Longitudinal functional principal component analysis

Sonja Greven[*][†]

Department of Statistics
Ludwig-Maximilians-University Munich
Ludwigstr. 33
80539 Munich, Germany
e-mail: sonja.greven@stat.uni-muenchen.de

Ciprian Crainiceanu[†‡] **and Brian Caffo**[†‡]

Department of Biostatistics
Johns Hopkins University
615 N. Wolfe Street
Baltimore, MD 21205, USA
e-mail: ccrainic@jhsph.edu*;* bcaffo@jhsph.edu

Daniel Reich[§]

Translational Neuroradiology Unit
Neuroimmunology Branch
National Institute of Neurological Disorders and Stroke
National Institutes of Health
Bethesda, MD 20814, USA
e-mail: reichds@ninds.nih.gov

Departments of Radiology and Neurology
Johns Hopkins Hospital
600 N. Wolfe Street
Baltimore, MD 21287, USA

Abstract: We introduce models for the analysis of functional data observed at multiple time points. The dynamic behavior of functional data is decomposed into a time-dependent population average, baseline (or static) subject-specific variability, longitudinal (or dynamic) subject-specific variability, subject-visit-specific variability and measurement error. The model can be viewed as the functional analog of the classical longitudinal mixed effects model where random effects are replaced by random processes. Methods have wide applicability and are computationally feasible for moderate

and large data sets. Computational feasibility is assured by using principal component bases for the functional processes. The methodology is motivated by and applied to a diffusion tensor imaging (DTI) study designed to analyze differences and changes in brain connectivity in healthy volunteers and multiple sclerosis (MS) patients. An R implementation is provided.

Contents

1. Introduction

Scientific studies now commonly collect functional or imaging data at multiple visits over time. In this paper we introduce a class of models and inferential methods for the analysis of longitudinal data where each repeated observation is functional.

Our motivating data set comes from a diffusion tensor imaging (DTI) study, which was designed to analyze cross-sectional and longitudinal differences in
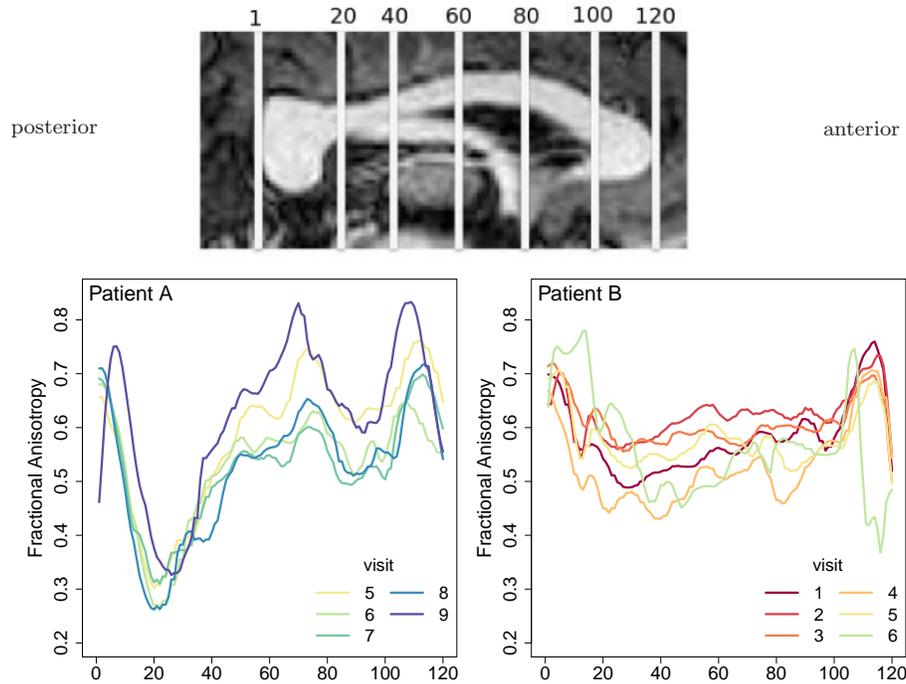
FIG 1. *Top: Sagittal image of the corpus callosum in one of the study subjects, a healthy 33-year-old man, showing the segmentation used [following 46] for construction of the tract profile. Values denote the bin number at the boundary point from the splenium (back of the head) to the genu/rostrum (closer to the eyes). Bottom: Two example subjects (both MS patients) from the tractography data with 5 and 6 complete visits, respectively. Shown are the fractional anisotropy along the corpus callosum, measured at the 120 sample points. Different visits for the same subject are indicated by color and overlaid.*

brain connectivity in healthy volunteers and multiple sclerosis (MS) patients. For each of 112 subjects and each visit, we have fractional anisotropy (FA) measurements along the corpus callosum in the brain. Figure 1 shows an image of the corpus callosum with 7 biological landmarks (denoted 1, 20, 40, 60, 80, 100, 120) used for registration of measurements across subjects. Each visit's data for a subject is a finely sampled function across the corpus callosum, with the argument of the function being the spatial distance along the tract. For illustration, the FA data is displayed for 2 subjects, one with 5 and one with 6 complete visits. Although change over time may be subtle in comparison with measurement error, accurate quantification of that change is crucial for applications ranging from powering clinical trials to understanding brain development. This data structure is not unique to this study. In fact, the tractography data is an example of many new data sets containing functions or images that are observed repeatedly over time (longitudinal functional data).

The common structure of these studies can be understood using an analogy with classical longitudinal data [8]. Longitudinal data is commonly analyzed using the very flexible class of linear mixed models [22, 44], which explicitly decompose the variation in the data into between- and within- subject variability. Similarly, we decompose the dynamic behavior of functional data into a time-dependent population average, baseline (or static) subject-specific variability, longitudinal (or dynamic) subject-specific variability, subject-visit-specific variability and measurement error. Technically this is achieved by replacing random effects with random functional effects.

We propose an estimation procedure that is based on an eigenanalysis and extends functional principal component analysis (FPCA) to the longitudinal setting. Computation is very efficient, even for very large data sets. The estimation procedure performed well both in an extensive simulation study and in the DTI application, where it uncovered subtle but potentially important subject-specific changes over time in a specific region of the corpus callosum, the isthmus. The character of these changes could conceivably be used as an early gauge of disease progression or response to neuroprotective therapies.

Our approach is different from functional mixed models based on the smoothing of fixed and random curves using splines or wavelets [3, 15, 16, 29]. In contrast to these methods focusing on the estimation of fixed and random curves, our approach is based on functional principal component analysis. In addition to the computational advantages of such an approach [compare also 19], we are able to extract the main differences between subjects in their average profiles and in how their profiles evolve over time. Such a signal extraction, which is not possible using smoothing methods alone, allows the relation of subject-specific scores to other variables such as disease status, age or disease progression. Our approach can be seen as an extension of functional principal component analysis for multilevel functional data [7]. Our methods apply to longitudinal data where each observation is functional, and should thus not be confused with nonparametric methods for the longitudinal profiles of scalar variables [17, 30, 31, 37, 41, 48, 50, 51]. For good introductions to functional data analysis in general, please see [10, 34].

The remainder of the paper is organized as follows. Section 2 introduces the longitudinal functional model and explains how dimension reduction via longitudinal functional principal component analysis (LFPCA) is achieved. Section 3 develops our estimation procedure and provides computational efficiency results. Section 4 shows the performance of our procedure in an extensive simulation study. Section 5 provides the application of LFPCA methods to the tractography data, while Section 6 concludes with a discussion. Theoretical results and proofs are given in the appendix. Supplementary material [13] providing an R function implementing LFPCA, simulation code and additional graphs is available in the archive that the Electronic Journal of Statistics maintains on Project Euclid.

2. The longitudinal functional model

In this section we introduce models for data sets where functional data are recorded at multiple time points or visits for the same observational unit or subject. The observed data are $\{Y_{ij}(d), d \in \mathcal{D}, T_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{V}_{ij}\}$, where $Y_{ij}(\cdot)$ is a random function in $L^2[0,1]$ observed at arguments d in some set $\mathcal{D}$, T_{ij} is the time of visit j for subject i, and $\boldsymbol{Z}_{ij}$ and $\boldsymbol{V}_{ij}$ are vectors of covariates for subject $i = 1, \ldots, I$ at visit $j = 1, \ldots, J_i$, where the number of visits J_i can vary with the subject, i. We assume that at least some subjects i have at least 3 visits, that is $J_i \geq 3$. The multi-level case when $J_i \leq 2$ for all i was fully addressed by [7] and [6].

2.1. The functional random intercept and random slope model

The data structure in this paper is similar to that of standard longitudinal data, with the exception that instead of observing scalars, Y_{ij}, one observes functions, $Y_{ij}(d)$, over time. We use this analogy to build up intuition and to introduce the functional equivalent of the standard longitudinal model. For simplicity we first extend the random intercept and slope model [35]. The functional analog is

$$Y_{ij}(d) = \eta(d, T_{ij}) + X_{i,0}(d) + X_{i,1}(d)T_{ij} + U_{ij}(d) + \varepsilon_{ij}(d), \tag{2.1}$$

where $\eta(d, T_{ij})$ is a fixed main effect surface, $X_{i,0}(d)$ is the random functional intercept for subject i, $X_{i,1}(d)$ is the random functional slope for subject i, T_{ij} is the time of visit j for subject i, $U_{ij}(d)$ is the random subject and visit-specific functional deviation, and $\varepsilon_{ij}(d)$ is random homoscedastic white noise. We make the following assumptions:

A.1 $\boldsymbol{X}_i(d) = \{X_{i,0}(d), X_{i,1}(d)\}$, $U_{ij}(d)$ and $\varepsilon_{ij}(d)$ are zero-mean, square-integrable, mutually uncorrelated random processes on $[0,1]$,

A.2 $X_{i,0}(d)$ and $X_{i,1}(d)$ have auto-covariance functions $K_0(d, d')$ and $K_1(d, d')$, respectively, and cross-covariance function $K_{01}(d, d')$,

A.3 $U_{ij}(d)$ has covariance function $K_U(d, d')$, and

A.4 $\varepsilon_{ij}(d)$ is white noise measurement error with variance σ^2.

There are several parallels between the scalar random intercept-random slope model and model (2.1). First, $Y_{ij}(d)$ is now a functional observation. Second, $X_{i,0}(d)$ and $X_{i,1}(d)$ replace the scalar random effects b_{i0} and b_{i1} as functional random intercept and random slope, respectively, capturing subject-to-subject variation. Third, the cross-covariance function $K_{01}(d, d')$ replaces the covariance between b_{i0} and b_{i1}. Fourth, the subject- and visit-specific deviation now consists of two parts. $U_{i,j}(d)$ is a visit-specific functional deviation from the subject-specific functional trend, capturing visit-to-visit functional variation on the same subject. $\varepsilon_{ij}(d)$ is additional white noise measurement error, capturing random uncorrelated variation within each curve. The overall mean trend is allowed to

be a smooth surface $\eta(d, T_{ij})$, which generalizes the linear mean $\beta_0 + \beta_1 T_{ij}$ often assumed in the scalar model.

Model (2.1) encompasses several simpler models that can be obtained as special cases. For example, visits may be of equal number $J_i = J$ per subject or equally spaced, $T_{ij} = j$ for all i and j. The mean function $\eta(d, T_{ij})$ may be time constant, $\eta(d)$, additive or linear in T, $\eta(s, T_{ij}) = \eta_1(d) + \eta_2(T_{ij})$, or $\eta(d, T_{ij}) = \eta_0(d) + T_{ij}\eta_1(d)$. The latter formulation is a direct extension of the linear population trend typically assumed in the scalar model.

Model (2.1) allows the decomposition of the variation in the observed curves into a) differences in subject's baseline functions; b) differences in subjects' average changes over time; c) visit-specific variation around these average trends; and d) measurement error. This decomposition is of interest in many applications. For example, in the tractography application we describe in Section 5, it is of interest to study both the population cross-sectional and the dynamic behavior of various measurements along neuronal tracts.

2.2. The general functional mixed model

While model (2.1) is rich enough for our application, it lends itself well to generalization, which could be useful in other applications. A more general form of the longitudinal functional model is

$$Y_{ij}(d) = \eta(d, \boldsymbol{Z}_{ij}) + \boldsymbol{V}'_{ij}\boldsymbol{X}_i(d) + U_{ij}(d) + \varepsilon_{ij}(d), \tag{2.2}$$

where we assume that

B.1 $\boldsymbol{X}_i(d)$, $U_{ij}(d)$ and $\varepsilon_{ij}(d)$ are zero-mean, square-integrable, mutually uncorrelated random processes on $[0, 1]$,

B.2 $\boldsymbol{X}_i(d)$ is a vector-valued random process with auto-covariance functions for the p components $K_{11}(d, d'), \ldots, K_{pp}(d, d')$, and cross-covariance functions $K_{12}(d, d'), \ldots, K_{1,p}(d, d'), \ldots, K_{p-1,p}(d, d')$,

B.3 $U_{ij}(d)$ is a random process with covariance function $K_U(d, d')$, and

B.4 $\varepsilon_{ij}(d)$ is white noise measurement error with variance σ^2.

