9-9-2010

LONGITUDINAL PENALIZED FUNCTIONAL REGRESSION

Jeff Goldsmith
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, jgoldsmi@jhsph.edu

Ciprian M. Crainiceanu
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Brian Caffo
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Daniel Reich
Translational Neurology Unit, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, Johns Hopkins University, Department of Neurology

Longitudinal Penalized Functional Regression

JEFF GOLDSMITH, CIPRIAN M. CRAINICEANU, BRIAN CAFFO AND DANIEL REICH

September 9, 2010

Abstract

We propose a new regression model and inferential tools for the case when both the outcome and the functional exposures are observed at multiple visits. This data structure is new but increasingly present in applications where functions or images are recorded at multiple times. This raises new inferential challenges that cannot be addressed with current methods and software. Our proposed model generalizes the Generalized Linear Mixed Effects Model (GLMM) by adding functional predictors. Smoothness of the functional coefficients is ensured using roughness penalties estimated by Restricted Maximum Likelihood (REML) in a corresponding mixed effects model. This method is computationally feasible and is applicable when the functional predictors are measured densely, sparsely or with error; code implementing the proposed procedure is freely available. Methods are applied to a longitudinal diffusion tensor imaging (DTI) study relating changes in the microstructure of intracranial white matter tracts to cognitive disability in multiple sclerosis patients, but we note that the discussed data structure is increasingly common and our methods apply generally. An online appendix compares two implementations, one likelihood-based and the other Bayesian, and provides the software used in simulations.

Keywords: Functional Regression, Longitudinal Regression, Smoothing Splines, Mixed Models.

1 Introduction

Traditionally, longitudinal studies have collected scalar measurement on subjects over time. As technologies for the collection and storage of larger measurements have become widely available, longitudinal studies have begun to collect functional observations on subjects over several visits. However, techniques for the incorporation of functional covariates in longitudinal regression models are not currently available. We address this problem by introducing a novel longitudinal functional regression model that

adds subject-specific random effects to the well-studied cross sectional functional regression model. We also develop inferential techniques for all parameters in this new model, and implement these methods in computationally efficient and publicly available software. Our proposed approach is based on the single-level functional regression model developed in [13] but is extended to the statistically novel and scientifically useful longitudinal setting. We distinguish our data structure an inferential framework from the rich work in single-level functional regression [6, 12, 20, 24, 27, 28], from Functional ANOVA methods which decompose observed functions into group and subject components [3, 15], and from the important work in which scalar longitudinal observations are treated as a sparsely observed functional covariates [16, 23, 32].

In this paper we are motivated by a longitudinal diffusion tensor imaging (DTI) study, designed to investigate the effect of multiple sclerosis on patient disability. Multiple sclerosis is an autoimmune disease that results in the formation of lesions in the cerebral white matter and leads to severe patient disability; DTI is a noninvasive imaging procedure that yields quantitative information about tissue microstructure in individual functional systems in the brain. By collecting longitudinal information about patient cognitive function and about disease progression via DTI, researchers hope to better understand the relationship between MS and disability.

From this study, we have densely sampled mean and parallel diffusivity measurements from several white matter tracts, as well as cognitive disability measurements for each patient over several visits. In Figure 1 we display a functional predictor and cognitive disability outcome for two subjects over time; the study contains 100 subjects with a median of 3 visits roughly 6 months apart. We stress that this data structure, with very large predictors and scalar outcomes observed longitudinally, is increasingly common. Moreover, we emphasize that our methods are motivated by this study, but are generally applicable. A single-level analysis of these data was presented in [13].

More formally, we consider the setting in which we observe for each subject $1 \leq i \leq I$ at each visit $1 \leq j \leq J_i$ data of the form $[Y_{ij}, W_{ij1}(s), \ldots, W_{ijL}(s), X_{ij}]$, where Y_{ij} is a scalar outcome, $W_{ijl}(s) \in \mathcal{L}^2[0,1]$, $1 \leq l \leq L$, are functional covariates, and X_{ij} is a row vector of scalar covariates. We propose the

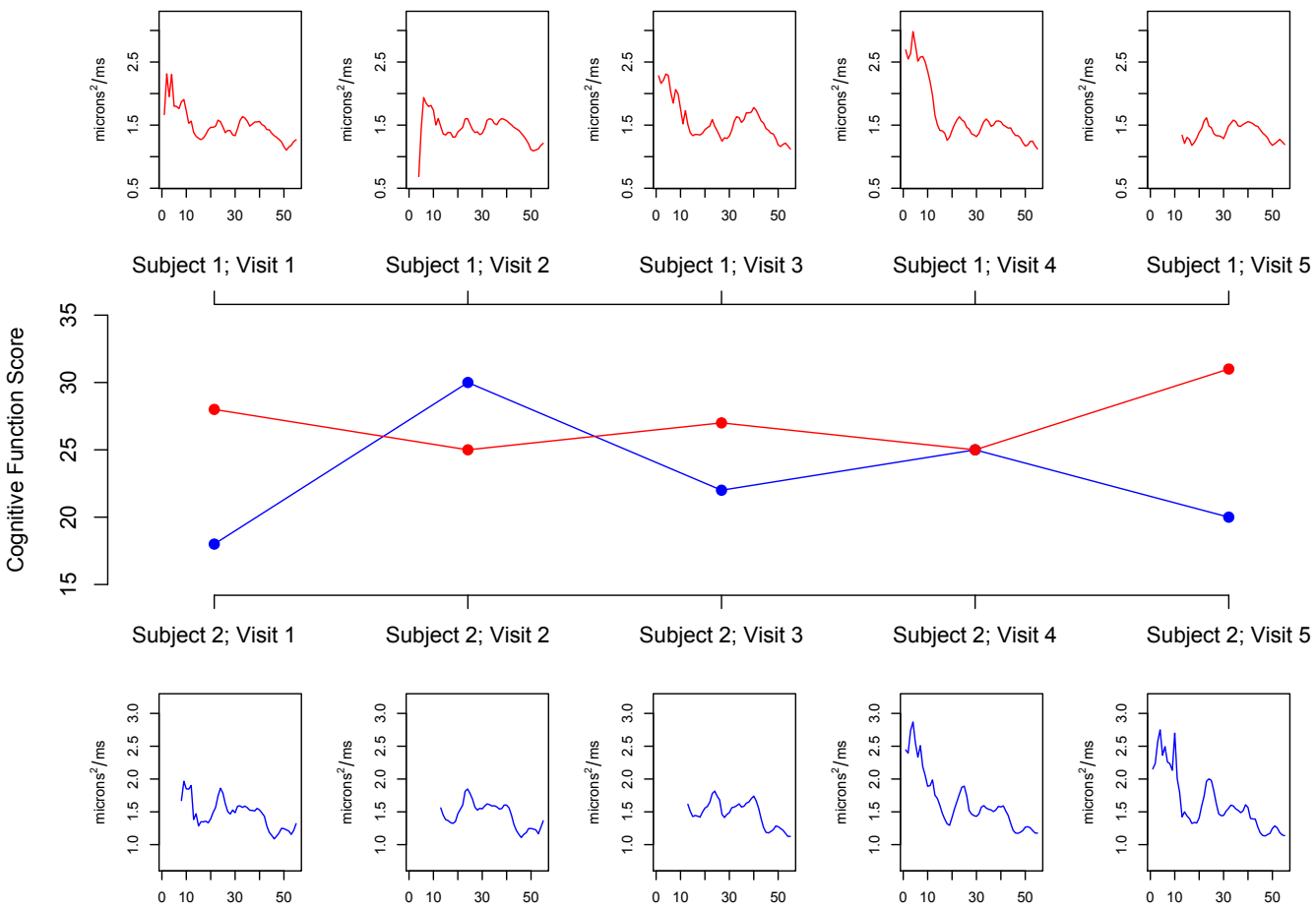

Figure 1: Structure of the data. The center panel displays the scalar PASAT-3 cognitive disability measure for two subjects over five visits; above and below are the parallel diffusivity tract profiles of the right corticospinal tract corresponding to each subject-visit outcome, which we use as a functional regressor.

