LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Jan Gertheiss, Jeff Goldsmith,
Ciprian Crainiceanu & Sonja Greven

Longitudinal Scalar-on-Function Regression with Application to Tractography Data

Technical Report Number 118, 2012
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Longitudinal Scalar-on-Function Regression with Application to Tractography Data

Jan Gertheiss[*†] Jeff Goldsmith[‡]
Ciprian Crainiceanu[‡] & Sonja Greven[†]

February 2, 2012

Abstract

We propose a class of estimation techniques for scalar-on-function regression in longitudinal studies where both outcomes, such as test results on motor functions, and functional predictors, such as brain images, may be observed at multiple visits. Our methods are motivated by a longitudinal brain diffusion tensor imaging (DTI) tractography study. One of the primary goals of the study is to evaluate the contemporaneous association between human function and brain imaging over time. The complexity of the study requires development of methods that can simultaneously incorporate: (1) multiple functional (and scalar) regressors; (2) longitudinal outcome and functional predictors measurements per patient; (3) Gaussian or non-Gaussian outcomes; and, (4) missing values within functional predictors. We review existing approaches designed to handle such types of data and discuss their limitations. We propose two versions of a new method, longitudinal functional principal components regression. These methods extend the well-known functional principal component regression and allow for different effects of subject-specific trends in curves and of visit-specific deviations from that trend. The different methods are compared in simulation studies, and the most promising approaches are used for analyzing the tractography data.

Keywords: Functional Principal Components, Functional Regression, Longitudinal Functional Principal Components Regression, Multiple Sclerosis, Repeated Measurements, Diffusion Tensor Imaging

[*]To whom correspondence should be addressed: `jan.gertheiss@stat.uni-muenchen.de`
[†]Department of Statistics, Ludwig-Maximilians-Universität Munich, Germany.
[‡]Department of Biostatistics, Johns Hopkins University, Baltimore, USA.

1 Introduction

Increasingly, longitudinal studies collect data, such as curves or images, that is functional in nature. Interest often centers on using these functional observations to predict longitudinal or time-invariant scalar outcomes. To be specific, we are motivated by a neurological study on disease progression and corresponding changes in diffusion tensor images of the brain in multiple sclerosis (MS) patients. Interest lies in relating changes in neuronal tract properties extracted from the diffusion tensor images to disability scores measured at each visit, as well as in discriminating between MS patients and controls. Figure 1 displays the fractional anisotropy along the corpus callosum and the left corticospinal tracts for one of the MS patients observed at 6 different visits over a period of 4 years.

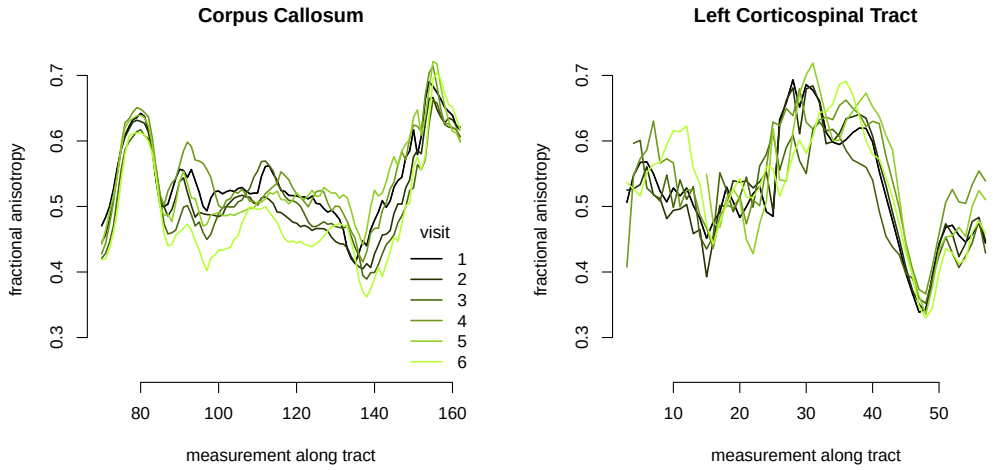

Figure 1: Fractional anisotropy along the corpus callosum and the left corticospinal tract for a multiple sclerosis patient, observed at 6 visits and measured at 93 and 55 sample points, respectively.

In addition to fractional anisotropy, several other measurements of water diffusivity including mean diffusivity, parallel diffusivity and magnetization transfer ratio are available; measurements for other white matter tracts in the brain are also given. The study is one example of a rapidly increasing number of biomedical studies where, in contrast to simpler scalar-on-function regression, both outcomes and functional predictors are observed repeatedly over time, outcomes may be non-gaussian, and there are multiple functional predictors. Any realistic method for such a problem will have to deal with additional non-functional covariates, such as age and sex, as well as partly missing or noisy functional predictors.

There exists a rich literature dedicated to scalar-on-function regression. For normally distributed outcomes the *functional* linear model (Ramsay and Silverman, 2005) is implemented in the R package `fda` (Ramsay et al., 2011). The

P-spline approach by Marx and Eilers (1999) can also be applied for cases when functional data are measured at the same equidistant points. In contrast to the `fda` package, P-splines can be used with outcomes that are not normally distributed. The functional linear model has been extended to non-gaussian data by James (2002), Müller and Stadtmüller (2005) and James and Silverman (2005). Ferraty and Vieu (2006) proposed methods for nonparametric functional regression and classification, and Reiss and Ogden (2007, 2010) developed generalized functional principal components regression and partial least squares. Goldsmith et al. (2011) used mixed models methodology for fitting generalized functional linear models. In Goldsmith et al. (2012) these penalized functional regression methods are extended to longitudinal data, that is, to the cases when both the outcome and the functional data are measured at multiple visits. R implementations of functional principal component regression (FPCR) and longitudinal penalized functional regression (LPFR) are found in the add-on package `refund` (Crainiceanu and Reiss, 2011), which is using `fda` (Ramsay et al., 2011) and `mgcv` (Wood, 2006, 2011) with all tuning parameters being estimated by REML.

Despite these important advances, scalar-on-function regression remains an active area of research. Indeed, in practice one is interested in methods that can work for: (1) outcomes that are normally or non-normally distributed; (2) multiple functional predictors that are measured with sizeable noise and with different patterns of missing data; and (3) sampling designs that are complex. The brain tractography application has all these features, as outcomes can be cognitive/motor outcomes (continuous, but slightly skewed) or multiple sclerosis status (0-1), there are multiple predictors due to the various types of water diffusion measurements along multiple tracts, measurements along tracts exhibit between 20% and 30% missing data, and brain images and outcomes are measured at multiple visits.

Before proceeding with the description of our approach we provide a one-stop description of the state-of-the-art of scalar-on-function regression; see Table 1. In what follows, we will focus only on methods that have associated published software or code and compare them with respect to their ability to handle: multiple functional and scalar predictors, non-gaussian outcomes, longitudinal data, and missing values. A check mark in parentheses in Table 1 indicates that such data could, in principle, be handled, but no software is currently available. Currently, only longitudinal penalized functional regression (LPFR) can handle regression where functional data are observed at multiple visits. We conclude that, in spite of the extensive research described here, the scalar on function regression methodology and software contains many gaps. These problems are especially serious with respect to availability of software, which is limited at best. For example, for FLiRTI (James et al., 2009) software is available (`http://www-bcf.usc.edu/~gareth/`), but methods are implemented only for the functional linear model. The blockwise boosting approach (Gertheiss and Tutz, 2009; Tutz and Gertheiss, 2010) is currently implemented for gaussian and

method	multiple functional/ scalar predictors	non-gaussian outcomes	longitudinal data	missing values in the curves
P-splines	✓	✓		
FLM	✓			✓
GFLM	(✓)	(✓)		(✓)
FAME	(✓)	(✓)		(✓)
FLiRTI		(✓)		
BB		✓		
NPFDR		✓		
FPCR	(✓)	✓		✓
LPFR	✓	✓	✓	✓

Table 1: Some functional regression methods and desired properties for typical biomedical data, including the tractography data. Considered are P-splines (Marx and Eilers, 1999), the functional linear model (FLM; Ramsay and Silverman, 2005), its generalization (GFLM; James, 2002; Müller and Stadtmüller, 2005), functional adaptive model estimation (FAME; James and Silverman, 2005), functional linear regression that's interpretable (FLiRTI; James et al., 2009), blockwise boosting (BB; Gertheiss and Tutz, 2009; Tutz and Gertheiss, 2010), nonparametric functional data regression (NPFDR; Ferraty and Vieu, 2006), functional principal component regression (FPCR; Reiss and Ogden, 2007, 2010), longitudinal penalized functional regression (LPFR; Goldsmith et al., 2012). A check mark indicates if the respective method can be applied to such data, parentheses indicate that such data could in principle be handled, but no software for doing so is currently available.

binary outcomes only (`http://www.statistik.lmu.de/~gertheiss/`). Extensions of functional principal components regression (FPCR) to multiple functional (and scalar) predictors, based on multivariate functional PCA (see Ramsay and Silverman, 2005), are currently not available in the `refund` package.

