Longitudinal functional data analysis

So Young Park and Ana-Maria Staicu*

Received 9 July 2015; Accepted 16 July 2015

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 modelling 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 behaviour 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 behaviour in finite samples. The proposed method is motivated by and applied to a diffusion tensor imaging study of multiple sclerosis. Copyright © 2015 John Wiley & Sons, Ltd.

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, 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 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 modelling 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 $Y_{ij}()$ 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 modelling 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 modelling 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 \textit{sparse longitudinal design}) and the subject profiles are observed at fine grids (hence \textit{dense functional design}). We propose to model $Y_{ij}(\cdot)$ as

$$Y_{ij}(s) = \mu(s, T_{ij}) + X_i(s, T_{ij}) + \epsilon_{ij}(s); \quad X_i(s, T_{ij}) = \sum_{k \geq 1} \xi_{ik}(T_{ij})\phi_k(s)$$

for $s \in S$ and $T_{ij} \in T$ (1)

where $S$ and $T$ 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 $\epsilon_{ij}(\cdot)$ 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 $\epsilon_{ij}$'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 $\epsilon_{ij}$. Here, $\{\phi_k(\cdot)\}_k$ is an orthogonal basis in $L^2(S)$, and the $\xi_{ik}(T_{ij})$'s are the corresponding basis coefficients that have zero mean and 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 $T$. 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 $\xi_{ik}(T_{ij}) = \xi_{0,ik} + T_{ij}\xi_{1,ik}$ for appropriately defined random terms $\xi_{0,ik}$ and $\xi_{1,ik}$, 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 $\lambda_k$, known correlation function $\rho_k(\cdot; \nu)$ with unknown parameter $\nu$, and $n = 1$, model (1) resembles to Gromenko et al. (2012) and Gromenko & Kokoszka (2013) for spatially indexed functional data. (iii) If $\xi_{ik}(T_{ij}) = \sum_{l \geq 1} \xi_{ikl}\psi_{ikl}(T_{ij})$ with orthogonal basis functions $\psi_{ikl}(T)$'s and the corresponding coefficients $\xi_{ikl}$'s, then model (1) is similar to Chen & Müller (2012) who used time-varying basis functions $\phi_k(\cdot|T)$ instead of our proposed $\phi_k(\cdot)$ in model (1) and assumed a white noise residual process $\epsilon_{ij}$.

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 $\{\phi_k(\cdot|T)\}_k$ 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 non-trivial. One option is to use a pre-specified basis; Zhou et al. (2008) considered this approach in modelling 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 $\{\phi_k(\cdot)\}_k$ using an appropriate \textit{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 $\Sigma(s, s') = \int_T c((s, T), (s', T'))g(T)dT$ for $s, s' \in S$: we show that this bivariate function is a proper covariance function (Horváth & Kokoszka, 2012). Section 2 shows that the proposed basis $\{\phi_k(\cdot)\}_k$ has optimal properties with respect to some appropriately defined criterion. From this viewpoint, 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. (2015), who considered independent functional data.
The rest of paper is organized as follows. Section 2 introduces the proposed modelling 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 Modelling longitudinal functional data

Let \{T_{ij}, Y_{ij}(s_j) : r = 1, \ldots, R; j = 1, \ldots, m_r\} be the observed data for the \(r\)th subject, where \(Y_{ij}(t)\) is the \(r\)th 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, \(m_r\), is relatively small to moderate and the set of time points of all subjects, \(\{T_{ij} : for all i,j\}\), is dense in \(T\). Without loss of generality, we set \(S = T = [0, 1]\).

We model the response \(Y_{ij}(\cdot)\) using (1), where we assume that \(e_{ij}(s)\) is the sum of independent components: \(e_{ij}(s) = e_{1,ij}(s) + e_{2,ij}(s)\). Here, \(e_{1,ij}(\cdot)\) is a random square integrable function, which has smooth covariance function \(\Gamma(\cdot, \cdot')\) with covariance \(\text{cov}(e_{1,ij}(s), e_{1,ij}(s')) = \sigma^2 e\) if \(s = s'\) and 0 otherwise.