Model (2.1) and its assumptions A.1 to A.4 are obtained as a special case of (2.2) and B.1 to B.4 by setting $p = 2$, $\boldsymbol{Z}_{ij} = T_{ij}$ and $\boldsymbol{V}_{ij} = (1, T_{ij})'$. Note that model (2.1) counts the components of $\boldsymbol{X}$ from 0 to 1 rather than from 1 to 2 to stress the analogy to the scalar random intercept-random slope model. The functional mixed-effects ANOVA model [7] results if we set $p = 1$, $\boldsymbol{Z}_{ij} = j$, $\boldsymbol{V}_{ij} = 1$, and $\eta(d, Z_{ij}) = \mu(d) + \eta_j(d)$. More generally, $\boldsymbol{Z}_{ij}$ and $\boldsymbol{V}_{ij}$ are vectors of known covariates for subject i at time T_{ij}. $\eta(d, \boldsymbol{Z}_{ij})$ is the fixed main effect surface, which can depend parametrically, semi-parametrically or non-parametrically on the covariates $\boldsymbol{Z}_{ij} = (Z_{ij,1}, \ldots, Z_{ij,m})$. The simplest parametric form is a linear mean $\eta(d, \boldsymbol{Z}_{ij}) = \alpha_0 + \alpha_1 d + \boldsymbol{Z}'_{ij}\boldsymbol{\beta}$, while the most complex nonparametric form is a $(p+1)$-dimensional smooth function $\eta(d, \boldsymbol{Z}_{ij})$. Intermediate semi-parametric models such as $\eta(d, \boldsymbol{Z}_{ij}) = \eta_1(d, Z_{ij,1}) + \cdots +$

$\eta_m(d, Z_{ij,m})$ or $\eta(d, \boldsymbol{Z}_{ij}) = \eta_0(d) + \boldsymbol{Z}'_{ij}\boldsymbol{\beta}$ could also be useful in particular applications.

Model (2.2) is the functional analog of the linear mixed model for longitudinal data [22]. It is similar to models used by [15, 29], but we do not assume Gaussianity and allow for more general fixed effects. The model of [15] also does not admit correlated random functional effects, as are present in (2.1). In addition, we follow quite a different modeling approach, using longitudinal functional principal component analysis instead of smoothing splines or wavelets for nested curves. This has large computational advantages compared to [29], especially when the number of random effects is large, see Section 4.3 and [19]. We are at the same time able to extract the main differences between subjects in how their profiles evolve over time, something not possible in these other approaches.

For notational simplicity in the remainder of the paper we will focus on model (2.1). In Appendix B we point out the small technical differences for fitting the more general model (2.2).

2.3. Dimension reduction via longitudinal FPCA

While models (2.1) and (2.2) are intuitive generalizations of linear mixed effects models, their computational feasibility is not obvious, especially for large numbers of subjects, visits and observations. We here propose an efficient modeling approach, Longitudinal Functional Principal Component Analysis (LFPCA). LFPCA is the longitudinal generalization of functional principal component analysis (FPCA) [34] and multilevel functional principal component analysis (MFPCA) [7]. The main idea of LFPCA is to extract the main directions of variation of the $\boldsymbol{X}$ and U processes using an eigen decomposition of their respective covariance operators on the basis of Mercer's theorem [27]. The Karhunen-Loève expansion [20, 26] is then used to obtain parsimonious expansions of $\boldsymbol{X}$ and U.

In the notation of model (2.1), we expand the covariance operator of the bivariate process $\boldsymbol{X}_i(d) = \{X_{i,0}(d), X_{i,1}(d)\}$ as

$$K_X(d, d') = \left(\begin{array}{cc} K_0(d, d') & K_{01}(d, d') \\ K_{01}(d', d) & K_1(d, d') \end{array} \right) = \sum_{k=1}^{\infty} \lambda_k \phi_k^X(d) \phi_k^X(d')',$$

where $\phi_k^X(d) = \{\phi_k^0(d), \phi_k^1(d)\}'$ are the ordered eigenfunctions of $K_X(d, d')$ corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Similarly, let $K_U(d, d') = \sum_{k=1}^{\infty} \nu_k \phi_k^U(d) \phi_k^U(d')$, where $\phi_k^U(d)$ are the ordered eigenfunctions of $K_U(d, d')$ corresponding to the eigenvalues $\nu_1 \geq \nu_2 \geq \cdots \geq 0$. The eigenfunctions $\{\phi_k^U, k \in \mathbb{N}\}$ form an orthonormal basis of $L^2[0, 1]$ with respect to the usual $L^2[0, 1]$ scalar product. The eigenfunctions $\{\phi_k^X, k \in \mathbb{N}\}$ form an orthonormal basis of $L^2[0, 1] \times L^2[0, 1]$ with respect to the additive scalar product

$$< (f_0, f_1), (g_0, g_1) >= \int_0^1 f_0(s)g_0(s)ds + \int_0^1 f_1(s)g_1(s)ds.$$

The function pairs (ϕ_k^0, ϕ_j^1), (ϕ_k^0, ϕ_j^U) or (ϕ_k^1, ϕ_j^U) are not required to be orthogonal in $L^2[0,1]$, nor will (ϕ_k^0, ϕ_j^0) or (ϕ_k^1, ϕ_j^1) be orthogonal in general for $k \neq j$. The Karhunen-Loève expansions of the random processes are

$$\boldsymbol{X}_i(d) = \sum_{k=1}^{\infty} \xi_{ik} \phi_k^X(d) \qquad \text{and} \qquad U_{ij}(d) = \sum_{k=1}^{\infty} \zeta_{ijk} \phi_k^U(d),$$

where the principal components scores

$$\xi_{ik} = \int_0^1 X_{i,0}(s)\phi_k^0(s)ds + \int_0^1 X_{i,1}(s)\phi_k^1(s)ds \quad \text{and} \quad \zeta_{ijk} = \int_0^1 U_{ij}(s)\phi_k^U(s)ds$$

are uncorrelated random variables with mean zero and variances λ_k and ν_k, respectively. Assumption A.1. is ensured by assuming that $\{\xi_{ik}, i = 1, \ldots, I, k \in \mathbb{N}\}$ and $\{\zeta_{ijk}, j = 1, \ldots, J_i, i = 1, \ldots, I, k \in \mathbb{N}\}$ are mutually uncorrelated. Because working with infinite expansions is impractical, we consider the finite-dimensional approximations of the $\boldsymbol{X}$ and U processes

$$\boldsymbol{X}_i(d) = \sum_{k=1}^{N_X} \xi_{ik} \phi_k^X(d) \qquad \text{and} \qquad U_{ij}(d) = \sum_{k=1}^{N_U} \zeta_{ijk} \phi_k^U(d),$$

where N_X and N_U will be estimated, as described in Section 3.4. Conditional on N_X and N_U the finite approximation to model (2.1) is

$$\left.\begin{aligned} Y_{ij}(d) &= \eta(d, T_{ij}) + \sum_{k=1}^{N_X} \xi_{ik} \boldsymbol{V}_{ij}' \phi_k^X(d) + \sum_{k=1}^{N_U} \zeta_{ijl} \phi_k^U(d) + \varepsilon_{ij}(d), \\ \xi_{ik} &\overset{unc}{\sim} (0, \lambda_k), \quad \zeta_{ijl} \overset{unc}{\sim} (0, \nu_l), \quad \varepsilon_{ij}(d) \overset{unc}{\sim} (0, \sigma^2), \end{aligned}\right\} (2.3)$$

$\boldsymbol{V}_{ij} = (1, T_{ij})'$, which is a linear mixed model [see also 7]. Here, $x_l \overset{unc}{\sim} (0, a)$ denotes uncorrelated variables with mean 0 and variance a. We are neither assuming normality of the processes in (2.1) nor of the scores in (2.3). LFPCA extends similarly to the more general model (2.2).

3. Estimation

For reasons of simplicity, we focus the presentation on model (2.1), but estimation is done similarly for model (2.2). The minor adjustments for fitting (2.2) are described in Appendix B. We assume that the mean, covariance operators and eigenfunctions are smooth. For presentation purposes, we assume that all functions $Y_{ij}(d)$ are measured at a finite number, D, of grid points $\mathcal{D} \subset [0,1]$. However, the method can easily handle missing data, both in terms of visits per subject or observations per visit. Estimation can be done using a few simple steps, which will be described in more detail in the following.

Step 1 The fixed effect surface η is estimated using the working independence model

$$Y_{ij}(d) = \eta(d, T_{ij}) + \epsilon_{ij}(d).$$

Smoothness selection is by REML, which is more robust to neglecting the correlations in the errors than prediction error methods [21].

Step 2 The autocovariance functions for the random processes $\boldsymbol{X}_i = (X_{i,0}, X_{i,1})$ and U_{ij} are estimated from the residuals $Y_{ij}(d) - \hat{\eta}(d, T_{ij})$, using a linear regression step.

Step 3 The 'raw' autocovariance function estimates from step 2 are subjected to bivariate smoothing, yielding also an estimate for σ^2.

Step 4 Eigen decompositions of the smoothed autocovariance functions provide bases for representing $\boldsymbol{X} = (X_{i,0}, X_{i,1})$ and U_{ij}, which are truncated to achieve parsimony.

Step 5 Estimated BLUPs then provide estimates for the subject- and visit-specific scores, which summarize the main differences in the dynamics of functions over time.

3.1. Estimation of the mean

The fixed effect population mean surface $\eta(d, T)$ can be estimated using a bivariate smoother in d and T under a working independence assumption. For discussions of smoothing for correlated data, see [21, 24]. Possibilities for smoothers include penalized splines [39], smoothing splines [11] and local polynomials [9]. Choice of a smoother and of the smoothing parameter or bandwidth is discussed extensively in the literature and is not the main focus here. It is our experience that most reasonable smoothers used judiciously will provide similar results. For simplicity and efficiency of the implementation for large data sets, we use penalized spline smoothing with REML estimation of the smoothing parameter. This choice has also been found to be relatively robust to misspecification of the error correlation structure in [21].

A bivariate smoother is appropriate when the collection of observations across visits and subjects is relatively dense. This need not be the case in general, and simpler choices might be more sensible. For example, $\eta(d, T_{ij}) = \eta_0(d) + T_{ij}\beta$ might be more appropriate if the T_{ij} form a sparser collection. In the case of equally spaced visits, $T_{ij} = T_j$, [7] used $\eta(d, T_{ij}) = \eta_j(d)$. Choices will depend on the particular application, available data and scientific problem. In most applications, estimating the mean function is quite easy and, even, routine. Once a consistent estimator of the mean function is available, data can be centered as follows $Y_{ij}(d) - \hat{\eta}(d, T_{ij})$ for all i, j and d. In the following we assume that the $Y_{ij}(d)$ are mean zero.

3.2. Estimation of the covariance operators

A crucial point of our proposed methodology is estimating the covariance operators $K_X(\cdot, \cdot)$ and $K_U(\cdot, \cdot)$. To estimate the covariance functions, we focus on the cross-products $Y_{ij}(d)Y_{ik}(d')$. Because $Y_{ij}(d)$ has zero mean, each product $Y_{ij}(d)Y_{ik}(d')$ is an estimator of the covariance between the function observed at

time T_{ij} evaluated at location d, and the function observed at time T_{ik} evaluated at location d'. Every subject thus contributes an estimator each for every available pair of observations at time T_{ij} evaluated at location d, and at time T_{ik} evaluated at location d'. Available pairs of observations may vary between subjects in their (d, d') and (T_{ij}, T_{ik}) combinations. The method described in the following can thus easily handle missing data, both in terms of visits per subject or observations per visit.

Under the assumptions of model (2.1),

$$\begin{aligned}
\mathrm{E}\{Y_{ij}(d)Y_{ik}(d')\} &= \mathrm{Cov}\{Y_{ij}(d), Y_{ik}(d')\} \qquad\qquad\qquad (3.1)\\
&= K_0(d, d') + T_{ik}K_{01}(d, d') + T_{ij}K_{01}(d', d) + T_{ij}T_{ik}K_1(d, d')\\
&\qquad\qquad\qquad\qquad\qquad + [K_U(d, d') + \sigma^2\delta_{dd'}]\delta_{jk},
\end{aligned}$$

for all d, d', i, j and k, where δ_{jk} is Kronecker's delta. Equation (3.1) suggests a straightforward solution for estimating the covariance operators: regress linearly the "outcome" $Y_{ij}(d)Y_{ik}(d')$ on the "covariates" $(1, T_{ik}, T_{ij}, T_{ik}T_{ij}, \delta_{jk})$, where the "parameters" are $\{K_0(d, d'), K_{01}(d, d'), K_{01}(d', d), K_1(d, d'), K_U(d, d'); d, d' \in \mathcal{D}; \sigma^2\}$.

While the intuition behind the method is simple, there are two potential pitfalls that should be carefully avoided. First, σ^2 is identifiable only under the assumption that $K_U(d, d')$ is a bivariate smooth function in d and d'. Second, in a straightforward implementation of the linear regression on the basis of (3.1), there are $D^2 \sum_{i=1}^{I} J_i^2$ observations and $4D^2 + 1$ variables. In our moderately sized tract data, this would correspond to 19 million observations and $57,500$ variables. In larger data sets the problem would be even more serious. Thus, careful implementation is required to ensure computational feasibility. We propose the following 3-step estimation procedure that avoids these problems.