Here we propose to use the longitudinal functional principal component analysis (LFPCA) proposed by Greven et al. (2010) to extend the FPCR framework to the case when functional data are observed at multiple visits. Two LFPCA-based regression methods will be developed. These approaches: (1) allow different subject- and visit-level effects on the outcome; (2) can be used when curves are observed with missings, or measured with error; (3) are applicable for both Gaussian and non-Gaussian outcomes; and (4) have freely available software implementations. The first point is highly relevant in applications where the interest will center on identifying a specific component of variability that is associated with the outcome. The second point will allow the efficient use of information and will avoid discarding predictors that exhibit missing data.

The paper is organized as follows. In Section 2 we review and extend existing techniques for longitudinal functional regression, and in Section 3 we propose two possible versions of longitudinal functional principal components regression. In Sections 4 and 5, we compare the different approaches via simulation studies and

analytically. The methods that performed well in the simulation studies are then used to analyze the tractography data in Section 6. We close with a discussion in Section 7. For all computations we used R (R Development Core Team, 2011), with code being provided in a web appendix.

2 Generalized Linear Mixed Models with Functional Predictors

We start with reviewing the generalized linear mixed effects model with functional predictors and the longitudinal penalized functional regression (LPFR) method for estimating the parameters in this model (Goldsmith et al., 2012).

We observe data of the form $(Y_{ij},\ X_{ij1},\ldots,X_{ijp},\ Z_{ij0},\ldots,Z_{ijg},\ W_{ij1}(s),\ldots$ $\ldots,W_{ijq}(s))$, where Y_{ij} denotes the response value for individual i at visit j $(i=1,\ldots,n;\ j=1,\ldots,n_i)$; $W_{ijm}(s)$ are functional predictor curves over domain $\mathcal{D}_m,\ m=1,\ldots,q;\ X_{ij}=(X_{ij1},\ldots,X_{ijp})^\top$ and $Z_{ij}=(Z_{ij0},\ldots,Z_{ijg})^\top$ denote (vectors of) additional scalar explanatory variables. For such data, Goldsmith et al. (2012) proposed the model

$$\mu_{ij}=h(\eta_{ij})\ \text{ and }\ \eta_{ij}=\alpha+\sum_{l=1}^{p}X_{ijl}\beta_l+\sum_{v=0}^{g}Z_{ijv}b_{vi}+\sum_{m=1}^{q}\int_{\mathcal{D}_m}W_{ijm}(s)\gamma_m(s)\,ds, \tag{1}$$

with fixed effects $\beta_1,\ldots,\beta_p$, and independent and identically distributed (iid) vectors of random effects $(b_{0i},\ldots,b_{gi})^\top=b_i\sim N(0,\Gamma)$. Function h is assumed to be a known link function. For given (nonfunctional and functional) covariates $X_{ij},Z_{ij},W_{ij1},\ldots\ \ldots,W_{ijq}$ and random effects b_i, the distribution of Y_{ij} is assumed to be from an exponential family with conditional mean $\mu_{ij}=E(Y_{ij}|b_i,X_{ij},Z_{ij},W_{ij1},\ldots\ \ldots,W_{ijq})$. For details about the generalized linear mixed model, see e.g. McCulloch et al. (2008).

Model (1) can be estimated using the LPFR approach developed in Goldsmith et al. (2012). Briefly, this method decomposes the functional predictors $W_{ijm}(s),\ m=1,\ldots,q$, using functional principal components analysis (FPCA) ignoring the repeated and non-independent observation of curves within subjects i across visits j. Next, coefficient functions $\gamma_m(s)$ are expressed using a flexible spline basis. With these expansions, model (1) can be expressed in a mixed model framework that induces smoothness in the coefficient functions and incorporates random effects b_i that account for correlation in the outcomes Y_{ij}. Easy-to-use functions for fitting longitudinal functional regression models with random intercepts and penalizing non-linear $\gamma_m(s)$ using the LPFR approach are implemented in the R package `refund` (Crainiceanu and Reiss, 2011).

Several advantages of the mixed model approach to longitudinal functional regression are apparent. First, well-developed software can be used to fit such

models either through the `refund` R package designed with functional data analysis in mind or through the general mixed model software in `mgcv` (Wood, 2006, 2011). The mixed model framework also allows the construction of confidence intervals for estimated coefficient functions; see, for example, Goldsmith et al. (2011), Ruppert et al. (2003) or Wood (2006) for details. Smoothing parameters that control the shape of the coefficient functions can be automatically estimated by ML or REML, and testing for constancy or linearity of the coefficient function is possible through testing whether the smoothing parameters are nonzero (Crainiceanu et al., 2005; Greven et al., 2008). For the LPFR approach in particular, the FPCA expansion of functional predictors allows one to borrow strength across subjects in estimating basis functions; this is particularly useful when curves are partially unobserved.

However, the LPFR approach to longitudinal functional regression has major drawbacks. Foremost, it does not account for the longitudinal structure of the data in modeling the effect of the functional predictor on the outcome, but only in the subject-specific random effects. In particular, the term $\int_{\mathcal{D}_m} W_{ijm}(s)\gamma_m(s)\,ds$ appearing in (1) does not separate the subject- and visit-level effects of the curve W_{ijm} which is of interest in many scientific settings. On the one hand, any impact of W_{ijm} on Y_{ij} may be driven by visit-specific deviations from a subject-specific mean; on the other hand, the visit-specific deviations may also be essentially noise masking the effect of the subject-specific mean. Similarly, the FPCA decomposition used to expand functional predictors ignores the longitudinal structure of the observations and may miss important sources of variability.

3 Longitudinal Functional Principal Components Regression

As an extension of principal components regression (PCR; Massy, 1965; see also Frank and Friedman, 1993) to functional data, functional principal components regression (FPCR) has been proposed; see Reiss and Ogden (2007) and references therein. In our case, however, functional predictors are not independent but are instead repeated observations on the same individuals. Therefore we will extend FPCR to longitudinal functional data. We will present two possible ways of conducting longitudinal functional principal components regression (LFPCR) based on longitudinal functional principal components analysis (LFPCA). The methods allow for different effects of subject-specific trends in curves and of visit-specific deviations from that trend. So it may be checked, for example, whether such visit-specific deviations are informative, or mostly constitute functional measurement error. Moreover, LFPCR is easy to implement (see web appendix).

3.1 Longitudinal Functional Principal Components Analysis

Functional principal components analysis can be used for decomposing the variability in functional data. Since, however, some curves are obtained from the same individual, measurements are dependent. Hence, we use the functional random intercept and random slope model (Greven et al., 2010) where for subject i at visit j measurement $W_{ij}(s)$ at location $s \in \mathcal{D}$ is modeled as

$$W_{ij}(s) = \eta(s, T_{ij}) + B_{i,0}(s) + T_{ij}B_{i,1}(s) + U_{ij}(s) + \varepsilon_{ij}(s). \tag{2}$$

Time point T_{ij} indicates the time of visit j for subject i, and $\eta(s,T)$ is the overall smooth mean surface. The random processes $B_i(s) = \{B_{i,0}(s), B_{i,1}(s)\}$, $U_{ij}(s)$ and $\varepsilon_{ij}(s)$ are assumed to be mean zero, squared-integrable and mutually uncorrelated. The components $B_{i,0}(s)$ and $B_{i,1}(s)$ of $B_i(s)$ denote a functional random intercept and a random slope, respectively, capturing between-subject variation. $U_{ij}(s)$ is a visit-specific functional deviation from the subject-specific functional trend, capturing visit-to-visit functional variation on the same subject ('within-subject variation'). $\varepsilon_{ij}(s)$ is white noise measurement error with variance ς^2, capturing random uncorrelated variation within each curve, cf. Greven et al. (2010). Thus, model (2) allows to decompose functional variation into three parts: subject-specific variation $B_i(s)$, visit-specific variation $U_{ij}(s)$, and measurement error.

The Karhunen-Loève expansions of the random processes $B_i(s)$ and $U_{ij}(s)$ are

$$B_{i,0}(s) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k^0(s), \quad B_{i,1}(s) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k^1(s), \quad U_{ij}(s) = \sum_{r=1}^{\infty} \zeta_{ijr}\phi_r^U(s), \tag{3}$$

where the principal component scores $\xi_{ik} = \int_{\mathcal{D}} B_{i,0}(s)\phi_k^0(s)\,ds + \int_{\mathcal{D}} B_{i,1}(s)\phi_k^1(s)\,ds$ and $\zeta_{ijr} = \int_{\mathcal{D}} U_{ij}(s)\phi_r^U(s)\,ds$ are uncorrelated random variables with mean zero and variances λ_k and ν_r, respectively. LFPCA estimates model (2) using a truncated version with N_B and N_U components of the expansions in (3), cf. Greven et al. (2010). Thus, model (2) becomes

$$W_{ij}(s) \approx \eta(s, T_{ij}) + \sum_{k=1}^{N_B} \xi_{ik}(\phi_k^0(s) + T_{ij}\phi_k^1(s)) + \sum_{r=1}^{N_U} \zeta_{ijr}\phi_r^U(s) + \varepsilon_{ij}(s). \tag{4}$$

For illustration, fractional anisotropy along the corticospinal tract is analyzed. Figure 2 shows the first two estimated functional principal components $\{\hat{\phi}_k^0, \hat{\phi}_k^1\}$, $k = 1, 2$, and the first two curves $\hat{\phi}_r^U$, $r = 1, 2$. Additional figures can be found in the web appendix. Apparently, for both within-subject and between-subject variation, the largest part of variation is explained by variation in the general level of fractional anisotropy in distinct regions along the corticospinal tract.

