Biostatistics (2015), **0**, 0, *pp.* 1–13
doi:10.1093/biostatistics/supp˙Biostat˙revision˙final

Supplementary Material to "Robust functional clustering of ERP data with application to a study of implicit learning in autism"

KYLE HASENSTAB

Department of Statistics, University of California, Los Angeles, CA, USA

CATHERINE SUGAR, DONATELLO TELESCA

Department of Biostatistics, University of California, Los Angeles, CA, USA

SHAFALI JESTE

Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA, USA

DAMLA ŞENTÜRK*

Department of Biostatistics, University of California, Los Angeles, CA, USA

dsenturk@ucla.edu

1. Estimation of FPCA model components

The mean functions of the functional trajectories are obtained by local linear smoothing; the smooth covariance surfaces are estimated via a two dimensional local least squares algorithm. To eliminate the effects of measurement error, the diagonal elements of the raw covariance matrix are removed before the two-dimensional smoothing step. A nonparametric functional principal

*To whom correspondence should be addressed.

component analysis step will be employed on the smooth estimate of the covariance surface by a standard discretization procedure to estimate the eigenfunctions and eigenvalues. In order to guarantee that the covariance matrix is non-negative definite, any eigenvalues with negative estimates and their corresponding eigenfunctions are removed from the functional principal component decompositions of the covariances. For explicit expressions of the estimated mean, eigenfunctions and eigenvalues, denoted by $\widehat{\mu}(t)$, $\widehat{\phi}_k(t)$ and $\widehat{\lambda}_k$, respectively, as well as the covariance surfaces, readers are referred to Şentürk and Nguyen (2011) and Şentürk *and others* (2013). Bandwidths for the smooths may be selected using cross-validation or generalized cross-validation (GCV) (Yao, Müller and Wang, 2005). For a computationally efficient bandwidth choice in the proposed smoothing procedures, we adopt the generalized cross-validation algorithm of Liu and Muller (2008).

The estimates of the mean, eigenfunctions, eigenvalues for trajectories in covariance subset v and cluster c, denoted by $\widehat{\mu}^{(c)}(t)$, $\widehat{\phi}_k^{(v)}(t)$, $\widehat{\lambda}_k^{(v)}$ and $\hat{\xi}_{ik}^{(c,v)}$, respectively, are obtained similarly based on trajectories in the specific subset and cluster in the iterative steps of the RFC algorithm. For dense functional data, subject-specific scores can be estimated using the projection, $\hat{\xi}_{ik}^{(c,v)} = \int\{y_i^{(c,v)}(t) - \widehat{\mu}^{(c)}(t)\}\widehat{\phi}_k^{(v)}(t)\mathrm{d}t$; for sparse data applications, the best linear unbiased prediction (BLUP) of the scores has been proposed by Yao, Müller and Wang (2005). The single-level FPCA model can be considered a linear mixed effects model with fixed effects mean $\mu^{(c)}(t)$ and random effects $\xi_{ik}^{(c,v)}$. Random effects model matrices are the eigenfunctions corresponding to the random effects scores. Under this framework, scores may be estimated using their best linear unbiased predictions based on the conditional expectation of the scores given the data. Let $y_{ip} = y_i(t_{ip})$, $p = 1, \ldots, T_i$ denote an observation on the i^{th} subject at time t_{ip} and let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT_i})^{\mathrm{T}}$ denote the $T_i \times 1$ observation vector across all time points. Further let $\hat{\boldsymbol{\mu}}_i^{(c)} = \{\hat{\mu}^{(c)}(t_{i1}), \ldots, \hat{\mu}^{(c)}(t_{iT_i})\}^{\mathrm{T}}$, $\widehat{\boldsymbol{\Phi}}_i^{(v)}$ denote the $T_i \times K_v$ matrix of estimated eigenfunctions whose k^{th} column is the vector $\{\hat{\phi}_k^{(v)}(t_{i1}), \ldots, \hat{\phi}_k^{(v)}(t_{iT_i})\}^{\mathrm{T}}$ and let $\widehat{\boldsymbol{\Lambda}}^{(v)}$ denote the $K_v \times K_v$

diagonal matrix $\widehat{\boldsymbol{\Lambda}}^{(v)} = \text{diag}(\hat{\lambda}_1^{(v)}, \ldots, \hat{\lambda}_{K_v}^{(v)})$. Then the $K_v \times 1$ estimator vector for the single-level functional principal component scores are given by

