Published in final edited form as:

J Comput Graph Stat. 2015 April 1; 24(2): 477–501. doi:10.1080/10618600.2014.901914.

Functional Additive Mixed Models

Fabian Scheipl[1], **Ana-Maria Staicu**[2], and **Sonja Greven**[1]

[1]Ludwig-Maximilians-Universität München

[2]North Carolina State University

Abstract

We propose an extensive framework for additive regression models for correlated functional responses, allowing for multiple partially nested or crossed functional random effects with flexible correlation structures for, e.g., spatial, temporal, or longitudinal functional data. Additionally, our framework includes linear and nonlinear effects of functional and scalar covariates that may vary smoothly over the index of the functional response. It accommodates densely or sparsely observed functional responses and predictors which may be observed with additional error and includes both spline-based and functional principal component-based terms. Estimation and inference in this framework is based on standard additive mixed models, allowing us to take advantage of established methods and robust, flexible algorithms. We provide easy-to-use open source software in the pffr() function for the R-package refund. Simulations show that the proposed method recovers relevant effects reliably, handles small sample sizes well and also scales to larger data sets. Applications with spatially and longitudinally observed functional data demonstrate the flexibility in modeling and interpretability of results of our approach.

Keywords

Functional data analysis; functional principal component analysis; P-splines; Smoothing; Varying coefficient models

---

In recent years, many scientific studies have collected functional data that exhibit correlation structures amenable to explicit modeling. Such structures may arise from a longitudinal study design (e.g. Goldsmith et al., 2012; Greven et al., 2010; Morris and Carroll, 2006), crossed designs (e.g. Aston et al., 2010), or spatial sampling of curves (e.g. Delicado et al., 2010; Giraldo et al., 2010; Gromenko et al., 2012; Nerini et al., 2010; Staicu et al., 2010). Simultaneously, regression for independent functional responses (e.g. Faraway, 1997) has made large advances, including both multiple scalar (e.g. Reiss et al., 2010) and multiple functional predictors in concurrent or more general relationships (e.g. Ivanescu et al., 2012). Our work is motivated by a longitudinal neuroimaging study containing repeated measurements of a functional proxy variable for neuronal health along 3 white matter tracts derived from diffusion tensor imaging (DTI). The goal of our analysis is to quantify the relationship of these function-valued proxy measures while accounting for the longitudinal correlation structure as well as the effects of patient characteristics like age and gender. This DTI study is an example of a longitudinal functional data set where models must account for the correlation structure of the data while including both scalar and functional covariates in the predictor.

To address these challenges, we propose conditional regression models for functional responses that accommodate general correlation structures via functional and scalar random effects as well as flexible linear or nonlinear effects of scalar and functional covariates. The major contributions of this paper thus consist in 1) developing a general inferential framework for additive mixed models for correlated functional responses that accommodates diverse correlation structures and flexible modeling of the mean structure extending Ivanescu et al. (2012), 2) unifying two previously separate strands of prior work, by subsuming both functional principal component- (FPC) and spline-based approaches, and 3) evaluating our implementation available in the the R-package refund (Crainiceanu et al., 2011) on real and simulated data.

Our goal is to describe and implement a framework that offers analysts of functional data similar flexibility in model specification to what is available in current implementations of (geo-)additive mixed models for scalar data. Specifically, we consider structured additive regression models of the general form

$$y_i(t) = \sum_{r=1}^{R} f_r(\mathscr{X}_{ri}, t) + \varepsilon_i(t), \quad (1)$$

for functional responses $y_i(t), i = 1,\ldots,n,$ observed over a domain $\mathscr{T}$. Each term in the additive predictor is a function of a) the index t of the response and b) a subset $\mathscr{X}_r$ of the complete covariate set $\mathscr{X}$ including scalar and functional covariates and (partially) nested or crossed grouping factors. To make this more concrete, Table 1 shows the most important combinations of $\mathscr{X}_r$ and effect shapes available in our framework. We assume a white noise error process independent of $\mathscr{X}$, such that the $\varepsilon_i(t)$ are independent and identically distributed (i.i.d.) Gaussian variables with mean zero and constant variance σ_ε^2 across $\mathscr{T}$. Additionally, smooth and potentially correlated error curves can be included via curve-specific random effects to model (co-)variance along t and dependence between functional observations. We assume all effects in Table 1 to be smooth but unknown functions in the covariates, and this smoothness assumptions on all components of the predictor ensures smoothness of $y_i(t)$ up to the white noise measurement error $\varepsilon_i(t)$. Scalar random effects b_g are mean zero Gaussian variables with general covariance structure between the different levels of g. Functional random effects $b_g(t)$ for a grouping variable g with M levels are modeled as realizations of a mean-zero Gaussian random process on $\{1,\ldots,M\} \times \mathscr{T}$ with a general covariance function $K^b(m,m',t,t') = \mathrm{Cov}(b_m(t), b_{m'}(t'))$ that is smooth in t, where m,m' denote different levels of g. Note that our model class and software admits multiple partially or completely nested or crossed grouping factors for both scalar and functional random effects, but different random effects $b_g(t), b_{g'}(t)$ are assumed to be mutually independent. We assume integrability for the effects of functional covariates. Our implementation for functional effects such as $\int x_i(s)\beta(s,t)\,ds$ also accommodates varying integration ranges with fixed, potentially observation-specific integration limits $l_i(t), u_i(t)$, similar to the historical functional model in Malfait and Ramsay (2003). Densely as well as sparsely observed functional responses and suitably preprocessed functional predictors with measurement error can be used in this framework. We approximate each term $f_r(\mathscr{X}_r, t)$ by a linear combination of basis functions defined by the tensor product of marginal bases for $\mathscr{X}_r$

and *t*. Since basis dimensions have to be sufficiently large to ensure enough flexibility, maximum likelihood estimation of model (1) is likely to lead to substantial overfitting. The penalized likelihood approach described in Section 1.2 stabilizes estimates by suppressing variability not strongly supported by the data and finds a data-driven compromise between goodness of fit and simplicity of the fitted effects.

Most existing work on functional random effects has considered only special cases such as the functional random intercept model (Abramovich and Angelini, 2006; Di et al., 2009; Krafty et al., 2011), functional random intercept and slope model (Greven et al., 2010), a single level of random effects functions (Antoniadis and Sapatinas, 2007; Guo, 2002; Qin and Guo, 2006), or a two or three-level hierarchy (Baladandayuthapani et al., 2008; Bigelow and Dunson, 2007; Brumback and Rice, 1998; Li et al., 2007; Morris et al., 2003; Scarpa and Dunson, 2009; Staicu et al., 2010; Zhou et al., 2010). Aston et al. (2010) consider a general functional random effects structure under the assumption of a joint functional principal component (FPC) basis for all functional random effects in the model, which are estimated under a working independence assumption between curves. It is unclear, however, how well this approach works if the latent processes do not share the same eigenfunctions and how the correlation between functional observations affects FPC estimation. FPC estimation for correlated observations is a topic of ongoing research (c.f. Hörmann and Kokoszka, 2010, 2011; Panaretos and Tavakoli, 2013a,b). Morris and Carroll (2006); Morris et al. (2003); Zhu et al. (2011) propose a general Bayesian functional linear mixed model based on a wavelet transformation of (usually very spiky) data observed on an equidistant grid. The model proposed by Morris and Carroll (2006) includes correlation between different random effects and heterogeneous residual errors, which we do not. Our approach, on the other hand, is well suited to smooth underlying curves and allows a more general mean structure than previous functional linear mixed models; in particular we are able to estimate smooth nonlinear or linear effects of scalar and/or functional covariates within the same framework. In addition, we are able to handle data on non-equidistant or sparse grids.

To the best of our knowledge, our proposal is the first publicly available implementation that allows such a high level of flexibility for a functional regression model – prior work either limits the predictor to the effect of a single functional covariate and a functional intercept, such as the linmod function in package **fda** (Ramsay et al., 2011) for the **R** language (**R** Development Core Team, 2011) or to linear effects of scalar covariates, such as the fosr function for function-on-scalar regression (Reiss et al., 2010) in the **R**-package **refund**. Like the linear function-on-function regression approach in Ivanescu et al. (2012) we build on, both approaches are limited to independent functional responses. Morris and Carroll (2006) provide a closed source implementation for wavelet-based functional linear mixed models in WFMM (Herrick, 2013) that allows very general random effect and residual structures, but implement neither effects of functional covariates nor nonlinear effects of scalar covariates. The **PACE** package (Fang et al., 2013) for **MATLAB** implements FPC based regression models where the predictor is limited to the effect of a single functional or scalar covariate. Our proposal has some similarities with the regression models for independent or longitudinal scalar responses in Goldsmith et al. (2011, 2012), implemented in the pfr and lpfr functions in **refund**, since we also base inference on additive mixed models for scalar-

on-scalar regression. However, the extension to functional responses and functional random effects with flexible correlation structure as well as the inclusion of FPC-based effects is non-trivial.

The paper is organized as follows: Section 1 develops our general approach and estimation framework for functional additive mixed models. Our method is evaluated in a simulation study and in an application to the motivating longitudinal DTI study in Section 2. Section 3 closes with a discussion and outlook.

1 Penalized regression for correlated functional data

Functional responses $y_i(t)$ are observed on a grid of T_i points $t_i=\left(t_{i1},\ldots,t_{iT_i}\right)^\mathsf{T}$. To simplify notation, we assume identical grids $t_i \equiv t=(t_1,\ldots,t_T)^\mathsf{T}$ for $i = 1,...,n$ in the following, but note that functional responses observed on irregular and/or sparse grids are naturally accommodated in the rephrased model formulation given in (2). Then, model (1) can be expressed as

$$y_{il}=\sum_{r=1}^{R} f_r\left(\mathscr{X}_{ri},t_l\right)+\varepsilon_{il} \quad (2)$$

for $i = 1,...,n$ and $l = 1,...,T$. The assumption of white noise errors translates to $\varepsilon_{il} \overset{\text{i.i.d.}}{\sim} N\left(0,\sigma_\varepsilon^2\right)$. The smoothness assumption on $\mathrm{E}(y_i(t))$ is preserved implicitly by enforcing smoothness across $\mathscr{I}$ for all $f_r\left(\mathscr{X}_r,t\right)$. To fit the model, we form $\boldsymbol{y}=\left(\boldsymbol{y}_1^\mathsf{T},\ldots,\boldsymbol{y}_n^\mathsf{T}\right)^\mathsf{T}$, an nT-vector that holds the concatenated function evaluation vectors $\boldsymbol{y}_i=(y_{i1},\ldots,y_{iT})^\mathsf{T}$. In the following, let $\mathscr{X}_r$ denote the vector or matrix containing rows of observations $\mathscr{X}_{ri}$. Let $f(\boldsymbol{t})$ denote the vector of function evaluations of f for each entry in the vector $\boldsymbol{t}$ and let $f(x,\boldsymbol{t})$ denote the vector of evaluations of f for each combination of rows in the vectors or matrices $x,\boldsymbol{t}$. Let $1_d=(1,\ldots,1)^\mathsf{T}$ denote a d-vector of ones. The row tensor product of an $m \times a$ matrix $\boldsymbol{A}$ and an $m \times b$ matrix $\boldsymbol{B}$ is defined as the $m \times ab$ matrix $\boldsymbol{A} \odot \boldsymbol{B}=\left(\boldsymbol{A}\otimes 1_b^\mathsf{T}\right)\cdot\left(1_a^\mathsf{T}\otimes\boldsymbol{B}\right)$, where $\cdot$ denotes element-wise multiplication.

