Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

Multilevel sparse functional principal component analysis

Chongzhi Di[a,*], Ciprian M. Crainiceanu[b] and Wolfgang S. Jank[c]

Received 3 March 2014; Accepted 24 March 2014

We consider analysis of sparsely sampled multilevel functional data, where the basic observational unit is a function and data have a natural hierarchy of basic units. An example is when functions are recorded at multiple visits for each subject. Multilevel functional principal component analysis was proposed recently for such data when functions are densely recorded. Here, we consider the case when functions are sparsely sampled and may contain only a few observations per function. We exploit the multilevel structure of covariance operators and achieve data reduction by principal component decompositions at both between-subject and within-subject levels. We address inherent methodological differences in the sparse sampling context to: (i) estimate the covariance operators; (ii) estimate the functional principal component scores; and (iii) predict the underlying curves. Through simulations, the proposed method is able to discover dominating modes of variations and reconstruct underlying curves well even in sparse settings. Our approach is illustrated by two applications, the Sleep Heart Health Study and eBay auctions.

Keywords: functional principal component analysis; multilevel models; smoothing

1 Introduction

Motivated by modern scientific studies, functional data analysis (FDA; Ramsay & Silverman, 2005) is becoming a popular and important area of research. The basic unit of functional data is a function, curve, or image, typically high dimensional, and the aim of FDA is to understand variations in a sample of functions. Functional principal component analysis (FPCA) plays a central role in FDA, and thus, we briefly review the most recent developments in this area. The fundamental aims of FPCA are to capture the principal directions of variation and to reduce dimensionality. Besides discussion by Ramsay & Silverman (2005), other relevant research in FPCA includes that by Ramsay & Dalzell (1991), Silverman (1996), James et al. (2000), and Yao et al. (2005), while important theoretical results can be found in the work of Hall & Hosseini-Nasab (2006).

The standard FPCA methodology was designed for a sample of *densely* recorded *independent* functions. Its scope was extended in two directions later. First, it was extended to *sparse* functional/longitudinal data (Yao et al., 2005; Müller, 2005). Sparsity is the characteristic of the sampling scheme that leads to a small number of observations per function and a dense collection of sampling points across all curves. Traditionally, these data were treated as

[a]**Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M2-B500, Seattle, WA 98115, USA**

[b]**Department of Biostatistics, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205, USA**

[c]**Department of Information Systems and Decision Sciences, University of South Florida, Tampa, FL 33620, USA**

*__Email: cdi@fhcrc.org__

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

longitudinal data (Diggle et al., 2002) and analyzed using parametric or semi-parametric models. The functional approach views them as sparse and noisy realizations of an underlying smooth process and aims to study the modes of variations of the process. Second, it was extended to functional data with a *hierarchical* structure, which led to multilevel FPCA (MFPCA; Di et al., 2009). In this paper, we propose methods for a sample of *sparsely* recorded functions with *hierarchical* structure and discuss issues arising from both sparsity and multilevel structure.

Our research was motivated by two applications, the Sleep Heart Health Study (SHHS) and online auctions from `eBay.com`. The SHHS is a multi-center cohort study of sleep and its impacts on health outcomes. A detailed description of the SHHS can be found in the work of Quan et al. (1997) and Crainiceanu et al. (2009a). Between 1995 and 1997, 6441 participants were recruited and underwent in-home polysomnograms (PSGs). A PSG is a quasi-continuous multi-channel recording of physiological signals acquired during sleep that include two surface electroencephalograms (EEG). Between 1999 and 2003, a second SHHS follow-up visit was undertaken on 3201 participants (47.8% of baseline cohort) and included a repeat PSG as well as other measurements on sleep habits and other health status variables. The sleep EEG percent δ-power series of the SHHS data is one of the main outcome variables. The data include a function of time in 30-second intervals per subject per visit, with each function having approximately 960 data points in an 8-hour interval of sleep. For this analysis, we conducted the analysis of sparsified data where each function is sub-sampled at a random set of time points, and compared this to full analysis of the dense data.

Our second application originates from online auctions, which are challenging because they involve *user-generated* data: Sellers decide when to post an auction, and bidders decide when to place bids. This can result in individual auctions that have extremely sparse observations, especially during the early parts of the auction. In fact, well-documented bidding strategies such as early bidding or last-minute bidding cause "bidding-draughts" (Bapna et al., 2004; Shmueli et al., 2007; Jank & Shmueli, 2009) during the middle, leaving the auction with barely any observations at all. Peng & Müller (2008), Liu & Müller (2009), Jank et al. (2010), and Reithinger et al. (2008) studied analysis of dynamics of such sparse auction data. Here, we study bidding records of 843 *digital camera auctions* that were listed on eBay between April 2007 and January 2008. These auctions were on 515 types of digital cameras, from 233 distinct sellers. On average, there were 11 bids per auction. The timing of the bids was irregular and often sparse: Some auctions contained as many as 56 bids, while others included as few as one to two bids. In this application, we are particularly interested in investigating the pattern of variation of an auction's bidding path and decomposing it into components that are attributable to the product and components attributable to the bidding process.

The remainder of this paper is organized as follows. Section 2 introduces MFPCA for sparse data. Section 3 provides mathematical details for predicting the principal component scores and curves. Section 4 describes extensive simulation studies for realistic settings. Section 5 describes applications of our methodology to the SHHS data and eBay auction data. Section 6 includes some discussion.

2 MFPCA for sparsely sampled functions

In this section, we briefly review the MFPCA technique proposed by Di et al. (2009) and then discuss and address methodological issues arising from sparsity.

2.1. MFPCA: review

The MFPCA was designed to capture dominant modes of variations and reduce dimensions for multilevel functional data. This method decomposes the total functional variation into between-subject and within-subject variations via functional analysis of variance (FANOVA) and conducts principal component analysis at both levels. More precisely,

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

The ISI's Journal for the Rapid Dissemination of Statistics Research

letting $Y_{ij}(t)$ denote the observed function for subject i at visit j, the two-way FANOVA decomposes the total variation as

$$Y_{ij}(t) = \mu(t) + \eta_j(t) + Z_i(t) + W_{ij}(t) + \epsilon_{ij}(t), \tag{1}$$

where $\mu(t)$ and $\eta_j(t)$ are fixed functional effects that represent the overall mean function and visit-specific shifts, respectively; $Z_i(t)$ and $W_{ij}(t)$ are the subject-specific and visit-specific deviations, respectively; and $\epsilon_{ij}(t)$ is measurement error process with mean 0 and variance σ^2. The level 1 and 2 processes, $Z_i(t)$ and $W_{ij}(t)$, are assumed to be centered stochastic processes that are uncorrelated with each other. The idea of MFPCA is to decompose both $Z_i(t)$ and $W_{ij}(t)$ using the Karhunen–Loève (KL) expansions (Karhunen, 1947; Loève, 1945; Ramsay & Silverman, 2005), i.e., $Z_i(t) = \sum_{k=1}^{N_1} \xi_{ik}\phi_k^{(1)}(t)$, $W_{ij}(t) = \sum_{l=1}^{N_2} \zeta_{ijl}\phi_l^{(2)}(t)$, where $\phi_k^{(1)}(t)$ and $\phi_l^{(2)}(t)$ are level 1 and level 2 eigenfunctions, respectively, and ξ_{ik} and ζ_{ijl} are mean zero random variables called principal component scores. The variances of ξ_{ik} and ζ_{ijl}, $\lambda_k^{(1)}$ and $\lambda_l^{(2)}$, respectively, are the level 1 and 2 eigenvalues that characterize the magnitude of variation in the direction of the corresponding eigenfunctions. The number of principal components at levels 1 and 2, N_1 and N_2, could be either finite integers or ∞. Combining model (1) with the KL expansions, one obtains the MFPCA model

$$Y_{ij}(t) = \mu(t) + \eta_j(t) + \sum_{k=1}^{N_1} \xi_{ik}\phi_k^{(1)}(t) + \sum_{l=1}^{N_2} \zeta_{ijl}\phi_l^{(2)}(t) + \epsilon_{ij}(t). \tag{2}$$

The MFPCA reduces high-dimensional hierarchical functional data $\{Y_{i1}(t), \ldots, Y_{in_i}(t)\}$ into the low dimensional principal component score vectors, including subject-level (level 1) scores $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iN_1})$ and subject/visit-level (level 2) scores $\boldsymbol{\zeta}_{ij} = (\zeta_{ij1}, \cdots, \zeta_{ijN_2})$, while retaining most information contained in the data. The detailed interpretation of this model is discussed by Di et al. (2009).

Equation (1) introduces FANOVA in full generality without details about the sampling design for t_{ijs}. The set of sampling points $\{t_{ijs} : s = 1, 2, \ldots, T_{ij}\}$ for subject i at visit j can be dense or sparse, regular or irregular, depending on the application. Although Di et al. (2009) discussed potential difficulties with sparse designs, they focused on densely and regularly recorded functional data. In the following, we will discuss and address new problems raised by the sparse sampling design. This will lead to methods that are related to but markedly different from MFPCA.

2.2. Estimating eigenvalues and eigenfunctions

Throughout the paper, we assume sparse and irregular grid points. More precisely, for each subject and visit, the number of grid points T_{ij} is relatively small, and the set of grid points $\{t_{ijs} : s = 1, 2, \ldots, T_{ij}\}$ is a random sample of $\mathcal{T}$. We also assume that the sets of grid points are different across subjects and visits and that the collection of grid points over all subjects and visits is a dense subset of $\mathcal{T}$.

