HHS Public Access

Author manuscript

Stat (Int Stat Inst). Author manuscript; available in PMC 2015 November 19.

Published in final edited form as:

Stat (Int Stat Inst). 2015 ; 4(1): 212–226. doi:10.1002/sta4.89.

Longitudinal Functional Data Analysis

So Young Park[a] and **Ana-Maria Staicu**[a,*]

[a] Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA

Abstract

We consider dependent functional data that are correlated because of a longitudinal-based design: each subject is observed at repeated times and at each time a functional observation (curve) is recorded. We propose a novel parsimonious modeling framework for repeatedly observed functional observations that allows to extract low dimensional features. The proposed methodology accounts for the longitudinal design, is designed to study the dynamic behavior of the underlying process, allows prediction of full future trajectory, and is computationally fast. Theoretical properties of this framework are studied and numerical investigations confirm excellent behavior in finite samples. The proposed method is motivated by and applied to a diffusion tensor imaging study of multiple sclerosis.

Keywords

Dependent functional data; Diffusion Tensor Imaging; Functional principal component analysis; Longitudinal design; Multiple Sclerosis

1. Introduction

Longitudinal functional data consist of functional observations (such as profiles or images) observed at several times for each subject of many. Examples of such data include the Baltimore Longitudinal Study of Aging (BLSA), where daily physical activity count profiles are observed for each subject at several consecutive days (Goldsmith et al., 2014; Xiao et al., 2015) and the longitudinal diffusion tensor imaging (DTI) study, where modality profiles along well-identified tracts are observed for each multiple sclerosis (MS) patient at several hospital visits (Greven et al., 2010). As a result of an increasing number of such applications, *longitudinal functional data analysis* has received much attention recently; see for example Morris et al. (2003); Morris & Carroll (2006); Baladandayuthapani et al. (2008); Di et al. (2009); Greven et al. (2010); Staicu et al. (2010); Li & Guan (2014).

Our motivation is the longitudinal DTI study, where the objective is to investigate the evolution of the MS disease as measured by the dynamics of a common DTI modality profile - fractional anisotropy (FA) - along the corpus callosum (CCA) of the brain. Every

* astaicu@ncsu.edu.

Supplementary Material

Detailed proofs of the theoretical results, additional numerical investigations and data analysis results are included in a supplementary material that is available online.

MS subject in the study is observed over possibly multiple hospital visits and at each visit the subject's brain is imaged using DTI. In this paper we consider summaries of FA at 93 equally spaced locations along the brain's CCA, which we refer to as CCA-FA profile. The change over time in the CCA-FA profiles is informative of the progression of the MS disease, and thus a model that accounts for all the dependence sources in the data has the potential to be a very useful tool in practice. We propose a modeling framework that captures the process dynamics over time and provides prediction of a full CCA-FA trajectory at a future visit.

Existing literature in longitudinal functional data can be separated into two categories, based on whether or not it accounts for the actual time T_{ij} at which the profile $Y_{ij}(\cdot)$ is observed; here i indexes the subjects and j indexes the repeated measures of the subject. Moreover, most methods that incorporate the time T_{ij} focus on modeling the process dynamics (Greven et al., 2010) and only few can do prediction of a future full trajectory. Chen & Müller (2012) considered the latter issue and introduced an interesting perspective, but their method is very computationally expensive and its application in practice is limited as a result. We propose a novel parsimonious modeling framework to study the process dynamics and prediction of future full trajectory in a computationally feasible manner.

In this paper we focus on the case where the sampling design of T_{ij}'s is sparse (hence *sparse longitudinal design*) and the subject profiles are observed at fine grids (hence *dense functional design*). We propose to model $Y_{ij}(\cdot)$ as:

$$Y_{ij}(s) = \mu(s, T_{ij}) + X_i(s, T_{ij}) + \epsilon_{ij}(s); \quad X_i(s, T_{ij}) = \sum_{k \geq 1} \xi_{ik}(T_{ij}) \phi_k(s) \qquad \text{for} \quad s \in \mathcal{S} \quad \text{and} \quad T_{ij} \in \mathcal{T} \tag{1}$$

where $\mathcal{S}$ and $\mathcal{T}$ are closed compact sets, $\mu(\cdot, T_{ij})$ is an unknown smooth mean response corresponding to T_{ij}, $X_i(\cdot, T_{ij})$ is a smooth random deviation from the mean at T_{ij}, and $\varepsilon_{ij}(\cdot)$ is a residual process with zero-mean and unknown covariance function to be described later. The bivariate processes $X_i(\cdot, \cdot)$'s are independent and identically distributed (iid), the error processes ε_{ij}'s are iid and furthermore are independent of X_i's. For identifiability we require that X_i comprises solely the random deviation that is specific to the subject; the repeated time-specific deviation is included in ε_{ij}. Here $\{\varphi_k(\cdot)\}_k$ is an orthogonal basis in $L^2(\mathcal{S})$ and $\xi_{ik}(T_{ij})$'s are the corresponding basis coefficients that have zero-mean, are uncorrelated over i, but correlated over j. We assume that the set of visit times of all subjects, $\{T_{ij} : i, j\}$, is dense in $\mathcal{T}$. Full model assumptions are given in Section 2.

The class of model (1) is rich and includes many existent models, as we illustrate now. (i) If $\xi_{ik}(T_{ij}) = \zeta_{0,ik} + T_{ij}\zeta_{1,ik}$ for appropriately defined random terms $\zeta_{0,ik}$ and $\zeta_{1,ik}$, model (1) can be represented as in Greven et al. (2010). (ii) If $\text{cov}(\xi_{ik}(T), \xi_{ik}(T')) = \lambda_k \rho_k(|T - T'|; \nu)$ for some unknown variance λ_k, known correlation function $\rho_k(\cdot; \nu)$ with unknown parameter ν, and $n = 1$, model (1) resembles to Gromenko et al. (2012) and Gromenko & Kokoszka (2013) for spatially indexed functional data. (iii) If $\xi_{ik}(T_{ij}) = \sum_{l \geq 1} \zeta_{ikl} \psi_{ikl}(T_{ij})$ with orthogonal basis functions $\psi_{ikl}(T)$'s and the corresponding coefficients ζ_{ikl}'s, then model (1) is similar to Chen & Müller (2012) who used time-varying basis functions $\varphi_k(\cdot|T)$ instead of our proposed $\varphi_k(\cdot)$ in model (1) and assumed a white noise residual process ε_{ij}.

The use of time-invariant orthogonal basis functions is one key difference between the proposed framework and Chen & Müller (2012); another important difference is the flexible error structure that our approach accommodates. The key difference leads to several major advantages of the proposed method. First, by using a time-invariant basis functions, the basis coefficients, $\xi_{ik}(T_{ij})$'s extract the low dimensional features of these massive data. The longitudinal dynamics is emphasized only through the time-varying coefficients $\xi_{ik}(T_{ij})$'s of (1) and, thus, this perspective makes the study of the process dynamics easier to understand. Second, our approach involves at most two dimensional smoothing and as a result is computationally very fast; in contrast, the time-varying basis functions $\{\varphi_k(\cdot|T)\}_k$ at each T, require three dimensional smoothing which is not only complex but also computationally intensive and slow.

Nevertheless, selecting the time-invariant basis is nontrivial. One option is to use a pre-specified basis; Zhou et al. (2008) considered this approach in modeling paired of sparse functional data. Another option is to use data-driven basis functions, such as eigenbasis of some covariance. The challenge is: what covariance to use ? We take the latter direction and propose to determine $\{\varphi_k(\cdot)\}_k$ using an appropriate *marginal* covariance. In this regard, let $c((s, T), (s', T'))$ be the covariance function of $X_i(s, T)$ and $g(T)$ be the density of T_{ij}'s. Define $\Sigma(s, s') = \int_{\mathcal{T}} c\left((s, T), \left(s', T\right)\right) g(T)\, dT$ for $s, s' \in \mathcal{S}$: we show that this bivariate function is a proper covariance function (Horváth & Kokoszka, 2012). Section 2 shows that the proposed basis $\{\varphi_k(\cdot)\}_k$ has optimal properties with respect to some appropriately defined criterion. From this view point, the model representation (1) is optimal. The idea of using the eigenbasis of the pooled covariance can be related to Jiang & Wang (2010) and Pomann et al. (2013), who considered independent functional data.

The rest of paper is organized as follows. Section 2 introduces the proposed modeling framework. Section 3 describes the estimation methods and implementation. The methods are studied theoretically in Section 4 and then numerically in Section 5. Section 6 discusses the application to the tractography DTI data.