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

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

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

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

where $\widehat{\boldsymbol{D}}_{i,\xi}^{(v)} = \{\widehat{\boldsymbol{\Lambda}}^{(1,v)} \widehat{\boldsymbol{\Phi}}_{i1}^{(1,v)^{\mathrm{T}}}, \ldots, \widehat{\boldsymbol{\Lambda}}^{(1,v)} \widehat{\boldsymbol{\Phi}}_{iJ}^{(1,v)^{\mathrm{T}}}\}$, $\widehat{\boldsymbol{D}}_{i,\zeta}^{(v)} = \{\widehat{\boldsymbol{\Lambda}}^{(2,v)} \widehat{\boldsymbol{\Phi}}_{i1}^{(2,v)^{\mathrm{T}}}, \ldots, \widehat{\boldsymbol{\Lambda}}^{(2,v)} \widehat{\boldsymbol{\Phi}}_{iJ}^{(2,v)^{\mathrm{T}}}\}$ and $\widehat{\boldsymbol{\Sigma}}_{T,i}^{(v)}$ denotes the estimated $(\sum_j T_{ij}) \times (\sum_j T_{ij})$ block covariance matrix with the (j, j') block equal to $\widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)} \widehat{\boldsymbol{\Lambda}}^{(1,v)} \widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)^{\mathrm{T}}} + \widehat{\boldsymbol{\Phi}}_{ij}^{(2,v)} \widehat{\boldsymbol{\Lambda}}^{(2,v)} \widehat{\boldsymbol{\Phi}}_{ij}^{(2,v)^{\mathrm{T}}} + \hat{\sigma^2}^{(v)} I_{T_{ij} \times T_{ij}}$ when $j = j'$ and $\widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)} \widehat{\boldsymbol{\Lambda}}^{(1,v)} \widehat{\boldsymbol{\Phi}}_{ij}^{(1,v)^{\mathrm{T}}}$ when $j \neq j'$.

2. Description of the multilevel RFC algorithm

Similar to the single-level case, we assume that $Y_{ij}(t)$ is sampled from a mixture of subprocesses with cluster means and induced covariance subsets. Motivated by similar low data quality patterns, and hence similar functional trajectories of number of averaged ERPs from sliding windows

across electrodes, covariances at both levels of the multilevel deviations are grouped in the same covariance subset. In addition, since mean functions within subjects across electrodes are quite similar in our application, cluster membership of different electrode trajectories within a subject are assumed to be the same, where clusters are determined more based on the shapes of the overall means $\mu(t)$. Hence, conditional on cluster c and covariance subset v, the means and covariances of the subprocesses are given by

$$E\{Y_{ij}(t)\} = \mu^{(c)}(t) + \eta_j^{(c)}(t), \quad \text{cov}\{Y_{ij}(s), Y_{ij'}(t)\} = \Sigma_B^{(v)}(s,t), \qquad j \neq j', \tag{2.1}$$

$$\text{cov}\{Y_{ij}(s), Y_{ij}(t)\} = \Sigma_T^{(v)}(s,t) = \Sigma_B^{(v)}(s,t) + \Sigma_W^{(v)}(s,t) + \sigma^{2^{(v)}} I(s=t), \tag{2.2}$$

where $\Sigma_T^{(v)}(s,t)$ is the overall covariance function, $\Sigma_B^{(v)}(s,t)$ and $\Sigma_W^{(v)}(s,t)$ are the between and within subunit covariance functions, respectively, and $\sigma^{2^{(v)}} I(s=t)$ represents the error variance on the diagonal of $\Sigma_T^{(v)}(s,t)$. The between and within structures may be decomposed further using their eigenvalues and eigenfunctions such that $\Sigma_B^{(v)}(s,t) = \sum_k \lambda_k^{(1,v)} \phi_k^{(1,v)}(s)\phi_k^{(1,v)}(t)$ and $\Sigma_W^{(v)}(s,t) = \sum_\ell \lambda_\ell^{(2,v)} \phi_\ell^{(2,v)}(s)\phi_\ell^{(2,v)}(t)$.