The first step of MFPCA is to estimate the eigenvalues and eigenfunctions. This can be carried out by the method of moments and eigen-analysis for dense functional data, but smoothing is needed for sparse functional data. Let $K_B(s,t) = \text{cov}\{Z_i(s), Z_i(t)\}$ be the covariance function for level 1 processes ("between" covariance), $K_W(s,t) = \text{cov}\{W_{ij}(s), W_{ij}(t)\}$ be the covariance function for level 2 processes ("within"). The total covariance function $K_T(s,t)$ contains three sources of variation, that is, $K_T(s,t) = K_B(s,t) + K_W(s,t) + \sigma^2 I(t = s)$. One can easily verify that $\text{E}\{Y_{ij}(t)\} = \mu(t) + \eta_j(t)$, $\text{cov}\{Y_{ij}(s), Y_{ij}(t)\} = K_B(s,t) + K_W(s,t) + \sigma^2 I(t = s)$, and $\text{cov}\{Y_{ij}(s), Y_{ik}(t)\} = K_B(s,t)$. These results suggest the following convenient algorithm to estimate the eigenvalues and eigenfunctions. Because functions were sparsely sampled over irregular grid points, scatter plot smoothing will be used repeatedly to estimate the underlying mean functions and covariance kernels.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

Sparse MFPCA algorithm

Step 1. Apply scatter plot smoothing using pairs $\{(t_{ijs}, Y_{ij}(t_{ijs})) : i = 1,\ldots,n; j = 1,\ldots,n_i; s = 1,\ldots,T_{ij}\}$ to obtain an estimate of $\mu(t)$, $\hat{\mu}(t)$ and apply scatter plot smoothing using pairs $\{(t_{ijs}, Y_{ij}(t_{ijs}) - \hat{\mu}(t_{ijs})) : i = 1,\ldots,n; s = 1,\ldots,T_{ij}\}$ to obtain an estimate of $\eta_j(t)$, $\hat{\eta}_j(t)$.

Step 2. Estimate $\hat{K}_B(s,t)$ by bivariate smoothing of all products $\{Y_{ij_1}(t_{ij_1 s}) - \hat{\mu}(t_{ij_1 s}) - \hat{\eta}_{j_1}(t_{ij_1 s})\}\{Y_{ij_2}(t_{ij_2 r}) - \hat{\mu}(t_{ij_2 r}) - \hat{\eta}_{j_2}(t_{ij_2 r})\}$ with respect to $(t_{ij_1 s}, t_{ij_2 r})$ for all i, j_1, j_2, r and s with $j_1 \neq j_2$.

Step 3. Estimate $\hat{K}_T(s,t)$ by bivariate smoothing of all products $\{Y_{ij}(t_{ijs}) - \hat{\mu}(t_{ijs}) - \hat{\eta}_j(t_{ijs})\}\{Y_{ij}(t_{ijr}) - \hat{\mu}(t_{ijr}) - \hat{\eta}_j(t_{ij_2 r})\}$ with respect to (t_{ijs}, t_{ijr}) for all i, j, r, s with $r \neq s$, and set $\hat{K}_W(s,t) = \hat{K}_T(s,t) - \hat{K}_B(s,t)$.

Step 4. Use eigen-analysis on $\hat{K}_B(s,t)$ to obtain $\hat{\lambda}_k^{(1)}$, $\hat{\phi}_k^{(1)}(t)$; use eigen-analysis on $\hat{K}_W(s,t)$ to obtain $\hat{\lambda}_l^{(2)}$, $\hat{\phi}_l^{(2)}(t)$.

Step 5. Estimate the nugget variance σ^2 by smoothing $\{Y_{ij}(t_{ijs}) - \hat{\mu}(t_{ijs}) - \hat{\eta}_j(t_{ijs})\}^2 - \hat{K}_T(t_{ijs}, t_{ijs})$ with respect to t_{ijs} for all possible i, j, s.

For univariate and bivariate smoothing, we use penalized spline smoothing (Ruppert et al., 2003) with the smoothing parameter estimated via restricted maximum likelihood. This technique has been implemented in statistical packages, such as the "SemiPar" package in R. Cross-validation (CV) or generalized CV (GCV) could also be used. Alternatively, one can also use local polynomial smoothing (Fan & Gijbels, 1996) with CV to choose the smoothing parameter. In step 3, the diagonal elements (when $s = r$) are dropped when estimating the total covariance $K_T(s,t)$, because they are contaminated by measurement error. In contrast, diagonal elements are included when estimating the between covariance $K_B(s,t)$.

In step 4, one needs to determine the dimensions of level 1 and 2 spaces, namely, N_1 and N_2, respectively. Although they are allowed to be ∞ in theory, in practice, a low-dimensional principal component space suffices to approximate the functional space. We will discuss this issue in more details later.

2.3. Principal component scores

Once the fixed functional effects $\mu(t)$, $\eta_j(t)$, the eigenvalues $\lambda_k^{(1)}$, $\lambda_l^{(2)}$, and the eigenfunctions $\phi_k^{(1)}(t)$, $\phi_l^{(2)}(t)$ are estimated, the MFPCA model can be re-written as a linear mixed model

$$\begin{cases} Y_{ijs} = \mu(t_{ijs}) + \eta_j(t_{ijs}) + \sum_{k=1}^{N_1} \xi_{ik}\phi_k^{(1)}(t_{ijs}) + \sum_{l=1}^{N_2} \zeta_{ijl}\phi_l^{(2)}(t_{ijs}) + \epsilon_{ijs} \\ \xi_{ik} \sim N\left\{0, \lambda_k^{(1)}\right\},\ \zeta_{ijl} \sim N\left\{0, \lambda_l^{(2)}\right\},\ \epsilon_{ijs} \sim N(0, \sigma^2), \end{cases} \tag{3}$$

where $Y_{ijs} := Y_{ij}(t_{ijs})$ and $\epsilon_{ijs} = \epsilon_{ij}(t_{ijs})$. The random effects ξ_{ik} and ζ_{ijl} are principal component scores that we are trying to estimate. Thus, one could use the mixed model inferential machinery to estimate the scores, for example, using the best linear unbiased prediction (BLUP). The BLUP gives point estimates of the scores, and one could also construct their 95% confidence intervals.

Note that the subject-specific effect, $Z_i(t)$, and the visit-specific effect, $W_{ij}(t)$, are linear functions of the random effects ξ_{ik} and ζ_{ijl}, respectively. Thus, $Z_i(t)$, $W_{ij}(t)$, and their variability can be estimated directly from the BLUP formulas for the random effects. The BLUPs of $Z_i(t)$ and $W_{ij}(t)$ are shrinkage estimators, which automatically combine information from different visits of the same subject and across subjects. More information is borrowed when the measurement error variance, σ, is large and when the number of observations at the subject level is small.

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

The ISI's Journal for the Rapid Dissemination of Statistics Research

2.4. Choosing the dimensions N_1 and N_2

Two popular methods for estimating the dimension of the functional space in the single level case are CV (Rice & Silverman, 1991) and Akaike's information criterion (as in Yao et al., 2005). These methods can be generalized to the multilevel setting. For example, one could use the leave-one-subject-out CV criterion to select the number of dimensions. Define the CV score as

$$\mathrm{CV}(N_1, N_2) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{s=1}^{T_{ij}} \left\{ Y_{ij}(t_{ijs}) - \hat{Y}_{ij}^{(-i),N_1,N_2}(t_{ijs}) \right\}^2,$$

where $\hat{Y}_{ij}^{(-i),N_1,N_2}(t_{ijs})$ is the predicted curve for subject i at visit j, computed after removing the data from subject i. The estimated number of dimensions N_1 and N_2 are the arguments that minimize $CV(N_1, N_2)$. In practice, the leave-one-subject-out CV method may be too computationally intensive, and an m-fold CV can serve as a fast alternative. This method divides the subjects into m groups, and the prediction error for the m^{th} group is calculated by fitting a model using the data from other groups. The number of dimensions is chosen as those that minimize the total prediction error. Similar criteria could be designed for leave-visits-out.

One could also use a fast method. More precisely, let P_1 and P_2 be two thresholds, and define

$$N_1 = \min\left\{k : \rho_k^{(1)} \geq P_1, \lambda_k^{(1)} < P_2\right\}, \; N_2 = \min\left\{k : \rho_k^{(2)} \geq P_1, \lambda_k^{(2)} < P_2\right\},$$

where $\rho_k^{(j)} = \left(\lambda_1^{(j)} + \cdots + \lambda_k^{(j)}\right) / \left(\lambda_1^{(j)} + \cdots + \lambda_k^{(j)} + \cdots\right), j = 1, 2$, is the proportion of variation explained by the first k principal components at level j. Intuitively, this method chooses the number of dimension at each level to be the smallest integer k such that the first k components explain more than P_1 of the total variation, while any component after the k^{th} explains less than P_2 of the variation. To use this method, the thresholds P_1 and P_2 need to be carefully tuned using simulation studies or CVs. In practice, we found that $P_1 = 90\%$ and $P_2 = 5\%$ often work well.

2.5. Iterative procedure to improve accuracy

Initial estimates of the mean functions and eigenfunctions via smoothing are typically accurate in the dense functional data, and less accurate for the sparse data. To improve estimation accuracy, one may adopt an iterative procedure, which iterates between updating fixed effects (mean, eigenfunctions, and eigenvalues) and principal component scores.

Iterative sparse MFPCA algorithm

Step 1. Obtain initial estimates, $\hat{\mu}^0(t)$, $\hat{\eta}_j^0(t)$, $\hat{\lambda}_k^{(1),0}$, $\hat{\lambda}_l^{(2),0}$, $\hat{\phi}_k^{(1),0}(t)$, $\hat{\phi}_l^{(2),0}(t)$, and $\hat{\sigma}^{2,0}$ for all j, k, l, using the sparse MFPCA algorithm described in Section 2.2; estimate principal component scores, $\hat{\boldsymbol{\xi}}_i^0$ and $\hat{\boldsymbol{\zeta}}_i^0$, using formulas that will be described in Section 3.