1.1 Tensor product representation of effects

Each of the R terms in model (2) can be represented as a weighted sum of basis functions defined on the product space of the covariates in $\mathscr{X}_r$ and t, where each marginal basis is associated with a corresponding marginal penalty. A very versatile method to construct basis functions on such a joint space is given by the row tensor product of marginal bases evaluated on $\mathscr{X}_r$ and $\boldsymbol{t}$ (e.g. De Boor, 1978; Wood, 2006, ch. 4.1.8). Specifically, for each of the terms,

$$f_r\left(\underset{nT\times 1}{\mathscr{X}_{\boldsymbol{r}}},\boldsymbol{t}\right)\approx\left(\underset{nT\times K_x}{\boldsymbol{\Phi_{xr}}}\odot\underset{nT\times K_t}{\boldsymbol{\Phi_{tr}}}\right)\underset{K_xK_t\times 1}{\boldsymbol{\theta}_r}=\boldsymbol{\Phi}_r\boldsymbol{\theta}_r. \quad (3)$$

$\boldsymbol{\Phi}_{xr}$ contains the evaluations of a suitable marginal basis for the covariate(s) in $\mathscr{X}_r$ and $\boldsymbol{\Phi}_{tr}$ contains the evaluations of a marginal basis in $\boldsymbol{t}$. The shape of the function is determined by the vector of coefficients $\boldsymbol{\theta}_r$. A corresponding penalty term can be defined by the Kronecker sum of the marginal penalty matrices $\boldsymbol{P}_{xr}$ and $\boldsymbol{P}_{tr}$ associated with each basis (Wood, 2006, ch. 4.1), i.e.

$$pen\left(\boldsymbol{\theta}_r|\lambda_{tr},_{xr}\right)=\boldsymbol{\theta}_r^T\left(\lambda_{xr}\underset{K_x\times K_x}{\boldsymbol{P}_{\boldsymbol{xr}}}\otimes\boldsymbol{I}_{K_t}+\lambda_{tr}\boldsymbol{I}_{K_x}\otimes\underset{K_t\times K_t}{\boldsymbol{P}_{\boldsymbol{tr}}}\right)\boldsymbol{\theta}_r=\boldsymbol{\theta}_r^T\boldsymbol{P}_r\left(\lambda_{tr},\lambda_{xr}\right)\boldsymbol{\theta}_r. \quad (4)$$

$\boldsymbol{P}_{xr}$ and $\boldsymbol{P}_{tr}$ are known and fixed positive (semi-)definite penalty matrices and λ_{tr} and λ_{xr} are positive smoothing parameters controlling the trade-off between goodness of fit and the smoothness of $f_r\left(\mathscr{X}_r,t\right)$ in $\mathscr{X}_r$ and $\boldsymbol{t}$, respectively. This flexible construction is valid for any combination of bases associated with quadratic penalties. Alternative constructions of the joint penalty such as a direct Kronecker product $\lambda_r\left(\boldsymbol{P}_{xr}\otimes\boldsymbol{P}_{tr}\right)$ associated with a single smoothing parameter λ_r are possible, see Wood (2006, ch. 4.1.8) for a discussion. Typically, $\boldsymbol{K}_x$ and $\boldsymbol{K}_t$ vary with $\boldsymbol{r}$ as well, but we drop the additional index for simplicity. In the following paragraphs, we will motivate and define $\boldsymbol{\Phi}_{xr}$, $\boldsymbol{\Phi}_{tr}$,$\boldsymbol{P}_{tr}$ and $\boldsymbol{P}_{xr}$ for the different types of terms available in our implementation. Effects that are constant over t are associated with $\boldsymbol{\Phi}_{tr}=1_{nT}$ and $\boldsymbol{P}_{tr}=0$, while users are free to choose any suitable marginal basis matrix $\boldsymbol{\Phi}_{tr}$ and penalty $\boldsymbol{P}_{tr}$ for terms that vary over t.

Spline basis representation of effects of scalar covariates—For scalar covariates, index-varying effects are very similar to varying coefficient terms in models for scalar responses, c.f. Ivanescu et al. (2012). For the functional intercept $\boldsymbol{\alpha}(t)$, $\boldsymbol{\Phi}_{xr}=1_{nT}$ and $\boldsymbol{P}_{xr}=0$. For effects like $z\boldsymbol{\delta}$ and $z\boldsymbol{\delta}(t)$ that are linear in a scalar covariate z, the marginal basis for the covariate direction reduces to $\Phi_{xr}=\boldsymbol{z}\otimes 1_T$ where $\boldsymbol{z}=(z_1,\ldots,z_n)^{\mathsf{T}}$, with penalty $\boldsymbol{P}_{xr}=0$. For nonlinear effects of scalar covariates like $\boldsymbol{\gamma}(z)$ or $\boldsymbol{\gamma}(z,t)$, $\boldsymbol{\Phi}_{xr}$ is a suitable marginal spline basis matrix over z and $\boldsymbol{P}_{xr}$ is the associated penalty.

Spline basis representation of functional effects—For linear effects of functional covariates $x(s)$, we model $\boldsymbol{\beta}(s,t)$ using tensor product splines with basis functions $\boldsymbol{\Phi}_{ks}(s),k_s=1,\ldots,\boldsymbol{K}_x$, over $\mathscr{S}$ and a spline basis defined over $\mathscr{T}$. We approximate the integral by numerical integration on the grid defined by the observation points $s_1,\ldots,s_H$ in $\mathscr{S}$. The effect in (2) then is

$$\int_{\mathscr{S}}x_i\left(s\right)\beta\left(s,t_l\right)ds\approx\sum_{h=1}^{H}w_hx_i\left(s_h\right)\sum_{k_s=1}^{K_x}\sum_{k_t=1}^{K_t}\Phi_{k_s}\left(s_h\right)\Phi_{k_t}\left(t_l\right)\boldsymbol{\theta}_{r,k_s,k_t}.$$

In the notation of (3),

$$\Phi_{\boldsymbol{xr}}=\left[\boldsymbol{x}\,diag\left(\boldsymbol{w}\right)\Phi_s\right]\otimes 1_T\approx\left[\int_{\mathscr{S}}x_j\left(s\right)\Phi_{k_s}\left(s\right)ds\right]_{\substack{i=1,\ldots,n\\k_s=1,\ldots,K_x}}\otimes 1_T,\ \text{where}$$

$\boldsymbol{w}=(w_1,\ldots,w_H)^{\mathsf{T}}$ contains the quadrature weights for a numerical integration scheme,

$\boldsymbol{x}=[x_i(s_h)]_{i=1,\ldots,n;\ h=1,\ldots,H}$ and $\Phi_s=[\Phi_{k_s}(s_h)]_{h=1,\ldots,H;\ k_s=1,\ldots,K_x}$. $\boldsymbol{P}_{xr}$ in the tensor product penalty (4) is the penalty associated with the $\boldsymbol{\Phi}_{ks}(s)$. We can extend this construction, which is equivalent to the one introduced in Ivanescu et al. (2012), to cover terms like $\int_{l_i(t)}^{u_i(t)} x_i(s)\beta(s,t)\,ds$ (e.g. Malfait and Ramsay, 2003) with fixed, potentially observation-specific integration limits $l_i(t), u_i(t) \in \mathscr{S}$. This is achieved by defining suitable weight matrices $w_{i,l}$ with zero entries for $s_h < l_i(t_l)$ and $s_h > u_i(t_l)$. Such effects will often be required for covariates and responses that are observed on the same time domain, where responses cannot be influenced by future covariate values. In the limit, this also includes the concurrent model with terms $\boldsymbol{X}(t)\boldsymbol{\beta}(t)$.

Our framework also extends to non-linear function-on-function effects $\int_{\mathscr{S}} F(x_i(s), s, t)\,ds$, which generalize the functional generalized additive model (McLean et al., 2012) from scalar to functional responses. They offer similar flexibility to purely nonparametric approaches like Febrero-Bande and Oviedo de la Fuente (2012); Ferraty et al. (2011); Ferraty and Vieu (2006), e.g. In our framework, such terms can be represented as

$$\int_{\mathscr{S}} F(x_i(s), s, t_l)\,ds \approx \sum_{h=1}^{H} w_h \sum_{k_s=1}^{K_x} \sum_{k_t=t}^{K_t} \Phi_{k_s}(x_i(s_h), s_h)\,\Phi_{k_t}(t_l)\,\boldsymbol{\theta}_{k_s k_t}$$

with $\Phi_{\boldsymbol{xr}}=\left[\left(\boldsymbol{w}^{\mathsf{T}} \otimes \boldsymbol{I_n}\right)\Phi_{\boldsymbol{s}}\right]\otimes 1_T$ and $\Phi_{\boldsymbol{s}}=[\Phi_{k_s}(x_i(s_h), s_h)]_{i=1,\ldots,n,\ h=1,\ldots,H;\ k_s=1,\ldots K_x}$, and $\boldsymbol{P}_{xr}$ the penalty associated with $\boldsymbol{\Phi}_s$. Basis functions $\boldsymbol{\Phi}_{ks}(x(s),s)$ can be tensor product basis functions derived from marginal bases for $x(s)$ and s or true bivariate basis functions.

FPC basis representation of functional effects—Consider a functional covariate expanded in the Karhunen-Loève expansion $x_i(s)=\sum_k \psi_k(s)\xi_{ik}$ with $\int \psi_k(s)\psi_{k'}(s)\,ds=\delta_{kk'}$; $\mathrm{E}(\xi_{ik})=0$; $Var(\xi_{jk})=\zeta_k$. Under the assumption that $\int_{\mathscr{S}} \sum_{k>K_x} \psi_k(s)\xi_{ik}\beta(s,t)\,ds \approx 0$ for some $\boldsymbol{K}_x$, i.e., that all smaller modes of variation of $x(s)$ only have a negligible effect on $y(t)$, we can write

$$\int_{\mathscr{S}} x_i(s)\beta(s,t)\,ds \approx \int_{\mathscr{S}} \sum_{k=1}^{K_x} \boldsymbol{\psi}_k(s)\xi_{ik}\beta(s,t)\,ds = \sum_{k=1}^{K_x} \xi_{ik}\tilde{\beta}_k(t)$$

with $\tilde{\beta}_k(t)=\int_{\mathscr{S}} \psi_k(s)\beta(s,t)\,ds$. Thus, a linear function-on-function effect can be represented as a sum of varying coefficient terms for the FPC loadings $\boldsymbol{\xi}_{ik}$. This representation extends FPC regression approaches (e.g. Reiss and Ogden, 2007) from scalar to functional responses. In the notation of the general framework, $\Phi_{\boldsymbol{xr}}=[\xi_{ik}]_{i=1,\ldots,n;\ k=1,\ldots,K_x}\otimes 1_T$ and $\boldsymbol{P}_{xr}$ = 0. For the t-direction, $\boldsymbol{\Phi}_{tr}$ and the associated $\boldsymbol{P}_{tr}$

can be chosen freely. An implicit assumption here is that all $\boldsymbol{\beta}_k(t), k = 1,\ldots,\boldsymbol{K}_x$ have similar smoothness, as they are all associated with the same smoothing parameter.

This FPC-based approach may be advantageous for functional covariates observed on irregular or sparse grids – for such data, the spline-based method requires a preprocessing step (c.f. Goldsmith et al., 2011, 2012; James, 2002) to impute the incomplete trajectories on a dense and regular grid, whereas FPCs can be estimated directly from sparse data (Yao et al., 2005). Additionally, if the shapes of the $\hat{\psi}_k(s)$ are meaningful to practitioners and $\boldsymbol{K}_x$ is small, the coefficient functions $\tilde{\beta}_k(t)$ may be easier to interpret than a coefficient surface $\boldsymbol{\beta}(s,t)$. On the other hand, since inference is performed *conditional* on the estimated FPCs $\hat{\psi}_k(s)$ and associated loadings $\hat{\xi}_{ik}$ coverage issues associated with these neglected sources of estimation variability (c.f. Goldsmith et al., 2013) and bias introduced by estimation error in the FPC analysis step may occur. Additionally, $\int_{\mathcal{S}} \sum_{k>K_x} \psi_k(s)\,\xi_{ik}\beta(s,t)\,ds \approx 0$ might be a strong assumption that is hard to check in applications, as is the choice of the discrete tuning parameter $\boldsymbol{K}_x$.

