Web-based Supplementary Materials for A Multi-Dimensional Functional Principal Components Analysis of EEG Data

Kyle Hasenstab, Aaron Scheffler, Donatello Telesca,
Catherine A. Sugar, Shafali Jeste, Charlotte DiStefano and Damla Şentürk

Web Appendix A: Estimation Algorithm

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

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

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

The decompositions given in (2) and (5) are truncated to include only the components containing the largest modes of variation. Specifically, the number of first-stage subject level (K) and subunit level (P) principal components are selected using percentage of variation explained. Let K_{s_q} and P_{s_q} be the smallest number of components satisfying the criteria $\{\sum_{k=1}^{K_{s_q}} \hat{\lambda}_k^{(1)}(s_q)\}/\{\sum_{k=1}^{M_1} \hat{\lambda}_k^{(1)}(s_q)\} > 0.9$ and $\{\sum_{p=1}^{P_{s_q}} \hat{\lambda}_p^{(2)}(s_q)\}/\{\sum_{p=1}^{M_2} \hat{\lambda}_k^{(2)}(s_q)\} > 0.9$ for first and second levels, respectively, where M_1 and M_2 are large. We select $K \equiv \max_q(K_{s_q})$ and $P \equiv \max_q(P_{s_q})$ to ensure the same number of principal components at each s_q for subsequent modeling of the first-stage eigenscores. The number of first (K'_k) and second level components (P'_p) of the second-stage decomposition are selected similarly to explain at least 90% of the total variation. Other methods for selecting the number of principal components include cross-validation (Rice and Wu, 2001) and Akaike's Information Criterion (Yao et al., 2005).

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

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

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

MD-FPCA Algorithm [`MultiLevelFuncLong.m`]

---

1. Estimate $\mu(t,s)$ with a two-dimensional smoother applied to all observed data $\{s_{ijq}, t_{ijq\ell}, X_{ij}(t_{ijq\ell}, s_{ijq}),\ i = 1,\dots,n;\ j = 1,\dots,J;\ q \in Q_{ij};\ \ell \in L_{ijq}\}$. [`MeanSmooth2D.m`]
2. Estimate $\eta_j(t,s)$ with a two-dimensional smoother applied to all mean-centered observation pairs $\{s_{ijq}, t_{ijq\ell},\ X_{ij}(t_{ijq\ell}, s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}),\ i = 1,\dots,n;\ q \in Q_{ij};\ \ell \in L_{ijq}\}$. [`MeanSmooth2D.m`]
3. First-stage Karhunen-Loève decomposition:
    (a) For each longitudinal time s_q, $q = 1,\dots,Q$, perform a multilevel FPCA for all data $\{t_{ijq\ell}, X_{ij}(t_{ijq\ell}|s_{ijq}),\ i = 1,\dots,n;\ j = 1,\dots,J;\ \ \ell \in L_{ijq}\}$ observed at $s_{ijq} = s_q$. [`MultilevelFPCA.m`]
        i. Estimate $\Sigma^{(1)}(t,t'|s_q)$ with a two-dimensional smoother applied to the products $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}\{X_{ij'}(t_{ij'q\ell'}|s_{ij'q}) - \hat{\mu}(t_{ij'q\ell'}, s_{ij'q}) - \hat{\eta}_{j'}(t_{ij'q\ell'}, s_{ij'q})\}$ over functional times $\{t_{ijq\ell},\ t_{ij'q\ell'}\}$ for all subjects and subunits from different electrodes, observed at $\ell \in L_{ijq}$, $\ell' \in L_{ij'q}$ and $j \neq j'$. [`CovarianceSmooth.m`]
        ii. Estimate $\widetilde{\Sigma}_T(t,t'|s_q)$ with a two-dimensional smoother applied to the products $\{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}\{X_{ij}(t_{ijq\ell'}|s_{ijq}) - \hat{\mu}(t_{ijq\ell'}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell'}, s_{ijq})\}$ over functional times $\{t_{ijq\ell},\ t_{ijq\ell'}\}$ for all subjects and subunits from the same electrode, observed at $\ell \neq \ell' \in L_{ijq}$. [`CovarianceSmooth.m`]
        iii. Estimate $\Sigma^{(2)}(t,t'|s_q) = \widetilde{\Sigma}_T(t,t'|s_q) - \Sigma^{(1)}(t,t'|s_q)$ by taking the difference of $\widehat{\widetilde{\Sigma}}_T(t,t'|s_q) - \widehat{\Sigma}^{(1)}(t,t'|s_q)$.
        iv. Estimate $\sigma^2_{s_q} = \{X_{ij}(t_{ijq\ell}|s_{ijq}) - \hat{\mu}(t_{ijq\ell}, s_{ijq}) - \hat{\eta}_j(t_{ijq\ell}, s_{ijq})\}^2 - \widetilde{\Sigma}_T(t_{ijq\ell}, t_{ijq\ell'}|s_q)$, observed at $\ell = \ell' \in L_{ijq}$.
    (b) Employ a FPCA on $\widehat{\Sigma}^{(1)}(t,t'|s_q)$ to estimate $\{\lambda^{(1)}_k(s_q), \phi^{(1)}_k(t|s_q)\}$.

(c) Employ a FPCA on $\widehat{\Sigma}^{(2)}(t, t'|s_q)$ to estimate $\{\lambda_p^{(2)}(s_q), \phi_p^{(2)}(t|s_q)\}$.

(d) For each longitudinal time s_q, $q = 1, \ldots, Q$, let K_{s_q} and P_{s_q} be the smallest number of components satisfying the criteria $\{\sum_{k=1}^{K_{s_q}} \hat{\lambda}_k^{(1)}(s_q)\}/\{\sum_{k=1}^{M_1} \hat{\lambda}_k^{(1)}(s_q)\} > 0.9$ and $\{\sum_{p=1}^{P_{s_q}} \hat{\lambda}_p^{(2)}(s_q)\}/\{\sum_{p=1}^{M_2} \hat{\lambda}_k^{(2)}(s_q)\} > 0.9$, where M_1, M_2 are large.

(e) Set $K = \max_q(K_{s_q})$ and $P = \max_q(P_{s_q})$.

(f) For each longitudinal time s_q, $q = 2, \ldots, Q$, determine the sign of $\hat{\phi}_k^{(1)}(\cdot|s_q)$ by minimizing the L^2 distance to the previous eigenfunction $\hat{\phi}_k^{(1)}(\cdot|s_{q-1})$. Repeat the procedure for $\hat{\phi}_p^{(2)}(\cdot|s_q)$.

(g) Estimate $\xi_{ik}(s_q)$ and $\zeta_{ijp}(s_q)$ using their multilevel best linear unbiased predictors (BLUP). [`ComputeScores.m`]

4. Second-stage Karhunen-Loève decomposition:

(a) Perform a two-dimensional smooth of the raw covariances $\{\hat{\xi}_{ik}(s_q)\hat{\xi}_{ik}(s_{q'})\}$ over $\{s_q, s_{q'}\}$, $q, q' = 1, \ldots, Q$, and $\{\hat{\zeta}_{ijp}(s_q)\hat{\zeta}_{ij'p}(s_{q'})\}$ over $\{s_q, s_{q'}\}$, $q = 1, \ldots, Q; j, j' = 1, \ldots, J$.