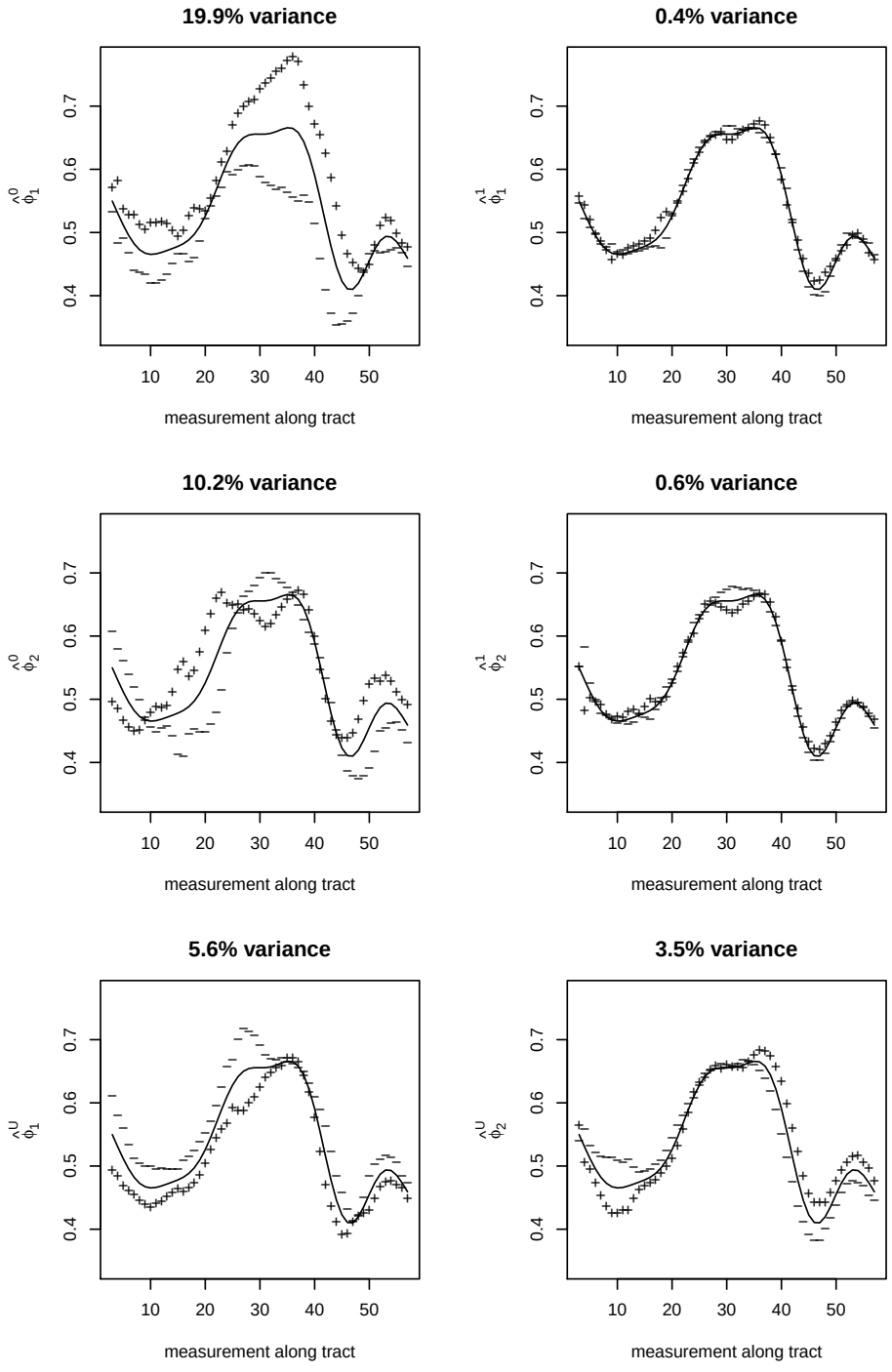

Figure 2: Estimated functional principal components $\{\hat{\phi}_k^0, \hat{\phi}_k^1\}$, $k = 1, 2$ (top/middle), and $\hat{\phi}_r^U$, $r = 1, 2$ (bottom). Depicted are the overall time-constant mean (solid line) $\pm 2\sqrt{\hat{\lambda}_k}$ times $\hat{\phi}_k^0$ or $\hat{\phi}_k^1$, and $\pm 2\sqrt{\hat{\nu}_r}$ times $\hat{\phi}_r^U$, respectively (+/−). Percentages give the proportion of overall variance that is explained by the respective component.

The B-process varies most strongly roughly between measurement points 25 and 50, top left); and for the U-process the interesting region is roughly the first half of the tract.

To choose the number of components N_B and N_U that are used to model the B- and U-processes, the proportion of explained variation can be used. Under some assumptions (in particular, for standardized visit times), total variation is given by $\int_{\mathcal{D}} \mathrm{Var}\{W_{ij}(s)\}\, ds = \sum_{k=1}^{\infty} \lambda_k + \sum_{r=1}^{\infty} \nu_r + \varsigma^2$. So N_B and N_U may be chosen as the minimum numbers such that $\{\sum_{k=1}^{N_B} \hat{\lambda}_k + \sum_{r=1}^{N_U} \hat{\nu}_r + \hat{\varsigma}^2\} / \int_{\mathcal{D}} \widehat{\mathrm{Var}\{W_{ij}(s)\}}\, ds \geq L$, where L is a pre-specified proportion of explained variance, such as $L = 0.90$ or $L = 0.95$, cf. Greven et al. (2010).

Another important feature of LFPCA is that missing values in the curves are imputed automatically. Similarly to the FPCA-based approach (see Section 2), all available observations are used to estimate the principal component bases for all model components, and best linear unbiased prediction yields estimates of corresponding scores and curves. With LFPCA, however, curves are not simply pooled across subjects, but the longitudinal structure of the data is taken into account, allowing the differential analysis of subject- and visit-level variability.

3.2 Regression Modeling using LFPCA Scores

The first and more intuitive of our LFPCA-based regression methods directly extends FPCR to the longitudinal setting. For modeling response Y_{ij} of subject i at visit j, we may use a PCR model where Y_{ij} is regressed on the scores ξ_{ik} and ζ_{ijr} from Section 3.1. To account for the repeated measures structure in Y_{ij}, we use a mixed model with subject-level random effects b_i. Furthermore, we note that scores in the LFPCA model (2) only refer to deviations from the mean surface $\eta(s, T_{ij})$. Therefore we include a time-varying intercept $\int \varphi(s)\eta(s, T_{ij})\, ds = \alpha(T_{ij})$, which can be estimated using penalized splines in the mixed models framework (see, e.g., Wood, 2011). Thus, our first LFPCR model is given by

$$\mu_{ij} = h(\eta_{ij}) \quad \text{and} \quad \eta_{ij} = \alpha(T_{ij}) + b_i + \sum_{l=1}^{p} \beta_l X_{ijl} + \sum_{k=1}^{N_B} \theta_k \xi_{ik} + \sum_{r=1}^{N_U} \delta_r \zeta_{ijr}, \qquad (5)$$

where μ_{ij} denotes the conditional mean of Y_{ij} given the covariates and random effects. We assume $b_i \sim N(0, \tau^2)$ and conditionally independent Y_{ij} with a distribution from a simple exponential family. In addition to the scores ξ_k ($k = 1, \ldots, N_B$) and ζ_r ($r = 1, \ldots, N_U$), we specify fixed effects β_l, e.g. for age and sex. Potential random effects (beyond b_i) may be added as done in the linear mixed model with functional covariate (1). Additional functional predictors would result in additional LFPCA scores and could thus be easily included. The model can also be simplified by focusing only on the scores from one level, as between-subject

(ξ_{ik}) or within-subject (ζ_{ijr}) variation. For estimation of model parameters, analogously to traditional scalar PCR, scores obtained from LFPCA are plugged in the regression model, and coefficients can, for example, be estimated in the generalized additive mixed models framework (Wood, 2011).

Due to the construction of scores, model (5) can also be interpreted as a functional linear model where predictor and coefficient curves are expressed in the same orthonormal basis; see also Section 5. As in cross-sectional FPCA, the choice of the truncation lag is an important consideration. In Sections 4 and 5, we investigate and discuss how the quality of our LFPCR model is influenced by the numbers of components N_B and N_U.