Let \(c((s, T), (s', T')) = E[X(s, T)X(s', T')]\) be the covariance function of the process \(X(\cdot, \cdot)\) 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); because of its definition, we call \(\Sigma\) as the marginal covariance function induced by \(X\). The unpublished work of Chen et al. (2015) independently considered a similar marginal covariance in a related setting. Let \(W(s, T_{ij}) = X(s, T_{ij}) + e_{ij}(s)\), \(W_t\) 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 \(\{\phi_k(s)\}; l_k\) be the eigenfunctions of \(\Xi(s, s')\), where \(\{\phi_k(\cdot) : k\}\) forms an orthogonal basis in \(L^2[0, 1]\) and \(\lambda_1 \geq \lambda_2 \geq \cdots \geq 0\). Using arguments similar to the standard functional principal component analysis, the eigenbasis functions \(\{\phi_k(\cdot) : k = 1, \ldots, K\}\) are optimal in the sense that they minimize the following weighted mean square error: \(\text{MSE}(\theta_1(\cdot), \ldots, \theta_K(\cdot)) = \int_0^1 \sum_{k=1}^K \lambda_k \theta_k(s) > \theta_k(s) ||^2 g(T)dT\), where \(<f_1, f_2> = \int_0^1 f_1(s)f_2(s)ds\) is the usual inner product in \(L^2[0, 1]\).

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

One way to model the dependence of the coefficients, \(\xi_k(T_{ij})\)'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) and Gromenko & Kokoszka (2013). We consider this approach in the analysis of the DTI data, Section 6. Another approach is to assume a non-parametric covariance structure and employ a common functional data analysis technique. We detail the latter approach in this section.

For each \(k\), denote by \(G_k(T, T') = \text{cov}(\xi_k(T), \xi_k(T'))\) the covariance function of \(\xi_k(T)\) and \(\xi_k(T')\); \(G_k(T, T')\) is a smooth bivariate function defined on \([0, 1] \times [0, 1]\). Mercer’s theorem provides the following convenient spectral decomposition \(G_k(T, T') = \sum_{k \geq 1} \eta_k \psi_k(T) \psi_k(T')\), where \(\eta_1 \geq \eta_2 \geq \cdots \geq 0\) and \(\{\psi_k(\cdot)\}\) is an orthogonal basis in \(L^2[0, 1]\). Using the Karhunen–Loève expansion, we represent \(\xi_k(\cdot)\) as \(\xi_k(T_{ij}) = \sum_{k=1}^{\infty} \xi_k(T_{ij}) \psi_k(T_{ij})\), where \(\xi_k(\cdot) = \int \xi_k(T) \psi_k(T)dT\), have zero mean, variance equal to \(\eta_k\) and are uncorrelated over \(i\). By collecting all the components, we represent the model (1) as \(Y_{ij}(s) = \mu(s, T_{ij}) + \sum_{k=1}^{\infty} \xi_k(T_{ij}) \psi_k(T_{ij}) \phi_k(s) + e_{ij}(s)\), for \(e_{ij}(s) = \sum_{k=1}^{\infty} e_{ij}(s) \phi_k(s) + e_{2,ij}(s)\). In practice, we truncate this expansion. Let \(K\) 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 \(\sum L_k\) respective basis functions.
\[ Y_i(s, T_{ij}) = \mu(s, T_{ij}) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} \xi_{ikl}(T_{ij}) \phi_k(s) + \epsilon_{ij}(s), \]  

(2)

where \( \epsilon_{ij}(s) \approx \sum_{k=1}^{K} e_{ijk} \phi_k(s) + e_{2,ij}(s) \). 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 \( \xi_{ik}() \).

### 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., 2015); here, we briefly describe it for completeness. Our focus and novelty is the estimation of the marginal covariance function and of the eigenfunctions \( \phi_k() \)'s (Section 3.2), as well as the estimation of the time-varying basis coefficients \( \xi_{ik}() \)'s (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. (2015), we estimate the mean function, \( \mu(s, T) \), using bivariate smoothing via bivariate tensor product splines (Wood, 2006) of the pooled data \( \hat{Y}_{ijr} = Y_{ijr}(s_r) \)'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 \( \mu(s, T) = \sum_{q_1=1}^{d_s} \sum_{q_2=1}^{d_T} B_{s,q_1}(s) B_{T,q_2}(T) \beta_{q_1q_2} = B(s, T)^\top \beta \), where \( B(s, T) \) is the known \( d_s, d_T \)-dimensional vector of \( B_{s,q_1}(s) B_{T,q_2}(T) \)'s, and \( \beta \) is the vector of unknown parameters, \( \beta_{q_1q_2} \)'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, that is, \( \int \int \{ \partial^2 \mu(s, T)/\partial s^2 \}^2 dTds = \beta^\top (P_s \otimes I_{d_T}) \beta \) in direction \( s \) and \( \int \int \{ \partial^2 \mu(s, T)/\partial T^2 \}^2 dTds = \beta^\top (I_{d_s} \otimes P_T) \beta \) in \( T \). The penalized criterion to be minimized is

\[ \sum_{i,j,r} \left[ \hat{Y}_{ijr} - B(s_r, T_{ij})^\top \beta \right]^2 + \beta^\top (\lambda_s P_s \otimes I_{d_T} + \lambda_T I_{d_s} \otimes P_T) \beta, \]

where \( \lambda_s \) and \( \lambda_T \) 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. The estimated mean function is