As in the spline-based case, we can extend this to FPC-based nonlinear function-on-function effects. The proposal corresponds to an extension of the functional additive model by Müller and Yao (2008) from scalar to functional responses, with the effect of the functional covariate given by $f_r(x_i(s),t) = \sum_{k=1}^{K_x'} F_k\left(\hat{\xi}_{ik}, t\right)$. In the notation of our general framework, $\Phi_{\boldsymbol{xr}} = \left[\Phi_{\xi_1}|\ldots|\Phi_{\xi_{K_x'}}\right]$ where $\Phi_{\xi_k} = \left[\Phi_a\left(\hat{\xi}_{ik}\right)\right]_{\substack{i=1,\ldots,n \\ a=1,\ldots,A}} \otimes 1_T$ for suitable spline basis functions $\boldsymbol{\Phi}_a(\cdot)$, such that $K_x = AK_x'$. The marginal penalty is given by $\boldsymbol{P}_{\boldsymbol{xr}} = \boldsymbol{I}_{K_x'} \otimes \boldsymbol{P}_\xi$, where $\boldsymbol{P}_{\boldsymbol{\xi}}$ is the penalty associated with the $\Phi_a\left(\hat{\xi}_{jk}\right)$. In t-direction, we are again free to choose any suitable basis $\boldsymbol{\Phi}_{tr}$ and penalty $\boldsymbol{P}_{tr}$. Further extensions to interaction effects of estimated FPC scores $f_r(x_i(s),t) = \sum_{k=1}^{K_x'} \sum_{k<k'\leq K_x'} F_{k,k'}\left(\hat{\xi}_{ik}, \hat{\xi}_{ik'}, t\right)$ are also obvious in this framework.

Spline basis representation of functional random effects—Functional random effects $b_g(t)$ are represented as smooth functions in t for each level $1,\ldots,M$ of the grouping variable g. In the notation of equation (3), functional random intercepts are associated with a marginal basis $\Phi_{\boldsymbol{xr}} = \left[\delta_{g(i)m}\right]_{\substack{i=1,\ldots,n \\ m=1,\ldots,M}} \otimes 1_T$, where $g(i)$ denotes the level of g for observation i. This yields an incidence matrix mapping the observations to the different levels of the grouping variable. For a functional random slope effect in a scalar covariate z, $\Phi_{\boldsymbol{xr}} = \left[z_i\delta_{g(i)m}\right]_{\substack{i=1,\ldots,n \\ m=1,\ldots,M}} \otimes 1_T$. In the notation of equation (4), the marginal penalty $\boldsymbol{P}_{xr}$ for functional random effects is a $M \times M$ precision matrix that defines the dependence

structure between the levels of g. The quadratic penalty (4) is mathematically equivalent to the distributional assumption $\theta_r \sim N\left(0, \sigma_\varepsilon^{-2}\boldsymbol{P}_r(\lambda_{tr}, \lambda_{xr})^{-1}\right)$ (Brum-back et al., 1999). Through the representation in (3), this induces a mean zero Gaussian process assumption $b_g(t) \sim GP(0, K^b(g(i), g(i'), t, t'))$, with covariance evaluated for all nT observations $K^b(g, g, t, t) = \sigma_\varepsilon^{-2}\Phi_r\boldsymbol{P}_r(\lambda_{tr}, \lambda_{xr})^{-1}\Phi_r^\top$. The smoothing parameter $\boldsymbol{\lambda}_{xr}$ controls the relative contribution of the inter-unit variability relative to the common roughness of the functional random effects controlled by $\boldsymbol{\lambda}_{tr}$.

If observations on different levels of the grouping factor are assumed independent, $\boldsymbol{P}_{xr} = \boldsymbol{I}_M$ is simply the identity matrix. More generally, $\boldsymbol{P}_{xr}$ can represent any fixed dependence structure between levels of g: It can be a (partially improper) precision matrix of a random field with known correlation structure, implied, for example, by the spatial or temporal arrangement of the different levels of g, such as a Gaussian Markov random field (GMRF) on geographical regions for conditionally auto-regressive (CAR) model terms. Alternatively, $(\boldsymbol{P}_{xr})^{-1}$ can be defined using any valid correlation function based on – for example – spatial, temporal, or genetic distances between levels of g. If the grouping variable is simply the index of observations (i.e., $g(i) = i$), this construction yields smooth residual curves with potential for spatial or temporal autocorrelation. This innovative definition of functional random effects admits very flexible model specifications, since any combination of spline basis, smoothness penalty and between-subject correlation can be used for functional random effects. This allows, for example, for spatially correlated functional residuals with periodicity constraints for the Canadian Weather data (see Appendix C of the online supplement). Multiple (partially) nested or crossed random effects can be constructed in this way and are implemented in pffr().

FPC basis representation of functional random intercepts—For functional random intercepts without between-unit correlation, i.e., for $b_g(t) \overset{\text{i.i.d.}}{\sim} GP\left(0, K^b\left(t, t'\right)\right)$, it can be advantageous to use the eigenfunctions of the covariance operator $\boldsymbol{K}^b(t,t')$ as basis functions in t. Specifically, we use the Karhunen-Loève expansion of random processes to represent $b_g(t) \approx \sum_{k=1}^{K_t} \eta_k(t)\nu_{gk}$ with $\boldsymbol{\kappa}_k, \boldsymbol{\eta}_k(t)$ the ordered eigenvalues and -functions of $\boldsymbol{K}^b(t,t')$, $\boldsymbol{\nu}_{gk}$ the associated FPC loadings, and $\boldsymbol{K}_t$ a suitable truncation lag. The marginal basis for the t-direction is then $\Phi_{\boldsymbol{tr}} = 1_{\boldsymbol{n}} \otimes \left[\hat{\eta}_1(t) \,|\, \cdots \,|\, \eta_{K_t}(t)\right]$. Since $\mathrm{E}(\boldsymbol{\nu}_{gk}) = 0$ and $\mathrm{Var}(\boldsymbol{\nu}_{gk}) = \boldsymbol{\kappa}_k$, a reasonable marginal penalty is $\boldsymbol{P}_{\boldsymbol{tr}} = diag\left(\hat{\kappa}_1, \ldots, \hat{\kappa}_{K_t}\right)^{-1}$. This encourages relative contributions of the FPCs to the random effect curves that are roughly proportional to their estimated magnitudes $\hat{\kappa}_k$. As for the spline-based functional random effects, $\Phi_{xr} = \left[\delta_{g(i)m}\right]_{\substack{i=1,\ldots,m \\ m=1,\ldots,M}} \otimes 1_T$ is an incidence matrix for the group levels, while $\boldsymbol{P}_{xr} = \boldsymbol{I}_M$.

In practice, $\boldsymbol{\eta}_k(t)$ and $\boldsymbol{\kappa}_k$ have to be estimated. An iterative procedure can be outlined as follows: (1) Use a fit without functional random effects under independence assumption to

$\boldsymbol{E}=[\varepsilon_{il}]_{\substack{i=1,\ldots,n\\ l=1,\ldots,T}}$ obtain working residuals. (2) Compute the group-level means of residual curves $\bar{\boldsymbol{E}}=\Delta^{\mathsf{T}}\boldsymbol{E}$ with $\Delta=\left[\delta_{g(i)m}/n_m\right]_{\substack{i=1,\ldots,n\\ m=1,\ldots,M}}$, where n_m is the number of observations for the m-th level of g. (3) Perform a (truncated) spectral decomposition of $\hat{\boldsymbol{K}}^b=\left[\hat{K}^b(t_l,t_{l'})\right]_{l,l'=1,\ldots,T}$ to obtain $\hat{\eta}_k(t)$, $\hat{\kappa}_k$ for $k = 1, ..., K_t$. A suitable estimate for $\hat{\boldsymbol{K}}^b$ can be derived from smoothing the entries in the matrix $M^{-1}\bar{\boldsymbol{E}}^{\mathsf{T}}\bar{\boldsymbol{E}}$ (without the diagonal) as in Yao et al. (2005). This approach for estimating $\hat{\boldsymbol{K}}^b$ can only be used for random intercepts for a single grouping variable. Compared to spline-based functional random effects, FPC-based modeling holds the promise of using the optimal, most parsimonious basis to represent $b_g(t)$. Computationally, it is expected to scale much better for large M, as the number of coefficients associated with a functional random effect is MK_t and K_t for FPCs will typically be much smaller than in a sufficiently flexible spline basis. On the other hand, the FPC approach requires a pilot estimate for $\boldsymbol{K}^b$. The subsequent performance will be sensitive to the quality of the estimation of the FPCs and to the choice of K_t.

1.2 Mixed model representation

Using the tensor product representation introduced in the previous subsection for all terms, model (1) can be re-written as

$$\boldsymbol{y}=\boldsymbol{\Phi\theta}+\epsilon;\qquad \epsilon\sim\mathbf{N}\left(\mathbf{0},\boldsymbol{\sigma}_{\varepsilon}^{2}\boldsymbol{I}_{\mathbf{nT}}\right), \quad (5)$$

where $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1|...|\boldsymbol{\Phi}_R]$ contains the concatenated $\boldsymbol{\Phi}_r$ associated with the different model terms and $\theta=\left(\theta_1^{\mathsf{T}},\ldots,\theta_R^{\mathsf{T}}\right)^{\mathsf{T}}$ the respective stacked coefficient vectors $\boldsymbol{\theta}_r$. To clear up notation, we assign a sequential index $v = 1,...,V$ to the smoothing parameters $\boldsymbol{\lambda_{xr}}$,$\boldsymbol{\lambda_{tr}}$ in (4), where V is the total number of smoothing parameters in the model. We pad $\boldsymbol{P_{tr}}\otimes\boldsymbol{I}_{K_x}$ and $\boldsymbol{I}_{K_t}\otimes\boldsymbol{P_{xr}}$ with rows and columns of zeros, denoting these matrices $\tilde{\boldsymbol{P}}_{v1}$ and $\tilde{\boldsymbol{P}}_{v2}$ such that the penalty $\theta_r^T\left(\lambda_{tr}\boldsymbol{P}_{tr}\otimes\boldsymbol{I}_{K_x}+\lambda_{xr}\boldsymbol{I}_{K'}\otimes\boldsymbol{P}_{xr}\right)\theta_r=\lambda_{v_1}\theta^T\tilde{\boldsymbol{P}}_{v_1}\theta+\lambda_{v2}\theta^T\tilde{\boldsymbol{P}}_{v_2}\theta$ refers to the full coefficient vector $\boldsymbol{\theta}$. The penalized likelihood criterion to be minimized then becomes

$$\frac{1}{\boldsymbol{\sigma}_{\varepsilon}^{2}}\|\boldsymbol{y}-\boldsymbol{\Phi\theta}\|^2+\sum_{v=1}^{V}\frac{\lambda_v}{\boldsymbol{\sigma}_{\varepsilon}^{2}}\boldsymbol{\theta}^T\tilde{\boldsymbol{P}}_v\boldsymbol{\theta}. \quad (6)$$

The total number of smoothing parameters is $V \leq 2R$, as some terms are constant over t or $\mathscr{X}_r$ and the corresponding $\tilde{\boldsymbol{P}}_v$ are zero. Let $\tau_v=\frac{\sigma_\varepsilon^2}{\lambda_v}$ and use similar arguments as in Ruppert et al. (2003, ch. 4.9) to obtain the solution $\hat{\theta}$ of (6) as the best linear unbiased predictor in the linear mixed effects model (MEM)

$$\boldsymbol{y} \sim N\left(\boldsymbol{\Phi\theta}, \boldsymbol{\sigma}_{\varepsilon}^{2}\boldsymbol{I}_{\mathbf{nT}}\right); \qquad \boldsymbol{\theta} \sim \mathbf{N}\left(\mathbf{0}, \left(\sum_{\mathbf{v}=\mathbf{1}}^{\mathbf{V}} \boldsymbol{\tau}_{\mathbf{v}}^{-1}\tilde{\boldsymbol{P}}_{\mathbf{v}}\right)^{-}\right), \quad (7)$$

where S^- denotes the generalized inverse of S, and $N(0,S^-)$ is a partially improper Gaussian distribution with positive semi-definite covariance matrix S. The impropriety results from rank deficiencies in some of the $\tilde{\boldsymbol{P}}_v$, since roughness penalties typically define a nullspace of maximally smooth functions. Numerical difficulties posed by the positive semi-definiteness are solved by another re-parameterization that separates the various model terms into their unpenalized and penalized components, i.e. into "fixed" effects and "random" effects with a proper distribution, respectively. These are well known issues in the literature on penalized regression splines described in detail e.g. in Ruppert et al. (2003, ch. 4.9),Wood (2006, ch. 6.6.1); recent developments for tensor product splines are in Wood et al. (2013).