3.3 Functional Regression using Decomposed Curves

In addition to estimates of the scores ξ_k and ζ_r, LFPCA also yields estimates of the functional principal components $\{\phi_k^0(s), \phi_k^1(s)\}$ and $\phi_r^U(s)$, $k = 1, \ldots, N_B$, $r = 1, \ldots, N_U$. Thus, between-subject variation $B_i(s, T_{ij}) = B_{i,0}(s) + T_{ij}B_{i,1}(s)$ over the domain of the functions, $\mathcal{D}$, and time T, as well as within-subject variation $U_{ij}(s)$ can be reconstructed using (4). Here, $B_i(s, T_{ij})$ represents the systematic trend in subject i over time, while $U_{ij}(s)$ denotes visit-specific deviations from this trend. Both parts may be important as predictors. For example, $B_i(s, T_{ij})$ may be more relevant if $U_{ij}(s)$ constitutes mostly measurement error, while $U_{ij}(s)$ might be the more important component if curves that are *unusual* for this person are highly predictive for the outcome Y_{ij}.

Functional covariates $B_i(s, T_{ij})$ and $U_{ij}(s)$ can now be used in a functional regression model for Y_{ij}. Since response values Y_{ij} from the same individual will still be correlated we include a random intercept b_i, as described in 3.2. Because $B_i(s, T_{ij})$ and $U_{ij}(s)$ only give deviations from the general trend $\eta(s, T_{ij})$ the intercept should be allowed to vary over time. Thus, our second LFPCR approach is given by the model

$$\mu_{ij} = h(\eta_{ij}) \text{ and } \eta_{ij} = \alpha(T_{ij}) + b_i + \int_{\mathcal{D}} \gamma_B(s) B_i(s, T_{ij}) ds + \int_{\mathcal{D}} \gamma_U(s) U_{ij}(s) ds, \quad (6)$$

with $B_i(s, T_{ij}) = B_{i,0}(s) + T_{ij}B_{i,1}(s)$ and $B_{i,0}(s) = \sum_{k=1}^{N_B} \xi_{ik}\phi_k^0(s)$, $B_{i,1}(s) = \sum_{k=1}^{N_B} \xi_{ik}\phi_k^1(s)$, $U_{ij}(s) = \sum_{r=1}^{N_U} \zeta_{ijr}\phi_r^U(s)$, and conditional mean μ_{ij}. As in (5), we assume a random intercept $b_i \sim N(0, \tau^2)$ and conditionally independent observations. Additional scalar covariates can be included as fixed (or random) effects. Smooth effects of scalar covariates can easily be included in the mixed models framework (see, e.g., Wood, 2011). Additional functional predictors would be included as additional B- and U-processes resulting from LFPCA of these curves. After LFPCA is carried out for the functional predictors and $B_i(s, T_{ij})$ and $U_{ij}(s)$ are obtained, regression model (6) becomes a generalized linear mixed model with functional predictors which is analogous to model (1). Hence, the same methods

as described in Section 2 can be used for fitting it – for example, functions from `mgcv` (Wood, 2006, 2011); see the web appendix for details.

Compared to Section 3.2, this approach is particularly advantageous if the coefficient functions cannot be well expanded in the first few principal components. Only such an expansion would reduce the model to one with the first few scores as sole predictors (see also Section 5).

4 Simulation Studies

To investigate and compare the performance of different longitudinal functional regression approaches presented in the previous sections, we perform several simulation studies. To judge the performance of each method, we look at the respective observed mean squared error (MSE) $\frac{1}{n}\sum_{i,j}(\mu_{ij} - \hat{\mu}_{ij})^2$ on the simulated data. This is the mean of squared differences between the true (conditional) mean μ_{ij} of individual i at visit j, given the covariates and random effects, and the corresponding estimated mean $\hat{\mu}_{ij}$, with n denoting the overall number of observations. We do not compare fitted values to observed ones, because this would favor over-fitting methods. The reason for considering the MSE as defined above is that (a) it takes prediction of random effects into account and (b) the estimated coefficient functions are not directly comparable between the different models and/or to the true underlying model structure because of different model assumptions (see also Section 5). A summary of all methods considered is found in Table 2. In addition to LFPCR, we consider two versions of LPFR and three simple benchmark methods. The first LPFR approach uses predictor curves directly and penalized B-splines for fitting the coefficient functions (LPFR_B). Observations with missing values in the predictor curves are omitted. The second approach uses a truncated power basis for the coefficient functions and FPCA of predictor curves to impute missing values (LPFR_TRi), as described in Section 2 and implemented in `refund` (Crainiceanu and Reiss, 2011). The three simple benchmark regression tools that are considered are (1) a saturated model, i.e., $\hat{\mu}_{ij} = y_{ij}$, where y_{ij} denotes the observed response value for individual i at visit j, (2) a simple random intercept model without any covariates, and (3) a random intercept model without covariates but with a fixed smooth trend $f(T_{ij})$ over time points T_{ij}.

In our first scenario we consider single longitudinal predictor curves $W_{ij}(s)$ that are constructed according to the LFPCA model (2). For $\phi_k^0(s)$ and $\phi_k^1(s)$ we use an orthonormal sine/cosine basis, and for $\phi_k^U(s)$ we take Legendre polynomials, as done by Greven et al. (2010). Also, the visit times T_{ij} are simulated analogously to Greven et al. (2010), such that the mean for each subject is zero, and increments $T_{ij} - T_{ij-1}$ are independent draws from $U[0,1]$; then times are scaled to have unit variance. Scores ξ_{ik} and ζ_{ijr} are assumed to be normal with ξ-variances $\lambda_k = 0.5^{k-2}$, $k = 1, \ldots, 6$, and zero otherwise; for ζ-variances ν_r, we

method name	description
OF	A model with complete over-fitting, i.e., $\hat{\mu}_{ij} = y_{ij}$.
RI	A simple random intercept model without any covariates.
RI_tv	As in RI, but with an additional fixed time-varying intercept.
LFPCR, B/U90	Longitudinal functional principal components regression based on the B- and U-processes from LFPCA with 90% variance explained; see 3.1 and 3.3.
LFPCR, B/U95	As LFPCR, B/U90, but with 95% variance explained; see 3.1 and 3.3.
LFPCR, S90	Longitudinal functional principal components regression based on the scores from LFPCA with 90% variance explained; see 3.1 and 3.2.
LFPCR, S95	As LFPCR, S90, but with 95% variance explained; see 3.1 and 3.2.
LPFR_B, l	A (generalized) functional linear model with random intercept where deviations from linear coefficient functions are penalized. Coefficient functions are estimated using B-splines, and predictor curves are used directly.
LPFR_B, c	As LPFR_B, l, but penalizing deviations from constant functions.
LPFR_TRi, l	As LPFR_B, l (i.e., penalizing deviations from linear functions), but using a truncated power basis for the coefficient functions and FPCA of predictor curves to impute missing values.
LPFR_TRi, c	As LPFR_TRi, l, but penalizing deviations from constant functions.

Table 2: Methods for longitudinal functional data regression that are compared in simulation studies.

have $\nu_r = 0.5^r$, $r = 1, \ldots, 4$, and zero otherwise. For the measurement error variance we assume $\varsigma^2 = 0.01$. Our design is unbalanced with on average four observations per individual $i = 1, \ldots, 100$. After generating the functional predictor curves, we simulate response values Y_{ij} according to the linear mixed model with functional covariate and random intercept with variance $\tau^2 = 2$. Response Y_{ij} is assumed to be (conditionally) normal with variance $\sigma^2 = 2$. For the true coefficient function $\gamma(s)$ we consider (a) a nonlinear function, (b) a linear and (c) a constant one (see Figure 3, left). The simulation scenario is designed such that the signal-to-noise ratio is similar to/smaller than the one found when analyzing the tractography data. Data generation, model estimation and evaluation of the MSE is independently repeated 100 times and the resulting errors for scenario (a) are summarized in Figure 4 (top left). One can see that methods which assume the true underlying model structure (namely LPFR_B/TRi) perform best, but that LFPCR also performs quite well (in particular when using the B/U-based

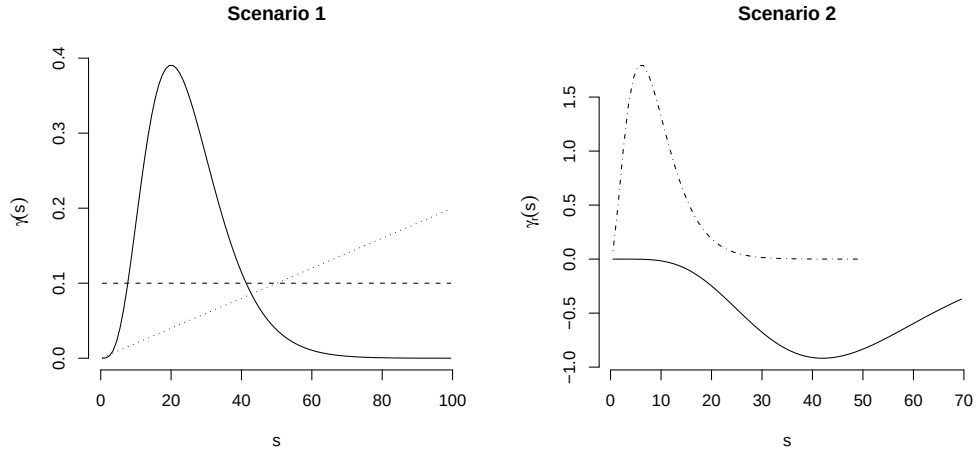

Figure 3: Assumed true coefficient functions for simulation scenarios 1a (solid), 1b (dashed), 1c (dotted), and $\gamma_1(s)$ (dashed/dotted) and $\gamma_2(s)$ (solid) for scenario 2.

approach). Results for (b) and (c) are similar (only shown in the web appendix). The over-fitting model of course produces errors around variance $\sigma^2 = 2$. If the B/U-based approach is applied it is apparently not important how much variability is exactly explained by LFPCA, as long as it is a large proportion such as 90% or 95%. This makes sense because regularization is imposed when coefficient functions are fit.