\( \hat{\mu}(s, T) = B(s, T)^\top \hat{\beta} \). 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: Data-based orthogonal basis

Once the mean function is estimated, let \( \bar{Y}_{ijr} = Y_{ijr} - \hat{\mu}(s_r, T_{ij}) \) 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 \( \hat{\Xi}(s, s') \) is obtained; the pooled sample covariance is a suitable choice, if all the curves are observed on the same grid of points:

\[ \hat{\Xi}(s_r, s'_{r'}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \bar{Y}_{ijr} \bar{Y}_{ijr'}/ \left( \sum_{i=1}^{n} m_i \right). \]  

(3)
As data $Y_{ij}$’s are observed with white noise, $\varepsilon_{2,ij}(s)$, the “diagonal” elements of the sample covariance, $\hat{\Sigma}(s, s)$, are inflated by the variance of the noise, $\sigma^2$. 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 zeroed. The resulting smoothed covariance function, $\hat{\Sigma}(s, s')$, is used as an estimator of $\Sigma(s, s')$. In Section 4, we show that $\hat{\Sigma}(s, s')$ is an unbiased and consistent estimator of $\Sigma(s, s')$ in two settings: (1) the data are observed fully and without noise, that is, $\varepsilon_{ij}(s) \equiv 0$ and (2) the data are observed fully and with measurement error of type $\varepsilon_{1,ij}(s)$, that is, $\varepsilon_{ij}(s) \equiv \varepsilon_{1,ij}(s)$.

Let $\hat{\Phi}(s), \hat{\lambda}_k$ be the pairs of eigenvalues/eigenfunctions obtained from the sample decomposition of the estimated covariance function, $\hat{\Sigma}(s, s')$. The truncation value $K$ is determined based on prespecified 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 prespecified PVE (Di et al., 2009; Staicu et al., 2010).

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

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

Consider now $\{(T_{ij}, \hat{\xi}_{W,ijk}) : j = 1, \ldots, m_i\}$ 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 $G_k(\cdot, \cdot)$ non-parametrically, by employing functional principal component analysis techniques for sparse functional data (Yao et al., 2005).

Let $\{(\psi_{kl}(\cdot), \eta_{kl})\}$ be the pairs of eigenfunctions and eigenvalues of the covariance $\Sigma_{ijkl}$; we model the proxy observations as $\tilde{\xi}_{W,ijkl} = \sum_{i=1}^{\infty} \eta_{ijkl} \psi_{kl}(T_{ij}) + \tilde{e}_{ijkl}$ where $\xi_{ijkl}$’s are random variables with zero mean and variances equal to $\eta_{ijkl}$, $\tilde{e}_{ijkl}$’s are iid with zero mean and variance equal to $\tilde{\sigma}_{ijkl}^2$ and independent of $\eta_{ijkl}$. Following Yao et al. (2005), we first obtain the raw sample covariance, $\tilde{G}_{ik}(T_{ij}, T_{ij}') = \tilde{\xi}_{W,ijkl} \tilde{\xi}_{W,ijkl}'$. Then the estimated smooth covariance surface, $\hat{G}_k(T, T')$, is obtained by using bivariate smoothing of $\{(T_{ij}, \tilde{\xi}_{W,ijkl}) : i, j = 1, \ldots, m_i\}$. 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 $\{\tilde{G}_k(T_{ij}, T_{ij}'): i, j = 1, \ldots, m_i\}$ are removed because the noise $\tilde{e}_{ijkl}$ leads to inflated variance function. Let $\{(\hat{\psi}_{kl}(\cdot), \hat{\eta}_{kl})\}$ be the pairs of eigenvalues/eigenfunctions of the estimated covariance surface, $\hat{G}_k(T, T')$. The truncation value, $L_k$, is determined based on prespecified PVE, using similar ideas as in Section 3.2. The variance $\hat{\sigma}_{ijkl}^2$ is estimated as the average of the difference between a smooth estimate of the variance based on $\{(T_{ij}, \tilde{\xi}_{W,ijkl})\}$ 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 $\{(\hat{\psi}_{kl}(\cdot))\}_{l=1}^{L_k}$, eigenvalues $\hat{\eta}_{kl}$’s and error variance $\hat{\sigma}_{ijkl}^2$ are estimated, the aforementioned model framework can be viewed as a mixed effects model, and the random components $\xi_{ijkl}$ can be predicted using conditional expectation and a jointly Gaussian assumption for $\hat{\xi}_{ijkl}$’s and $\tilde{e}_{ijkl}$’s. In particular, $\hat{\xi}_{ijkl} = \mathbb{E}[\xi_{ijkl}|\tilde{\xi}_{W,ijkl}] = \hat{\eta}_{ijkl} \hat{\psi}_{kl}(\cdot) \tilde{\xi}_{W,ijkl}$, where $\hat{\psi}_{kl} = \{\hat{\psi}_{kl}(T_{ij}), \ldots, \hat{\psi}_{kl}(T_{im})\}^T$ is the $m_i$-dimensional column vector of the evaluations of
Theorem 1
Assume (A1)–(A3) hold. Then \(|\hat{\Sigma}(s, s') - \Sigma(s, s')| \overset{p}{\to} 0\) as \(n\) diverges. If in addition (A4) holds, then

