained in estimation, which avoids the need for bootstrapping.

In this chapter, we propose the use of central posterior envelopes (CPEs) to capture uncertainty in the posterior samples of the marginal longitudinal and functional eigenfunctions and mean function estimated using the B-LFPCA model. A traditional method for uncertainty quantification utilized in Shamshoian et al. (2022) is the $(1-\alpha)100\%$ simultaneous parametric credible interval, which is determined via pointwise estimates of the mean and standard deviation of a posterior sample (Crainiceanu et al. (2007); Baladandayuthapani et al. (2005)). Although straightforward, simultaneous parametric credible intervals have disadvantages: they rely on the distributional assumption of normality, and the bounds of the credible intervals formed are symmetric around the mean and calculated via pointwise estimates, requiring discretization of the functional observations. The proposed CPEs provide a flexible and alternative way to capture uncertainty in a posterior sample that avoids these disadvantages. In particular, CPEs delineated by subsets of a posterior sample at specified α-level contours are employed as a descriptive visualization tool to capture variation in the posterior sample. These subsets are obtained via the use of functional depth, which provides a center-outward ordering of a sample of functions, extending order statistics and percentiles to functional data. Functional depth has been utilized in a wide variety of other applications including visual aids as well as point estimates in FDA. Some examples include the construction of a median or trimmed mean function, functional boxplots (Sun and Genton (2012)), surface boxplots (Genton et al. (2014)), outlier detection via the outliergram (Arribas-Gil and Romo (2014)), robust rank, and permutation and location tests for distributional and dispersion differences in two-sample functional data groups (López-Pintado and Wrobel (2017); López-Pintado and Qian (2020)). López-Pintado and Romo (2009) introduced the notion of modified band depth (MBD), extending the definition of band depth based on a graph-based approach, by measuring the proportion of time that a curve lies in the band delimited by a subsample of curves. Sun et al. (2012) derived a computationally efficient algorithm for calculating MBD that can rank millions of curves in seconds. Genton et al. (2014) further extended MBD to higher dimensional functional data through modified volume depth (MVD). Recently, the use of depth has been explored for non-Euclidean object data in Dai et al. (2022) and random objects residing in finite- and infinite-dimensional manifolds in Dubey et al. (2022). We utilize both MBD and MVD to obtain point estimates and functional depth based CPEs for the marginal longitudinal and functional eigenfunctions and mean function obtained from the B-LFPCA model. By extending MVD to order posterior samples of the marginal covariances and covariance kernel, we are able to obtain point estimates of these model components and marginal eigenfunctions as well as CPEs of the latter that capture increasing levels of variation in the model. It is important to note that CPEs are not tools for inference but are fully data-driven visual aids for capturing uncertainty in a sample of functional posterior estimates.

The chapter is organized as follows. The B-LFPCA model considered is introduced in Section 3.2 based on the Bayesian multidimensional functional model that was shown to exhibit computational feasibility, data-adaptive behavior, and good operating characteristics proposed in Shamshoian et al. (2022). B-LFPCA utilizes a random tensor product basis expansion of the longitudinal functional data followed by modeling the basis coefficients from the expansion with a latent factor model. Through the use of multiplicative gamma process shrinkage priors (Bhattacharya and Dunson (2011); Montagna et al. (2012)) placed on the factor loadings, posterior estimation of the low-dimensional features and mean function is achieved. This is followed by an outline of the traditional posterior summaries for the B-LFPCA model components. The proposed CPEs and point estimates based on functional depth are outlined in Section 3.3, followed by simulation studies to highlight their finite sample performance under violations of key model assumptions and in the presence of different types of functional outliers (Section 3.4). Section 3.5 details an application of CPEs to the analysis of ERPs captured during our motivating visual implicit learning experiment. Further, the credible interval method employed in Shamshoian et al. (2022) for uncertainty quantification is compared to the proposed CPE method in our analysis. Finally, a brief discussion summarizing key findings and future avenues of research is included in Section 3.6.

3.2 Bayesian longitudinal functional principal component analysis

3.2.1 Model specification

Let $Y_i(s,t) = f_i(s,t) + \epsilon_i(s,t)$ denote the observed noisy response for subject i, $i = 1, \ldots, n$, at longitudinal time $s \in \mathcal{S}$ and functional time $t \in \mathcal{T}$, where $\mathcal{S}$ and $\mathcal{T}$ are compact subspaces in $\mathbb{R}$, represented as a sum of a smooth underlying function $f_i(s,t)$, and measurement error $\epsilon_i(s,t)$. The measurement error $\epsilon_i(s,t)$ is assumed to be i.i.d. with mean zero and variance σ^2_ϵ. The smooth function $f_i(s,t)$ is assumed to exist in the space of square integrable functions $L^2(\mathcal{S} \times \mathcal{T})$ with mean $\mu(s,t)$ and covariance kernel $K\{(s,t),(s',t')\} = \text{Cov}\{f_i(s,t), f_i(s',t')\}$. The multidimensional Karhunen-Loève (KL) expansion of $f_i(s,t)$ detailed in Ramsay and Silverman (2005) is defined as

$$f_i(s,t) = \mu(s,t) + \sum_{h=1}^{\infty} Z_{ih} \varsigma_h(s,t), \tag{3.1}$$

where $Z_{ih} = \int_{\mathcal{S}} \int_{\mathcal{T}} \{f_i(s,t) - \mu(s,t)\} \varsigma_h(s,t) dsdt$ denotes the hth uncorrelated subject-specific score with mean zero and variance ρ_h, and $\varsigma_h(s,t)$ are the orthonormal eigenfunctions of $K\{(s,t),(s',t')\}$ that form a basis of $L^2(\mathcal{S} \times \mathcal{T})$. The first H terms in expansion (3.1) form a H-dimensional approximation of $f_i(s,t)$ with the optimality property that when compared to any other H-dimensional orthonormal basis, reconstructing $f_i(s,t)$ using the eigenfunctions results in an approximation of $f_i(s,t)$ that explains the most variation. Due to the complexities involved in modeling the four-dimensional covariance kernel $K\{(s,t),(s',t')\}$ and characterizing the two-dimensional eigenfunctions $\varsigma_h(s,t)$. Chen et al. (2017) proposed product functional principal component analysis (FPCA) which achieves dimension reduction through factorization of $f_i(s,t)$ into its longitudinal and functional components under the assumption of weak separability of the longitudinal and functional dimensions. Weak separability implies that the direction of variation in the functional dimension of the data stays the same for fixed slices along the longitudinal dimension, and vice versa. This assumption is weaker than the commonly assumed strong separability, which implies constant covariance along one dimension for fixed values of the other dimensions of the data. Due to this property, weak separability motivates the concept of the marginal longitudinal and functional covariance, which are obtained by integrating the covariance kernel over the functional and longitudinal dimensions, respectively. Let the marginal longitudinal and functional covariance functions of $f_i(s,t)$, denoted $K_{\mathcal{S}}(s,s')$ and $K_{\mathcal{T}}(t,t')$, respectively, be defined as

$$K_{\mathcal{S}}(s,s') = \int_{\mathcal{T}} K\{(s,t),(s',t)\}dt = \sum_{j=1}^{\infty} \tau_j \psi_j(s)\psi_j(s'),$$

and

$$K_{\mathcal{T}}(t,t') = \int_{\mathcal{S}} K\{(s,t),(s,t')\}ds = \sum_{k=1}^{\infty} \vartheta_k \phi_k(t)\phi_k(t'),$$

where $\tau_1 \geqslant \tau_2 \geqslant \ldots$ denote the ordered nonnegative longitudinal eigenvalues corresponding to longitudinal eigenfunctions $\psi_1(t), \psi_2(t), \ldots$ of $K_{\mathcal{S}}(s,s')$ that form an orthonormal basis in $L^2(\mathcal{S})$, and $\vartheta_1 \geqslant \vartheta_2 \geqslant \ldots$ denote the ordered nonnegative functional eigenvalues corresponding to functional eigenfunctions $\phi_1(t), \phi_2(t), \ldots$ of $K_{\mathcal{T}}(t,t')$ that form an orthonormal basis in $L^2(\mathcal{T})$. Then, let $\chi_{ijk} = \int_{\mathcal{S}}\int_{\mathcal{T}}\{f_i(s,t) - \mu(s,t)\}\psi_j(s)\phi_k(t)dsdt$ be the subject-specific product score from the projection of the mean-centered smooth underlying function onto the tensor product of the jth longitudinal and kth functional eigenfunctions, denoted $\psi_j(s) \otimes \phi_k(t)$, with mean zero and variance $\mathrm{Var}(\chi_{ijk}) = v_{jk}$. Under the assumption of weak separability, the product scores are uncorrelated over the longitudinal and functional dimensions, or more specifically $\mathrm{Cov}(\chi_{ijk}, \chi_{ij'k'}) = 0$ when $j \neq j'$ or $k \neq k'$. The product KL expansion of $f_i(s,t)$ is then given by

$$f_i(s,t) = \mu(s,t) + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \chi_{ijk}\psi_j(s)\phi_k(t). \tag{3.2}$$

Further under weak separability, the longitudinal and functional eigenvalues are given as $\tau_j = \sum_{k=1}^{\infty} v_{jk}$ and $\vartheta_k = \sum_{j=1}^{\infty} v_{jk}$, respectively. Properties of weak separability were further explored in Lynch and Chen (2018), which showed that if weak separability holds the covariance kernel $K\{(s,t),(s',t')\}$ can be written as the tensor product of the longitudinal and functional eigenfunctions $\psi_j(s) \otimes \phi_k(t)$ as

$$K\{(s,t),(s',t')\} = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} \psi_j(s) \psi_j(s') \phi_k(t) \phi_k(t'). \tag{3.3}$$

In practice, the product KL expansion is truncated to include J and K longitudinal and functional eigencomponents, $f_i(s,t) \approx \mu(s,t) + \sum_{j=1}^{J} \sum_{k=1}^{K} \chi_{ijk} \psi_j(s) \phi_k(t)$, based on fraction of variance explained (FVE) of the respective longitudinal and functional marginal covariances. The covariance kernel is then approximated by $K\{(s,t),(s',t')\} \approx \sum_{j=1}^{J} \sum_{k=1}^{K} v_{jk} \psi_j(s) \psi_j(s') \phi_k(t) \phi_k(t')$, and the marginal covariances are approximated by $K_{\mathcal{S}}(s,s') \approx \sum_{j=1}^{J} \tau_j \psi_j(s) \psi_j(s')$ and $K_{\mathcal{T}}(t,t') \approx \sum_{k=1}^{K} \vartheta_k \phi_k(t) \phi_k(t')$. A total of $J = 2$ and $K = 2$ or 3 longitudinal and functional eigencomponents are typically retained in most applications (explaining more than 80% of the total variation in the longitudinal and functional marginal covariances separately), thus achieving effective dimension reduction with the smallest amount of unexplained variation.

The Bayesian longitudinal functional principal component analysis (B-LFPCA) model considered is adapted from the probability model for longitudinal functional data given in Shamshoian et al. (2022). The B-LFPCA approach starts with the expansion of the smooth underlying function $f_i(s,t)$ on a set of p_1 B-spline basis functions $\{b_1^{(1)}(s), \ldots, b_{p_1}^{(1)}(s)\}$ and p_2 B-spline basis functions $\{b_1^{(2)}(t), \ldots, b_{p_2}^{(2)}(t)\}$ in the longitudinal and functional dimensions, respectively. Using a random tensor product, the expansion is constructed as $f_i(s,t) = \sum_{\ell=1}^{p_1} \sum_{m=1}^{p_2} \theta_{i\ell m} b_\ell^{(1)}(s) b_m^{(2)}(t)$, where $\theta_{i\ell m}$ denote the subject-specific random basis coefficients, and p_1 and p_2 are chosen to be large enough to ensure adequate smoothing of $f_i(s,t)$. Next for $\ell = 1, \ldots, p_1$, $m = 1, \ldots, p_2$, a latent factor model is assumed for the subject-specific random basis coefficients as $\theta_{i\ell m} = \sum_{j=1}^{q_1} \sum_{k=1}^{q_2} \eta_{ijk} \gamma_{\ell j} \lambda_{mk} + \zeta_{i\ell m}$, where η_{ijk}, $j = 1, \ldots, q_1, (q_1 < p_1)$, $k = 1, \ldots, q_2, (q_2 < p_2)$, denote the Gaussian-distributed subject-specific latent scores corresponding to $\gamma_{\ell j}$, the jth longitudinal factor loading for $b_\ell^{(1)}(s)$, and λ_{mk}, the kth functional factor loading for $b_m^{(2)}(t)$, and $\zeta_{i\ell m} \sim N(0, \sigma^2_{\zeta_{\ell m}})$ denote the uncorrelated subject-specific reconstruction errors. This leads to the following expansion of

$$\begin{aligned} f_i(s,t) &= \sum_{\ell=1}^{p_1} \sum_{m=1}^{p_2} \theta_{i\ell m} b_\ell^{(1)}(s) b_m^{(2)}(t), \\ &= \sum_{\ell=1}^{p_1} \sum_{m=1}^{p_2} \left(\sum_{j=1}^{q_1} \sum_{k=1}^{q_2} \eta_{ijk} \gamma_{\ell j} \lambda_{mk} + \zeta_{i\ell m} \right) b_\ell^{(1)}(s) b_m^{(2)}(t), && (3.4) \\ &= \sum_{j=1}^{q_1} \sum_{k=1}^{q_2} \eta_{ijk} \ddot{\psi}_j(s) \ddot{\phi}_k(t) + r_i(s,t), && (3.5) \end{aligned}$$

where $\ddot{\psi}_j(s) = \sum_{\ell=1}^{p_1} \gamma_{\ell j} b_\ell^{(1)}(s)$, $j = 1, \ldots, q_1$, denote the data-adaptive longitudinal basis functions, $\ddot{\phi}_k(t) = \sum_{m=1}^{p_2} \lambda_{mk} b_m^{(2)}(t)$, $k = 1, \ldots, q_2$, denote the data-adaptive functional basis functions, and $r_i(s,t) = \sum_{\ell=1}^{p_1} \sum_{m=1}^{p_2} b_\ell^{(1)}(s) b_m^{(2)}(t) \zeta_{i\ell m}$ denotes the uncorrelated subject-specific reconstruction error. Therefore, the expansion (3.5) models $f_i(s,t)$ as a random tensor product of q_1 and q_2 data-adaptive basis functions corresponding to Gaussian-distribution latent scores η_{ijk}, and any residual variability in the expansion given q_1 and q_2 is captured in $r_i(s,t)$. It is further assumed that $\eta_{ijk} \sim N(\beta_{jk}, \sigma^2_{\eta_{jk}})$, where β_{jk} is the mean coefficient and $\sigma^2_{\eta_{jk}}$ variance of is the variance of the latent scores for $j = 1, \ldots, q_1$, $k = 1, \ldots, q_2$. Marginalizing over $\theta_{i\ell m}$, the mean function is determined as

$$\mu(s,t) = \sum_{j=1}^{q_1} \sum_{k=1}^{q_2} \beta_{jk} \ddot{\psi}_j(s) \ddot{\phi}_k(t).$$

Structural assumptions of the covariance process in the B-LFPCA model are formed based on the covariance of the latent scores. In particular, weak separability of the longitudinal and functional dimensions in this expansion is induced when it is assumed that $\mathrm{Cov}(\eta_{ijk}, \eta_{ij'k'}) =$

0 when $j \neq j'$ or $k \neq k'$. Thus, the covariance kernel is determined as

$$K\{(s,t),(s',t')\} = \sum_{j=1}^{q_1}\sum_{k=1}^{q_2} \sigma^2_{\eta_{jk}} \ddot{\psi}_j(s)\ddot{\psi}_j(s')\ddot{\phi}_k(t)\ddot{\phi}_k(t') + \sum_{\ell=1}^{p_1}\sum_{m=1}^{p_2} \sigma^2_{\zeta_{\ell m}} b^{(1)}_\ell(s) b^{(1)}_\ell(s') b^{(2)}_m(t) b^{(2)}_m(t').$$

The first term in the above equation is analogous to the covariance kernel under the assumption weak separability of the product FPCA model given in (3.3) with additional variation added along the diagonal in the second term that captures residual variation not captured by the q_1 and q_2 data-adaptive basis functions, constituting the B-LFPCA model along with the expansion (3.5). Similarly, if $\ddot{\psi}_j(s)$ and $\ddot{\phi}_k(t)$ were chosen to be the eigenfunctions of the marginal longitudinal and function covariances $\psi_j(s)$ and $\phi_k(t)$, respectively, the expansion (3.5) is reminiscent of the KL product expansion (3.2) when truncated to J and K eigencomponents. Therefore, the reconstruction error $r_i(s,t)$ would capture any residual variation not explained by the $J \times K$ tensor products of the longitudinal and functional eigenfunctions, while the latent scores η_{ijk} would model the mean function and covariance process under the structural assumption of weak separability.

A mixed effects model for posterior estimation can be formulated by representing the latent factor model given in (3.4) in vector form. Let $\Gamma = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p_1})^\top$ denote the $p_1 \times q_1$ longitudinal factor loading matrix with $\boldsymbol{\gamma}_\ell = (\gamma_{\ell 1}, \ldots, \gamma_{\ell q_1})^\top$, and $\Lambda = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{p_2})^\top$ denote the $p_2 \times q_2$ functional factor loading matrix with $\boldsymbol{\lambda}_m = (\lambda_{m1}, \ldots, \lambda_{mq_2})^\top$. Further, if η_i denotes the $q_1 \times q_2$ subject-specific latent score matrix where η_{ijk} is the element in the jth row and kth column of η_i, then $\boldsymbol{\eta}_i = \text{vec}(\eta_i)$ is the $q_1 q_2 \times 1$ vector of the subject-specific latent scores obtained from the vec$(\cdot)$ operator, which stacks the columns of a given matrix. Similarly, let ζ_i denote the $p_1 \times p_2$ subject-specific reconstruction error matrix where $\zeta_{i\ell m}$ is the element in the ℓth row and mth column of ζ_i, and $\boldsymbol{\zeta}_i = \text{vec}(\zeta_i)$ be the resulting $p_1 p_2 \times 1$ vector. If $\otimes$ represents the Kronecker product, the full mixed effects model can be written in vector form as

$$\begin{aligned}
\boldsymbol{Y}_i &= \boldsymbol{f}_i + \boldsymbol{\epsilon}_i = (B_1 \otimes B_2)\left\{(\Lambda \otimes \Gamma)\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i\right\} + \boldsymbol{\epsilon}_i, \qquad (3.6)\\
\boldsymbol{\eta}_i &\sim N_{q_1q_2}(\boldsymbol{\beta}, \Sigma_\eta), \quad \boldsymbol{\zeta}_i \sim N_{p_1p_2}(\mathbf{0}_{p_1p_2}, \Sigma_\zeta),\\
\boldsymbol{\epsilon}_i &\sim N_{n_sn_t}(\mathbf{0}_{n_sn_t}, \sigma^2_\epsilon I_{n_sn_t}), \quad i = 1, \dots, n,
\end{aligned}$$

where $\boldsymbol{Y}_i = \{Y_i(s_1,t_1), \dots, Y_i(s_1,t_{n_t}), \dots, Y_i(s_{n_s},t_1), \dots, Y_i(s_{n_s},t_{n_t})\}^\top$ denotes the response observed at a total of n_s longitudinal and n_t functional time points, $\boldsymbol{f}_i = \{f_i(s_1,t_1), \dots, f_i(s_1,t_{n_t}), \dots, f_i(s_{n_s},t_1), \dots, f_i(s_{n_s},t_{n_t})\}^\top$ denotes the $n_sn_t \times 1$ vector of underlying smooth functions, and $\boldsymbol{\epsilon}_i = \{\epsilon_i(s_1,t_1), \dots, \epsilon_i(s_1,t_{n_t}), \dots, \epsilon_i(s_{n_s},t_1), \dots, \epsilon_i(s_{n_s},t_{n_t})\}^\top$ denotes the $n_sn_t \times 1$ vector of measurement error. Moreover, in (3.6), $B_1 = (\mathbf{b}^1_1, \dots, \mathbf{b}^1_{p_1})$ denotes the $n_s \times p_1$ matrix of B-spline longitudinal basis functions with $\mathbf{b}^1_\ell = \{b^{(1)}_\ell(s_1), \dots, b^{(1)}_\ell(s_{n_s})\}^\top$, $B_2 = (\mathbf{b}^2_1, \dots, \mathbf{b}^2_{p_2})$ denotes the $n_t \times p_2$ matrix of B-spline functional basis functions with $\mathbf{b}^2_m = \{b^{(2)}_m(t_1), \dots, b^{(2)}_m(t_{n_t})\}^\top$. The vector of subject-specific latent scores $\boldsymbol{\eta}_i$ is q_1q_2-variate normally distributed with $q_1q_2 \times 1$ vector of mean coefficients $\boldsymbol{\beta} = (\beta_{11}, \dots, \beta_{1q_2}, \dots, \beta_{q_11}, \dots, \beta_{q_1q_2})^\top$ and $q_1q_2 \times q_1q_2$ diagonal covariance $\Sigma_\eta = \mathrm{diag}(\sigma^2_{\eta_{11}}, \dots, \sigma^2_{\eta_{1q_2}}, \dots, \sigma^2_{\eta_{q_11}}, \dots, \sigma^2_{\eta_{q_1q_2}})$. The vector of subject-specific reconstruction errors $\boldsymbol{\zeta}_i$ is p_1p_2-variate normally distributed with mean $\mathbf{0}_{p_1p_2}$, denoting the $p_1p_2 \times 1$ vector of zeros, and $p_1p_2 \times p_1p_2$ diagonal covariance $\Sigma_\zeta = \mathrm{diag}(\sigma^2_{\zeta_{11}}, \dots, \sigma^2_{\zeta_{1p_2}}, \dots, \sigma^2_{\zeta_{p_11}}, \dots, \sigma^2_{\zeta_{p_1p_2}})$. Lastly, $\mathbf{0}_{n_sn_t}$ is used to denote the $n_sn_t \times 1$ vector of zeros, and $I_{n_sn_t}$ is used to denote the $n_sn_t \times n_sn_t$ identity matrix used in defining the n_sn_t-variate normal distribution of the measurement errors $\boldsymbol{\epsilon}_i$. After selection of an appropriate number of B-spline basis functions p_1 and p_2 as well as a sufficiently large enough number of latent factors $q_1 < p_1$ and $q_2 < p_2$, conditionally conjugate priors allow for efficient posterior estimation of through the use Markov chain Monte Carlo (MCMC). The priors distributions for the B-LFPCA model and subsequent posterior distributions and MCMC sampling algorithm are detailed in Appendix C.1.

3.2.2 Traditional posterior summaries for B-LFPCA components

Posterior estimates of longitudinal factor loadings $\gamma_{\ell j}^{(m)}$, $\ell = 1, \ldots, p_1$, $j = 1, \ldots, q_1$, where the superscript m, $m = 1, \ldots, M$, is used to index the posterior estimates obtained from the MCMC sampler after burn-in and thinning, lead to the posterior estimates of the longitudinal and functional data-adaptive basis functions $\ddot{\psi}_j^{(m)}(s) = \sum_{\ell=1}^{p_1} \gamma_{\ell j}^{(m)} b_\ell^{(1)}(s)$, $j = 1, \ldots, q_1$. Similarly, posterior estimates of functional factor loadings $\lambda_{mk}^{(m)}$, $m = 1, \ldots, p_2$, $k = 1, \ldots, q_2$, leads to the posterior estimates of the functional data-adaptive basis functions $\ddot{\phi}_k^{(m)}(t) = \sum_{m=1}^{p_2} \lambda_{mk}^{(m)} b_m^{(2)}(t)$, $k = 1, \ldots, q_2$. Then, the posterior estimates of the data-adaptive basis functions and the mean coefficient vector $\boldsymbol{\beta}^{(m)} = (\beta_{11}^{(m)}, \ldots, \beta_{q_1 q_2}^{(m)})^\top$ lead to the posterior estimates of the mean function as $\mu^{(m)}(s,t) = \sum_{j=1}^{q_1} \sum_{k=1}^{q_2} \beta_{jk}^{(m)} \ddot{\psi}_j^{(m)}(s) \ddot{\phi}_k^{(m)}(t)$. Further, the posterior estimates of the variance of the latent factors $\sigma_{\eta_{jk}}^{2(m)}$, $j = 1, \ldots, q_1$, $k = 1, \ldots, q_2$, and variance of the reconstruction errors $\sigma_{\zeta_{\ell m}}^{2(m)}$, $\ell = 1, \ldots, p_1$, $m = 1, \ldots, p_2$, are used to obtain the posterior estimates of the covariance kernel as $K^{(m)}\{(s,t),(s',t')\} = \sum_{j=1}^{q_1} \sum_{k=1}^{q_2} \sigma_{\eta_{jk}}^{2(m)} \ddot{\psi}_j^{(m)}(s) \ddot{\psi}_j^{(m)}(s') \ddot{\phi}_k^{(m)}(t) \ddot{\phi}_k^{(m)}(t') + \sum_{\ell=1}^{p_1} \sum_{m=1}^{p_2} \sigma_{\zeta_{\ell m}}^{2(m)} b_\ell^{(1)}(s) b_\ell^{(1)}(s') b_m^{(2)}(t) b_m^{(2)}(t')$. Direct computation of the posterior estimates of the longitudinal marginal covariance, which has better scalability than integrating over the posterior estimates of the covariance kernel, is employed as

$$K_{\mathcal{S}}^{(m)}(s,s') = \sum_{j=1}^{q_1} \ddot{\psi}_j^{(m)}(s) \ddot{\psi}_j^{(m)}(s') \sum_{k=1}^{q_2} \sigma_{\eta_{jk}}^{2(m)} \omega_k^{(m)} + \sum_{\ell=1}^{p_1} b_\ell^{(1)}(s) b_\ell^{(1)}(s') \sum_{m=1}^{p_2} \sigma_{\zeta_{\ell m}}^{2(m)} \Omega_m,$$

where $\omega_k^{(m)} = \int_{\mathcal{T}} \ddot{\phi}_k^{(m)}(t) \ddot{\phi}_k^{(m)}(t) dt$ and $\Omega_m = \int_{\mathcal{T}} b_m^{(2)}(t) b_m^{(2)}(t) dt$. Similarly, the posterior estimates of the functional marginal covariance are determined as

$$K_{\mathcal{T}}^{(m)}(t,t') = \sum_{k=1}^{q_2} \ddot{\phi}_k^{(m)}(t) \ddot{\phi}_k^{(m)}(t') \sum_{j=1}^{q_1} \sigma_{\eta_{jk}}^{2(m)} \omega_j^{(m)} + \sum_{m=1}^{p_2} b_m^{(2)}(t) b_m^{(2)}(t') \sum_{\ell=1}^{p_1} \sigma_{\zeta_{\ell m}}^{2(m)} \Omega_\ell,$$

where $\omega_j^{(m)} = \int_{\mathcal{S}} \ddot{\psi}_j^{(m)}(s) \ddot{\psi}_j^{(m)}(s) ds$ and $\Omega_\ell = \int_{\mathcal{S}} b_\ell^{(1)}(s) b_\ell^{(1)}(s) ds$. The proof for the formulations of the posterior estimates of the longitudinal and functional marginal covariance using direct computation over integration is provided in Appendix C.2. We consider the singular value decomposition (SVD) of the posterior estimates of the marginal covariances to recover the lower dimensional interpretations of the longitudinal and functional eigenfunctions provided in the product FPCA decomposition. The posterior estimates of the longitudinal components are obtained via SVD as $K_{\mathcal{S}}^{(m)}(s,s') \approx \sum_{j=1}^{J} \tau_j^{(m)} \psi_j^{(m)}(s)\psi_j^{(m)}(s')$, where $\psi_j^{(m)}(s)$ and $\tau_j^{(m)}$ are the posterior estimates of the longitudinal eigenfunctions and eigenvalues, respectively. Similarly, the SVD of $K_{\mathcal{T}}^{(m)}(t,t') \approx \sum_{k=1}^{K} \vartheta_k^{(m)} \phi_k^{(m)}(t)\phi_k^{(m)}(t')$, targets the posterior estimates of the functional eigenfunctions, $\phi_k^{(m)}(t)$, and functional eigenvalues, $\vartheta_k^{(m)}$. The SVD of the posterior estimates of the discretized marginal covariances are obtained, and the resulting posterior estimates of the eigencomponents are appropriately standardized (Capra and Müller (1997)). The mean or median FVE calculated across the M posterior samples is used to choose the total number of eigencomponents retained in the longitudinal and functional dimensions, J and K, respectively. Lastly, an additional alignment step is implemented to obtain the signs of the longitudinal and functional eigenfunctions and detailed in Appendix C.2.

Following estimation, the traditional point estimates for the B-LFPCA components are obtained through averaging of the posterior estimates. The mean estimate is obtained as $\widehat{\mu}(s,t) = (1/M)\sum_{m=1}^{M} \mu^{(m)}(s,t)$. The longitudinal and functional marginal covariance estimates are obtained as $\widetilde{K}_{\mathcal{S}}(s,s') = (1/M)\sum_{m=1}^{M} K_{\mathcal{S}}^{(m)}(s,s')$ and $\widetilde{K}_{\mathcal{T}}(t,t') = (1/M)\sum_{m=1}^{M} K_{\mathcal{T}}^{(m)}(t,t')$, respectively. Then the covariance kernel estimate is targeted as $\overline{K}\{(s,t),(s',t')\} = (1/M)\sum_{m=1}^{M} K^{(m)}\{(s,t),(s',t')\}$. The longitudinal and functional eigenfunctions and eigenvalues can be targeted in two ways. The first approach is to average the posterior estimates, $\widehat{\psi}_j(s) = (1/M)\sum_{m=1}^{M} \psi_j^{(m)}(s)$, $\widehat{\tau}_j = (1/M)\sum_{m=1}^{M} \tau_j^{(m)}$, $j = 1,\ldots,J$, and $\widehat{\phi}_k(t) = (1/M)\sum_{m=1}^{M} \phi_k^{(m)}(t)$, $\widehat{\vartheta}_k = (1/M)\sum_{m=1}^{M} \vartheta_k^{(m)}$, $k = 1,\ldots,K$. The second approach is to perform SVD on the longitudinal marginal covariance estimate as $\widetilde{K}_{\mathcal{S}}(s,s') \approx \sum_{j=1}^{J} \widetilde{\tau}_j \widetilde{\psi}_j(s)\widetilde{\psi}_j(s')$ followed by appropriate standardization of the eigencomponent estimates leading to the longitudinal eigenfunction and eigenvalue estimates obtained via marginal covariance estimation, denoted by $\widetilde{\psi}_j(s)$ and $\widetilde{\tau}_j$, respectively. Similarly, the functional eigenfunction and eigenvalue estimates obtained via marginal covariance estimation, denoted by $\widetilde{\phi}_k(t)$ and $\widetilde{\vartheta}_k$, respectively, are calculated as $\widetilde{K}_{\mathcal{T}}(t, t') \approx \sum_{k=1}^{K} \widetilde{\vartheta}_k \widetilde{\phi}_k(t) \widetilde{\phi}_k(t')$ followed by appropriate standardization. The finite sample performance of the point estimates is evaluated via simulations in Section 3.4.

Traditional Bayesian analysis relies on $(1-\alpha)100\%$ credible intervals as a posterior summary of spread in the posterior distribution. In the functional data setting in Shamshoian et al. (2022), simultaneous parametric credible intervals under the assumption of normality were employed following the methodology detailed in Crainiceanu et al. (2007) and Baladandayuthapani et al. (2005). These credible intervals are formed using symmetric bounds from estimates of the pointwise mean and variance obtained from a functional posterior sample. Pitfalls arise when using simultaneous parametric intervals due to the symmetric bounds that may be too restrictive in cases of highly skewed posterior samples as well as distributional assumptions that may not hold under violations of modeling assumptions. Another method for obtaining credible intervals is detailed in Krivobokova et al. (2010) and relies on pointwise estimates of the sample quantiles to form the bounds. In this chapter, the use of functional depth-based central posterior envelopes (CPEs) are proposed for the B-LFPCA functional model components and further detailed in Section 3.3. Rather than forming a single $(1-\alpha)100\%$ credible interval, CPEs are a descriptive visual approach for capturing the spread of a functional posterior distribution through the use of multiple CPE contours plotted for a grid of α values. Due to the properties of functional depth, this results in a visualization of the most central regions of the posterior distribution and provides a flexible nonparametric tool for capturing variation in a posterior sample. The notations used for the traditional and depth based posterior summaries outlined above and in Section 3.3 are summarized for the reader's reference in Table 3.1.

Table 3.1: Notation used for point estimates of the B-LFPCA components (mean function $\mu(s,t)$, longitudinal eigenfunction $\psi_j(s)$, functional eigenfunction $\phi_k(t)$, longitudinal eigenvalue τ_j, functional eigenvalue ϑ_k, longitudinal marginal covariance $K_{\mathcal{S}}(s,s')$, functional marginal covariance $K_{\mathcal{T}}(t,t')$, and covariance kernel $K\{(s,t),(s',t')\}$) and the proposed CPEs for $\mu(s,t)$, $\psi_j(s)$, and $\phi_k(t)$.

	Traditional Posterior Summaries		Proposed Posterior Summaries			
B-LFPCA Component	Point Estimate		Point Estimate		Central Posterior Envelopes (CPEs)	
$\mu(s,t)$	$\widehat{\mu}(s,t)$	Mean estimate	$\widetilde{m}\{\mu(s,t)\}$	MVD median	$D_1^{\star} - \alpha\{\mu(s,t)\}$	MVD-CPE
$\psi_j(s)$	$\widehat{\psi}_j(s)$	Longitudinal eigenfunction estimate	$\widehat{m}\{\psi_j(s)\}$	MBD median	$D_1 - \alpha\{\psi_j(s)\}$	MBD-CPE
			$\widetilde{m}\{\psi_j(s)\}$	MVD median	$D_1^{\star} - \alpha\{\psi_j(s)\}$	MVD-CPE
	$\widetilde{\psi}_j(s)$	Longitudinal eigenfunction estimate via marginal covariance estimation	$\overline{m}\{\psi_j(s)\}$	Kernel MVD median	$D_1^{\dagger} - \alpha\{\psi_j(s)\}$	Kernel MVD-CPE
τ_j	$\widehat{\tau}_j$	Longitudinal eigenvalue estimate				
	$\widetilde{\tau}_j$	Longitudinal eigenvalue estimate via marginal covariance estimation				
$\phi_k(t)$	$\widehat{\phi}_k(t)$	Functional eigenfunction estimate	$\widehat{m}\{\phi_k(t)\}$	MBD median	$D_1 - \alpha\{\phi_k(t)\}$	MBD-CPE
			$\widetilde{m}\{\phi_k(t)\}$	MVD median	$D_{1-\alpha}^{\star}\{\phi_k(t)\}$	MVD-CPE
	$\widetilde{\phi}_k(t)$	Functional eigenfunction estimate via marginal covariance estimation	$\overline{m}\{\phi_k(t)\}$	Kernel MVD median	$D_{1-\alpha}^{\dagger}\{\phi_k(t)\}$	Kernel MVD-CPE
ϑ_k	$\widehat{\vartheta}_k$	Functional eigenvalue estimate				
	$\widetilde{\vartheta}_k$	Functional eigenvalue estimate via marginal covariance estimation				
$K_{\mathcal{S}}(s,s')$	$\widetilde{K}_{\mathcal{S}}(s,s')$	Longitudinal marginal covariance estimate	$\widetilde{m}\{K_{\mathcal{S}}(s,s')\}$	MVD median		
			$\overline{m}\{K_{\mathcal{S}}(s,s')\}$	Kernel MVD median		
$K_{\mathcal{T}}(t,t')$	$\widetilde{K}_{\mathcal{T}}(t,t')$	Functional marginal covariance estimate	$\widetilde{m}\{K_{\mathcal{T}}(t,t')\}$	MVD median		
			$\overline{m}\{K_{\mathcal{T}}(t,t')\}$	Kernel MVD median		
$K\{(s,t),(s',t')\}$	$\overline{K}\{(s,t),(s',t')\}$	Covariance kernel estimate	$\overline{m}[K\{(s,t),(s',t')\}]$	Kernel MVD median		

3.3 Proposed functional depth based CPEs for B-LFPCA components

Functional depth orders a sample of functions from the center-outward such that the 'deepest' curve, with the highest functional depth value, is defined as the median curve to the most outlying curve, with the lowest functional depth value. There are multiple functional depth measures that have been detailed in the current literature (see Zuo and Serfling (2000) and Gijbels and Nagy (2017)), but we center our focus on modified band depth, a popular graph-based measure of functional depth introduced in López-Pintado and Romo (2009). Let $g^{(1)}(u), \ldots, g^{(M)}(u)$, $u \in \mathcal{I}$, denote a sample of M posterior curves defined on a compact interval $\mathcal{I} \in \mathbb{R}$. A band in $\mathbb{R}^2$ delineated by r, $2 \leqslant r \leqslant M$, posterior estimates, $g^{(m_1)}(u), \ldots, g^{(m_r)}(u)$, constituting a subset of r curves from the full posterior sample $\{g^{(1)}(u), \ldots, g^{(M)}(u)\}$ is defined as

$$\mathcal{B}\left\{g^{(m_1)}(u), \ldots, g^{(m_r)}(u)\right\} = \left[\{u, g(u)\} : u \in \mathcal{I}, \min_{w=m_1,\ldots,m_u} g^{(w)}(u) \leqslant g(u) \leqslant \max_{w=m_1,\ldots,m_u} g^{(w)}(u)\right].$$

Band depth of López-Pintado and Romo (2009) considers the proportion of bands $\mathcal{B}\{g^{(m_1)}(u), \ldots, g^{(m_r)}(u)\}$ determined by r different curves $g^{(m_1)}(u), \ldots, g^{(m_r)}(u)$ containing the graph of $g(u)$. Rather than the proportion of bands that contain the entire graph of $g(u)$, modified band depth (MBD) extends band depth to consider the proportion of time that the graph of $g(u)$ lies inside the bands. More specifically, let the set in the interval $\mathcal{I}$ where the function $g^{(m)}(u)$ lies inside the band $\mathcal{B}\{g^{(m_1)}(u), \ldots, g^{(m_r)}(u)\}$ be denoted as $\mathcal{A}_r\{g^{(m)}(u); g^{(m_1)}(u), \ldots, g^{(m_r)}(u)\} = \{u \in \mathcal{I} : \min_{w=m_1,\ldots,m_r} g^{(w)}(u) \leqslant g^{(m)}(u) \leqslant \max_{w=m_1,\ldots,m_u} g^{(w}(u)\}$. Further, if $\mathbb{L}(\cdot)$ denotes the Lebesgue measure on $\mathcal{I} \in \mathbb{R}$, then proportion of time that the curve $g^{(m)}(u)$ lies inside the band $\mathcal{B}\{g^{(m_1)}(u), \ldots, g^{(m_r)}(u)\}$ is defined as $\mathcal{A}_r^*\{g^{(m)}(u); g^{(m_1)}(u), \ldots, g^{(m_r)}(u)\} = \mathbb{L}[\mathcal{A}_r\{g^{(m)}(u); g^{(m_1)}(u), \ldots, g^{(m_r)}(u)\}]/\mathbb{L}(\mathcal{I})$. The modified band depth (MBD) for the curve $g^{(m)}(u)$ in $g^{(1)}(u), \ldots, g^{(M)}(u)$ given R, where $2 \leqslant R \leqslant M$ denotes a fixed total number of curves used to delineate a band, is

$$MBD_{M,R}\left\{g^{(m)}(u)\right\} = \sum_{r=2}^{R}\left[\binom{M}{2}^{-1}\sum_{1\leqslant m_1<m_2<\ldots<m_r\leqslant M}\mathcal{A}_r^*\left\{g^{(m)}(u);g^{(m_1)}(u),\ldots,g^{(m_r)}(u)\right\}\right].$$

MBD ranks each curve $g^{(m)}(u)$ as the sum of $\mathcal{A}_r^*\{g^{(m)}(u);g^{(m_1)}(u),\ldots,g^{(m_r)}(u)\}$ over all possible combinations of bands delineated by $2 \leqslant R \leqslant M$ total curves that can be formed from the sample of M total curves. The higher the MBD value for a curve means there is a larger proportion of bands that the curve lies in and translates to a curve closer to the center of the sample. Similarly, the lower the MBD value for a curve means there is a smaller proportion of bands that the curve lies in which translates to a more outlying observation. Therefore, MBD results in an effective ordering of the sample from the center outward. We follow common practice and set $R = 2$ in applications for computational efficiency (and drop R from the MBD notation), where bands delineated by all combinations of two curves are considered in the MBD definition (i.e. $MBD_M\{g^{(m)}(u)\} \equiv MBD_{M,2}\{g^{(m)}(u)\}$).