(b) Employ a FPCA on the smooth covariance surface obtained in 4 (a) to estimate $\{\lambda_{kk'}, \psi_{kk'}^{(1)}(s)\}$.

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

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

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

Web Appendix B: The Meta-preprocessing

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

Web Appendix C: Additional Data Analysis Results: Electrode Level Variation and Subject-specific Eigenscores

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

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

ASD is a highly heterogeneous disorder. In order to study subgroups with similar learning patterns within the diagnostic groups, we plot the two leading subject level eigenscores from the second-stage of MD-FPCA in Web Figures 9 (a) and (b) for the ASD and TD groups (the subject level eigenscores from the first-stage decompositions are displayed in Web Figure 8). In addition, the smoothed subject-specific amplitude difference trajectories (averaged over the four electrodes) across trials at the peak location ($t = 0$) partitioned by the median of the leading scores ξ'_{i11} are given in Web Figures 9 (c) and (d) for ASD and TD groups, respectively. The variation in the leading eigenscore in the ASD group represents the largest component of the total variation in the ASD group that is observed in the intermediate and later trials. While no major clusters seem to emerge in either diagnostic group, there seems to be a small cluster around zero in the ASD group. Specifically the small cluster around zero observed in Web Figure 9 (a), falling in the below median partition, are plotted in black in Web Figure 9 (c) with minimal variation along trials, signaling little or no implicit learning. The second group plotted in gray shows a positive condition differentiation at intermediate trials, implying implicit learning is taking place. In contrast, the variation in the leading eigenscore in the TD group represents variation at early and intermediate trials, corresponding to two groups in TD with positive (plotted in gray) and negative (plotted in black) condition differentiation. These subgroups with distinct patterns of condition differentiation at the P3 peak location within the two diagnostic groups are similar to insights gained in our previous works (Hasenstab et al. (2016)) based on a clustering of subjects according to longitudinal trends over trials, ignoring the functional dimension.

Web Appendix D: Simulation

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

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

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

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

$\int f(s)^2 ds$ and $||f(t,s)||^2 = \int\int f(t,s)^2 dsdt$, respectively. For assessment of estimated scalar model components, we utilize mean squared error (MSE).

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

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

References

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

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

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

Hasenstab, K., Sugar, C., Telesca, D., Jeste, S. and Şentürk, D. (2016) Robust functional clustering of longitudinal ERP trends. *Biostatistics*, **17(3),** 484–498.

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

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

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

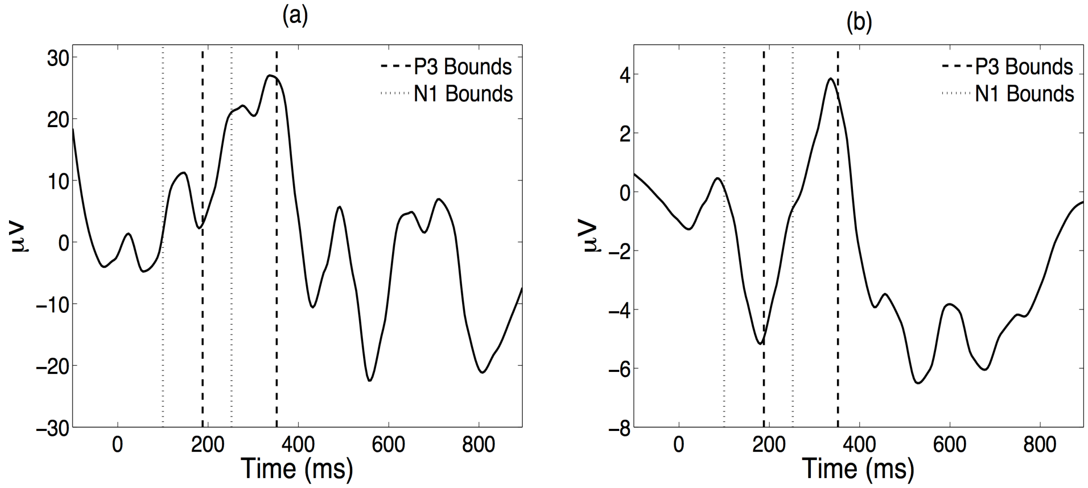
Yao, F., Müller, H. and Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100,** 577–590.

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

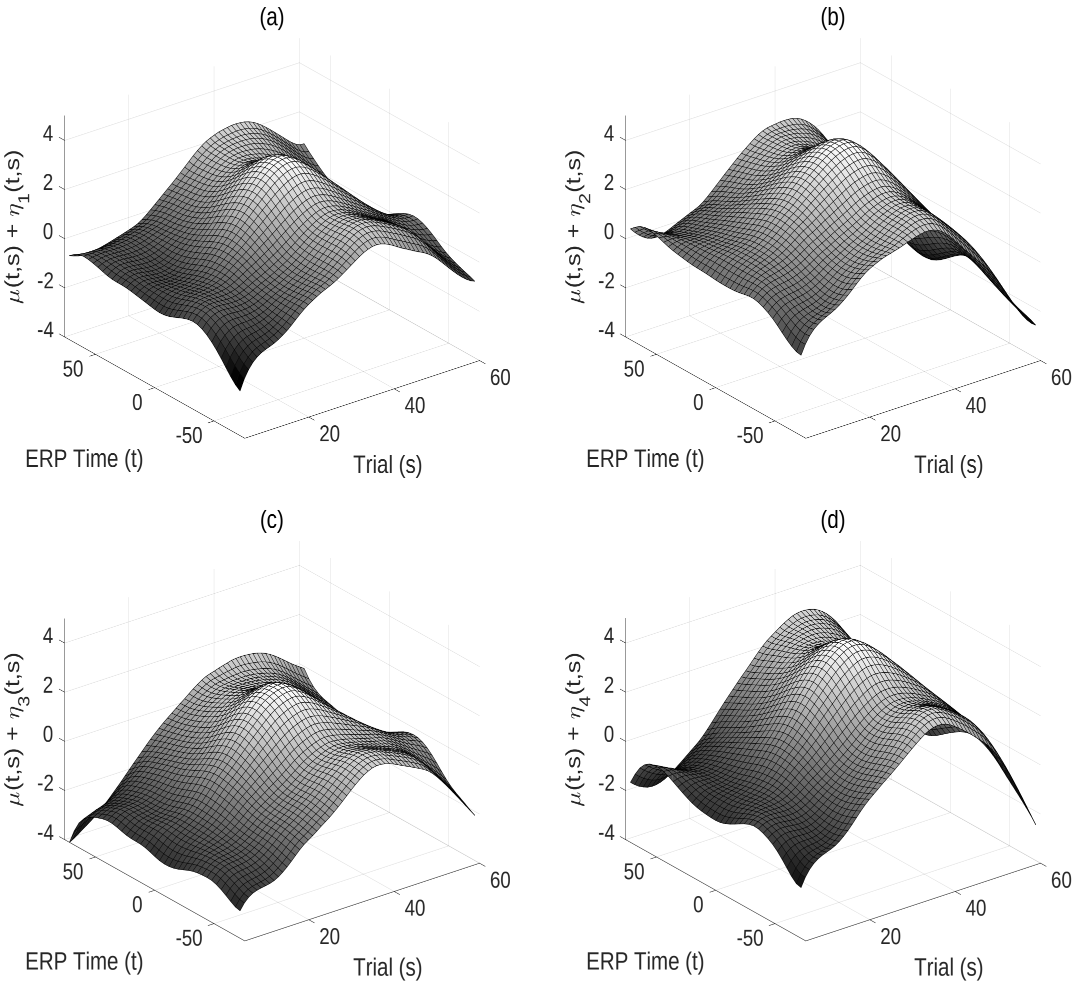
	Dense		Sparse	
	N=30	N=100	N=30	N=100
Low SNR				
$\mu(t, s)$	.002 (.001, .004)	.001 (.000, .001)	.002 (.001, .004)	.001 (.000, .001)
$X_{ij}(t\|s)$	.134 (.130, .138)	.131 (.129, .134)	.140 (.136, .145)	.134 (.131, .138)
ρ(s)	.013 (.005, .032)	.003 (.001, .008)	.023 (.014, .038)	.006 (.003, .010)
ρ	.003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

