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

Emilie Campos

Department of Biostatistics
University of California, Los Angeles

Joint work with:

Aaron Scheffler, Donatello Telesca, Catherine Sugar, Charlotte DiStefano, Shafali Jeste, April Levin, Adam Naples, Sara J. Webb, James C. McPartland, Damla Şentürk

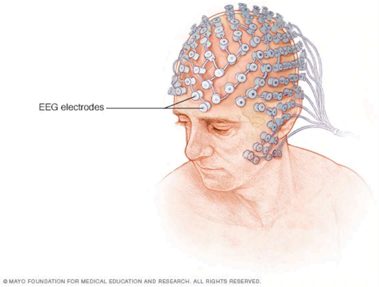
# Outline

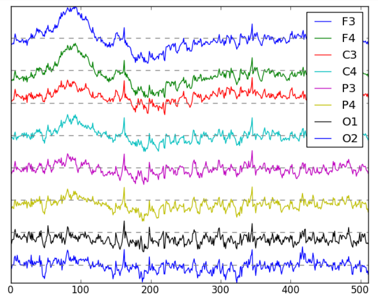
- Motivation
- M-HPCA Algorithm
- Applications
  - Audio odd-ball paradigm
  - Day-to-day test-retest reliability of PSD
- Simulation
- Concluding Remarks

# Motivation

# Data collection

- Electroencephalography (EEG): non-invasive and low-cost modality to analyze the brain and its behavior
- Categorized into two groups:
  - Resting state: analyzed on the frequency domain via the power spectral density (PSD)
  - Event-related: analyzed on the time domain via the event-related potential (ERP) created by time-locking the brain response to the onset of the stimulus



# Data structure

- Generated data viewed as functional objects collected across the scalp across varying experimental conditions within a single longitudinal visit or across multiple visits
  - Termed region-referenced multilevel functional data

- Common analysis of EEG reduces the data complexity by collapsing one of the dimensions
  - Functional: analyze peak or mean amplitude of the ERP or power within a specific frequency band
  - Regional: averaging functions across electrodes to create a scalp-wide average

# Current methods

- Hybrid PCA: uses the concept of **weak separability** to decompose variation into regional and functional dimensions (Scheffler, 2020)
    - Weak separability: assumes the direction of variation along one of the dimensions stays constant across slices of the other dimension so estimates directions of variation utilizing marginal lower-dimensional covariances
    - Uses both vector and functional PCA
    - Assumes regions are non-exchangeable
- Multilevel FPCA: **functional ANOVA model** that decomposes total variation into between- and within-subjects variation (Di, 2009)
    - Uses FPCA
    - Assumes levels are exchangeable
- These approaches can be extended for exchangeable levels and non-exchangeable regions to decompose multilevel region-referenced functional data

# M-HPCA Algorithm

# M-HPCA algorithm

- Proposed M-HPCA borrows ideas from HPCA and M-FPCA: decomposes the total variation into between- and within-subjects variation then assumes weak separability on both of these covariance processes

- Results in participant-specific and repetition-specific eigenvectors and eigenfunctions that are highly interpretable as participant-specific and repetition-specific regional and functional directions of variation, respectively

- Enables computationally efficient estimation and inference by representing covariance matrices as a weighted sum of lower dimensional building blocks, and through the use of a minorization-maximization (MM) algorithm

# Model

- $Y_{dij}(r,t)$ denotes the functional observation for subject $i$, $i = 1, \ldots, n_d$, from group $d$, $d = 1, \ldots, D$, in region $r$, $r = 1, \ldots, R$, at time $t$, $t \in \mathcal{T}$ and is modeled as

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

  - $\mu(t)$: overall mean function
  - $\eta_{dj}(r,t)$: group-region-repetition-specific shift from the overall mean
  - $Z_{di}(r,t)$: subject-region-specific deviation
  - $W_{dij}(r,t)$: subject-region-repetition deviation
  - $\epsilon_{dij}(r,t)$: independent measurement error

# Covariances

- Total covariance:

$$K_{d,\text{Total}}\{(r,t),(r',t')\} = \text{cov}\{Y_{dij}(r,t), Y_{dij}(r',t')\}$$

- Between-subject covariance:

$$K_{d,B}\{(r,t),(r',t')\} = \text{cov}\{Y_{dij_1}(r,t), Y_{dij_2}(r',t')\} = \text{cov}\{Z_{di}(r,t), Z_{di}(r',t')\}$$

- Within-subject covariance

$$\begin{aligned} K_{d,W}\{(r,t),(r',t')\} &:= K_{d,\text{Total}}\{(r,t),(r',t')\} - K_{d,B}\{(r,t),(r',t')\} \\ &= \text{cov}\{W_{dij}(r,t), W_{dij}(r',t')\} \end{aligned}$$

# Marginal covariances

- Functional marginal between and within covariance surfaces

$$\Sigma_{d,\mathcal{T},B}(t,t') = \sum_{r=1}^{R} K_{d,B}\{(r,t),(r,t')\} = \sum_{\ell=1}^{\infty} \tau_{d\ell,\mathcal{T}}^{(1)} \phi_{d\ell}^{(1)}(t) \phi_{d\ell}^{(1)}(t')$$

$$\Sigma_{d,\mathcal{T},W}(t,t') = \sum_{r=1}^{R} K_{d,W}\{(r,t),(r,t')\} = \sum_{m=1}^{\infty} \tau_{dm,\mathcal{T}}^{(2)} \phi_{dm}^{(2)}(t) \phi_{dm}^{(2)}(t'),$$