Step 2. Apply steps 1 and 2 of the sparse MFPCA algorithm on $Y_{ij}(t_{ijs}) - \sum_{k=1}^{N_1} \hat{\xi}_{ik}^0 \hat{\phi}_k^{(1),0}(t_{ijs}) - \sum_{l=1}^{N_2} \hat{\zeta}_{ijl}^0 \hat{\phi}_l^{(2),0}(t_{ijs})$ and obtain updated mean functions $\hat{\mu}^1(t)$ and $\hat{\eta}_j^1(t)$.

Step 3. Apply steps 3–6 of the sparse MFPCA algorithm on $Y_{ij}(t_{ijs}) - \hat{\mu}^1(t_{ijs}) - \hat{\eta}_j^1(t_{ijs})$, and obtain updated eigenvalues $\hat{\lambda}_k^{(1),1}$, $\hat{\lambda}_l^{(2),1}$, eigenfunctions $\hat{\phi}_k^{(1),1}(t)$, $\hat{\phi}_l^{(2),1}(t)$, and $\hat{\sigma}^{2,1}$ for all j, k, l.

Step 4. Update principal component scores, $\hat{\boldsymbol{\xi}}_i^1$ and $\hat{\boldsymbol{\zeta}}_i^1$, based on the new estimates from steps 2 and 3.

Step 5. Stop if certain criteria are met. Otherwise, set $\hat{\mu}^1(t)$, $\hat{\eta}_j^1(t)$, $\hat{\lambda}_k^{(1),1}$, $\hat{\lambda}_l^{(2),1}$, $\hat{\phi}_k^{(1),1}(t)$, $\hat{\phi}_l^{(2),1}(t)$, $\hat{\boldsymbol{\xi}}_i^1$, and $\hat{\boldsymbol{\zeta}}_i^1$ as initial estimates and repeat steps 2–5 until the algorithm converges.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

One reasonable set of the stopping criteria could be $\|\hat{\mu}^1(t)-\hat{\mu}^0(t)\| < \varepsilon_1$, $\|\hat{\eta}_j^1(t)-\hat{\eta}_j^0(t)\| < \varepsilon_1$, $\|\hat{\phi}_k^{(1),1}(t)-\hat{\phi}_k^{(1),0}(t)\| < \varepsilon_1$, $\|\hat{\phi}_l^{(2),1}(t)-\hat{\phi}_l^{(2),0}(t)\| < \varepsilon_1$, $|\hat{\lambda}_k^{(1),1}-\hat{\lambda}_k^{(1),0}| < \varepsilon_2$, and $|\hat{\lambda}_l^{(2),1}-\hat{\lambda}_l^{(2),0}| < \varepsilon_2$ for all j, k, and l, where $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ are small prespecified thresholds.

3 Prediction of principal component scores and curves

This section provides BLUP calculation results for principal component scores and function prediction at various levels. Some heavy notation is unavoidable, but results are crucial for the implementation of our quick algorithms.

3.1. Prediction of principal component scores

We introduce some notations before presenting the formulas for the principal component scores. Let $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{iN_1})^T$ be an $N_1 \times 1$ vector, $\boldsymbol{\zeta}_{ij} = (\zeta_{ij1}, \zeta_{ij2}, \ldots, \zeta_{ijN_2})^T$ be an $N_2 \times 1$ vector, $\boldsymbol{\zeta}_i = \left(\boldsymbol{\zeta}_{i1}^T, \boldsymbol{\zeta}_{i2}^T, \ldots, \boldsymbol{\zeta}_{in_i}^T\right)^T$ be an $(N_2 n_i) \times 1$ vector, $\boldsymbol{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ijT_{ij}})$ be a $T_{ij} \times 1$ vector, $\boldsymbol{t}_{ij} = (t_{ij1}, \ldots, t_{ijT_{ij}})$ be a $T_{ij} \times 1$ vector, and $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}^T, \ldots, \boldsymbol{Y}_{in_i}^T)^T$ be a $(\sum_j T_{ij}) \times 1$ vector. Calculations will be carried out conditionally on $\mu(t)$, $\eta_j(t)$, $\lambda_k^{(1)}$, $\lambda_l^{(2)}$, $\phi_k^{(1)}(t)$, and $\phi_l^{(2)}(t)$, which can be estimated using methods described in the previous sections. It is straightforward to evaluate these functions at observed grid points, i.e., $\boldsymbol{\mu}_{ij} = \{\mu(t_{ij1}), \ldots, \mu(t_{ijT_{ij}})\}^T$, $\boldsymbol{\mu}_i = \left(\boldsymbol{\mu}_{ij}^T, \ldots, \boldsymbol{\mu}_{in_i}^T\right)^T$, $\boldsymbol{\eta}_{ij} = \{\eta_j(t_{ij1}), \ldots, \eta_j(t_{ijT_{ij}})\}^T$, $\boldsymbol{\eta}_i = \left(\boldsymbol{\eta}_{ij}^T, \ldots, \boldsymbol{\eta}_{in_i}^T\right)^T$, $\boldsymbol{\phi}_{k,ij}^{(1)} = \left\{\phi_k^{(1)}(t_{ij1}), \ldots, \phi_k^{(1)}(t_{ijT_{ij}})\right\}^T$, and $\boldsymbol{\phi}_{l,ij}^{(2)} = \left\{\phi_l^{(2)}(t_{ij1}), \ldots, \phi_l^{(2)}(t_{ijT_{ij}})\right\}^T$. Let $\boldsymbol{\Phi}_{ij}^{(1)}$ denote a $T_{ij} \times N_1$ matrix whose k^{th} column is given by $\boldsymbol{\phi}_{k,ij}^{(1)}$, $\boldsymbol{\Phi}_{ij}^{(2)}$ denote a $T_{ij} \times N_2$ matrix whose l^{th} column is given by $\boldsymbol{\phi}_{l,ij}^{(2)}$, $\boldsymbol{\Lambda}^{(1)}$ denote an $N_1 \times N_1$ diagonal matrix with diagonal elements $\left(\lambda_1^{(1)}, \ldots, \lambda_{N_1}^{(1)}\right)$, and $\boldsymbol{\Lambda}^{(2)}$ denote an $N_2 \times N_2$ diagonal matrix with diagonal elements $\left(\lambda_1^{(2)}, \ldots, \lambda_{N_2}^{(2)}\right)$. The following proposition gives the point estimates and variance for the principal component scores.

Proposition 1

Under the MFPCA model (3), the BLUP for principal component scores $\left(\boldsymbol{\xi}_i^T, \boldsymbol{\zeta}_i^T\right)$ *has the following form:*

$$\begin{pmatrix} \hat{\boldsymbol{\xi}}_i \\ \hat{\boldsymbol{\zeta}}_i \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_i \\ \boldsymbol{B}_i \end{pmatrix} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i - \boldsymbol{\eta}_i), \tag{4}$$

and their covariance matrix, $\text{cov}\{(\hat{\boldsymbol{\xi}}_i^T - \boldsymbol{\xi}_i^T, \hat{\boldsymbol{\zeta}}_i^T - \boldsymbol{\zeta}_i^T) \mid \boldsymbol{Y}_i\}$*, is given by*

$$\begin{pmatrix} \boldsymbol{\Lambda}^{(1)} & 0 \\ 0 & \boldsymbol{\Lambda}^{(2)} \otimes \boldsymbol{I}_{n_i \times n_i} \end{pmatrix} - \begin{pmatrix} \boldsymbol{A}_i \\ \boldsymbol{B}_i \end{pmatrix} \boldsymbol{\Sigma}_i^{-1} \left(\boldsymbol{A}_i^T, \boldsymbol{B}_i^T\right), \tag{5}$$

where $\otimes$ *denotes the Kronecker product,* $\boldsymbol{A}_i := \text{cov}(\boldsymbol{\xi}_i, \boldsymbol{Y}_i)$ *is an* $N_1 \times \left(\sum_j T_{ij}\right)$ *matrix,* $\boldsymbol{B}_i := \text{cov}(\boldsymbol{\zeta}_i, \boldsymbol{Y}_i)$ *is an* $(N_2 n_i) \times \left(\sum_j T_{ij}\right)$ *matrix, and* $\boldsymbol{\Sigma}_i := \text{cov}(\boldsymbol{Y}_i)$ *is a* $\left(\sum_j T_{ij}\right) \times \left(\sum_j T_{ij}\right)$ *matrix. The matrix* $\boldsymbol{A}_i$ *has the form* $\boldsymbol{A}_i = \left(\boldsymbol{\Lambda}^{(1)}\boldsymbol{\Phi}_{i1}^{(1)^T}, \boldsymbol{\Lambda}^{(1)}\boldsymbol{\Phi}_{i2}^{(1)^T}, \ldots, \boldsymbol{\Lambda}^{(1)}\boldsymbol{\Phi}_{in_i}^{(1)^T}\right)$*, and* $\boldsymbol{B}_i$ *is a block diagonal matrix with diagonal elements* $\left\{\boldsymbol{\Lambda}^{(2)}\boldsymbol{\Phi}_{i1}^{(2)^T}, \boldsymbol{\Lambda}^{(2)}\boldsymbol{\Phi}_{i2}^{(2)^T}, \ldots, \boldsymbol{\Lambda}^{(2)}\boldsymbol{\Phi}_{in_i}^{(2)^T}\right\}$*. Let* $\boldsymbol{\Sigma}_{i,jk} := \text{cov}(\boldsymbol{Y}_{ij}, \boldsymbol{Y}_{ik})$ *be the* (j,k) *block of* $\boldsymbol{\Sigma}_i$ *with size* $T_{ij} \times T_{ik}$*. When* $j = k$*,*

$$\boldsymbol{\Sigma}_{i,jj} = \boldsymbol{\Phi}_{ij}^{(1)}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\Phi}_{ij}^{(1)^T} + \boldsymbol{\Phi}_{ij}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\Phi}_{ij}^{(2)^T} + \sigma^2 \boldsymbol{I}_{T_{ij} \times T_{ij}},$$

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

The ISI's Journal for the Rapid Dissemination of Statistics Research

where $\boldsymbol{I}_{T_{ij}\times T_{ij}}$ *is an identity matrix, and when* $j \neq k$, $\boldsymbol{\Sigma}_{i,jk} = \boldsymbol{\Phi}_{ij}^{(1)} \boldsymbol{\Lambda}^{(1)} \boldsymbol{\Phi}_{ik}^{(1)^T}$.