longitudinal functional regression outcome model

$$\begin{aligned} Y_{ij} &\sim \text{EF}(\mu_{ij}, \eta) \\ g(\mu_{ij}) &= X_{ij}\beta + Z_{ij}b_i + \sum_{l=1}^{L} \int_0^1 W_{ijl}(s)\gamma_l(s)ds \end{aligned} \tag{1}$$

where "$\text{EF}(\mu_{ij}, \eta)$" denotes an exponential family distribution with mean μ_{ij} and dispersion parameter η. Here $X_{ij}\beta$ is the standard fixed effects component, $Z_{ij}b_i$ is the standard random effects component, $b_i \sim \text{Normal}(0, G)$ are subject-specific random effects, and the $\int_0^1 W_{ijl}(s)\gamma_l(s)ds$ are the subject/visit-specific functional effects. Model (1) is novel in that it adds subject-specific random effects to the standard cross-sectional functional regression model; the random effects are used to account for the correlation between

repeated observations at the subject level. Many different structures of random effects, b_i, will be needed and used. Of course, in practice the $W_{ijl}(s)$ are not truly functional but are observed on a dense (or sparse) grid and often with error; moreover, the predictors are not necessarily observed over the same domain. In fact, while the measurement error is negligible in our application, there are some missing values in the functional predictor in Figure 1, and the mean and parallel diffusivity functional predictors have different domains. We point out that solutions to these problems are well known [11, 27, 29, 31], and do not discuss them here.

Our proposed method to fit model (1) extends the approach to single-level functional regression in [8, 13] to the current longitudinal setting. Specifically, we use functional principal components decomposition to parsimoniously express the $W_{ijl}(s)$ and a large spline basis to estimate the coefficient function $\gamma_l(s)$. Smoothness of $\gamma_l(s)$ is induced via penalization in a mixed effects model. Some strengths of this approach are that: 1) it extends functional regression to model the dynamic association between outcomes and functional predictors in longitudinal studies; 2) it casts a novel functional model in terms of well-understood mixed models; 3) it is applicable in any situation in which the $W_{ijl}(s)$ are observed or can be estimated, including when they are observed sparsely or with error; 4) it can be fit using standard, well-developed statistical software; and 5) it provides confidence or credible intervals for all the parameters of the model, including the functional ones. We emphasize this final point, as confidence intervals are rarely discussed in the functional regression literature, and in penalized approaches to functional regression they are typically bootstrap or empirical intervals. The connection to mixed models provides a simple, statistically principled approach for constructing confidence or credible intervals. Thus our software implementations provide the confidence intervals as a byproduct of the fitting procedure.

We contrast the setting of this paper with the large body of existing functional regression work. Foremost, existing functional regression work [5, 6, 7, 4, 10, 18, 17, 21, 24, 27, 28] deals only with cross sectional regression. Here, we are focused on longitudinally observed functional predictors and outcomes, which necessitates the addition of random effects to the standard functional regression model. To the best of our knowledge, this setting has not been considered previously. Moreover, the addition of random effects increases the complexity of the functional regression model, requiring new methodology implemented in efficient software.

Additionally, we point out that several alternative penalized approaches to single-level functional regression exist [6, 7, 28], but that each of these incorporates a computationally expensive cross-validation procedure. How and whether such methods would extend to fitting longitudinal models of type (1) remains to be elucidated. The additional complexity of a longitudinal model may increase the computational burden, perhaps prohibitively. Low-dimension approaches to single-level functional regression [24], in which the smoothness of $\gamma_l(s)$ depends on the dimension of its basis, are less automated than penalized approaches but are, notably, generalizable to the longitudinal setting. In fact, such models are a special case of the method we propose in Section 2.

It is also important to distinguish the proposed longitudinal functional regression model with well developed Functional ANOVA models [3, 15]. Here, one observes functions organized into groups; the goal is to express the functions as a combination of a group mean, a subject-specific deviation from the group mean, and possibly a subject-visit-specific deviation. Much of this work has focused on the estimation of and inference for the group means, and has employed penalized splines in expressing these functions. A scalar-on-function regression that incorporates random effects to account for group effects, as proposed in this paper, is not considered; we however point out that the work in FANOVA suggests that numerous datasets necessitating such a model exist and are under investigation. Others [11, 14] have extended functional principal components analysis to the multilevel and longitudinal settings, emphasizing parsimonious and computationally efficient methods for the expression of subject-visit specific curves. While these methods are the state of the art in describing the variability in observed multilevel and longitudinal functions, they do not consider accompanying longitudinal outcomes.

Finally, we highlight the distinction between the treatment of scalars observed longitudinally as sparse functional covariates [16, 23, 32] and the current setting, in which functions are observed longitudinally.

The introduction of a longitudinal functional regression model, therefore, fills a gap in the functional data analysis literature. While the proposed approach is based on the framework described in [13] and arises naturally therein, the longitudinal extension developed here defines a broadly useful class of functional regression models. Moreover, computationally feasible software for the estimation and inference related to longitudinal functional regression models is freely available online.

The manuscript is organized in the following way. Section 2 describes the proposed general method

for longitudinal functional regression. In Section 3 we pursue a simulation study to examine the viability of the proposed method and in Section 4 we apply our method to the DTI data. We end with a discussion in Section 5. Implementations of the method proposed in Section 2, provided in both likelihood-based and Bayesian frameworks and accompanied by a discussion of the the advantages and disadvantages of each, are available in an online appendix which also contains all software used in the simulation exercise.

2 LPFR

The longitudinal penalized functional regression (LPFR) method builds on an approach to single-level functional regression in which the penalization is achieved via a mixed model [13]. The addition of random effects to this model therefore arises naturally and with a minimal increase in complexity.

2.1 Single-level Penalized Functional Regression

We briefly describe the penalized functional regression (PFR) approach for single-level data with outcome model

$$\begin{aligned} Y_i &\sim \text{EF}(\mu_i, \eta) \\ g(\mu_i) &= \alpha + X_i\beta + \int_0^1 W_i(s)\gamma(s)ds\ . \end{aligned} \tag{2}$$

The first step in PFR is to express the functional covariate in terms of a truncated Karhunen-Loéve decomposition. That is, we let $\Sigma^W(s,t) = \text{Cov}[W_i(s), W_i(s)]$ be the covariance operator on the observed functions and $\sum_{k=1}^{\infty}\lambda_k\psi_k(s)\psi_k(s)$ be the spectral decomposition of $\Sigma^W(s,t)$, where $\lambda_1 \geq \lambda_2 \geq \ldots$ are the non-increasing eigenvalues and $\boldsymbol{\psi}(\cdot) = \{\psi_k(\cdot) : k \in \mathbb{Z}^+\}$ are the corresponding orthonormal eigenfunctions. Note that if $W_i(s)$ is observed with error then a good estimator of $\Sigma^W(s,t)$ is obtained by smoothing the off-diagonal elements of the observed covariance operator [29, 31]. Moreover, in the case that the $W_i(s)$ are sampled sparsely or over different grids, one can construct an estimate of the covariance operator using the following two-stage procedure [11]. First, use a fine grid to bin each subject's observations to construct a rough estimate of the covariance operator based on these under-smoothed functions; second, smooth the rough covariance operator estimated in the previous step. A truncated

Karhunen-Loéve approximation for $W_i(s)$ is given by $W_i(s) = \mu(s) + \sum_{k=1}^{K_w} c_{ik}\psi_k(s)$, where K_w is the truncation lag, the $c_{ik} = \int_0^1 \{W_i(s) - \mu(s)\}\, \psi_k(s)ds$ are uncorrelated random variables with variance λ_k, and $\mu(s)$ is the mean function over all subjects and visits. Unbiased estimators of c_{ik} can be obtained either as the Riemann sum approximation to the integral $\int_0^1 \{W_i(s) - \mu(s)\}\, \psi_j(s)ds$ or via the mixed effects model ([9, 11])

$$\begin{aligned} W_i(s) &= \mu(s) + \textstyle\sum_{k=1}^{K_w} c_{ik}\psi_k(s) + \epsilon(s) \\ \boldsymbol{c}_i &\sim \mathrm{N}(0, \Lambda),\ \epsilon(s) \sim \mathrm{N}(0, \sigma_\epsilon^2), \end{aligned} \tag{3}$$

where Λ is a $K_w \times K_w$ matrix with $(k,k)^{th}$ entry λ_k and 0 elsewhere, and the c_{ik}, ϵ_{ij} are mutually independent for every i, k.