One of the main advantages of formulating the penalized likelihood optimization as estimation in an MEM is that the smoothing parameters $\lambda_v = \frac{\sigma_\varepsilon^2}{\tau_v}$ can be treated as variance component parameters and thus can be estimated using restricted maximum likelihood (REML). In particular, Reiss and Ogden (2009) and Wood (2011) have shown that smoothing parameter selection with REML is more stable and results in somewhat lower MSE than generalized cross-validation (GCV) and Krivobokova and Kauermann (2007) have shown that REML estimation of penalized splines is more robust to error correlation mis-specification than AIC-based criteria. A second advantage this approach offers is that the representation of our model class (1) results in a fit criterion (7) equivalent to that of conventional additive mixed models for scalar data. This means much of the powerful and versatile inference machinery developed for scalar linear and additive mixed models (AMMs) over the last years can be applied directly to the proposed model class of functional AMMs, due to their close structural similarity. Specifically, 1) pointwise, bias-corrected confidence bands (Marra and Wood, 2012; Nychka, 1988; Ruppert et al., 2003) are available for the functional effects, 2) tests for random effects as well as tests for constant or linear effects versus more general alternatives developed for scalar responses (Crainiceanu and Ruppert, 2004; Crainiceanu et al., 2005; Greven et al., 2008; Scheipl et al., 2008; Wood, 2013), and 3) model selection approaches that have recently been proposed for scalar-response AMMs (Greven and Kneib, 2010; Marra and Wood, 2011) are directly applicable to the proposed model class. Finally, the proposed approach accommodates a large variety of effects, at no increase in the level of complexity of the algorithm itself. The tensor product representation given in Section 1.1 combined with the MEM representation (7) allows for a unified framework for smoothness parameter selection and estimation of all model components in model (1), including functional random effects and FPC-based effects.

Implementation

The full framework for functional additive models we describe here is implemented in the pffr-function in the refund package for R. The underlying inference engine is the mgcv package (Wood, 2011) for generalized additive models which also supplies most of the functionality for constructing basis and penalty matrices. pffr offers a formula-based interface similar to the established formula syntax of mgcv and returns a rich model object

whose fit can be summarized, plotted and compared with other model formulations without any programming effort by the user through convenient utility functions.

2 Empirical evaluation

The following section describes an extensive simulation study and results for the motivating application to a longitudinal DTI study. A fully reproducible example analysis of the well known Canadian Weather data showcasing the flexibility of pffr can be found in Section C of the online appendix.

2.1 Simulation study

Simulation setup—We simulate data with repeated measures structure for a model $y_{ij}(t) = \sum_{r=1}^{R} f_r(\mathcal{X}_{rij}, t) + \varepsilon_{ij}(t)$ for the following four scenarios to investigate the sensitivity of the estimates to varying model complexity, noise levels and number of observations:

1. Functional random intercept, functional random slope:
   $\sum f_r(\mathcal{X}_{rij}, t) = \alpha(t) + b_{i0}(t) + b_{i1}(t) u_{ij}$
2. Functional random intercept, functional covariate:
   $\sum f_r(\mathcal{X}_{rij}, t) = \alpha(t) + \int x_{1,ij}(s) \beta_1(s,t) ds + b_{i0}(t)$
3. Functional random intercept, two functional covariates:
   $\sum f_r(\mathcal{X}_{rij}, t) = \alpha(t) + \int x_{1,ij}(s) \beta_1(s,t) ds + \int x_{2,ij}(s) \beta_2(s,t) ds + b_{i0}(t)$
4. Functional random intercept, functional covariate, smooth scalar covariate effect, varying coefficient term:
   $\sum f_r(\mathcal{X}_{rij}, t)) = \alpha(t) + \int x_{1,ij}(s) \beta_1(s,t) ds + \gamma_1(z_{1,ij}, t) + \delta_2(t) z_{2,ij} + b_{i0}(t)$

Definitions of the various effect functions and descriptions of the data generating processes used for the covariates can be found in section B of the online supplement, along with unabridged simulation results and graphical displays of data and estimated effects for the replications with minimal, maximal and median error for each scenario.

For each of the four scenarios, we run 10 replications for each combination of the following settings, yielding 1920 model fits in total:

– number of subjects: $M \in \{10,100\}$

– mean number of observations per subject: $n_i \in \{3,20\}$. Subject labels $i \in \{1,...,M\}$ are drawn from a multinomial distribution with probabilities $P\left(i=i'\right) \propto \sqrt{i'}$ to generate unbalanced designs.

– number of grid points for t: $T \in \{30,60\}$

– relative importance of random effects: $\text{SNR}_B \in \{0.2,1,5\}$, where SNR_B is the ratio of the standard deviation of the additive predictor without random effects divided by the standard deviation of the random effect functions; e.g. for $\text{SNR}_B = 5$, the contribution of

each functional random effect to the variability in $y(t)$ is about 5 times smaller than that of the non-random effects.

– signal-to-noise ratio: $\text{SNR}_\varepsilon \in \{1,5\}$, where SNR_ε is the ratio of the standard deviation of the additive predictor divided by the standard deviation of the residuals σ_ε.

Our results show that 10 replications for each combination are sufficient to derive precise estimates of effects of the setting parameters on estimation errors and computation times, c.f. Figures 1 and 2. Fits are obtained with the defaults in pffr(), i.e., cubic B-spline bases with 20 basis functions and first order difference penalty for the functional intercept, tensor products of cubic B-spline bases with five marginal basis functions for the tensor product terms, with first order difference penalties for the *t*- and *s*-directions and second order difference penalties for the covariate direction (if applicable). The smoothing parameters are REML estimates as returned by mgcv. The models for settings 1 to 4 include $K = 120$ to 1020 coefficients and 5 to 8 smoothing parameters.

Estimation error—We use the relative integrated mean squared error defined as

$$rIMSE\left(\hat{f}_r(\mathcal{X}_r,t)\right) = n^{-1}\sum_{i=1}^{n}\frac{\int\left(\hat{f}_r(\mathcal{X}_{ri},t) - f_r(\mathcal{X}_{ri},t)\right)^2 dt}{\int f_r(\mathcal{X}_{ri},t)^2 dt}$$

to evaluate the accuracy of the estimates. Relative errors allow comparisons across different scenarios and noise levels regardless of the ranges of the true $f_r(\mathcal{X}_r,t)$. Note that we evaluate the estimation accuracy of the effects on the scale of the response, not on the scale of the coefficient function itself to make errors directly comparable across effects. Detailed analysis of results (see Appendix B in the supplement) shows that there are no relevant interaction effects between the setting parameters M, n_i, T, SNR_B and $\mathbf{SNR}_{\boldsymbol{\varepsilon}}$ on the observed errors within scenarios, so we fit log-linear models with main effects for the setting parameters in each scenario to observed r**IMSE** values and proceed to analyze the estimated effects. Figure 1 shows baseline levels and the estimated multiplicative effects of the simulation parameters on the r**IMSE**s. The effect of increasing the number of grid points T for $y(t)$ from 30 to 60 is not shown, as it decreased relative errors for all quantities by a factor of about 0.7 to 0.5. Baseline r**IMSE** values (top left panel) are given for data with $\mathbf{SNR}_{\boldsymbol{\varepsilon}} = 1$, $M = 10$, $n_i = 3$, $T = 30$, $\mathbf{SNR}_{\boldsymbol{B}} = 0.2$. In this very noisy setting with small sample size and dominant random effects, covariate effect estimates are not very accurate, with relative errors mostly in the vicinity of one. Since the random effects are estimated with little error, however, the error for the responses in this difficult setting is small as well. Increasing $\mathbf{SNR}_{\boldsymbol{\varepsilon}}$ from 1 to 5 (top right panel) decreases relative errors about 16-fold, with smaller 8-fold reductions for the random effects in scenario 1. Increasing the number of groups from $\boldsymbol{M} = 10$ to $\boldsymbol{M} = 100$ (second row, left panel) has no substantial effect on the overall estimation accuracy of the $y_i(t)$. Estimation accuracy of the functional random effects is not improved either due to the commensurate increase in the number of parameters, while errors for the covariate effects decrease about 8-fold. An increase in the average number of observations per group from $n_i = 3$ to $n_i = 20$ (second row, right panel) results in a similar reduction of relative errors for the covariate effects, and also a marked four- to sixfold decrease in the errors for the response trajectories. A reduction of the relative contribution of the random effects to the linear predictor, i.e. increasing $\mathbf{SNR}_B$ from 0.2 to 1 [5] (bottom row, left [right] panel), improves

the overall estimation accuracy of $y(t)$ only slightly if at all [factor 0.7 to 0.8]. This overall improvement is due to the large reduction of errors for the covariate effects, which compensates for the observed deterioration of random effect estimates. While the errors for the former decrease about 8-fold [16-fold], the errors for the latter increase about 1.5- to twofold [five- to 16-fold].

Comparison to other approaches—Appendix C summarizes additional results for comparisons between spline-based and FPC-based terms implemented for function-on-function effects and functional random effects in pffr as well as the wavelet-based approach for functional linear mixed models of Morris and Carroll (2006) implemented in WFMM (Herrick, 2013).

Coverage—We also evaluate coverage of approximate point-wise empirical Bayes confidence intervals (CIs) (c.f. Wood, 2006, eq. (4.35)) with constraint correction (Marra and Wood, 2012) for a nominal level of 95%. For each fitted model, we record the proportion of point-wise intervals covering the true value of each quantity evaluated on a fine grid. Note that the coverages of neighboring grid points are not independent, but for the computationally intensive models we fit this is a feasible alternative to coverage estimates based on hundreds of replicates of each setting. CI coverage was consistently very close to the nominal level for $\hat{y}(t)$ (Median$_{10\%}$ quantile$-$90% quantile : $0.95_{(0.92-0.97)}$) and $\hat{b_0}(t)$ ($0.95_{(0.9-0.98)}$), while $\hat{\alpha}(t)$ $\left(0.97_{(0.9-1)}\right)$ showed some overcoverage as well as a few replicates with coverage below 0.7 for small and noisy data. Coverage for functional random slopes $\hat{b_1}(t)$ was below nominal for small groups, but close to nominal for larger datasets (n_i = 3: $0.9_{(0.81-0.96)}$; n_i = 20: $0.95_{(0.9-0.99)}$). Similarly for $\hat{\beta}_1(s,t)$ and $\hat{\beta}_2(s,t)$, overall coverage was close to the nominal level ($0.95_{(0.85-0.99)}$), with systematic undercoverage in small and noisy data sets with dominating random effects. Both $\hat{\gamma}_1(z_1,t)$ $\left(0.99_{(0.94-1)}\right)$ and $\hat{\delta}_2(t)$ $\left(1_{(0.75-1)}\right)$ had overcoverage, the latter with many outliers with observed coverages below 0.8.

Computation times—Figure 2 shows computation times on an 2.2 GHz AMD Opteron 6174 processor for the different scenarios and sample sizes. Especially for models with multiple random effects (scenario 1) computation times increase dramatically in $\boldsymbol{M}$. Smaller models are fit rapidly, and even for the largest data sets with $nT = 1.2 \cdot 10^5$, computation times are not prohibitively long. Speed gains for REML inference on large data sets can be achieved by using the pffr()-option to use mgcv's bam() routine for estimating additive models on data sets that do not fit into memory, as in Section 2.2. Using GCV optimization (Wood, 2004) instead of REML-based inference in pffr() can yield up to 10-fold speedups especially for large data sets, but tends to be less stable.