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

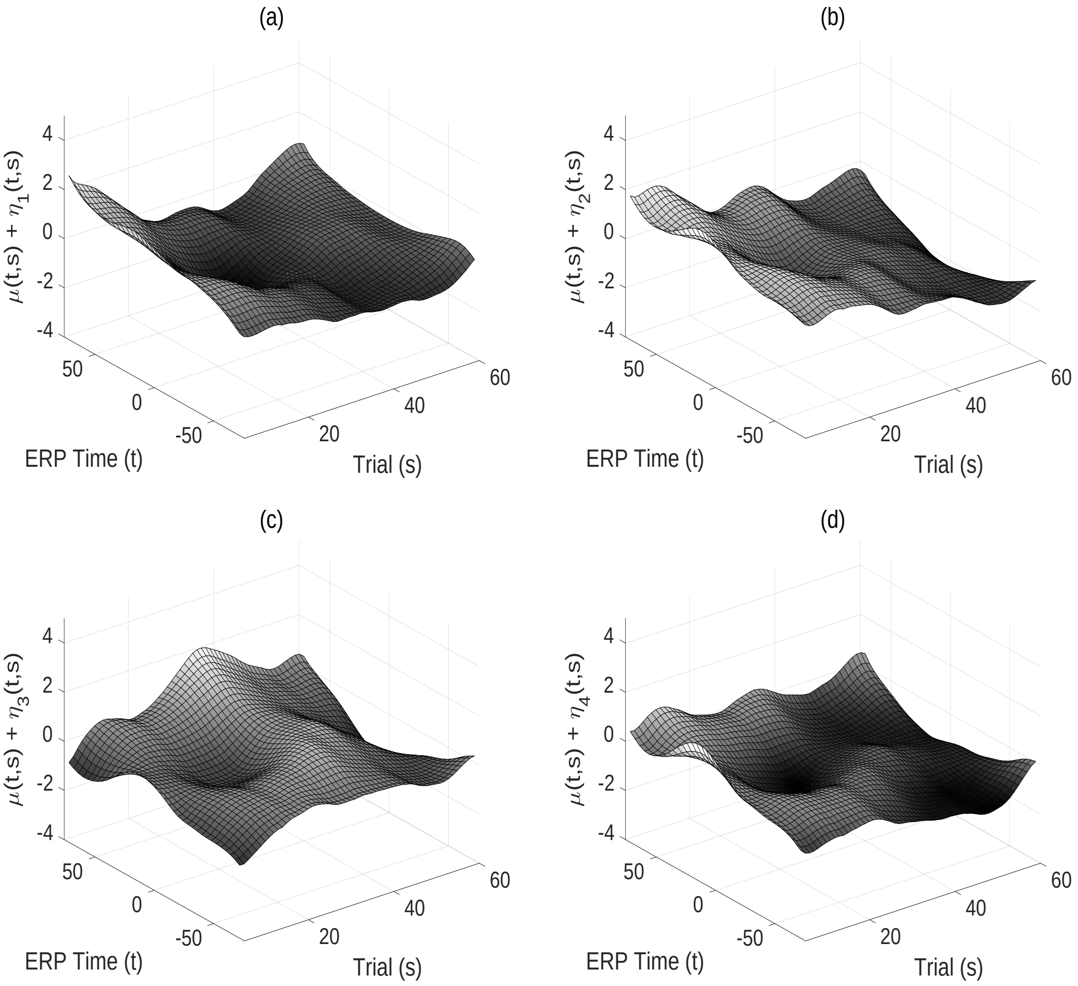
	Dense		Sparse	
	N=30	N=100	N=30	N=100
High SNR				
$\mu(t,s)$	.002 (.001, .004)	.001 (.000, .001)	.002 (.001, .005)	.001 (.000, .001)
$X_{ij}(t\|s)$	.006 (.005, .008)	.003 (.003, .006)	.011 (.009, .015)	.006 (.005, .009)
ρ(s)	.013 (.005, .033)	.003 (.001, .009)	.022 (.014, .040)	.006 (.004, .011)
ρ	.001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



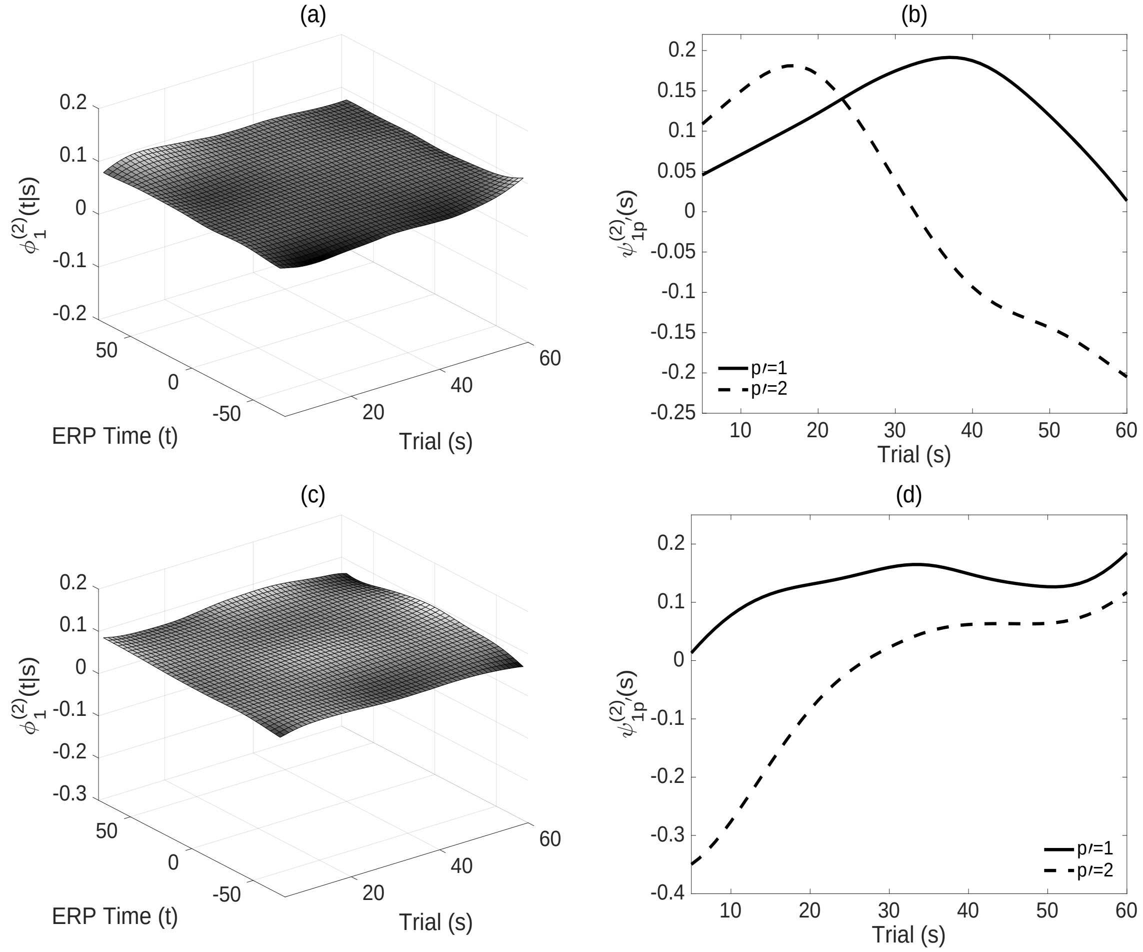
Web Figure 1: (a) ERP waveform from a single subject, condition, electrode and trial in the right frontal region of the scalp. (b) The average of the first 30 consecutive ERP waveforms for the same subject, electrode and condition after preprocessing.