- Regional marginal between and within covariance matrices

$$(\Sigma_{d,\mathcal{R},B})_{r,r'} = \int_{\mathcal{T}} K_{d,B}\{(r,t),(r',t)\}dt = \sum_{k=1}^{R} \tau_{dk,\mathcal{R}}^{(1)} \mathrm{v}_{dk}^{(1)}(r) \mathrm{v}_{dk}^{(1)}(r')$$

$$(\Sigma_{d,\mathcal{R},W})_{r,r'} = \int_{\mathcal{T}} K_{d,W}\{(r,t),(r',t)\}dt = \sum_{p=1}^{R} \tau_{dp,\mathcal{R}}^{(2)} \mathrm{v}_{dp}^{(2)}(r) \mathrm{v}_{dp}^{(2)}(r'),$$

- $\phi_{d\ell}^{(1)}(t)$ and $\phi_{dm}^{(2)}(t)$ are the level 1 and level 2 eigenfunctions, $\mathrm{v}_{dk}^{(1)}(r)$ and $\mathrm{v}_{dp}^{(2)}(r)$ are the level 1 and level 2 eigenvectors, and $\tau_{d\ell,\mathcal{T}}^{(1)}$, $\tau_{dm,\mathcal{T}}^{(2)}$, $\tau_{dk,\mathcal{R}}^{(1)}$, and $\tau_{dp,\mathcal{R}}^{(2)}$ are the respective eigenvalues

# Decomposed model

- Utilizing the marginal eigenfunctions and eigenvectors

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

# Decomposed variance

- Total covariance is decomposed as

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

where $\lambda_{d,k\ell}^{(1)} = \mathrm{var}(\zeta_{di,k\ell})$ and $\lambda_{d,pm}^{(2)} = \mathrm{var}(\xi_{dij,pm})$

# Converting to single indices

- Each of the indices are truncated to $K$, $L$, $P$, and $M$ and we use single indices $g = (k-1) + K(\ell - 1) + 1$ and $h = (p-1) + P(m-1) + 1$

$$Y_{dij}(r,t) = \mu(t) + \eta_{dj}(r,t) + \sum_{g=1}^{G} \zeta_{dig}\varphi^{(1)}_{dg}(r,t) + \sum_{h=1}^{H} \xi_{dijh}\varphi^{(2)}_{dh}(r,t) + \epsilon_{dij}(r,t)$$

$$\zeta_{dig} \sim \mathrm{N}\left(0, \lambda^{(1)}_{dg}\right),\ \xi_{dijh} \sim \mathrm{N}\left(0, \lambda^{(2)}_{dh}\right),\ \epsilon_{dij}(r,t) \sim \mathrm{N}(0, \sigma^2_d)$$

  where

  - $G = KL$
  - $\varphi^{(1)}_{dg}(r,t) = \mathrm{v}^{(1)}_{dk}(r)\phi^{(1)}_{d\ell}(t)$
  - $\zeta_{dig} = \langle Z_{di}(r,t), \varphi^{(1)}_{dg}(r,t)\rangle$
  - $\mathrm{var}(\zeta_{dig}) = \lambda^{(1)}_{dg}$
  - $H = PM$
  - $\varphi^{(2)}_{dh}(r,t) = \mathrm{v}^{(2)}_{dp}(r)\phi^{(2)}_{dm}(t)$
  - $\xi_{dijh} = \langle W_{dij}(r,t), \varphi^{(2)}_{dh}(r,t)\rangle$
  - $\mathrm{var}(\xi_{dijh}) = \lambda^{(2)}_{dh}$

- Use an MM algorithm to estimate the variance components ($\lambda^{(1)}_{dg}$, $\lambda^{(2)}_{dh}$, and $\sigma^2_d$) and scores ($\zeta_{dig}$ and $\xi_{dijh}$)

# Vectorize

$$Y_{di} = Z_{di}\boldsymbol{\zeta}_{di} + W_{di}\boldsymbol{\xi}_{di} + \boldsymbol{\epsilon}_{di} \text{ for } i = 1, \dots, n_d$$

$$\boldsymbol{\zeta}_{di} \sim \text{MVN}\left(0, \boldsymbol{\Lambda}_d^{(1)}\right), \; \boldsymbol{\xi}_{di} \sim \text{MVN}\left(0, I_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)}\right), \; \boldsymbol{\epsilon}_{di} \sim \text{MVN}\left(0, \sigma_d^2 I_{TRJ_i}\right)$$

- $Y_{di} \sim \text{MVN}\left(\mathbf{0}, \boldsymbol{\Sigma}_{di}\right)$, where $\boldsymbol{\Sigma}_{di} = Z_{di}\boldsymbol{\Lambda}_d^{(1)} Z_{di}^{\mathrm{T}} + W_{di}\left(I_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)}\right) W_{di}^{\mathrm{T}} + \sigma_d^2 I_{TRJ_i}$
- Log-likelihood: $\ell_d\left(\boldsymbol{\Lambda}_d^{(1)}, \boldsymbol{\Lambda}_d^{(2)}, \sigma_d^2\right) = -\frac{1}{2}\sum_{i=1}^{n_d} \log\det \boldsymbol{\Sigma}_{di} + Y_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1} Y_{di}$

# MM Algorithm

- Minorizing function of log-likelihood:

$$\sum_{i=1}^{n_d} f_{di}\left(\boldsymbol{\Lambda}_d^{(1)}, \boldsymbol{\Lambda}_d^{(2)}, \sigma_d^2 \middle| \boldsymbol{\Lambda}_d^{(1)(c)}, \boldsymbol{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)}\right)$$

$$= \sum_{i=1}^{n_d} -\frac{1}{2}\Bigg[\mathrm{tr}\left(Z_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} Z_{di} \boldsymbol{\Lambda}_d^{(1)(c)}\right) + \boldsymbol{\zeta}_{di}^{(c)\mathrm{T}} \boldsymbol{\Lambda}_d^{-(1)(c)} \boldsymbol{\zeta}_{di}^{(c)}$$

$$+ \mathrm{tr}\left\{W_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} W_{di}\left(I_{J_i} \otimes \Lambda_d^{(2)(c)}\right)\right\} + \boldsymbol{\xi}_{di}^{(c)\mathrm{T}}\left(I_{J_i} \otimes \boldsymbol{\Lambda}_d^{-1(2)(c)}\right) \xi_{di}^{(c)}$$

$$+ \mathrm{tr}\left(\sigma_d^2 \boldsymbol{\Sigma}_{di}^{-1(c)}\right) + \frac{\sigma_d^{4(c)}}{\sigma_d^2} Y_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-2(c)} Y_{di}\Bigg] + q^{(c)}$$