The multilevel RFC adopts the same structure as the single-level algorithm. Details on the multilevel RFC algorithm are described in the algorithm table below. Since we are interested in clustering individuals largely by subject-level (level 1) differences and level 2 is mainly used to incorporate dependencies within subjects across electrodes, initial clustering is performed by applying k-means to the estimated level 1 scores (ξ_{ik}) based on the entire sample (Step 1).

Cluster membership updates still utilize functional predictions based on the estimated non-parametric truncated multilevel random effects model

$$\widehat{y}_{ij}^{(c,v)}(t_{ijp}) = \widehat{\mu}_{(-i)}^{(c)}(t_{ijp}) + \widehat{\eta}_{j(-i)}^{(c)}(t_{ijp}) + \sum_{k=1}^{K_v} \widehat{\xi}_{ik}^{(c,v)} \widehat{\phi}_{k(-i)}^{(1,v)}(t_{ijp}) + \sum_{\ell=1}^{L_v} \widehat{\zeta}_{ijl}^{(c,v)} \widehat{\phi}_{\ell(-i)}^{(2,v)}(t_{ijp}), \tag{2.3}$$

where the $(-i)$ notation denotes that the observed multilevel functional data for subject i has been left out while obtaining the estimates. In (2.3), K_v and L_v refer to the number of eigen components selected for the between and within levels, respectively. First and second level eigenfunctions in

Multilevel RFC Algorithm

1. Perform multilevel FPCA on the entire sample and cluster level 1 scores $\widehat{\xi}_{ik}$ to obtain the initial mean cluster memberships $c_i^{(0)}$ for $i = 1, \ldots, n$.
2. For each subject i belonging to covariance subset v and assigned to mean cluster c during iteration r:
   (a) Estimate $\widehat{\mu}^{(c)}_{(-i)}(t)$, $\widehat{\eta}^{(c)}_{j(-i)}(t)$, $c = 1, \ldots, C$, using all subjects assigned to mean cluster c at iteration r while leaving out the i^{th} subject.
   (b) Estimate $\widehat{\phi}^{(1,v)}_{k(-i)}(t)$, $k = 1, \ldots, K_v$, and $\widehat{\phi}^{(2,v)}_{\ell(-i)}(t)$, $\ell = 1 \ldots, L_v$, for covariance subset v using all mean centered trajectories belonging to covariance v while leaving out the i^{th} subject.
   (c) Estimate $\widehat{\xi}^{(c,v)}_{ik}$, $\widehat{\zeta}^{(c,v)}_{ijl}$, for $c = 1 \ldots, C$ and covariance subset v.
   (d) Calculate predictions for $c = 1, \ldots, C$ and covariance subset v via

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

   (e) Assign i^{th} subject to mean cluster

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

3. Repeat Step 2 until no curve is reclassified.

(2.3) are estimated based on subunit trajectories within subjects in covariance subset v, centered by subtraction of their overall mean $\widehat{\mu}^{(c)}_{(-i)}(t)$ and the subunit-specific mean deviation $\widehat{\eta}^{(c)}_{j(-i)}(t)$ for subjects in cluster c. First and second level subject-specific scores are estimated as projections, similar to the single-level case for dense functional data (Di *and others*, 2009) and via BLUP estimates for sparse cases (Di, Crainiceanu and Jank, 2014), where explicit forms are provided in Section 1 of the online Supplementary Material. In contrast to the single-level case, the criterion in Step 2(e) also sums over J subunits within a subject.

3. Comparison of clustering results in the data application

We compare the three clustering algorithms according to multilevel extensions of the Calinski-Harabasz index (CHI) (Calinski and Harabasz, 1974), Davies-Bouldin Index (DBI) (Davies and Bouldin, 1979) and Silhouette index (SI) (Rousseeuw, 1987). The CHI is an internal cluster evaluation metric that incorporates information on within-cluster variation and differences between mean cluster estimates and global mean estimates across the entire sample. Higher CHI values indicate better cluster quality. Define $B = \sum_{c=1}^{C} N_c \sum_{j=1}^{J} \int_t [\{\mu^{(c)}(t) + \eta_j^{(c)}(t)\} - \{\mu(t) + \eta_j(t)\}]^2 \mathrm{d}t$ to be a separation measure between cluster means and the mean of the entire sample, where N_c represents the number of subjects in cluster c. Define $W = \sum_{c=1}^{C} \sum_{i \in c} \sum_{j=1}^{J} \sum_{p=1}^{T_{ij}} \{y_{ij}^{(c,v)}(t_{ijp}) - \mu^{(c)}(t_{ijp}) - \eta_j^{(c)}(t_{ijp})\}^2$ to be a measure of within cluster similarity. The CHI is defined to be $\{(n-C)B\}/\{(C-1)W\}$.