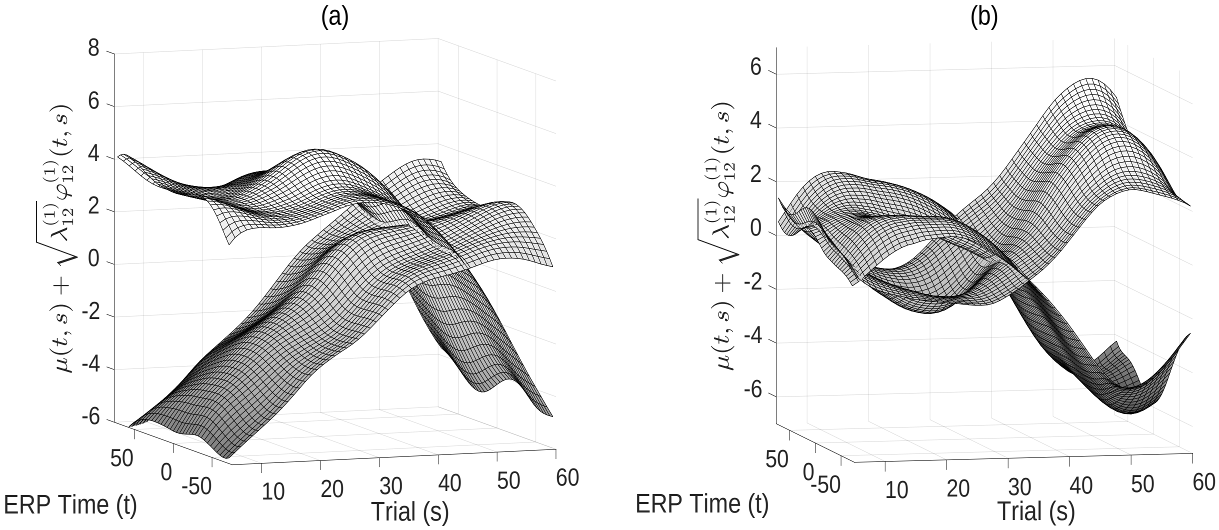
Web Figure 2: Estimated electrode-specific mean surfaces for the ASD group.



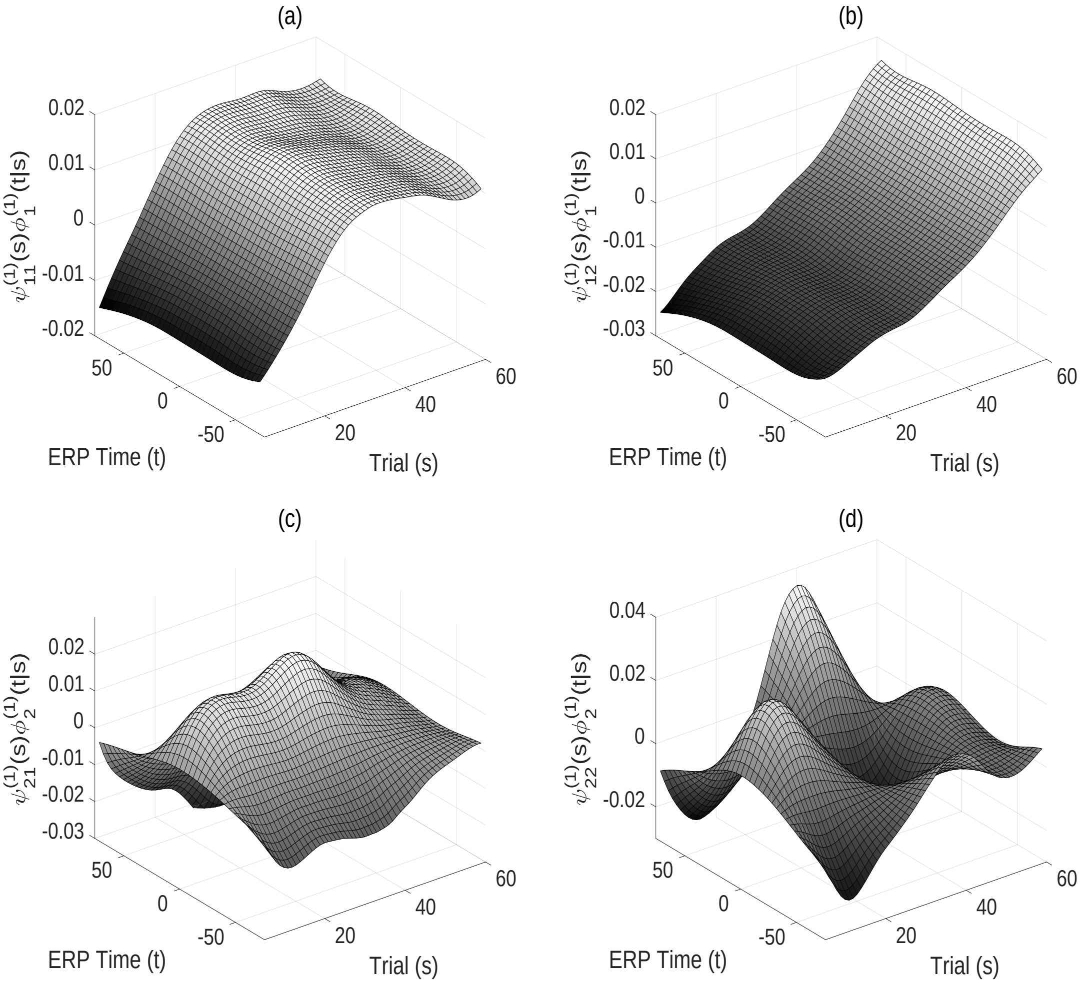
Web Figure 3: Estimated electrode-specific mean surfaces for the TD group.