# MM Algorithm

- Minorizing function of log-likelihood:

$$
\begin{aligned}
&\sum_{i=1}^{n_d} f_{di}\left(\boldsymbol{\Lambda}_d^{(1)}, \boldsymbol{\Lambda}_d^{(2)}, \sigma_d^2 \middle| \boldsymbol{\Lambda}_d^{(1)(c)}, \boldsymbol{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)}\right) \\
&\quad = \sum_{i=1}^{n_d} -\frac{1}{2}\Bigg[\operatorname{tr}\left(Z_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} Z_{di} \boldsymbol{\Lambda}_d^{(1)(c)}\right) + \zeta_{di}^{(c)\mathrm{T}} \boldsymbol{\Lambda}_d^{-(1)(c)} \zeta_{di}^{(c)} \\
&\qquad + \operatorname{tr}\left\{W_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} W_{di}\left(I_{J_i} \otimes \Lambda_d^{(2)(c)}\right)\right\} + \xi_{di}^{(c)\mathrm{T}}\left(I_{J_i} \otimes \boldsymbol{\Lambda}_d^{-1(2)(c)}\right) \xi_{di}^{(c)} \\
&\qquad + \operatorname{tr}\left(\sigma_d^2 \boldsymbol{\Sigma}_{di}^{-1(c)}\right) + \frac{\sigma_d^{4(c)}}{\sigma_d^2} Y_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-2(c)} Y_{di}\Bigg] + q^{(c)}
\end{aligned}
$$

Level 1 variance components

# MM Algorithm

- Minorizing function of log-likelihood:

$$
\begin{aligned}
&\sum_{i=1}^{n_d} f_{di}\left(\boldsymbol{\Lambda}_d^{(1)}, \boldsymbol{\Lambda}_d^{(2)}, \sigma_d^2 \middle| \boldsymbol{\Lambda}_d^{(1)(c)}, \boldsymbol{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)}\right) \\
&\quad = \sum_{i=1}^{n_d} -\frac{1}{2}\Bigg[\operatorname{tr}\left(Z_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} Z_{di} \boldsymbol{\Lambda}_d^{(1)(c)}\right) + \zeta_{di}^{(c)\mathrm{T}} \boldsymbol{\Lambda}_d^{-(1)(c)} \zeta_{di}^{(c)} \\
&\qquad + \operatorname{tr}\left\{W_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} W_{di}\left(I_{J_i} \otimes \Lambda_d^{(2)(c)}\right)\right\} + \xi_{di}^{(c)\mathrm{T}}\left(I_{J_i} \otimes \boldsymbol{\Lambda}_d^{-1(2)(c)}\right) \xi_{di}^{(c)} \\
&\qquad + \operatorname{tr}\left(\sigma_d^2 \boldsymbol{\Sigma}_{di}^{-1(c)}\right) + \frac{\sigma_d^{4(c)}}{\sigma_d^2} Y_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-2(c)} Y_{di}\Bigg] + q^{(c)}
\end{aligned}
$$

Level 2 variance components

# MM Algorithm

- Minorizing function of log-likelihood:

$$
\begin{aligned}
&\sum_{i=1}^{n_d} f_{di}\left(\boldsymbol{\Lambda}_d^{(1)}, \boldsymbol{\Lambda}_d^{(2)}, \sigma_d^2 \middle| \boldsymbol{\Lambda}_d^{(1)(c)}, \boldsymbol{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)}\right) \\
&= \sum_{i=1}^{n_d} -\frac{1}{2}\Bigg[\operatorname{tr}\left(Z_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} Z_{di} \boldsymbol{\Lambda}_d^{(1)(c)}\right) + \zeta_{di}^{(c)\mathrm{T}} \boldsymbol{\Lambda}_d^{-(1)(c)} \zeta_{di}^{(c)} \\
&\quad + \operatorname{tr}\left\{W_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} W_{di}\left(I_{J_i} \otimes \Lambda_d^{(2)(c)}\right)\right\} + \xi_{di}^{(c)\mathrm{T}}\left(I_{J_i} \otimes \boldsymbol{\Lambda}_d^{-1(2)(c)}\right) \xi_{di}^{(c)} \\
&\quad + \operatorname{tr}\left(\sigma_d^2 \boldsymbol{\Sigma}_{di}^{-1(c)}\right) + \frac{\sigma_d^{4(c)}}{\sigma_d^2} Y_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-2(c)} Y_{di}\Bigg] + q^{(c)}
\end{aligned}
$$