Summary—Important effects that contribute relevantly to the predictor are estimated with good to excellent accuracy. Only a single replicate resulted in an r**IMSE** for $y(t)$ greater than 0.1 – even in the most challenging data situations with few noisy observations and small group sizes, our approach is able to reproduce the true structure of the data well. Our results

indicate that estimation accuracy of covariate effects is affected most strongly by changes in the noise levels $\mathbf{SNR}_{\boldsymbol{\varepsilon}}$ and especially $\mathbf{SNR}_B$, and less strongly by changes in the available number of observations $\boldsymbol{M}$,n_i and T. The patterns of relative change in accuracy are identical for simple functional regression coefficients, index-varying smooth effects or effect surfaces for functional covariates. The estimation accuracy of the functional random effects is affected strongly by the relative importance of the random effects $\mathbf{SNR}_B$ and the group size n_i, and little by the number of groups $\boldsymbol{M}$. **FPC**-based random effects seem to require a sufficiently large number of groups and low noise level to obtain usable **FPC** estimates. Spline-based approaches yielded superior results to **FPC**-based and wavelet-based approaches, but it should be noted that the data-generating process for the simulation study was spline-based itself. Overall, the observed coverage of the approximate pointwise CI was very close to the nominal level except for very small or noisy data.

2.2 Modeling spatial association of demyelination in a longitudinal MS study

Our motivating tractography study comprises 162 MS patients and 42 healthy controls who are observed at one to eight visits, spread over up to four years, with 476 visits in total. MS damages white matter tracts (WMT) in the brain due to lesions, axonal damage and demyelination. Diffusion tensor imaging (DTI) is a magnetic resonance imaging technique that is able to resolve individual WMTs in the central nervous system (Basser et al., 2000), and is thus a very useful tool in monitoring disease progression in MS patients. At each visit, fractional anisotropy (FA) was determined via DTI along the corpus callosum (CCA, connecting the left and right hemispheres of the brain), the left corticospinal tract (CST, connecting the brain and the spinal cord), and the left optic radiation tract (OPR, connecting visual cortex and thalamus). FA is derived from the estimated diffusion tensor and is equal to zero if water diffuses perfectly isotropically (Brownian motion) and to one if it diffuses with perfectly organized movement of all molecules in one direction for a given voxel. It may be decreased in MS patients and thus serves as a marker of disease progression here. Tracts are registered within and between subjects using standard biological landmarks identified by an experienced neuroradiologist. Figure 3 displays registered tract profiles as functions of tract location; profiles corresponding to four different subjects at first visit are highlighted.

Various aspects of this complex tractography dataset have been explored in a sequence of papers including Goldsmith et al. (2011), Goldsmith et al. (2012), Staicu et al. (2011), Ivanescu et al. (2012). This study was first introduced by Greven et al. (2010), who modeled longitudinal variability in trajectories FPC-based, but could not take into account any covariate effects. Our goal here is to better understand the spatial course of the demyelination process via its FA proxy and investigate possible differences therein between MS and healthy subjects. Ivanescu et al. (2012) considered a similar question, but used only the first measurement of each subject since their approach is unable to handle the longitudinal structure of the data. We assume a functional linear dependence between the FA along the CCA and the two functional covariates – FA along the OPR and FA along the CTS – while adjusting for the effects of other relevant covariates such as gender, age at visit, and disease status. Specifically, if $y_ij(t)$ is the FA profile at location t on the CCA tract observed at visit j for subject i, we assume that the conditional mean of $y_ij(t)$, $\mu_ij(t)$, is

$$\mu_{ij}(t) = \alpha_{d_i}(t) + \delta_{g_i}(t) + \nu(u_{ij}, t) + \int x_{1,ij}(s)\,\beta_{1,d_i}(s,t)\,ds + \int x_{2,ij}(r)\,\beta_{2,d_i}(r,t)\,dr \quad (8)$$

where $x_{1,ij}(s)$ and $x_{2,ij}(r)$ are the FA profiles at locations s and r along the OPR and CTS tracts, respectively, observed at the jth visit of the ith subject. Here d_i is the disease status of the subject, with $d_i = 1$ for MS patients, and $d_i = 0$ for healthy subjects; g_i indicates the gender: $g_i = 1$ for males and 0 for females; and u_ij is the age (in years) at the jth visit of the ith subject. Note that the effects of FA-OPR and FA-CST at the current visit are disease group-specific, with $\beta_{1,0}(s,t), \beta_{2,0}(r,t)$ for controls and $\beta_{1,1}(s,t), \beta_{2,1}(r,t)$ for MS patients. Neither age nor gender effects were found to differ between disease groups in the model-building process.

Effect estimates for a naïve model (8) along the lines of Ivanescu et al. (2012) under assumed independence with measures of uncertainty are provided in Appendix B, Figure 7 for completeness. Due to the inappropriate conditional independence assumption, this approach underestimates the variability of the estimates. We use our proposed functional additive mixed model to account for the within-subject correlation, which is the key advantage of our approach over available function-on-function regression methods. Specifically, a more appropriate model is

$$y_{ij}(t) = \mu_{ij}(t) + b_{i0}(t) + \varepsilon_{ijt}, \quad (9)$$

where $b_{i0}(t)$ are subject-specific functional random intercepts. Model (9) can be fit using the pffr() function in the refund package. Estimating (9) took about 17 hours on an 2 GHz AMD Opteron processor.

Since there are subjects with a few missing locations along the tracts and since the FA measurements are observed with noise, we preprocess the functional covariates. (Note that missing values in the functional response are not an issue for our approach.) The FA profiles are first detrended by subtracting the disease group-specific mean function to make the estimated effects easily interpretable (see Appendix A.1). They are then smoothed, which also imputes missing values. For smoothing, we use functional principal component analysis (Di et al., 2009; Yao et al., 2005) for all tract-specific FA curves under a working independence assumption between profiles on the same subject. Since the observed FA-CCA profiles exhibit a lot of small scale structure at locations $5 - 20$ and > 85, spline based functional random intercepts would require a very large basis to provide sufficient flexibility. Instead, we use the residual curves from model (8) fitted under an independence assumption to obtain an unsmoothed FPC-basis for the random intercepts, as described on page 10.

Figure 4 shows the estimated mean of the FA profiles along the CCA tract (anterior to posterior, i.e. front of the head to the back) for female subjects with and without MS (left panel) and for male and female MS patients (second from left). The estimated mean FA profiles have similar shapes, with a sharp increase in the rostrum/genu (front), a plateau in the middle section, followed by a decrease near the isthmus and a rapid increase towards the splenium (back). As expected, MS patients tend to have lower FA-CCA, especially in the posterior section from the rostral body to the splenium. The effect of gender seems to be

negligible. The estimated age effect, $\hat{\nu}(u_{ij}, t)$, indicates that FA-CCA decreases almost linearly with age over the entire tract, particularly in the anterior part, but this effect is fairly small. Not accounting for the longitudinal data structure (c.f. Figure 7), differences between MS and healthy subjects would be found to be much larger and statistically significant along the entire tract. The corresponding estimate for the age effect seems implausible. Due to the misplaced independence assumption, the variability of the estimates shown in Figure 7 is underestimated, but should be approximately correct in Figures 4 and 5. The rightmost panels in Figure 8 give covariances and correlations for $\varepsilon_{ij}(t)$ and show that the white-noise-error assumption is reasonable for model (9), but severely violated for (8). They also show that spline-based random intercepts are less successful in removing all structure from the residuals in this case, especially in the rostrum/genu.

In healthy controls, FA values at the ends of the OPR tract (towards lateral geniculate nucleus and visual cortex, respectively) and in its middle section show a positive association with FA values along the entire CCA tract (see Figure 5). For the CTS tract, there is some indication of a positive association between FA values in the beginning of the CTS tract (medulla) and the end of the CCA (splenium) and between the end of the CTS tract (subcortical white matter) and the beginning of the CCA (rostrum/genu), the latter corresponding with spatial proximity. These patterns should be indicative of the normal ageing process, while the observed associations mostly vanish for MS patients or become much weaker. It should be noted, however, that simulation results indicate potentially low estimation accuracy of fixed effects in settings such as this one in which the random effects dominate the predictor. Figure 6 displays the predicted intercept curves $\widehat{b_{i0}(t)}$ (left panel) and observed residuals $\hat{\varepsilon}_{ij}(t) = y_{ij}(t) - \hat{\mu}_{ij}(t) - \hat{b}_{i0}(t)$ (right panel). The large variation in the predicted functional intercepts reveals large inter-subject variability. By accounting for the between-subject variability the observed integrated root mean square error of the responses with the proposed method (0.027) reduces to half of its magnitude compared to Ivanescu et al. (2012) under an independence assumption (0.05). Model (9) explains about 90% of the observed variability, while (8) explains only about 63%.

In conclusion, using our flexible modeling framework for the FA profiles along the CCA tract shows that a large fraction of the variability in the data is captured by subject-specific random effects. Modeling the dependence on FA profiles at other well identified tracts, OPR and CTS, can provide new insights into the spatial association in normal ageing and disease processes. Interestingly, our results indicate that the associations between demyelination along the left CTS and left OPR tracts and the CCA tract are weaker for MS patients than for healthy controls. A possible interpretation of this finding could be that demyelination processes in MS patients are more strongly localized, consistent with the development of localized lesions during MS. By properly accounting for the longitudinal structure of the data the estimation uncertainty of all effects increases compared to model (8) under an independence assumption.

3 Discussion and Outlook

We propose a general framework for flexible additive regression models for correlated functional responses, allowing for multiple functional random effects with, for example,

spatial, temporal, spatio-temporal or longitudinal (Section 2.2) correlation structures. Dependence structures can be modeled either implicitly by specifying smooth temporal, spatial or tempo-spatial effects or explicitly by including functional random effects with marginal between-unit correlation structures given by the precision matrices of Gaussian (Markov) random fields. Estimation and inference is performed by standard additive mixed model software, allowing us to take advantage of established robust and flexible algorithms. The approach is implemented as fully documented open-source software in the pffr()-function in the refund package (Crainiceanu et al., 2011) for R. Effects of functional covariates and functional random effects are available in both FPC- and spline-based variants and both scalar and functional covariates can have linear or more general smooth effects on the outcome trajectories, allowing analysts to choose the most suitable tools for the task at hand.

Simulation experiments show that the proposed method recovers relevant effects reliably and handles small group sizes and/or low numbers of replications well. Data sets of considerable size can be fit in acceptable time. Two applications demonstrate that our approach makes it possible to fit flexible models that do justice to complex data situations and yet still yield interpretable results that can help to understand the underlying processes.

This work opens up a number of interesting avenues for further research. A first direction concerns the covariance structure of the residuals. Since our present inference algorithms do not exploit the extreme sparsity of the design matrices for smooth observation-specific residual terms, estimating such terms dramatically increases computation time and memory requirements for large data sets. On the other hand, simply assuming i.i.d. errors ε_{it} will often be unrealistic since some degree of auto-correlation and heteroscedasticity over the index of the functional response is usually encountered in practice. We are currently investigating an iterative procedure similar to the approach in Reiss et al. (2010), where observed residuals from an initial model estimated under a working independence assumption are used to estimate a working covariance structure and the model is then re-estimated based on de-correlated data. If successful, such a marginal model specification could offer an efficient alternative to the conditional modeling approach outlined in the present paper. In a second direction, we are currently developing diagnostic measures to identify potential problems with low-rank functional covariates (c.f. Appendix A) as well as practical model-building strategies regarding the estimation of corresponding regression surfaces. A closely related avenue of inquiry are more in-depth comparisons of spline- and FPC-based approaches for modeling function-on-function terms as well as functional random effects in order to evaluate their relative strengths and weaknesses. The unifying framework implemented in pffr() will greatly facilitate such comparisons. In addition, we have begun implementing a dedicated toolbox for REML-based inference tailored to function-on-function regression. This effort is based on the computationally efficient array regression approach of Currie et al. (2006), which is expected to speed up inference for large scale problems and help to generalize the proposed methods for multidimensional functional responses and image regression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Danny Reich and Peter Calabresi for supplying and explaining the DTI tractography data. Richard Herrick and Jeffrey S. Morris provided a Linux-version of WFMM. Sonja Greven and Fabian Scheipl were funded by Emmy Noether grant GR 3793/1-1 from the German Research Foundation. Ana-Maria Staicu's research was supported by U.S. National Science Foundation grant number DMS 1007466 and by US National Institute of Health grant number 1R01 NS085211-01.

A Identifiability

A.1 Imposing suitable identifiability constraints

Additive models for scalar responses ensure identifiability by imposing suitable constraints on the functions that make up the additive predictor $\beta_0+\sum_s f_s(x_s)$, such as a sum-to-zero constraint $\sum_{i=1}^n f_s(x_{si})=0$ for each function $f_s(x_s)$ (Wood, 2006). Otherwise, any constant could be added to one function and subtracted from the others without changing the fit criterion.