Equation (4) provides the best prediction of the principal component scores under the Gaussian assumptions, and the best linear prediction otherwise. Thus, the Gaussian assumptions in model (3) can be relaxed. Crainiceanu et al. (2009b) also provided formulae for the BLUPs of the principal component scores. Their results are applicable to balanced and dense designs only, i.e., to cases when each function is measured at exactly the same set of grid points; the formulae in Proposition 1 are applicable for both balanced and unbalanced designs. When data are balanced and dense, the results of Crainiceanu et al. (2009b) are preferable because they avoid inverting large matrices. Otherwise, one should use results in Proposition 1.

Once estimates of principal component scores are obtained, they can be used in further analysis as either outcome or predictor variables. For example, Di et al. (2009) explored the distribution of subject-specific principal component scores in different sex and age groups. Di et al. (2009) and Crainiceanu et al. (2009b) considered generalized multilevel functional regression, which modeled principal component scores as predictors for health outcomes, such as hypertension. These analyses can be extended to the sparse case.

3.2. Prediction of functional effects

Our approach for sparsely recorded functions allows estimation of the covariance structure of the population of functions at various levels; it also provides a simple algorithm for predicting subject-level or subject/visit-level curves. This process intrinsically borrows information both across subjects and across visits within subjects and yields surprisingly accurate predictions; see our simulation results in Section 4 and data analysis in Section 5 for demonstration. The following proposition provides the formulas for prediction of various functions and their confidence intervals. Results are derived based on the MFPCA model (3) and Proposition 1.

Proposition 2

Let $\boldsymbol{\Phi}^{(1)}(t) := \left\{\phi_1^{(1)}(t), \phi_2^{(1)}(t), \cdots, \phi_{N_1}^{(1)}(t)\right\}^T$ *and* $\boldsymbol{\Phi}^{(2)}(t) := \left\{\phi_1^{(2)}(t), \phi_2^{(2)}(t), \ldots, \phi_{N_2}^{(2)}(t)\right\}^T$. *Under the MFPCA model (3), the BLUP for the subject-specific curve,* $Z_i(t)$, *is given by*

$$\hat{Z}_i(t) = \boldsymbol{\Phi}^{(1)}(t)^T \hat{\boldsymbol{\xi}}_i = \boldsymbol{\Phi}^{(1)}(t)^T \boldsymbol{A}_i \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i - \boldsymbol{\eta}_i),$$

with variance, $\text{var}\left\{\hat{Z}_i(t) - Z_i(t) \mid \boldsymbol{Y}_i\right\} = \boldsymbol{\Phi}^{(1)}(t)^T \left(\boldsymbol{\Lambda}^{(1)} - \boldsymbol{A}_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{A}_i^T\right) \boldsymbol{\Phi}^{(1)}(t)$. *The BLUP for the visit-specific curve,* $W_{ij}(t)$, *is given by*

$$\hat{W}_{ij}(t) = \boldsymbol{\Phi}^{(2)}(t)^T \hat{\boldsymbol{\zeta}}_{ij} = \boldsymbol{\Phi}^{(2)}(t)^T \boldsymbol{H}_j \boldsymbol{B}_i \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i - \boldsymbol{\eta}_i),$$

with variance, $\text{var}\left\{\hat{W}_{ij}(t) - W_{ij}(t) \mid \boldsymbol{Y}_i\right\} = \boldsymbol{\Phi}^{(2)}(t)^T \left(\boldsymbol{\Lambda}^{(2)} \bigotimes \boldsymbol{I}_{n_i\times n_i} - \boldsymbol{H}_j \boldsymbol{B}_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{A}_i^T \boldsymbol{H}_j^T\right) \boldsymbol{\Phi}^{(2)}(t)$, *where the matrix* $\boldsymbol{H}_j = (\boldsymbol{H}_{j,1}, \ldots, \boldsymbol{H}_{j,n_j})$ *is* $N_2 \times (N_2 n_i)$ *with* $\boldsymbol{H}_{j,k} = \boldsymbol{I}_{N_2\times N_2}$ *if* $j = k$ *and* $\boldsymbol{H}_{j,k} = \boldsymbol{0}$ *otherwise. The individual curve* $Y_{ij}(t)$ *can be predicted by* $\hat{Y}_{ij}(t) = \hat{\mu}(t) + \hat{\eta}_j(t) + \hat{Z}_i(t) + \hat{W}_{ij}(t)$, *with variance,* $\text{var}\{\hat{Y}_{ij}(t) - Y_{ij}(t) \mid \boldsymbol{Y}_i\}$, *estimated by*

$$\boldsymbol{\Phi}(t)^T \boldsymbol{G}_j \left\{ \begin{pmatrix} \boldsymbol{\Lambda}^{(1)} & 0 \\ 0 & \boldsymbol{\Lambda}^{(2)} \bigotimes \boldsymbol{I}_{n_i\times n_i} \end{pmatrix} - \begin{pmatrix} \boldsymbol{A}_i \\ \boldsymbol{B}_i \end{pmatrix} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{A}_i^T, \boldsymbol{B}_i^T) \right\} \boldsymbol{G}_j^T \boldsymbol{\Phi}(t),$$

where $\boldsymbol{G}_j = (\boldsymbol{I}_{N_1\times N_1}, \boldsymbol{H}_j)$ *and* $\boldsymbol{\Phi}(t) = (\boldsymbol{\Phi}^{(1)}(t)^T, \boldsymbol{\Phi}^{(2)}(t)^T)^T$.

Proofs of Proposition 1 and 2 can be obtained by BLUP calculations (see, e.g., Ruppert et al., 2003) for general linear mixed models and are omitted. This is a two-step procedure, first estimating in the fixed functional effects

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

$\mu(t)$ and $\eta_j(t)$, eigenvalues $\lambda_k^{(1)}$ and $\lambda_l^{(2)}$, eigenfunctions $\phi_k^{(1)}(t)$ and $\phi_k^{(2)}(t)$, and variance σ^2 and then plugging in these estimators in the BLUP estimation. Note that similar to those of Yao et al. (2005), the variance formulas in Proposition 1 and 2 do not account for the uncertainty associated to estimating the quantities in the first step. Goldsmith et al. (2013) proposed bootstrap-based procedures to obtain corrected confidence bands for functional data using principal components, and their approach can be adopted to estimate the total variation.

4 Simulations

To evaluate finite sample performance, we conducted simulation studies under a variety of settings. The data were generated from a true model used by Di et al. (2009), except that the curves are sampled on a sparse set of grid points. More precisely, the true model is the MFPCA model with with $\xi_{ik} \sim N\left(0, \lambda_k^{(1)}\right)$, $\zeta_{ijl} \sim N\left(0, \lambda_l^{(2)}\right)$, $\epsilon_{ij}(t_{ijm}) \sim N(0, \sigma^2)$, and $(t_{ijm} : m = 1, 2, \ldots, T_{ij})$ is a set of grid points in the interval $[0, 1]$. The set of grid points is generated uniformly in the interval $[0, 1]$ and is different across subjects and visits. For simplicity, we assume that the number of visits is the same across subjects and that the number of grid points per curve is the same across subjects and visits, i.e., $n_i = J$ and $T_{ij} = N$ for $i = 1, \ldots, n; j = 1, \ldots, n_i$, although our methods apply to the case of varying n_i and T_{ij} easily. The true mean function is $\mu(t) = 8t(1-t)$. The true eigenvalues are $\lambda_k^{(1)} = 0.5^{k-1}$ and $\lambda_l^{(2)} = 0.5^{l-1}$, $k, l = 1, 2, 3, 4$, with eigenfunctions $\phi_k^{(1)}(t) = \left\{\sqrt{2}\sin(2\pi t), \sqrt{2}\cos(2\pi t), \sqrt{2}\sin(4\pi t), \sqrt{2}\cos(4\pi t)\right\}$ at level 1 and $\phi_1^{(2)}(t) = 1$, $\phi_2^{(2)}(t) = \sqrt{3}(2t-1)$, $\phi_3^{(2)}(t) = \sqrt{5}(6t^2-6t+1)$, $\phi_4^{(2)}(t) = \sqrt{7}(20t^3-30t^2+12t-1)$ at level 2.

Table I. Root (integrated) mean square errors for eigenvalues and eigenfunctions in simulations.

		Eigenvalues				Eigenfunctions			
n	N	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
				Level 1					
100	3	0.25	0.39	0.69	1.16	0.45	0.66	1.03	1.07
100	6	0.29	0.36	0.76	1.26	0.56	0.81	1.00	1.21
100	9	0.19	0.25	0.35	0.48	0.38	0.54	0.83	0.98
100	12	0.21	0.26	0.36	0.54	0.42	0.66	0.85	1.08
200	3	0.18	0.22	0.26	0.36	0.34	0.48	0.73	0.92
200	6	0.19	0.23	0.30	0.41	0.35	0.56	0.76	0.97
300	3	0.17	0.20	0.23	0.31	0.32	0.46	0.66	0.87
				Level 2					
100	3	0.14	0.18	0.28	0.36	0.25	0.37	0.67	0.90
100	6	0.15	0.21	0.30	0.42	0.31	0.51	0.71	0.95
100	9	0.15	0.23	0.45	0.64	0.27	0.39	0.81	0.98
100	12	0.17	0.25	0.37	0.64	0.36	0.62	0.83	1.06
200	3	0.12	0.16	0.39	0.50	0.21	0.30	0.67	0.90
200	6	0.14	0.22	0.32	0.51	0.30	0.53	0.74	0.97
300	3	0.09	0.10	0.16	0.20	0.15	0.21	0.33	0.51

Simulation settings vary according to sample size ($n = 100, 200, 300$) and the number of grid points per function ($N = 3, 6, 9, 12$). In simulations, the first four principal components (PC) at both levels were compared to their underlying true counterparts. Root mean square errors and root integrated mean square errors are used to measure estimation accuracy for eigenvalues and eigenfunctions, respectively.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

We considered several scenarios corresponding to various choices of n, N, and σ^2 and simulated 1000 data sets for each scenario. We considered $n = 100, 200, 300$ subjects, $J = 2$ visits per subject, $N = 3, 6, 9, 12$ measurements per function, and magnitude of noise $\sigma = 0.01, 0.5, 1, 2$. Due to space limitations, we report results for a few scenarios and present the others in the Supporting information.