\[ \|\hat{\Sigma}(\cdot, \cdot) - \Sigma(\cdot, \cdot)\|_s \overset{p}{\to} 0 \text{ as } n \to \infty, \]

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). Because the mean estimation has been studied previously, we assume that the response trajectories, \(Y_j(\cdot)\)'s, have zero mean and focus on the estimation of the model covariance. Throughout this section, we assume that \(Y_j(\cdot)\) is observed fully as a function over the domain, \(S = [0, 1]\). Section 4.1 discusses the main theoretical results when data are observed without error, that is, \(e_{ij}(s) = 0\) for \(s \in [0, 1]\). Section 4.2 extends the results to the case when the data are corrupted with smooth error process \(e_{ij}(s) \equiv e_{ij}(s)\). The proofs are detailed in the Supplementary Information; also in the Supplementary Information, we include a discussion on how to relax some of the assumptions. Throughout this section, we use \(S\) and \(T\) to distinguish between the domains.

We assume that the bivariate process \(X_j(s, T)\) 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:

\begin{enumerate}
  \item[(A1)] \(X = \{X(s, T) : (s, T) \in S \times T\}\) is a square integrable element of the \(L^2(S \times T)\), that is, \(E[ \int \int X^2(s, T)dsdT] < \infty\), where \(S\) and \(T\) are compact sets.
  \item[(A2)] The sampling density \(g(T)\) is continuous and \(\sup_{T \in T} g(T) < \infty\).
\end{enumerate}

Under (A1) and (A2), the function \(\Sigma(s, s')\) defined earlier (i) is symmetric, (ii) is positive definite and (iii) has eigenvalues \(\lambda_k\)'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 \(e_{ij}(s) = 0\) and thus \(Y_j(s) = X_j(s, T_j)\) for \(s \in S\). The sample covariance of \(Y_j(s)\) is \(\hat{\Sigma}(s, s') = \sum_{j=1}^{n} \sum_{m=1}^{m_j} Y_j(s)Y_j(s')/(\sum_{j=1}^{n} m_j)\). The following assumptions regard the moment behaviour of \(X\) and are commonly used in functional data analysis (Yao et al., 2005; Chen & Müller, 2012); we require them in our study.

\begin{enumerate}
  \item[(A3)] \(E[\|X(s, T)X(s', T)X(s, T')\|S] < \infty\) for arbitrary \(s, s' \in S\) and \(T, T' \in T\).
  \item[(A4)] \(E[\|X(\cdot, T)\|^[n]} < \infty\) for each \(T \in T\).
\end{enumerate}

Theorem 1
Assume (A1)–(A3) hold. Then \(\|\hat{\Sigma}(s, s') - \Sigma(s, s')\|_s \overset{p}{\to} 0\) as \(n\) diverges. If in addition (A4) holds, then

\[ \|\hat{\Sigma}(\cdot, \cdot) - \Sigma(\cdot, \cdot)\|_s \overset{p}{\to} 0 \text{ as } n \to \infty, \]
where \( \|k(\cdot, \cdot)\|_2 = \left\{ \int \int k^2(s, s') ds ds' \right\}^{1/2} \) is the Hilbert–Schmidt norm of \( k(\cdot, \cdot) \).