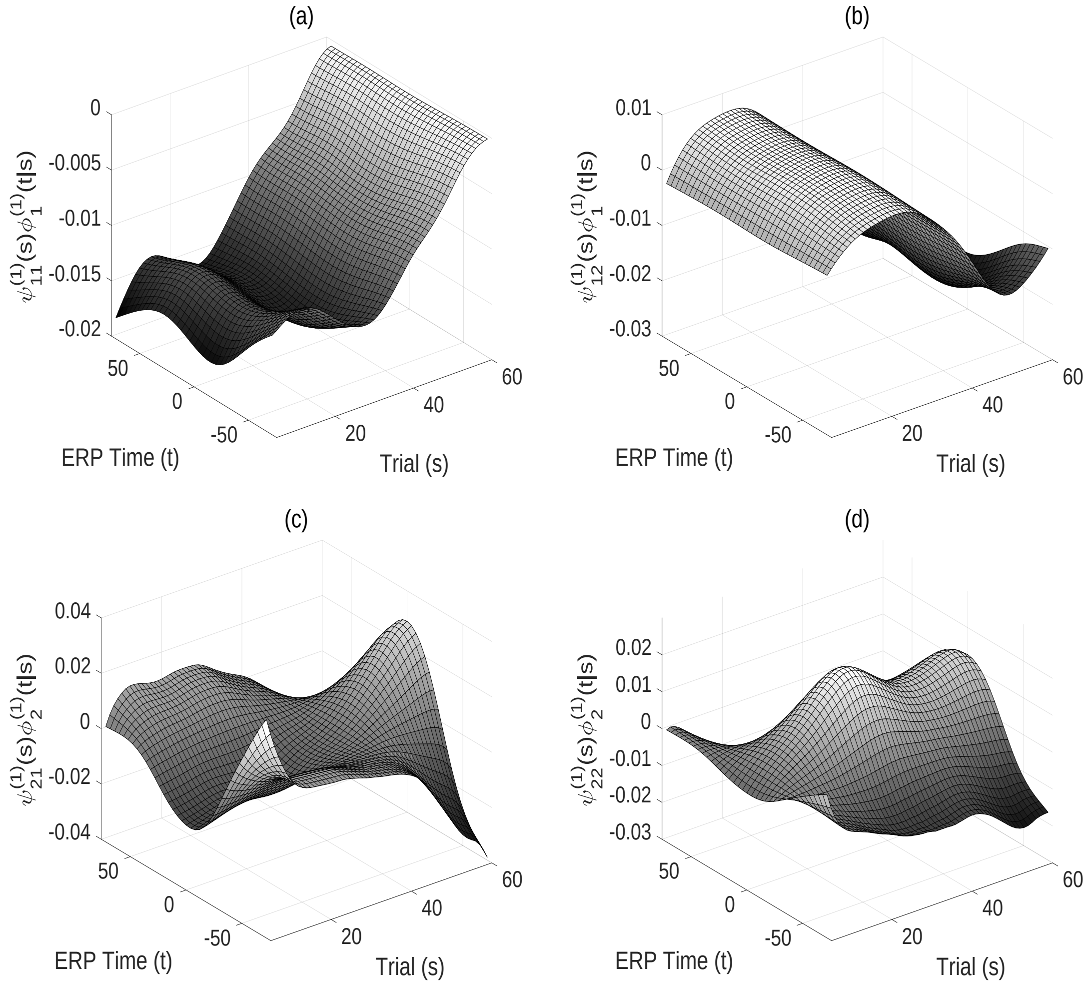
Web Figure 4: (a), (c) Estimated leading electrode level first-stage eigenfunctions, $\{\phi_1^{(2)}(t|s)\}$ and (b), (d) estimated leading electrode level second-stage eigenfunctions, $\{\psi_{1p'}^{(2)}(s)\}$, $p' = 1, 2$, for the ASD and TD groups, respectively.



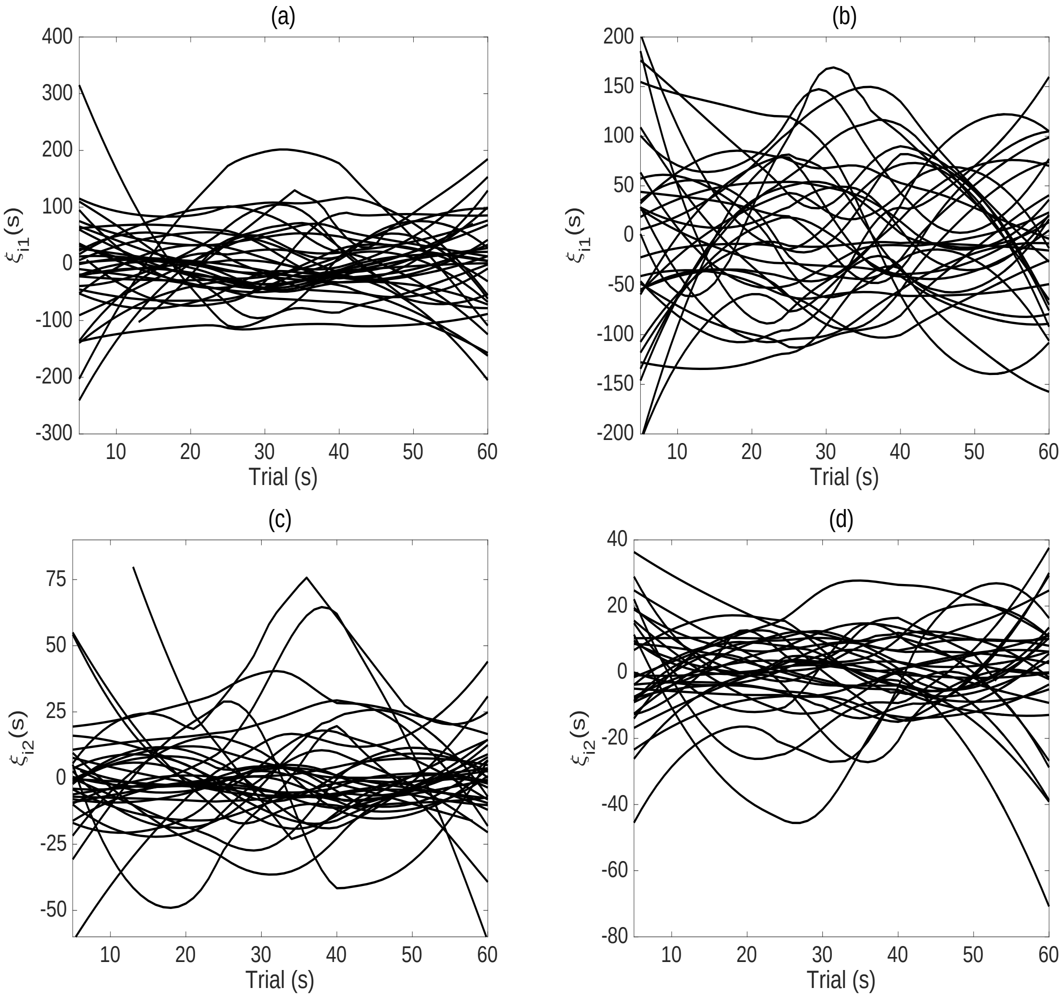
Web Figure 5: (a)-(b) Estimated surface intervals, $\mu(t,s) \pm \sqrt{\lambda_{12}^{(1)}}\varphi_{12}^{(1)}(t,s)$, for the ASD and TD groups, respectively.