Measurement error variance component

# MM Algorithm

- Minorizing function of log-likelihood:

$$\sum_{i=1}^{n_d} f_{di}\left(\boldsymbol{\Lambda}_d^{(1)}, \boldsymbol{\Lambda}_d^{(2)}, \sigma_d^2 \middle| \boldsymbol{\Lambda}_d^{(1)(c)}, \boldsymbol{\Lambda}_d^{(2)(c)}, \sigma_d^{2(c)}\right)$$

$$= \sum_{i=1}^{n_d} -\frac{1}{2}\Bigg[\operatorname{tr}\left(Z_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} Z_{di} \boldsymbol{\Lambda}_d^{(1)(c)}\right) + \boldsymbol{\zeta}_{di}^{(c)\mathrm{T}} \boldsymbol{\Lambda}_d^{-(1)(c)} \boldsymbol{\zeta}_{di}^{(c)}$$

$$+ \operatorname{tr}\left\{W_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-1(c)} W_{di}\left(I_{J_i} \otimes \Lambda_d^{(2)(c)}\right)\right\} + \boldsymbol{\xi}_{di}^{(c)\mathrm{T}}\left(I_{J_i} \otimes \boldsymbol{\Lambda}_d^{-1(2)(c)}\right) \xi_{di}^{(c)}$$

$$+ \operatorname{tr}\left(\sigma_d^2 \boldsymbol{\Sigma}_{di}^{-1(c)}\right) + \frac{\sigma_d^{4(c)}}{\sigma_d^2} Y_{di}^{\mathrm{T}} \boldsymbol{\Sigma}_{di}^{-2(c)} Y_{di}\Bigg] + q^{(c)}$$

- Variance components separated $\Rightarrow$ derivatives are easy!

## Simplifying covariance

- Denote by $K_{d,B} = \mathbf{\Phi}_d^{(1)}\mathbf{\Lambda}_d^{(1)}\mathbf{\Phi}_d^{(1)\mathrm{T}}$ and denote by $K_{d,W} = \mathbf{\Phi}_d^{(2)}\mathbf{\Lambda}_d^{(2)}\mathbf{\Phi}_d^{(2)\mathrm{T}}$, both of dimension $TR \times TR$

$$
\begin{aligned}
\mathbf{\Sigma}_{di} &= Z_{di}\mathbf{\Lambda}_d^{(1)}Z_{di}^{\mathrm{T}} + W_{di}\left(I_{J_i} \otimes \mathbf{\Lambda}_d^{(2)}\right)W_{di}^{\mathrm{T}} + \sigma_d^2 I_{TRJ_i} \\
&= \mathbf{1}_{J_i \times J_i} \otimes \mathbf{\Phi}_d^{(1)}\mathbf{\Lambda}_d^{(1)}\mathbf{\Phi}_d^{(1)\mathrm{T}} + I_{J_i \times J_i} \otimes \left(\sigma_d^2 I_{TR} + \mathbf{\Phi}_d^{(2)}\mathbf{\Lambda}_d^{(2)}\mathbf{\Phi}_d^{(2)\mathrm{T}}\right) \\
&= \mathbf{1}_{J_i \times J_i} \otimes K_{d,B} + I_{J_i \times J_i} \otimes \left(\sigma_d^2 I_{TR} + K_{d,W}\right) \\
\mathbf{\Sigma}_{di}^{-1} &= I_{J_i} \otimes \left(\sigma_d^2 I_{TR} + K_{d,W}\right)^{-1} \\
&\qquad - \mathbf{1}_{J_i \times J_i} \otimes \left\{\left(\sigma_d^2 I_{TR} + K_{d,W}\right)^{-1} K_{d,B}\left(\sigma_d^2 I_{TR} + K_{d,W} + J_i K_{d,B}\right)^{-1}\right\}
\end{aligned}
$$

# Simplifying covariance

- Denote by $K_{d,B} = \boldsymbol{\Phi}_d^{(1)} \boldsymbol{\Lambda}_d^{(1)} \boldsymbol{\Phi}_d^{(1)\mathrm{T}}$ and denote by $K_{d,W} = \boldsymbol{\Phi}_d^{(2)} \boldsymbol{\Lambda}_d^{(2)} \boldsymbol{\Phi}_d^{(2)\mathrm{T}}$, both of dimension $TR \times TR$

$$\begin{aligned}
\boldsymbol{\Sigma}_{di} &= Z_{di} \boldsymbol{\Lambda}_d^{(1)} Z_{di}^{\mathrm{T}} + W_{di} \left( I_{J_i} \otimes \boldsymbol{\Lambda}_d^{(2)} \right) W_{di}^{\mathrm{T}} + \sigma_d^2 I_{TRJ_i} \\
&= \mathbf{1}_{J_i \times J_i} \otimes \boldsymbol{\Phi}_d^{(1)} \boldsymbol{\Lambda}_d^{(1)} \boldsymbol{\Phi}_d^{(1)\mathrm{T}} + I_{J_i \times J_i} \otimes \left( \sigma_d^2 I_{TR} + \boldsymbol{\Phi}_d^{(2)} \boldsymbol{\Lambda}_d^{(2)} \boldsymbol{\Phi}_d^{(2)\mathrm{T}} \right) \\
&= \mathbf{1}_{J_i \times J_i} \otimes K_{d,B} + I_{J_i \times J_i} \otimes \left( \sigma_d^2 I_{TR} + K_{d,W} \right) \\
\boldsymbol{\Sigma}_{di}^{-1} &= I_{J_i} \otimes \left( \sigma_d^2 I_{TR} + K_{d,W} \right)^{-1} \\
&\qquad - \mathbf{1}_{J_i \times J_i} \otimes \left\{ \left( \sigma_d^2 I_{TR} + K_{d,W} \right)^{-1} K_{d,B} \left( \sigma_d^2 I_{TR} + K_{d,W} + J_i K_{d,B} \right)^{-1} \right\}
\end{aligned}$$

- Only two $TR \times TR$ matrices need to be inverted, but we can simplify even further