We found that the estimation accuracy of eigenvalues and eigenfunctions increases with the number of subjects n, the number of visits J, and the number of grid points N. With $n = 100$ subjects and $J = 2$ visits, the first two components at both levels can be recovered well when $N = 3$; more precisely, the shapes of estimated eigenfunctions approximate well the true eigenfunctions, and the estimated eigenvalues are close to their true values. When the number of grid points increases to $N = 9$, all four principal components at both levels can be estimated well. Table I reports the root mean square errors for eigenvalues and root integrated mean square errors for eigenfunctions in various simulation settings. More details on these results can be found in the Supporting information.

We also evaluated the finite sample performance of predictions of subject and subject/visit-level curves. We discuss our results for the case when $n = 200$, $J = 2$ with different levels of noise level, σ, and number of observations per visit, N. Figure 1 shows predictions for subject-specific and visit-specific curves for the first subject under various scenarios. In the sparsest case, $N = 3$, the BLUPs can still capture the rough shape of the curve, and the confidence bands are

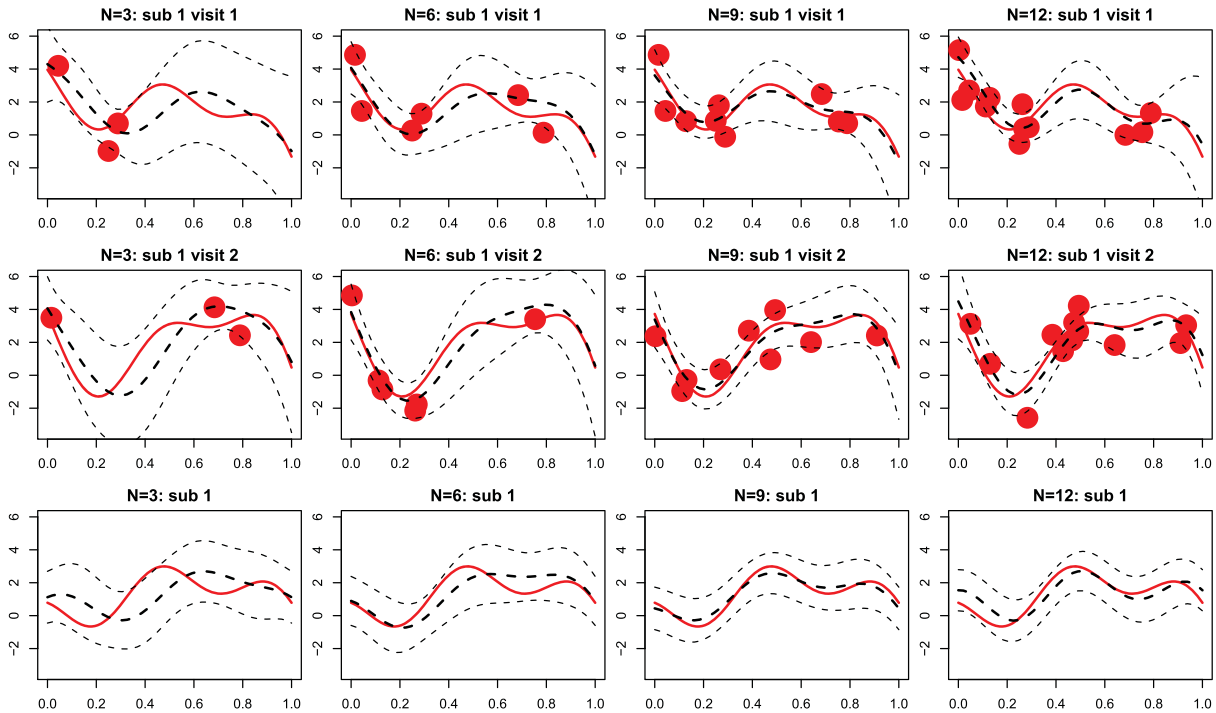

Figure 1. Prediction of curves for the first subject, in simulation setting with $n = 200$ subjects, $J = 2$ visits per subject, $N = 3, 6, 9, 12$ grid points per curve, and noise variance $\sigma^2 = 1$. Different columns correspond to different levels of sparsity, with the number of grid points varying from three to 12. The first and second rows show predictions of subject/visit-specific curves, at visits 1 and 2, respectively. In these subfigures, red solid lines correspond to the true underlying curves, $Y_{ij}(t)$, and red dots are observed sparse data, $Y_{ij}(t_{ijs})$. The thick black dashed lines are the predictions of curves, $\hat{Y}_{ij}(t)$, and thin black dashed lines give their 95% pointwise confidence bands. The third row displays subject-level curves, $Z_i(t)$, and their predictions $\hat{Z}_i(t)$.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

relatively wide because of the large amount of uncertainty but cover the true curve in most cases. For example, the predicted curve corresponding to subject 1 visit 2 (first panel in the middle row) identifies a local minimum at $t = 0.2$, even though there are no observations around the area. This is probably due to the additional information provided by the data for the same subject at visit 1 (first panel in the top row). The bottom row shows results for subject-specific curve, $Z_1(t)$, indicating that the BLUP estimates captures the major trend, although misses some of the details. When the number of grid points increases, predictions of both individual curves and subject-specific curves improve. When $N = 9$, these predictions are already very close to the true curves.

In summary, the sparse MFPCA algorithm is able to capture dominating modes of variations at both levels for sparse data, in a typical setting with hundreds of subjects and a few visits per subject. Predictions of curves via BLUPs also perform very well in finite samples.

5 Applications

5.1. SHHS

We now illustrate the use of sparse MFPCA on the SHHS data. The data contain dense EEG series for 3201 subjects at two visits per subject. Di et al. (2009) analyzed the full SHHS data using the MFPCA methodology and extracted dominant modes of variations at both between-subject and within-subject levels. To illustrate the comparison between dense and sparse analysis, we take a sparse random sample of the SHHS data even though the true data are dense. More precisely, for each subject and each visit, we take a random sample of N data points from the sleep EEG δ-power time series. We perform analysis with $N = 3, 6, 12, 24$ observations per visit and compare the results using the proposed sparse MFPCA method on the sub-sampled sparse data with those using the MFPCA method on the full data set. This will provide insights into the performance of our method and build confidence into using these methods even when N is small. The analysis based on the full data (henceforth denoted "full analysis") provides a "gold standard", which can be compared with the analysis based on sparse data. The sparse MFPCA analyses with increased number of observations, N, will further illustrate how much information is gained at different level of sparsity, in realistic settings.

Figure 2 displays the estimated mean functions, including the overall mean function and visit-specific mean functions, in the dense case and four different sparse cases. Even with $N = 3$, the estimated mean functions are similar to those from the full analysis. By borrowing information from 3201 subjects, the sparse MFPCA captures the trend of mean functions well and correctly identifies peaks and valleys, compared to those from the full analysis. When the number of grid points increases to six, 12, and 24, the estimated mean functions become indistinguishable from the full analysis. Table II displays estimated eigenvalues, and Figure 3 shows estimated eigenfunctions at the subject level (level 1). The dashed lines represent eigenfunctions estimated from the dense analysis, while solid lines correspond to estimated eigenfunctions from the sparse analysis. The first principal components were recovered very well in each case, even when $N = 3$. The shapes of the second and third components are roughly captured with three grid points and become closer to those from full analysis as N increases. When $N = 24$, all components agree well with their full analysis counterparts. For the visit level (level 2), similar results are observed, and the details are reported in the Supporting information. Similar patterns can be observed for eigenvalues and percent variance explained. The surprisingly good performance on estimation mean functions and eigenvalues is due to the ability of sparse MFPCA to borrow information over all subjects and visits.

We next evaluate predictions in each of the four sparse scenarios and illustrate the results on the first two subjects in Figure 4. The thick solid lines are predictions using sparse MFPCA methods on sparse data, while the thin solid lines are smooth estimates of the functions using the full data set. The dots are the actual sampled points. With $N = 3$ (first row of panels), predictions of curves can only capture the rough trend and miss refined features such as peaks and valleys, even though mean functions and eigenfunctions can be recovered well. This is not surprising, because