The DBI is another internal cluster evaluation metric that assesses within cluster variation and between cluster separation, where lower DBI values correspond to better clustering results and greater cluster separation. Define $S_c = (1/N_c J) \sum_{i=1}^{N_c} \sum_{j=1}^{J} [\sum_{p=1}^{T_{ij}} \{y_{ij}^{(c,v)}(t_{ijp}) - \mu^{(c)}(t_{ijp}) - \eta_j^{(c)}(t_{ijp})\}^2]^{1/2}$ to be a measure of within cluster variation for mean cluster c. Further define $M_{c,c'} = \sum_{j=1}^{J} (\sum_{p=1}^{T_{ij}} [\{\mu^{(c)}(t_{ijp}) + \eta_j^{(c)}(t_{ijp})\} - \{\mu^{(c')}(t_{ijp}) + \eta_j^{(c')}(t_{ijp})\}]^2)^{1/2}$ to be a measure of separation between two clusters c and c'. The DBI is then defined to be $(1/n_c) \sum_{c=1}^{n_c} \max_{c \neq c'} \{(S_c + S_{c'})/M_{c,c'}\}$, $c' = 1, \ldots, C$, where the definition reduces to $(S_c + S_{c'})/M_{c,c'}$ when $C = 2$.

The SI incorporates the average dissimilarity of individual trajectories across different clusters and within the same cluster. SI values range from -1 to 1, where -1 indicates poor clustering results and 1 indicates that the data is appropriately clustered. An SI of zero suggests trajectories are on cluster borders. For subjects i, i' in cluster c, define $a(i) = \{1/(N_c - 1)\} \sum_{i' \in c, i \neq i'} \sum_{j=1}^{J} [\sum_{p=1}^{T_{ij}} \{y_{ij}^{(c,v)}(t_{ijp}) - y_{i'j}^{(c,v)}(t_{i'jp})\}^2]^{1/2}$ to be a dissimilarity measure of the trajectories for subject i to the trajectories of other subjects within the same cluster. For subject i in cluster c and subject i' in cluster c', define $d_{c'}(i) = (1/N_{c'}) \sum_{i \in c'} \sum_{j=1}^{J} [\sum_{p=1}^{T_{ij}} \{y_{ij}^{(c,v)}(t_{ijp}) - y_{i'j}^{(c',v)}(t_{i'jp})\}^2]^{1/2}$ and $b(i) =$

$\min_{c \neq c'} d_{c'}(i)$ to be a dissimilarity measure of the trajectories for subject i to the trajectories of other subjects in other clusters. Define $s(i) = \{b(i) - a(i)\} / \max\{a(i), b(i)\}$ and $s_c = (1/N_c) \sum_{i \in c} s(i)$. The SI is defined to be $(1/C) \sum_{c=1}^{C} s_c$. Note that the SI involves finding distances between trajectories across pairs of subjects within and between clusters that may not be observed across the same set of trials. Therefore, distances are calculated across the set of trials that are observed for both subjects.

The results of the three internal cluster validation indices for the RFC, SFC and FC algorithms are shown in Table 1. Indices show that RFC achieves a better cluster separation over the other two algorithms within each group. Differences in the index values across the three algorithms appear greater in the ASD group.

Table 1: Internal cluster validation indices for the RFC, SFC and FC algorithms within the TD and ASD groups. A superscript h or l indicates that higher or lower index values, respectively, correspond to better cluster quality.