## Simplifying covariance further

- Woodbury matrix identity:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

- Apply to our two matrices

$$\begin{aligned}\left(\sigma_d^2 I_{TR} + K_{d,W}\right)^{-1} &= \left(\sigma_d^2 I_{TR} + \mathbf{\Phi}_d^{(2)}\mathbf{\Lambda}_d^{(2)}\mathbf{\Phi}_d^{(2)\mathrm{T}}\right)^{-1} \\ &= \frac{1}{\sigma_d^2} I_{TR} - \frac{1}{\sigma_d^4}\mathbf{\Phi}_d^{(2)}\left\{\left(\mathbf{\Lambda}_d^{(2)}\right)^{-1} + \tfrac{1}{\sigma_d^2}\mathbf{\Phi}_d^{(2)\mathrm{T}}\mathbf{\Phi}_d^{(2)}\right\}^{-1}\mathbf{\Phi}_d^{(2)\mathrm{T}}\end{aligned}$$

and denoting $(\sigma_d^2 I_{TR} + K_{d,W})^{-1}$ by $Q$ for simplicity, we expand the second inverse as

$$\begin{aligned}\left(\sigma_d^2 I_{TR} + K_{d,W} + J_i K_{d,B}\right)^{-1} &= \left(Q^{-1} + J_i\mathbf{\Phi}_d^{(1)}\mathbf{\Lambda}_d^{(1)}\mathbf{\Phi}_d^{(1)\mathrm{T}}\right)^{-1} \\ &= Q - J_i Q\mathbf{\Phi}_d^{(1)}\left\{\left(\mathbf{\Lambda}_d^{(1)}\right)^{-1} + J_i\mathbf{\Phi}_d^{(1)\mathrm{T}}Q\mathbf{\Phi}_d^{(1)}\right\}^{-1}\mathbf{\Phi}_d^{(1)\mathrm{T}}Q.\end{aligned}$$

# Simplifying covariance further

- Woodbury matrix identity:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

- Apply to our two matrices

$G \times G$

$$\left(\sigma_d^2 I_{TR} + K_{d,W}\right)^{-1} = \left(\sigma_d^2 I_{TR} + \mathbf{\Phi}_d^{(2)}\mathbf{\Lambda}_d^{(2)}\mathbf{\Phi}_d^{(2)\mathrm{T}}\right)^{-1}$$
$$= \frac{1}{\sigma_d^2} I_{TR} - \frac{1}{\sigma_d^4}\mathbf{\Phi}_d^{(2)}\left\{\left(\mathbf{\Lambda}_d^{(2)}\right)^{-1} + \tfrac{1}{\sigma_d^2}\mathbf{\Phi}_d^{(2)\mathrm{T}}\mathbf{\Phi}_d^{(2)}\right\}^{-1}\mathbf{\Phi}_d^{(2)\mathrm{T}}$$

and denoting $(\sigma_d^2 I_{TR} + K_{d,W})^{-1}$ by $Q$ for simplicity, we expand the second inverse as

$$\left(\sigma_d^2 I_{TR} + K_{d,W} + J_i K_{d,B}\right)^{-1} = \left(Q^{-1} + J_i \mathbf{\Phi}_d^{(1)}\mathbf{\Lambda}_d^{(1)}\mathbf{\Phi}_d^{(1)\mathrm{T}}\right)^{-1}$$
$$= Q - J_i Q \mathbf{\Phi}_d^{(1)}\left\{\left(\mathbf{\Lambda}_d^{(1)}\right)^{-1} + J_i \mathbf{\Phi}_d^{(1)\mathrm{T}} Q \mathbf{\Phi}_d^{(1)}\right\}^{-1}\mathbf{\Phi}_d^{(1)\mathrm{T}} Q.$$

# Simplifying covariance further

- Woodbury matrix identity:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

- Apply to our two matrices

$G \times G$

$$\begin{aligned}\left(\sigma_d^2 I_{TR} + K_{d,W}\right)^{-1} &= \left(\sigma_d^2 I_{TR} + \boldsymbol{\Phi}_d^{(2)} \boldsymbol{\Lambda}_d^{(2)} \boldsymbol{\Phi}_d^{(2)\mathrm{T}}\right)^{-1} \\ &= \frac{1}{\sigma_d^2} I_{TR} - \frac{1}{\sigma_d^4} \boldsymbol{\Phi}_d^{(2)} \left\{ \left(\boldsymbol{\Lambda}_d^{(2)}\right)^{-1} + \tfrac{1}{\sigma_d^2} \boldsymbol{\Phi}_d^{(2)\mathrm{T}} \boldsymbol{\Phi}_d^{(2)} \right\}^{-1} \boldsymbol{\Phi}_d^{(2)\mathrm{T}}\end{aligned}$$

and denoting $(\sigma_d^2 I_{TR} + K_{d,W})^{-1}$ by $Q$ for simplicity, we expand the second inverse as

$H \times H$

$$\begin{aligned}\left(\sigma_d^2 I_{TR} + K_{d,W} + J_i K_{d,B}\right)^{-1} &= \left(Q^{-1} + J_i \boldsymbol{\Phi}_d^{(1)} \boldsymbol{\Lambda}_d^{(1)} \boldsymbol{\Phi}_d^{(1)\mathrm{T}}\right)^{-1} \\ &= Q - J_i Q \boldsymbol{\Phi}_d^{(1)} \left\{ \left(\boldsymbol{\Lambda}_d^{(1)}\right)^{-1} + J_i \boldsymbol{\Phi}_d^{(1)\mathrm{T}} Q \boldsymbol{\Phi}_d^{(1)} \right\}^{-1} \boldsymbol{\Phi}_d^{(1)\mathrm{T}} Q.\end{aligned}$$

# Functional intraclass correlation

- Advantage of multilevel models: quantifying similarity between repetitions within subjects

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

- High fICC $\Rightarrow$ more of the variability is explained by differences across subjects and repetitions are similar within a subject
- Different from fICC proposed in Di et al. which did not have a region dimension
- Point estimate and percentile bootstrap confidence intervals