In a second scenario, we keep variances σ^2 and τ^2 as before, but consider two functional predictors defined on $(0, s_{\max})$ and generated by $W_{ij.}(s) = \frac{1}{100}(15 + \sum_{t=1}^{5} \theta_{ijt} \sin\{2\pi s(3-\theta_{ijt})/s_{\max}\} - \vartheta_{ijt})$, where $\theta_{ijt} = \tilde{\theta}_{it} + \psi_{ijt}$ and $\vartheta_{ijt} = \tilde{\vartheta}_{it} + \upsilon_{ijt}$. $\tilde{\theta}_{it}$, ψ_{ijt}, $\tilde{\vartheta}_{it}$, υ_{ijt} are independent random variables with $\tilde{\theta}_{it} \sim U[0, 4]$, $\tilde{\vartheta}_{it} \sim U[0, 2\pi]$, $\psi_{ijt} \sim U[-4/5, 4/5]$, $\upsilon_{ijt} \sim U[-2\pi/5, 2\pi/5]$ for predictor curve $W_{ij1}(s)$, and $\tilde{\theta}_{it} \sim U[0, 6]$, $\tilde{\vartheta}_{it} \sim U[0, 2\pi]$, $\psi_{ijt} \sim U[-6/10, 6/10]$, $\upsilon_{ijt} \sim U[-2\pi/10, 2\pi/10]$ for $W_{ij2}(s)$. For functional covariates $W_{ij1}(s)$ and $W_{ij2}(s)$ we have $s_{\max} = 50$ and $s_{\max} = 70$, respectively. Here, the LFPCA model (2) is not used for generating the functional predictors, but curves are directly simulated (in a manner similar to Tutz and Gertheiss (2010)). The assumed true coefficient functions $\gamma_1(s)$ and $\gamma_2(s)$ are shown in Figure 3 (right). The design is now balanced with five functional observations per subject $i = 1, \ldots, 100$. Before fitting the regression models, we add white noise measurement error with variances 0.008^2 and 0.004^2 to $W_{ij1}(s)$ and $W_{ij2}(s)$, respectively. As before, data generation, model estimation and evaluation of errors is independently repeated 100 times. Results (see web appendix) are similar as before: LPFR_B/TRi, which assume the correct model, perform best, but superiority over LFPCR is only moderate. The only big difference to scenario 1 is that now RI performs as good/bad as RI_tv.

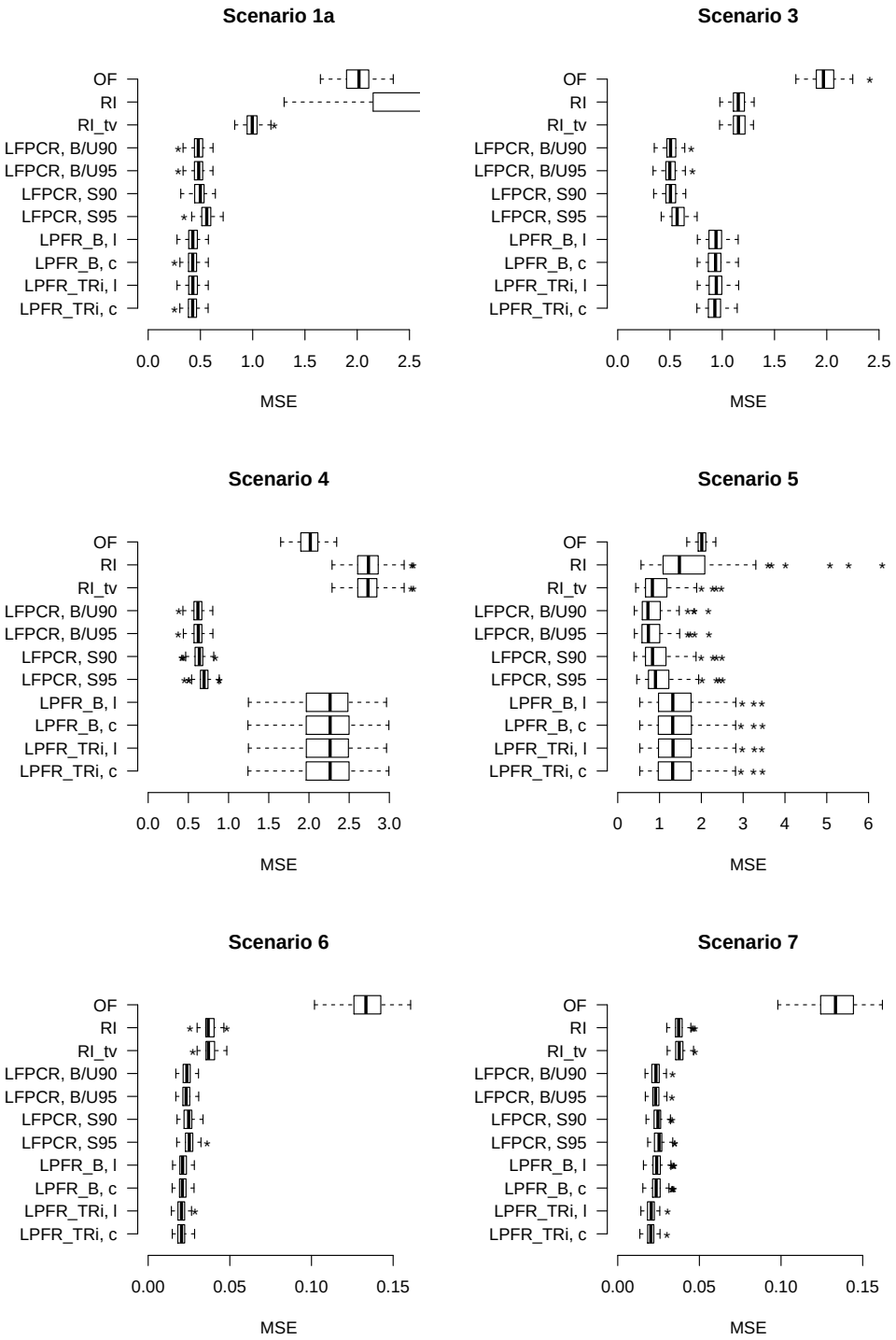

Figure 4: Results of simulation scenarios 1a, 3–7 in terms of the (observed) mean squared error (MSE) $\frac{1}{n}\sum_{i,j}(\mu_{ij} - \hat{\mu}_{ij})^2$. Abbreviations are defined in Table 2. For 1a, the box for RI is not completely shown because a number of error values were too extreme.

When the data generating process deviates from this simple model, however, LFPCR is distinctly superior to regression modeling where model (1) is assumed; see Figure 4 (top right, middle). Data generation for those and further scenarios is summarized below:

- *Scenario 3*: The same specifications as in scenario 1, but now the ξ- and ζ-scores are used as predictors with all regression coefficients equal to 1.
- *Scenario 4*: Again the same specifications as in scenario 1, but now only the U-process from (2) is used as functional predictor in a functional linear model with random intercept and with the true coefficient function having a shape like the nonlinear function from scenario 1 (see Figure 3, left). That means, between-subject variation is due only to the random intercept.
- *Scenario 5*: As in scenario 4, but now the U-process is seen as additional measurement error, and the only relevant functional predictor is the B-process $B_{i,0}(s) + T_{ij}B_{i,1}(s)$.
- *Scenario 6*: For generating the linear predictor the same specifications as in scenario 2 are used, but after employing the logistic function as the link binary outcomes are sampled.
- *Scenario 7*: As scenario 6, but in both sets of predictor curves 50 blocks of 2–4 missings are randomly distributed.

It can be seen that in such cases where the general functional trend $\eta(s, T)$ is irrelevant but only deviations from that trend are informative (scenarios 3–5), a linear mixed model with functional covariate is inadequate. Furthermore, LFPCR based on B- and U-processes seems to be superior to the score-based approach. In Section 5 we will have a closer look at connections between the different regression approaches.

For the logit model (Figure 4, bottom) results are qualitatively the same as for linear modeling (scenario 2). If missings are found in the predictor curves (Figure 4, bottom right), imputing missing values is apparently superior to omitting curves with missing values. The latter is done by LPFR_B, whereas LPFR_TRi uses imputation as described in Section 2.

5 Comparing Different Approaches

Deeper insight into connections and differences between the methods considered will help us understand some of the findings from the simulations. In Figure 4 we saw, for example, that B/U-based LFPCR tends to perform better than an approach using scores.