2. Modeling longitudinal functional data

Let $[\{T_{ij}, Y_{ij}(s_r) : r = 1, \ldots, R\} : j = 1, \ldots, m_i,]$ be the observed data for the ith subject, where $Y_{ij}(\cdot)$ is the jth profile at random time T_{ij} for subject i, and each profile is observed at the fine grid of points $\{s_1, \ldots, s_R\}$. For convenience we use the generic index s instead of s_r. The number of 'profiles' per subject, m_i is relatively small to moderate and the set of time points of all subjects, $\{T_{ij}$: for all $i, j\}$, is dense in $\mathcal{T}$. Without loss of generality, we set $\mathcal{S} = \mathcal{T} = [0, 1]$. We model the response $Y_{ij}(\cdot)$ using (1), where we assume that $\varepsilon_{ij}(s)$ is the sum of independent components $\varepsilon_{ij}(s) = \varepsilon_{1,ij}(s) + \varepsilon_{2,ij}(s)$. Here $\varepsilon_{1,ij}(\cdot)$ is a random square integrable function which has smooth covariance function $\Gamma(s, s') = \text{cov}\{\varepsilon_{1,ij}(s), \varepsilon_{1,ij}(s')\}$ and $\varepsilon_{2,ij}(s)$ is white noise with covariance $\text{cov}\{\varepsilon_{2,ij}(s), \varepsilon_{2,ij}(s')\} = \sigma^2$ if $s = s'$ and 0 otherwise.

Let $c((s, T), (s', T')) = \text{E}[X_i(s, T)X_i(s', T')]$ be the covariance function of the process $X_i(\cdot,\cdot)$ and let $\Sigma(s, s') = \int c((s, T), (s', T))g(T)dT$, where $g(\cdot)$ is the sampling density of T_{ij}. In Section 4 we show that $\Sigma(s, s')$ is a *proper* covariance function (Horváth & Kokoszka, 2012); due to its definition we call Σ as the *marginal* covariance function induced by X_i. The

unpublished work Chen et al. (2015) independently considered a similar marginal covariance in a related setting. Denote by $W_i(s, T_{ij}) = X_i(s, T_{ij}) + \varepsilon_{1,ij}(s)$; W_i is a bivariate process defined on $[0, 1]^2$ and its induced marginal covariance is $\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')$. Let $\{\varphi_k(s), \lambda_k\}_k$ be the eigencomponents of $\Xi(s, s')$, where $\{\varphi_k(\cdot) : k\}$ forms an orthogonal basis in $L^2[0, 1]$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$. Using arguments similar to the standard functional principal component analysis (FPCA), the eigenbasis functions $\{\varphi_k(\cdot) : k = 1, \ldots, K\}$ are optimal in the sense that they minimize the following weighted mean square error:

$$\mathrm{MSE}(\theta_1(\cdot), \ldots, \theta_K(\cdot)) = \int_0^1 \mathsf{E} \left\| W_i(\cdot, T) - \sum_{k=1}^{K} <W_i(\cdot, T), \theta_k(\cdot)> \theta_k(\cdot) \right\|^2 g(T)\, dT,$$

where $<f_1(\cdot), f_2(\cdot)> = \int_0^1 f_1(s)\, f_2(s)\, ds$ is the usual inner product in $L^2[0, 1]$.

Using the orthogonal basis in $L^2[0, 1]$ $\{\varphi_k(\cdot)\}_k$, we can represent the square integrable smooth process $W_i(\cdot, T)$ as $W_i(s, T_{ij}) = \sum_{k=1}^{\infty} \xi_{W,ijk} \phi_k(s)$, where $\xi_{W,ijk} = \int W_i(s, T_{ij}) \varphi_k(s) ds = \xi_{ik}(T_{ij}) + e_{ijk}$, and $\xi_{W,ijk}$ are not necessarily uncorrelated over k. Here $\xi_{ik}(T_{ij}) = \int X_i(s, T_{ij}) \varphi_k(s) ds$ and $e_{ijk} = \int \varepsilon_{1,ij}(s) \varphi_k(s) ds$ are specified by the definition of W_i; for fixed k these terms are mutually independent due to the independence of the processes X_i and $\varepsilon_{1,ij}$. For each k, one can easily show that, $\xi_{ik}(\cdot)$ is a smooth zero-mean random process in $L^2[0, 1]$ and is iid over i. Furthermore e_{ijk} are zero-mean iid random variables over i, j; denote by $\sigma^2_{e,k}$ their finite variance.

One way to model the dependence of the coefficients, $\xi_{ik}(T_{ij})$'s, is by using common techniques in longitudinal data analysis; for example by assuming a parametric covariance structure. As we discussed in Section 1, this leads to models similar to Greven et al. (2010); Gromenko et al. (2012); Gromenko & Kokoszka (2013). We consider this approach in the analysis of the DTI data, Section 6. Another approach is to assume a nonparametric covariance structure and employ a common functional data analysis technique. We detail the latter approach in this section.

For each $k \geq 1$ denote by $G_k(T, T') = \mathrm{cov}\{\xi_{,ik}(T), \xi_{,ik}(T')\}$ the smooth covariance function in $[0, 1] \times [0, 1]$. Mercer's theorem provides the following convenient spectral decomposition $G_k(T, T') = \sum_{l \geq 1} \eta_{kl} \psi_{kl}(T) \psi_{kl}(T')$, where $\eta_{k1} \geq \eta_{k2} \geq \ldots \geq 0$ and $\{\psi_{kl}(\cdot)\}_{l \geq 1}$ is an orthogonal basis in $L^2[0, 1]$. Using the Karhunen-Loève (KL) expansion, we represent $\xi_{ik}(\cdot)$ as: $\xi_{ik}(T_{ij}) = \sum_{l=1}^{\infty} \zeta_{ikl} \psi_{kl}(T_{ij})$, where $\zeta_{ikl} = \int \xi_{ik}(T) \psi_{kl}(T) dT$, have zero-mean, variance equal to η_{kl}, and are uncorrelated over l. By collecting all the components, we represent the model (1) as $Y_{ij}(s) = \mu(s, T_{ij}) + \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \zeta_{ikl} \psi_{kl}(T_{ij}) \phi_k(s) + \epsilon_{ij}(s)$, for $\epsilon_{ij}(s) = \sum_{k=1}^{\infty} e_{ijk} \phi_k(s) + \epsilon_{2,ij}(s)$. In practice we truncate this expansion. Lek K and $L_1, \ldots, L_K$ such that $Y_{ij}(s)$ is well approximated by the following truncated model based on the leading K and $\sum_k L_k$ respective basis functions

$$Y_i(s, T_{ij}) = \mu(s, T_{ij}) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} \zeta_{ikl} \psi_{kl}(T_{ij}) \phi_k(s) + \epsilon_{ij}(s), \quad (2)$$

where $\epsilon_{ij}(s) \cong \sum_{k=1}^{K} e_{ijk}\phi_k(s) + \epsilon_{2,ij}(s)$. The truncated model (2) gives a parsimonious representation of the longitudinal functional data. It allows to study its dependence through two sets of eigenfunctions: one dependent solely on s and one solely on T_{ij}. This approach involves two main challenges: first, determining consistent estimator of the marginal covariance and second determining consistent estimators of the time-varying coefficients $\xi_{ik}(\cdot)$.

3. Estimation of model components

We discuss estimation of all model components. The mean estimation is carried out using existing methods (Chen & Müller, 2012; Scheipl et al., 2014); here we briefly describe it for completeness. Our focus and novelty is the estimation of the marginal covariance function and of the eigenfunctions $\varphi_k(\cdot)$'s (see Section 3.2), as well as the the estimation of the time-varying basis coefficients $\xi_{ik}(\cdot)$'s (see Section 3.3). Prediction of $Y_i(s, T)$ is detailed in Section 3.4.

3.1. Step 1: Mean function

As in Scheipl et al. (2014) we estimate the mean function, $\mu(s, T)$, using bivariate smoothing via bivariate tensor product splines (Wood, 2006) of the pooled data $Y_{ijr} = Y_{ij}(s_r)$'s. Consider two univariate B-spline bases, $\{B_{s,1}(s), \ldots, B_{s,d_s}(s)\}$ and $\{B_{T,1}(T), \ldots, B_{T,d_T}(T)\}$, where d_s and d_T are their respective dimensions. The mean surface is represented as a linear combination of a tensor product of the two univariate B-spline bases $\mu(s, T) = \sum_{q_1=1}^{d_s}\sum_{q_2=1}^{d_T} B_{s,q1}(s)B_{T,q2}(T)\beta_{q1q2} = \mathbf{B}(s, T)^T\boldsymbol{\beta}$, where $\mathbf{B}(s, T)$ is the known $d_s d_T$-dimensional vector of $B_{s,q1}(s)B_{T,q2}(T)$'s, and $\boldsymbol{\beta}$ is the vector of unknown parameters, β_{q1q2}'s. The bases dimensions, d_s and d_T, are set to be sufficiently large to accommodate the complexity of the true mean function, and the roughness of the function is controlled through the size of the curvature in each direction separately, i.e.