(A5) Let \( a_1 = (\lambda_1 - \lambda_2) \) and \( a_k = \max(\lambda_{k-1} - \lambda_k, (\lambda_k - \lambda_{k+1})) \) for \( k \geq 2 \), where \( \lambda_k \) is the \( k \)th 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 (A5) holds further, the eigen-elements of \( \hat{\Sigma}(s, s') \) are consistent estimators of the corresponding eigen-elements of \( \Sigma(s, s') \).

**Corollary 1**

Under the assumptions (A1)–(A5), for each \( k \), we have \( |\hat{\lambda}_k - \lambda_k| \overset{p}{\to} 0 \) and \( \|\hat{\phi}_k(\cdot) - \phi_k(\cdot)\|_2 \overset{p}{\to} 0 \) 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 \( \hat{\xi}_{W,jk} \); the result follows if \( \sup_{s,T} |Y(s, T)| \) 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 \( \hat{\xi}_{W,jk} \)'s is asymptotically identical to that based on \( \hat{\xi}_{W,jk} \). 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 \( \hat{\xi}_{ikl} \).

(A6) \( E[\sup_{s,T} |X(s, T)|^2] \leq M_n \) for a constant, \( M > 0 \), and an arbitrary integer, \( n \geq 1 \); this is equivalent to assume that \( X(s, T) \) is absolutely bounded almost surely.

(A7) Let \( b_{k1} = (\eta_{k1} - \eta_{k2}) \) and \( b_{ki} = \max(\eta_{k(i-1)} - \eta_{ki}, (\eta_{ki} - \eta_{k(i+1)})) \) for \( i \geq 2 \), where \( \eta_{ki} \) is the \( i \)th largest eigenvalues of \( G_k(T, T') \). Assume that \( 0 < b_{ki} < \infty \) and \( \eta_{ki} > 0 \) for all \( k \) and \( i \).

(A8) \( \hat{\xi}_{ikl} \) and \( \hat{\psi}_{ikl} \) are jointly Gaussian.

**Theorem 2**

Under the assumptions (A1)–(A6), for each \( k \) \( \sup_{T,T'} |\hat{\xi}_{W,jk} - \xi_{W,jk}| \overset{p}{\to} 0 \) and \( \|\hat{\Phi}_k(. \cdot, \cdot) - G_k(\cdot, \cdot)\|_2 \overset{p}{\to} 0 \) as \( n \) diverges. In fact, a stronger result also holds, namely, \( \sup_{T,T'} |\hat{\Phi}_k(T, T') - G_k(T, T')| \overset{p}{\to} 0 \) 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 \( \hat{\Phi}_k(\cdot, \cdot) \) satisfy \( |\hat{\eta}_{kl} - \eta_{kl}| \overset{p}{\to} 0 \), and \( \|\hat{\psi}_{kl}(\cdot) - \psi_{kl}(\cdot)\|_2 \overset{p}{\to} 0 \) as \( n \) diverges. Uniform convergence of \( \hat{\psi}_{kl}(\cdot) \) also holds: \( \sup_{T,T'} |\hat{\psi}_{kl}(T) - \psi_{kl}(T)| \overset{p}{\to} 0 \). Furthermore, as \( n \) diverges, we have \( |\hat{\sigma}_{e,k}^2 - \sigma_{e,k}^2| \overset{p}{\to} 0 \) and \( |\hat{\zeta}_{ikl} - \zeta_{ikl}| \overset{p}{\to} 0 \), where \( \zeta_{ikl} = E[\hat{\xi}_{ikl}\xi_{W,ik}] \) and \( \hat{\xi}_{W,ik} \) is the \( m_i \)-dimensional column vector of \( \xi_{W,ik} \)'s.

The consistency results for all model components imply prediction consistency.

**Theorem 3**

Assume (A1)–(A8), for each \( (s, T) \in S \times T \). Then \( \hat{Y}(s, T) \overset{p}{\to} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \zeta_{ikl} \hat{\psi}_{ikl}(T) \phi_k(s) \) as \( n, K \) and \( L_k \)'s \( \to \infty \).