5.1 Comparing Different LFPCR Approaches

First, we compare the two presented versions of LFPCR, the one using LFPCA scores as predictors and the functional version based on B- and U-processes. The "functional part" in model (5) is

$$\begin{aligned}\sum_{k=1}^{N_B}\theta_k\xi_{ik}+\sum_{r=1}^{N_U}\delta_r\zeta_{ijr} &= \sum_{k=1}^{N_B}\theta_k\left(\int_{\mathcal{D}}B_{i,0}(s)\phi_k^0(s)ds+\int_{\mathcal{D}}B_{i,1}(s)\phi_k^1(s)ds\right)\\ &+ \sum_{r=1}^{N_U}\delta_r\int_{\mathcal{D}}U_{ij}(s)\phi_r^U(s)ds\\ &= \int_{\mathcal{D}}B_{i,0}(s)\sum_{k=1}^{N_B}\theta_k\phi_k^0(s)ds+\int_{\mathcal{D}}B_{i,1}(s)\sum_{k=1}^{N_B}\theta_k\phi_k^1(s)ds\\ &+ \int_{\mathcal{D}}U_{ij}(s)\sum_{r=1}^{N_U}\delta_r\phi_r^U(s)ds.\end{aligned}$$

This is a functional linear model with predictors $B_{i,0}(s)$, $B_{i,1}(s)$ and $U_{ij}(s)$, and corresponding coefficient functions restricted to spaces spanned by the first eigenfunctions. These restrictions may explain problems in some situations. A more flexible approach would be to estimate these coefficient functions directly, which gives a third way to do LFPCR – a version that also uses B- and U-processes as predictors, but $B_{i,0}(s)$ and $B_{i,1}(s)$ separately. Since, however, $B_i(s,T_{ij})=B_{i,0}(s)+T_{ij}B_{i,1}(s)$ as the systematic trend in subject i over time can be nicely interpreted, we prefer (6). From the considerations above it becomes clear that the latter approach and the score-based approach are neither equivalent nor is one a special case of the other.

Another important difference between the B/U-based and the score-based approach is that in the latter case regularization is imposed by selecting the number of functional principal components. This number is tied to explaining the variability in the predictor curves and not to the parameter functions. If the B/U-based approach is applied, however, regularization through penalized splines is directly applied to the coefficient curves. This makes the approach more robust to over-fitting; over-fitting may become a problem for the score-based approach if a too high proportion of explained variance is chosen or if the B- or U-process is not associated with the response, as for example in simulation scenario 5. A possible way to alleviate these problems could be to use regularization when fitting the regression coefficients θ_k and δ_r.

5.2 Comparing LPFR and LFPCR

The score-based LFPCR model can be seen to be a special case of the LPFR model only in particular cases. Consider the case where variability over time in

$\hat{\eta}(s,T)$ is very low and the influence of $B_{i,1}(s)$ can be neglected (see, e.g., $\hat{\phi}_k^1$ in Figure 2). Furthermore, assume that only the score ζ_{ij1} is associated with the outcome. Then, it can be shown (see the web appendix) that the linear predictor has approximately the form $\tilde{\alpha} + \tilde{b}_i + \sum_{l=1}^{p} X_{ijl}\beta_l + \int_{\mathcal{D}} W_{ij}(s)\delta_1\phi_1^U(s)\,ds$. This is an LPFR model with random intercept $\tilde{b}_i$ and coefficient function $\delta_1\phi_1^U(s)$. In general, however, the score-based LFPCR model and LPFR are different.

Next, we investigate the differences between LPFR and the B/U-based LFPCR seen in the simulation studies. Assume that model (1) and decomposition (2) hold. Then we have

$$\begin{aligned}\int_{\mathcal{D}} W_{ij}(s)\gamma(s)ds &= \int_{\mathcal{D}} \gamma(s)\eta(s,T_{ij})ds + \int_{\mathcal{D}} \gamma(s)(B_{i,0}(s) + T_{ij}B_{i,1}(s))ds \\ &+ \int_{\mathcal{D}} \gamma(s)U_{ij}(s) + \int_{\mathcal{D}} \gamma(s)\varepsilon_{ij}(s)ds \\ &= \alpha(T_{ij}) + \int_{\mathcal{D}} \gamma(s)B_i(s,T_{ij})ds + \int_{\mathcal{D}} \gamma(s)U_{ij}(s)ds + \tilde{\varepsilon}_{ij},\end{aligned}$$

where $\tilde{\varepsilon}_{ij}$ is noise with mean zero. If $W_{ij}(s)$ is assumed to be a smooth curve without measurement error $\varepsilon_{ij}(s)$, noise $\tilde{\varepsilon}_{ij}$ disappears and the (generalized) functional linear model (1) can be seen as a special case of LFPCR using B- und U-processes where $\gamma_B(s) = \gamma_U(s) = \gamma(s)$ and $\alpha(T_{ij})$ has a specific form. Thus, if (1) holds, LFPCR using B- und U-processes will also be an adequate modeling approach as long as (2) is a reasonable approximation to the (functional) data generating process, as seen in simulation scenarios 1, 6 and 7. By contrast, if the LFPCR model is correct and the overall mean trend $\eta(s,T_{ij})$ is not relevant for the response, or if $\gamma_B(s) \neq \gamma_U(s)$, the functional linear part in (1) is not appropriate, as observed in scenarios 3–5. In summary, LFPCR is more general than methods based on model (1), such as LPFR. If the latter model is correct, the performance of LFPCR suffers slightly due to its generality. On the other hand, if specific LPFR assumptions are incorrect, LPFR does not adequately estimate the association between functional predictors and outcomes. In the web appendix, we additionally compare the B/U-based LFPCR model to much simpler mixed models.

6 Application to the Tractography Data

Multiple sclerosis (MS) is a neurological disease that affects the central nervous system and in particular damages white matter tracts in the brain through lesions, myelin loss and axonal damage. Diffusion tensor imaging (DTI) is a magnetic resonance imaging technique that allows the extraction of information on individual tracts and thus allows a better understanding of damages in neuronal tracts and how these relate to disease progression. In our study, 176 MS patients were repeatedly scanned over time for an average of 1.27 years to follow disease

progression, measured by disability scores such as the 9 hole peg test (*peg9*), and corresponding changes in DTI measurements. The *peg9* measures the time required to put nine pegs into nine holes and then remove them (cf. Cutter et al., 1999). Several summary indices including the fractional anisotropy and the magnetization transfer ratio were extracted from the DTI images along several important tracts, including the corpus callosum, the corticospinal tract and the optic radiations tract. Our primary goal is to relate changes in disability to corresponding changes in tract profiles.

As seen in Section 4, methods based on the (generalized) functional linear model perform well as long as this model corresponds to the true underlying regression structure. For the tractography data it seems reasonable to assume that the functional covariates as a whole (for example fractional anisotropy along a tract of interest) contain relevant information. Therefore the linear mixed model with functional predictors may be an adequate tool for analyzing the relationship between measures from tractography and disability of MS patients.

To investigate whether the assumption of equal effects for trend and deviation from the trend holds, we will also consider our new LFPCR approach. Following results from the simulation study, we will use the better-performing estimation approach based on B- and U-processes.

6.1 Generalized Functional Linear Models with Random Intercept

We consider the fractional anisotropy (FA) along the corpus callosum, the corticospinal tract that is contralateral to the dominant hand, and the optic radiations tract as potential functional predictors for the *peg9* score for the dominant hand. In addition to functional predictors, we consider scalar covariates sex and age, and a dummy variable indicating whether this is the patient's first visit or not. The latter is done to account for a potential learning effect with respect to the conducted test (see Goldsmith et al., 2012). We first fit a generalized functional linear model with random intercept. Since the (conditional) distribution of the *peg9* scores is slightly skewed and scores are positive, we assume a Gamma distribution with log-link, instead of a normal distribution with identity-link. We found that there is only a clear dependence between measures along the corticospinal tract and *peg9* (see the web appendix), which is biologically plausible in that the corticospinal tract connects the motor cortex to the opposite side of the body, and therefore mediates motor signals.