$\int\int\{\partial^2\mu(s,T)/\partial s^2\}^2 dT\, ds = \boldsymbol{\beta}^T\left(\mathbf{p}_s \otimes \mathbf{I}_{d_T}\right)\boldsymbol{\beta}$ in direction s, and

$\int\int\{\partial^2\mu(s,T)/\partial T^2\}^2 dT ds = \boldsymbol{\beta}^T(\mathbf{I}_{d_s} \otimes \mathbf{P}_T)\boldsymbol{\beta}$ in T. The penalized criterion to be minimized is $\sum_{i,j,r}\left[Y_{ijr} - \mathbf{B}(s_r, T_{ij})^T\boldsymbol{\beta}\right]^2 + \boldsymbol{\beta}^T\left(\lambda_s\mathbf{P}_s \otimes \mathbf{I}_{d_T} + \lambda_T\mathbf{I}_{d_s} \otimes \mathbf{P}_T\right)\boldsymbol{\beta}$,, where λ_s and λ_T are smoothing parameters that control the trade-off between the smoothness of the fit and the goodness of fit. The smoothing parameters can be selected by the restricted maximum likelihood (REML) or generalized cross-validation (GCV). The estimated mean function is $\hat{\mu}(s,T) = \mathbf{B}(s,T)^T\hat{\boldsymbol{\beta}}$. This method is a very popular smoothing technique of bivariate data.

Other available bivariate smoothers can be used to estimate the mean $\mu(s, T)$: for example, kernel-based local linear smoother (Hastie et al., 2009), bivariate penalized spline smoother (Marx & Eilers, 2005) and the sandwich smoother (Xiao et al., 2013). The sandwich smoother (Xiao et al., 2013) is especially useful in the case of very high dimensional data for its appealing computational efficiency, in addition to its estimation accuracy.

3.2. Step 2: Marginal covariance. Data-based orthogonal basis

Once the mean function is estimated, let $\widetilde{Y}_{ijr}=Y_{ijr}-\hat{\mu}(s_r,T_{ij})$ be the demeaned data. We use the demeaned data to estimate the marginal covariance function induced by $W_i(s, T_{ij})$, $\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')$. The estimation of $\Xi(s, s')$ consists of two steps. In the first step, a raw covariance estimator $\widetilde{\Xi}\left(s,s'\right)$ is obtained; the pooled sample covariance is a suitable choice, if all the curves are observed on the same grid of points:

$$\widetilde{\Xi}(s_r,s_{r'})=\sum_{i=1}^{n}\sum_{j=1}^{m_i}\tilde{Y}_{ijr}\tilde{Y}_{ijr'}/\left(\sum_{i=1}^{n}m_i\right). \quad (3)$$

As data Y_{ijr}'s are observed with white noise, $\varepsilon_{2,ij}(s_r)$, the 'diagonal' elements of the sample covariance, $\widetilde{\Xi}(s_r,s_r)$, are inflated by the variance of the noise, σ^2. In the second step, the preliminary covariance estimator is smoothed by ignoring the 'diagonal' terms; see also Staniswalis & Lee (1998) and Yao et al. (2005) who used similar technique for the case of independent functional data. In our simulation and data application we use the sandwich smoother (Xiao et al., 2015). To ensure the positive semi-definiteness of the estimator the negative eigenvalues are zero-ed. The resulting smoothed covariance function, $\hat{\Xi}\left(s,s'\right)$, is used as an estimator of $\Xi\left(s,s'\right)$. In Section 4, we show that $\hat{\Xi}\left(s,s'\right)$ is an unbiased and consistent estimator of $\Xi\left(s,s'\right)$ in two settings: 1) the data are observed fully and without noise, i.e. $\in_{ij}(s) \equiv 0$ and 2) the data are observed fully and with measurement error of type $\in_{1,ij}(s)$, i.e $\in_{ij}(s) \equiv \in_{1,ij}(s)$.

Let $\left\{\hat{\phi}_k(s),\hat{\lambda}_k\right\}_k$ be the pairs of eigenvalues/eigenfunctions obtained from the spectral decomposition of the estimated covariance function, $\hat{\Xi}\left(s,s'\right)$. The truncation value K is determined based on pre-specified percentage of variance explained (PVE); specifically, K can be chosen as the smallest integer such that $\left(\sum_{k=1}^{K}\hat{\lambda}_k/\sum_{k=1}^{\infty}\hat{\lambda}_k\right)$ is greater than the pre-specified PVE (Di et al. 2009; Staicu et al., 2010).

3.3. Step 3: Covariance of the time-varying coefficients

Let $\widetilde{\xi}_{W,ijk}=\int\widetilde{Y}_{ij}(s)\,\hat{\phi}_k(s)\,ds$ be the projection of the jth repeated demeaned curve of the ith subject onto the direction $\hat{\phi}_k(\cdot)$ for $k = 1, \ldots, K$. Since $\widetilde{Y}_{ij}(\cdot)$ is observed at dense grids of points $\{s_r : r = 1, \ldots, R\}$ in [0, 1] for all i and j, $\widetilde{\xi}_{W,ijk}$ is approximated accurately through numerical integration. It is easy to see that the version of $\widetilde{\xi}_{W,ijk}$ that uses $\mu(s, T_{ij})$ in place of $\hat{\mu}(s,T_{ij})$ and $\varphi_k(s)$ in place of $\hat{\phi}_k(s)$ converges to $\xi_{W,ijk}$ with probability one, as R diverges. The time-varying terms $\widetilde{\xi}_{W,ijk}$ are proxy measurements of $\xi_{ik}(T_{ij})$; they will be used to study the temporal dependence along the direction $\varphi_k(\cdot)$, $G_k(T, T') = \text{cov}\{\xi_{ik}(T), \xi_{ik}(T')\}$, and furthermore to obtain prediction for all times $T \in [0, 1]$.

Consider now $\left\{\left(T_{ij}, \widetilde{\xi}_{W,ijk}\right) : j=1, \ldots, m_i\right\}_i$ as the 'observed data'. One viable approach is to assume a parametric structure for $G_k(\cdot, \cdot)$ such as AR(1) or a random effects model framework; this is typically preferable when m_i is very small and the longitudinal design is balanced. We discuss random effects model for estimating the longitudinal covariance in the data application. Here we consider a more flexible approach and estimate the covariance $G_k(\cdot,\cdot)$ nonparametrically, by employing FPCA techniques for sparse functional data (Yao et al., 2005).

Let $\{\psi_{kl}(\cdot), \eta_{kl}\}_l$ be the pairs of eigenfunctions and eigenvalues of the covariance G_k; we model the proxy observations as $\widetilde{\xi}_{W,ijk} = \sum_{l=1}^{\infty} \zeta_{ikl}\psi_{kl}(T_{ij}) + \widetilde{e}_{ijk}$ where ζ_{ikl}'s are random variables with zero mean and variances equal to η_{kl}, $\widetilde{e}_{ijk}$'s are iid with zero-mean and variance equal to $\widetilde{\sigma}^2_{e,k}$ and independent of ζ_{ikl}. Following Yao et al. (2005), we first obtain the raw sample covariance, $\widetilde{G}_{ik}\left(T_{ij}, T_{ij'}\right) = \widetilde{\xi}_{W,ijk}\widetilde{\xi}_{W,ij'k}$. Then the estimated smooth covariance surface, $\hat{G}_k\left(T, T'\right)$, is obtained by using bivariate smoothing of $\left\{\left(T_{ij}, T_{ij'}\right), \widetilde{G}_{ik}\left(T_{ij}, T_{ij'}\right) : i, j \neq j'\right\}$. Kernel-based local linear smoothing (Yao et al., 2005) or penalized tensor product spline smoothing (Wood, 2006) can be used at this step. The diagonal terms $\left\{\widetilde{G}_{ik}\left(T_{ij}, T_{ij'}\right) : i, j=j'\right\}$ are removed because the noise $\widetilde{e}_{ijk}$ leads to inflated variance function. Let $\left\{\hat{\psi}_{kl}(\cdot), \hat{\eta}_{kl}\right\}_l$ be the pairs of eigenvalues/eigenfunctions of the estimated covariance surface, $\hat{G}_k\left(T, T'\right)$. The truncation value, L_k, is determined based on pre-specified PVE; using similar ideas as in Section 3.2. The variance $\widetilde{\sigma}^2_{e,k}$ is estimated as the average of the difference between a smooth estimate of the variance based on $\left\{T_{ij}, \widetilde{\xi}^2_{W,ijk}\right\}$ and $\hat{G}_k(T, T)$; Yao et al. (2005) discusses an alternative that dismisses the terms at the boundary when estimating the error variance.