MBD of the posterior estimates of the longitudinal eigenfunctions $\psi_j(s)$ and functional eigenfunctions $\phi_k(t)$, which are both denoted as $g(u) \equiv \{\psi_j(s), \phi_k(t)\}$, are utilized to obtain point estimates (i.e. MBD median) and functional band depth central posterior envelopes (i.e. MBD-CPEs). While the MBD median, denoted by $\widehat{m}\{g(u)\}$, equals the curve with the largest MBD value in the posterior sample, the $(1-\alpha)100\%$ MBD-CPE, denoted by $D_{1-\alpha}\{g(u)\}$, is formed by the band delineated by the $(1-\alpha)100\%$ deepest posterior estimates in the sample. Both the MBD median and MBD-CPEs for $g(u)$ are targeted via Algorithm 3.

Point estimates and CPEs can also be formed based on the ranking of a posterior sample of d-dimensional surfaces through the use of modified volume depth (MVD). Let $\boldsymbol{g}^{(1)}(\boldsymbol{u}),\ldots,\boldsymbol{g}^{(M)}(\boldsymbol{u})$, $\boldsymbol{u} \in \mathcal{D}$, denote a sample of M d-dimensional posterior surfaces that is defined on $\mathcal{D} \in \mathbb{R}^d$. Further let $\mathcal{A}_r\{\boldsymbol{g}^{(m)}(\boldsymbol{u});\boldsymbol{g}^{(m_1)}(\boldsymbol{u}),\ldots,\boldsymbol{g}^{(m_r)}(\boldsymbol{u})\} = \{\boldsymbol{u} \in \mathcal{D} : \min_{w=m_1,\ldots,m_r} \boldsymbol{g}^{(w)}(\boldsymbol{u}) \leqslant \boldsymbol{g}^{(m)}(\boldsymbol{u}) \leqslant \max_{w=m_1,\ldots,m_r} \boldsymbol{g}^{(w)}(\boldsymbol{u})\}$ denote the region in $\mathcal{D}$ where the d-dimensional

Algorithm 3: MBD Median $\widehat{m}\{g(u)\}$ and MBD-CPE $D_{1-\alpha}\{g(u)\}$ for $g(u) \equiv \{\psi_j(s), \phi_k(t)\}$

```
Step 1: Calculate the MBD of the posterior samples:
    $$MBD_M\{g^{(1)}(t)\}, \ldots, MBD_M\{g^{(M)}(u)\}.$$
Step 2: Order the MBD values of the posterior sample from the smallest to the largest and
    denote the corresponding ordered samples as $g^{[1]}(u), \ldots, g^{[M]}(u)$.
Step 3: Calculate the $(1-\alpha)100\%$ MBD-CPE as
    $$D_{1-\alpha}\{g(u)\} = B\left\{g^{[[\alpha M]+1]}(u), \ldots, g^{[M]}(u)\right\},$$
    and the MBD median as $\widehat{m}\{g(u)\} = g^{[M]}(u)$.
```

surface $\boldsymbol{g}^{(m)}(\boldsymbol{u})$ lies inside the simplex delineated by the d-dimensional surfaces $\boldsymbol{g}^{(m_1)}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{(m_r)}(\boldsymbol{u})$. If $\mathbb{L}_d(\cdot)$ denotes the Lebesgue measure extended to $\mathbb{R}^d$, then $\mathcal{A}_r^*\{\boldsymbol{g}^{(m)}(\boldsymbol{u}); \boldsymbol{g}^{(m_1)}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{(m_r)}(\boldsymbol{u})\} = \mathbb{L}_d[\mathcal{A}_r\{\boldsymbol{g}^{(m)}(\boldsymbol{u}); \boldsymbol{g}^{(m_1)}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{(m_r)}(\boldsymbol{u})\}]/\mathbb{L}_d(\mathcal{D})$ is used to denote the proportion of time that the d-dimensional surface $\boldsymbol{g}^{(m)}(\boldsymbol{u})$ lies inside the simplex formed by the d-dimensional surfaces $\boldsymbol{g}^{(m_1)}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{(m_r)}(\boldsymbol{u})$. Using $R = 2$ total d-dimensional surfaces to delineate a simplex (similar to the definition of MBD), MVD for the d-dimensional surface $\boldsymbol{g}^{(m)}(\boldsymbol{u})$ in $\boldsymbol{g}^{(1)}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{(M)}(\boldsymbol{u})$, is given as

$$MVD_{M,d}\left\{\boldsymbol{g}^{(m)}(\boldsymbol{u})\right\} = \binom{M}{2}^{-1} \sum_{1 \leqslant m_1 < m_2 \leqslant M} \mathcal{A}_2^*\left\{\boldsymbol{g}^{(m)}(\boldsymbol{u}); \boldsymbol{g}^{(m_1)}(\boldsymbol{u}), \boldsymbol{g}^{(m_2)}(\boldsymbol{u})\right\}.$$

When $d = 2$, MVD is used to order either the posterior estimates of the mean function, longitudinal marginal covariance, or functional marginal covariance (i.e. $\boldsymbol{g}(\boldsymbol{u}) \equiv \{\mu(s,t), K_{\mathcal{S}}(s,s'), K_{\mathcal{T}}(t,t')\}$) to obtain the MVD median, denoted by $\widetilde{m}\{\boldsymbol{g}(\boldsymbol{u})\}$. Additionally, an alternative way for obtaining point estimates and CPEs for $g(u) = \{\psi_j(s), \phi_k(t)\}$ is through the ranking of the posterior sample of $\boldsymbol{g}(\boldsymbol{u}) \equiv \{K_{\mathcal{S}}(s,s'), K_{\mathcal{T}}(t,t')\}$ based on MVD. The MVD median and $(1-\alpha)100\%$ MVD-CPEs formed for $g(u)$ are denoted by $\widetilde{m}\{g(u)\}$ and $D_{1-\alpha}^{\star}\{g(u)\}$, respectively. By first ordering the posterior sample of respective marginal covariances, the ordered corresponding posterior sample of eigenfunctions is obtained and used to calculate the MVD median and MVD-CPEs. The MVD median equals the functional MVD median of $g(u)$, while the $(1-\alpha)100\%$ MVD-CPE for $g(u)$ is formed by the band delineated by the $M(1-\alpha)100\%$ deepest longitudinal or functional eigenfunction estimates in the sample, where the estimates are ordered according to MVD of their respective and corresponding posterior marginal covariances. The $(1-\alpha)100\%$ MVD-CPEs can also be obtained for the mean function, denoted by $D^{\star}_{1-\alpha}\{\mu(s,t)\}$, by calculating the bands formed from the $(1-\alpha)100\%$ deepest mean functions via MVD at fixed slices across $s \in \mathcal{S}$ and/or $t \in \mathcal{T}$ (i.e. $\mu(t|s)$ and $\mu(s|t)$). When $d = 4$, the posterior estimates of the covariance kernel (i.e. $\boldsymbol{g}(\boldsymbol{u}) \equiv K\{(s,t),(s',t')\}$) are ordered using MVD to obtain the kernel MVD median, denoted by $\overline{m}\{\boldsymbol{g}(\boldsymbol{u})\}$. This ordering based on the MVD of the posterior covariance kernels can also be used to order the longitudinal and functional covariances to obtain the kernel MVD medians denoted by $\overline{m}\{\boldsymbol{g}^*(\boldsymbol{u})\}$ for $\boldsymbol{g}^*(\boldsymbol{u}) \equiv \{K_{\mathcal{S}}(s,s'), K_{\mathcal{T}}(t,t')\}$. Similar to the procedure described above, the ordering of the respective marginal covariances via the MVD of the covariance kernels results in an ordering of the corresponding posterior eigenfunctions to obtain the kernel MVD median and $(1-\alpha)100\%$ kernel MVD-CPEs denoted by $\overline{m}\{g(u)\}$ and $D^{\dagger}_{1-\alpha}\{g(u)\}$, respectively, for $g(u) = \{\psi_j(s), \phi_k(t)\}$. The point estimates and CPEs based on the ordering of high-dimensional surfaces via MVD are targeted in Algorithm 4.

Algorithm 4: MVD and Kernel MVD Median $\{\widetilde{m}(\cdot), \overline{m}(\cdot)\}$ and MVD-CPE and Kernel MVD-CPE $\{D^{\star}_{1-\alpha}(\cdot), D^{\dagger}_{1-\alpha}(\cdot)\}$

```
Step 1: Calculate the MVD of the posterior $d$-dimensional surfaces:
    $MVD_{M,d}\{\boldsymbol{g}^{(1)}(\boldsymbol{u})\}, \ldots, MBD_{M,d}\{\boldsymbol{g}^{(M)}(\boldsymbol{u})\}$.
Step 2: Order the MVD values of the posterior sample from the smallest to the largest and
    denote the corresponding ordered posterior sample as $\boldsymbol{g}^{[1]}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{[M]}(\boldsymbol{u})$.
When $d = 2$ and $\boldsymbol{g}(\boldsymbol{u}) \equiv \mu(s,t)$:
    Step 3: Calculate the MVD median as $\widetilde{m}\{\boldsymbol{g}(\boldsymbol{u})\} = \boldsymbol{g}^{[M]}(\boldsymbol{u})$.
    Step 4: Calculate the $(1-\alpha)100\%$ MVD-CPEs across fixed slices of $s \in \mathcal{S}$ as
        $$D^{\star}_{1-\alpha}\{\mu(s,t)\} = \left[s \in \mathcal{S} : B\left\{\mu^{[\lfloor \alpha M \rfloor + 1]}(t|s), \ldots, \mu^{[M]}(t|s)\right\}\right],$$
    and/or across fixed slices of $t \in \mathcal{T}$ as
        $$D^{\star}_{1-\alpha}\{\mu(s,t)\} = \left[t \in \mathcal{T} : B\left\{\mu^{[\lfloor \alpha M \rfloor + 1]}(s|t), \ldots, \mu^{[M]}(s|t)\right\}\right].$$
When $d = 2$ and $\boldsymbol{g}(\boldsymbol{u}) \equiv \{K_{\mathcal{S}}(s,s'), K_{\mathcal{T}}(t,t')\}$:
    Step 3: Calculate the MVD median as $\widetilde{m}\{\boldsymbol{g}(\boldsymbol{u})\} = \boldsymbol{g}^{[M]}(\boldsymbol{u})$.
    Step 4: The SVD of the ordered $\boldsymbol{g}^{[1]}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{[M]}(\boldsymbol{u})$ lead to their corresponding ordered
        $g^{[1]}(u), \ldots, g^{[M]}(u)$ for $g(u) = \{\psi_j(s), \phi_k(t)\}$.
    Step 5: Calculate the $(1-\alpha)100\%$ MVD-CPE
        $$D^{\star}_{1-\alpha}\{g(u)\} = B\left\{g^{[\lfloor \alpha M \rfloor + 1]}(u), \ldots, g^{[M]}(u)\right\},$$
    and the MVD median as $\widetilde{m}\{g(u)\} = g^{[M]}(u)$.
When $d = 4$ and $\boldsymbol{g}(\boldsymbol{u}) \equiv K\{(s,t),(s',t')\}$:
    Step 3: Calculate the kernel MVD median as $\overline{m}\{\boldsymbol{g}(\boldsymbol{u})\} = \boldsymbol{g}^{[M]}(\boldsymbol{u})$.
    Step 4: Based on ordering from Step 2: $m = 1, \ldots, M \mapsto m = [1], \ldots, [M]$, order the
        sample $\boldsymbol{g}^{*[1]}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{*[M]}(\boldsymbol{u})$ for $\boldsymbol{g}^{*}(\boldsymbol{u}) \equiv \{K_{\mathcal{S}}(s,s'), K_{\mathcal{T}}(t,t')\}$.
    Step 5: Obtain the kernel MVD medians as $\overline{m}\{\boldsymbol{g}^{*}(\boldsymbol{u})\} = \boldsymbol{g}^{*[M]}(\boldsymbol{u})$.
    Step 6: The SVD of the ordered $\boldsymbol{g}^{*[1]}(\boldsymbol{u}), \ldots, \boldsymbol{g}^{*[M]}(\boldsymbol{u})$ lead to their corresponding
        ordered $g^{[1]}(u), \ldots, g^{[M]}(u)$ for $g(u) = \{\psi_j(s), \phi_k(t)\}$.
    Step 7: Calculate the $(1-\alpha)100\%$ kernel MVD-CPE
        $$D^{\dagger}_{1-\alpha}\{g(u)\} = B\left\{g^{[\lfloor \alpha M \rfloor + 1]}(u), \ldots, g^{[M]}(u)\right\},$$
    and the kernel MVD median as $\overline{m}\{g(u)\} = g^{[M]}(u)$.
```

Functional depth-based medians proposed above ($\widehat{m}(\cdot)$, $\widetilde{m}(\cdot)$ and $\overline{m}(\cdot)$) estimate the central tendency in the posterior distributions of the B-LFPCA components and provide a realistic estimate equal to one of the observed functional posterior samples rather than relying on pointwise averages as is done in most of the traditional posterior summaries. In addition, the proposed functional depth based CPEs (denoted by $D_{1-\alpha}(\cdot)$ $D^{\star}_{1-\alpha}(\cdot)$, and $D^{\dagger}_{1-\alpha}(\cdot)$, corresponding to MBD, MVD, and kernel-MVD based summaries, respectively), are formed from a band delineated from the $1-\alpha$ deepest subset of the functional posterior estimates. Hence, the proposed methods use functional depth measures to construct fully data-driven summaries that capture the uncertainty in the posterior estimates in a flexible way. *It is important to note that CPEs are not credible intervals and, at the present, are not intended to be utilized for inference. Instead, CPEs provide a descriptive visualization tool for capturing the variation in the posterior sample, and when plotted for a grid of α values, allow for visualization of the most central regions of the functional posterior distributions.* Notations used for point estimates and CPEs for both the traditional and proposed functional depth summaries based are summarized in Table 3.1.

3.4 Simulation Studies

Eight simulation scenarios are showcased to display the use of CPEs in describing the variation in the posterior samples under violation of the weak separability assumption and in the presence of functional outliers. Additionally, finite sample properties of the traditional and depth based point estimates from Section 3.2.2 and Section 3.3, respectively, are studied under the different simulation scenarios. The eight simulation scenarios proposed are: Case 1 – no outliers, Case 2 – violation of weak separability assumption, Case 3 – magnitude outlier, Case 4 – amount of variation outlier, Case 5 – time-shifted longitudinal eigenfunction outlier, Case 6 – higher-frequency longitudinal eigenfunction outlier, Case 7 – time-shifted functional eigenfunction outlier, and Case 8 – higher-frequency functional eigenfunction outlier. In Case 1, we generate data using the product FPCA model (3.2) with added measurement error for $s \in [0,1]$ and $t \in [0,1]$ observed at a uniform grid of $n_s = 20$ longitudinal and $n_t = 20$ functional time points. The performance of the posterior summaries is evaluated in Case 2 when the assumption of weak separability is violated through generation of observations such that the product scores χ_{jk} are correlated over the functional and longitudinal dimensions. Magnitude outliers are generated by adding a constant deviation (with a random sign) to $\mu(s,t)$ for $t \geqslant T_i$ and $s \geqslant S_i$, where $T_i \sim \text{Unif}[0,1]$ and $S_i \sim \text{Unif}[0,1]$. This adds constant variation to a randomly selected portion of the unit time domain in both the functional and longitudinal dimensions, with greater variation added when $s > 1/2$ and $t > 1/2$. Amount of variation outliers are generated using larger variances of the product scores v_{jk}, which result in added variation throughout the unit interval in both the longitudinal and functional dimensions. Longitudinal eigenfunction outliers are generated from time-shifted (Case 5) or higher frequency (Case 6) longitudinal eigenfunctions, where additional variation is added along the direction of the time-shifted or higher frequency longitudinal eigenfunctions used to generate the outliers. Similarly, functional eigenfunction outliers are generated from time-shifted (Case 7) or higher frequency (Case 8) functional eigenfunctions. Results are reported for outliers generated as $q = 10$ and 20% of the sample ($n = 30$) for cases 3-8, where details of data generation under the eight simulation scenarios are deferred to Appendix C.3. The B-LFPCA model is fit using $p_1 = 11$ and $p_2 = 11$ B-spline basis functions in both longitudinal and functional dimensions with knots at $s = (1/8, 1/4, 3/8, 1/2, 5/8, 3/4, 7/8)$ and $t = (1/8, 1/4, 3/8, 1/2, 5/8, 3/4, 7/8)$ using a total of $q_1 = 4$ and $q_2 = 4$ longitudinal and functional latent factors. Previous work of Boland et al. (2023) using a similar Bayesian functional principal component analysis (BFPCA) model suggested selecting the number of basis functions as half the number of functional time points (e.g. $p_1 = \lfloor n_s/2 \rfloor$) and the number of latent factors as a fraction of the number of selected basis functions (e.g. $q_1 = max(6, \lfloor p_1/4 \rfloor)$). The B-LFPCA model may require fewer basis functions and latent factors as needed to avoid overfitting for interpretability of the CPEs. An example of this in practice is given in our data application in Section 3.5. Results are reported based on a total of 200 Monte Carlo with $10,000$ MCMC iterations ($2,500$ for burn-in and thinning at every 5th iteration) of four independent Markov chains, and $M = 6,000$ posterior estimates for each Monte Carlo run.

Finite sample performance of functional-valued point estimates containing a longitudinal and/or functional argument (e.g. longitudinal eigenfunctions and covariance kernel) are assessed via the standardized integrated mean squared error (IMSE), $IMSE_{\widehat{g}(u)} = [\int_u \{\widehat{g}(u) - g(u)\}^2 du] / \int_u g^2(u) du$, where $u \in \mathcal{D}$ can be multidimensional. Point estimates of scalars (e.g. longitudinal eigenvalues) are assessed via the standardized mean squared error (MSE), $MSE_{\widehat{\rho}} = (\widehat{\rho} - \rho)^2/\rho^2$, where ρ is a scalar value. The mean IMSE and MSE values from 200 Monte Carlo runs for the eight simulation scenarios are summarized in Tables 3.2 and 3.3, respectively, when the number of outliers is $q = 20\%$ and in Appendix Tables C.1 and C.2, respectively, when $q = 10\%$. The traditional and proposed point estimates for the longitudinal and functional eigenfunctions from the Monte Carlo run with the median IMSE are given in Figure 3.1 for Cases 1 and 2, and, respectively, in Figures C.1 and C.2 for Cases 3 through 8. Violation of the weak separability assumption in Case 2 leads to bias in the eigenfunctions as observed in Table 3.2 with larger IMSE values as compared to Case 1 and results in bias of the marginal longitudinal and functional covariances. In Case 3, the magnitude outliers lead to eigenfunction misspecification as the constant deviation is estimated as both the first longitudinal and functional eigenfunctions, leading to bias in every point estimate except for the mean function. The increase in IMSE for the mean function (Table 3.2) is instead due to increased variation in mean function estimation. The use of larger longitudinal and functional eigenvalues clearly results in bias of these estimates (Table 3.3) in Case 4 as well as estimates of the marginal covariances and covariance kernel (Table 3.2). Although it does not bias eigenfunction estimation, there is a slight increase in the IMSE of the eigenfunction estimates due to increased variation in eigenfunction estimation. Higher IMSE values of the marginal longitudinal eigenfunctions and covariance in Case 5 and marginal functional eigenfunctions and covariance in Case 7 are due to the added time-shifted eigenfunction outliers, leading to more variation in eigenfunction estimation in the longitudinal and functional time intervals in Cases 5 and 7, respectively. Bias in the second longitudinal and functional eigenfunctions in Cases 6 and 8, respectively, is a result of eigenfunction misspecification as the higher-frequency outlier eigenfunctions are estimated as the second leading direction of variations. In effect, this leads to bias in the estimation of the marginal longitudinal and functional covariances in Cases 6 and 8, respectively. In general, there does not appear to be an advantage for the traditional over the proposed depth-based point estimates given the similar IMSE and MSE values in Tables 3.2, 3.3, C.1, and C.2 in most of the simulation cases. This similar performance is potentially due to the data-adaptive behavior, capturing the effects of outliers and violations of model assumptions, of the B-LFPCA model reported in Shamshoian et al. (2022). Further in-depth and case-specific results for the effects the generated outliers have on the performance of the point estimates and CPEs for Cases 3-8 are deferred to Appendix C.3.

Figures 3.2 and 3.3 display CPEs for the two leading longitudinal and functional eigenfunctions, respectively, from a single Monte Carlo run overlaying $M = 6,000$ posterior estimates (given in gray) from α cutoffs ranging from 0.05 to 0.95 for Cases 1 and 2. Similar plots containing CPEs in outlier Cases 3 to 8 (when outlier percentage is equal to $q = 20\%$) for the longitudinal eigenfunctions, $\psi_1(s)$ and $\psi_2(s)$, are given in Appendix Figures C.5 and C.6, respectively, and functional eigenfunctions, $\phi_1(t)$ and $\phi_2(t)$, in Appendix Figures C.7 and C.8, respectively. Additionally, the MVD-CPEs of the mean function for Cases 1 and 3 ($q = 20\%$) are given in Appendix Figures C.3 and C.4, respectively. CPEs from increasing α cutoffs help visualize regions with the most central functional posterior observations in the sample. Note that while MBD-CPEs of the marginal eigenfunctions, which directly order the posterior longitudinal and functional eigenfunctions, are nested in each other for increasing α values, MVD-CPEs and kernel MVD-CPEs are not necessarily nested. The MVD-CPEs and kernel MVD-CPEs are based on the functional depth rankings of the posterior longitudinal and functional marginal covariances and covariance kernels, respectively, which may not result in a nested ordering of the longitudinal and functional eigenfunctions. This is also the reason why MVD-CPEs are typically wider than MBD-CPEs, incorporating variation

Table 3.2: The mean standardized integrated mean squared error (IMSE) for both the traditional and functional depth-based point estimates from the 200 Monte Carlo runs. The eight simulation cases correspond to: Case 1 – no outliers, Case 2 – violation of weak separability, Case 3 – magnitude outlier, Case 4 – amount of variation outlier, Case 5/6 – time-shifted/higher-frequency longitudinal eigenfunction outlier, Case 7/8 – time-shifted/higher-frequency functional eigenfunction outlier, with $q = 20\%$ outliers.

			$q = 20\%$					
Point Estimate	Case 1	Case 2*	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
	IMSE		**IMSE**					
$\widehat{\mu}(t)$	0.0003	0.0005	0.0023	0.0006	0.0004	0.0004	0.0005	0.0004
$\widetilde{m}\{\mu(t)\}$	0.0003	0.0005	0.0024	0.0007	0.0004	0.0005	0.0005	0.0005
$\widehat{\psi}_1(s)$	0.0129	0.1555	1.9079	0.0627	0.0346	0.0215	0.0236	0.0174
$\widetilde{\psi}_1(s)$	0.0128	0.1587	1.9199	0.0632	0.0344	0.0198	0.0238	0.0174
$\widehat{m}\{\psi_1(s)\}$	0.0129	0.1593	1.9190	0.0655	0.0348	0.0207	0.0241	0.0177
$\widetilde{m}\{\psi_1(s)\}$	0.0137	0.1604	1.9182	0.0631	0.0363	0.0208	0.0246	0.0190
$\overline{m}\{\psi_1(s)\}$	0.0138	0.1602	1.8988	0.0667	0.0364	0.0220	0.0250	0.0185
$\widehat{\psi}_2(s)$	0.0258	0.1669	1.5821	0.0639	0.0427	0.7630	0.0362	0.0313
$\widetilde{\psi}_2(s)$	0.0163	0.1613	1.6241	0.0643	0.0371	0.9557	0.0260	0.0213
$\widehat{m}\{\psi_2(s)\}$	0.0172	0.1640	1.6326	0.0643	0.0385	0.9119	0.0275	0.0225
$\widetilde{m}\{\psi_2(s)\}$	0.0205	0.1670	1.6280	0.0652	0.0414	0.9112	0.0292	0.0263
$\overline{m}\{\psi_2(s)\}$	0.0211	0.1671	1.6273	0.0697	0.0420	0.8879	0.0329	0.0269
$\widehat{\phi}_1(t)$	0.0078	0.1087	1.7919	0.0401	0.0070	0.0089	0.0287	0.0136
$\widetilde{\phi}_1(t)$	0.0077	0.1099	1.8021	0.0407	0.0070	0.0089	0.0283	0.0118
$\widehat{m}\{\phi_1(t)\}$	0.0080	0.1112	1.8009	0.0415	0.0071	0.0092	0.0291	0.0125
$\widetilde{m}\{\phi_1(t)\}$	0.0083	0.1110	1.8040	0.0417	0.0078	0.0095	0.0289	0.0130
$\overline{m}\{\phi_1(t)\}$	0.0086	0.1107	1.8010	0.0420	0.0081	0.0102	0.0309	0.0157
$\widehat{\phi}_2(t)$	0.0343	0.1328	1.9319	0.0422	0.0321	0.0332	0.0397	1.0371
$\widetilde{\phi}_2(t)$	0.0295	0.1292	1.9872	0.0426	0.0263	0.0304	0.0358	1.3491
$\widehat{m}\{\phi_2(t)\}$	0.0197	0.1222	1.9843	0.0424	0.0191	0.0209	0.0343	1.2572
$\widetilde{m}\{\phi_2(t)\}$	0.0291	0.1305	1.9801	0.0449	0.0269	0.0299	0.0376	1.3828
$\overline{m}\{\phi_1(t)\}$	0.0286	0.1268	1.9842	0.0463	0.0260	0.0305	0.0395	1.5042
$\widetilde{K}_{\mathcal{S}}(s,s')$	0.0627	0.2083	17.684	0.5173	0.0873	0.1501	0.0711	0.0685
$\widetilde{m}\{K_{\mathcal{S}}(s,s')\}$	0.0656	0.2058	16.648	0.4575	0.0913	0.1511	0.0716	0.0698
$\overline{m}\{K_{\mathcal{S}}(s,s')\}$	0.0690	0.2089	15.966	0.4626	0.0948	0.1570	0.0746	0.0767
$\widetilde{K}_{\mathcal{T}}(t,t')$	0.0571	0.1765	16.075	0.4544	0.0598	0.0605	0.0947	0.1589
$\widetilde{m}\{K_{\mathcal{T}}(t,t')\}$	0.0603	0.1775	15.083	0.4049	0.0623	0.0630	0.0991	0.1635
$\overline{m}\{K_{\mathcal{T}}(t,t')\}$	0.0650	0.1803	14.551	0.4053	0.0672	0.0721	0.1052	0.1696
$\overline{K}\{(s,t),(s',t')\}$	0.1005	0.1922	17.182	0.5114	0.1354	0.1841	0.1246	0.2007
$\overline{m}[K\{(s,t),(s',t')\}]$	0.1076	0.2054	15.740	0.4831	0.1437	0.1949	0.1349	0.2126

*IMSE for point estimates of $K\{(s,t),(s',t')\}$ are calculated in comparison to the covariance kernel used for data generation in Case 2 (Appendix C.3).

Table 3.3: The mean standardized integrated mean squared error (IMSE) and standardized mean squared error (MSE) for both the traditional and functional depth-based point estimates from the 200 Monte Carlo runs. The eight simulation cases correspond to: Case 1 – no outliers, Case 2 – violation of weak separability, Case 3 – magnitude outlier, Case 4 – amount of variation outlier, Case 5 – time-shifted longitudinal eigenfunction outlier, Case 6 – higher-frequency longitudinal eigenfunction outlier, Case 7 – time-shifted functional eigenfunction outlier, Case 8 – higher-frequency functional eigenfunction outlier with q = 20% outliers.

			q = 20%					
Point Estimate	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
	MSE		**MSE**					
$\widehat{\tau}_1$	0.0488	0.0591	15.495	0.6121	0.0499	0.0577	0.0545	0.0510
$\widetilde{\tau}_1$	0.0487	0.0592	15.315	0.5838	0.0510	0.0632	0.0541	0.0509
$\widehat{\tau}_2$	0.0378	0.0535	9.0287	1.4118	0.0510	0.0375	0.0363	0.0439
$\widetilde{\tau}_2$	0.0426	0.0632	9.0209	1.6163	0.0596	0.0649	0.0420	0.0502
$\widehat{\vartheta}_1$	0.0424	0.0494	12.937	0.5113	0.0484	0.0466	0.0450	0.0683
$\widetilde{\vartheta}_1$	0.0424	0.0499	12.806	0.5838	0.0484	0.0465	0.0466	0.0733
$\widehat{\vartheta}_2$	0.0333	0.0377	18.493	2.9308	0.0356	0.0356	0.2399	0.1281
$\widetilde{\vartheta}_2$	0.0403	0.0417	18.287	3.1965	0.0413	0.0412	0.2571	0.1106

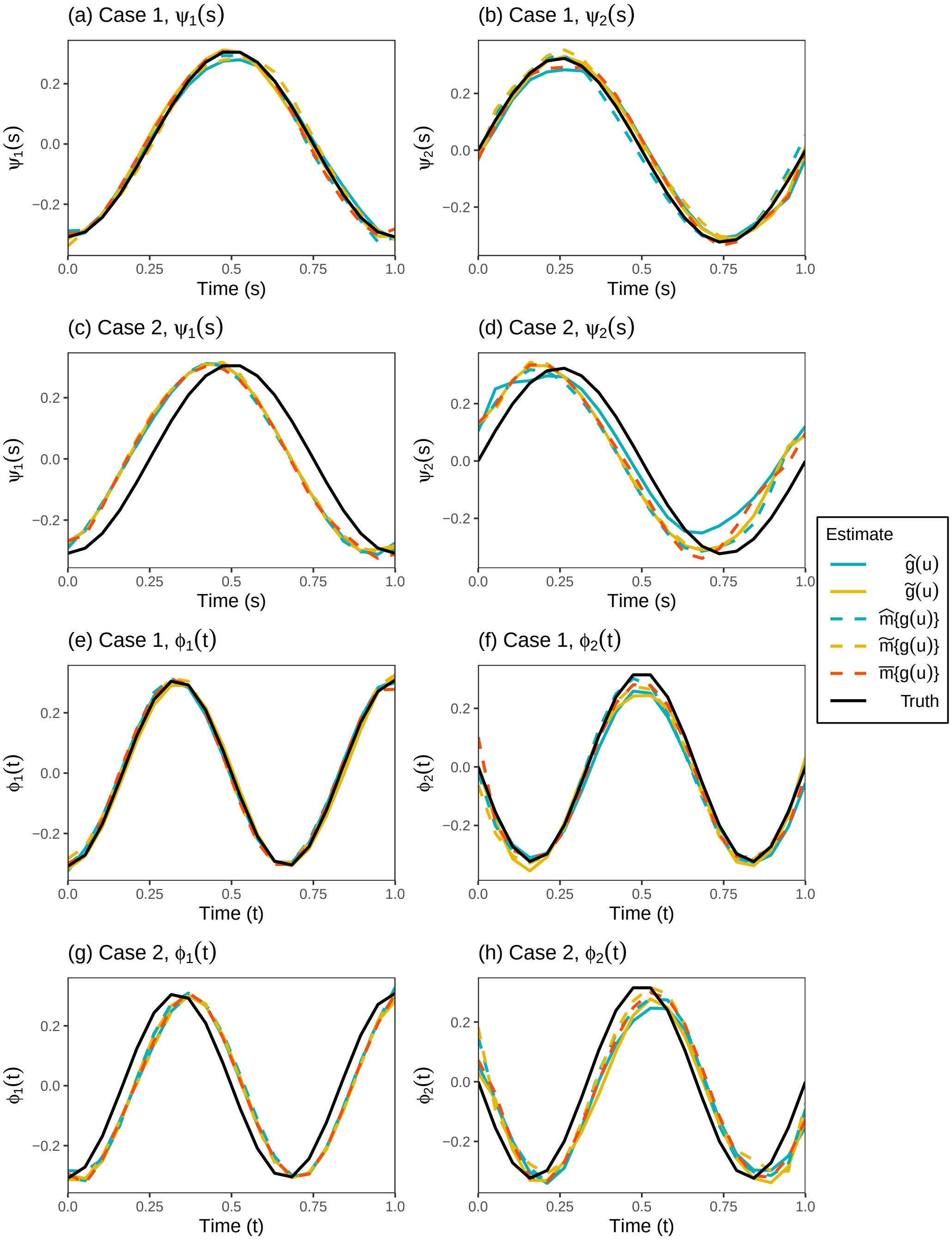


Figure 3.1: Point estimates of $\psi_1(s)$, $\psi_2(s)$, $\phi_1(t)$, and $\phi_2(t)$ in the are displayed from the top to bottom rows for simulation Cases 1 and 2 from runs with 50th percentile IMSE values. Longitudinal/functional eigenfunction estimates, longitudinal/functional eigenfunctions estimates via longitudinal/functional marginal covariance estimation, MBD median, MVD median, and kernel MVD median are given in solid blue, solid yellow, dashed blue, dashed yellow, and dashed red overlaying the true function given in solid black.

from the marginal longitudinal/functional covariance processes rather than only longitudinal/functional eigenfunction-specific variation. Similarly, this is also the reason why generally the kernel MVD-CPEs are often even wider than MVD-CPEs, incorporating variation from the entire covariance process across both the longitudinal and functional dimensions rather than only the longitudinal or functional level variation captured in the respective marginal covariances. This is particularly apparent in Cases 3-8, where added variation due to functional outliers leads to additional contamination of the covariance kernel. CPEs utilized in B-LFPCA appear to provide a flexible representation of the shape and spread of the posterior samples even in the presence of functional outliers and can capture increasing levels of variation across multiple dimensions, contributing to the visualization of the posterior samples.

While MVD-CPEs are typically more narrow than kernel MVD-CPEs in most of the simulation cases, the kernel MVD-CPEs in Case 2, which generates functional observations using product scores that violate the key model assumption of weak separability, often behave in the opposite manner. In Figure 3.2(k), the posterior sample space covered by the MVD-CPE up to $\alpha = 0.45$ is approximately the same covered by the kernel MVD-CPE up to $\alpha = 0.75$ in Figure 3.2(ℓ). Similarly, in Figure 3.3(k), the posterior sample space covered by the MVD-CPE up to $\alpha = 0.65$ is covered up to $\alpha = 0.85$ by the kernel MVE-CPE in Figure 3.2(ℓ), while the kernel MVD-CPE contour up to $\alpha = 0.05$ in Case 1 (Figure 3.3(i)) is much wider than the corresponding MVD-CPE contour (Figure 3.3(h)). Narrower kernel MVD-CPE contours compared to the MVD-CPE contours observed in Case 2 are potentially a consequence of the generated functional data violating the assumption of weak separability. To recall, under the assumption of weak separability the covariance kernel, $K\{(s,t),(s',t')\}$, can be written as a linear combination of the variance of the product scores, v_{jk}, and eigenfunctions of the longitudinal and functional marginal covariances, $\psi_j(s)$ and $\phi_k(t)$. When this assumption is violated, the covariance kernel can still be estimated, but integration over the functional/longitudinal dimension to obtain the functional/longitudinal marginal covari-

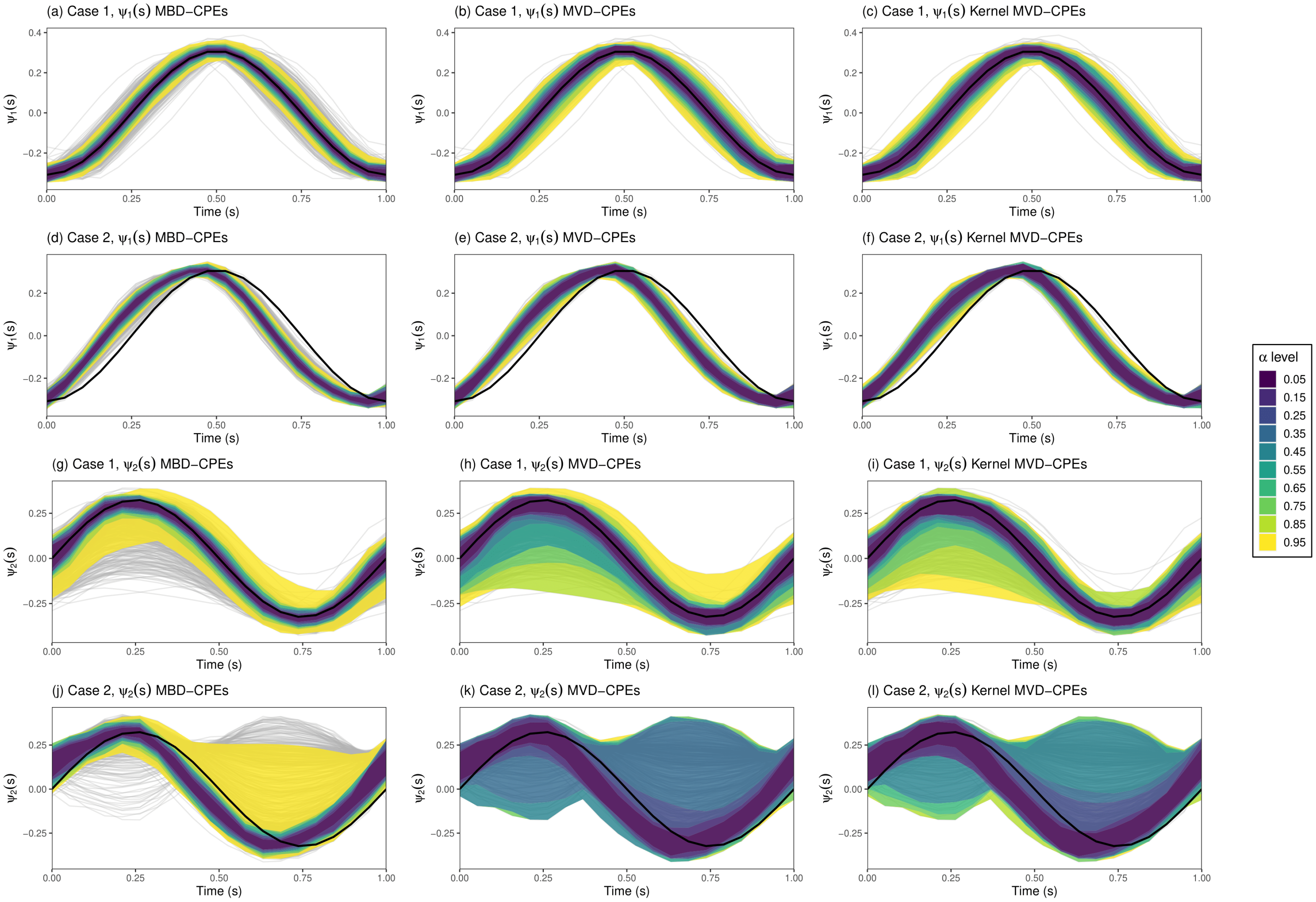


Figure 3.2: CPE contours of the longitudinal eigenfunctions, $\psi_1(s)$ (rows 1-2) and $\psi_2(s)$ (rows 3-4), for simulation Cases 1 and 2. The light grey solid lines, overlaying the true function in solid black, represent the sample of $M = 6000$ posterior estimates. The left to right hand columns display the MBD, MVD-CPEs, and kernel MVD-CPEs, denoted by $D_{1-\alpha}\{\psi_j(s)\}$, $D^{\star}_{1-\alpha}\{\psi_j(s)\}$, and $D^{\dagger}_{1-\alpha}\{\psi_j(s)\}$, respectively, at a grid of α levels marked by varying contour colors.