<u>Step A:</u> $\{K_0(d, d'), K_{01}(d, d'), K_{01}(d', d), K_1(d, d'), K_U(d, d') + \sigma^2\delta_{dd'}\}$ is estimated for each pair $d \leq d' \in \mathcal{D}$ using least squares estimation based on (3.1). Symmetry constraints yield $K_0(d, d') = K_0(d', d)$, $K_1(d, d') = K_1(d', d)$ and $K_U(d, d') = K_U(d', d)$ for $d > d'$. Denote estimates by $\widetilde{K}_0(d, d')$, $\widetilde{K}_{01}(d, d')$, $\widetilde{K}_1(d, d')$ and $\widetilde{K}_U(d, d')$.

<u>Step B:</u> Bivariate smoothing in d and d' over $\widetilde{K}_0(d, d')$, $\widetilde{K}_{01}(d, d')$ and $\widetilde{K}_1(d, d')$ yields smooth estimates $\widehat{K}_0(d, d')$, $\widehat{K}_{01}(d, d')$ and $\widehat{K}_1(d, d')$. Bivariate smoothing over $\widetilde{K}_U(d, d')$, leaving out the diagonal elements as proposed by [41, 50], also yields estimates $\widehat{K}_U(d, d')$.

Please see Section 3.1 for a discussion of bivariate smoothing.

<u>Step C:</u> σ^2 can be estimated as $\widehat{\sigma}^2 = \frac{1}{D} \sum_{d=1}^{D} \{\widetilde{K}_U(d, d) - \widehat{K}_U(d, d)\}$, if positive, and as zero otherwise.

Estimation in Step A can be done using efficient matrix-vector computations as detailed in Theorem 1 in Appendix A. The following is a consequence of that theorem.

Corollary 1. *The computational effort for estimation of the covariance functions in Step A for the general model* (2.2) *is of the order $O\{\max(p^6, p^2D^2g)\}$, where $g = \sum J_i^2$ and p is the dimension of the vector-valued random process $\boldsymbol{X}_i(d)$ in* (2.2).

All proofs can be found in Appendix A. For model (2.1), $p = 2$ is small. Note that p^2D^2 is the order of the number of unknown parameters in the covariance functions, and g is the number of observation pairs contributing to the estimation. The effort thus is linear in both. Our software implementation is so efficient that the computational effort is dominated by the bivariate smoothing of the mean and covariance functions; see Section 4.3 for a detailed investigation of efficiency.

Our procedure does not guarantee that $\widehat{K}_X(\cdot, \cdot)$ and $\widehat{K}_U(\cdot, \cdot)$ are positive definite. We correct this problem by trimming the eigenvalue-eigenvector pairs corresponding to negative eigenvalues, a method that has been found to increase the L^2 accuracy [17] and has been shown to work well in practice [51].

3.3. Estimation of the eigenfunctions and scores

In the previous section we showed how to obtain the estimated covariance matrices $\widehat{\boldsymbol{K}}_0 = \{\widehat{K}_0(d, d')\}_{d;d' \in \mathcal{D}}$, $\widehat{\boldsymbol{K}}_{01} = \{\widehat{K}_{01}(d, d')\}_{d;d' \in \mathcal{D}}$, $\widehat{\boldsymbol{K}}_1 = \{\widehat{K}_1(d, d')\}_{d;d' \in \mathcal{D}}$, and $\widehat{\boldsymbol{K}}_U = \{\widehat{K}_U(d, d')\}_{d;d' \in \mathcal{D}}$. Estimates of the eigenvalues and of the eigenfunctions $\phi_k^X(\cdot)$, $k = 1, 2, \ldots, D$, and $\phi_k^U(\cdot)$, $k = 1, 2, \ldots, D$, at the grid points $\mathcal{D}$ can then be obtained using the spectral decomposition of $\boldsymbol{K}_X$ and $\boldsymbol{K}_U$, $\widehat{\boldsymbol{K}}_X = \sum_{k=1}^{2D} \widehat{\lambda}_k \widehat{\phi}_k^X \{\widehat{\phi}_k^X\}'$ and $\widehat{\boldsymbol{K}}_U = \sum_{k=1}^{D} \widehat{\nu}_k \widehat{\phi}_k^U \{\widehat{\phi}_k^U\}'$, where $\widehat{\phi}_k^X = \{(\widehat{\phi}_k^0(d), \widehat{\phi}_k^1(d)); d \in \mathcal{D}\}$ and $\widehat{\phi}_k^U = \{\widehat{\phi}_k^U(d); d \in \mathcal{D}\}$, are orthonormal vectors in $\mathbb{R}^{2D}$ and $\mathbb{R}^D$, respectively. The estimation of the number of eigenfunctions retained for further analysis, N_X and N_U, is described in Section 3.4.

In Section 2.3, equation (2.3), we showed that for fixed N_X and N_U, model (2.1) is a linear mixed model. Thus, we can use best linear unbiased prediction (BLUP) to obtain predictions of the subject- and subject/visit-specific scores, ξ_{ik} and ζ_{ijk}, respectively. BLUP calculation does not require a normality assumption and is a generalization of the conditional expectations used by [50].

For given eigenfunctions, mean function $\eta(d, T)$, and variances λ_k, $k = 1, \ldots, N_X$, ν_k, $k = 1, \ldots, N_U$, the BLUP for $\boldsymbol{b} = (\xi_{11}, \ldots, \xi_{1N_X}, \ldots, \xi_{I1}, \ldots, \xi_{IN_X}, \zeta_{111}, \ldots, \zeta_{11N_U}, \ldots, \zeta_{IJ_I1}, \ldots, \zeta_{IJ_IN_U})$ in model (2.3) is given in the usual form by

$$\widehat{\boldsymbol{b}} = \boldsymbol{D}\boldsymbol{Z}'(\sigma^2\boldsymbol{I} + \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}')^{-1}(\boldsymbol{Y} - \boldsymbol{\eta}), \tag{3.2}$$

where $\boldsymbol{Z} = [\boldsymbol{Z}_X|\boldsymbol{Z}_U]$, $\boldsymbol{Z}_X = \boldsymbol{E}_{\boldsymbol{I}} \otimes \boldsymbol{\Phi}^0 + \boldsymbol{T} \otimes \boldsymbol{\Phi}^1$, $\boldsymbol{Z}_U = \boldsymbol{I_n} \otimes \boldsymbol{\Phi}^U$, $\boldsymbol{E_I} = (\delta_{ih})_{ij=11,\ldots,IJ_I;h=1,\ldots,I}$, $\boldsymbol{T} = (T_{ij}\delta_{ih})_{ij=11,\ldots,IJ_I;h=1,\ldots,I}$, $\boldsymbol{\Phi}^0 = \{\phi_k^0(d)\}_{d \in \mathcal{D};k=1,\ldots,N_X}$, $\boldsymbol{\Phi}^1 = \{\phi_k^1(d)\}_{d \in \mathcal{D};k=1,\ldots,N_X}$, $\boldsymbol{\Phi}^U = \{\phi_k^U(d)\}_{d \in \mathcal{D};k=1,\ldots,N_U}$, $\boldsymbol{D} = \text{blockdiag}(\boldsymbol{D}_X, \boldsymbol{D}_U) = \text{blockdiag}\{\boldsymbol{I_I} \otimes \text{diag}(\lambda_1, \ldots, \lambda_{N_X}), \boldsymbol{I_n} \otimes \text{diag}(\nu_1, \ldots, \nu_{N_U})\}$, $\boldsymbol{Y} = \{Y_{11}(1), \ldots, Y_{11}(D), \ldots, Y_{1J_1}(1), \ldots, Y_{1J_1}(D), \ldots, Y_{IJ_I}(1), \ldots,$

$Y_{IJ_I}(D)\}$, and $\boldsymbol{\eta} = \{\eta(1, T_{11}), \ldots, \eta(D, T_{11}), \ldots, \eta(D, T_{1J_1}), \ldots, \eta(1, T_{IJ_I}), \ldots, \eta(D, T_{IJ_I})\}$. Here, $\otimes$ denotes the Kronecker product of matrices, and $(a_{ijh})_{ij=11,\ldots,IJ_I;h=1,\ldots,I}$ denotes a matrix with entries a_{ijh}, rows $ij, j = 1, \ldots, J_i$, $i = 1, \ldots, I$, and columns $h = 1, \ldots, I$.

We can obtain estimated BLUPs (EBLUPs) using the estimated functions and variances $\widehat{\eta}(\cdot, \cdot)$, $\widehat{\sigma}^2$, $\widehat{\phi}_k^0(\cdot)$, $\widehat{\phi}_k^1(\cdot)$, $\widehat{\lambda}_k$, $k = 1, \ldots, N_X$, and $\widehat{\phi}_k^U(\cdot)$, $\widehat{\nu}_k$, $k = 1, \ldots, N_U$, from Sections 3.1 and 3.2. This does not require fitting the model (2.3), which greatly increases computational efficiency. While straightforward implementation of (3.2) requires inverting $nD \times nD$ matrices, which would result in computational effort of the order $O(n^3 D^3)$, we make use of common matrix rules and of the model structure to obtain a more efficient representation, as detailed in Theorem 2 in Appendix A. The following result of Theorem 2 confirms the manageable computational effort even for very large data sets.

Corollary 2. *Computational effort for calculation of the estimated BLUPs in* (3.2) *is of the order* $O\{\max(nDf, IN_X^3)\}$, *where* $n = \sum_{i=1}^{I} J_i$ *and* $f = N_U + N_X I/n$.

The proofs are provided in Appendix A. N_X and N_U are typically small, and much smaller than either D or the number of observed curves n. These results and efficient block matrix manipulation make the models proposed here feasible even for very large data sets. For example, one of the simulation examples in Section 4 uses $1,000$ subjects, who were observed at 8 visits and had 200 observations per visit.

3.4. Decomposition of variance and choice of the number of components

There are several possible ways to choose the numbers of eigenfunctions N_X and N_U. Two alternatives that have been used before are leave-one-curve-out cross validation [38] and an AIC-type criterion [50]. Alternatively, one can make use of the fact that (2.3) is a linear mixed model, with N_X and N_U corresponding to the number of random effects. The conditional Akaike information criterion (cAIC), proposed for the selection of random effects in linear mixed models [12, 23, 43] could thus be employed. [40] and [6] point out that choosing the number of eigenfunctions corresponds to step-wise testing for zero variance components. They propose using a Restricted Likelihood Ratio Test (RLRT) for this zero variance. The null distribution can be easily approximated using methods introduced by [14] based on the null distribution derived in [5].

Here we follow a simpler approach based on the proportion of variance explained. This approach has several advantages: a) popularity; b) simplicity and interpretability; c) quantification of the contribution of the different processes to the variability in $Y_{ij}(d)$.

To better understand variance partitioning, we give the following result.

Lemma 1. *Let* $Y_{ij}(d) \in L^2[0, 1]$ *be a process that follows model* (2.1) *with zero mean,* $\eta(d, T_{ij}) \equiv 0$. *Let* T_{ij} *be independently distributed as* T *for all* i *and* j,

where $\mathrm{E}(T^2) < \infty$, and let T_{ij} be independent of $\boldsymbol{X}_i$, U_{ij} and $\varepsilon_{ij}(d), d \in \mathcal{D}$. Then, the average variance of $Y_{ij}(d)$ can be written as

$$
\int_0^1 Var\{Y_{ij}(s)\}ds = \int_0^1 \left(\sum_{k=1}^{\infty} \lambda_k \left[\{\phi_k^0(s)\}^2 + 2\,\mathrm{E}(T_{ij})\phi_k^0(s)\phi_k^1(s) \right. \right.
$$
$$
\left. \left. + \mathrm{E}(T_{ij}^2)\{\phi_k^1(s)\}^2 \right] + \sum_{k=1}^{\infty} \nu_k \{\phi_k^U(s)\}^2 + \sigma^2 \right) ds.
$$

The proof can be found in Appendix A. Given the usual interpretation of eigenvalues as variance explained in FPCA, one could be tempted to interpret λ_k similarly in the longitudinal context. The variance decomposition that we just described indicates that in LFPCA, λ_k can be interpreted as a variance component only if the time variable is standardized to have zero mean and unit variance. In this case, the two components of the ϕ_k^X eigenfunction, ϕ_k^0 and ϕ_k^1, will be on the same scale. We can then directly discuss λ_k as the "variance explained" by the eigenfunction ϕ_k^X of K_X. Thus, we recommend standardizing the time variable. The variation in $Y_{ij}(d)$ then has the following simple additive decomposition.

Corollary 3. *In the case when* $\mathrm{E}(T_{ij}) = 0$ *and* $Var(T_{ij}) = 1$, *the expression in Lemma 1 reduces to*

$$
\int_0^1 Var\{Y_{ij}(s)\}ds = \sum_{k=1}^{\infty} \lambda_k + \sum_{k=1}^{\infty} \nu_k + \sigma^2.
$$

Thus, for standardized T_{ij}, the variation in $Y_{ij}(d)$ can be decomposed additively into the contributions from the random intercept and random slope process, $\sum_{k=1}^{\infty} \lambda_k$, from the visit-specific deviation process, $\sum_{k=1}^{\infty} \nu_k$, and from the additional random noise, σ^2. This leads to a simple decision rule for N_X and N_U: choose ϕ_k^X and ϕ_k^U corresponding to λ_k and ν_k in decreasing order, until

$$
\{\sum_{k=1}^{N_X} \lambda_k + \sum_{k=1}^{N_U} \nu_k + \sigma^2\}/\{\sum_{k=1}^{\infty} \lambda_k + \sum_{k=1}^{\infty} \nu_k + \sigma^2\} \geq L,
$$

where L is a pre-specified proportion of explained variation, such as $L = 0.95$. $\sum_{k=1}^{\infty} \lambda_k$ and $\sum_{k=1}^{\infty} \nu_k$ provide quantifications of the relative importance of the $\boldsymbol{X}$ and U processes.