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

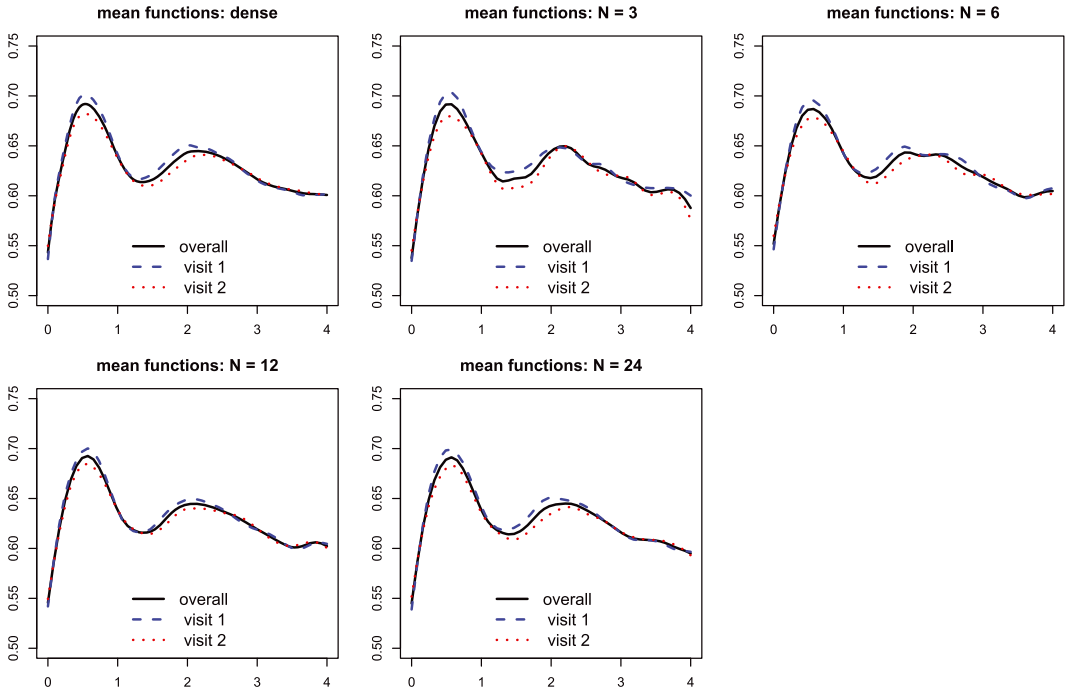

Figure 2. Estimated mean functions from MFPCA for the SHHS data: dense and sparse cases. The upper left panel shows estimated mean functions from the full analysis using dense data. The four remaining panels correspond to results from sparse cases, with the number of grid points per curve $N = 3, 6, 12, 24$, respectively. In each subfigure, solid lines correspond to overall mean functions, $\hat{\mu}(t)$, while dashed and dotted lines represent visit-specific mean functions, $\hat{\mu}(t) + \hat{\eta}_j(t)$, at visits 1 and 2, respectively.

Table II. Estimated eigenvalues (first three components at level 1 and first five components at level 2) for SHHS, from dense and four sparse cases.

		Level 1			Level 2				
		PC1	PC2	PC3	PC1	PC2	PC3	PC4	PC5
Dense	Eigenvalue	1.30	0.10	0.10	1.30	0.80	0.70	0.60	0.60
	Percent	80.80	7.60	3.30	21.80	12.80	12.50	10.90	9.60
$N = 3$	Eigenvalue	1.20	0.00	0.00	1.30	0.70	0.50	0.50	0.40
	Percent	98.90	1.10	0.00	34.00	18.50	14.40	13.20	10.30
$N = 6$	Eigenvalue	1.30	0.10	0.10	1.30	0.80	0.60	0.60	0.40
	Percent	86.40	5.60	4.30	27.40	15.60	13.50	11.80	9.10
$N = 12$	Eigenvalue	1.30	0.10	0.10	1.30	0.80	0.70	0.60	0.50
	Percent	80.30	7.30	4.30	24.60	14.70	13.80	11.50	9.40
$N = 24$	Eigenvalue	1.30	0.10	0.10	1.30	0.80	0.70	0.70	0.50
	Percent	81.80	7.10	3.50	23.60	13.80	13.40	12.10	9.70

"Percent" means percentage of variation explained by the corresponding to the principal component, relative to the total variation at the corresponding level. "N" is the number of grid points per function and reflects levels of sparsity.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

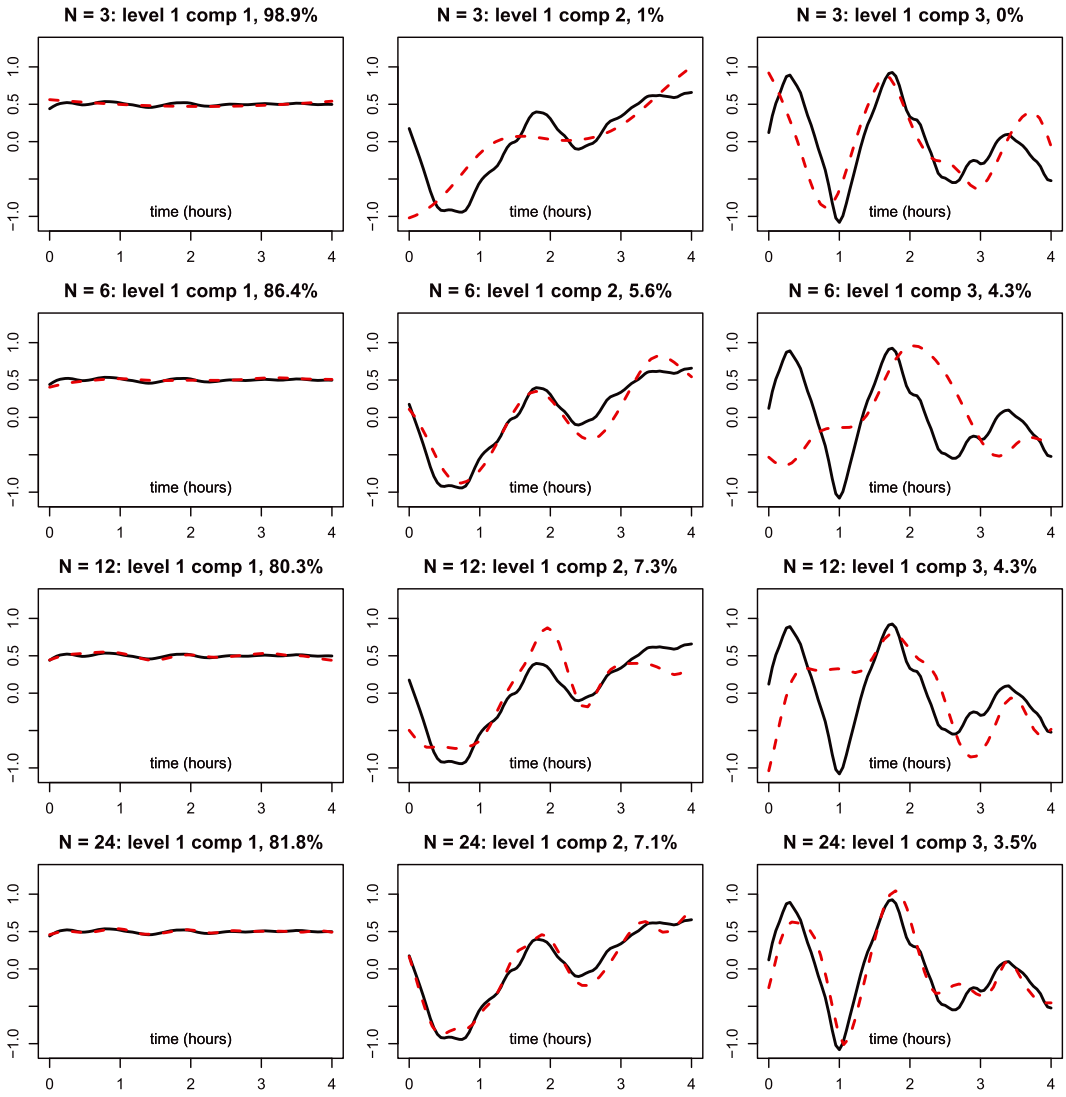

Figure 3. Estimated eigenfunctions (first three components at level 1) from MFPCA for the SHHS data: dense and sparse cases. The four rows correspond to different levels of sparsity, with the number of grid points per function $N = 3, 6, 12, 24$, respectively. The three columns represent the first three principal components, respectively. In each subfigure, solid black lines correspond to estimates from the dense case, while dashed red lines represent estimates from sparse cases.

much subject-specific information loss should be expected. When the number of sampled points increases (next rows of panels), the sparse MFPCA method estimates more detailed features.

To summarize, the sparse MFPCA methods can estimate the mean and principal component functions well even with sparse samples, provided that there are many subjects. The prediction accuracy greatly depends on the amount of available information, or equivalently, the level of sparsity. One can expect to recover a rough trend with few grid points, but more observations are needed to estimate detailed features, if those features exist.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