Once the eigenbasis functions $\{\psi_{kl}(\cdot)\}_{l=1}^{L_k}$, eigenvalues η_{kl}'s, and error variance $\widetilde{\sigma}^2_{e,k}$ are estimated, the above model framework can be viewed as a mixed effects model and the random components ζ_{ikl} can be predicted using conditional expectation and a jointly Gaussian assumption for ζ_{ijk}'s and e_{ijk}'s. In particular, $\hat{\xi}_{ikl} = \mathsf{E}\left[\zeta_{ikl} | \widetilde{\boldsymbol{\xi}}_{W,ik}\right] = \hat{\eta}_{kl}\hat{\boldsymbol{\psi}}^T_{ikl}\hat{\Sigma}^{-1}_{\xi_{W,ik}}\widetilde{\xi}_{W,il}$, where $\hat{\boldsymbol{\psi}}_{ikl} = \left\{\hat{\psi}_{kl}(T_{i1}), \ldots, \hat{\psi}_{kl}(T_{im_i})\right\}^T$ is the m_i-dimensional column vector of the evaluations of $\hat{\psi}_{kl}(\cdot)$ at $\{T_{ij} : j = 1, \ldots, m_i\}$, $\hat{\Sigma}_{\xi_{W,ik}}$ is a $m_i \times m_i$ - matrix with (j, j')th element equal to $\hat{G}_k\left(T_{ij}, T_{ij'}\right) + \hat{\widetilde{\sigma}}^2_{e,k}$, for $j = j'$ and $\hat{G}_k\left(T_{ij}, T_{ij'}\right)$ otherwise, and $\widetilde{\boldsymbol{\xi}}_{W,ik}$ is the m_i - dimensional column vector of $\widetilde{\xi}_{W,ijk}$'s. The predicted time-varying coefficients corresponding to a generic time T are obtained as $\hat{\xi}_{ik}(T) = \sum_{l=1}^{L_k}\hat{\zeta}_{ikl}\hat{\psi}_{kl}(T)$. Yao et al. (2005) proved the consistency of the eigenfunctions and predicted trajectories when $\xi_{W,ijk}$'s are observed. In Section 4 we extend these results to the case when the proxy $\widetilde{\xi}_{W,ijk}$'s are

used instead and when the profiles $Y_{ij}(\cdot)$ are fully observed and the noise is of the type $\varepsilon_{ij}(s) = \varepsilon_{1,ij}(s)$; i.e. the data $Y_{ij}(\cdot)$ are observed with smooth error.

3.4. Step 4: Trajectories reconstruction

We are now able to predict the full response curve at any time point $T \in [0, 1]$ by:

$\hat{Y}_i(s,T) = \hat{\mu}(s,T) + \sum_{k=1}^{K} \hat{\xi}_{ik}(T)\,\hat{\phi}_k(s)$, where $s \in [0, 1]$. In Section 4 we show the consistency of $\hat{Y}_i(s,T)$.

4. Theoretical properties

Next we discuss the asymptotic properties of the estimators and the predicted trajectories. Our setting - sparse longitudinal design and dense functional design - requires new techniques than the ones commonly used for theoretical investigation of repeated functional data such as Chen & Müller (2012). Since the mean estimation has been studied previously, we assume that the response trajectories, $Y_{ij}(\cdot)$'s, have zero-mean and focus on the estimation of the model covariance. Throughout this section we assume that $Y_{ij}(\cdot)$ is observed fully as a function over the domain, $\mathscr{S} = [0, 1]$. Section 4.1 discusses the main theoretical results when data are observed without error, i.e. $\varepsilon_{ij}(s) \equiv 0$ for $s \in [0, 1]$. Section 4.2 extends the results to the case when the data are corrupted with smooth error process $\varepsilon_{ij}(s) \equiv \varepsilon_{1,ij}(s)$. The proofs are detailed in the Supplementary Material; also in the Supplementary Material we include a discussion on how to relax some of the assumptions. Throughout this section we use $\mathscr{S}$ and $\mathscr{T}$ to distinguish between the domains.

We assume that the bivariate process $X_i(s, T)$ is a realization of a true random process, $X(s, T)$, with zero-mean and smooth covariance function, $c((s, T),(s', T'))$, which satisfies some regularity conditions:

- **(A1.)** $X = \{X(s, T) : (s, T) \in \mathscr{S} \times \mathscr{T}$ is a square integrable element of the $L^2(\mathscr{S} \times \mathscr{T})$, i.e. $\mathsf{E}\left[\int\int X^2(s,T)\,dsdT\right] < \infty$, where $\mathscr{S}$ and $\mathscr{T}$ are compact sets.
- **(A2.)** The sampling density $g(T)$ is continuous and $\sup_{T \in \mathscr{T}} |g(T)| < \infty$.

Under (A1.) and (A2.), the function $\Sigma(s, s')$ defined above (i) is symmetric, (ii) is positive definite, and (iii) has eigenvalues λ_k's with $\sum_{k=1}^{\infty} \lambda_k < \infty$. Thus $\Sigma(\cdot, \cdot)$ is a *proper* covariance function (Horváth & Kokoszka, 2012, p.24).

4.1. Response curves measured without error

Assume $\varepsilon_{ij}(s) \equiv 0$ and thus $Y_{ij}(s) = X_i(s, T_{ij})$ for $s \in \mathscr{S}$. The sample covariance of $Y_{ij}(s)$ is $\hat{\Sigma}(s, s') = \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{ij}(s)\, Y_{ij}(s') \,/\, \left(\sum_{i=1}^{n} m_i\right)$. The following assumptions regard the moment behavior of X and are commonly used in functional data analysis (Yao et al., 2005; Chen & Müller, 2012); we require them in our study.

- **(A3.)** $\mathrm{E}[X(s, T)X(s', T)X(s, T')X(s', T')] < \infty$ for arbitrary $s, s' \in \mathscr{S}$ and $T, T' \in \mathscr{T}$.
- **(A4.)** $\mathrm{E}[\|X(\cdot, T)\|^4] < \infty$ for each $T \in \mathscr{T}$.

Theorem 1—*Assume (A1.) - (A3.) hold. Then* $|\hat{\Sigma}\left(s,s'\right)-\Sigma\left(s,s'\right)|\xrightarrow{p}0$ *as n diverges. If in addition (A4.) holds, then*

$$\|\hat{\Sigma}(\cdot,\cdot)-\Sigma(\cdot,\cdot)\|_s\xrightarrow{p}0\quad as\quad n\to\infty,\quad (4)$$

where $\|k(\cdot,\cdot)\|_s=\left\{\int\int k^2\left(s,s'\right)dsds'\right\}^{1/2}$ *is the Hilbert-Schmidt norm of k*(·,·).

(A5.) Let $a_1=(\lambda_1-\lambda_2)$ and $a_k=max\left[(\lambda_{k-1}-\lambda_k),(\lambda_k-\lambda_{k+1}))\right]$ for $k\geq 2$, where $_k$ is the kth largest eigenvalues of $\Sigma(s, s')$. Assume that $0<a_k<\infty$ and $\lambda_k>0$ for all k (no crossing or ties among eigenvalues).

Using Theorem 4.4 and Lemma 4.3 of Bosq (2000, p.104), the consistency result (4) implies that, if furthermore (A5.) holds, the eigen-elements of $\hat{\Sigma}\left(s,s'\right)$ are consistent estimators of the corresponding eigen-elements of $\Sigma(s, s')$.

Corollary 1—Under the assumptions (A1.)-(A5.), for each k we have $|\hat{\lambda}_k-\lambda_k|\xrightarrow{p}0$, *and* $\|\hat{\phi}_k(\cdot)-\phi_k(\cdot)\|_s\xrightarrow{p}0$ *as n diverges.*

Next, we focus on the estimation of the covariance $G_k(T, T')$, which describes the longitudinal dynamics. We first show the uniform consistency of $\hat{\xi}_{W,ijk}$; the result follows if $\sup_{j,s}|Y_i(s, T_{ij})|$ is bounded almost surely. which is ensured if (A6.) holds. Then, we use this result to show that the estimator of $G_k(T, T')$ based on $\hat{\xi}_{W,ijk}$'s is asymptotically identical to that based on $\xi_{W,ijk}$. Consistency results of the remaining model components follow directly from Yao et al. (2005). The Gaussian assumption (A8.) is needed to show the consistency of $\hat{\zeta}_{ikl}$.

- **(A6.)** $\mathrm{E}[\sup_{s,T}|X(s, T)|^a]\leq M^a$ for a constant, $M>0$, and an arbitrary integer, $a\geq 1$; This is equivalent to assume that $X(s, T)$ is absolutely bounded almost surely.
- **(A7.)** Let $b_{k1}=(\eta_{k1}-\eta_{k2})$ and $b_{kl}=\max[(\eta_{k(l-1)}-\eta_{kl}),(\eta_{kl}-\eta_{k(l+1)}))]$ for $l\geq 2$, where η_{kl} is the lth largest eigenvalues of $G_k(T, T')$. Assume that $0<b_{kl}<\infty$ and $\eta_{kl}>0$ for all k and l.
- **(A8.)** ζ_{ikl} and e_{ijk} are jointly Gaussian.

Theorem 2—*Under the assumptions (A1.) - (A6.), for each k* $sup_j|\widetilde{\xi}_{W,ijk}-\xi_{W,ijk}|\xrightarrow{p}0$ *and* $\|\hat{G}_k(\cdot,\cdot)-G_k(\cdot,\cdot)\|_s\xrightarrow{p}0$ as n diverges. In fact a stronger result also holds, namely $sup_{T,T'}|\hat{G}_k\left(T,T'\right)-G_k\left(T,T'\right)|\xrightarrow{p}0$ *as n diverges.*

Corollary 2—Assume (A1.) - (A8.) hold for each k and l. Then the eigenvalues $\hat{\eta}_{kl}$ *and eigenfunctions* $\hat{\psi}_{kl}(\cdot)$ *of* $\hat{G}_k(\cdot,\cdot)$ *satisfy* $|\hat{\eta}_{kl}-\eta_{kl}|\xrightarrow{p}0$, *and* $\|\hat{\psi}_{kl}(\cdot)-\psi_{kl}(\cdot)\|_s\xrightarrow{p}0$ *as n diverges. Uniform convergence of* $\hat{\psi}_{kl}(\cdot)$ *also holds:* $sup_T|\hat{\psi}_{kl}(T)-\psi_{kl}(T)|\xrightarrow{p}0$.