4. Simulations

4.1. Simulation design

To investigate the performance of our estimation procedure, we conduct an extensive simulation study. The design combines and extends scenarios used by [7] and [50]. For all settings, we generate 1000 data sets from model (2.3),

where $N_X = N_U = 4$. We set the mean function to $\eta(d, T) = 0.5(T/4 - d)^2$. The unequally spaced time points T_{ij} are simulated such that the mean for each subject is zero, and increments $T_{ij} - T_{ij-1}$ are independent draws from a uniform distribution on $[0, 1]$. The time variable is then standardized to have unit variance. The curves $Y_{ij}(d)$ are taken to be observed for $d \in \mathcal{D} = \{(k - 0.5)/D, k = 1, \ldots, D\}$, $D = 120$, as in the tract data. We set the variances to be $\lambda_k = \nu_k = 2^{1-k}$, $k = 1, \ldots, 4$, and $\sigma = 0.05$. This choice corresponds to 0.07% of the overall average variance explained by the error variance σ^2, higher than in the tract data (0.02%, please see Table 2).

We consider all possible combinations of the following scenarios:

1. number of subjects (a) $I = 50$ (b) $I = 100$ (c) $I = 200$ and (d) $I = 500$, including both smaller and larger numbers than in the tract data,
2. (a) balanced design with $J_i = 4$ for all i,

   (b) unbalanced design with $J_i \in \{1, \ldots, 9\}$, (a multiple of 8, 8, 9, 6, 5, 5, 4, 3, 2 times, respectively), giving 4 observations per subject on average,
3. (a) normal scores $\xi_{ik} \sim \mathcal{N}(0, \lambda_k)$ and $\zeta_{ijk} \sim \mathcal{N}(0, \nu_k)$ for all i, j and k,

   (b) non-normal scores; ξ_{ik} drawn from a mixture of two normals, with equal probability from either $\mathcal{N}(\sqrt{\lambda_k/2}, \lambda_k/2)$ or from $\mathcal{N}(-\sqrt{\lambda_k/2}, \lambda_k/2)$; ζ_{ijk} drawn with equal probability from either $\mathcal{N}(\sqrt{\nu_k/2}, \nu_k/2)$ or $\mathcal{N}(-\sqrt{\nu_k/2}, \nu_k/2)$,
4. (a) eigenfunctions $\phi_k^X = (\phi_k^0, \phi_k^1)$ with ϕ_k^0 and ϕ_k^1 orthogonal and of equal norm $\sqrt{1/2}$; ϕ_k^U are not orthogonal to either ϕ_k^0 or ϕ_k^1,

   (b) eigenfunctions $\phi_k^X = (\phi_k^0, \phi_k^1)$ with ϕ_k^0 and ϕ_k^1 non-orthogonal and of unequal norms $\sqrt{3/4}$ and $\sqrt{1/4}$. ϕ_k^U are equal to ϕ_j^0 or ϕ_j^1 for some j for all k,
5. (a) estimation does not include bivariate smoothing of the covariance functions. In this case, smoothing is only used to obtain an estimate of the diagonal $K_U(d, d), d \in \mathcal{D}$, and of σ^2,

   (b) estimation includes bivariate smoothing of the covariance functions.

This gives 64 different combinations overall. The eigenfunctions for setting 4. (a) are

$$
\begin{aligned}
\phi_1^0(d) &= \sin(2\pi d) & \phi_1^1(d) &= 1/\sqrt{2} & \phi_1^U(d) &= 1 \\
\phi_2^0(d) &= \cos(2\pi d) & \phi_2^1(d) &= \sin(6\pi d) & \phi_2^U(d) &= \sqrt{3}(2d - 1) \\
\phi_3^0(d) &= \sin(4\pi d) & \phi_3^1(d) &= \cos(6\pi d) & \phi_3^U(d) &= \sqrt{5}(6d^2 - 6d + 1) \\
\phi_4^0(d) &= \cos(4\pi d) & \phi_4^1(d) &= \sin(8\pi d) & \phi_4^U(d) &= \sqrt{7}(20d^3 - 30d^2 + 12d - 1).
\end{aligned}
$$

Note that while ϕ_k^0 and ϕ_k^1 are orthogonal, they are not orthogonal to ϕ_j^U for all k and $j \neq 1$. The eigenfunctions for setting 4. (b) are

$$\phi_1^0(d) = \sin(2\pi d)\sqrt{2/3} \qquad\qquad \phi_1^1(d) = 1/2$$
$$\phi_2^0(d) = \cos(2\pi d)\sqrt{2/3} \qquad\qquad \phi_2^1(d) = \sqrt{3}(2d-1)/2$$
$$\phi_3^0(d) = \sin(4\pi d)\sqrt{2/3} \qquad\qquad \phi_3^1(d) = \sqrt{5}(6d^2 - 6d + 1)/2$$
$$\phi_4^0(d) = \cos(4\pi d)\sqrt{2/3} \qquad\qquad \phi_4^1(d) = \sqrt{7}(20d^3 - 30d^2 + 12d - 1)/2$$
$$\phi_1^U(d) = \sqrt{4}\phi_1^1(d)$$
$$\phi_2^U(d) = \sqrt{4/3}\phi_1^0(d)$$
$$\phi_3^U(d) = \sqrt{4/3}\phi_2^1(d)$$
$$\phi_4^U(d) = \sqrt{4/3}\phi_3^1(d).$$

Note that now ϕ_k^0 and ϕ_j^1 are non-orthogonal, and ϕ_k^0 has a larger norm than ϕ_k^1. Also, ϕ_k^U is equal to one of the ϕ_j^0 or ϕ_j^1, $j = 1, \ldots, 4$, for each k, making separation of the two processes $\boldsymbol{X}$ and U much more difficult. For bivariate smoothing of the mean and covariance functions, we use tensor product penalized cubic regression splines with 10 knots per dimension, where the smoothing parameters are estimated using REML estimation, as implemented in the R package `mgcv` [47].

To investigate the sensitivity of our results to our choices for η and σ, we also consider four variations on the balanced design with $I = 100$ and $J_i = 4$ for all i (1b and 2a), non-orthogonal ϕ_k^0, ϕ_k^1 and ϕ_k^U with unequal weight on ϕ_k^1 and ϕ_k^0 (4b), a mixture distribution for the scores ξ_{ij} and ζ_{ijk} (3b), and bivariate smoothing of the covariance functions (5b). For these four settings, we vary $\eta(d,T) = (T/4 - d/D + 1/2)(T/4 + d/D - 1/2)$, $\eta(d,T) = \sin(\pi T/2)d/D$, $\sigma = 0.5$ (corresponding to 6.25% of the overall average variance explained by the error variance σ^2) or $\sigma = 1$ (21.05%), respectively.

For each of the 1000 replications and for each of the 68 settings, our estimation procedure from Section 3 with $N_X = N_U = 4$ is used to obtain estimates of the mean function, the covariance functions, the eigenfunctions, the scores, and all variances.

4.2. Simulation results

We now discuss results for one of the 68 settings in detail, and point out differences across settings. The complete simulation results can be found in the supplementary material.

Figure 2 and Table 1 show the main results of simulations based on a balanced design with $I = 100$ and $J_i = 4$ for all i (1b and 2a), non-orthogonal ϕ_k^0, ϕ_k^1 and ϕ_k^U with unequal weight on ϕ_k^1 and ϕ_k^0 (4b), a mixture distribution for the scores ξ_{ij} and ζ_{ijk} (3b), and no bivariate smoothing of the covariance functions (5a). A plot of the true and estimated mean functions can be found in the supplementary material, illustrating that the mean is well and unbiasedly estimated.

Figure 2 shows the true and estimated eigenfunctions $\phi_k^X = (\phi_k^0, \phi_k^1)$ and ϕ_k^U, $k = 1, \ldots, 4$. Results for ϕ_k^0, ϕ_k^1 and ϕ_k^U are displayed in the left, middle
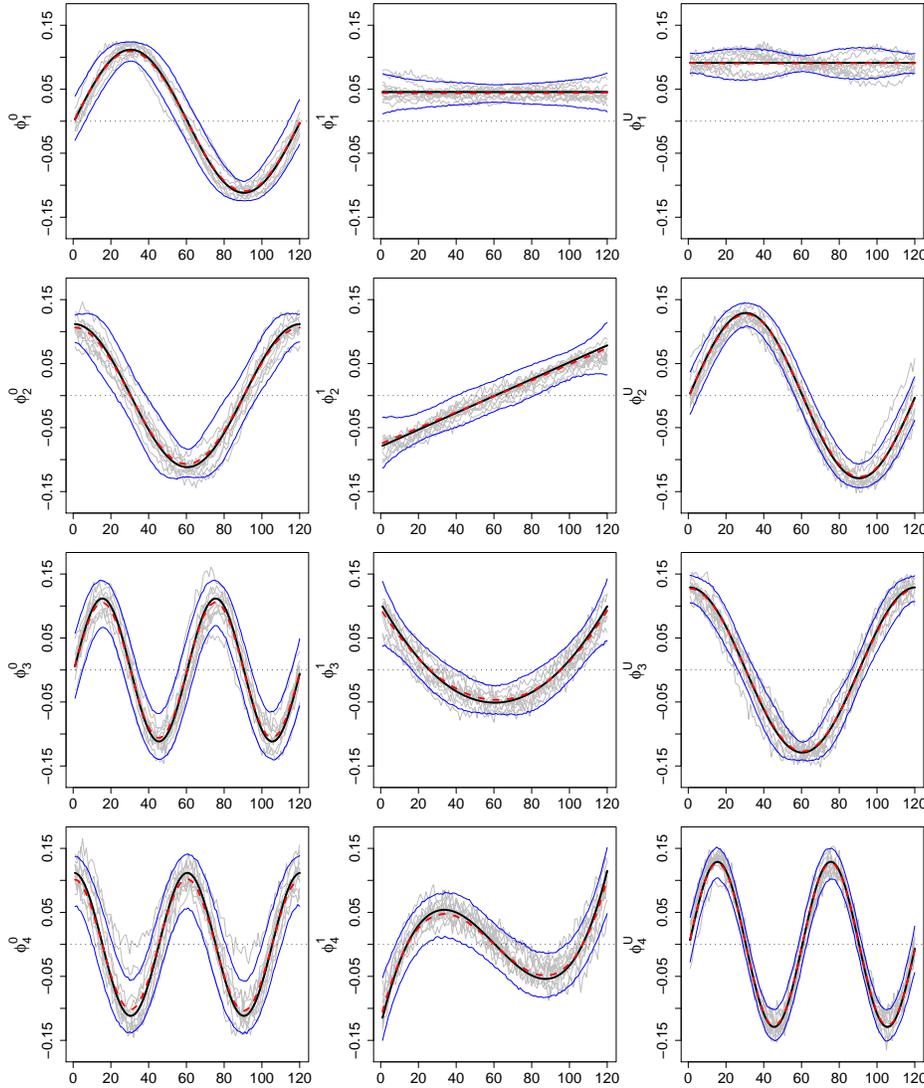
FIG 2. *True and estimated eigenfunctions* $\phi_k^X = (\phi_k^0, \phi_k^1)$ *and* ϕ_k^U, $k = 1, \ldots, 4$. *The left column gives results for the part* ϕ_k^0 *corresponding to the random functional intercept* $X_{i,0}$, *the middle column for the part* ϕ_k^1 *corresponding to the random functional slope* $X_{i,1}$, *and the right column for the component* ϕ_k^U *corresponding to the visit-specific functional deviation* U_{ij}. *Shown are the true function (thick black line), the mean of the estimated functions over 1000 simulations (dashed red line), the pointwise 5th and 95th percentiles of the estimated functions (blue), and the estimated functions from the first 50 simulations (grey). Simulations were based on model* (2.3) *with* $N_X = N_U = 4$, *a balanced design with* $I = 100$ *and* $J_i = 4$ *for all* i, *non-orthogonal* ϕ_k^0, ϕ_k^1 *and* ϕ_k^U *with unequal weight on* ϕ_k^1 *and* ϕ_k^0, *a mixture distribution for the scores* ξ_{ij} *and* ζ_{ijk}, *and no bivariate smoothing of the covariance functions.*

TABLE 1

True and estimated subject-specific and visit-specific scores ξ_{ik} and ζ_{ijk}. Given are summary statistics of the scaled differences $(\widehat{\xi}_{ik} - \xi_{ik})/\sqrt{\lambda_k}$ (top) and $(\widehat{\zeta}_{ijk} - \zeta_{ijk})/\sqrt{\nu_k}$ (bottom), $k = 1, \ldots, 4$. Simulations were based on model (2.3) with $N_X = N_U = 4$, a balanced design with $I = 100$ and $J_i = 4$ for all i, non-orthogonal ϕ_k^0, ϕ_k^1 and ϕ_k^U with unequal weight on ϕ_k^1 and ϕ_k^0, a mixture distribution for the scores ξ_{ij} and ζ_{ijk}, and no bivariate smoothing of the covariance functions.

Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
-2.39	-0.31	0.00	0.00	0.30	2.56
-2.52	-0.24	0.00	0.00	0.23	3.13
-3.42	-0.31	0.00	0.00	0.31	3.39
-4.17	-0.20	0.00	0.00	0.20	4.58
-2.80	-0.18	0.00	0.00	0.18	3.11
-3.68	-0.39	0.00	0.00	0.39	4.26
-2.77	-0.30	0.01	0.01	0.31	2.54
-3.38	-0.34	0.00	0.00	0.34	4.16

and right panels, respectively. Shown are the true function (thick black line), the mean of the estimated functions over 1000 simulations (dashed red line), the pointwise 5th and 95th percentiles of the estimated functions (blue), and the estimated functions from the first 100 simulations (grey). Note that the covariance functions, and thus the eigenfunctions, are not smoothed in this setting.

For all functions, the mean of the estimated functions is very close to the true function, and the variability around it is small. It can be noted that the ϕ_k^U are slightly better estimated. This is due to the fact that estimation of the covariance function $K_U(d, d')$ is based on $n = \sum J_i$ visits, while estimation of the covariance function $K_X(d, d')$ is based on only I subjects, with $I = n/4$ in this setting. In this case, ϕ_k^0 has a larger norm than ϕ_k^1, making estimation of this component easier. This is noticeable in a smaller variance for ϕ_k^0. Nevertheless, estimation of the ϕ_k^1 is also remarkably good. Overall, estimation of all functions is very good, even though: a) ϕ_k^0 and ϕ_k^1 are not mutually orthogonal; and b) each ϕ_k^U is actually identical to either ϕ_j^0 or ϕ_j^1 for some j. Our estimation procedure effectively separates the $\boldsymbol{X}$ and U processes, even in the most difficult of circumstances and with a moderate sample size.

Table 1 provides results for the scores ξ_{ik} and ζ_{ijk}, $k = 1, \ldots, 4$. Shown are summary statistics for the scaled differences between estimated and true scores, $(\widehat{\xi}_{ik} - \xi_{ik})/\sqrt{\lambda_k}$ and $(\widehat{\zeta}_{ijk} - \zeta_{ijk})/\sqrt{\nu_k}$, $k = 1, \ldots, 4$. The table illustrates that the majority of estimates lies close to the true scores, relative to the standard deviation of the score in question, even if the distribution of the estimates is more heavy tailed than in a normal distribution. This might be expected from the fact that the principal components ϕ_k^X and ϕ_k^U in model (2.3) are estimated and not observed.

Further figures in the supplementary material show results for the estimation of the variances, σ^2, λ_k and ν_k, $k = 1, \ldots, 4$. The $\widehat{\nu}_k$ are centered at the true values ν_k, with about 70% within 10% and more than 95% within 20% of the true value. The $\widehat{\lambda}_k$ show a slight downward bias and somewhat larger varia-

tion, reflecting the smaller effective sample size for estimation of these variance components. σ^2 is estimated almost as well as the ν_k.

Overall, the estimation procedure performed very well in a wide range of scenarios. Across simulations, we found the following similarities and differences. First, results improve for an increasing number of subjects I. As expected, a larger I decreases the variability of the estimated eigenfunctions, mean function, scores and variances. The slight downward bias in the $\widehat{\lambda}_k$ disappears with increasing I. Second, a balanced design (2a) improves results compared to an unbalanced design (2b) with the same number of subjects and visits. A balanced design leads to a) decreased variability in the estimated mean $\widehat{\eta}(d, T)$, as we estimate the mean under a working independence assumption before estimating the complex covariance structure b) decreased variability in the estimated eigenfunctions $\widehat{\phi}_K^X$ and decreased variability and small sample bias in the variances $\widehat{\lambda}_k$, $k = 1, \ldots, N_X$. This is similarly due to the fact that we estimate the covariance functions using least squares under a working independence assumption. Estimation of the ϕ_k^U is not much affected by how balanced the design is, although there is some indication that small estimates $\widehat{\lambda}_k$ are compensated for by a slight increase in the $\widehat{\nu}_k$. Third, results for normal scores (3a) and non-normal scores (3b) where virtually identical. This is expected, as BLUPs do not rely on a normality assumption and are thus robust to departures from normality. Still, it is reassuring to see this confirmed in practice. Forth, non-orthogonality of ϕ_k^0 and ϕ_k^1 (4b) does not affect results compared to orthogonality (4a). Even though in (4b), each ϕ_k^U is equal to either ϕ_j^0 or ϕ_j^1 for some j, estimation of the ϕ_k^U is equally good in both cases. The only consistent difference between the two designs is, that as in (4b) ϕ_k^0 has a larger part in the norm of ϕ_k^X, estimation of ϕ_k^0 improves somewhat, while estimation of ϕ_k^1 deteriorates slightly. Fifth, results excluding (5a) and including (5b) bivariate smoothing of the covariance functions were very similar, with the smooth version more effective at filtering out the measurement errors $\varepsilon_{ij}(d)$ and obtaining smooth eigenfunctions ϕ_k^X and ϕ_k^U.

Our sensitivity analyses indicate that results are not very sensitive to the choice of the mean function, with all three considered mean functions estimated well and unbiasedly. Large error variances increase the variability of all estimates. When signal-to-noise ratios become small due to very large error variances and small variances λ_k or ν_k, this leads to some underestimation of the magnitude (but not shape) of the eigenfunctions. This is due to the unbiasedness of estimation for the covariance functions, which are quadratic in the eigenfunctions, and the attenuation resulting from large variances in $|\operatorname{E}(X)| = \sqrt{\operatorname{E}(X^2) - \operatorname{Var}(X)} < \sqrt{\operatorname{E}(X^2)}$.

4.3. Computational efficiency

To investigate computation time we considered different combinations of number of subjects $I \in \{25, 50, 100, 200, 500, 1000\}$, number of observations per subject $J \in \{4, 8\}$, and number of sample points per curve $D \in \{50, 100, 200, 500\}$. All
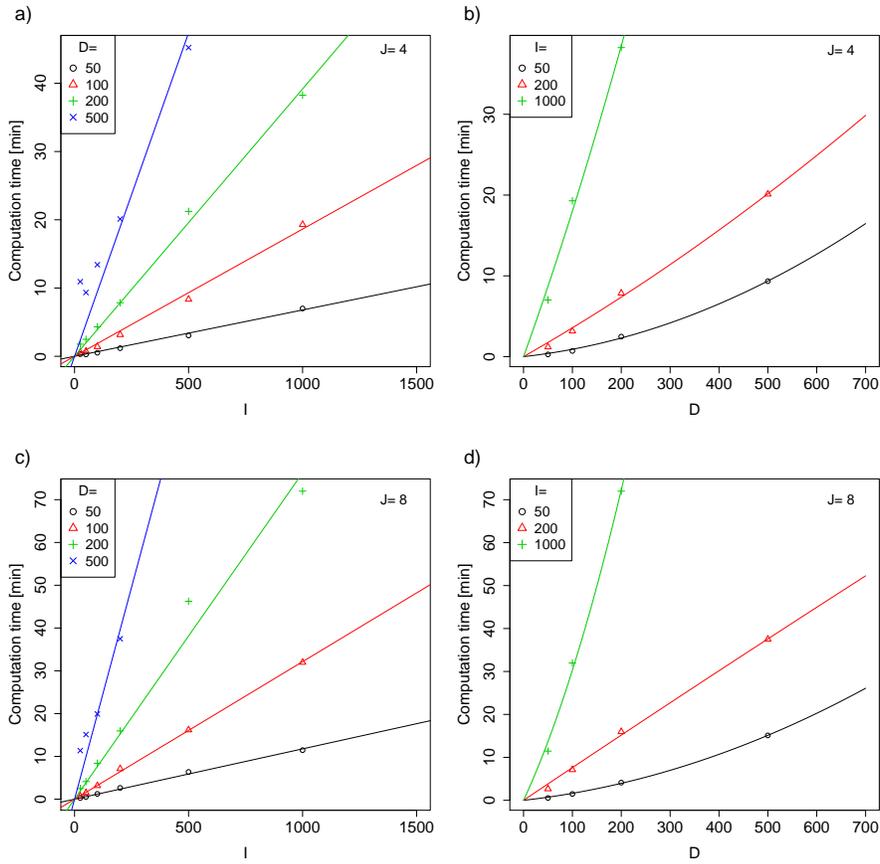
FIG 3. *Computation time for LFPCA for a simulated data set with the given number of subjects I and number of observations per subject J, and with D sample points per curve. Specifics of how computation time was measured are given in Section 4.3.*

other parameters were chosen as for the simulations described in Section 4.1, settings combination 2(a), 3(a), 4(a), 5(a).

Figure 3 provides the computation times. System times were (for practical reasons) measured on three different cluster nodes running 64-bit Red Hat Linux, with 2.3/ 2.6/3.0 GHz AMD Opteron Processors and 32 GB random access memory. Figures 3 a) and c) for $J = 4$ and $J = 8$ display computation time versus the number of subjects I, stratified by D. For example, computation time for $I = 100$ subjects with $J = 4$ visits and $D = 100$ points per curve was just 1.4 minutes, while computation for $I = 1,000$, $J = 8$ and $D = 200$ took 72 minutes. Figures 3 b) and d) display computation times versus sample points per curve, D, stratified by I. As suggested by Corollaries 1 and 2, computation time is roughly linear in I and between linear and quadratic in D. A linear

regression of the log computation time $\log(C)$ on $\log(I)$, $\log(J)$ and $\log(D)$ yields $\widehat{\log(C)} = -7.12 + 0.93\log(I) + 1.38\log(D) + 0.82\log(J)$. The coefficient for $\log(D)$ is indeed between 1 and 2, as expected from Corollaries 1 and 2. The coefficients for $\log(I)$ and $\log(J)$ are close to 1, but somewhat lower, reflecting that there are some parts of the estimation procedure not depending on I and J. The adjusted R^2 for the model is a high 0.98, indicating that good estimates of computation time on comparable machines can be obtained from this regression equation also for parameter combinations not considered here. Note, however, that for very large I and, especially, D, memory might be more of a concern than computational efficiency. In that case, one can replace our efficient matrix computations by less efficient methods that optimize memory usage.

5. Application

In this section, we use LFPCA to decompose the variability in the tractography data. We first provide the scientific background.

5.1. Background and scientific questions

Multiple sclerosis (MS) is a disorder of the central nervous system (CNS) [e.g. 4]. MS causes typical abnormalities on magnetic resonance imaging (MRI) scans of the brain and spinal cord, and consequently MRI has become the primary diagnostic tool for MS. MRI scanning is increasingly used to monitor disease progression and response to therapy and has become an important surrogate outcome measure in clinical trials.

Diffusion tensor imaging (DTI), in contrast to conventional MRI techniques, is able to resolve individual functional tracts within the CNS white matter, the primary target of MS. DTI is sensitive to diffusion anisotropy, which, in the brain and spinal cord, corresponds to the tendency of water to diffuse along axonal tracts [1]. A focus on one or several tracts with specific functional correlates can then help in understanding the neuroanatomical basis of disability in MS. Quantitative measures derived from DTI data include fractional anisotropy (FA), measuring the degree of anisotropy between 0 and 1 [2]. FA can be decreased in MS due to lesions, loss of myelin and axon damage [42, 45].

Measurement of tract-specific MRI indices has traditionally worked with averages along tracts, ignoring the spatial variation of those indices within tracts [25, 32, 36]. However, that spatial variation can be considerable. The extent to which accounting for this spatial variation can improve detection of abnormality, correlation with disability, or sensitivity to change across time, remains uncertain. The last of these is particularly relevant for monitoring individual patients in the clinic and for the design and powering of clinical trials of new drugs.

We are interested in using the full spatial information to gain a better understanding of differences between subjects both with respect to their mean tract profiles over time (static behavior) and to the changes in those tract profiles over time (dynamic behavior). Our data set includes measurements for 84 MS

patients and 28 controls with 1 to 8 complete visits, giving 308 visits overall. At each visit, we have measurements of FA and the diffusivities along several tracts in the brain, which were reconstructed using the tract finding algorithm of [28]. We will focus here on the corpus callosum, a tract connecting the two hemispheres of the brain. The 120 sample points - from the splenium (back of the head) to the genu/rostrum (closer to the eyes) - were chosen equidistantly between certain landmarks on the tract used for registration of curves between subjects [33]. The corpus callosum and its segmentation are illustrated in Figure 1 (top). Figure 1 (bottom) shows example profiles from two MS patients, illustrating the variability of profiles between subjects and within subjects over time. Visual examination of these tract profiles reveals variation within subjects across both space and time but no clear and consistent trend over time.

5.2. Application of LFPCA to the tractography data

As changes in MRI hardware over the five years of the study caused some variation in the measured MRI indices, we use a preprocessing step to remove differences due to variation in scanning technique. For each of the five scanning epochs, we estimate a mean profile for cases alone, using one visit per subject. This avoids confounding of disease status with epoch due to uneven distribution of cases and controls among epochs, and confounding by disease progression. We then subtract the difference of the epoch mean profile to the overall mean profile from all functional observations.