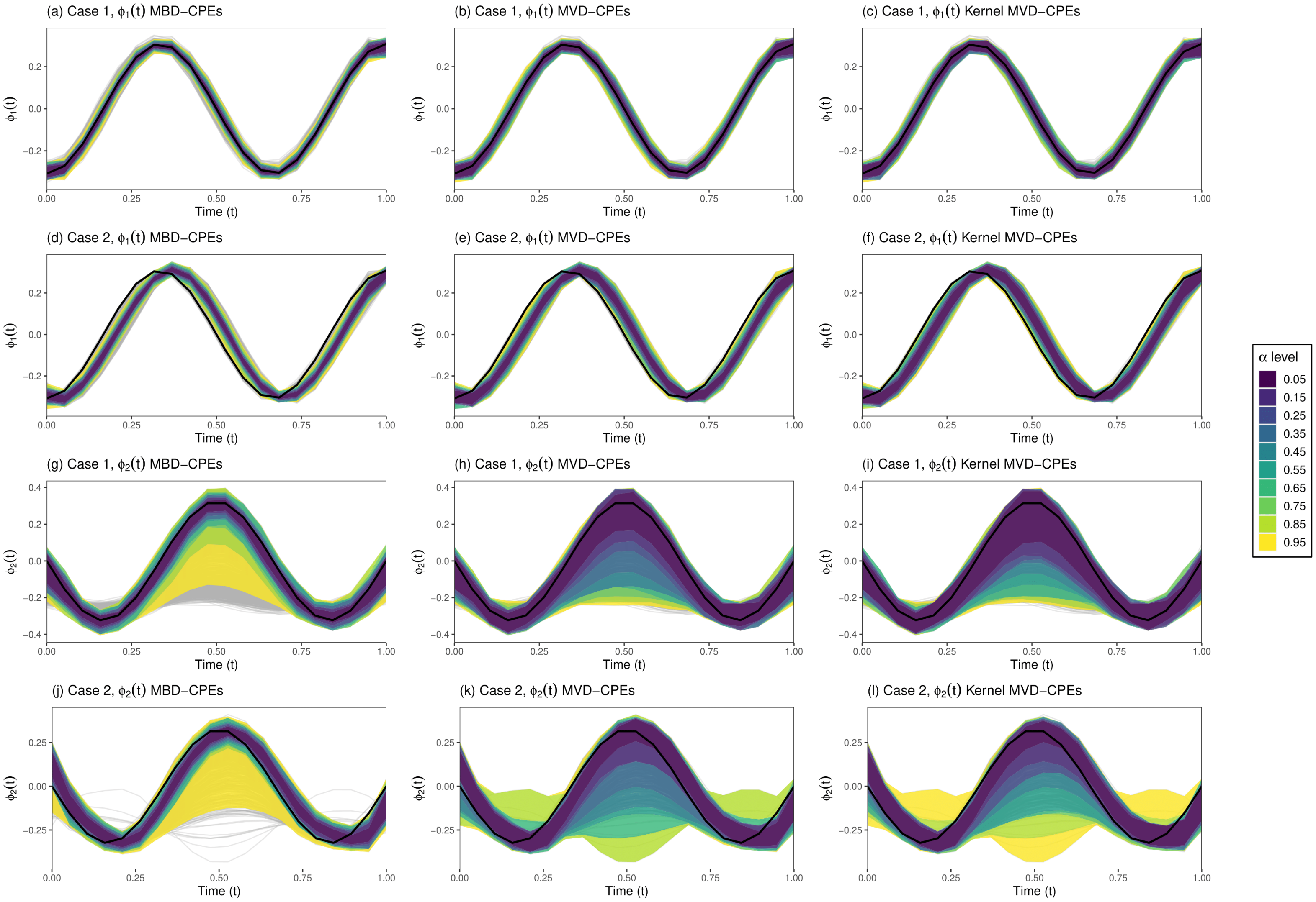


Figure 3.3: CPE contours of the functional eigenfunctions, $\phi_1(t)$ (rows 1-2) and $\phi_2(t)$ (rows 3-4), for simulation Cases 1 and 2. The light grey solid lines, overlaying the true function in solid black, represent the sample of $M = 6000$ posterior estimates. The left to right hand columns display the MBD, MVD-CPEs, and kernel MVD-CPEs, denoted by $D_{1-\alpha}\{\phi_k(t)\}$, $D^{\star}_{1-\alpha}\{\phi_k(t)\}$, and $D^{\dagger}_{1-\alpha}\{\phi_k(t)\}$, respectively, at a grid of α levels marked by varying contour colors.

ance does not result in the functional/longitudinal eigenfunctions used to generate the data when SVD is applied. This is observed in the inflated IMSE values in Table 3.2 as compared to Case 1 for these estimates, and Figure 3.1 indicates that, although the eigenfunctions retain their general shape, they become warped and stretched across the unit time interval in both longitudinal and functional dimensions. Hence, the ordering and visualization of the eigenfunctions given by the kernel MVD-CPEs rather than the MVD-CPEs can lead to a more conservative representation of uncertainty in a posterior sample when weak separability is violated due to the unreliability of the posterior marginal covariances and is a potentially useful tool worth the added computational burden if suspected.

3.5 Data application

3.5.1 Implicit learning experiment

Our motivating study was conducted at UCLA by our collaborator Dr. Shafali Jeste on neural correlates of implicit learning for children with autism spectrum disorder (ASD), a heterogeneous neurodevelopmental disorder characterized by social interaction and communication impairments (Lord et al. (2000); Jeste et al. (2015)). Implicit learning is characterized as the detection of irregularities in one's environment without conscious awareness or intention to learn and is thought to play a role in language and social behavior, which can provide insights into non-verbal cognition and adaptive social functioning in ASD (Jeste et al. (2015)). A group of 37 children with a prior diagnosis of ASD aged two to five years old and 34 age-matched typically developing (TD) peers participated in the visual implicit learning study during which electroencephalography (EEG) signals measured spontaneous electrical activity across the scalp.

During the experiment, the participants were shown a continuous stream of six-colored geometric shapes (pink squares, blue crosses, yellow circles, turquoise diamonds, gray triangles, and red octagons), grouped into three shape pairs that were presented in a random order on a computer screen (see Figure 3.4(a)). Presentation of a single shape constituted an experimental trial and resulted in an event-related potential (ERP) waveform recorded as a time-locked EEG signal in response to the visual stimuli at electrodes placed across the scalp. As the experiment progressed, transitions within a shape pair stayed fixed, and hence the within-pair ordering could be learned forming the “expected” condition, while the between-pair transitions were random and unpredictable forming the “unexpected” condition. Shape pairs were presented 120 times resulting in ERP waveforms recorded for 120 trials for both the expected and unexpected conditions. Implicit learning was detected as condition differences in the recorded ERP waveforms between the expected and unexpected conditions. As learning was anticipated to occur over trials, capturing the longitudinal trends in condition differences was a key goal to reflect the evolution of implicit learning for comparison between diagnostic groups (ASD vs. TD).

In studies such as the one described above, recorded ERPs result in functional data with commonly studied paradigm-specific components. The implicit learning paradigm leads to two well-known ERP components thought to be related to cognitive processes and early category recognition: the P3 peak and N1 dip (see Figure 3.4(c)) (Jeste et al. (2015)). Using the same pre-processed ERP data studied in Hasenstab et al. (2015, 2017) and Shamshoian et al. (2022), our analysis focuses on the condition difference (expected - unexpected) of the ERP waveforms. Due to the low SNR of the pre-processed ERP, the meta-preprocessing procedure (MAP-ERP) detailed in Hasenstab et al. (2015) is applied to the pre-processed ERPs to boost the SNR. MAP-ERP allows for the P3 peak locations to be identified while still retaining the longitudinal information in the ERPs (common to collapse the longitudinal dimension to boost SNR by averaging ERP over trials). For each subject, electrode, and condition, MAP-ERP averages the ERP waveforms in a sliding window of overlapping sets of trials with a maximum number of 30 trials within a set and applies a peak detection algorithm (Jeste et al. (2015)) to the averaged ERPs to identify the P3 peak location for a given trial. Once the P3 locations are obtained, the averaged ERPs are centered around the

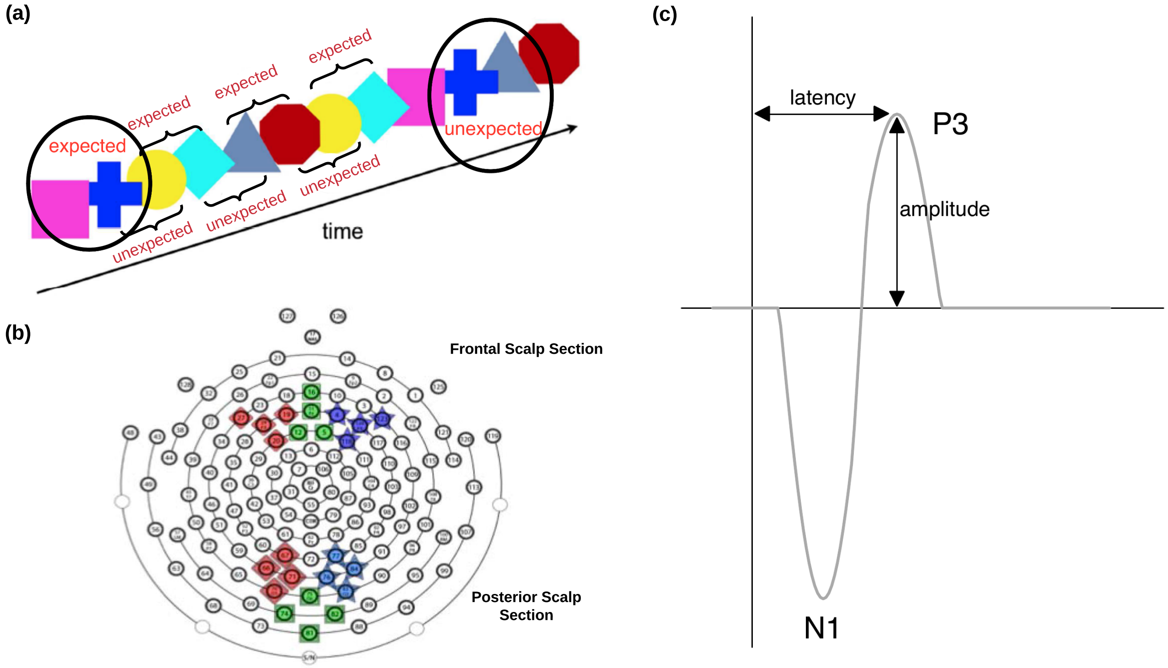
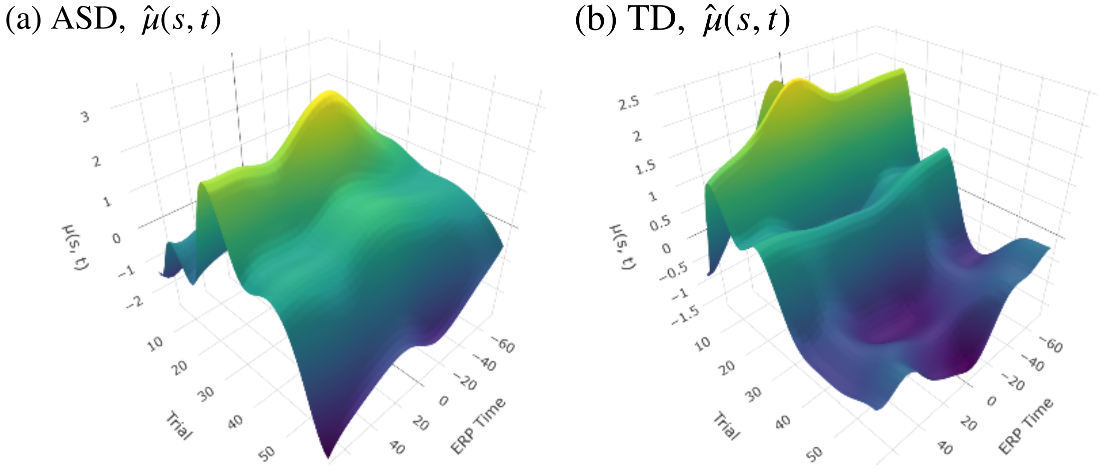


Figure 3.4: (a) Visualization of the implicit learning paradigm. The continuous stream of six-colored shapes is organized into three shape pairs. The "expected" condition is defined as the transition between shapes within a shape pair, and the "unexpected" condition is defined as the transition between shape pairs. (b) The four electrodes located in the right frontal portion of the scalp of interest are denoted by the dark blue stars in the labeled frontal scalp section. (c) A depiction of the ERP phasic components P3 and N1 in the implicit learning paradigm.

P3 peak by examining a 144 ms symmetric window around the P3 peak (i.e., $t \in [-72$ ms, 72 ms$]$). This results in ERP curves with $n_t = 37$ functional time points and an enhanced SNR that are aligned by the P3 peak across all subjects, trials, electrodes, and conditions at $t = 0$. Since our interest lies in characterizing implicit learning, the condition differences are obtained by subtracting the meta-preprocessed ERP corresponding to the unexpected condition from the expected condition for each subject, trial, and electrode. Lastly, the condition differences are averaged across the four electrodes in the right frontal region of the scalp (see Figure 3.4(b)) and constitute our functional dimension with $n_t = 37$ time points for each subject observed over $n_s = 56$ longitudinal trials $s \in [5, 60]$, where maximal condition differentiation was detected in previous studies (Hasenstab et al. (2015)). Four ASD subjects and two TD subjects were removed prior to analysis due to their trial-level missingness profiles, and all other missing trial-level observations are imputed as an update in the MCMC algorithm. Further details on pre-processing of the ERP data, subject removal due to missingness, imputation of missing data, and hyperparameters used in fitting the B-LFPCA model are deferred to Appendix C.4. The B-LFPCA model was fit separately for the ASD and TD groups with $(p_1, p_2) = (10, 10)$ B-spline basis functions and $(q_1, q_2) = (5, 5)$ latent factors, which was a fewer amount of basis functions than the rule of thumb discussed in Section 3.4 but avoided overfitting and led to interpretable CPEs. Additionally, the 95% simultaneous parametric credible intervals (Krivobokova et al. (2010); Crainiceanu et al. (2007)) were calculated for the posterior samples of the longitudinal and functional eigenfunctions to highlight our visualization approach via the proposed CPEs in comparison to traditional methods for uncertainty quantification.

3.5.2 Data analysis results

Figure 3.5 displays a three-dimensional visualization of the estimated mean of the condition difference for ASD and TD groups across ERP time and trials. Additionally, the MVD-CPEs of the mean function at α contours ranging from 0.05 to 0.95 given at fixed slices of the lon-



(a) ASD, $\hat{\mu}(s,t)$

(b) TD, $\hat{\mu}(s,t)$

Figure 3.5: Point estimates of the mean condition differentiation $\widehat{\mu}(s,t)$ for ASD and TD

gitudinal dimension $s = (5, 35, 60)$ (rows 1 and 3) and fixed slices of the functional dimension $t = (-40, 0, 40)$ (rows 3 and 4) for both diagnostic groups, overlaying $M = 6,000$ posterior estimates in gray, are displayed in Figure C.9. Similar to the results seen in Shamshoian et al. (2022), the ASD group tends to have positive condition differentiation between trials 23 to 50 with maximum positive condition differentiation at trial 29 (Figure 3.5(a)). Whilst the TD group tends to have positive condition differentiation at earlier trials of the experiment compared to the ASD group (Figure 3.5(b)). As positive condition differentiation is believed to be indicative of implicit learning, these findings suggest that the TD group may be differentiating between the expected and unexpected conditions earlier than the ASD group. However, further exploration into the covariance structure can provide further insight into condition differentiation between the ASD and TD groups.

MBD-CPEs from α contours ranging from 0.05 to 0.95 for the leading three longitudinal (rows 1 and 2) and functional (row 3 and 4) eigenfunctions for both diagnostic groups, overlaying $M = 6,000$ posterior estimates in gray, in Figure 3.6. The MVD-CPEs and kernel

MVD-CPEs of the leading three longitudinal eigenfunctions for both diagnostic groups are deferred to Appendix Figures C.10 and C.11, respectively, and both are wider than the MBD-CPEs as expected. Decomposition of the longitudinal marginal covariance over trials provides insight into how implicit learning evolved through the experiment and the differences between the ASD and TD groups. The three leading longitudinal eigencomponents explain more than 80% of the total variation in both groups (median and (2.5th, 97.5th) percentiles of FVE at 80.2% (75.7%, 84.3%) and 84.1% (80.8%, 87.2%) in the ASD and TD groups, respectively). The leading longitudinal eigenfunction in the ASD (FVE: 37.1% (30.9%, 46.8%)) and TD (FVE: 40.5% (35.0%, 48.4%)) groups are flat across trials, representing the overall variability in condition differentiation trial-to-trial (Figure 3.6(a) and (d)). The second leading longitudinal eigenfunction in the ASD (FVE: 24.9% (19.8%, 30.4%)) and TD (FVE: 29.3% (23.3%, 34.5%)) groups represents boundary variation (Figures 3.6(b) and (e)). This is in part due to the MAP-ERP pre-processing step as the number of trials contained in the overlapping sets of trials in the sliding window shrinks linearly towards one at the boundaries. Trial 5 contains the smallest number of trials within its sliding window which results in increased variability, but variation at the boundaries can also be a result of difficulties in modeling trials located at the boundaries (i.e. trial 60). This is consistent with the wider MVD-CPEs mean functions displayed in Figure C.9(a) and (g) for trial 5 and Figure C.9(c) and (i) where the ASD group has more variability at trial 60 than the TD group. Finally, the third longitudinal eigenfunction represents variation in condition differentiation in both ASD (FVE: 17.6% (13.6%, 21.9%)) and TD (FVE: 14.0% (10.3%, 18.7%)) groups around trials 25 to 30 (Figures 3.6(c) and (f)). In particular, there is a larger sample space covered by the CPEs for the ASD group compared to the TD group, indicating more heterogeneity between subjects in the variability of condition differentiation within the ASD group as compared to the TD group.

Decomposition of the functional marginal covariance over ERP time provides insight into within-trial P3 peak differences between the ASD and TD groups. The three leading func-

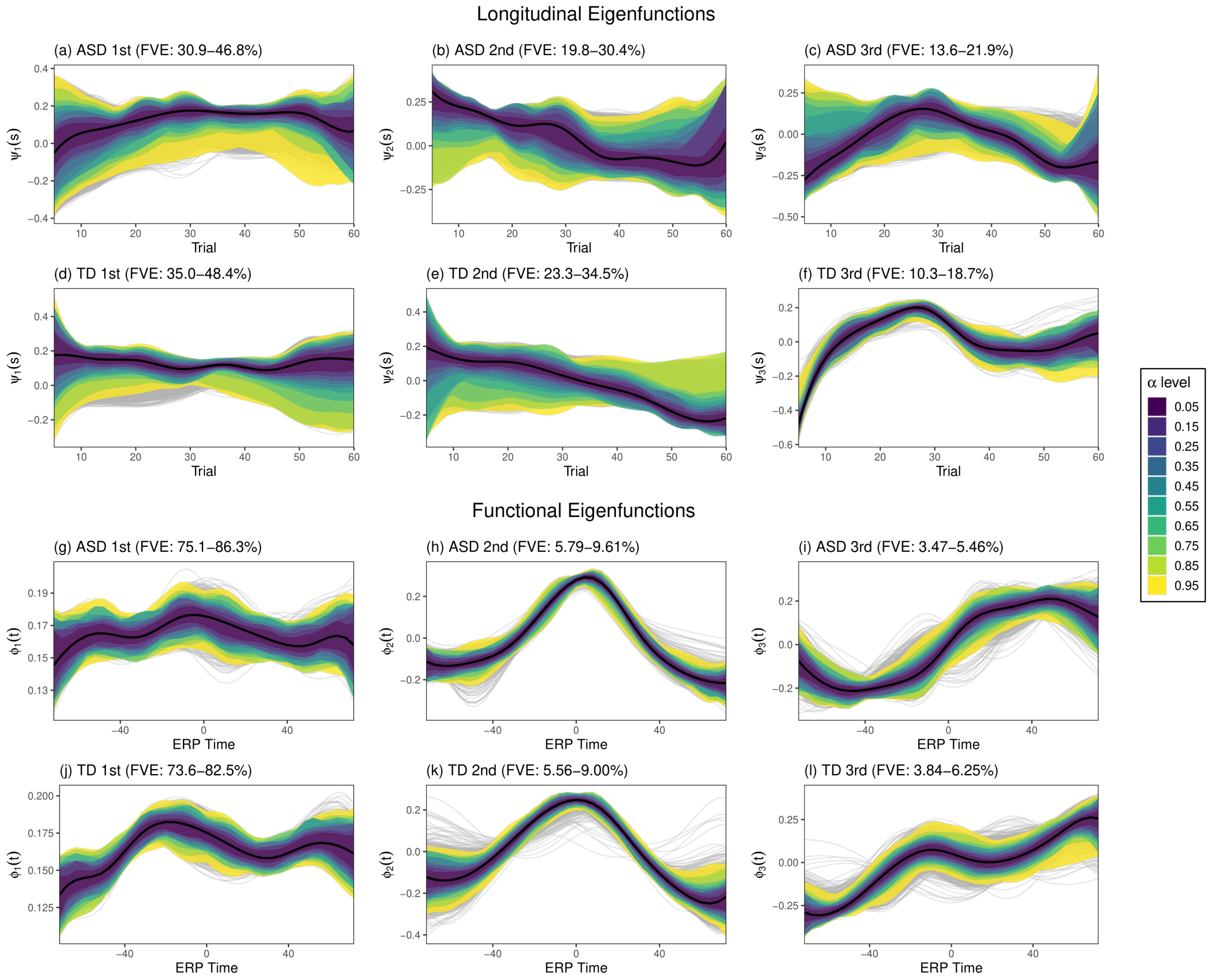


Figure 3.6: MBD-CPE contours of the three leading longitudinal and functional eigenfunctions for ASD and TD displayed from the left- to right-hand columns at a grid of α levels marked by varying contour colors, overlaying the posterior estimates given in gray. The first and second rows display the MBD-CPEs of the longitudinal eigenfunctions, and the third and fourth rows display the MBD-CPEs of the functional eigenfunctions. The MBD medians are given in solid black.

tional eigencomponents explain more than 90% of the total variation in both groups (median and (2.5th, 97.5th) percentiles of FVE at 91.4% (89.6%, 93.2%) and 90.2% (88.2%, 92.2%) in the ASD and TD groups, respectively). For both the ASD and TD groups, the leading functional eigenfunction captures overall variability around the P3 peak (Figure 3.6(g) and (j)). The second leading functional eigenfunction represents variation in the P3 peak amplitude (Figure 3.6(h) and (k)). Lastly, the third leading functional eigenfunction characterizes variation in the shape of the P3 peak around 70ms before and after (Figure 3.6(i) and (ℓ)).

Figure C.12 displays the 95% simultaneous parametric credible interval given as dashed black lines and pointwise mean as the solid black line for the leading three longitudinal (rows 1 and 2) and functional (row 3 and 4) eigenfunctions for both diagnostic groups in Figure 3.6. For the longitudinal eigenfunctions, the 95% credible intervals are slightly wider than the 95% MBD-CPEs (Figures C.12(a)-(e)). Additionally, the MBD-CPEs aid in the interpretation of the longitudinal eigenfunctions by adding a more precise look into the posterior distributions based on the contours that is lacking in the 95% credible intervals. The 95% credible intervals of the functional eigenfunctions appear to cover around the same area as the 95% α-level contours of the MBD-CPEs, likely due to smaller variation in the posterior samples, and both lead to the same interpretation of the functional eigenfunctions. Overall the CPEs provide a more granular look at the behavior of the eigenfunctions due to the grid of α-level contours as only a single 95% credible interval is typically reported. If instead the $(1-\alpha)\%$ credible intervals were reported at a grid of α-level contours similar to the CPEs, they would be fully nested within each other but with no variation in shape, unlike the CPEs which more flexibly capture shape differences at varying contours.

3.6 Discussion

We expand upon the descriptive tools to visualize variation in the posterior samples proposed in Boland et al. (2023) to a higher multidimensional framework for longitudinal and functional model components in B-LFPCA. Relying on the work in Shamshoian et al. (2022) for the development of the computationally efficient, accurate, and data-adaptive model for B-LFPCA, we obtain an estimation framework for direct inference on model components under the assumption of weak separability of the longitudinal and functional dimensions. Through the implementation of B-LFPCA, we recover highly interpretable marginal eigenfunctions as well as the mean surface, marginal covariances, and covariance kernel and propose depth-based summaries for these quantities. The proposed summaries lead to a data-driven approach in capturing variability in the longitudinal and functional model components, particularly the marginal eigenfunctions, without applying symmetry constraints or relying on estimation methods that fail to treat the posterior samples as truly functional data. In particular, the proposed summaries are based on the rankings of the posterior samples using functional depth for single-dimensional functions as well as extensions of functional depth up to four-dimensional surfaces. Three functional depth-based summaries are considered: one based on the direct ranking of the entire sample of longitudinal and functional eigenfunctions, another through the ranking of the associated two-dimensional longitudinal and functional marginal covariances, and lastly through the ranking of the four-dimensional covariance kernels. All three approaches were shown to lead to flexible modeling of the variation in the posterior samples, with increasingly conservative/wider CPEs when more variation is incorporated. More specifically, capturing variation from a sample of a single marginal eigenfunction (MBD-CPEs) to greater variation captured in covariance processes a single dimension (MVD-CPEs) to incorporating variation from both functional and longitudinal dimensions (kernel MVD-CPEs). Although, it was shown in simulations that if the assumption of weak separability is violated, this trend of increasingly wider CPEs may not be consistent.

CPEs in the multidimensional setting in application to EEG experiments appear to be a useful tool in visualizing uncertainty in the posterior estimates of the data in FPCA-based models, where the weak separability of the dimensions is assumed and holds. Adding the use of CPEs in the implicit learning paradigm enriched the results of the analysis and lead to more flexible and detailed visual representations of learning trends between the two groups that were not captured using solely traditional credible intervals for uncertainty quantification. Given the utility of CPEs based on our analysis, further research into applying depth measures for more complicated object data, such as the work by Dai et al. (2022), is of interest, particularly for higher dimensional objects and surfaces. In our proposed CPEs, we faced limitations when visualizing the MVD-CPEs of the mean surface, and as a solution discretized either the longitudinal or functional dimensions and obtained two-dimensional representations of a three-dimensional surface, which does not treat the surface as a truly functional estimate. Therefore, a visualization tool in higher dimensions is needed and an interesting open problem for future research.

APPENDIX A

Chapter 1: Appendices

A.1 Eigenvectors of marginal covariances from subgroups

To justify merging of TFT vectors (across diagnostic groups, conditions and scalp regions) in obtaining trial-specific covariances, leading the the marginal covariance defined in Section 2.3, we assessed whether eigenvectors of the marginal covariances from the eight subgroups were sufficiently similar. The leading six eigenvectors are displayed in Figures A.1-A.6 for the delta frequency band and Figures A.7-A.12 for the theta frequency band. We find that the estimated eigenvectors (i.e. directions of variation) are sufficiently similar across the eight TFT groupings, which justifies estimation of the marginal covariance by merging data in our data application.

A.2 Eigenvectors of trial-specific covariances

To assess the constant direction of functional variation across trials of the longitudinal dimension (an assumption needed in defining the marginal covariance as an average of trial-specific covariances in Section 2.3), we include plots of the estimated six leading eigenvectors from trial-specific covariances defined at trials $20, 40$ and 60 for analysis of the delta and theta frequency bands in Figures A.13-A.18. Similarity of eigenvectors signal no violation of the assumption of constant variation across trials in our data application.

A.3 Histograms of MDPCA scores

For mixed effects modeling of the eigenscores, we provide in Figure A.19 the histograms of the eigenscores, displaying relatively symmetric distributions in our data application, signaling no violation of the normality assumption.

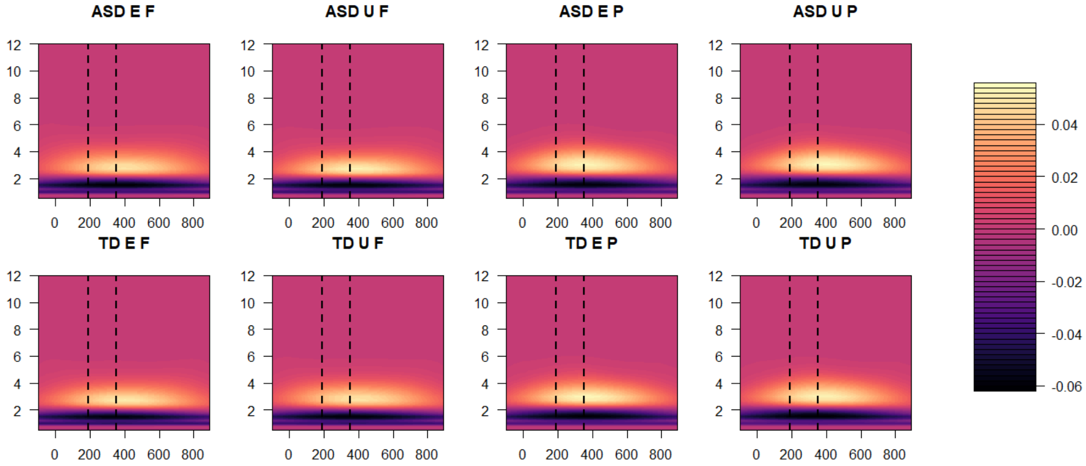


Figure A.1: Estimated PC1 for the marginal covariances in the delta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

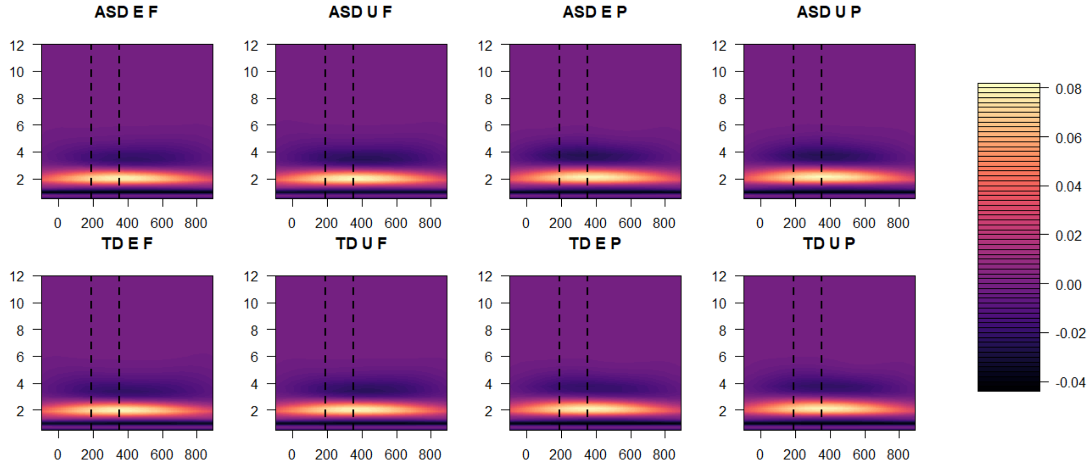


Figure A.2: Estimated PC2 for the marginal covariances in the delta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

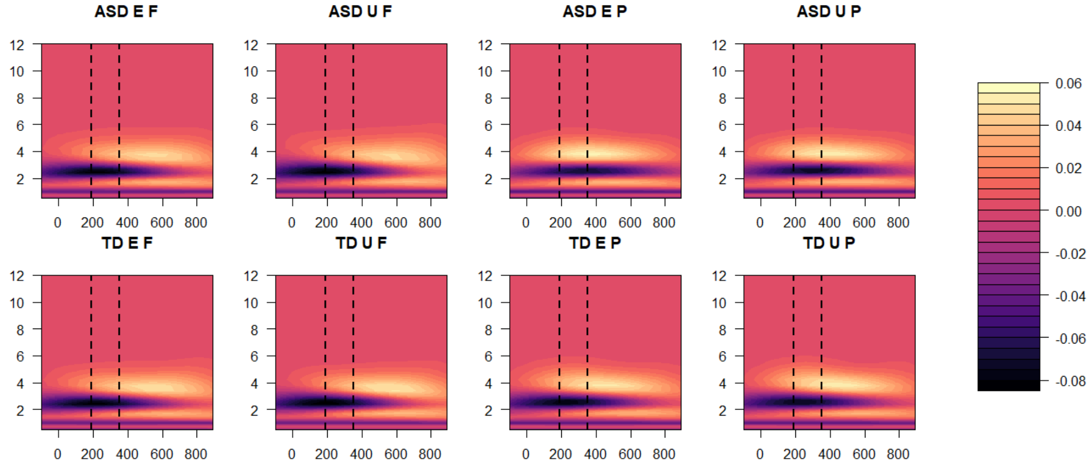


Figure A.3: Estimated PC3 for the marginal covariances in the delta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

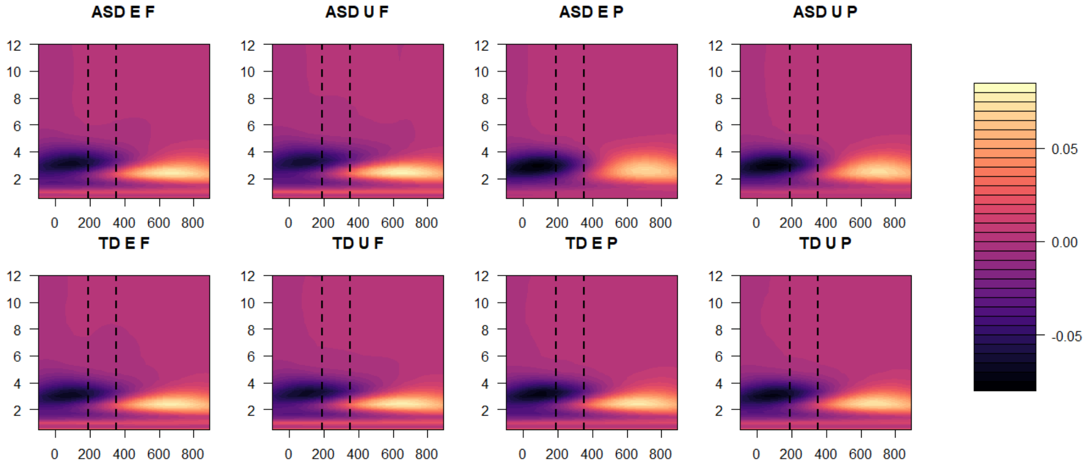


Figure A.4: Estimated PC4 for the marginal covariances in the delta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

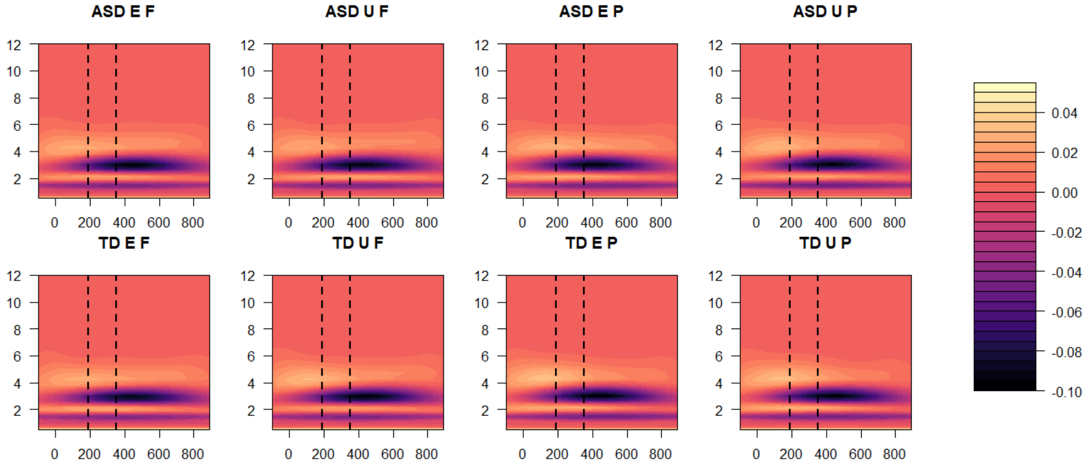


Figure A.5: Estimated PC5 for the marginal covariances in the delta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

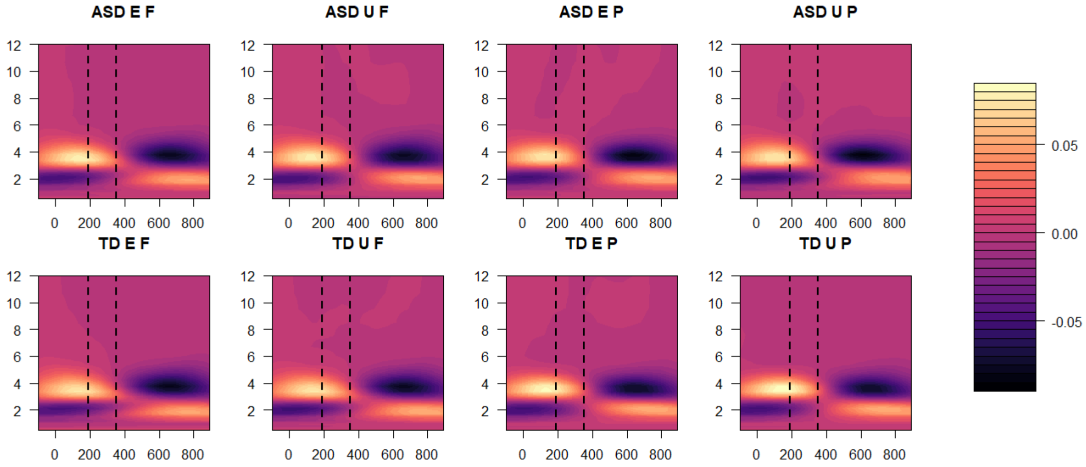


Figure A.6: Estimated PC6 for the marginal covariances in the delta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