Furthermore, as n diverges, we have $|\widehat{\widetilde{\sigma}}^2_{e,k} - \sigma^2_{e,k}| \xrightarrow{p} 0$ *and* $|\hat{\zeta}_{ikl} - \widetilde{\zeta}_{ikl}| \xrightarrow{p} 0$, *where* $\widetilde{\zeta}_{ikl} = E\left[\zeta_{ikl} | \boldsymbol{\xi}_{W,ik}\right]$ *and* $\boldsymbol{\xi}_{W,ik}$ *is the* m_i-dimensional column vector of $\xi_{W,ijk}$'s.

The consistency results for all model components imply prediction consistency.

Thorem 3—*Assume (A1.) - (A8.), for each (s, T)* $\in \mathscr{S} \times \mathscr{T}$, Then $\hat{Y}_i(s,T) \xrightarrow{p} \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)$ as n, K and L_k's $\to \infty$.

4.2. Response curves measured with smooth error

Assume next that $Y_{ij}(s)$ are observed with smooth error $\varepsilon_{ij}(s) \equiv \varepsilon_{1,ij}(s)$ and thus $Y_{ij}(s) = X_i(s, T_{ij}) + \varepsilon_{1,ij}(s)$ for $s \in \mathscr{S}$ and $\varepsilon_{1,i}(\cdot) \in L^2(\mathscr{S})$. The main difference from Section 4.1 is that the sample covariance of $Y_{ij}(s)$ is an estimator of $\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')$, not of $(\Sigma s, s')$; we denote the sample covariance of $Y_{ij}(s)$ by $\hat{\Xi}\left(s, s'\right) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} Y_{ij}(s) Y_{ij}\left(s'\right) / \left(\sum_{i=1}^{n} m_i\right)$. Using similar arguments as earlier, we show that $\hat{\Xi}\left(s, s'\right)$ is an unbiased estimator of $\Xi(s, s')$. Moreover similar arguments can be used to show the pointwise consistency as well as the Hilbert-Schmidt norm consistency of $\hat{\Xi}\left(s, s'\right)$. Additional assumptions are required.

(A9.) Assume $\varepsilon_{ij}(\cdot)$ is realization of $\varepsilon = \{\varepsilon(s) : s \in \mathscr{S}\}$, which is square integrable process in $L^2(\mathscr{S})$.

(A10.) $\mathsf{E}\left[\|\epsilon(\cdot)\|^4\right] < \infty$

(A11.) $\mathsf{E}\left[sup_s |\epsilon(s)|^a\right] \leq M^a$ for a constant, $M > 0$, and an arbitrary integer, $a \geq 1$.

Corollary 3—Under the assumptions (A1.) - (A3.), and (A9.), for each(s, s') – $|\hat{\Xi}\left(s, s'\right) - \Xi\left(s, s'\right)| \xrightarrow{p} 0$ as n diverges. And under the assumptions (A1.)-(A4.), (A9.)-(A11.), $\|\hat{\Xi}(\cdot,\cdot) - \Xi(\cdot,\cdot)\|_s \xrightarrow{p} 0$ *and* $sup_j |\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0 \quad as \quad n \to \infty$.

The proofs of these results are detailed in the Supplementary Material. As the smooth error process $\varepsilon_{1,ij}(s)$ is correlated only along the functional argument, s, and $\varepsilon_{1,ij}(s)$ are iid over i, j, it follows that the theoretical properties of the predictions - of the time-varying coefficients and the response curve - hold without any modification.

The theoretical results are based on the assumptions that data $Y_{ij}(s)$'s are observed fully, without white noise, $\varepsilon_{2,ij}(s) \equiv 0$ for all s, and have mean zero. Some of these assumptions are quite common in theoretical study involving functional data; Cardot et al. (2003, 2004); Chen & Müller (2012). They are discussed in the Supplementary Material.

5. Simulation study

We study our approach in finite samples and compare its performance with Chen & Müller (2012) denoted by CM. We generate $N_{sim} = 1000$ samples from model (1) with $K = 2$, $Y_{ij}(s)$

$= \mu(s, T_{ij}) + \xi_{i1}(T_{ij})\varphi_1(s) + \xi_{i2}(T_{ij})\varphi_2(s) + \varepsilon_{ij}(s)$, where $\mu(s, T) = 1 + 2s + 3T + 4sT$, and $\varphi_1(s) = 1$ and $\varphi_2(s) = \sqrt{2}sin(2\beta s)$. The grid of points for s is the set of 101 equispaced points in [0, 1]. For each i, there are m_i profiles associated with visit times, $\{T_{ij} : j = 1, \ldots, m_i\}$; T_{ij}'s are randomly sampled from 41 equally spaced points in [0, 1]. $\xi_{ik}(T)$ are generated from various covariance structures: (a) non-parametric covariance (NP) where $\xi_{ik}(T) = \zeta_{ik1}\ \psi_{k1}(T) + \zeta_{ik2}\ \psi_{k2}(T)$; (b) random effects model (REM) $\xi_{ik}(T) = b_{ik0} + b_{ik1}T$, and (c) exponential autocorrelation (Exp) $cov\left\{_{,ik}(\mathsf{T}), \quad _{,ik}\left(\mathsf{T}'\right)\right\} = \check{\mathrm{k}}\text{æ}_{\mathrm{k}}^{|\mathsf{T}-\mathsf{T}'|}$. Errors are generated from $\varepsilon_{ij}(s) = e_{ij1}\varphi_1(s) + e_{ij2}\varphi_2(s) + \varepsilon_{2,ij}(s)$, where e_{ij1}, e_{ij2} and $\varepsilon_{2,ij}(s)$ are mutually independent with zero- mean and variances equal to $\sigma^2_{e,1}, \sigma^2_{e,2}$ and σ^2, respectively; the white noise variance, σ^2, is set based on the signal to noise ratio (SNR). The details of the models are specified in the Supplementary Material. For each sample of size n we form a training set and a test set. The test set contains 10 profiles and is obtained as follows: randomly select 10 subjects from the sample and collect the subjects' last profile. The remaining profiles for the 10 subjects and the data corresponding to the rest (n – 10) of the subjects form the training set. Our model is fitted using the training set and the methods of Section 3. The mean function, $\mu(s, T)$, is modeled using 50 cubic spline basis functions obtained from the tensor product of $d_S = 10$ basis functions in direction s and $d_T = 5$ in T . The smoothing parameters are selected via REML. The finite truncations K and L_k's are all estimated using the pre-specified level PVE = 0.95.

Estimation accuracy for the model components is evaluated using integrated mean squared errors (IMSE), while prediction performance is assessed through in-sample integrated prediction errors (IN-IPE) and out-of-sample IPE (OUT-IPE). Table 1 shows the results for different covariance models for $\xi_{ik}(T)$, different number of repeated curve measurements per subject, different SNRs, complex error process, and varying sample sizes. The performance of the proposed estimation (see columns for μ, φ_1, and φ_2 of this table) is slightly affected by the covariance structure of $\xi_{ik}(T)$'s and m_i, but in general is quite robust to the factors we investigated. As expected the estimation accuracy improves with larger sample size; see the 3 × 3 top left block of IMSE results corresponding to n = 100, n = 300, and n = 500. Moreover both the prediction of $\xi_{ik}(T)$'s and that of $Y_{ij}(\cdot)$ are considered; see columns labeled ξ_1, ξ_2, IN-IPE and OUT-IPE of Table 1. The underlying covariance structure of $\xi_{ik}(T)$'s affects the prediction accuracy. Furthermore increasing the number of repeated curve measurements m_i improves the accuracy more than increasing the sample size n. This observation should not be surprising, as with larger number of repeated measurements the estimation of the covariance of the longitudinal process $\xi_{ik}(T)$'s improves and as a result it yields superior prediction. We compared our results with another, rather naïve approach: predict a subject's profile by the average of all previously observed profiles for that subject. The naïve approach (see columns $\text{IN-IPE}_{\text{naive}}$ and $\text{OUT-IPE}_{\text{naive}}$) is very sensitive to the covariance structure of $\xi_{ik}(T)$. In all the cases studied the prediction accuracy is inferior to the proposed method.

Table 2 shows the comparison with CM, when the kernel bandwidth is fixed to h = 0.1 for both mean and covariance smoothing. The prediction using CM is more sensitive to the covariance structure of the underlying time-varying coefficients $\xi_{ik}(T)$ and its accuracy can

be improved by up to 50% using our proposed approach. Computation-wise, there is an order of magnitude difference in the computational cost between the methods: when $n = 100$ CM takes over 16 minutes, while our approach takes about 7 seconds. The overall conclusion is that the proposed approach provides an improved prediction performance over the existing methods in a computationally efficient manner.