The second step in PFR is to model the coefficient function $\gamma(s)$ using a large spline basis with smoothness induced explicitly via a mixed effects model. For example, let $\boldsymbol{\phi}(s) = \{\phi_1(s), \phi_2(s), \ldots, \phi_{K_g}(s)\}$ be a truncated power series spline basis, so that $\gamma(s) = \boldsymbol{\phi}(s)\mathbf{g} = g_0 + g_1 t + \sum_{k=3}^{K_g} g_k(t - \kappa_k)_+$ where $\mathbf{g} = \{g_1, \ldots, g_{K_g}\}^T$ and $\{\kappa_k\}_{k=3}^{K_g}$ are knots. Thus,

$$\int_0^1 W_i(s)\gamma(s)ds = a + \int_0^1 \boldsymbol{c}_i'\boldsymbol{\psi}^T(s)\boldsymbol{\phi}(s)\boldsymbol{g}\, ds = a + \boldsymbol{c}_i'\boldsymbol{M}_{\psi\phi}\boldsymbol{g},$$

where $\boldsymbol{c}_i' = (c_{i1}, \ldots, c_{iK_w})$, $\boldsymbol{M}_{\psi\phi}$ is a $K_w \times K_g$ dimensional matrix with the $(k,l)^{th}$ entry equal to $\int_0^1 \psi_k(s)\phi_l(s)ds$ and $a = \int_0^1 \mu(s)\gamma(s)ds$. Let $\boldsymbol{C}$ be the $I \times K_w$ matrix of PC loadings with i^{th} row $\boldsymbol{c}_i'$; the outcome model (2) is posed as

$$\begin{aligned} \boldsymbol{Y} \mid \boldsymbol{W}(s) &\sim EF(\boldsymbol{\mu}, \eta) \\ g(\boldsymbol{\mu}) &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} \\ \boldsymbol{u} &\sim \mathrm{N}[0, \sigma_g^2\boldsymbol{I}], \end{aligned} \tag{4}$$

where $\boldsymbol{X} = [1\ X\ (\boldsymbol{C}\boldsymbol{M}_{\psi\phi})^{[,1:2]}]$ is a matrix consisting of non-functional covariates and the first two columns of $\boldsymbol{C}\boldsymbol{M}_{\psi\phi}$, which are fixed effects used in the modeling of $\gamma(s)$, $\boldsymbol{Z} = (\boldsymbol{C}\boldsymbol{M}_{\psi\phi})^{[,3:K_g]}$ consists of the columns of $\boldsymbol{C}\boldsymbol{M}_{\psi\phi}$ which are random effects used in the modeling of $\gamma(s)$, $\boldsymbol{\beta} = [\alpha, \beta, g_1, g_2]$ is the vector of fixed effects, and $\boldsymbol{u} = \{g_k\}_{k=3}^{K_g}$ is the vector of random effects; the term $a = \int_0^1 \mu(s)\gamma(s)ds$ is incorporated into the overall model intercept α. We note that other spline bases for $\gamma(s)$ can (and, in

the Bayesian software implementation, will) be used with appropriate changes to the specification of the random effects $\boldsymbol{u}$ used to model $\gamma(s)$.

2.2 Longitudinal Penalized Functional Regression

Longitudinal penalized functional regression is natural in this framework. Given data of the form $[Y_{ij}, W_{ij1}(s), ..., W_{ij}$ for subjects $1 \leq i \leq I$ over visits $1 \leq j \leq J$, we model the functional effect in the following way. First, express the $W_{ijl}(s)$ using a functional principal components decomposition with eigenfunctions $\boldsymbol{\psi}_l(\cdot)$ and loadings $\boldsymbol{C}_l$. Then, estimate the coefficient functions $\gamma_l(s)$ using a flexible spline basis $\boldsymbol{\phi}(\cdot)$ so that $\gamma_l(s) = \boldsymbol{\phi}(s)\boldsymbol{g}_l$. Smoothness of the coefficient functions is induced via a mixed model, and subject-specific random effects are included as additional random components.

Specifically, define Z_1 to be the $(IJ) \times I$ design matrix obtained by row-stacking the design vectors Z_{ij} of subject-specific random effects in (1) and $\mathbf{M}_{\psi\phi l}$ to be the $K_w \times K_g$ matrix with $(k,j)^{th}$ entry $\int_0^1 \psi_{kl}(s)\phi_j(s)ds$. We write the longitudinal outcome model (1) as

$$
\begin{aligned}
\boldsymbol{Y} \mid \boldsymbol{W}(s) &\sim EF(\boldsymbol{\mu}, \boldsymbol{\gamma}) \\
g(\boldsymbol{\mu}) &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} \\
\boldsymbol{u} &\sim \mathrm{N}\left[\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_b^2 \boldsymbol{I}_I & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \sigma_{\boldsymbol{g}_1}^2 \boldsymbol{I}_{K_g-2} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \sigma_{\boldsymbol{g}_L}^2 \boldsymbol{I}_{K_g-2} \end{bmatrix}\right].
\end{aligned}
$$

Here, $\boldsymbol{X} = [1\ X\ (\boldsymbol{C}_1\mathbf{M}_{\psi\phi 1})^{[,1:2]} \dots (\boldsymbol{C}_L\mathbf{M}_{\psi\phi L})^{[,1:2]}]$ is the design matrix consisting of scalar covariates and fixed effects used to model the $\gamma_l(s)$; $\boldsymbol{Z} = [Z_1\ (\boldsymbol{C}_1\mathbf{M}_{\psi\phi 1})^{[,3:K_g]} \dots (\boldsymbol{C}_L\mathbf{M}_{\psi\phi L})^{[,3:K_g]}]$ is the design matrix consisting of subject-specific random effects and random effects used to model the coefficient functions; $\boldsymbol{\beta} = [\alpha, \beta, g_{11}, g_{12}, \dots, g_{L1}, g_{L2}]$ is the vector of fixed effect parameters; and $\boldsymbol{u} = [\{b_i\}_{i=1}^{I}, \{g_{1k}\}_{k=3}^{K_g}, \dots, \{g_{Lk}\}_{k=3}^{K_g}]$ is the vector of subject-specific random effects and the random effects used to model the $\gamma_l(s)$. In this way longitudinal functional regression models can be flexibly estimated using standard, yet carefully constructed, mixed effects models. As in [8], it is possible to model jointly the principal component loadings c_{ilk} and the outcome, though this is not necessary if good estimates of the c_{ilk} are available. Note that we

have assumed the same spline basis $\phi(\cdot)$ for each functional parameter $\gamma_l(s)$ and the same truncation lags K_w, K_g for each functional predictor and coefficient, but that these assumptions are easily relaxed.

A related approach, which can be advantageous in the setting in which the shape of $\gamma_l(s)$ is known, takes $\phi_l(\cdot)$ as a collection of parametric functions, i.e. $\phi_l(s) = \{\mathbf{1}, t\}$ if $\gamma_l(s)$ is a linear function, and uses random effects to model the longitudinal structure of the data but not to induce smoothness on $\gamma_l(s)$. However, absent setting-specific knowledge of the shape of $\gamma_l(s)$ we typically advocate the more flexible penalized approach described above.