We obtain a decomposition of the variance using LFPCA. The time variable is centered by subject and standardized. For bivariate smoothing of the mean and covariance functions, we use tensor product penalized cubic regression splines with 30 knots per dimension, with smoothing parameters estimated using REML. A graph of the mean function $\eta(d, T)$ is given in the supplementary material. The mean profile is roughly constant over time, with some variation near areas of high curvature.

For a pre-specified level $L = 90\%$ of explained average variance $\int_0^1 \text{Var}\{Y_{ij}(s)\}ds$, LFPCA gives $N_X = N_U = 6$ principal components ϕ_k^X and ϕ_k^U for the $\boldsymbol{X}$ and U processes. The decomposition of the average variance is given in Table 2. 38% of the variation is explained by the first principal component for $\boldsymbol{X}$, ϕ_1^X, another 23% by the first principal component for U, ϕ_1^U. Overall, the first six components ϕ_k^X, $k = 1, \ldots, 6$, explain 55% of the average variance, indicating that the $\boldsymbol{X}$ process captures most of the variation in the data. Within $\boldsymbol{X}$, most of the variation is explained by the random functional intercept $X_{i,0}$, but the variance due to the subject-specific random slope is still large compared to the measurement error. Note also that the study period is much shorter than the disease duration for some of the patients, such that $X_{i,1}$ might still be of large practical relevance over many years. Within-curve measurement error is negligible due to a smoothing step during profile construction, explaining only 0.02%. Estimated variances were $\widehat{\lambda}_k = 0.316, 0.060, 0.030, 0.021, 0.015, 0.010$, $k = 1, \ldots, 6$, $\widehat{\nu}_k = 0.189, 0.045, 0.025, 0.017, 0.012, 0.011$, $k = 1, \ldots, 6$, and $\widehat{\sigma}^2 = 0.0002$.

k	ϕ_k^0	ϕ_k^1	ϕ_k^U	σ^2	cumulative
1	37.97	0.21	22.81	0.02	61.01
2	6.76	0.55	5.41		73.73
3	3.33	0.32	2.96		80.34
4	2.11	0.43	2.07		84.95
5	1.50	0.28	1.51		88.24
6	0.98	0.19	1.32		90.73
	52.65	1.98	36.08	0.02	90.73

Figure 4 shows the first three estimated principal components for the random intercept and slope process $\boldsymbol{X}$. The left column gives estimates for the ϕ_k^0, corresponding to the random functional intercept $X_{i,0}$. Depicted are estimates for the overall mean $\eta(d)$ (solid line), and for $\eta(d) \pm 2\sqrt{\lambda_k}\phi_k^0$, $k = 1, 2, 3$ (+ and -, respectively). The middle column gives the corresponding results for the random functional slope $X_{i,1}$. The right column shows boxplots for the estimates of the scores ξ_{ik} corresponding to (ϕ_k^0, ϕ_k^1), $k = 1, 2, 3$, by case/control group. Estimated scores for the two example patients with tract profiles shown in Figure 1 are indicated by A and B, respectively.

Positive loadings $\widehat{\xi}_{i1} > 0$ on the first component correspond to a lower mean function with a particularly deep FA dip in the isthmus (around 20), but only to small changes of profiles over time. For example, patient A with a much lower dip than B loads positively on this component, while B's loading is roughly zero. The second component is a mean contrast, with positive scores corresponding to a lower dip around 20 and a higher plateau around 60. The corresponding change over time is similar, if smaller in magnitude, and could explain how the differences in mean profiles evolved over time. The large positive score $\widehat{\xi}_{i2}$ in patient A corresponds to the large contrast between low dip and high plateau in this patient, which is very unpronounced in patient B (roughly zero score). The third component shows a shift of the location of the dip, which might reflect differences in anatomy that affect the tractography. This goes hand in hand with a slight further shifting and deepening (for negative scores) of the dip over time. For example, patient A, in contrast to patient B, exhibits more of a deepening of the dip and a shift to the left, with corresponding negative score $\widehat{\xi}_{i3}$. As mentioned above, these consistent changes over time are not immediately apparent from an examination of the tract profiles in Figure 1 but are clearly revealed by loadings on the principal components derived from the LFPCA analysis. In future work, we plan to examine whether these changes can portend disease course.

Figure 5 shows the corresponding results for the visit-specific functional deviation U. $\widehat{\phi}_1^U$ is similar in shape to $\widehat{\phi}_1^0$. Patient A at visit 8, for example, shows a lower profile than would be expected from the average evolution in this pa-

FIG 4. *The first three estimated principal components for the random intercept and slope process X. The left column gives estimates for the ϕ_k^0, corresponding to the random functional intercept $X_{i,0}$. Depicted are estimates for the overall mean $\eta(d)$ (solid line), and for $\eta(d) \pm 2\sqrt{\lambda_k}\phi_k^0$, $k = 1, 2, 3$ (+ and ·, respectively). The middle column gives the corresponding results for the random functional slope $X_{i,1}$. The right column shows boxplots for the estimates of the scores ξ_{ik} corresponding to (ϕ_k^0, ϕ_k^1), $k = 1, 2, 3$, by case/control group. Estimated scores for the two example patients with tract profiles shown in Figure 1 are indicated by A and B, respectively.*

FIG 5. *The first three estimated principal components for the visit-specific deviation process U. The left column gives results for the principal components ϕ_k^U, depicting estimates for the overall mean $\eta(d)$ (solid line), and for $\eta(d) \pm 2\sqrt{\nu_k}\phi_k^U$, $k = 1, 2, 3$ (+ and -, respectively). The right column shows boxplots for the estimates of the scores ζ_{ik} corresponding to ϕ_k^U, $k = 1, 2, 3$, by case/control group. Estimated scores for example visits of the two patients with tract profiles shown in Figure 1 are indicated by A (visit 8) and B (visit 2), respectively.*

tient over time, and consequently has a positive score $\widehat{\zeta}_{ij1}$, with the converse being true for patient B at visit 2. Components $\widehat{\phi}_2^U$ and $\widehat{\phi}_3^U$ seem to pick up variation at the ends of the tract as well as visit-to-visit shifts of the location of the dip, which might be due to measurement error. Note that the U process captures both measurement error and true biological fluctuations, which are impossible to separate without additional subject-matter insight. Filtering out these processes allows us to study the systematic trends modeled by $\boldsymbol{X}$.

Our model allows straightforward inclusion of additional covariates such as case/control status, disease severity, medication or age in the mean function η. In this study, however, we were interested in how the main variations in tract profiles and their changes over time differed by case/control group, i.e. in the covariance part of the model. When fixed effect group specific means are the target of inference, our approach could be used to improve the confidence band estimators on the group-specific mean difference. For example, the group-specific means could be estimated first under independence, then the covariances can be estimated using LFPCA, and then estimates of the group-specific mean differences can be further improved using the estimated covariance structure. The process can even be iterated until convergence.

Focusing on the covariances, we find a statistically significant difference in the distribution of the estimated scores $\widehat{\xi}_{i1}$ between MS patients and controls (p=0.0056 in a Mann-Whitney-Wilcoxon test; also in a linear regression adjusting for age and sex). The patient group in particular seems to have a higher mean and a heavier right tail. This could be an indication of a mixture in this group of patients who are more or less affected by MS along this particular tract. Potential loading-based clustering into patient subgroups will be of interest in future work. Interestingly, FA for this component is not decreased uniformly along the tract, but only posterior to the genu (ca. 1-100), with the decrease being especially pronounced in the area of the isthmus (ca. 20). Our results thus identify the region of the corpus callosum (the isthmus) where MS seems to take its greatest toll and also define the ways in which that portion of the tract changes from one year to the next. In future work, we plan to examine whether these changes can portend disease course. This result could not have been obtained by using the average FA instead of our functional approach.

6. Discussion

We have introduced methods for functional data that is observed at multiple time points for the same subject. Our methods can be viewed as extending longitudinal mixed effects models by replacing the random effects with random processes. Models are designed to decompose the longitudinal functional data into a time-dependent population average, baseline (or static) subject-specific variability, longitudinal (or dynamic) subject-specific variability, subject/visit-specific variability and measurement error. We propose an estimation procedure based on an eigen expansion that is highly computationally efficient and performs well in a wide range of simulations and in our application.

Our work is different from functional data methodology applied to the analysis of longitudinally observed scalar data [41, among others], but builds on methods from both functional and longitudinal data analysis [8, 34]. While the considered model shares similarities with models used by [15, 29], we do not assume Gaussianity and our approach is based on functional principal component analysis. In addition to computational advantages (compare Section 4.3 and [19]), this allows the extraction of the main differences between subjects

in the dynamics of their profiles over time, something of interest in many applications including our tractography study. Also, our work is different from methods for the 3-D analysis of subject-specific DTI studies [see for example 18]. It takes a functional data approach to tract data, as has recently been done for non-longitudinal DTI tractography data in [52].

Our approach can serve as a stepping-stone for further developments in the field of longitudinally observed functional data, and lends itself well to extensions. As our estimation procedure performs best when the number of time points per subject is balanced, it might be interesting to investigate further improvements, such as via iterations between mean and covariance estimation. Using an iterative approach, [49] find improvements to the integrated mean squared errors that are most pronounced for sparse functional data, where the number of sample points per curve is small. We will pursue such an approach in the future, in particular if dealing with sparse longitudinal functional data. While the functional random intercept-random slope model was sufficient for our application, it would also be interesting to apply our general model in more complex settings. And as our methods extract the main modes of variation in longitudinal functional data, including differences in mean curves and changes in curves over time, the associated scores could be of interest for further use in regression or classification.

Acknowledgements

Appendix A: Theoretical results and proofs

Theorem 1. *Estimates of the covariance functions in Step A can be expressed as* $\widehat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{c}_1$ *and* $\widehat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{c}_2$. *Here,* $\boldsymbol{\beta}_1$ *is the* $5 \times \{D(D-1)/2\}$ *matrix with column* $\{K_0(d,d'), K_{01}(d,d'), K_{01}(d',d), K_1(d,d'), K_U(d,d')\}$ *corresponding to* $d < d' \in \mathcal{D}$, *and* $\boldsymbol{\beta}_2$ *is the* $4 \times D$ *matrix with column* $\{K_0(d,d), K_{01}(d,d), K_1(d,d'), K_U(d,d) + \sigma^2\}$ *corresponding to* $d \in \mathcal{D}$. $\boldsymbol{X}_1$ *is the* $(\sum_{i=1}^I J_i^2) \times 5$ *matrix with rows* $(1, T_{ik}, T_{ij}, T_{ij}T_{ik}, \delta_{jk})$, $j, k = 1, \ldots, J_i, i = 1, \ldots, I$, *and* $\boldsymbol{X}_2$ *is the* $\{\sum_{i=1}^I J_i(J_i+1)/2\} \times 4$ *matrix with rows* $(1, T_{ik} + T_{ij}, T_{ij}T_{ik}, \delta_{jk})$, $j \leq k = 1, \ldots, J_i, i = 1, \ldots, I$. $\boldsymbol{c}_1$ *is the* $(\sum_{i=1}^I J_i^2) \times \{D(D-1)\}/2)$ *matrix with column* $(Y_{ij}(d)Y_{ik}(d'), j, k = 1, \ldots, J_i, i = 1, \ldots, I)$ *corresponding to* $d < d' \in \mathcal{D}$, *and* $\boldsymbol{c}_2$ *the* $\{\sum_{i=1}^I J_i(J_i+1)/2\} \times D$ *matrix with column* $(Y_{ij}(d)Y_{ik}(d), j \leq k = 1, \ldots, J_i, i = 1, \ldots, I)$ *corresponding to* $d \in \mathcal{D}$.