# Inference

- A larger number of components are estimated in the eigendecompositions of the marginal covariances in order to increase the accuracy of the total variation
- For inference on the means, a more parsimonious model is desired $\Rightarrow$ retain enough leading product components to explain 80% of the variation at each level
- Define the FVE at each level as

$$FVE_{dB}(G') = \frac{\sum_{g=1}^{G'} \widehat{\lambda}_{dg}^{(1)}}{\sum_{g=1}^{G} \widehat{\lambda}_{dg}^{(1)}} \qquad \text{and} \qquad FVE_{dW}(H') = \frac{\sum_{h=1}^{H'} \widehat{\lambda}_{dh}^{(2)}}{\sum_{h=1}^{H} \widehat{\lambda}_{dh}^{(2)}}$$

  where $G'$ and $H'$ denote the numbers of product components chosen at each level such that $FVE_{dB}(G')$ and $FVE_{dW}(H')$ are greater than 0.8
- Variance components are re-estimated by including only the leading components and group-level inference can be drawn via parametric bootstrap

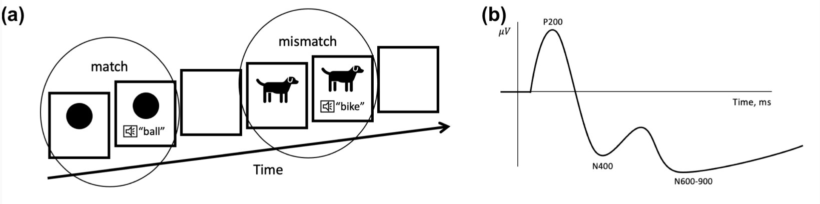
# Applications

## Audio odd-ball paradigm

# Application: Language impairment in autism

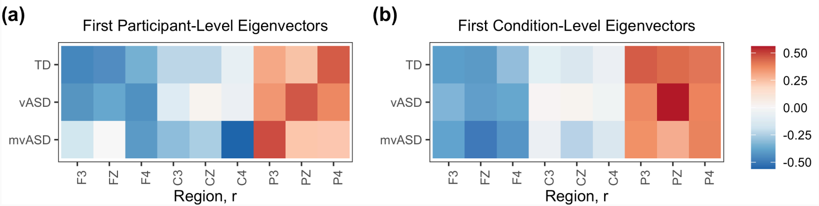
- Autism spectrum disorder (ASD): developmental disorder that affects communication and behavior
- Goal: study the neural mechanisms underlying language impairment in children with ASD (DiStefano, 2019)
- Study cohort: 31 children aged 5-11 years old were recruited
  - Typically Developing (TD): n = 14
  - Verbal ASD (vASD): n = 10
  - Minimally Verbal (mvASD): n = 7

# Audio odd-ball paradigm




- Multilevel: A picture was presented and an audio recording of a spoken word was played that either matched or did not match (match vs. mismatch)
- Regional: 9 electrodes
- Functional: 137 equally spaced time points between -96ms to 992ms
- Three ERP components of interest:
  - P200: attention
  - N400: semantic processing
  - N600-900 or late negative component (LNC): semantic integration

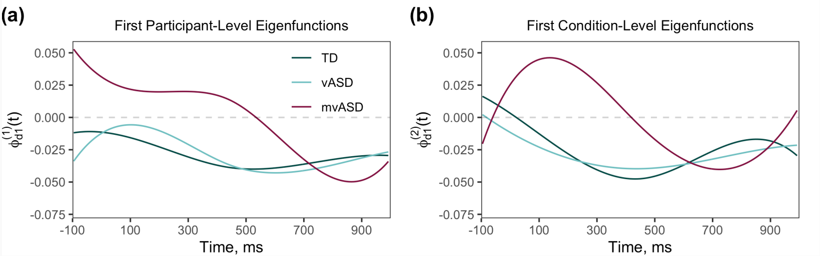
# Marginal eigenvectors




- Leading eigenvectors at both levels represent a contrast between electrodes at the front and back of the scalp for all groups

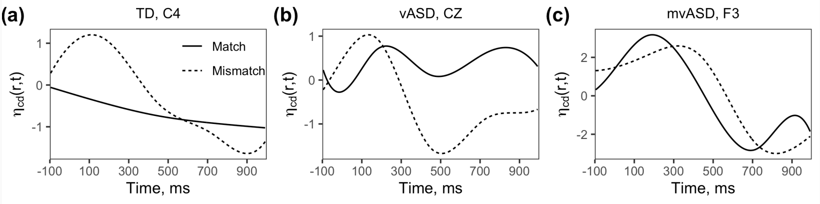
- Related to the dipole effect: observed signals are exact opposites on opposing sides of the scalp

# Marginal eigenfunctions




- TD and vASD: leading participant- and condition-level eigenfunctions in the TD and vASD groups signal variability in N400 and LNC components
- mvASD: most of the variation at the participant level is observed in the late LNC component and most of the variation at the condition level is observed in the contrast between P200 and LNC components

# Condition differentiation




- TD: only trending condition differentiation was detected at C4 ($p = 0.055$), but estimated condition-specific mean functions from almost all regions visually showed a condition difference
- vASD: condition differentiation strongest at Cz ($p = 0.025$), difference between conditions was mostly driven by the N400 and LNC
- mvASD: detected at F3, partly due to the late LNC ($p = 0.045$)

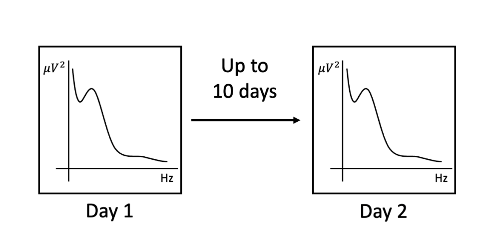
# Applications

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

# Application: Day-to-day test-retest reliability of PSD

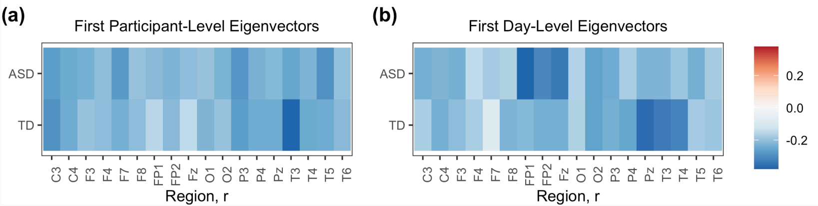
- Goal: study the day-to-day test-retest reliability of PSD as part of the Feasibility study of ABC-CT (Levin, 2020)
- Study cohort: 47 children aged 5-11 years old were recruited
    - Typically Developing (TD): n = 25
    - ASD (ASD): n = 22