3 Simulations

We pursue a simulation study to test the effectiveness of our proposed method in estimating one or more functional coefficients in longitudinal regression models. We use two implementations, a likelihood-based approach and a Bayesian approach, and briefly note a few differences between the two. First, the likelihood based approach estimates the matrix of PC loadings $\boldsymbol{C}$ using a Riemann sum approximation and then treats this matrix as fixed, while the Bayesian approach estimates $\boldsymbol{C}$ as in equation (3). Second, the Bayesian implementation uses a b-spline basis for $\gamma(s)$ to improve mixing of the MCMC chains. A more detailed discussion of the two implementations is available in an online appendix. To ensure reproducibility, full code for the simulation exercise is also available online.

3.1 Univariate Simulations

We first generate samples from the model

$$\begin{aligned} Y_{ij} &= b_i + \int_0^{10} W_{ij}(s)\gamma(s)ds + \epsilon_{ij}, \quad i = 1, \ldots, 100 \\ W'_{ij}(s) &= W_{ij}(s) + \delta_{ij}(s) \\ W_{ij}(s) &= u_{ij1} + u_{ij2}t + \sum_{k=1}^{10} \left\{ v_{ijk1} \sin\left(\frac{\pi k}{5}t\right) + v_{ijk2} \cos\left(\frac{\pi k}{5}t\right) \right\} \end{aligned}$$

where $\epsilon_{ij} \sim \text{N}[0, \sigma_Y^2]$, $\delta_{ij}(s) \sim \text{N}[0, \sigma_W^2]$, $b_i \sim \text{N}[0, \sigma_b^2]$, $u_{ij1} \sim \text{N}[0, 25]$, $u_{ij2} \sim \text{N}[0, 0.04]$, $v_{ijk1}, v_{ijk2} \sim \text{N}[0, 1/k^2]$ and $W'_{ij}(s)$ denotes the observed functional predictor for subject i at visit j. By construction the $W_{ij}(s)$ are a combination of a vertical shift, a slope, and sine and cosine terms of various periods.

The $W_{ij}(s)$ are observed on the dense grid $[t_g = \frac{g}{10} : g = 0, \ldots, 100]$. We assume $I = 100$ subjects, and select $\gamma_1(s) = 2\sin(\pi t/5)$ and $\gamma_2(s) = \sqrt{t}$. Note that the first true coefficient function is selected to be an early principal component of the observed functions, and the second is an arbitrary smooth function. We take $J \in \{3, 10\}$, $\sigma^2_\epsilon \in \{5, 10\}$, $\sigma^2_W \in \{0, .5\}$ and $\sigma^2_b \in \{5, 50\}$, which gives a total of 16 possible parameter combinations and two coefficient functions. For each of these combinations, we generate 100 datasets and fit model 5 using both likelihood-based and Bayesian approaches. For the Bayesian implementation, we used chains of length 500 with the first 100 as burn-in.

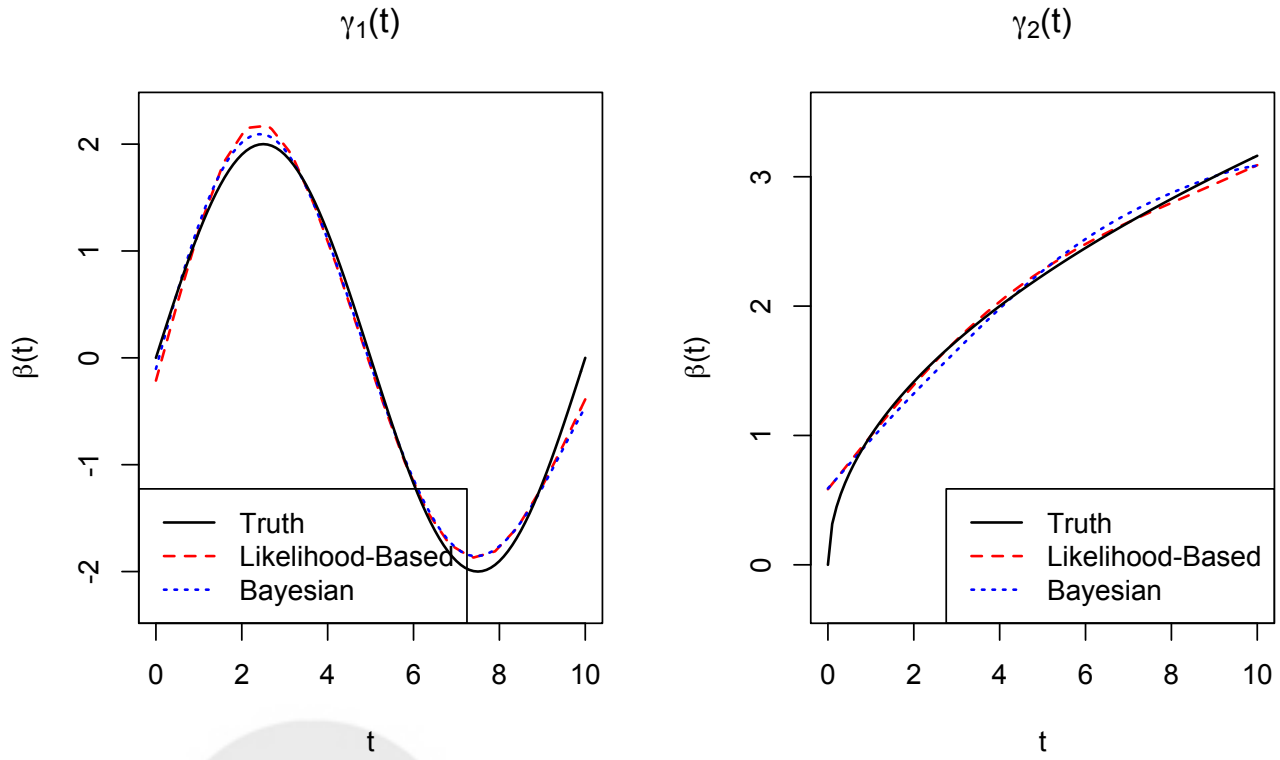

Figure 2: For both true coefficient functions and both implementations, we show the estimate $\hat{\gamma}(s)$ with median MSE. Median MSEs for likelihood-based and Bayesian approaches are, respectively, 0.016 and 0.012 for $\gamma_1(s)$ and 0.006 and 0.009 for $\gamma_2(s)$.

We calculate the mean squared error of the estimated coefficient function $\hat{\gamma}(s)$ as

$$\text{MSE} = \sum_{g=0}^{100} \{\hat{\gamma}(t_g) - \gamma(t_g)\}^2 .$$

Table 1 provides the average mean squared error (AMSE) using the both implementations taken over the

	Likelihood-based				Bayesian			
	$\gamma_1(\cdot)$		$\gamma_2(\cdot)$		$\gamma_1(\cdot)$		$\gamma_2(\cdot)$	
J=3								
σ^2_W	=0	=.5	=0	=.5	=0	=.5	=0	=.5
$\sigma^2_Y = 5$								
σ^2_b = 5	0.019	0.020	0.007	0.008	0.013	0.015	0.010	0.013
σ^2_b = 50	0.020	0.021	0.007	0.009	0.014	0.017	0.011	0.014
$\sigma^2_Y = 10$								
σ^2_b = 5	0.028	0.028	0.009	0.010	0.025	0.028	0.015	0.017
σ^2_b = 50	0.030	0.031	0.009	0.010	0.031	0.036	0.016	0.019
J=10								
$\sigma^2_Y = 5$								
σ^2_b = 5	0.008	0.009	0.005	0.006	0.006	0.009	0.004	0.006
σ^2_b = 50	0.008	0.009	0.005	0.006	0.006	0.008	0.004	0.006
$\sigma^2_Y = 10$								
σ^2_b = 5	0.012	0.013	0.006	0.007	0.008	0.009	0.006	0.008
σ^2_b = 50	0.012	0.013	0.006	0.007	0.008	0.009	0.006	0.008

Table 1: Average MSE for the likelihood-based and Bayesian implementations in the univariate setting over 100 repetitions for each possible parameter combination and true coefficient function.

100 simulated datasets for each possible combination of J, σ^2_W, σ^2_Y, σ^2_b and coefficient function. In Figure 2, we show the estimates resulting in the median MSE for $J = 3$, $\sigma^2_W = 0$, $\sigma^2_Y = 5$, $\sigma^2_b = 5$ under both coefficient functions.