A similar issue arises in the context of our proposed model. For arbitrary functions $\bar{\gamma}_z(t), \bar{b}_g(t)$

$$\mathrm{E}(y_i(t)) = \alpha(t) + \gamma(z_i,t) + b_{g(i)}(t) = \left(\alpha(t) + \bar{\gamma}_z(t) + \bar{b}_g(t)\right) + \left(\gamma(z_i,t) - \bar{\gamma}_z(t)\right) + \left(b_{g(i)}(t) - \bar{b}_g(t)\right)$$

obtains the same fit with two different parameterizations. To avoid this, we impose sum-to-zero constraints for each t so that $n^{-1}\sum_{i=1}^n b_{g(i)}(t) = n^{-1}\sum_{i=1}^n \gamma(z_i,t) = 0 \quad \forall t$.

We also center covariate trajectories $x_i(s)$ by subtracting the mean function $\bar{x}(s) = n^{-1}\sum_i x_i(s)$. If both the sum-to-zero constraints for each t are imposed and functional covariates are centered, all effects that vary over the index of the response are directly interpretable as deviations from the overall mean trajectory $\alpha(t)$. Standard sum-to-zero constraints implemented in mgcv, which would correspond to $\Sigma_{i,t}\ \gamma(z_i,t) = 0$, yield neither identifiable models nor effects that are interpretable like this. Implementationwise, we use the method described in Wood (2006, ch. 1.8.1) to absorb the sum-to-zero-for-each-t constraints into the design matrices of all effects varying over t, see section A in the online supplement for details and examples.

A.2 Limits on the identifiability of complex regression surfaces for low-rank functional covariates

For function-on-function-regression terms $\int_{\mathcal{S}} x(s)\beta(s,t)\,ds$, identifiability of $\beta(s,t)$ is guaranteed under conditions derived in He et al. (2003), Chiou et al. (2004) and Prchal and

Sarda (2007), which are hard to verify empirically. In practice, an important quantity in this regard for the stability of spline-based estimates is the effective rank of the covariance operator of $x(s)$, which can be defined as the number of eigenvalues that together account for at least 99.5% of the covariate's variability. If this effective rank is low, the kernel of the functional covariate's covariance operator is large. Scheipl and Greven (2012) have shown that spline-based regression surface estimates can be unstable if the kernel of the functional covariate's covariance operator overlaps the function space spanned by parameter vectors in the nullspace of the tensor product spline's roughness penalty. Based on theoretical considerations and simulation results (c.f. Scheipl and Greven, 2012), we recommend that practitioners check the effective rank of the observed covariance matrix of functional covariates and the amount of overlap between the kernel of the functional covariate's covariance operator and the nullspace of the associated roughness penalty. Utility functions to perform these checks and constructors for modified roughness penalties without nullspaces are included in refund.

B Supplementary Details for the DTI Data Analysis

C Supplementary Simulation Study Results

Comparison with FPC-based approaches—We fit models with an FPC-based function-on-function term (c.f. page 8) and models with FPC-based functional random intercepts (c.f. page 10) to each dataset generated for the second scenario. rIMSEs for the FPC-based function-on-function term were larger than those of the spline-based estimates by a mean factor of $1.5_{(1.2–2.1)}$, while computation time was about the same for $\boldsymbol{M} = 10$ ($1_{(0.8–1.4)}$) and somewhat longer for $\boldsymbol{M} = 100$ ($1.3_{(1.1–1.6)}$). Results for the FPC-based functional random intercept were more different from the spline-based option. Specifically, the FPC-based functional random intercept showed fairly little improvement for $\mathbf{SNR}_{\varepsilon} = 5$ compared to $\mathbf{SNR}_{\varepsilon} = 1$. For the latter, the FPC performance was fairly similar ($\boldsymbol{M} = 10$: factor of $2.4_{(1.3–9.4)}$, $\boldsymbol{M} = 100$: factor of $1.1_{(1.0–1.5)}$), while it was much less precise for the former: $3.4_{(1.1–17)}$ for $\boldsymbol{M} = 100$ and $17_{(2.5–106)}$ for $\boldsymbol{M} = 10$. As expected, however, FPC-based functional random intercepts scaled much better than spline-based ones for larger datasets in terms of computation time due to their more compact optimal basis representation – for $\boldsymbol{M} = 100$, the iterative FPC procedure was faster than spline-based random effect models by a factor of $0.3_{(0.2–0.5)}$. Also, our spline-based data generating process corresponding to five non-zero FPCs (c.f. Appendix B of the supplement) may be more difficult for FPC based approaches: previous simulation studies of FPC-based functional regression have typically used data generating processes with lower effective rank (e.g. Chen and Müller, 2012; Müller and Yao, 2008; Wu et al., 2010, with 2, 3, and 4 eigenfunctions, respectively) and simpler coefficient shapes.

Comparison with WFMM—We compare our approach to the available implementation of the wavelet-based functional linear mixed models of Morris and Carroll (2006) in WFMM (Herrick, 2013). We can only provide this comparison for scenario 1 as the other scenarios feature terms that are not available in WFMM, which can only fit random effect curves and functional linear effects $z_{ij}\beta(t)$ of scalar covariates z. Note that, differing from the results for pffr in the remainder of the article, these results are for balanced data, as the WFMM

algorithm seems to fail whenever there are any subjects with 1 or 2 observations only, and 10 replicates per setting. We used the default hyper- and tuning parameters for WFMM, with 2000 iterations of burn-in followed by 10000 iterations of sampling. In general, the IMSEs for WFMM are about double to three times those of pffr. Specifically, the IMSE for $y(t)$ is increased by a median factor of $2.5_{(2.1–3.1)}$ (IMSE($b_0(t)$: $2.3_{(1.6–3.8)}$, IMSE($b_1(t)$: $2.4_{(1.5–4)}$). Note, however, that this comparison is not entirely fair to WFMM, as it is designed for spiky data e.g. from spectrometry (i.e., it assumes sparsity in a suitable wavelet domain), not the smooth functional data that we assume and correspondingly simulated here. Although WFMM is much slower (4- to 16-fold) than pffr for small and intermediate data sizes, its computation time increases much slower than pffr for larger data sets due to its efficient data representation in the wavelet domain and its very fast C++ implementation. Detailed results for this comparison are provided in Section B.6 of the online appendix.

References

Abramovich F, Angelini C. Testing in mixed-effects FANOVA models. Journal of Statistical Planning and Inference. 2006; 136(12):4326–4348.

Antoniadis A, Sapatinas T. Estimation and inference in functional mixed-effects models. Computational Statistics & Data Analysis. 2007; 51(10):4793–4813.

Aston JAD, Chiou JM, Evans JP. Linguistic pitch analysis using functional principal component mixed effect models. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2010; 59(2): 297–317.

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

Basser P, Pajevic S, Pierpaoli C, Duda J. In vivo fiber tractography using DT-MRI data. Magnetic Resonance in Medicine. 2000; 44:625–632. [PubMed: 11025519]

Bigelow JL, Dunson DB. Bayesian adaptive regression splines for hierarchical data. Biometrics. 2007; 63(3):724–732. [PubMed: 17403106]

Brumback B, Rice JA. Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). Journal of the American Statistical Association. 1998; 93:961–994.

Brumback B, Ruppert D, Wand MP. Comment. Journal of the American Statistical Association. 1999; 94(447):794–797.

Chen K, Müller H-G. Modeling repeated functional observations. Journal of the American Statistical Association. 2012; 107:1599–1609.

Chiou J, Müller H, Wang J. Functional response models. Statistica Sinica. 2004; 14(3):675–694.

Crainiceanu CM, Reiss PT, Goldsmith J, Greven S, Huang L, Scheipl F. refund: Regression with Functional Data. R package version 0.1-5. 2011

Crainiceanu CM, Ruppert D. Likelihood ratio tests in linear mixed models with one variance component. Journal of the Royal Statistical Society. Series B. 2004; 66(1):165–185.

Crainiceanu CM, Ruppert D, Claeskens G, Wand MP. Exact likelihood ratio tests for penalised splines. Biometrika. 2005; 92(1):91–103.

Currie ID, Durban M, Eilers PHC. Generalized linear array models with applications to multidimensional smoothing. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2006; 68(2):259–280.

De Boor, C. A practical guide to splines. Springer; New York: 1978.

Delicado P, Giraldo R, Comas C, Mateu J. Statistics for spatial functional data: some recent contributions. Environmetrics. 2010; 21(3-4):224–239.

Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel functional principal component analysis. Annals of Applied Statistics. 2009; 3(1):458–488. [PubMed: 20221415]

Fang Y, Liu B, Müller H-G, Wang J-L. PACE: Principal Analysis by Conditional Expectation. Version 2.16. 2013

Faraway JJ. Regression analysis for a functional response. Technometrics. 1997; 39(3):254–261.

Febrero-Bande M, Oviedo de la Fuente M. Statistical computing in functional data analysis: The R package fda.usc. Journal of Statistical Software. 2012; 51(4):1–28. [PubMed: 23504300]

Ferraty, F.; Goia, A.; Salinelli, E.; Vieu, P. Recent advances on functional additive regression.. In: Ferraty, F., editor. Recent Advances in Functional Data Analysis and Related Topics. Springer; 2011. p. 97-102.

Ferraty, F.; Vieu, P. Nonparametric Functional Data Analysis: Theory and Practice. Springer; 2006.

Giraldo R, Delicado P, Mateu J. Continuous time-varying kriging for spatial prediction of functional data: An environmental application. Journal of Agricultural, Biological, and Environmental Statistics. 2010; 15(1):66–82.

Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D. Penalized functional regression. Journal of Computational and Graphical Statistics. 2011; 20:830–851. [PubMed: 22368438]

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

Goldsmith J, Greven S, Crainiceanu CM. Corrected confidence bands for functional data using principal components. Biometrics. 2013; 69(1):41–51. [PubMed: 23003003]

Greven S, Crainiceanu CM, Caffo BS, Reich D. Longitudinal functional principal component analysis. Electronic Journal of Statistics. 2010; 4:1022–1054. [PubMed: 21743825]

Greven S, Crainiceanu CM, Küchenhoff H, Peters A. Restricted likelihood ratio testing for zero variance components in linear mixed models. Journal of Computational and Graphical Statistics. 2008; 17(4):870–891.

Greven S, Kneib T. On the behaviour of marginal and conditional aic in linear mixed models. Biometrika. 2010; 97(4):773–789.

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

Guo W. Functional mixed effects models. Biometrics. 2002; 58(1):121–128. [PubMed: 11890306]

He, G.; Müller, H-G.; Wang, JL. Extending correlation and regression from multivariate to functional data.. In: Puri, M., editor. Asymptotics in statistics and probability. VSP International Science Publishers; 2003. p. 301-315.

Herrick, R. WFMM (Version 3.0 ed.). The University of Texas M.D. Anderson Cancer Center; 2013.

Hörmann S, Kokoszka P. Weakly dependent functional data. The Annals of Statistics. 2010; 38(3): 1845–1884.

Hörmann, S.; Kokoszka, P. Consistency of the mean and the principal components of spatially distributed functional data.. In: Ferraty, F., editor. Recent Advances in Functional Data Analysis and Related Topics. Springer; 2011. p. 169-175.

Ivanescu, AE.; Staicu, A-M.; Scheipl, F.; Greven, S. Penalized function-on-function regression. Technical Report 254. Johns Hopkins University, Dept. of Biostatistics Working Papers; 2012.

James GM. Generalized linear models with functional predictors. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002; 64(3):411–432.

Krafty RT, Hall M, Guo W. Functional mixed effects spectral analysis. Biometrika. 2011; 98(3):583–598.

Krivobokova T, Kauermann G. A note on penalized spline smoothing with correlated errors. Journal of the American Statistical Association. 2007; 102(480):1328–1337.

Li Y, Wang N, Hong M, Turner ND, Lupton JR, Carroll RJ. Nonparametric estimation of correlation functions in longitudinal and spatial data, with application to colon carcinogenesis experiments. The Annals of Statistics. 2007; 35(4):1608–1643.