# Resting-state paradigm



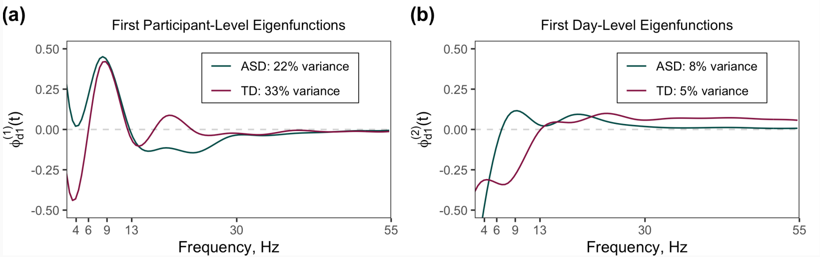

- Multilevel: Participants were shown screensaver-like videos on two separate days, a median of 6 days apart
- Regional: 18 electrodes
- Functional: 55 equally spaced frequency points between 0 to 55 Hz

# Marginal eigenvectors




- Signal constant variation across the scalp for both groups and levels
- Modeling data also along the regional dimension, provides independent support for the previous study of Levin et al. using the scalp-wide average of PSD

# Marginal eigenfunctions




- Leading participant-level eigenfunctions: most of the variation *across subjects* is observed in the alpha peak amplitude in both groups
- Leading day-level eigenfunctions: variation *across days* is in the dominant peak location within alpha to beta bands in ASD, within theta to alpha in TD

# Functional ICC

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

- This fICC is defined using the variances of the product components which capture the variation along the regional dimension as well as functional dimension
- Estimated to be 0.673 [95% CI (0.626, 0.793)] for ASD and 0.656 for TD [95% CI (0.639, 0.776)]
- Signals good agreement in within subject day-to-day PSD, where most of the variation is due to differences across subjects, not differences across days

# Simulation

# Data generation

- Generated data from the model

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

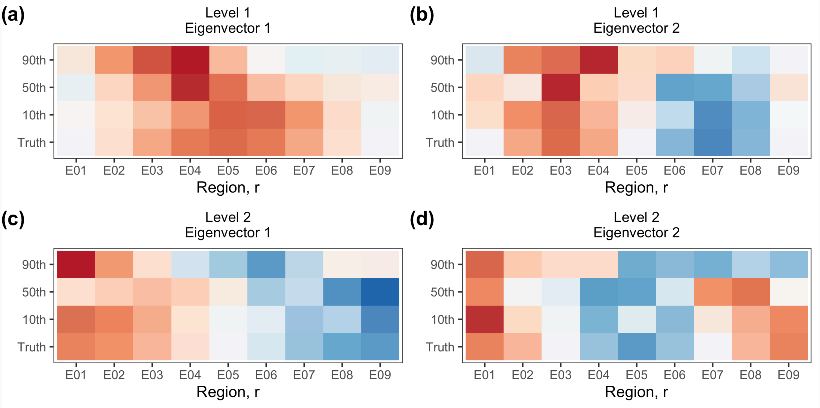
- Simulations were conducted for
    - Two sample sizes: $n_d = 15$ and $n_d = 50$
    - Two noise levels: $\sigma_d^2 = 0.25$ and $\sigma_d^2 = 1$
    - Two data sparsity levels: dense and sparse

# Effects of sample size, noise, and sparsity

- Relative square errors (RSEs) and mean square errors (MSEs) generally decrease with increasing sample size and are lower for the dense settings than the sparse settings

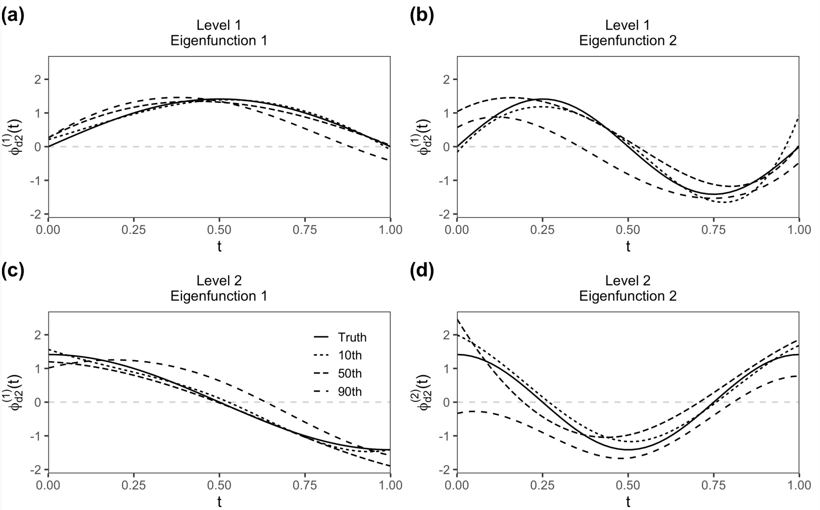
- RSEs increase with additional noise

# Eigenvector estimation



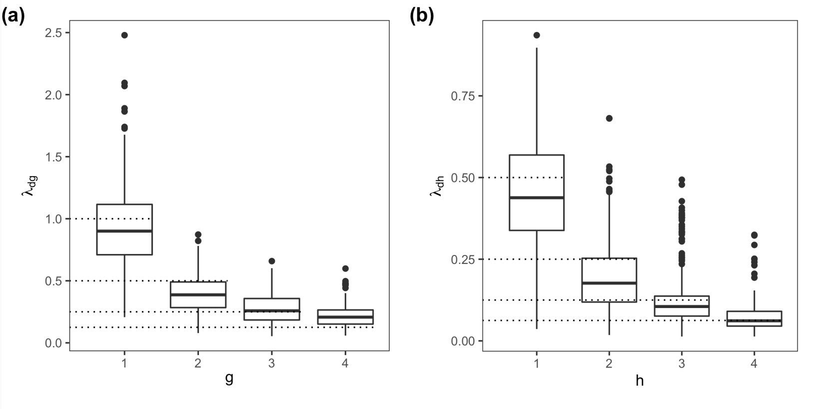

- Eigenvectors are generally estimated very well, with occasional confusion between the first and second eigenvector at each level due to small variances