Proof. Consider least squares estimation of $(K_0(d,d'), K_{01}(d,d'), K_{01}(d',d), K_1(d,d'), K_U(d,d') + \sigma^2\delta_{dd'}; d \leq d' \in \mathcal{D})$ on the basis of (3.1). First, note that the design matrix in the corresponding linear regression is block diagonal, with

blocks corresponding to $(d, d'), d < d'$, containing entries $(1, T_{ik}, T_{ij}, T_{ij}T_{ik}, \delta_{jk})$ in the row corresponding to $Y_{ij}(d)Y_{ik}(d')$, and blocks corresponding to (d, d) containing entries $(1, T_{ik} + T_{ij}, T_{ij}T_{ik}, \delta_{jk})$ in the row corresponding to $Y_{ij}(d)Y_{ik}(d)$. Second, note that the blocks are identical for all pairs $(d, d'), d < d'$, respectively all pairs (d, d). Least squares estimates thus can be expressed as $\widehat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{c}_1$ and $\widehat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{c}_2$, where $\boldsymbol{X}_1'\boldsymbol{X}_1$ and $\boldsymbol{X}_2'\boldsymbol{X}_2$ are 5×5 and 4×4 matrices, respectively. $\square$

Proof of Corollary 1. From Theorem 1 and the analogous argument for the general model, only $(p^2+1) \times (p^2+1)$ matrices need to be inverted. Matrix inversion thus is of order $O(p^6)$. Matrix multiplication is of order $O(p^2 D^2 g)$, giving overall computational effort of order $O(\max(p^6, p^2 D^2 g))$. $\square$

Theorem 2. *The estimated BLUPs in* (3.2) *can be expressed as*

$$\widehat{\boldsymbol{b}} = \begin{pmatrix} \boldsymbol{I}_{N_X I} & -\boldsymbol{B}\boldsymbol{C}' \\ 0 & \boldsymbol{I}_{N_U n} \end{pmatrix} \begin{pmatrix} \boldsymbol{B} & 0 \\ 0 & \boldsymbol{R} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_{N_X I} & 0 \\ -\boldsymbol{C}\boldsymbol{B} & \boldsymbol{I}_{N_U n} \end{pmatrix} \boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{\eta}),$$

with $\boldsymbol{C} = \boldsymbol{E_I} \otimes \boldsymbol{\phi^{U'}}\boldsymbol{\phi^0} + \boldsymbol{T} \otimes \boldsymbol{\phi^{U'}}\boldsymbol{\phi^1}$, $\boldsymbol{R} = \boldsymbol{I_n} \otimes \mathrm{diag}(\nu_k/(\nu_k + \sigma^2)) + \boldsymbol{G}\boldsymbol{F}\boldsymbol{G}'$, $\boldsymbol{G} = \boldsymbol{E_I} \otimes \mathrm{diag}(\nu_k/(\nu_k+\sigma^2))\boldsymbol{\phi^{U'}}\boldsymbol{\phi^0} + \boldsymbol{T} \otimes \mathrm{diag}(\nu_k/(\nu_k+\sigma^2))\boldsymbol{\phi^{U'}}\boldsymbol{\phi^1}$, *and where* $\boldsymbol{B}$ *and* $\boldsymbol{F}$ *are block-diagonal with blocks*

$$\boldsymbol{B}_i = [J_i\boldsymbol{\Phi^{0'}}\boldsymbol{\Phi^0} + T_{i\bullet}(\boldsymbol{\Phi^{0'}}\boldsymbol{\Phi^1} + \boldsymbol{\Phi^{1'}}\boldsymbol{\Phi^0}) + T_{i\bullet}^2\boldsymbol{\Phi^{1'}}\boldsymbol{\Phi^1}$$
$$+ \mathrm{diag}(\sigma^2/\lambda_1, \ldots, \sigma^2/\lambda_{N_X})]^{-1},$$

$$\boldsymbol{F}_i = [J_i\boldsymbol{\Phi^{0'}}\boldsymbol{L}\boldsymbol{\Phi^0} + T_{i\bullet}(\boldsymbol{\Phi^{0'}}\boldsymbol{L}\boldsymbol{\Phi^1} + \boldsymbol{\Phi^{1'}}\boldsymbol{L}\boldsymbol{\Phi^0}) + T_{i\bullet}^2\boldsymbol{\Phi^{1'}}\boldsymbol{L}\boldsymbol{\Phi^1}$$
$$+ \mathrm{diag}(\sigma^2/\lambda_1, \ldots, \sigma^2/\lambda_{N_X})]^{-1},$$

$i = 1, \ldots, I$, *denoting* $T_{i\bullet} = \sum_{j=1}^{J_i} T_{ij}$, $T_{i\bullet}^2 = \sum_{j=1}^{J_i} T_{ij}^2$, *and* $\boldsymbol{L} = \boldsymbol{I_D} - \boldsymbol{\Phi^U}\mathrm{diag}(\nu_k/(\nu_k+\sigma^2))\boldsymbol{\Phi^{U'}}$. *Here,* $\mathrm{diag}(a_k)$ *denotes a diagonal matrix with entries* a_k, $k = 1, \ldots, N_U$, *and we suppress for simplicity hat notation that indicates estimated quantities.*

Proof. For simplicity, we suppress hat notation that indicates estimated quantities in the computation of the EBLUPs. Using the Woodbury formula, we obtain

$$\widehat{\boldsymbol{b}} = (\boldsymbol{Z}'\boldsymbol{Z} + \sigma^2\boldsymbol{D}^{-1})^{-1}\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{\eta}).$$

Using the Schur complement $\boldsymbol{S}$, write

$$(\boldsymbol{Z}'\boldsymbol{Z} + \sigma^2\boldsymbol{D}^{-1})^{-1} = \begin{pmatrix} \boldsymbol{Z}_X'\boldsymbol{Z}_X + \sigma^2\boldsymbol{D}_X^{-1} & \boldsymbol{Z}_X'\boldsymbol{Z}_U \\ \boldsymbol{Z}_U'\boldsymbol{Z}_X & \boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2\boldsymbol{D}_U^{-1} \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} \boldsymbol{I}_{N_X I} & -\boldsymbol{A}^{-1}\boldsymbol{Z}_X'\boldsymbol{Z}_U \\ 0 & \boldsymbol{I}_{N_U n} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}^{-1} & 0 \\ 0 & \boldsymbol{S}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_{N_X I} & 0 \\ -\boldsymbol{Z}_U'\boldsymbol{Z}_X\boldsymbol{A}^{-1} & \boldsymbol{I}_{N_U n} \end{pmatrix},$$

where $\boldsymbol{A} = \boldsymbol{Z}_X'\boldsymbol{Z}_X + \sigma^2\boldsymbol{D}_X^{-1}$ and $\boldsymbol{S} = (\boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2\boldsymbol{D}_U^{-1}) - \boldsymbol{Z}_U'\boldsymbol{Z}_X\boldsymbol{A}^{-1}\boldsymbol{Z}_X'\boldsymbol{Z}_U$.

Using properties of the Kronecker product, we have

$$
\begin{aligned}
\boldsymbol{A} \;=\;& \operatorname{diag}(J_1,\ldots,J_I) \otimes \boldsymbol{\Phi^{0\prime}\Phi^0} + \operatorname{diag}(T_{1\bullet},\ldots,T_{I\bullet}) \otimes (\boldsymbol{\Phi^{0\prime}\Phi^1} + \boldsymbol{\Phi^{1\prime}\Phi^0}) \\
&+ \operatorname{diag}(T_{1\bullet}^2,\ldots,T_{I\bullet}^2) \otimes \boldsymbol{\Phi^{1\prime}\Phi^1} + \boldsymbol{I_I} \otimes \operatorname{diag}(\sigma^2/\lambda_1,\ldots,\sigma^2/\lambda_{N_X}).
\end{aligned}
$$

Thus, $\boldsymbol{A}$ is a block-diagonal matrix with I blocks $\boldsymbol{A}_i$ of size $N_X \times N_X$, and $\boldsymbol{B} = \boldsymbol{A}^{-1}$ can be computed as a block-diagonal matrix with the ith block of size $N_X \times N_X$ of the form

$$
\begin{aligned}
\boldsymbol{B}_i \;=\;& [J_i \boldsymbol{\Phi^{0\prime}\Phi^0} + T_{i\bullet}(\boldsymbol{\Phi^{0\prime}\Phi^1} + \boldsymbol{\Phi^{1\prime}\Phi^0}) + T_{i\bullet}^2 \boldsymbol{\Phi^{1\prime}\Phi^1} \\
&\qquad + \operatorname{diag}(\sigma^2/\lambda_1,\ldots,\sigma^2/\lambda_{N_X})]^{-1}.
\end{aligned}
$$

Analogously,

$$
\begin{aligned}
(\boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2 \boldsymbol{D}_U^{-1})^{-1} \;=\;& (\boldsymbol{I_n} \otimes \boldsymbol{\Phi^{U\prime}\Phi^U} + \boldsymbol{I_n} \otimes \operatorname{diag}(\sigma^2/\nu_1,\ldots,\sigma^2/\nu_{N_U}))^{-1} \\
=\;& \boldsymbol{I_n} \otimes \operatorname{diag}(\nu_1/(\nu_1+\sigma^2),\ldots,\nu_{N_U}/(\nu_{N_U}+\sigma^2))
\end{aligned}
$$

can be computed explicitly, as the columns of $\boldsymbol{\Phi_U}$ are orthonormal by construction. And finally, using the Woodbury formula again, the Schur complement $\boldsymbol{S}$ can be inverted as

$$
\begin{aligned}
\boldsymbol{R} = \boldsymbol{S}^{-1} \;=\;& (\boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2 \boldsymbol{D}_U^{-1})^{-1} + (\boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2 \boldsymbol{D}_U^{-1})^{-1} \boldsymbol{Z}_U'\boldsymbol{Z}_X \times \\
& [\boldsymbol{A} - \boldsymbol{Z}_X'\boldsymbol{Z}_U(\boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2 \boldsymbol{D}_U^{-1})^{-1} \times \\
& \boldsymbol{Z}_U'\boldsymbol{Z}_X]^{-1} \boldsymbol{Z}_X'\boldsymbol{Z}_U(\boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2 \boldsymbol{D}_U^{-1})^{-1} \\
=\;& \boldsymbol{I_n} \otimes \operatorname{diag}(\nu_k/(\nu_k+\sigma^2)) + \boldsymbol{GFG'},
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{C} \;=\;& \boldsymbol{Z}_U'\boldsymbol{Z}_X = \boldsymbol{E_I} \otimes \phi^{U\prime}\phi^0 + \boldsymbol{T} \otimes \phi^{U\prime}\phi^1, \\
\boldsymbol{G} \;=\;& (\boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2 \boldsymbol{D}_U^{-1})^{-1} \boldsymbol{C} \\
=\;& \boldsymbol{E_I} \otimes \operatorname{diag}(\nu_k/(\nu_k+\sigma^2))\phi^{U\prime}\phi^0 + \boldsymbol{T} \otimes \operatorname{diag}(\nu_k/(\nu_k+\sigma^2))\phi^{U\prime}\phi^1,
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{H} \;=\;& \boldsymbol{Z}_X'\boldsymbol{Z}_U(\boldsymbol{Z}_U'\boldsymbol{Z}_U + \sigma^2 \boldsymbol{D}_U^{-1})^{-1}\boldsymbol{Z}_U'\boldsymbol{Z}_X \\
=\;& (\boldsymbol{E_I}' \otimes \boldsymbol{\Phi^{0\prime}\Phi^U} + \boldsymbol{T}' \otimes \boldsymbol{\Phi^{1\prime}\Phi^U}) \times \\
& \qquad\qquad (\boldsymbol{I_n} \otimes \operatorname{diag}(\nu_k/(\nu_k+\sigma^2)))(\boldsymbol{E_I} \otimes \boldsymbol{\Phi^{U\prime}\Phi^0} + \boldsymbol{T} \otimes \boldsymbol{\Phi^{U\prime}\Phi^1}) \\
=\;& \operatorname{diag}(J_i) \otimes \boldsymbol{\Phi^{0\prime}\Phi^U} \operatorname{diag}(\nu_k/(\nu_k+\sigma^2))\boldsymbol{\Phi^{U\prime}\Phi^0} \\
& + \operatorname{diag}(T_{i\bullet}) \otimes \boldsymbol{\Phi^{0\prime}\Phi^U} \operatorname{diag}(\nu_k/(\nu_k+\sigma^2))\boldsymbol{\Phi^{U\prime}\Phi^1} \\
& + \operatorname{diag}(T_{i\bullet}) \otimes \boldsymbol{\Phi^{1\prime}\Phi^U} \operatorname{diag}(\nu_k/(\nu_k+\sigma^2))\boldsymbol{\Phi^{U\prime}\Phi^0} \\
& + \operatorname{diag}(T_{i\bullet}^2) \otimes \boldsymbol{\Phi^{1\prime}\Phi^U} \operatorname{diag}(\nu_k/(\nu_k+\sigma^2))\boldsymbol{\Phi^{U\prime}\Phi^1}
\end{aligned}
$$

is again block-diagonal with $N_X \times N_X$ blocks, as is $\boldsymbol{A}$, such that $\boldsymbol{F} = [\boldsymbol{A} - \boldsymbol{H}]^{-1}$ can be computed by inverting each block

$$
\begin{aligned}
\boldsymbol{A}_i - \boldsymbol{H}_i \;=\;& [J_i \boldsymbol{\Phi^{0\prime}L\Phi^0} + T_{i\bullet}(\boldsymbol{\Phi^{0\prime}L\Phi^1} + \boldsymbol{\Phi^{1\prime}L\Phi^0}) + T_{i\bullet}^2 \boldsymbol{\Phi^{1\prime}L\Phi^1} \\
&\qquad + \operatorname{diag}(\sigma^2/\lambda_1,\ldots,\sigma^2/\lambda_{N_X})],
\end{aligned}
$$

with $\boldsymbol{L} = \boldsymbol{I_D} - \boldsymbol{\Phi}^{\boldsymbol{U}} \operatorname{diag}(\nu_k/(\nu_k + \sigma^2))\boldsymbol{\Phi}^{\boldsymbol{U}'}$, separately. $\qquad\square$