6. DTI application

DTI is a magnetic resonance imaging technique, which provides different measures of water diffusivity along brain white matter tracts; its use is instrumental especially in diseases that affect the brain white matter tissue, such as MS (see Alexander et al. (2007), Basser et al. (1994), Basser et al. (2000), Basser & Pierpaoli (2011)). In this paper we consider the DTI measure called FA along CCA; specifically we consider one-dimensional summaries of FA along CCA (CCA-FA). The DTI study involves 162 MS patients, which are observed at between one and eight hospital visits, with a total of 421 visits and a median of two visits per subject. At each visit, FA profile is recorded at 93 locations along the CCA. The measurements are registered within and between subjects using standard biological landmarks identified by an experienced neuroradiologist (Scheipl et al., 2014).

Our main objective is twofold: (i) to understand the dynamic behavior of the CCA-FA profile in MS patients over time and (ii) to make accurate predictions of the CCA-FA profile of a patient at their next visit. Various aspects of the DTI study have been also considered in Goldsmith et al. (2011), Staicu et al. (2012), Pomann et al. (2013), and Scheipl et al. (2014). Greven et al. (2010) used an earlier version of the DTI study consisting of data from fewer and possibly different patients and obtained through a different registration technique. They studied the dynamic behavior of CCA-FA over time in MS; however, their method cannot provide prediction of the entire CCA-FA profile at the subject's next visit. By being able to predict the full CCA-FA profile at the subject's future visit, our approach has the potential to shed lights on the understanding of the MS progression over time as well as its response to treatment.

To start with, for each subject we define the hospital visit time T_{ij} by the difference between the reported visit time and the subject's baseline visit time; thus $T_{i1} = 0$ for all subjects i. Then the resulting values are scaled by the maximum value in the study so that $T_{ij} \in [0, 1]$ for all i and j. The sampling distribution of the visit times is right-skewed with rather strong skewness; for example there are only few observations T_{ij}'s close to 1. The strong skewness of the sampling distribution of T_{ij}'s has serious implications on the estimation of the bivariate mean $\mu(s, T)$; a completely nonparametric bivariate smoothing would results in unstable and highly variable estimation. This is probably why Greven et al. (2010) first centered the times for each patient i, $\{T_{ij} : j = 1, \ldots, m_i\}$, and then standardized the overall set $\{T_{ij} : i, j\}$ to have unit variance. However, such subject-specific transformation of T_{ij}'s loses interpretability and it is not suited for prediction at unobserved times - which is crucial in our analysis. One way to bypass this issue is to assume a simpler parametric structure along the longitudinal direction, T, for the mean function; based on exploratory analysis we assume linearity in T . Specifically we consider $\mu(s, T_{ij}) = \mu_0(s) + \beta_T(s)T_{ij}$, where $\mu_0(\cdot)$ and $\beta_T(\cdot)$ are unknown, smooth functions of s. We estimate $\mu_0(\cdot)$ and $\beta_T(\cdot)$ using a penalized

univariate cubic spline regression with 10 basis functions; the smoothing parameters are estimated using REML. The estimates $\hat{\mu}(s,T)$ and $\hat{\beta}_T(s)$ are displayed in Figure S1 of the Supplementary Material. Using the bootstrap of subjects - based methods of Park et al. (2015) and $B = 1000$ bootstrap samples we construct 95% joint confidence bands for $\hat{\beta}_T(s)$; see Figure 1. The confidence band contains zero for all s, indicating evidence that a mean model $\mu(s, T_{ij}) = \mu_0(s)$ is more appropriate.

Next we demean the data and estimate the marginal covariance; using a preset level $PVE = 0.95$ we obtain $K = 10$ eigenfunctions. Figure 2 shows the leading 3 eigenfunctions that explain in turn 62.69%, 8.37% and 6.77% of the total variance; the rest of the estimated eigenfunctions are given in Figure S3 of the Supplementary Material. Preliminary investigation (not shown here) indicates a simpler model for the longitudinal covariance: a random effects model $\xi_{ik}(T_{ij}) = b_{0ik} + b_{1ik}T_{ij}$, where $\text{var}(b_{lik}) = \text{œ}^2_{lk}$ for $l = 0, 1$ and $\text{cov}(b_{0ik}, b_{1ik}) = \text{œ}_{01k}$. This resulting model is similar to Greven et al. (2010). The fitted time-varying coefficient functions, $\hat{\xi}_{ik}(T)$, for $k = 1, 2$ and 3 are shown in Figure 3, and the rest are shown in Figure S4 of the Supplementary Material. The estimated $\hat{\xi}_{i1}(T)$ suggest some longitudinal changes, but the signs generally remain constant across time. The results imply that a subject mean profile tends to stay lower than the population mean, if the first eigenfunction corresponding to that individual is positively loaded at baseline, and vise versa. In contrast, $\hat{\xi}_{i2}(T)$, are mostly constant across visit times and imply little changes over time.

Finally, we assess the goodness-of-fit and prediction accuracy of our final model. For the goodness-of-fit we use the in-sample integrated prediction error (IN-IPE): IN-IPE= $\sum_{i=1}^{162}\sum_{j=1}^{m_i}\int\left\{Y_{ij}(s)-\hat{Y}_{ij}(s)\right\}^2 ds/\left\{\sum_{i=1}^{162}m_i\right\}$, where $\hat{Y}_{ij}(s)=\hat{\mu}_0(s)+\sum_{k=1}^{K}\left(\hat{b}_{0ik}+\hat{b}_{1ik}T_{ij}\right)\hat{\phi}_k(s)$, and $Y_{ij}(\cdot)$'s are the observed curve data. The square root of the IN-IPE is 2.31×10^{-2} for our model; for comparison Greven et al. (2010) yields 2.66×10^{-2} and Chen & Müller (2012) gives 3.76×10^{-2}. For prediction accuracy we use leave-the last-curve-out integrated prediction error (OUT-IPE) calculated for the 106 subjects observed at two hospital visits or more: OUT-IPE = $\sum_{i=1}^{106}\int\left\{Y_{im_i}(s)-\hat{Y}_{im_i}^{[-im_i]}(s)\right\}^2 ds/106$, where $\hat{Y}_{im_i}^{[-im_i]}(s)$ is the predicted curve at time T_{im_i} for the ith subject using the fitted model based on all the data less the m_ith curve of the ith subject. Figure 4 shows such predicted curves $\hat{Y}_{im_i}^{[-im_i]}(s)$ obtained using our model and the naive model for three randomly selected subjects at their last visit. The square root of OUT-IPE is 3.48×10^{-2} for our model; for comparison Chen & Müller (2012) gives 8.71×10^{-2} and the naïve approach gives 3.52×10^{-2}. These results suggest that, in this short term study of MS, there is a small variation of CCA-FA profiles over time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Staicu's research was supported by NSF grant number DMS 1454942 and NIH grant R01 NS085211. We thank Daniel Reich and Peter Calabresi for the DTI tractography data.

References

Alexander AL, Lee JE, Lazar M, Field AS. Diffusion tensor imaging of the brain. Neurotherapeutics. 2007; 4(3):316–329. [PubMed: 17599699]

Baladandayuthapani V, Mallick BK, Young Hong M, Lupton JR, Turner ND, Carroll RJ. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. Biometrics. 2008; 64(1):64–73. [PubMed: 17608780]

Basser PJ, Mattiello J, LeBihan D. Mr diffusion tensor spectroscopy and imaging. Biophysical journal. 1994; 66(1):259. [PubMed: 8130344]

Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A. In vivo fiber tractography using dt-mri data. Magnetic resonance in medicine. 2000; 44(4):625–632. [PubMed: 11025519]

Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. Journal of magnetic resonance. 2011; 213(2):560–570. [PubMed: 22152371]

Bosq, D. Linear processes in function spaces: theory and applications. Vol. 149. Springer; 2000.

Cardot H, Ferraty F, Mas A, Sarda P. Testing hypotheses in the functional linear model. Scandinavian Journal of Statistics. 2003; 30(1):241–255.

Cardot H, Goia A, Sarda P. Testing for no effect in functional linear regression models, some computational approaches. Communications in Statistics-Simulation and Computation. 2004; 33(1): 179–199.

Chen K, Delicado P, Müller HG. Modeling function-valued stochastic processes, with applications to fertility dynamics. 2015 Manuscript submitted.

Chen K, Müller HG. Modeling repeated functional observations. Journal of the American Statistical Association. 2012; 107(500):1599–1609. doi:10.1080/01621459.2012.734196.

Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel functional principal component analysis. The annals of applied statistics. 2009; 3(1):458. [PubMed: 20221415]

Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D. Penalized functional regression. Journal of Computational and Graphical Statistics. 2011; 20(4)

Goldsmith J, Zipunnikov V, Schrack J. Generalized multilevel functional-on-scalar regression and principal component analysis. 2014

Greven S, Crainiceanu C, Caffo B, Reich D. Longitudinal functional principal component analysis. Electronic Journal of Statistics. 2010:1022–1054. doi:10.1214/10-EJS575. [PubMed: 21743825]

Gromenko O, Kokoszka P. Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination. Computational Statistics & Data Analysis. 2013; 59:82–94.