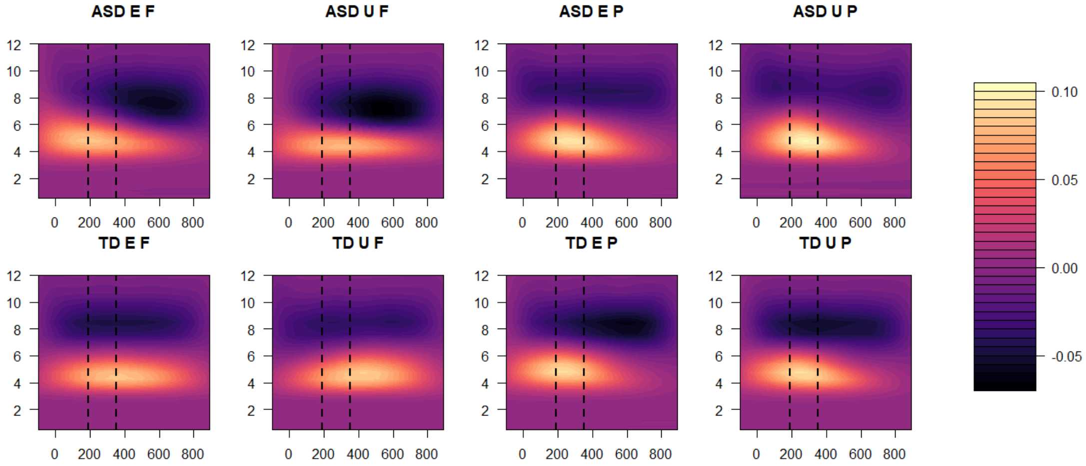


Figure A.7: Estimated PC1 for the marginal covariances in the theta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

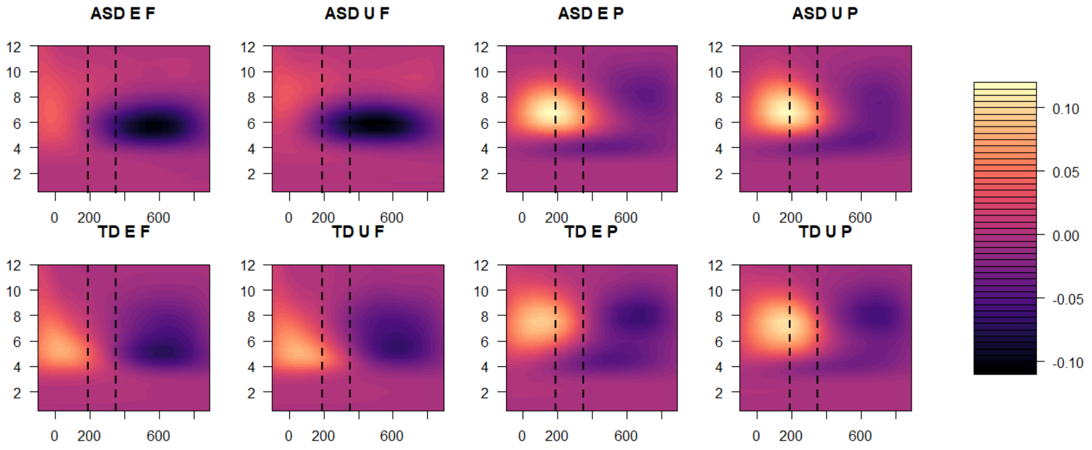


Figure A.8: Estimated PC2 for the marginal covariances in the theta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

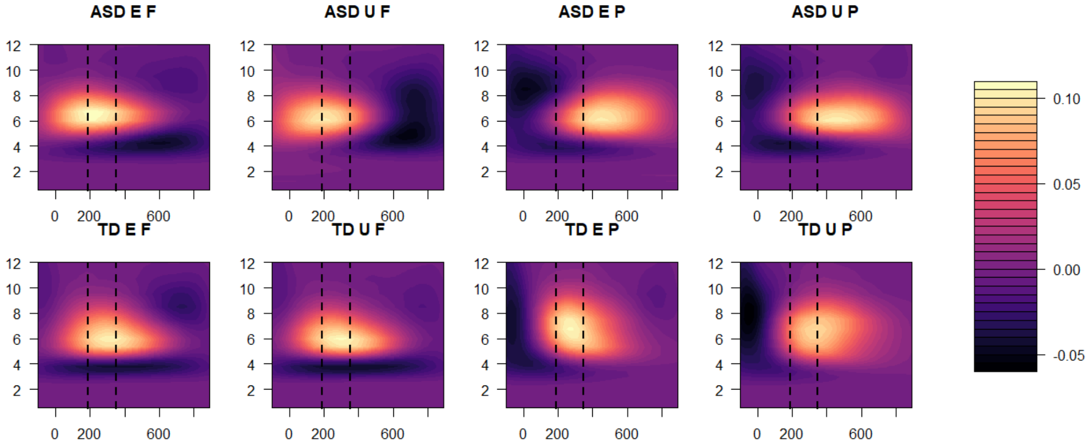


Figure A.9: Estimated PC3 for the marginal covariances in the theta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

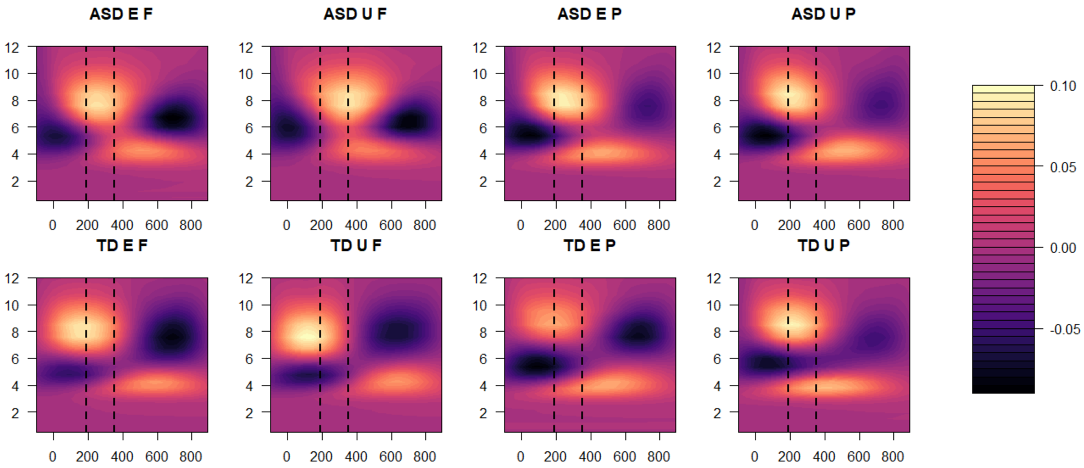


Figure A.10: Estimated PC4 for the marginal covariances in the theta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

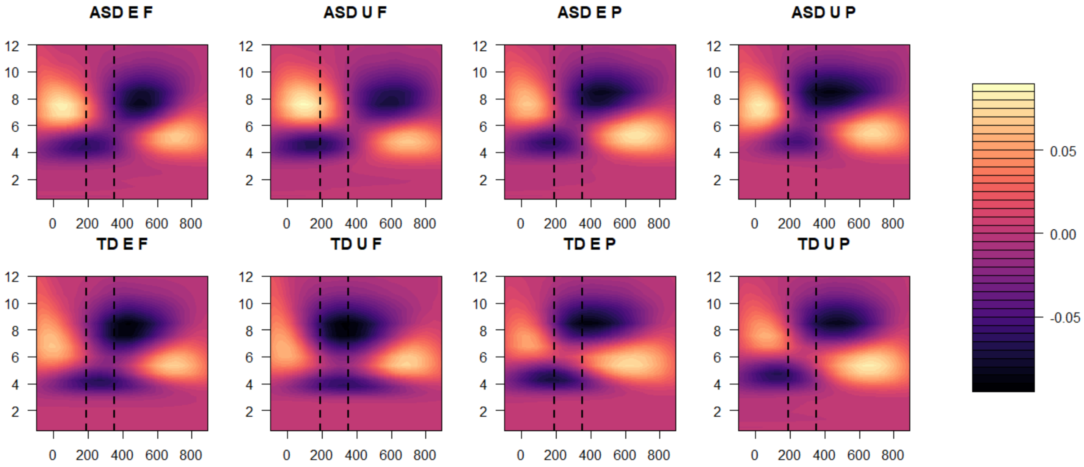


Figure A.11: Estimated PC5 for the marginal covariances in the theta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

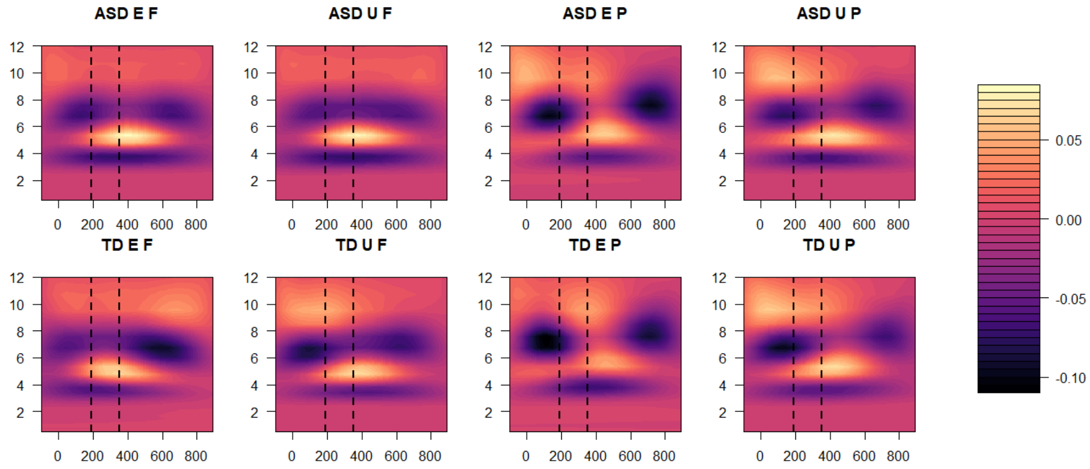


Figure A.12: Estimated PC6 for the marginal covariances in the theta frequency band in two diagnostic groups (ASD, TD), two conditions (expected (E), unexpected (U)) and two scalp sections (frontal (F) and (posterior (P)).

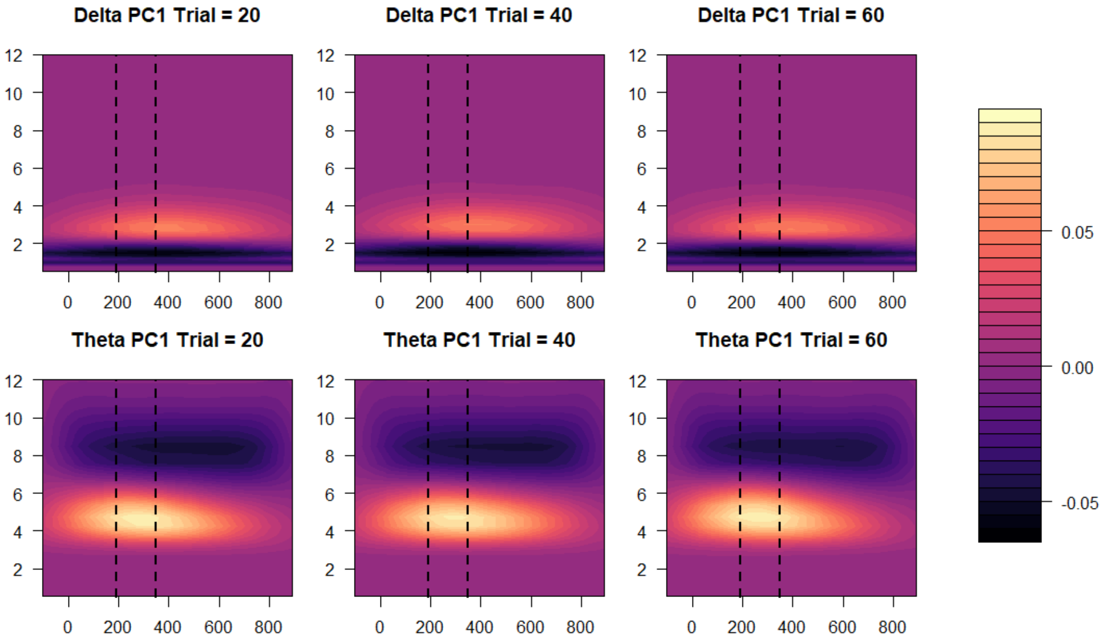


Figure A.13: The estimated PC1 of trial-specific covariances for both the delta and theta frequency bands at trials $s = 20, 40$, and 60.

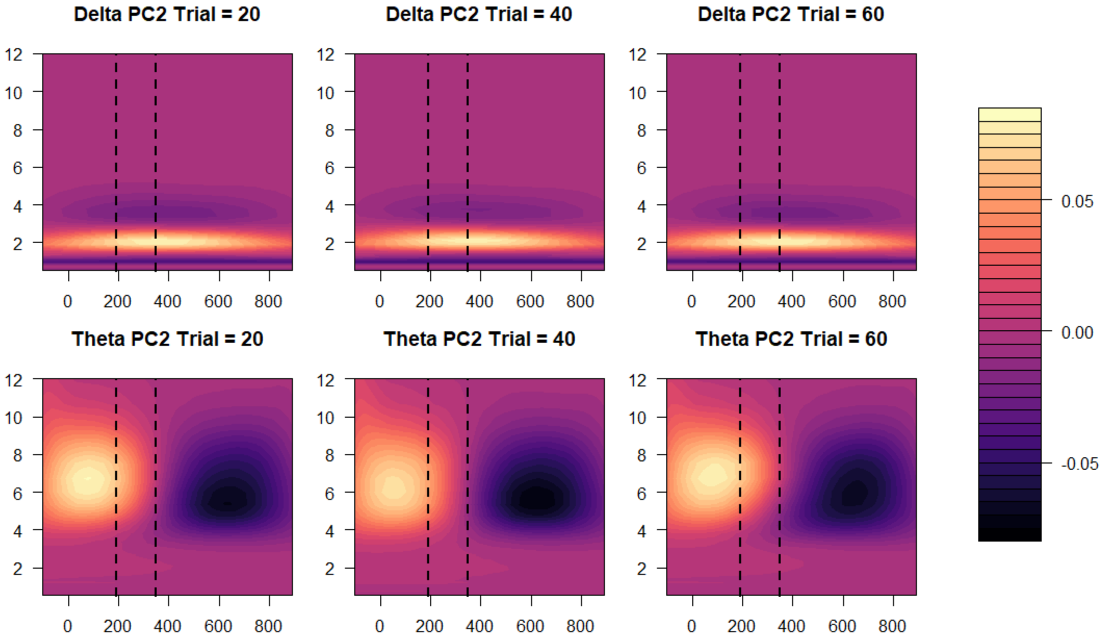


Figure A.14: The estimated PC2 of trial-specific covariances for both the delta and theta frequency bands at trials $s = 20, 40$, and 60.

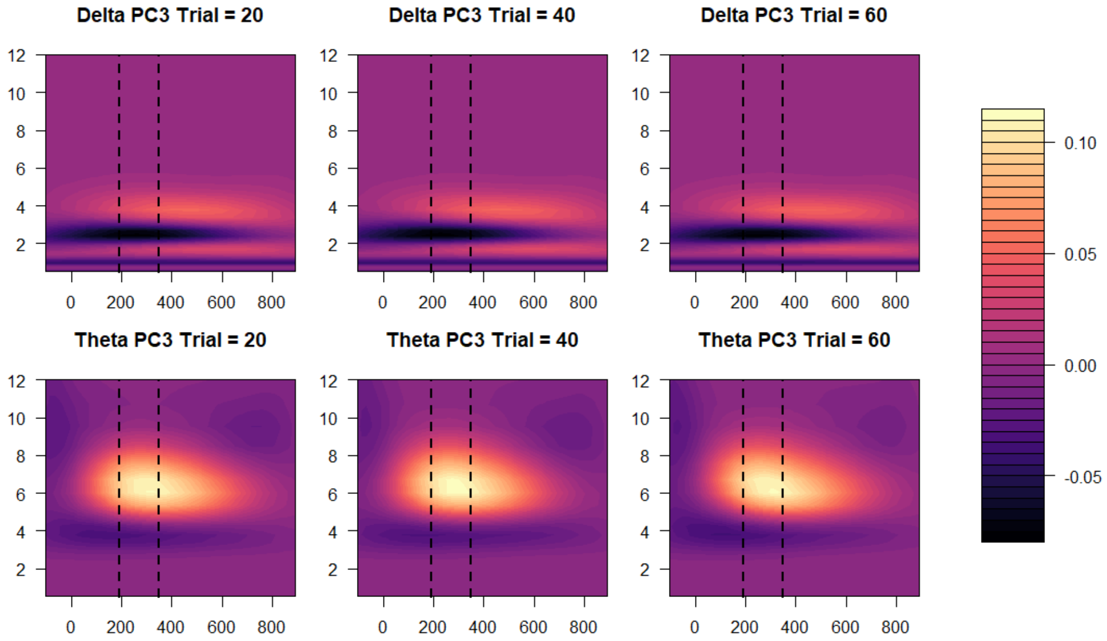


Figure A.15: The estimated PC3 of trial-specific covariances for both the delta and theta frequency bands at trials $s = 20, 40$, and 60.

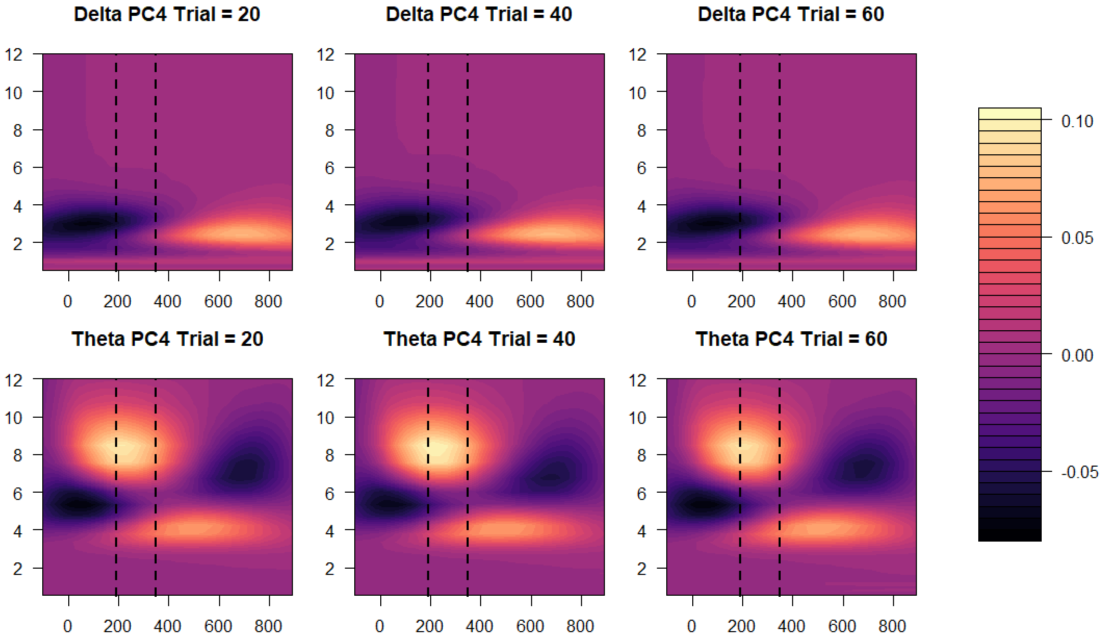


Figure A.16: The estimated PC4 of trial-specific covariances for both the delta and theta frequency bands at trials $s = 20, 40$, and 60.

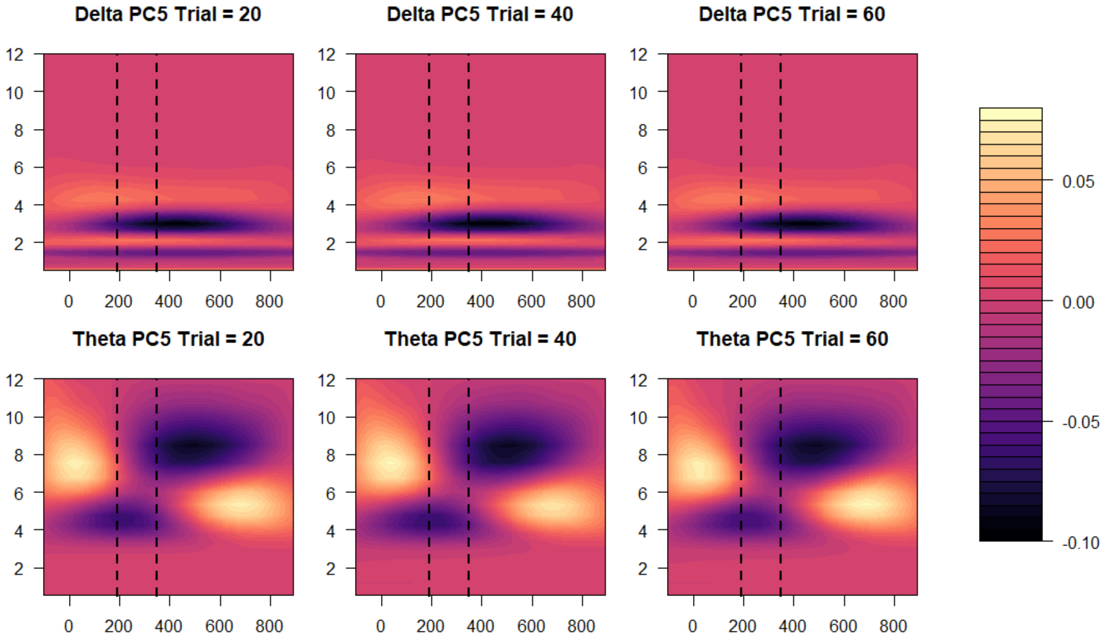


Figure A.17: The estimated PC5 of trial-specific covariances for both the delta and theta frequency bands at trials $s = 20, 40$, and 60.

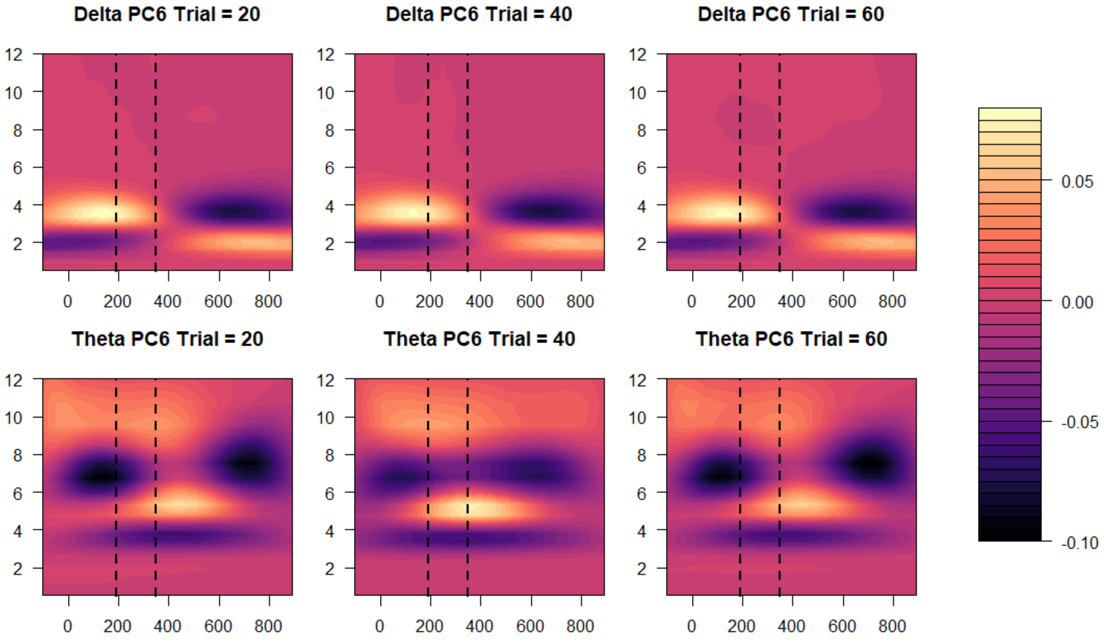


Figure A.18: The estimated PC6 of trial-specific covariances for both the delta and theta frequency bands at trials $s = 20, 40$, and 60.

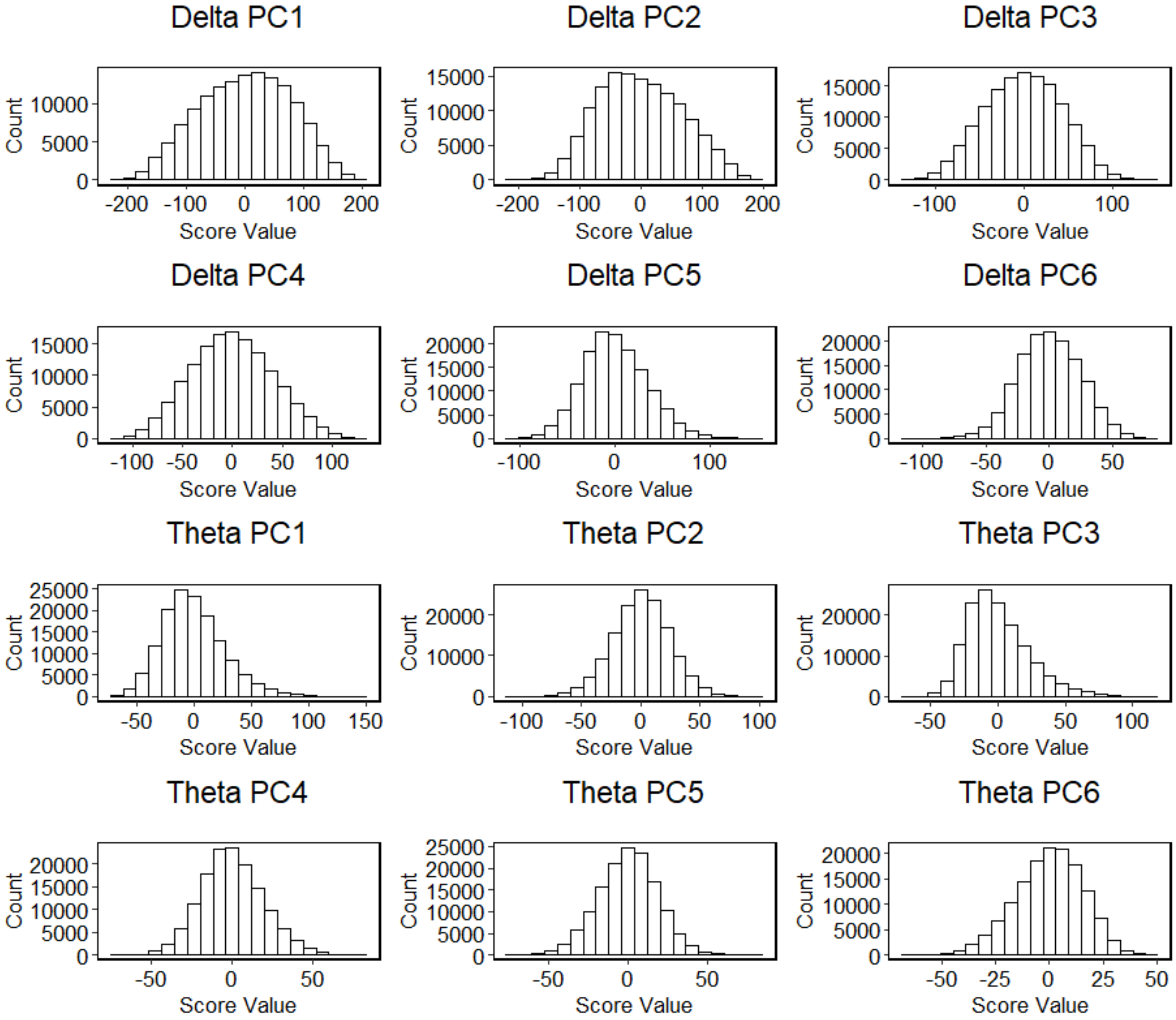


Figure A.19: Histograms of the estimated MDPCA scores corresponding to the leading six eigencomponents for analysis in the delta and theta frequency bands in the top two rows and bottom two rows, respectively.

APPENDIX B

Chapter 2 Appendices

B.1 Posterior Distributions for BFPCA

Using Gaussian priors for the mean coefficients $\boldsymbol{\beta}$ and factor loadings $\boldsymbol{\lambda}_\ell$, noninformative prior for the error variance σ_ϵ^2 (proportional to a constant c), gamma prior for the variance of the mean coefficients σ_β^2 and a modified multiplicative gamma process shrinkage (MMGPS) prior for the variance components of the factor loading matrix $\sigma_{\lambda_{r\ell}}^2$, the model can be given in matrix form as:

$$
\begin{aligned}
&\boldsymbol{Y}_i = \boldsymbol{f}_i + \boldsymbol{\epsilon}_i = B(\boldsymbol{\beta} + \Lambda\boldsymbol{\eta}_i) + \boldsymbol{\epsilon}_i, \\
&\boldsymbol{\eta}_i \sim N_L(\mathbf{0}_L, I_L), \quad \boldsymbol{\epsilon}_i \sim N_T(\mathbf{0}_T, \sigma_\epsilon^2 I_T), \quad i = 1, \ldots, n, \\
&\boldsymbol{\beta} \sim N_R\left(\mathbf{0}_R, \frac{1}{\sigma_\beta^2}\Omega^{-1}\right), \quad \sigma_\beta^2 \sim \text{Gamma}\left(\frac{a_\beta}{2}, \frac{a_\beta}{2}\right), \quad \frac{1}{\sigma_\epsilon^2} \propto c \\
&\boldsymbol{\lambda}_\ell \sim N_R\left(\mathbf{0}_R, \Sigma_{\lambda_\ell}\right), \quad \Sigma_{\lambda_\ell} = \text{diag}(\sigma_{\lambda_{1\ell}}^2, \ldots, \sigma_{\lambda_{R\ell}}^2), \quad \sigma_{\lambda_{r\ell}}^2 = \varphi_{r\ell}^{-1}\tau_\ell^{-1}, \\
&\varphi_{r\ell} \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \quad \tau_\ell = \prod_{h=1}^{\ell} \delta_h, \quad \delta_1 \sim \text{Gamma}(a_1, 1), \\
&\delta_h \sim \text{Gamma}(a_2, 1) I(\delta_h > 1), \quad h \geqslant 2.
\end{aligned}
$$

Let $\widetilde{\boldsymbol{\Lambda}} = \text{Vec}(\Lambda)$ be the $RL \times 1$ vector stacking the L columns of Λ and $H_\epsilon = 1/\sigma_\epsilon^2$. The full conditional distributions are as follows:

1. $\boldsymbol{\beta}|\text{others} \sim N_R(\boldsymbol{\mu}_\beta^{post}, v_\beta^{post})$ where $v_\beta^{post} = (1/n)\{B^\top \Sigma_y^{-1} B + (\sigma_\beta^2/n)\Omega\}^{-1}$, $\boldsymbol{\mu}_\beta^{post} = v_\beta^{post} \times$

$nB^\top \Sigma_y^{-1}\overline{\boldsymbol{Y}}$, where $\Sigma_y^{-1} = B\Lambda\Lambda^\top B^\top + \sigma_\epsilon^2 I_T$ and $\overline{\boldsymbol{Y}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{Y}_i$.

2. $\widetilde{\boldsymbol{\Lambda}}|\text{others} \sim N_{RL}(\boldsymbol{\mu}_\Lambda^{post}, v_\Lambda^{post})$ where $v_\Lambda^{post} = \sigma_\epsilon^2\{(\sum_{i=1}^n \boldsymbol{\eta}_i\boldsymbol{\eta}_i^\top)\otimes B^\top B + \sigma_\epsilon^2\Sigma_\Lambda^{-1}\}^{-1}$, $\boldsymbol{\mu}_\Lambda^{post} = v_\Lambda^{post} \times (1/\sigma_\epsilon^2)\sum_{i=1}^n(\boldsymbol{\eta}_i^\top \otimes B)^\top \boldsymbol{Y}_i^c$, where $\otimes$ is the Kronecker product, Σ_Λ is a diagonal matrix, denoted as $\Sigma_\Lambda = \text{diag}(\varphi_{11}^{-1}\tau_1^{-1}, \ldots, \varphi_{R1}^{-1}\tau_1^{-1}, \ldots, \varphi_{1L}^{-1}\tau_L^{-1}, \ldots, \varphi_{RL}^{-1}\tau_L^{-1})$, and $\boldsymbol{Y}_i^c = \boldsymbol{Y}_i - B\boldsymbol{\beta}$.

3. $H_\epsilon|\text{others} \sim \text{Gamma}(a_{\sigma_\epsilon^2}^{post}, b_{\sigma_\epsilon^2}^{post})$, where $a_{\sigma_\epsilon^2}^{post} = (nT)/2$, $b_{\sigma_\epsilon^2}^{post} = RSS/2$, and $RSS = \sum_{i=1}^n\{\boldsymbol{Y}_i - B(\boldsymbol{\beta} + \Lambda\boldsymbol{\eta}_i)\}^\top\{\boldsymbol{Y}_i - B(\boldsymbol{\beta} + \Lambda\boldsymbol{\eta}_i)\}$.

4. $\boldsymbol{\eta}_i|\text{others} \sim N_L(\boldsymbol{\mu}_{\eta_i}^{post}, v_{\eta_i}^{post})$, where $v_{\eta_i}^{post} = \sigma_\epsilon^2(\Lambda^\top B^\top B\Lambda + \sigma_\epsilon^2 I_L)^{-1}$, $\boldsymbol{\mu}_{\eta_i}^{post} = v_{\eta_i}^{post} \times (1/\sigma_\epsilon^2)\Lambda^\top B^\top(\boldsymbol{Y}_i - B\boldsymbol{\beta})$, for $i = 1, \ldots, n$.

5. $\sigma_\beta^2|\text{others} \sim \text{Gamma}(a_{\sigma_\beta^2}^{post}, b_{\sigma_\beta^2}^{post})$, where $a_{\sigma_\beta^2}^{post} = (R + a_\beta)/2$, $b_{\sigma_\beta^2}^{post} = (a_\beta + \boldsymbol{\beta}^\top\Omega\boldsymbol{\beta})/2$

6. $\varphi_{r\ell}|\text{others} \sim \text{Gamma}(a_{\varphi_{r\ell}}^{post}, b_{\varphi_{r\ell}}^{post})$, where $a_{\varphi_{r\ell}}^{post} = (\nu + 1)/2$, $b_{\varphi_{r\ell}}^{post} = (\nu + \tau_\ell\lambda_{r\ell}^2)/2$.

7. $\tau_\ell|\text{others} = \prod_{h=1}^\ell \delta_h|\text{others}$

   (a) $\delta_1|\text{others} \sim \text{Gamma}(a_{\delta_1}^{post}, b_{\delta_1}^{post})$, where $a_{\delta_1}^{post} = a_1 + (1/2)RL$, $b_{\delta_1}^{post} = 1 + (\gamma_1 + \sum_{\ell=2}^L \zeta_\ell^{(1)}\gamma_\ell)/2$, where $\zeta_\ell^{(1)} = \prod_{h=2}^\ell \delta_h$ and $\gamma_\ell = \sum_{r=1}^R \varphi_{r\ell}\lambda_{r\ell}^2$.

   (b) $\delta_h|\text{others} \sim \text{Gamma}(a_{\delta_h}^{post}, b_{\delta_h}^{post})I\,(\delta_h \geqslant 1)$, $h \geqslant 2$, where $a_{\delta_h}^{post} = a_2 + R(L-h+1)/2$, $b_{\delta_h}^{post} = 1 + (\gamma_h + \sum_{\ell=(h+1)}^L \zeta_\ell^{(h)}\gamma_\ell)/2$, and $\zeta_\ell^{(h)} = \prod_{h'\neq h}\delta_{h'}$.

A Gibbs sampler is used to sample from the posterior distributions given above. For the variance component σ_β^2, the hyperparameter for the prior is set to $a_\beta = 2$. For the factor loading matrix, the hyperparameter of the prior for $\varphi_{r\ell}$ is set to $\nu = 10$, and the hyperparameters of the priors for τ_ℓ are set to $a_1 = 1$ and $a_2 = 2$. A small constant $\varsigma = 0.00001$ is added to the diagonal elements of the penalty matrix Ω to guarantee positive-definiteness, where Ω is an $R \times R$ matrix with elements

$$\Omega = \begin{bmatrix} 1 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & \dots & \dots & 0 & -1 & 1 \end{bmatrix} + \varsigma I_R,$$

and I_R denotes the $R \times R$ identity matrix.

B.2 Alignment of Eigenfunction Estimates

Let $\psi_k^{(m)}(t)$ denote the mth posterior sample, $m = 1, \dots, M$, of the kth eigenfunction, $\psi_k(t)$, $k = 1, \dots, K$. To align the sign of the eigenfunction estimates across the MCMC samples, we utilize the below alignment algorithm. In the proposed alignment algorithm, $\{\psi_k^{(1)\star}(t), \dots, \psi_k^{(M)\star}(t)\}$ denotes the aligned posterior sample with $\psi_k^{(m)\star}(t) = a^{(m)}\psi_k^{(m)}(t)$ and $a^{(m)} \in \{-1, 1\}$, and $\overline{\psi}_k^{(m)}(t) = (1/m)\sum_{j=1}^{m} \psi_k^{(j)\star}(t)$ denotes the ergodic mean of the aligned sample.

Algorithm 5 Alignment of the posterior eigenfunctions

```
Step 1: Set $a^{(1)} = 1$ and $\psi_k^{(1)^\star} = \psi_k^{(1)}$.
Step 2: For $m = 2, \dots, M$;
    a. Calculate the aligned ergodic mean $\overline{\psi}_k^{(m-1)}(t)$ and compute
           $d_+^{(m)} := \int \left| \overline{\psi}_k^{(m-1)}(t) - \psi_k^{(m)}(t) \right| dt,$
       and
           $d_-^{(m)} := \int \left| \overline{\psi}_k^{(m-1)}(t) + \psi_k^{(m)}(t) \right| dt.$
    b. Set
           $a^{(m)} = I\left(d_+^{(m)} \leqslant d_-^{(m)}\right) - I\left(d_+^{(m)} > d_-^{(m)}\right),$
       and $\psi_k^{(m)^\star}(t) = a^{(m)}\psi_k^{(m)}(t)$, where $I(\cdot)$ denotes the indicator function.
```

B.3 Simulation Cases

Five simulation scenarios are considered to display the use of CPEs in describing the variation in the posterior samples in the presence of functional outliers (results given in main chapter Section 2.4).