Malfait N, Ramsay J. The historical functional linear model. Canadian Journal of Statistics. 2003; 31(2):115–128.

Marra G, Wood SN. Practical variable selection for generalized additive models. Computational Statistics & Data Analysis. 2011; 55(7):2372–2387.

Marra G, Wood SN. Coverage properties of confidence intervals for generalized additive model components. Scandinavian Journal of Statistics. 2012; 39(1):53–74.

McLean MW, Hooker G, Staicu A-M, Scheipl F, Ruppert D. Functional generalized additive models. Journal of Computational and Graphical Statistics. 2012 (to appear).

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

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

Müller H-G, Yao F. Functional additive models. Journal of the American Statistical Association. 2008; 103(484):1534–1544.

Nerini D, Monestiez P, Manté C. Cokriging for spatial functional data. Journal of Multivariate Analysis. 2010; 101(2):409–418.

Nychka D. Confidence intervals for smoothing splines. Journal of the American Statistical Association. 1988; 83:1134–1143.

Panaretos VM, Tavakoli S. Cramér-Karhunen-Loève representation and harmonic principal component analysis of functional time series. Stochastic Processes and their Applications. 2013a; 123(7):2279–2807.

Panaretos VM, Tavakoli S. Fourier analysis of stationary time series in function space. The Annals of Statistics. 2013b; 41(2):568–603.

Prchal L, Sarda P. Spline estimator for functional linear regression with functional response. 2007 unpublished.

Qin L, Guo W. Functional mixed-effects model for periodic data. Biostatistics. 2006; 7(2):225–234. [PubMed: 16207823]

Ramsay JO, Wickham H, Graves S, Hooker G. fda: Functional Data Analysis. R package version 2.2.7. 2011

R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2011.

Reiss PT, Huang L, Mennes M. Fast function-on-scalar regression with penalized basis expansions. The International Journal of Biostatistics. 2010; 6(1):28.

Reiss PT, Ogden T. Functional principal component regression and functional partial least squares. Journal of the American Statistical Association. 2007; 102:984–996.

Reiss PT, Ogden T. Smoothing parameter selection for a class of semiparametric linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2009; 71(2):505–523.

Ruppert, D.; Carroll, RJ.; Wand, MP. Semiparametric Regression. Cambridge University Press; Cambridge, UK: 2003.

Scarpa B, Dunson DB. Bayesian hierarchical functional data analysis via contaminated informative priors. Biometrics. 2009; 65(3):772–780. [PubMed: 19173703]

Scheipl, F.; Greven, S. Identifiability in penalized function-on-function regression models. Technical Report 125. LMU München; 2012.

Scheipl F, Greven S, Küchenhoff H. Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. Computational Statistics & Data Analysis. 2008; 52(7):3283–3299.

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

Staicu A-M, Crainiceanu CM, Ruppert D, Reich D. Modeling functional data with spatially heterogeneous shape characteristics. Biometrics. 2011; 68(2):331–343. [PubMed: 22050118]

Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association. 2004; 99(467):673–686.

Wood, SN. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC; 2006.

Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B. 2011; 73(1):3–36.

Wood SN. On p-values for smooth components of an extended generalized additive model. Biometrika. 2013; 100(1):221–228.

Wood SN, Scheipl F, Faraway JJ. Straightforward intermediate rank tensor product smoothing in mixed models. Statistics and Computing. 2013; 23:341–360.

Wu Y, Fan J, Müller H-G. Varying-coefficient functional linear regression. Bernoulli. 2010; 16(3): 730–758.

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

Zhou L, Huang JZ, Martinez JG, Maity A, Baladandayuthapani V, Carroll RJ. Reduced rank mixed effects models for spatially correlated hierarchical functional data. Journal of the American Statistical Association. 2010; 105(489):390–400. [PubMed: 20396628]

Zhu H, Brown PJ, Morris JS. Robust, adaptive functional regression in functional mixed model framework. Journal of the American Statistical Association. 2011; 106(495):1167–1179. [PubMed: 22308015]

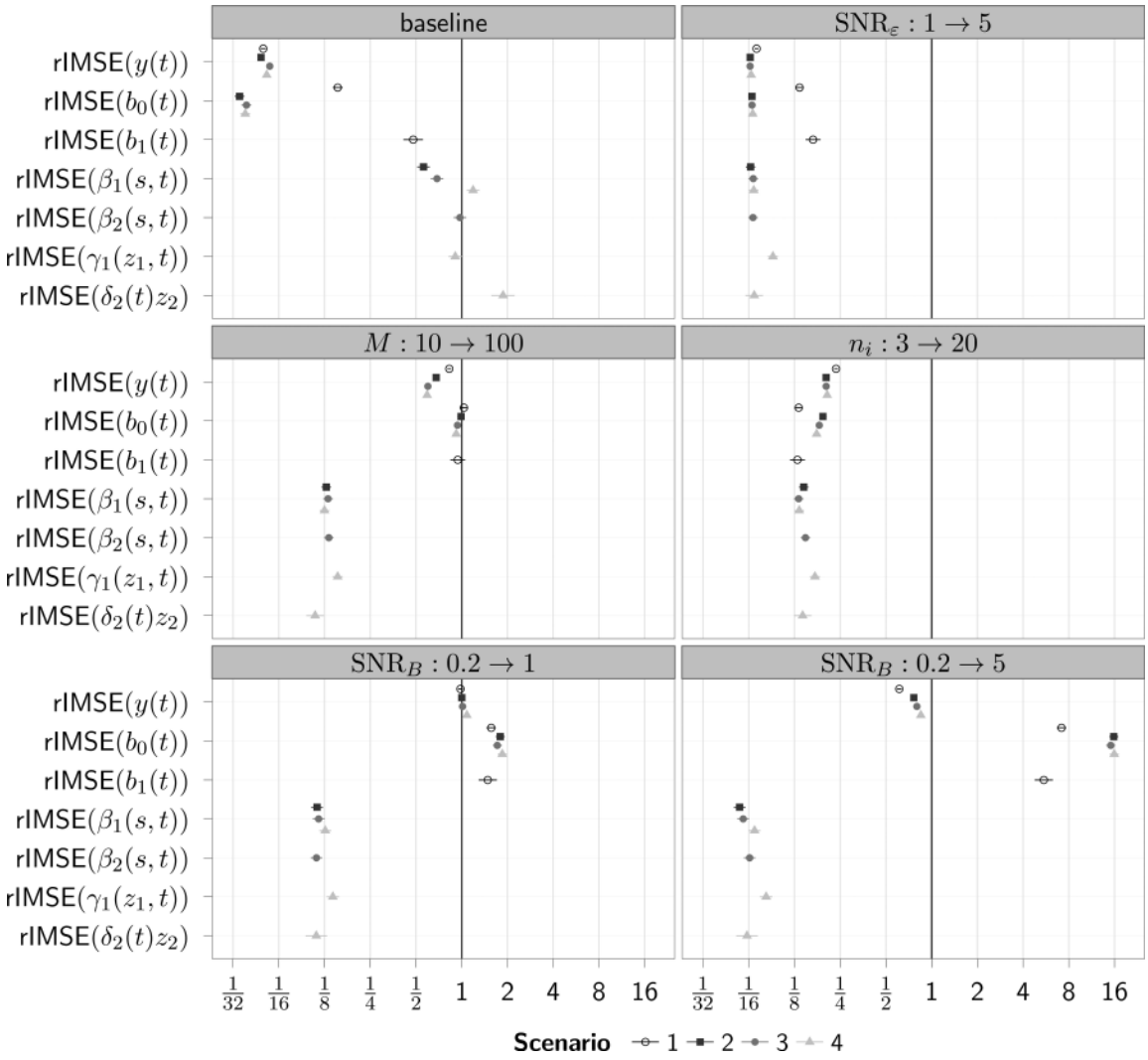

Figure 1.

Baseline levels and estimated multiplicative change in rIMSE for the 4 scenarios. The scenarios are depicted with different symbols, and the segments accompanying the symbols correspond to the estimated effect ± 2 standard errors. Effects other than $b_0(t)$ only occur in a subset of scenarios. Horizontal axis on $\log_2$.

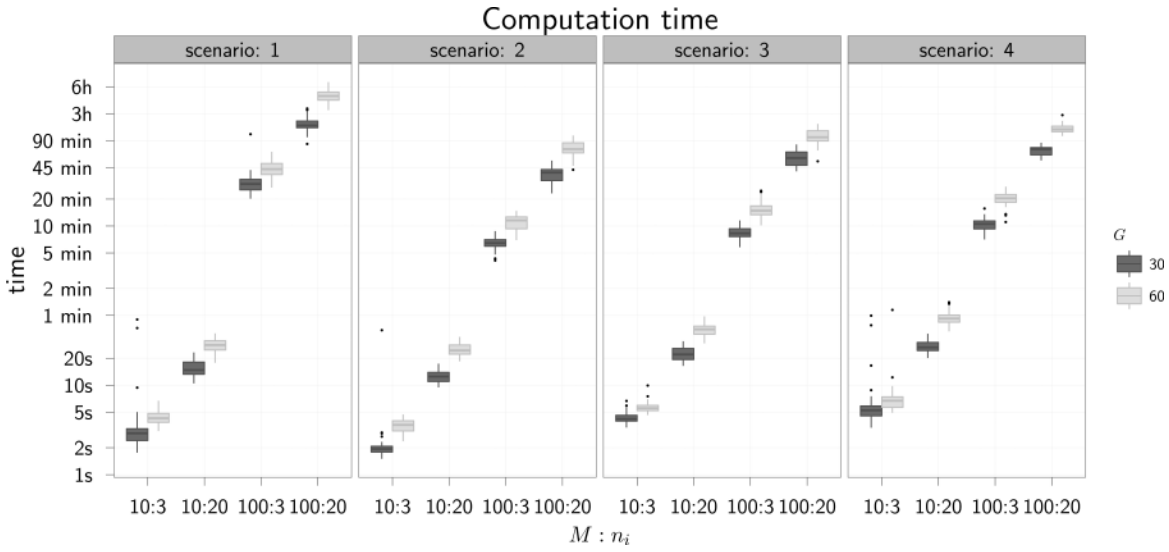

Figure 2.

Computation times for scenarios 1 to 4 (from left to right). Vertical axis on $\log_{10}$-scale. Horizontal axis for the various combinations of numbers of subjects $\boldsymbol{M}$ and average number of replications per subject n_i. Results for $T = 30$ in dark grey and in light grey for $T = 60$. Timings are wall-clock time taken on an 2.2 GHz AMD Opteron 6174.

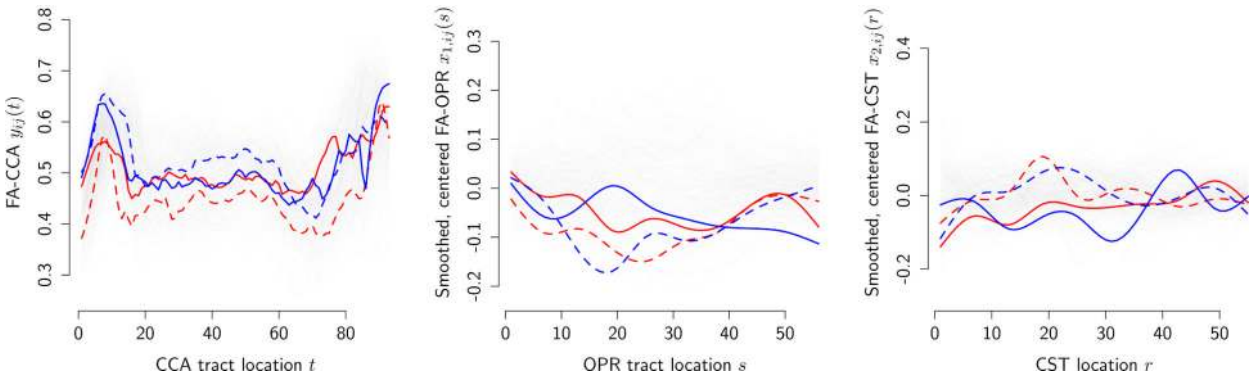

Figure 3.