From Table 1 and Figure 2, we see that the estimation of $\gamma(s)$ is very accurate regardless of the magnitude of the random effect variance or, notably, the presence or absence of measurement error. As expected, there is a substantial decrease in AMSE when one observes 10 visits per subject compared to 3 visits per subject. Doubling the error variance on the outcome has the largest impact on the AMSE, but in many situations this impact is small.

We note that several differences between the results for the likelihood-based and Bayesian implementations are apparent for $J = 3$. Particularly, notice that the likelihood-based implementation generally has better performance for $\gamma_2(s)$. A possible reason for this is that the Bayesian implementation uses a smaller basis for both the functional predictors and the coefficient function and uses a b-spline basis for $\gamma(s)$, rather than a truncated power series. With fewer observations, the smaller b-spline basis may lack the flexibility to adequately represent $\gamma_2(s)$. A second difference in the results for the different implementations is the larger impact of measurement error on AMSE for the Bayesian implementation. Recall that this approach jointly estimates the model parameters and the matrix of PC loadings $\boldsymbol{C}$. The added

variability in estimating $\boldsymbol{C}$ in the presence of measurement error may lead to more variable estimation of the functional coefficient $\gamma(s)$; however, it is possible that using longer chains would minimize the impact of observing the predictor functions with error. We also note that for $J = 10$, these differences largely disappear.

Finally, in Figure 3 we show the coverage probabilities for the 95% confidence and credible intervals produced by a subset of the simulations described above. We show the two extreme situations: first, we let $J = 3$, $\sigma_W^2 = 0$, $\sigma_Y^2 = 5$, and $\sigma_b^2 = 5$; second, we let $J = 10$, $\sigma_W^2 = .5$, $\sigma_Y^2 = 10$, and $\sigma_b^2 = 50$. For $\gamma_1(s)$, we generally see that the confidence intervals have coverage probabilities somewhat lower than the nominal level, while the credible intervals are slightly conservative with the exception of one region. A more interesting situation is apparent for $\gamma_2(s)$. We note from Figure 2 that both implementations tend to oversmooth the leftmost tail of the coefficient function; this is reflected in the very low coverage probabilities there. The Bayesian implementation recovers from this initial oversmoothing and has conservative credible intervals over the remainder of the domain, but the likelihood-based implementation has a second dip corresponding to a second region of oversmoothing before achieving coverage probabilities more similar to those for $\gamma_1(s)$.

3.2 Multivariate Simulations

Next, we generate samples from the model

$$\begin{aligned} Y_{ij} &= b_i + \int_0^{10} W_{ij1}(s)\gamma_1(s)ds + \int_0^{10} W_{ij2}(s)\gamma_2(s)ds + \epsilon_{ij}, \quad i = 1, \ldots, 100 \\ W'_{ijl}(s) &= W_{ijl}(s) + \delta_{ijl}(s) \\ W_{ijl}(s) &= u_{ijl1} + u_{ijl2}t + \sum_{k=1}^{10} \left\{ v_{ijkl1} \sin\left(\frac{\pi k}{5}t\right) + v_{ijkl2} \cos\left(\frac{\pi k}{5}t\right) \right\} \end{aligned}$$

where $\epsilon_{ijl} \sim \mathrm{N}[0, \sigma_Y^2]$, $\delta_{ijl}(s) \sim \mathrm{N}[0, \sigma_W^2]$, $b_i \sim \mathrm{N}[0, \sigma_b^2]$, $u_{ijl1} \sim \mathrm{N}[0, 25]$, $u_{ijl2} \sim \mathrm{N}[0, 0.04]$, $v_{ijkl1}, v_{ijkl2} \sim \mathrm{N}[0, 1/k^2]$ and $W'_{ijl}(s)$ denotes the l^{th}observed functional predictor for subject i at visit j. Thus the $W_{ijl}(s)$ are independent functions constructed in the same manner as in the univariate simulations. The $W_{ij}(s)$ are observed on the dense grid $[t_g = \frac{g}{10} : g = 0, \ldots, 100]$ and we set $I = 100$. Again we choose $\gamma_1(s) = 2\sin(\pi t/5)$ and $\gamma_2(s) = \sqrt{t}$ and we take $J \in \{3, 10\}$, $\sigma_\epsilon^2 \in \{5, 10\}$, $\sigma_W^2 \in \{0, .5\}$ and $\sigma_b^2 \in \{5, 50\}$. For

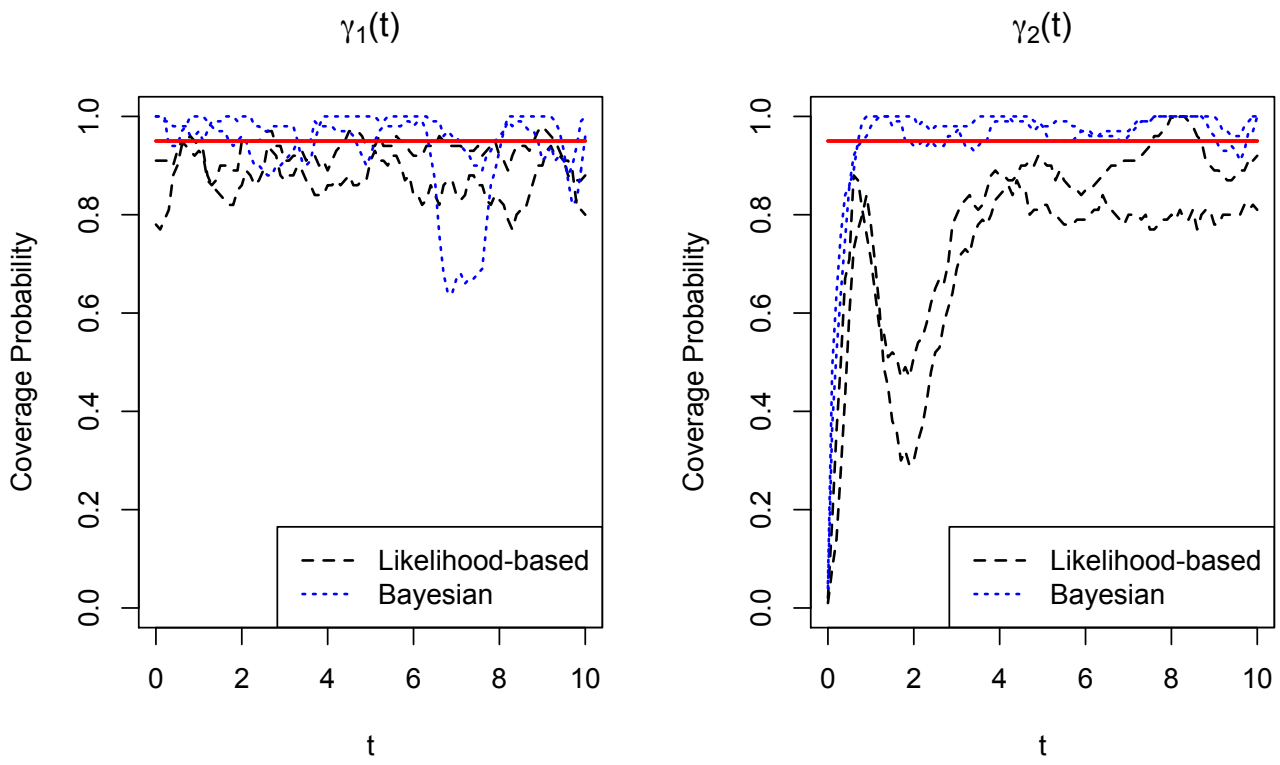

Figure 3: Coverage probabilities for 95% confidence and credible intervals for both true coefficient functions.

each of these combinations, we generate 100 datasets and fit model 5 using both of the implementations described in Section A. Due to the added complexity of the model, we used chains of length 1000 with the first 500 as burn-in.