Case 1 (no outliers): For case 1 with no outliers, a sample of functional data are generated according to the FPCA model with $K = 2$ eigencomponents. More specifically, functional data is generated according to $Y_i(t_j) = \mu(t_j) + \sum_{k=1}^{K} \xi_{ik}\psi_k(t_j) + \epsilon_i(t_j)$, for $i = 1, \dots, n = 50$ subjects at an equidistant grid of $j = 1, \dots, T = 40$ time points in the unit interval 0 to 1. The mean function and the two mutually orthonormal eigenfunctions equal $\mu(t) = 10\sqrt{1 - 2(t - 0.5)^2}$, $\psi_1(t) = \sqrt{2}\sin(2\pi t)$ and $\psi_2(t) = \sqrt{2}\cos(2\pi t)$, respectively. The subject-specific FPCA scores, $(\xi_{i1}, \xi_{i2})^\top$, are generated from independent normal distributions with mean zero and variances $\rho_1 = 15$ and $\rho_2 = 5$, respectively. Lastly, the measurement error, $\epsilon_i(t_j)$, is generated independently from $N(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2 = 15$.

Case 2 (magnitude outlier): Case 3 generates magnitude outliers by adding a constant deviation to the mean function (with a random sign): $Z_i(t) = Y_i(t) + \omega_i W(t)$ for $t > T_i$ and

$Z_i(t) = Y_i(t)$ for $t < T_i$, where $T_i \sim \text{Unif}[0,1]$, $W(t) = 20, t \in [0,1]$, and ω_i is generated as a discrete variable with values -1 or 1 with probability 1/2.

Case 3 (amount of variation outlier): Case 3 generates amount of variation outliers by using larger eigenvalues: $Z_i(t) = \mu(t) + \sum_{k=1}^{2} \zeta_{ik}\psi_k(t) + \epsilon_i(t)$, where ζ_{i1} and ζ_{i2} are generated independently from $N(0,30)$ and $N(0,20)$, respectively.

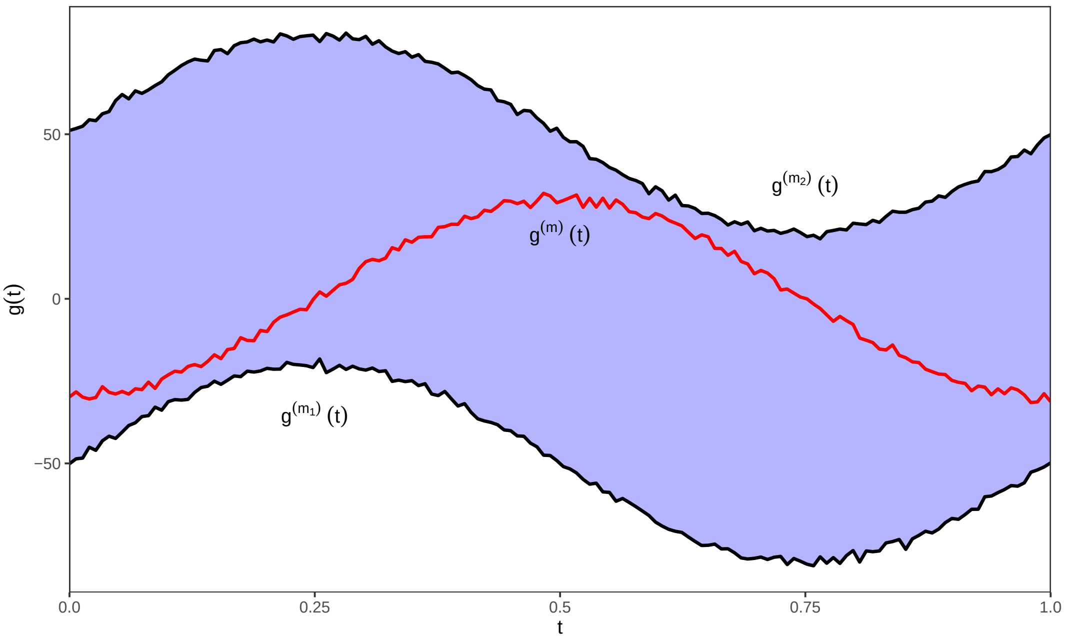
Case 4 (time-shifted eigenfunction outlier): Case 4 generates time-shifted eigenfunction outliers: $Z_i(t) = \mu(t) + \sum_{k=1}^{2} \xi_{ik}\kappa_k(t) + \epsilon_i(t)$, where $\kappa_1(t) = \sqrt{2}\sin\{2\pi(t-0.25)\}$ and $\kappa_2(t) = \sqrt{2}\cos\{2\pi(t-0.25)\}$.

Case 5 (higher-frequency eigenfunction outlier): For case 5, higher-frequency eigenfunction outliers are generated according to $Z_i(t) = \mu(t) + \sum_{k=1}^{2} \xi_{ik}\kappa_k(t) + \epsilon_i(t)$, where $\kappa_1(t) = \sqrt{2}\sin\{4\pi t\}$ and $\kappa_2(t) = \sqrt{2}\cos\{4\pi t\}$.

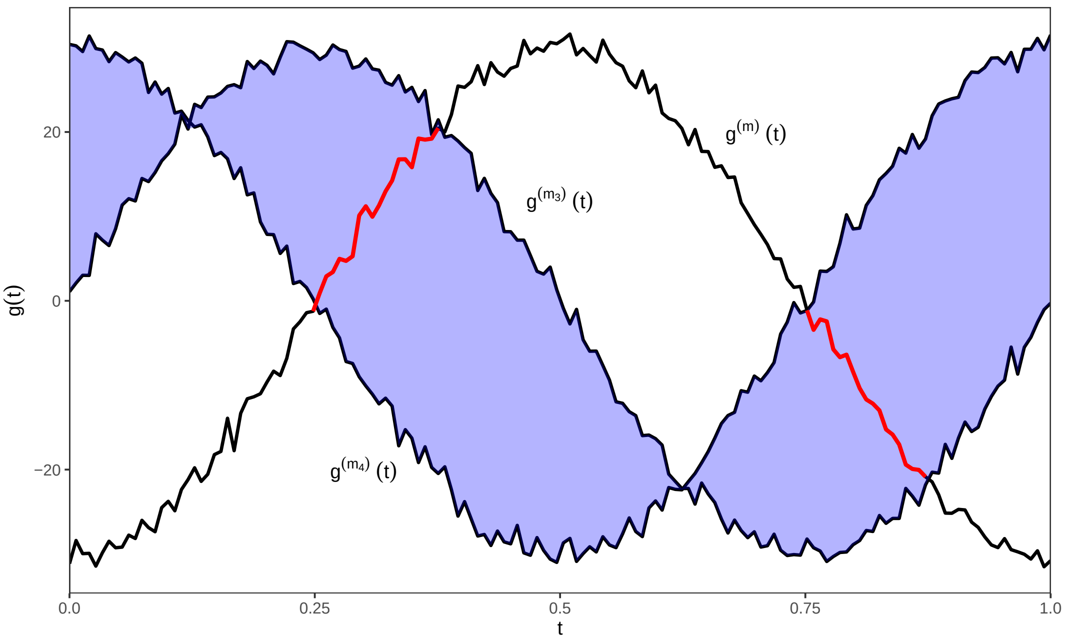
In cases 2 through 5, the percent of outliers generated equal $q = 10$ or 20% of the functional sample.

B.4 EEG Data

Our motivating study collected electroencephalogram (EEG) data sampled at 500Hz for 2 minutes using a 128-channel HydroCel Geodesic Sensor Net on 58 children with autism spectrum disorder (ASD) and 39 of their typically developed (TD) peers. Four electrodes near the eyes were removed prior to recording to improve the comfort of the participants. The data was then interpolated to the international 10-20 system 25 channel montage via spherical interpolation, and independent component analysis (ICA) was used for identification of artifacts. Specifically, the EEG signals were reconstructed without components attributed to nonneural sources of the signals, such as the electromyogram (EMG) or other non-stereotyped artifacts, and then re-referenced to an average of all channels. The first 38 seconds of the artifact-free EEG data was used for each subject for spectral power analysis as the 38 seconds of the recordings represented the minimum amount of artifact-free data available across all participants and was deemed an appropriate minimum threshold to gain reliable estimates of the signals following previous literature. Spectral density estimates of the 38 second EEG recordings were obtained using Welch's method by dividing the data into 2-second Hanning windows with 50% overlap and transforming into the frequency domain via a fast Fourier transformation (FFT). For each electrode, the spectral densities were averaged at each overlapping segment, resulting in electrode-specific estimates of the spectral density, which were then averaged across the 25 electrodes to obtain scalp-wide spectral densities for each participant.



(a) Band delineated by $g^{(m_1)}(t)$ and $g^{(m_2)}(t)$



(b) Band delineated by $g^{(m_3)}(t)$ and $g^{(m_4)}(t)$

Figure B.1: The bands given, $\mathcal{B}\left\{g^{(1)}(t), g^{(2)}(t)\right\}$ and $\mathcal{B}\left\{g^{(3)}(t), g^{(4)}(t)\right\}$, are represented as the blue shaded region. The proportion of the curve $g^{(m)}(t)$ that lies within the respective bands is given in red, where $\mathcal{A}_2^*\left\{g^{(m)}(t); g^{(1)}(t), g^{(2)}(t)\right\} = 1$ and $\mathcal{A}_2^*\left\{g^{(m)}(t); g^{(3)}(t), g^{(4)}(t)\right\} = 0.253$.

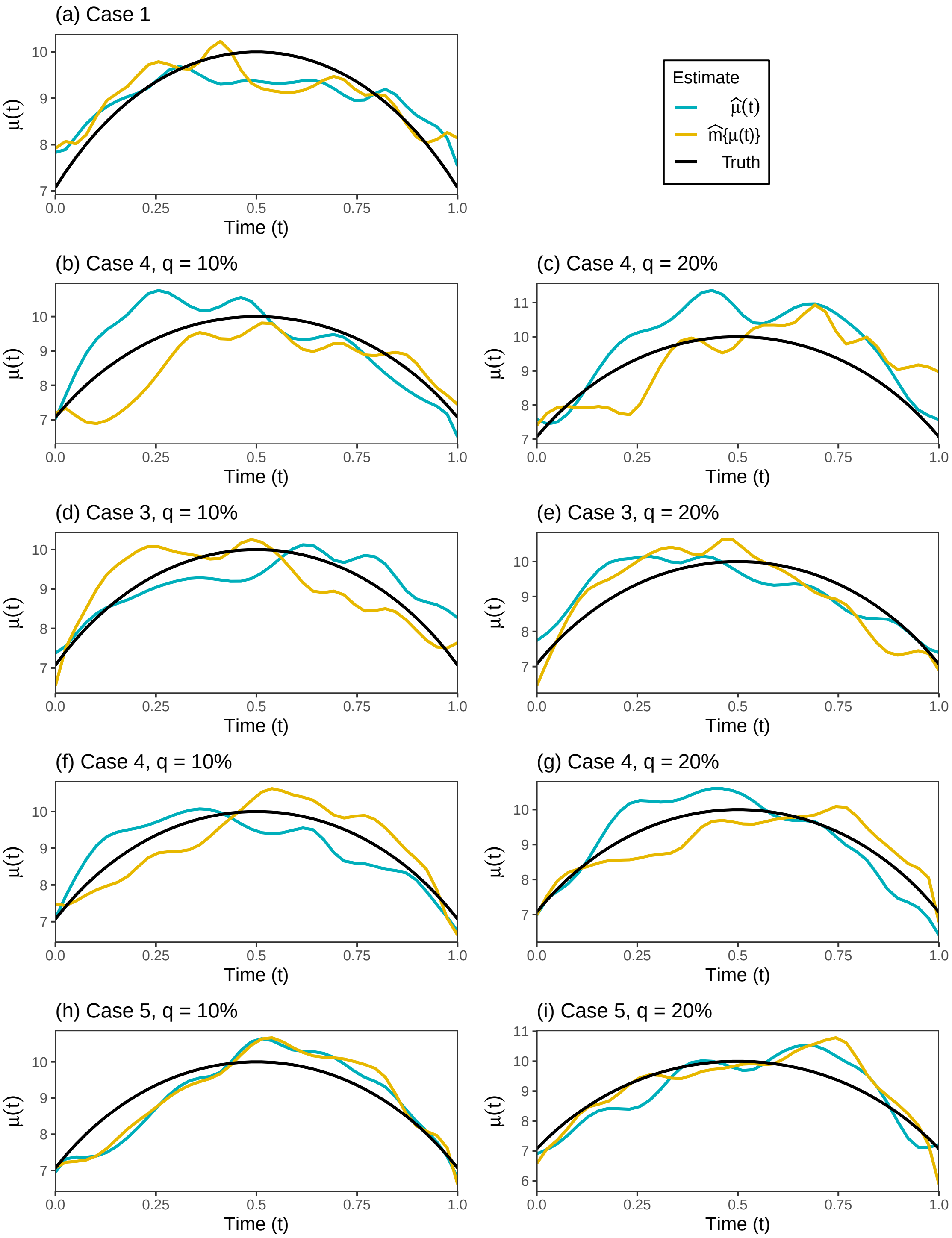


Figure B.2: Point estimates of $\mu(t)$ for each simulation case with $q = 10$ and $q = 20\%$ outliers from runs with 50th percentile IMSE values. Mean estimate and MBD median are given in blue and solid yellow, respectively, overlaying the true function given in black.

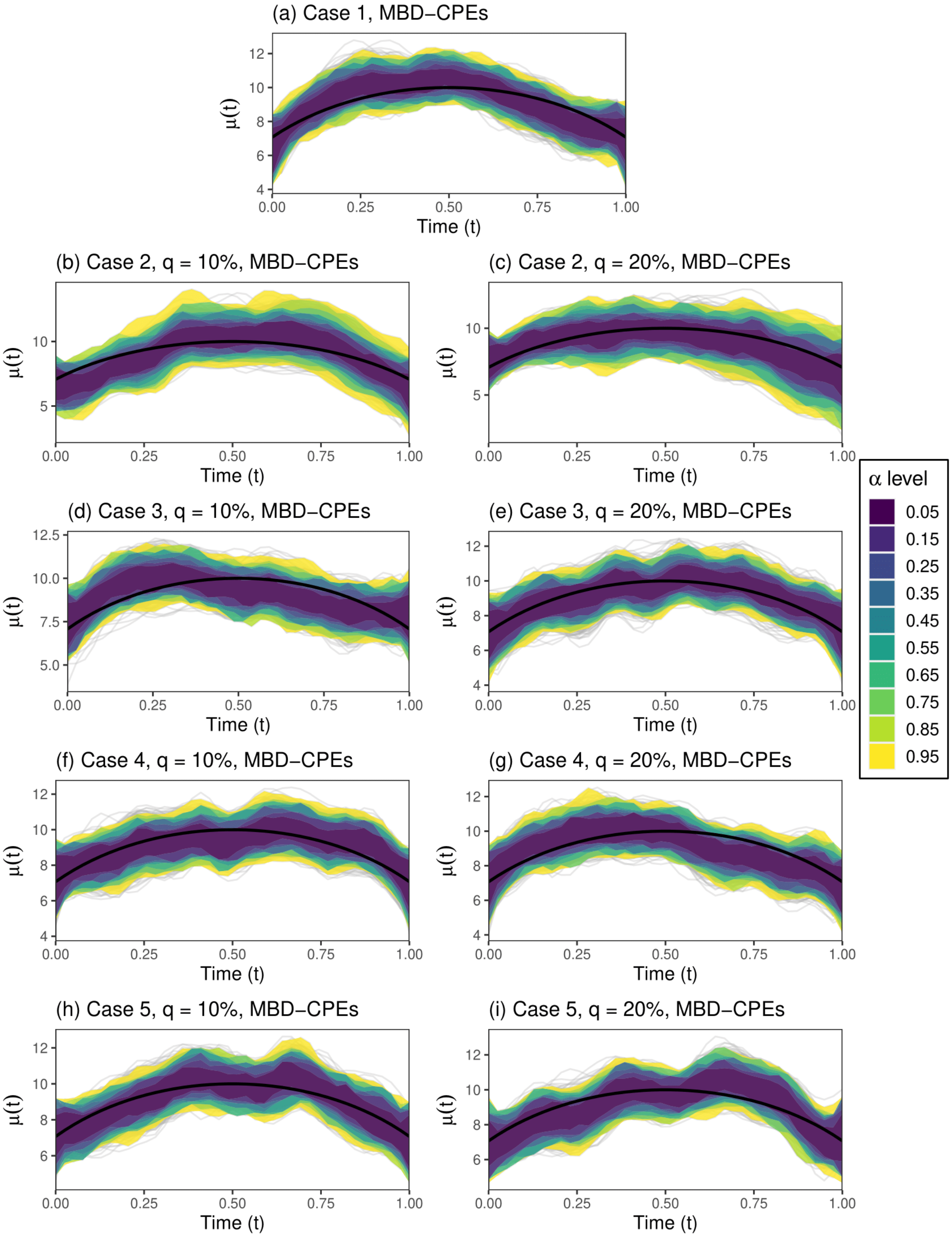


Figure B.3: MBD-CPE contours of $\mu(t)$, denoted by $D_{1-\alpha}\{\mu(t)\}$, for each simulation case, overlaying the posterior estimates and the true function given in gray and black, respectively. The left and right hand columns (excluding the first row) display the MBD-CPEs for $q = 10\%$ and $q = 20\%$ outliers, respectively, at a grid of α levels marked by varying contour colors.

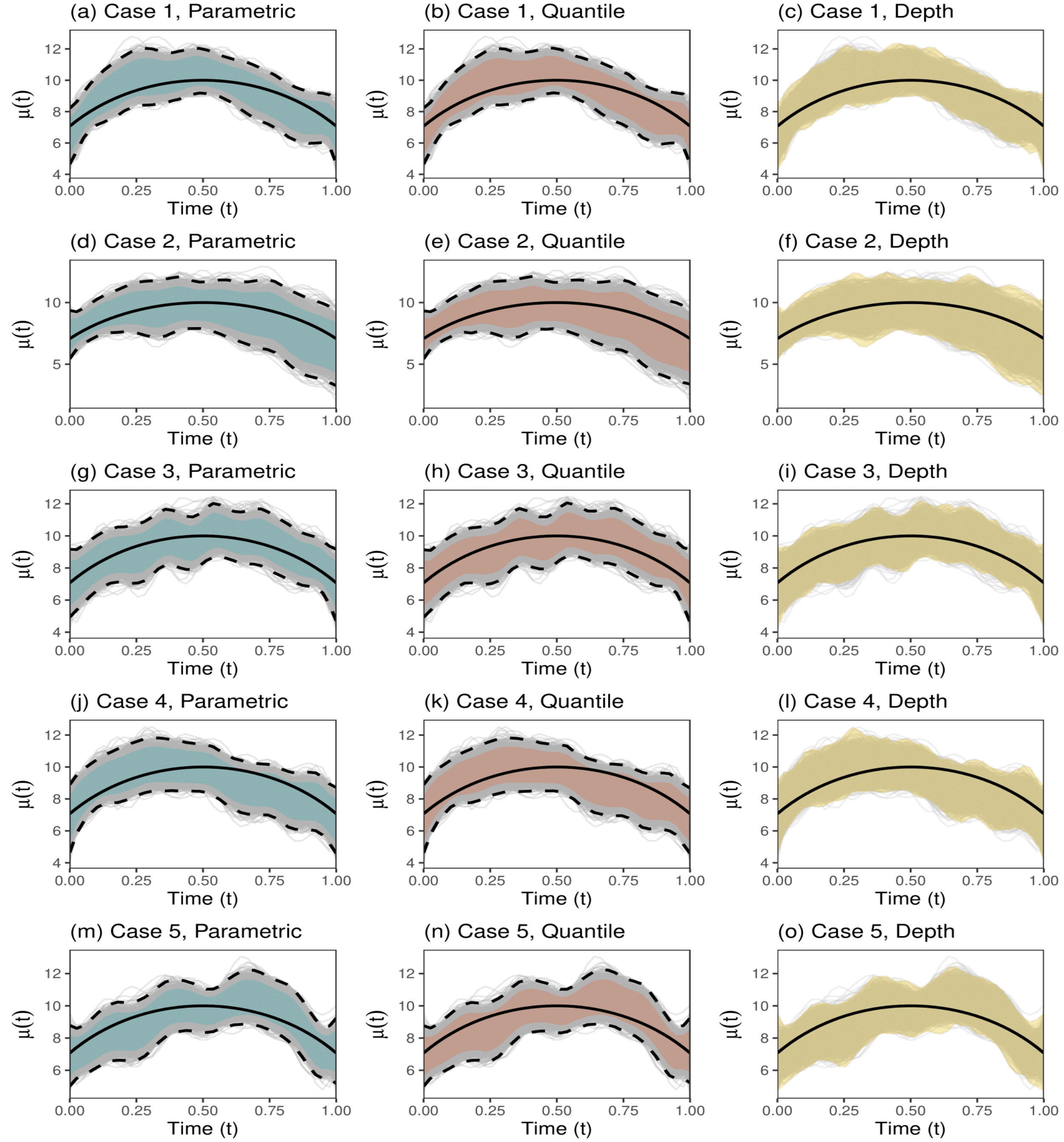
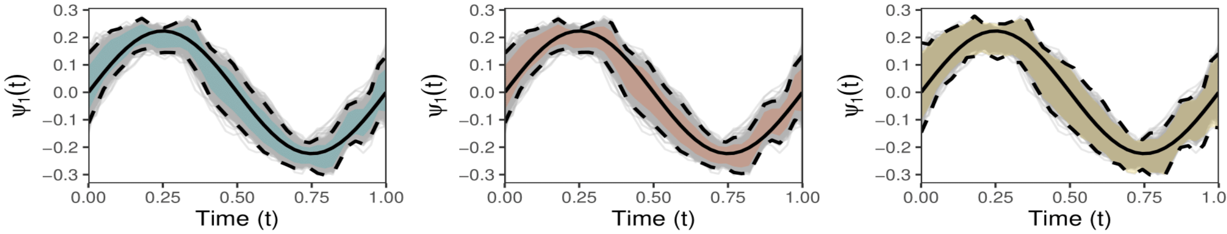
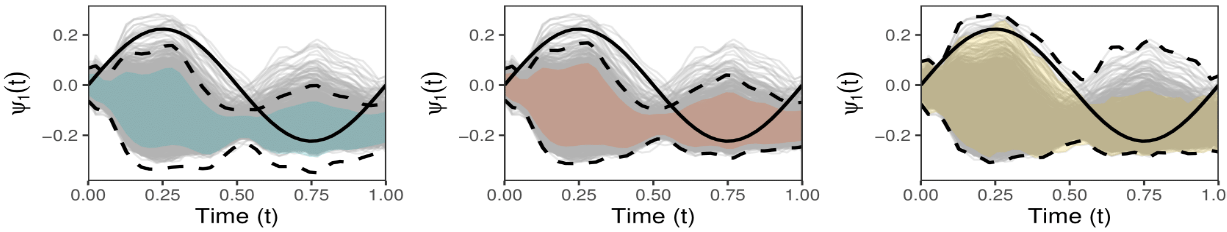


Figure B.4: 95% parametric and quantile credible intervals along with 95% CPEs for $\mu(t)$ from a single Monte Carlo run for all simulation cases with $q = 20\%$ outliers. The light grey solid lines represent the sample of $M = 4000$ posterior estimates. The blue, red, and yellow shaded regions represent $P^p_{.95}\{\mu(t)\}$, $Q^p_{.95}\{\mu(t)\}$, and $D_{.95}\{\mu(t)\}$, respectively. The black dashed lines in the left and middle columns represent $P^s_{.95}\{\mu(t)\}$ and $Q^s_{.95}\{\mu(t)\}$, respectively, while the true function is given in solid black.


(a) Case 1, Parametric

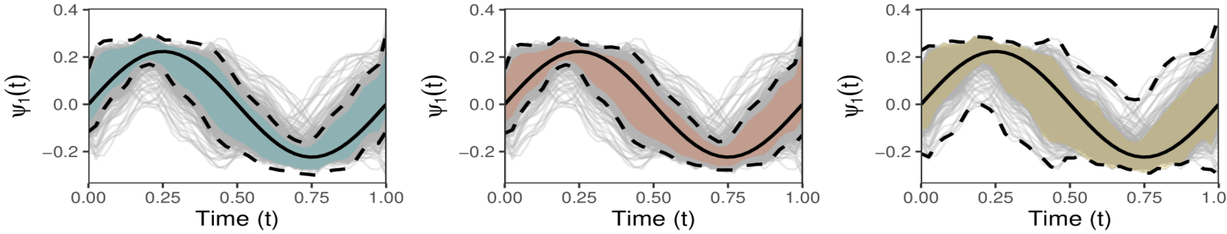
(b) Case 1, Quantile

(c) Case 1, Depth


(d) Case 2, Parametric

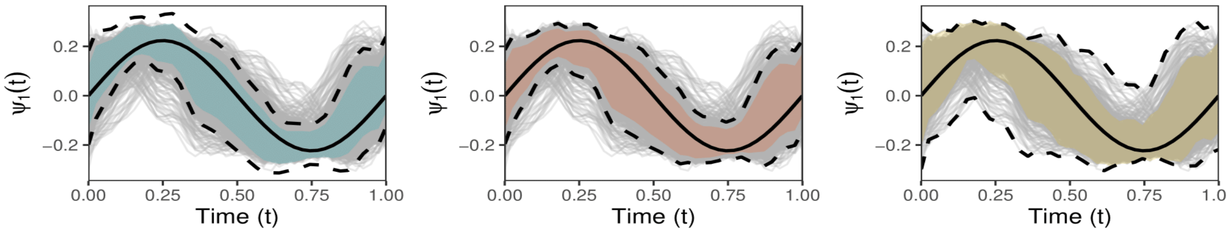
(e) Case 2, Quantile

(f) Case 2, Depth


(g) Case 3, Parametric

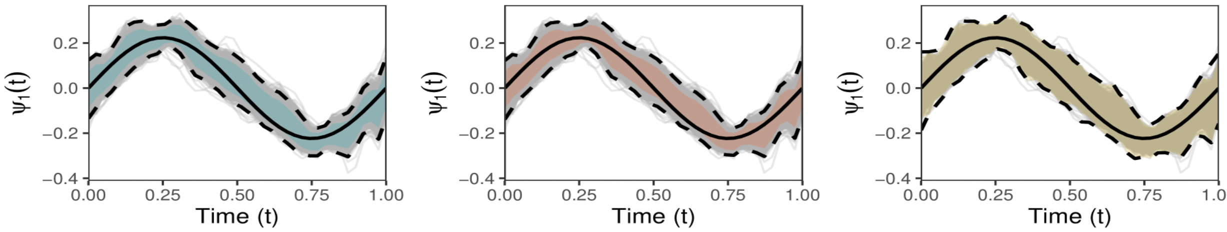
(h) Case 3, Quantile

(i) Case 3, Depth


(j) Case 4, Parametric

(k) Case 4, Quantile

(l) Case 4, Depth


(m) Case 5, Parametric

(n) Case 5, Quantile

(o) Case 5, Depth

Figure B.5: 95% parametric and quantile credible intervals along with 95% CPEs for $\psi_1(t)$ from a single Monte Carlo run for all simulation cases with $q = 20\%$ outliers. The light grey solid lines represent the sample of $M = 4000$ posterior estimates. The blue, red, and yellow shaded regions represent $P^p_{.95}\{\psi_1(t)\}$, $Q^p_{.95}\{\psi_1(t)\}$, and $D_{.95}\{\psi_1(t)\}$, respectively. The black dashed lines in the left, middle and right columns represent $P^s_{.95}\{\psi_1(t)\}$, $Q^s_{.95}\{\psi_1(t)\}$, and $D^\star_{.95}\{\psi_1(t)\}$, respectively, while the true function is given in solid black.

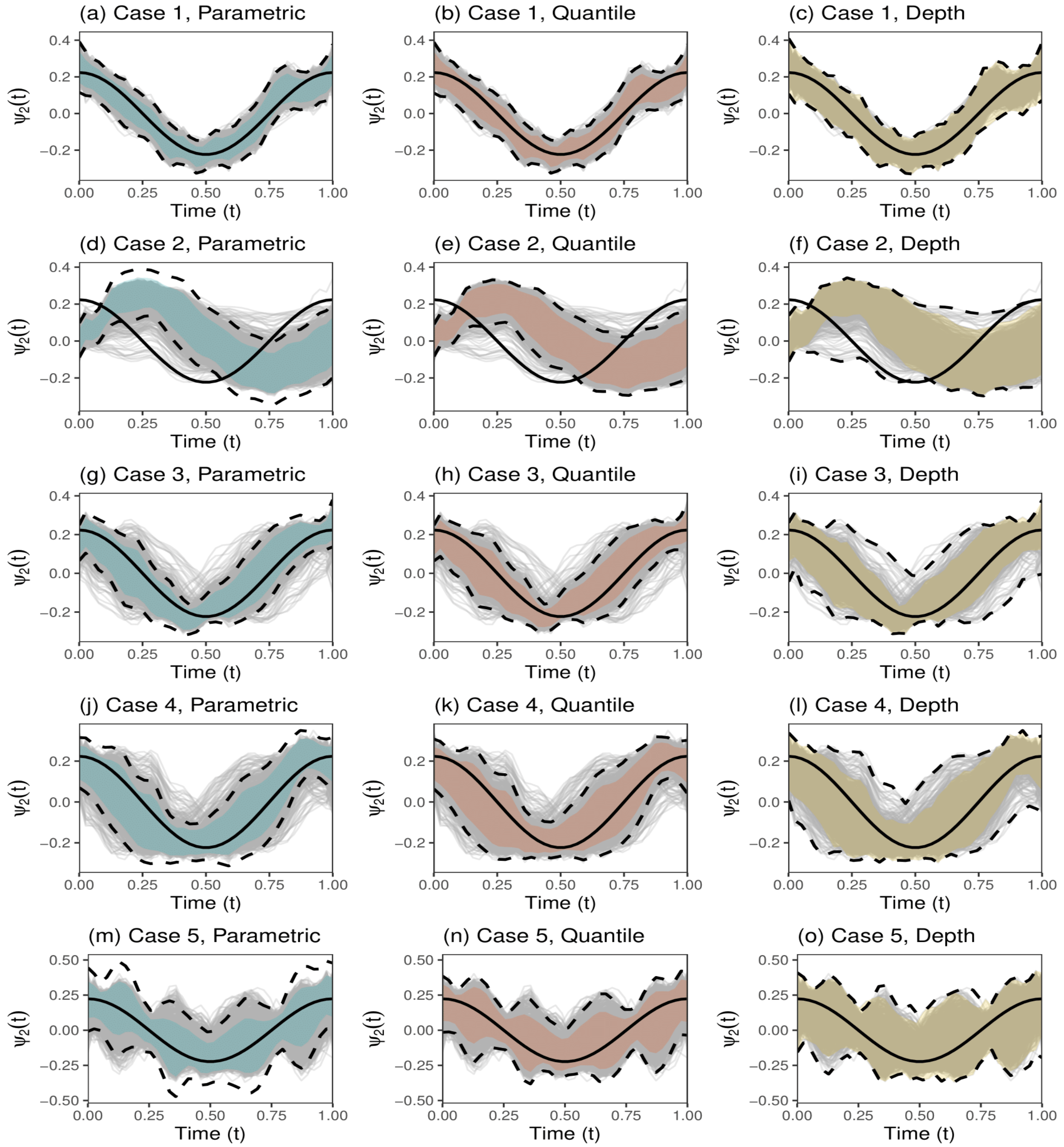


Figure B.6: 95% parametric and quantile credible intervals along with 95% CPEs for $\psi_2(t)$ from a single Monte Carlo run for all simulation cases with $q = 20\%$ outliers. The light grey solid lines represent the sample of $M = 4000$ posterior estimates. The blue, red, and yellow shaded regions represent $P^p_{.95}\{\psi_2(t)\}$, $Q^p_{.95}\{\psi_2(t)\}$, and $D_{.95}\{\psi_2(t)\}$, respectively. The black dashed lines in the left, middle and right columns represent $P^s_{.95}\{\psi_2(t)\}$, $Q^s_{.95}\{\psi_2(t)\}$, and $D^{\star}_{.95}\{\psi_2(t)\}$, respectively, while the true function is given in solid black.

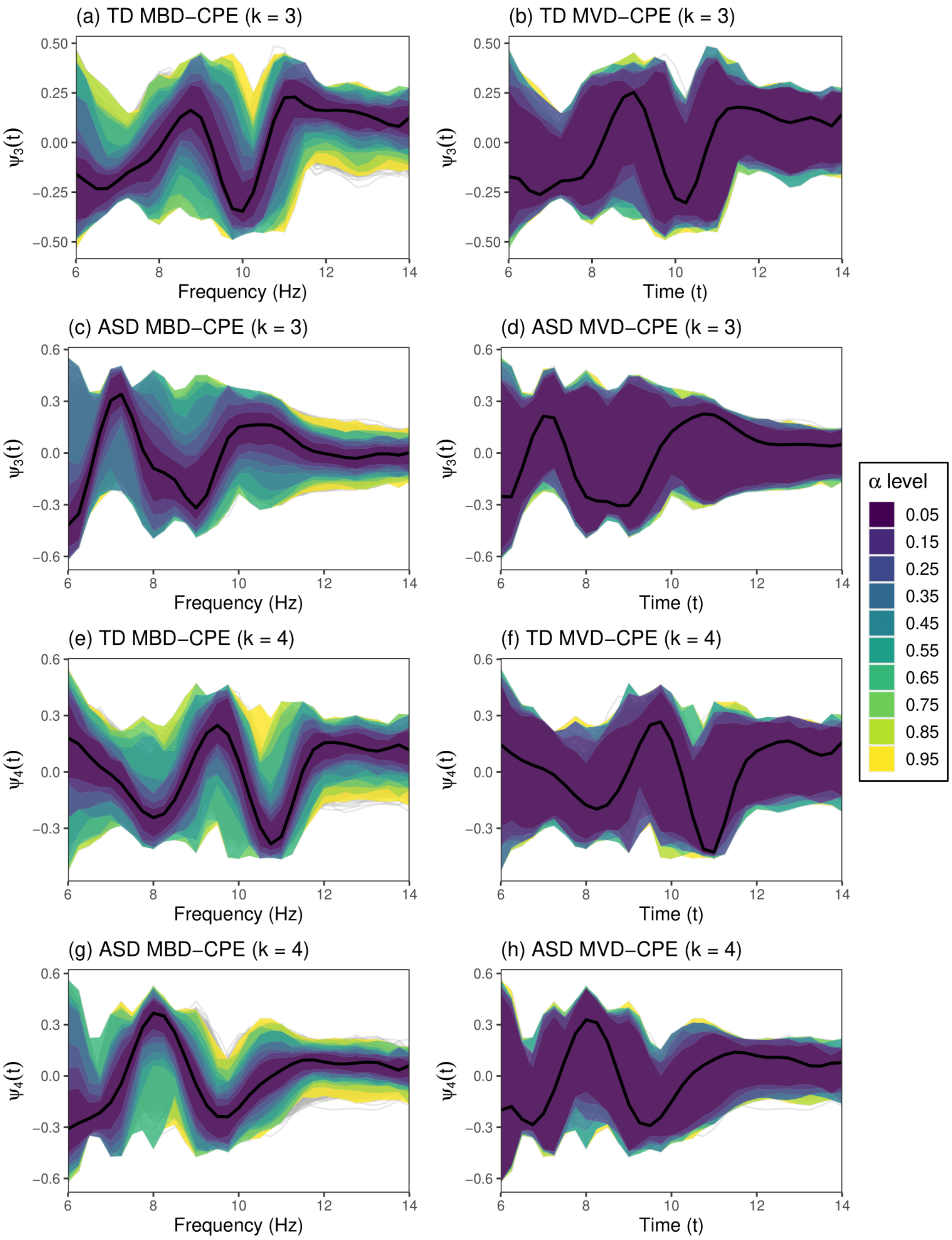


Figure B.7: CPE contours of the third and fourth leading eigenfunctions for both ASD and TD groups in our data application, overlaying the posterior estimates given in gray. The left and right hand columns display the MBD and MVD-CPEs, denoted by $D_{1-\alpha}\{\psi_1(t)\}$ and $D^{\star}_{1-\alpha}\{\psi_1(t)\}$, respectively, at a grid of α levels marked by varying contour colors. The estimated MBD and MVD median are given in solid black in the right and left columns, respectively.

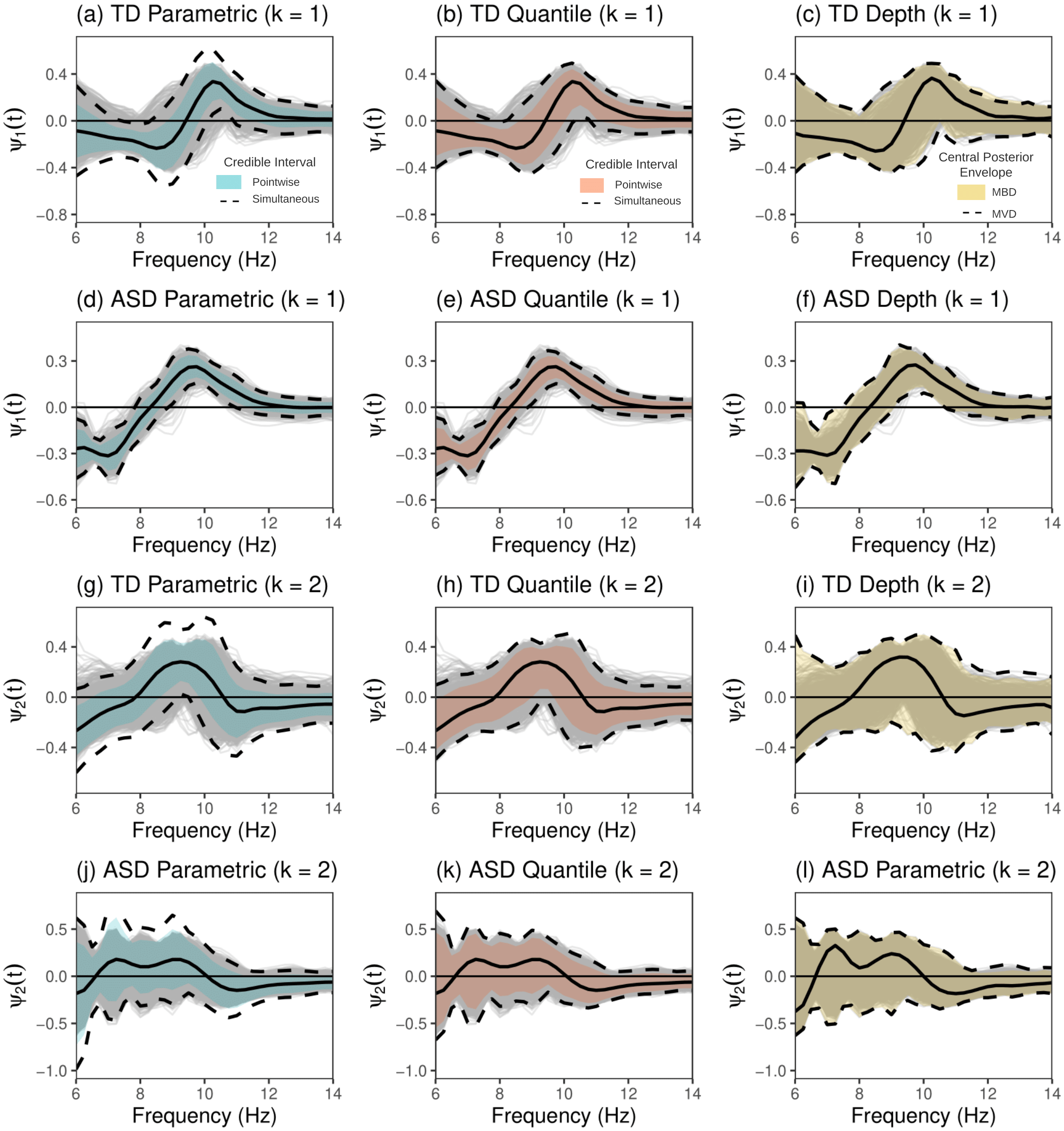


Figure B.8: 95% parametric and quantile credible intervals along with 95% CPEs for the leading two eigenfunctions in our data application. The light grey solid lines represent the sample of $M = 4000$ posterior estimates. The blue, red, and yellow shaded regions represent $P^p_{.95}\{\psi_k(t)\}$, $Q^p_{.95}\{\psi_k(t)\}$, and $D_{.95}\{\psi_k(t)\}$, respectively. The black dashed lines in the left, middle and right columns represent $P^s_{.95}\{\psi_1(t)\}$, $Q^s_{.95}\{\psi_k(t)\}$, and $D^\star_{.95}\{\psi_k(t)\}$, respectively, while the estimated eigenfunctions (eigenfunction estimate and MBD median) are given in solid black.