	TD			ASD		
	RFC	SFC	FC	RFC	SFC	FC
CHI^h	5.26	4.37	2.80	3.83	2.16	1.64
DBI^l	2.16	2.44	2.53	2.65	3.47	3.90
SI^h	0.13	0.09	0.10	0.08	0.03	-0.01

4. Simulation studies

The goals of the simulations are to study the performance of the proposed RFC compared to FC and SFC, and to study the performance of the algorithm under the second non-identifiability condition outlined in Section 3.2 that the cluster mean functions lie in the same or different covariance subset eigenspaces. As in the data applications, we consider two clusters and covariance subsets with a total sample size of 35 subjects and $J = 4$ subunits. Response trajectories $y_{ij}^{(c,v)}(t)$

for cluster c and covariance group v are generated using the random effects model

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

where subunit-specific mean functions $\eta_j^{(c)}(t)$ are taken to be zero for simplicity. The number of eigen components are taken to be two at both levels ($K_v = L_v = 2$) where eigenscores $\xi_{ik}^{(c,v)}$ and $\zeta_{ijl}^{(c,v)}$ are independently sampled from $N(0, \lambda_k^{(1,v)})$ and $N(0, \lambda_\ell^{(2,v)})$, respectively. The measurement error $\epsilon_{ij}^{(v)}(t)$ is sampled from $N(0, \sigma^{2^{(v)}})$ and observation times t are regular and equally spaced in the interval $\mathcal{T} = [0, 1]$ with $T_{ij} = 40$. The eigenvalues and error variance are selected as $\lambda^{(1,1)} = (500, 200)/d$, $\lambda_1^{(2,1)} = (250, 100)/d$, $\lambda^{(1,2)} = (700, 300)/d$, $\lambda^{(2,2)} = (500, 250)/d$, $\sigma^{2^{(1)}} = 10/d$, $\sigma^{2^{(2)}} = 50/d$ with the divisor d chosen specifically for the particular simulation case considered. The means and eigenfunctions used are defined as follows: $\mu^{(1)}(t) = 2\exp\{-10(t - 0.3)^2\} + \exp\{-1(t-0.75)^2\} - 1.5$, $\mu^{(2)}(t) = -3\exp\{-10(t-0.35)^2\} - \exp\{(t-0.5)^2\} + 2$, $\phi_1^{(1,1)}(t) = \sqrt{2}\sin(\pi t)$, $\phi_2^{(1,1)}(t) = \sqrt{2}\cos(\pi t)$, $\phi_1^{(2,1)}(t) = \mathbf{1}t$, $\phi_2^{(2,1)}(t) = \sqrt{3}(2t - 1)$, $\phi_1^{(1,2)}(t) = \sqrt{2}\sin(2\pi t)$, $\phi_2^{(1,2)}(t) = \sqrt{2}\cos(2\pi t)$, $\phi_1^{(2,2)}(t) = \sqrt{5}(6t^2 - 6t + 1)$, $\phi_2^{(2,2)}(t) = \sqrt{7}(20t^3 - 30t^2 + 12t - 1)$.

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

entire sample across simulation scenarios).

Performance of the clustering algorithms is evaluated using three common measures of cluster quality, the correct classification rate (CCR), the adjusted Rand index (ARI) and the normalized mutual information (NMI). The correct classification rate measures the proportion of subjects correctly classified in reference to the true external clustering. CCR is defined as the maximum proportion of correctly classified subjects among all cluster label correspondences between the current cluster output and the true external cluster reference. The second measure, ARI, quantifies the degree of similarity between two partitions by comparing clustering results against an external true clustering and takes into account the number of object pairs in the same set or in different sets in the two partitions (Steinley, 2004). ARI corrects the original Rand index by enforcing a constant expected value of 0 and an upper bound of 1 under the assumption of a generalized hypergeometric model. The mutual information quantifies the information shared across a pair of cluster partitions by accounting for the number of subjects assigned to the same cluster or different clusters. NMI normalizes the mutual information using the geometric mean of the entropies from each partition to bound the index between 0 and 1 (Strehl and Ghosh, 2002). Higher CCR, ARI and NMI indicate higher cluster quality.

The means and percentiles of the CCR, ARI and NMI values from 200 Monte Carlo runs for the RFC, SFC and FC algorithms under the five simulation cases are given in Table 2. We also plot the medians of the cluster mean estimates across the 200 Monte Carlo runs along with the true cluster means in Figure 1 for different algorithms and simulation cases. We conducted a preliminary simulation study to select smoothing bandwidths for the mean function and the covariance surface. The optimal bandwidths were selected using generalized cross-validation at each iteration from candidates in $[.05, .3]$ and $[.15, .3]$ for the mean function and covariance surface, respectively. Note that the interval of candidate bandwidths is larger for the two-dimensional covariance smoothing to achieve stable estimates at the considered small sample size. We found