### 4.2 Response curves measured with smooth error

Assume next that \( Y_j(s) \) are observed with smooth error \( \epsilon_{ij}(s) = \epsilon_{1,ij}(s) \), and thus, \( Y_j(s) = X_i(s, T_{ij}) + \epsilon_{1,ij}(s) \) for \( s \in S \) and \( \epsilon_{1,ij}(s) \in L^2(S) \). The main difference from Section 4.1 is that the sample covariance of \( Y_j(s) \) is an estimator of \( \Sigma(s, s') = \Sigma(s, s') + \Gamma(s, s') \), not of \( \Sigma(s, s') \); we denote the sample covariance of \( Y_j(s) \) by \( \hat{\Sigma}(s, s') = \sum_{i=1}^{n} Y_j(s)y_j(s') / (\sum_{i=1}^{n} m_i) \). Using similar arguments as earlier, we show that \( \hat{\Sigma}(s, s') \) is an unbiased estimator of \( \Sigma(s, s') \). Moreover, similar arguments can be used to show the pointwise consistency as well as the Hilbert–Schmidt norm consistency of \( \hat{\Sigma}(s, s') \). Additional assumptions are required.
The theoretical results are based on the assumptions that data of the predictions – of the time-varying coefficients and the response curve - hold without any modification. Furthermore, increasing the number of repeated curve measurements corresponding to the rest subjects from the sample and collect the subjects’ last profile. The remaining profiles for the 10 subjects and the data form a training set and a test set. The test set contains 10 profiles and is obtained as follows: randomly select 10 variances equal to structures: (i) non-parametric covariance (NP) where

\[
e_{ij}(\cdot) = \xi_{ik}(T) = \xi_{ik}1\psi_{k1}(T) + \xi_{ik}2\psi_{k2}(T) \text{; (ii) random effects model (REM)}
\]

and (iii) exponential autocorrelation (Exp) covariance structure of \(\psi_{k1} \text{ and } \psi_{k2}\). The details of the models are specified in the Supplementary Information. 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 modelled 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 prespecified level PVE = 0.95.

Estimation accuracy for the model components is evaluated using integrated mean squared errors, while prediction performance is assessed through in-sample integrated prediction errors (IN-IPE) and out-of-sample IPE (OUT-IPE). Table I shows the results for different covariance models for \(\xi_{ik}(T)\), different numbers of repeated curve measurements per subject, different SNRs, complex error process and varying sample sizes. The performance of the proposed estimation (see columns for \(\mu, \phi_1\) and \(\phi_2\) 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 x 3 top left block of integrated mean squared errors results corresponding to \(n = 100, n = 300\) and \(n = 500\). Moreover, both the prediction of \(\xi_{ik}(T)\)’s and \(f Y_{ij}(\cdot)\) are considered; see columns labelled \(\xi_1, \xi_2\, \text{IN-IPE and OUT-IPE}\) in Table I. 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
Table I. Estimation and prediction accuracy results based on $N_{\text{sim}} = 1000$ simulations.

|       | $\mu$ | $\phi_1$ | $\phi_2$ | $\xi_1$ | $\xi_2$ | IN-IPE naive | OUT-IPE naive | OUT-IPE naive |
|-------|-------|----------|----------|---------|---------|--------------|---------------|---------------|
| NP (a) | $n=100$ | 0.092 | 0.003 | 0.011 | 0.338 | 0.224 | 0.406 | 7.790 | 0.988 | 11.478 | $m_i \overset{iid}{\sim} \{8, \ldots, 12\}$ and SNR = 1 |
|       | $n=300$ | 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 | $m_i \overset{iid}{\sim} \{8, \ldots, 12\}$ and SNR = 1 |
|       | $n=300$ | 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 | 1.143 | 2.498 | $m_i \overset{iid}{\sim} \{8, \ldots, 12\}$ and SNR = 1 |
|       | $n=300$ | 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 |

The estimation of the covariance of the longitudinal process $\xi_{ik}(T)$ 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 IN-IPE_{naive} and OUT-IPE_{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 I. Continued.

|       | $\phi_1$ | $\phi_2$ | $\xi_1$ | $\xi_2$ | IN-IPE $\text{naive}$ | OUT-IPE | OUT-IPE $\text{naive}$ |
|-------|----------|----------|---------|---------|----------------------|---------|-----------------------|
| REM b | $n=100$  | 0.097    | 0.035   | 0.300   | 0.293               | 0.205   | 0.552                |
|       | $n=300$  | 0.034    | 0.010   | 0.160   | 0.140               | 0.178   | 0.552                |
|       | $n=500$  | 0.020    | 0.006   | 0.136   | 0.114               | 0.172   | 0.554                |
| Exp c | $n=100$  | 0.080    | 0.033   | 0.330   | 0.451               | 0.434   | 0.895                |
|       | $n=300$  | 0.027    | 0.009   | 0.010   | 0.236               | 0.313   | 0.410                |
|       | $n=500$  | 0.016    | 0.005   | 0.006   | 0.221               | 0.284   | 0.403                |

$n_i \overset{iid}{\sim} (15, \ldots, 20)$ and SNR = 5

Table II 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 min, while our approach takes about 7 s. The overall conclusion is that the proposed approach provides an improved prediction performance over the existing methods in a computationally efficient manner.