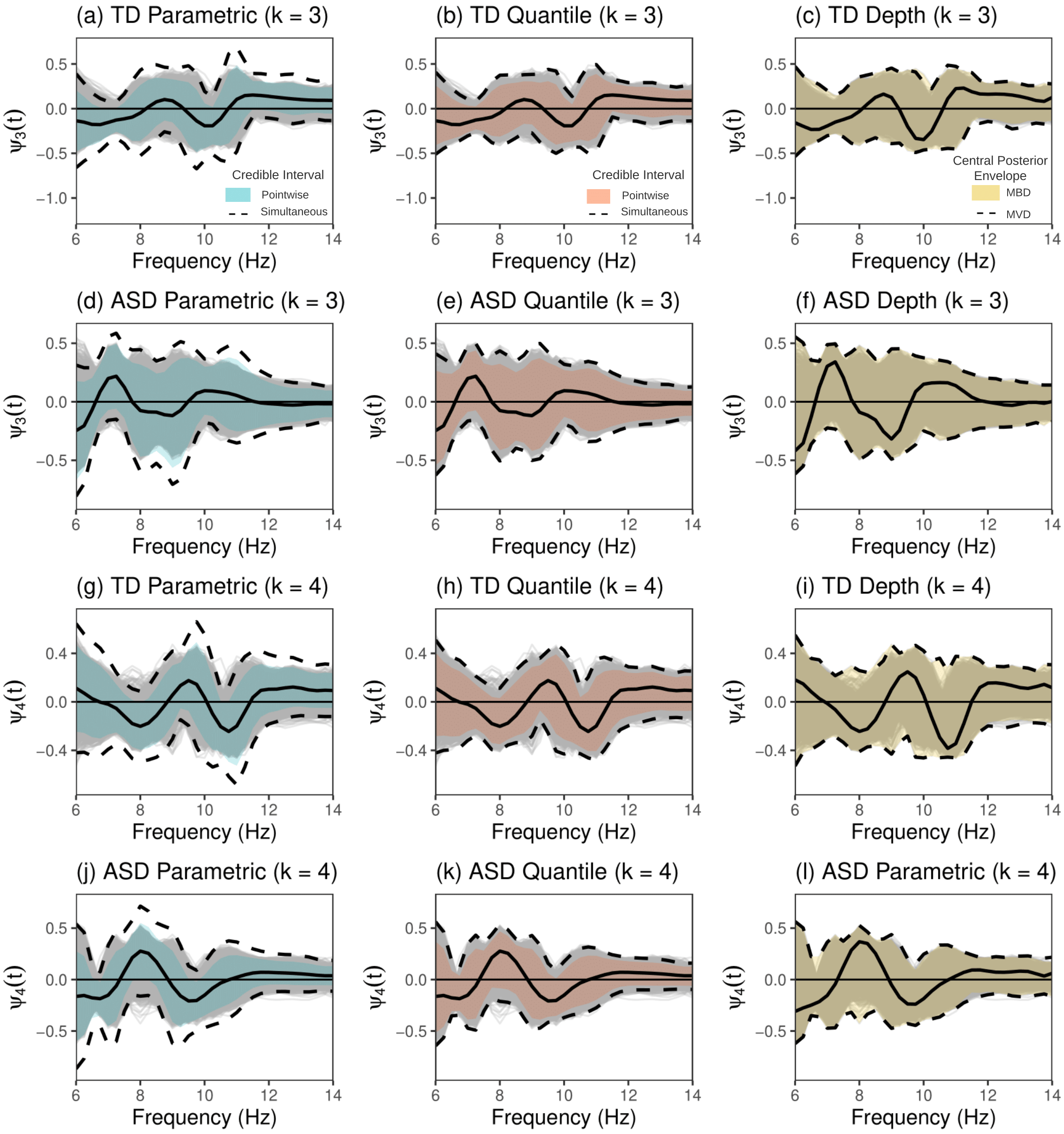


Figure B.9: 95% parametric and quantile credible intervals along with 95% CPEs for the leading third and fourth eigenfunctions in our data application. The light grey solid lines represent the sample of $M = 4000$ posterior estimates. The blue, red, and yellow shaded regions represent $P^{p}_{.95}\{\psi_k(t)\}$, $Q^{p}_{.95}\{\psi_k(t)\}$, and $D_{.95}\{\psi_k(t)\}$, respectively. The black dashed lines in the left, middle and right columns represent $P^{s}_{.95}\{\psi_k(t)\}$, $Q^{s}_{.95}\{\psi_k(t)\}$, and $D^{\star}_{.95}\{\psi_k(t)\}$, respectively, while the estimated eigenfunctions (eigenfunction estimate and MBD median) are given in solid black.

APPENDIX C

Chapter 3 Appendices

C.1 Prior and Posterior Distributions and Gibbs Sampling for B-LFPCA

To recall, posterior estimation for the BLFPCA model is achieved by first specifying a mixed effects model as:

$$\begin{aligned}
&\boldsymbol{Y}_i = \boldsymbol{f}_i + \boldsymbol{\epsilon}_i = (B_1 \otimes B_2)\left\{(\Lambda \otimes \Gamma)\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i\right\} + \boldsymbol{\epsilon}_i,\\
&\boldsymbol{\eta}_i \sim N_{q_1 q_2}(\boldsymbol{\beta}, \Sigma_\eta), \quad \boldsymbol{\zeta}_i \sim N_{p_1 p_2}(\boldsymbol{0}_{p_1 p_2}, \Sigma_\zeta),\\
&\boldsymbol{\epsilon}_i \sim N_{n_s n_t}(\boldsymbol{0}_{n_s n_t}, \sigma_\epsilon^2 I_{n_s n_t}), \quad i = 1, \ldots, n.
\end{aligned}$$

Conditionally conjugate priors provide efficient posterior estimation of the mean coefficients and variance components through the use of Markov chain Monte Carlo (MCMC). More specifically, Gaussian priors for the longitudinal factor loadings $\boldsymbol{\gamma}_\ell$ (ℓth row of Γ) and functional factor loadings $\boldsymbol{\lambda}_m$ (mth row of Λ) are employed. This is followed by an adaptive regularization approach for the variance components of the longitudinal factor loadings $\sigma^2_{\gamma_{\ell j}}$ and functional factor loadings $\sigma^2_{\lambda_{mk}}$, where the number of latent factors q_1 and q_2 are chosen to be sufficiently large and a modified multiplicative gamma process shrinkage (MMGPS) prior Bhattacharya and Dunson (2011); Shamshoian et al. (2022) is employed. The prior for the longitudinal factor loadings $\boldsymbol{\gamma}_\ell$, $\ell = 1, \ldots, p_1$, is

$$\begin{aligned}
&\boldsymbol{\gamma}_\ell \sim N_{q_1}\left(\mathbf{0}_{q_1}, \Sigma_{\gamma_\ell}\right), \quad \Sigma_{\gamma_\ell} = \operatorname{diag}\left(\sigma^2_{\gamma_{\ell 1}}, \ldots, \sigma^2_{\gamma_{\ell q_1}}\right), \quad \sigma^2_{\gamma_{\ell j}} = \varphi^{-1}_{1\ell j}\pi^{-1}_{1j}, \\
&\quad \varphi_{1\ell j} \sim \text{Gamma}\left(\frac{\nu_1}{2}, \frac{\nu_1}{2}\right), \quad \pi_{1j} = \prod_{h=1}^{j} \delta_{1v}, \\
&\quad \delta_{11} \sim \text{Gamma}\left(a_{11}, 1\right), \quad \delta_{1h} \sim \text{Gamma}\left(a_{12}, 1\right) I\left(\delta_{1h} > 1\right), \; h \geqslant 2, \\
&\quad a_{11} \sim \text{Gamma}\left(r_1, 1\right), \quad a_{12} \sim \text{Gamma}\left(r_2, 1\right),
\end{aligned}$$

where Σ_{γ_ℓ} is a $q_1 \times q_1$ diagonal matrix comprised of the variance components, $\sigma^2_{\gamma_{\ell 1}}, \ldots, \sigma^2_{\gamma_{\ell q_1}}$, for the the ℓth longitudinal factor loading $\boldsymbol{\gamma}_\ell$, and $I(\cdot)$ denotes the indicator function. Adaptive shrinkage is induced through the gamma hyper-priors placed on a_{11} and a_{12}, and for more details on the prior formulation and mechanics defer the reader to Section 3 of Shamshoian et al. Shamshoian et al. (2022). Similarly, the prior for the functional factor loadings $\boldsymbol{\lambda}_m$, $m = 1, \ldots, p_2$, is

$$\begin{aligned}
&\boldsymbol{\lambda}_m \sim N_{q_2}\left(\mathbf{0}_{q_2}, \Sigma_{\lambda_m}\right), \quad \Sigma_{\lambda_m} = \operatorname{diag}\left(\sigma^2_{\lambda_{m1}}, \ldots, \sigma^2_{\lambda_{mq_1}}\right), \quad \sigma^2_{\lambda_{mk}} = \varphi^{-1}_{2mk}\pi^{-1}_{2k}, \\
&\quad \varphi_{2mk} \sim \text{Gamma}\left(\frac{\nu_2}{2}, \quad \frac{\nu_2}{2}\right), \quad \pi_{2k} = \prod_{h=1}^{k} \delta_{2h}, \\
&\quad \delta_{21} \sim \text{Gamma}\left(a_{21}, 1\right), \quad \delta_{2h} \sim \text{Gamma}\left(a_{22}, 1\right) I\left(\delta_{2h} > 1\right), \; h \geqslant 2, \\
&\quad a_{21} \sim \text{Gamma}\left(r_1, 1\right), \quad a_{22} \sim \text{Gamma}\left(r_2, 1\right),
\end{aligned}$$

where Σ_{λ_m} is a $q_2 \times q_2$ diagonal matrix comprised of the variance components, $\sigma^2_{\lambda_{m1}}, \ldots, \sigma^2_{\lambda_{mq_2}}$, for the mth functional factor loading $\boldsymbol{\lambda}_m$. The priors for the B-LFPCA model are completed by employing a Gaussian prior for the mean coefficients $\boldsymbol{\beta}$ and gamma priors for the variance of the mean coefficients $\sigma^2_{\beta_{jk}}$, variance of the latent factors $\sigma^2_{\eta_{jk}}$, variance of the reconstruction errors $\sigma^2_{\zeta_{\ell m}}$, and variance of the measurement error σ^2_ϵ. Specifically, we assume

$$\boldsymbol{\beta} \sim N_{q_1q_2}\left(\mathbf{0}_{q_1q_2}, \Sigma_\beta\right), \quad \Sigma_\beta = \text{diag}\left(\sigma^2_{\beta_{11}}, \ldots, \sigma^2_{\beta_{q_1q_2}}\right), \quad \frac{1}{\sigma^2_{\beta_{jk}}} \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right),$$

where Σ_β is a $q_1q_2 \times q_1q_2$ diagonal matrix comprised of the variance components, $\sigma^2_{\beta_{11}}, \ldots,$ $\sigma^2_{\beta_{q_1q_2}}$, and for $\ell = 1, \ldots, p_1$, $m = 1, \ldots, p_2$, $j = 1, \ldots, q_1$, $k = 1, \ldots, q_2$,

$$\frac{1}{\sigma^2_{\eta_{jk}}} \sim \text{Gamma}\left(a_\eta, b_\eta\right), \quad \frac{1}{\sigma^2_{\zeta_{\ell m}}} \sim \text{Gamma}\left(a_\zeta, b_\zeta\right), \quad \frac{1}{\sigma^2_\epsilon} \sim \text{Gamma}\left(a_\epsilon, b_\epsilon\right).$$

We target the posterior distributions in model (3.6) using a Gibbs sampler with the exception of the shrinkage parameters a_{11}, a_{12}, a_{21}, and a_{22}, which are updated via a Metropolis-Hastings step. In what follows, let $\Theta_i = (\boldsymbol{\theta}_{i1}, \ldots, \boldsymbol{\theta}_{ip_1})^\top = (\boldsymbol{\theta}^*_{i1}, \ldots, \boldsymbol{\theta}^*_{ip_2})$ denote the $p_1 \times p_2$ matrix of subject-specific basis coefficients with ℓth row $\boldsymbol{\theta}_{i\ell} = (\theta_{i\ell 1}, \ldots, \theta_{i\ell p_2})^\top$ and mth column $\boldsymbol{\theta}^*_{im} = (\theta_{i1m}, \ldots, \theta_{ip_1m})^\top$, and let $\boldsymbol{\Theta}_i = Vec(\Theta_i)$ be the $p_1p_2 \times 1$ vector obtained from stacking the columns of Θ_i. Further, let Gamma* denote the rate parameterization of the Gamma probability density function (i.e., if $a \sim$ Gamma(b, c), then $E[a] = bc$), and $\Phi(.)$ denote the standard normal cumulative density function. The steps of the MCMC chain are as follows:

1. $\boldsymbol{\gamma_\ell}$|others $\sim N_{q_1}(\boldsymbol{\mu}^{post}_{\gamma_\ell}, v^{post}_{\gamma_\ell})$ where $v^{post}_{\gamma_\ell} = (\sum_{i=1}^n \eta_i \Lambda^\top \Sigma^{-1}_{\zeta_\ell} \Lambda \eta_i^\top + \Sigma^{-1}_{\gamma_\ell})^{-1}$, $\boldsymbol{\mu}^{post}_{\gamma_\ell} = v^{post}_{\gamma_\ell} \times$ $\sum_{i=1}^n \eta_i \Lambda^\top \Sigma^{-1}_{\zeta_\ell} \boldsymbol{\theta}_{i\ell}$, where $\Sigma^{-1}_{\zeta_\ell}$ is a $p_2 \times p_2$ diagonal matrix denoted as $\Sigma^{-1}_{\zeta_\ell} = \text{diag}(\sigma^{-2}_{\zeta_{\ell 1}}, \ldots,$ $\sigma^{-2}_{\zeta_{\ell p_2}})$, for $\ell = 1, \ldots, p_1$.

2. $\varphi_{1\ell j}$|others $\sim$ Gamma$(a^{post}_{\varphi_{1\ell j}}, b^{post}_{\varphi_{1\ell j}})$, where $a^{post}_{\varphi_{1\ell j}} = (\nu_1 + 1)/2$, $b^{post}_{\varphi_{1\ell j}} = (\nu_1 + \pi_{1j}\gamma^2_{\ell j})/2$, for $\ell = 1, \ldots, p_1$, $j = 1, \ldots, q_1$.

3. π_{1j}|others $= \prod_{h=1}^j \delta_{1h}$|others, for $j = 1, \ldots, q_1$.

    (a) δ_{11}|others $\sim$ Gamma$(a^{post}_{\delta_{11}}, b^{post}_{\delta_{1h}})$, where $a^{post}_{\delta_{11}} = a_{11} + (p_1q_1)/2$, $b^{post}_{\delta_{11}} = 1 + (\sum_{j=1}^{q_1} \pi^{(1)}_{1j}$ $\sum_{\ell=1}^{p_1} \varphi_{1\ell j}\gamma^2_{\ell j})/2$, where $\pi^{(1)}_{1j} = \prod_{h=2}^j \delta_{1h}$.

(b) $\delta_{1h}|\text{others} \sim \text{Gamma}(a^{post}_{\delta_{1h}}, b^{post}_{\delta_{1h}})$, where $a^{post}_{\delta_{1h}} = a_{12} + p_1(q_1 - h + 1)/2$, $b^{post}_{\delta_{1h}} = 1 + (\sum_{j=(h+1)}^{q_1} \pi_{1j}^{(h)} \sum_{\ell=1}^{p_1} \varphi_{1\ell j}\gamma_{\ell j}^2)/2$, where $\pi_{1j}^{(h)} = \prod_{h' \neq h} \delta_{1h'}$.

4. Sample a_{11}:

    (a) Sample $u \sim \text{Uniform}(0,1)$.

    (b) Sample $x \sim N(0,1)I(x + a_{11} > 0)$, where $I(.)$ is the indicator function, and set the proposal value of a_{11} to $a_{11}^* = x + a_{11}$.

    (c) Compute
    $$A = \frac{\text{Gamma}^*(\delta_{11}, a_{11}^*, 1)\text{Gamma}^*(a_{11}^*, r_1, 1)\Phi(a_{11})}{\text{Gamma}^*(\delta_{11}, a_{11}, 1)\text{Gamma}^*(a_{11}, r_1, 1)\Phi(a_{11}^*)}.$$

    (d) If $A > u$, then set the new value of a_{11} to a_{11}^*. Otherwise, keep the previous value of a_{11}.

5. Sample a_{12}:

    (a) Sample $u \sim \text{Uniform}(0,1)$.

    (b) Sample $x \sim N(0,1)I(x+a_{12} > 0)$, and set the proposal value of a_{12} to $a_{12}^* = x+a_{12}$.

    (c) Compute
    $$A = \frac{\text{Gamma}^*(a_{12}^*, r_2, 1)\prod_{h=2}^{q_1}\text{Gamma}^*(\delta_{1h}, a_{12}^*, 1)\Phi(a_{12})}{\text{Gamma}^*(a_{12}, r_2, 1)\prod_{h=2}^{q_1}\text{Gamma}^*(\delta_{1h}, a_{12}, 1)\Phi(a_{12}^*)}.$$

    (d) If $A > u$, then set the new value of a_{12} to a_{12}^*. Otherwise, keep the previous value of a_{12}.

6. $\boldsymbol{\lambda}_m|\text{others} \sim N_{q_2}(\boldsymbol{\mu}^{post}_{\lambda_m}, v^{post}_{\lambda_m})$, where $v^{post}_{\lambda_m} = (\sum_{i=1}^n \eta_i \Gamma^\top \Sigma_{\zeta_m}^{-1}\Gamma\eta_i^\top + \Sigma_{\lambda_m}^{-1})^{-1}$, where $\Sigma_{\zeta_m}^{-1} = \text{diag}(\sigma_{\zeta_{1m}}^{-2}, \ldots, \sigma_{\zeta_{p_1 m}}^{-2})$, $\boldsymbol{\mu}^{post}_{\lambda_m} = v^{post}_{\lambda_m} \times \sum_{i=1}^n \eta_i\Gamma^\top\Sigma_{\zeta_m}^{-1}\boldsymbol{\theta}_{im}^*$, where $\Sigma_{\zeta_m}^{-1}$ is a $p_1 \times p_1$ diagonal matrix denoted as $\Sigma_{\zeta_m}^{-1} = \text{diag}(\sigma_{\zeta_{1m}}^{-2}, \ldots, \sigma_{\zeta_{p_1 m}}^{-2})$, for $m = 1, \ldots, p_2$.

7. $\varphi_{2mk}|\text{others} \sim \text{Gamma}(a^{post}_{\varphi_{2mk}}, b^{post}_{\varphi_{2mk}})$, where $a^{post}_{\varphi_{2mk}} = (\nu_2+1)/2$, $b^{post}_{\varphi_{2mk}} = (\nu_2+\pi_{2k}\lambda_{mk}^2)/2$, for $m = 1, \ldots, p_2$, $k = 1, \ldots, q_2$.

8. π_{2k}|others $= \prod_{h=1}^{k} \delta_{2h}$|others, for $k = 1, \dots, q_2$.

    (a) δ_{21}|others $\sim$ Gamma$(a^{post}_{\delta_{21}}, b^{post}_{\delta_{21}})$, where $a^{post}_{\delta_{21}} = a_{21} + (p_2 q_2)/2$, $b^{post}_{\delta_{21}} = 1 + (\sum_{k=1}^{q_2} \pi_{2k}^{(1)} \sum_{m=1}^{p_2} \varphi_{2mk} \lambda_{mk}^2)/2$, where $\pi_{2k}^{(1)} = \prod_{h=2}^{k} \delta_{2h}$.

    (b) δ_{2h}|others $\sim$ Gamma$(a^{post}_{\delta_{2h}}, b^{post}_{\delta_{2h}})$, where $a^{post}_{\delta_{2h}} = a_{22} + p_2(q_2 - h + 1)/2$, $b^{post}_{\delta_{2h}} = 1 + (\sum_{k=(h+1)}^{q_2} \pi_{2k}^{(h)} \sum_{m=1}^{p_2} \varphi_{2mk} \lambda_{mk}^2)/2$, where $\pi_{2k}^{(h)} = \prod_{h' \neq h} \delta_{2h'}$.

9. Sample a_{21}:

    (a) Sample $u \sim$ Uniform$(0, 1)$.

    (b) Sample $x \sim N(0, 1)I(x + a_{21} > 0)$, and set the proposal value of a_{21} to $a_{21}^* = x + a_{21}$.

    (c) Compute

    $$A = \frac{\text{Gamma}^*(\delta_{21}, a_{21}^*, 1)\text{Gamma}^*(a_{21}^*, r_1, 1)\Phi(a_{21})}{\text{Gamma}^*(\delta_{21}, a_{21}, 1)\text{Gamma}^*(a_{21}, r_1, 1)\Phi(a_{21}^*)}.$$

    (d) If $A > u$, then set the new value of a_{21} to a_{21}^*. Otherwise, keep the previous value of a_{21}.

10. Sample a_{22}:

    (a) Sample $u \sim$ Uniform$(0, 1)$.

    (b) Sample $x \sim N(0, 1)I(x + a_{22} > 0)$, and set the proposal value of a_{22} to $a_{22}^* = x + a_{22}$.

    (c) Compute

    $$A = \frac{\text{Gamma}^*(a_{22}^*, r_2, 1)\prod_{h=2}^{q_2} \text{Gamma}^*(\delta_{2h}, a_{22}^*, 1)\Phi(a_{22})}{\text{Gamma}^*(a_{22}, r_2, 1)\prod_{h=2}^{q_2} \text{Gamma}^*(\delta_{2h}, a_{22}, 1)\Phi(a_{22}^*)}.$$

    (d) If $A > u$, then set the new value of a_{22} to a_{22}^*. Otherwise, keep the previous value of a_{22}.

11. $\sigma^2_{\zeta_{\ell m}}$|others $\sim$ Gamma$(a^{post}_{\sigma^2_{\zeta_{\ell m}}}, b^{post}_{\sigma^2_{\zeta_{\ell m}}})$, where $a^{post}_{\sigma^2_{\zeta_{\ell m}}} = a_\zeta + n/2$, $b^{post}_{\sigma^2_{\zeta_{\ell m}}} = b_\zeta + (1/2)\sum_{i=1}^{n}(\theta_{i\ell m} - \boldsymbol{\gamma}_\ell^\top \eta_i \boldsymbol{\lambda}_m)^2$, for $\ell = 1, \dots, p_1$, $m = 1, \dots, p_2$.

12. σ^2_ϵ|others $\sim$ Gamma$(a^{post}_{\sigma^2_\epsilon}, b^{post}_{\sigma^2_\epsilon})$, where $a^{post}_{\sigma^2_\epsilon} = a_\epsilon + n_{tot}/2$, $b^{post}_{\sigma^2_\epsilon} = b_\epsilon + (1/2)\sum_{i=1}^{n}\{\boldsymbol{Y}_i - (B_1 \otimes B_2)\boldsymbol{\Theta}_i\}^\top\{\boldsymbol{Y}_i - (B_1 \otimes B_2)\boldsymbol{\Theta}_i\}$, where n_{tot} is the total number of non-missing observed time points across the longitudinal and functional dimensions ($n_{tot} = n_s \times n_t \times n$ when there is no sparsity in $Y_i(s,t)$).

13. $\boldsymbol{\eta_i}$|others $\sim N_{q_1q_2}(\boldsymbol{\mu}^{post}_{\eta_i}, v^{post}_{\eta_i})$, where $v^{post}_{\eta_i} = \{(\Lambda^\top \otimes \Gamma^\top)\Sigma_\zeta^{-1}(\Lambda \otimes \Gamma)\}^{-1}$, $\boldsymbol{\mu}^{post}_{\eta_i} = v^{post}_{\eta_i} \times (\Lambda^\top \otimes \Gamma^\top)\Sigma_\zeta^{-1}\boldsymbol{\Theta}_i$, for $i = 1, \dots, n$.

14. $\boldsymbol{\Theta_i}$|others $\sim N_{p_1p_2}(\boldsymbol{\mu}^{post}_{\theta_i}, v^{post}_{\theta_i})$, where $v^{post}_{\theta_i} = \{\sigma_\epsilon^{-2}(B_1^\top B_1 \otimes B_2^\top B_2) + \Sigma_\zeta^{-1}\}^{-1}$, $\boldsymbol{\mu}^{post}_{\theta_i} = v^{post}_{\theta_i} \times \{\sigma^2_\epsilon(B_1 \otimes B_2)\boldsymbol{Y}_i + \Sigma_\zeta^{-1}(\Lambda \otimes \Gamma)\boldsymbol{\eta}_i\}$, for $i = 1, \dots, n$.

15. $\sigma^2_{\eta_{jk}}$|others $\sim$ Gamma$(a^{post}_{\sigma^2_{\eta_{jk}}}, b^{post}_{\sigma^2_{\eta_{jk}}})$, where $a^{post}_{\sigma^2_{\eta_{jk}}} = a_\eta + n/2$, $b^{post}_{\sigma^2_{\eta_{jk}}} = b_\eta + (1/2) + \sum_{i=1}^{n}(\eta_{ijk} - \beta_{jk})^2$, for $j = 1, \dots, q_1$, $k = 1, \dots, q_2$.

16. $\boldsymbol{\beta}$|others $\sim N(\boldsymbol{\mu}^{post}_\beta, v^{post}_\beta)$, where $v^{post}_\beta = \{n\Sigma_\eta^{-1}I_{q_1q_2} + \Sigma_\beta^{-1}\}^{-1}$, $\boldsymbol{\mu}^{post}_\beta = v^{post}_\beta \times \Sigma_\eta^{-1}\sum_{i=1}^{n}\boldsymbol{\eta}_i$, where $I_{q_1q_2}$ is the $q_1q_2 \times q_1q_2$ identity matrix.

17. $\sigma^2_{\beta_{jk}}$|others $\sim$ Gamma$(a^{post}_{\sigma^2_{\beta_{jk}}}, b^{post}_{\sigma^2_{\beta_{jk}}})$, where $a^{post}_{\sigma^2_{\beta_{jk}}} = 1$, $b^{post}_{\sigma^2_{\beta_{jk}}} = (1 + \beta^2_{jk})/2$, for $j = 1, \dots, q_1$, $k = 1, \dots, q_2$.

Gibbs sampler updates, as well as Metropolis-Hastings steps, are employed to obtain posterior samples of the above parameters. For the variance components of the element-wise precisions of the longitudinal and functional factor loading matrices $\varphi_{1\ell j}$ and φ_{2mk}, respectively, the hyperparameters for the priors are set to $\nu_1 = 5$ and $\nu_2 = 5$. For the adaptive shrinkage parameters a_{11}, a_{12}, a_{21} and a_{22}, the hyperparameters of the priors are set to $r_1 = 1$ and $r_2 = 2$. For the variance component of the latent factors, $\sigma^2_{\eta_{jk}}$, the hyperparameters of the prior are set to $a_\eta = 1$ and $b_\eta = 1$. For the variance component of the reconstruction errors, $\sigma^2_{\zeta_{\ell m}}$, the hyperparameters of the prior are set to $a_\zeta = 0.5$ and $b_\zeta = 0.25$. For the variance component of the measurement error, the hyperparameters of the prior are set to $a_\epsilon = 0.0001$ and $b_\epsilon = 0.0001$. Modifications of these hyperparameters for the data application (Section 3.5) are detailed in Appendix C.4.

C.2 Post-processing of MCMC Samples

In this appendix, we give details on the post-processing of the posterior estimates of the longitudinal and functional eigenfunctions detailed in Section 3.2.2. First, we provide the proof for the direct computation of the functional and longitudinal marginal covariances, $K_{\mathcal{S}}(s,s')$ and $K_{\mathcal{T}}(t,t')$, respectively. Recall the covariance kernel $K\{(s,t),(s',t')\}$ can be expressed as

$$K\{(s,t),(s',t')\} = \sum_{j=1}^{q_1}\sum_{k=1}^{q_2}\sigma^2_{\eta_{jk}}\ddot{\psi}_j(s)\ddot{\psi}_j(s')\ddot{\phi}_k(t)\ddot{\phi}_k(t') + \sum_{\ell=1}^{p_1}\sum_{m=1}^{p_2}\sigma^2_{\zeta_{\ell m}}b^{(1)}_\ell(s)b^{(1)}_\ell(s')b^{(2)}_m(t)b^{(2)}_m(t').$$

Direction computation of the posterior longitudinal marginal covariance function $K_{\mathcal{S}}(s,s')$ yields

$$\begin{aligned}
K_{\mathcal{S}}(s,s') &= \int_{\mathcal{T}} K\{(s,t),(s',t)\}dt \\
&= \int_{\mathcal{T}}\left\{\sum_{j=1}^{q_1}\sum_{k=1}^{q_2}\sigma^2_{\eta_{jk}}\ddot{\psi}_j(s)\ddot{\psi}_j(s')\ddot{\phi}_k(t)\ddot{\phi}_k(t) + \sum_{\ell=1}^{p_1}\sum_{m=1}^{p_2}\sigma^2_{\zeta_{\ell m}}b^{(1)}_\ell(s)b^{(1)}_\ell(s')b^{(2)}_m(t)b^{(2)}_m(t)\right\}dt \\
&= \sum_{j=1}^{q_1}\ddot{\psi}_j(s)\ddot{\psi}_j(s')\sum_{k=1}^{q_2}\sigma^2_{\eta_{jk}}\int_{\mathcal{T}}\ddot{\phi}_k(t)\ddot{\phi}_k(t)dt + \sum_{\ell=1}^{p_1}b^{(1)}_\ell(s)b^{(1)}_\ell(s')\sum_{m=1}^{p_2}\sigma^2_{\zeta_{\ell m}}\int_{\mathcal{T}}b^{(2)}_m(t)b^{(2)}_m(t)dt \\
&= \sum_{j=1}^{q_1}\ddot{\psi}_j(s)\ddot{\psi}_j(s')\sum_{k=1}^{q_2}\sigma^2_{\eta_{jk}}\omega_k + \sum_{\ell=1}^{p_1}b^{(1)}_\ell(s)b^{(1)}_\ell(s')\sum_{m=1}^{p_2}\sigma^2_{\zeta_{\ell m}}\Omega_m,
\end{aligned}$$

where $\omega_k = \int_{\mathcal{T}}\ddot{\phi}_k(t)\ddot{\phi}_k(t)dt$ and $\Omega_m = \int_{\mathcal{T}}b^{(2)}_m(t)b^{(2)}_m(t)dt$. A similar expression for the functional marginal covariance function $K_{\mathcal{T}}(t,t')$ is derived as

$$\begin{aligned}
K_{\mathcal{T}}(t,t') &= \int_{\mathcal{S}} K\{(s,t),(s,t')\}ds \\
&= \int_{\mathcal{S}} \left\{ \sum_{j=1}^{q_1}\sum_{k=1}^{q_2} \sigma^2_{\eta_{jk}} \ddot{\psi}_j(s)\ddot{\psi}_j(s)\ddot{\phi}_k(t)\ddot{\phi}_k(t') + \sum_{\ell=1}^{p_1}\sum_{m=1}^{p_2} \sigma^2_{\zeta_{\ell m}} b^{(1)}_{\ell}(s) b^{(1)}_{\ell}(s) b^{(2)}_{m}(t) b^{(2)}_{m}(t') \right\} ds \\
&= \sum_{k=1}^{q_2} \ddot{\phi}_k(t)\ddot{\phi}_k(t') \sum_{j=1}^{q_1} \sigma^2_{\eta_{jk}} \int_{\mathcal{S}} \ddot{\psi}_j(s)\ddot{\psi}_j(s)ds + \sum_{m=1}^{p_2} b^{(2)}_{m}(t) b^{(2)}_{m}(t') \sum_{\ell=1}^{p_1} \sigma^2_{\zeta_{\ell m}} \int_{\mathcal{S}} b^{(1)}_{\ell}(s) b^{(1)}_{\ell}(s)ds \\
&= \sum_{k=1}^{q_2} \ddot{\phi}_k(t)\ddot{\phi}_k(t') \sum_{j=1}^{q_1} \sigma^2_{\eta_{jk}} \omega_j + \sum_{m=1}^{p_2} b^{(2)}_{m}(t) b^{(2)}_{m}(t') \sum_{\ell=1}^{p_1} \sigma^2_{\zeta_{\ell m}} \Omega_{\ell},
\end{aligned}$$

where $\omega_j = \int_{\mathcal{S}} \ddot{\psi}_j(s)\ddot{\psi}_j(s)ds$ and $\Omega_{\ell} = \int_{\mathcal{S}} b^{(1)}_{\ell}(s) b^{(1)}_{\ell}(s)ds$. Computing the marginal covariance functions using the above expressions has much better scalability than first calculating the covariance kernel $K\{(s,t),(s',t')\}$ and computing the required integrals $\int_{\mathcal{T}} K\{(s,t),(s',t)\}dt$ and $\int_{\mathcal{S}} K\{(s,t),(s,t')\}ds$. The 'trapz' function the R package 'pracma' is used to numerically approximate ω_m, ω_j, Ω_m, and Ω_{ℓ}.

Next, we detail the alignment algorithm for the posterior longitudinal and functional eigenfunctions. Let $g^{(m)}(u)$ denote the mth posterior sample, $m = 1, \ldots, M$, of either the jth longitudinal or kth functional eigenfunction, $\psi_j(s)$, $j = 1, \ldots, J$, or $\phi_k(t)$, $k = 1, \ldots, K$, respectively. To align the sign of the longitudinal and functional eigenfunction estimates across the MCMC samples, we utilize the below alignment algorithm. In the proposed alignment algorithm, $\{g^{(1)^\star}(u), \ldots, g^{(M)^\star}(u)\}$ denotes the aligned posterior sample with $g(m)^\star(u) = a^{(m)} g(m)(u)$ and $a^{(m)} \in \{-1, 1\}$, and $\overline{g}^{(m)}(u) = (1/m)\sum_{j=1}^{m} g^{(j)^\star}(u)$ denotes the ergodic mean of the aligned sample.

Algorithm 6 Alignment of the posterior longitudinal and functional eigenfunctions

```
Step 1: Set $a^{(1)} = 1$ and $g^{(1)\star}(u) = g^{(1)}(u)$.
Step 2: For $m = 2, \ldots, M$;
    a. Calculate the aligned ergodic mean $\overline{g}^{(m-1)}(u)$ and compute
        $$d_{+}^{(m)} := \left\{ \int \left( \overline{g}^{(m-1)}(u) - g^{(m)}(u) \right)^2 du \right\}^{1/2},$$
       and
        $$d_{-}^{(m)} := \left\{ \int \left( \overline{g}^{(m-1)}(u) + g^{(m)}(u) \right)^2 du \right\}^{1/2}.$$
    b. Set
        $$a^{(m)} = I\left( d_{+}^{(m)} \leqslant d_{-}^{(m)} \right) - I\left( d_{+}^{(m)} > d_{-}^{(m)} \right),$$
       and $g^{(m)\star}(u) = a^{(m)} g^{(m)}(u)$, where $I(\cdot)$ denotes the indicator function.
```

C.3 Simulation Details and Outlier Results

Eight simulation scenarios are considered to display the use of CPEs in describing the variation in the posterior samples in the presence of functional outliers and violation of the assumption of weak separability (results given in Section 3.4). In cases 3 through 8, the percent of outliers generated equal $q = 10$ or 20% of the functional sample.

Case 1 (no outliers): For case 1 with no outliers, a sample of functional data are generated according to the product FPCA model with $J = 2$ and $K = 2$ longitudinal and functional eigencomponents, respectively. More specifically, functional data is generated according to $Y_i(s_u, t_v) = \mu(s_u, t_v) + \sum_{j=1}^{J} \sum_{k=1}^{K} \chi_{ijk} \psi_j(s_u) \phi_k(t_v) + \epsilon_i(s_u, t_v)$, for $i = 1, \ldots, n = 30$ subjects at an equidistant grid of $u = 1, \ldots, n_s = 20$, and $v = 1, \ldots, n_t = 20$, longitudinal and functional time points, respectively, in the unit interval 0 to 1. The mean function is equal to $\mu(s,t) = 20\sqrt{1 - (s - 0.5)^2 - (t - 0.5)^2}$. The two mutually orthonormal longitudinal eigenfunctions are equal to $\psi_1(s) = -\sqrt{2}\cos(2\pi s)$ and $\psi_2(s) = \sqrt{2}\sin(2\pi t)$, and the two mutually orthonormal functional eigenfunctions are equal to $\phi_1(t) = -\sqrt{2}\cos(3\pi s)$ and $\phi_2(t) = -\sqrt{2}\sin(3\pi t)$. The subject-specific product scores, $(\chi_{i11}, \chi_{i12}, \chi_{i21}, \chi_{i22})^\top$, are generated from multivariate normal distributions with mean zero and covariance $V_1 = \mathrm{diag}(v_{11} = 2.557, v_{12} = 0.770, v_{21} = 0.516, v_{22} = 0.156)$. Lastly, the measurement error, $\epsilon_i(s_u, t_v)$, is generated independently from $N(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2 = 5$.

Case 2 (violation of weak separability): For case 2, functional observations are generated that violate the assumption of weak separability as $Z_i(s,t) = \mu(s,t) + \sum_{j=1}^{2}\sum_{k=1}^{2}\{\chi_{ijk}^{*}\psi_j(s)\phi_k(t)\} + \epsilon_i(s,t)$, where subject-specific product scores, $(\chi_{i11}^{*}, \chi_{i12}^{*}, \chi_{i21}^{*}, \chi_{i22}^{*})^\top$, are generated from multivariate normal distributions with mean zero and covariance matrix

$$V_2 = \begin{bmatrix} 2.604 & 0.829 & 0.616 & -0.033 \\ 0.829 & 0.912 & -0.264 & 0.175 \\ 0.616 & -0.264 & 0.666 & 0.072 \\ -0.033 & 0.175 & 0.072 & 0.283 \end{bmatrix}.$$

The covariance matrix V_2 violates the assumption of weak separability as the product scores generated are correlated across the longitudinal and functional dimensions.