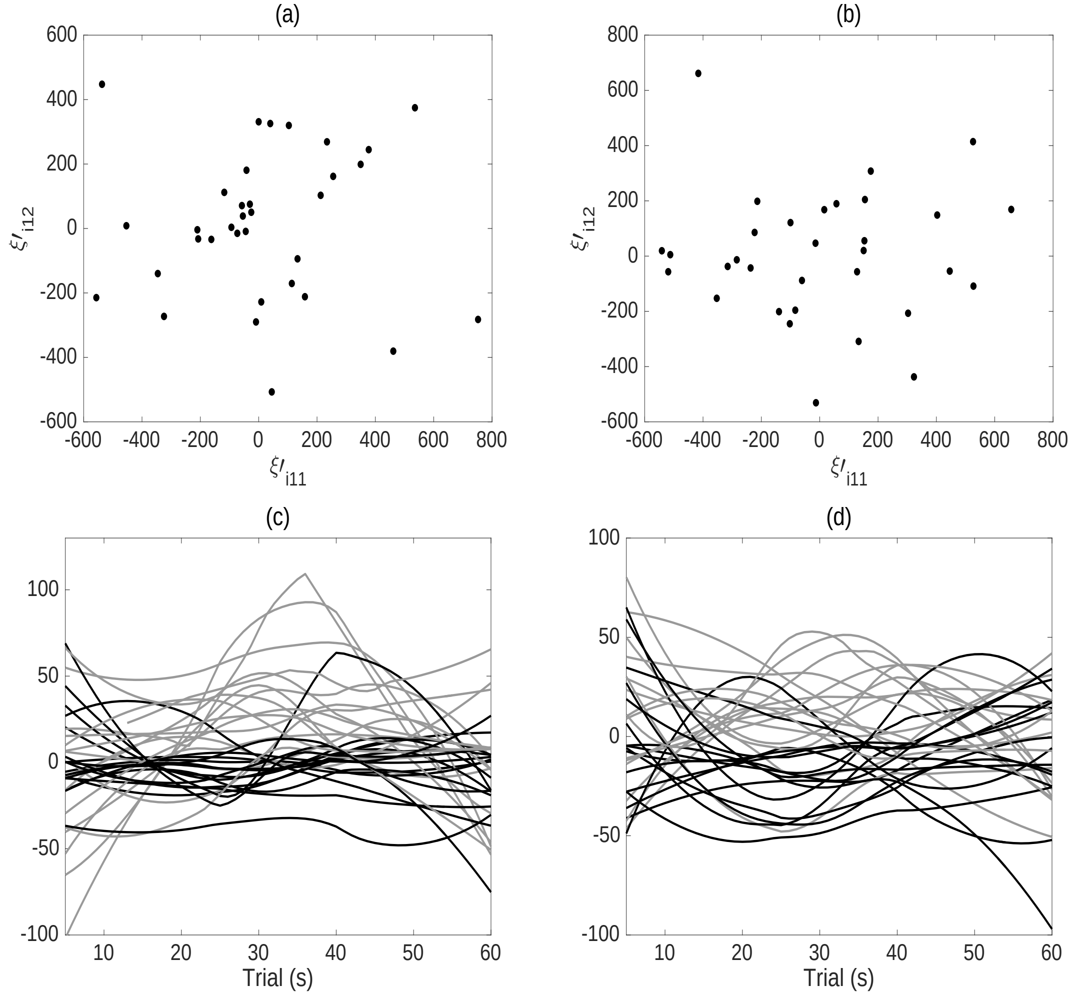
Web Figure 6: Estimated subject level principal surfaces $\varphi^{(1)}_{kk'}(t,s) = \psi^{(1)}_{kk'}(s)\phi^{(1)}_{k}(t|s)$ for (a) $k = 1, k' = 1$, (b) $k = 1, k' = 2$, (c) $k = 2, k' = 1$ and (d) $k = 2, k' = 2$ for the ASD group.



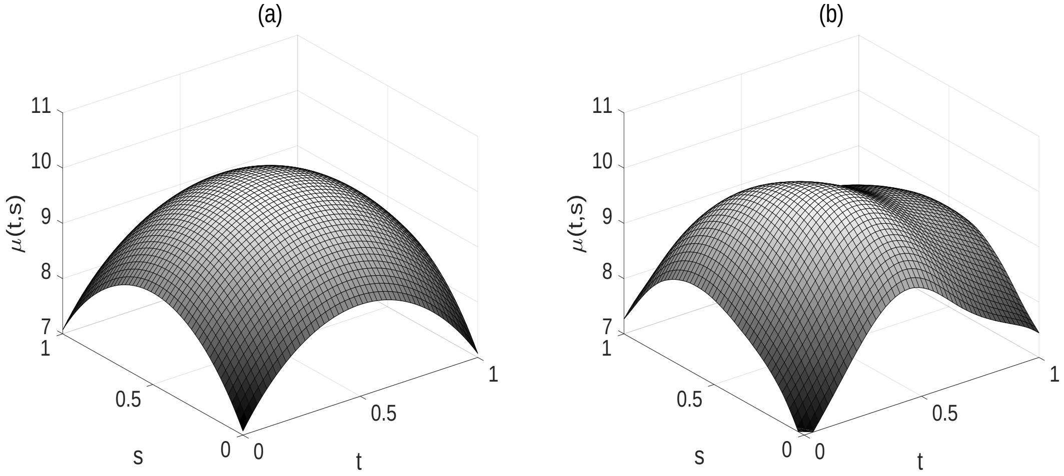
Web Figure 7: Estimated subject level principal surfaces $\varphi^{(1)}_{kk'}(t,s) = \psi^{(1)}_{kk'}(s)\phi^{(1)}_{k}(t|s)$ for (a) $k=1, k'=1$, (b) $k=1, k'=2$, (c) $k=2, k'=1$ and (d) $k=2, k'=2$ for the TD group.