In addition to fractional anisotropy, we may also consider the magnetization transfer ratio (MTR) along the corticospinal tract as a potential predictor for *peg9*. Figure 5 shows estimated coefficient functions for a generalized functional linear model with random intercept and these two functional predictors (and scalar ones as above). The solid black line is obtained for LPFR with a

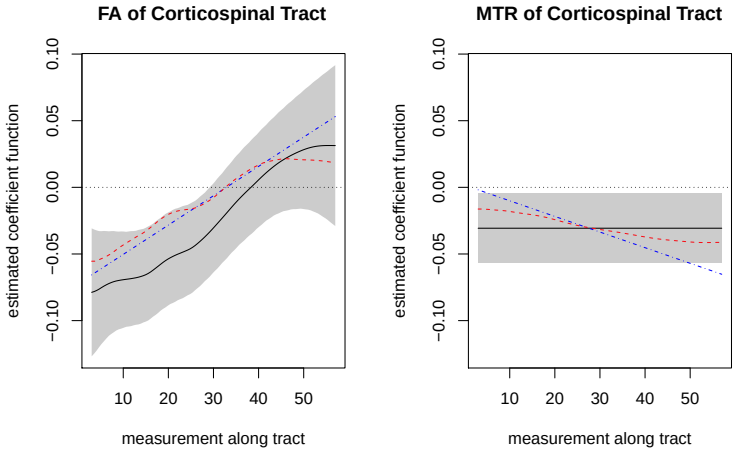

Figure 5: Estimated coefficient functions when a generalized functional linear model with random intercept is fit to data with functional predictors fractional anisotropy and magnetization transfer ratio of the corticospinal tract (and scalar predictors visit > 1, sex, age). Considered are the results of a complete-case analysis using LPFR with a B-spline basis for the coefficient functions and penalizing deviations from a constant line (solid black), the same penalty but using a truncated power spline basis with knots at each observation point and imputing missing values using FPCA (dashed red), and the `refund` implementation, where deviations from a linear function are penalized and missings are imputed (dashed/dotted blue). The shaded region corresponds to 90% pointwise confidence intervals as provided by `mgcv`.

penalized B-spline basis for the coefficient functions and penalizing deviations from a constant line. The red dashed line refers to the same penalty but using a truncated power spline basis with knots at each observation point. The blue dashed/dotted line is the `refund` implementation `lpfr()`, where deviations from a linear function are penalized. In the latter two cases missing observations are imputed as described in Section 2. In the first case, curves with missing values are omitted (reducing sample size by 38%). Since missings typically occur for technical reasons, it can be assumed that curves are missing at random, and hence a complete-case analysis is reasonable. The shaded region corresponds to 90% pointwise confidence intervals as provided by `mgcv`. Apparently there is a dependence between measures along the corticospinal tract and *peg9*, where high values of fractional anisotropy in the first half of the tract result in lower disability scores. This makes sense because decreasing fractional anisotropy indicates disease progression (see Harrison et al., 2011). When looking at the solid and dashed lines for MTR, where deviations from a constant are penalized, it seems that in this case using mean MTR is a sufficient way to include anatomical information. Note that if deviations from linearity (dashed/dotted blue) are penalized, both coefficient functions are estimated to be non-constant, but linear. This, however, seems to us an artifact of the penalization.

	GFLM		LFPCR	
variable	estimated coefficient	p-value	estimated coefficient	p-value
(Intercept)	4.3772	0.0000	2.8877	0.0000
visit > 1	-0.0436	0.0111	-0.0598	0.0024
sex	0.2377	0.0007	0.2049	0.0009
age	0.0065	0.0165	0.0063	0.0098

Table 3: Estimated fixed effects for scalar predictors when a generalized functional linear model with random intercept is directly fit to data with functional predictors fractional anisotropy and magnetization transfer ratio along the corticospinal tract (GFLM, left), or when the model is fit to data with LFPCA B-and U-processes of the fractional covariates being used as functional predictors (LFPCR, right).

Results for the scalar covariates (complete case analysis) are given in Table 3 (left). Apparently there is a learning effect after the first visit. Older patients tend to have higher disability scores, but the largest effect is found for sex, with males having distinctly higher scores than females.

6.2 Longitudinal Functional Principal Components Regression

We carry out LFPCA of fractional anisotropy (compare Figure 2) and magnetization transfer ratio along the corticospinal tract. The scores and functional principal components are then used to reconstruct the respective B- and U-processes $B_i(s, T_{ij})$ and $U_{ij}(s)$ for each patient i at visit j. Resulting curves for the first five patients are shown in Figure 6, with colors corresponding to the patients' *peg9* scores. These curves are then used to build a generalized functional linear model with random intercept, and additional scalar predictors age, sex and visit > 1 and time-varying intercept (see 3.3). Estimates for scalar covariates are given in Table 3 (right), estimated functions in Figure 7. Since missing values are already imputed when constructing B- and U-processes, differences between different fitting procedures for the regression model can be neglected, as only the bases differ. Effects of scalar covariates are estimated to be similar to estimates above (Table 3, left). The large difference between intercepts occurs because in the (generalized) functional linear model predictor curves are not centered, whereas the B- and U-processes are centered by construction (see 3.1). The time-varying intercept in Figure 7 (top left) indicates that the disability score increases over time, as expected in a diseased population. From the coefficient functions for both the B- and U-process of fractional anisotropy (Figure 7, middle), it follows that patients with higher FA values than the 'average patient' in the first half of the tract tend to have lower disability scores, which is in accordance with findings from 6.1. Assuming equal coefficient functions for these two processes, as done

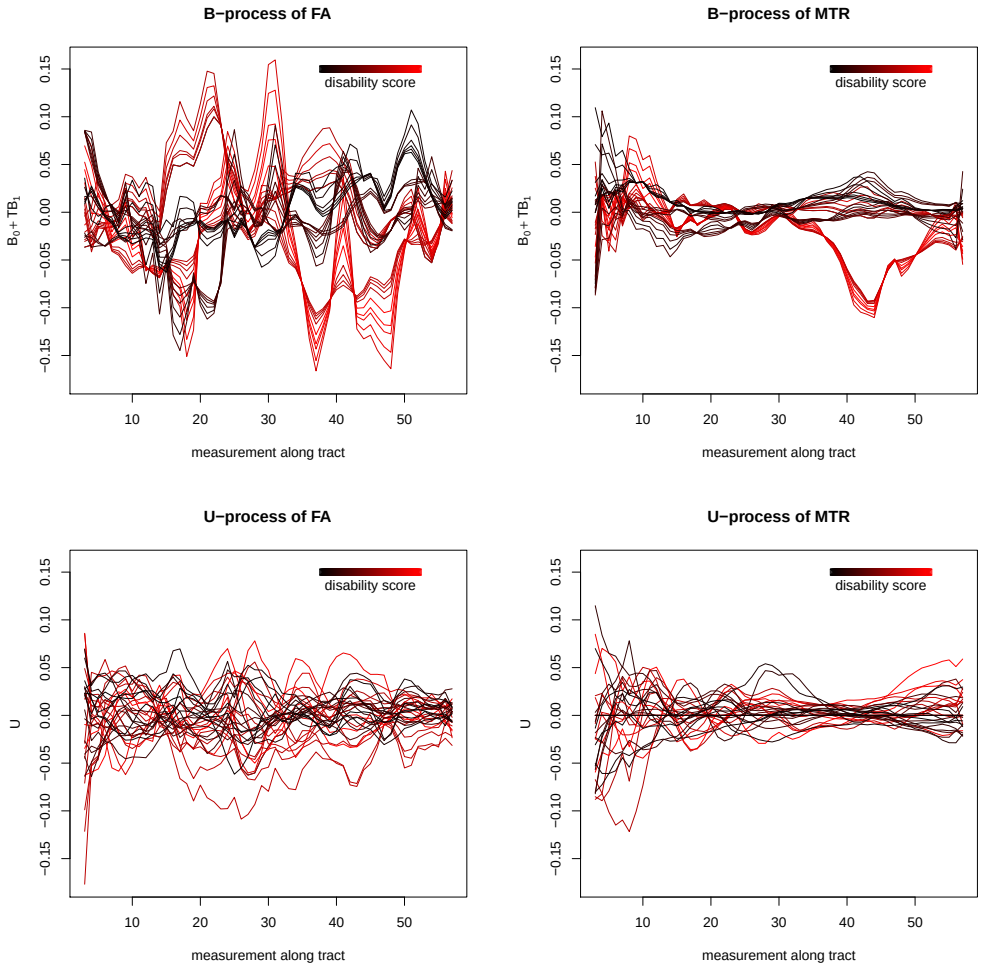

Figure 6: B- and U-processes estimated by LFPCA for FA and MTR along the corticospinal tract. For illustration, the results for the first five patients are shown, colors correspond to the patients' *peg9* disability scores at the corresponding visit.

in 6.1, seems reasonable. Estimated coefficient curves for magnetization transfer ratio (Figure 7, right) indicate that mean MTR along the tract is highly informative with respect to *peg9* when, for example, different patients are compared (between-patients variation). Within-patient variation (the U-process) of MTR seems to be less informative and may constitute mostly measurement error.