Proof of Corollary 2. From Theorem 2, only matrices of size $N_X \times N_X$ need to be inverted to compute the EBLUPs, giving computational effort of order $O(IN_X^3)$. Usage of the block structure for all matrices reduces computation for the matrix multiplications. For example, multiplication of the $(nN_U + IN_X) \times nD$ and $nD \times 1$ matrices $\boldsymbol{Z}'$ and $(\boldsymbol{Y} - \boldsymbol{\eta})$, usually an $O(nD(nN_U + IN_X))$ operation, here reduces to I multiplications of $N_X \times D$ with $D \times 1$ matrices and n multiplications of $N_U \times D$ with $D \times 1$ matrices. Similarly bookkeeping for the other operations leads to the overall effort of order $O(nD(N_U + N_X I/n))$. $\qquad\square$

Proof of Lemma 1 and Corollary 3. Iterated expectations give us

$$\operatorname{Var}\{Y_{ij}(s)\} = \operatorname{E}\{\operatorname{Var}[Y_{ij}(s)|T_{ij}]\} + \operatorname{Var}\{\operatorname{E}[Y_{ij}(s)|T_{ij}]\}.$$

As $\operatorname{E}\{Y_{ij}(s)|T_{ij}\} = 0$,

$$
\begin{aligned}
\int_0^1 \operatorname{Var}\{Y_{ij}(s)\}ds \;&=\; \int_0^1 \operatorname{E}\{\sum_{k=1}^{\infty} \lambda_k[\phi_k^0(s) + T_{ij}\phi_k^1(s)]^2 + \sum_{k=1}^{\infty} \nu_k(\phi_k^U(s))^2 + \sigma^2\}ds \\
&=\; \int_0^1 \sum_{k=1}^{\infty} \lambda_k[(\phi_k^0(s))^2 + 2\operatorname{E}(T_{ij})\phi_k^0(s)\phi_k^1(s) + \operatorname{E}(T_{ij}^2)(\phi_k^1(s))^2] \\
&\qquad\qquad\qquad\qquad\qquad\quad + \sum_{k=1}^{\infty} \nu_k(\phi_k^U(s))^2 + \sigma^2 ds.
\end{aligned}
$$

Now consider the case where $\operatorname{E}(T_{ij}) = 0$ and $\operatorname{Var}(T_{ij}) = 1$. In this case, we have

$$
\begin{aligned}
\int_0^1 \operatorname{Var}\{Y_{ij}(s)\}ds &= \sum_{k=1}^{\infty} \lambda_k \int_0^1 (\phi_k^0(s))^2 + (\phi_k^1(s))^2 ds + \sum_{k=1}^{\infty} \nu_k \int_0^1 (\phi_k^U(s))^2 ds + \sigma^2 \\
&= \sum_{k=1}^{\infty} \lambda_k + \sum_{k=1}^{\infty} \nu_k + \sigma^2
\end{aligned}
$$

due to the orthonormality of the eigenfunctions. $\qquad\square$

Appendix B: Estimation of the general functional mixed model

Estimation for model (2.2) proceeds in the same way as for model (2.1). In this section, we briefly point out the necessary minor adjustments. An estimate of the mean function $\eta(d, \boldsymbol{Z}_{ij})$ can again be obtained under a working independence assumption and under the specified model. For example, under the specification $\eta(d, \boldsymbol{Z}_{ij}) = \eta_1(d, Z_{ij,1}) + \cdots + \eta_m(d, Z_{ij,m})$, bivariate smoothing in an additive model can be used.

In the estimation of the covariance functions, equation (3.1) now becomes

$$
\begin{aligned}
\mathrm{E}\{Y_{ij}(d)Y_{ik}(d')\} &= \boldsymbol{V}'_{ij}\boldsymbol{K}_X(d,d')\boldsymbol{V}_{ij} + [K_U(d,d') + \sigma^2\delta_{dd'}]\delta_{jk} \\
&= \sum_{l=1}^{p}\sum_{m=1}^{p} V_{ijl}V_{ijm}K_{lm}(d,d') + [K_U(d,d') + \sigma^2\delta_{dd'}]\delta_{jk},
\end{aligned}
$$

where $\boldsymbol{V}_{ij} = (V_{ij1}, \ldots, V_{ijp})$, and the three step estimation procedure for the covariance functions extends straightforwardly. The size of the matrix to be inverted during step 1 increases from 5×5 to $(p^2 + 1) \times (p^2 + 1)$.

Similarly, estimation of the eigenfunctions using the spectral decomposition, and estimation of the scores using best linear unbiased prediction, proceeds completely analogously, keeping in mind that $\{\phi_k^X(\cdot) = (\phi_k^1(\cdot), \ldots, \phi_k^p(\cdot)), k = 1, 2, \ldots\}$ now form an orthonormal basis for the $(L^2[0, 1])^p$.

Choice of N_X and N_U can again proceed using the proportion of variance explained. Standardization of variables in $\boldsymbol{V}_{ij}$ is recommended. Note, however, an additional complication in the higher-dimensional case. If some of the covariates in $\boldsymbol{V}_{ij}$ are correlated, corresponding additional terms will appear in the expansion of $\int \mathrm{Var}\{Y_{ij}(d)\}$. The eigenvalues λ_k might then somewhat over- or underrepresent the relative importance of the corresponding component ϕ_k^X in explaining the variation in $Y_{ij}(d)$. If strong correlations are a concern, additional measures, such as the use of orthogonal polynomials in the case of polynomial $\boldsymbol{V}_{ij}$, should be taken.

Supplementary Material

Supplementary materials to "Longitudinal functional principal component analysis" by S. Greven, C. Crainiceanu, B. Caffo and D. Reich (doi: 10.1214/10-EJS575SUPP).

References

[1] BASSER, P., MATTIELLO, J. and LEBIHAN, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical Journal* **66** 259–267.

[2] BASSER, P. J. and PIERPAOLI, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance, Series B* **111** 209–219.

[3] BRUMBACK, B. A. and RICE, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association* 961–976. MR1649194

[4] CALABRESI, P. A. (2008). Multiple sclerosis and demyelinating conditions of the central nervous system. In *Cecil Medicine* 23rd ed. (L. Goldman and D. A. Ausiello, eds.) Saunders Elsevier.

[5] CRAINICEANU, C. and RUPPERT, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B* **66** 165-185. MR2035765

[6] CRAINICEANU, C. M., STAICU, A. M. and DI, C. Z. (2009). Generalized Multilevel Functional Regression. *Journal of the American Statistical Association* **104** 1550–1561.

[7] DI, C. Z., CRAINICEANU, C. M., CAFFO, B. S. and PUNJABI, N. M. (2008). Multilevel functional principal component analysis. *Annals of Applied Statistics* **3** 458-488.

[8] DIGGLE, P., HEAGERTY, P., LIANG, K. Y. and ZEGER, S. (2002). *Analysis of longitudinal data.* Oxford University Press, USA. MR2049007

[9] FAN, J. and GIJBELS, I. (1996). *Local polynomial modelling and its applications.* CRC Press. MR1383587

[10] FERRATY, F. and VIEU, P. (2006). *Nonparametric functional data analysis: theory and practice.* Springer Verlag. MR2229687

[11] GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach.* Chapman & Hall Ltd. MR1270012

[12] GREVEN, S. and KNEIB, T. (2010). On the Behaviour of Marginal and Conditional Akaike Information Criteria in Linear Mixed Models. *Biometrika* to appear.

[13] GREVEN, S., CRAINICEANU, C., CAFFO, B. and REICH D. (2010). Supplement to "Longitudinal functional principal component analysis." DOI: 10.1214/10-EJS575SUPP

[14] GREVEN, S., CRAINICEANU, C. M., KÜCHENHOFF, H. and PETERS, A. (2008). Restricted Likelihood Ratio Testing for Zero Variance Components in Linear Mixed Models. *Journal of Computational and Graphical Statistics* **17** 870–891.

[15] GUO, W. (2002). Functional mixed effects models. *Biometrics* **58** 121–128. MR1891050

[16] GUO, W. (2004). Functional data analysis in longitudinal settings using smoothing splines. *Statistical methods in medical research* **13** 49. MR2037869

[17] HALL, P., MÜLLER, H. G. and YAO, F. (2008). Modelling sparse generalized longitudinal observations with latent Gaussian processes. *Journal of the Royal Statistical Society: Series B* **70** 703–723. MR2523900

[18] HEIM, S., FAHRMEIR, L., EILERS, P. and MARX, B. (2007). 3D space-varying coefficient models with application to diffusion tensor imaging. *Computational Statistics & Data Analysis* **51** 6212–6228. MR2407709

[19] HERRICK, R. C. and MORRIS, J. S. (2006). Wavelet-Based Functional Mixed Model Analysis: Computation Considerations. In *Proceedings, Joint Statistical Meetings, ASA Section on Statistical Computing.*

[20] KARHUNEN, K. (1947). Über Lineare Methoden in der Wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae* **37** 1-79. MR0023013

[21] KRIVOBOKOVA, T. and KAUERMANN, G. (2007). A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association* **102** 1328–1337. MR2412553

[22] LAIRD, N. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963-974.

[23] Liang, H., Wu, H. and Zou, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika* **95** 773–778.

[24] Lin, X. and Carroll, R. J. (2000). Nonparametric Function Estimation for Clustered Data When the Predictor Is Measured Without/With Error. *Journal of the American Statistical Association* **95** 520-534. MR1803170

[25] Lin, F., Yu, C., Jiang, T., Li, K., Li, X., Qin, W., Sun, H. and Chan, P. (2006). Quantitative analysis along the pyramidal tract by length-normalized parameterization based on diffusion tensor tractography: application to patients with relapsing neuromyelitis optica. *NeuroImage* **33** 154–160.

[26] Loeve, M. (1945). Fonctions aléatoires du second ordre. *Comptes Rendus Académie des Sciences* **220** 380.

[27] Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A.* 415–446.

[28] Mori, S., Crain, B. J., Chacko, V. and Van Zijl, P. C. M. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology* **45** 265–269.

[29] Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68** 179-199. MR2188981

[30] Müller, H. G. (2005). Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics* **32** 223–240. MR2188671

[31] Müller, H. G. and Zhang, Y. (2005). Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* **61** 1064-1075. MR2216200

[32] Oh, J. S., Song, I. C., Lee, J. S., Kang, H., Park, K. S., Kang, E. and Lee, D. S. (2007). Tractography-guided statistics (TGIS) in diffusion tensor imaging for the detection of gender difference of fiber integrity in the midsagittal and parasagittal corpora callosa. *Neuroimage* **36** 606–616.

[33] Ozturk, A., Smith, S., Gordon-Lipkin, E., Harrison, D., Shiee, N., Pham, D., Caffo, B., Calabresi, P. and Reich, D. (2009). MRI of the corpus callosum in multiple sclerosis: association with disability. *Multiple Sclerosis* to appear.

[34] Ramsay, J. O. and Silverman, B. (2005). *Functional data analysis*, 2nd ed. Springer. MR2168993

[35] Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52** 447–458. MR0205392

[36] Reich, D. S., Smith, S. A., Zackowski, K. M., Gordon-Lipkin, E. M., Jones, C. K., Farrell, J. A. D., Mori, S., van Zijl, P. C. M. and Calabresi, P. A. (2007). Multiparametric magnetic resonance imaging analysis of the corticospinal tract in multiple sclerosis. *Neuroimage* **38** 271–279.

[37] Rice, J. A. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statistica Sinica* **14** 631-647. MR2087966

[38] RICE, J. A. and SILVERMAN, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B* **53** 233–243. MR1094283

[39] RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression.* Cambridge University Press. MR1998720

[40] STAICU, A. M., CRAINICEANU, C. M. and CARROLL, R. J. (2010). Fast Methods for Spatially Correlated Multilevel Functional Data. *Biostatistics* to appear.

[41] STANISWALIS, J. G. and LEE, J. J. (1998). Nonparametric Regression Analysis of Longitudinal Data. *Journal of the American Statistical Association* **93** 1403–1404. MR1666636

[42] TIEVSKY, A. L., PTAK, T. and FARKAS, J. (1999). Investigation of apparent diffusion coefficient and diffusion tensor anisotropy in acute and chronic multiple sclerosis lesions. *American Journal of Neuroradiology* **20** 1491-1499.

[43] VAIDA, F. and BLANCHARD, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92** 351-370. MR2201364

[44] VERBEKE, G. and MOLENBERGHS, G. (2009). *Linear mixed models for longitudinal data.* Springer. MR1880596

[45] WERRING, D., CLARK, C., BARKER, G., THOMPSON, A. and MILLER, D. (1999). Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology* **52** 1626-1632.

[46] WITELSON, S. F. (1989). Hand and sex differences in the isthmus and genu of the human corpus callosum: a postmortem morphological study. *Brain* **112** 799-835.

[47] WOOD, S. N. (2006). *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC. MR2206355

[48] WU, H. and ZHANG, J. T. (2006). *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches.* Wiley-Blackwell. MR2216899

[49] YAO, F. and LEE, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society, Series B* **68** 3-25. MR2212572

[50] YAO, F., MÜLLER, H. G. and WANG, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100** 577–590. MR2160561

[51] YAO, F., CLIFFORD, A. J., DUEKER, S. R., FOLLETT, J., LIN, Y., BUCHHOLZ, B. A. and VOGEL, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* **59** 676–685. MR2004273

[52] ZHU, H., STYNER, M., TANG, N., LIU, Z., LIN, W. and GILMORE, J. (2010). FRATS: Functional Regression Analysis of DTI Tract Statistics. *IEEE Transactions on Medical Imaging* **29** 1039–1049.