Case 3 (magnitude outlier): Case 3 generates magnitude outliers by adding a constant deviation to the mean function (with a random sign): $Z_i(s,t) = Y_i(s,t) + \omega_i W(s)W(t)$ for $s > S_i$ and $t > T_i$ and $Z_i(s,t) = Y_i(s,t)$ otherwise, where $S_i \sim \mathrm{Unif}[0,1]$, $T_i \sim \mathrm{Unif}[0,1]$, $W(s) = \sqrt{20}, s \in [0,1]$, $W(t) = \sqrt{20}, t \in [0,1]$, and ω_i is generated as a discrete variable with values -1 or 1 with probability $1/2$.

Case 4 (amount of variation outlier): Case 4 generates amount of variation outliers by using larger variance of the product scores: $Z_i(s,t) = \mu(s,t) + \sum_{j=1}^{2}\sum_{k=1}^{2}\xi_{iik}\psi_j(s)\phi_k(t) + \epsilon_i(s,t)$, where $(\xi_{i11}, \xi_{i12}, \xi_{i21}, \xi_{i22})^\top$ are generated from a multivariate normal distribution with mean 0 and covariance $V_3 = \mathrm{diag}(v_{11} = 8.959, v_{12} = 5.434, v_{21} = 4.917, v_{22} = 2.982)$.

Case 5 (time-shifted longitudinal eigenfunction outlier): Case 5 generates time-shifted longitudinal eigenfunction outliers: $Z_i(s,t) = \mu(s,t) + \sum_{j=1}^{2}\sum_{k=1}^{2}\chi_{ijk}\kappa_j(s)\phi_k(t) + \epsilon_i(s,t)$, where $\kappa_1(s) = -\sqrt{2}\cos\{2\pi(s - 1/8)\}$ and $\kappa_2(s) = \sqrt{2}\sin\{2\pi(s - 1/8)\}$.

Case 6 (higher-frequency longitudinal eigenfunction outlier): For case 6, higher-frequency longitudinal eigenfunction outliers are generated according to $Z_i(s,t) = \mu(s,t) + \sum_{j=1}^{2}\sum_{k=1}^{2}\{\chi_{ijk}$ $\kappa_j(s)\phi_k(t)\} + \epsilon_i(s,t)$, where $\kappa_1(s) = -\sqrt{2}\cos(4\pi s)$ and $\kappa_2(s) = \sqrt{2}\sin(4\pi s)$.

Case 7 (time-shifted functional eigenfunction outlier): Case 7 generates time-shifted functional eigenfunction outliers: $Z_i(s,t) = \mu(s,t) + \sum_{j=1}^{2}\sum_{k=1}^{2}\chi_{ijk}\psi_j(s)\kappa_k(t) + \epsilon_i(s,t)$, where $\kappa_1(t) = -\sqrt{2}\cos\{3\pi(t-1/9)\}$ and $\kappa_2(t) = -\sqrt{2}\sin\{3\pi(t-1/9)\}$.

Case 8 (higher-frequency functional eigenfunction outlier): For case 8, higher-frequency functional eigenfunction outliers are generated according to $Z_i(s,t) = \mu(s,t) + \sum_{j=1}^{2}\sum_{k=1}^{2}\{\chi_{ijk}$ $\psi_j(s)\kappa_k(t)\} + \epsilon_i(s,t)$, where $\kappa_1(t) = -\sqrt{2}\cos(5\pi t)$ and $\kappa_2(t) = -\sqrt{2}\sin(5\pi t)$.

The magnitude outliers in the third simulation scenario add a constant deviation from the mean function over a random portion of the longitudinal and functional time domains ($s \in [S_i, 1]$ and $t \in [T_i, 1]$) with a random sign. Due to the random sign of the constant deviation, rather than biasing the mean function, they increase the variation in the mean function estimation (IMSE values for mean estimation are higher for Case 3 than other simulation scenarios in Table 3.2 and C.1). This can also be seen in the MVD-CPEs of the mean function in Figure C.4 compared to the MVD-CPES for Case 1 in Figure C.3 as there is greater variation in the posterior samples when $t > 1/2$ that increases in severity as we go across the fixed slices of s, particularly so after $s > 1/2$. The addition of the constant deviation also leads to identifiability issues in longitudinal and functional eigenfunction estimation, where CPEs of both the first longitudinal and functional eigenfunctions portray the constant variation, especially in the second half of the unit time domain. MBD-CPEs up to $\alpha = 0.65$, MVD-CPEs up to $\alpha = 0.45$, and kernel MVD-CPEs up to $\alpha = 0.05$ are constant in the second half of the unit interval for the first longitudinal eigenfunction in Figure C.5. Similarly, MBD-CPEs up to $\alpha = 0.65$ and MVD-CPEs and kernel MVD-CPEs up to $\alpha = 0.05$ are constant in the second half of the unit interval for the first functional eigenfunction in Figure C.7. CPEs of the second longitudinal and functional eigenfunctions capture variation along the first longitudinal and functional eigenfunction as the second major direction of variation (following the constant variation in the second half of the unit time interval) (Figure C.5 and C.8). This leads to higher IMSE values in both longitudinal and functional eigenfunction estimation for Case 3 compared to other simulation cases. Note that the traditional point estimates $\widehat{\psi}_j(s)$ and $\widehat{\phi}_k(t)$ have the smallest IMSE values for Case 3, possibly due to the cancellation of the effects of the constant deviation with a random sign while averaging over the posterior samples.

Case 4 generates outliers with larger longitudinal and functional eigenvalues, which increase the variation along the longitudinal and functional eigenfunctions. Due to the shapes of the four eigenfunctions considered, this adds variation across the entire unit interval in both longitudinal and functional dimensions, as is detected through the CPEs in Figures C.5 and C.6(d)-(f) and wider CPEs in C.7 and C.8(d)-(f). While this additional variation does not bias the point estimates of the longitudinal and functional eigenfunctions as can be seen in point estimates given in Figures C.1 and C.2, respectively, the IMSE in estimation of the eigenfunctions are higher than Case 1 with no outliers due to the added variation. Tables 3.3 and C.2 show that the added variation leads to bias in the eigenvalues, τ_j and ϑ_k, and, as a consequence, bias in the marginal covariances, $K_{\mathcal{S}}(s, s')$ and $K_{\mathcal{T}}(t, t')$, and covariance kernel, $K\{(s, t), (s', t')\}$, with larger IMSE values compared to Case 1.

Cases 5 and 7 consider the direction of variation outliers in which the functional data are generated under either longitudinal (Case 5) or functional (Case 7) eigenfunctions shifted in time. Case 5 leads to added variation in the estimation of the longitudinal eigenfunctions, $\psi_j(s)$, throughout the longitudinal unit time interval (Figures C.5 and C.6(g)-(i)) and impacts estimation of the longitudinal marginal covariance, $K_{\mathcal{S}}(s, s')$. Case 7 leads to added variation in the estimation of the functional eigenfunctions, $\phi_k(t)$, throughout the functional unit time interval (Figures C.7 and C.8(m)-(o)) and impacts estimation of the functional marginal covariance, $K_{\mathcal{T}}(t, t')$. This is consistent with higher IMSE for the longitudinal eigenfunctions and marginal covariance point estimates in Case 5 and functional eigenfunctions and marginal covariance point estimates in Case 7 compared to Case 1 in Table 3.2

and C.1. Finally, direction of variation outliers in which the frequency of the longitudinal and functional eigenfunctions is increased are considered for Cases 6 and 8, respectively. The added variation due to the outliers with higher frequency particularly impacts the estimation of the second eigenfunctions $\psi_2(s)$ (Case 6) and $\phi_2(t)$ (Case 8) seen in the noticeably larger IMSE values as compared to Case 1. This is apparent in Figure C.6 for Case 6 with higher frequency posterior estimates of $\psi_2(s)$ captured in the $\alpha = 0.75$ MBD-CPES and $\alpha = 0.05$ MVD-CPES and kernel MVD-CPEs. Similarly, this is apparent in Figure C.8 for Case 8 with higher frequency posterior estimates of $\phi_2(t)$ captured in also the $\alpha = 0.75$ MBD-CPES and $\alpha = 0.05$ MVD-CPES and kernel MVD-CPEs.

C.4 Data Analysis

The visual implicit learning experiment recorded EEG for 120 trials per expected and unexpected condition for each subject in the ASD (n = 37) and TD (n = 34) groups. The EEGs were measured over 128-electrode Geodesic Sensor Net at 250Hz resulting in 250 within-trial time points per ERP over 1000ms and preprocessed using NetStation 4.4.5 software (Electrical Geodesics, Inc.). Further details on artifact detection, bad channel replacement, referencing, filtering, smoothing, and base-line corrections can be found in Hasenstab et al. (2015). Prior to fitting the B-LFPCA model for the ASD and TD groups separately, 4 ASD subjects and 2 TD subjects were removed. Subjects with more than 25% of their longitudinal observations missing were removed leading to leading to 3 ASD subjects and 1 TD subject. Then, subjects, where more than 50% of the longitudinal data were at least 2 standard deviations away from the group-level mean condition differences, were removed leading to 1 ASD and 1 TD subject. After the removal of the 6 subjects, there were a total of 27 missing longitudinal observations across 6 subjects in the ASD group and 24 missing longitudinal observations across 9 subjects in the TD group. The missing observations were imputed as an update in the MCMC chain and treated as observed data in the Gibbs sampler, further details of which can be found in Shamshoian et al. (2022). The MCMC chain is employed for the ASD and TD groups with 50,000 posterior draws, 20,000 of which are used for burn-in and thinning at every 5th iteration leading to $M = 6000$ posterior samples. The hyperparameters are selected as $\nu_1 = \nu_2 = 1$, $r_1 = r_2 = 1$, $a_\eta = b_\eta = 1$, $a_\zeta = b_\zeta = 0.5$, and $a_\epsilon = b_\epsilon = 10,000$.

C.5 Tables and Figures

Table C.1: The mean standardized integrated mean squared error (IMSE) for the point estimates from the 200 Monte Carlo runs. The eight simulation cases correspond to: Case 1 – no outliers, Case 2 – violation of weak separability, Case 3 – magnitude outlier, Case 4 – amount of variation outlier, Case 5/6 – time-shifted/higher-frequency longitudinal eigenfunction outlier, Case 7/8 – time-shifted/higher-frequency functional eigenfunction outlier, with $q = 10\%$ outliers.

			$q = 10\%$					
Point Estimate	Case 1	Case 2*	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
	IMSE		**IMSE**					
$\widehat{\mu}(t)$	0.0003	0.0005	0.0012	0.0005	0.0004	0.0004	0.0003	0.0004
$\widetilde{m}\{\mu(t)\}$	0.0003	0.0005	0.0012	0.0005	0.0004	0.0004	0.0004	0.0005
$\widehat{\psi}_1(s)$	0.0129	0.1555	1.3119	0.0451	0.0179	0.0141	0.0148	0.0185
$\widetilde{\psi}_1(s)$	0.0128	0.1587	1.3602	0.0459	0.0178	0.0124	0.0147	0.0185
$\widehat{m}\{\psi_1(s)\}$	0.0129	0.1593	1.3564	0.0467	0.0182	0.0128	0.0151	0.0190
$\widetilde{m}\{\psi_1(s)\}$	0.0137	0.1604	1.3534	0.0466	0.0188	0.0135	0.0155	0.0195
$\overline{m}\{\psi_1(s)\}$	0.0138	0.1602	1.3652	0.0461	0.0199	0.0146	0.0161	0.0211
$\widehat{\psi}_2(s)$	0.0258	0.1669	2.0145	0.0489	0.0292	0.1732	0.0289	0.0326
$\widetilde{\psi}_2(s)$	0.0163	0.1613	2.1153	0.0478	0.0211	0.1387	0.0182	0.0223
$\widehat{m}\{\psi_2(s)\}$	0.0172	0.1640	2.1084	0.0480	0.0220	0.1460	0.0192	0.0240
$\widetilde{m}\{\psi_2(s)\}$	0.0205	0.1670	2.1328	0.0503	0.0253	0.1605	0.0225	0.0263
$\overline{m}\{\psi_2(s)\}$	0.0211	0.1671	2.1180	0.0500	0.0266	0.1174	0.0237	0.0295
$\widehat{\phi}_1(t)$	0.0078	0.1087	1.3375	0.0224	0.0072	0.0100	0.0105	0.0082
$\widetilde{\phi}_1(t)$	0.0077	0.1099	1.3727	0.0219	0.0072	0.0101	0.0105	0.0078
$\widehat{m}\{\phi_1(t)\}$	0.0080	0.1112	1.3520	0.0222	0.0076	0.0104	0.0107	0.0083
$\widetilde{m}\{\phi_1(t)\}$	0.0083	0.1110	1.3572	0.0217	0.0079	0.0109	0.0113	0.0092
$\overline{m}\{\phi_1(t)\}$	0.0086	0.1107	1.3688	0.0245	0.0084	0.0118	0.0122	0.0098
$\widehat{\phi}_2(t)$	0.0343	0.1328	1.8349	0.0305	0.0364	0.0357	0.0289	0.2372
$\widetilde{\phi}_2(t)$	0.0295	0.1292	1.9154	0.0279	0.0327	0.0314	0.0249	0.3223
$\widehat{m}\{\phi_2(t)\}$	0.0197	0.1222	1.9095	0.0268	0.0202	0.0227	0.0198	0.2248
$\widetilde{m}\{\phi_2(t)\}$	0.0291	0.1305	1.9102	0.0295	0.0295	0.0304	0.0261	0.3195
$\overline{m}\{\phi_1(t)\}$	0.0286	0.1268	1.9389	0.0318	0.0309	0.0361	0.0251	0.3329
$\widetilde{K}_{\mathcal{S}}(s, s')$	0.0627	0.2083	5.0570	0.2467	0.0712	0.0907	0.0622	0.0688
$\widetilde{m}\{K_{\mathcal{S}}(s, s')\}$	0.0656	0.2058	4.7407	0.2178	0.0744	0.0898	0.0646	0.0715
$\overline{m}\{K_{\mathcal{S}}(s, s')\}$	0.0690	0.2089	4.5832	0.2078	0.0780	0.0950	0.0685	0.0789
$\widetilde{K}_{\mathcal{T}}(t, t')$	0.0571	0.1765	4.6149	0.1963	0.0592	0.0563	0.0679	0.0904
$\widetilde{m}\{K_{\mathcal{T}}(t, t')\}$	0.0603	0.1775	4.3438	0.1705	0.0623	0.0582	0.0725	0.0964
$\overline{m}\{K_{\mathcal{T}}(t, t')\}$	0.0650	0.1803	4.1864	0.1643	0.0663	0.0654	0.0783	0.1011
$\overline{K}\{(s, t), (s', t')\}$	0.1005	0.1922	5.2699	0.2936	0.1149	0.1282	0.1013	0.1386
$\overline{m}[K\{(s, t), (s', t')\}]$	0.1076	0.2054	4.8858	0.2664	0.1230	0.1351	0.1114	0.1503

*IMSE for point estimates of $K\{(s, t), (s', t')\}$ are calculated in comparison to the covariance kernel used for data generation in Case 2 (see Appendix C.3).

Table C.2: The mean standardized mean squared error (MSE) for the scalar point estimates from the 200 Monte Carlo runs. The eight simulation cases correspond to: Case 1 – no outliers, Case 2 – violation of weak separability, Case 3 – magnitude outlier, Case 4 – amount of variation outlier, Case 5 – time-shifted longitudinal eigenfunction outlier, Case 6 – higher-frequency longitudinal eigenfunction outlier, Case 7 – time-shifted functional eigenfunction outlier, Case 8 – higher-frequency functional eigenfunction outlier with $q = 10\%$ outliers.

			$q = 10\%$					
Point Estimate	Case 1	Case 2*	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
	MSE		**MSE**					
$\widehat{\tau}_1$	0.0488	0.0591	3.2820	0.3081	0.0469	0.0452	0.0467	0.0439
$\widetilde{\tau}_1$	0.0487	0.0592	3.2236	0.2977	0.0474	0.0472	0.0464	0.0444
$\widehat{\tau}_2$	0.0378	0.0535	4.9695	0.3635	0.0377	0.0360	0.0447	0.0455
$\widetilde{\tau}_2$	0.0426	0.0632	5.1648	0.4121	0.0427	0.0574	0.0498	0.0530
$\widehat{\vartheta}_1$	0.0424	0.0494	2.6322	0.2355	0.0454	0.0449	0.0414	0.0433
$\widetilde{\vartheta}_1$	0.0424	0.0499	2.5882	0.2282	0.0454	0.0446	0.0421	0.0452
$\widehat{\vartheta}_2$	0.0333	0.0377	12.535	0.9638	0.0350	0.0435	0.1001	0.0352
$\widetilde{\vartheta}_2$	0.0403	0.0417	13.114	1.0527	0.0420	0.0490	0.1053	0.0456

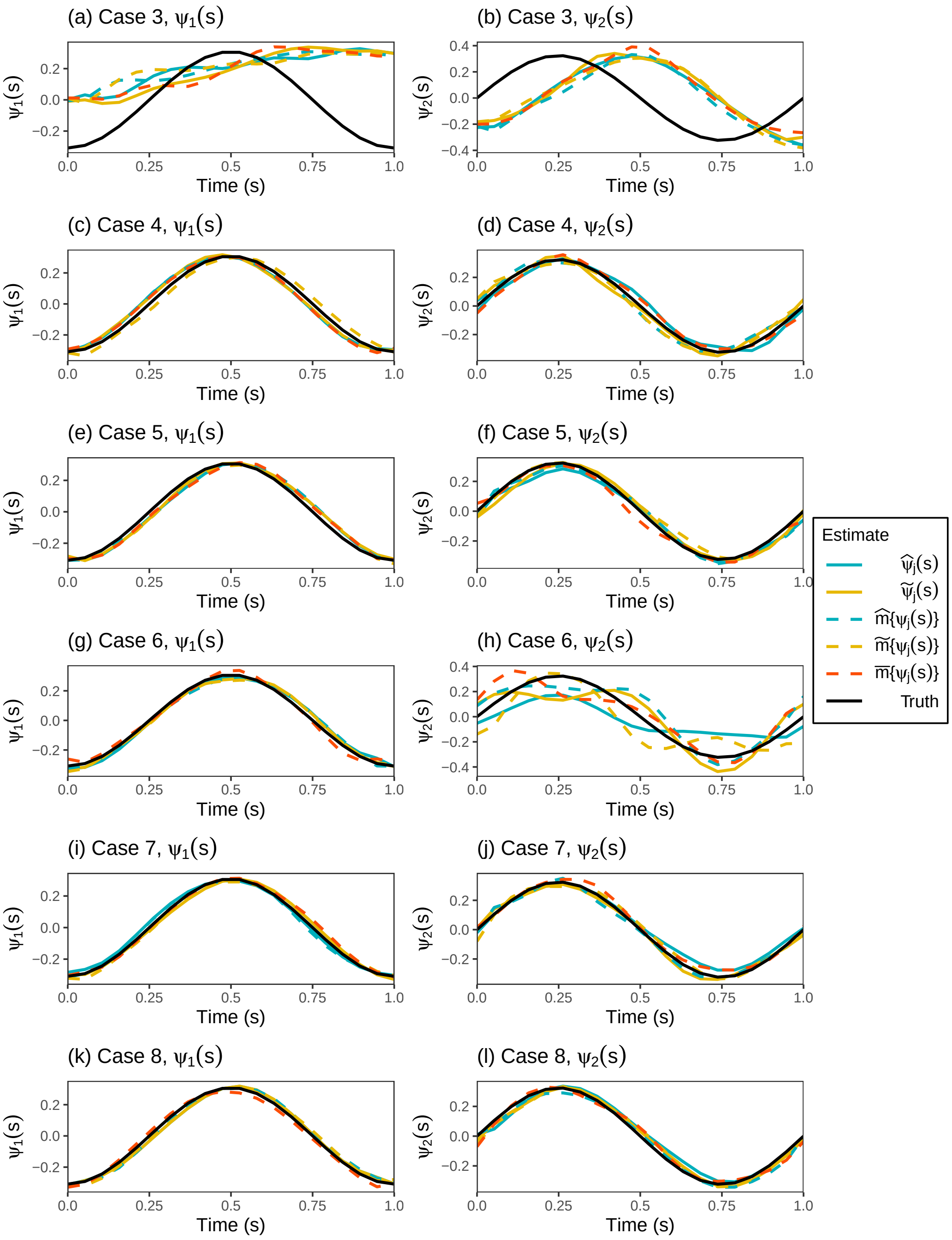

(a) Case 3, $\psi_1(s)$ (b) Case 3, $\psi_2(s)$ (c) Case 4, $\psi_1(s)$ (d) Case 4, $\psi_2(s)$ (e) Case 5, $\psi_1(s)$ (f) Case 5, $\psi_2(s)$ (g) Case 6, $\psi_1(s)$ (h) Case 6, $\psi_2(s)$ (i) Case 7, $\psi_1(s)$ (j) Case 7, $\psi_2(s)$ (k) Case 8, $\psi_1(s)$ (l) Case 8, $\psi_2(s)$

Figure C.1: Point estimates of $\psi_1(s)$ and $\psi_2(s)$ in the first and section columns, respectively, for simulation Cases 3 through 8 with $q = 20\%$ outliers from runs with 50th percentile IMSE values. Longitudinal eigenfunction estimates, longitudinal eigenfunctions estimates via longitudinal marginal covariance estimation, MBD median, MVD median, and kernel MVD median are given in solid blue, solid yellow, dashed blue, dashed yellow, and dashed red overlaying the true function given in solid black.

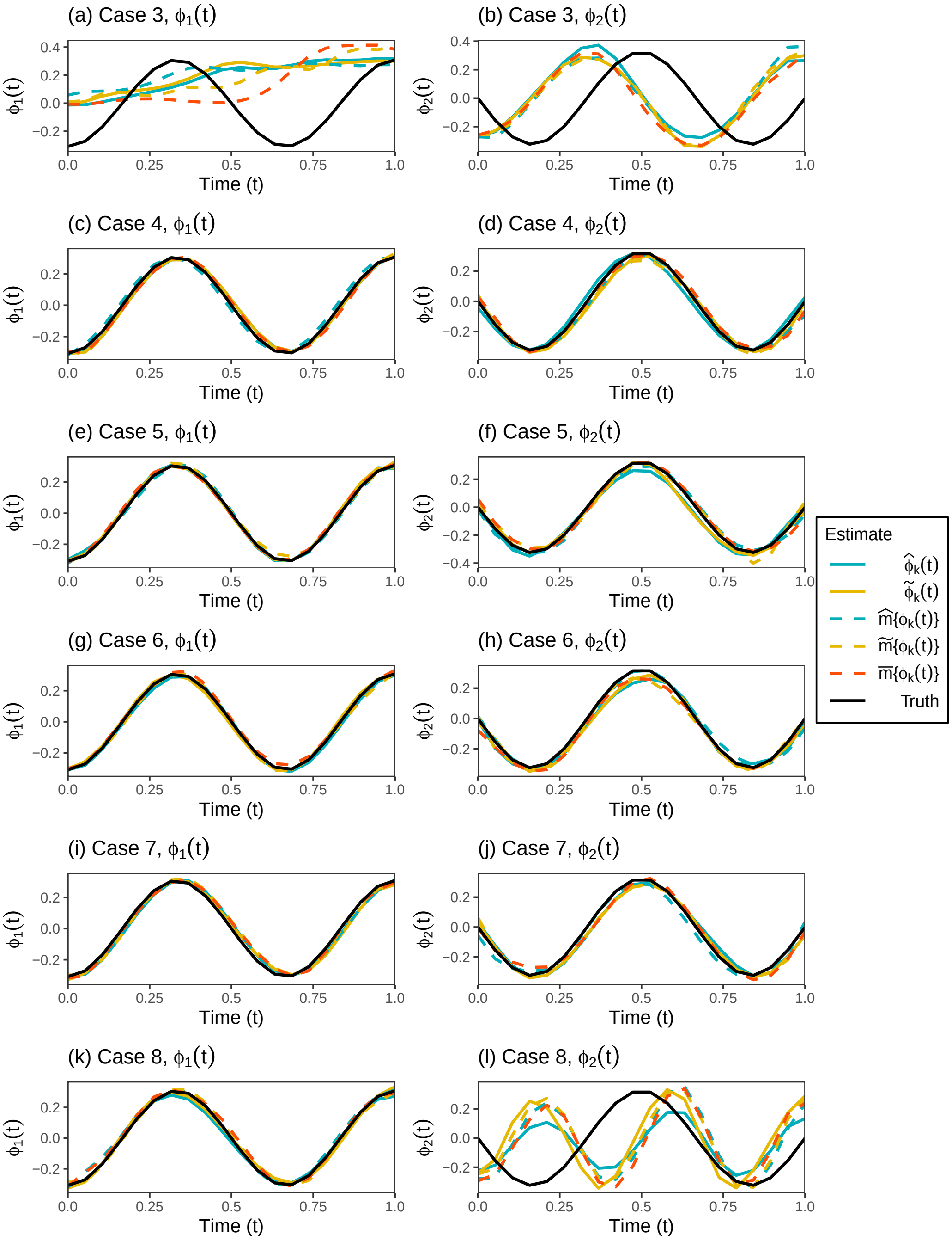


Figure C.2: Point estimates of $\phi_1(t)$ and $\phi_2(t)$ in the first and section columns, respectively, for simulation Cases 3 through 8 with $q = 20\%$ outliers from runs with 50th percentile IMSE values. Functional eigenfunction estimates, functional eigenfunctions estimates via functional marginal covariance estimation, MBD median, MVD median, and kernel MVD median are given in solid blue, solid yellow, dashed blue, dashed yellow, and dashed red overlaying the true function given in solid black.

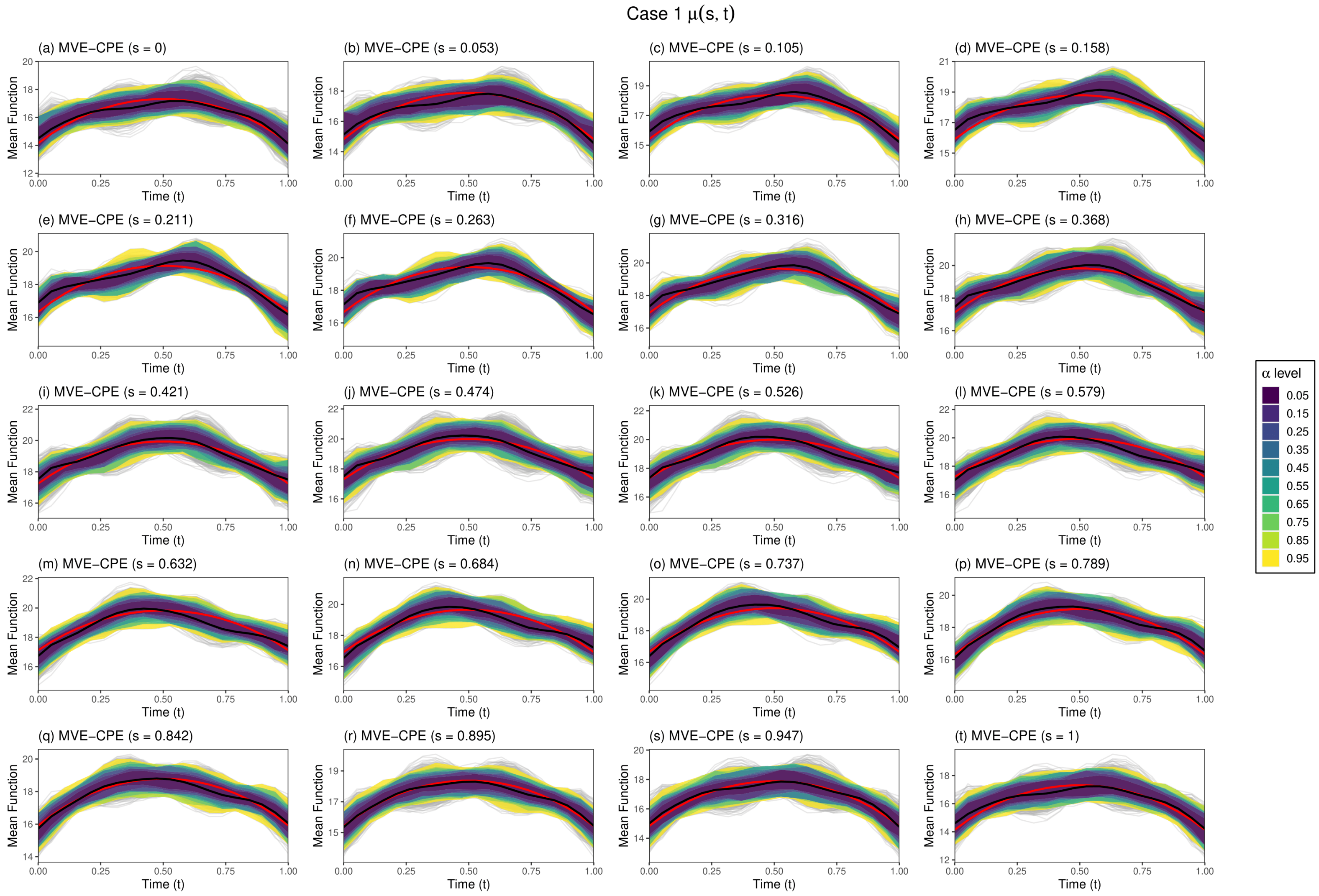


Figure C.3: MVD-CPE contours of the mean function, denoted by $D^{\star}_{1-\alpha}\{\mu(s,t)\}$, for simulation Case 1 at a grid of α levels marked by varying contour colors, overlaying the sample of $M = 6000$ posterior estimates given in gray. The MVD-CPEs of the mean function are displayed in functional time (t) at fixed slices of longitudinal time $s \in \mathcal{S}$ (i.e. $(\mu(t|s))$ for the $n_s = 20$ total time points. The solid black line represents the MVD median, denoted by $\widetilde{m}\{\mu(s,t)\}$, and the solid red line represents the true function.

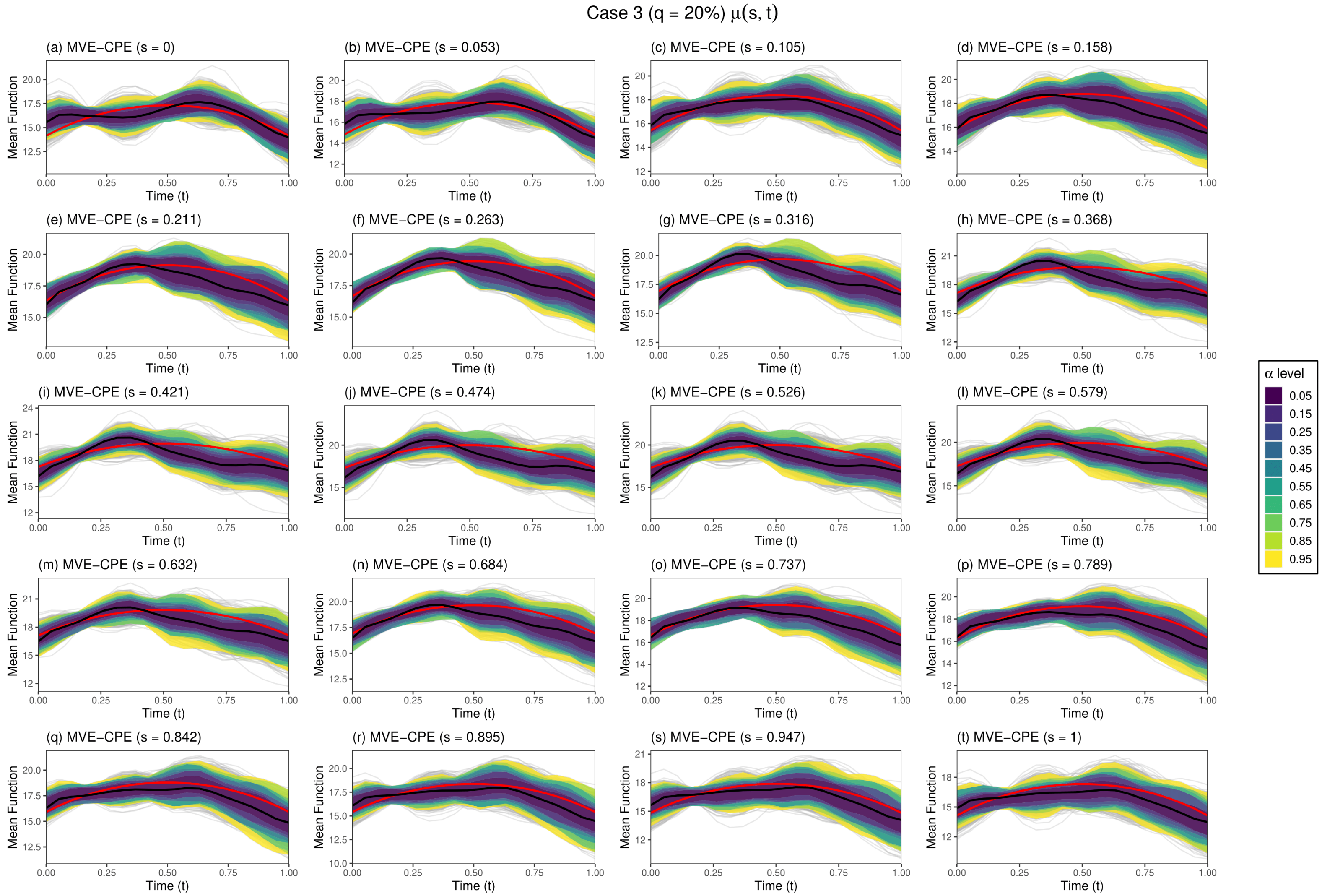


Figure C.4: MVD-CPE contours of the mean function, denoted by $D^{\star}_{1-\alpha}\{\mu(s,t)\}$, for simulation Case 3 with q = 20% outliers at a grid of α levels marked by varying contour colors, overlaying the sample of $M = 6000$ posterior estimates given in gray. The MVD-CPEs of the mean function are displayed in functional time (t) at fixed slices of longitudinal time $s \in \mathcal{S}$ (i.e. $(\mu(t|s))$ for the $n_s = 20$ total time points. The solid black line represents the MVD median, denoted by $\widetilde{m}\{\mu(s,t)\}$, and the solid red line represents the true function.

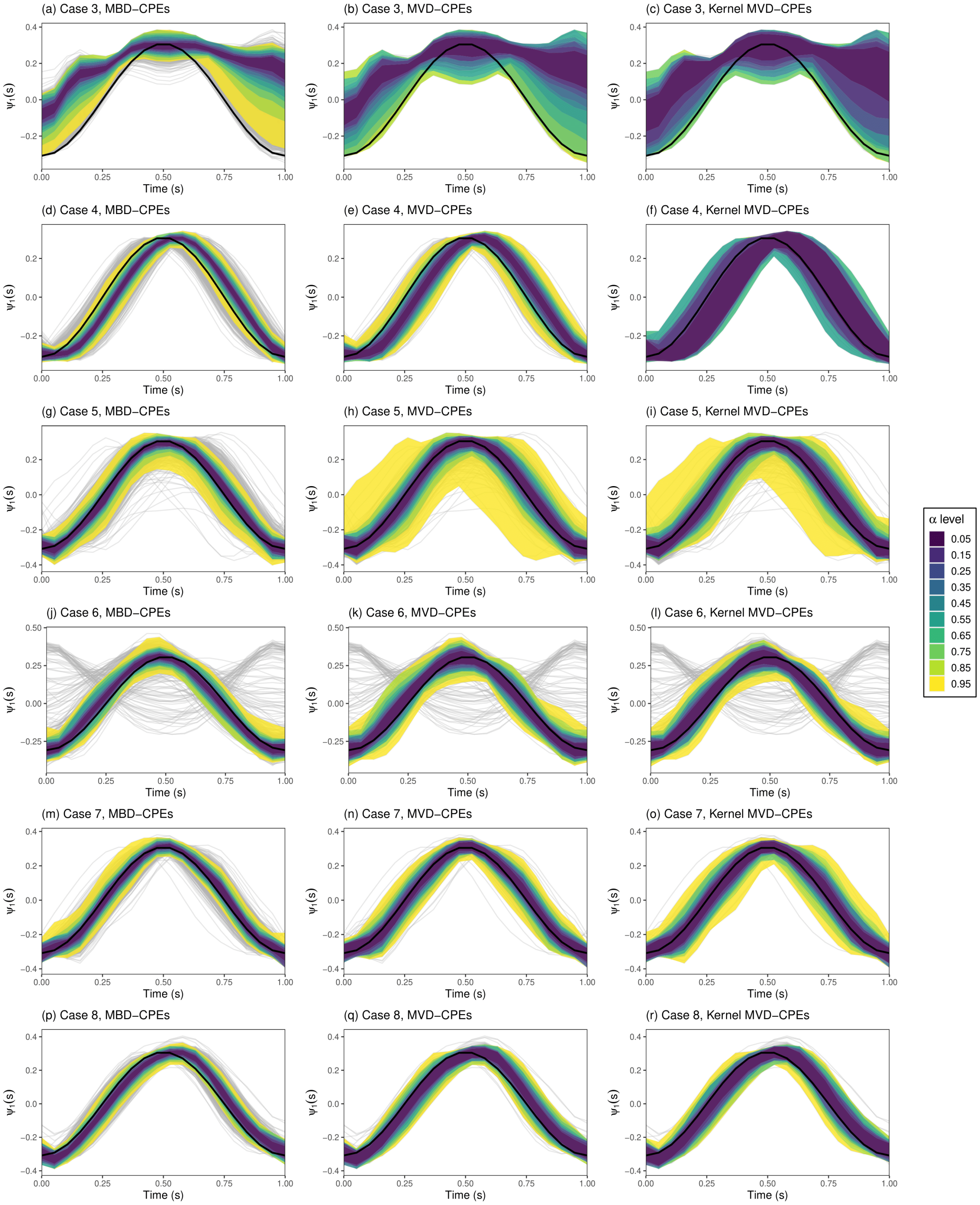


Figure C.5: CPE contours of $\psi_1(s)$ for simulation Case 3–8 with $q = 20\%$ outliers. The light grey solid lines, overlaying the true function in solid black, represent the sample of $M = 6000$ posterior estimates. The left- to right-hand columns display the MBD, MVD-CPES, and kernel MVD-CPEs, denoted by $D_{1-\alpha}\{\psi_1(s)\}$, $D^{\star}_{1-\alpha}\{\psi_1(s)\}$, and $D^{\dagger}_{1-\alpha}\{\psi_1(s)\}$, respectively, at a grid of α levels marked by varying contour colors.