### 6 Diffusion tensor imaging 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 (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., 2015).

Our main objective is twofold: (i) to understand the dynamic behaviour 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 by Goldsmith et al. (2011), Staicu et al. (2012), Pomann et al. (2015) and Scheipl et al. (2015). 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 behaviour 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 ISI’s Journal for the Rapid Dissemination of Statistics Research

S. Y. Park and A.-M. Staicu

(wileyonlinelibrary.com) DOI: 10.1002/sta4.89

Table II. Comparison between the proposed method and Chen & Müller (2012) in the presence of correlated errors.

|                      | Chen & Müller (2012) | Proposed method (from Tables I and S2) |
|----------------------|----------------------|-----------------------------------------|
|                      | IN-IPE | OUT-IPE | Time (s) | IN-IPE | OUT-IPE | Time (s) |
| 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   |

Results based on \( N_{\text{sim}} = 1000 \) simulations. NP, non-parametric; SNR, signal to noise; IN-IPE, in-sample integrated prediction errors; OUT-IPE, out-sample integrated prediction errors; REM, random effects model. iid, independent and identically distributed; Exp, exponential autocorrelation.

the sampling distribution of \( T_{ij} \)'s has serious implications on the estimation of the bivariate mean \( \mu(s, T) \); a completely non-parametric bivariate smoothing would result in unstable and highly variable estimation. This is probably why Greven et al. (2010) first centred 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 Supporting Information. Using the bootstrap of subjects – based methods of Park et al. (2015) and \( B = 1000 \) bootstrap samples, we construct 95% joint confidence bands for \( \hat{\beta}_T(s) \) (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 three 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 Supporting Information. 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 \( \text{var}(b_{1ik}) = \sigma_{ik}^2 \) for \( l = 0, 1 \) and \( \text{cov}(b_{0ik}, b_{1ik}) = \sigma_{01k} \). 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 Supporting Information. The estimated \( \hat{\xi}_{11}(T) \) 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, \( \hat{\xi}_{22}(T) \) 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-IPE: \( \text{IN-IPE} = \sum_{i=1}^{162} \sum_{j=1}^{m_i} \int (Y_{ij}(s) - \hat{Y}_{ij}(s))^2 ds / \sum_{i=1}^{162} m_i \), where \( \hat{Y}_{ij}(s) = \hat{\mu}_0(s) + \sum_{k=1}^{K} (\hat{b}_{0k} + \hat{b}_{1kT_{ij}}) \hat{\phi}_k(s) \), and \( Y_{ij}(s) \)’s are the observed curve data. The square root of the IN-IPE is \( 2.31 \times 10^{-2} \) for our model; for comparison, Greven et al. (2010) yields \( 2.66 \times 10^{-2} \), and Chen & Müller (2012) give \( 3.76 \times 10^{-2} \). 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: \( \text{OUT-IPE} = \sum_{i=1}^{106} \int (\hat{Y}_{imi}(s) - Y_{imi}(s))^2 ds / 106 \), where \( \hat{Y}_{imi}(s) \) is the predicted curve at time \( T_{imi} \) for the \( i \)th subject using the fitted model based on all the data less the \( m_i \)th curve of the \( i \)th subject. Figure 4 shows such predicted curves \( \hat{Y}_{imi}(s) \) obtained using our model and the naive model for three randomly selected subjects at their last visit. The square root of OUT-IPE is \( 3.48 \times 10^{-2} \) for our model; for comparison, Chen & Müller (2012) give \( 8.71 \times 10^{-2} \), and the naïve approach gives \( 3.52 \times 10^{-2} \). These results suggest that, in this short-term study of MS, there is a small variation of CCA-FA profiles over time.