Stat
The ISI's Journal for the Rapid Dissemination of Statistics Research

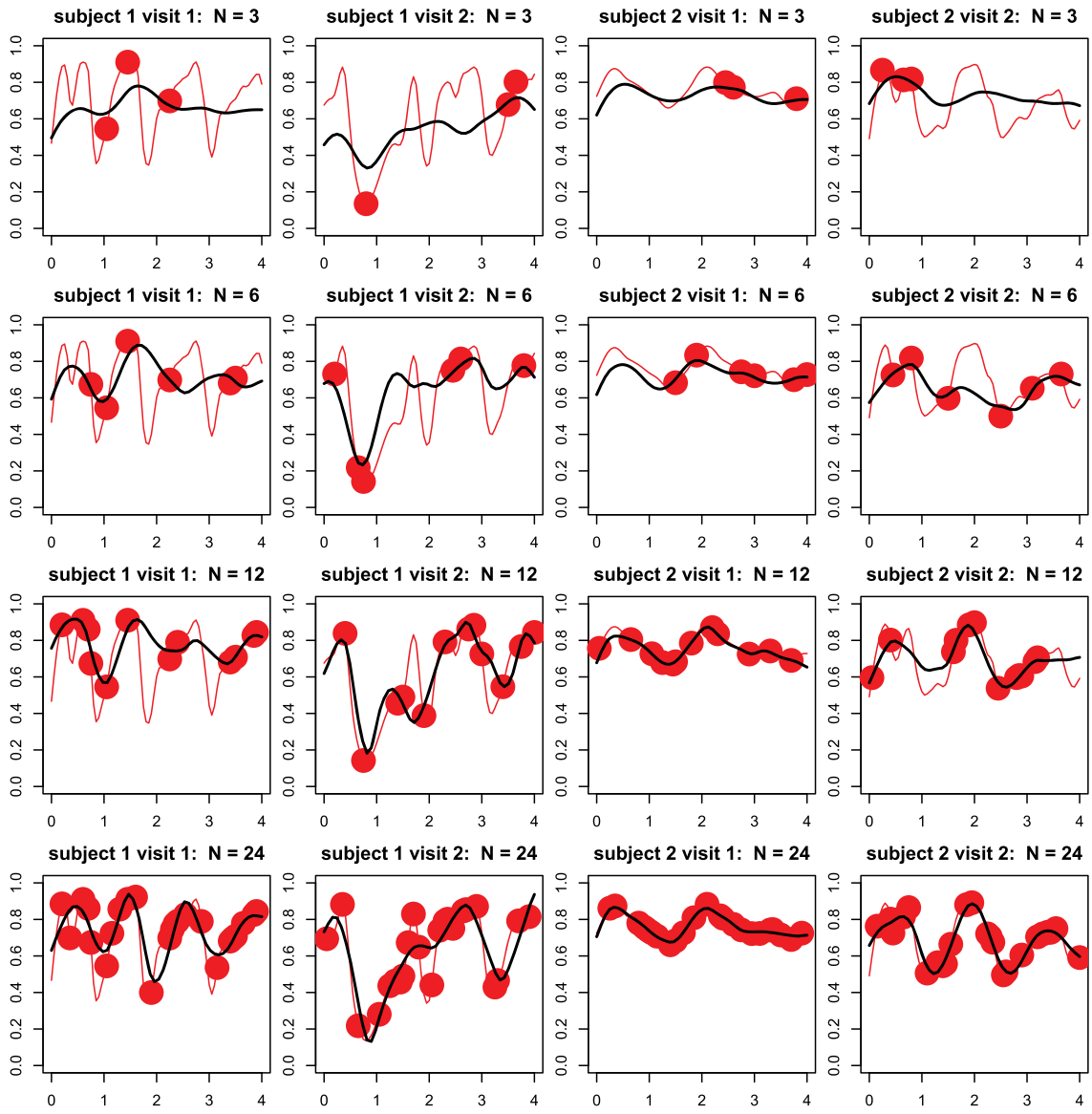

Figure 4. Prediction of sleep EEG curves for the first two subjects from the SHHS. The first and second columns correspond to visits 1 and 2 for the first subject, respectively. The third and fourth columns correspond to visits 1 and 2 for the second subject, respectively. Different rows represent different levels of sparsity, with the number of grid points N varying from three to 12. In each subfigure, red lines represent smoothed sleep EEG curves, while red dots are sparsified data at certain level of sparsity. Black lines are predictions of sleep EEG curves from the MFPCA model.

5.2. eBay auction data

In this section, we analyzed a subset of our online auction data that consist of 40 pairs of auctions of digital cameras. Every pair contains two auctions of exactly the same camera, so any observed differences in the auction outcome must be due to differences in the seller or the bidding process. Figure 5 displays the raw auction data for the first three cameras. Auction time is normalized to to $[0, 1]$, where $t = 0$ and $t = 1$ correspond to the beginning and end of an

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

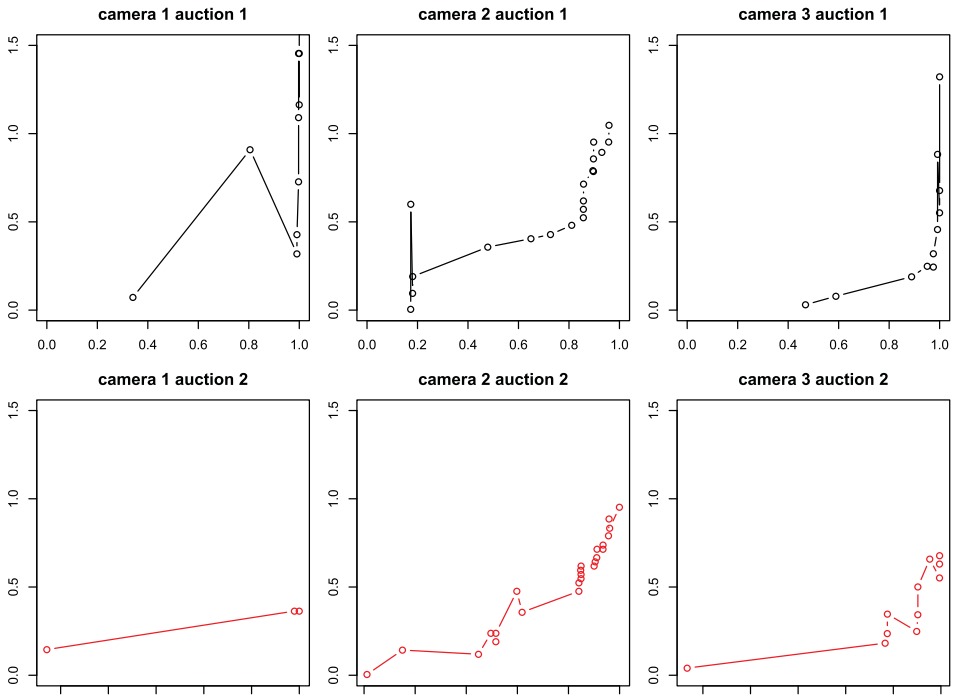

Figure 5. Auction trajectories for three digital cameras with two auctions per camera. The "*x*" axis represents normalized bidding time with range $t \in [0, 1]$, and the "*y*" axis corresponds to re-scaled bidding prices. The two subfigures in each column show two auctions from a specific type of digital camera.

auction, respectively. We also re-scale the *y*-axis ("price") to the average final price so that all auctions are comparable with respect to an item's value. Figure 5 shows that some auctions include as few as three bids, while others have as many as 20–30 bids. The bid timing is irregular and unbalanced—often it is quite sparse at the beginning and middle phases of the auction but rather dense towards the end. We point out that the bids in Figure 5 are not monotonically increasing as would be expected from an ascending auction. The reason lies in eBay's *proxy bidding system* (explained in detail in the Supporting information).

We look at level 1 and 2 principal components, which extract dominating modes of variations at the between and within camera level, respectively. Eigenvalues and eigenfunctions are shown in Table III and Figure 6, respectively. Based on Table III, levels 1 and 2 explain 35.9% and 64.1% of the total variation, respectively. Thus, there is substantial amount of variation at both levels.

Next, we look at estimated eigenfunctions at both levels, which capture modes of variations across different cameras and across auctions within the same camera. In Figure 6, the first row shows the shapes of three leading eigenfunctions at level 1, $\left\{\phi_1^{(1)}(t), \phi_2^{(1)}(t), \phi_3^{(1)}(t)\right\}$, while the second row displays the types of variations resulting from them. The first eigenfunction (PC1) is mostly positive, indicating that auctions loading positively (negatively) on this component will always have higher (lower) values than average. Its magnitude is small at the beginning, increases and reaches maximum around $t = 0.2$, then gradually decreases to 0 around $t = 0.8$, and is close to 0 after $t = 0.8$. The price for an auction loading positively on PC1 (the "+" line in row 2 column 1) increases rapidly at the beginning, almost flattens in the middle phase, and increases somewhat towards the end, while an auction with a negative loading on PC1 (the "-" line in row 2 column 1) has prices flat at the beginning but increasing after $t = 0.2$. This component explains

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

Stat
The ISI's Journal for the Rapid Dissemination of Statistics Research

Table III. Estimated eigenvalues from eBay auction data.

	Level 1 (Proportion explained: 35.9%)			Level 2 (Proportion explained: 64.1%)	
Component	1	2	3	1	2
Eigenvalue ($\times 10^{-3}$)	6.2	3.6	2.3	19.4	3.4
% var	48.2	27.8	18.1	84.6	14.6
Cum. % var	48.2	76.0	94.1	84.6	99.2

Levels 1 and 2 correspond to between and within camera variations, respectively. "% var" means percentage of variance explained by corresponding components, while "cum. % var" means cumulative percentage of variance explained by current and prior components.