From left to right: FA profiles along CCA, OPR and CST for MS patients (red) and controls (blue). Solid line: females; dashed: males. FA-OPR and FA-CST are de-trended and smoothed.

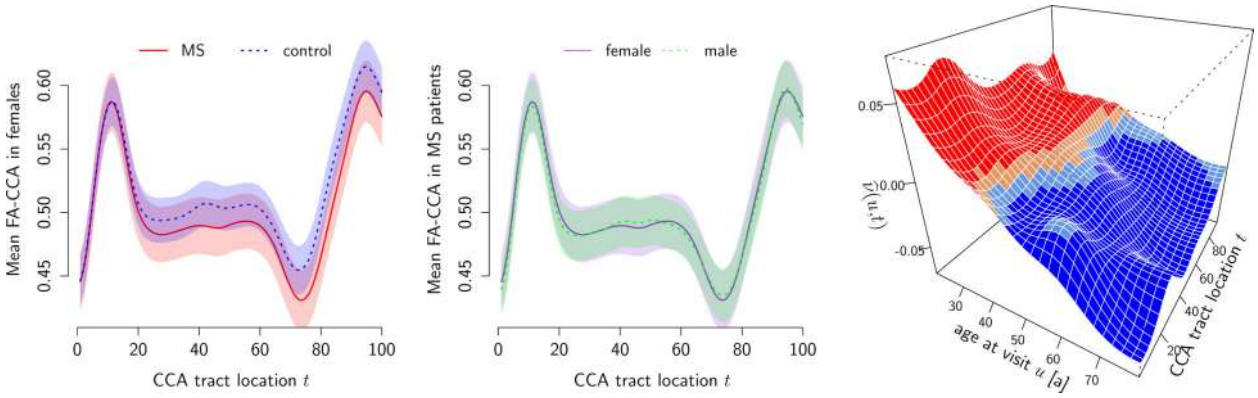

Figure 4.

Estimated components of model (9) with ±2 pointwise standard errors. Coefficient surfaces are color-coded for sign and approximate pointwise significance (95%): blue if sig. < 0, light blue if < 0, light red if > 0, red if sig. > 0. Left to right: mean FA-CCA for healthy (blue, dotted) versus MS (red, solid) females; mean FA-CCA for female (purple, solid) and male (green, dotted) MS patients; estimated smooth index-varying age effect $\hat{\nu}(u,t)$.

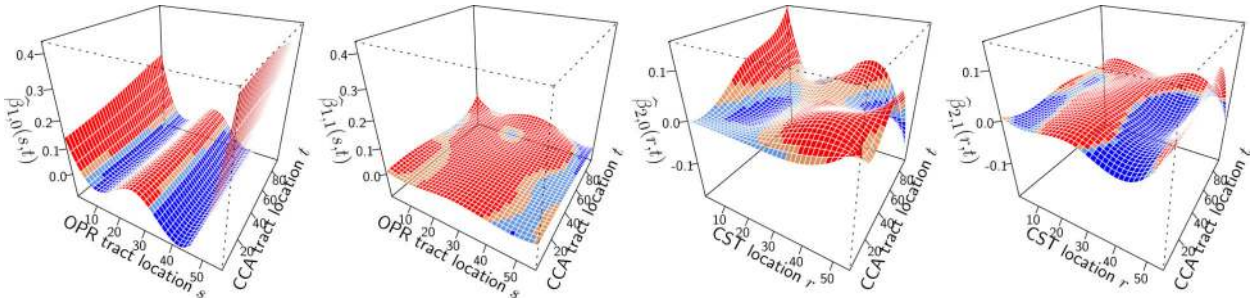

Figure 5.

Left to right: Estimated coefficient surfaces $\hat{\beta}_{1,0}(s,t)$, $\hat{\beta}_{1,1}(s,t)$, $\hat{\beta}_{2,0}(r,t)$, $\hat{\beta}_{2,1}(r,t)$.

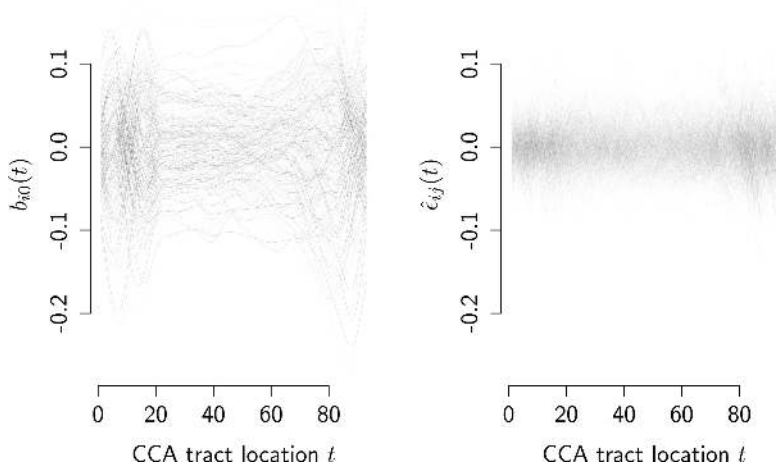

Figure 6.
Predicted functional intercepts $\hat{b_{i0}}(t)$ and observed residuals $\hat{e}_{ij}(t)$ for model (9).

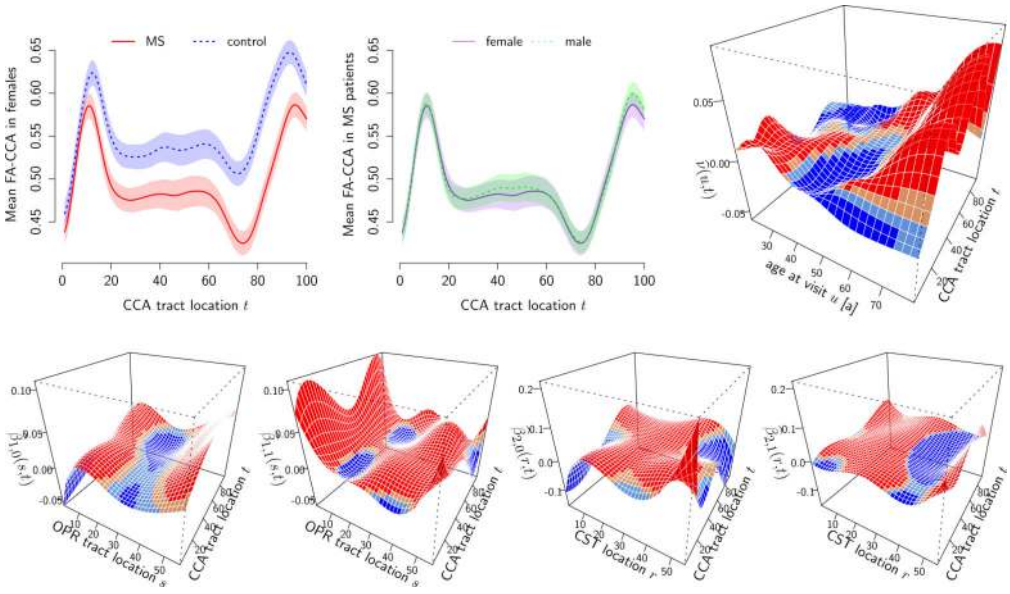

Figure 7.

Estimated components of model (8) with ±2 pointwise standard errors, using Ivanescu et al. (2012). Coefficient surfaces are color-coded for sign and pointwise significance (95%): blue if sig. < 0, light blue if < 0, light red if > 0, red if sig. > 0. Top row, left to right: mean FA-CCA for healthy (blue, dotted) versus MS (red, solid) females; mean FA-CCA for female (purple, solid) and male (green, dotted) MS patients; estimated effect of age-at-visit $\hat{\nu}(u,t)$. Bottom row, left to right: Estimated coefficient surfaces $\hat{\beta}_{1,0}(s,t)$, $\hat{\beta}_{1,1}(s,t)$, $\hat{\beta}_{2,0}(r,t)$, $\hat{\beta}_{2,1}(r,t)$.

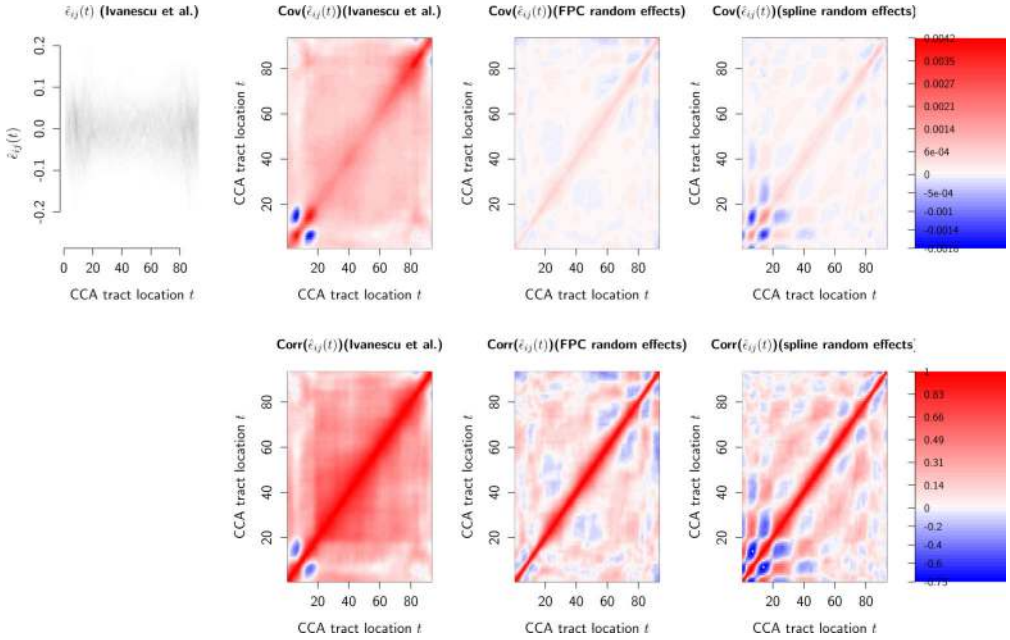

Figure 8.

Top row, left to right: Observed residuals $\hat{\varepsilon}_{ij}(t)$ for model (8); empirical covariance for $\hat{\varepsilon}_{ij}(t)$ for model (8); empirical covariance for $\hat{\varepsilon}_{ij}(t)$ for model (9) with FPC-based random intercepts; empirical covariance for $\hat{\varepsilon}_{ij}(t)$ for model (9) with spline-based random intercepts; legend for covariance values. Bottom row: Empirical correlations.

Table 1

Forms of $f_r(\mathcal{X}_r, t)$ depending on the covariates in $\mathcal{X}_r$ and linearity or smoothness in these covariates (rows), and on whether the effect is constant or varying over t (columns). For scalar categorical covariates, synthetic scalar covariates in effect or reference category coding are created. Note that effects can become interaction effects if $\mathcal{X}_r$ additionally contains such scalar categorical covariates. For example, we estimate group-specific effects of the functional covariates for MS patients and healthy controls in our DTI application.

X_r	$f_r(X_r, t)$ **constant over** t	$f_r(X_r, t)$ **varying over** t
$\emptyset$ (no covariates)	scalar intercept a	functional intercept $a(t)$
functional covariate $x(s)$	linear functional effect $\int_T x(s)\beta(s)ds$	linear functional effect $\int_T x(s)\beta(s,t)ds$
	smooth functional effect $\int_T F(x(s), s)ds$	smooth functional effect $\int_T F(x(s), s, t)ds$
scalar covariate z	linear effect $z\delta$	functional linear effect $z\delta(t)$
	smooth effect $\gamma(z)$	smooth effect $\gamma(z, t)$
vector of scalar covariates $\mathbf{z}$	interaction effect $z_1 z_2 \delta$	functional interaction effect $z_1 z_2 \delta(t)$
	varying coefficient $z_1\delta(z_2)$	functional varying coefficient $z_1\delta(z_2, t)$
	smooth effect $\gamma(z)$	smooth effect $\gamma(z,t)$
grouping variable g	random intercept b_g	functional random intercept $b_g(t)$
grouping variable g and scalar covariate z	random slope zb_g	functional random slope $zb_g(t)$