that the selected bandwidths were similar across iterations, therefore we fixed the bandwidth choice for the full simulation study at 0.05 and 0.15 for the mean and covariance smooths, respectively, to reduce computation time. All three algorithms perform poorly in the first two simulation cases of non-identifiability conditions (Figures 1(a)-(b)), since the cluster means lying in the eigenspace of the covariance subsets is also a non-identifiable case for SFC and FC with non-overlapping cluster and covariance subset memberships. RFC outperforms SFC and FC in the third simulation case (Figure 1(c)), improving cluster quality by incorporating the known covariance heterogeneity into the clustering of the mean trends. When the covariance groups are highly similar (simulation case 4), all three algorithms perform equally well as expected (Figure 1(d)). In case (5), RFC is almost as effective in finding clusters as FC, where cluster and covariance subset membership overlap. As in case (3), SFC is unable to recover clusters under the multiple covariance subsets of case (5) (Figure 1(e)). Due to the small sample size, all three algorithms perform poorly occasionally, as reflected in the lower CCR, ARI and NMI values for the 5th percentile.

References

Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics* **3**, 1–27.

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

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

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

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

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics* **20**, 53–65.

Şentürk, D. and Nguyen, D. V. (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. V. (2013). Modeling time varying effects with generalized and unsynchronized longitudinal data. *Statistics in Medicine* **32**, 2971–2987.

Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psyhological Methods* **9**, 386–396.

Strehl, A. and Ghosh, J. (2002). Cluster ensembles - A knowledge reuse framework for combining multiple partitions *Journal of Machine Learning Research* **3**, 583–617.

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

Table 2: The correct classification rate (CCR), the adjusted Rand index (ARI) and the normalized mutual information (NMI) means, medians and (5th, 95th) percentiles for the RFC, SFC and FC algorithms over 200 Monte Carlo runs. While the first two simulation cases correspond to the non-identifiability conditions, the last three correspond to the assumptions of RFC, SFC and FC algorithms, respectively.

		RFC		SFC		FC	
		Mean	Percentile	Mean	Percentile	Mean	Percentile
Case 1	CCR	0.79	0.80 (0.54, 0.97)	0.84	0.86 (0.59, 1.00)	0.59	0.56 (0.51, 0.84)
	ARI	0.38	0.34 (-0.02, 0.89)	0.51	0.50 (0.00, 1.00)	0.04	-0.01 (-0.03, 0.46)
	NMI	0.38	0.37 (0.02, 0.83)	0.47	0.44 (0.02, 1.00)	0.07	0.01 (0.00, 0.50)
Case 2	CCR	0.78	0.80 (0.51, 1.00)	0.62	0.60 (0.51, 0.89)	0.58	0.57 (0.51, 0.69)
	ARI	0.43	0.34 (-0.03, 1.00)	0.08	0.01 (-0.03, 0.58)	0.01	-0.01 (-0.03, 0.11)
	NMI	0.43	0.27 (0.00, 1.00)	0.09	0.03 (0.00, 0.50)	0.04	0.01 (0.00, 0.16)
Case 3	CCR	0.90	1.00 (0.54, 1.00)	0.76	0.74 (0.51, 1.00)	0.58	0.57 (0.51, 0.81)
	ARI	0.73	1.00 (-0.02, 1.00)	0.35	0.21 (-0.03, 1.00)	0.03	-0.01 (-0.03, 0.38)
	NMI	0.72	1.00 (0.01, 1.00)	0.33	0.22 (0.00, 1.00)	0.06	0.01 (0.00, 0.45)
Case 4	CCR	0.93	1.00 (0.56, 1.00)	0.96	1.00 (0.57, 1.00)	0.97	1.00 (0.69, 1.00)
	ARI	0.81	1.00 (-0.01, 1.00)	0.89	1.00 (-0.01, 1.00)	0.93	1.00 (0.10, 1.00)
	NMI	0.80	1.00 (0.01, 1.00)	0.89	1.00 (0.02, 1.00)	0.93	1.00 (0.16, 1.00)
Case 5	CCR	0.98	1.00 (0.86, 1.00)	0.77	0.77 (0.54, 0.97)	0.99	1.00 (0.97, 1.00)
	ARI	0.94	1.00 (0.49, 1.00)	0.33	0.27 (-0.02, 0.89)	0.97	1.00 (0.89, 1.00)
	NMI	0.94	1.00 (0.49, 1.00)	0.36	0.33 (0.01, 0.84)	0.97	1.00 (0.84, 1.00)