Table 2 provides the AMSEs resulting from both likelihood-based and Bayesian implementations of the multivariate longitudinal functional regression model taken over the 100 simulated data sets. We see that the results in the multivariate setting are remarkably similar to those in the univariate case, despite the additional complexity of the model. Specifically, the AMSEs are negligibly affected and the comparisons between the likelihood-based and Bayesian implementations remain valid. Though not presented, figures examining the coverage probabilities of confidence and credible intervals are also largely unchanged.

	Likelihood-based				Bayesian			
	$\gamma_1(\cdot)$		$\gamma_2(\cdot)$		$\gamma_1(\cdot)$		$\gamma_2(\cdot)$	
J=3								
σ^2_W	=0	=.5	=0	=.5	=0	=.5	=0	=.5
$\sigma^2_Y = 5$								
σ^2_b = 5	0.018	0.019	0.007	0.008	0.015	0.012	0.010	0.010
σ^2_b = 50	0.028	0.019	0.007	0.008	0.015	0.015	0.011	0.010
$\sigma^2_Y = 10$								
σ^2_b = 5	0.025	0.028	0.009	0.010	0.026	0.025	0.016	0.015
σ^2_b = 50	0.026	0.038	0.010	0.011	0.031	0.029	0.017	0.016
J=10								
$\sigma^2_Y = 5$								
σ^2_b = 5	0.009	0.009	0.005	0.004	0.008	0.008	0.005	0.004
σ^2_b = 50	0.009	0.009	0.005	0.004	0.007	0.008	0.005	0.004
$\sigma^2_Y = 10$								
σ^2_b = 5	0.013	0.013	0.006	0.006	0.010	0.010	0.006	0.006
σ^2_b = 50	0.013	0.014	0.006	0.006	0.010	0.009	0.006	0.006

Table 2: Average MSE for the likelihood-based and Bayesian implementations in the multivariate setting over 100 repetitions for each possible parameter combination and true coefficient function.

4 Application to Longitudinal DTI Regression

Our application explores the relationship between cerebral white matter tracts in multiple sclerosis (MS) patients and cognitive impairment over time. White matter tracts are made up of myelinated axons: axons are the long projections of a neuron that transmit electrical signals and myelin is a fatty substance surround axons in white matter that enables these signals to be carried very quickly. Multiple sclerosis, an autoimmune disease which results in demyelinated lesions in white matter tracts, leads to significant disability in patients.

Diffusion tensor imaging (DTI) is a magnetic resonance imaging (MRI) based modality that traces the diffusion of water in the brain. Because water diffuses anisotropically in the white matter and isotropically elsewhere, DTI is used to generate images of the white matter specifically [1, 2, 19, 22]. Several measurements of water diffusion are provided by DTI, including fractional anisotropy and mean diffusivity, and continuous summaries of white matter tracts, parameterized by distance along the tract and called tract profiles, can be derived from diffusion tensor images.

In this study, 100 patients are scanned approximately once per year and undergo a collection of tests to assess cognitive and motor function; patients are seen between 2 and 8 times, with a median of 3 visits

per subject. Here we focus on the mean diffusivity profile of the corpus callosum tract and the parallel diffusivity profile of the right corticospinal tract as our functional predictors and the Paced Auditory Serial Addition Test (PASAT), a commonly used examination of cognitive function affected by MS with scores ranging between 0 and 60, as our scalar outcome.

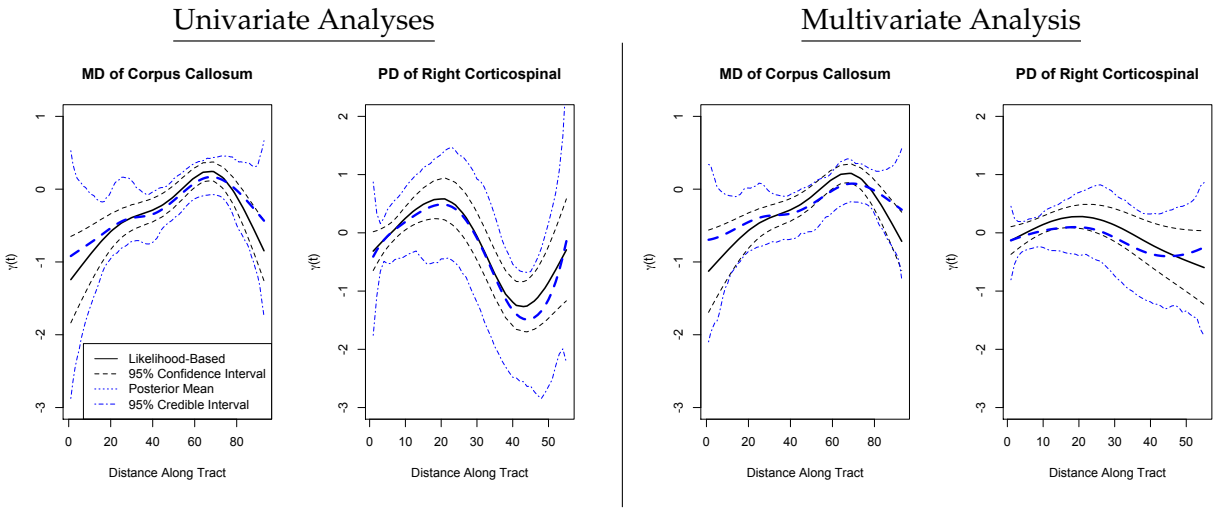

Figure 4: Results of univariate and multivariate analyses. On the left, we show the estimated coefficient function the mean diffusivity (MD) profile of the corpus callosum and the parallel diffusivity (PD) profile of the right corticospinal tract treated separately in univariate analyses. On the right we show the results of a multivariate analysis including both profiles.

We begin by fitting univariate models of the form

$$\begin{aligned} Y_{ij} &= \alpha + X_{ij}\beta + u_i + \int_0^1 W_{ij}(s)\gamma(s)ds + \epsilon_{ij} \\ u_i &\sim \mathrm{N}[0, \sigma_u^2],\ \epsilon_{ij} \sim \mathrm{N}[0, \sigma_\epsilon^2] \end{aligned} \tag{5}$$

where Y_{ij} is the PASAT score for subject i at visit l, $W_{ij}(s)$ is functional predictor for subject i at visit l, and the variable $X_{ij} = I(l > 1)$ is used to account for a learning effect that causes PASAT scores to generally rise between the first and second visit. Two such models are fit, one using the mean diffusivity profile of the corpus callosum and another using the parallel diffusivity profile of the right corticospinal tract. These models are fit using both likelihood-based and Bayesian implementations of the method describe in Section 2. Estimates of the functional coefficient $\gamma(s)$, along with credible and confidence intervals, are presented in the left panels of Figure 4. Note that in both cases the credible and confidence intervals are quite different; we recall from our simulations that the credible intervals were conservative, while the

confidence intervals generally did not achieve nominal coverage. However, for the corpus callosum both intervals indicate that the first half of the tract and region 60-80 have a significant impact on the PASAT outcome. For the corticospinal tract, the region from 40-50 is indicated as having a significant impact on the outcome.

In Table 3 we show the percent of the outcome variance explained in each of several models. We note that the largest source of variation in the outcome is the subject-specific random effect, but that the univariate regression models improve on this. In fact, it appears the two functional predictors explain a similar amount of variability beyond the random intercept only model. However, the right corticospinal tract mediates movement and strength, not cognition, and therefore should not have a direct impact on the PASAT score. We hypothesize that this tract profile is correlated with overall disease burden and, therefore, decreased cognitive ability but does not itself effect the PASAT outcome. On the other hand, the corpus callosum is plausibly linked to cognitive performance; degradation of this tract should affect the PASAT score [25].

	Random Int. Only	Univariate: MD	Univariate: PD	Multivariate
PVE	81.2 %	88.6%	88.7	88.8%

Table 3: Percent of the variance in the PASAT outcome explained by a random intercept only model, two univariate longitudinal functional regression models, and a multivariate longitudinal functional regression model.