# Eigenfunction estimation



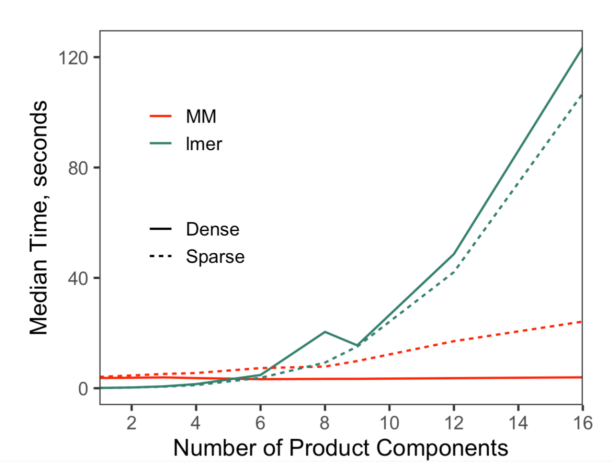

- Eigenfunction are estimated very well

# Eigenvalue estimation



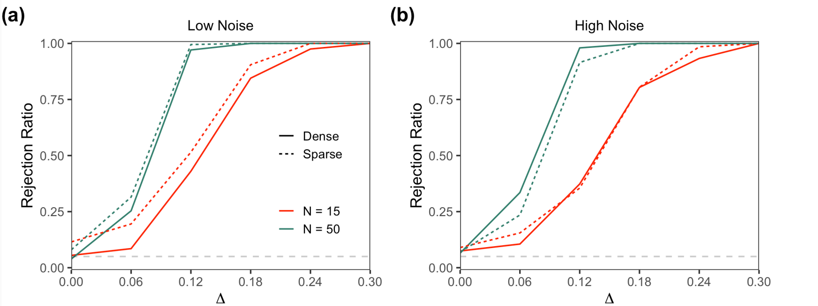

- Eigenvalues are on average estimated well, even with high noise and a small sample size

# Comparisons to `lme4`



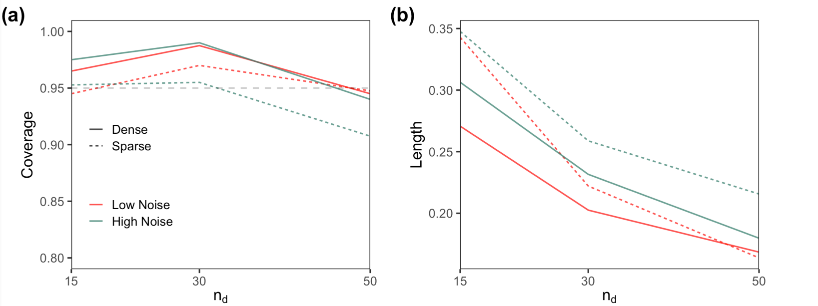

- M-HPCA's MM algorithm step is much faster than using the popular `lme4` package

# Power of bootstrap tests




- Tested the equality of the group-region-repetition shifts within the first region, $H_0 : \eta_{1j}(1, t) = \eta_1(1, t)$
- Generate data with the modification $\eta_{dj}(r, t) = 0$ for $r \neq 1$ and $d \neq 1$ and $\eta_{1j}(1, t) = (-1)^j \Delta$ for $r = 1$ and $d = 1$
- Level of the test hovers around 0.05 for varying levels of noise, sample size, and sparsity
- Power increases faster for the larger sample size

# Coverage and length of fICC confidence intervals




- Coverage of 95% CI hovers around 95% in all simulation designs
- Length of the confidence intervals is shorter for larger sample sizes, less noise, and dense data

# Concluding Remarks

- What did we do
  - Proposed multilevel hybrid principal components analysis (M-HPCA) that efficiently combines techniques used for multilevel functional data and region-referenced functional data
- Extensions
  - Can easily be extended to data with two functional dimensions and one regional dimension by decomposing the between- and within-subjects variation into three marginal covariances rather than two using weak separability
- Limitations
  - limitation?

# Thank you!

Slides available at bit.ly/m-hpca-talk
R package available at bit.ly/m-hpca

# References


- DiStefano C, Şentürk D, Jeste SS. ERP evidence of semantic processing in children with ASD. Developmental Cognitive Neuroscience 2019; 36: 100640. doi: 10.1016/j.dcn.2019.100640
- Levin AR, Naples AJ, Scheffler AW, et al. Day-to-Day Test-Retest Reliability of EEG Profiles in Children With Autism Spectrum Disorder and Typical Development. Frontiers in Integrative Neuroscience 2020; 14(April): 1–12. doi: 10.3389/fnint.2020.00021
- Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel Functional Principal Component Analysis. Ann Appl Stat. 2009;3(1):458-488. doi:10.1214/08-AOAS206SUPP
- Scheffler A, Telesca D, Li Q, et al. Hybrid principal components analysis for region-referenced longitudinal functional EEG data. Biostatistics 2020; 21(1): 139–157. doi: 10.1093/biostatistics/kxy034