Figure 1. Left panel: 95% pointwise and joint confidence bands (CBs) of the slope function \( \hat{\beta}_T(s) \) 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 \( \hat{\mu}_0(s) \) (grey line) \( \pm 2 \sqrt{\hat{\lambda}_k \hat{\phi}_k(s)} \) (+ and − signs, respectively).
Figure 3. Estimated time-varying coefficients $\hat{\beta}_{ik}(T)$ for $k = 1, 2$ and $3$ using random effects model.

Figure 4. Predicted values of fractional anisotropy for the last visits of three randomly selected subjects; actual observations (grey); predictions using our model (black solid) and using the naive approach (black dashed).

Acknowledgements

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 (2007), ‘Diffusion tensor imaging of the brain’, Neurotherapeutics, 4(3), 316–329.

Baladandayuthapani, V, Mallick, BK, Young Hong, M, Lupton, JR, Turner, ND & Carroll, RJ (2008), ‘Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis’, Biometrics, 64(1), 64–73.

Basser, PJ, Mattiello, J & LeBihan, D (1994), ‘MR diffusion tensor spectroscopy and imaging’, Biophysical Journal, 66(1), 259–267.

Basser, PJ, Pajevic, S, Pierpaoli, C, Duda, J & Aldroubi, A (2000), ‘In vivo fiber tractography using DT-MRI data’, Magnetic Resonance in Medicine, 44(4), 625–632.
Basser, PJ & Pierpaoli, C (2011), ‘Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI’, Journal of Magnetic Resonance, 213(2), 560–570.

Bosq, D (2000), Linear Processes in Function Spaces: Theory and Applications, Vol. 149, Springer, New York.

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

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

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

Chen, K & Müller, HG (2012), ‘Modeling repeated functional observations’, Journal of the American Statistical Association, 107(500), 1599–1609.

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

Goldsmith, J, Bobb, J, Crainiceanu, CM, Caffo, B & Reich, D (2011), ‘Penalized functional regression’, Journal of Computational and Graphical Statistics, 20(4), 830–851.

Goldsmith, J, Zipunnikov, V & Schrack, J (2014), ‘Generalized multilevel functional-on-scalar regression and principal component analysis’, Biometrics, in press.

Greven, S, Crainiceanu, C, Caffo, B & Reich, D (2010), ‘Longitudinal functional principal component analysis’, Electronic Journal of Statistics, 4, 1022–1054.

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

Gromenko, O, Kokoszka, P, Zhu, L & Sojka, J (2012), ‘Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends’, The Annals of Applied Statistics, 6(2), 669–696.

Hastie, T, Tibshirani, R & Friedman, J (2009), The Elements of Statistical Learning, Vol. 2, Springer, New York.

Horváth, L & Kokoszka, P (2012), Inference for Functional Data with Applications, Vol. 200, Springer, New York.

Jiang, CR & Wang, JL (2010), ‘Covariate adjusted functional principal components analysis for longitudinal data’, The Annals of Statistics, 38, 1194–1226.

Li, Y & Guan, Y (2014), ‘Functional principal component analysis of spatio-temporal point processes with applications in disease surveillance’, Journal of the American Statistical Association, 109, 1205–1215.

Marx, BD & Eilers, PH (2005), ‘Multidimensional penalized signal regression’, Technometrics, 47(1), 13–22.

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

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

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

Pomann, GM, Staicu, AM & Ghosh, S (2015), ‘A two sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis’, Journal of the Royal Statistical Society: Series C (Applied Statistics). To appear.
Scheipl, F, Staicu, AM & Greven, S (2015), ‘Functional additive mixed models’, *Journal of Computational and Graphical Statistics*, in press.

Staicu, AM, Crainiceanu, CM & Carroll, RJ (2010), ‘Fast methods for spatially correlated multilevel functional data’, *Biostatistics*, 11(2), 177–194.

Staicu, AM, Crainiceanu, CM, Reich, DS & Ruppert, D (2012), ‘Modeling functional data with spatially heterogeneous shape characteristics’, *Biometrics*, 68(2), 331–343.

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

Wood, SN (2006), ‘Low-rank scale-invariant tensor product smooths for generalized additive mixed models’, *Biometrics*, 62(4), 1025–1036.

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

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

Xiao, L, Ruppert, D, Zipunnikov, V & Crainiceanu, C (2015), ‘Fast covariance estimation for high-dimensional functional data’, *Statistics and Computing*, in press.

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

Zhou, L, Huang, JZ & Carroll, RJ (2008), ‘Joint modelling of paired sparse functional data using principal components’, *Biometrika*, 95(3), 601–619.

Supplementary material

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