Finally, we carry out a multivariate analysis that includes both the corpus callosum mean diffusivity profile and the corticospinal tract parallel diffusivity profile. The estimated functional coefficients in this model are given in the right panels of Figure 4. We see that the estimate for the corpus callosum and accompanying confidence intervals are largely unchanged from the univariate regression, but that the estimate for the corticospinal tract is quite different. Rather than a peak from 10-20 and a dip from 40-50, the coefficient for the corticospinal tract is near zero over its entire range, and no regions appear particularly important in predicting the outcome. Based on the univariate and multivariate analyses, it seems that the right corticospinal tract is a functional confounder in that it is correlated with disease burden and cognitive ability but not biologically linked to the outcome, while the direct relationship between the corpus callosum and the PASAT score appears statistically plausible.

5 Discussion

In this paper we have proposed a novel longitudinal functional regression model and explored a general approach to estimate the parameters of this model, including scalar and functional coefficients and subject-specific random effects. This approach is appealing in that in casts difficult longitudinal functional regression problems in terms of the popular and well-known mixed model framework. The method is: 1) very general, allowing for the functional coefficient to be fit using a flexible penalized approach or modeled parametrically; 2) applicable in whether the functional covariates are sparsely or densely sampled, or measured with error; and 3) computationally efficient and tractable. To the last point, we developed two implementations for this approach using common statistical software and discussed their comparative advantages and disadvantages in an online appendix.

Additionally, we have carefully tested our proposed method under both likelihood-based and Bayesian approaches in a detailed simulation study considering both univariate and multivariate settings. The results of this study demonstrate the effectiveness of the method in estimating functional coefficients under a variety of conditions, including few or many visits per subject, different levels of outcome and random effect variance, and in the presence or absence of measurement error on the functional predictor. We have also applied our method to a scientifically interesting and clinically important data set exploring the relationship between intracranial white matter tracts and cognitive disability in multiple sclerosis patients.

Future work may take several directions. Basing longitudinal functional regression on functional decomposition techniques tailored to the structure of the data may give results that are simpler to implement or interpret. The treatment of longitudinal functional observations as outcomes will continue to be scientifically interesting and statistically challenging. Penalized approaches that use cross validation to impose smoothness, though computationally expensive, could provide a useful alternative to the mixed model approach presented here. Finally, functional regression techniques for data sets in which the predictor is very large, either from extremely dense observation or because the predictor is a 2- or 3-dimensional image, will be needed as such data sets are collected and analyzed.

6 Acknowledgments

The research of Goldsmith, Crainiceanu and Caffo was supported by Award Number R01NS060910 from the National Institute Of Neurological Disorders And Stroke. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of Neurological Disorders And Stroke or the National Institutes of Health. This research was partially supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke. Scans were funded by grants from the National Multiple Sclerosis Society and EMD Serono to Peter Calabresi, whose support we gratefully acknowledge.

References

[1] P. Basser, J. Mattiello, and D. LeBihan. MR Diffusion Tensor Spectroscopy and Imaging. *Biophysical Journal*, 66:259–267, 1994.

[2] P. Basser, S. Pajevic, C. Pierpaoli, and J. Duda. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44:625–632, 2000.

[3] B. Brumback and J. Rice. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93:961–976, 1998.

[4] H Cardot, C Crambes, A Kneip, and P Sarda. Smoothing splines estimators in functional linear regression with errors-in-variables. *Computational Statistics and Data Analysis*, 2007.

[5] H. Cardot, F. Ferraty, and P. Sarda. Functional Linear Model. *Statistics and Probability Letters*, 45:11–22, 1999.

[6] H. Cardot, F. Ferraty, and P. Sarda. Spline Estimators for the Functional Linear Model. *Statistica Sinica*, 13:571–591, 2003.

[7] H. Cardot and P. Sarda. Estimation in Generalized Linear Model for Functional Data via Penalized Likelihood. *Journal of Multivariate Analysis*, 92:24–41, 2005.

[8] C.M. Crainiceanu and J. Goldsmith. Bayesian Functional Data Analysis using WinBUGS. *Journal of Statistical Software*, 32:1–33, 2010.

[9] C.M. Crainiceanu, A.M. Staicu, and C. Di. Generalized Multilevel Functional Regression. *Journal of the American Statistical Association*, 104:1550–1561, 2009.

[10] C Crambes, A Kneip, and P Sarda. Smoothing splines estimators for functional linear regression. *Annals of Statistics*, 2009.

[11] C.Z. Di, C.M. Crainiceanu, B.S. Caffo, and N.M. Punjabi. Multilevel Functional Principal Component Analysis. *Annals of Applied Statistics*, 4:458–288, 2009.

[12] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, 2006.

[13] J. Goldsmith, J. Bobb, C. Crainiceanu, B. Caffo, and D. Reich. Penalized Functional Regression. *Journal of Computational and Graphical Statistics*, 2010.

[14] S. Greven, C. Crainiceanu, B. Caffo, and D. Reich. Longitudinal Functional Prin- cipal Component Analysis. *Electronic Journal of Statistics*, 2010.

[15] W. Guo. Functional mixed effects models. *Biometrics*, 58:121–128, 2002.

[16] P. Hall, H.G. Müller, and J. L. Wang. Properties of Principal Component Methods for Functional and Longitudinal Data Analysis. *Annals of Statistics*, 34:14931517, 2006.

[17] G James, J Wang, and J Zhu. Functional Linear Regression That's Interpretable. *Annals of Statistics*, 37:2083–2108, 2009.

[18] G.M. James. Generalized linear models with functional predictors. *Journal Of The Royal Statistical Society Series B*, 64:411–432, 2002.

[19] D. LeBihan, J. Mangin, C. Poupon, and C. Clark. Diffusion Tensor Imaging: Concepts and Applications. *Journal of Magnetic Resonance Imaging*, 13:534–546, 2001.

[20] M. Lindquist and I. McKeague. Logistic Regression with Brownian-like Predictors. *Journal of the American Statistical Association*, 104:1575–1585, 2009.

[21] B. D. Marx and P.H.C. Eilers. Generalized Linear Regression on Sampled Signals and Curves: a P-spline Approach. *Technometrics*, 41:1–13, 1999.

[22] S. Mori and P. Barker. Diffusion magnetic resonance imaging: its principle and applications. *The Anatomical Record*, 257:102–109, 1999.

[23] H.G. Müller. Functional Modelling and Classification of Longitudinal Data. *Scandinavian Journal of Statistics*, 32:223–240, 2005.

[24] H.G. Müller and U. Stadtmüller. Generalized functional linear models. *Annals of Statistics*, 33:774–805, 2005.

[25] A. Ozturk, S. Smith, E. Gordon-Lipkin, D. Harrison, N. Shiee, D. Pham, B. Caffo, P. Calabresi, and D. Reich. MRI of the corpus callosum in multiple sclerosis: association with disability. *Multiple Sclerosis*, 16:166–177, 2010.

[26] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and the R Core team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2009. R package version 3.1-96.

[27] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, 2005.

[28] P. Reiss and R. Ogden. Functional Principal Component Regression and Functional Partial Least Squares. *Journal of the American Statistical Association*, 102:984–996, 2007.

[29] J. Staniswalis and J. Lee. Nonparametric Regression Analysis of Longitudinal Data. *Journal of the American Statistical Association*, 444:1403–1418, 1998.

[30] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, Fourth Edition, 2002.

[31] F. Yao, H. Müller, A. Clifford, S. Dueker, J. Follett, Y. Lin, B. Buchholz, and J. Vogel. Shrinkage Estimation for Functional Principal Component Scores with Application to the Population. *Biometrics*, 59:676–685, 2003.

[32] F. Yao, H. Muller, and J. Wang. Functional Linear Regression Analysis for Longitudinal Data. *Annals of Statistics*, 100:577–590, 2005.

A Computational Issues

The models described in the Section 2 are necessarily complex to account for the size and complexity of the data set. This raises reasonable concerns about computational feasibility, especially because many competing methods are slow even in the single-level case. To dispel these concerns we present key segments of code from two implementations, a likelihood-based approach in `R` and a Bayesian approach in `WinBUGS`, of our method for longitudinal functional regression. We also provide overview of the benefits and drawbacks of each. Complete code for each is available online.