Gromenko O, Kokoszka P, Zhu L, Sojka J. Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. The Annals of Applied Statistics. 2012; 6(2): 669–696. doi: 10.1214/11-AOAS524.

Hastie, T.; Tibshirani, R.; Friedman, J.; Hastie, T.; Friedman, J.; Tibshirani, R. The elements of statistical learning. Vol. 2. Springer; 2009.

Horváth, L.; Kokoszka, P. Inference for functional data with applications. Vol. 200. Springer; 2012.

Jiang CR, Wang JL. Covariate adjusted functional principal components analysis for longitudinal data. The Annals of Statistics. 2010:1194–1226.

Li Y, Guan Y. Functional principal component analysis of spatio-temporal point processes with applications in disease surveillance. Journal of the American Statistical Association. 2014; 0(ja) null, doi:10.1080/01621459.2014. 885434.

Marx BD, Eilers PH. Multidimensional penalized signal regression. Technometrics. 2005; 47(1):13–22.

Morris JS, Carroll RJ. Wavelet-based functional mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2006; 68(2):179–199. [PubMed: 19759841]

Morris JS, Vannucci M, Brown PJ, Carroll RJ. Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. Journal of the American Statistical Association. 2003; 98(463): 573–583.

Park SY, Staicu AM, Xiao L, Crainiceanu CM. Simple fixed effects inference for complex functional models. 2015 Manuscript submitted.

Pomann GM, Staicu AM, Ghosh S. Two sample hypothesis testing for functional data. 2013

Scheipl F, Staicu AM, Greven S. Functional additive mixed models. Journal of Computational and Graphical Statistics. 2014:00–00. just-accepted.

Staicu AM, Crainiceanu CM, Carroll RJ. Fast methods for spatially correlated multilevel functional data. Biostatistics. 2010; 11(2):177–194. [PubMed: 20089508]

Staicu AM, Crainiceanu CM, Reich DS, Ruppert D. Modeling functional data with spatially heterogeneous shape characteristics. Biometrics. 2012; 68(2):331–343. [PubMed: 22050118]

Staniswalis JG, Lee JJ. Nonparametric regression analysis of longitudinal data. Journal of the American Statistical Association. 1998; 93(444):1403–1418.

Wood SN. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. Biometrics. 2006; 62(4):1025–1036. doi:10.1111/j.1541-0420.2006.00574.x. [PubMed: 17156276]

Xiao L, Huang L, Schrack JA, Ferrucci L, Zipunnikov V, Crainiceanu CM. Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. Biostatistics. 2015; 16(2):352–367. [PubMed: 25361695]

Xiao L, Li Y, Ruppert D. Fast bivariate p-splines: the sandwich smoother. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2013; 75(3):577–599.

Xiao L, Ruppert D, Zipunnikov V, Crainiceanu C. Fast Covariance Estimation for High-dimensional Functional Data. Statistics and Computing. 2015 to appear.

Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association. 2005; 100(470):577–590.

Zhou L, Huang JZ, Carroll RJ. Joint modelling of paired sparse functional data using principal components. Biometrika. 2008; 95(3):601–619. [PubMed: 19396364]

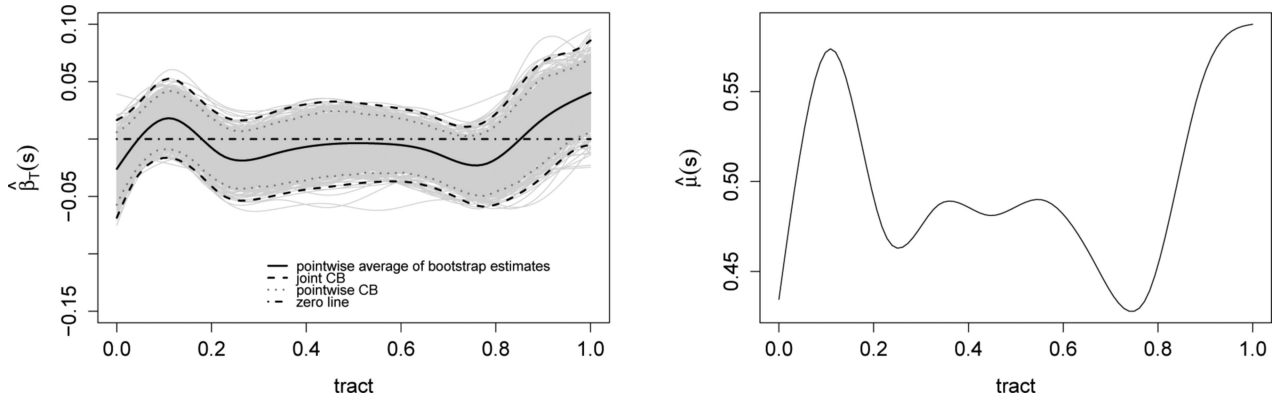

Figure 1.

Left panel: 95% pointwise and joint confidence bands of the slope function $\beta_T(s)$ of $\mu(s, T)$ using bootstrap; Right: final mean estimate, $\hat{\mu}(s,T) = \hat{\mu}_0(s)$

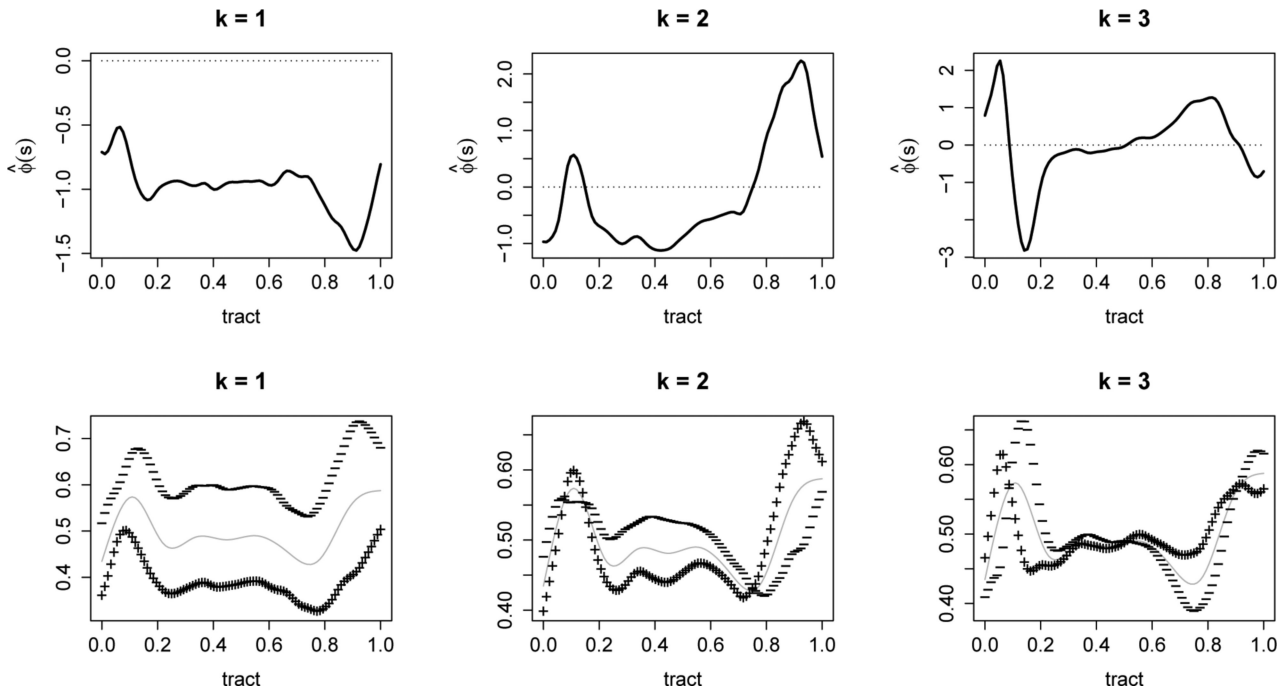

Figure 2.

Top: First three eigenfunctions of the estimated marginal covariance; Bottom: estimated mean function $\hat{\mu}_0(s)$ (gray line) $\pm 2\sqrt{\hat{\lambda}_k}\hat{\phi}_k(s)$ (+ and − signs, respectively)

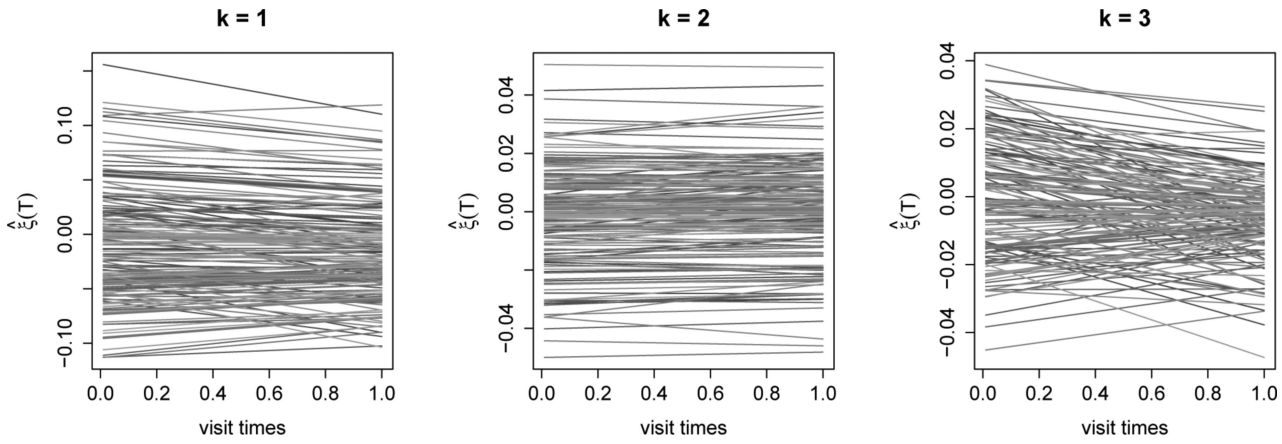

Figure 3.