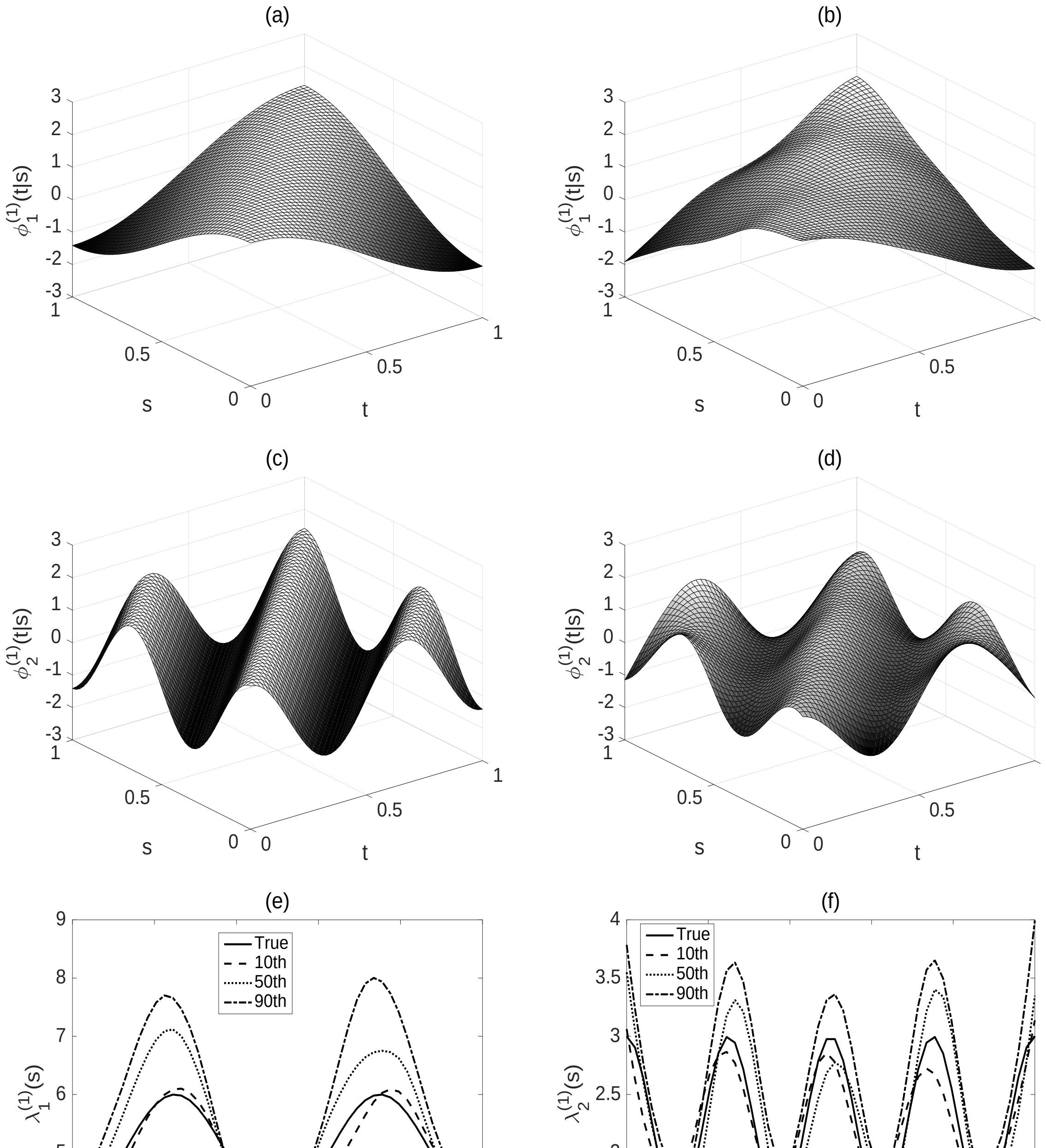
Web Figure 8: Estimated subject-specific eigenscores from the first-stage decompositions for the ASD ((a), (c)) and TD ((b), (d)) groups.



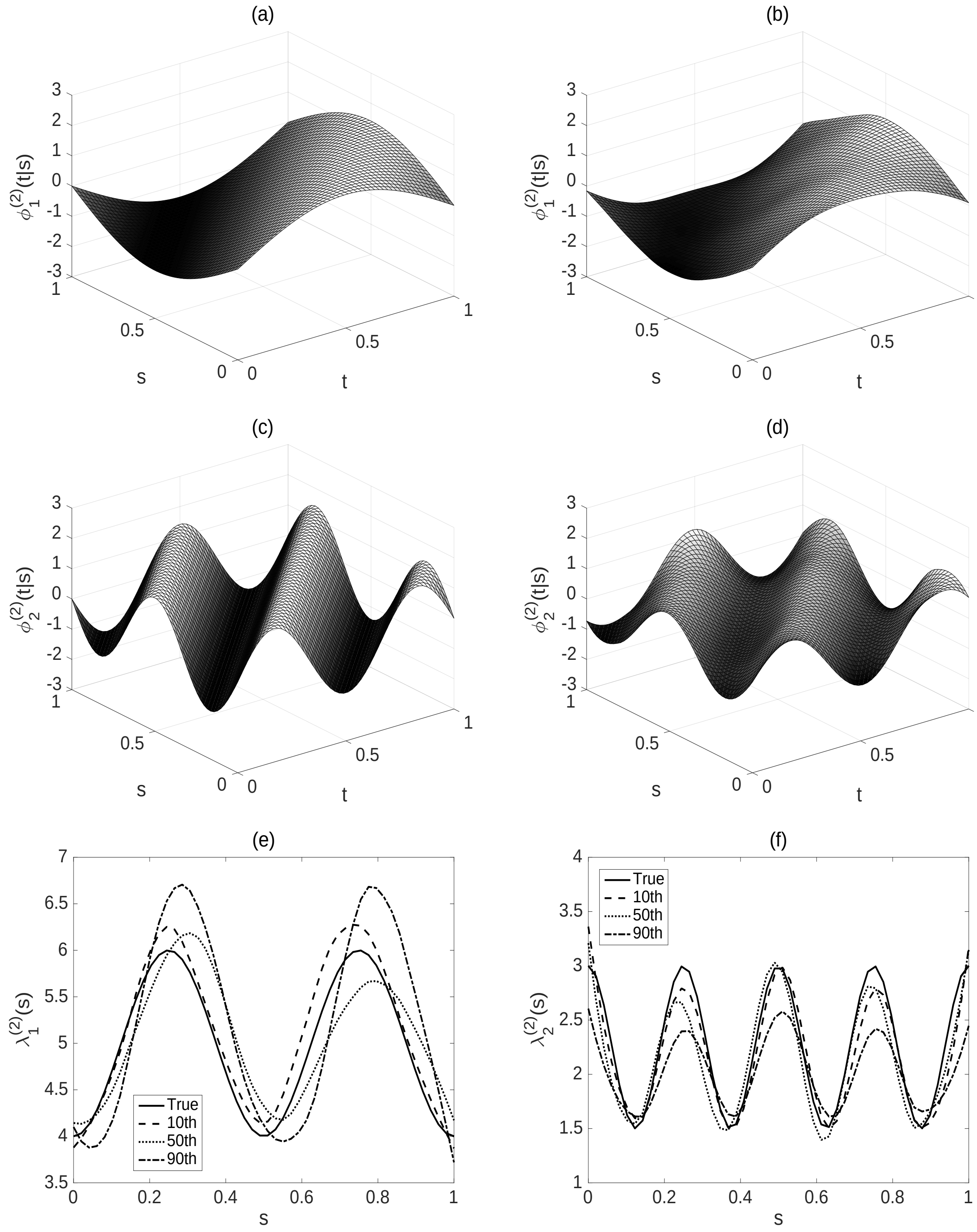
Web Figure 9: (a)-(b) The two leading subject level eigenscores from the second-stage decompositions for the ASD and TD groups, respectively. (c)-(d) The smoothed subject-specific amplitude difference trajectories at P3 peak location $t = 0$, partitioned by the median of the leading scores ξ'_{i11} for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