For simplicity, the code given below assumes Gaussian outcomes and subject-specific random intercepts, with the I subjects observed J times each. The functional coefficient is modeled using a large spline basis with explicit penalty. Straightforward alterations to the basic code presented here allow for more complex models.

A.1 Likelihood-based R Implementation

Because LPFR takes advantage of standard mixed effects models to express longitudinal functional models, we can take advantage of standard mixed effects software to fit such models. Here, we use the `lme` function in the `nlme` package [26] to fit Gaussian functional models using the following code:

```
# evaluate the J matrix using numeric integration; assumes
# t is observed over an even grid
M.mat <- t(psi) %*% phi *(max(s) - min(s))/(length(s)-1)

# compute the CJ matrix
CJ <- C[,1:Kw] %*% M.mat

# the following code creates the design matrices so that
# the lme function performs the correct estimation
X=cbind(1, CJ[,1:2])

Z1=matrix(rep(c(rep(1, J), rep(0, I*J)), I), nrow=length(Y), ncol=I)
Z2=as.matrix(CJ[,3:Kw])
```

```
Z=cbind(Z1, Z2)

# give the columns of the design matrices meaningful names
colnames(X)=c("intercept", "1", "t")
colnames(Z)=c(paste("int.",1:dim(Z1)[2], sep=""), paste("spline.",3:Kg,sep="" ))

# these commands organize the inputs for the lme() function
re.block.inds=list(1:dim(Z1)[2], (dim(Z1)[2]+1):(dim(Z)[2]))
Z.block=list(length=2)
for (i in 1:length(re.block.inds))
  Z.block[[i]] <- as.formula(paste("~Z[,c(",paste( re.block.inds[[i]],
    collapse=","),")]-1")) group=rep(1, I*J)
grouped=data.frame(X,Y,Z)
model.data=groupedData(Y~X|group, data=grouped)

# fit the longitudinal functional regression model
fit.smooth=lme(Y~-1+X, random=list(group=pdBlocked(Z.block,
  pdClass=rep("pdIdent",length(Z.block)))))
```

The quantities `I`$=I$, `J`$=J$, `M.mat`$=\boldsymbol{M}_{\psi\phi}$, `C`$=\boldsymbol{C}$, `Y`$=Y$, `X`$=\boldsymbol{Z}$, `Z1`$=Z_1$, `Z2`$=Z_2$, and `Z`$=\boldsymbol{Z}$, are as defined in Section 2. We note that general scalar outcomes can be fit by using the `glmmPQL` function in the `MASS` package [30] in place of the `lme` function here.

This approach has several advantages, especially as compared to the `WinBUGS` implementation. The first, and perhaps most important, is familiarity: `R` is commonly used among statisticians, and this implementation will be readily understood. Another advantage is that this code executes very quickly, so that computation time is significantly shorter. However, there is an important disadvantage of this implementation. After our best efforts, we have been unable to use standard mixed effects software to fit models with $n > 120$, though we have not found a limit on the number of visits per subject. For larger studies, this constraint is very important. This is a problem related to the `R` implementation of mixed effects models and not a problem with our methodology. The problem is that the matrix of random effects explodes in size, which is not well handled by `R`.

A.2 Bayesian WinBUGS Implementation

WinBUGS is a powerful alternative when fitting mixed effects models, and as such can be used to implement the LPFR method. We note two minor changes that arise in this implementation: 1) we use a cubic b-spline basis for $\gamma(s)$ with a first order random walk prior to give better mixing properties; and 2) we take advantage of the ability to jointly model the exposure and outcome models (3 and 5). Further, to increase the computation efficiency of this implementation, we use a variation on the projection method described in [11]. In particular, let

$$A_{ijk} = \int_0^1 \{W_{ij}(s) - \mu(s)\}\, \psi_k(s) ds = c_{ijk} + \epsilon'_{ijk}$$

where ϵ'_{ijk} includes the variance from the excluded dimension in the truncated Karhunen-Loéve decomposition and the integrated measurement error. Note that A_{ijk} can be estimated using numerical integration, and that by letting $\boldsymbol{A}_{ij} = (A_{ij1}, \ldots, A_{ijK_w})^T$ we can replace the exposure model 3 with

$$\begin{aligned} \boldsymbol{A}_{ij} &= \boldsymbol{c}_{ij} + \epsilon'_{ij} \\ \boldsymbol{c}_{ij} &\sim \mathrm{N}(0, \Lambda),\ \epsilon'_{ij} \sim \mathrm{N}(0, \sigma^2_{\epsilon'} \boldsymbol{I}), \end{aligned} \tag{6}$$

These changes are implemented in the following code segment; note that we do not use the function inprod to substantially reduce computation time, and that the quantity JxGam $= \boldsymbol{M}_{\psi\phi}\boldsymbol{g}$ is updated once per model iteration, outside the main loop over subjects.

```
model
{ # Start model
for (i in 1:I) { # Begin loop over subjects
  for (j in 1:J) { # Begin loop over visits
    Y[((i-1)*J+j)]~dnorm(fitted.val[((i-1)*J+j)], tau_Y)
    fitted.val[((i-1)*J+j)]<-alpha + u[i] + eta[((i-1)*J+j)]
    eta[((i-1)*J+j)]<-C[((i-1)*J+j),1]*JxGam[1]+C[((i-1)*J+j),2]*JxGam[2]+
      C[((i-1)*J+j),3]*JxGam[3]+C[((i-1)*J+j),4]*JxGam[4]+
      C[((i-1)*J+j),5]*JxGam[5]+C[((i-1)*J+j),6]*JxGam[6]+
      C[((i-1)*J+j),7]*JxGam[7]+C[((i-1)*J+j),8]*JxGam[8]+
```

```
      C[((i-1)*J+j),9]*JxGam[9]+C[((i-1)*J+j),10]*JxGam[10]

      for (k in 1:Kw) {# Begin loop over observations within subject and visit
        A[((i-1)*J+j), k]~dnorm(C[((i-1)*J+j),k], tau)
        C[((i-1)*J+j),k]~dnorm(0,ll[k])
      } # End loop over observations within subject and visit

    } # End loop over visits

    u[i]~dnorm(0, tau_u)

  } # End loop over subjects

  for(l in 1:Kw){ # Begin loop over rows of M.mat matrix
    JxGam[l]<-M.mat[l,1]*g[1]+M.mat[l,2]*g[2]+M.mat[l,3]*g[3]+
      M.mat[l,4]*g[4]+M.mat[l,5]*g[5]+M.mat[l,6]*g[6]+
      M.mat[l,7]*g[7]+M.mat[l,8]*g[8]+M.mat[l,9]*g[9]+
      M.mat[l,10]*g[10]
  } # End loop over rows of M.mat

  for (i in 1:Kw){ # Begin loop over precisions of PC loadings
    ll[i]~dgamma(0.1,0.1)
    lambda[i]<-1/ll[i]
  } # End loop over precisions of PC loadings

  for (l in 2:Kg){ # Begin loop over g coefficients
    g[l]~dnorm(g[l-1],tau_g)
  }
  ... # Here are the prior distributions for non-functional covariates
}#End model
```

There are several important advantages that arise in this implementation. Most importantly, this im-

plementation allows for the easy construction of posterior credible intervals. As noted above, the functional regression literature has focused on estimation of functional coefficients, while very little attention has been paid to confidence intervals. Even then, among papers using a penalized approach to estimation, confidence intervals are generally empirical or bootstrap intervals. The direct computation of posterior credible intervals therefore addresses a major need in the functional regression literature. Another advantage is joint modeling, which accounts for the uncertainty in estimating the PC loadings; while in many cases this variability in negligible, it can be very important in the presence of large measurement error or when the functional predictors are sparsely observed. That said, the WinBUGS implementation also has the disadvantage that hyperparamters sometimes must be chosen with care for the program to successfully execute.