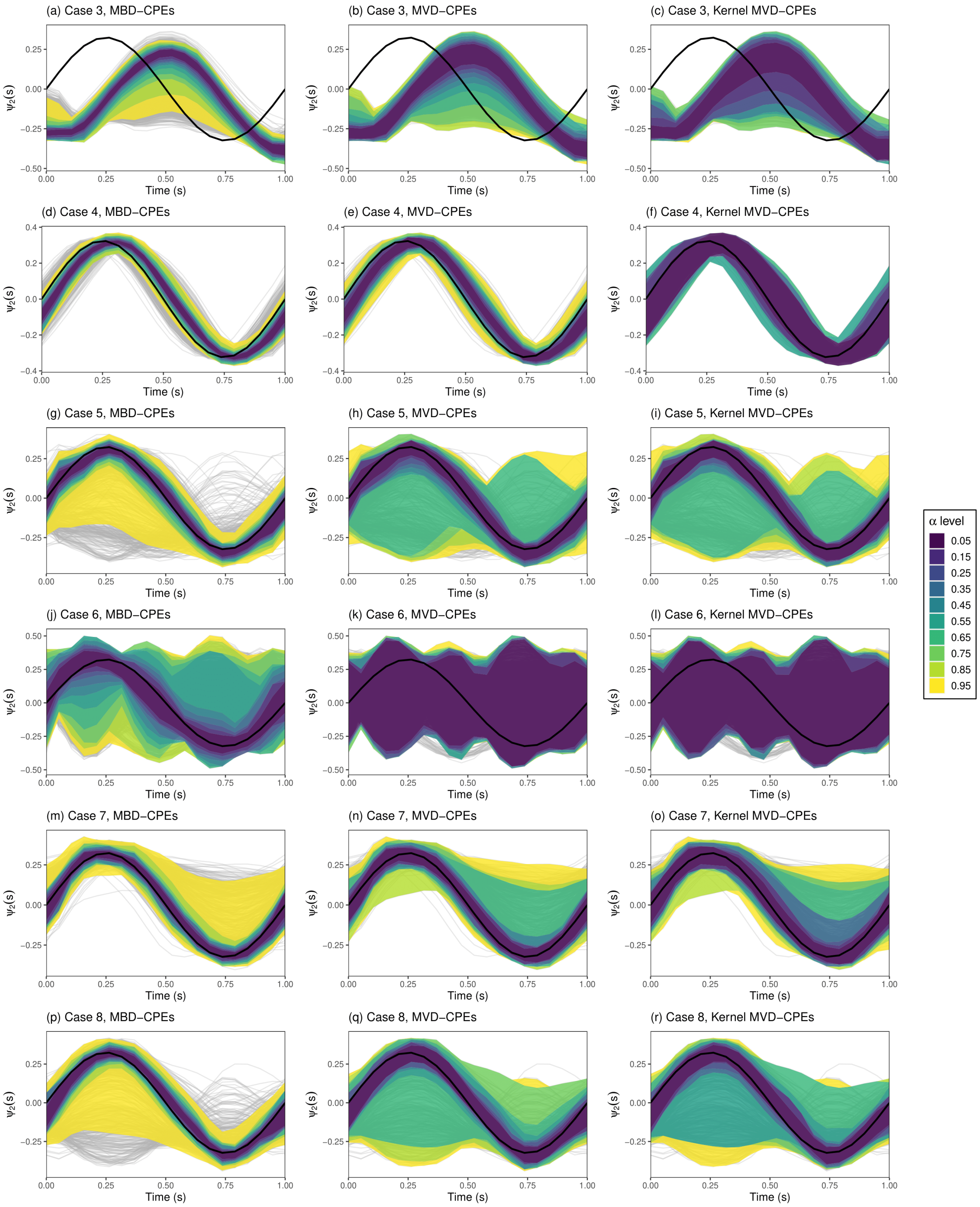


Figure C.6: CPE contours of $\psi_2(s)$ for simulation Case 3–8 with $q = 20\%$ outliers. The light grey solid lines, overlaying the true function in solid black, represent the sample of $M = 6000$ posterior estimates. The left- to right-hand columns display the MBD, MVD-CPES, and kernel MVD-CPEs, denoted by $D_{1-\alpha}\{\psi_2(s)\}$, $D^{\star}_{1-\alpha}\{\psi_2(s)\}$, and $D^{\dagger}_{1-\alpha}\{\psi_2(s)\}$, respectively, at a grid of α levels marked by varying contour colors.

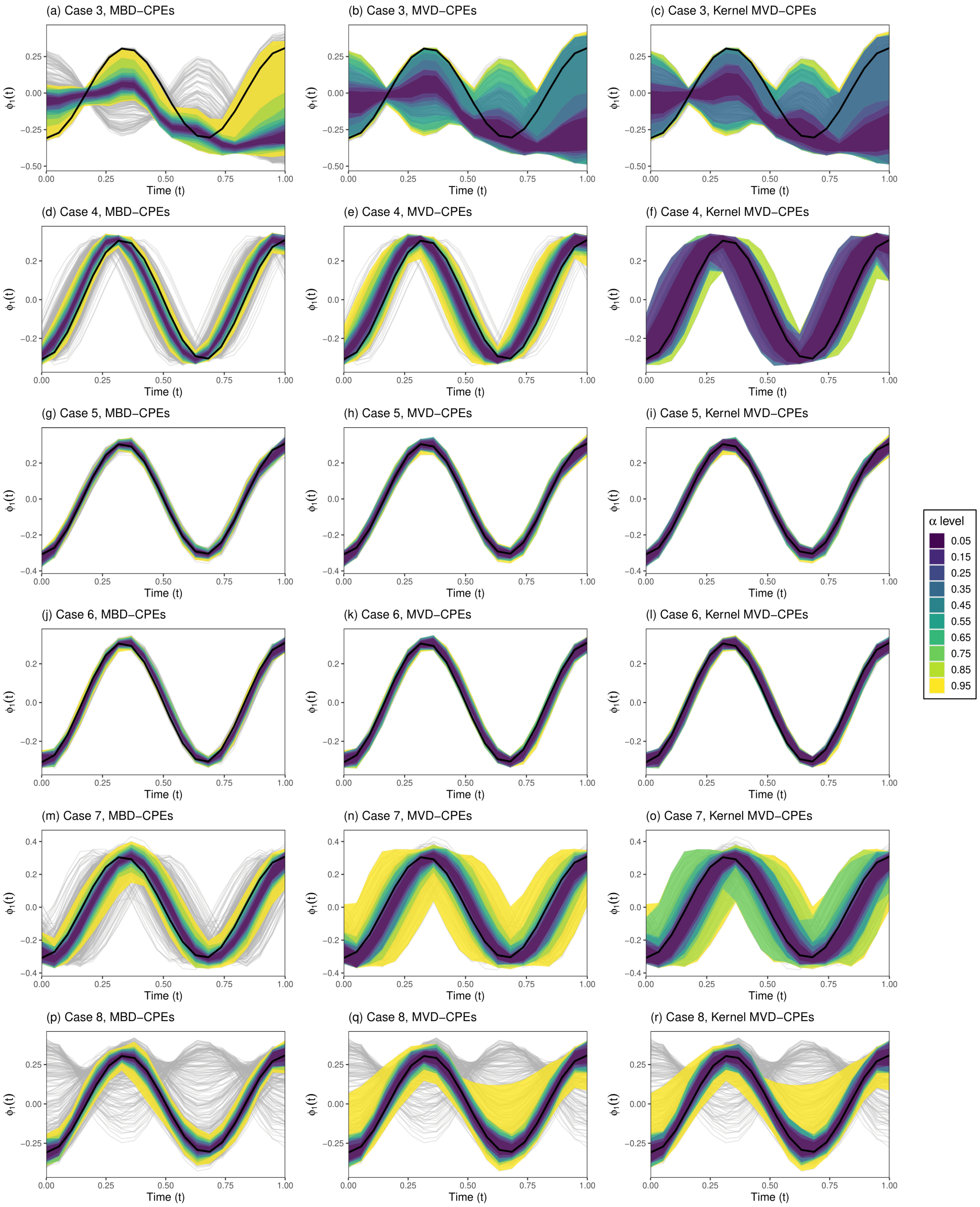


Figure C.7: CPE contours of $\phi_1(t)$ for simulation Case 3–8 with $q = 20\%$ outliers. The light grey solid lines, overlaying the true function in solid black, represent the sample of $M = 6000$ posterior estimates. The left- to right-hand columns display the MBD, MVD-CPES, and kernel MVD-CPEs, denoted by $D_{1-\alpha}\{\phi_1(t)\}$, $D^{\star}_{1-\alpha}\{\phi_1(t)\}$, and $D^{\dagger}_{1-\alpha}\{\phi_1(t)\}$, respectively, at a grid of α levels marked by varying contour colors.

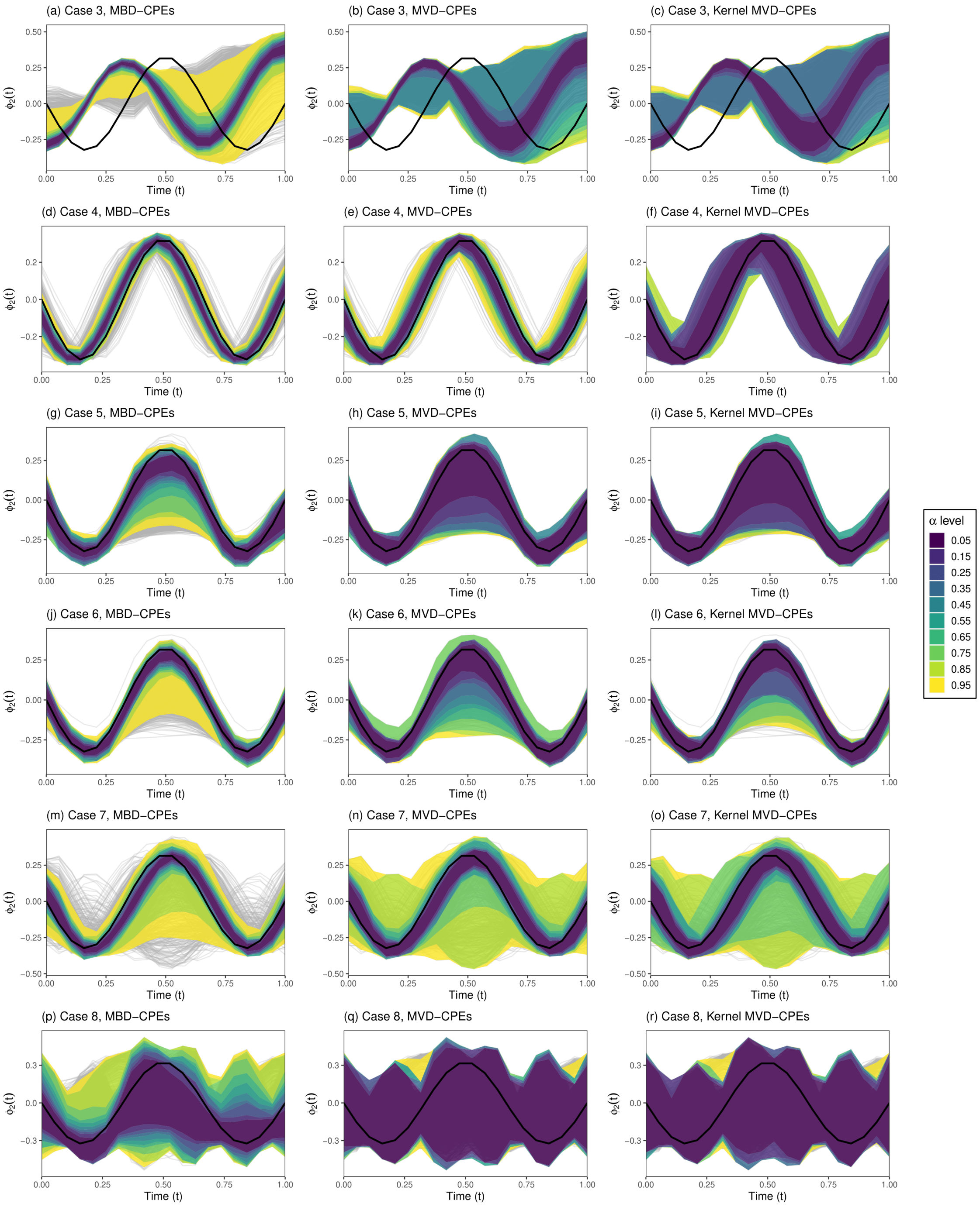


Figure C.8: CPE contours of $\phi_2(t)$ for simulation Case 3–8 with $q = 20\%$ outliers. The light grey solid lines, overlaying the true function in solid black, represent the sample of $M = 6000$ posterior estimates. The left- to right-hand columns display the MBD, MVD-CPES, and kernel MVD-CPEs, denoted by $D_{1-\alpha}\{\phi_2(t)\}$, $D^{\star}_{1-\alpha}\{\phi_2(t)\}$, and $D^{\dagger}_{1-\alpha}\{\phi_2(t)\}$, respectively, at a grid of α levels marked by varying contour colors.

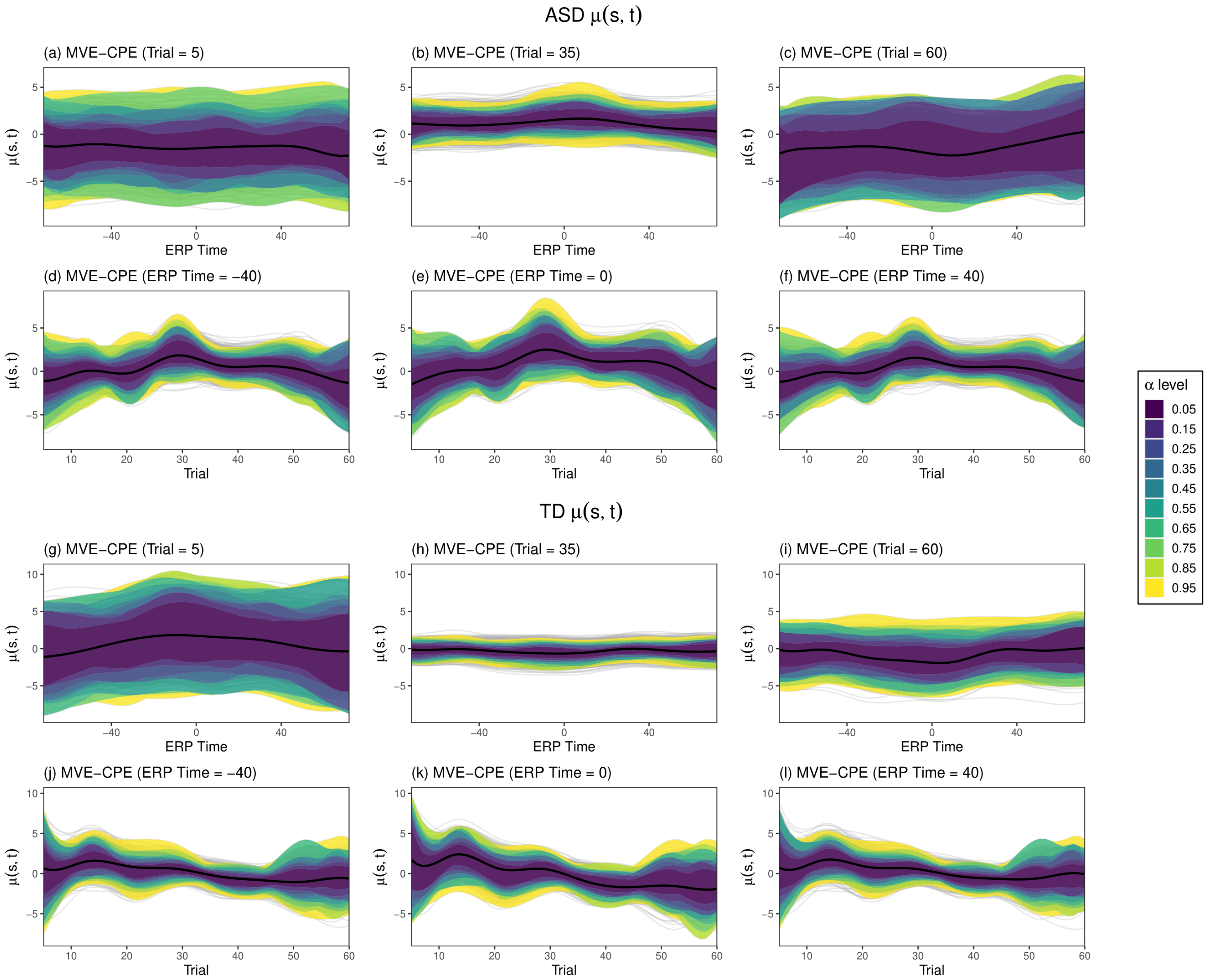


Figure C.9: MVD-CPE contours of the mean function for the ASD and TD groups at a grid of α levels marked by varying contour colors, overlaying the posterior estimates given in gray. The first and third rows display the MVD-CPEs in ERP time at fixed slices of $s = (5, 35, 60)$ (i.e. $(\mu(t|s))$, and the second and fourth rows display the MVD-CPEs in trials at fixed slices of $t = (-40, 0, 40)$ (i.e. $(\mu(s|t))$. The MVD medians are given in solid black.

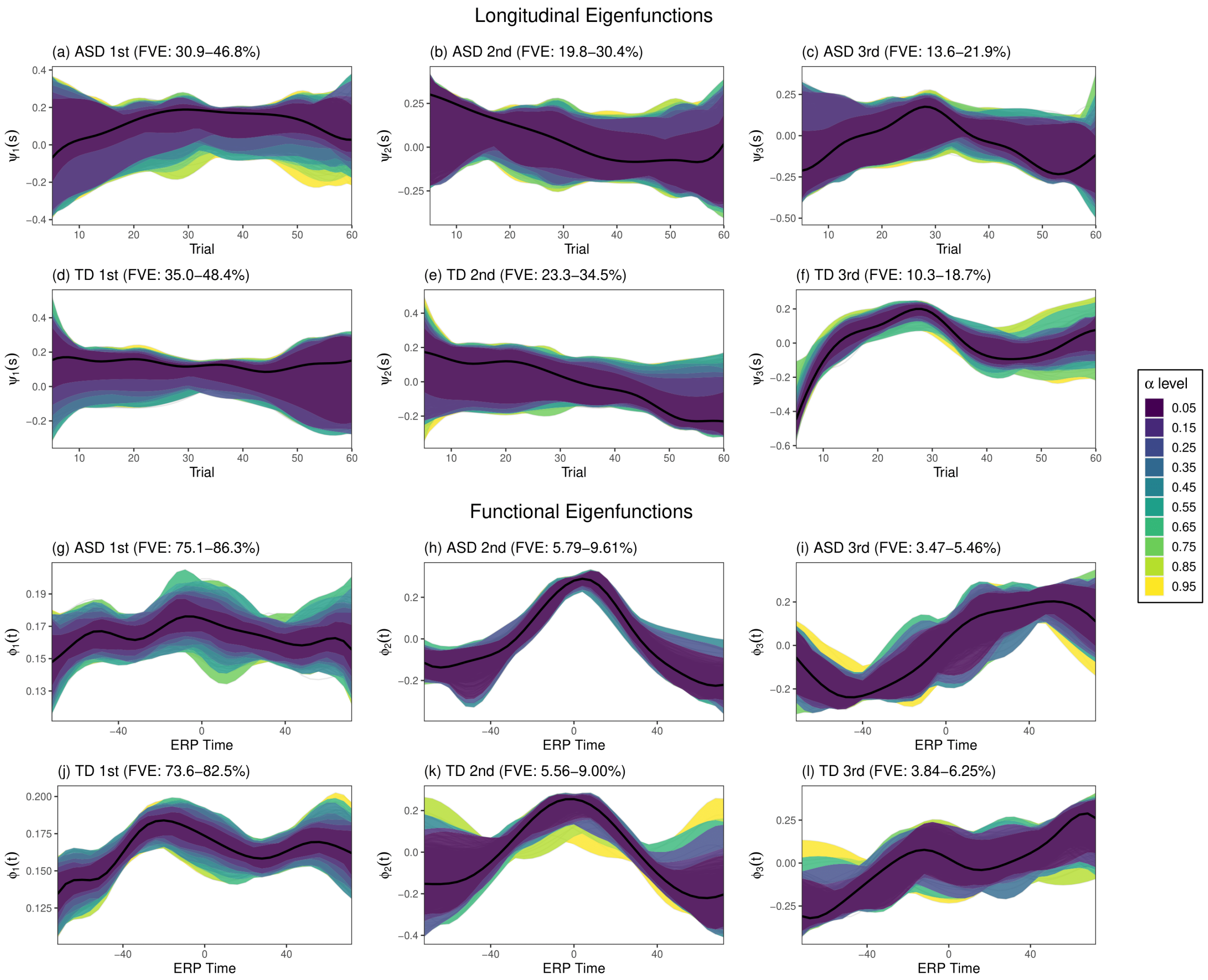


Figure C.10: MVD-CPE contours of the three leading longitudinal and functional eigenfunctions for the ASD and TD groups displayed at a grid of α levels marked by varying contour colors, overlaying the posterior estimates given in gray. The first and second rows display the MVD-CPEs of the longitudinal eigenfunctions, and the third and fourth rows display the MVD-CPEs of the functional eigenfunctions. The MVD medians are given in solid black.

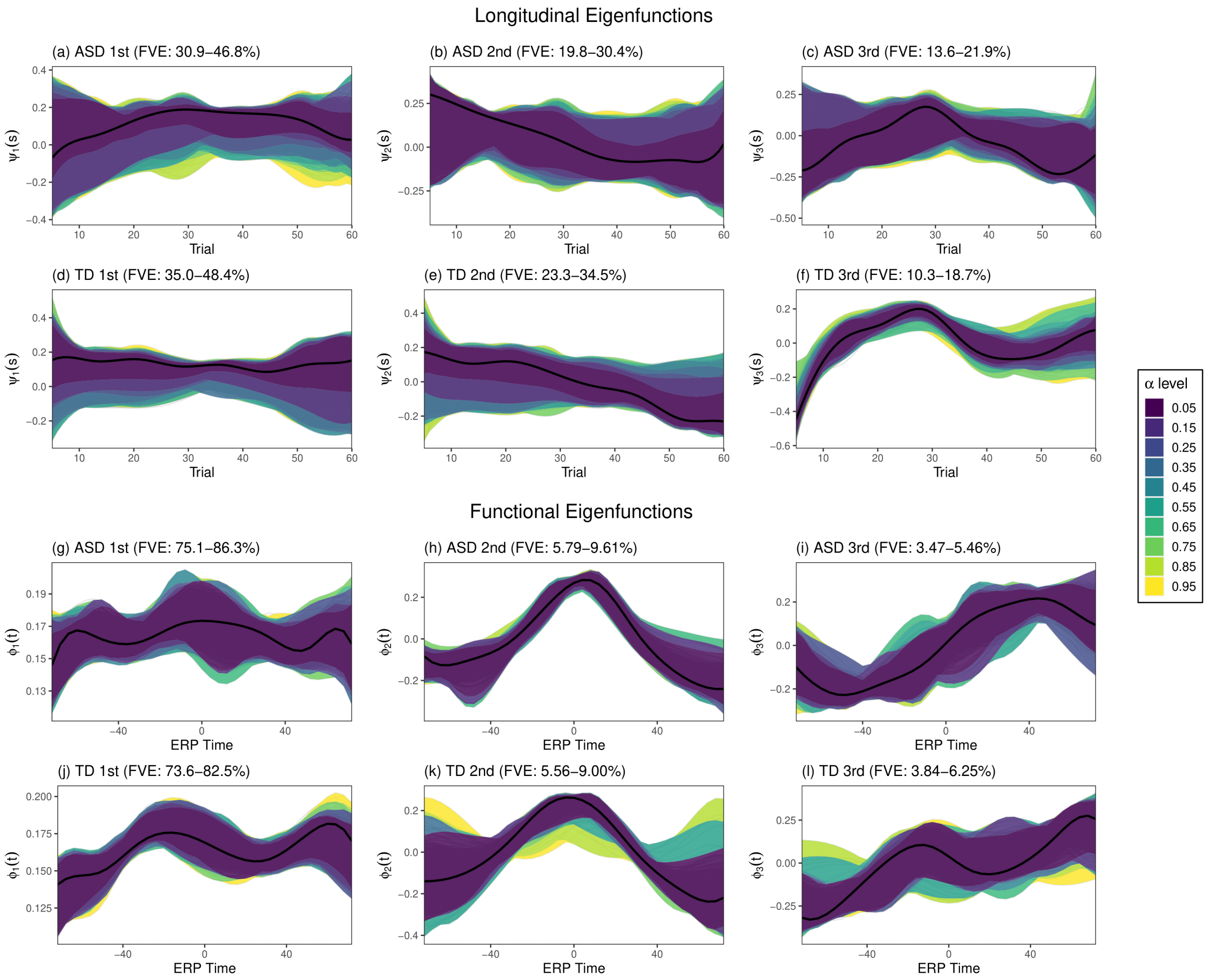


Figure C.11: Kernel MVD-CPE contours of the three leading longitudinal and functional eigenfunctions for the ASD and TD groups displayed from at a grid of α levels marked by varying contour colors, overlaying the posterior estimates given in gray. The first and second rows display the kernel MVD-CPEs of the longitudinal eigenfunctions, and the third and fourth rows display the kernel MVD-CPEs of the functional eigenfunctions. The MVD medians are given in solid black.

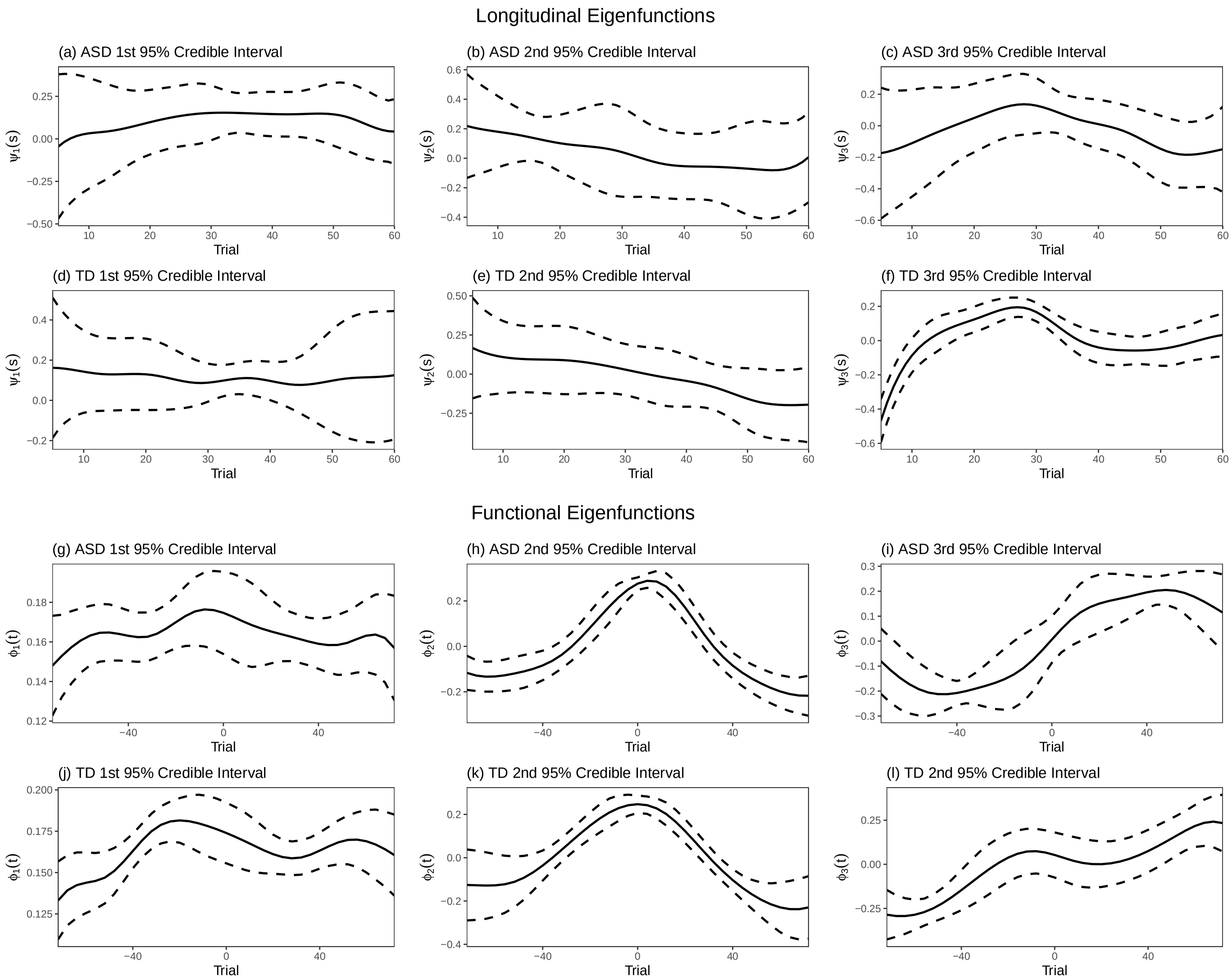


Figure C.12: 95% simultaneous parametric credible intervals represented as the solid black dashed lines of the three leading longitudinal and functional eigenfunctions displayed for ASD and TD groups. The first and second rows display the credible intervals of the longitudinal eigenfunctions, and the third and fourth rows display the credible intervals of the functional eigenfunctions. The pointwise means are given as the solid black line.

Bibliography

Arribas-Gil, A. and Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, 15(4):603–619.

Baladandayuthapani, V., Mallick, B. K., and Caroll, R. J. (2005). Spatially adaptive bayesian regression splines. *Journal of Computational and Graphical Statistics*, 14(2):378–394.

Baladandayuthapani, V., Mallick, B. K., Hong, M. Y., Lupton, J. R., Turner, N. D., and Carroll, R. J. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics*, 64(1):64–73.

Battista, T. D. and Gattone, S. A. (2004). Multivariate bootstrap confidence regions for abundance vector using data depth. *Environmental and Ecological Statistics*, 11:355–365.

Bernat, E. M., Malone, S. M., Williams, W. J., Patrick, C. J., and Iacono, W. G. (2007). Decomposing delta, theta, and alpha time-frequency ERP activity from a visual oddball task using PCA. *International Journal of Psychophysiology*, 64(1):62–74.

Bernat, E. M., Williams, W. J., and Gehring, W. J. (2005). Decomposing ERP time-frequency energy using PCA. *Clinical Neurophysiology*, 116(6):1314–1334.

Bhattacharya, A. and Dunson, D. B. (2011). Sparse bayesian infinite factor models. *Biometrika*, 98(2):291–306.

Boland, J., Telesca, D., Sugar, C. A., Jeste, S. S., Dickinson, A., DiStefano, C., and Şentürk, D. (2023). Central posterior envelopes for bayesian functional principal component analysis. *Journal of Data Science, in review.*

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

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

Capra, W. B. and Müller, H.-G. (1997). An accelerated-time model for response curves. *Journal of the American Statistical Association*, 92(437):72–83.

Cardot, H. (2007). Conditional functional principal component analysis. *Scandinavian Journal of Statistics*, 34(2):317–335.

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

Chen, K., Delicado, P., and Müller, H. G. (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.

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

Chiang, A. K. I., Rennie, C. J., Robinson, P. A., van Albada, S. J., and Kerr, C. C. (2011). Age trends and sex differences of alpha rhythms including split alpha peaks. *Clinical Neurophysiology*, 122(8):1505–1517.

Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2003). Journal of the Royal Statistical Society, Series B (Statistical Methodology). *Statistica Sinica*, 65:405–423.

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

Cragg, L., Kovacevic, N., McIntosh, A. R., Poulsen, C., Martinu, K., Leonard, G., and Paus, T. (2011). Maturation of EEG power spectra in early adolescence: a longitudinal study. *Developmental Science*, 14(5):935–943.

Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). Spatially adaptive bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics*, 16(2):265–288.

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

Dai, X., Lopez-Pintado, S., and Initiative, A. D. N. (2022). Tukey's depth for object data. *Journal of the American Statistical Association*, 0(0):1–13.

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

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

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

Dong, J., Estes, J. P., Li, G., and Şentürk, D. (2016). A two-step estimation approach for logistic varying coefficient modeling of longitudinal data. *Journal of Statistical Planning and Inference*, 174:38–51.

Dubey, D., Chen, Y., and Müller, H.-G. (2022). Depth profiles and the geometric exploration of random objects through optimal transport. *Arxiv preprint*.

Dustman, R. E., E., S. D., and Emmerson, R. Y. (1999). Life-span changes in EEG spectral amplitude, amplitude variability and mean frequency. *Clinical Neurophysiology*, 110(8):1399–1409.

Edgar, J., Heiken, K., Chen, Y., and Herrington, J. (2015). Resting-state alpha in autism spectrum disorder and alpha associations with thalamic volume. *Journal of Autism and Developmental Disorders*, 45(3):795–804.

Farge, M. (1992). Wavelet transforms and their applications to turbulence. *Annual Review of Fluid Mechanics*, 24(1):395–458.

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

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

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

Gijbels, I. and Nagy, S. (2017). On a general definition of depth for functional data. *Statistical Science*, 32(4):630–639.

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

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

Harper, J., Malone, S. M., and Bernat, E. M. (2014). Theta and delta band activity explain N2 and P3 ERP component activity in a go/no-go task. *Clinical Neurophysiology*, 125(1):124–132.

Hasenstab, K., Scheffler, A., Telesca, D., Sugar, C. A., Jeste, S., DiStefano, C., and Şentürk, D. (2017). A multidimensional functional principal component analysis of EEG data. *Biometrics*, 73(3):999–1009.

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

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

Jiang, C.-R. and Wang, J.-L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *Computational Statistics & Data Analysis*, 38(2):1194–1226.

Karhunen, K. (1946). Zur spektraltheorie stochastischer prozesse. *Annales Academiae Scientiarum Fennicae, Series A. 1*, 34:1–37.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Kotz, S., Kozubowski, T., and Podgórski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkháuser, Boston.

Krafty, R. T., Hall, M., and Guo, W. (2011). Functional mixed effects spectral analysis. *Biometrika*, 98(3):583–598.

Krafty, R. T., Rosen, O., Stoffer, D. S., Buysse, D. J., and Hall, M. H. (2017). Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology. *Journal of the American Statistical Association*, 112(520):1405–1416.

Krivobokova, T., Kneib, T., and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association*, 105(490):852–863.

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

Levin, A. R., Naples, A. J., Scheffler, A. W., Webb, S. J., Shic, F., Sugar, C. A., Murias, M., Bernier, R. A., Chawarska, K., Dawson, G., Faja, S., Jeste, S., Nelson, C. A., McPartland, J. C., and Şentürk, D. (2020). Day-to-day test-retest reliability of EEG profiles in children with autism spectrum disorder and typical development. *Frontiers in Integrative Neuroscience*, 14(21).

Li, Q., Shamshoian, J., Şentürk, D., Sugar, C., Jeste, S., DiStefano, C., and Telesca, D. (2020). Region-referenced spectral power dynamics of EEG signals: A hierarchical modeling approach. *Annals of Applied Statistics*, 14(4):2053–2068.

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

Loève, M. (1946). Décomposition harmonique des fonctions aléatoires stationnaires du second ordre. *La Revue Scientifique*, 84:159–162.

López-Pintado, S. and Qian, K. (2020). A depth-based global envelope test for comparing two groups of functions with applications to biomedical data. *Statistics in Medicine*, 40(7):1639–1652.

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

López-Pintado, S. and Wrobel, J. (2017). Robust non-parametric tests for imaging data based on data depth. *Stat*, 6(1):405–419.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., and Rutter, M. (2000). The autism diagnostic observation schedule—-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3):205—-223.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., and Rutter, M. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 38(2):179—199.

Lynch, B. and Chen, K. (2018). A test of weak separability for multi-way functional data, with application to brain connectivity studies. *Biometrika*, 105(4):815–831.

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

Montagna, S., Tokdar, S. T., Neelon, B., and Dunson, D. B. (2012). Bayesian latent factor regression for functional and longitudinal data. *Biometrics*, 68(4):1064–1073.

Motta, G. and Ombao, H. (2012). Evolutionary factor analysis of replicated time series. *Biometrics*, 68(3):825–836.

Ngo, D., Sun, Y., Genton, M. G., Wu, J., Srinivasan, R., Cramer, S. C., and Ombao, H. (2015). An exploratory data analysis of electroencephalograms using the functional boxplots approach. *Frontiers in Neuroscience*, 9(282).

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

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

Pozdnyakov, V., Elbroch, L. M., Labarga, A., Meyer, T., and Yan, J. (2019). Discretely observed Brownian motion governed by telegraph process: Estimation. *Methodology and Computing in Applied Probability*, 21(3):907–920.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, New York.

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

Scheffler, A., Telesca, D., Li, Q., Sugar, C. A., Distefano, C., Jeste, S., and Şentürk, D. (2020). Hybrid principal component analysis for region-referenced longitudinal functional EEG data. *Biostatistics*, 21(1):139–157.

Scheffler, A. W., Dickinson, A., DiStefano, C., Jeste, S. S., and Şentürk, D. (2022). Covariate-adjusted hybrid principal components analysis for region-referenced functional EEG data. *Statistics and Its Interface*, 15(2):209–223.

Scheffler, A. W., Telesca, D., Sugar, C. A., Jeste, S. S., Dickinson, A., DiStefano, C., and Şentürk, D. (2019). Covariate-adjusted region-referenced generalized functional linear model for EEG data. *Statistics in Medicine*, 38(30):5587–5602.

Shamshoian, J., Şentürk, D., and Telesca, D. (2022). Bayesian analysis of longitudinal and multidimensional functional data. *Biostatistics*, 23(2):558–573.

Shi, M., Weiss, R. E., and Taylor, J. M. G. (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(2):151–163.

Sohn, A. (2016). *acid: Analysing Conditional Income Distributions*. R package version 1.1.

Somsen, R. J., van't Klooster, B. J., van der Molen, M. W., van Leeuwen, H. M., and Licht, R. (1997). Growth spurs in brain maturation during middle childhood as indexed by EEG power spectra. *Biological Psychology*, 44(3):187–209.

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

Stroganova, T. A., Orekhova, E. V., and Posikera, I. N. (1999). EEG alpha rhythm in infants. *Clinical Neurophysiology*, 110(6):997–1012.

Suarez, A. and Ghosal, S. (2017). Bayesian estimation of principal components for functional data. *Bayesian Analysis*, 12(2):311–333.

Sun, Y. and Genton, M. G. (2012). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

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

Tierney, A. L., Gabard-Durnam, L., Vogel-Farley, V., Tager-Flusberg, H., and Nelson, C. A. (2012). Developmental trajectories of resting EEG power: an endophenotype of autism spectrum disorder. *PLoS ONE*, 7(6):e39127.

Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.

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

Welch, P. D. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73.

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

Xiao, L., Li, C., Checkley, W., and Crainiceanu, C. M. (2018). Fast covariance estimation for sparse functional data. *Statistics & Computing*, 28:511–522.

Yao, F., Müller, H.-G., and Wang, J.-L. (2012). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Yeh, A. B. and Singh, K. (1997). Balanced confidence regions based on Tukey's depth and the bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(3):639–652.

Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V., and Carroll, R. J. (2012). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association*, 105(489):390–400.

Zipunnikov, V., Caffo, B., Yousem, D. M., Davatzikos, C., Schwartz, B. S., and Crainiceanu, C. (2011). Multilevel functional principal component analysis for high-dimensional data. *Journal of Computational and Graphical Statistics*, 20(4):852–873.

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

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482.