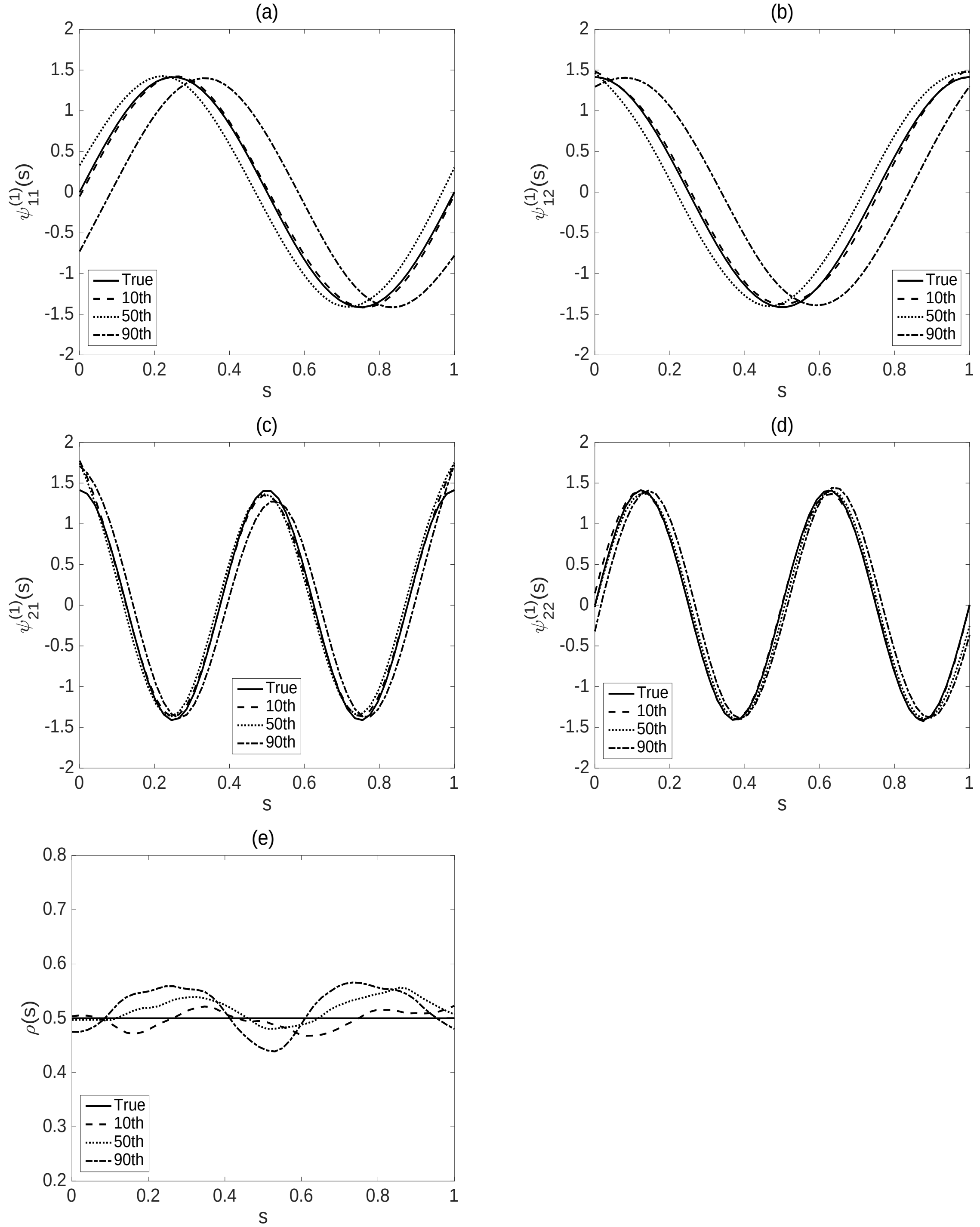
Web Figure 10: The true (a) and estimated (b) mean surfaces, $\mu(t, s)$, based on 200 Monte Carlo runs from the dense design scenario at $N = 100$ and low SNR. The estimated function corresponds to the Monte Carlo run with RSE value at the 50th percentile.



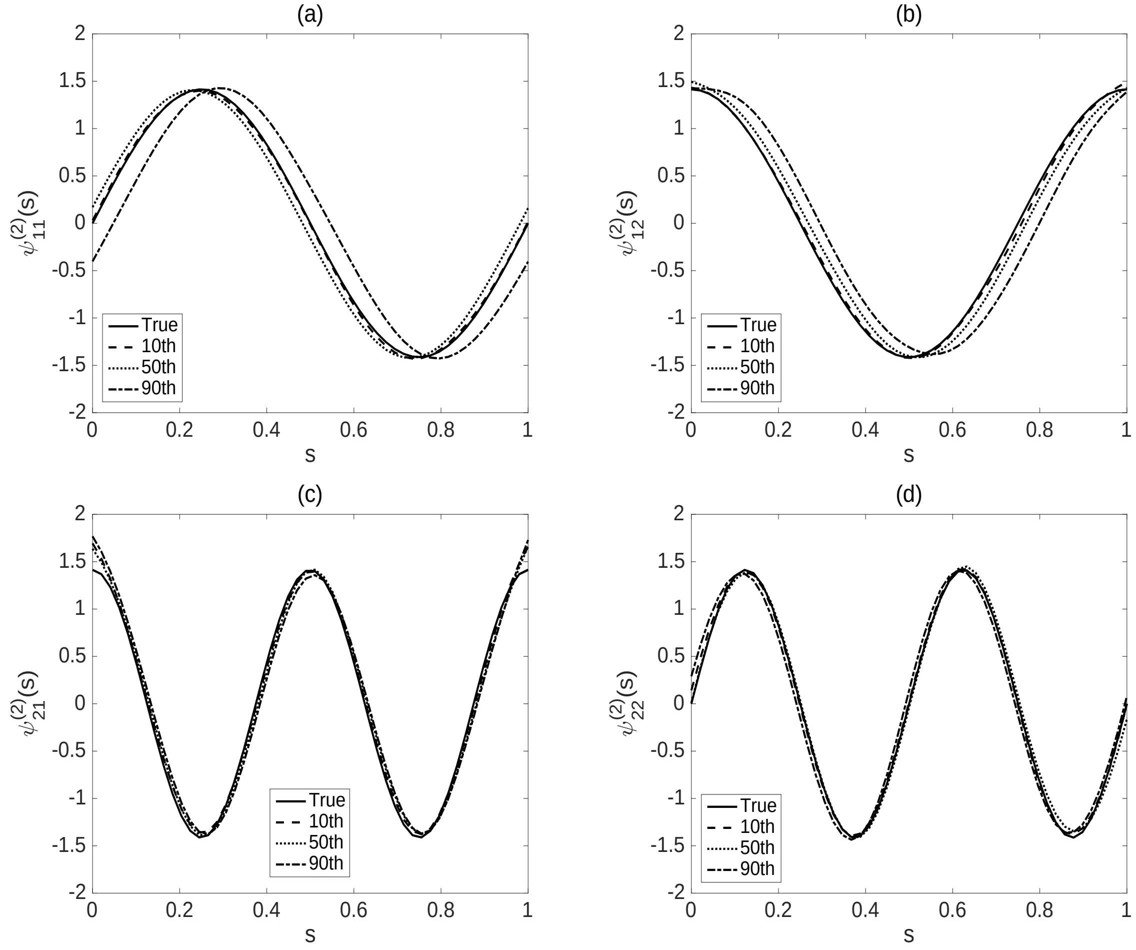
Web Figure 11: The true and estimated model components based on 200 Monte Carlo runs from the dense design scenario at $N = 100$ and low SNR. The true ((a), (c)) and estimated ((b), (d)) first-stage subject level eigenfunctions are displayed in the first two rows. The estimated functions correspond to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



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



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



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