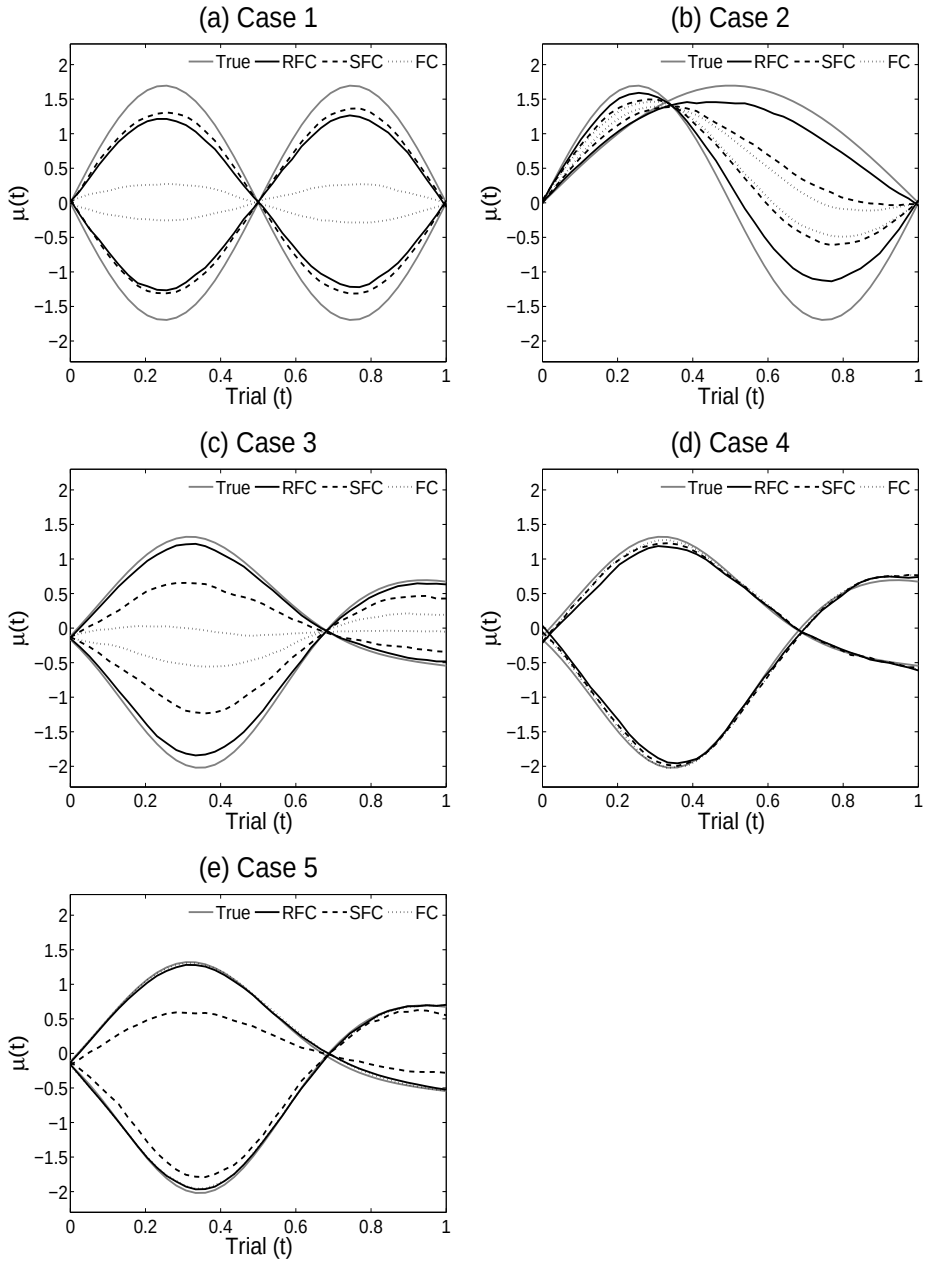

Fig. 1: The estimated cluster mean functions for the RFC, SFC and FC algorithms over 200 Monte Carlo runs overlaying the true cluster mean functions.