Estimated time-varying coefficients $\hat{\xi}_{ik}(T)$ for k = 1, 2 and 3 using REM

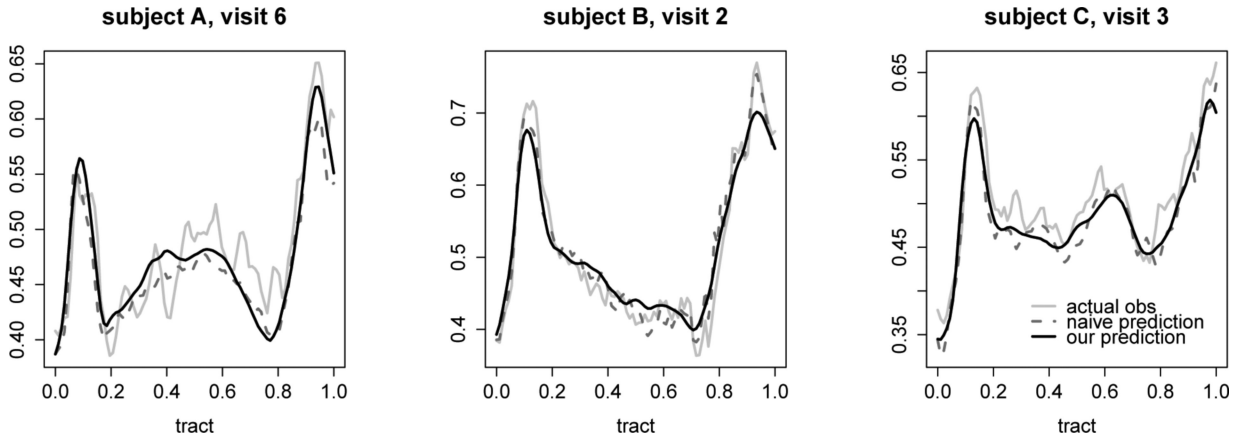

Figure 4.

Predicted values of FA for the last visits of three randomly selected subjects; actual observations (gray); predictions using our model (black solid) and using the naive approach (black dashed)

Table 1

Estimation and prediction accuracy results based on N_{sim} = 1000 simulations

		μ	φ_1	φ_2	ξ_1	ξ_2	IN-IPE	IN-IPE$_{naive}$	OUT-IPE	OUT-IPE$_{naive}$
iid $m_i \sim$ {8, …, 12} and SNR = 1										
NP (a)	n = 100	0.092	0.003	0.011	0.338	0.224	0.406	7.790	0.988	11.478
	n = 300	0.031	0.001	0.009	0.226	0.138	0.313	7.773	0.559	11.349
	n = 500	0.019	0.001	0.009	0.199	0.117	0.288	7.779	0.455	11.262
REM (b)	n = 100	0.114	0.027	0.033	0.376	0.314	0.328	1.199	1.011	2.160
	n = 300	0.040	0.008	0.013	0.216	0.162	0.265	1.197	0.675	2.160
	n = 500	0.024	0.005	0.010	0.181	0.133	0.247	1.197	0.571	2.150
Exp (c)	n = 100	0.095	0.022	0.030	0.399	0.540	0.554	1.528	1.426	2.520
	n = 300	0.031	0.007	0.015	0.289	0.412	0.508	1.531	1.143	2.498
	n = 500	0.019	0.004	0.013	0.266	0.383	0.494	1.530	1.074	2.492
iid $m_i \sim$ {15, …, 20} and SNR = 1										
NP (a)	n = 100	0.076	0.002	0.010	0.180	0.101	0.238	7.807	0.477	10.666
	n = 300	0.026	< 0. 001	0.009	0.120	0.065	0.183	7.796	0.282	10.728
	n = 500	0.016	< 0. 001	0.009	0.108	0.058	0.173	7.797	0.242	10.772
REM (b)	n = 100	0.097	0.025	0.031	0.272	0.252	0.232	0.897	0.612	1.833
	n = 300	0.034	0.008	0.013	0.156	0.132	0.201	0.896	0.462	1.841
	n = 500	0.020	0.005	0.010	0.135	0.110	0.194	0.897	0.440	1.836
Exp (c)	n = 100	0.080	0.022	0.030	0.308	0.417	0.467	1.240	1.048	2.147
	n = 300	0.026	0.006	0.015	0.233	0.309	0.444	1.245	0.938	2.155
	n = 500	0.016	0.004	0.012	0.221	0.285	0.438	1.246	0.886	2.129

$m_i \sim$ iid {8, …, 12} and SNR = 5										
		μ	φ_1	φ_2	ξ_1	ξ_2	IN-IPE	IN-IPE$_{naive}$	OUT-IPE	OUT-IPE$_{naive}$
NP (a)	n = 100	0.092	0.005	0.005	0.328	0.213	0.363	7.184	0.958	10.795
	n = 300	0.031	0.001	0.002	0.213	0.124	0.268	7.170	0.506	10.662
	n = 500	0.019	0.001	0.001	0.187	0.103	0.242	7.178	0.402	10.585
REM (b)	n = 100	0.114	0.037	0.037	0.404	0.355	0.293	0.594	0.958	1.478
	n = 300	0.040	0.010	0.011	0.218	0.167	0.235	0.595	0.627	1.476
	n = 500	0.024	0.006	0.007	0.180	0.135	0.219	0.596	0.529	1.467
Exp (c)	n = 100	0.095	0.033	0.033	0.420	0.573	0.513	0.922	1.419	1.838
	n = 300	0.031	0.010	0.010	0.290	0.412	0.466	0.929	1.109	1.814
	n = 500	0.019	0.006	0.006	0.264	0.378	0.453	0.929	1.033	1.807

$m_i \sim$ iid {15, …, 20} and SNR = 5										
		μ	φ_1	φ_2	ξ_1	ξ_2	IN-IPE	IN-IPE$_{naive}$	OUT-IPE	OUT-IPE$_{naive}$
NP (a)	n = 100	0.076	0.003	0.003	0.174	0.095	0.205	7.462	0.441	10.300
	n = 300	0.026	0.001	0.001	0.113	0.057	0.147	7.453	0.239	10.359
	n = 500	0.016	< 0. 001	0.001	0.101	0.050	0.136	7.454	0.200	10.406
REM (b)	n = 100	0.097	0.035	0.035	0.300	0.293	0.205	0.552	0.568	1.464
	n = 300	0.034	0.010	0.010	0.160	0.140	0.178	0.552	0.426	1.473
	n = 500	0.020	0.006	0.007	0.136	0.114	0.172	0.554	0.405	1.470
Exp (c)	n = 100	0.080	0.033	0.033	0.330	0.451	0.434	0.895	1.012	1.779
	n = 300	0.027	0.009	0.010	0.236	0.313	0.410	0.902	0.901	1.785
	n = 500	0.016	0.005	0.006	0.221	0.284	0.403	0.902	0.851	1.763

Table 2

Comparison between the proposed method and Chen & Muller (2012) in the presence of correlated errors. Results based on N_{sim} = 1000 simulations

$m_i \overset{iid}{\sim} \{8, \ldots, 12\}$ and SNR = 1							
		Chen & Muller (2012)			Proposed method (from Tables 1 and S2)		
		IN-IPE	OUT-IPE	time (seconds)	IN-IPE	OUT-IPE	time (seconds)
NP (a)	n = 100	0.880	2.221	983.872	0.406	0.988	7.369
	n = 300	0.622	1.468	1659.611	0.313	0.559	15.892
	n = 500	0.556	1.298	2502.462	0.288	0.455	21.418
REM (b)	n = 100	0.424	1.359	1084.753	0.328	1.011	9.282
	n = 300	0.289	0.729	1955.193	0.265	0.675	11.347
	n = 500	0.257	0.614	2947.126	0.247	0.571	22.559
Exp (c)	n = 100	0.634	1.642	1556.182	0.554	1.426	7.514
	n = 300	0.549	1.251	1959.219	0.508	1.143	16.229
	n = 500	0.531	1.155	2865.041	0.494	1.074	17.109