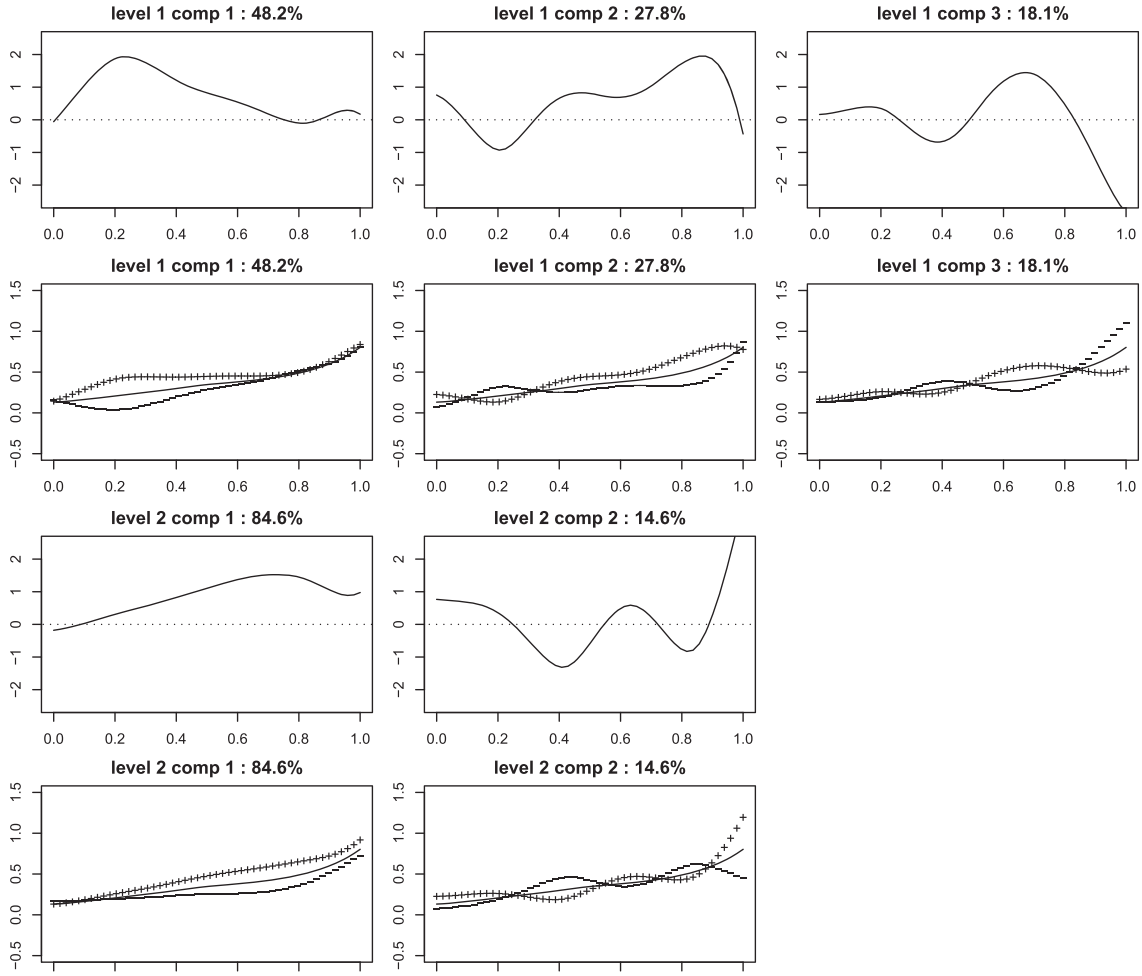

Figure 6. Estimated eigenfunctions at level 1 (Rows 1-2) and level 2 (Rows 3-4). In Row 1 and 3, the solid lines represent estimated eigenfunctions. In Row 2 and 4, solid lines represent the overall mean function $\mu(t)$, and the lines indicated by "+" and "-" are mean function plus or minus a multiple of $\phi_1^{(1)}(t)$, respectively. Namely, "+" represents $\mu(t) + c\,\phi_1^{(1)}(t)$, and "-" represents $\mu(t) + c\,\phi_1^{(1)}(t)$ for some constant c.

48.2% of variation at the between camera level. In the auction context, this PC suggests that while price (or more precisely, willingness to pay) increases over the course of auctions, there is significant variation in bidder's valuations during the early auction stages. This is in line with earlier research that has documented that the early auction phase is often swamped with "bargain hunters". The second principal component that explains 27.8% of variation characterizes mostly the variation at the later parts ($t = 0.4$ to $t = 0.95$) of auctions. This is in line with auctions that experience different early-stage and late-stage dynamics (Jank & Shmueli, 2009). One can interpret other eigenfunctions in a similar manner. The third and fourth rows of Figure 6 show the shapes of two leading eigenfunctions at level 2, e.g., with camera level. More detailed explanations of the auction application can be found at the Supporting information.

Our analysis allows one to partition bidders' willingness to pay into temporally different segments. In fact, level 1 analysis shows that early differences in willingness to pay can be attributed mostly to product differences. Our level 2 analysis suggests that bidders' uncertainty about the valuation is largest during mid to late auction.

6 Discussion

We considered sparsely sampled multilevel functional data and proposed a sparse MFPCA methodology for such data. We incorporated smoothing to deal with sparsity. Simulation studies show that our methods perform well. In an application to the SHHS, we compared the full analysis with sparse analysis under different levels of sparsity. The results show that the sparse MFPCA methodology works well in extracting the principal components, while prediction accuracy of functions depends on the level of sparsity. We also applied our methods to auction data from eBay and identified interesting bidding patterns. Di et al. (2009) are the first to attempt to generalize FPCA to multilevel functional data, and the current paper further extends its scope to sparsely sampled hierarchical functions. To the best of our knowledge, this is the first FPCA-based approach for sparsely sampled functional data. The proposed sparse MFPCA not only helps understand modes of variation at different levels but also reduces dimension for further analysis.

Multilevel FDA is a rich research area motivated by an explosion of studies that generate functional data sets. The explosion is mainly due to improved technologies that generate a new type of data sets. There are few directions for future methodological development. First, it would be interesting to extend the generalized multilevel functional regression Crainiceanu et al. (2009b) to the sparse case. Second, more efficient methods for estimating eigenvalues and eigenfunctions (e.g., generalizations of James et al., 2000; Peng & Paul, 2009) can be extended to the multilevel functional data. Another direction is to consider non-FPCA-based methods, e.g., extensions of wavelet-based functional mixed models of Morris et al. (2003) and Morris & Carroll (2006).

Acknowledgements

Chongzhi Di's research was supported by grants R21ES022332, P01CA53996, R01HG006124, and R01AG014358 from the National Institute of Health. Ciprian Crainiceanu's research was supported by Award Number R01NS060910 from the National Institute Of Neurological Disorders and Stroke.

References

Bapna, R, Goes, P, Gupta, A & Jin, Y (2004), 'User heterogeneity and its impact on electronic auction market design: an empirical exploration', *MIS Quarterly*, **28**(1), 21–43.

Crainiceanu, CM, Caffo, BS, Di, CZ & Punjabi, NM (2009a), 'Nonparametric signal extraction and measurement error in the analysis of electroencephalographic activity during sleep', *Journal of the American Statistical Association*, **104**(486), 541–555.

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

The ISI's Journal for the Rapid Dissemination of Statistics Research

Crainiceanu, CM, Staicu, AM & Di, CZ (2009b), 'Generalized multilevel functional regression', *Journal of the American Statistical Association*, **104**(488), 1550–1561.

Di, CZ, Crainiceanu, CM, Caffo, BS & Punjabi, NM (2009), 'Multilevel functional principal component analysis', *Annals of Applied Statistics*, **3**(1), 458–488.

Diggle, PJ, Heagerty, P, Liang, KY & Zeger, SL (2002), *Analysis of Longitudinal Data*, second edn, Oxford Statistical Science Series, vol. 25, *Oxford University Press*, Oxford.

Fan, J & Gijbels, I (1996), *Local Polynomial Modelling and its Applications*, Monographs on Statistics and Applied Probability, *Chapman & Hall*, London.

Goldsmith, J, Greven, S & Crainiceanu, C (2013), 'Corrected confidence bands for functional data using principal components', *Biometrics*, **69**(1), 41–51.

Hall, P & Hosseini-Nasab, M (2006), 'On properties of functional principal components analysis', *Journal of the Royal Statistical Society—Series B*, **68**(1), 109–126.

James, GM, Hastie, TJ & Sugar, CA (2000), 'Principal component models for sparse functional data', *Biometrika*, **87**(3), 587–602.

Jank, W & Shmueli, G (2009), 'Studying heterogeneity of price evolution in eBay auctions via functional clustering', Handbook of Information Systems Series: Business Computing, Emerald Group Publishing Limited, Bingley, United Kingdom, 237–261.

Jank, W, Shmueli, G & Zhang, S (2010), 'A flexible model for price dynamics in online auctions', *Journal of the Royal Statistical Society—Series C*, **59**(5), 781–804.

Karhunen, K (1947), *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, *Suomalainen Tiedeakatemia*, Helsinki, Finland.

Liu, B & Müller, H (2009), 'Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics', *Journal of the American Statistical Association*, **104**(486), 704–717.

Loève, M (1945), 'Fonctions aléatoires de second ordre', *Comptes Rendus Académie des Sciences*, **220**, 469.

Morris, JS & Carroll, RJ (2006), 'Wavelet-based functional mixed models', *Journal of the Royal Statistical Society—Series B*, **68**(2), 179–199.

Morris, JS, Vannucci, M, Brown, PJ & Carroll, RJ (2003), 'Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis', *Journals American Statistical Association*, **98**(463), 573–584.

Müller, HG (2005), 'Functional modelling and classification of longitudinal data', *Scandinavian Journal of Statistics*, **32**, 223–240.

Peng, J & Müller, H (2008), 'Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions', *Annals of Applied Statistics*, **2**(3), 1056–1077.

Peng, J & Paul, D (2009), 'A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data', *Journal of Computational and Graphical Statistics*, **18**(4), 995–1015.

Quan, SF, Howard, BV, Iber, C, Kiley, JP, Nieto, FJ, O'Connor, GT, Rapoport, DM, Redline, S, Robbins, J, Samet, JM & Wahl, PW (1997), 'The Sleep Heart Health Study: design, rationale, and methods', *Sleep*, **20**, 1077–1085.

Ramsay, JO & Dalzell, CJ (1991), 'Some tools for functional data analysis', *Journal of the Royal Statistical Society—Series B*, **53**(3), 539–572.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.50

Ramsay, JO & Silverman, BW (2005), *Functional Data Analysis*, second edn, Springer Series in Statistics, *Springer Verlag*, New York.

Reithinger, F, Jank, W, Tutz, G & Shmueli, G (2008), 'Smoothing sparse models, unevenly sampled curves using semiparametric mixed: an application to online auctions', *Journal of the Royal Statistical Society—Series C*, **57**(2), 127–148.

Rice, JA & Silverman, BW (1991), 'Estimating the mean and covariance structure nonparametrically when the data are curves', *Journal of the Royal Statistical Society—Series B*, **53**(1), 233–243.

Ruppert, D, Wand, MP & Carroll, RJ (2003), *Semiparametric Regression*, Cambridge Series in Statistical and Probabilistic Mathematics, *Cambridge University Press*, Cambridge.

Shmueli, G, Russo, R & Jank, W (2007), 'The BARISTA: a model for bid arrivals in online auctions', *The Annals of Applied Statistics*, **1**(2), 412–441.

Silverman, BW (1996), 'Smoothed functional principal components analysis by choice of norm', *Annals of Statistics*, **24**(1), 1–24.

Yao, F, Müller, HG & Wang, JL (2005), 'Functional data analysis for sparse longitudinal data', *Journals of the American Statistical Association*, **100**(470), 577–591.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.