Correlation between *peg9* scores predicted by the generalized functional linear model with random intercept and the scores predicted by LFPCR is greater than 0.99. Variances of scores predicted by the two methods are similar (232, resp. 237). That implies that *peg9* predictions using either method are very similar (see also the web appendix). In the generalized functional linear model, the

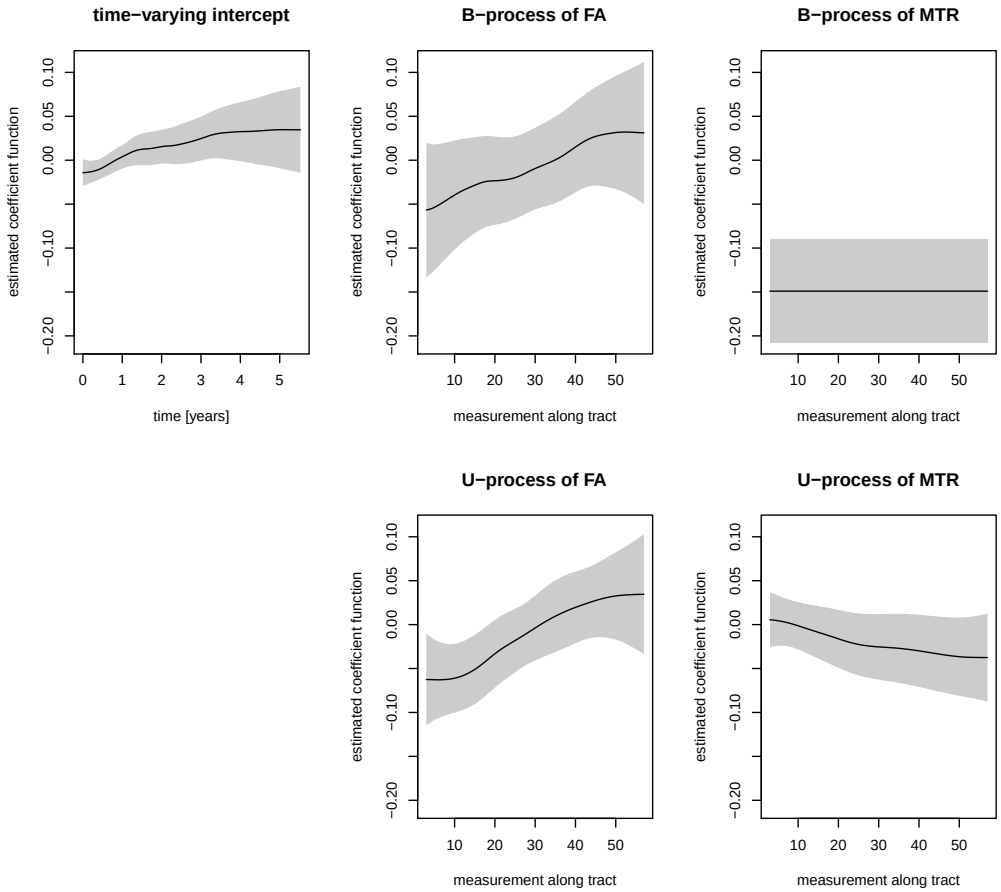

Figure 7: Estimated coefficient functions when a generalized functional linear model with random intercept is fit to *peg9* scores with LFPCA B- and U-processes of fractional anisotropy (FA) and magnetization transfer ratio (MTR) of the corticospinal tract being used as functional predictors, and with scalar predictors visit > 1, sex, age.

fitted function for MTR (see dashed line in Figure 5) seems to be a compromise between the coefficient functions of the B- and U-processes of MTR (see Figure 7, right).

7 Summary and Discussion

We presented and compared different tools for scalar-on-function regression that can be applied when observations are taken repeatedly over time. We proposed two novel versions of principal components regression: longitudinal functional principal components regression (LFPCR), based on longitudinal functional principal components analysis (LFPCA). The first approach uses scalar scores ob-

tained from LFPCA as explanatory variables in a longitudinal mixed model. The second version proposed uses the LFPCA decomposition of functional covariates to construct a mixed model with (multiple) functional predictors. For each functional covariate, two processes are obtained that are then used as predictors. (1) a process describing systematic trends within subjects over time and (2) visit-specific deviations from that trend. By contrast to the score-based approach, the second method also uses the covariates' functional character when building the regression model, and it has been shown in simulation studies that it tends to perform better. Furthermore, it yields nice interpretations. For example, if deviations from subject-specific functional trends are just measurement error, irrelevant for the response variable, the corresponding coefficient function will be around zero. But as in a standard functional linear model, the coefficient function for the subject-specific trend will indicate the interesting regions in the signals' domain, and the functional shape of the influence on the response. LFPCR distinctly outperforms mixed models that use functional covariates directly when the overall trend in the functional predictors is not important for the response. On the other hand, it is competitive if the (generalized) functional linear model is (close to) the true model and the LFPCA model is a good approximation to the functional data generating process.

The presented LFPCR approach can also be applied when only the functional predictors vary over time, but the response does not change from visit to visit. For example, consider the subjects' case status when it is to be discriminated between MS patients and controls. In a case like this, we focus on the subject-specific deviations from the overall trend, and use this processes as functional predictors in an adequate regression model; for example, in a logit model, if the aim of the analysis is binary classification (as case/control). As the trend in the functions might be an important predictor, advantages over taking just the functions' average can be expected. Such an analysis of the tractography data is provided in the web appendix.

All the proposed methods result in a (generalized) additive mixed model with scalar or vector-valued random effects, scalar and functional fixed effects and potentially smooth effects of covariates such as time. Such models can be fit using R package `mgcv` (Wood, 2006, 2011). If LFPCA is to be applied before fitting the regression model, R-code available at `http://www.statistik.lmu.de/institut/ag/fda/research.html` can be used. With LFPCA, missing values in the predictor curves are imputed automatically. If curves are used directly, imputation and model fitting is also possible, as described here and implemented in `refund` (Crainiceanu and Reiss, 2011).

Since the final regression model considered here is additive, it can easily be extended, for example by two-dimensional surfaces describing interactions of scalar predictors, modeling higher-dimensional functional predictors such as images, or spatial effects. All such models can be implemented within the mixed models framework in `mgcv` (Wood, 2006, 2011).

Acknowledgements

The research of Gertheiss and Greven was supported by the German Research Foundation through Emmy Noether grant GR3793/1-1. The work of Goldsmith and Crainiceanu 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 National Institute Of Neurological Disorders And Stroke or the National Institute of Health. We thank Daniel Reich from the National Institute Of Neurological Disorders And Stroke, the Departments of Radiology and Neurology at Johns Hopkins Hospital, and the Department of Biostatistics at Johns Hopkins University for providing the data and related information.

References

Crainiceanu, C. M. and P. Reiss (2011). *refund: Regression with Functional Data*. R package version 0.1-5.

Crainiceanu, C. M., D. Ruppert, G. Claeskens, and M. P. Wand (2005). Exact likelihood ratio test for penalised splines. *Biometrika 92*, 91–103.

Cutter, G. R., M. L. Baier, R. A. Rudick, D. L. Cookfair, J. S. Fischer, J. P. an K. Syndulko, B. G. Weinshenker, J. P. Antel, C. Confavreux, G. W. Ellison, F. Lublin, A. E. Miller, S. M. Rao, S. Reingold, A. Thompson, and E. Willoughby (1999). Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain 122*, 871–882.

Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis*. New York: Springer.

Frank, I. E. and J. H. Friedman (1993). A statistical view of some chemometrics regression tools (with discussion). *Technometrics 35*, 109–148.

Gertheiss, J. and G. Tutz (2009). Supervised feature selection in mass spectrometry-based proteomic profiling by blockwise boosting. *Bioinformatics 25*, 1076–1077.

Goldsmith, J., J. Bobb, C. Crainiceanu, B. Caffo, and D. Reich (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics 20*, 830–851.

Goldsmith, J., C. Crainiceanu, B. Caffo, and D. Reich (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society C (Applied Statistics)*. (to appear).

Greven, S., C. Crainiceanu, B. Caffo, and D. Reich (2010). Longitudinal functional principal components analysis. *Electronic Journal of Statistics 4*, 1022–1054.

Greven, S., C. Crainiceanu, H. Küchenhoff, and A. Peters (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics 17*, 870–891.

Harrison, D. M., B. S. Caffo, N. Shiee, J. A. D. Farrell, P.-L. Bazin, S. K. Farrell, J. N. Ratchford, P. A. Calabresi, and D. S. Reich (2011). Longitudinal changes in diffusion tensor-based quantitative MRI in multiple sclerosis. *Neurology 76*, 179–186.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society B 64*, 411–432.

James, G. M. and B. W. Silverman (2005). Functional adaptive model estimation. *Journal of the American Statistical Association 100*, 565–576.

James, G. M., J. Wang, and J. Zhu (2009). Functional linear regression that's interpretable. *The Annals of Statistics 37*, 2083–2108.

Marx, B. D. and P. H. C. Eilers (1999). Genaralized linear regression on sampled signals and curves: A p-spline approach. *Technometrics 41*, 1–13.

Massy, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association 60*, 234–256.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus (2008). *Genealized, Linear, and Mixed Models* (2nd ed.). Hoboken, NJ: Wiley.

Müller, H.-G. and U. Stadtmüller (2005). Generalized functional linear models. *The Annals of Statistics 33*, 774–805.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2nd ed.). New York: Springer.

Ramsay, J. O., H. Wickham, S. Graves, and G. Hooker (2011). *fda: Functional Data Analysis*. R package version 2.2.7.

Reiss, P. T. and R. T. Ogden (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association 102*, 984–996.

Reiss, P. T. and R. T. Ogden (2010). Functional generalized linear models with images as predictors. *Biometrics 66*, 61–69.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.

Tutz, G. and J. Gertheiss (2010). Feature extraction in signal regression: A boosting technique for functional data regression. *Journal of Computational and Graphical Statistics 19*, 154–174.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. London: Chapman & Hall.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society B 73*, 3–36.